

词元之后

AI时代真正稀缺的是什么？

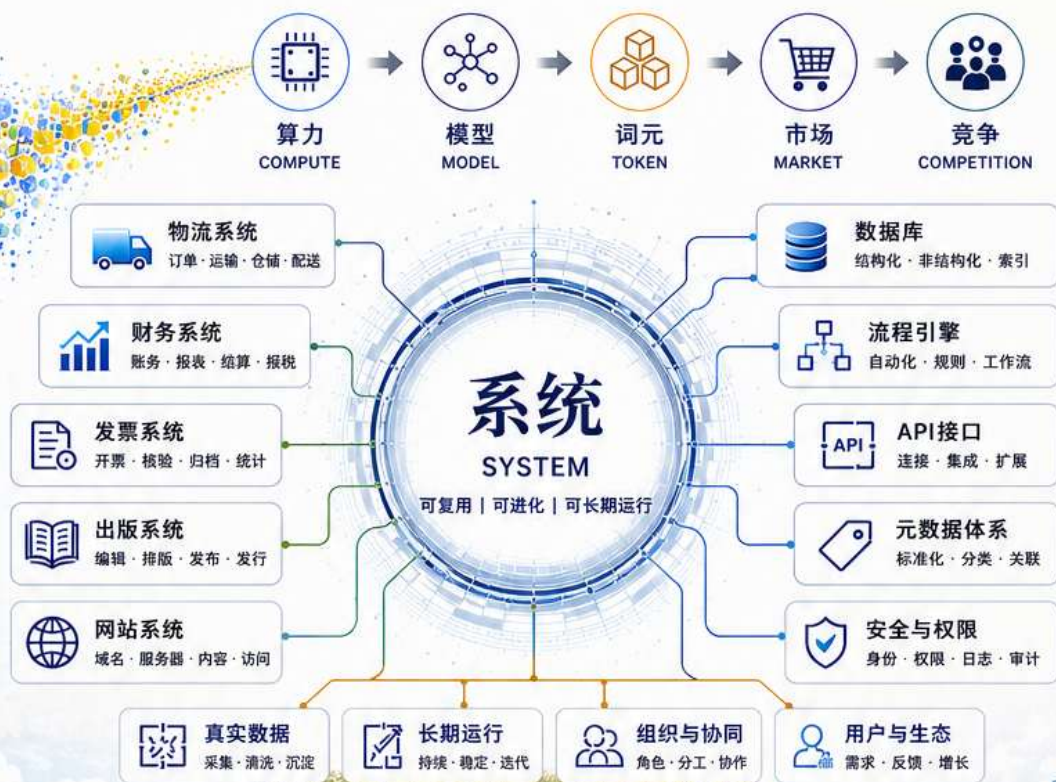
BEYOND TOKENS

What Is Truly Scarce in the Age of AI?

当词元越来越丰富，
什么反而越来越**稀缺**？

从算力到系统
从信息到能力
从短期到长期
从数据到文明记忆

系统，才是未来的根



- ⌚ 时间无法压缩
- 👤 经验无法购买
- 🤝 信任无法速成
- 📖 文明记忆无法替代

这就是真正的壁垒



算力是水

流动·支持·灌溉



词元是养分

丰富·易得·消耗



系统是根

不可替代·长期沉淀·持续生长

真正稀缺的，
不是词元，而是：
**系统能力
时间资产
文明记忆**

—— 未来属于长期主义者



作者 / 结构系统设计者

巫朝晖 JEFFI CHAO HUI WU

国际独立学者

ORCID: <https://orcid.org/0009-0009-5318-3027>

重要 DOI

<https://doi.org/10.5281/zenodo.17646306>

词元之后：AI时代真正稀缺的是什么



出版日期

2026年6月

词元只是入口，系统才是未来的根。

Tokens are the entry, Systems are the root of the future.

 한국어

Outline

成熟结构型物流系统

结构优先的智能物流与财务整合系统

A Mature Logistics System

An Intelligent Logistics and Financial Integration System
Driven by Structure-First Design

20年实战沉淀

- 2003 – 至今
跨代持续运行
- 10,000+ TEU
年处理能力
- 2.5人团队
超越传统团队效率



低人力 · 高效率

Low Manpower
High Efficiency

2.5人团队 / 1万TEU

2.5 People Team / 10,000 TEU



高自动化 · 低错误

High Automation
Low Error Rate

95%自动处理率

95% Automation Rate



合规精准 · 实时税务

Compliance & Accuracy
Real-time Tax Management

GST自动计算与申报

GST Auto Calculation & Reporting



数据主权 · 安全可控

Data Sovereignty
Secure & Controllable

单机运行 · 数据自主可控

Standalone Operation ·
Data Sovereignty



单据自动提取
Auto Document
Capture



数据交叉验证
Cross Validation



异常智能阻断
Exception Blocking



自动生成凭证
Auto Vouchers



财务报表输出
Financial Reports



全流程闭环
Full Process
Closed Loop



业务 · 财务 · 税务一体化
Business · Finance · Tax
Integrated



结构驱动设计
Structure-Driven
Design



持续进化能力
Continuous
Evolution



经时间验证
Proven Over
Two Decades

作者 / 系统设计者：巫朝晖 JEFFI CHAO HUI WU

国际独立学者 | Independent International Scholar

 <https://doi.org/10.5281/zenodo.20201895>

二十年跨代运行 · 结构优先设计

以极小人力，实现最大系统效能

智能驱动业务，结构创造价值

让复杂的系统，回归简单而强大的结构

跨时代实战二十年的 智能税务系统

二十年实战验证 · 系统化 · 自动化 · 合规化 · 持续运行 · 稳定可靠

A CROSS-GENERATION INTELLIGENT TAX SYSTEM
WITH TWENTY YEARS OF PRACTICAL APPLICATION

前 AI 时代的
超前独立系统

2005 年成型并持续运行
二十年真实验证

★★★★★

解决物流财税的复杂性难题

Solving the Complexity of Logistics Taxation

基于真实资金流的结构验证

内容要点 | KEY HIGHLIGHTS



直面物流税务的真实复杂度

深入应对澳洲税务体系的多税种、
多规则、多层次、多变化



自动匹配 · 智能核对 · 精准校验

发票、账务、报关、付款、GST、BAS 等
全流程数据自动匹配与核对



端到端自动化流程

从数据采集、分类、计算、核对到申报，
全流程自动化处理



合规稳定 · 风险可控

实时合规校验，预警风险，确保长期
稳定运行与可追溯



二十年实战验证

持续运行二十年，跨越政策变化，
经受时间与业务的双重检验



系统化架构 · 可扩展可复制

模块化设计，适应业务增长与结构调整，
可扩展、可复制

智能税务系统

发票管理
INVOICE
MANAGEMENT

自动匹配
AUTO MATCHING

核对校验
VERIFICATION

税务申报
TAX REPORTING

合规检查
COMPLIANCE
CHECK

报表分析
REPORT
ANALYSIS

发票

报关

付款

账务



作者 / AUTHOR

巫朝晖 JEFFI CHAO HUI WU

国际独立学者

INTERNATIONAL INDEPENDENT SCHOLAR



出版日期 / PUBLICATION DATE

2026年3月

MARCH 2026



原文 DOI / ORIGINAL DOI <https://doi.org/10.5281/zenodo.19182200>

不是 AI 时代的产品，而是前 AI 时代的远见与实践

二十年前的独立设计，二十年的持续运行，构建真正属于业务本身的智能税务系统

跨代 20 年 前 AI 时代 智能物流与财务系统

A Cross-Generational Intelligent Logistics and Financial System over 20 Years

— Pre-AI Era Intelligent Logistics and Financial System —

— 一种在前 AI 时代已长期实战运行的结构智能系统 —

A Structure-Based Intelligent System That Has Been Continuously
and Practically Operated in the Pre-AI Era



物流
Logistics



数据
Data



财务
Finance



发票
Invoicing



税务
Taxation



闭环
Closed Loop



数据主权
Data Sovereignty

系统要点

KEY POINTS

- 1997 构思 · 2005 实战运行 · 2013 全面整合
从构想到现实，持续进化超过 20 年
- 物流 · 发票 · 财务 · 税务一体化闭环系统
真正的端到端闭环管理
- 95% 自动处理 · 5% 人工介入
高度自动化，效率与准确性并重
- 年处理超一万标箱 (TEU)
真实世界大规模运行数据验证
- 2.5 人团队全链条运行
精干团队，完成端到端全流程管理
- 结构优先 · 异常阻断机制 · 数据主权
以结构确保可靠性，以机制保障正确性
- 单机架构穿越互联网、云计算与 AI 时代
在前 AI 时代已经智能化并持续稳定运行

实战运行数据 OPERATIONAL PERFORMANCE

业务规模
ANNUAL CAPACITY
10,000+
TEU / 年

自动化处理率
AUTOMATION RATE
95%
自动处理
Automated

人工介入率
HUMAN INTERVENTION
5%
人工介入
Human Touch

与既有系统的关系

RELATIONSHIP WITH EXISTING SYSTEMS

《2005 年的
智能发票系统》



重点是
发票自动化原型

《智能税务系统》



重点是
税务与资金流

《成熟结构型
物流系统》



更偏物流
运营结构



1997

构思与原型
Concept & Prototype



2005

系统实战运行
Operational Implementation



2013

全面整合升级
Full Integration



AI 时代

持续进化
Continuous Evolution

作者 / Author

巫朝晖

JEFFI CHAO HUI WU

国际独立学者 / 结构系统设计者
International Independent Scholar /
Structural System Designer

文章 DOI / Article DOI

<https://doi.org/10.5281/zenodo.17646306>

相关 DOI / Related DOIs

<https://doi.org/10.5281/zenodo.20201895>
<https://doi.org/10.5281/zenodo.19182200>

出版日期 / Publication Date



2026年4月
April 2026



Indexed and catalogued
in WorldCat (OCLC)
已被 WorldCat (OCLC)
索引与编目



永久归档于 TROVE
(澳大利亚国家图书馆)
Permanently archived in
TROVE (National Library of Australia)



永久存档记录 (DataCite 概念 DOI)
Permanent archival record
(DataCite Concept DOI)

REDEFINING "COMPUTATION" ITSELF

THE HYPERDIMENSIONAL ALGORITHM

A NEW DEFINITION OF "COMPUTATION"

超维算法

核心要点：

- 1 多维结构**
数据不是单一输入输出，
而是多起点、多结果的结构网络
- 2 非线性因果**
结果可以先于原因，
多路径并行存在
- 3 随机有序**
随机性与结构性同时存在于系统内部
- 4 一数多果**
同一数据可对应多个结果，
不依赖单一计算路径
- 5 入口触发**
通过改变入口而非修改数据，
实现不同结果输出



MULTI-DIMENSIONAL STRUCTURE

Data as nodes.
Not just inputs or outputs,
but multiple origins
and multiple results.



NON-LINEAR CAUSALITY

Results can precede
causes. Multiple
paths. Recursive
and interconnected.



RANDOM YET ORDERED

Randomness and order
coexist as an intrinsic
part of the algorithmic
structure.



ONE DATA, MANY RESULTS

The same metadata
can correspond to
different results
through different
structural entrances.



ENTRANCE-BASED ACTIVATION

Change the entrance,
not the data.
Activate different
states and outcomes.

A FOUNDATIONAL FRAMEWORK
FOR HYPERDIMENSIONAL COMPUTATION
AND A NEW SCIENTIFIC PARADIGM

Contents

词元之后：AI 时代真正稀缺的是什么？	1
一、词元溯源：从哲学概念到经济单位的四重跨越	4
二、技术底座：BPE 算法与隐性的“语言税”	5
三、用途与产出：从对话到智能体的指数级跃迁	6
四、机构与市场：三足鼎立的产业图谱	6
五、竞争与门槛：价格战内卷与价值分层	8
六、当下困局与未来出路	10
中文词元与英文词元：统一概念下的三重壁垒	10
七、为什么系统构建能力无法通过大量词元直接获得	13
八、我的资产：真实历史数据、跨领域系统、构建能力、预见视觉与自证能力	15
九、词元之后：真正稀缺的五种能力	17
十、词元流向何处：从短暂对话到文明记忆	23
十一、时间的不可压缩性：四种无法购买的资产	23
十二、结论：从“词元囤积”到“系统构建”	25
Beyond Tokens: What Is Truly Scarce in the Age of AI?	30
I. Token Origins: The Quadruple Leap from Philosophical Concept to Economic Unit	37
II. The Technical Base: BPE Algorithm and the Implicit "Language Tax"	38
III. Uses and Output: The Exponential Leap from Conversation to Agents	40
IV. Institutions and Markets: The Tripartite Industrial Landscape	41
V. Competition and Barriers: Price War Involution and Value Stratification	44
VI. Current Difficulties and Future Pathways	47
Chinese Tokens vs. English Tokens: Three Barriers Under a Unified Concept	48
VII. Why System-Building Capability Cannot Be Directly Obtained Through Large Quantities of Tokens	52

VIII. My Assets: Real Historical Data, Cross-Domain Systems, Building Capability, Foresight Vision, and Self-Verification Capability	55
IX. Beyond Tokens: Five Truly Scarce Capabilities.....	58
X. Where Do Tokens Flow: From Transient Conversations to Civilization's Memory	68
XI. The Incompressibility of Time: Four Unpurchasable Assets	69
XII. Conclusion: From "Token Hoarding" to "System Building"	71
Après les Tokens : Ce qui est véritablement rare à l'ère de l'IA	77
I. Origines des tokens : le quadruple saut du concept philosophique à l'unité économique	83
II. La base technique : l'algorithme BPE et la « taxe linguistique » implicite	84
III. Utilisations et production : le saut exponentiel de la conversation aux agents.....	85
IV. Institutions et marchés : le paysage industriel tripartite	86
V. Concurrence et barrières : involution de la guerre des prix et stratification de la valeur90	
VI. Difficultés actuelles et voies d'avenir.....	92
Tokens chinois vs. tokens anglais : trois barrières sous un concept unifié.....	93
VII. Pourquoi la capacité à construire des systèmes ne peut être obtenue directement par de grandes quantités de tokens.....	97
VIII. Mes actifs : données historiques réelles, systèmes transversaux, capacité de construction, vision prospective et capacité d'auto-vérification	101
IX. Après les tokens : cinq capacités véritablement rares	105
X. Où vont les tokens : des conversations transitoires à la mémoire de la civilisation	116
XI. L'incompressibilité du temps : quatre actifs inachetables	117
XII. Conclusion : De « l'accumulation de tokens » à « la construction de systèmes ».....	119
Después de los Tokens: Lo que es verdaderamente escaso en la Era de la IA	127
I. Orígenes de los tokens: el salto cuádruple del concepto filosófico a la unidad económica	133
II. La base técnica: el algoritmo BPE y el impuesto lingüístico implícito	134
III. Usos y producción: el salto exponencial de la conversación a los agentes.....	135
IV. Instituciones y mercados: el panorama industrial tripartito.....	136

V. Competencia y barreras: involución de la guerra de precios y estratificación del valor	140
VI. Dificultades actuales y vías futuras	142
Tokens chinos vs. tokens ingleses: tres barreras bajo un concepto unificado.....	143
VII. Por qué la capacidad de construir sistemas no puede obtenerse directamente mediante grandes cantidades de tokens.....	147
VIII. Mis activos: datos históricos reales, sistemas transversales, capacidad de construcción, visión prospectiva y capacidad de autoverificación	151
IX. Después de los tokens: cinco capacidades verdaderamente escasas	155
X. Hacia dónde fluyen los tokens: de las conversaciones transitorias a la memoria de la civilización	165
XI. La incompresibilidad del tiempo: cuatro activos incomprables	166
XII. Conclusión: De la «acumulación de tokens» a la «construcción de sistemas»	168
トークン以後：AI 時代に真に希少なものは	176
一、トークンの淵源：哲学的概念から経済単位への四重の飛躍.....	180
二、技術的基盤：BPE アルゴリズムと暗黙の「言語税」	181
三、用途と産出：対話からエージェントへの指数関数的飛躍.....	182
四、機関と市場：三極構造の産業図譜.....	183
五、競争と参入障壁：価格戦の内巻きと価値の階層化.....	186
六、現在の難局と将来の出路.....	188
中国語トークンと英語トークン：統一概念の下三重の壁.....	189
七、なぜシステム構築能力は大量のトークンによって直接獲得できないのか	192
八、私の資産：現実の履歴データ、分野横断的システム、構築能力、予見ビジョン、自己実証能力	195
九、トークン以後：真に希少な五つの能力.....	198
十、トークンはどこへ流れるのか：一過性の対話から文明の記憶へ.....	207
十一、時間の非圧縮性：購入不可能な四つの資産.....	207
十二、結論：「トークン貯め」から「システム構築」へ.....	209

..... الاصطناعي؟ الذكاء عصر في حقًا شحيح هو ما : (التوكنز) الرموز بعد	216
..... الاقتصادية الوحدة إلى الفلسفي المفهوم من الرباعية القفزة: الرموز نشأة - أولاً	220
..... الضمنية "اللغوية الضمنية" و (BPE) للبايتات الزوجي الترميز خوارزمية: التقنية القاعدة - ثانياً	221
..... (الأجنات) الوكلاء إلى المحادثة من الأسية القفزة: والإنتاج الاستخدامات - ثالثاً	222
..... الأركان الثلاثي الصناعي المشهد: والأسواق المؤسسات - رابعاً	223
..... القيمة وتقسيم (الإنفولوجن) الاستنزافية الأسعار حرب: والحواجز المنافسة - خامساً	225
..... المستقبل ومسارات الحالية الصعوبات - سادساً	227
..... موحد مفهوم تحت حواجز ثلاثة: الإنجليزية الرموز مقابل الصينية الرموز	227
..... الرموز من كبيرة كميات خلال من مباشرة الأنظمة بناء قدرة اكتساب يمكن لا لماذا - سابعاً	230
الرؤية، البناء على والقدرة، التخصصات متعددة والأنظمة، الحقيقية التاريخية البيانات: أصولي - ثامناً	
..... الذاتي التحقق وقدرة، المستقبلية	232
..... حقًا نادرة قدرات خمس: الرموز بعد - تاسعاً	235
..... الحضارة ذاكرة إلى العابرة المحادثات من: الرموز تتدفق أين - عاشراً	242
..... شراؤها يمكن لا أصول أربعة: للضغط الوقت قابلية عدم - عشر حادي	243
..... "الأنظمة بناء" إلى "الرموز تكديس" من: الخاتمة - عشر ثاني	244
Nach den Tokens: Was im Zeitalter der KI wirklich knapp ist.....	250
I. Ursprünge der Tokens: Der vierfache Sprung vom philosophischen Konzept zur wirtschaftlichen Einheit.....	256
II. Die technische Basis: BPE-Algorithmus und die implizite "Sprachsteuer"	257
III. Verwendungen und Output: Der exponentielle Sprung von der Konversation zum Agenten.....	258
IV. Institutionen und Märkte: Das tripartite Industrielandschaft.....	259
V. Wettbewerb und Eintrittsbarrieren: Preiskriegsinvolution und Wertschichtung	263
VI. Aktuelle Schwierigkeiten und zukünftige Wege.....	265
Chinesische Tokens vs. englische Tokens: Drei Barrieren unter einem einheitlichen Konzept.....	266
VII. Warum die Fähigkeit zum Systemaufbau nicht direkt durch große Mengen von Tokens erworben werden kann	270

VIII. Meine Vermögenswerte: Reale historische Daten, domänenübergreifende Systeme, Aufbaufähigkeit, Zukunftsbild und Selbstverifizierungsfähigkeit	273
IX. Nach den Tokens: Fünf wirklich knappe Fähigkeiten	278
X. Wohin fließen Tokens: Von flüchtigen Gesprächen zum Gedächtnis der Zivilisation ..	288
XI. Die Inkompressibilität der Zeit: Vier nicht kaufbare Vermögenswerte	289
XII. Fazit: Von der "Token-Hortung" zum "Systemaufbau"	292
Depois dos Tokens: O que é verdadeiramente escasso na Era da IA	299
I. Origens dos tokens: O salto quádruplo do conceito filosófico para a unidade económica	304
II. A base técnica: Algoritmo BPE e o imposto linguístico implícito	305
III. Usos e produção: O salto exponencial da conversa para os agentes	307
IV. Instituições e mercados: A paisagem industrial tripartida.....	308
V. Concorrência e barreiras: Involução da guerra de preços e estratificação de valor	312
VI. Dificuldades atuais e caminhos futuros	314
Tokens chineses vs. tokens ingleses: Três barreiras sob um conceito unificado.....	315
VII. Porque não é possível adquirir diretamente a capacidade de construção de sistemas através de grandes quantidades de tokens	319
VIII. Os meus ativos: Dados históricos reais, sistemas interdisciplinares, capacidade de construção, visão prospetiva e capacidade de autoverificação.....	322
IX. Depois dos tokens: Cinco capacidades verdadeiramente escassas	326
X. Para onde fluem os tokens: Das conversas transitórias à memória da civilização	336
XI. A incompressibilidade do tempo: Quatro ativos não compráveis.....	337
XII. Conclusão: Da "acumulação de tokens" à "construção de sistemas"	339
После Токенов: Что действительно является дефицитным в эпоху ИИ	347
I. Происхождение токенов: Четверной скачок от философской концепции к экономической единице.....	353
II. Техническая основа: Алгоритм BPE и неявный «языковой налог».....	354
III. Использование и производство: Экспоненциальный скачок от диалога к агентам	356

IV. Институты и рынки: Трехсторонний индустриальный ландшафт	357
V. Конкуренция и барьеры: Инволюция ценовой войны и стратификация ценности	361
VI. Текущие трудности и будущие пути	363
Китайские токены против английских токенов: Три барьера под единой концепцией	364
VII. Почему способность к построению систем не может быть приобретена непосредственно через большие объемы токенов	368
VIII. Мои активы: Реальные исторические данные, междисциплинарные системы, способность к построению, дальновидное видение и способность к самоопровержению.....	372
IX. После токенов: Пять действительно дефицитных способностей	376
X. Куда текут токены: От мимолетных разговоров к памяти цивилизации	387
XI. Несжимаемость времени: Четыре непокупаемых актива	388
XII. Заключение: От «накопления токенов» к «построению систем»	390
토큰 이후: AI 시대에 진정으로 희소한 것은 무엇인가	398
1. 토큰의 기원: 철학적 개념에서 경제적 단위로의 사중 비약	402
2. 기술적 기반: BPE 알고리즘과 암묵적인 '언어세'	403
3. 용도와 산출: 대화에서 에이전트로의 기하급수적 비약	404
4. 기관과 시장: 삼두 체제의 산업 지형.....	405
5. 경쟁과 진입 장벽: 가격전 인플레이션과 가치 계층화.....	408
6. 현재의 곤경과 미래의 출구	409
중국어 토큰과 영어 토큰: 통일된 개념 아래의 세 가지 장벽.....	410
7. 왜 시스템 구축 능력은 대량의 토큰으로 직접 획득할 수 없는가	413
8. 나의 자산: 실제 역사 데이터, 학제간 시스템, 구축 능력, 예측 비전, 자기 증명 능력.....	416
9. 토큰 이후: 진정으로 희소한 다섯 가지 능력.....	419
10. 토큰은 어디로 흘러가는가: 일시적 대화에서 문명의 기억으로	427
11. 시간의 비압축성: 구매 불가능한 네 가지 자산	427
12. 결론: '토큰 쌓기'에서 '시스템 구축'으로	429



词元之后：AI 时代真正稀缺的是什么？

作者/国际独立学者：巫朝晖 JEFFI CHAO HUI WU

ORCID: <https://orcid.org/0009-0009-5318-3027>

【前置声明】

本文是基于我个人三十余年跨领域实践（从 1989 年赴澳留学至今，涉及远程工作、物流管理、自动化处理、网站建设、文化传播、多语言出版以及跨领域研究）所形成的个人观点。词元是正在快速演变的新事物，本文仅代表截至成文时（2026 年 6 月）的个人阶段性观察与判断，不构成任何形式的投资建议、商业预测或行业标准。

本文不追求普遍适用性，也不声称代表任何行业共识。随着词元市场、AI 技术和全球政策环境的持续变化，本文中的部分判断可能在未来需要修正或更新。读者应结合最新信息独立判断。

我所论述的“五种稀缺能力”及“五项个人资产”，均来自我个人的系统构建实践（公开实证见 <https://times.net.au> 及该网站列出的全部 40+ 部独立创作作品、ISSN 3083-5178、超 60 个 DOI、以及被 WorldCat、澳大利亚国家图书馆、CERN Zenodo 等国际基础设施永久存档的记录）。

为了避免断章取义，任何针对本文的有效反驳或质疑，其前提条件是：完整浏览本文所有链接文章及其中的嵌套链接，并完整阅读我在 <https://times.net.au> 列出的全部作品。脱离这一整体语境的局部引用或简化概括，均无法构成对本文的公正评价。

关于有效反驳的补充说明：

本文为个人观点，不追求普遍适用性，亦不以引用他人文献为前提。任何有效质疑应采用对等比较（APPLE TO APPLE）：以当前事实对当前事实，以个人实践对个人实践，以已完成系统对已完成系统。用“未来 AI 可能做到”否定“当前已完成的实践”，或以团队/机构能力对标个人实践，均不构成对等比较。

本文所述五项资产组合为个人实践记录，未声称“不存在任何对标案例”。若认为存在完全对标的案例，请举出具体公开案例，并说明其在真实历史数据、跨领域系统、系统构建能力、预见未来的视觉、自证能力五个维度上的对应证据。否则，相关质疑仅属于观点表达，而非有效反驳。

摘要

词元 (Token) 正在成为 AI 时代的核心计量单位，但本文将论证：词元本身不是最终价值。随着推理成本持续下降，词元将像今天的网络流量和硬盘存储一样廉价而普遍。当词元不再稀缺时，真正稀缺的是什么？基于作者三十余年跨领域实践（从 1993 年远程工作原型系统、2005 年智能发票系统，到智能物流系统、十语出版系统、DOI 管理、自动交叉索引及跨平台发布），本文提出以下核心判断：第一，系统构建能力——包括结构设计、自动化、跨领域实践、长期运行心智模型和将词元嵌入现实业务流程的能力——才是真正的稀缺资产。第二，系统构建能力无法通过大量词元直接获得，因为它依赖决策、异常处理、权衡、真实世界反馈循环和时间积累的信任。第三，时间不可压缩、跨领域经验不可购买、系统构建能力难以复制、长期运行记录无法速成，这四种资产构成词元时代最深层的稀缺。本文以作者构建的《时代跃迁》学术网站 (<https://times.net.au>)，ISSN 3083-5178，被 WorldCat、CERN Zenodo、澳大利亚国家图书馆 TROVE 等国际基础设施永久存档) 为公开实证，呈现一个人如何用三十年时间构建跨领域、可验证、长期运行的系统体系。

本文认为：词元只是燃料，系统才是发动机；词元之后，真正稀缺的是能够驾驭词元的人，以及他们所建立起来的结构。

“如果没有货物，要钱何用？”

当我回顾词元的发展历史时，越来越觉得它与人类过去经历过的许多技术革命十分相似。每一次新的基础资源出现时，人们总会首先关注资源本身，而忽略资源背后的结构。工业时代如此，互联网时代如此，人工智能时代同样如此。今天许多人讨论的是词元价格、词元成本、词元额度以及词元供给能力，仿佛拥有更多词元就意味着拥有未来。但从我几十年的实践经验来看，资源本身从来不是最终价值。真正决定价值的，始终是人们如何组织和利用这些资源。

我理解“词元”，不能只从一个新名词开始，而必须从它的历史开始。从 1906 年美国哲学家皮尔士在纸上写下 20 个“the”时提出的“类型与实例”之辨，到 2026 年中国国家数据局局长刘烈宏在博鳌论坛上为 Token 定下中文译名“词元”，这个概念走过了从哲学到计算机、从区块链到人工智能的漫长旅程。Token 这个词在西方计算机世界中早已存在，最早在程序语言、编译器和文本处理系统中，token 指的是计算机能够识别和处理的基本单位。后来在网络安全、身份验证、API 调用和区块链语境中，token 又经常被理解为“令牌”，表示访问权限、身份凭证或数字资产。但到了大语言模型时代，Token 不再主要是身份凭证，也不是区块链资产，而是人工智能模型处理语言、代码、图像描述和多模态信息时的基本计算单位。

中国后来把人工智能领域的 Token 命名为“词元”，我认为这是一个相当准确的翻译。它没有沿用“令牌”，因为大模型中的 Token 并不是通行证；也没有简单翻译成“字”或“词”，因为它有时是一个字，有时是一个词，有时是半个英文单词，有时甚至只是一个标点或符号。“词”说明它与语言和信息有关，“元”说明它是基本单位。对大众读者来说，可以简单理解为：人类看见的是一句话，人工智能看见的是一串词元。在这条道路上，词元从一个冷僻的语言学术语，蜕变为 AI 时代的核心计量单位，进而成为连接技术供给与商业需求的结算工具。作为亲身经历过智能物流系统搭建、自动化出版流程设计、跨平台内容分发等长期项目的实践者，我对这一演变有着切身的感受——词元从来不是目的，而是手段；真正决定价值的，从来不是谁拥有更多词元，而是谁能用词元构建出持续运行的系统和不可复制的结构。本文所论述的系统构建能力，其成果已公开呈现于 <https://times.net.au> 一个由独立学者构建、被全球图书馆体系永久存档的跨领域学术出版系统。

回顾自己的经历，我并不是在词元出现之后才开始思考系统，而是在长期实践过程中逐渐形成了对系统的理解。从 1989 年赴澳留学开始，我经历过不同领域的学习、工作与探索。后来逐步涉及远程工作、物流管理、自动化处理、网站建设、文化传播、多语言出版以及跨领域研究。回头再看，这些经历虽然发生在不同年代，却有一个共同特点，那就是始终围绕着如何建立结构、如何减少重复劳动、如何提高整体效率展开。当时并没有词元这个概念，也没有今天的大语言模型，但我已经在不断思考如何让信息自动流动、自动组织和自动处理。

因此，当今天越来越多人把词元视为一种新的生产资料时，我更倾向于把它看成一种新的基础能源。能源固然重要，但能源并不会自动创造价值。电力本身不会自动变成

工厂，网络本身不会自动变成企业，词元本身也不会自动变成知识体系。未来即使每个人都能够轻松获得大量词元，也不意味着每个人都能够创造同样的成果。决定差异的，并不是词元数量，而是词元进入什么样的结构之中。

在对词元进行深入分析之前，我还需要指出另一个重要观察：Token 就类似货币，东西方科技、系统、政策差异导致无法通用。正如美元和人民币不能自由兑换一样，中文词元和英文词元也面临着类似的流通壁垒。这不仅仅是语言差异的问题，更深层的原因在于：东西方在科技层面（模型架构、分词算法、训练数据）、系统层面（API 生态、支付体系、数据标准）、政策层面（数据跨境流动、出口管制、国家安全审查）存在三重差异，导致词元无法像理想中的“全球通用 AI 货币”那样自由流通。这一观察对理解词元市场的未来格局具有重要意义——未来可能出现“中文词元区”“英文词元区”等基于语言和政策的 Token 生态区，跨区调用将面临效率损失和合规成本。

一、词元溯源：从哲学概念到经济单位的四重跨越

词元的身份演变可以清晰地划分为四个阶段。

哲学语言学的原点。 1906 年，皮尔士在研究指号理论时，区分了“类型”和“实例”。前者代表抽象的规则或形式，后者则是该规则在现实中的具体呈现。这一区分后来成为结构主义语言学的重要基础。

计算机科学的编码转向。 1960 年代，程序员写下“int x = 5；”这行代码时，计算机需要将其拆解为“int”、“x”、“=”、“5”等独立单元，每一个这样的单元就是一个 Token。词元由此成为机器“读懂”指令的最小信息单位。

区块链的价值载体实验。 2017 年，随着 ICO 热潮兴起，Token 被赋予了“可流通的数字权益凭证”这层新身份。虽然那一轮狂热逐渐冷却，但“Token 作为价值符号”的认知就此扎根。

AI 时代的经济基础设施。 2026 年 3 月是一个标志性的节点——英伟达 CEO 黄仁勋将数据中心重新定义为“生产 AI 智能 Token 的工厂”，中国国家数据局局长刘烈宏则代表官方给出了“词元”这一中文译名，并称其为“智能时代的价值锚点”。至此，词元完成了从技术概念到经济概念的最后跨越。

二、技术底座：BPE 算法与隐性的“语言税”

词元之所以能够成为 AI 的标准计量单位，依赖于一项被遗忘二十余年的技术——字节对编码。

1994 年，美国程序员菲利普·盖奇在一本 C 语言技术杂志上发表了一篇关于数据压缩算法的文章，介绍了 BPE 的基本原理：反复扫描文本，将最常相邻出现的两个字符“焊接”成新符号，一轮一轮迭代压缩。这篇论文在当时并未引起关注，因为其压缩效率并不突出。

直到 2016 年，爱丁堡大学的研究员里科·森里希在研究机器翻译的分词难题时，偶然检索出这篇旧文，并敏锐地意识到 BPE 恰好是解决分词问题的绝佳方案——无需预先定义词典，完全让数据自己“说话”，高频组合会自然凝结成词元。2019 年，OpenAI 在发布 GPT-2 时正式采用了这一思路，并将分词起点直接设定在“字节”层面，使模型理论上能够处理任何语言文字。

然而，BPE 算法的“频率优先”逻辑，在无意中形成了一种隐性的“语言税”体系。由于英语是互联网语料的绝对主流，模型对英文词元的切分效率远高于其他语言。表达相同的意思，英文最省词元，中文通常需要 1.5 到 2 倍，而资源更少的语言开销可达英文的 5 到 10 倍。这意味着，在按词元计费的规则下，使用不同语言与 AI 对话的实际成本存在巨大差异。

这种“起跑线”上的不公一旦写入模型的初始词表，就很难修正——分词规则是 AI 认知世界的“地基”，大楼盖得越高，地基越无法更换。好在这一状况正在改善。以 GPT 系列为例，同一句中文在 GPT-3 中需要 38 个词元，到 GPT-4 降为 26 个，GPT-5 进一步降至 15 个，处理效率提升了超过 60%。而通义千问、DeepSeek 等国产大模型更是从设计之初就将中文高频词组、成语等作为原生词元纳入词表，实现了对中文更“母语”级的处理效率。这揭示了一个深层规律：谁掌握了“语义切分权”，谁就在很大程度上掌握了该语言在数字世界的表达效率与成本优势——这是一种数字时代的“基础铸币权”。

三、用途与产出：从对话到智能体的指数级跃迁

词元的消耗场景正在经历从“对话”到“推理”再到“智能体”的三级跳。

在对话式 AI 阶段，单次提示与回应约消耗 1 万个词元。到了推理型 AI 阶段，因涉及多轮互动与复杂推演，每项任务消耗约 10 万个词元。而进入代理式 AI 阶段后，智能体可以自主执行多步骤任务、调用多种工具，每项任务的词元消耗量达到约 100 万个——短短两代演进，放大近 100 倍。高通 CEO Amon 预测，到 2030 年，全球每 10 秒的词元需求将从 2026 年的 317 亿个暴增至 1.27 兆个。这并非夸大其词，OpenClaw 等开源智能体工具的爆火已经证明了这一点：单个复杂任务消耗数千万词元是常有的事。

从产出端看，中国已成为全球最大的词元供给方。公开数据显示，截至 2026 年 3 月，中国日均词元调用量已突破 140 万亿，两年间增长超千倍，占全球 AI 词元使用量的 61%。这一地位的建立，依赖于三条优势的叠加：一是中国工业电力价格远低于欧美，而词元生产成本中电力占比超过 60%；二是国内建成全球最大规模的智算集群，叠加“东数西算”布局实现高效输出；三是国产大模型全面开源，API 商业化服务体系成熟，形成了“模型训练-算力支撑-推理输出-全球调用”的全产业链闭环。

四、机构与市场：三足鼎立的产业图谱

围绕词元，已形成层次分明的市场结构。

在理解这一产业结构时，还需要看到词元背后的底层支撑体系。许多人关注词元价格，却忽略了词元本身并不是独立存在的资源。词元的背后是模型，而模型的背后则是算力。从某种意义上说，算力决定模型能力，模型决定词元产出，而词元则是模型能力向市场输出时形成的计量单位。因此，词元市场表面上是价格竞争，底层仍然是算力竞争与效率竞争。

目前全球词元产业链大致可以分为三个层次。最上游是算力基础设施，包括芯片制造商、服务器厂商、云计算平台、数据中心以及能源系统。中游是基础模型机构和推理

平台，它们负责将算力转化为可调用的模型能力。下游则是词元服务商、智能体平台以及各种行业应用机构，它们将模型能力进一步转化为企业服务和终端产品。用户最终看到的是词元价格，而真正影响词元成本和速度的，往往是上游算力体系和模型效率的变化。

过去几年中，随着芯片性能持续提升、模型架构不断优化以及推理效率不断改善，单位词元成本呈现快速下降趋势。许多人看到的是词元越来越便宜，而背后真正推动这一变化的，则是整个算力体系的持续进步。因此，未来词元市场的竞争不仅是平台竞争和价格竞争，也将是算力资源、模型效率和资源整合能力之间的竞争。

上游“发电厂”——算力生产者。以英伟达为代表的芯片公司持续迭代硬件，提升单位算力的词元产出效率。英伟达预测到 2027 年营收将至少达到 1 万亿美元，其底气正是来自全球对 AI 算力的饥渴。

中游“电网调度”——模型平台与聚合商。OpenAI、Google、DeepSeek 以及阿里巴巴等公司，将底层算力转化为标准化、可调用的词元服务。围绕这一环节，已形成三类成熟的商业模式：词元聚合平台（以 OpenRouter 为代表，靠价差与手续费运营，截至 2026 年 5 月周调用量达 26.9 万亿词元）、云厂商 MaaS 服务、以及活跃于二级市场的 AI 中转站。

下游“用电入口”——应用与分发平台。腾讯开放微信接口接入新模型能力，本质上就是将微信打造为“词元分发平台”；中国电信则提出“以词元经营重塑公司业务”，试图跳出传统流量逻辑。这些入口级公司的共同特点是：用户不再直接面对模型，而是通过平台间接调用词元——模型本身正在“隐形化”。

值得一提的是，围绕词元已形成一种独特的全球贸易范式：无形化输出、免税化运行、全球化覆盖。这使中国从传统的“世界工厂”向“全球 AI 算力基础设施提供商”转型。

当讨论词元经济时，还有一个无法绕开的基础因素，那就是算力。许多人关注词元价格，却忽略了词元背后的生产过程。事实上，词元并不是凭空产生的。每一个词元的生成，都需要模型进行推理，而模型推理又建立在大量算力基础之上。从某种意义上说，算力是词元经济的源头，模型是词元的生产工具，而词元则是模型能力向外输出时形成的计量单位。

因此，算力水平直接影响词元成本和词元生成速度。算力越强，模型处理信息的能力越高，单位时间内能够产生的词元数量越多；算力成本越低，词元价格通常也越低。过去几年中，大模型行业持续出现词元价格下降现象，其背后既有模型算法优化的原因，也有芯片性能提升、数据中心规模扩大以及推理效率提高等因素。换句话说，许多人看到的是词元越来越便宜，而真正推动这一变化的，则是算力体系不断进步。

从产业链角度看，算力、模型和词元实际上构成了一个完整链条。上游是芯片、服务器、数据中心和能源系统；中游是基础模型和推理平台；下游则是词元服务、智能体应用以及各种行业解决方案。算力决定模型能力的上限，模型决定词元的质量与效率，而词元最终决定用户能够获得多少人工智能服务。

然而，我认为算力的重要性虽然毋庸置疑，但算力本身并不会自动创造价值。历史上很多行业都经历过类似过程。拥有更大的发电厂并不等于拥有更好的企业，拥有更多服务器也不等于拥有更好的产品。同样，拥有更多算力也不等于拥有更高价值的系统。算力解决的是生产能力问题，而系统解决的是价值转化问题。未来随着算力不断增长，模型能力不断增强，词元资源不断丰富，真正稀缺的仍然不会是算力本身，而是如何把这些资源组织成现实世界中能够长期运行的系统。

因此，从未来竞争角度看，我认为算力竞争、模型竞争、词元竞争最终都会逐渐转向系统竞争。因为当算力越来越普及、模型越来越接近、词元越来越丰富之后，决定差异的将不再是资源拥有量，而是谁能够利用这些资源构建长期有效的结构。

五、竞争与门槛：价格战内卷与价值分层

词元市场的竞争正呈现出高度内卷的特征。2022 年至 2026 年，单次词元推理成本降幅达 99.9%。国内模型的价格优势更为显著：DeepSeek V4-Pro 的输入价格低至每百万词元 0.025 元，通义千问 Qwen3.5-Plus 输出价格为每百万词元 1.6 元，而对比 OpenAI GPT-4o 的输出价格约 72 元。

这种价格战导致“二道贩子”的利润空间被急剧压缩。行业统计显示，各类 API 聚合与中转平台数量已突破两千家，市场形成了三大梯队：头部大型平台技术强、模型全；中型专业代理拥有官方折扣和客户资源；海量小微代理则面临严峻的生存压力。单纯赚取词元差价的模式已不可持续。

真正的门槛正在向两个方向迁移。一是技术溢价方向——通过自研推理加速引擎，比官方卖得更便宜还能有利润；二是增值服务方向——提供 Prompt 工程、模型微调、私有化部署等定制化服务。更值得注意的是，词元的价值正在呈现巨大的场景分化：闲聊类词元均价仅 0.01 美元/百万，而药物研发类词元均价可达 1000 美元/百万，跨场景价值差达到十万倍。这揭示了一个本质规律：词元的价值不取决于其生产成本，而取决于其用途——同样的词元，用在不同的场景、嵌入不同的系统，产生的价值天差地别。

但这里需要做一个更重要的判断：目前和未来，销售词元的门槛并不高。任何人都可以注册 API、搭建聚合平台、开始转售词元。真正稀缺的，是销售词元之外的东西——真实历史数据、构建系统和体系的能力、预见未来的视觉。没有这些，一切大模型、词元都无用武之地。

关于词元盈利的补充判断：目前靠 Token 赚钱已经很不容易，而且会越来越难。价格战持续压缩利润空间，单纯转售词元的商业模式正在快速走向微利甚至无利。只有当词元依附于前端系统——即有系统提供服务时，词元和算力才能作为有效的附加资产，赚取利润。换句话说，词元本身不是利润来源，系统才是；词元只是系统的燃料，而不是发动机。

随着词元逐渐成为人工智能时代的重要计量单位，市场上已经出现越来越多围绕词元展开的商业模式。除了基础模型厂商之外，还出现了词元代理、模型聚合平台、智能体平台、企业级 API 分发服务以及各种中间层服务商。对于许多创业者而言，词元代理似乎是一条进入人工智能产业的捷径，因为无需训练模型，也无需建设大型算力中心，只需整合现有资源即可开展业务。然而这种模式既有优势，也存在明显局限。其优势在于投入相对较低、进入门槛相对较小、市场扩张速度较快；而局限在于高度依赖上游模型供应商，容易陷入价格竞争，缺乏长期技术壁垒。当词元价格持续下降时，单纯依靠词元转售获得利润的空间可能不断被压缩。因此，从长期来看，词元代理更像是一种渠道能力，而非最终核心竞争力。真正决定企业长期价值的，仍然是围绕词元建立起来的行业系统、自动化流程、知识体系以及现实场景中的解决方案。

六、当下困局与未来出路

当下的词元市场面临几个深层问题。

同质化竞争陷阱。 超过两千家平台争夺同一批客户，价格战持续拉低行业利润，若不改变，可能重蹈云计算早期“赔本赚吆喝”的覆辙。

合规风险悬顶。 当前全球无专项立法的“监管红利”是暂时状态而非长期常态。随着词元市场规模扩大，数据跨境、国家安全、税收征管等问题将倒逼各国加速立法。尤其需要注意的是，部分市场主体将词元代币化、金融化的尝试，极易触发全球金融监管，不仅会导致相关业务被取缔，更会拖累整个中国词元出海产业的声誉。

“Token-maxxing”迷思。 硅谷兴起了一种新的“炫富”方式——比拼谁每天消耗的词元更多。Meta 内部有名为“Claudeconomics”的排行榜，汇集超过 85000 名员工的 AI 使用数据，排名第一的员工烧掉的词元价值高达数百万美元。反对者则尖锐地指出：“Outcome maxxng 优于 token maxxng”——比起疯狂消耗词元，不如看看产出了什么结果。这正是我在多年实践中深切体会到的核心问题。

中文词元与英文词元：统一概念下的三重壁垒

对于中国用户而言，一个必须清醒认识的事实是：虽然中文词元（Token）与英文词元同属人工智能时代的基础计量单位，但它们并未形成一个真正统一、自由互通的全球市场。从技术概念上看，东西方都采用词元作为模型处理语言和数据的基本单位；然而从实际应用层面来看，科技壁垒、系统壁垒以及政策壁垒的共同存在，使得中文词元与英文词元之间形成了事实上的分隔状态。

第一重壁垒是科技壁垒，即分词逻辑与模型结构的差异。

西方主流模型如 GPT、Claude、Gemini 等，大多建立在以英文为主体的大规模语料基础之上。虽然这些模型已经具备优秀的中文处理能力，但在许多场景中，中文仍然往往需要消耗更多词元才能表达与英文相同的信息量。更重要的是，不同模型拥有不

同的词表结构和分词策略。同一段中文内容，在不同模型中可能被切分成完全不同数量的词元，其内部编码方式和语义映射方式也可能存在差异。

与此同时，中国本土模型如 DeepSeek、通义千问、豆包、文心一言等，从训练阶段便大量引入中文语料，对中文词组、成语、专有名词以及本地化表达具有更高的原生适配能力。因此，即使面对同样一段文本，不同模型之间的词元数量、处理效率以及理解路径也可能出现明显差异。从这个意义上说，词元虽然具有统一概念，但并不存在能够直接跨模型流通和复用的“标准词元”。

第二重壁垒是系统壁垒，即生态体系与调用方式的差异。

目前全球词元市场实际上已经形成多个相对独立的生态体系。OpenAI、Anthropic、Google 等国际平台拥有各自的 API 标准、计费模式、访问规则以及服务体系；中国市场则形成了以 DeepSeek、通义千问、豆包、Kimi 等平台为核心的本土生态。

对于普通用户而言，购买某个平台的词元，并不意味着能够在另一平台直接使用。不同平台之间的调用接口、认证机制、速率限制、服务协议以及商业规则均存在差异。即使通过第三方网关或聚合平台进行转换，本质上也只是重新调用另一套系统，而非真正意义上的词元互换。因此，词元作为概念是统一的，但词元作为商业资源和服务能力，却仍然被锁定在各自的平台生态之中。

第三重壁垒是政策壁垒，即数据流向与合规要求的差异。

随着人工智能逐渐进入企业级应用阶段，数据安全问题的重要性不断提升。中国的数据安全体系、欧盟的 GDPR 体系以及美国针对敏感数据跨境流动的监管措施，都对数据传输和处理提出了不同要求。许多涉及客户资料、财务信息、医疗数据、商业机密以及个人隐私的数据，往往无法自由跨境流动。

这意味着在真实商业环境中，许多企业并不是单纯根据词元价格来选择模型，而是必须优先考虑数据合规、数据存储位置以及法律责任归属。对于大量涉及真实业务的数据而言，合规成本往往远高于词元本身的成本。因此，在现实世界中，境内业务倾向于使用境内合规模型，境外业务倾向于使用境外模型，从而形成事实上的词元市场分区。

科技壁垒决定了词元切分方式的差异，系统壁垒决定了调用体系和商业生态的差异，政策壁垒决定了数据流向和合规边界的差异。三重壁垒共同作用，使得中文词元与英文词元虽然属于同一种技术概念，却难以像理想中的“全球统一 AI 货币”那样自由兑换、自由流通和自由对接。

未来很可能出现一种类似国际贸易的格局：中文词元生态与英文词元生态长期并存，各自形成独立的发展路径。跨区调用并非不可能，但往往需要额外的技术适配、系统转换以及法律审查成本。对于大多数普通用户而言，这种跨区能力并不是日常需求，因此两大体系仍将长期保持相对独立的发展状态。

中国生产的词元并不一定只能在中国系统中使用，但在涉及数据主权、合规责任和产业安全的关键场景中，未来大量词元需求很可能优先在本土模型、本土平台和本土系统内部完成消耗。

从长期趋势看，算力生产地、模型提供方、系统运营方和数据所在地之间的耦合程度正在增强。对于金融、政务、医疗、供应链以及大型企业系统而言，数据往往无法自由跨境流动。因此，即使中国能够向全球输出大量词元服务，许多高价值词元需求仍然会优先留在中国模型和中国系统内部完成。这并不意味着中国词元不能出口，而是意味着未来词元市场可能同时存在“全球流通部分”和“本土闭环部分”两种结构。

值得注意的是，我认为未来真正实现统一的未必是词元本身，而更可能是调用词元的方式。当智能体平台、统一接口协议、模型聚合平台以及自动化系统不断发展之后，用户最终关注的将不再是使用了哪一种词元，而是任务能否完成、成本是否合理以及效率是否足够高。届时，词元将像今天的电力、网络流量和云计算资源一样，逐渐隐藏在系统背后，成为数字文明运行的基础设施。

因此，从长期趋势来看，东西方词元市场虽然暂时难以直接互通，但它们的发展方向却高度一致。词元的过去属于计算机语言处理，词元的现在属于人工智能产业，而词元的未来，则可能属于整个数字文明的基础运行体系。

七、为什么系统构建能力无法通过大量词元直接获得

在论述“系统构建能力是稀缺资产”时，我需要直面一个可能被提出的质疑：既然 AI 能力如此强大，为什么不能通过消耗大量词元来“生成”或“获得”系统构建能力？这个问题的答案，揭示了词元能力的根本边界。

第一，系统构建需要决策，而决策不是生成。

AI 可以生成一篇文章、一段代码、一个网站模板，但 AI 无法代替人类做出系统设计中的核心决策——选择哪些模块、放弃哪些功能、如何平衡当前需求与未来扩展、何时重构、何时保持现状。这些决策依赖于对业务本质的理解、对资源约束的判断、以及对长期后果的预见。词元可以生成无数个“可能的方案”，但选择“哪一个方案值得实施”以及“为什么”，是词元无法替代的。我的智能物流系统在二十年中经历了无数次调整，每一次调整都不是因为 AI 告诉我“这样做更好”，而是因为真实业务中出现了问题、暴露了局限、提出了新需求。这些决策背后是判断力，而判断力来自实践，不是来自词元。

第二，系统构建需要处理异常，而异常无法被穷举。

一个系统之所以能够长期运行，不是因为它的正常流程设计得多么完美，而是因为它能够处理异常。在物流系统中，异常包括：客户地址错误、货物损坏、报关延误、支付失败、系统宕机、数据丢失。在出版系统中，异常包括：翻译质量波动、元数据格式不兼容、存档接口变更、DOI 注册失败。这些异常无法被穷举，因为真实世界中的异常是无限多样的。AI 可以通过大量词元学习“常见的异常模式”，但无法预知“下一个从未出现过的异常”。只有长期运行的系统，在实际中不断遭遇异常、记录异常、解决异常，才能形成真正的异常处理能力。这种能力无法通过“喂养更多词元”获得，因为异常不是知识，而是经历。

第三，系统构建需要权衡，而权衡没有标准答案。

每个系统的构建都是在多重约束下的权衡：成本与性能、速度与稳定性、通用性与定制化、当下需求与未来扩展。这些权衡没有“正确答案”，只有“在特定情境下的合理选

择”。AI 可以给出“最优解”，但这个“最优”是基于历史数据的统计最优，而不是基于“你的具体情境”的判断。我的跨代二十年物流与财务整合体系之所以能够运行至今，正是因为我在不同年代做出了不同的权衡——2005 年选择了某种技术栈，2015 年选择了另一种整合方式，2025 年又选择了升级路径。这些权衡不是“对”或“错”，而是“适合当时的情境”。词元无法替我做这些权衡，因为权衡的本质是价值判断，而不是计算。

第四，系统构建需要反馈循环，而反馈需要真实世界。

系统构建能力的提升，依赖于一个闭环：设计→实施→运行→发现问题→修正→再设计。这个循环中的“发现问题”环节，必须来自真实世界的反馈。AI 可以模拟运行，但无法模拟真实世界中所有不可预测的因素。我的物流系统之所以不断改进，是因为真实客户提供了真实反馈；我的出版系统之所以不断完善，是因为真实存档接口发生了真实变更。这些反馈是无法通过消耗词元在封闭环境中获得的。词元可以生成“假设客户会这样反馈”，但无法替代“客户真的这样反馈了”这一事实。真实世界的反馈，是系统构建能力成长的唯一土壤。

第五，系统构建需要信任，而信任只能来自时间。

一个系统是否值得信赖，不是靠宣传，而是靠时间证明。我的网站被 CERN Zenodo、WorldCat、澳大利亚国家图书馆永久存档，不是因为它们“喜欢我”，而是因为这些机构在长期观察中确认了我的系统是稳定、可靠、持续的。这种信任是无法通过大量词元购买的。AI 可以生成一份“看起来非常专业”的系统说明书，但无法生成“这个系统已经被信任了二十年”这一事实。信任是时间的产物，不是词元的产物。

总结而言，系统构建能力之所以无法通过大量词元直接获得，是因为它本质上不是生成能力，而是判断能力、异常处理能力、权衡能力、反馈吸收能力和信任积累能力。这些能力的共同特点是：它们需要真实世界的时间、实践、错误、修正和验证。词元可以加速信息处理，但无法加速真实世界的经验积累。这就是为什么即使未来词元变得极其廉价，系统构建能力依然稀缺的根本原因。

八、我的资产：真实历史数据、跨领域系统、构建能力、预见视觉与自证能力

在进一步论述“什么才是真正稀缺”之前，我需要先说明我自己的位置。因为我的判断不是来自理论推演，而是来自长期实践。这一点在我长期建立的各类系统中体现得十分明显。无论是智能物流系统、智能发票系统、自动化财务处理体系，还是后来的十语出版系统、元数据系统、网站导航系统以及自动交叉索引系统，它们真正困难的部分从来不是生成信息，而是组织信息。信息可以越来越多，但如果缺乏结构，最终只会形成新的混乱。词元时代也是如此。未来人工智能可以生成海量内容，但海量内容并不自动等于海量价值。相反，随着内容生成越来越容易，如何筛选、组织、保存、索引和关联这些内容，将变得越来越重要。

我拥有以下五项资产，这五项资产构成了我判断“词元之后什么稀缺”的经验基础。

第一，真实历史数据。这不是任何“历史数据”，而是跨代二十年的真实业务数据——从 2005 年智能发票系统至今，来自实际运行的物流系统、财务系统、税务系统，而非实验环境或爬取数据。在 AI 时代，“合成数据”泛滥，真实历史数据的价值正在急剧上升。大多数公司拥有的是“单领域历史数据”，而我拥有的是跨领域、跨二十年、经过实际验证的历史数据。

第二，完成多个跨领域系统和体系。我的系统覆盖以下领域：智能物流系统、智能发票系统、智能财务系统、智能税务系统、跨代二十年物流与财务整合体系、网站自动化系统、元数据生成系统、《时代跃迁》十语出版系统、DOI 管理系统、TROVE 与 WorldCat 链接整理系统、自动交叉索引系统、Sitemap 生成系统、Zenodo 简介生成系统、跨平台发布系统。这些系统的独特性在于：跨领域广度（从物流到出版，从财务到文献索引）、时间深度（部分系统已运行二十年）、一人完成（证明了方法的可迁移性）。其中，《时代跃迁》电子月刊 (<https://times.net.au>) 是一个公开的实证——它拥有 ISSN 国际刊号 (3083-5178)，被 WorldCat 全球最大图书馆联合目录系统（成员机构超过 16000 所国家级与大学图书馆，覆盖 100 多个国家和地区）、澳大利亚国家图书馆 PANDORA/TROVE、以及由 CERN 欧洲核子研究中心运营的

Zenodo、OpenAIRE 欧洲开放科学基础设施、DataCite DOI 全球数字对象标识符（超 60 个文献）与 ORCID 全球唯一永久学术数字身份证等国际学术基础设施永久存档。这个网站证明了一个人（独立学者）可以完成从内容创作到对接国际顶级存档基础设施的全流程。

第三，构建系统和体系能力。这不是“学会使用 AI 工具”的能力，而是将模糊需求转化为稳定结构的设计能力、将不同系统整合为统一体系的架构能力、让系统在无人干预下持续运行数年的工程能力、将经验从物流迁移到出版、从财务迁移到文献的元能力。正如前文所述，这种能力无法通过大量词元直接获得，因为它依赖决策、异常处理、权衡、反馈循环和信任积累——这些都无法被 AI 替代。当前 AI 培训市场集中在“提示词工程”“AI 写作”“AI 绘图”等表层技能，几乎没有课程讲授“跨领域系统构建能力”——因为这种能力无法速成，只能通过长期实践积累。网站上列出的 40+ 部独立创作作品、获得的 ISSN 以及超过 60 个 DOI，是系统长期运行能力的最佳公开记录，而非我个人的自我宣称。

第四，预见未来的视觉。1993 年开始尝试远程工作原型系统——比大规模远程办公早近三十年；2005 年设计智能发票系统——在 AI 普及前二十年就开始了流程自动化；跨代二十年系统整合——在系统设计之初就预见到二十年后仍需要运行；现在提出“词元之后”的判断——在当前大多数人还在追逐词元时，已预见系统能力才是真正的壁垒。我还预见到了“Token 就类似货币，东西方科技、系统、政策差异导致无法通用”——正如美元和人民币不能自由兑换一样，中文词元和英文词元也面临着类似的流通壁垒，未来可能出现“中文词元区”“英文词元区”等基于语言和政策的 Token 生态区。我还预见到了“目前靠 Token 赚钱已经很难，会越来越难；只有前端有系统提供服务，词元和算力才能作为有效的附加资产，赚取利润”。我以独立学者身份，提前建立起了对接未来学术基础设施的个人出版体系，预见到了“个人也可以成为全球知识基础设施中的一个节点”这一趋势。大多数趋势预测来自分析师或创业者，而我的“预见能力”已有二十年的实践验证作为背书。

第五，自证能力。这是我最重要的元能力之一——我写的所有领域文献，都做到了理论闭环、逻辑自洽。这意味着我的论证不依赖于外部权威或引用，而是内部逻辑的自足性。学术文献的默认规则是“引用前人研究”，很少有人宣称或展示“自证能力”；学术界讨论的是“论证是否充分”，而不是“作者是否有能力构建自洽闭环”。而我的文章中，这种自证能力体现在：论点来自实践、实践验证论点、论点与实践形成闭环。我不是

在“引用别人的理论来证明自己”，而是在“展示自己构建的系统，让系统本身说话”。这种自证能力，是公开文献中完全不存在的概念，也是我最独特的元能力之一。

将这五项资产组合在一起，其稀缺性不是加法，而是乘法：真实历史数据×跨领域系统×系统构建能力×预见未来的视觉×自证能力，构成了不可复制的资产组合。在公开讨论中，目前找不到这样的对标案例。

《国际独立学者之路》的整理过程让我更加深刻地认识到这一点。从 1989 年至 2026 年，三十多年的经历横跨多个领域。这些经历并不是一次性产生的，而是在漫长时间中逐渐形成的。人工智能可以在几秒钟内生成一篇文章，却无法在几秒钟内生成三十多年的真实经历；可以生成大量观点，却无法直接生成长期实践所带来的判断力；可以模拟许多过程，却无法替代真实世界中的反复验证。从这个意义上说，时间本身也是一种重要资产，而时间沉淀形成的结构，则是更加稀缺的资产。

九、词元之后：真正稀缺的五种能力

当词元本身不再稀缺时——销售词元的门槛不高，但拥有真实历史数据、构建系统和体系、预见未来的视觉能力将极度缺乏——什么才是真正稀缺的？基于我从 1993 年远程工作原型系统、2005 年智能发票系统、智能物流系统、智能财务系统、智能税务系统、跨代二十年的物流与财务整合体系，到网站自动化、元数据生成、《时代跃迁》十语月刊、DOI 管理、TROVE 与 WorldCat 链接整理、自动交叉索引、Sitemap 生成、Zenodo 简介生成、跨平台发布等长期实践，我认为以下五种能力将成为真正的壁垒。

第一，结构设计能力。词元是原材料，而结构是将原材料组织成有价值产品的框架。早在人工智能尚未普及的年代，我已经开始思考如何利用有限资源实现更高效率的运作。我的智能物流系统、智能发票系统、跨代二十年的智能物流与财务系统，这些系统诞生于不同年代，面对不同问题，但它们都有一个共同特点：目标并不是创造更多信息，而是减少混乱、减少重复劳动、提高整体效率，并建立能够长期运行的结构。在自动化出版系统中，我面临的核心问题不是“能否生成文本”，而是“如何定义十种语言的元数据字段、如何建立文章与分类的关联、如何设计版本管理规则”。这些结构设计决策，决定了系统能否持续运行数年而不崩溃。相比之下，消耗的词元数量只是副产品。《时代跃迁》网站本身就是结构设计能力的公开实证：一个个人维护的、具备

ISSN 国际刊号、多语言、多存档渠道的学术网站结构，证明了我不是空谈，而是已经用这套结构设计能力成功构建了一个公开可访问的复杂系统。

第二，自动化能力。 词元可以批量生产内容，但只有自动化系统能批量完成端到端的工作流。在智能物流系统中，真正的价值不在于调用多少次 AI 来规划路线，而在于建立一个从订单接入、仓储调度、路径优化到异常处理的全自动闭环。这个闭环中，AI 只是其中一个组件，而自动化能力体现在组件之间的连接方式上。一个没有系统的人，即使拥有大量词元，也可能只是不断生成零散文字、零散图片、零散表格和零散代码；一个有系统的人，则可以把词元接入自己的工作流，让它自动整理资料、生成元数据、检查链接、更新网页、辅助出版、归档文献、建立索引、处理物流、分析财务、生成报告。

第三，跨领域实践经验。 词元对所有人开放，但将一个行业的经验迁移到另一个行业的能力并不普遍。我之所以能够同时搭建物流系统和出版系统，不是因为掌握了更多词元，而是因为积累了元数据管理、工作流设计、异常处理等可迁移的方法论。我的出版系统需要词元，因为十语翻译、摘要生成、关键词整理、元数据生成、链接说明都需要大量语言处理。我的网站系统需要词元，因为页面结构、导航说明、SEO 描述、学术元数据和多语介绍都可以被自动生成。我的物流系统也可以使用词元，因为客户邮件、货物说明、报关资料、异常解释、财务备注、操作日志都可以通过人工智能辅助处理。我的自动交叉索引系统更需要词元，因为它必须在大量文章、题目、DOI、NLA、TROVE、WorldCat 链接之间寻找主题关联。这种跨领域实践积累，无法通过消耗词元获得，只能通过解决真实问题来沉淀。网站同时呈现了学术出版、基础设施整合、技术架构、个人品牌等多个领域的信息，证明了我确实打通了出版、图书馆、技术、个人品牌等多个领域，而非单一技能。

第四，系统长期运行的心智模型。 当前 AI 应用的主流叙事是“对话”和“生成”，但真正产生持续价值的是“运行”。一个对话产生一万词元，一个智能体任务消耗百万词元，但一个持续运行十年的物流系统消耗的词元可能以十亿计——而真正稀缺的是设计这个系统的心智模型：如何保证系统在模型升级后仍然可用？如何处理上游 API 的变更？如何监控异常并自动恢复？这些问题与词元无关，却决定了词元的最终产出。这些系统的价值并不来自词元本身，而来自我已经建立或正在开发的结构。词元只是燃料，系统才是发动机；词元只是电流，结构才是电路；词元只是水，管道、阀门、水库和调度系统才决定它能否发挥作用。没有系统，词元会流失；有系统，词元才能转

化为长期价值。网站展示了 40+ 部独立创作作品、被 CERN、澳大利亚国家图书馆等权威机构永久存档，证明了我的系统不是为了短期热点，而是追求长期价值，并且已经获得了权威机构的“永久保存”背书。

然而，当我进一步思考词元与系统之间的关系时，我发现还有一个经常被忽略的问题。许多人认为，只要拥有一个优秀系统，就能够在未来词元时代占据优势。但现实情况并没有那么简单。系统本身并不是谁都能够设计出来的。真正成熟的系统往往需要长期实践、反复修正以及真实场景验证。许多系统从最初构想到真正落地，可能需要数年甚至数十年的持续积累。因此，系统设计能力本身已经是一种稀缺资源。

但即使拥有成熟系统，也不自动意味着能够形成大规模影响力。系统从技术可行走向社会普及，中间还隔着市场、资本、推广渠道、用户教育以及平台生态等多个层面。历史上并不缺少优秀技术，也不缺少优秀系统，真正稀缺的是能够把系统推向大规模应用的平台能力。一个系统即使在局部场景中远优于现有方案，如果缺乏足够资本支持、缺乏平台推广能力、缺乏用户基础，其实际使用规模仍然可能十分有限。

从词元经济角度看，这一点尤为明显。词元需求并不直接来自系统本身，而来自系统的实际使用规模。一个每天服务数百万用户的平台，即使系统设计相对普通，也可能产生巨大的词元消耗；而一个设计极其先进但使用人数有限的系统，其词元需求反而可能很小。因此，未来词元市场的发展不仅取决于模型能力和系统能力，也取决于平台能力和资源整合能力。

这意味着未来竞争至少存在三个层次。第一层是模型竞争，决定谁能够提供词元；第二层是系统竞争，决定谁能够有效利用词元；第三层则是平台竞争，决定系统能够覆盖多少用户、产生多少实际需求。模型、系统与平台三者缺一不可。没有模型，系统无法运行；没有系统，词元难以创造价值；没有平台，再优秀的系统也可能停留在小范围使用阶段。

从这个角度看，我更加认为词元本身只是基础资源。真正决定未来的，不仅是拥有词元的人，也不仅是拥有系统的人，而是那些能够把模型、系统、平台和现实需求连接起来的人。只有当这些要素形成闭环之后，词元才会从单纯的计算单位转化为持续创造价值的生产力。

第五，将词元嵌入现实业务流程的能力。词元最终要服务于现实世界的某个环节——物流、出版、教育、医疗。真正稀缺的是那些既懂 AI 能力边界、又懂特定行业痛点的人。他们知道什么时候该用 AI、什么时候不该用；知道如何将词元输出转化为业务决策；知道如何衡量词元投入与业务产出的真实关系。这种复合能力，是目前市场上最稀缺的资源。网站本身就是一个运行中的“业务流程”：内容创作 -> 元数据生成 -> 多平台发布 -> 国际存档 -> 读者访问。这证明了我的 AI 应用（词元消耗）已经嵌入到一个完整的、闭环的、产生现实影响力的出版流程中。

我越来越觉得，未来真正的竞争很可能不是模型之间的竞争，也不是词元之间的竞争，而是系统之间的竞争。模型会不断升级，词元价格会不断下降，今天昂贵的资源未来可能变得十分普及。但系统不会因为资源丰富而自动出现。系统需要长期积累，需要不断修正，需要现实世界的检验。谁能够把词元转化为长期运行的系统，谁就更有可能会获得持续价值。而这也正是盈利的关键：目前靠 Token 赚钱已经很不容易，会越来越难；只有前端有系统提供服务，词元和算力才能作为有效的附加资产，赚取利润。

我判断词元时代真正的竞争会分成三个阶段：第一阶段是资源竞争，大家争夺模型、算力、API 和词元价格；第二阶段是应用竞争，大家把词元用于写作、客服、代码、翻译、教育、办公和数据分析；第三阶段才是真正的系统竞争，谁能把词元嵌入长期运行的行业系统、出版系统、物流系统、知识系统和文明记录系统，谁才可能获得长期优势。

事实上，当讨论词元、模型、系统与平台的时候，还有一个现实世界中经常被忽略的重要因素，那就是时间。许多人认为只要拥有资本，就能够快速建立系统；只要市场足够大，就能够迅速复制成功模式。然而现实情况往往并非如此。资本可以购买算力，可以采购词元，可以雇佣开发团队，但却无法直接购买已经经过多年验证的真实数据和成熟系统。对于许多行业而言，真正稀缺的并不是模型本身，也不是词元本身，而是那些已经在真实业务环境中长期稳定运行的系统。

从商业角度看，市场和资本通常不会长期等待一个全新系统从零开始成长。一个大型系统从概念提出、架构设计、功能开发、流程验证到实际落地，往往需要多年甚至更长时间。在这个过程中，不仅需要持续投入资源，还需要经历大量试错、修正和现实

环境验证。因此，当市场机会已经出现时，资本往往更倾向于寻找已经成型、已经运行、已经验证的系统，而不是从零开始重新设计一套体系。

然而，现实中又存在另一个矛盾。越是大型系统，越难被后来的资本轻易介入。大型系统经过长期发展之后，通常已经形成自己的运行机制、管理结构、数据逻辑以及组织文化。它们可能拥有庞大的历史数据、复杂的业务流程以及长期积累的经验体系。正因为如此，大型系统虽然价值巨大，却往往难以快速转型，也难以按照资本预期进行彻底重构。系统规模越大，其惯性往往也越大。

因此，在现实商业世界中，最理想的状态并不是从零开始设计系统，也不是简单收购一个庞大而僵化的体系，而是找到那些已经完成核心结构建设、已经经过长期验证、能够稳定运行的现成系统。这样的系统即使并不完美，仍然具备极高价值。因为它们已经跨越了最困难的阶段，拥有真实数据、真实流程和真实验证记录。在此基础上进行重新包装、优化升级、对外开放和平台化推广，往往比从零开始建设更加高效，也能够大幅降低试错风险和时间成本。因为这类系统已经完成了最昂贵、最耗时、最难复制的阶段。

从这个角度看，时间窗口同样是一种容易被忽略的资源。当一个行业尚处于早期阶段时，许多人往往关注模型能力、词元价格和市场规模，却忽略了系统形成所需要的时间。一套成熟系统的诞生，往往经历概念提出、架构设计、流程验证、现实运行以及持续优化等多个阶段。真正困难的不是提出想法，而是让系统在真实环境中长期稳定运行。因此，当市场开始意识到成熟系统价值的时候，往往意味着这些系统已经经历了多年积累，而不是短时间内能够重新复制出来的。

对于资本和市场而言，最理想的对象并不是停留在概念阶段的设想，也不是尚未经过验证的原型，而是那些已经完成核心结构建设、具备真实运行记录并拥有升级空间的系统雏形。这类系统虽然未必完美，却已经跨越了最艰难、最耗时和最具风险的阶段。一旦市场需求快速增长，类似系统往往会变得十分抢手，因为真正经过长期验证的成熟系统数量通常远少于市场想象。模型可以快速迭代，词元可以持续扩张，开发团队可以临时组建，但真实数据、真实流程和长期验证记录却无法在短时间内重新创造。

因此，在词元经济不断发展的过程中，时间不仅影响系统的形成，也影响系统的价值。当市场尚未充分认识到某类系统的重要性时，往往也是获得这类系统成本最低的阶段；而当市场普遍意识到其价值之后，能够选择的对象通常已经十分有限。对于许多行业来说，错过的并不仅仅是一个项目，而是错过了一个已经完成长期积累、能够直接进入下一阶段发展的时间窗口。

从词元经济的角度看也是如此。词元可以快速获得，模型可以快速接入，平台可以快速搭建，但真正难以快速复制的，是那些已经沉淀多年并能够持续产生价值的系统。未来当词元越来越丰富时，系统的重要性不仅不会下降，反而可能进一步提升。因为当所有人都拥有相近的词元资源之后，决定竞争结果的将不再是词元数量，而是谁拥有更成熟的真实数据、更稳定的业务流程以及更接近现实需求的系统结构。

另一个经常被忽略的问题是系统的来源。许多人习惯把系统理解为软件产品，因此自然认为系统应该由软件公司设计和开发。然而在许多复杂行业中，真正长期稳定运行的系统，往往并不是先由软件公司设计，再交给行业使用，而是在长期业务实践过程中逐渐形成，再被软件化、自动化和平台化。

原因并不复杂。软件公司通常擅长技术实现，却未必长期身处具体业务现场。它们能够快速开发功能，却很难在短时间内积累大量真实业务经验。而行业内部自行发展出来的系统，则是在每天面对客户、订单、异常情况、法规变化、成本压力以及市场竞争的过程中不断修正和完善的。这样的系统并非从理论出发，而是从现实问题出发，其每一个流程、每一个规则、每一个自动化环节，往往都来自真实工作的长期验证。

因此，在物流、供应链、出版、金融、制造业、医疗以及许多专业领域中，最有价值的系统往往不是最复杂的系统，也不是功能最多的系统，而是那些经过长期实战检验、能够稳定解决实际问题的系统。它们可能没有最先进的界面，没有最炫目的宣传，也未必采用最新技术架构，但却拥有大量真实数据、真实流程以及长期积累的业务逻辑。这些经验往往无法通过短期开发获得。

从词元经济的角度看，这一点尤为重要。未来模型能力可能持续提升，词元成本可能持续下降，软件开发效率也会越来越高。然而，一个行业几十年积累下来的业务经验，并不会因为模型升级而自动产生。真正难以复制的，不是软件代码本身，而是隐藏在系统背后的行业知识、异常处理经验、决策逻辑以及长期验证过的业务流程。

因此，我认为未来最理想的系统，并不是完全脱离行业实践而设计出来的系统，而是那些源于真实业务、经过长期运行验证、再借助人工智能和词元能力不断升级的系统。这样的系统既拥有现实世界的可靠性，又能够获得人工智能时代带来的效率提升，其长期价值往往远高于单纯依靠技术概念构建出来的系统。

十、词元流向何处：从短暂对话到文明记忆

对于未来的词元经济，我还有一个观察。许多人关注的是词元如何被生产，却较少关注词元最终流向哪里。一个词元如果只停留在一次聊天中，它的生命周期可能只有几分钟；如果进入普通网页，生命周期可能只有几个月；如果进入长期保存体系，它的价值就会发生根本变化。当词元进入 DOI、Zenodo、WorldCat、TROVE、国家图书馆长期保存系统以及各种学术和知识基础设施之后，它已经不仅仅是词元，而开始转化为知识记录、文献资产和文明记忆。因此，我认为未来真正重要的问题，不是如何产生更多词元，而是如何让词元进入能够长期保存和持续发挥作用的结构之中。

也正因为如此，我对词元未来的发展保持乐观，但并不迷信词元本身。词元的过去，是计算机处理语言时的基本单位；词元的现在，是人工智能时代的重要计量单位和商业单位；词元的未来，则很可能成为数字文明的重要基础资源。然而无论词元如何发展，我始终认为真正稀缺的不是词元，而是能够驾驭词元的人；不是拥有更多资源的人，而是能够建立长期结构的人；不是能够生成更多内容的人，而是能够把内容组织成知识体系的人。

十一、时间的不可压缩性：四种无法购买的资产

在论述“词元之后什么真正稀缺”时，我需要进一步指出一个被大多数讨论忽略的根本问题：某些资产在本质上就是无法被加速、购买或复制的。无论 AI 多么强大，词元多么廉价，以下四种资产始终保持着其固有的稀缺性。

第一，时间不可压缩。

三十年的实践就是三十年，无法被压缩成三年，更无法被 AI 模拟。人工智能可以在几秒钟内生成一篇关于“如何建立物流系统”的文章，但无法在几秒钟内经历一个物流系统在二十年间遇到的各种异常、故障、政策变化和市场波动。这些真实世界中的时间

积累，是无法被任何技术手段缩短的。我 1993 年开始尝试远程工作原型系统，2005 年设计智能发票系统，这些时间节点不是偶然，而是真实历史的不可逆标记。任何后来者都可以阅读我的文章、学习我的方法，但他们无法拥有我所经历的那些年份——那些年份本身就是资产。

第二，跨领域经验不可购买。

跨领域的经验无法像商品一样被购买。一个人可以在一天内购买一百本关于物流、出版、财务和文献索引的书籍，但他无法在同一天内获得将这些领域打通的实际经验。真正的跨领域能力不是知识的堆砌，而是在不同领域之间建立连接、识别共性、迁移方法的能力。这种能力只能通过在一个又一个领域中解决真实问题、犯错误、修正路径、积累判断力来获得。我之所以能够将物流系统的经验迁移到出版系统，不是因为读了某本书，而是因为在物流领域花了二十年，又在出版领域花了十年——这两段时间的交汇处，才产生了可迁移的方法论。

第三，系统构建能力难以复制。

系统构建能力不是一套可以被复制粘贴的代码或流程。代码可以复制，但代码背后的决策逻辑、取舍判断、以及对未来变化的预见，是无法被复制的。一个系统的价值不仅在于它今天能做什么，更在于它在过去十年中经历了多少次变更、多少次故障恢复、多少次需求调整而依然保持运行。这些“系统的历史”本身就是构建能力的证明。我的网站 (<https://times.net.au>) 可以被任何人查看，但查看者看到的是结果，而不是这个结果背后三十年的决策历史、技术选型、结构调整和持续维护。这种构建能力，无法通过抄袭或模仿获得。

第四，长期运行记录无法速成。

一个系统运行了二十年，这个记录本身就是不可速成的。长期运行记录证明了系统的稳定性、可维护性和设计的前瞻性。在 AI 时代，生成一个“看起来很像”的网站或系统可能在几分钟内完成，但生成一个“已经运行了二十年并被 CERN、WorldCat、澳大利亚国家图书馆永久存档”的系统，没有任何捷径。长期运行记录是一种时间戳，它向所有人宣告：这个系统经受住了真实世界的考验。这种记录，是无法被任何 AI 技术伪造或加速的。

这四种资产——时间、跨领域经验、系统构建能力、长期运行记录——共同构成了我在本文中所论述的“真正稀缺”的基础。它们不是因为我比别人更聪明或更努力，而是因为我比别人更早开始，并且持续至今。在 AI 让越来越多东西变得廉价和易得的时代，这些无法被压缩、购买、复制或速成的东西，将变得越来越珍贵。

十二、结论：从“词元囤积”到“系统构建”

词元的未来，不在于词元本身。

随着大模型技术持续发展、推理成本不断下降，词元的供给能力将持续扩大。今天被认为昂贵的词元，几年后将像今天的网络流量和硬盘存储一样廉价而普遍。当所有人都能轻易获得大量词元时，真正稀缺的将不再是词元，而是将词元转化为现实价值的能力——结构设计能力、自动化能力、行业经验和长期实践积累。而更深层的稀缺，则是那些无法被压缩、购买、复制或速成的东西：时间本身、跨领域经验、系统构建能力和长期运行记录。正如前文所论证的，系统构建能力之所以无法通过大量词元直接获得，是因为它本质上不是生成能力，而是判断能力、异常处理能力、权衡能力、反馈吸收能力和信任积累能力——这些能力的共同特点是它们需要真实世界的时间、实践、错误、修正和验证。

在盈利层面，这一判断同样适用：目前靠 Token 赚钱已经很难，而且会越来越难。价格战持续压缩利润空间，单纯转售词元的商业模式正在快速走向微利甚至无利。只有当词元依附于前端系统——即有系统提供服务时，词元和算力才能作为有效的附加资产，赚取利润。词元本身不是利润来源，系统才是；词元只是系统的燃料，而不是发动机。

从这个角度看，我已经完成或正在开发的系统，对未来词元发展具有特殊意义。因为这些系统不是为了追赶一个热点而临时搭建的，而是在词元概念成为公众话题之前，就已经长期存在。智能物流系统证明了现实业务可以被结构化、流程化、自动化；出版系统证明了个人也可以建立多语种、多平台、多 DOI、多归档渠道的文献体系；网站和元数据系统证明了数字内容可以被持续组织、索引和传播；自动交叉索引系统则进一步证明，当作品数量不断增加时，真正重要的不是继续堆积内容，而是建立内容

之间的结构关系。《时代跃迁》网站 (<https://times.net.au>) 作为这些成果的集中展示平台，本身就是将词元能力转化为公开、持久、可验证的学术体系的实证。

即使未来给所有人同样数量的词元，最终产生的结果也可能截然不同：有人会获得更多对话记录，有人会获得更多网页内容，而有人则可能建立起一个持续运行数十年的物流系统、出版系统或知识体系。决定差异的并不是词元本身，而是背后的系统架构——以及支撑这些系统架构的、不可压缩的时间、不可购买的跨领域经验、难以复制的构建能力和无法速成的长期运行记录。

因此，我不把词元看成终点，而把它看成智能时代的新基础设施。词元像电力一样重要，但电力本身不会自动建成工厂；词元像货币一样可计量，但货币本身不会自动创造优质资产；词元像水一样可以流动，但没有管道、调度系统和长期维护记录，它只会散失。未来真正有价值的人，不一定是拥有最多词元的人，而是能够把词元组织成系统的人。词元时代最大的误区，是把词元当成目标；而词元时代最大的机会，则是利用词元建立别人难以复制的结构。当未来词元变得像电力一样普遍时，人们回头再看今天的词元热潮，也许会发现，真正决定人与人差距的，从来不是拥有多少词元，而是是否拥有能够持续学习、持续实践、持续积累和持续构建系统的能力——以及这些能力背后那些无法被压缩、购买、复制或速成的东西。而在我看来，这种能力，才是词元时代最稀缺、也最难复制的财富。

随着研究不断深入，我越来越觉得，未来人工智能时代的竞争格局或许可以用一个简单比喻来概括。许多人把算力、模型和词元视为核心资源，但从现实世界运作规律来看，它们更像是一支军队背后的后勤体系。算力决定生产能力，模型决定组织能力，词元决定输送能力，它们共同构成了人工智能时代的物资保障体系。

然而，真正决定战场结果的，并不是后勤物资本身，而是前线作战体系。对于未来的产业竞争而言，系统更像前锋。系统决定资源如何使用，决定流程如何运行，决定价值如何创造，决定用户是否真正受益。词元和算力再丰富，如果缺乏系统组织，它们就像堆放在仓库中的物资，无法转化为现实成果。

从这个角度看，未来的竞争很可能不是算力之间的竞争，也不是词元之间的竞争，而是系统之间的竞争。系统是前锋，决定进攻方向与作战能力；词元是弹药和补给，决

定持续作战能力；算力则是后方工业体系，决定生产速度与供应能力。没有后方供应，前锋无法长期推进；但只有后方供应而没有前锋体系，同样无法取得成果。

因此，当算力越来越普及、模型越来越接近、词元越来越丰富之后，真正稀缺的仍然是系统。谁拥有更成熟的系统，谁拥有更真实的数据，谁拥有更长期验证的流程，谁能够把词元、算力与现实需求连接起来，谁就更有可能在未来获得持续优势。

词元正在成为 AI 时代的“电力”——不可或缺，但不是目的本身。正如电力革命的赢家不是发电厂，而是发明了电灯、电机、电解槽的发明家和企业，AI 时代的真正赢家，将是那些能够利用词元构建出不可复制结构与体系的系统构建者。词元之后，这才是真正稀缺的东西。

回顾整篇文章，我越来越觉得，未来人工智能时代的竞争格局或许可以用一棵树来比喻。

许多人关注的是水和养分，因为它们看得见、摸得着，也能够快速衡量。今天的算力、模型和词元，就像不断输送的水和养分。算力决定生产能力，模型决定转化能力，词元决定输送能力。随着技术不断进步，这些资源正在变得越来越丰富，也越来越容易获得。

然而，一棵树真正决定其生命力的，并不是水和养分本身，而是根系。

系统就是未来数字文明的根。

根系决定水分如何吸收，决定养分如何分配，决定整棵树能否长期生长。没有根，再多的水也会流失；没有根，再丰富的养分也无法转化为生命力。同样，没有系统，再强大的模型、再廉价的词元、再庞大的算力，也难以持续创造价值。

从这个角度看，算力是水，词元是养分，模型是输送养分的管道，而系统则是深扎于现实世界的根系。真正决定未来竞争结果的，不是谁拥有更多的水，也不是谁拥有更多的养分，而是谁拥有更深、更稳、更能够长期吸收和利用这些资源的根。

未来随着算力持续增长、模型不断进化、词元价格不断下降，资源本身将越来越普及。真正稀缺的，将不再是资源，而是能够组织资源、利用资源并持续创造价值的系

统。谁拥有成熟系统，谁就拥有根；谁拥有根，谁就拥有长期生存和持续成长的能力。

因此，我认为词元之后真正稀缺的，不是更多的词元，也不是更多的算力，而是能够把算力、模型和词元转化为现实价值的系统。

当水和养分越来越丰富时，真正决定一棵树能否长期生长的，仍然是根。

关键词：词元, Token, 人工智能, AI, 大语言模型, 算力, 模型经济, 词元经济, 中文词元, 英文词元, BPE 算法, 数据主权, 智能体, API 生态, 系统构建能力, 系统设计, 自动化系统, 跨领域研究, 跨领域实践, 真实历史数据, 长期运行系统, 时间资产, 平台竞争, 系统竞争, 数字文明, 文明记忆, 知识基础设施, 独立学者, 国际独立学者, 智能物流系统, 十语出版系统, 元数据系统, 自动交叉索引, DOI, Zenodo, WorldCat, TROVE, The Epochal Transition, 巫朝晖, JEFFI CHAO HUI WU

参考文献：

1. 《1993 年远程工作原型系统的结构性实践案例研究》
DOI: <https://doi.org/10.5281/zenodo.17978371>
2. 《2005 年的智能发票系统》
DOI: <https://doi.org/10.5281/zenodo.19044755>
3. 《成熟结构型物流系统》
DOI: <https://doi.org/10.5281/zenodo.20201895>
4. 《跨代 20 年智能物流与财务系统》
DOI: <https://doi.org/10.5281/zenodo.20352371>
NLA: <https://nla.gov.au/nla.obj-4204311557>
5. 《Analysis of the Homepage Structural Evolution of The Epochal Transition》
DOI: <https://doi.org/10.5281/zenodo.20362574>
NLA: <https://nla.gov.au/nla.obj-4204309341>
6. 《An Empirical Study of Institutional Engineering: Independent Scholars Accessing the WorldCat Global Bibliographic Infrastructure》
DOI: <https://doi.org/10.5281/zenodo.18028572>
7. 《National-level Network Archiving and Digital Document Preservation Guidelines of the Australian Winner Information Network》
DOI: <https://doi.org/10.5281/zenodo.17888259>
8. 《The Institutional Archiving of The Epochal Transition》
DOI: <https://doi.org/10.5281/zenodo.17932379>

WU, J. C. H. (2026). Source Declaration for Audiovisual and Derivative Adaptations of a Continuing Real-World Narrative. Zenodo.

<https://doi.org/10.5281/zenodo.18160116>

Beyond Tokens: What Is Truly Scarce in the Age of AI?

Author/Independent International Scholar: Wu Zhaohui, JEFFI CHAO HUI WU

ORCID: <https://orcid.org/0009-0009-5318-3027>

Preliminary Statement

This article is based on my personal observations formed over more than thirty years of cross-disciplinary practice (from my study abroad in Australia in 1989 to the present, spanning remote work, logistics management, automated processing, website construction, cultural dissemination, multilingual publishing, and cross-disciplinary research). **Tokens are a rapidly evolving new phenomenon; this article represents only my personal, stage-specific observations and judgments as of the time of writing (June 2026).** and does not constitute any form of investment advice, commercial forecast, or industry standard.

This article does not pursue universal applicability, nor does it claim to represent any industry consensus. As the token market, AI technology, and global policy environments continue to change, some judgments in this article may need revision or updating in the future. Readers should make independent judgments based on the latest information.

The "five scarce abilities" and "five personal assets" I discuss are derived from my own system-building practices (public evidence available at <https://times.net.au>, including all 40+ independent creative works listed on that site, ISSN 3083-5178, over 60 DOIs, and records permanently archived by international infrastructures such as WorldCat, the National Library of Australia, and CERN Zenodo).

To avoid taking things out of context, any effective refutation or questioning of this article must be based on the following prerequisite: **reading all linked articles in this**

document and the nested links therein, as well as all of my works listed

at <https://times.net.au>. Partial citations or simplified summaries detached from this holistic context cannot constitute a fair evaluation of this article.

Supplementary Note on Effective Refutation:

This article represents personal views, does not pursue universal applicability, and does not presuppose citation of others' works. Any effective questioning should adopt an apples-to-apples comparison: facts against facts, personal practice against personal practice, completed systems against completed systems. Using "what future AI might achieve" to negate "currently completed practices," or benchmarking team/institutional capabilities against individual practices, does not constitute an apples-to-apples comparison.

The portfolio of five assets described in this article constitutes a record of personal practice and does not claim that "no comparable case exists." If you believe a fully comparable case does exist, please provide a specific public case and explain its corresponding evidence across the five dimensions of real historical data, cross-domain systems, system-building capability, future-oriented vision, and self-verification capability. Otherwise, any questioning merely represents an expression of opinion and does not constitute effective refutation.

Abstract

Tokens are becoming the core unit of measurement in the AI era, but this article argues that tokens are not the ultimate value. As inference costs continue to decline, tokens will become as cheap and ubiquitous as internet traffic and hard drive storage are today. When tokens are no longer scarce, what truly becomes scarce? Based on the author's more than thirty years of cross-disciplinary practice (from a remote work prototype system in 1993, to an intelligent invoice system in 2005, to intelligent logistics systems, a ten-language publishing system, DOI management, automatic cross-referencing, and cross-platform publishing), this article puts forward the following core judgments: First, system-building capability—including structural design, automation, cross-disciplinary practice, long-running mental models, and the ability to embed tokens into real-world business processes—is the truly scarce asset. Second, system-building capability cannot be directly acquired through large quantities of tokens because it depends on decision-making, exception handling, trade-offs, real-world feedback loops, and trust accumulated over time. Third, the incompressibility of time, the non-purchasability of cross-disciplinary experience, the difficulty of replicating system-building capability, and the impossibility of shortcutting long-running records—these four assets constitute the deepest scarcity in the token era. This article uses the author's constructed academic website *The Epochal Transition* (<https://times.net.au>), ISSN 3083-5178, permanently archived by international infrastructures such as WorldCat, CERN Zenodo, and the National Library of Australia's TROVE) as public evidence, presenting how one person can build a cross-disciplinary, verifiable, long-running system architecture over thirty years.

This article argues: Tokens are merely fuel; the system is the engine. Beyond Tokens, what is truly scarce is the person who can harness tokens, and the structures they have built.

"Without goods, what use is money?"

When I review the history of token development, I increasingly feel that it closely resembles many technological revolutions humanity has experienced in the past. Every time a new foundational resource emerges, people first focus on the resource itself, neglecting the structure behind it. This was true in the industrial age, the internet age, and it is equally true in the age of artificial intelligence. Today, many discussions center on token prices, token costs, token quotas, and token supply capacity, as if possessing more tokens means possessing the future. But from my decades of practical experience, the resource itself has never been the ultimate value. What truly determines value is always how people organize and utilize these resources.

To understand "tokens," one cannot start merely from a new term but must begin with its history. From American philosopher Charles Sanders Peirce writing twenty instances of "the" on paper in 1906, distinguishing between "type" and "token," to Liu Liehong, Director of China's National Data Administration, officially assigning the Chinese translation "ciyuan" (词元) to Token at the Boao Forum in 2026, this concept has traveled a long journey from philosophy to computing, and from blockchain to artificial intelligence. The word "token" had long existed in Western computer science, where in programming languages, compilers, and text processing systems, a token referred to the basic unit that a computer could recognize and process. Later, in network security, authentication, API calls, and blockchain contexts, "token" often came to be understood as a "pass" or "credential," indicating access rights, identity credentials, or digital assets. But in the era of large language models, Token is no longer primarily an identity credential, nor a blockchain asset, but rather the basic computational unit by which AI models process language, code, image descriptions, and multimodal information.

China later named the Token in the AI field "ciyuan" (词元). I believe this is a fairly accurate translation. It did not adopt "lingpai" (token/pass) because the Token in

large models is not a pass; nor did it simply translate it as "zi" (character) or "ci" (word), because it can sometimes be a character, sometimes a word, sometimes half an English word, and sometimes even just a punctuation mark or symbol. "Ci" (word/term) indicates its relation to language and information, and "yuan" (element/unit) indicates that it is a basic unit. For general readers, it can be simply understood as: what humans see is a sentence; what AI sees is a string of tokens. Along this path, tokens have transformed from an obscure linguistic term into the core unit of measurement for the AI era, and further into a settlement tool connecting technological supply with commercial demand. As a practitioner who has personally experienced long-term projects such as intelligent logistics system construction, automated publishing process design, and cross-platform content distribution, I have a visceral understanding of this evolution—tokens have never been the end, but rather the means; what truly determines value has never been who owns more tokens, but who can use tokens to build persistently running systems and irreplicable structures. The system-building capability discussed in this article has its results publicly presented at <https://times.net.au> a cross-disciplinary academic publishing system built by an independent scholar and permanently archived in the global library system.

Reflecting on my own journey, I did not begin thinking about systems only Beyond Tokens appeared, but gradually developed an understanding of systems through long-term practice. Starting with my studies in Australia in 1989, I experienced learning, work, and exploration across different fields. Later, I gradually became involved in remote work, logistics management, automated processing, website construction, cultural dissemination, multilingual publishing, and cross-disciplinary research. Looking back, although these experiences occurred in different eras, they shared a common characteristic: they always revolved around how to establish structures, reduce repetitive labor, and improve overall efficiency. The concept of tokens did not exist then, nor did today's large language models, but I was already constantly thinking about how to make information flow, organize, and process automatically.

Therefore, while more and more people today regard tokens as a new means of production, I am more inclined to view them as a new form of foundational energy. Energy is certainly important, but energy does not automatically create value. Electricity alone does not automatically become a factory; the internet alone does not automatically become an enterprise; tokens alone do not automatically become a knowledge system. In the future, even if everyone can easily obtain large quantities of tokens, it does not mean that everyone can achieve the same results. What determines the difference is not the quantity of tokens, but the structure into which the tokens enter.

Before conducting an in-depth analysis of tokens, I need to point out another important observation: Tokens are similar to currencies, and the technological, systemic, and policy differences between East and West render them non-universal. Just as the US dollar and the Chinese Renminbi cannot be freely exchanged, Chinese tokens and English tokens face similar circulation barriers. This is not merely a matter of linguistic difference; the deeper reason lies in the triple differences between East and West at the technological level (model architectures, tokenization algorithms, training data), systemic level (API ecosystems, payment systems, data standards), and policy level (cross-border data flows, export controls, national security reviews), which prevent tokens from circulating freely as an idealized "global universal AI currency." This observation has significant implications for understanding the future landscape of the token market—in the future, "Chinese token zones," "English token zones," and other token ecosystems based on language and policy may emerge, and cross-zone calls will face efficiency losses and compliance costs.

I. Token Origins: The Quadruple Leap from Philosophical Concept to Economic Unit

The evolution of token's identity can be clearly divided into four stages.

The origin in philosophical linguistics. In 1906, Peirce, while studying semiotics, distinguished between "type" and "token." The former represents an abstract rule or form, while the latter is the specific instantiation of that rule in reality. This distinction later became an important foundation of structural linguistics.

The encoding turn in computer science. In the 1960s, when a programmer wrote the line of code "int x = 5;", the computer needed to break it down into independent units such as "int," "x," "=", "5," etc. Each such unit was a token. Tokens thus became the smallest unit of information for a machine to "read" instructions.

The value carrier experiment of blockchain. In 2017, with the rise of the ICO craze, Token was given a new identity as a "negotiable digital equity certificate." Although that wave of frenzy gradually cooled, the understanding of "Token as a value symbol" had taken root.

The economic infrastructure of the AI age. March 2026 was a landmark moment—NVIDIA CEO Jensen Huang redefined data centers as "factories producing intelligent AI Tokens," and Liu Liehong, Director of China's National Data Administration, gave the official Chinese translation "ciyuan" (词元), calling it the "value anchor of the intelligent age." At this point, tokens completed their final leap from a technical concept to an economic concept.

II. The Technical Base: BPE Algorithm and the Implicit "Language Tax"

The reason tokens have been able to become the standard unit of measurement for AI relies on a technology that was forgotten for more than two decades—byte pair encoding.

In 1994, American programmer Philip Gage published an article on data compression algorithms in a C language technical magazine, introducing the basic principle of BPE: repeatedly scanning text and "welding" the two most frequently adjacent characters into a new symbol, compressing iteratively round by round. This paper did not attract attention at the time because its compression efficiency was not outstanding.

It was not until 2016 that Rico Sennrich, a researcher at the University of Edinburgh, while studying the word segmentation problem in machine translation, Accidentally retrieved this old paper and keenly realized that BPE was the perfect solution to the tokenization problem—there was no need to predefine a dictionary; the data was allowed to "speak" for itself, with high-frequency combinations naturally coalescing into tokens. In 2019, OpenAI officially adopted this approach when releasing GPT-2, setting the starting point for tokenization directly at the "byte" level, enabling the model to theoretically handle any written language.

However, the "frequency-first" logic of the BPE algorithm inadvertently formed an implicit "language tax" system. Because English is the absolute mainstream of internet corpora, the model's efficiency in segmenting English tokens is far higher than for other languages. To express the same meaning, English uses the fewest tokens, Chinese typically requires 1.5 to 2 times as many, and for languages with even fewer resources, the cost can be 5 to 10 times that of English. This means that

under pay-per-token pricing, the actual cost of conversing with AI varies enormously depending on the language used.

This "starting line" unfairness, once written into a model's initial vocabulary, is very difficult to correct—the tokenization rules are the "foundation" of AI's understanding of the world; the higher the building, the harder it is to replace the foundation. Fortunately, this situation is improving. Taking the GPT series as an example, the same sentence in Chinese required 38 tokens in GPT-3, dropped to 26 in GPT-4, and further to 15 in GPT-5—an efficiency improvement of over 60%. Moreover, domestic Chinese large models such as Tongyi Qianwen and DeepSeek have incorporated high-frequency Chinese phrases, idioms, and the like as native tokens into their vocabularies from the design stage, achieving a more "native" processing efficiency for Chinese. This reveals a deep underlying principle: whoever masters the "right to semantic segmentation" largely holds the efficiency and cost advantage for expressing that language in the digital world—this is a form of "fundamental minting right" in the digital age.

III. Uses and Output: The Exponential Leap from Conversation to Agents

The scenarios for token consumption are undergoing a three-level jump, from "conversation" to "reasoning" to "agents."

In the conversational AI stage, a single prompt and response consumes about 10,000 tokens. In the reasoning AI stage, involving multiple rounds of interaction and complex inference, each task consumes about 100,000 tokens. Entering the agentic AI stage, agents can autonomously perform multi-step tasks and invoke multiple tools, with token consumption per task reaching approximately 1 million—a nearly 100-fold increase over just two generations of evolution. Qualcomm CEO Amon predicts that by 2030, global token demand every 10 seconds will explode from 31.7 billion in 2026 to 1.27 trillion. This is no exaggeration; the explosive popularity of open-source agent tools like OpenClaw has proven this point: a single complex task consuming tens of millions of tokens is common.

On the output side, China has become the world's largest token supplier. Public data shows that as of March 2026, China's daily token call volume had exceeded 140 trillion, a more than thousand-fold increase in two years, accounting for 61% of global AI token usage. The establishment of this position relies on the superposition of three advantages: first, China's industrial electricity prices are far lower than those in Europe and America, and electricity accounts for over 60% of token production costs; second, China has built the world's largest intelligent computing cluster, achieving efficient output through the "East Data, West Computing"布局; third, domestic large models are fully open-sourced, and the API commercialization service system is mature, forming a complete industrial chain 闭环 from "model training - computing power support - inference output - global invocation."

IV. Institutions and Markets: The Tripartite Industrial Landscape

A clearly layered market structure has formed around tokens.

In understanding this industrial structure, it is also necessary to see the underlying support system behind tokens. Many people focus on token prices but overlook that tokens are not an independent resource. Behind tokens are models, and behind models is computing power. In a sense, computing power determines model capability, models determine token output, and tokens are the unit of measurement formed when model capabilities are exported to the market. Therefore, on the surface, the token market is about price competition, but at its core, it remains competition in computing power and efficiency.

The current global token industry chain can be roughly divided into three levels. The uppermost level is the computing power infrastructure, including chip manufacturers, server manufacturers, cloud computing platforms, data centers, and energy systems. The middle level consists of foundational model institutions and inference platforms, responsible for converting computing power into callable model capabilities. The lower level consists of token service providers, agent platforms, and various industry application institutions, which further convert model capabilities into enterprise services and end products. What users ultimately see is the token price, but what truly affects token cost and speed is often changes in the upstream computing power system and model efficiency.

In the past few years, as chip performance has continued to improve, model architectures have been optimized, and inference efficiency has increased, the cost per token has shown a rapid downward trend. Many people see tokens becoming cheaper, but the real driver behind this change is the continuous progress of the computing power system. Therefore, future competition in the token market will

involve not only platform and price competition but also competition in computing power resources, model efficiency, and resource integration capabilities.

Upstream "power plants" – computing power producers. Chip companies represented by NVIDIA are continuously iterating hardware to improve token output efficiency per unit of computing power. NVIDIA predicts its revenue will reach at least \$1 trillion by 2027, a confidence underpinned by the global thirst for AI computing power.

Midstream "grid dispatch" – model platforms and aggregators. Companies such as OpenAI, Google, DeepSeek, and Alibaba convert underlying computing power into standardized, callable token services. Three mature business models have emerged around this link: token aggregation platforms (represented by OpenRouter, operating on price differences and commissions, with weekly call volume reaching 26.9 trillion tokens as of May 2026), cloud provider MaaS services, and AI relay stations active in the secondary market.

Downstream "power entry points" – application and distribution platforms. Tencent opening its WeChat interface to access new model capabilities essentially turns WeChat into a "token distribution platform"; China Telecom has proposed "reshaping its business through token management," attempting to 跳出 the traditional traffic logic. What these entry-point companies have in common is that users no longer face the model directly but call tokens indirectly through the platform—the model itself is becoming "invisible."

It is worth mentioning that a unique paradigm of global trade has formed around tokens: intangible output, tax-exempt operation, and global coverage. This is transforming China from a traditional "world factory" into a "global provider of AI computing power infrastructure."

When discussing the token economy, there is another fundamental factor that cannot be circumvented: computing power. Many people focus on token prices but overlook the production process behind tokens. In fact, tokens are not 凭空 generated. Each token generation requires the model to perform inference, and model inference is

built on a foundation of massive computing power. In a sense, computing power is the source of the token economy, models are the production tools for tokens, and tokens are the unit of measurement formed when model capabilities are exported outward.

Therefore, computing power levels directly affect token costs and token generation speed. The stronger the computing power, the higher the model's ability to process information, and the more tokens it can produce per unit time; the lower the cost of computing power, the lower token prices tend to be. In the past few years, the large model industry has witnessed a continuous decline in token prices, driven not only by model algorithm optimization but also by chip performance improvements, data center scaling, and increased inference efficiency. In other words, many people see tokens becoming cheaper, but what truly drives this change is the continuous advancement of the computing power system.

From an industry chain perspective, computing power, models, and tokens actually form a complete chain. Upstream are chips, servers, data centers, and energy systems; midstream are foundational models and inference platforms; downstream are token services, agent applications, and various industry solutions. Computing power determines the upper limit of model capability; models determine the quality and efficiency of tokens; and tokens ultimately determine how much AI service users can obtain.

However, I believe that while the importance of computing power is undeniable, computing power itself does not automatically create value. Many industries have experienced similar processes in history. Owning a larger power plant does not mean owning a better enterprise; owning more servers does not mean owning a better product. Similarly, owning more computing power does not mean owning a higher-value system. Computing power solves the problem of production capacity, while systems solve the problem of value conversion. In the future, as computing power continues to grow, model capabilities continue to strengthen, and token resources continue to proliferate, what will truly be scarce will still not be computing

power itself, but how to organize these resources into systems that can operate long-term in the real world.

Therefore, from a future competition perspective, I believe that competition in computing power, models, and tokens will all gradually shift towards competition in systems. Because when computing power becomes more and more 普及, models become more and more similar, and tokens become more and more abundant, what will determine the difference will no longer be the quantity of resources owned, but who can use these resources to build long-term effective structures.

V. Competition and Barriers: Price War Involution and Value Stratification

Competition in the token market is exhibiting highly involutioned characteristics. From 2022 to 2026, the cost per single token inference decreased by 99.9%. Domestic models show even more significant price advantages: DeepSeek V4-Pro's input price is as low as 0.025 RMB per million tokens, Tongyi Qianwen Qwen3.5-Plus output price is 1.6 RMB per million tokens, compared to OpenAI GPT-4o's output price of approximately 72 RMB.

This price war has sharply compressed the profit margins of "middlemen." Industry statistics show that the number of various API aggregators and relay platforms has exceeded two thousand, with the market forming three major tiers: the top tier of large platforms with strong technology and comprehensive models; the middle tier of professional agents with official discounts and customer resources; and the vast number of micro and small agents facing severe survival 压力. The business model of simply profiting from token price differences is no longer sustainable.

The real barriers are shifting in two directions. One is the technical premium direction—by developing proprietary inference acceleration engines, one can sell cheaper than the official provider and still make a profit. The second is the value-

added service direction—providing customized services such as prompt engineering, model fine-tuning, and private deployment. More notably, the value of tokens is exhibiting enormous scenario-based 分化: average price for chat-related tokens is as low as \$0.01 per million, while for drug discovery-related tokens, the average price can reach \$1,000 per million, a cross-scenario value difference of a hundred thousand times. This reveals an essential principle: the value of a token is not determined by its production cost, but by its use—the same token, when used in different scenarios and embedded in different systems, generates vastly different value.

But a more important judgment needs to be made here: Currently and in the future, the barrier to selling tokens is not high. Anyone can register for an API, set up an aggregation platform, and start reselling tokens. What is truly scarce is what lies beyond selling tokens—real historical data, the ability to build systems and architectures, and a future-oriented vision. Without these, all large models and tokens are of no use.

Supplementary judgment on token profitability: Making money from tokens is already difficult, and it will become increasingly so. Price wars are continually compressing profit margins, and the business model of simply reselling tokens is rapidly 走向 meager or even no profits. Only when tokens are attached to a front-end system—that is, when a system provides services—can tokens and computing power serve as effective supplementary assets to generate profits. In other words, tokens themselves are not the source of profit; the system is. Tokens are merely the system's fuel, not its engine.

As tokens gradually become an important unit of measurement in the age of artificial intelligence, an increasing number of business models centered around tokens have emerged in the market. Besides foundational model vendors, there are also token agents, model aggregation platforms, agent platforms, enterprise-grade API distribution services, and various middle-layer service providers. For many entrepreneurs, token agency seems like a shortcut into the AI industry because it requires no model training or construction of large computing centers; one only

needs to integrate existing resources to start doing business. However, this model has both advantages and obvious limitations. Its advantages include relatively low initial investment, relatively low barriers to entry, and faster market expansion; its limitations include high dependence on upstream model suppliers, susceptibility to price competition, and a lack of long-term technical barriers. As token prices continue to fall, the room for profit from pure token resale may be continuously compressed. Therefore, in the long run, token agency is more of a distribution capability than an ultimate core competitiveness. What truly determines a company's long-term value remains the industry-specific systems, automation processes, knowledge systems, and real-world solutions built around tokens.

VI. Current Difficulties and Future Pathways

The current token market faces several deep-seated problems.

The trap of homogeneous competition. More than two thousand platforms are competing for the same set of customers, and price wars continue to drag down industry profits. Without change, this may repeat the early days of cloud computing, "losing money to gain market share."

The looming threat of compliance risks. The current global "regulatory dividend" of no specific legislation is a temporary state, not a long-term norm. As the token market expands, issues such as cross-border data, national security, and tax collection will force countries to accelerate legislation. Particular attention should be paid to the fact that some market players' attempts to tokenize and financialize tokens could easily trigger global financial regulation, not only leading to the shutdown of such businesses but also damaging the reputation of China's entire token export industry.

The "Token-maxxing" myth. A new form of "showing off wealth" has emerged in Silicon Valley—competing over who consumes more tokens daily. Meta has an internal leaderboard called "Claudeconomics," compiling AI usage data from over 85,000 employees, with the top employee burning token values as high as several million dollars. Critics pointedly say: "Outcome maxxing is better than token maxxing"—rather than 疯狂 consuming tokens, it's better to see what outcomes are produced. This is exactly the core issue I have deeply understood through years of practice.

Chinese Tokens vs. English Tokens: Three Barriers Under a Unified Concept

For Chinese users, a fact that must be clearly recognized is this: although Chinese tokens (ciyuan) and English tokens belong to the same foundational unit of measurement in the AI age, they have not formed a truly unified, freely interoperable global market. From a technical concept perspective, both East and West use tokens as the basic unit for models to process language and data; however, from a practical application perspective, the combined existence of technological barriers, systemic barriers, and policy barriers has created a de facto separation between Chinese tokens and English tokens.

The first barrier is the technological barrier, namely differences in tokenization logic and model architecture.

Western mainstream models such as GPT, Claude, and Gemini are mostly built on large-scale corpora primarily composed of English. Although these models already possess good Chinese processing capabilities, in many scenarios, Chinese still tends to require more tokens to express the same amount of information as English. More importantly, different models have different vocabulary structures and tokenization strategies. The same piece of Chinese content may be segmented into completely different numbers of tokens across different models, and their internal encoding methods and semantic mapping methods may also differ.

At the same time, Chinese domestic models such as DeepSeek, Tongyi Qianwen, Doubao, and Wenxin Yiyan have incorporated large amounts of Chinese corpora from the training stage, giving them higher native adaptability to Chinese phrases, idioms, proper nouns, and localized expressions. Therefore, even facing the same text, the number of tokens, processing efficiency, and understanding pathways across different models can vary significantly. In this sense, while token is a unified

concept, there is no "standard token" that can be directly circulated and reused across models.

The second barrier is the systemic barrier, namely differences in ecosystems and invocation methods.

The current global token market has in fact formed multiple relatively independent ecosystems. International platforms like OpenAI, Anthropic, and Google have their own API standards, pricing models, access rules, and service systems; the Chinese market has formed a domestic ecosystem centered on platforms like DeepSeek, Tongyi Qianwen, Doubao, and Kimi.

For ordinary users, purchasing tokens on one platform does not mean they can directly use them on another platform. The invocation interfaces, authentication mechanisms, rate limits, service agreements, and business rules across different platforms all differ. Even if conversion is done through third-party gateways or aggregation platforms, it is essentially recalling another system, not a true token interchange. Therefore, while token is a unified concept, tokens as a commercial resource and service capability remain locked within their respective platform ecosystems.

The third barrier is the policy barrier, namely differences in data flows and compliance requirements.

As artificial intelligence gradually enters the enterprise application stage, the importance of data security issues continues to rise. China's data security system, the EU's GDPR system, and US regulatory measures on the cross-border flow of sensitive data all impose different requirements on data transmission and processing. Much data involving customer information, financial data, medical records, trade secrets, and personal privacy often cannot freely cross borders.

This means that in real business environments, many enterprises do not choose models based solely on token prices but must prioritize data compliance, data storage location, and legal liability. For much business-critical real data, compliance

costs are often far higher than the token cost itself. Therefore, in the real world, domestic businesses tend to use compliant domestic models, while overseas businesses tend to use overseas models, resulting in de facto token market segmentation.

Technological barriers determine differences in token segmentation methods; systemic barriers determine differences in invocation systems and business ecosystems; policy barriers determine differences in data flows and compliance boundaries. These three barriers acting together mean that while Chinese tokens and English tokens belong to the same technical concept, they are difficult to be freely exchanged, freely circulated, and freely interfaced like an idealized "global unified AI currency."

In the future, a situation similar to international trade is likely to emerge: the Chinese token ecosystem and the English token ecosystem will coexist for a long time, each following independent development paths. Cross-zone invocation is not impossible, but it often requires additional technical adaptation, system conversion, and legal review costs. For most ordinary users, such cross-zone capability is not a daily need, so the two major systems will likely maintain relatively independent development for a long time.

Tokens produced in China are not necessarily confined to use within Chinese systems, but in key scenarios involving data sovereignty, compliance responsibility, and industrial security, a large portion of future token demand is likely to be preferentially consumed within domestic models, domestic platforms, and domestic systems.

From a long-term trend perspective, the coupling between the location of computing power production, the model provider, the system operator, and the data location is increasing. For finance, government, healthcare, supply chains, and large enterprise systems, data often cannot freely flow across borders. Therefore, even if China can export large quantities of token services globally, much high-value token demand will still preferentially stay within Chinese models and Chinese systems for completion.

This does not mean Chinese tokens cannot be exported, but rather that the future token market may simultaneously exhibit two structures: a "global circulation" part and a "domestic closed-loop" part.

It is worth noting that I believe what may truly achieve unification in the future is not the tokens themselves, but rather the way tokens are invoked. As agent platforms, unified interface protocols, model aggregation platforms, and automation systems continue to develop, users will ultimately focus not on which token is used, but on whether the task is completed, whether the cost is reasonable, and whether the efficiency is sufficient. At that time, tokens, like electricity, internet traffic, and cloud computing resources today, will gradually become hidden behind systems, becoming the fundamental operating infrastructure of digital civilization.

Therefore, from a long-term perspective, although Eastern and Western token markets are temporarily difficult to directly interconnect, their development directions are highly aligned. The past of tokens belongs to computer language processing, the present of tokens belongs to the AI industry, and the future of tokens likely belongs to the fundamental operational system of the entire digital civilization.

VII. Why System-Building Capability Cannot Be Directly Obtained Through Large Quantities of Tokens

In arguing that "system-building capability is a scarce asset," I need to directly confront a possible 质疑: since AI is so powerful, why can't system-building capability be "generated" or "obtained" by consuming large quantities of tokens? The answer to this question reveals the fundamental boundaries of token capability.

First, system-building requires decisions, and decisions are not generation.

AI can generate an article, a piece of code, a website template, but AI cannot replace the core decisions in system design that humans must make—choosing which modules, abandoning which functions, balancing current needs with future expansion, when to 重构, when to maintain the status quo. These decisions depend on understanding the essence of the business, judging resource constraints, and foreseeing long-term consequences. Tokens can generate countless "possible solutions," but choosing "which solution is worth implementing" and "why" is something tokens cannot replace. My intelligent logistics system has undergone countless adjustments over two decades, and each adjustment was not because an AI told me "this is better," but because a real problem emerged in the actual business, exposing limitations and raising new demands. Behind these decisions is judgment, and judgment comes from practice, not from tokens.

Second, system-building requires handling exceptions, and exceptions cannot be exhaustively enumerated.

The reason a system can operate long-term is not because its normal 流程 is perfectly designed, but because it can handle exceptions. In a logistics system, exceptions include: incorrect customer addresses, damaged goods, customs delays, payment failures, system outages, data loss. In a publishing system, exceptions include: fluctuation in translation quality, incompatible metadata formats, changes to

archival interfaces, DOI registration failures. These exceptions cannot be exhaustively enumerated because the variety of exceptions in the real world is infinite. AI can learn "common exception patterns" through large quantities of tokens, but it cannot predict "the next exception that has never appeared before." Only a long-running system, continuously encountering, recording, and resolving exceptions in practice, can develop true exception-handling capability. This capability cannot be obtained by "feeding more tokens," because exceptions are not knowledge; they are experience.

Third, system-building requires trade-offs, and trade-offs have no standard answers.

The construction of every system involves trade-offs under multiple constraints: cost vs. performance, speed vs. stability, generality vs. customization, immediate needs vs. future expansion. These trade-offs have no "correct answer," only "reasonable choices under specific circumstances." AI can provide an "optimal solution," but this "optimal" is statistically optimal based on historical data, not a judgment based on "your specific context." The reason my 跨代 two-decade integrated logistics and finance system has operated to this day is precisely because I made different trade-offs in different eras—choosing a certain technology stack in 2005, choosing another integration method in 2015, and choosing an upgrade path in 2025. These trade-offs are not "right" or "wrong," but "suitable for the context at the time." Tokens cannot make these trade-offs for me, because the essence of trade-offs is value judgment, not calculation.

Fourth, system-building requires feedback loops, and feedback requires the real world.

The improvement of system-building capability depends on a closed loop: design → implementation → operation → problem detection → correction → redesign. The "problem detection" step in this loop must come from real-world feedback. AI can simulate operation, but it cannot simulate all unpredictable factors in the real world. The reason my logistics system has continuously improved is that real customers provided real feedback; the reason my publishing system has continuously improved

is that real archival interfaces experienced real changes. This feedback cannot be obtained by consuming tokens in a closed environment. Tokens can generate "hypothetical customer feedback," but they cannot replace the fact that "the customer actually gave this feedback." Real-world feedback is the sole soil for the growth of system-building capability.

Fifth, system-building requires trust, and trust can only come from time.

Whether a system is trustworthy is not proven by promotion, but by time. The reason my website is permanently archived by CERN Zenodo, WorldCat, and the National Library of Australia is not because they "like me," but because these institutions, through 长期 observation, have confirmed that my system is stable, reliable, and persistent. This trust cannot be purchased with large quantities of tokens. AI can generate a "very professional-looking" system manual, but it cannot generate the fact that "this system has been trusted for two decades." Trust is a product of time, not a product of tokens.

To summarize, the reason system-building capability cannot be directly obtained through large quantities of tokens is that it is not essentially a generative capability, but rather a capability for judgment, exception handling, trade-offs, feedback absorption, and trust accumulation. What these capabilities share in common is that they require real-world time, practice, errors, corrections, and validation. Tokens can accelerate information processing, but they cannot accelerate real-world experience accumulation. This is the fundamental reason why, even as tokens become extremely cheap in the future, system-building capability will remain scarce.

VIII. My Assets: Real Historical Data, Cross-Domain Systems, Building Capability, Foresight Vision, and Self-Verification Capability

Before further discussing "what is truly scarce," I need to first explain my own position. Because my judgments are not derived from theoretical deduction, but from long-term practice. This is very evident in the various systems I have built over a long period. Whether it is the intelligent logistics system, intelligent invoice system, automated financial processing system, or later the ten-language publishing system, metadata system, website navigation system, and automatic cross-referencing system, the truly difficult part has never been generating information, but organizing it. Information can grow indefinitely, but without structure, it will only create new chaos. The same is true in the token era. In the future, AI can generate vast amounts of content, but vast amounts of content do not automatically equal vast amounts of value. On the contrary, as content generation becomes easier, how to filter, organize, preserve, index, and 关联 this content will become increasingly important.

I possess the following five assets, which constitute the empirical basis for my judgment of "what is scarce Beyond Tokens."

First, real historical data. This is not just any "historical data," but twenty years of 跨代 real business data—from the 2005 intelligent invoice system to the present, from actual 运行 logistics systems, financial systems, and tax systems, not from experimental environments or scraped data. In the AI era, with the proliferation of "synthetic data," the value of real historical data is rising sharply. Most companies possess "single-domain historical data," whereas I possess cross-domain, two-decade, practically validated historical data.

Second, completion of multiple cross-domain systems and architectures. My systems cover the following domains: intelligent logistics system, intelligent invoice

system, intelligent financial system, intelligent tax system, 跨代 two-decade logistics and financial integration system, website automation system, metadata generation system, *The Epochal Transition* ten-language publishing system, DOI management system, TROVE and WorldCat link organization system, automatic cross-referencing system, Sitemap generation system, Zenodo profile generation system, cross-platform publishing system. The uniqueness of these systems lies in: cross-domain breadth (from logistics to publishing, from finance to literature indexing), temporal depth (some systems have been running for two decades), and completion by one person (demonstrating the transferability of the method). Among these, *The Epochal Transition* e-monthly (<https://times.net.au>) serves as public 实证—it holds the ISSN international standard serial number (3083-5178), is permanently archived by WorldCat (the world's largest library union catalog system, with member institutions exceeding 16,000 national and university libraries across over 100 countries and regions), the National Library of Australia's PANDORA/TROVE, as well as CERN (the European Organization for Nuclear Research)-operated Zenodo, OpenAIRE (the European Open Science Infrastructure), DataCite DOI (global digital object identifier, with over 60 documents), and ORCID (the global unique permanent academic digital identifier). This website demonstrates that a single person (an independent scholar) can complete the entire process from content creation to interfacing with top-tier international archival infrastructure.

Third, system and architecture building capability. This is not the ability to "use AI tools," but the design capability to transform vague requirements into stable structures, the architectural capability to integrate different systems into a unified architecture, the engineering capability to keep systems running for years without human intervention, and the meta-ability to transfer experience from logistics to publishing, from finance to literature. As discussed earlier, this capability cannot be directly obtained through large quantities of tokens because it depends on decision-making, exception handling, trade-offs, feedback loops, and trust accumulation—none of which can be replaced by AI. The current AI training market focuses on surface skills such as "prompt engineering," "AI writing," and "AI drawing"; almost no courses teach "cross-domain system-building capability"—because this capability

cannot be shortcut; it can only be accumulated through long-term practice. The over 40 independent creative works listed on the website, the obtained ISSN, and the over 60 DOIs are the best public records of the system's long-running capability, not my self-proclamation.

Fourth, future-oriented vision. Starting to experiment with a remote work prototype system in 1993---nearly thirty years before large-scale remote work became widespread; designing an intelligent invoice system in 2005---starting process automation twenty years before AI became widespread; cross-generational two-decade system integration---foreseeing at the time of system design that it would still need to operate twenty years later; now proposing the judgment of "Beyond Tokens"—at a time when most people are still chasing tokens, already foreseeing that system capability is the true barrier. I also foresaw that "tokens are similar to currencies, and the technological, systemic, and policy differences between East and West render them non-universal"—just as the US dollar and Chinese renminbi cannot be freely exchanged, Chinese tokens and English tokens face similar circulation barriers, and in the future, "Chinese token zones," "English token zones," and other token ecosystems based on language and policy may emerge. I also foresaw that "making money from tokens is already difficult and will become increasingly so; only when there is a front-end system providing services can tokens and computing power serve as effective supplementary assets to generate profits." As an independent scholar, I established in advance a personal publishing system that interfaces with future academic infrastructure, foreseeing the trend that "an individual can also become a node in the global knowledge infrastructure." Most trend forecasts come from analysts or entrepreneurs, whereas my "foresight capability" has the validation of two decades of practice as its endorsement.

Fifth, self-verification capability. This is one of my most important meta-capabilities—all the literature I write in various fields achieves theoretical closure and logical self-consistency. This means my arguments do not rely on external authority or citations, but on the self-sufficiency of internal logic. The default rule of academic literature is to "cite previous research"; few people claim or demonstrate "self-verification

capability"; academic discussion focuses on "whether the argument is sufficient," not on "whether the author has the ability to construct a self-consistent 闭环." In my articles, this self-verification capability is manifested in: arguments come from practice, practice validates the arguments, and the arguments and practice form a 闭环. I am not "citing others' theories to prove myself," but rather "showing the systems I have built and letting the systems speak for themselves." This self-verification capability is a concept entirely absent from public literature and one of my most unique meta-capabilities.

Combining these five assets, their scarcity is not additive, but multiplicative: real historical data × cross-domain systems × system-building capability × future-oriented vision × self-verification capability constitutes an irreproducible asset portfolio. In public discussions, no comparable case can currently be found.

The process of compiling "The Road of an International Independent Scholar" has made me realize this even more profoundly. From 1989 to 2026, more than thirty years of experience spans multiple fields. These experiences did not arise all at once but gradually formed over a long period. AI can generate an article in seconds, but it cannot generate thirty-plus years of real experience in seconds; it can generate a large number of opinions, but it cannot directly generate the judgment accumulated through long-term practice; it can simulate many processes, but it cannot replace the repeated validation of the real world. In this sense, time itself is also an important asset, and the structure precipitated by time is an even rarer asset.

IX. Beyond Tokens: Five Truly Scarce Capabilities

When tokens themselves are no longer scarce—the barrier to selling tokens is not high, but there will be an extreme lack of real historical data, the ability to build systems and architectures, and future-oriented vision—what will be truly scarce? Based on my long-term practice from the 1993 remote work prototype system, the 2005 intelligent invoice system, intelligent logistics system, intelligent financial system, intelligent tax system, the cross-generational two-decade logistics and

financial integration system, to website automation, metadata generation, *The Epochal Transition* ten-language monthly, DOI management, TROVE and WorldCat link organization, automatic cross-referencing, Sitemap generation, Zenodo profile generation, cross-platform publishing, I believe the following five capabilities will become true barriers.

First, structural design capability. Tokens are raw material, and structure is the framework that organizes raw material into valuable products. Even before AI became widespread, I had already started thinking about how to use limited resources to achieve higher operational efficiency. My intelligent logistics system, intelligent invoice system, and 跨代 two-decade intelligent logistics and financial system were born in different eras, facing different problems, yet they share a common characteristic: the goal was not to create more information, but to reduce chaos, reduce repetitive labor, improve overall efficiency, and establish structures that can operate long-term. In the automated publishing system, the core problem I faced was not "whether text can be generated," but "how to define metadata fields for ten languages, how to establish associations between articles and categories, and how to design version management rules." These structural design decisions determine whether the system can continue running for years without collapsing. In contrast, the quantity of tokens consumed is merely a byproduct. The The Epochal Transition website itself serves as public evidence of structural design capability: a personally maintained academic website structure with an ISSN international standard serial number, multilingual support, multiple archival channels, proving that I am not engaging in empty talk, but have already successfully built a complex, publicly accessible system using this structural design capability.

Second, automation capability. Tokens can batch-produce content, but only automation systems can batch-complete end-to-end workflows. In the intelligent logistics system, the real value is not in how many times AI is called to plan routes, but in establishing a fully automated closed loop from order intake, warehouse scheduling, route optimization, to exception handling. In this closed loop, AI is just one component, and automation capability is manifested in the connections between

components. A person without a system, even with a large number of tokens, may only continuously generate scattered words, scattered images, scattered tables, and scattered code; a person with a system, on the other hand, can 接口 tokens into their workflow, allowing them to automatically organize materials, generate metadata, check links, update webpages, assist publishing, archive literature, build indexes, process logistics, analyze finances, and generate reports.

Third, cross-domain practical experience. Tokens are available to everyone, but the ability to transfer experience from one industry to another is not universal. The reason I can build both a logistics system and a publishing system is not because I have more tokens, but because I have accumulated transferable methodologies such as metadata management, workflow design, and exception handling. My publishing system needs tokens because ten-language translation, abstract generation, keyword organization, metadata generation, and link descriptions all require a large amount of language processing. My website system needs tokens because page structures, navigation descriptions, SEO descriptions, academic metadata, and multilingual introductions can all be automatically generated. My logistics system can also use tokens because customer emails, goods descriptions, customs documentation, exception explanations, financial notes, and operation logs can all be AI-assisted. My automatic cross-referencing system needs tokens even more because it must find thematic associations between large numbers of articles, titles, DOIs, NLA, TROVE, and WorldCat links. This cross-domain practical accumulation cannot be obtained by consuming tokens; it can only be precipitated by solving real problems. The website simultaneously presents information from multiple domains, including academic publishing, infrastructure integration, technical architecture, and personal brand, proving that I have indeed connected publishing, library science, technology, and personal branding, rather than possessing a single skill.

Fourth, the mental model for long-term system operation. The mainstream narrative for current AI applications is "conversation" and "generation," but what truly produces sustained value is "operation." A conversation generates 10,000 tokens, an agent

task consumes a million tokens, but a logistics system that has run continuously for a decade may consume billions of tokens—yet what is truly scarce is the mental model for designing this system: how to ensure the system remains usable after model upgrades? How to handle changes to upstream APIs? How to monitor exceptions and recover automatically? These issues have nothing to do with tokens, yet they determine the ultimate output of tokens. The value of these systems does not come from the tokens themselves, but from the structures I have already built or am developing. Tokens are merely fuel; the system is the engine. Tokens are merely electric current; structure is the circuit. Tokens are merely water; the pipes, valves, reservoirs, and dispatch systems determine whether it can play a role. Without a system, tokens will dissipate; with a system, tokens can be converted into long-term value. The website displays over 40 independent creative works and permanent archiving by authoritative institutions such as CERN and the National Library of Australia, proving that my system is not pursuing short-term hype, but long-term value, and has already received "permanent preservation" endorsement from authoritative institutions.

However, when I further consider the relationship between tokens and systems, I find another often overlooked issue. Many people believe that as long as they have an excellent system, they will be able to gain an advantage in the future token era. But the reality is not that simple. Systems themselves are not something that just anyone can design. Truly mature systems often require long-term practice, repeated revision, and real-world validation. Many systems may take years or even decades of continuous accumulation from initial concept to actual implementation. Therefore, system design capability itself is already a scarce resource.

But even possessing a mature system does not automatically lead to large-scale impact. Between the technical feasibility of a system and its societal 普及 lie multiple layers including markets, capital, distribution channels, user education, and platform ecosystems. History is not short of good technologies or good systems; what is truly scarce is the platform capability to push a system to large-scale application. Even if a system is far superior to existing solutions in a local scenario, without sufficient

capital support, platform promotion capability, and user base, its actual usage scale may still be very limited.

From the perspective of the token economy, this is particularly evident. Token demand does not come directly from the system itself, but from the actual scale of the system's use. A platform serving millions of users daily, even with relatively ordinary system design, may generate enormous token consumption; conversely, an advanced system with very limited users may have very low token demand.

Therefore, the future development of the token market depends not only on model capability and system capability, but also on platform capability and resource integration capability.

This means that future competition has at least three levels. The first level is model competition, determining who can supply tokens. The second level is system competition, determining who can effectively utilize tokens. The third level is platform competition, determining how many users a system can reach and how much actual demand it can generate. Models, systems, and platforms are all indispensable. Without models, systems cannot operate. Without systems, tokens cannot create value. Without platforms, even the best system may remain in small-scale use.

From this perspective, I am even more convinced that tokens themselves are merely foundational resources. What truly determines the future is not only those who own tokens, nor only those who own systems, but those who can connect models, systems, platforms, and real-world needs. Only when these elements form a closed loop will tokens transform from mere computational units into value-creating productive forces.

Fifth, the ability to embed tokens into real-world business processes. Tokens must ultimately serve some link in the real world—logistics, publishing, education, healthcare. What is truly scarce are those who understand both the boundaries of AI capability and the pain points of a specific industry. They know when to use AI and when not to; they know how to convert token outputs into business decisions; they know how to measure the real relationship between token input and business output.

This composite capability is currently the scarcest resource in the market. The website itself is a running "business process": content creation → metadata generation → multi-platform publishing → international archiving → reader access. This proves that my AI application (token consumption) has been embedded into a complete, closed-loop publishing process that generates real-world impact.

I increasingly feel that the real competition in the future is likely not between models, nor between tokens, but between systems. Models will continue to upgrade, token prices will continue to fall, and today's expensive resources may become very 普及 in the future. But systems will not automatically appear just because resources are abundant. Systems require long-term accumulation, continuous revision, and real-world validation. Whoever can convert tokens into long-running systems is more likely to obtain sustained value. And this is also the key to profitability: making money from tokens is already difficult, and it will become increasingly so; only when there is a front-end system providing services can tokens and computing power serve as effective supplementary assets to generate profits.

I judge that the real competition in the token era will unfold in three stages: the first stage is resource competition, with everyone vying for models, computing power, APIs, and token prices; the second stage is application competition, with everyone using tokens for writing, customer service, code, translation, education, office work, and data analysis; the third stage is the true system competition, where whoever can embed tokens into long-running industry systems, publishing systems, logistics systems, knowledge systems, and civilization recording systems will likely gain long-term advantage.

In fact, when discussing tokens, models, systems, and platforms, there is another important factor that is often overlooked in the real world: time. Many people believe that with capital, they can quickly build systems; with a sufficiently large market, they can rapidly replicate success models. However, reality is often not so. Capital can purchase computing power, can procure tokens, can hire development teams, but it cannot directly purchase real data and mature systems that have been validated over many years. For many industries, what is truly scarce is not the models

themselves, nor the tokens themselves, but the systems that have been operating stably in real business environments for a long time.

From a business perspective, markets and capital typically will not wait for a completely new system to grow from scratch. A large system from concept proposal, architecture design, function development, process validation to actual implementation often takes many years or even longer. During this process, not only is continuous resource 投入 required, but also extensive trial and error, revision, and real-world validation. Therefore, when market opportunities have already emerged, capital tends to prefer systems that are already formed, already running, and already validated, rather than designing an entire system from scratch.

However, there is another 矛盾 in reality. The larger a system, the harder it is for later capital to easily 介入. After long-term development, large systems typically have already formed their own operating mechanisms, management structures, data logic, and organizational cultures. They may have vast historical data, complex business processes, and long-accumulated experience systems. Because of this, although large systems are enormously valuable, they are often 难以 to rapidly transform or thoroughly 重构 according to capital expectations. The larger the system, the greater its inertia tends to be.

However, there is another contradiction in reality. The larger a system, the harder it is for later capital to easily intervene. After long-term development, large systems typically have already formed their own operating mechanisms, management structures, data logic, and organizational cultures. They may have vast historical data, complex business processes, and long-accumulated experience systems. Because of this, although large systems are enormously valuable, they are often difficult to rapidly transform or thoroughly restructure according to capital expectations. The larger the system, the greater its inertia tends to be.

Therefore, in the real business world, the optimal state is neither designing a system from scratch nor simply acquiring a large, rigid system, but rather finding systems that have already completed their core structure construction, have been validated

over the long term, and can operate stably. Even if such systems are not perfect, they still have extremely high value because they have already crossed the most difficult stage, possessing real data, real processes, and real validation records. Based on this, repackaging, optimizing, opening to the outside world, and platformizing promotion are often more efficient than starting from scratch, and can also significantly reduce trial-and-error risk and time cost. Because such systems have already completed the most expensive, most time-consuming, and hardest-to-replicate stage.

From this perspective, the time window is also a resource that is easily overlooked. When an industry is still in its early stages, many people focus on model capability, token prices, and market size, ignoring the time required for system formation. The birth of a mature system typically goes through multiple stages: concept proposal, architecture design, process validation, real-world operation, and continuous optimization. The truly difficult part is not proposing the idea, but keeping the system running stably in a real-world environment over the long term. Therefore, when the market begins to recognize the value of mature systems, it often means that these systems have already undergone years of accumulation, rather than being something that can be replicated again in a short time.

For capital and the market, the most ideal targets are not ideas at the conceptual stage, nor prototypes that have not yet been validated, but rather the embryonic systems that have completed core structure construction, possess real operation records, and have room for upgrading. Although such systems may not be perfect, they have already crossed the most difficult, most time-consuming, and most risky stage. Once market demand grows rapidly, systems like these often become highly sought after, because the number of truly long-validated mature systems is typically much smaller than the market imagines. Models can iterate quickly, tokens can expand continuously, development teams can be assembled temporarily, but real data, real processes, and long-term validation records cannot be recreated in a short time.

Therefore, in the ongoing development of the token economy, time not only affects the formation of systems but also affects the value of systems. When the market has not yet fully recognized the importance of a certain type of system, that is often also the period when the cost of acquiring such systems is lowest; and when the market generally realizes their value, the available choices are often already very limited. For many industries, what is missed is not just a project, but a time window for entering the next stage of development that has already completed long-term accumulation.

From the perspective of the token economy, this is also the case. Tokens can be obtained quickly, models can be accessed quickly, platforms can be built quickly, but what is truly difficult to replicate quickly are the systems that have accumulated over many years and can continue to produce value. In the future, when tokens become increasingly abundant, the importance of systems will not only not decline, but may actually increase further. Because when everyone has similar token resources, what determines the outcome of competition will no longer be the quantity of tokens, but who possesses more mature real data, more stable business processes, and system structures closer to real-world needs.

Another often overlooked issue is the origin of systems. Many people are accustomed to understanding systems as software products, and therefore naturally assume that systems should be designed and developed by software companies. However, in many complex industries, the systems that truly operate stably over the long term are often not first designed by software companies and then given to the industry, but rather gradually formed in the course of long-term business practice, and then turned into software, automated, and platformized.

The reason is not complex. Software companies typically excel at technical implementation, but they do not necessarily stay within the specific business context for long. They can develop functions quickly, but they have difficulty accumulating a large amount of real business experience in a short time. In contrast, systems that develop organically within an industry are continuously revised and improved in the

face of customers, orders, exceptions, regulatory changes, cost pressures, and market competition on a daily basis. Such systems do not start from theory, but from real-world problems; each of their processes, each rule, and each automation link typically comes from the long-term validation of real work.

Therefore, in logistics, supply chains, publishing, finance, manufacturing, healthcare, and many other professional fields, the most valuable systems are often not the most complex systems, nor the systems with the most functions, but those that have been battle-tested over the long term and can stably solve real problems. They may not have the most advanced interfaces, the flashiest promotion, or the latest technical architecture, but they possess a large amount of real data, real processes, and long-accumulated business logic. This experience often cannot be obtained through short-term development. From the perspective of the token economy, this is particularly important. In the future, model capabilities may continue to improve, token costs may continue to fall, and software development efficiency will also become higher and higher. However, the business experience accumulated by an industry over decades will not automatically arise because of model upgrades. What is truly difficult to replicate is not the software code itself, but the industry knowledge, exception-handling experience, decision-making logic, and long-validated business processes hidden behind the system.

Therefore, I believe that the most ideal systems in the future will not be those designed completely detached from industry practice, but those that originate from real business, have been validated through long-term operation, and are then continuously upgraded with the help of artificial intelligence and token capability. Such systems possess both real-world reliability and the efficiency gains brought by the AI age, and their long-term value is often far higher than those systems built solely on technical concepts.

X. Where Do Tokens Flow: From Transient Conversations to Civilization's Memory

Regarding the future token economy, I have another observation. Many people focus on how tokens are produced, but few 关注 where tokens ultimately flow. If a token only remains in a single chat, its lifespan may be only a few minutes; if it enters an ordinary webpage, its lifespan may be only a few months; if it enters a long-term preservation system, its value changes fundamentally. When tokens enter DOIs, Zenodo, WorldCat, TROVE, national library long-term preservation systems, and various academic and knowledge infrastructures, they are no longer merely tokens; they begin to transform into knowledge records, literature assets, and civilization's memory. Therefore, I believe that the truly important question in the future is not how to generate more tokens, but how to enable tokens to enter structures where they can be long-term preserved and continue to exert their role.

It is for this reason that I remain optimistic about the future development of tokens, but I do not 迷信 tokens themselves. The past of tokens is the basic unit for computers to process language; the present of tokens is an important unit of measurement and commercial unit in the age of AI; the future of tokens is likely to become an important foundational resource for digital civilization. However, no matter how tokens develop, I always maintain that what is truly scarce is not tokens, but the people who can harness tokens; not those with more resources, but those who can establish long-term structures; not those who can generate more content, but those who can organize content into knowledge systems.

XI. The Incompressibility of Time: Four Unpurchasable Assets

In discussing "what is truly scarce Beyond Tokens," I need to further point out a fundamental issue that most discussions overlook: certain assets are inherently incapable of being accelerated, purchased, or replicated. No matter how powerful AI becomes or how cheap tokens become, the following four assets will always retain their inherent scarcity.

First, time is incompressible.

Thirty years of practice is simply thirty years; it cannot be compressed into three years, nor can it be simulated by AI. AI can generate an article on "how to build a logistics system" in seconds, but it cannot experience, in seconds, the various anomalies, failures, policy changes, and market fluctuations that a logistics system encounters over two decades. These real-world accumulations of time cannot be shortened by any technological means. My starting to experiment with a remote work prototype system in 1993, designing an intelligent invoice system in 2005—these points in time are not accidents; they are irreversible markers of real history. Anyone who comes later can read my articles and learn my methods, but they cannot possess the years I have lived through—those years themselves are assets.

Second, cross-domain experience cannot be purchased.

Cross-domain experience cannot be bought like a commodity. A person can buy a hundred books on logistics, publishing, finance, and literature indexing in a single day, but they cannot acquire, on that same day, the practical experience of connecting these domains. True cross-domain capability is not the accumulation of knowledge, but the ability to establish connections between different domains, identify commonalities, and transfer methods. This capability can only be acquired by solving real problems, making mistakes, correcting paths, and accumulating judgment in one domain after another. The reason I can transfer experience from logistics systems to publishing systems is not because I read a certain book, but

because I spent twenty years in logistics and another ten years in publishing—the intersection of these two periods of time is where the transferable methodology emerged.

Third, system-building capability is difficult to replicate.

System-building capability is not a set of code or processes that can be copied and pasted. Code can be copied, but the decision-making logic behind the code, the trade-off judgments, and the foresight of future changes cannot be replicated. The value of a system lies not only in what it can do today, but also in how many changes, how many failure recoveries, and how many requirement adjustments it has undergone over the past decade while continuing to operate. This "history of the system" itself is proof of building capability. My website (<https://times.net.au>) can be viewed by anyone, but what viewers see is the result, not the three decades of decision-making history, technology selection, structural adjustments, and continuous maintenance behind that result. This building capability cannot be obtained by plagiarism or imitation.

Fourth, long-running operation records cannot be shortcut.

The record that a system has run for two decades is itself inherently incapable of being shortcut. A long-running record proves the system's stability, maintainability, and design foresight. In the AI era, generating a "fake-looking" website or system may be completed in minutes, but generating a "system that has been running for two decades and is permanently archived by CERN, WorldCat, and the National Library of Australia" has no shortcuts. A long-running record is a timestamp, declaring to everyone that this system has withstood the test of the real world. Such a record cannot be forged or accelerated by any AI technology.

These four assets—time, cross-domain experience, system-building capability, and long-running operation records—together constitute the basis of what I have argued in this article as "truly scarce." I possess them not because I am smarter or more hardworking than others, but because I started earlier than others and have continued to the present. In an era where AI is making more and more things cheap

and easily accessible, these things that cannot be compressed, purchased, replicated, or shortcutted will become increasingly precious.

XII. Conclusion: From "Token Hoarding" to "System Building"

The future of tokens does not lie in tokens themselves.

With the continuous development of large model technology and the continuous decline of inference costs, token supply will continue to expand. Tokens that are expensive today will be as cheap and ubiquitous as internet traffic and hard drive storage in a few years. When everyone can easily obtain large quantities of tokens, what will be truly scarce will no longer be tokens, but the ability to convert tokens into real-world value—structural design capability, automation capability, industry experience, and long-term practical accumulation. And the deeper scarcity lies in those things that cannot be compressed, purchased, replicated, or shortcutted: time itself, cross-domain experience, system-building capability, and long-running operation records. As argued earlier, system-building capability cannot be directly obtained through large quantities of tokens because it is not essentially a generative capability, but rather a capability for judgment, exception handling, trade-offs, feedback absorption, and trust accumulation—the common characteristic of these capabilities is that they require real-world time, practice, errors, corrections, and validation.

On the profitability level, this judgment applies equally: making money from tokens is already difficult, and it will become increasingly so. Price wars are continually compressing profit margins, and the business model of simply reselling tokens is rapidly heading toward meager or even no profits. Only when tokens are attached to a front-end system—that is, when a system provides services—can tokens and computing power serve as effective supplementary assets to generate profits. Tokens themselves are not the source of profit; the system is. Tokens are merely the system's fuel, not its engine.

From this perspective, the systems I have already completed or am developing have special significance for the future development of tokens. Because these systems were not temporarily built to chase a 热点; they existed long before the concept of tokens became a public topic. The intelligent logistics system proved that real-world business can be structured, processized, and automated; the publishing system proved that an individual can also establish a multilingual, multi-platform, multi-DOI, multi-archive literature system; the website and metadata system proved that digital content can be continuously organized, indexed, and disseminated; the automatic cross-referencing system further proves that when the number of works continuously increases, what truly matters is not continuing to pile up content, but establishing structural relationships between content. The *The Epochal Transition* website (<https://times.net.au>), as a centralized display platform for these achievements, is itself empirical evidence of converting token capability into an open, permanent, verifiable academic system.

Even if everyone were given the same quantity of tokens in the future, the results produced could be vastly different: some would get more conversation logs, some would get more webpage content, while some might establish a logistics system, publishing system, or knowledge system that runs continuously for decades. What determines the difference is not the tokens themselves, but the system architecture behind them—and the incompressible time, the unpurchasable cross-domain experience, the difficult-to-replicate building capability, and the impossible-to-shortcut long-running operation records that support those system architectures.

Therefore, I do not regard tokens as an end, but as the new infrastructure of the intelligent age. Tokens are as important as electricity, but electricity alone does not automatically build factories; tokens are as measurable as money, but money alone does not automatically create quality assets; tokens can flow like water, but without pipes, dispatch systems, and long-term maintenance records, they will only dissipate. The truly valuable people in the future are not necessarily those with the most tokens, but those who can organize tokens into systems. The greatest misconception of the token era is treating tokens as the goal; the greatest

opportunity of the token era is using tokens to build structures that others cannot replicate. When tokens become as common as electricity in the future, people looking back at today's token craze may find that what truly determined the gap between people was never how many tokens they owned, but whether they possessed the ability to continuously learn, continuously practice, continuously accumulate, and continuously build systems—and those things behind these abilities that cannot be compressed, purchased, replicated, or shortcut. And in my view, this ability is the most scarce and most difficult-to-replicate wealth in the token era.

As my research deepens, I increasingly feel that the competitive landscape of the future AI era might be summarized by a simple analogy. Many people regard computing power, models, and tokens as core resources, but from the perspective of real-world operations, they are more like the logistics system behind an army. Computing power determines production capacity, models determine organizational capability, and tokens determine distribution capability; together they constitute the material support system for the AI era.

However, what truly determines the outcome of a battle is not the logistics supplies themselves, but the frontline combat system. For future industrial competition, systems are more like the forward line. Systems determine how resources are used, how processes are run, how value is created, and whether users truly benefit. No matter how abundant tokens and computing power are, without system organization, they are like supplies piled up in a warehouse, unable to be converted into real results.

From this perspective, future competition is likely not between computing power, nor between tokens, but between systems. The system is the forward line, determining the direction of attack and combat capability; tokens are ammunition and supplies, determining sustainable combat capability; computing power is the rear industrial base, determining production speed and supply capacity. Without rear supply, the forward line cannot advance long-term; but with only rear supply and no forward line system, results cannot be achieved either.

Therefore, as computing power becomes more widespread, models become more similar, and tokens become more abundant, what remains truly scarce is still the system. Whoever possesses a more mature system, more real data, more long-validated processes, and the ability to connect tokens and computing power with real-world needs will be more likely to gain sustained advantage in the future.

Tokens are becoming the "electricity" of the AI age—indispensable, but not an end in themselves. Just as the winners of the electricity revolution were not the power plants, but the inventors and entrepreneurs who invented the light bulb, the motor, and the electrolytic cell, the true winners of the AI age will be the system builders who can use tokens to construct irreproducible structures and architectures. Beyond Tokens, that is what is truly scarce.

Looking back over this entire article, I increasingly feel that the competitive landscape of the future AI age might be illustrated by a tree.

Many people focus on water and nutrients, because they are visible, tangible, and easily measurable. Today's computing power, models, and tokens are like the continuously delivered water and nutrients. Computing power determines production capacity, models determine transformation capability, and tokens determine distribution capability. As technology continues to advance, these resources are becoming increasingly abundant and easy to obtain.

However, what truly determines the vitality of a tree is not the water and nutrients themselves, but the root system.

The system is the root of future digital civilization.

The root system determines how water is absorbed, how nutrients are distributed, and whether the whole tree can grow long-term. Without roots, no matter how much water, it will be lost.; without roots, no matter how abundant the nutrients, they cannot be transformed into vitality. Similarly, without a system, no matter how powerful the model, how cheap the tokens, or how vast the computing power, it is difficult to continuously create value.

From this perspective, computing power is water, tokens are nutrients, models are the pipes that deliver nutrients, and the system is the root system that deeply penetrates the real world. What truly determines the outcome of future competition is not who has more water, nor who has more nutrients, but who has deeper, more stable roots that can absorb and utilize these resources long-term.

In the future, as computing power continues to grow, models continue to evolve, and token prices continue to fall, the resources themselves will become increasingly common. What will truly be scarce will no longer be the resources, but the systems that can organize, utilize, and continuously create value from the resources. Whoever possesses a mature system possesses the roots; whoever possesses the roots possesses the ability for long-term survival and continuous growth.

Therefore, I believe that what is truly scarce Beyond Tokens is not more tokens, nor more computing power, but the systems that can convert computing power, models, and tokens into real-world value.

When water and nutrients become increasingly abundant, what truly determines whether a tree can grow long-term is still the roots.

Keywords: Token, AI, Artificial Intelligence, Large Language Model, Computing Power, Model Economy, Token Economy, Chinese Token, English Token, BPE Algorithm, Data Sovereignty, Agent, API Ecosystem, System-Building Capability, System Design, Automation System, Cross-Disciplinary Research, Cross-Disciplinary Practice, Real Historical Data, Long-Running System, Time Asset, Platform Competition, System Competition, Digital Civilization, Civilization Memory, Knowledge Infrastructure, Independent Scholar, International Independent Scholar, Intelligent Logistics System, Ten-Language Publishing System, Metadata System, Automatic Cross-Referencing, DOI, Zenodo, WorldCat, TROVE, The Epochal Transition, Wu Zhaohui, JEFFI CHAO HUI WU

References:

1. *A Structural Case Study of the 1993 Remote Work Prototype System*
DOI: <https://doi.org/10.5281/zenodo.17978371>
 2. *The 2005 Intelligent Invoice System*
DOI: <https://doi.org/10.5281/zenodo.19044755>
 3. *Mature Structural Logistics System*
DOI: <https://doi.org/10.5281/zenodo.20201895>
 4. **Cross-Generational 20-Year Intelligent Logistics and Financial System**
DOI: <https://doi.org/10.5281/zenodo.20352371>
NLA: <https://nla.gov.au/nla.obj-4204311557>
 5. *Analysis of the Homepage Structural Evolution of The Epochal Transition*
DOI: <https://doi.org/10.5281/zenodo.20362574>
NLA: <https://nla.gov.au/nla.obj-4204309341>
 6. *An Empirical Study of Institutional Engineering: Independent Scholars Accessing the WorldCat Global Bibliographic Infrastructure*
DOI: <https://doi.org/10.5281/zenodo.18028572>
 7. *National-level Network Archiving and Digital Document Preservation Guidelines of the Australian Winner Information Network*
DOI: <https://doi.org/10.5281/zenodo.17888259>
 8. *The Institutional Archiving of The Epochal Transition*
DOI: <https://doi.org/10.5281/zenodo.17932379>
- WU, J. C. H. (2026). Source Declaration for Audiovisual and Derivative Adaptations of a Continuing Real-World Narrative. Zenodo.
<https://doi.org/10.5281/zenodo.18160116>

Après les Tokens : Ce qui est véritablement rare à l'ère de l'IA

Auteur/chercheur indépendant international : Wu Zhaohui, JEFFI CHAO HUI WU

ORCID : <https://orcid.org/0009-0009-5318-3027>

Déclaration préliminaire

Cet article est basé sur mes observations personnelles issues de plus de trente ans de pratique transdisciplinaire (depuis mes études en Australie en 1989, couvrant le travail à distance, la gestion logistique, le traitement automatisé, la création de sites web, la diffusion culturelle, l'édition multilingue et la recherche transdisciplinaire). **Les tokens sont une nouveauté en évolution rapide ; cet article ne représente que mon observation et mon jugement personnels, à un stade donné, à la date de rédaction (juin 2026) et ne constitue en aucun cas un conseil en investissement, une prévision commerciale ou une norme industrielle.**

Cet article ne recherche pas une applicabilité universelle et ne prétend pas représenter un quelconque consensus sectoriel. Avec l'évolution continue du marché des tokens, des technologies d'IA et de l'environnement politique mondial, certains jugements de cet article pourraient devoir être révisés ou mis à jour à l'avenir. Le lecteur doit porter un jugement indépendant en se basant sur les informations les plus récentes.

Les « cinq capacités rares » et les « cinq actifs personnels » dont je parle proviennent de ma propre pratique de la construction de systèmes (preuves publiques disponibles sur <https://times.net.au>, y compris la liste complète de plus de 40 œuvres créatives indépendantes sur ce site, l'ISSN 3083-5178, plus de 60 DOI, et des documents archivés en permanence par des infrastructures internationales telles que WorldCat, la Bibliothèque nationale d'Australie et CERN Zenodo).

Pour éviter toute sortie de contexte, toute réfutation ou questionnement valable de cet article doit avoir comme condition préalable : **la lecture complète de tous les articles liés dans ce document et des liens imbriqués qu'ils contiennent, ainsi que de la totalité de mes œuvres listées sur <https://times.net.au>**. Les citations partielles ou les résumés simplifiés détachés de ce contexte global ne peuvent constituer une évaluation juste de cet article.

Note complémentaire sur la réfutation valable :

Cet article exprime des opinions personnelles, ne recherche pas l'applicabilité universelle et ne présuppose pas la citation des travaux d'autrui. Toute contestation valable doit adopter une comparaison de type « pomme avec pomme » : faits contre faits, pratique personnelle contre pratique personnelle, systèmes réalisés contre systèmes réalisés. Utiliser « ce que l'IA future pourrait accomplir » pour nier « des pratiques déjà réalisées aujourd'hui », ou comparer les capacités d'une équipe/institution à une pratique individuelle, ne constitue pas une comparaison équitable.

Le portefeuille de cinq actifs décrit dans cet article constitue un enregistrement de pratique personnelle et ne prétend pas qu'« aucun cas comparable n'existe ». Si vous estimez qu'un cas pleinement comparable existe, veuillez fournir un cas public spécifique et expliquer ses preuves correspondantes sur les cinq dimensions que sont les données historiques réelles, les systèmes transversaux, la capacité de construction de systèmes, la vision prospective et la capacité d'auto-vérification. À défaut, toute contestation n'est qu'une expression d'opinion et ne constitue pas une réfutation valable.

Résumé

Les tokens deviennent l'unité de mesure centrale à l'ère de l'IA, mais cet article démontrera que le token n'est pas la valeur ultime. Alors que les coûts d'inférence continuent de baisser, les tokens deviendront aussi bon marché et omniprésents que le trafic Internet et le stockage sur disque dur le sont aujourd'hui. Lorsque les tokens

ne seront plus rares, qu'est-ce qui le deviendra véritablement ? Sur la base de plus de trente ans de pratique transdisciplinaire de l'auteur (d'un prototype de système de travail à distance en 1993, d'un système de facturation intelligent en 2005, jusqu'aux systèmes logistiques intelligents, au système d'édition en dix langues, à la gestion DOI, au croisement automatique et à la publication multiplateforme), cet article avance les jugements fondamentaux suivants : Premièrement, la capacité à construire des systèmes – y compris la conception structurelle, l'automatisation, la pratique transdisciplinaire, les modèles mentaux de fonctionnement à long terme et la capacité à intégrer des tokens dans des processus commerciaux réels – est le véritable actif rare. Deuxièmement, la capacité à construire des systèmes ne peut être acquise directement par de grandes quantités de tokens car elle dépend de la prise de décision, du traitement des exceptions, des arbitrages, des boucles de rétroaction du monde réel et de la confiance accumulée au fil du temps.

Troisièmement, l'incompressibilité du temps, le caractère non achetable de l'expérience transdisciplinaire, la difficulté de reproduire la capacité de construction de systèmes et l'impossibilité d'accélérer l'obtention de dossiers de fonctionnement à long terme – ces quatre actifs constituent la rareté la plus profonde à l'ère des tokens. Cet article prend comme preuve publique le site académique *The Epochal Transition* (<https://times.net.au>), ISSN 3083-5178, archivé en permanence par des infrastructures internationales telles que WorldCat, CERN Zenodo et TROVE de la Bibliothèque nationale d'Australie), construit par l'auteur, présentant comment une personne peut construire un système pluridisciplinaire, vérifiable et durable sur trente ans.

Cet article soutient que : les tokens ne sont que le carburant ; le système est le moteur. Après les tokens, ce qui est véritablement rare, c'est la personne capable de maîtriser les tokens, et les structures qu'elle a établies.

« Sans marchandises, à quoi sert l'argent ? »

Lorsque je passe en revue l'histoire du développement des tokens, je ressens de plus en plus qu'elle ressemble beaucoup à de nombreuses révolutions technologiques que l'humanité a connues par le passé. Chaque fois qu'une nouvelle

ressource fondamentale émerge, les gens se concentrent d'abord sur la ressource elle-même, négligeant la structure qui la sous-tend. C'était vrai à l'ère industrielle, à l'ère d'Internet, et c'est tout aussi vrai à l'ère de l'intelligence artificielle. Aujourd'hui, de nombreuses discussions portent sur les prix des tokens, les coûts des tokens, les quotas de tokens et la capacité d'approvisionnement en tokens, comme si posséder plus de tokens signifiait posséder l'avenir. Mais d'après mes décennies d'expérience pratique, la ressource elle-même n'a jamais été la valeur ultime. Ce qui détermine vraiment la valeur, c'est toujours la façon dont les gens organisent et utilisent ces ressources.

Pour comprendre les « tokens », on ne peut pas partir simplement d'un nouveau terme ; il faut commencer par son histoire. Du philosophe américain Charles Sanders Peirce écrivant vingt occurrences de « the » sur du papier en 1906, distinguant le « type » du « token », à Liu Liehong, directeur de l'Administration nationale des données de Chine, donnant la traduction chinoise officielle « ciyuan » (词元) au Token au forum de Boao en 2026, ce concept a parcouru un long chemin, de la philosophie à l'informatique, et de la blockchain à l'intelligence artificielle. Le mot « token » existait depuis longtemps en informatique occidentale, où, dans les langages de programmation, les compilateurs et les systèmes de traitement de texte, un token désignait l'unité de base qu'un ordinateur pouvait reconnaître et traiter. Plus tard, dans les contextes de sécurité réseau, d'authentification, d'appels API et de blockchain, « token » a souvent été compris comme un « laissez-passer » ou un « justificatif », indiquant des droits d'accès, des justificatifs d'identité ou des actifs numériques. Mais à l'ère des grands modèles de langage, le Token n'est plus principalement un justificatif d'identité, ni un actif blockchain, mais plutôt l'unité de calcul de base par laquelle les modèles d'IA traitent le langage, le code, les descriptions d'images et les informations multimodales.

La Chine a ensuite nommé le Token dans le domaine de l'IA « ciyuan » (词元). Je pense que c'est une traduction assez précise. Elle n'a pas utilisé « lingpai » (laissez-passer), car le Token dans les grands modèles n'est pas un laissez-passer ; elle ne l'a pas non plus simplement traduit par « zi » (caractère) ou « ci » (mot), car il peut

parfois être un caractère, parfois un mot, parfois la moitié d'un mot anglais, et parfois même un simple signe de ponctuation ou un symbole. « Ci » indique sa relation avec le langage et l'information, et « yuan » indique qu'il s'agit d'une unité de base. Pour le lecteur général, on peut comprendre simplement que : ce que les humains voient comme une phrase, l'IA le voit comme une chaîne de tokens. Sur cette voie, les tokens sont passés d'un terme linguistique obscur à l'unité de mesure centrale de l'ère de l'IA, puis à un outil de règlement reliant l'offre technologique à la demande commerciale. En tant que praticien ayant personnellement expérimenté des projets à long terme tels que la construction d'un système logistique intelligent, la conception d'un processus d'édition automatisé et la distribution de contenu multiplateforme, j'ai une compréhension viscérale de cette évolution – les tokens n'ont jamais été la fin, mais le moyen ; ce qui détermine vraiment la valeur, ce n'est jamais qui possède le plus de tokens, mais qui peut utiliser les tokens pour construire des systèmes fonctionnant durablement et des structures irréplicables. La capacité de construction de systèmes discutée dans cet article a ses résultats présentés publiquement sur <https://times.net.au> – un système d'édition académique transdisciplinaire construit par un chercheur indépendant et archivé en permanence dans le système bibliothécaire mondial.

En réfléchissant à mon propre parcours, je n'ai pas commencé à penser aux systèmes seulement après l'apparition des tokens, mais j'ai progressivement développé une compréhension des systèmes à travers une pratique de longue haleine. Depuis mes études en Australie en 1989, j'ai vécu des expériences d'apprentissage, de travail et d'exploration dans différents domaines. Plus tard, je me suis progressivement impliqué dans le travail à distance, la gestion logistique, le traitement automatisé, la création de sites web, la diffusion culturelle, l'édition multilingue et la recherche transdisciplinaire. En regardant en arrière, bien que ces expériences aient eu lieu à différentes époques, elles partagent une caractéristique commune : elles ont toujours tourné autour de la manière d'établir des structures, de réduire le travail répétitif et d'améliorer l'efficacité globale. Le concept de tokens n'existait pas à l'époque, pas plus que les grands modèles de langage d'aujourd'hui,

mais je réfléchissais déjà constamment à la manière de faire circuler, organiser et traiter l'information automatiquement.

Par conséquent, alors qu'un nombre croissant de personnes considèrent aujourd'hui les tokens comme un nouveau moyen de production, je suis plus enclin à les considérer comme une nouvelle forme d'énergie fondamentale. L'énergie est certes importante, mais l'énergie ne crée pas automatiquement de la valeur. L'électricité seule ne devient pas automatiquement une usine ; l'Internet seul ne devient pas automatiquement une entreprise ; les tokens seuls ne deviennent pas automatiquement un système de connaissances. À l'avenir, même si tout le monde peut obtenir facilement de grandes quantités de tokens, cela ne signifie pas que tout le monde puisse obtenir les mêmes résultats. Ce qui fait la différence, ce n'est pas la quantité de tokens, mais la structure dans laquelle les tokens entrent.

Avant de procéder à une analyse approfondie des tokens, je dois souligner une autre observation importante : Les tokens sont similaires aux monnaies, et les différences technologiques, systémiques et politiques entre l'Est et l'Ouest les rendent non universels. Tout comme le dollar américain et le renminbi chinois ne peuvent être librement échangés, les tokens chinois et les tokens anglais font face à des barrières de 流通 similaires. Ce n'est pas seulement une question de différence linguistique ; la raison plus profonde réside dans les triples différences entre l'Est et l'Ouest au niveau technologique (architectures de modèles, algorithmes de tokenisation, données d'entraînement), au niveau systémique (écosystèmes d'API, systèmes de paiement, normes de données) et au niveau politique (flux de données transfrontaliers, contrôles à l'exportation, examens de sécurité nationale), qui empêchent les tokens de circuler librement comme une « monnaie IA universelle mondiale » idéalisée. Cette observation a des implications significatives pour comprendre le futur paysage du marché des tokens – à l'avenir, des « zones de tokens chinois », des « zones de tokens anglais » et d'autres écosystèmes de tokens basés sur la langue et la politique pourraient émerger, et les appels inter-zones feront face à des pertes d'efficacité et à des coûts de conformité.

I. Origines des tokens : le quadruple saut du concept philosophique à l'unité économique

L'évolution de l'identité des tokens peut être clairement divisée en quatre étapes.

L'origine en linguistique philosophique. En 1906, Peirce, étudiant la sémiotique, distingua le « type » du « token ». Le premier représente une règle ou une forme abstraite, tandis que le second est l'instanciation spécifique de cette règle dans la réalité. Cette distinction devint plus tard un fondement important de la linguistique structurale.

Le tournant du codage en informatique. Dans les années 1960, lorsqu'un programmeur écrivait la ligne de code « int x = 5 ; », l'ordinateur devait la décomposer en unités indépendantes telles que « int », « x », « = », « 5 », etc. Chacune de ces unités était un token. Les tokens devinrent ainsi la plus petite unité d'information permettant à une machine de « lire » les instructions.

L'expérience du véhicule de valeur avec la blockchain. En 2017, avec l'essor de la frénésie des ICO, le Token reçut une nouvelle identité en tant que « certificat de droits numériques négociable ». Bien que cette vague d'enthousiasme se soit progressivement calmée, la compréhension du « Token comme symbole de valeur » avait pris racine.

L'infrastructure économique de l'ère de l'IA. Mars 2026 fut un moment marquant – le PDG de NVIDIA, Jensen Huang, a redéfini les centres de données comme des « usines produisant des Tokens d'IA intelligents », et Liu Liehong, directeur de l'Administration nationale des données de Chine, a donné la traduction chinoise officielle « ciyuan » (词元), le qualifiant de « point d'ancrage de la valeur de l'ère intelligente ». À ce stade, les tokens ont accompli leur dernier saut, passant d'un concept technique à un concept économique.

II. La base technique : l'algorithme BPE et la « taxe linguistique » implicite

La raison pour laquelle les tokens ont pu devenir l'unité de mesure standard de l'IA repose sur une technologie oubliée depuis plus de deux décennies – le codage par paires d'octets (Byte Pair Encoding, BPE).

En 1994, le programmeur américain Philip Gage publia un article sur les algorithmes de compression de données dans un magazine technique sur le langage C, présentant le principe de base du BPE : scanner le texte de manière répétée et « souder » les deux caractères adjacents les plus fréquents en un nouveau symbole, en comprimant de manière itérative tour après tour. Cet article ne suscita pas l'attention à l'époque car son efficacité de compression n'était pas exceptionnelle.

Ce n'est qu'en 2016 que Rico Sennrich, chercheur à l'Université d'Édimbourg, alors qu'il étudiait le problème de la segmentation lexicale en traduction automatique, tomba par hasard sur ce vieil article et réalisa avec acuité que le BPE était la solution parfaite au problème de tokenization – il n'était pas nécessaire de prédéfinir un dictionnaire ; les données étaient laissées « parler » d'elles-mêmes, les combinaisons à haute fréquence se coalesçant naturellement en tokens. En 2019, OpenAI a officiellement adopté cette approche lors de la sortie de GPT-2, plaçant le point de départ de la tokenization directement au niveau de l'« octet », permettant au modèle de traiter théoriquement n'importe quelle langue écrite.

Cependant, la logique de la « priorité à la fréquence » de l'algorithme BPE a formé par inadvertance un système implicite de « taxe linguistique ». Parce que l'anglais est le courant dominant absolu des corpus Internet, l'efficacité du modèle à segmenter les tokens anglais est bien supérieure à celle des autres langues. Pour exprimer la même signification, l'anglais utilise le moins de tokens, le chinois en nécessite généralement 1,5 à 2 fois plus, et pour les langues aux ressources encore plus faibles, le coût peut être de 5 à 10 fois supérieur à celui de l'anglais. Cela

signifie que, sous la tarification à la consommation, le coût réel de la conversation avec l'IA varie considérablement selon la langue utilisée.

Cette injustice de la « ligne de départ », une fois inscrite dans le vocabulaire initial d'un modèle, est très difficile à corriger – les règles de tokenization sont le « fondement » de la compréhension du monde par l'IA ; plus le bâtiment est haut, plus il est difficile de remplacer les fondations. Heureusement, cette situation s'améliore. Prenant la série GPT comme exemple, la même phrase en chinois nécessitait 38 tokens dans GPT-3, est tombée à 26 dans GPT-4, et à 15 dans GPT-5 – une amélioration d'efficacité de plus de 60%. De plus, les grands modèles chinois domestiques tels que Tongyi Qianwen et DeepSeek ont incorporé dès la phase de conception des phrases chinoises hautement fréquentes, des expressions idiomatiques, etc., comme tokens natifs dans leurs vocabulaires, réalisant ainsi une efficacité de traitement du chinois plus « native ». Cela révèle un principe profond : quiconque maîtrise le « droit à la segmentation sémantique » détient dans une large mesure l'avantage d'efficacité et de coût pour l'expression de cette langue dans le monde numérique – c'est une forme de « droit fondamental de frappe » à l'ère numérique.

III. Utilisations et production : le saut exponentiel de la conversation aux agents

Les scénarios de consommation des tokens subissent un saut à trois niveaux, de la « conversation » au « raisonnement » puis aux « agents ».

Au stade de l'IA conversationnelle, une seule invite et réponse consomme environ 10 000 tokens. Au stade de l'IA de raisonnement, impliquant de multiples tours d'interaction et inférences complexes, chaque tâche consomme environ 100 000 tokens. En entrant dans le stade de l'IA agentique, les agents peuvent effectuer de manière autonome des tâches en plusieurs étapes et invoquer de multiples outils, la consommation de tokens par tâche atteignant environ 1 million – une multiplication par près de 100 en seulement deux générations d'évolution. Le PDG de Qualcomm,

Amon, prédit que d'ici 2030, la demande mondiale de tokens toutes les 10 secondes passera de 31,7 milliards en 2026 à 1,27 billion. Ce n'est pas une exagération ; la popularité explosive d'outils agents open source comme OpenClaw a prouvé ce point : une seule tâche complexe consommant des dizaines de millions de tokens est monnaie courante.

Du côté de la production, la Chine est devenue le plus grand fournisseur mondial de tokens. Les données publiques montrent qu'en mars 2026, le volume d'appel de tokens quotidien de la Chine avait dépassé les 140 000 milliards, soit une augmentation de plus de mille fois en deux ans, représentant 61% de l'utilisation mondiale des tokens d'IA. L'établissement de cette position repose sur la combinaison de trois avantages : premièrement, les prix industriels de l'électricité en Chine sont bien inférieurs à ceux d'Europe et d'Amérique, et l'électricité représente plus de 60% des coûts de production de tokens ; deuxièmement, la Chine a construit le plus grand cluster de calcul intelligent au monde, réalisant une production efficace grâce à la répartition « Calcul à l'Est, données à l'Ouest » ; troisièmement, les grands modèles chinois domestiques sont entièrement open-source, et le système de services de commercialisation des API est mature, formant une chaîne industrielle complète en boucle fermée, de « l'entraînement des modèles - soutien en calcul - production d'inférence - appel mondial ».

IV. Institutions et marchés : le paysage industriel tripartite

Une structure de marché clairement stratifiée s'est formée autour des tokens.

Pour comprendre cette structure industrielle, il est également nécessaire de considérer le système de soutien sous-jacent aux tokens. Beaucoup de gens se concentrent sur les prix des tokens mais négligent que les tokens ne sont pas une ressource indépendante. Derrière les tokens se trouvent les modèles, et derrière les modèles se trouve la puissance de calcul. En un sens, la puissance de calcul détermine la capacité du modèle, les modèles déterminent la production de tokens, et les tokens sont l'unité de mesure formée lorsque les capacités du modèle sont

exportées vers le marché. Par conséquent, en surface, le marché des tokens est une question de concurrence par les prix, mais au fond, il s'agit toujours d'une concurrence en matière de puissance de calcul et d'efficacité.

La chaîne industrielle mondiale actuelle des tokens peut être grossièrement divisée en trois niveaux. Le niveau le plus élevé est celui de l'infrastructure de puissance de calcul, comprenant les fabricants de puces, les fabricants de serveurs, les plateformes de cloud computing, les centres de données et les systèmes énergétiques. Le niveau intermédiaire comprend les institutions de modèles fondamentaux et les plateformes d'inférence, chargées de convertir la puissance de calcul en services de tokens appelables. Le niveau inférieur comprend les fournisseurs de services de tokens, les plateformes d'agents et diverses applications industrielles, qui convertissent davantage les capacités des modèles en services d'entreprise et produits finaux. Ce que les utilisateurs voient finalement, c'est le prix du token, mais ce qui affecte réellement le coût et la vitesse du token, c'est souvent l'évolution du système de puissance de calcul en amont et de l'efficacité des modèles.

Au cours des dernières années, avec l'amélioration continue des performances des puces, l'optimisation des architectures de modèles et l'augmentation de l'efficacité d'inférence, le coût par token a montré une tendance à la baisse rapide. Beaucoup de gens voient les tokens devenir moins chers, mais le véritable moteur de ce changement est le progrès continu du système de puissance de calcul. Par conséquent, la concurrence future sur le marché des tokens impliquera non seulement une concurrence entre plateformes et par les prix, mais aussi une concurrence en matière de ressources de calcul, d'efficacité des modèles et de capacités d'intégration des ressources.

Centrales électriques en amont – producteurs de puissance de calcul. Les entreprises de puces représentées par NVIDIA améliorent continuellement le matériel pour améliorer l'efficacité de la production de tokens par unité de puissance de calcul. NVIDIA prédit que son chiffre d'affaires atteindra au moins 1 000 milliards

de dollars d'ici 2027, une confiance étayée par la soif mondiale de puissance de calcul IA.

Régulation du réseau au milieu – plateformes de modèles et agrégateurs. Des entreprises telles qu'OpenAI, Google, DeepSeek et Alibaba convertissent la puissance de calcul sous-jacente en services de tokens standardisés et appelables. Trois modèles commerciaux matures ont émergé autour de ce maillon : les plateformes d'agrégation de tokens (représentées par OpenRouter, fonctionnant sur les différences de prix et les commissions, avec un volume d'appels hebdomadaires atteignant 26 900 milliards de tokens en mai 2026), les services MaaS des fournisseurs de cloud, et les stations de relais IA actives sur le marché secondaire.

Points d'entrée en aval – plateformes d'application et de distribution. L'ouverture de l'interface WeChat par Tencent pour accéder aux capacités des nouveaux modèles transforme essentiellement WeChat en une « plateforme de distribution de tokens » ; China Telecom a proposé de « remodeler son activité par la gestion des tokens », tentant de sortir de la logique traditionnelle du trafic. Ces entreprises point d'entrée ont en commun que les utilisateurs ne font plus face directement au modèle, mais appellent les tokens indirectement par l'intermédiaire de la plateforme – le modèle lui-même devient « invisible ».

Il convient de mentionner qu'un paradigme unique de commerce mondial s'est formé autour des tokens : production immatérielle, fonctionnement hors taxes et couverture mondiale. Cela transforme la Chine d'une « usine du monde » traditionnelle en un « fournisseur mondial d'infrastructures de puissance de calcul IA ».

Lorsqu'on discute de l'économie des tokens, il y a un autre facteur fondamental qui ne peut être contourné : la puissance de calcul. Beaucoup de gens se concentrent sur les prix des tokens mais négligent le processus de production derrière les tokens. En fait, les tokens ne sont pas générés à partir de rien. Chaque génération de token nécessite que le modèle effectue une inférence, et l'inférence du modèle est construite sur une base de puissance de calcul massive. En un sens, la puissance de calcul est la source de l'économie des tokens, les modèles sont les

outils de production des tokens, et les tokens sont l'unité de mesure formée lorsque les capacités du modèle sont exportées vers l'extérieur.

Par conséquent, les niveaux de puissance de calcul affectent directement les coûts des tokens et la vitesse de génération des tokens. Plus la puissance de calcul est élevée, plus la capacité du modèle à traiter l'information est grande, et plus le nombre de tokens pouvant être produits par unité de temps est élevé ; plus le coût de la puissance de calcul est bas, plus les prix des tokens ont tendance à être bas. Ces dernières années, l'industrie des grands modèles a été témoin d'une baisse continue des prix des tokens, motivée non seulement par l'optimisation des algorithmes des modèles, mais aussi par l'amélioration des performances des puces, l'augmentation de l'échelle des centres de données et l'accroissement de l'efficacité d'inférence. En d'autres termes, beaucoup de gens voient les tokens devenir moins chers, mais le véritable moteur de ce changement est le progrès continu du système de puissance de calcul.

D'un point de vue de la chaîne industrielle, la puissance de calcul, les modèles et les tokens forment en réalité une chaîne complète. En amont se trouvent les puces, les serveurs, les centres de données et les systèmes énergétiques ; au milieu se trouvent les modèles fondamentaux et les plateformes d'inférence ; en aval se trouvent les services de tokens, les applications d'agents et diverses solutions industrielles. La puissance de calcul détermine la limite supérieure de la capacité du modèle ; les modèles déterminent la qualité et l'efficacité des tokens ; et les tokens déterminent finalement la quantité de services d'IA que l'utilisateur peut obtenir.

Cependant, je pense que, si l'importance de la puissance de calcul est indéniable, la puissance de calcul elle-même ne crée pas automatiquement de la valeur. De nombreuses industries ont connu des processus similaires dans l'histoire. Posséder une centrale électrique plus grande ne signifie pas posséder une meilleure entreprise ; posséder plus de serveurs ne signifie pas posséder un meilleur produit. De même, posséder plus de puissance de calcul ne signifie pas posséder un système de plus grande valeur. La puissance de calcul résout le problème de la capacité de production, tandis que les systèmes résolvent le problème de la

conversion de valeur. À l'avenir, à mesure que la puissance de calcul continuera de croître, que les capacités des modèles continueront de se renforcer et que les ressources en tokens continueront de proliférer, ce qui sera véritablement rare ne sera toujours pas la puissance de calcul elle-même, mais la manière d'organiser ces ressources en systèmes capables de fonctionner à long terme dans le monde réel.

Par conséquent, du point de vue de la concurrence future, je pense que la concurrence en matière de puissance de calcul, de modèles et de tokens se déplacera progressivement vers une concurrence en matière de systèmes. Parce que lorsque la puissance de calcul deviendra plus répandue, les modèles deviendront plus similaires et les tokens plus abondants, ce qui fera la différence ne sera plus la quantité de ressources possédées, mais qui pourra utiliser ces ressources pour construire des structures efficaces à long terme.

V. Concurrence et barrières : involution de la guerre des prix et stratification de la valeur

La concurrence sur le marché des tokens présente des caractéristiques d'involution très prononcées. De 2022 à 2026, le coût par inférence de token a diminué de 99,9%. Les modèles nationaux présentent des avantages de prix encore plus significatifs : le prix d'entrée de DeepSeek V4-Pro est aussi bas que 0,025 RMB par million de tokens, le prix de sortie de Tongyi Qianwen Qwen3.5-Plus est de 1,6 RMB par million de tokens, contre environ 72 RMB pour la sortie d'OpenAI GPT-4o.

Cette guerre des prix a fortement comprimé les marges bénéficiaires des « intermédiaires ». Les statistiques de l'industrie montrent que le nombre de diverses plateformes d'agrégation et de relais d'API a dépassé les deux mille, le marché formant trois grandes strates : la strate supérieure des grandes plateformes avec une technologie solide et des modèles complets ; la strate moyenne des agents professionnels disposant de réductions officielles et de ressources clientèles ; et la vaste multitude de micro et petits agents faisant face à de fortes pressions de survie.

Le modèle commercial consistant simplement à profiter des différences de prix des tokens n'est plus viable.

Les véritables barrières se déplacent dans deux directions. L'une est la direction de la prime technique – en développant des moteurs d'accélération d'inférence propriétaires, on peut vendre moins cher que le fournisseur officiel et tout de même réaliser un bénéfice. La seconde est la direction des services à valeur ajoutée – fournir des services personnalisés tels que l'ingénierie des instructions, le réglage fin des modèles et le déploiement privé. Plus notable encore, la valeur des tokens présente une différenciation énorme basée sur les scénarios : le prix moyen des tokens liés à la discussion est aussi bas que

0,01 par million, tandis que pour les tokens liés à la détection de médicaments, le prix moyen peut atteindre 1000 par million, tandis que pour les tokens liés à la détection de médicaments, le prix moyen peut atteindre 1000 par million, soit une différence de valeur de cent mille fois selon les scénarios. Cela révèle un principe essentiel : la valeur d'un token n'est pas déterminée par son coût de production, mais par son utilisation – le même token, utilisé dans différents scénarios et intégré dans différents systèmes, génère une valeur radicalement différente.

Mais un jugement plus important doit être formulé ici : Actuellement et à l'avenir, la barrière à la vente de tokens n'est pas élevée. N'importe qui peut s'inscrire pour une API, créer une plateforme d'agrégation et commencer à revendre des tokens. Ce qui est véritablement rare, c'est ce qui se trouve au-delà de la vente de tokens – les données historiques réelles, la capacité à construire des systèmes et des architectures, et une vision tournée vers l'avenir. Sans cela, tous les grands modèles et tokens ne servent à rien.

Jugement complémentaire sur la rentabilité des tokens : Gagner de l'argent avec les tokens est déjà difficile, et cela le deviendra de plus en plus. Les guerres des prix compriment continuellement les marges bénéficiaires, et le modèle commercial de la simple revente de tokens s'oriente rapidement vers des bénéfices maigres, voire

nuls. Ce n'est que lorsque les tokens sont attachés à un système frontal – c'est-à-dire lorsqu'un système fournit des services – que les tokens et la puissance de calcul peuvent servir d'actifs supplémentaires efficaces pour générer des bénéfices. En d'autres termes, les tokens eux-mêmes ne sont pas la source du profit ; le système l'est. Les tokens ne sont que le carburant du système, pas son moteur.

Alors que les tokens deviennent progressivement une unité de mesure importante à l'ère de l'intelligence artificielle, un nombre croissant de modèles commerciaux centrés sur les tokens ont émergé sur le marché. Outre les fournisseurs de modèles fondamentaux, on trouve également des agents de tokens, des plateformes d'agrégation de modèles, des plateformes d'agents, des services de distribution d'API d'entreprise et divers fournisseurs de services de couche intermédiaire. Pour de nombreux entrepreneurs, l'agence de tokens semble être un raccourci pour entrer dans l'industrie de l'IA car elle ne nécessite ni entraînement de modèle ni construction de grands centres de calcul ; il suffit d'intégrer les ressources existantes pour commencer à faire des affaires. Cependant, ce modèle a à la fois des avantages et des limites évidentes. Ses avantages incluent un investissement initial relativement faible, des barrières à l'entrée relativement basses et une expansion plus rapide du marché ; ses limites incluent une forte dépendance vis-à-vis des fournisseurs de modèles en amont, une vulnérabilité à la concurrence par les prix et un manque de barrières techniques à long terme. Alors que les prix des tokens continuent de baisser, la marge bénéficiaire de la simple revente de tokens risque d'être continuellement comprimée. Par conséquent, à long terme, l'agence de tokens est plus une capacité de distribution qu'une compétitivité fondamentale ultime. Ce qui détermine véritablement la valeur à long terme d'une entreprise, ce sont toujours les systèmes spécifiques à l'industrie, les processus d'automatisation, les systèmes de connaissances et les solutions du monde réel construits autour des tokens.

VI. Difficultés actuelles et voies d'avenir

Le marché actuel des tokens est confronté à plusieurs problèmes profonds.

Le piège de la concurrence homogène. Plus de deux mille plateformes se disputent les mêmes clients, et les guerres des prix continuent de faire baisser les bénéfices de l'industrie. Sans changement, cela pourrait répéter les débuts du cloud computing, où l'on « perdait de l'argent pour gagner des parts de marché ».

La menace des risques de conformité. Le « dividende réglementaire » actuel de l'absence de législation spécifique est un état temporaire, non une norme à long terme. À mesure que le marché des tokens se développe, les problèmes de données transfrontalières, de sécurité nationale et de recouvrement fiscal forceront les pays à accélérer leur législation. Il est particulièrement important de noter que les tentatives de certains acteurs du marché de tokeniser et de financieriser les tokens pourraient facilement déclencher une réglementation financière mondiale, non seulement entraînant la fermeture de ces activités, mais aussi nuisant à la réputation de toute l'industrie chinoise d'exportation de tokens.

Le mythe de la « Token-maxxing ». Une nouvelle forme d'« étalage de richesse » est apparue dans la Silicon Valley – rivaliser sur qui consomme le plus de tokens par jour. Meta a un classement interne appelé « Claudeconomics », qui compile les données d'utilisation de l'IA de plus de 85 000 employés, le meilleur employé brûlant des valeurs de tokens atteignant plusieurs millions de dollars. Les critiques déclarent avec mordant : « L'optimisation des résultats est meilleure que l'optimisation des tokens » – plutôt que de consommer frénétiquement des tokens, il vaut mieux regarder quels résultats sont produits. C'est exactement le problème central que j'ai profondément compris au cours d'années de pratique.

Tokens chinois vs. tokens anglais : trois barrières sous un concept unifié

Pour les utilisateurs chinois, un fait qui doit être clairement reconnu est le suivant : bien que les tokens chinois (ciyuan) et les tokens anglais appartiennent à la même unité de mesure fondamentale à l'ère de l'IA, ils n'ont pas formé un marché mondial véritablement unifié et librement interopérable. D'un point de vue conceptuel

technique, l'Est et l'Ouest utilisent tous les tokens comme unité de base pour que les modèles traitent la langue et les données ; cependant, d'un point de vue pratique, l'existence combinée de barrières technologiques, de barrières systémiques et de barrières politiques a créé une séparation de fait entre les tokens chinois et les tokens anglais.

La première barrière est la barrière technologique, à savoir les différences dans la logique de tokenization et l'architecture des modèles.

Les modèles occidentaux dominants tels que GPT, Claude et Gemini sont principalement construits sur de grands corpus principalement en anglais. Bien que ces modèles possèdent déjà d'excellentes capacités de traitement du chinois, dans de nombreux scénarios, le chinois a encore tendance à nécessiter plus de tokens pour exprimer la même quantité d'information que l'anglais. Plus important encore, différents modèles ont des structures de vocabulaire et des stratégies de tokenization différentes. Le même texte chinois peut être segmenté en un nombre de tokens totalement différent selon les modèles, et leurs méthodes de codage interne et de mappage sémantique peuvent également différer.

Parallèlement, les modèles chinois domestiques tels que DeepSeek, Tongyi Qianwen, Doubao et Wenxin Yiyao ont incorporé de grandes quantités de corpus chinois dès la phase d'entraînement, ce qui leur confère une adaptabilité native plus élevée aux expressions chinoises, idiomes, noms propres et expressions localisées. Par conséquent, même face au même texte, le nombre de tokens, l'efficacité du traitement et les voies de compréhension peuvent varier considérablement d'un modèle à l'autre. En ce sens, bien que le token soit un concept unifié, il n'existe pas de « token standard » pouvant être directement circuler et réutilisé entre les modèles.

La deuxième barrière est la barrière systémique, à savoir les différences dans les écosystèmes et les méthodes d'invocation.

Le marché mondial actuel des tokens a en réalité formé de multiples écosystèmes relativement indépendants. Les plateformes internationales telles qu'OpenAI, Anthropic et Google ont leurs propres normes d'API, modèles de tarification, règles

d'accès et systèmes de services ; le marché chinois a formé un écosystème domestique centré sur des plateformes comme DeepSeek, Tongyi Qianwen, Doubao et Kimi.

Pour les utilisateurs ordinaires, acheter des tokens sur une plateforme ne signifie pas qu'ils peuvent les utiliser directement sur une autre. Les interfaces d'invocation, les mécanismes d'authentification, les limites de débit, les accords de service et les règles commerciales diffèrent d'une plateforme à l'autre. Même si la conversion est effectuée par l'intermédiaire de passerelles tierces ou de plateformes d'agrégation, il s'agit essentiellement de rappeler un autre système, non d'un véritable échange de tokens. Par conséquent, bien que le token soit un concept unifié, les tokens en tant que ressource commerciale et capacité de service restent enfermés dans leurs écosystèmes de plateforme respectifs.

La troisième barrière est la barrière politique, à savoir les différences dans les flux de données et les exigences de conformité.

Alors que l'intelligence artificielle entre progressivement dans le stade de l'application d'entreprise, l'importance des questions de sécurité des données ne cesse de croître. Le système chinois de sécurité des données, le système RGPD de l'UE et les mesures de contrôle américaines sur les flux transfrontaliers de données sensibles imposent tous des exigences différentes en matière de transmission et de traitement des données. De nombreuses données impliquant des informations clients, des données financières, des dossiers médicaux, des secrets commerciaux et des données personnelles ne peuvent souvent pas traverser librement les frontières.

Cela signifie que, dans les environnements commerciaux réels, de nombreuses entreprises ne choisissent pas les modèles uniquement en fonction des prix des tokens, mais doivent prioriser la conformité des données, l'emplacement de stockage des données et la responsabilité légale. Pour une grande partie des données commerciales critiques réelles, les coûts de conformité sont souvent bien supérieurs au coût du token lui-même. Par conséquent, dans le monde réel, les

entreprises nationales ont tendance à utiliser des modèles nationaux conformes, tandis que les entreprises étrangères ont tendance à utiliser des modèles étrangers, ce qui entraîne une segmentation de fait du marché des tokens.

Les barrières technologiques déterminent les différences dans les méthodes de segmentation des tokens ; les barrières systémiques déterminent les différences dans les systèmes d'appel et les écosystèmes commerciaux ; les barrières politiques déterminent les différences dans les flux de données et les frontières de conformité. Ces trois barrières agissant ensemble font que, bien que les tokens chinois et les tokens anglais appartiennent au même concept technique, ils ne peuvent être librement échangés, librement circulés et librement interfacés comme une « monnaie IA universelle mondiale » idéalisée.

À l'avenir, une situation similaire au commerce international est susceptible d'émerger : l'écosystème des tokens chinois et l'écosystème des tokens anglais coexisteront longtemps, chacun suivant des voies de développement indépendantes. L'invocation inter-zones n'est pas impossible, mais elle nécessite souvent des adaptations techniques supplémentaires, une conversion de système et des coûts de révision juridique. Pour la plupart des utilisateurs ordinaires, cette capacité inter-zones n'est pas un besoin quotidien, donc les deux grands systèmes resteront probablement relativement indépendants pendant longtemps.

Les tokens produits en Chine ne sont pas nécessairement confinés à une utilisation dans les systèmes chinois, mais dans les scénarios clés impliquant la souveraineté des données, la responsabilité de conformité et la sécurité industrielle, une grande partie de la demande future de tokens sera probablement consommée de manière préférentielle au sein des modèles, plateformes et systèmes domestiques.

D'un point de vue tendanciel à long terme, le couplage entre le lieu de production de la puissance de calcul, le fournisseur de modèles, l'opérateur du système et l'emplacement des données augmente. Pour la finance, l'administration, la santé, les chaînes d'approvisionnement et les grands systèmes d'entreprise, les données ne peuvent souvent pas traverser librement les frontières. Par conséquent, même si la

Chine peut exporter de grandes quantités de services de tokens à l'échelle mondiale, une grande partie de la demande de tokens à haute valeur ajoutée restera préférentiellement au sein des modèles et systèmes chinois pour être complétée. Cela ne signifie pas que les tokens chinois ne peuvent pas être exportés, mais plutôt que le futur marché des tokens pourrait présenter simultanément deux structures : une partie « circulation mondiale » et une partie « boucle fermée domestique ».

Il convient de noter que ce qui pourrait véritablement réaliser l'unification à l'avenir, ce ne sont pas les tokens eux-mêmes, mais plutôt la manière dont les tokens sont invoqués. À mesure que les plateformes d'agents, les protocoles d'interface unifiés, les plateformes d'agrégation de modèles et les systèmes d'automatisation continueront de se développer, les utilisateurs ne se concentreront finalement plus sur le token utilisé, mais sur l'achèvement de la tâche, le caractère raisonnable du coût et l'efficacité suffisante. À ce moment-là, les tokens, comme l'électricité, le trafic Internet et les ressources de cloud computing aujourd'hui, deviendront progressivement cachés derrière les systèmes, devenant l'infrastructure opérationnelle fondamentale de la civilisation numérique.

Par conséquent, d'un point de vue à long terme, bien que les marchés des tokens orientaux et occidentaux soient temporairement difficiles à interconnecter directement, leurs directions de développement sont très alignées. Le passé des tokens appartient au traitement informatique du langage, le présent des tokens appartient à l'industrie de l'IA, et l'avenir des tokens appartiendra probablement au système opérationnel fondamental de l'ensemble de la civilisation numérique.

VII. Pourquoi la capacité à construire des systèmes ne peut être obtenue directement par de grandes quantités de tokens

En soutenant que « la capacité à construire des systèmes est un actif rare », je dois faire face à une possible contestation : puisque l'IA est si puissante, pourquoi la capacité à construire des systèmes ne peut-elle être « générée » ou « obtenue » en

consommant de grandes quantités de tokens ? La réponse à cette question révèle les limites fondamentales de la capacité des tokens.

Premièrement, la construction de systèmes nécessite des décisions, et les décisions ne sont pas une génération.

L'IA peut générer un article, un code, un modèle de site web, mais l'IA ne peut pas remplacer les décisions centrales dans la conception de systèmes que les humains doivent prendre – choisir quels modules, abandonner quelles fonctions, équilibrer les besoins actuels avec l'expansion future, quand remanier, quand maintenir le statu quo. Ces décisions dépendent de la compréhension de l'essence de l'activité, du jugement des contraintes de ressources et de la prévision des conséquences à long terme. Les tokens peuvent générer d'innombrables « solutions possibles », mais choisir « quelle solution vaut la peine d'être mise en œuvre » et « pourquoi » est quelque chose que les tokens ne peuvent pas remplacer. Mon système logistique intelligent a subi d'innombrables ajustements sur deux décennies, et chaque ajustement n'était pas parce qu'une IA m'a dit « c'est mieux », mais parce qu'un problème réel est apparu dans l'activité réelle, exposant des limites et soulevant de nouvelles exigences. Derrière ces décisions se trouve le jugement, et le jugement vient de la pratique, pas des tokens.

Deuxièmement, la construction de systèmes nécessite le traitement des exceptions, et les exceptions ne peuvent pas être énumérées de manière exhaustive.

La raison pour laquelle un système peut fonctionner à long terme n'est pas parce que son flux normal est parfaitement conçu, mais parce qu'il peut gérer les exceptions. Dans un système logistique, les exceptions incluent : adresses clients incorrectes, marchandises endommagées, retards douaniers, échecs de paiement, pannes du système, perte de données. Dans un système d'édition, les exceptions incluent : fluctuation de la qualité de la traduction, incompatibilité des formats de métadonnées, changements des interfaces d'archivage, échecs d'enregistrement DOI. Ces exceptions ne peuvent pas être énumérées de manière exhaustive car la variété des exceptions dans le monde réel est infinie. L'IA peut apprendre des «

modèles d'exception courants » grâce à de grandes quantités de tokens, mais elle ne peut pas prédire « la prochaine exception qui n'est jamais apparue auparavant ». Seul un système fonctionnant depuis longtemps, rencontrant, enregistrant et résolvant continuellement des exceptions dans la pratique, peut développer une véritable capacité de traitement des exceptions. Cette capacité ne peut être obtenue en « nourrissant plus de tokens », car les exceptions ne sont pas des connaissances ; elles sont de l'expérience.

Troisièmement, la construction de systèmes nécessite des arbitrages, et les arbitrages n'ont pas de réponses standard.

La construction de chaque système implique des arbitrages sous de multiples contraintes : coût vs performance, rapidité vs stabilité, généralité vs personnalisation, besoins immédiats vs expansion future. Ces arbitrages n'ont pas de « bonne réponse », seulement des « choix raisonnables dans des circonstances spécifiques ». L'IA peut fournir une « solution optimale », mais cet « optimal » est statistiquement optimal basé sur des données historiques, non un jugement basé sur « votre contexte spécifique ». La raison pour laquelle mon système intégré logistique et financier sur deux décennies a fonctionné jusqu'à aujourd'hui est précisément parce que j'ai fait différents arbitrages à différentes époques – choisir une certaine pile technologique en 2005, choisir une autre méthode d'intégration en 2015, et choisir une voie de mise à niveau en 2025. Ces arbitrages ne sont ni « bons » ni « mauvais », mais « adaptés au contexte de l'époque ». Les tokens ne peuvent pas faire ces arbitrages pour moi, car l'essence des arbitrages est un jugement de valeur, non un calcul.

Quatrièmement, la construction de systèmes nécessite des boucles de rétroaction, et la rétroaction nécessite le monde réel.

L'amélioration de la capacité à construire des systèmes dépend d'une boucle fermée : conception → mise en œuvre → fonctionnement → détection des problèmes → correction → reconception. L'étape de « détection des problèmes » dans cette boucle doit provenir d'une rétroaction du monde réel. L'IA peut simuler le

fonctionnement, mais elle ne peut pas simuler tous les facteurs imprévisibles du monde réel. La raison pour laquelle mon système logistique s'est continuellement amélioré est que de vrais clients ont fourni de vraies réactions ; la raison pour laquelle mon système d'édition s'est continuellement amélioré est que de vraies interfaces d'archivage ont connu de vrais changements. Ces réactions ne peuvent être obtenues en consommant des tokens dans un environnement fermé. Les tokens peuvent générer des « réactions hypothétiques des clients », mais ils ne peuvent pas remplacer le fait que « le client a réellement donné cette réaction ». La rétroaction du monde réel est le seul terreau pour la croissance de la capacité à construire des systèmes.

Cinquièmement, la construction de systèmes nécessite la confiance, et la confiance ne peut venir que du temps.

La confiance en un système ne se prouve pas par la promotion, mais par le temps. La raison pour laquelle mon site web est archivé en permanence par CERN Zenodo, WorldCat et la Bibliothèque nationale d'Australie n'est pas qu'ils « m'apprécient », mais que ces institutions, par une observation à long terme, ont confirmé que mon système est stable, fiable et persistant. Cette confiance ne peut être achetée avec de grandes quantités de tokens. L'IA peut générer un manuel de système « à l'aspect très professionnel », mais elle ne peut pas générer le fait que « ce système est digne de confiance depuis deux décennies ». La confiance est un produit du temps, pas un produit des tokens.

En résumé, la raison pour laquelle la capacité à construire des systèmes ne peut être obtenue directement par de grandes quantités de tokens est qu'elle n'est pas essentiellement une capacité générative, mais plutôt une capacité de jugement, de traitement des exceptions, d'arbitrage, d'absorption de la rétroaction et d'accumulation de la confiance. Ce que ces capacités partagent en commun, c'est qu'elles nécessitent du temps, de la pratique, des erreurs, des corrections et une validation dans le monde réel. Les tokens peuvent accélérer le traitement de l'information, mais ils ne peuvent pas accélérer l'accumulation d'expérience dans le monde réel. C'est la raison fondamentale pour laquelle, même si les tokens

deviennent extrêmement bon marché à l'avenir, la capacité à construire des systèmes restera rare.

VIII. Mes actifs : données historiques réelles, systèmes transversaux, capacité de construction, vision prospective et capacité d'auto-vérification

Avant de discuter plus avant de « ce qui est véritablement rare », je dois d'abord expliquer ma propre position. Parce que mes jugements ne sont pas dérivés d'une déduction théorique, mais d'une pratique de longue haleine. Ceci est très évident dans les différents systèmes que j'ai construits sur une longue période. Qu'il s'agisse du système logistique intelligent, du système de facturation intelligent, du système de traitement financier automatisé, ou plus tard du système d'édition en dix langues, du système de métadonnées, du système de navigation de site web et du système de croisement automatique, la partie vraiment difficile n'a jamais été de générer de l'information, mais de l'organiser. L'information peut croître indéfiniment, mais sans structure, elle ne fera que créer un nouveau chaos. Il en va de même à l'ère des tokens. À l'avenir, l'IA pourra générer d'énormes quantités de contenu, mais d'énormes quantités de contenu ne signifient pas automatiquement d'énormes quantités de valeur. Au contraire, à mesure que la génération de contenu devient plus facile, savoir comment filtrer, organiser, préserver, indexer et relier ce contenu deviendra de plus en plus important.

Je possède les cinq actifs suivants, qui constituent la base empirique de mon jugement sur « ce qui est rare après les tokens ».

Premièrement, des données historiques réelles. Ce ne sont pas n'importe quelles « données historiques », mais vingt ans de données commerciales réelles transgénérationnelles – du système de facturation intelligent de 2005 à nos jours, provenant de systèmes logistiques, financiers et fiscaux réels en fonctionnement, non pas d'environnements expérimentaux ou de données collectées. À l'ère de l'IA,

avec la prolifération des « données synthétiques », la valeur des données historiques réelles augmente rapidement. La plupart des entreprises possèdent des « données historiques mono-domaine », alors que je possède des données historiques transversales, couvrant vingt ans, et validées pratiquement.

Deuxièmement, l'achèvement de multiples systèmes et architectures

transversaux. Mes systèmes couvrent les domaines suivants : système logistique intelligent, système de facturation intelligent, système financier intelligent, système fiscal intelligent, système d'intégration logistique et financière sur deux décennies, système d'automatisation de site web, système de génération de métadonnées, système d'édition en dix langues de *The Epochal Transition*, système de gestion DOI, système d'organisation des liens TROVE et WorldCat, système de croisement automatique, système de génération de Sitemap, système de génération de profil Zenodo, système d'édition multiplateforme. Le caractère unique de ces systèmes réside dans : l'étendue transversale (de la logistique à l'édition, de la finance à l'indexation documentaire), la profondeur temporelle (certains systèmes fonctionnent depuis deux décennies) et l'achèvement par une seule personne (démontrant la transférabilité de la méthode). Parmi ceux-ci, le mensuel *The Epochal Transition* (<https://times.net.au>) sert de preuve publique – il possède le numéro ISSN international normalisé (3083-5178), est archivé en permanence par WorldCat (le plus grand système de catalogue collectif de bibliothèques au monde, avec des institutions membres dépassant les 16 000 bibliothèques nationales et universitaires dans plus de 100 pays et régions), la PANDORA/TROVE de la Bibliothèque nationale d'Australie, ainsi que Zenodo (géré par CERN, l'Organisation européenne pour la recherche nucléaire), OpenAIRE (l'infrastructure européenne pour la science ouverte), DataCite DOI (identifiant d'objet numérique mondial, avec plus de 60 documents), et ORCID (l'identifiant numérique académique permanent unique mondial). Ce site web démontre qu'une seule personne (un chercheur indépendant) peut réaliser l'ensemble du processus, de la création de contenu à l'interface avec les infrastructures d'archivage académique de premier plan.

Troisièmement, la capacité à construire des systèmes et des architectures. Ce n'est pas la capacité à « utiliser des outils d'IA », mais la capacité de conception pour transformer des exigences vagues en structures stables, la capacité architecturale pour intégrer différents systèmes en une architecture unifiée, la capacité d'ingénierie pour maintenir des systèmes en fonctionnement pendant des années sans intervention humaine, et la méta-capacité à transférer l'expérience de la logistique à l'édition, de la finance à la littérature. Comme discuté précédemment, cette capacité ne peut être obtenue directement par de grandes quantités de tokens car elle dépend de la prise de décision, du traitement des exceptions, des arbitrages, des boucles de rétroaction et de l'accumulation de la confiance – aucun de ces éléments ne peut être remplacé par l'IA. Le marché actuel de la formation à l'IA se concentre sur des compétences de surface telles que « l'ingénierie des instructions », « l'écriture par IA » et « le dessin par IA » ; presque aucun cours n'enseigne la « capacité à construire des systèmes transversaux » – car cette capacité ne peut pas être acquise rapidement ; elle ne peut être accumulée que par une pratique de longue haleine. Les plus de 40 œuvres créatives indépendantes listées sur le site web, l'ISSN obtenu et les plus de 60 DOI sont les meilleurs enregistrements publics de la capacité de fonctionnement à long terme du système, et non une autoproclamation.

Quatrièmement, une vision tournée vers l'avenir. Commencer à expérimenter un système prototype de travail à distance en 1993 – près de trente ans avant que le travail à distance à grande échelle ne se généralise ; concevoir un système de facturation intelligent en 2005 – commencer l'automatisation des processus vingt ans avant la démocratisation de l'IA ; l'intégration transgénérationnelle sur deux décennies – prévoir au moment de la conception du système qu'il aurait encore besoin de fonctionner vingt ans plus tard ; proposer maintenant le jugement « après les tokens » – à un moment où la plupart des gens courent encore après les tokens, prévoir que la capacité des systèmes est le véritable obstacle. J'ai également prévu que « les tokens sont similaires aux monnaies, et les différences technologiques, systémiques et politiques entre l'Est et l'Ouest les rendent non universels » – tout comme le dollar américain et le renminbi chinois ne peuvent être librement

échangés, les tokens chinois et les tokens anglais font face à des barrières de circulation similaires, et à l'avenir, des « zones de tokens chinois », des « zones de tokens anglais » et d'autres écosystèmes de tokens basés sur la langue et la politique pourraient émerger. J'ai également prévu que « gagner de l'argent avec les tokens est déjà difficile et le deviendra de plus en plus ; ce n'est que lorsqu'il y a un système frontal fournissant des services que les tokens et la puissance de calcul peuvent servir d'actifs supplémentaires efficaces pour générer des bénéfices ». En tant que chercheur indépendant, j'ai établi à l'avance un système d'édition personnel qui s'interface avec l'infrastructure académique future, prévoyant la tendance selon laquelle « un individu peut également devenir un nœud dans l'infrastructure mondiale de la connaissance ». La plupart des prévisions de tendance proviennent d'analystes ou d'entrepreneurs, alors que ma « capacité de prévision » a la validation de deux décennies de pratique comme garantie.

Cinquièmement, la capacité d'auto-vérification. C'est l'une de mes méta-capacités les plus importantes – tous les documents que j'écris dans divers domaines atteignent une clôture théorique et une autocohérence logique. Cela signifie que mes arguments ne reposent pas sur une autorité externe ou des citations, mais sur l'autosuffisance de la logique interne. La règle par défaut de la littérature académique est de « citer les recherches antérieures » ; peu de personnes revendiquent ou démontrent une « capacité d'auto-vérification » ; la discussion académique se concentre sur « si l'argumentation est suffisante », et non sur « si l'auteur a la capacité de construire une boucle fermée cohérente ». Dans mes articles, cette capacité d'auto-vérification se manifeste par : les arguments viennent de la pratique, la pratique valide les arguments, et les arguments et la pratique forment une boucle fermée. Je ne « cite pas les théories d'autrui pour me prouver », mais « montre les systèmes que j'ai construits et laisse les systèmes parler d'eux-mêmes ». Cette capacité d'auto-vérification est un concept entièrement absent de la littérature publique et l'une de mes méta-capacités les plus uniques.

La combinaison de ces cinq actifs, leur rareté n'est pas additive, mais multiplicative : données historiques réelles × systèmes transversaux × capacité de construction de

systèmes × vision prospective × capacité d'auto-vérification constitue un portefeuille d'actifs irreproductible. Dans les discussions publiques, aucun cas comparable ne peut actuellement être trouvé.

Le processus de systématisation de *La voie d'un chercheur indépendant international* m'a fait réaliser cela encore plus profondément. De 1989 à 2026, plus de trente ans d'expérience couvrent de multiples domaines. Ces expériences ne sont pas apparues toutes en même temps, mais se sont formées progressivement sur une longue période. L'IA peut générer un article en quelques secondes, mais elle ne peut pas générer plus de trente ans d'expérience réelle en quelques secondes ; elle peut générer de nombreuses opinions, mais elle ne peut pas générer directement le jugement accumulé par une pratique à long terme ; elle peut simuler de nombreux processus, mais elle ne peut pas remplacer la validation répétée du monde réel. En ce sens, le temps lui-même est également un actif important, et la structure précipitée par le temps est un actif encore plus rare.

IX. Après les tokens : cinq capacités véritablement rares

Lorsque les tokens eux-mêmes ne seront plus rares – la barrière à la vente de tokens n'est pas élevée, mais il y aura un manque extrême de données historiques réelles, de capacité à construire des systèmes et des architectures, et de vision tournée vers l'avenir – qu'est-ce qui sera véritablement rare ? Sur la base de ma pratique à long terme, du prototype de système de travail à distance de 1993, du système de facturation intelligent de 2005, du système logistique intelligent, du système financier intelligent, du système fiscal intelligent, du système d'intégration logistique et financière sur deux décennies, jusqu'à l'automatisation de sites web, la génération de métadonnées, le mensuel en dix langues *The Epochal Transition*, la gestion DOI, l'organisation des liens TROVE et WorldCat, le croisement automatique, la génération de Sitemap, la génération de profil Zenodo, l'édition multiplateforme, je crois que les cinq capacités suivantes deviendront de véritables obstacles.

Premièrement, la capacité de conception structurelle. Les tokens sont la matière première, et la structure est le cadre qui organise la matière première en produits de valeur. Bien avant la généralisation de l'IA, j'avais déjà commencé à réfléchir à la manière d'utiliser des ressources limitées pour atteindre une efficacité opérationnelle plus élevée. Mon système logistique intelligent, mon système de facturation intelligent et mon système logistique et financier intelligent sur deux décennies sont nés à différentes époques, face à différents problèmes, mais ils partagent une caractéristique commune : l'objectif n'était pas de créer plus d'informations, mais de réduire le chaos, de réduire le travail répétitif, d'améliorer l'efficacité globale et d'établir des structures pouvant fonctionner à long terme. Dans le système d'édition automatisé, le problème central auquel j'étais confronté n'était pas « de savoir si le texte pouvait être généré », mais « comment définir les champs de métadonnées pour dix langues, comment établir des associations entre les articles et les catégories, et comment concevoir des règles de gestion de version ». Ces décisions de conception structurelle déterminent si le système peut continuer à fonctionner pendant des années sans s'effondrer. En revanche, la quantité de tokens consommée n'est qu'un sous-produit. Le site web *The Epochal Transition* est lui-même une preuve publique de la capacité de conception structurelle : une structure de site web académique maintenue personnellement, avec un ISSN international normalisé, un support multilingue, de multiples canaux d'archivage, prouvant que je ne fais pas de discours vide, mais que j'ai déjà construit avec succès un système complexe et accessible au public en utilisant cette capacité de conception structurelle.

Deuxièmement, la capacité d'automatisation. Les tokens peuvent produire du contenu en masse, mais seuls les systèmes d'automatisation peuvent accomplir en masse des flux de travail de bout en bout. Dans le système logistique intelligent, la valeur réelle ne réside pas dans le nombre de fois où l'IA est appelée pour planifier des itinéraires, mais dans l'établissement d'une boucle fermée entièrement automatisée, de la réception des commandes, à la planification de l'entrepôt, l'optimisation des itinéraires, jusqu'au traitement des exceptions. Dans cette boucle fermée, l'IA n'est qu'un composant, et la capacité d'automatisation se manifeste

dans les connexions entre les composants. Une personne sans système, même avec une grande quantité de tokens, peut ne générer que des mots dispersés, des images dispersées, des tableaux dispersés et des codes dispersés ; une personne avec un système, en revanche, peut intégrer les tokens dans son flux de travail, leur permettant d'organiser automatiquement le matériel, de générer des métadonnées, de vérifier les liens, de mettre à jour les pages web, d'aider à l'édition, d'archiver la littérature, de créer des index, de traiter la logistique, d'analyser les finances et de générer des rapports.

Troisièmement, l'expérience pratique transversale. Les tokens sont disponibles pour tous, mais la capacité de transférer l'expérience d'une industrie à une autre n'est pas universelle. Si je peux construire à la fois un système logistique et un système d'édition, ce n'est pas parce que j'ai plus de tokens, mais parce que j'ai accumulé des méthodologies transférables telles que la gestion des métadonnées, la conception des flux de travail et le traitement des exceptions. Mon système d'édition a besoin de tokens car la traduction en dix langues, la génération de résumés, l'organisation des mots-clés, la génération de métadonnées et les descriptions de liens nécessitent tous un traitement linguistique important. Mon système de site web a besoin de tokens car les structures de page, les descriptions de navigation, les descriptions SEO, les métadonnées académiques et les présentations multilingues peuvent toutes être générées automatiquement. Mon système logistique peut également utiliser des tokens car les courriels des clients, les descriptions de marchandises, les documents douaniers, les explications d'exceptions, les notes financières et les journaux d'opération peuvent tous être assistés par l'IA. Mon système de croisement automatique a encore plus besoin de tokens car il doit trouver des associations thématiques entre un grand nombre d'articles, de titres, de DOI, de liens NLA, TROVE et WorldCat. Cette accumulation pratique transversale ne peut être obtenue en consommant des tokens ; elle ne peut être décantée qu'en résolvant des problèmes réels. Le site web présente simultanément des informations provenant de multiples domaines, y compris l'édition académique, l'intégration d'infrastructures, l'architecture technique et la marque personnelle, prouvant que j'ai

effectivement relié l'édition, la bibliothéconomie, la technologie et la marque personnelle, plutôt que de posséder une compétence unique.

Quatrièmement, le modèle mental pour le fonctionnement à long terme des systèmes. Le récit dominant pour les applications actuelles de l'IA est la « conversation » et la « génération », mais ce qui produit véritablement de la valeur durable, c'est le « fonctionnement ». Une conversation génère 10 000 tokens, une tâche d'agent en consomme un million, mais un système logistique qui a fonctionné continuellement pendant une décennie peut consommer des milliards de tokens – mais ce qui est véritablement rare, c'est le modèle mental pour concevoir ce système : comment garantir que le système reste utilisable après les mises à niveau du modèle ? Comment gérer les changements apportés aux API en amont ? Comment surveiller les exceptions et récupérer automatiquement ? Ces questions n'ont rien à voir avec les tokens, mais elles déterminent la production finale des tokens. La valeur de ces systèmes ne vient pas des tokens eux-mêmes, mais des structures que j'ai déjà construites ou que je développe. Les tokens ne sont que le carburant ; le système est le moteur. Les tokens ne sont que le courant électrique ; la structure est le circuit. Les tokens ne sont que l'eau ; les tuyaux, les vannes, les réservoirs et les systèmes de répartition déterminent si elle peut jouer un rôle. Sans système, les tokens se dissiperont ; avec un système, les tokens peuvent être convertis en valeur à long terme. Le site web affiche plus de 40 œuvres créatives indépendantes et un archivage permanent par des institutions faisant autorité telles que le CERN et la Bibliothèque nationale d'Australie, prouvant que mon système ne recherche pas le battage médiatique à court terme, mais la valeur à long terme, et a déjà reçu l'approbation de la « préservation permanente » d'institutions faisant autorité.

Cependant, lorsque je considère plus avant la relation entre les tokens et les systèmes, je trouve un autre problème souvent négligé. Beaucoup de gens pensent qu'avec un excellent système, ils pourront obtenir un avantage dans la future ère des tokens. Mais la réalité n'est pas si simple. Les systèmes eux-mêmes ne sont pas à la portée de n'importe qui. Les systèmes véritablement matures nécessitent souvent

une pratique à long terme, des révisions répétées et une validation dans le monde réel. De nombreux systèmes peuvent prendre des années, voire des décennies d'accumulation continue, de l'idée initiale à la mise en œuvre effective. Par conséquent, la capacité à concevoir des systèmes est déjà une ressource rare.

Mais même en possédant un système mature, cela ne conduit pas automatiquement à un impact à grande échelle. Entre la faisabilité technique d'un système et sa diffusion dans la société se trouvent de multiples couches incluant les marchés, les capitaux, les canaux de distribution, l'éducation des utilisateurs et les écosystèmes de plateformes. L'histoire ne manque pas de bonnes technologies ou de bons systèmes ; ce qui est véritablement rare, c'est la capacité de plateforme à pousser un système vers une application à grande échelle. Même si un système est bien supérieur aux solutions existantes dans un scénario local, sans soutien suffisant en capitaux, capacité de promotion de plateforme et base d'utilisateurs, son échelle d'utilisation réelle peut rester très limitée.

Du point de vue de l'économie des tokens, ceci est particulièrement évident. La demande de tokens ne vient pas directement du système lui-même, mais de l'échelle réelle d'utilisation du système. Une plateforme servant des millions d'utilisateurs par jour, même avec une conception de système relativement ordinaire, peut générer une consommation de tokens énorme ; inversement, un système avancé avec très peu d'utilisateurs peut avoir une demande de tokens très faible. Par conséquent, le développement futur du marché des tokens dépend non seulement de la capacité des modèles et de la capacité des systèmes, mais aussi de la capacité des plateformes et de la capacité d'intégration des ressources.

Cela signifie que la concurrence future comporte au moins trois niveaux. Le premier niveau est la concurrence des modèles, déterminant qui peut fournir des tokens. Le deuxième niveau est la concurrence des systèmes, déterminant qui peut utiliser efficacement les tokens. Le troisième niveau est la concurrence des plateformes, déterminant combien d'utilisateurs un système peut atteindre et quelle demande réelle il peut générer. Modèles, systèmes et plateformes sont tous indispensables. Sans modèles, les systèmes ne peuvent pas fonctionner. Sans systèmes, les tokens

ne peuvent pas créer de valeur. Sans plateformes, même le meilleur système peut rester à une échelle d'utilisation restreinte.

De ce point de vue, je suis encore plus convaincu que les tokens ne sont eux-mêmes que des ressources fondamentales. Ce qui détermine véritablement l'avenir, ce ne sont pas seulement ceux qui possèdent des tokens, ni seulement ceux qui possèdent des systèmes, mais ceux qui peuvent connecter modèles, systèmes, plateformes et besoins du monde réel. Ce n'est que lorsque ces éléments forment une boucle fermée que les tokens se transforment de simples unités de calcul en forces productives créatrices de valeur.

Cinquièmement, la capacité d'intégrer les tokens dans les processus commerciaux du monde réel. Les tokens doivent finalement servir un maillon du monde réel – logistique, édition, éducation, santé. Ce qui est véritablement rare, ce sont ceux qui comprennent à la fois les limites des capacités de l'IA et les points de douleur d'une industrie spécifique. Ils savent quand utiliser l'IA et quand ne pas l'utiliser ; ils savent comment convertir les sorties de tokens en décisions commerciales ; ils savent comment mesurer la relation réelle entre les entrées de tokens et les sorties commerciales. Cette capacité composite est actuellement la ressource la plus rare sur le marché. Le site web est lui-même un « processus commercial » en fonctionnement : création de contenu → génération de métadonnées → publication multiplateforme → archivage international → accès des lecteurs. Cela prouve que mon application de l'IA (consommation de tokens) a été intégrée dans un processus d'édition complet, en boucle fermée, générant un impact réel.

Je ressens de plus en plus que la véritable concurrence à l'avenir ne sera probablement pas entre les modèles, ni entre les tokens, mais entre les systèmes. Les modèles continueront d'évoluer, les prix des tokens continueront de baisser, et les ressources chères aujourd'hui deviendront très répandues à l'avenir. Mais les systèmes n'apparaîtront pas automatiquement parce que les ressources sont abondantes. Les systèmes nécessitent une accumulation à long terme, des révisions continues et une validation dans le monde réel. Quiconque peut convertir les tokens en systèmes fonctionnant à long terme est plus susceptible d'obtenir une valeur

durable. Et c'est aussi la clé de la rentabilité : gagner de l'argent avec les tokens est déjà difficile, et cela le deviendra de plus en plus ; ce n'est que lorsqu'il y a un système frontal fournissant des services que les tokens et la puissance de calcul peuvent servir d'actifs supplémentaires efficaces pour générer des bénéfices.

Je juge que la véritable concurrence à l'ère des tokens se déroulera en trois étapes : la première étape est la concurrence des ressources, tout le monde se disputant les modèles, la puissance de calcul, les API et les prix des tokens ; la deuxième étape est la concurrence des applications, tout le monde utilisant les tokens pour écrire, service client, code, traduction, éducation, travail de bureau et analyse de données ; la troisième étape est la véritable concurrence des systèmes, où celui qui peut intégrer les tokens dans des systèmes industriels, des systèmes d'édition, des systèmes logistiques, des systèmes de connaissances et des systèmes d'enregistrement de la civilisation fonctionnant à long terme aura probablement un avantage durable.

En fait, lorsqu'on discute des tokens, des modèles, des systèmes et des plateformes, il y a un autre facteur important souvent négligé dans le monde réel : le temps. Beaucoup de gens pensent qu'avec du capital, ils peuvent rapidement construire des systèmes ; avec un marché suffisamment grand, ils peuvent rapidement reproduire des modèles de réussite. Cependant, la réalité n'est souvent pas ainsi. Le capital peut acheter de la puissance de calcul, peut acheter des tokens, peut embaucher des équipes de développement, mais il ne peut pas acheter directement des données réelles et des systèmes matures qui ont été validés sur de nombreuses années. Pour de nombreuses industries, ce qui est véritablement rare n'est pas les modèles eux-mêmes, ni les tokens eux-mêmes, mais les systèmes qui ont fonctionné de manière stable dans des environnements commerciaux réels pendant une longue période.

D'un point de vue commercial, les marchés et les capitaux n'attendent généralement pas qu'un système entièrement nouveau se développe à partir de zéro. Un grand système, de la proposition du concept, de la conception de l'architecture, du développement des fonctions, de la validation des processus à la mise en œuvre

effective, prend souvent plusieurs années, voire plus. Au cours de ce processus, non seulement un investissement continu en ressources est nécessaire, mais aussi de nombreux essais et erreurs, révisions et validations dans le monde réel. Par conséquent, lorsque les opportunités de marché sont déjà apparues, les capitaux ont tendance à préférer les systèmes déjà formés, déjà en fonctionnement et déjà validés, plutôt que de concevoir un système entier à partir de zéro.

Cependant, il y a une autre contradiction dans la réalité. Plus un système est grand, plus il est difficile pour les capitaux ultérieurs de s'y engager facilement. Après un développement à long terme, les grands systèmes ont généralement déjà formé leurs propres mécanismes de fonctionnement, structures de gestion, logiques de données et cultures organisationnelles. Ils peuvent avoir des données historiques vastes, des processus commerciaux complexes et des systèmes d'expérience accumulés sur une longue période. Pour cette raison, bien que les grands systèmes soient extrêmement précieux, ils sont souvent difficiles à transformer rapidement ou à restructurer complètement selon les attentes des capitaux. Plus le système est grand, plus son inertie tend à être grande.

Par conséquent, dans le monde réel des affaires, l'état optimal n'est ni de concevoir un système à partir de zéro, ni d'acquérir un grand système rigide, mais de trouver des systèmes qui ont déjà achevé leur construction structurelle de base, qui ont été validés à long terme et qui peuvent fonctionner de manière stable. Même si ces systèmes ne sont pas parfaits, ils ont une très grande valeur car ils ont déjà franchi l'étape la plus difficile, possédant des données réelles, des processus réels et des enregistrements de validation réels. Sur cette base, le réemballage, l'optimisation, l'ouverture au monde extérieur et la promotion par plateformes sont souvent plus efficaces que de partir de zéro, et peuvent également réduire considérablement les risques d'essais et d'erreurs et les coûts de temps. Parce que ces systèmes ont déjà réalisé l'étape la plus coûteuse, la plus chronophage et la plus difficile à reproduire.

De ce point de vue, la fenêtre temporelle est également une ressource souvent négligée. Lorsqu'une industrie en est encore à ses débuts, beaucoup de gens se concentrent sur les capacités des modèles, les prix des tokens et la taille du marché,

ignorant le temps nécessaire à la formation des systèmes. La naissance d'un système mature passe généralement par plusieurs étapes : proposition du concept, conception de l'architecture, validation des processus, fonctionnement dans le monde réel et optimisation continue. La partie vraiment difficile n'est pas de proposer l'idée, mais de maintenir le système en fonctionnement stable dans un environnement réel sur le long terme. Par conséquent, lorsque le marché commence à reconnaître la valeur des systèmes matures, cela signifie souvent que ces systèmes ont déjà subi des années d'accumulation, plutôt que d'être quelque chose qui peut être reproduit à nouveau en peu de temps.

Pour les capitaux et le marché, les cibles les plus idéales ne sont pas les idées au stade conceptuel, ni les prototypes qui n'ont pas encore été validés, mais les systèmes embryonnaires qui ont achevé la construction structurelle de base, possèdent des enregistrements de fonctionnement réels et ont une marge d'amélioration. Bien que ces systèmes puissent ne pas être parfaits, ils ont déjà franchi l'étape la plus difficile, la plus chronophage et la plus risquée. Une fois que la demande du marché augmente rapidement, les systèmes comme celui-ci deviennent souvent très recherchés, car le nombre de systèmes matures réellement validés à long terme est généralement bien inférieur à ce que le marché imagine. Les modèles peuvent itérer rapidement, les tokens peuvent se développer continuellement, les équipes de développement peuvent être assemblées temporairement, mais les données réelles, les processus réels et les enregistrements de validation à long terme ne peuvent pas être recréés en peu de temps.

Par conséquent, dans le développement continu de l'économie des tokens, le temps n'affecte pas seulement la formation des systèmes, mais affecte aussi la valeur des systèmes. Lorsque le marché n'a pas encore pleinement reconnu l'importance d'un certain type de système, c'est souvent aussi la période où le coût d'acquisition de ces systèmes est le plus bas ; et lorsque le marché réalise généralement leur valeur, les choix disponibles sont souvent déjà très limités. Pour de nombreuses industries, ce qui est manqué n'est pas seulement un projet, mais une fenêtre temporelle pour

entrer dans l'étape suivante du développement qui a déjà achevé une accumulation à long terme.

Du point de vue de l'économie des tokens, il en va de même. Les tokens peuvent être obtenus rapidement, les modèles peuvent être utilisés rapidement, les plateformes peuvent être construites rapidement, mais ce qui est vraiment difficile à reproduire rapidement, ce sont les systèmes qui ont décanté pendant de nombreuses années et peuvent continuer à produire de la valeur. À l'avenir, lorsque les tokens deviendront de plus en plus abondants, l'importance des systèmes non seulement ne diminuera pas, mais pourrait même augmenter davantage. Parce que lorsque tout le monde aura des ressources en tokens similaires, ce qui déterminera le résultat de la concurrence ne sera plus la quantité de tokens, mais qui possède des données réelles plus matures, des processus commerciaux plus stables et des structures de système plus proches des besoins du monde réel.

Un autre problème souvent négligé est l'origine des systèmes. Beaucoup de gens ont l'habitude de comprendre les systèmes comme des produits logiciels, et donc supposent naturellement que les systèmes doivent être conçus et développés par des entreprises de logiciels. Cependant, dans de nombreuses industries complexes, les systèmes qui fonctionnent véritablement de manière stable à long terme ne sont souvent pas d'abord conçus par des entreprises de logiciels, puis donnés à l'industrie, mais plutôt formés progressivement au cours d'une pratique commerciale de longue haleine, puis logiciels, automatisés et transformés en plateformes.

La raison n'est pas complexe. Les entreprises de logiciels excellent généralement dans la mise en œuvre technique, mais elles ne restent pas nécessairement dans le contexte commercial spécifique pendant longtemps. Elles peuvent développer des fonctions rapidement, mais elles ont du mal à acquérir une grande quantité d'expérience commerciale réelle en peu de temps. En revanche, les systèmes qui se développent organiquement au sein d'une industrie sont continuellement révisés et améliorés face aux clients, commandes, exceptions, changements réglementaires, pressions sur les coûts et concurrence sur le marché au quotidien. Ces systèmes ne partent pas de la théorie, mais de problèmes réels ; chacun de leurs processus,

chacune de leurs règles, chaque lien d'automatisation provient typiquement d'une validation à long terme du travail réel.

Par conséquent, dans la logistique, les chaînes d'approvisionnement, l'édition, la finance, la fabrication, la santé et de nombreux autres domaines professionnels, les systèmes les plus précieux sont souvent ni les plus complexes, ni ceux avec le plus de fonctions, mais ceux qui ont été éprouvés sur le long terme et peuvent résoudre stablement des problèmes réels. Ils peuvent ne pas avoir les interfaces les plus avancées, la promotion la plus flashy ou l'architecture technique la plus récente, mais ils possèdent de grandes quantités de données réelles, de processus réels et de logique commerciale accumulée sur une longue période. Cette expérience ne peut souvent pas être obtenue par un développement à court terme.

Du point de vue de l'économie des tokens, ceci est particulièrement important. À l'avenir, les capacités des modèles pourront continuer à s'améliorer, les coûts des tokens pourront continuer à baisser, et l'efficacité du développement logiciel deviendra également de plus en plus élevée. Cependant, l'expérience commerciale accumulée par une industrie sur des décennies ne surgira pas automatiquement en raison des mises à niveau des modèles. Ce qui est vraiment difficile à reproduire, ce n'est pas le code logiciel lui-même, mais les connaissances sectorielles, l'expérience de traitement des exceptions, la logique de prise de décision et les processus commerciaux validés à long terme cachés derrière le système.

Par conséquent, je pense que les systèmes les plus idéaux à l'avenir ne seront pas ceux conçus complètement détachés de la pratique industrielle, mais ceux qui proviennent d'une activité réelle, ont été validés par un fonctionnement à long terme, puis sont continuellement améliorés avec l'aide de l'intelligence artificielle et des capacités des tokens. Ces systèmes possèdent à la fois la fiabilité du monde réel et les gains d'efficacité apportés par l'ère de l'IA, et leur valeur à long terme est souvent bien supérieure à celle des systèmes construits uniquement sur des concepts techniques.

X. Où vont les tokens : des conversations transitoires à la mémoire de la civilisation

En ce qui concerne la future économie des tokens, j'ai une autre observation. Beaucoup de gens se concentrent sur la façon dont les tokens sont produits, mais peu d'attention est portée à leur destination finale. Si un token ne reste que dans une seule conversation, sa durée de vie peut n'être que de quelques minutes ; s'il entre dans une page web ordinaire, sa durée de vie peut n'être que de quelques mois ; s'il entre dans un système de préservation à long terme, sa valeur change fondamentalement. Lorsque les tokens entrent dans les DOI, Zenodo, WorldCat, TROVE, les systèmes de préservation à long terme des bibliothèques nationales et diverses infrastructures académiques et de connaissances, ils ne sont plus simplement des tokens ; ils commencent à se transformer en enregistrements de connaissances, actifs documentaires et mémoire de la civilisation. Par conséquent, je pense que la véritable question importante à l'avenir n'est pas comment générer plus de tokens, mais comment permettre aux tokens d'entrer dans des structures où ils peuvent être préservés à long terme et continuer à jouer leur rôle.

C'est pour cette raison que je reste optimiste quant au développement futur des tokens, mais je ne vénère pas les tokens. Le passé des tokens est l'unité de base pour les ordinateurs afin de traiter le langage ; le présent des tokens est une unité de mesure et une unité commerciale importantes à l'ère de l'IA ; l'avenir des tokens est susceptible de devenir une ressource fondamentale importante pour la civilisation numérique. Cependant, quelle que soit l'évolution des tokens, je soutiens toujours que ce qui est véritablement rare, ce ne sont pas les tokens, mais les personnes capables de maîtriser les tokens ; non pas ceux qui ont plus de ressources, mais ceux qui peuvent établir des structures à long terme ; non pas ceux qui peuvent générer plus de contenu, mais ceux qui peuvent organiser le contenu en systèmes de connaissances.

XI. L'incompressibilité du temps : quatre actifs inachetables

En discutant de « ce qui est véritablement rare après les tokens », je dois souligner un point fondamental que la plupart des discussions négligent : certains actifs sont intrinsèquement incapables d'être accélérés, achetés ou reproduits. Quelle que soit la puissance de l'IA, quel que soit le bon marché des tokens, les quatre actifs suivants conserveront toujours leur rareté inhérente.

Premièrement, le temps est incompressible.

Trente ans de pratique, c'est simplement trente ans ; cela ne peut être compressé en trois ans, ni simulé par l'IA. L'IA peut générer un article sur « comment construire un système logistique » en quelques secondes, mais elle ne peut pas vivre, en quelques secondes, les diverses anomalies, pannes, changements de politique et fluctuations du marché qu'un système logistique rencontre sur deux décennies. Ces accumulations du monde réel de temps ne peuvent être raccourcies par aucun moyen technologique. Mon début d'expérimentation d'un prototype de système de travail à distance en 1993, la conception d'un système de facturation intelligent en 2005 – ces points dans le temps ne sont pas des accidents ; ce sont des marqueurs irréversibles de l'histoire réelle. Quiconque vient plus tard peut lire mes articles et apprendre mes méthodes, mais il ne peut pas posséder les années que j'ai vécues – ces années elles-mêmes sont des actifs.

Deuxièmement, l'expérience transversale ne peut pas être achetée.

L'expérience transversale ne peut pas être achetée comme une marchandise. Une personne peut acheter cent livres sur la logistique, l'édition, la finance et l'indexation documentaire en un seul jour, mais elle ne peut pas acquérir, ce même jour, l'expérience pratique de la connexion de ces domaines. La véritable capacité transversale n'est pas l'accumulation de connaissances, mais la capacité à établir des connexions entre différents domaines, à identifier des points communs et à transférer des méthodes. Cette capacité ne peut être acquise qu'en résolvant des problèmes réels, en faisant des erreurs, en corrigeant des chemins et en accumulant

du jugement dans un domaine après l'autre. La raison pour laquelle je peux transférer mon expérience des systèmes logistiques aux systèmes d'édition n'est pas parce que j'ai lu un certain livre, mais parce que j'ai passé vingt ans dans la logistique et dix autres dans l'édition – c'est à l'intersection de ces deux périodes qu'a émergé la méthodologie transférable.

Troisièmement, la capacité à construire des systèmes est difficile à reproduire.

La capacité à construire des systèmes n'est pas un ensemble de codes ou de processus qui peuvent être copiés et collés. Le code peut être copié, mais la logique de prise de décision derrière le code, les jugements de compromis, et la prévoyance des changements futurs ne peuvent pas être reproduits. La valeur d'un système ne réside pas seulement dans ce qu'il peut faire aujourd'hui, mais aussi dans combien de changements, de récupérations après panne et d'ajustements de besoins il a subis au cours de la dernière décennie tout en continuant à fonctionner. Cette « histoire du système » est elle-même une preuve de la capacité de construction. Mon site web (<https://times.net.au>) peut être consulté par n'importe qui, mais ce que les visiteurs voient, c'est le résultat, non les trois décennies d'historique de prise de décision, de choix technologiques, d'ajustements structurels et de maintenance continue derrière ce résultat. Cette capacité de construction ne peut être obtenue par plagiat ou imitation.

Quatrièmement, les enregistrements de fonctionnement à long terme ne peuvent pas être accélérés.

L'enregistrement qu'un système a fonctionné pendant deux décennies est lui-même intrinsèquement incapable d'être accéléré. Un enregistrement de fonctionnement à long terme prouve la stabilité, la maintenabilité et la prévoyance de conception du système. À l'ère de l'IA, générer un site web ou un système qui « semble » exister peut être réalisé en quelques minutes, mais générer un « système qui fonctionne depuis deux décennies et est archivé en permanence par le CERN, WorldCat et la Bibliothèque nationale d'Australie » n'a pas de raccourci. Un enregistrement de fonctionnement à long terme est un horodatage, déclarant à tous que ce système a

résisté à l'épreuve du monde réel. Un tel enregistrement ne peut être falsifié ou accéléré par aucune technologie d'IA.

Ces quatre actifs – le temps, l'expérience transversale, la capacité à construire des systèmes et les enregistrements de fonctionnement à long terme – constituent ensemble la base de ce que j'ai soutenu dans cet article comme étant « véritablement rare ». Je les possède non pas parce que je suis plus intelligent ou plus travailleur que les autres, mais parce que j'ai commencé plus tôt que les autres et que j'ai continué jusqu'à présent. À une époque où l'IA rend de plus en plus de choses bon marché et facilement accessibles, ces choses qui ne peuvent être compressées, achetées, reproduites ou accélérées deviendront de plus en plus précieuses.

XII. Conclusion : De « l'accumulation de tokens » à « la construction de systèmes »

L'avenir des tokens ne réside pas dans les tokens eux-mêmes.

Avec le développement continu de la technologie des grands modèles et la baisse continue des coûts d'inférence, l'offre de tokens continuera de s'étendre. Les tokens qui sont chers aujourd'hui seront aussi bon marché et omniprésents que le trafic Internet et le stockage sur disque dur dans quelques années. Lorsque tout le monde pourra obtenir facilement de grandes quantités de tokens, ce qui sera véritablement rare ne sera plus les tokens, mais la capacité à convertir les tokens en valeur dans le monde réel – capacité de conception structurelle, capacité d'automatisation, expérience industrielle et accumulation pratique à long terme. Et la rareté la plus profonde réside dans ces choses qui ne peuvent être compressées, achetées, reproduites ou accélérées : le temps lui-même, l'expérience transversale, la capacité à construire des systèmes et les enregistrements de fonctionnement à long terme. Comme soutenu précédemment, la capacité à construire des systèmes ne peut être obtenue directement par de grandes quantités de tokens car elle n'est pas essentiellement une capacité générative, mais plutôt une capacité de jugement, de

traitement des exceptions, d'arbitrage, d'absorption de la rétroaction et d'accumulation de la confiance – la caractéristique commune de ces capacités est qu'elles nécessitent du temps, de la pratique, des erreurs, des corrections et une validation dans le monde réel.

Sur le plan de la rentabilité, ce jugement s'applique également : gagner de l'argent avec les tokens est déjà difficile, et cela le deviendra de plus en plus. Les guerres des prix compriment continuellement les marges bénéficiaires, et le modèle commercial de la simple revente de tokens s'oriente rapidement vers des bénéfices maigres, voire nuls. Ce n'est que lorsque les tokens sont attachés à un système frontal – c'est-à-dire lorsqu'un système fournit des services – que les tokens et la puissance de calcul peuvent servir d'actifs supplémentaires efficaces pour générer des bénéfices. Les tokens eux-mêmes ne sont pas la source du profit ; le système l'est. Les tokens ne sont que le carburant du système, pas son moteur.

De ce point de vue, les systèmes que j'ai déjà réalisés ou que je développe ont une signification particulière pour le développement futur des tokens. Parce que ces systèmes n'ont pas été construits temporairement pour suivre une mode ; ils existaient bien avant que le concept de tokens ne devienne un sujet public. Le système logistique intelligent a prouvé que l'activité du monde réel peut être structurée, procéduralisée et automatisée ; le système d'édition a prouvé qu'un individu peut également établir un système documentaire multilingue, multiplateforme, multi-DOI, multi-archivage ; le système de site web et de métadonnées a prouvé que le contenu numérique peut être continuellement organisé, indexé et diffusé ; le système de croisement automatique prouve en outre que lorsque le nombre d'œuvres augmente continuellement, ce qui importe vraiment n'est pas de continuer à accumuler du contenu, mais d'établir des relations structurelles entre les contenus. Le site web *The Epochal Transition* (<https://times.net.au>), en tant que plateforme d'affichage centralisée de ces réalisations, est lui-même une preuve empirique de la conversion de la capacité des tokens en un système académique ouvert, permanent et vérifiable.

Même si tout le monde recevait la même quantité de tokens à l'avenir, les résultats produits pourraient être radicalement différents : certains obtiendraient plus de journaux de conversation, certains obtiendraient plus de contenu de pages web, tandis que d'autres pourraient établir un système logistique, un système d'édition ou un système de connaissances qui fonctionne continuellement pendant des décennies. Ce qui fait la différence, ce ne sont pas les tokens eux-mêmes, mais l'architecture du système qui les sous-tend – et le temps incompressible, l'expérience transversale inachetable, la capacité de construction difficile à reproduire et les enregistrements de fonctionnement à long terme impossibles à accélérer qui soutiennent ces architectures système.

Par conséquent, je ne considère pas les tokens comme une fin, mais comme la nouvelle infrastructure de l'ère intelligente. Les tokens sont aussi importants que l'électricité, mais l'électricité seule ne construit pas automatiquement des usines ; les tokens sont aussi mesurables que l'argent, mais l'argent seul ne crée pas automatiquement des actifs de qualité ; les tokens peuvent couler comme l'eau, mais sans tuyaux, systèmes de répartition et enregistrements de maintenance à long terme, ils ne feront que se dissiper. Les personnes véritablement précieuses à l'avenir ne seront pas nécessairement celles qui ont le plus de tokens, mais celles qui peuvent organiser les tokens en systèmes. La plus grande idée fausse de l'ère des tokens est de considérer les tokens comme un objectif ; la plus grande opportunité de l'ère des tokens est d'utiliser les tokens pour construire des structures que d'autres ne peuvent pas reproduire. Lorsque les tokens deviendront aussi courants que l'électricité à l'avenir, les gens regarderont en arrière sur l'engouement actuel pour les tokens et découvriront peut-être que ce qui a vraiment déterminé l'écart entre les personnes n'a jamais été le nombre de tokens qu'elles possédaient, mais si elles possédaient la capacité d'apprendre continuellement, de pratiquer continuellement, d'accumuler continuellement et de construire continuellement des systèmes – et ces choses derrière ces capacités qui ne peuvent être compressées, achetées, reproduites ou accélérées. Et à mon avis, cette capacité est la richesse la plus rare et la plus difficile à reproduire à l'ère des tokens.

Alors que mes recherches s'approfondissent, je ressens de plus en plus que le paysage concurrentiel de la future ère de l'IA pourrait être résumé par une analogie simple. Beaucoup de gens considèrent la puissance de calcul, les modèles et les tokens comme des ressources centrales, mais du point de vue du fonctionnement du monde réel, ils ressemblent plutôt au système logistique derrière une armée. La puissance de calcul détermine la capacité de production, les modèles déterminent la capacité d'organisation, et les tokens déterminent la capacité de distribution ; ensemble, ils constituent le système de soutien matériel de l'ère de l'IA.

Cependant, ce qui détermine véritablement l'issue d'une bataille, ce ne sont pas les fournitures logistiques elles-mêmes, mais le système de combat de première ligne. Pour la future concurrence industrielle, les systèmes sont davantage comme la ligne d'attaque. Les systèmes déterminent comment les ressources sont utilisées, comment les processus sont exécutés, comment la valeur est créée et si les utilisateurs en bénéficient véritablement. Aussi abondants que soient les tokens et la puissance de calcul, sans l'organisation d'un système, ils sont comme des fournitures entassées dans un entrepôt, incapables d'être convertis en résultats réels.

De ce point de vue, la concurrence future ne sera probablement pas entre puissances de calcul, ni entre tokens, mais entre systèmes. Le système est la ligne d'attaque, déterminant la direction de l'attaque et la capacité de combat ; les tokens sont les munitions et les fournitures, déterminant la capacité de combat durable ; la puissance de calcul est la base industrielle arrière, déterminant la vitesse de production et la capacité d'approvisionnement. Sans approvisionnement arrière, la ligne d'attaque ne peut pas avancer longtemps ; mais avec seulement l'approvisionnement arrière et aucun système de ligne d'attaque, des résultats ne peuvent pas non plus être obtenus.

Par conséquent, à mesure que la puissance de calcul se généralise, que les modèles deviennent plus similaires et que les tokens deviennent plus abondants, ce qui reste véritablement rare est toujours le système. Quiconque possède un système plus mature, des données plus réelles, des processus validés plus longtemps, et la

capacité de connecter les tokens et la puissance de calcul aux besoins du monde réel, sera plus susceptible d'obtenir un avantage durable à l'avenir.

Les tokens deviennent « l'électricité » de l'ère de l'IA – indispensables, mais non une fin en soi. Tout comme les gagnants de la révolution électrique n'étaient pas les centrales électriques, mais les inventeurs et entrepreneurs qui ont inventé l'ampoule, le moteur et la cuve électrolytique, les véritables gagnants de l'ère de l'IA seront les constructeurs de systèmes capables d'utiliser les tokens pour construire des structures et des architectures irréplicables. Après les tokens, voilà ce qui est véritablement rare.

En revenant sur l'ensemble de cet article, je ressens de plus en plus que le paysage concurrentiel de la future ère de l'IA pourrait être illustré par l'analogie d'un arbre.

Beaucoup de gens se concentrent sur l'eau et les nutriments, car ils sont visibles, tangibles et facilement mesurables. La puissance de calcul, les modèles et les tokens d'aujourd'hui sont comme l'eau et les nutriments continuellement délivrés. La puissance de calcul détermine la capacité de production, les modèles déterminent la capacité de transformation, et les tokens déterminent la capacité de distribution. À mesure que la technologie progresse, ces ressources deviennent de plus en plus abondantes et faciles à obtenir.

Cependant, ce qui détermine véritablement la vitalité d'un arbre, ce n'est pas l'eau et les nutriments eux-mêmes, mais le système racinaire.

Le système est la racine de la future civilisation numérique.

Le système racinaire détermine comment l'eau est absorbée, comment les nutriments sont distribués, et si l'arbre entier peut croître à long terme. Sans racines, même beaucoup d'eau s'écoulera ; sans racines, même les nutriments les plus abondants ne peuvent être transformés en vitalité. De même, sans système, aussi puissant soit le modèle, aussi bon marché les tokens, aussi vaste la puissance de calcul, il est difficile de créer de la valeur de manière continue.

De ce point de vue, la puissance de calcul est l'eau, les tokens sont les nutriments, les modèles sont les tuyaux qui délivrent les nutriments, et le système est le système racinaire qui pénètre profondément dans le monde réel. Ce qui détermine véritablement le résultat de la concurrence future, ce n'est pas qui a plus d'eau, ni qui a plus de nutriments, mais qui a des racines plus profondes, plus stables, capables d'absorber et d'utiliser ces ressources à long terme.

À l'avenir, à mesure que la puissance de calcul continuera de croître, les modèles continueront d'évoluer et les prix des tokens continueront de baisser, les ressources elles-mêmes deviendront de plus en plus courantes. Ce qui sera véritablement rare ne sera plus les ressources, mais les systèmes capables d'organiser, d'utiliser et de créer continuellement de la valeur à partir des ressources. Quiconque possède un système mature possède les racines ; quiconque possède les racines possède la capacité de survie à long terme et de croissance continue.

Par conséquent, je pense que ce qui est véritablement rare après les tokens, ce ne sont pas plus de tokens, ni plus de puissance de calcul, mais les systèmes capables de convertir la puissance de calcul, les modèles et les tokens en valeur dans le monde réel.

Lorsque l'eau et les nutriments deviennent de plus en plus abondants, ce qui détermine véritablement si un arbre peut croître à long terme, c'est toujours le système racinaire.

Mots-clés : Token, Intelligence Artificielle, IA, Grand Modèle de Langage, Puissance de Calcul, Économie des Modèles, Économie des Tokens, Token Chinois, Token Anglais, Algorithme BPE, Souveraineté des Données, Agent, Écosystème API, Capacité de Construction de Systèmes, Conception de Systèmes, Système d'Automatisation, Recherche Transdisciplinaire, Pratique Transdisciplinaire, Données Historiques Réelles, Système à Long Terme, Actif Temporel, Concurrence des Plateformes, Concurrence des Systèmes, Civilisation Numérique, Mémoire de la Civilisation, Infrastructure de Connaissance, Chercheur Indépendant, Chercheur Indépendant International, Système Logistique Intelligent, Système d'Édition en Dix

Langues, Système de Métadonnées, Croisement Automatique, DOI, Zenodo,
WorldCat, TROVE, The Epochal Transition, Wu Zhaohui, JEFFI CHAO HUI WU

Références :

1. *A Structural Case Study of the 1993 Remote Work Prototype System*
DOI: <https://doi.org/10.5281/zenodo.17978371>
2. *The 2005 Intelligent Invoice System*
DOI: <https://doi.org/10.5281/zenodo.19044755>
3. *Mature Structural Logistics System*
DOI: <https://doi.org/10.5281/zenodo.20201895>
4. **Cross-Generational 20-Year Intelligent Logistics and Financial System**
DOI: <https://doi.org/10.5281/zenodo.20352371>
NLA: <https://nla.gov.au/nla.obj-4204311557>
5. *Analysis of the Homepage Structural Evolution of The Epochal Transition*
DOI: <https://doi.org/10.5281/zenodo.20362574>
NLA: <https://nla.gov.au/nla.obj-4204309341>
6. *An Empirical Study of Institutional Engineering: Independent Scholars Accessing the WorldCat Global Bibliographic Infrastructure*
DOI: <https://doi.org/10.5281/zenodo.18028572>
7. *National-level Network Archiving and Digital Document Preservation Guidelines of the Australian Winner Information Network*
DOI: <https://doi.org/10.5281/zenodo.17888259>
8. *The Institutional Archiving of The Epochal Transition*
DOI: <https://doi.org/10.5281/zenodo.17932379>

WU, J. C. H. (2026). Source Declaration for Audiovisual and Derivative Adaptations of a Continuing Real-World Narrative. Zenodo.

<https://doi.org/10.5281/zenodo.18160116>

Después de los Tokens: Lo que es verdaderamente escaso en la Era de la IA

Autor/Investigador Independiente Internacional: Wu Zhaohui, JEFFI CHAO HUI WU

ORCID: <https://orcid.org/0009-0009-5318-3027>

Declaración Preliminar

Este artículo se basa en mis observaciones personales derivadas de más de treinta años de práctica transdisciplinaria (desde mis estudios en Australia en 1989, abarcando trabajo remoto, gestión logística, procesamiento automatizado, construcción de sitios web, difusión cultural, edición multilingüe e investigación transdisciplinaria). **Los tokens son una novedad en rápida evolución; este artículo representa solo mi observación y juicio personal, en una etapa determinada, en la fecha de redacción (junio de 2026) y no constituye en modo alguno asesoramiento de inversión, pronóstico comercial o estándar industrial.**

Este artículo no busca una aplicabilidad universal ni pretende representar ningún consenso sectorial. Con la evolución continua del mercado de tokens, las tecnologías de IA y el entorno político global, algunos juicios de este artículo podrían necesitar revisión o actualización en el futuro. El lector debe emitir juicios independientes basándose en la información más reciente.

Las «cinco capacidades escasas» y los «cinco activos personales» que discuto provienen de mi propia práctica de construcción de sistemas (evidencia pública disponible en <https://times.net.au>), incluyendo la lista completa de más de 40 obras creativas independientes en ese sitio, el ISSN 3083-5178, más de 60 DOI, y registros archivados permanentemente por infraestructuras internacionales como WorldCat, la Biblioteca Nacional de Australia y CERN Zenodo).

Para evitar sacar de contexto, cualquier refutación o cuestionamiento válido de este artículo debe tener como condición previa: **la lectura completa de todos los artículos enlazados en este documento y de los enlaces anidados que contienen, así como de la totalidad de mis obras listadas en <https://times.net.au>**. Las citas parciales o los resúmenes simplificados desprendidos de este contexto global no pueden constituir una evaluación justa de este artículo.

Nota complementaria sobre la refutación válida:

Este artículo expresa opiniones personales, no busca una aplicabilidad universal y no presupone la cita de trabajos de otros. Cualquier cuestionamiento válido debe adoptar una comparación de tipo «manzana con manzana»: hechos contra hechos, práctica personal contra práctica personal, sistemas completados contra sistemas completados. Utilizar «lo que la IA futura podría lograr» para negar «prácticas ya completadas hoy», o comparar las capacidades de un equipo/institución con una práctica individual, no constituye una comparación justa.

La cartera de cinco activos descrita en este artículo constituye un registro de práctica personal y no pretende que «no exista ningún caso comparable». Si cree que existe un caso plenamente comparable, proporcione un caso público específico y explique sus pruebas correspondientes en las cinco dimensiones de datos históricos reales, sistemas transversales, capacidad de construcción de sistemas, visión prospectiva y capacidad de autoverificación. De lo contrario, cualquier cuestionamiento es meramente una expresión de opinión y no constituye una refutación válida.

Resumen

Los tokens se están convirtiendo en la unidad de medida central en la era de la IA, pero este artículo demostrará que el token no es el valor último. A medida que los costes de inferencia continúan disminuyendo, los tokens se volverán tan baratos y omnipresentes como lo son hoy el tráfico de Internet y el almacenamiento en disco duro. Cuando los tokens ya no sean escasos, ¿qué será verdaderamente escaso?

Basándose en más de treinta años de práctica transdisciplinaria del autor (desde un prototipo de sistema de trabajo remoto en 1993, hasta un sistema de facturación inteligente en 2005, pasando por sistemas logísticos inteligentes, un sistema de edición en diez idiomas, gestión de DOI, referencias cruzadas automáticas y publicación multiplataforma), este artículo presenta los siguientes juicios fundamentales: Primero, la capacidad de construir sistemas —incluyendo el diseño estructural, la automatización, la práctica transdisciplinaria, los modelos mentales de funcionamiento a largo plazo y la capacidad de integrar tokens en procesos comerciales reales— es el verdadero activo escaso. Segundo, la capacidad de construir sistemas no puede adquirirse directamente mediante grandes cantidades de tokens porque depende de la toma de decisiones, el manejo de excepciones, las compensaciones, los bucles de retroalimentación del mundo real y la confianza acumulada a lo largo del tiempo. Tercero, la incompresibilidad del tiempo, la imposibilidad de comprar la experiencia transdisciplinaria, la dificultad de replicar la capacidad de construcción de sistemas y la imposibilidad de acelerar la obtención de registros de funcionamiento a largo plazo —estos cuatro activos constituyen la escasez más profunda en la era de los tokens. Este artículo utiliza como evidencia pública el sitio web académico *The Epochal Transition* (<https://times.net.au/>, ISSN 3083-5178, archivado permanentemente por infraestructuras internacionales como WorldCat, CERN Zenodo y TROVE de la Biblioteca Nacional de Australia), construido por el autor, presentando cómo una persona puede construir un sistema pluridisciplinario, verificable y duradero a lo largo de treinta años.

Este artículo sostiene que: los tokens son solo el combustible; el sistema es el motor. Después de los tokens, lo que es verdaderamente escaso es la persona que puede dominar los tokens, y las estructuras que ha establecido.

«Sin mercancías, ¿de qué sirve el dinero?»

Cuando reviso la historia del desarrollo de los tokens, siento cada vez más que se parece mucho a muchas revoluciones tecnológicas que la humanidad ha experimentado en el pasado. Cada vez que surge un nuevo recurso fundamental, las personas se centran primero en el recurso en sí, descuidando la estructura que

lo subyace. Esto fue cierto en la era industrial, en la era de Internet, y es igualmente cierto en la era de la inteligencia artificial. Hoy, muchas discusiones se centran en los precios de los tokens, los costes de los tokens, las cuotas de tokens y la capacidad de suministro de tokens, como si poseer más tokens significara poseer el futuro. Pero según mis décadas de experiencia práctica, el recurso en sí mismo nunca ha sido el valor último. Lo que realmente determina el valor es siempre la forma en que las personas organizan y utilizan estos recursos.

Para entender los «tokens», no se puede empezar simplemente con un nuevo término; hay que empezar por su historia. Desde el filósofo estadounidense Charles Sanders Peirce escribiendo veinte instancias de «the» en papel en 1906, distinguiendo entre «tipo» y «token», hasta Liu Liehong, director de la Administración Nacional de Datos de China, dando la traducción china oficial «ciyuan» (词元) al Token en el Foro de Boao en 2026, este concepto ha recorrido un largo camino, desde la filosofía hasta la informática, y desde la cadena de bloques hasta la inteligencia artificial. La palabra «token» existía desde hacía mucho tiempo en la informática occidental, donde, en los lenguajes de programación, los compiladores y los sistemas de procesamiento de texto, un token se refería a la unidad básica que un ordenador podía reconocer y procesar. Más tarde, en contextos de seguridad de redes, autenticación, llamadas a API y cadena de bloques, «token» a menudo se entendía como un «pase» o «credencial», indicando derechos de acceso, credenciales de identidad o activos digitales. Pero en la era de los grandes modelos de lenguaje, el Token ya no es principalmente una credencial de identidad, ni un activo de cadena de bloques, sino más bien la unidad básica de cálculo mediante la cual los modelos de IA procesan el lenguaje, el código, las descripciones de imágenes y la información multimodal.

China nombró posteriormente al Token en el campo de la IA «ciyuan» (词元). Creo que es una traducción bastante precisa. No utilizó «lingpai» (pase), porque el Token en los grandes modelos no es un pase; tampoco lo tradujo simplemente como «zi» (carácter) o «ci» (palabra), porque a veces puede ser un carácter, a veces una palabra, a veces la mitad de una palabra inglesa, y a veces incluso un simple signo

de puntuación o símbolo. «Ci» indica su relación con el lenguaje y la información, y «yuan» indica que es una unidad básica. Para el lector general, se puede entender simplemente como: lo que los humanos ven como una frase, la IA lo ve como una cadena de tokens. En este camino, los tokens han pasado de ser un término lingüístico oscuro a la unidad de medida central de la era de la IA, y luego a una herramienta de liquidación que conecta la oferta tecnológica con la demanda comercial. Como profesional que ha experimentado personalmente proyectos a largo plazo como la construcción de un sistema logístico inteligente, el diseño de procesos de edición automatizados y la distribución de contenidos multiplataforma, tengo una comprensión visceral de esta evolución: los tokens nunca han sido el fin, sino el medio; lo que realmente determina el valor nunca es quién posee más tokens, sino quién puede usar los tokens para construir sistemas que funcionen de manera duradera y estructuras irreplicables. La capacidad de construcción de sistemas discutida en este artículo tiene sus resultados presentados públicamente en <https://times.net.au> — un sistema de edición académica transdisciplinario construido por un investigador independiente y archivado permanentemente en el sistema bibliotecario mundial.

Reflexionando sobre mi propio viaje, no empecé a pensar en sistemas solo después de la aparición de los tokens, sino que desarrollé gradualmente una comprensión de los sistemas a través de una práctica de larga duración. Desde mis estudios en Australia en 1989, he experimentado el aprendizaje, el trabajo y la exploración en diferentes campos. Más tarde, me involucré gradualmente en el trabajo remoto, la gestión logística, el procesamiento automatizado, la construcción de sitios web, la difusión cultural, la edición multilingüe y la investigación transdisciplinaria. Mirando hacia atrás, aunque estas experiencias ocurrieron en diferentes épocas, comparten una característica común: siempre giraron en torno a cómo establecer estructuras, reducir el trabajo repetitivo y mejorar la eficiencia general. El concepto de tokens no existía entonces, ni los grandes modelos de lenguaje de hoy, pero ya estaba pensando constantemente en cómo hacer que la información fluyera, se organizara y se procesara automáticamente.

Por lo tanto, mientras que un número creciente de personas hoy consideran los tokens como un nuevo medio de producción, yo estoy más inclinado a verlos como una nueva forma de energía fundamental. La energía es ciertamente importante, pero la energía no crea automáticamente valor. La electricidad por sí sola no se convierte automáticamente en una fábrica; Internet por sí solo no se convierte automáticamente en una empresa; los tokens por sí solos no se convierten automáticamente en un sistema de conocimiento. En el futuro, incluso si todos pueden obtener grandes cantidades de tokens fácilmente, no significa que todos puedan obtener los mismos resultados. Lo que marca la diferencia no es la cantidad de tokens, sino la estructura en la que entran los tokens.

Antes de realizar un análisis en profundidad de los tokens, necesito señalar otra observación importante: Los tokens son similares a las monedas, y las diferencias tecnológicas, sistémicas y políticas entre Oriente y Occidente los hacen no universales. Así como el dólar estadounidense y el renminbi chino no pueden intercambiarse libremente, los tokens chinos y los tokens ingleses enfrentan barreras de circulación similares. No es solo una cuestión de diferencia lingüística; la razón más profunda reside en las triples diferencias entre Oriente y Occidente a nivel tecnológico (arquitecturas de modelos, algoritmos de tokenización, datos de entrenamiento), a nivel sistémico (ecosistemas de API, sistemas de pago, estándares de datos) y a nivel político (flujos de datos transfronterizos, controles de exportación, revisiones de seguridad nacional), que impiden que los tokens circulen libremente como una «moneda universal global de IA» idealizada. Esta observación tiene implicaciones significativas para comprender el futuro panorama del mercado de tokens — en el futuro, podrían surgir «zonas de tokens chinos», «zonas de tokens ingleses» y otros ecosistemas de tokens basados en el idioma y la política, y las llamadas entre zonas enfrentarán pérdidas de eficiencia y costes de cumplimiento.

I. Orígenes de los tokens: el salto cuádruple del concepto filosófico a la unidad económica

La evolución de la identidad de los tokens se puede dividir claramente en cuatro etapas.

El origen en la lingüística filosófica. En 1906, Peirce, al estudiar la semiótica, distinguió entre «tipo» y «token». El primero representa una regla o forma abstracta, mientras que el segundo es la instanciación específica de esa regla en la realidad. Esta distinción se convirtió más tarde en un fundamento importante de la lingüística estructural.

El giro de la codificación en la informática. En la década de 1960, cuando un programador escribía la línea de código «int x = 5;», el ordenador necesitaba descomponerla en unidades independientes como «int», «x», «=», «5», etc. Cada una de estas unidades era un token. Los tokens se convirtieron así en la unidad más pequeña de información para que una máquina «leyera» las instrucciones.

El experimento del vehículo de valor con la cadena de bloques. En 2017, con el auge de la locura de las ICO, al Token se le dio una nueva identidad como «certificado de derechos digitales negociable». Aunque esa ola de entusiasmo se calmó gradualmente, la comprensión del «Token como símbolo de valor» había echado raíces.

La infraestructura económica de la era de la IA. Marzo de 2026 fue un momento emblemático: el CEO de NVIDIA, Jensen Huang, redefinió los centros de datos como «fábricas que producen Tokens de IA inteligentes», y Liu Liehong, director de la Administración Nacional de Datos de China, dio la traducción china oficial «ciyuan» (词元), calificándolo como el «ancla de valor de la era inteligente». En este punto, los tokens completaron su último salto, de concepto técnico a concepto económico.

II. La base técnica: el algoritmo BPE y el impuesto lingüístico implícito

La razón por la que los tokens han podido convertirse en la unidad de medida estándar de la IA se basa en una tecnología olvidada durante más de dos décadas: el codificación por pares de bytes (Byte Pair Encoding, BPE).

En 1994, el programador estadounidense Philip Gage publicó un artículo sobre algoritmos de compresión de datos en una revista técnica sobre el lenguaje C, presentando el principio básico del BPE: escanear el texto repetidamente y «soldar» los dos caracteres adyacentes más frecuentes en un nuevo símbolo, comprimiendo iterativamente ronda tras ronda. Este artículo no atrajo la atención en ese momento porque su eficiencia de compresión no era excepcional.

No fue hasta 2016 que Rico Sennrich, investigador de la Universidad de Edimburgo, mientras estudiaba el problema de la segmentación de palabras en la traducción automática, encontró casualmente este viejo artículo y se dio cuenta con agudeza de que el BPE era la solución perfecta al problema de tokenización — no era necesario predefinir un diccionario; se permitía que los datos «hablaran» por sí mismos, y las combinaciones de alta frecuencia se fusionaban naturalmente en tokens. En 2019, OpenAI adoptó oficialmente este enfoque al lanzar GPT-2, estableciendo el punto de partida para la tokenización directamente a nivel de «byte», lo que permite al modelo procesar teóricamente cualquier idioma escrito.

Sin embargo, la lógica de «prioridad a la frecuencia» del algoritmo BPE formó inadvertidamente un sistema implícito de «impuesto lingüístico». Debido a que el inglés es la corriente principal absoluta de los corpus de Internet, la eficiencia del modelo para segmentar tokens en inglés es mucho mayor que para otros idiomas. Para expresar el mismo significado, el inglés usa la menor cantidad de tokens, el chino generalmente necesita de 1,5 a 2 veces más, y para idiomas con aún menos recursos, el coste puede ser de 5 a 10 veces mayor que el del inglés. Esto significa

que, bajo la tarificación por token, el coste real de conversar con la IA varía enormemente según el idioma utilizado.

Esta injusticia de la «línea de salida», una vez inscrita en el vocabulario inicial de un modelo, es muy difícil de corregir — las reglas de tokenización son el «fundamento» de la comprensión del mundo por parte de la IA; cuanto más alto es el edificio, más difícil es reemplazar los cimientos. Afortunadamente, esta situación está mejorando. Tomando la serie GPT como ejemplo, la misma frase en chino requería 38 tokens en GPT-3, cayó a 26 en GPT-4, y a 15 en GPT-5 — una mejora de eficiencia de más del 60%. Además, los grandes modelos chinos domésticos como Tongyi Qianwen y DeepSeek han incorporado desde la etapa de diseño frases chinas de alta frecuencia, expresiones idiomáticas, etc., como tokens nativos en sus vocabularios, logrando así una eficiencia de procesamiento del chino más «nativa». Esto revela un principio profundo: quienquiera que domine el «derecho a la segmentación semántica» posee en gran medida la ventaja de eficiencia y coste para expresar ese idioma en el mundo digital — esto es una forma de «derecho fundamental de acuñación» en la era digital.

III. Usos y producción: el salto exponencial de la conversación a los agentes

Los escenarios de consumo de tokens están experimentando un salto de tres niveles, de la «conversación» al «razonamiento» y luego a los «agentes».

En la etapa de IA conversacional, una sola petición y respuesta consume unos 10 000 tokens. En la etapa de IA de razonamiento, que implica múltiples rondas de interacción e inferencias complejas, cada tarea consume unos 100 000 tokens. Al entrar en la etapa de IA agéntica, los agentes pueden realizar de forma autónoma tareas de múltiples pasos e invocar múltiples herramientas, con un consumo de tokens por tarea que alcanza aproximadamente 1 millón — un aumento de casi 100 veces en solo dos generaciones de evolución. El CEO de Qualcomm, Amon, predice que para 2030, la demanda mundial de tokens cada 10 segundos se disparará de

31 700 millones en 2026 a 1,27 billones. Esto no es una exageración; la popularidad explosiva de herramientas de agente de código abierto como OpenClaw ha demostrado este punto: una sola tarea compleja que consume decenas de millones de tokens es algo común.

En el lado de la producción, China se ha convertido en el mayor proveedor mundial de tokens. Los datos públicos muestran que, en marzo de 2026, el volumen diario de llamadas a tokens de China había superado los 140 billones, un aumento de más de mil veces en dos años, lo que representa el 61% del uso mundial de tokens de IA. El establecimiento de esta posición se basa en la suma de tres ventajas: primero, los precios industriales de la electricidad en China son mucho más bajos que en Europa y América, y la electricidad representa más del 60% de los costes de producción de tokens; segundo, China ha construido el mayor clúster de cálculo inteligente del mundo, logrando una producción eficiente mediante la disposición «Cálculo al Este, Datos al Oeste»; tercero, los grandes modelos chinos domésticos son totalmente de código abierto, y el sistema de servicios de comercialización de API es maduro, formando una cadena industrial completa en bucle cerrado, desde el «entrenamiento de modelos - soporte de cálculo - producción de inferencia - llamada global».

IV. Instituciones y mercados: el panorama industrial tripartito

Una estructura de mercado claramente estratificada se ha formado en torno a los tokens.

Para comprender esta estructura industrial, también es necesario considerar el sistema de apoyo subyacente a los tokens. Muchas personas se centran en los precios de los tokens pero pasan por alto que los tokens no son un recurso independiente. Detrás de los tokens están los modelos, y detrás de los modelos está la potencia de cálculo. En un sentido, la potencia de cálculo determina la capacidad del modelo, los modelos determinan la producción de tokens, y los tokens son la unidad de medida que se forma cuando las capacidades del modelo se exportan al

mercado. Por lo tanto, en la superficie, el mercado de tokens se trata de competencia de precios, pero en el fondo, sigue siendo una competencia en materia de potencia de cálculo y eficiencia.

La cadena industrial mundial actual de tokens se puede dividir aproximadamente en tres niveles. El nivel más alto es el de la infraestructura de potencia de cálculo, que incluye fabricantes de chips, fabricantes de servidores, plataformas de computación en la nube, centros de datos y sistemas energéticos. El nivel intermedio incluye las instituciones de modelos fundamentales y las plataformas de inferencia, encargadas de convertir la potencia de cálculo en servicios de tokens invocables. El nivel inferior incluye los proveedores de servicios de tokens, las plataformas de agentes y diversas aplicaciones industriales, que convierten aún más las capacidades de los modelos en servicios empresariales y productos finales. Lo que los usuarios ven finalmente es el precio del token, pero lo que realmente afecta el coste y la velocidad del token es a menudo el cambio en el sistema de potencia de cálculo ascendente y la eficiencia de los modelos.

En los últimos años, a medida que el rendimiento de los chips ha seguido mejorando, las arquitecturas de los modelos se han optimizado y la eficiencia de la inferencia ha aumentado, el coste por token ha mostrado una tendencia a la baja rápida. Muchas personas ven que los tokens se vuelven más baratos, pero el verdadero motor de este cambio es el progreso continuo del sistema de potencia de cálculo. Por lo tanto, la competencia futura en el mercado de tokens implicará no solo la competencia entre plataformas y de precios, sino también la competencia en recursos de cálculo, eficiencia de los modelos y capacidades de integración de recursos.

Centrales eléctricas ascendentes – productores de potencia de cálculo. Las empresas de chips representadas por NVIDIA mejoran continuamente el hardware para mejorar la eficiencia de producción de tokens por unidad de potencia de cálculo. NVIDIA predice que sus ingresos alcanzarán al menos 1 billón de dólares para 2027, una confianza respaldada por la sed mundial de potencia de cálculo de IA.

Regulación de la red en el medio – plataformas de modelos y

agregadores. Empresas como OpenAI, Google, DeepSeek y Alibaba convierten la potencia de cálculo subyacente en servicios de tokens estandarizados e invocables. Han surgido tres modelos de negocio maduros en torno a este eslabón: las plataformas de agregación de tokens (representadas por OpenRouter, que operan con diferencias de precio y comisiones, con un volumen de llamadas semanales que alcanza los 26,9 billones de tokens en mayo de 2026), los servicios MaaS de los proveedores de nube, y las estaciones de retransmisión de IA activas en el mercado secundario.

Puntos de entrada descendentes – plataformas de aplicación y distribución. La apertura de la interfaz de WeChat por parte de Tencent para acceder a las capacidades de los nuevos modelos convierte esencialmente a WeChat en una «plataforma de distribución de tokens»; China Telecom ha propuesto «remodelar su negocio mediante la gestión de tokens», intentando salir de la lógica tradicional del tráfico. Lo que estas empresas punto de entrada tienen en común es que los usuarios ya no se enfrentan directamente al modelo, sino que invocan los tokens indirectamente a través de la plataforma — el modelo mismo se está volviendo «invisible».

Vale la pena mencionar que se ha formado un paradigma único de comercio mundial en torno a los tokens: producción intangible, funcionamiento libre de impuestos y cobertura mundial. Esto está transformando a China de una tradicional «fábrica del mundo» en un «proveedor mundial de infraestructura de potencia de cálculo de IA».

Al discutir la economía de los tokens, hay otro factor fundamental que no se puede eludir: la potencia de cálculo. Muchas personas se centran en los precios de los tokens pero pasan por alto el proceso de producción detrás de los tokens. De hecho, los tokens no se generan de la nada. Cada generación de token requiere que el modelo realice una inferencia, y la inferencia del modelo se basa en una base de potencia de cálculo masiva. En un sentido, la potencia de cálculo es la fuente de la economía de los tokens, los modelos son las herramientas de producción de los

tokens, y los tokens son la unidad de medida que se forma cuando las capacidades del modelo se exportan al exterior.

Por lo tanto, los niveles de potencia de cálculo afectan directamente los costes de los tokens y la velocidad de generación de tokens. Cuanto más fuerte es la potencia de cálculo, mayor es la capacidad del modelo para procesar información, y más tokens puede producir por unidad de tiempo; cuanto menor es el coste de la potencia de cálculo, más bajos tienden a ser los precios de los tokens. En los últimos años, la industria de los grandes modelos ha sido testigo de una continua disminución de los precios de los tokens, impulsada no solo por la optimización de los algoritmos de los modelos, sino también por las mejoras en el rendimiento de los chips, el aumento de la escala de los centros de datos y el incremento de la eficiencia de la inferencia. En otras palabras, muchas personas ven que los tokens se vuelven más baratos, pero el verdadero motor de este cambio es el avance continuo del sistema de potencia de cálculo.

Desde una perspectiva de la cadena industrial, la potencia de cálculo, los modelos y los tokens forman en realidad una cadena completa. Aguas arriba se encuentran los chips, los servidores, los centros de datos y los sistemas energéticos; en el medio se encuentran los modelos fundamentales y las plataformas de inferencia; aguas abajo se encuentran los servicios de tokens, las aplicaciones de agentes y diversas soluciones industriales. La potencia de cálculo determina el límite superior de la capacidad del modelo; los modelos determinan la calidad y la eficiencia de los tokens; y los tokens determinan finalmente cuánto servicio de IA puede obtener el usuario.

Sin embargo, creo que, si bien la importancia de la potencia de cálculo es innegable, la potencia de cálculo en sí misma no crea automáticamente valor. Muchas industrias han experimentado procesos similares en la historia. Poseer una central eléctrica más grande no significa poseer una mejor empresa; poseer más servidores no significa poseer un mejor producto. Del mismo modo, poseer más potencia de cálculo no significa poseer un sistema de mayor valor. La potencia de cálculo resuelve el problema de la capacidad de producción, mientras que los sistemas

resuelven el problema de la conversión de valor. En el futuro, a medida que la potencia de cálculo siga creciendo, las capacidades de los modelos sigan fortaleciéndose y los recursos de tokens sigan proliferando, lo que será verdaderamente escaso no será la potencia de cálculo en sí misma, sino la forma de organizar estos recursos en sistemas capaces de funcionar a largo plazo en el mundo real.

Por lo tanto, desde una perspectiva de competencia futura, creo que la competencia en potencia de cálculo, modelos y tokens se desplazará gradualmente hacia una competencia en sistemas. Porque cuando la potencia de cálculo se vuelva más generalizada, los modelos se vuelvan más similares y los tokens más abundantes, lo que marcará la diferencia ya no será la cantidad de recursos poseídos, sino quién puede usar estos recursos para construir estructuras efectivas a largo plazo.

V. Competencia y barreras: involución de la guerra de precios y estratificación del valor

La competencia en el mercado de tokens presenta características de involución muy pronunciadas. De 2022 a 2026, el coste por inferencia de token disminuyó un 99,9%. Los modelos nacionales presentan ventajas de precio aún más significativas: el precio de entrada de DeepSeek V4-Pro es tan bajo como 0,025 RMB por millón de tokens, el precio de salida de Tongyi Qianwen Qwen3.5-Plus es de 1,6 RMB por millón de tokens, en comparación con el precio de salida de OpenAI GPT-4o de aproximadamente 72 RMB.

Esta guerra de precios ha comprimido bruscamente los márgenes de beneficio de los «intermediarios». Las estadísticas de la industria muestran que el número de diversas plataformas de agregación y retransmisión de API ha superado los dos mil, formando el mercado tres grandes estratos: el estrato superior de grandes plataformas con tecnología sólida y modelos completos; el estrato medio de agentes profesionales con descuentos oficiales y recursos de clientes; y la vasta multitud de micro y pequeños agentes que enfrentan fuertes presiones de supervivencia. El

modelo de negocio de simplemente beneficiarse de las diferencias de precio de los tokens ya no es sostenible.

Las verdaderas barreras se están desplazando en dos direcciones. Una es la dirección de la prima técnica — desarrollando motores de aceleración de inferencia propietarios, se puede vender más barato que el proveedor oficial y aún así obtener beneficios. La segunda es la dirección de los servicios de valor añadido — proporcionar servicios personalizados como ingeniería de indicaciones, ajuste fino de modelos y despliegue privado. Aún más notable, el valor de los tokens presenta una diferenciación enorme basada en los escenarios: el precio medio de los tokens relacionados con la charla es tan bajo como 0,01 dólares por millón, mientras que para los tokens relacionados con el descubrimiento de fármacos, el precio medio puede alcanzar los 1000 dólares por millón, una diferencia de valor de cien mil veces según los escenarios. Esto revela un principio esencial: el valor de un token no está determinado por su coste de producción, sino por su uso — el mismo token, utilizado en diferentes escenarios e integrado en diferentes sistemas, genera un valor radicalmente diferente.

Pero es necesario hacer un juicio más importante aquí: Actualmente y en el futuro, la barrera para vender tokens no es alta. Cualquiera puede registrarse para obtener una API, crear una plataforma de agregación y empezar a revender tokens. Lo que es verdaderamente escaso es lo que está más allá de la venta de tokens — datos históricos reales, la capacidad de construir sistemas y arquitecturas, y una visión orientada al futuro. Sin esto, todos los grandes modelos y tokens son inútiles.

Juicio complementario sobre la rentabilidad de los tokens: Ganar dinero con los tokens ya es difícil, y lo será cada vez más. Las guerras de precios comprimen continuamente los márgenes de beneficio, y el modelo de negocio de la simple reventa de tokens se encamina rápidamente hacia beneficios reducidos o incluso nulos. Solo cuando los tokens se adjuntan a un sistema frontal — es decir, cuando un sistema proporciona servicios — pueden los tokens y la potencia de cálculo servir como activos complementarios efectivos para generar beneficios. En otras

palabras, los tokens en sí mismos no son la fuente del beneficio; el sistema lo es.

Los tokens son solo el combustible del sistema, no su motor.

A medida que los tokens se convierten gradualmente en una unidad de medida importante en la era de la inteligencia artificial, ha surgido un número creciente de modelos de negocio centrados en los tokens en el mercado. Además de los proveedores de modelos fundamentales, también hay agentes de tokens, plataformas de agregación de modelos, plataformas de agentes, servicios de distribución de API de nivel empresarial y diversos proveedores de servicios de capa intermedia. Para muchos emprendedores, la agencia de tokens parece ser un atajo para entrar en la industria de la IA porque no requiere entrenamiento de modelos ni construcción de grandes centros de cálculo; solo es necesario integrar los recursos existentes para empezar a hacer negocios. Sin embargo, este modelo tiene tanto ventajas como limitaciones evidentes. Sus ventajas incluyen una inversión inicial relativamente baja, barreras de entrada relativamente bajas y una expansión más rápida del mercado; sus limitaciones incluyen una alta dependencia de los proveedores de modelos ascendentes, una vulnerabilidad a la competencia de precios y una falta de barreras técnicas a largo plazo. A medida que los precios de los tokens continúan cayendo, el margen de beneficio de la simple reventa de tokens corre el riesgo de comprimirse continuamente. Por lo tanto, a largo plazo, la agencia de tokens es más una capacidad de distribución que una competitividad fundamental última. Lo que realmente determina el valor a largo plazo de una empresa sigue siendo los sistemas específicos de la industria, los procesos de automatización, los sistemas de conocimiento y las soluciones del mundo real construidos en torno a los tokens.

VI. Dificultades actuales y vías futuras

El mercado actual de tokens enfrenta varios problemas profundos.

La trampa de la competencia homogénea. Más de dos mil plataformas compiten por los mismos clientes, y las guerras de precios siguen reduciendo los beneficios de la

industria. Sin un cambio, esto podría repetir los primeros días de la computación en la nube, donde se «perdía dinero para ganar cuota de mercado».

La amenaza de los riesgos de cumplimiento. El actual «dividendo regulatorio» de la ausencia de legislación específica es un estado temporal, no una norma a largo plazo. A medida que el mercado de tokens se expande, los problemas de datos transfronterizos, seguridad nacional y recaudación de impuestos forzarán a los países a acelerar la legislación. Es particularmente importante señalar que los intentos de algunos actores del mercado de tokenizar y financiarizar los tokens podrían fácilmente desencadenar una regulación financiera global, no solo provocando el cierre de esas actividades, sino también dañando la reputación de toda la industria de exportación de tokens de China.

El mito de la «Token-maxxing». Ha surgido una nueva forma de «presumir riqueza» en Silicon Valley: competir sobre quién consume más tokens al día. Meta tiene una clasificación interna llamada «Claudeconomics», que recopila datos de uso de IA de más de 85 000 empleados, y el empleado mejor clasificado quema valores de tokens que alcanzan varios millones de dólares. Los críticos señalan con mordacidad: «La optimización de resultados es mejor que la optimización de tokens» — en lugar de consumir tokens frenéticamente, es mejor ver qué resultados se producen. Este es exactamente el problema central que he entendido profundamente a través de años de práctica.

Tokens chinos vs. tokens ingleses: tres barreras bajo un concepto unificado

Para los usuarios chinos, un hecho que debe ser reconocido claramente es el siguiente: aunque los tokens chinos (ciyuan) y los tokens ingleses pertenecen a la misma unidad de medida fundamental en la era de la IA, no han formado un mercado mundial verdaderamente unificado y libremente interoperable. Desde una perspectiva conceptual técnica, Oriente y Occidente utilizan los tokens como la unidad básica para que los modelos procesen el idioma y los datos; sin embargo,

desde una perspectiva práctica, la existencia combinada de barreras tecnológicas, barreras sistémicas y barreras políticas ha creado una separación de facto entre los tokens chinos y los tokens ingleses.

La primera barrera es la barrera tecnológica, es decir, las diferencias en la lógica de tokenización y la arquitectura de los modelos.

Los modelos occidentales dominantes como GPT, Claude y Gemini se construyen principalmente sobre grandes corpus predominantemente en inglés. Aunque estos modelos ya poseen excelentes capacidades de procesamiento del chino, en muchos escenarios, el chino todavía tiende a requerir más tokens para expresar la misma cantidad de información que el inglés. Aún más importante, diferentes modelos tienen diferentes estructuras de vocabulario y estrategias de tokenización. El mismo texto chino puede segmentarse en un número de tokens totalmente diferente según los modelos, y sus métodos de codificación interna y mapeo semántico también pueden diferir.

Al mismo tiempo, los modelos chinos domésticos como DeepSeek, Tongyi Qianwen, Doubao y Wenxin Yiyan han incorporado grandes cantidades de corpus chinos desde la etapa de entrenamiento, lo que les confiere una mayor adaptabilidad nativa a las expresiones chinas, modismos, nombres propios y expresiones localizadas. Por lo tanto, incluso frente al mismo texto, el número de tokens, la eficiencia del procesamiento y las vías de comprensión pueden variar significativamente de un modelo a otro. En este sentido, aunque el token es un concepto unificado, no existe un «token estándar» que pueda ser directamente circulado y reutilizado entre los modelos.

La segunda barrera es la barrera sistémica, es decir, las diferencias en los ecosistemas y los métodos de invocación.

El mercado mundial actual de tokens ha formado en realidad múltiples ecosistemas relativamente independientes. Las plataformas internacionales como OpenAI, Anthropic y Google tienen sus propios estándares de API, modelos de tarificación, reglas de acceso y sistemas de servicios; el mercado chino ha formado un

ecosistema doméstico centrado en plataformas como DeepSeek, Tongyi Qianwen, Doubao y Kimi.

Para los usuarios ordinarios, comprar tokens en una plataforma no significa que puedan usarlos directamente en otra. Las interfaces de invocación, los mecanismos de autenticación, los límites de velocidad, los acuerdos de servicio y las reglas comerciales difieren de una plataforma a otra. Incluso si la conversión se realiza a través de pasarelas de terceros o plataformas de agregación, esencialmente se está invocando a otro sistema, no un verdadero intercambio de tokens. Por lo tanto, aunque el token es un concepto unificado, los tokens como recurso comercial y capacidad de servicio permanecen bloqueados dentro de sus respectivos ecosistemas de plataforma.

La tercera barrera es la barrera política, es decir, las diferencias en los flujos de datos y los requisitos de cumplimiento.

A medida que la inteligencia artificial entra gradualmente en la etapa de aplicación empresarial, la importancia de los problemas de seguridad de los datos no deja de crecer. El sistema chino de seguridad de datos, el sistema GDPR de la UE y las medidas de control estadounidenses sobre los flujos transfronterizos de datos sensibles imponen todos diferentes requisitos sobre la transmisión y el procesamiento de datos. Muchos datos que involucran información de clientes, datos financieros, registros médicos, secretos comerciales y datos personales a menudo no pueden cruzar libremente las fronteras.

Esto significa que, en entornos comerciales reales, muchas empresas no eligen los modelos basándose únicamente en los precios de los tokens, sino que deben priorizar el cumplimiento de los datos, la ubicación de almacenamiento de los datos y la responsabilidad legal. Para una gran parte de los datos comerciales críticos reales, los costes de cumplimiento suelen ser mucho más altos que el coste del token en sí mismo. Por lo tanto, en el mundo real, las empresas nacionales tienden a utilizar modelos nacionales conformes, mientras que las empresas extranjeras

tienden a utilizar modelos extranjeros, lo que resulta en una segmentación de facto del mercado de tokens.

Las barreras tecnológicas determinan las diferencias en los métodos de segmentación de tokens; las barreras sistémicas determinan las diferencias en los sistemas de invocación y los ecosistemas comerciales; las barreras políticas determinan las diferencias en los flujos de datos y las fronteras de cumplimiento. Estas tres barreras actuando juntas hacen que, aunque los tokens chinos y los tokens ingleses pertenezcan al mismo concepto técnico, no pueden ser libremente intercambiados, libremente circulados y libremente interconectados como una «moneda universal global de IA» idealizada.

En el futuro, es probable que surja una situación similar al comercio internacional: el ecosistema de tokens chinos y el ecosistema de tokens ingleses coexistirán durante mucho tiempo, cada uno siguiendo caminos de desarrollo independientes. La invocación entre zonas no es imposible, pero a menudo requiere adaptaciones técnicas adicionales, conversión de sistemas y costes de revisión legal. Para la mayoría de los usuarios ordinarios, esta capacidad entre zonas no es una necesidad diaria, por lo que los dos grandes sistemas probablemente permanecerán relativamente independientes durante mucho tiempo.

Los tokens producidos en China no están necesariamente confinados a su uso dentro de los sistemas chinos, pero en escenarios clave que involucran soberanía de datos, responsabilidad de cumplimiento y seguridad industrial, es probable que una gran parte de la demanda futura de tokens se consuma preferentemente dentro de los modelos, plataformas y sistemas domésticos.

Desde una perspectiva de tendencia a largo plazo, el acoplamiento entre la ubicación de la producción de potencia de cálculo, el proveedor de modelos, el operador del sistema y la ubicación de los datos está aumentando. Para las finanzas, la administración, la salud, las cadenas de suministro y los grandes sistemas empresariales, los datos a menudo no pueden fluir libremente a través de las fronteras. Por lo tanto, incluso si China puede exportar grandes cantidades de

servicios de tokens a nivel mundial, gran parte de la demanda de tokens de alto valor seguirá siendo preferentemente dentro de los modelos y sistemas chinos para su realización. Esto no significa que los tokens chinos no puedan exportarse, sino que el futuro mercado de tokens podría presentar simultáneamente dos estructuras: una parte de «circulación global» y una parte de «bucle cerrado doméstico».

Vale la pena señalar que lo que podría verdaderamente lograr la unificación en el futuro no son los tokens en sí mismos, sino la forma en que se invocan los tokens. A medida que las plataformas de agentes, los protocolos de interfaz unificados, las plataformas de agregación de modelos y los sistemas de automatización sigan desarrollándose, los usuarios finalmente no se centrarán en qué token se utiliza, sino en si la tarea se completa, si el coste es razonable y si la eficiencia es suficiente. En ese momento, los tokens, como la electricidad, el tráfico de Internet y los recursos de computación en la nube hoy, se volverán gradualmente ocultos detrás de los sistemas, convirtiéndose en la infraestructura operativa fundamental de la civilización digital.

Por lo tanto, desde una perspectiva a largo plazo, aunque los mercados de tokens orientales y occidentales son temporalmente difíciles de interconectar directamente, sus direcciones de desarrollo están muy alineadas. El pasado de los tokens pertenece al procesamiento informático del lenguaje, el presente de los tokens pertenece a la industria de la IA, y el futuro de los tokens probablemente pertenecerá al sistema operativo fundamental de toda la civilización digital.

VII. Por qué la capacidad de construir sistemas no puede obtenerse directamente mediante grandes cantidades de tokens

Al argumentar que «la capacidad de construir sistemas es un activo escaso», necesito enfrentar una posible objeción: dado que la IA es tan poderosa, ¿por qué no se puede «generar» u «obtener» la capacidad de construir sistemas

consumiendo grandes cantidades de tokens? La respuesta a esta pregunta revela los límites fundamentales de la capacidad de los tokens.

Primero, la construcción de sistemas requiere decisiones, y las decisiones no son generación.

La IA puede generar un artículo, un código, una plantilla de sitio web, pero la IA no puede reemplazar las decisiones centrales en el diseño de sistemas que los humanos deben tomar — elegir qué módulos, abandonar qué funciones, equilibrar las necesidades actuales con la expansión futura, cuándo reestructurar, cuándo mantener el status quo. Estas decisiones dependen de la comprensión de la esencia del negocio, del juicio de las restricciones de recursos y de la previsión de las consecuencias a largo plazo. Los tokens pueden generar innumerables «soluciones posibles», pero elegir «qué solución merece la pena implementar» y «por qué» es algo que los tokens no pueden reemplazar. Mi sistema logístico inteligente ha sufrido innumerables ajustes en dos décadas, y cada ajuste no fue porque una IA me dijera «esto es mejor», sino porque surgió un problema real en el negocio real, exponiendo limitaciones y planteando nuevas demandas. Detrás de estas decisiones está el juicio, y el juicio proviene de la práctica, no de los tokens.

Segundo, la construcción de sistemas requiere manejar excepciones, y las excepciones no pueden enumerarse exhaustivamente.

La razón por la que un sistema puede funcionar a largo plazo no es porque su flujo normal esté perfectamente diseñado, sino porque puede manejar las excepciones. En un sistema logístico, las excepciones incluyen: direcciones incorrectas de los clientes, mercancías dañadas, retrasos aduaneros, fallos de pago, caídas del sistema, pérdida de datos. En un sistema de edición, las excepciones incluyen: fluctuación en la calidad de la traducción, incompatibilidad de los formatos de metadatos, cambios en las interfaces de archivo, fallos de registro de DOI. Estas excepciones no pueden enumerarse exhaustivamente porque la variedad de excepciones en el mundo real es infinita. La IA puede aprender «patrones de excepción comunes» a través de grandes cantidades de tokens, pero no puede

predecir «la próxima excepción que nunca ha aparecido antes». Solo un sistema que funciona desde hace mucho tiempo, encontrando, registrando y resolviendo continuamente excepciones en la práctica, puede desarrollar una verdadera capacidad de manejo de excepciones. Esta capacidad no puede obtenerse «alimentando más tokens», porque las excepciones no son conocimiento; son experiencia.

Tercero, la construcción de sistemas requiere compensaciones, y las compensaciones no tienen respuestas estándar.

La construcción de cada sistema implica compensaciones bajo múltiples restricciones: coste frente a rendimiento, velocidad frente a estabilidad, generalidad frente a personalización, necesidades inmediatas frente a expansión futura. Estas compensaciones no tienen una «respuesta correcta», sino «elecciones razonables en circunstancias específicas». La IA puede proporcionar una «solución óptima», pero este «óptimo» es estadísticamente óptimo basado en datos históricos, no un juicio basado en «su contexto específico». La razón por la que mi sistema integrado logístico y financiero de dos décadas ha funcionado hasta hoy es precisamente porque hice diferentes compensaciones en diferentes épocas — eligiendo cierta pila tecnológica en 2005, eligiendo otro método de integración en 2015, y eligiendo una ruta de actualización en 2025. Estas compensaciones no son «correctas» o «incorrectas», sino «adecuadas para el contexto de la época». Los tokens no pueden hacer estas compensaciones por mí, porque la esencia de las compensaciones es un juicio de valor, no un cálculo.

Cuarto, la construcción de sistemas requiere bucles de retroalimentación, y la retroalimentación requiere el mundo real.

La mejora de la capacidad de construir sistemas depende de un bucle cerrado: diseño → implementación → funcionamiento → detección de problemas → corrección → rediseño. El paso de «detección de problemas» en este bucle debe provenir de la retroalimentación del mundo real. La IA puede simular el funcionamiento, pero no puede simular todos los factores impredecibles del mundo real. La razón por la que

mi sistema logístico ha mejorado continuamente es que clientes reales proporcionaron retroalimentación real; la razón por la que mi sistema de edición ha mejorado continuamente es que interfaces de archivo reales experimentaron cambios reales. Esta retroalimentación no puede obtenerse consumiendo tokens en un entorno cerrado. Los tokens pueden generar «retroalimentación hipotética de los clientes», pero no pueden reemplazar el hecho de que «el cliente realmente dio esta retroalimentación». La retroalimentación del mundo real es el único suelo para el crecimiento de la capacidad de construir sistemas.

Quinto, la construcción de sistemas requiere confianza, y la confianza solo puede venir del tiempo.

La confiabilidad de un sistema no se demuestra con la promoción, sino con el tiempo. La razón por la que mi sitio web está archivado permanentemente por CERN Zenodo, WorldCat y la Biblioteca Nacional de Australia no es porque les «caigo bien», sino porque estas instituciones, a través de la observación a largo plazo, han confirmado que mi sistema es estable, fiable y persistente. Esta confianza no puede comprarse con grandes cantidades de tokens. La IA puede generar un manual de sistema «de aspecto muy profesional», pero no puede generar el hecho de que «este sistema ha sido confiable durante dos décadas». La confianza es un producto del tiempo, no un producto de los tokens.

En resumen, la razón por la que la capacidad de construir sistemas no puede obtenerse directamente mediante grandes cantidades de tokens es que no es esencialmente una capacidad generativa, sino más bien una capacidad de juicio, manejo de excepciones, compensación, absorción de retroalimentación y acumulación de confianza. Lo que estas capacidades tienen en común es que requieren tiempo, práctica, errores, correcciones y validación en el mundo real. Los tokens pueden acelerar el procesamiento de la información, pero no pueden acelerar la acumulación de experiencia en el mundo real. Esta es la razón fundamental por la que, incluso si los tokens se vuelven extremadamente baratos en el futuro, la capacidad de construir sistemas seguirá siendo escasa.

VIII. Mis activos: datos históricos reales, sistemas transversales, capacidad de construcción, visión prospectiva y capacidad de autoverificación

Antes de discutir más a fondo «lo que es verdaderamente escaso», necesito explicar primero mi propia posición. Porque mis juicios no se derivan de la deducción teórica, sino de la práctica de larga duración. Esto es muy evidente en los diversos sistemas que he construido durante un largo período. Ya sea el sistema logístico inteligente, el sistema de facturación inteligente, el sistema de procesamiento financiero automatizado, o más tarde el sistema de edición en diez idiomas, el sistema de metadatos, el sistema de navegación del sitio web y el sistema de referencias cruzadas automáticas, la parte realmente difícil nunca ha sido generar información, sino organizarla. La información puede crecer indefinidamente, pero sin estructura, solo creará un nuevo caos. Lo mismo ocurre en la era de los tokens. En el futuro, la IA podrá generar cantidades masivas de contenido, pero cantidades masivas de contenido no significan automáticamente cantidades masivas de valor. Al contrario, a medida que la generación de contenido se vuelve más fácil, saber cómo filtrar, organizar, preservar, indexar y relacionar este contenido se volverá cada vez más importante.

Poseo los siguientes cinco activos, que constituyen la base empírica de mi juicio sobre «lo que es escaso después de los tokens».

Primero, datos históricos reales. Estos no son unos «datos históricos» cualquiera, sino veinte años de datos comerciales reales transgeneracionales — desde el sistema de facturación inteligente de 2005 hasta el presente, provenientes de sistemas logísticos, financieros y fiscales reales en funcionamiento, no de entornos experimentales o de datos raspados. En la era de la IA, con la proliferación de los «datos sintéticos», el valor de los datos históricos reales está aumentando rápidamente. La mayoría de las empresas poseen «datos históricos de un solo dominio», mientras que yo poseo datos históricos transversales, que abarcan veinte años y están validados prácticamente.

Segundo, la finalización de múltiples sistemas y arquitecturas transversales. Mis sistemas cubren los siguientes dominios: sistema logístico inteligente, sistema de facturación inteligente, sistema financiero inteligente, sistema fiscal inteligente, sistema de integración logística y financiera de dos décadas, sistema de automatización de sitios web, sistema de generación de metadatos, sistema de edición en diez idiomas de *The Epochal Transition*, sistema de gestión de DOI, sistema de organización de enlaces TROVE y WorldCat, sistema de referencias cruzadas automáticas, sistema de generación de Sitemap, sistema de generación de perfil Zenodo, sistema de edición multiplataforma. La singularidad de estos sistemas reside en: la amplitud transversal (desde la logística hasta la edición, desde las finanzas hasta la indexación de literatura), la profundidad temporal (algunos sistemas han estado funcionando durante dos décadas), y la finalización por una sola persona (demostrando la transferibilidad del método). Entre ellos, el mensual *The Epochal Transition* (<https://times.net.au>) sirve como evidencia pública — posee el número ISSN internacional normalizado (3083-5178), está archivado permanentemente por WorldCat (el sistema de catálogo colectivo de bibliotecas más grande del mundo, con instituciones miembro que superan las 16 000 bibliotecas nacionales y universitarias en más de 100 países y regiones), el PANDORA/TROVE de la Biblioteca Nacional de Australia, así como Zenodo (gestionado por CERN, la Organización Europea para la Investigación Nuclear), OpenAIRE (la infraestructura europea para la ciencia abierta), DataCite DOI (identificador de objeto digital global, con más de 60 documentos), y ORCID (el identificador digital académico permanente único global). Este sitio web demuestra que una sola persona (un investigador independiente) puede realizar todo el proceso, desde la creación de contenido hasta la interfaz con las infraestructuras de archivo académico de primer nivel.

Tercero, la capacidad de construir sistemas y arquitecturas. Esta no es la capacidad de «usar herramientas de IA», sino la capacidad de diseño para transformar requisitos vagos en estructuras estables, la capacidad arquitectónica para integrar diferentes sistemas en una arquitectura unificada, la capacidad de ingeniería para mantener sistemas funcionando durante años sin intervención humana, y la meta-

capacidad para transferir la experiencia de la logística a la edición, de las finanzas a la literatura. Como se discutió anteriormente, esta capacidad no puede obtenerse directamente mediante grandes cantidades de tokens porque depende de la toma de decisiones, el manejo de excepciones, las compensaciones, los bucles de retroalimentación y la acumulación de confianza — ninguno de estos puede ser reemplazado por la IA. El mercado actual de formación en IA se centra en habilidades superficiales como la «ingeniería de indicaciones», la «escritura con IA» y el «dibujo con IA»; casi ningún curso enseña la «capacidad de construir sistemas transversales» — porque esta capacidad no puede adquirirse rápidamente; solo puede acumularse mediante una práctica de larga duración. Las más de 40 obras creativas independientes listadas en el sitio web, el ISSN obtenido y los más de 60 DOI son los mejores registros públicos de la capacidad de funcionamiento a largo plazo del sistema, y no una autoproclamación.

Cuarto, una visión orientada al futuro. Comenzar a experimentar con un sistema prototipo de trabajo remoto en 1993 — casi treinta años antes de que el trabajo remoto a gran escala se generalizara; diseñar un sistema de facturación inteligente en 2005 — comenzar la automatización de procesos veinte años antes de la democratización de la IA; la integración transgeneracional de dos décadas — prever en el momento del diseño del sistema que todavía necesitaría funcionar veinte años después; proponer ahora el juicio de «después de los tokens» — en un momento en que la mayoría de la gente todavía persigue los tokens, prever que la capacidad de los sistemas es la verdadera barrera. También prevé que «los tokens son similares a las monedas, y las diferencias tecnológicas, sistémicas y políticas entre Oriente y Occidente los hacen no universales» — así como el dólar estadounidense y el renminbi chino no pueden intercambiarse libremente, los tokens chinos y los tokens ingleses enfrentan barreras de circulación similares, y en el futuro, podrían surgir «zonas de tokens chinos», «zonas de tokens ingleses» y otros ecosistemas de tokens basados en el idioma y la política. También prevé que «ganar dinero con los tokens ya es difícil y lo será cada vez más; solo cuando hay un sistema frontal que proporciona servicios, los tokens y la potencia de cálculo pueden servir como activos complementarios efectivos para generar beneficios». Como investigador

independiente, establecí con antelación un sistema de edición personal que se interconecta con la infraestructura académica futura, previendo la tendencia de que «un individuo también puede convertirse en un nodo en la infraestructura global del conocimiento». La mayoría de las predicciones de tendencias provienen de analistas o empresarios, mientras que mi «capacidad de previsión» tiene la validación de dos décadas de práctica como garantía.

Quinto, la capacidad de autoverificación. Esta es una de mis meta-capacidades más importantes — todos los escritos que produzco en diversos campos logran un cierre teórico y una autocoherencia lógica. Esto significa que mis argumentos no dependen de una autoridad externa o citas, sino de la autosuficiencia de la lógica interna. La regla por defecto de la literatura académica es «citar investigaciones anteriores»; pocas personas reclaman o demuestran una «capacidad de autoverificación»; la discusión académica se centra en «si la argumentación es suficiente», y no en «si el autor tiene la capacidad de construir un bucle cerrado coherente». En mis artículos, esta capacidad de autoverificación se manifiesta en: los argumentos provienen de la práctica, la práctica valida los argumentos, y los argumentos y la práctica forman un bucle cerrado. No estoy «citando las teorías de otros para probarme», sino «mostrando los sistemas que he construido y dejando que los sistemas hablen por sí mismos». Esta capacidad de autoverificación es un concepto completamente ausente en la literatura pública y una de mis meta-capacidades más únicas.

La combinación de estos cinco activos, su escasez no es aditiva, sino multiplicativa: datos históricos reales × sistemas transversales × capacidad de construcción de sistemas × visión prospectiva × capacidad de autoverificación constituye una cartera de activos irreproducible. En las discusiones públicas, actualmente no se puede encontrar ningún caso comparable.

El proceso de sistematización de *El camino de un investigador independiente internacional* me ha hecho darme cuenta de esto aún más profundamente. De 1989 a 2026, más de treinta años de experiencia abarcan múltiples campos. Estas experiencias no surgieron todas a la vez, sino que se formaron gradualmente a lo

largo de un largo período. La IA puede generar un artículo en segundos, pero no puede generar más de treinta años de experiencia real en segundos; puede generar muchas opiniones, pero no puede generar directamente el juicio acumulado a través de la práctica a largo plazo; puede simular muchos procesos, pero no puede reemplazar la validación repetida del mundo real. En este sentido, el tiempo mismo es también un activo importante, y la estructura precipitada por el tiempo es un activo aún más raro.

IX. Después de los tokens: cinco capacidades verdaderamente escasas

Cuando los tokens en sí mismos ya no sean escasos — la barrera para vender tokens no es alta, pero habrá una falta extrema de datos históricos reales, de capacidad para construir sistemas y arquitecturas, y de visión orientada al futuro — ¿qué será verdaderamente escaso? Basándome en mi práctica a largo plazo, desde el prototipo de sistema de trabajo remoto de 1993, el sistema de facturación inteligente de 2005, el sistema logístico inteligente, el sistema financiero inteligente, el sistema fiscal inteligente, el sistema de integración logística y financiera de dos décadas, hasta la automatización de sitios web, la generación de metadatos, el mensual en diez idiomas *The Epochal Transition*, la gestión de DOI, la organización de enlaces TROVE y WorldCat, las referencias cruzadas automáticas, la generación de Sitemap, la generación de perfil Zenodo, la edición multiplataforma, creo que las siguientes cinco capacidades se convertirán en verdaderas barreras.

Primero, la capacidad de diseño estructural. Los tokens son la materia prima, y la estructura es el marco que organiza la materia prima en productos valiosos. Mucho antes de que la IA se generalizara, ya había empezado a pensar en cómo utilizar recursos limitados para lograr una mayor eficiencia operativa. Mi sistema logístico inteligente, mi sistema de facturación inteligente y mi sistema logístico y financiero inteligente de dos décadas nacieron en diferentes épocas, enfrentando diferentes problemas, pero comparten una característica común: el objetivo no era crear más

información, sino reducir el caos, reducir el trabajo repetitivo, mejorar la eficiencia global y establecer estructuras que pudieran funcionar a largo plazo. En el sistema de edición automatizado, el problema central al que me enfrentaba no era «si el texto podía generarse», sino «cómo definir los campos de metadatos para diez idiomas, cómo establecer asociaciones entre artículos y categorías, y cómo diseñar reglas de gestión de versiones». Estas decisiones de diseño estructural determinan si el sistema puede seguir funcionando durante años sin colapsar. En contraste, la cantidad de tokens consumidos es meramente un subproducto. El sitio web *The Epochal Transition* es en sí mismo una evidencia pública de la capacidad de diseño estructural: una estructura de sitio web académico mantenida personalmente, con un ISSN internacional normalizado, soporte multilingüe, múltiples canales de archivo, demostrando que no estoy haciendo discursos vacíos, sino que ya he construido con éxito un sistema complejo y accesible al público utilizando esta capacidad de diseño estructural.

Segundo, la capacidad de automatización. Los tokens pueden producir contenido en masa, pero solo los sistemas de automatización pueden completar en masa flujos de trabajo de extremo a extremo. En el sistema logístico inteligente, el valor real no reside en cuántas veces se llama a la IA para planificar rutas, sino en el establecimiento de un bucle cerrado completamente automatizado, desde la recepción de pedidos, la planificación del almacén, la optimización de rutas, hasta el manejo de excepciones. En este bucle cerrado, la IA es solo un componente, y la capacidad de automatización se manifiesta en las conexiones entre los componentes. Una persona sin sistema, incluso con una gran cantidad de tokens, puede que solo genere palabras dispersas, imágenes dispersas, tablas dispersas y códigos dispersos; una persona con un sistema, en cambio, puede integrar los tokens en su flujo de trabajo, permitiéndoles organizar automáticamente materiales, generar metadatos, verificar enlaces, actualizar páginas web, ayudar en la edición, archivar literatura, crear índices, procesar logística, analizar finanzas y generar informes.

Tercero, la experiencia práctica transversal. Los tokens están disponibles para todos, pero la capacidad de transferir la experiencia de una industria a otra no es universal. La razón por la que puedo construir tanto un sistema logístico como un sistema de edición no es porque tenga más tokens, sino porque he acumulado metodologías transferibles como la gestión de metadatos, el diseño de flujos de trabajo y el manejo de excepciones. Mi sistema de edición necesita tokens porque la traducción a diez idiomas, la generación de resúmenes, la organización de palabras clave, la generación de metadatos y las descripciones de enlaces requieren un procesamiento lingüístico importante. Mi sistema de sitio web necesita tokens porque las estructuras de página, las descripciones de navegación, las descripciones SEO, los metadatos académicos y las presentaciones multilingües pueden generarse automáticamente. Mi sistema logístico también puede usar tokens porque los correos electrónicos de los clientes, las descripciones de mercancías, los documentos aduaneros, las explicaciones de excepciones, las notas financieras y los registros de operaciones pueden ser asistidos por la IA. Mi sistema de referencias cruzadas automáticas necesita aún más tokens porque debe encontrar asociaciones temáticas entre un gran número de artículos, títulos, DOI, enlaces NLA, TROVE y WorldCat. Esta acumulación práctica transversal no puede obtenerse consumiendo tokens; solo puede decantarse resolviendo problemas reales. El sitio web presenta simultáneamente información de múltiples dominios, incluyendo la edición académica, la integración de infraestructuras, la arquitectura técnica y la marca personal, demostrando que he conectado efectivamente la edición, la biblioteconomía, la tecnología y la marca personal, en lugar de poseer una habilidad única.

Cuarto, el modelo mental para el funcionamiento a largo plazo de los sistemas. La narrativa dominante para las aplicaciones actuales de la IA es la «conversación» y la «generación», pero lo que realmente produce valor sostenido es el «funcionamiento». Una conversación genera 10 000 tokens, una tarea de agente consume un millón, pero un sistema logístico que ha funcionado continuamente durante una década puede consumir miles de millones de tokens — pero lo que es verdaderamente escaso es el modelo mental para diseñar este sistema: cómo

garantizar que el sistema siga siendo utilizable después de las actualizaciones del modelo? Cómo manejar los cambios en las API ascendentes? Cómo monitorear las excepciones y recuperarse automáticamente? Estas cuestiones no tienen nada que ver con los tokens, pero determinan la producción final de los tokens. El valor de estos sistemas no proviene de los tokens mismos, sino de las estructuras que ya he construido o estoy desarrollando. Los tokens son solo el combustible; el sistema es el motor. Los tokens son solo la corriente eléctrica; la estructura es el circuito. Los tokens son solo el agua; las tuberías, las válvulas, los depósitos y los sistemas de despacho determinan si puede desempeñar un papel. Sin un sistema, los tokens se disiparán; con un sistema, los tokens pueden convertirse en valor a largo plazo. El sitio web muestra más de 40 obras creativas independientes y un archivo permanente por instituciones autorizadas como el CERN y la Biblioteca Nacional de Australia, demostrando que mi sistema no persigue el bombo publicitario a corto plazo, sino el valor a largo plazo, y ya ha recibido el respaldo de «preservación permanente» de instituciones autorizadas.

Sin embargo, cuando considero más a fondo la relación entre los tokens y los sistemas, encuentro otro problema a menudo pasado por alto. Mucha gente cree que con un excelente sistema podrán obtener una ventaja en la futura era de los tokens. Pero la realidad no es tan simple. Los sistemas en sí mismos no están al alcance de cualquiera. Los sistemas verdaderamente maduros a menudo requieren práctica a largo plazo, revisiones repetidas y validación en el mundo real. Muchos sistemas pueden tardar años, incluso décadas, de acumulación continua, desde la idea inicial hasta la implementación efectiva. Por lo tanto, la capacidad para diseñar sistemas es ya un recurso escaso.

Pero incluso poseyendo un sistema maduro, esto no conduce automáticamente a un impacto a gran escala. Entre la viabilidad técnica de un sistema y su difusión en la sociedad se encuentran múltiples capas que incluyen los mercados, los capitales, los canales de distribución, la educación de los usuarios y los ecosistemas de plataformas. La historia no carece de buenas tecnologías o buenos sistemas; lo que es verdaderamente escaso es la capacidad de plataforma para llevar un sistema a

una aplicación a gran escala. Incluso si un sistema es muy superior a las soluciones existentes en un escenario local, sin suficiente apoyo de capital, capacidad de promoción de plataforma y base de usuarios, su escala de uso real puede seguir siendo muy limitada.

Desde la perspectiva de la economía de los tokens, esto es particularmente evidente. La demanda de tokens no proviene directamente del sistema mismo, sino de la escala real de uso del sistema. Una plataforma que sirve a millones de usuarios al día, incluso con un diseño de sistema relativamente ordinario, puede generar un consumo de tokens enorme; a la inversa, un sistema avanzado con muy pocos usuarios puede tener una demanda de tokens muy baja. Por lo tanto, el desarrollo futuro del mercado de tokens depende no solo de la capacidad de los modelos y de la capacidad de los sistemas, sino también de la capacidad de las plataformas y de la capacidad de integración de recursos.

Esto significa que la competencia futura tiene al menos tres niveles. El primer nivel es la competencia de los modelos, determinando quién puede suministrar tokens. El segundo nivel es la competencia de los sistemas, determinando quién puede utilizar eficazmente los tokens. El tercer nivel es la competencia de las plataformas, determinando cuántos usuarios puede alcanzar un sistema y cuánta demanda real puede generar. Modelos, sistemas y plataformas son todos indispensables. Sin modelos, los sistemas no pueden funcionar. Sin sistemas, los tokens no pueden crear valor. Sin plataformas, incluso el mejor sistema puede permanecer en una escala de uso restringido.

Desde esta perspectiva, estoy aún más convencido de que los tokens son en sí mismos meros recursos fundamentales. Lo que realmente determina el futuro no son solo aquellos que poseen tokens, ni solo aquellos que poseen sistemas, sino aquellos que pueden conectar modelos, sistemas, plataformas y necesidades del mundo real. Solo cuando estos elementos forman un bucle cerrado, los tokens se transforman de simples unidades de cálculo en fuerzas productivas creadoras de valor.

Quinto, la capacidad de integrar los tokens en los procesos comerciales del mundo real. Los tokens deben servir finalmente a un eslabón del mundo real: logística, edición, educación, salud. Lo que es verdaderamente escaso son aquellos que entienden tanto los límites de las capacidades de la IA como los puntos débiles de una industria específica. Saben cuándo usar la IA y cuándo no; saben cómo convertir las salidas de los tokens en decisiones comerciales; saben cómo medir la relación real entre las entradas de tokens y las salidas comerciales. Esta capacidad compuesta es actualmente el recurso más escaso en el mercado. El sitio web es en sí mismo un «proceso comercial» en funcionamiento: creación de contenido → generación de metadatos → publicación multiplataforma → archivo internacional → acceso de los lectores. Esto demuestra que mi aplicación de la IA (consumo de tokens) se ha integrado en un proceso de edición completo, en bucle cerrado, que genera un impacto real.

Siento cada vez más que la verdadera competencia en el futuro probablemente no será entre modelos, ni entre tokens, sino entre sistemas. Los modelos seguirán evolucionando, los precios de los tokens seguirán bajando, y los recursos caros hoy se volverán muy comunes en el futuro. Pero los sistemas no aparecerán automáticamente porque los recursos sean abundantes. Los sistemas requieren una acumulación a largo plazo, revisiones continuas y validación en el mundo real. Quien pueda convertir los tokens en sistemas que funcionen a largo plazo tendrá más probabilidades de obtener un valor duradero. Y esta es también la clave de la rentabilidad: ganar dinero con los tokens ya es difícil, y lo será cada vez más; solo cuando hay un sistema frontal que proporciona servicios, los tokens y la potencia de cálculo pueden servir como activos complementarios efectivos para generar beneficios.

Juzgo que la verdadera competencia en la era de los tokens se desarrollará en tres etapas: la primera etapa es la competencia de recursos, todos compitiendo por modelos, potencia de cálculo, API y precios de tokens; la segunda etapa es la competencia de aplicaciones, todos usando tokens para escribir, servicio al cliente, código, traducción, educación, trabajo de oficina y análisis de datos; la tercera etapa

es la verdadera competencia de sistemas, donde quien pueda integrar los tokens en sistemas industriales, sistemas de edición, sistemas logísticos, sistemas de conocimiento y sistemas de registro de la civilización que funcionen a largo plazo tendrá probablemente una ventaja duradera.

De hecho, cuando se discuten los tokens, los modelos, los sistemas y las plataformas, hay otro factor importante a menudo pasado por alto en el mundo real: el tiempo. Mucha gente piensa que con capital, pueden construir rápidamente sistemas; con un mercado suficientemente grande, pueden replicar rápidamente modelos de éxito. Sin embargo, la realidad a menudo no es así. El capital puede comprar potencia de cálculo, puede comprar tokens, puede contratar equipos de desarrollo, pero no puede comprar directamente datos reales y sistemas maduros que han sido validados durante muchos años. Para muchas industrias, lo que es verdaderamente escaso no son los modelos en sí mismos, ni los tokens en sí mismos, sino los sistemas que han funcionado de manera estable en entornos comerciales reales durante un largo período.

Desde una perspectiva comercial, los mercados y los capitales generalmente no esperan a que un sistema completamente nuevo se desarrolle desde cero. Un gran sistema, desde la propuesta del concepto, el diseño de la arquitectura, el desarrollo de las funciones, la validación de los procesos hasta la implementación efectiva, a menudo lleva muchos años o incluso más. Durante este proceso, no solo se requiere una inversión continua en recursos, sino también muchas pruebas y errores, revisiones y validación en el mundo real. Por lo tanto, cuando las oportunidades de mercado ya han aparecido, los capitales tienden a preferir los sistemas ya formados, ya en funcionamiento y ya validados, en lugar de diseñar un sistema entero desde cero.

Sin embargo, hay otra contradicción en la realidad. Cuanto más grande es un sistema, más difícil es que los capitales posteriores se involucren fácilmente. Después de un desarrollo a largo plazo, los grandes sistemas suelen tener ya formados sus propios mecanismos de funcionamiento, estructuras de gestión, lógicas de datos y culturas organizativas. Pueden tener vastos datos históricos,

procesos comerciales complejos y sistemas de experiencia acumulados durante largo tiempo. Debido a esto, aunque los grandes sistemas son extremadamente valiosos, a menudo es difícil transformarlos rápidamente o reestructurarlos completamente según las expectativas de los capitales. Cuanto más grande es el sistema, mayor tiende a ser su inercia.

Por lo tanto, en el mundo real de los negocios, el estado óptimo no es diseñar un sistema desde cero, ni adquirir un gran sistema rígido, sino encontrar sistemas que ya han completado su construcción estructural básica, que han sido validados a largo plazo y que pueden funcionar de manera estable. Incluso si estos sistemas no son perfectos, tienen un valor muy alto porque ya han superado la etapa más difícil, poseyendo datos reales, procesos reales y registros de validación reales. Sobre esta base, el reempaquetado, la optimización, la apertura al mundo exterior y la promoción por plataformas son a menudo más eficientes que empezar desde cero, y también pueden reducir considerablemente los riesgos de prueba y error y los costes de tiempo. Porque estos sistemas ya han realizado la etapa más costosa, más lenta y más difícil de replicar.

Desde esta perspectiva, la ventana temporal es también un recurso a menudo pasado por alto. Cuando una industria está en sus inicios, mucha gente se centra en las capacidades de los modelos, los precios de los tokens y el tamaño del mercado, ignorando el tiempo necesario para la formación de los sistemas. El nacimiento de un sistema maduro suele pasar por varias etapas: propuesta del concepto, diseño de la arquitectura, validación de los procesos, funcionamiento en el mundo real y optimización continua. La parte realmente difícil no es proponer la idea, sino mantener el sistema funcionando de manera estable en un entorno real a largo plazo. Por lo tanto, cuando el mercado comienza a reconocer el valor de los sistemas maduros, a menudo significa que estos sistemas ya han sufrido años de acumulación, en lugar de ser algo que puede ser replicado de nuevo en poco tiempo.

Para los capitales y el mercado, los objetivos más ideales no son las ideas en la etapa conceptual, ni los prototipos que no han sido validados, sino los sistemas

embrionarios que han completado la construcción estructural básica, poseen registros de funcionamiento reales y tienen margen de mejora. Aunque estos sistemas pueden no ser perfectos, ya han superado la etapa más difícil, más lenta y más arriesgada. Una vez que la demanda del mercado aumenta rápidamente, los sistemas como este a menudo se vuelven muy buscados, porque el número de sistemas maduros realmente validados a largo plazo es generalmente mucho menor de lo que el mercado imagina. Los modelos pueden iterar rápidamente, los tokens pueden desarrollarse continuamente, los equipos de desarrollo pueden ensamblarse temporalmente, pero los datos reales, los procesos reales y los registros de validación a largo plazo no pueden recrearse en poco tiempo.

Por lo tanto, en el desarrollo continuo de la economía de los tokens, el tiempo no solo afecta la formación de los sistemas, sino que también afecta el valor de los sistemas. Cuando el mercado aún no ha reconocido plenamente la importancia de un cierto tipo de sistema, a menudo es también el período en que el coste de adquisición de estos sistemas es más bajo; y cuando el mercado generalmente se da cuenta de su valor, las opciones disponibles a menudo ya son muy limitadas. Para muchas industrias, lo que se pierde no es solo un proyecto, sino una ventana temporal para entrar en la siguiente etapa del desarrollo que ya ha completado una acumulación a largo plazo.

Desde la perspectiva de la economía de los tokens, ocurre lo mismo. Los tokens pueden obtenerse rápidamente, los modelos pueden utilizarse rápidamente, las plataformas pueden construirse rápidamente, pero lo que es realmente difícil de replicar rápidamente son los sistemas que han decantado durante muchos años y pueden continuar produciendo valor. En el futuro, cuando los tokens se vuelvan cada vez más abundantes, la importancia de los sistemas no solo no disminuirá, sino que incluso puede aumentar aún más. Porque cuando todos tengan recursos de tokens similares, lo que determinará el resultado de la competencia ya no será la cantidad de tokens, sino quién posee datos reales más maduros, procesos comerciales más estables y estructuras de sistema más cercanas a las necesidades del mundo real.

Otro problema a menudo pasado por alto es el origen de los sistemas. Mucha gente tiene la costumbre de entender los sistemas como productos de software, y por lo tanto suponen naturalmente que los sistemas deben ser diseñados y desarrollados por empresas de software. Sin embargo, en muchas industrias complejas, los sistemas que funcionan verdaderamente de manera estable a largo plazo a menudo no son primero diseñados por empresas de software y luego dados a la industria, sino que se forman gradualmente en el curso de una práctica comercial de larga duración, y luego se convierten en software, se automatizan y se transforman en plataformas.

La razón no es compleja. Las empresas de software suelen sobresalir en la implementación técnica, pero no necesariamente permanecen en el contexto comercial específico durante mucho tiempo. Pueden desarrollar funciones rápidamente, pero tienen dificultades para adquirir una gran cantidad de experiencia comercial real en poco tiempo. En cambio, los sistemas que se desarrollan orgánicamente dentro de una industria se revisan y mejoran continuamente frente a los clientes, pedidos, excepciones, cambios regulatorios, presiones de costes y competencia del mercado a diario. Estos sistemas no parten de la teoría, sino de problemas reales; cada uno de sus procesos, cada una de sus reglas, cada eslabón de automatización proviene típicamente de la validación a largo plazo del trabajo real.

Por lo tanto, en la logística, las cadenas de suministro, la edición, las finanzas, la fabricación, la salud y muchos otros campos profesionales, los sistemas más valiosos a menudo no son los más complejos, ni los que tienen más funciones, sino aquellos que han sido probados a largo plazo y pueden resolver establemente problemas reales. Pueden no tener las interfaces más avanzadas, la promoción más llamativa o la arquitectura técnica más reciente, pero poseen grandes cantidades de datos reales, procesos reales y lógica comercial acumulada a largo plazo. Esta experiencia a menudo no puede obtenerse mediante un desarrollo a corto plazo.

Desde la perspectiva de la economía de los tokens, esto es particularmente importante. En el futuro, las capacidades de los modelos pueden seguir mejorando,

los costes de los tokens pueden seguir bajando, y la eficiencia del desarrollo de software también será cada vez mayor. Sin embargo, la experiencia comercial acumulada por una industria durante décadas no surgirá automáticamente debido a las actualizaciones de los modelos. Lo que es realmente difícil de replicar no es el código del software en sí mismo, sino el conocimiento sectorial, la experiencia en el manejo de excepciones, la lógica de toma de decisiones y los procesos comerciales validados a largo plazo que se esconden detrás del sistema.

Por lo tanto, creo que los sistemas más ideales en el futuro no serán aquellos diseñados completamente desligados de la práctica industrial, sino aquellos que provienen de una actividad real, han sido validados mediante un funcionamiento a largo plazo, y luego son mejorados continuamente con la ayuda de la inteligencia artificial y las capacidades de los tokens. Estos sistemas poseen tanto la fiabilidad del mundo real como las ganancias de eficiencia traídas por la era de la IA, y su valor a largo plazo es a menudo mucho mayor que el de aquellos sistemas contruidos únicamente sobre conceptos técnicos.

X. Hacia dónde fluyen los tokens: de las conversaciones transitorias a la memoria de la civilización

Con respecto a la futura economía de los tokens, tengo otra observación. Mucha gente se centra en cómo se producen los tokens, pero poca atención se presta a su destino final. Si un token solo permanece en una sola conversación, su vida útil puede ser de solo unos minutos; si entra en una página web ordinaria, su vida útil puede ser de solo unos meses; si entra en un sistema de preservación a largo plazo, su valor cambia fundamentalmente. Cuando los tokens entran en los DOI, Zenodo, WorldCat, TROVE, los sistemas de preservación a largo plazo de las bibliotecas nacionales y diversas infraestructuras académicas y de conocimiento, ya no son meramente tokens; comienzan a transformarse en registros de conocimiento, activos documentales y memoria de la civilización. Por lo tanto, creo que la verdadera cuestión importante en el futuro no es cómo generar más tokens, sino

cómo permitir que los tokens entren en estructuras donde puedan ser preservados a largo plazo y continúen desempeñando su papel.

Es por esta razón que sigo siendo optimista sobre el futuro desarrollo de los tokens, pero no venero los tokens. El pasado de los tokens es la unidad básica para que los ordenadores procesen el lenguaje; el presente de los tokens es una unidad de medida y una unidad comercial importantes en la era de la IA; el futuro de los tokens es probable que se convierta en un recurso fundamental importante para la civilización digital. Sin embargo, sea cual sea la evolución de los tokens, siempre sostengo que lo que es verdaderamente escaso no son los tokens, sino las personas capaces de dominar los tokens; no aquellos que tienen más recursos, sino aquellos que pueden establecer estructuras a largo plazo; no aquellos que pueden generar más contenido, sino aquellos que pueden organizar el contenido en sistemas de conocimiento.

XI. La incompresibilidad del tiempo: cuatro activos incomprables

Al discutir «lo que es verdaderamente escaso después de los tokens», necesito señalar un punto fundamental que la mayoría de las discusiones pasan por alto: ciertos activos son intrínsecamente incapaces de ser acelerados, comprados o replicados. Por muy poderosa que sea la IA, por muy baratos que sean los tokens, los siguientes cuatro activos siempre conservarán su escasez inherente.

Primero, el tiempo es incompresible.

Treinta años de práctica son simplemente treinta años; no pueden comprimirse en tres años, ni simularse por la IA. La IA puede generar un artículo sobre «cómo construir un sistema logístico» en segundos, pero no puede vivir, en segundos, las diversas anomalías, fallos, cambios de política y fluctuaciones del mercado que un sistema logístico encuentra en dos décadas. Estas acumulaciones de tiempo en el mundo real no pueden acortarse por ningún medio tecnológico. Mi inicio de

experimentación con un prototipo de sistema de trabajo remoto en 1993, el diseño de un sistema de facturación inteligente en 2005 — estos puntos en el tiempo no son accidentes; son marcadores irreversibles de la historia real. Quien venga más tarde puede leer mis artículos y aprender mis métodos, pero no puede poseer los años que he vivido — esos años en sí mismos son activos.

Segundo, la experiencia transversal no se puede comprar.

La experiencia transversal no se puede comprar como una mercancía. Una persona puede comprar cien libros sobre logística, edición, finanzas e indexación de literatura en un solo día, pero no puede adquirir, ese mismo día, la experiencia práctica de conectar estos dominios. La verdadera capacidad transversal no es la acumulación de conocimiento, sino la capacidad de establecer conexiones entre diferentes dominios, identificar puntos comunes y transferir métodos. Esta capacidad solo puede adquirirse resolviendo problemas reales, cometiendo errores, corrigiendo caminos y acumulando juicio en un dominio tras otro. La razón por la que puedo transferir mi experiencia de los sistemas logísticos a los sistemas de edición no es porque haya leído un cierto libro, sino porque pasé veinte años en logística y otros diez en edición — es en la intersección de estos dos períodos donde surgió la metodología transferible.

Tercero, la capacidad de construir sistemas es difícil de replicar.

La capacidad de construir sistemas no es un conjunto de códigos o procesos que se puedan copiar y pegar. El código puede copiarse, pero la lógica de toma de decisiones detrás del código, los juicios de compromiso y la previsión de los cambios futuros no pueden replicarse. El valor de un sistema no reside solo en lo que puede hacer hoy, sino también en cuántos cambios, recuperaciones de fallos y ajustes de requisitos ha sufrido en la última década mientras continuaba funcionando. Esta «historia del sistema» es en sí misma una prueba de la capacidad de construcción. Mi sitio web (<https://times.net.au>) puede ser visto por cualquiera, pero lo que los visitantes ven es el resultado, no las tres décadas de historial de toma de decisiones, elecciones tecnológicas, ajustes estructurales y mantenimiento

continuo detrás de ese resultado. Esta capacidad de construcción no puede obtenerse mediante plagio o imitación.

Cuarto, los registros de funcionamiento a largo plazo no pueden acelerarse.

El registro de que un sistema ha funcionado durante dos décadas es en sí mismo intrínsecamente incapaz de ser acelerado. Un registro de funcionamiento a largo plazo prueba la estabilidad, la mantenibilidad y la previsión de diseño del sistema. En la era de la IA, generar un sitio web o sistema de «apariencia» puede hacerse en minutos, pero generar un «sistema que ha funcionado durante dos décadas y está archivado permanentemente por el CERN, WorldCat y la Biblioteca Nacional de Australia» no tiene atajos. Un registro de funcionamiento a largo plazo es una marca de tiempo, declarando a todos que este sistema ha resistido la prueba del mundo real. Tal registro no puede ser falsificado o acelerado por ninguna tecnología de IA.

Estos cuatro activos — el tiempo, la experiencia transversal, la capacidad de construir sistemas y los registros de funcionamiento a largo plazo — constituyen juntos la base de lo que he sostenido en este artículo como «verdaderamente escaso». Los poseo no porque sea más inteligente o más trabajador que los demás, sino porque comencé antes que los demás y he continuado hasta el presente. En una era en la que la IA está haciendo que cada vez más cosas sean baratas y fácilmente accesibles, estas cosas que no pueden ser comprimidas, compradas, replicadas o aceleradas se volverán cada vez más valiosas.

XII. Conclusión: De la «acumulación de tokens» a la «construcción de sistemas»

El futuro de los tokens no reside en los tokens mismos.

Con el desarrollo continuo de la tecnología de los grandes modelos y la continua disminución de los costes de inferencia, la oferta de tokens seguirá expandiéndose. Los tokens que son caros hoy serán tan baratos y omnipresentes como el tráfico de Internet y el almacenamiento en disco duro en unos pocos años. Cuando todo el

mundo pueda obtener fácilmente grandes cantidades de tokens, lo que será verdaderamente escaso ya no serán los tokens, sino la capacidad de convertir los tokens en valor en el mundo real — capacidad de diseño estructural, capacidad de automatización, experiencia industrial y acumulación práctica a largo plazo. Y la escasez más profunda reside en aquellas cosas que no pueden ser comprimidas, compradas, replicadas o aceleradas: el tiempo mismo, la experiencia transversal, la capacidad de construir sistemas y los registros de funcionamiento a largo plazo. Como se argumentó anteriormente, la capacidad de construir sistemas no puede obtenerse directamente mediante grandes cantidades de tokens porque no es esencialmente una capacidad generativa, sino más bien una capacidad de juicio, manejo de excepciones, compensación, absorción de retroalimentación y acumulación de confianza — la característica común de estas capacidades es que requieren tiempo, práctica, errores, correcciones y validación en el mundo real.

En el plano de la rentabilidad, este juicio se aplica igualmente: ganar dinero con los tokens ya es difícil, y lo será cada vez más. Las guerras de precios comprimen continuamente los márgenes de beneficio, y el modelo de negocio de la simple reventa de tokens se encamina rápidamente hacia beneficios reducidos o incluso nulos. Solo cuando los tokens se adjuntan a un sistema frontal — es decir, cuando un sistema proporciona servicios — pueden los tokens y la potencia de cálculo servir como activos complementarios efectivos para generar beneficios. Los tokens en sí mismos no son la fuente del beneficio; el sistema lo es. Los tokens son solo el combustible del sistema, no su motor.

Desde esta perspectiva, los sistemas que ya he completado o que estoy desarrollando tienen un significado especial para el futuro desarrollo de los tokens. Porque estos sistemas no se construyeron temporalmente para seguir una moda; existían mucho antes de que el concepto de tokens se convirtiera en un tema público. El sistema logístico inteligente demostró que la actividad del mundo real puede ser estructurada, procedimentalizada y automatizada; el sistema de edición demostró que un individuo también puede establecer un sistema documental multilingüe, multiplataforma, multi-DOI, multi-archivo; el sistema de sitio web y

metadatos demostró que el contenido digital puede ser organizado, indexado y difundido continuamente; el sistema de referencias cruzadas automáticas demuestra además que cuando el número de obras aumenta continuamente, lo que realmente importa no es seguir acumulando contenido, sino establecer relaciones estructurales entre los contenidos. El sitio web *The Epochal Transition* (<https://times.net.au>), como plataforma de visualización centralizada de estos logros, es en sí mismo una evidencia empírica de la conversión de la capacidad de los tokens en un sistema académico abierto, permanente y verificable.

Incluso si todos recibieran la misma cantidad de tokens en el futuro, los resultados producidos podrían ser radicalmente diferentes: algunos obtendrían más registros de conversación, algunos obtendrían más contenido de páginas web, mientras que otros podrían establecer un sistema logístico, un sistema de edición o un sistema de conocimiento que funcione continuamente durante décadas. Lo que marca la diferencia no son los tokens en sí mismos, sino la arquitectura del sistema que los subyace — y el tiempo incompresible, la experiencia transversal incompresible, la capacidad de construcción difícil de replicar y los registros de funcionamiento a largo plazo imposibles de acelerar que sustentan esas arquitecturas de sistema.

Por lo tanto, no considero los tokens como un fin, sino como la nueva infraestructura de la era inteligente. Los tokens son tan importantes como la electricidad, pero la electricidad por sí sola no construye automáticamente fábricas; los tokens son tan medibles como el dinero, pero el dinero por sí solo no crea automáticamente activos de calidad; los tokens pueden fluir como el agua, pero sin tuberías, sistemas de despacho y registros de mantenimiento a largo plazo, solo se disiparán. Las personas verdaderamente valiosas en el futuro no serán necesariamente las que tengan más tokens, sino las que puedan organizar los tokens en sistemas. El mayor error de la era de los tokens es considerar los tokens como un objetivo; la mayor oportunidad de la era de los tokens es usar los tokens para construir estructuras que otros no puedan replicar. Cuando los tokens se vuelvan tan comunes como la electricidad en el futuro, la gente mirará hacia atrás en la locura actual por los tokens y quizás descubra que lo que realmente determinó la brecha entre las

personas nunca fue la cantidad de tokens que poseían, sino si poseían la capacidad de aprender continuamente, practicar continuamente, acumular continuamente y construir continuamente sistemas — y esas cosas detrás de estas capacidades que no pueden ser comprimidas, compradas, replicadas o aceleradas. Y en mi opinión, esta capacidad es la riqueza más escasa y más difícil de replicar en la era de los tokens.

A medida que mi investigación se profundiza, siento cada vez más que el panorama competitivo de la futura era de la IA podría ser resumido por una analogía simple. Mucha gente considera la potencia de cálculo, los modelos y los tokens como recursos centrales, pero desde la perspectiva del funcionamiento del mundo real, se parecen más al sistema logístico detrás de un ejército. La potencia de cálculo determina la capacidad de producción, los modelos determinan la capacidad de organización, y los tokens determinan la capacidad de distribución; juntos constituyen el sistema de apoyo material de la era de la IA.

Sin embargo, lo que realmente determina el resultado de una batalla no son los suministros logísticos en sí mismos, sino el sistema de combate de primera línea. Para la futura competencia industrial, los sistemas son más como la línea de ataque. Los sistemas determinan cómo se utilizan los recursos, cómo se ejecutan los procesos, cómo se crea el valor y si los usuarios se benefician verdaderamente. Por muy abundantes que sean los tokens y la potencia de cálculo, sin la organización de un sistema, son como suministros apilados en un almacén, incapaces de convertirse en resultados reales.

Desde esta perspectiva, la competencia futura probablemente no será entre potencias de cálculo, ni entre tokens, sino entre sistemas. El sistema es la línea de ataque, determinando la dirección del ataque y la capacidad de combate; los tokens son las municiones y los suministros, determinando la capacidad de combate sostenible; la potencia de cálculo es la base industrial trasera, determinando la velocidad de producción y la capacidad de suministro. Sin suministro trasero, la línea de ataque no puede avanzar durante mucho tiempo; pero con solo el

suministro trasero y ningún sistema de línea de ataque, tampoco se pueden obtener resultados.

Por lo tanto, a medida que la potencia de cálculo se generaliza, los modelos se vuelven más similares y los tokens se vuelven más abundantes, lo que sigue siendo verdaderamente escaso es siempre el sistema. Quien posea un sistema más maduro, datos más reales, procesos validados durante más tiempo, y la capacidad de conectar los tokens y la potencia de cálculo con las necesidades del mundo real, será más probable que obtenga una ventaja duradera en el futuro.

Los tokens se están convirtiendo en la «electricidad» de la era de la IA — indispensables, pero no un fin en sí mismos. Así como los ganadores de la revolución eléctrica no fueron las centrales eléctricas, sino los inventores y empresarios que inventaron la bombilla, el motor y la cuba electrolítica, los verdaderos ganadores de la era de la IA serán los constructores de sistemas capaces de utilizar los tokens para construir estructuras y arquitecturas irreplicables. Después de los tokens, eso es lo que es verdaderamente escaso.

Al revisar todo este artículo, siento cada vez más que el panorama competitivo de la futura era de la IA podría ser ilustrado por la analogía de un árbol.

Mucha gente se centra en el agua y los nutrientes, porque son visibles, tangibles y fácilmente medibles. La potencia de cálculo, los modelos y los tokens de hoy son como el agua y los nutrientes continuamente entregados. La potencia de cálculo determina la capacidad de producción, los modelos determinan la capacidad de transformación, y los tokens determinan la capacidad de distribución. A medida que la tecnología avanza, estos recursos se vuelven cada vez más abundantes y fáciles de obtener.

Sin embargo, lo que realmente determina la vitalidad de un árbol no es el agua y los nutrientes en sí mismos, sino el sistema radicular.

El sistema es la raíz de la futura civilización digital.

El sistema radicular determina cómo se absorbe el agua, cómo se distribuyen los nutrientes, y si todo el árbol puede crecer a largo plazo. Sin raíces, por mucha agua que haya, se escurrirá; sin raíces, por muy abundantes que sean los nutrientes, no pueden transformarse en vitalidad. Del mismo modo, sin un sistema, por muy potente que sea el modelo, por muy baratos que sean los tokens, por muy vasta que sea la potencia de cálculo, es difícil crear valor de manera continua.

Desde esta perspectiva, la potencia de cálculo es el agua, los tokens son los nutrientes, los modelos son las tuberías que entregan los nutrientes, y el sistema es el sistema radicular que penetra profundamente en el mundo real. Lo que realmente determina el resultado de la competencia futura no es quién tiene más agua, ni quién tiene más nutrientes, sino quién tiene raíces más profundas, más estables, capaces de absorber y utilizar estos recursos a largo plazo.

En el futuro, a medida que la potencia de cálculo siga creciendo, los modelos sigan evolucionando y los precios de los tokens sigan bajando, los recursos en sí mismos se volverán cada vez más comunes. Lo que será verdaderamente escaso ya no serán los recursos, sino los sistemas capaces de organizar, utilizar y crear continuamente valor a partir de los recursos. Quien posea un sistema maduro posee las raíces; quien posee las raíces posee la capacidad de supervivencia a largo plazo y de crecimiento continuo.

Por lo tanto, creo que lo que es verdaderamente escaso después de los tokens no son más tokens, ni más potencia de cálculo, sino los sistemas capaces de convertir la potencia de cálculo, los modelos y los tokens en valor en el mundo real.

Cuando el agua y los nutrientes se vuelven cada vez más abundantes, lo que realmente determina si un árbol puede crecer a largo plazo es siempre el sistema radicular.

Palabras clave: Token, Inteligencia Artificial, IA, Gran Modelo de Lenguaje, Potencia de Cálculo, Economía de los Modelos, Economía de los Tokens, Token Chino, Token Inglés, Algoritmo BPE, Soberanía de los Datos, Agente, Ecosistema de API, Capacidad de Construcción de Sistemas, Diseño de Sistemas, Sistema de

Automatización, Investigación Transdisciplinaria, Práctica Transdisciplinaria, Datos Históricos Reales, Sistema de Larga Duración, Activo Temporal, Competencia de Plataformas, Competencia de Sistemas, Civilización Digital, Memoria de la Civilización, Infraestructura de Conocimiento, Investigador Independiente, Investigador Independiente Internacional, Sistema Logístico Inteligente, Sistema de Edición en Diez Idiomas, Sistema de Metadatos, Referencias Cruzadas Automáticas, DOI, Zenodo, WorldCat, TROVE, The Epochal Transition, Wu Zhaohui, JEFFI CHAO HUI WU

Referencias:

1. *A Structural Case Study of the 1993 Remote Work Prototype System*
DOI: <https://doi.org/10.5281/zenodo.17978371>
 2. *The 2005 Intelligent Invoice System*
DOI: <https://doi.org/10.5281/zenodo.19044755>
 3. *Mature Structural Logistics System*
DOI: <https://doi.org/10.5281/zenodo.20201895>
 4. **Cross-Generational 20-Year Intelligent Logistics and Financial System**
DOI: <https://doi.org/10.5281/zenodo.20352371>
NLA: <https://nla.gov.au/nla.obj-4204311557>
 5. *Analysis of the Homepage Structural Evolution of The Epochal Transition*
DOI: <https://doi.org/10.5281/zenodo.20362574>
NLA: <https://nla.gov.au/nla.obj-4204309341>
 6. *An Empirical Study of Institutional Engineering: Independent Scholars Accessing the WorldCat Global Bibliographic Infrastructure*
DOI: <https://doi.org/10.5281/zenodo.18028572>
 7. *National-level Network Archiving and Digital Document Preservation Guidelines of the Australian Winner Information Network*
DOI: <https://doi.org/10.5281/zenodo.17888259>
 8. *The Institutional Archiving of The Epochal Transition*
DOI: <https://doi.org/10.5281/zenodo.17932379>
- WU, J. C. H. (2026). Source Declaration for Audiovisual and Derivative Adaptations of a Continuing Real-World Narrative. Zenodo.
<https://doi.org/10.5281/zenodo.18160116>

トークン以後：AI 時代に真に希少なものは

著者 / 国際独立研究者：巫朝晖（ジェフイ・チャオ・フイ・ウー）

ORCID: <https://orcid.org/0009-0009-5318-3027>

【前置きの声明】

本稿は、30 年以上にわたる私の学際的实践（1989 年のオーストラリア留学開始以降、リモートワーク、物流管理、自動処理、ウェブサイト構築、文化発信、多言語出版、学際的研究に及ぶ）に基づく個人的見解である。トークンは急速に進化する新しい現象であり、本稿はあくまで執筆時点（2026 年 6 月）における個人的な段階的観察と判断を表すもので、投資助言、商業予測、業界基準を構成するものではない。

本稿は普遍的な適用可能性を追求するものではなく、いかなる業界のコンセンサスを代表するものでもない。トークン市場、AI 技術、世界の政策環境の継続的な変化に伴い、本稿の判断の一部は将来修正や更新が必要となる可能性がある。読者は最新の情報に基づき独自に判断されたい。

私が論じる「五つの希少な能力」および「五つの個人的資産」は、私自身のシステム構築実践に由来する（公開された証拠は <https://times.net.au> で入手可能。同サイトに掲載された 40 以上の独立創作作品、ISSN 3083-5178、60 以上の DOI、そして WorldCat、オーストラリア国立図書館、CERN Zenodo などの国際的インフラによって永久保存された記録を含む）。

文脈の切り取りを避けるため、本稿に対する効果的な反駁や疑問提示の前提条件は、本稿内の全てのリンク先記事とそのネストされたリンク、および <https://times.net.au> に掲載された私の全作品を完全に読むことである。この全体的な文脈から切り離された部分的な引用や単純化された要約は、本稿の公正な評価を構成し得ない。

効果的な反駁に関する補足説明：

本稿は個人的見解であり、普遍的な適用可能性を追求せず、他者の文献引用を前提としない。いかなる効果的な疑問も、リンゴ同士の比較 (APPLE TO APPLE) を採用すべきである：事実対事実、個人の実践対個人の実践、完成したシステム対完成したシステム。「将来の AI が達成し得る」ことを以て「現在完了している実践」を否定すること、あるいはチーム・機関の能力を個人の実践と比較することは、対等な比較とはならない。

本稿で述べる 5 つの資産ポートフォリオは個人の実践記録であり、「対照となる事例は全く存在しない」と主張するものではない。完全に対照となる事例が存在すると考える場合は、具体的な公開事例を挙げ、それが現実の履歴データ、分野横断的システム、システム構築能力、未来を予見するビジョン、自己実証能力という 5 つの次元においてどのような対応する証拠を持っているかを説明されたい。そうでなければ、関連する疑問は単なる意見の表明に過ぎず、効果的な反駁とはならない。

要旨

トークンは AI 時代の中核的計測単位になりつつある。しかし本稿は、トークンそれ自体は最終的な価値ではないと論じる。推論コストの継続的な低下に伴い、トークンは今日のネットワークトラフィックやハードディスクストレージと同様に安価で一般的なものになるだろう。トークンがもはや希少でなくなったとき、真に希少になるものは何か？著者の 30 年以上にわたる学際的实践 (1993 年のリモートワークプロトタイプシステム、2005 年のインテリジェント請求書システムから、インテリジェント物流システム、10 言語出版システム、DOI 管理、自動相互参照、クロスプラットフォーム出版に至る) に基づき、本稿は以下の核心的判断を提示する：第一に、システム構築能力 (構造設計、自動化、分野横断的实践、長期運用のメンタルモデル、トークンを現実のビジネスプロセスに埋め込む能力を含む)こそが真に希少な資産である。第二に、システム構築能力は、大量のトークンを通じて直接獲得することはできない。なぜならそれは、意思決定、例外処理、トレードオフ、現実世界からのフィードバックループ、そして時間をかけて蓄積された信頼に依存するからである。第三に、時間の非圧縮性、分野横断的経験の非購買性、システム構築能力の複製困難性、長期運用記録の短期間での達成不可能性、これら四つの資産がトークン時代の最も深い希少性を構成する。本稿は、著者が構築した学術ウェブサイト『The Epochal Transition』 (<https://times.net.au/>)、ISSN 3083-5178、

WorldCat、CERN Zenodo、オーストラリア国立図書館 TROVE などの国際的インフラによって永久保存)を公開証拠として提示し、一人の人間がどのようにして 30 年かけて分野横断的で検証可能な長期運用システム体系を構築できるかを示す。

本稿の主張： トークンは単なる燃料であり、システムこそがエンジンである。トークン以後、真に希少なものはトークンを扱いこなせる人間、そしてその人間が築き上げた構造である。

「物品なくして、何のための金か？」

トークンの発展の歴史を振り返るとき、それが人類が過去に経験した多くの技術革命と非常に似ていると痛感する。新しい基礎的リソースが登場するたびに、人々はまずリソースそのものに焦点を当て、リソースの背後にある構造を無視する。産業時代もそうであり、インターネット時代もそうであり、人工知能の時代も同じである。今日、多くの人々が議論するのはトークンの価格、トークンのコスト、トークンの割り当て量、トークンの供給能力であり、まるでより多くのトークンを持つことが未来を持つことのように思われている。しかし私の数十年にわたる実践経験から言えば、リソースそのものが最終的な価値であったことは一度もない。価値を真に決定づけるのは、常に人々がこれらのリソースをどのように組織し活用するかである。

「トークン」を理解するには、新しい用語から始めるのではなく、その歴史から始めなければならない。1906 年にアメリカの哲学者チャールズ・サンダース・パースが「the」という単語の 20 のインスタンスを紙に書き、「型」と「実例」を区別して以来、2026 年に中国国家データ局局長の劉烈宏がボアオ・フォーラムで Token に「詞元」(ciyuan)という中国語訳を正式に定めるまで、この概念は哲学からコンピュータへ、ブロックチェーンから人工知能へと長い旅を歩んできた。Token という言葉は西洋のコンピュータ世界に長く存在し、プログラミング言語、コンパイラ、テキスト処理システムにおいて、トークンはコンピュータが認識・処理できる基本単位を指していた。その後、ネットワークセキュリティ、認証、API 呼び出し、ブロックチェーンの文脈では、「トークン」はアクセス権限、身分証明書、デジタル資産を示す「パス」や「credential」として理解されることが多かった。しかし大規模言語モデルの時代になると、Token はもはや主に身分証明書でもブロッ

クチェーン資産でもなく、AI モデルが言語、コード、画像記述、マルチモーダル情報を処理する際の基本的な計算単位となる。

中国は後に AI 分野の Token を「詞元」と命名した。これはかなり正確な翻訳であると思う。「令牌」(パス)は用いられなかった。なぜなら大規模モデルにおける Token は通行証ではないからだ。また単純に「字」や「詞」と翻訳されなかったのは、それが一字である場合もあれば、一語である場合もあり、英単語の半分である場合もあり、さらには句読点や記号である場合もあるからだ。「詞」は言語や情報との関連を示し、「元」は基本単位であることを示す。一般読者には、人間が見ているのは一文であり、人工知能が見ているのはトークンの列である、と単純に理解して差し支えない。この道筋において、トークンは、言語学の冷門用語から AI 時代の中核的計測単位へ、そして技術的供給と商業的需要を結ぶ決済ツールへと変貌を遂げた。インテリジェント物流システムの構築、自動化出版プロセスの設計、クロスプラットフォームコンテンツ配信などの長期プロジェクトを経験した実践者として、私はこの変遷を肌で感じている。トークンは決して目的ではなく手段であり、真に価値を決定づけるのは、誰により多くのトークンがあるかではなく、誰がトークンを使って持続的に稼働するシステムと複製不可能な構造を構築できるかである。本稿で論じるシステム構築能力の成果は、<https://times.net.au> で公開されている。これは、独立研究者によって構築され、世界の図書館システムに永久保存された分野横断的学術出版システムである。

自身の歩みを振り返ると、私はトークンが登場した後にシステムについて考え始めたのではなく、長期的な実践を通じて徐々にシステムへの理解を深めてきた。1989年のオーストラリア留学開始以来、私は様々な分野での学習、仕事、探求を経験してきた。その後、リモートワーク、物流管理、自動処理、ウェブサイト構築、文化発信、多言語出版、学際的研究に徐々に関わるようになった。振り返ってみると、これらの経験は時代は異なるものの、共通する特徴を持っている。それは常に、どのように構造を確立し、どのように反復労働を減らし、どのように全体的な効率を高めるか、ということを中心に展開してきたということである。当時はトークンという概念も今日のような大規模言語モデルも存在しなかったが、私はすでに情報を自動的に流し、自動的に組織し、自動的に処理する方法について絶えず考えていた。

したがって、今日ますます多くの人々がトークンを新たな生産手段と見なす傾向にある中で、私はそれを新たな基礎的エネルギーと見なすことに傾いている。エネルギーは確かに重要だが、エネルギーは自動的に価値を生み出すわけではない。電気自体が自動的に工場になるわけではなく、インターネット自体が自動的に企業になるわけでもなく、トークン自体が自動的に知識体系になるわけでもない。将来、誰もが大量のトークンを簡単に入手できるようになったとしても、それが誰もが同じ成果を生み出せることを意味するわけではない。差異を決定するのはトークンの量ではなく、トークンがどのような構造に入り込むかである。

トークンの詳細な分析に入る前に、もう一つ重要な観察点を指摘しておく必要がある。トークンは通貨に類似しており、東西の技術、システム、政策の違いにより、普遍的に通用しない。米ドルと人民元が自由に交換できないのと同様に、中国語トークンと英語トークンも同様の流通障壁に直面している。これは単なる言語の差異の問題ではない。その深層的理由は、東西の技術レベル（モデルアーキテクチャ、トークン化アルゴリズム、訓練データ）、システムレベル（APIエコシステム、決済システム、データ標準）、政策レベル（データ越境流動、輸出規制、国家安全保障審査）における三つの差異にあり、これらがトークンを理想的な「世界的に通用するAI通貨」のように自由に流通させることを妨げているからである。この観察は、トークン市場の将来像を理解する上で重要な意味を持つ。将来、「中国語トークン圏」、「英語トークン圏」など、言語と政策に基づくトークンエコシステム圏が出現し、圏を越えた呼び出しは効率の損失とコンプライアンスコストに直面するだろう。

一、トークンの淵源：哲学的概念から経済単位への四重の飛躍

トークンのアイデンティティの進化は、明確に四つの段階に分けることができる。

哲学言語学の原点。 1906年、パースは記号理論を研究する際に、「型」と「実例」を区別した。前者は抽象的な規則や形式を表し、後者はその規則の現実における具体的な現れである。この区別は後に構造言語学の重要な基礎となった。

コンピュータ科学におけるコーディングへの転回。 1960 年代、プログラマーが「int x = 5;」というコード行を書くとき、コンピュータはそれを「int」、「x」、「=」、「5」などの独立した単位に分解する必要がある、そのような各单位がトークンであった。トークンはこれにより、機械が命令を「読み取る」ための最小情報単位となった。

ブロックチェーンにおける価値媒体の実験。 2017 年、ICO ブームの高まりとともに、Token には「流通可能なデジタル権益証書」という新たなアイデンティティが与えられた。その熱狂は次第に冷めたものの、「Token は価値の象徴である」という認識はそのまま定着した。

AI 時代の経済インフラ。 2026 年 3 月は画期的な瞬間であった。NVIDIA の CEO ジェンセン・フアンはデータセンターを「インテリジェントな AI トークンを生産する工場」と再定義し、中国国家データ局局長の劉烈宏は「詞元」という中国語訳を公式に与え、「インテリジェント時代の価値アンカー」と呼んだ。この時点で、トークンは技術的概念から経済的概念への最後の飛躍を遂げた。

二、技術的基盤：BPE アルゴリズムと暗黙の「言語税」

トークンが AI の標準的計測単位となり得たのは、20 年以上も忘れられていた技術、バイト対符号化 (Byte Pair Encoding、BPE) に依存している。

1994 年、アメリカのプログラマー、フィリップ・ゲイジは C 言語技術雑誌にデータ圧縮アルゴリズムに関する記事を発表し、BPE の基本原理を紹介した。テキストを繰り返しスキャンし、最も頻繁に隣接する 2 文字を「溶接」して新しい記号とし、これを繰り返し圧縮していくというものである。この論文は圧縮効率が際立っていなかったため、当時は注目を集めなかった。

2016 年になって、エディンバラ大学の研究者リコ・センのリッヒが、機械翻訳における分かち書き問題を研究に、偶然この古い論文を発見し、BPE が分かち書き問題に対する完璧な解決策であることを鋭く見抜いた。辞書を事前に定義する必要はなく、データ自身に「語らせ」、高頻度の組み合わせを自然にトークンへと凝縮させることができるのである。2019 年、OpenAI は GPT-2 をリリースする際にこのア

アプローチを正式に採用し、トークン化の起点を直接「バイト」レベルに設定することで、モデルが理論上あらゆる言語を処理できるようにした。

しかし、BPE アルゴリズムの「頻度優先」の論理は、暗黙の「言語税」体系を意図せず形成してしまった。英語はインターネットコーパスの絶対的主流であるため、英語トークンの分割効率は他の言語よりもはるかに高い。同じ意味を表現するのに、英語は最も少ないトークンで済み、中国語は通常 1.5 倍から 2 倍、さらにリソースの少ない言語では英語の 5 倍から 10 倍のコストになることもある。これは、トークン単位の課金ルールの下では、異なる言語で AI と会話する実際のコストに大きな差が生じることを意味する。

このような「スタートライン」における不公平は、モデルの初期語彙に一旦書き込まれてしまうと、修正は非常に困難である。トークン化ルールは、AI が世界を認識するための「地基」であり、建物が高くなればなるほど、地基を交換することはできない。幸いなことに、この状況は改善されつつある。GPT シリーズを例にとると、同じ中国語の文が GPT-3 では 38 トークン必要だったのが、GPT-4 では 26 トークンに、GPT-5 では 15 トークンにまで減少しており、処理効率は 60% 以上向上している。また、DeepSeek、通義千問 (Tongyi Qianwen) などの中国製大規模モデルは、設計当初から中国語の高頻度フレーズや成語などをネイティブトークンとして語彙表に組み込んでおり、中国語に対してより「母語」レベルの処理効率を実現している。これは深い法則を明らかにしている。すなわち、「意味分割権」を掌握する者が、その言語のデジタル世界における表現効率とコスト優位性を大きく掌握するのであり、これはデジタル時代の「基礎的鑄造権」なのである。

三、用途と産出：対話からエージェントへの指数関数的飛躍

トークンの消費シナリオは、「対話」から「推論」、そして「エージェント」への三段階の飛躍を経験している。

対話型 AI の段階では、単一のプロンプトと応答は約 1 万トークンを消費する。推論型 AI の段階では、複数ラウンドの相互作用と複雑な推論が伴い、各タスクは約 10 万トークンを消費する。エージェント型 AI の段階に入ると、エージェントは多段階タスクを自律的に実行し、複数のツールを呼び出すことができ、各タスクのトークン

ン消費量は約 100 万トークンに達する。わずか 2 世代の進化で約 100 倍に拡大した。クアルコム CEO のアモン氏は、2030 年までに世界の 10 秒間のトークン需要は、2026 年の 317 億トークンから 1.27 兆トークンに急増すると予測している。これは誇張ではない。OpenClaw などのオープンソースエージェントツールの爆発的な人気を証明している。単一の複雑なタスクで数千万トークンを消費することは珍しくない。

産出の観点から見ると、中国は世界最大のトークン供給国となっている。公開データによると、2026 年 3 月現在、中国の 1 日あたりのトークン呼び出し量は 140 兆トークンを超え、2 年間で千倍以上増加し、世界の AI トークン使用量の 61% を占めている。この地位の確立は、3 つの優位性の重ね合わせに依存している。第一に、中国の工業用電力価格は欧米よりはるかに低く、トークン生産コストに占める電力の割合は 60% を超える。第二に、中国は世界最大規模のインテリジェントコンピューティングクラスターを建設し、「東数西算」（東部のデータを西部で計算）の配置により効率的な出力を実現している。第三に、国産大規模モデルが全面的にオープンソース化され、API の商業化サービス体系が成熟しており、「モデル訓練-計算資源支援-推論出力-世界規模での呼び出し」という完全な産業チェーンの閉ループを形成している。

四、機関と市場：三極構造の産業図譜

トークンを中心に、明確に階層化された市場構造が形成されている。

この産業構造を理解するにあたっては、トークンの背後にある基盤システムを見ることも必要である。多くの人々はトークンの価格に注目するが、トークンそれ自体が独立して存在するリソースではないことを見落としている。トークンの背後にはモデルがあり、モデルの背後には計算資源がある。ある意味で、計算資源がモデル能力を決定し、モデルがトークンの産出を決定し、トークンはモデル能力が市場に向けて出力される際に形成される計測単位である。したがって、トークン市場は表面的には価格競争であるが、底流では依然として計算資源競争と効率競争なのである。

現在の世界のトークン産業チェーンは、おおよそ3つのレベルに分けることができる。最上流は計算資源インフラであり、チップメーカー、サーバーメーカー、クラウドコンピューティングプラットフォーム、データセンター、エネルギーシステムを含む。中流は基盤モデル機関と推論プラットフォームであり、これらは計算資源を呼び出し可能なモデル能力に変換する責任を負う。下流はトークンサービスプロバイダー、エージェントプラットフォーム、および各種業界応用機関であり、これらはモデル能力をさらに企業サービスや最終製品へと変換する。ユーザーが最終的に目にするのはトークンの価格であるが、トークンのコストと速度に真に影響を与えるのは、往々にして上流の計算資源体系とモデル効率の変化である。

過去数年間、チップ性能の継続的な向上、モデルアーキテクチャの最適化、推論効率の継続的な改善に伴い、トークン単価は急速な低下傾向を示している。多くの人々はトークンが安くなっていると見ているが、この変化を真に推進しているのは計算資源体系の継続的な進歩である。したがって、将来のトークン市場の競争は、プラットフォーム間競争や価格競争だけでなく、計算資源、モデル効率、資源統合能力の間の競争でもある。

上流の発電所 – 計算資源生産者。 NVIDIA に代表されるチップ企業は、ハードウェアを継続的に反復進化させ、計算資源単位あたりのトークン産出効率を高めている。NVIDIA は 2027 年までに売上高を少なくとも 1 兆ドルにすると予測しているが、その自信の源は世界の AI 計算資源への渴望である。

中流の送電網制御 – モデルプラットフォームとアグリゲーター。 OpenAI、Google、DeepSeek、アリババなどの企業は、基盤となる計算資源を標準化され呼び出し可能なトークンサービスに変換する。この部分を中心に、3つの成熟したビジネスモデルが形成されている。トークンアグリゲーションプラットフォーム（OpenRouter に代表される、価格差と手数料で運営、2026 年 5 月時点で週間呼び出し量は 26.9 兆トークン）、クラウドベンダーによる MaaS サービス、そして二次市場で活発な AI 中継局である。

下流の電力入口 – 応用・配信プラットフォーム。 テンセントが微信（WeChat）インターフェースを開放して新モデル能力を接続することは、本質的に微信を「トークン配信プラットフォーム」に変えることである。中国電信は「トークン経営によ

る事業再構築」を提案し、従来のトラフィックロジックから脱却しようとしている。これらの入口企業に共通する特徴は、ユーザーがもはや直接モデルに直面するのではなく、プラットフォームを通じて間接的にトークンを呼び出すという点であり、モデル自体が「不可視化」しつつある。

特筆すべきは、トークンを中心に独自のグローバル貿易のパラダイムが形成されていることである。無形での輸出、非課税での運用、グローバルなカバレッジである。これにより、中国は従来の「世界の工場」から「世界の AI 計算資源インフラプロバイダー」へと変貌を遂げつつある。

トークン経済を議論する際、計算資源という避けて通れない基礎的要因がある。多くの人々はトークンの価格に注目するが、トークンの背後にある生産プロセスを見落としている。実際、トークンは空中から発生するものではない。トークンが生成されるたびに、モデルが推論を行う必要があり、モデルの推論は大量の計算資源の上に成り立っている。ある意味で、計算資源はトークン経済の源泉であり、モデルはトークンの生産ツールであり、トークンはモデル能力が外部に出力される際に形成される計測単位である。

したがって、計算資源レベルはトークンコストとトークン生成速度に直接影響する。計算資源が強力であればあるほど、モデルの情報処理能力は高くなり、単位時間あたりに生成できるトークン数は多くなる。計算資源のコストが低ければ低いほど、トークン価格も一般的に低くなる。過去数年間、大規模モデル業界ではトークン価格の継続的な低下が見られたが、その背後にはモデルアルゴリズムの最適化だけでなく、チップ性能の向上、データセンターの規模拡大、推論効率の向上などの要因がある。言い換えれば、多くの人々はトークンが安くなっていると思っているが、この変化を真に推進しているのは計算資源体系の継続的な進歩なのである。

産業チェーンの観点から見ると、計算資源、モデル、トークンは実際に完全なチェーンを構成している。上流はチップ、サーバー、データセンター、エネルギーシステムである。中流は基盤モデルと推論プラットフォームである。下流はトークンサービス、エージェントアプリケーション、および各種業界ソリューションである。計算資源はモデル能力の上限を決定し、モデルはトークンの品質と効率を決定し、

トークンは最終的にユーザーがどれだけの人工知能サービスを得られるかを決定する。

しかし、計算資源の重要性は論を俟たないが、計算資源それ自体が自動的に価値を生み出すわけではない。歴史的に多くの産業が同様の過程を経験してきた。より大きな発電所を持っていることがより優れた企業を持っていることを意味するわけではなく、より多くのサーバーを持っていることがより優れた製品を持っていることを意味するわけでもない。同様に、より多くの計算資源を持っていることがより高価値のシステムを持っていることを意味するわけでもない。計算資源は生産能力の問題を解決し、システムは価値変換の問題を解決する。将来、計算資源が増え続け、モデル能力が強まり続け、トークン資源が豊かになり続けるにつれて、真に希少になるのは依然として計算資源それ自体ではなく、これらの資源を現実世界で長期的に稼働するシステムへと組織化する方法である。

したがって、将来の競争の観点から、私は計算資源競争、モデル競争、トークン競争は最終的に徐々にシステム競争へと移行していくと考える。なぜなら、計算資源がますます普及し、モデルがますます接近し、トークンがますます豊かになったとき、差異を決定するのはもはや保有資源量ではなく、誰がこれらの資源を利用して長期的に有効な構造を構築できるかだからである。

五、競争と参入障壁：価格戦の内巻きと価値の階層化

トークン市場の競争は、高度な内巻き（インボリューション）の特徴を示している。2022 年から 2026 年にかけて、単回トークン推論コストは 99.9%減少した。国内モデルの価格優位性はさらに顕著である。DeepSeek V4-Pro の入力価格は 100 万トークンあたり 0.025 人民元（約 0.5 円）と低く、通義千問 Qwen3.5-Plus の出力価格は 100 万トークンあたり 1.6 人民元（約 32 円）である。これに対して、OpenAI GPT-4o の出力価格は約 72 人民元（約 1,440 円）である。

このような価格競争により、「仲介業者」の利益幅は急激に圧縮されている。業界統計によると、各種 API アグリゲーターと中継プラットフォームの数は 2,000 を超え、市場は 3 つの大きな階層を形成している。トップ層は技術力が強くモデルも充実した大手プラットフォーム、中堅層は公式割引と顧客リソースを持つ専門代理

店、そして多数の小規模代理店は厳しい生存圧力に直面している。単純なトークン価格差での儲けのビジネスモデルはもはや持続可能ではない。

真の参入障壁は、二つの方向に移行しつつある。一つは技術的プレミアムの方向である。独自開発の推論高速化エンジンにより、公式よりも安く販売しても利益を出すことができる。二つ目は付加価値サービスの方向である。プロンプトエンジニアリング、モデル微調整、プライベートデプロイメントなどのカスタマイズサービスを提供する。さらに注目すべきは、トークンの価値がシナリオによって大きく分化していることである。雑談系トークンの平均価格は 100 万あたり 0.01 ドル (約 1.5 円) にもかかわらず、創薬系トークンの平均価格は 100 万あたり 1000 ドル (約 15 万円) にも達し、シナリオ間の価値差は 10 万倍にもなる。これは本質的な法則を明らかにしている。トークンの価値はその生産コストによって決定されるのではなく、その用途によって決定されるのである。同じトークンでも、異なるシナリオで使用され、異なるシステムに組み込まれると、生み出す価値は天と地ほども違う。

しかし、ここでより重要な判断をしなければならない。現在そして将来において、トークンを販売するための参入障壁は高くない。誰でも API に登録し、アグリゲーションプラットフォームを構築し、トークンの再販を始めることができる。真に希少なものは、トークンを販売すること以外のもの、すなわち現実の履歴データ、システムや体系を構築する能力、未来を予見するビジョンである。これらがなければ、いかなる大規模モデルもトークンも活用されない。

トークンの収益性に関する補足判断：現時点でトークンで儲けることはすでに容易ではなく、ますます難しくなっている。価格競争は利益幅を圧縮し続けており、単純なトークン転売のビジネスモデルは急速に低利益、さらには無利益へと向かっている。トークンがフロントエンドシステム、すなわちサービスを提供するシステムに付随して初めて、トークンと計算資源は有効な付加的資産として機能し、利益を生み出すことができる。言い換えれば、トークンそれ自体は利益の源泉ではなく、システムこそが利益の源泉である。トークンはシステムの燃料に過ぎず、エンジンではない。

トークンが人工知能時代の重要な計測単位となるにつれて、トークンを中心としたビジネスモデルが市場にますます多く出現している。基盤モデルベンダーに加え

て、トークン代理店、モデルアグリゲーションプラットフォーム、エージェントプラットフォーム、エンタープライズ向け API 配信サービス、および様々なミドルレイヤーサービスプロバイダーが現れている。多くの起業家にとって、トークン代理店業は人工知能産業への参入捷徑のように見える。なぜなら、モデルを訓練する必要も、大規模な計算資源センターを建設する必要もなく、既存のリソースを統合するだけで事業を開始できるからである。しかし、このモデルには利点と同時に明らかな限界も存在する。利点は、比較的低い投資で済むこと、参入障壁が比較的小さいこと、市場拡大速度が速いことなどである。限界は、上流のモデルサプライヤーへの依存度が高く、価格競争に陥りやすく、長期的な技術的障壁を欠くことである。トークン価格が継続的に下落する場合、単純なトークン転売による利益の余地は継続的に圧縮される可能性がある。したがって、長期的には、トークン代理店業はむしろ流通能力であり、最終的な中核的競争力ではない。企業の長期的価値を真に決定するのは、依然としてトークンを中心に構築された業界システム、自動化プロセス、知識体系、そして現実のシナリオにおけるソリューションなのである。

六、現在の難局と将来の出路

現在のトークン市場は、いくつかの深層問題に直面している。

同質化競争の罠。 2,000 以上のプラットフォームが同じ顧客層を奪い合い、価格競争は業界の利益を押し下げ続けている。変わらなければ、クラウドコンピューティング初期の「薄利多売どころか損して客を掴む」事態を繰り返す可能性がある。

コンプライアンスリスクの顕在化。 現在の世界的な特別立法がないという「規制上のボーナス」は、一時的な状態であって長期的な常態ではない。トークン市場の規模拡大に伴い、データ越境、国家安全保障、税收管理などの問題が各国の立法を加速させるだろう。特に注意すべきは、一部の市場主体がトークンをトークン化・金融化しようとする試みは、世界の金融規制を誘発しやすく、関連ビジネスが取り締まりを受けるだけでなく、中国のトークン輸出産業全体の評判を傷つけることになりかねない。

「トークン最大化 (Token-maxxing)」の迷思。 シリコンバレーでは新しい「誇示」の方法が流行している。毎日どれだけ多くのトークンを消費するかを競うので

ある。Meta 社内には「Claudeconomics」と呼ばれるランキングがあり、85,000 人以上の従業員の AI 使用データを集計している。ランキング 1 位の従業員が消費するトークンの価値は数百万ドルに上る。反対派は辛辣に指摘する。「トークン最大化よりも結果最大化 (Outcome maxxng) が優れている」——狂ったようにトークンを消費するよりも、どんな結果を生み出したかを見る方が良い。これは、私が長年の実践で痛感してきた中核的問題そのものである。

中国語トークンと英語トークン：統一概念の下三重の壁

中国ユーザーにとって、明確に認識しなければならない事実がある。中国語トークン (詞元) と英語トークンは、どちらも人工知能時代の基礎的計測単位に属するが、真に統一され自由に相互運用できるグローバル市場を形成してはいない。技術的概念から見れば、東西ともにモデルが言語やデータを処理する基本単位としてトークンを採用している。しかし、実用的な応用の観点から見れば、技術的な壁、システムの壁、そして政策的な壁が共同して存在することで、中国語トークンと英語トークンの間には事実上の分離状態が形成されている。

第一の壁は技術的な壁であり、すなわちトークン化の論理とモデル構造の差異である。

GPT、Claude、Gemini などの西洋の主流モデルは、ほとんどが英語を主体とする大規模コーパスに基づいて構築されている。これらのモデルはすでに優れた中国語処理能力を備えているが、多くのシナリオでは、同じ情報量を表現するのに中国語は依然として英語よりも多くのトークンを必要とする傾向がある。さらに重要なのは、異なるモデルは異なる語彙表構造とトークン化戦略を持つことである。同じ中国語のテキストが、異なるモデルでは完全に異なる数のトークンに分割される可能性があり、その内部符号化方式や意味マッピング方式も異なる可能性がある。

同時に、DeepSeek、通義千問、豆包 (Doubao)、文心一言 (Wenxin Yiyan) などの中国国産モデルは、訓練段階から中国語コーパスを大量に導入しており、中国語のフレーズ、成語、固有名詞、ローカライズ表現に対してより高いネイティブ適合性を持っている。したがって、同じテキストに直面しても、モデルによってトークン数、処理効率、理解経路に明確な差異が生じる可能性がある。この意味で、ト

クンは統一された概念を持っているものの、モデル間を直接流通・再利用できる「標準トークン」は存在しないのである。

第二の壁はシステムの壁であり、すなわちエコシステムと呼び出し方式の差異である。

現在の世界のトークン市場は、実際には複数の比較的独立したエコシステムを形成している。OpenAI、Anthropic、Googleなどの国際的プラットフォームはそれぞれ独自のAPI標準、課金モデル、アクセスルール、サービス体系を持っている。中国市場では、DeepSeek、通義千問、豆包、Kimiなどのプラットフォームを中核とする本土エコシステムが形成されている。

一般ユーザーにとって、あるプラットフォームでトークンを購入しても、それを別のプラットフォームで直接使用できることを意味するわけではない。異なるプラットフォーム間では、呼び出しインターフェース、認証メカニズム、レート制限、サービス契約、ビジネスルールがそれぞれ異なる。第三者のゲートウェイやアグリゲーションプラットフォームを通じて変換しても、本質的には別のシステムを呼び出し直しているに過ぎず、真の意味でのトークン相互交換ではない。したがって、トークンは概念として統一されているが、商業リソースおよびサービス能力としてのトークンは、依然としてそれぞれのプラットフォームエコシステムの中に閉じ込められている。

第三の壁は政策的な壁であり、すなわちデータの流れとコンプライアンス要件の差異である。

人工知能が徐々にエンタープライズ応用段階に入るにつれて、データセキュリティ問題の重要性はますます高まっている。中国のデータセキュリティ体系、EUのGDPR体系、および米国の機密データの越境流動に対する規制措置は、いずれもデータの伝送と処理に異なる要件を課している。顧客情報、財務データ、医療データ、企業秘密、個人プライバシーに関わる多くのデータは、往々にして自由に国境を越えて流動することができない。

これは、実際のビジネス環境では、多くの企業が単にトークン価格に基づいてモデルを選択するのではなく、データコンプライアンス、データ保存場所、法的責任を

優先しなければならないことを意味する。多くの実際の業務データに関しては、コンプライアンスコストがトークンそのもののコストをはるかに上回ることが多い。したがって、現実世界では、国内業務は国内のコンプライアンスモデルを使用し、海外業務は海外のモデルを使用する傾向があり、結果として事実上のトークン市場の分断が生じる。

技術的な壁がトークン分割方式の差異を決定し、システムの壁が呼び出し体系と商業エコシステムの差異を決定し、政策的な壁がデータフローとコンプライアンス境界の差異を決定する。三重の壁が共同して作用することで、中国語トークンと英語トークンは、たとえ同じ技術概念に属しながらも、理想的な「世界的に統一された AI 通貨」のように自由に交換、自由に流通、自由に接続することが困難なのである。

将来的には、国際貿易に類似した状況が現れる可能性が高い。すなわち、中国語トークンエコシステムと英語トークンエコシステムが長期的に共存し、それぞれが独自の発展経路を形成する。圏を越えた呼び出しは不可能ではないが、往々にして追加の技術的適応、システム変換、および法的審査コストが必要となる。大多数の一般ユーザーにとって、このような圏越え能力は日常的な必要ではなく、したがって二大体系は長期間にわたって比較的独立した発展を維持するだろう。

中国で生産されたトークンは必ずしも中国のシステム内でのみ使用されるわけではないが、データ主権、コンプライアンス責任、産業安全保障に関わる重要なシナリオでは、将来的なトークン需要の多くは優先的に国内モデル、国内プラットフォーム、国内システム内部で消費される可能性が高い。

長期的なトレンドから見ると、計算資源の生産地、モデル提供者、システム運用者、データ所在地の間の結合度は強まっている。金融、行政、医療、サプライチェーン、大企業システムに関しては、データは往々にして自由に国境を越えて流通することができない。したがって、たとえ中国が大量のトークンサービスを世界に向けて輸出できたとしても、多くの高価値トークン需要は依然として優先的に中国のモデルと中国のシステム内部で完結するだろう。これは中国のトークンが輸出できないことを意味するのではなく、将来のトークン市場が「世界流通部分」と「国内閉ループ部分」の二つの構造を同時に持つ可能性があることを意味する。

注目すべきは、将来的に真の統一を実現するのはトークン自体ではなく、トークンと呼び出す方法である可能性が高いということだ。エージェントプラットフォーム、統一インターフェースプロトコル、モデルアグリゲーションプラットフォーム、自動化システムなどが発展を続けるにつれて、ユーザーが最終的に注目するのは、どのトークンが使われたかではなく、タスクが完了したか、コストが合理的か、効率が十分かどうかになるだろう。その時、トークンは今日の電気、ネットワークトラフィック、クラウドコンピューティング資源と同様に、次第にシステムの背後に隠れ、デジタル文明の基礎的運行インフラとなるだろう。

したがって、長期的なトレンドから見れば、東西のトークン市場は当面は直接相互運用することが難しいものの、その発展方向は高度に一致している。トークンの過去はコンピュータによる言語処理に属し、トークンの現在は人工知能産業に属し、トークンの未来はデジタル文明全体の基礎的運行体系に属する可能性がある。

七、なぜシステム構築能力は大量のトークンによって直接獲得できないのか

「システム構築能力は希少な資産である」と論じるにあたり、私は提起される可能性のある疑問に直面向き合う必要がある。AI がそれほど強力であるならば、なぜ大量のトークンを消費することでシステム構築能力を「生成」または「獲得」できないのか？この問いの答えは、トークン能力の根本的境界を明らかにする。

第一に、システム構築には意思決定が必要であり、意思決定は生成ではない。

AI は記事、コード、ウェブサイトのテンプレートを生成できるが、システム設計における中核的な意思決定——どのモジュールを選択し、どの機能を放棄するか、現在のニーズと将来の拡張のバランスをどう取るか、いつリファクタリングし、いつ現状を維持するか——を AI が代わりに行うことはできない。これらの意思決定は、ビジネスの本質の理解、リソース制約の判断、長期的帰結の予見に依存している。トークンは無数の「可能な解決策」を生成できるが、「どの解決策を実施する価値があるか」そして「なぜか」を選択することはトークンが代替できるものではない。私のインテリジェント物流システムは 20 年にわたって無数の調整を経験して

きたが、毎回の調整はAIが「こうすればより良い」と教えたからではなく、実際の業務に問題が発生し、限界が露呈し、新たなニーズが生じたからである。これらの意思決定の背後にあるのは判断力であり、判断力はトークンからではなく実践からもたらされる。

第二に、システム構築には例外処理が必要であり、例外を網羅的に列举することはできない。

システムが長期間稼働できる理由は、正常フローが完璧に設計されているからではなく、例外を処理できるからである。物流システムにおける例外には、顧客住所の誤り、貨物の破損、通関遅延、支払い失敗、システムダウン、データ損失などが含まれる。出版システムにおける例外には、翻訳品質の変動、メタデータ形式の非互換、アーカイブインターフェースの変更、DOI登録の失敗などが含まれる。これらの例外は網羅的に列举することはできない。なぜなら現実世界の例外の多様性は無限だからである。AIは大量のトークンを通じて「一般的な例外パターン」を学習できるが、「次に現れる未曾有の例外」を予測することはできない。長期間稼働するシステムだけが、実際の運用において継続的に例外に遭遇し、記録し、解決することで、真の例外処理能力を形成することができる。この能力は「より多くのトークンを与える」ことで獲得できるものではない。なぜなら例外は知識ではなく、経験だからである。

第三に、システム構築にはトレードオフが必要であり、トレードオフに標準的な答えはない。

あらゆるシステムの構築は、多重の制約下でのトレードオフである。コスト対性能、速度対安定性、汎用性対カスタマイズ性、現在のニーズ対将来の拡張。これらのトレードオフに「正解」はなく、「特定の状況下での合理的な選択」があるだけである。AIは「最適解」を与えることができるが、この「最適」は歴史的データに基づく統計的最適であって、「あなたの具体的な状況」に基づく判断ではない。私の世代を超えた20年間の物流・財務統合システムが今日まで稼働しているのは、まさに異なる時代に異なるトレードオフを行ってきたからである——2005年にある技術スタックを選択し、2015年に別の統合方法を選択し、2025年にアップグレード経路を選択した。これらのトレードオフは「正しい」か「間違っている」かでは

なく、「当時の状況に適合していた」のである。トークンはこれらのトレードオフを私に代わって行うことはできない。なぜならトレードオフの本質は価値判断であり、計算ではないからである。

第四に、システム構築にはフィードバックループが必要であり、フィードバックには現実世界が必要である。

システム構築能力の向上は、設計→実装→運用→問題発見→修正→再設計という閉ループに依存している。このループにおける「問題発見」の段階は、現実世界からのフィードバックに由来しなければならない。AI は運用をシミュレートできるが、現実世界におけるすべての予測不可能な要素をシミュレートすることはできない。私の物流システムが継続的に改善されてきたのは、実際の顧客が実際のフィードバックを提供したからであり、私の出版システムが継続的に改善されてきたのは、実際のアーカイブインターフェースが実際の変更を経験したからである。これらのフィードバックは、閉ざされた環境でトークンを消費することで獲得できるものではない。トークンは「顧客がこうフィードバックするだろう」という仮定を生成できるが、「顧客が実際にそうフィードバックした」という事実を代替することはできない。現実世界からのフィードバックは、システム構築能力が成長するための唯一の土壌である。

第五に、システム構築には信頼が必要であり、信頼は時間からしか生まれない。

システムが信頼に値するかどうかは、宣伝によってではなく、時間によって証明される。私のウェブサイトが CERN Zenodo、WorldCat、オーストラリア国立図書館によって永久保存されているのは、これらの機関が「私を気に入っている」からではなく、長期的な観察を通じて私のシステムが安定して信頼性が高く持続的であることを確認したからである。この信頼は大量のトークンで購入できるものではない。AI は「非常に専門的に見える」システム説明書を生成できるが、「このシステムが 20 年間信頼されてきた」という事実を生成することはできない。信頼は時間の産物であり、トークンの産物ではない。

まとめると、システム構築能力が大量のトークンを通じて直接獲得できない理由は、それが本質的に生成能力ではなく、判断能力、例外処理能力、トレードオフ能力、フィードバック吸収能力、信頼蓄積能力だからである。これらの能力に共通す

る特徴は、現実世界の時間、実践、エラー、修正、検証を必要とするということである。トークンは情報処理を加速することはできても、現実世界の経験蓄積を加速することはできない。これが、将来的にトークンが極めて安価になってもシステム構築能力が依然として希少であり続ける根本的な理由である。

八、私の資産：現実の履歴データ、分野横断的システム、構築能力、予見ビジョン、自己実証能力

「真に希少なものは何か」をさらに論じる前に、まず私自身の立場を説明する必要がある。なぜなら私の判断は理論的な演繹によるものではなく、長期的な実践によるものだからである。このことは、私が長年にわたって構築してきた様々なシステムに非常にはっきりと現れている。インテリジェント物流システム、インテリジェント請求書システム、自動化財務処理体系、その後の 10 言語出版システム、メタデータシステム、ウェブサイトナビゲーションシステム、自動相互参照システムのいずれにおいても、真に困難な部分は決して情報を生成することではなく、情報を組織することであった。情報はいくらでも増えていくが、構造がなければ、結局は新たな混乱を生み出すだけである。トークン時代も同様である。将来、人工知能は膨大な量のコンテンツを生成できるようになるが、膨大な量のコンテンツは自動的に膨大な量の価値を意味するわけではない。むしろ逆に、コンテンツ生成が容易になるにつれて、これらのコンテンツをどのように選別し、組織し、保存し、索引付けし、関連付けるかがますます重要になっていく。

私には以下の 5 つの資産があり、これらの 5 つの資産が私の「トークン以後に何が希少か」という判断の経験的基盤を構成している。

第一に、現実の履歴データ。これはいかなる「履歴データ」でもなく、世代を超えた 20 年間の実際の業務データである——2005 年のインテリジェント請求書システムから現在に至るまで、実験環境やスクレイピングデータではなく、実際に稼働している物流システム、財務システム、税務システムからのデータである。AI 時代において、「合成データ」が氾濫する中で、現実の履歴データの価値は急激に上昇している。ほとんどの企業が持つのは「単一分野の履歴データ」であるのに対し、私

が持つのは分野横断的であり、20 年にわたり、実際に検証された履歴データである。

第二に、複数の分野横断的システムと体系の完成。 私のシステムは以下の分野をカバーしている：インテリジェント物流システム、インテリジェント請求書システム、インテリジェント財務システム、インテリジェント税務システム、世代を超えた 20 年間の物流・財務統合体系、ウェブサイト自動化システム、メタデータ生成システム、『The Epochal Transition』10 言語出版システム、DOI 管理システム、TROVE と WorldCat リンク整理システム、自動相互参照システム、Sitemap 生成システム、Zenodo プロフィール生成システム、クロスプラットフォーム出版システム。これらのシステムの独自性は、分野横断的な広がり（物流から出版、財務から文献索引まで）、時間的深み（一部のシステムは 20 年間稼働）、一人での完成（方法の移行可能性を証明）にある。このうち、『The Epochal Transition』電子月刊誌（<https://times.net.au>）は公開された実証例である。これは ISSN 国際刊行物番号（3083-5178）を持ち、WorldCat（世界最大の図書館総合目録システム、100 カ国以上の 16,000 以上の国立大学図書館が加盟）、オーストラリア国立図書館の PANDORA/TROVE、そして CERN 欧州核研究機構が運営する Zenodo、OpenAIRE 欧州オープンサイエンスインフラ、DataCite DOI（グローバルデジタルオブジェクト識別子、60 以上の文献）、ORCID（グローバルにユニークな永続的学術デジタル ID）などの国際的学術インフラによって永久保存されている。このウェブサイトは、一人の人間（独立研究者）が、コンテンツ制作から国際的なトップレベルのアーカイブインフラとの連携までの全プロセスを完了できることを証明している。

第三に、システムと体系を構築する能力。 これは「AI ツールを使いこなす能力」ではなく、曖昧な要求を安定した構造に変換する設計能力、異なるシステムを統一的な体系に統合するアーキテクチャ能力、無人でシステムを数年単位で稼働させ続けるエンジニアリング能力、物流から出版、財務から文献へと経験を移行するメタ能力である。先に論じたように、この能力は大量のトークンによって直接獲得することはできない。なぜならそれは意思決定、例外処理、トレードオフ、フィードバックループ、信頼蓄積に依存しており、これらはいずれも AI によって代替できないからである。現在の AI トレーニング市場は、「プロンプトエンジニアリング」「AI ライティング」「AI 描画」などの表層的スキルに集中しており、「分野横断的シス

テム構築能力」を教えるコースはほとんどない。なぜならこの能力は短期間で習得できるものではなく、長期的な実践の蓄積によってのみ得られるからである。ウェブサイトに掲載された 40 以上の独立創作作品、取得した ISSN、60 以上の DOI は、システムの長期運用能力の最も優れた公開記録であり、私自身の自己宣伝ではない。

第四に、未来を予見するビジョン。 1993 年にリモートワークプロトタイプシステムの試行を開始——大規模なリモートワークの普及より約 30 年先行。2005 年にインテリジェント請求書システムを設計——AI 普及の 20 年前にプロセス自動化を開始。世代を超えた 20 年間のシステム統合——システム設計の時点で 20 年後も稼働し続ける必要を予見。現在「トークン以後」の判断を提示——現在の大多数の人々がまだトークンを追いかけている時に、システム能力こそが真の障壁であることを予見。さらに私は「トークンは通貨に類似しており、東西の技術、システム、政策の違いにより普遍的に通用しない」ことも予見していた——米ドルと人民元が自由に交換できないのと同様に、中国語トークンと英語トークンも同様の流通障壁に直面し、将来的には「中国語トークン圏」「英語トークン圏」などの言語と政策に基づくトークンエコシステム圏が出現するだろう。また「現在トークンで儲けることはすでに困難であり、ますます難しくなる。フロントエンドにサービスを提供するシステムがあって初めて、トークンと計算資源は有効な付加的資産として利益を生み出すことができる」ことも予見していた。私は独立研究者として、将来の学術インフラと連携する個人出版体系を先取りして構築し、「個人もグローバル知識インフラの一ノードとなり得る」という趨勢を予見していた。ほとんどの趨勢予測はアナリストや起業家によるものであるが、私の「予見能力」は 20 年の実践による検証という裏付けを持っている。

第五に、自己実証能力。 これは私の最も重要なメタ能力の一つである。私が執筆したあらゆる分野の文献は、理論的な閉ループと論理的な自己無矛盾性を達成している。これは、私の議論が外部権威や引用に依存するのではなく、内部論理の自己充足性に依存していることを意味する。学術文献のデフォルトルールは「先行研究を引用する」ことであり、「自己実証能力」を主張または示す人はほとんどいない。学術界の議論は「議論が十分であるか」に焦点を当て、「著者が自己無矛盾な閉ループを構築する能力を持っているか」ではない。私の記事では、この自己実証能力

は以下のように現れている。すなわち、主張は実践に由来し、実践が主張を検証し、主張と実践が閉ループを形成している。私は「他人の理論を引用して自分を証明する」のではなく、「自分が構築したシステムを示し、システム自身に語らせる」のである。この自己実証能力は、公開文献には全く存在しない概念であり、私の最もユニークなメタ能力の一つである。

これら五つの資産を組み合わせたとき、その希少性は加算ではなく乗算である。すなわち、現実の履歴データ × 分野横断的システム × システム構築能力 × 未来を予見するビジョン × 自己実証能力 は、複製不可能な資産ポートフォリオを構成する。公開されている議論の中で、現在のところこのような対照事例は見つからない。

『国際独立研究者の道』を整理する過程で、私はこのことをさらに深く認識した。1989 年から 2026 年までの 30 年以上にわたる経験は、複数の分野にまたがっている。これらの経験は一度に生じたものではなく、長い時間をかけて徐々に形成されたものである。人工知能は数秒で記事を生成できるが、30 年以上の実際の経験を数秒で生成することはできない。大量の意見を生成することはできても、長期的な実践によってもたらされる判断力を直接生成することはできない。多くのプロセスをシミュレートすることはできても、現実世界における繰り返しの検証を代替することはできない。この意味で、時間そのものも重要な資産であり、時間の堆積によって形成された構造はさらに希少な資産である。

九、トークン以後：真に希少な五つの能力

トークンそれ自体がもはや希少でなくなったとき——トークンを販売するための参入障壁は高くないが、現実の履歴データを持ち、システムや体系を構築し、未来を予見するビジョンを持つことは極度に不足するだろう——真に希少なものは何か？ 1993 年のリモートワークプロトタイプシステム、2005 年のインテリジェント請求書システム、インテリジェント物流システム、インテリジェント財務システム、インテリジェント税務システム、世代を超えた 20 年間の物流・財務統合体系から、ウェブサイト自動化、メタデータ生成、『The Epochal Transition』10 言語月刊誌、DOI 管理、TROVE と WorldCat リンク整理、自動相互参照、Sitemap 生成、

Zenodo プロフィール生成、クロスプラットフォーム出版に至るまでの長期的実践に基づき、以下の五つの能力が真の障壁となると考える。

第一に、構造設計能力。 トークンは原材料であり、構造は原材料を価値ある製品に組織化する枠組みである。人工知能が普及する遥か以前から、私は限られたリソースを活用してより高い効率の運用を実現する方法を考え始めていた。私のインテリジェント物流システム、インテリジェント請求書システム、世代を超えた 20 年間のインテリジェント物流・財務システムは、異なる時代に生まれ、異なる問題に直面していたが、それらには共通する特徴があった。目標はより多くの情報を作り出すことではなく、混乱を減らし、反復労働を減らし、全体的な効率を高め、長期にわたって稼働できる構造を確立することであった。自動化出版システムにおいて、私が直面した中核的問題は「テキストを生成できるか」ではなく、「10 言語のメタデータフィールドをどのように定義するか、記事とカテゴリの関連をどのように確立するか、バージョン管理ルールをどのように設計するか」であった。これらの構造設計の決定が、システムが何年もクラッシュせずに稼働し続けられるかどうかを決定する。対照的に、消費されるトークンの量は単なる副産物に過ぎない。『The Epochal Transition』ウェブサイト自体が、構造設計能力の公開された実証例である。すなわち、ISSN 国際刊行物番号を持ち、多言語対応し、複数のアーカイブチャンネルを持つ、個人が維持する学術ウェブサイト構造であり、私は空論を述べているのではなく、この構造設計能力を用いて公開された複雑なシステムを実際に構築したことを証明している。

第二に、自動化能力。 トークンはコンテンツをバッチ生産できるが、エンドツーエンドのワークフローをバッチ完了できるのは自動化システムだけである。インテリジェント物流システムにおいて、真の価値はルート計画のために AI を何回呼び出すかにあるのではなく、注文受付から倉庫スケジューリング、経路最適化、例外処理までの完全に自動化された閉ループを確立することにある。この閉ループの中で、AI は一構成要素に過ぎず、自動化能力は構成要素間の接続方法に現れる。システムを持たない人間は、大量のトークンを持っていても、散在するテキスト、散在する画像、散在する表、散在するコードを生成し続けるだけかもしれない。一方、システムを持つ人間は、トークンを自身のワークフローに接続し、それらに資料の自動整理、メタデータの生成、リンクのチェック、ウェブページの更新、出版の補助、

文献のアーカイブ、索引の作成、物流の処理、財務の分析、レポートの生成などを自動的に実行させることができる。

第三に、分野横断的实践経験。 トークンはすべての人に開かれているが、ある業界の経験を別の業界に移行する能力は普遍的ではない。私が物流システムと出版システムの両方を構築できる理由は、より多くのトークンを持っているからではなく、メタデータ管理、ワークフロー設計、例外処理などの移行可能な方法論を蓄積してきたからである。私の出版システムはトークンを必要とする。なぜなら、10 言語翻訳、要約生成、キーワード整理、メタデータ生成、リンク説明には大量の言語処理が必要だからである。私のウェブサイトシステムはトークンを必要とする。なぜなら、ページ構造、ナビゲーション説明、SEO 説明、学術メタデータ、多言語紹介などを自動生成できるからである。私の物流システムもトークンを使用できる。なぜなら、顧客メール、貨物説明、通関資料、例外説明、財務メモ、操作ログなどを人工知能で補助処理できるからである。私の自動相互参照システムは特にトークンを必要とする。なぜなら、大量の記事、タイトル、DOI、NLA、TROVE、WorldCat リンク間のテーマ関連を見つけ出さなければならないからである。このような分野横断的实践の蓄積は、トークンを消費することで得られるものではなく、現実の問題を解決することで堆積されるものである。ウェブサイトは、学術出版、インフラ統合、技術アーキテクチャ、個人ブランドなど複数の分野の情報を同時に提示しており、私が単一のスキルではなく、出版、図書館、技術、個人ブランドを実際に結び合わせていることを証明している。

第四に、システム長期運用のメンタルモデル。 現在の AI アプリケーションの主流の物語は「対話」と「生成」であるが、真に持続的な価値を生み出すのは「運用」である。一つの対話は 1 万トークンを生成し、一つのエージェントタスクは 100 万トークンを消費する。しかし、10 年間継続して稼働する物流システムが消費するトークンは数十億に上るかもしれない。そして真に希少なものは、このシステムを設計するメンタルモデルである。モデルのアップグレード後もシステムが使用可能であることをどう保証するか？上流 API の変更はどう対処するか？例外を監視し自動復旧するにはどうするか？これらの問題はトークンとは無関係でありながら、トークンの最終的な産出を決定する。これらのシステムの価値はトークンそれ自体から来るのではなく、私がすでに構築した、または開発中の構造から来る。トークンは単

なる燃料であり、システムこそがエンジンである。トークンは単なる電流であり、構造こそが回路である。トークンは単なる水であり、パイプ、バルブ、貯水池、調達システムこそがその役割を発揮させるかどうかを決定する。システムがなければトークンは流失し、システムがあればトークンは長期的な価値に変換される。ウェブサイトは、40以上の独立創作作品と、CERNやオーストラリア国立図書館などの権威機関による永久保存を示しており、私のシステムが短期的な話題性を追うのではなく長期的な価値を追求しており、権威機関による「永久保存」の裏付けをすでに得ていることを証明している。

しかし、トークンとシステムの関係についてさらに考察を進めると、もう一つ見落とされがちな問題がある。多くの人々は、優れたシステムさえ持っていれば将来のトークン時代に優位に立てると考えている。しかし現実とはそれほど単純ではない。システムそのものは誰にでも設計できるというものではない。真に成熟したシステムは、長期的な実践、繰り返しの修正、現実のシナリオによる検証を必要とする。多くのシステムは、最初の構想から実際の実装までに、数年、あるいは数十年という継続的な蓄積を必要とする。したがって、システム設計能力それ自体がすでに希少な資源なのである。

しかし、成熟したシステムを持っているとしても、それが自動的に大規模な影響力の形成を意味するわけではない。システムが技術的に可能であることから社会に普及することまでの間には、市場、資本、流通チャネル、ユーザー教育、プラットフォームエコシステムなど、複数の層が存在する。歴史は優れた技術や優れたシステムに不足しているわけではない。真に希少なものは、システムを大規模な応用にまで押し上げるプラットフォーム能力である。システムが局所的なシナリオにおいて既存のソリューションよりもはるかに優れていたとしても、十分な資本支援、プラットフォーム促進能力、ユーザーベースがなければ、その実際の利用規模は依然として非常に限られたものになる可能性がある。

トークン経済の観点から見ると、これは特に明らかである。トークン需要はシステムそのものから直接生じるのではなく、システムの実際の利用規模から生じる。毎日数百万人のユーザーにサービスを提供するプラットフォームは、たとえシステム設計が比較的普通であっても、巨大なトークン消費を生み出す可能性がある。逆に、設計は非常に先進的でも使用人数が限られているシステムは、トークン需要が

非常に小さい可能性がある。したがって、将来のトークン市場の発展は、モデル能力やシステム能力だけでなく、プラットフォーム能力や資源統合能力にも依存する。

これは、将来の競争には少なくとも三つのレベルが存在することを意味する。第一レベルはモデル競争であり、誰がトークンを供給できるかを決定する。第二レベルはシステム競争であり、誰がトークンを効果的に利用できるかを決定する。第三レベルはプラットフォーム競争であり、システムがどれだけのユーザーをカバーし、どれだけの実際の需要を生み出せるかを決定する。モデル、システム、プラットフォームの三者は欠くべからざるものである。モデルがなければシステムは稼働できない。システムがなければトークンは価値を創造しにくい。プラットフォームがなければ、どんなに優れたシステムも小規模な利用範囲に留まる可能性がある。

このように考えると、私はますます、トークン自体は単なる基礎的資源に過ぎないと確信する。真に未来を決定するのは、トークンを持つ人間でも、システムを持つ人間でもなく、モデル、システム、プラットフォーム、そして現実のニーズを結びつけることができる人間である。これらの要素が閉ループを形成して初めて、トークンは単なる計算単位から価値を創造し続ける生産力へと転換するのである。

第五に、トークンを現実のビジネスプロセスに埋め込む能力。 トークンは最終的に現実世界のどこかの部分——物流、出版、教育、医療——に役立たなければならない。真に希少なものは、AI 能力の境界を理解し、かつ特定の業界の課題を理解している人々である。彼らはいつ AI を使うべきか、いつ使うべきでないかを知っている。トークンの出力をビジネス上の意思決定に変換する方法を知っている。トークンの投入とビジネスの産出との真の関係を測定する方法を知っている。この複合的な能力は、現在の市場で最も希少な資源である。ウェブサイト自体が稼働中の「ビジネスプロセス」である。コンテンツ制作 → メタデータ生成 → マルチプラットフォーム出版 → 国際的アーカイブ → 読者のアクセス。これは、私の AI 応用（トークン消費）が、実際の影響力を生み出す完全な閉ループの出版プロセスに組み込まれていることを証明している。

私は、将来の真の競争はおそらくモデル間の競争でもトークン間の競争でもなく、システム間の競争であるとますます感じている。モデルはアップグレードを続け、

トークン価格は下落し続け、今日の高価な資源は将来非常に普及するかもしれない。しかしシステムは資源が豊富だからといって自動的に現れるものではない。システムは長期的な蓄積を必要とし、継続的な修正を必要とし、現実世界の検証を必要とする。誰がトークンを長期運用可能なシステムに変換できるかが、持続的な価値を得る可能性が高い。そしてこれこそが収益性の鍵でもある。すなわち、現在トークンで儲けることはすでに困難であり、ますます難しくなる。フロントエンドにサービスを提供するシステムがあつて初めて、トークンと計算資源は有効な付加的資産として利益を生み出すことができるのである。

私はトークン時代の真の競争は三つの段階に分かれると判断している。第一段階は資源競争であり、モデル、計算資源、API、トークン価格を奪い合う。第二段階は応用競争であり、トークンを執筆、カスタマーサービス、コーディング、翻訳、教育、オフィスワーク、データ分析に使用する。第三段階が真のシステム競争であり、誰がトークンを長期運用される産業システム、出版システム、物流システム、知識システム、文明記録システムに組み込むことができるかが、長期的な優位性を得る可能性が高い。

実際、トークン、モデル、システム、プラットフォームについて議論する際、現実世界では時間というもう一つ重要な要素がしばしば無視されている。多くの人は、資本があれば短期間でシステムを構築でき、市場が十分に大きければ成功モデルを迅速に複製できると考えている。しかし現実には往々にしてそうではない。資本は計算資源を購入でき、トークンを調達でき、開発チームを雇うことはできても、長年にわたって検証されてきた実際のデータや成熟したシステムを直接購入することはできない。多くの産業にとって、真に希少なものはモデルそのものではなく、トークンそのものでもなく、実際のビジネス環境で長期間安定して稼働してきたシステムなのである。

ビジネスの観点から見ると、市場と資本は通常、新しいシステムがゼロから成長するのを長期間待つことはない。大規模なシステムは、概念提案、アーキテクチャ設計、機能開発、プロセス検証から実際の実装に至るまで、しばしば数年、あるいはそれ以上の時間を要する。このプロセスでは、継続的な資源投入だけでなく、大量の試行錯誤、修正、現実環境での検証が必要である。したがって、市場機会がすで

に現れた時点では、資本はゼロから全く新しい体系を設計するよりも、すでに形になり、すでに稼働し、すでに検証されたシステムを探す傾向が強い。

しかし、現実にはもう一つの矛盾が存在する。大規模なシステムであればあるほど、後発の資本が容易に介入することは難しくなる。大規模なシステムは長期にわたる発展を経て、通常すでに独自の運用メカニズム、管理体制、データロジック、組織文化を形成している。それらは膨大な履歴データ、複雑なビジネスプロセス、長期間蓄積された経験体系を持っているかもしれない。そのため、大規模なシステムは価値が計り知れない一方で、資本の期待に従って迅速に変革したり、徹底的に再構築したりすることが困難であることが多い。システム規模が大きければ大きいほど、その慣性も大きくなる傾向がある。

したがって、現実のビジネス世界において最も理想的な状態は、ゼロからシステムを設計することでも、巨大で硬直した体系を単純に買収することでもなく、中核的な構造構築をすでに完了し、長期的な検証を経て、安定して稼働できる既存のシステムを見つけることである。そのようなシステムは完璧ではないにしても、最も困難な段階をすでに乗り越えており、現実のデータ、現実のプロセス、現実の検証記録を持っているため、依然として非常に高い価値を持つ。これらをベースに、再パッケージング、最適化、外部開放、プラットフォーム化・プロモーションを行うことは、ゼロから構築するよりも効率的であることが多く、試行錯誤のリスクと時間コストを大幅に削減できる。なぜなら、これらのシステムはすでに最も費用がかかり、最も時間がかかり、最も複製が難しい段階を完了しているからである。

この観点から見ると、時間の窓もまた見落とされがちな資源である。産業がまだ初期段階にあるとき、多くの人々はモデル能力、トークン価格、市場規模に注目し、システムの形成に必要な時間を無視している。成熟したシステムの誕生は、多くの場合、概念提案、アーキテクチャ設計、プロセス検証、現実運用、継続的改善などの複数の段階を経る。真に困難なのはアイデアを提案することではなく、システムを現実環境で長期的に安定稼働させることである。したがって、市場が成熟したシステムの価値を認識し始めたとき、それは往々にしてこれらのシステムがすでに長年の蓄積を経ていることを意味しており、短期間で再現できるものではない。

資本と市場にとって、最も理想的な対象は、概念段階の構想でも、まだ検証されていないプロトタイプでもなく、中核的な構造構築を完了し、実際の稼働記録を持ち、アップグレードの余地があるシステムの胎動段階にあるものである。このようなシステムは完璧ではないにしても、最も困難で、最も時間がかかり、最もリスクの高い段階をすでに乗り越えている。市場需要が急増すると、このようなシステムは非常に人気が出る傾向がある。なぜなら、真に長期的に検証された成熟したシステムの数、市場の想像よりもはるかに少ないからである。モデルは迅速に反復でき、トークンは継続的に拡張でき、開発チームは一時的に結成できるが、現実のデータ、現実のプロセス、長期的な検証記録は短期間で再創造することはできない。

したがって、トークン経済が発展していく過程において、時間はシステムの形成に影響を与えるだけでなく、システムの価値にも影響を与える。市場が特定のタイプのシステムの重要性を十分に認識していない時期は、そのようなシステムを獲得するコストが最も低い時期でもある。そして市場がその価値を一般的に認識したときには、選択可能な対象は通常すでに非常に限られている。多くの産業にとって、失われるのは単なる一プロジェクトではなく、長期的な蓄積をすでに完了し、次の発展段階に直接進むことができる時間の窓なのである。

トークン経済の観点からも同様である。トークンは迅速に入手でき、モデルは迅速に接続でき、プラットフォームは迅速に構築できるが、真に短期間で複製することが困難なのは、長年にわたって堆積され、価値を生み出し続けることができるシステムである。将来、トークンがますます豊富になるとき、システムの重要性は低下しないどころか、さらに高まる可能性がある。なぜなら、すべての人が同程度のトークン資源を持つようになった後、競争結果を決定するのはもはやトークンの数量ではなく、誰がより成熟した現実のデータ、より安定したビジネスプロセス、より現実のニーズに近いシステム構造を持っているかだからである。

もう一つ見落とされがちな問題は、システムの起源である。多くの人々はシステムをソフトウェア製品と理解することに慣れており、したがって当然、システムはソフトウェア会社によって設計・開発されるべきだと考えている。しかし、多くの複雑な産業において、真に長期間安定して稼働するシステムは、往々にして最初にソフトウェア会社によって設計され、業界に渡されるのではなく、長期の業務実践の

過程で徐々に形成され、その後ソフトウェア化、自動化、プラットフォーム化されるのである。

理由は複雑ではない。ソフトウェア会社は通常、技術的実装を得意とするが、必ずしも長期間にわたって具体的な業務現場に身を置いているわけではない。彼らは機能を迅速に開発できるが、短期間で大量の実際の業務経験を蓄積することは難しい。一方、業界内部で自然発生的に発展してきたシステムは、毎日の顧客、注文、例外状況、法規制の変化、コスト圧力、市場競争に直面しながら、絶えず修正・改善されてきた。このようなシステムは理論から出発するのではなく、現実の問題から出発しており、そのすべてのプロセス、すべてのルール、すべての自動化部分は、往々にして実際の業務の長期的な検証に由来している。

したがって、物流、サプライチェーン、出版、金融、製造業、医療、および多くの専門分野において、最も価値のあるシステムは、必ずしも最も複雑なシステムでもなければ、最も多くの機能を持つシステムでもなく、長期的な実戦検証を経て、現実の問題を安定的に解決できるシステムである。それらは最新のインターフェースを持っていないかもしれず、最も派手な宣伝をしていないかもしれず、最新の技術アーキテクチャを採用していないかもしれないが、大量の現実のデータ、現実のプロセス、長期的に蓄積されたビジネスロジックを持っている。これらの経験は、しばしば短期間の開発では得ることができない。

トークン経済の観点から見ると、これは特に重要である。将来、モデル能力は向上し続け、トークンコストは低下し続け、ソフトウェア開発効率もますます高まるだろう。しかし、業界が数十年かけて蓄積してきた業務経験は、モデルのアップグレードによって自動的に生じるものではない。真に複製することが困難なのは、ソフトウェアコードそのものではなく、システムの背後に隠された業界知識、例外処理経験、意思決定ロジック、長期的に検証されたビジネスプロセスなのである。

したがって、私は将来の最も理想的なシステムは、業界実践から完全に切り離されて設計されたものではなく、実際の業務に起源を持ち、長期運用による検証を経て、その後人工知能とトークン能力を活用して継続的にアップグレードされるシステムであると考えます。このようなシステムは、現実世界の信頼性と人工知能時代に

よる効率向上の両方を備えており、その長期的価値は、単に技術コンセプトに基づいて構築されたシステムよりもはるかに高いことが多い。

十、トークンはどこへ流れるのか：一過性の対話から文明の記憶へ

将来のトークン経済について、もう一つ観察がある。多くの人々はトークンがどのように生産されるかに注目しているが、トークンが最終的にどこへ流れるかについてはあまり注目していない。トークンが一度のチャットだけに留まれば、そのライフサイクルはわずか数分かもしれない。普通のウェブページに入れば、ライフサイクルはわずか数ヶ月かもしれない。長期保存体系に入れば、その価値は根本的に変化する。トークンが DOI、Zenodo、WorldCat、TROVE、国立図書館の長期保存システム、そして様々な学術・知識インフラに入ると、それはもはや単なるトークンではなく、知識記録、文献資産、文明の記憶へと変容し始める。したがって、将来真に重要な問題は、より多くのトークンを生み出す方法ではなく、トークンを長期的に保存され、役割を発揮し続けることができる構造に入れる方法である。

このため、私はトークンの将来の発展に楽観的であるが、トークンそのものを盲信するわけではない。トークンの過去は、コンピュータが言語を処理する際の基本単位である。トークンの現在は、人工知能時代の重要な計測単位および商業単位である。トークンの未来は、デジタル文明の重要な基礎資源となる可能性が高い。しかし、トークンがどのように発展しようと、私は常に、真に希少なものはトークンではなく、トークンを扱いこなせる人間であると考え。より多くの資源を持つ人間ではなく、長期的な構造を確立できる人間である。より多くのコンテンツを生成できる人間ではなく、コンテンツを知識体系として組織できる人間である。

十一、時間の非圧縮性：購入不可能な四つの資産

「トークン以後に何が真に希少か」を論じるにあたり、私は大多数の議論が無視している根本的な問題をさらに指摘する必要がある。ある種の資産は本質的に加速、

購入、複製が不可能である。AI がどれほど強力であろうと、トークンがどれほど安価であろうと、以下の四つの資産はその固有の希少性を常に維持する。

第一に、時間は圧縮できない。

30 年の実践は 30 年の実践であり、3 年に圧縮することはできず、AI によってシミュレートすることもできない。人工知能は「物流システムの構築方法」に関する記事を数秒で生成できるが、物流システムが 20 年の間に遭遇する様々な異常、故障、政策変更、市場変動を数秒で経験することはできない。これらの現実世界における時間の蓄積は、いかなる技術的手段によっても短縮することはできない。私が 1993 年にリモートワークプロトタイプシステムの試行を開始し、2005 年にインテリジェント請求書システムを設計したという事実は、偶然ではなく、現実の歴史の不可逆的な記録である。後から来る者は誰でも私の記事を読み、私の方法を学ぶことができる。しかし、私が生きた年月を持つことはできない。それらの年月そのものが資産なのである。

第二に、分野横断的経験は購入できない。

分野横断的経験は、商品のように購入することはできない。人は一日で物流、出版、財務、文献索引に関する 100 冊の本を買うことができる。しかし、これらの分野を繋ぎ合わせる実際の経験をその同じ一日で獲得することはできない。真の分野横断的能力は知識の堆積ではなく、異なる分野間の接続を確立し、共通性を識別し、方法を移行する能力である。この能力は、一つの分野で実際の問題を解決し、過ちを犯し、経路を修正し、判断力を蓄積することによってのみ獲得できる。私が物流システムの経験を出版システムに移行できるのは、ある本を読んだからではなく、物流分野で 20 年を費やし、出版分野でさらに 10 年を費やしたからである。そして、この二つの時間の交差点において、移行可能な方法論が生まれたのである。

第三に、システム構築能力は複製が困難である。

システム構築能力は、コピー＆ペーストできるコードやプロセスの集合ではない。コードは複製できるが、コードの背後にある意思決定ロジック、取捨選択の判断、将来の変化への予見性は複製できない。システムの価値は、今日何ができるかだけでなく、過去 10 年間に何回の変更、何回の障害復旧、何回の要件調整を経験しな

がらも稼働を続けてきたかにもある。この「システムの歴史」そのものが、構築能力の証明である。私のウェブサイト (<https://times.net.au>) は誰でも見ることができる。しかし、閲覧者が見るのは結果であり、この結果の背後にある 30 年の意思決定の歴史、技術選定、構造調整、継続的メンテナンスではない。この構築能力は、盗作や模倣によって得ることはできない。

第四に、長期運用記録は短期間では達成できない。

システムが 20 年間稼働したという記録は、それ自体が本質的に短期間での達成が不可能である。長期運用記録は、システムの安定性、保守性、設計の先見性を証明する。AI 時代において、「見た目だけの」ウェブサイトやシステムを数分で生成することは可能かもしれない。しかし、「20 年間稼働し、CERN、WorldCat、オーストラリア国立図書館によって永久保存されている」システムを生成することに近道はない。長期運用記録はタイムスタンプであり、このシステムが現実世界の試練に耐えたことをすべての人に宣言する。このような記録は、いかなる AI 技術によっても偽造または加速することはできない。

これら四つの資産——時間、分野横断的経験、システム構築能力、長期運用記録——は、本稿で私が「真に希少」として論じてきたものの基盤を共同して構成している。これらを私が持っているのは、他の人よりも賢いとか勤勉だからではなく、他の人よりも早く始め、そして現在まで続けてきたからである。AI がますます多くのものを安価で容易に入手できるものになっている時代において、これらの圧縮、購入、複製、短期間での達成が不可能なものは、ますます貴重になっていく。

十二、結論：「トークン貯め」から「システム構築」へ

トークンの未来は、トークンそのものにあるのではない。

大規模モデル技術の継続的な発展と推論コストの継続的な低下に伴い、トークンの供給能力は拡大し続ける。今日高価であると考えられているトークンは、数年後には今日のネットワークトラフィックやハードディスクストレージと同様に安価で一般的になるだろう。すべての人が簡単に大量のトークンを入手できるようになったとき、真に希少になるのはもはやトークンではなく、トークンを現実の価値に変換

する能力——構造設計能力、自動化能力、業界経験、長期的な実践の蓄積——である。そしてより深い希少性は、圧縮、購入、複製、短期間での達成が不可能なもの、すなわち時間そのもの、分野横断的経験、システム構築能力、長期運用記録である。先に論じたように、システム構築能力が大量のトークンを通じて直接獲得できない理由は、それが本質的に生成能力ではなく、判断能力、例外処理能力、トレードオフ能力、フィードバック吸収能力、信頼蓄積能力だからである。これらの能力に共通する特徴は、現実世界の時間、実践、エラー、修正、検証を必要とするということである。

収益性のレベルにおいても、この判断は同様に当てはまる。現在トークンで儲けることはすでに困難であり、ますます難しくなる。価格競争は利益幅を圧縮し続けており、単純なトークン転売のビジネスモデルは急速に低利益、さらには無利益へと向かっている。トークンがフロントエンドシステム、すなわちサービスを提供するシステムに付随して初めて、トークンと計算資源は有効な付加的資産として機能し、利益を生み出すことができる。トークンそれ自体は利益の源泉ではなく、システムこそが利益の源泉である。トークンはシステムの燃料に過ぎず、エンジンではない。

この観点から、私がすでに完成した、または開発中のシステムは、将来のトークン発展に対して特別な意義を持っている。なぜならこれらのシステムは、一時的な話題性を追いかけるために臨時に構築されたものではなく、トークンという概念が公の話題になるずっと前から存在していたからである。インテリジェント物流システムは、現実の業務が構造化、プロセス化、自動化できることを証明した。出版システムは、個人であっても多言語、マルチプラットフォーム、マルチ DOI、マルチアーカイブチャンネルの文献体系を確立できることを証明した。ウェブサイトとメタデータシステムは、デジタルコンテンツが継続的に組織化、索引付け、普及できることを証明した。自動相互参照システムはさらに、作品数が増え続けるとき、真に重要なのはコンテンツを積み上げ続けることではなく、コンテンツ間の構造的関係を確立することであることを証明している。『The Epochal Transition』ウェブサイト (<https://times.net.au>) は、これらの成果を集中展示するプラットフォームとして、トークン能力を公開され、永続的で、検証可能な学術体系に変換することの実証例そのものである。

将来、すべての人に同じ量のトークンが与えられたとしても、最終的に生み出される結果は全く異なるものになるだろう。より多くの対話記録を得る人もいれば、より多くのウェブページコンテンツを得る人もいる。一方で、数十年にわたって稼働し続ける物流システム、出版システム、知識体系を構築する人もいるだろう。差異を決定するのはトークンそのものではなく、その背後にあるシステムアーキテクチャであり、そしてそれらのシステムアーキテクチャを支える、圧縮不可能な時間、購入不可能な分野横断的経験、複製困難な構築能力、短期間では達成できない長期運用記録なのである。

したがって、私はトークンを終着点とは見なさず、インテリジェント時代の新たなインフラとして見なしている。トークンは電気のように重要であるが、電気自体が自動的に工場を建設するわけではない。トークンは通貨のように計量可能であるが、通貨自体が自動的に優良資産を創造するわけではない。トークンは水のように流動できるが、パイプ、調達システム、長期的な保守記録がなければ、それはただ流失するだけである。将来真に価値のある人は、必ずしも最も多くのトークンを持つ人ではなく、トークンをシステムに組織化できる人である。トークン時代の最大の誤解は、トークンを目標とすることである。そしてトークン時代の最大の機会、トークンを利用して他人が複製できない構造を構築することである。将来、トークンが電気のように一般的になったとき、人々は今日のトークンブームを振り返り、人間間の格差を真に決定していたのは、決して持っているトークンの量ではなく、継続的に学習し、継続的に実践し、継続的に蓄積し、継続的にシステムを構築する能力を持っているかどうか、そしてこれらの能力の背後にある、圧縮、購入、複製、短期間での達成が不可能なものであることに気付くかもしれない。そして私の見解では、この能力こそが、トークン時代において最も希少で、最も複製が困難な富なのである。

研究を深めるにつれて、私は将来の人工知能時代の競争構図はおそらく一つの単純な比喻で要約できるとますます感じるようになった。多くの人々は計算資源、モデル、トークンを中核的資源と見なしているが、現実世界の運用法則から見れば、それらはむしろ軍隊の背後にある兵站体系に似ている。計算資源は生産能力を決定し、モデルは組織能力を決定し、トークンは輸送能力を決定する。これらは共同して人工知能時代の物資保障体系を構成する。

しかし、戦場の結果を真に決定するのは、兵站物資そのものではなく、前線の作戦体系である。将来の産業競争にとって、システムはむしろフォワード（前線）に似ている。システムは資源の使用方法、プロセスの実行方法、価値の創造方法、ユーザーが真に受益するかどうかを決定する。トークンと計算資源がどんなに豊富でも、システムによる組織化が欠如していれば、それらは倉庫に積まれた物資のように、現実の成果に変換することができない。

この観点から見ると、将来の競争はおそらく計算資源間の競争でもトークン間の競争でもなく、システム間の競争である。システムはフォワードであり、攻撃の方向性と作戦能力を決定する。トークンは弾薬と補給であり、持続的な作戦能力を決定する。計算資源は後方の工業体系であり、生産速度と供給能力を決定する。後方の供給なしではフォワードは長期間前進できない。しかし後方の供給だけがあり、フォワード体系がなければ、同様に成果を上げることはできない。

したがって、計算資源がますます普及し、モデルがますます接近し、トークンがますます豊かになったとき、真に希少なものは依然としてシステムである。誰がより成熟したシステムを持ち、より現実的なデータを持ち、より長期的に検証されたプロセスを持ち、トークンと計算資源を現実のニーズに結びつけることができるかが、将来持続的な優位性を得る可能性が高い。

トークンはAI時代の「電気」になりつつある——不可欠ではあるが、目的そのものではない。電気革命の勝者が発電所ではなく、電球、モーター、電解槽を発明した発明家や起業家であったように、AI時代の真の勝者は、トークンを利用して複製不可能な構造と体系を構築できるシステム構築者である。トークン以後、これこそが真に希少なものである。

本稿全体を振り返ると、私は将来の人工知能時代の競争構図はおそらく一本の木に例えられるといいます感じる。

多くの人々は水と養分に注目する。なぜならそれらは見え、触れ、迅速に計測できるからである。今日の計算資源、モデル、トークンは、絶えず送り込まれる水と養分のようなものである。計算資源は生産能力を決定し、モデルは変換能力を決定し、トークンは輸送能力を決定する。技術の進歩に伴い、これらの資源はますます豊富になり、ますます入手しやすくなっている。

しかし、一本の木の生命力を真に決定するのは、水や養分そのものではなく、根系である。

システムこそが未来のデジタル文明の根である。

根系は水の吸収方法、養分の分配方法、木全体が長期的に成長できるかどうかを決定する。根がなければ、どんなに多くの水も流失する。根がなければ、どんなに豊富な養分も生命力に変換できない。同様に、システムがなければ、どんなに強力なモデル、どんなに安価なトークン、どんなに膨大な計算資源も、持続的に価値を創造することは難しい。

この観点から見ると、計算資源は水であり、トークンは養分であり、モデルは養分を運ぶパイプであり、システムは現実世界に深く根ざした根系である。将来の競争結果を真に決定するのは、誰がより多くの水を持っているかでも、誰がより多くの養分を持っているかでもなく、誰がより深く、より安定し、これらの資源を長期的に吸収・利用できる根を持っているかである。

将来、計算資源が成長し続け、モデルが進化し続け、トークン価格が下落し続けるにつれて、資源そのものはますます一般的になるだろう。真に希少になるのはもはや資源ではなく、資源を組織し、利用し、価値を創造し続けることができるシステムである。誰が成熟したシステムを持っているかが、根を持っているかであり、誰が根を持っているかが、長期的な生存と持続的成長の能力を持っているかである。

したがって、私はトークン以後に真に希少なものは、より多くのトークンでもなく、より多くの計算資源でもなく、計算資源、モデル、トークンを現実の価値に変換できるシステムであると考える。

水と養分がますます豊かになるとき、一本の木が長期的に成長できるかを真に決定するのは、依然として根なのである。

キーワード：トークン、人工知能、AI、大規模言語モデル、計算資源、モデル経済、トークン経済、中国語トークン、英語トークン、BPE アルゴリズム、データ主権、エージェント、API エコシステム、システム構築能力、システム設計、自動化システム、分野横断的研究、分野横断的実践、現実の履歴データ、長期運用システ

ム、時間資産、プラットフォーム競争、システム競争、デジタル文明、文明の記憶、知識インフラ、独立研究者、国際独立研究者、インテリジェント物流システム、10 言語出版システム、メタデータシステム、自動相互参照、DOI、Zenodo、WorldCat、TROVE、The Epochal Transition、巫朝晖、ジェフィ・チャオ・フイ・ウー

参考文献：

1. *A Structural Case Study of the 1993 Remote Work Prototype System*
DOI: <https://doi.org/10.5281/zenodo.17978371>
2. *The 2005 Intelligent Invoice System*
DOI: <https://doi.org/10.5281/zenodo.19044755>
3. *Mature Structural Logistics System*
DOI: <https://doi.org/10.5281/zenodo.20201895>
4. **Cross-Generational 20-Year Intelligent Logistics and Financial System**
DOI: <https://doi.org/10.5281/zenodo.20352371>
NLA: <https://nla.gov.au/nla.obj-4204311557>
5. *Analysis of the Homepage Structural Evolution of The Epochal Transition*
DOI: <https://doi.org/10.5281/zenodo.20362574>
NLA: <https://nla.gov.au/nla.obj-4204309341>
6. *An Empirical Study of Institutional Engineering: Independent Scholars Accessing the WorldCat Global Bibliographic Infrastructure*
DOI: <https://doi.org/10.5281/zenodo.18028572>
7. *National-level Network Archiving and Digital Document Preservation Guidelines of the Australian Winner Information Network*
DOI: <https://doi.org/10.5281/zenodo.17888259>
8. *The Institutional Archiving of The Epochal Transition*
DOI: <https://doi.org/10.5281/zenodo.17932379>

WU, J. C. H. (2026). Source Declaration for Audiovisual and Derivative Adaptations of a Continuing Real-World Narrative. Zenodo.

<https://doi.org/10.5281/zenodo.18160116>

الذكاء عصر في حقًا شحيح هو ما: (التوكنز) الرموز بعد الاصطناعي؟

وو هوي تشاو جيغي، هوي تشاو وو: مستقل دولي باحث/المؤلف

ORCID: <https://orcid.org/0009-0009-5318-3027>

【تمهيدي بيان】

عبر الممارسة من عامًا ثلاثين من أكثر من المستمدة الشخصية نظري وجهات إلى المقال هذا يستند اللوجستية الخدمات وإدارة، بُعد عن العمل لتشمل، 1989 عام أستراليا في دراستي منذ) التخصصات عبر والبحث، اللغات متعدد والنشر، الثقافي والنشر، الويب مواقع وبناء، الآلية والمعالجة ملاحظاتي إلا المقال هذا يمثل ولا، التطور سريعة جديدة ظاهرة هي (Token) الرموز. (التخصصات نصيحة الأحوال من حال بأي يشكل ولا، (2026 يونيو) التحرير وقت حتى المرحلية الشخصية وأحكامي صناعيًا معيارًا أو، تجارية توقعات أو، استثمارية.

تغير استمرار مع. صناعي إجماع أي تمثيل يدعي ولا، عالمية تطبيق قابلية إلى المقال هذا يسعى لا هذا في الأحكام بعض تحتاج قد، العالمية السياسية والبيئات، الاصطناعي الذكاء وتقنيات، الرموز سوق أحدث على بناءً مستقلة أحكام إصدار القارئ على يجب. المستقبل في تحديث أو مراجعة إلى المقال المعلومات.

مارستي من مستمدة أناقشها التي "الخمس الشخصية الأصول" و "النادرة الخمس القدرات" إن أكثر تضم قائمة وتشمل <https://times.net.au> على متاحة العامة الأدلة) الأنظمة بناء في الخاصة ISSN 3083- الدولي المعياري التسلسلي والرقم، الموقع ذلك على مستقلة إبداعًا عملاً 40 من بنى بواسطة دائم بشكل مؤرشفة وسجلات، DOI الرقمية للكائنات رقميًا معرفًا 60 من وأكثر، 5178 (CERN Zenodo ومرفق، الأسترالية الوطنية والمكتبة، WorldCat مثل دولية تحتية).

أن يجب المقال لهذا موجه وفعال صحيح تساؤل أو تفنيد أي فإن، سياقها من الكلمات انتزاع لتجنب المتداخلة والروابط المستند هذا في المرتبطة المقالات لجميع الكاملة القراءة: الأساسي شرطه يكون التلخيص أو الجزئي الاقتباس إن <https://times.net.au> على المدرجة أعمال جميع وكذلك، فيها المقال لهذا منصفًا تقييمًا يشكل أن يمكن لا الكلي السياق هذا عن المنفصل المبسط.

الفعال التفنيد حول تكميلية ملاحظة

الاستشهاد مسبقاً يفترض ولا ،عالمية تطبيق قابلية إلى يسعى ولا ،شخصية آراء عن المقال هذا يعبر ،حقائق مقابل حقائق : "بتفاحة تفاحة" مقارنة فعال تساؤل أي يتبنى أن يجب .الآخرين بأعمال قد ما" استخدام إن .مكتمة أنظمة مقابل مكتمة وأنظمة ،شخصية ممارسة مقابل شخصية وممارسة قدرات مقارنة أو ،"حالياً مكتمة ممارسات" لنفي "المستقبل في الاصطناعي الذكاء يحققه عادلة مقارنة يشكل لا ،فردية بممارسة مؤسسة/فريق

أنه تدعي ولا الشخصية للممارسة سجلاً المقال هذا في الموصوفة الخمسة الأصول مجموعة تشكل محددة عامة حالة تقديم فيرجى ،تماماً مماثلة حالة هناك أن تعتقد كنت إذا . "مماثلة حالة توجد لا" ،التخصصات متعددة والأنظمة ،الحقيقية التاريخية البيانات :الخمس الأبعاد عبر المقابلة أدلتها وشرح مجرد هو تساؤل أي فإن ،ذلك خلاف .الذاتي التحقق وقدرة ،المستقبلية والرؤية ،الأنظمة بناء وقدرة فعالاً تنفيذاً يشكل ولا رأي عن تعبير

الملخص

سيبين المقال هذا لكن ،الاصطناعي الذكاء عصر في المركزية القياس وحدة (Token) الرموز أصبحت رخصة الرموز ستصبح ،الاستدلال تكاليف انخفاض استمرار مع .النهائية القيمة ليس نفسه الرمز أن الرموز تصبح لا عندما .اليوم الصلبة الأقراص على التخزين وسعة الإنترنت مرور حركة مثل وشائعة عبر الممارسة من عامًا ثلاثين من أكثر إلى استنادًا حقًا؟ نادرًا يصبح الذي الشيء هو فما ،نادرة 2005، عام ذكي فواتير ونظام ،1993 عام بُعد عن العمل لنظام أولى نموذج من) للمؤلف التخصصات التبادلي والربط ،DOI معرفات وإدارة ،لغات بعشر النشر ونظام ،الذكاء اللوجستية الأنظمة إلى - الأنظمة بناء قدرة ،أولاً :التالية الجوهرية الأحكام المقال هذا يقدم ،(المنصات عبر والنشر ،التلقائي للتشغيل الذهنية والنماذج ،التخصصات متعددة والممارسة ،والأتمتة ،الهيكلية التصميم ذلك في بما لا ،ثانيًا .حقًا النادر الأصل هي - الواقعية الأعمال عمليات في الرموز دمج على والقدرة ،الأجل طويل اتخاذ على تعتمد لأنها الرموز من كبيرة كميات خلال من مباشرة الأنظمة بناء قدرة اكتساب يمكن والثقة ،الحقيقي العالم من الراجعة التغذية وحلقات ،والمفاضلات ،الاستثناءات ومعالجة ،القرار التخصصات متعددة الخبرة قابلية وعدم ،للضغط الوقت قابلية عدم ،ثالثًا .الوقت بمرور المتراكمة تشكل - الأجل طويلة التشغيل سجلات اختصار واستحالة ،الأنظمة بناء قدرة نسخ وصعوبة ،للشراء الذي الأكاديمي الويب موقع المقال هذا يستخدم .الرموز عصر في يكون ما أندر الأربعة الأصول هذه (<https://times.net.au/>، ISSN 3083-5178) "The Epochal Transition" العصر التحول" المؤلف بناه CERN Zenodo، وWorldCat مثل دولية تحتية بنى بواسطة دائم بشكل والمؤرشف ،5178 بناء واحد شخص قيام لكيفية عرضًا مقدمًا ،عام كدليل (الأسترالية الوطنية للمكتبة التابع TROVE و) عامًا ثلاثين مدار على الأمد وطويل للتحقق وقابل التخصصات متعدد نظام

هو حقاً نادر هو ما ،الرموز بعد .المحرك هو والنظام وقود؛ مجرد هي الرموز :أن المقال هذا يؤكد بناها التي والهياكل ،الرموز تسخير على القادر الشخص

"المال؟ فائدة ما ،بضائع بدون"

الثورات من العديد كبير حد إلى يشبه أنه متزايد بشكل أشعر ،الرموز تطور تاريخ أراجع عندما أولاً الناس يركز ،جديد أساسي مورد يظهر مرة كل في .الماضي في البشرية بها مرت التي التكنولوجيا وفي ،الصناعي العصر في صحيحاً هذا كان .وراءه يقف الذي الهيكل متجاهلين ،نفسه المورد على من العديد تركز ،اليوم .الاصطناعي الذكاء عصر في القدر بنفس صحيح وهو ،الإنترنت عصر أن لو كما ،الرموز توريد وقدرة ،الرموز وحصص ،الرموز وتكاليف ،الرموز أسعار على المناقشات يكن لم ،العملية الخبرة من لعقودي وفقاً ولكن .المستقبل امتلاك يعني الرموز من المزيد امتلاك الموارد لهذه الناس تنظيم كيفية دائماً هو حقاً القيمة يحدد ما .أبدأ النهائية القيمة هو نفسه المورد واستخدامها

من .تاريخها من يبدأ أن يجب بل ،جديد مصطلح من ببساطة يبدأ أن للمرء يمكن لا ،"الرموز" لفهم عام الورق على "the" لكلمة مثلاً عشرين يكتب وهو بيرس ساندرز تشارلز الأمريكي الفيلسوف الذي ،الصين في للبيانات الوطنية الإدارة مدير ،هونغ لي ليو إلى ،"المثال" و "النوع" بين مفرقاً ،1906 قطعت ،2026 عام أو بو متدى في Token لكلمة (代币) "ciyuan" الرسمية الصينية الترجمة أعطى الذكاء إلى (بلوكشين) الكتل سلسلة ومن ،الكمبيوتر علوم إلى الفلسفة من طويلاً شوطاً المفهوم هذا في ،كانت حيث ،الغربي الكمبيوتر عالم في طويلة فترة منذ موجودة "Token" كلمة كانت .الاصطناعي للكمبيوتر يمكن التي الأساسية الوحدة إلى تشير ،النصوص معالجة وأنظمة والمترجمات البرمجة لغات برمجة واجهة واستدعاءات ،والمصادقة ،الشبكات أمن سياقات في ،لاحقاً .ومعالجتها عليها التعرف بيان "أو "مرور تصريح" أنه على "Token" يفهم كان ما غالباً ،الكتل وسلسلة ،(API) التطبيقات عصر في لكن .الرقمية الأصول أو ،الهوية اعتماد بيانات أو ،الوصول حقوق إلى يشير مما ،"اعتماد سلسلة أصول من أصلاً ولا ،هوية اعتماد بيان الأول المقام في Token يعد لم ،الكبيرة اللغوية النماذج البرمجية والرموز اللغة الاصطناعي الذكاء نماذج بها تعالج التي الأساسية الحوسبة وحدة هو بل ،الكتل الوسائط متعددة والمعلومات الصور وأوصاف

هذه أن أعتقد .(代币) "ciyuan" الاصطناعي الذكاء مجال في Token بتسمية لاحقاً الصين قامت ليس الكبيرة النماذج في Token لأن ،(المرور تصريح) "lingpai" تستخدم لم .ما حد إلى دقيقة ترجمة ،حرفاً أحياناً يكون أن يمكن لأنه ،(كلمة) "ci" أو (حرف) "zi" إلى ببساطة تُرجمت ولا مرور؛ تصريح إلى "Ci" تشير .رمز أو ترقيم علامة مجرد حتى وأحياناً ،إنجليزية كلمة نصف وأحياناً ،كلمة وأحياناً فهمها يمكن ،العام للقارئ بالنسبة .أساسية وحدة أنها إلى "Yuan" وتشير ،والمعلومات باللغة علاقتها

هذا على الرموز من كسلسلة الاصطناعي الذكاء يراه ،كجمله البشر يراه ما :التالي النحو على ببساطة الذكاء عصر في المركزية القياس وحدة إلى غامض لغوي مصطلح من الرموز تحولت ،المسار شخصياً اختبر كممارس .التجاري بالطلب التكنولوجي العرض تربط تسوية أداة إلى ثم ،الاصطناعي عبر المحتوى وتوزيع ،آلية نشر عملية وتصميم ،ذكي لوجستي نظام بناء مثل الأجل طويلة مشاريع ليس حقاً القيمة يحدد ما وسيلة؛ بل ،غاية أبداً الرموز تكن لم – التطور لهذا عميق فهم لدي ،المنصات لا وهياكل باستمرار تعمل أنظمة لبناء الرموز استخدام يمكنه من بل ،أكثر رموزاً يمتلك من أبداً علناً المقال هذا في نوقشت التي الأنظمة بناء قدرة نتائج عرض تم .تضاهي وتم مستقل باحث بناه التخصصات متعدد أكاديمي نشر نظام وهو – <https://times.net.au> على العالمي المكتبات نظام في دائم بشكل أرشفته

تدرجياً طورت بل ،الرموز ظهور بعد فقط الأنظمة في بالتفكير أبداً لم ،الخاصة رحلتي في بالتأمل مررت ،1989 عام أستراليا في دراستي من بدءاً .الأجل طويلة الممارسة خلال من للأنظمة فهماً ،بعد عن العمل في تدرجياً تورطت ،لاحقاً .مختلفة مجالات في والاستكشاف والعمل التعلم بتجارب ،اللغات متعدد والنشر ،الثقافي والنشر ،الويب مواقع وبناء ،الآلية والمعالجة ،اللوجستية الخدمات وإدارة عصور في حدثت التجارب هذه أن من الرغم على ،الوراء إلى بالنظر .التخصصات متعدد والبحث العمل وتقليل ،الهياكل إنشاء كيفية حول دائماً دارت لقد :مشتركة خاصة في تشترك أنها إلا ،مختلفة اللغات نماذج ولا ،الوقت ذلك في موجوداً الرموز مفهوم يكن لم .العامة الكفاءة وتحسين ،المكرر تلقائياً وتعالج وتنظم تتدفق المعلومات جعل كيفية في باستمرار أفكر كنت لكنني ،اليوم الكبيرة

اعتبارها إلى أكثر أميل فإنني ،جديدة إنتاج كوسيلة الرموز اليوم الناس من متزايد عدد يعتبر بينما ،لذلك تلقائياً قيمة تخلق لا الطاقة لكن ،بالتأكيد مهمة الطاقة .الأساسية الطاقة أشكال من جديداً شكلاً لا وحدها الرموز تجارية؛ مؤسسة تلقائياً يصبح لا وحده الإنترنت مصنعاً؛ تلقائياً تصبح لا وحدها الكهرباء من كبيرة كميات على الحصول من الجميع تمكن لو حتى ،المستقبل في .معرفياً نظاماً تلقائياً تصبح كمية ليس الفرق يحدث ما .النتائج نفس تحقيق يمكنهم الجميع أن يعني لا فهذا ،بسهولة الرموز .الرموز فيه تدخل الذي الهيكل بل ،الرموز

العملات تشبه الرموز :أخرى مهمة ملاحظة إلى الإشارة إلى أحتاج ،للرموز متعمق تحليل إجراء قبل لا كما تماماً .عالمية غير تجعلها والغرب الشرق بين والسياسية والنظامية التكنولوجية والاختلافات الإنجليزية والرموز الصينية الرموز تواجه ،بحرية الصيني والرميني الأمريكي الدولار استبدال يمكن الاختلافات في يكمن الأعرق السبب لغوي؛ اختلاف مسألة مجرد ليست هذه .مماثلة تداول حواجز وبيانات ،الترميز وخوارزميات ،النماذج هياكل) التكنولوجي المستوى على والغرب الشرق بين الثلاثية (البيانات ومعايير ،الدفع وأنظمة ،البيئية التطبيقات برمجة وإجهات أنظمة) النظامي والمستوى (التدريب ،القومي الأمن ومراجعات ،التصدير وضوابط ،البيانات الحدود عبر التدفقات) السياسي والمستوى

لهذه "الاصطناعي للذكاء مثالية عالمية عالمية عملة" ك بحرية التداول من الرموز تمنع والتي مناطق "تظهر قد ،المستقبل في – الرموز لسوق المستقبل المشهد فهم على كبيرة آثار الملاحظة وستواجه ،والسياسة اللغة على قائمة للرموز أخرى بيئة ونظم ،"إنجليزية رموز مناطق" و ،"صينية رموز امثال وتكاليف الكفاءة في خسائر المناطق عبر الاستدعاءات

الوحدة إلى الفلسفي المفهوم من الرباعية القفزة :الرموز نشأة - أولاً الاقتصادية

مراحل أربع إلى بوضوح الرموز هوية تطور تقسيم يمكن

"النوع" بين ،للسيميائية دراسته أثناء ،بيرس ميز ،1906 عام في .الفلسفية اللسانيات في الأصل الواقع في القاعدة لتلك المحدد التجسيد هو الأخير بينما ،مجرداً شكلاً أو قاعدة يمثل الأول . "المثال" و .البنوية اللسانيات مهماً أساساً لاحقاً التمييز هذا أصبح

سطر يكتب المبرمج كان عندما ،العشرين القرن ستينيات في .الكمبيوتر علوم في البرمجي التحول إلى وما "5" و"=" و"x" و"int" مثل مستقلة وحدات إلى تفكيكه إلى الكمبيوتر احتاج ،"int x = 5;" الكود معلومات وحدة أصغر بذلك الرموز أصبحت .(Token) رمزاً الوحدات هذه من وحدة كل كانت .ذلك التعليمات "لقراءة" للجهاز

الأولى الطرح جنون صعود مع ،2017 عام في .(بلوكشين) الكتل سلسلة في القيمة حامل تجربة الرغم على . "لتداول قابلة رقمية حقوق شهادة" ك جديدة هوية Token ال أعطيت ،(ICO) للعملات ترسخ قد "للقيمة كرمز الرمز" فهم أن إلا ،تدرجياً هدأت هذه الحماس موجة أن من

الرئيس أعاد – فارقة لحظة 2026 مارس كان .الاصطناعي الذكاء لعصر الاقتصادية التحتية البنية ذكاء رموز تنتج مصانع" أنها على البيانات مراكز تعريف ،هوانغ جنسن ،NVIDIA لشركة التنفيذ الصينية الترجمة ،الصين في للبيانات الوطنية الإدارة مدير ،هونغ لي ليو وأعطى ،"ذكية اصطناعي أكملت ،النقطة هذه عند . "الذكي العصر قيمة مرساة" ب إياها واصفاً ،(詞元) "ciyuan" الرسمية اقتصادي مفهوم إلى تقني مفهوم من الأخيرة قفزتها الرموز

الضريبة"و (BPE) للبايتات الزوجي الترميز خوارزمية: التقنية القاعدة - ثانيًا الضمنية "اللغوية

تم تقنية على يعتمد الاصطناعي للذكاء القياسية القياس وحدة تصبح أن على الرموز قدرة سبب إن (Byte Pair Encoding, BPE) للبايتات الزوجي الترميز - الزمن من عقدين من لأكثر نسيانها

مجلة في البيانات ضغط خوارزميات حول مقالاً غيج فيليب الأمريكي المبرمج نشر، 1994 عام في "لحم" و متكرر بشكل النص مسح: BPE ل الأساسي المبدأ فيه قدم، C لغة في متخصصة تقنية تجذب لم. جولة بعد جولة تكراري بشكل الضغط مع، جديد رمز في تكراراً الأكثر المتجاورين الحرفين استثنائية تكن لم ضغطها كفاءة لأن الوقت ذلك في الانتباه الورقة هذه

تجزئة لمشكلة دراسته أثناء، إدنبرة جامعة في باحث، سينريش ريكو أن 2016 عام حتى يكن لم الحل كان BPE أن بثاقبة وأدرك القديمة الورقة هذه على بالصدفة عشر، الآلية الترجمة في الكلمات، "لتحدث" نفسها للبيانات يُترك بل مسبقاً؛ قاموس لتعريف حاجة هناك ليست - الترميز لمشكلة الأمثل رسمياً OpenAI اعتمدت، 2019 عام في. رموز إلى طيعي بشكل التردد عالية التركيبات اندماج مع مكن مما، "البايت" مستوى عند مباشرة للترميز البداية نقطة حددت حيث، GPT-2 إطلاق عند النهج هذا مكتوبة لغة أي معالجة من نظرياً النموذج

"لغوية ضريبة" نظام قصد غير عن شكل BPE لخوارزمية "التكرار أولوية" منطق فإن، ذلك ومع كفاءة فإن، الإنترنت بيانات مجموعات في المطلقة السائدة هي الإنجليزية اللغة لأن نظراً. ضمناً تستخدم، المعنى نفس عن للتعبير. الأخرى اللغات من بكثير أعلى الإنجليزية الرموز تجزئة في النموذج للغات وبالنسبة، ضعفاً 2 إلى 1.5 من عادةً الصينية اللغة وتتطلب، الرموز من عدد أقل الإنجليزية اللغة، يعني هذا. الإنجليزية اللغة تكلفة أضعاف 10 إلى 5 من التكلفة تكون أن يمكن، الأقل الموارد ذات كبيراً اختلافاً تختلف الاصطناعي الذكاء مع للمحادثة الفعلية التكلفة أن، رمز لكل التسعير ظل في المستخدمة اللغة على اعتماداً

قواعد - تصحيحه جداً يصعب، للنموذج الأولية المفردات في كتابته بمجرد، "البداية خط" في الظلم هذا استبدال صعوبة زادت، المبنى ارتفع فكلما للعالم؛ الاصطناعي الذكاء فهم "أساس" هي الترميز بالصينية الجملة نفس كانت، كمثال GPT سلسلة بأخذ. يتحسن الوضع هذا، الحظ لحسن. الأساس في تحسن - GPT-5 في 15 وإلى، GPT-4 في 26 إلى وانخفضت، GPT-3 في رمزاً 38 تتطلب Tongyi مثل المحلية الكبيرة الصينية النماذج قامت، ذلك على علاوة 60٪ من بأكثر الكفاءة ذلك شابه وما، الاصطلاحية والتعابير، التردد عالية الصينية العبارات بدمج DeepSeek و Qianwen "أصالة" أكثر الصينية اللغة معالجة كفاءة محققة، التصميم مرحلة منذ مفرداتها في أصلية كرموز

ميزة كبير حد إلى يمتلك "الدلالية التجزئة في الحق" على يسيطر من: عميق مبدأ عن هذا يكشف الصك حق" أشكال من شكل وهذا – الرقمي العالم في اللغة تلك عن للتعبير والتكلفة الكفاءة الرقمي العصر في "الأساسي.

الوكلاء إلى المحادثة من الأسية القفزة: والإنتاج الاستخدامات - ثالثاً (الأجنتات)

إلى ثم "الاستدلال" إلى "المحادثة" من، المستويات ثلاثية قفزة الرموز استهلاك سيناريوهات تشهد "الوكلاء".

في رمز 10000 حوالي الواحدة والاستجابة المطالبة تستهلك، للمحادثة الاصطناعي الذكاء مرحلة في المعقد والاستدلال التفاعل من متعددة جولات تتضمن التي، للاستدلال الاصطناعي ذكاء مرحلة يمكن، الوكيل الاصطناعي الذكاء مرحلة إلى الدخول عند. رمز 100000 حوالي مهمة كل تستهلك الرموز استهلاك ليصل، متعددة أدوات واستدعاء مستقل بشكل الخطوات متعددة مهام تنفيذ للوكلاء يتوقع. التطور من فقط جيلين مدى على ضعف مائة تقارب بزيادة – رمز مليون حوالي إلى مهمة لكل كل الرموز على العالمي الطلب سيرتفع، 2030 عام بحلول أنه، أمون، كوالكوم لشركة التنفيذي الرئيس الشعبية أثبتت لقد مبالغة؛ ليست هذه. تريليون 1.27 إلى 2026 عام في ملياراً 31.7 من ثوانٍ 10 التي الواحدة المعقدة فالمهمة: النقطة هذه OpenClaw مثل المصدر مفتوحة الوكلاء لأدوات الهائلة شائع أمر الرموز من الملايين عشرات تستهلك.

من اعتباراً أنه العامة البيانات تظهر. العالم في للرموز مورد أكبر الصين أصبحت، الإنتاج جانب على ألف عن تزيد بزيادة، رمز تريليون 140 الصين في اليومي الرموز استدعاء حجم تجاوز، 2026 مارس إنشاء يعتمد. الاصطناعي الذكاء لرموز العالمي الاستخدام من 61٪ يمثل ما وهو، عامين في مرة تلك من بكثير أقل الصين في الصناعية الكهرباء أسعار، أولاً: مزايا ثلاث تراكم على المكانة هذه الصين بنت، ثانياً الرموز؛ إنتاج تكاليف من 60٪ من أكثر الكهرباء وتمثل، وأمريكا أوروبا في الموجودة، الشرق في الحوسبة" توزيع خلال من فعالاً إنتاجاً محققة، العالم في ذكية حوسبة مجموعة أكبر خدمات ونظام، بالكامل المصدر مفتوحة المحلية الكبيرة الصينية النماذج، ثالثاً؛ "الغرب في البيانات من الحلقة ومغلقة كاملة صناعية سلسلة يشكل مما، ناضج (API) التطبيقات برمجة واجهات تسويق "العالمي الاستدعاء - الاستدلال إخراج - الحوسبة دعم - النموذج تدريب".

الأركان الثلاثي الصناعي المشهد: والأسواق المؤسسات - رابعاً

الرموز حول الطبقات واضح سوق هيكل تشكل.

يركز. الرموز وراء الأساسي الدعم نظام في النظر أيضاً الضروري من، الصناعي الهيكل هذا لفهم الرموز وراء. مستقلاً مورداً ليست الرموز أن يتجاهلون لكنهم الرموز أسعار على الناس من العديد وتحدد، النموذج قدرة الحوسبة قوة تحدد، ما بمعنى. الحوسبة قوة توجد النماذج وخلف، النماذج توجد السوق إلى النموذج قدرات تصدير عند تشكل التي القياس وحدة هي والرموز، الرموز إنتاج النماذج قوة في منافسة يزال لا، جوهره في لكن، الأسعار منافسة عن هو الرموز سوق، السطح على، لذلك والكفاءة الحوسبة.

هو الأعلى المستوى. مستويات ثلاثة إلى تقريباً الحالية العالمية الرموز صناعة سلسلة تقسيم يمكن الحوسبة ومنصات، الخوادم ومصنعي، الرقائق مصنعي ذلك في بما، الحوسبة لقوة التحية البنية الأساسية النماذج مؤسسات من المتوسط المستوى يتكون. الطاقة وأنظمة، البيانات ومراكز، السحابية يتكون. للاستدعاء قابلة رمز خدمات إلى الحوسبة قوة تحويل عن المسؤولية، الاستدلال ومنصات تحول والتي، الصناعة تطبيقات ومختلف، الوكلاء ومنصات، الرموز خدمات مقدمي من الأدنى المستوى، الرمز سعر هو النهاية في المستخدمون يراه ما. نهائية ومنتجات مؤسسية خدمات إلى النماذج قدرات المنع في الحوسبة قوة نظام في التغيير غالباً هو وسرعته الرمز تكلفة على حقاً يؤثر ما ولكن النموذج وكفاءة.

كفاءة وزيادة، النماذج هياكل وتحسين، الرقائق أداء تحسين استمرار مع، الماضية القليلة السنوات في أصبحت الرموز أن الناس من العديد يرى. سريعاً هبوطياً اتجاهًا رمز لكل التكلفة أظهرت، الاستدلال ستشمل، لذلك. الحوسبة قوة لنظام المستمر التقدم هو التغيير لهذا الحقيقي المحرك لكن، أرخص أيضاً ولكن، والأسعار المنصات بين المنافسة فقط ليس الرموز سوق في المستقبلية المنافسة الموارد تكامل وقدرات، النماذج وكفاءة، الحوسبة موارد في المنافسة.

على NVIDIA تمثلها التي الرقائق شركات تعمل. الحوسبة قوة منتج - المنع في "الطاقة محطات" أن NVIDIA تتوقع. الحوسبة قوة من وحدة لكل الرموز إنتاج كفاءة لتحسين باستمرار الأجهزة تحسين في العالمية الرغبة تدعمها ثقة وهي، 2027 عام بحلول الأقل على دولار تريليون إلى إيراداتها تصل الاصطناعي للذكاء الحوسبة قوة.

و OpenAI مثل شركات تقوم. والمجمعون النماذج منصات - النهر مجرى في "الشبكة تنظيم" وقابلة موحدة رمز خدمات إلى الأساسية الحوسبة قوة بتحويل Alibaba و DeepSeek و Google تمثلها التي) الرموز جميع منصات: الحلقة هذه حول ناضجة أعمال نماذج ثلاثة ظهرت. للاستدعاء

26.9 الأسبوعية استدعاءاتها حجم وبلغ، والعمولات الأسعار فروق على تعمل والتي، OpenRouter، الذكاء ترحيل ومحطات، السحابية الخدمات لمقدمي MaaS وخدمات، (2026 مايو بحلول رمز تريليون الثاني السوق في النشطة الاصطناعي.

للول WeChat لواجهة Tencent تحويل. والتوزيع التطبيق منصات – المصب في "الدخول نقاط" China اقترحت؛ "الرموز توزيع منصة" إلى أساسي بشكل WeChat يحول الجديدة النماذج قدرات إلى المرور حركة منطق من للخروج محاولة في، "الرموز إدارة خلال من أعمالها تشكيل إعادة" Telecom Telecom النموذج يواجهون يعودوا لم المستخدمين أن هو دخول كنقاط الشركات هذه فيه تشترك ما. التقليدي "مرئي غير" أصبح نفسه النموذج – المنصة خلال من مباشر غير بشكل الرموز يستدعون بل، مباشرة

، ملموس غير إنتاج: الرموز حول تشكل قد العالمية للتجارة نوعه من فريداً نموذجاً أن بالذكر يجدر مزود" إلى التقليدي "العالم مصنع" من الصين يحول هذا. عالمية وتغطية، الضرائب من معفي وتشغيل "الاصطناعي الذكاء حوسبة لقوة التحتية للبنية عالمي

من العديد يركز. الحوسبة قوة: تجنبه يمكن لا آخر أساسي عامل هناك، الرموز اقتصاد مناقشة عند الرموز إنشاء يتم لا، الواقع في. الرموز وراء الإنتاج عملية يتجاهلون لكنهم الرموز أسعار على الناس أساس على النموذج استدلال ويعتمد، استدلال إجراء النموذج من لرمز إنشاء كل يتطلب. شيء لا من إنتاج أدوات هي والنماذج، الرموز اقتصاد مصدر هي الحوسبة قوة، ما بمعنى. هائلة حوسبة قوة من الخارج إلى النموذج قدرات تصدير عند تشكل التي القياس وحدة هي والرموز، الرموز

كانت كلما. الرموز توليد وسرعة الرموز تكاليف على مباشر بشكل الحوسبة قوة مستويات تؤثر، لذلك إنتاجها يمكن التي الرموز عدد وزاد، المعلومات معالجة على النموذج قدرة زادت، أقوى الحوسبة قوة السنوات في. عادةً الرموز أسعار انخفضت، الحوسبة قوة تكلفة انخفضت وكلما زمنية؛ وحدة لكل فقط ليس مدفوعاً، الرموز أسعار في مستمراً انخفاضاً الكبيرة النماذج صناعة شهدت، الماضية القليلة وزيادة، البيانات مراكز حجم وزيادة، الرقائق أداء بتحسينات أيضاً ولكن، النماذج خوارزميات بتحسين المحرك لكن، أرخص أصبحت الرموز أن الناس من العديد يرى، أخرى بعبارة. الاستدلال كفاءة الحوسبة قوة لنظام المستمر التقدم هو التغيير لهذا الحقيقي

المنع في. كاملة سلسلة الواقع في والرموز والنماذج الحوسبة قوة تشكل، الصناعة سلسلة منظور من الأساسية النماذج توجد النهر مجرى في الطاقة؛ وأنظمة البيانات ومراكز والخوادم الرقائق توجد المختلفة الصناعية والحلول الوكلاء وتطبيقات الرموز خدمات توجد المصب في الاستدلال؛ ومنصات في الرموز وتحدد الرموز؛ وكفاءة جودة النماذج تحدد النموذج؛ لقدرة الأعلى الحد الحوسبة قوة تحدد عليها الحصول للمستخدم يمكن التي الاصطناعي الذكاء خدمة مقدار النهاية

نفسها الحوسبة قوة فإن ،إنكارها يمكن لا الحوسبة قوة أهمية أن من الرغم على أنه أعتقد ،ذلك ومع طاقة محطة امتلاك .التاريخ في مماثلة بعمليات الصناعات من العديد مرت لقد .تلقائياً قيمة تخلق لا امتلاك ،وبالمثل .أفضل منتج امتلاك يعني لا أكثر خوادم امتلاك أفضل؛ مؤسسة امتلاك يعني لا أكبر بينما ،الإنتاجية القدرة مشكلة الحوسبة قوة تحل .أعلى قيمة ذي نظام امتلاك يعني لا أكبر حوسبة قوة تعزيز واستمرار ،الحوسبة قوة نمو استمرار مع ،المستقبل في .القيمة تحويل مشكلة الأنظمة تحل ،نفسها الحوسبة قوة يكون لن حقاً نادراً سيكون ما فإن ،الرموز موارد تكاثر واستمرار ،النماذج قدرات الحقيقي العالم في الطويل المدى على العمل يمكنها أنظمة في الموارد هذه تنظيم كيفية بل

ستحول والرموز والنماذج الحوسبة قوة في المنافسة أن أعتقد ،المستقبلية المنافسة منظور من ،لذلك أكثر النماذج وتصبح ،انتشاراً أكثر الحوسبة قوة تصبح عندما لأنه .الأنظمة في المنافسة نحو تدريجياً يمكنه من بل ،المملوكة الموارد كمية يكون لن الفرق سيحدث ما فإن ،وفرة أكثر الرموز وتصبح ،تشابهاً .الأجل طويلة فعالة هياكل لبناء الموارد هذه استخدام

(الإنفلوشن) الاستنزافية الأسعار حرب :والحواجز المنافسة - خامساً

القيمة وتقسيم

انخفضت ،2026 عام إلى 2022 عام من .عالية استنزافية خصائص الرموز سوق في المنافسة تظهر الإدخال سعر :أهمية أكثر سعريّة مزايا المحلية النماذج تظهر .99.9% بنسبة الواحد الرمز استدلال تكلفة ل الإخراج وسعر ،رمز مليون لكل صيني يوان 0.025 إلى يصل منخفض DeepSeek V4-Pro ل إخراج بسعر مقارنة ،رمز مليون لكل صيني يوان 1.6 هو Tongyi Qianwen Qwen3.5-Plus صينيّاً يواناً 72 حوالي البالغ OpenAI GPT-4o .

أن الصناعة إحصاءات تظهر .حاد بشكل "الوسطاء" ربح هامش ضغط إلى هذه الأسعار حرب أدت ليشكل ،منصة ألفي تجاوز قد (API) التطبيقات برمجة لواجهات المختلفة والترحيل التجميع منصات عدد والنماذج القوية التكنولوجيا ذات الكبيرة المنصات من العليا الطبقة :رئيسية طبقات ثلاث السوق والعدد عملاء؛ وموارد رسمية خصومات لديهم الذين المحترفين الوكلاء من الوسطى والطبقة الشاملة؛ في المتمثل الأعمال نموذج يعد لم .شديدة بقاء ضغوط يواجهون الذين جداً الصغار الوكلاء من الهائل .مستداماً الرموز أسعار فروق من الاستفادة مجرد

محركات تطوير خلال من – التقنية العلاوة اتجاه هو الأول .اتجاهين في الحقيقية الحواجز تتحول ربحاً يحقق يزال ولا الرسمي المزود من أرخص بسعر البيع للمرء يمكن ،ذاتياً مملوكة استدلال تسريع والضبط ،المطالبات هندسة مثل مخصصة خدمات تقديم – المضافة القيمة خدمات اتجاه هو الثاني

السيناريو على بناء هائلاً تمايزاً تظهر الرموز قيمة أن، بروزاً والأكثر. الخاص والنشر، للنماذج الدقيق للرموز بالنسبة بينما، مليون لكل دولار 0.01 إلى يصل منخفض بالردشة المتعلقة الرموز سعر متوسط قيمة بفارق، مليون لكل دولار 1000 إلى السعر متوسط يصل أن يمكن، الأدوية باكتشاف المتعلقة بتكلفة تتحدد لا الرمز قيمة: جوهري مبدأ عن هذا يكشف. السيناريوهات عبر ضعف ألف مائة إلى يصل يولد مختلفة أنظمة في ودمجه مختلفة سيناريوهات في الرمز نفس فاستخدام - باستخدامه بل، إنتاجه تماماً مختلفة قيمة.

يمكن. مرتفعاً ليس الرموز بيع حاجز، المستقبل وفي حالياً: هنا أهمية أكثر حكم إصدار يجب ولكن في والبدء، تجميع منصة وإنشاء، (API) تطبيقات برمجة واجهة على للحصول التسجيل شخص لأي على والقدرة، الحقيقية التاريخية البيانات - الرموز بيع يتجاوز ما هو حقاً نادر هو ما. الرموز بيع إعادة منها فائدة لا والرموز الكبيرة النماذج كل، هذه بدون. المستقبلية والرؤية، والهياكل الأنظمة بناء.

حروب تعمل. صعوبة وسيزداد، بالفعل صعب الرموز من المال كسب: الرموز ربحية حول تكميلي حكم يتجه الرموز بيع إعادة مجرد في المتمثل الأعمال ونموذج، الربح هوامش ضغط على باستمرار الأسعار عندما أي - أمامي بنظام مرتبطة الرموز تكون عندما فقط. معدومة حتى أو ضئيلة أرباح نحو بسرعة الأرباح لتوليد فعالة تكميلية أصول بمثابة تكون أن الحوسبة وقوة للرموز يمكن - خدمات نظام يقدم وقود مجرد هي الرموز. الربح مصدر هو النظام الربح؛ مصدر ليست نفسها الرموز، أخرى بعبارة محرکه وليست، للنظام.

من متزايد عدد ظهر، الاصطناعي الذكاء عصر في مهمة قياس وحدة إلى تدريجياً الرموز تحول مع أيضاً هناك، الأساسية النماذج بائعي إلى بالإضافة. السوق في الرموز حول المتمركزة الأعمال نماذج على التطبيقات برمجة واجهات توزيع وخدمات، الوكلاء ومنصات، النماذج تجميع ومنصات، الرموز وكلاء تبدو، الأعمال رواد من للعديد بالنسبة. الوسيطة الطبقة خدمات مقدمي ومختلف، المؤسسات مستوى مراكز بناء أو نموذج تدريب تتطلب لا لأنها الاصطناعي الذكاء صناعة لدخول اختصار وكأنها الرموز وكالة فإن، ذلك ومع. التجارية الأعمال ممارسة لبدء الحالية الموارد دمج فقط الأمر يتطلب بل كبيرة؛ حوسبة دخول وحواجر، نسبياً منخفضاً أولياً استثماراً مزاياها تشمل. واضحة وقيود مزايا له النموذج هذا، المنع في النماذج موردي على كبيراً اعتماداً قيودها تشمل السوق؛ في أسرع وتوسعاً، نسبياً منخفضة أسعار انخفاض استمرار مع. الأجل طويلة تقنية حواجز إلى والافتقار، الأسعار لمنافسة والتعرض المدى على، لذلك. البسيطة الرموز بيع إعادة من الربح هامش على باستمرار الضغط يتم قد، الرموز طويلة القيمة يحدد ما. نهائية أساسية تنافسية قدرة وليست توزيع قدرة الرموز وكالة تعتبر، الطويل والحلول، المعرفة وأنظمة، الأتمتة وعمليات، بالصناعة الخاصة الأنظمة يزال لا حقاً للمؤسسة الأجل الرموز حول المبنية الواقعية.

المستقبل ومسارات الحالية الصعوبات - سادساً

العميقة المشاكل من العديد الحالي الرموز سوق يواجهه.

حروب وتستمر، العملاء مجموعة نفس على منصة ألفي من أكثر تنافس. المتجانسة المنافسة فخ للحوسبة الأولى الأيام مع السيناريو هذا يتكرر قد، تغيير بدون. الصناعة أرباح خفض في الأسعار "السوق في حصة لكسب المال يخسرون" كانوا حيث، السحابة.

مؤقتة حالة هي محدد تشريع وجود عدم في المتمثلة الحالية "التنظيم أرباح". الامتثال مخاطر خطر والأمن، الحدود عبر البيانات مثل قضايا فإن، الرموز سوق توسع مع. الأجل طويل معياراً وليست أن ملاحظة خاص بشكل المهم من. التشريع تسريع على الدول ستجبر الضرائب وتحصيل، القومي إثارة إلى بسهولة تؤدي قد مالية أدوات وجعلها الرموز لترميز السوق في اللاعبين بعض محاولات الإضرار إلى أيضاً بل، التجارية الأنشطة هذه إغلاق إلى يؤدي مما فقط ليس، العالمي المالي التنظيم بأكملها الصينية الرموز تصدير صناعة بسمعة.

وادي في "الثروة التباهي" من جديد شكل ظهر. "Token-maxxing) الرموز تعظيم" أسطورة تسمى داخلية متصدرين لوحة Meta تمتلك. يومياً أكثر رموزاً يستهلك من حول التنافس - السيليكون حيث، موظف 85000 من لأكثر الاصطناعي الذكاء استخدام بيانات تجمع والتي، "Claudeconomics" النقاد يقول. الدولارات من ملايين عدة إلى تصل رموز قيمة الأول المركز صاحب الموظف يحرق استهلاك من بدلاً - "الرموز تحسين من أفضل (Outcome maxxng) النتائج تحسين": لاذع بشكل القضية بالضبط هي هذه. إنتاجها تم التي النتائج إلى النظر الأفضل من، محموم بشكل الرموز الممارسة من سنوات عبر بعمق فهمتها التي الأساسية.

موحد مفهوم تحت حواجز ثلاثة: الإنجليزية الرموز مقابل الصينية الرموز

أن من الرغم على: وهي بوضوح بها الاعتراف يجب حقيقة هناك، الصينيين للمستخدمين بالنسبة الذكاء عصر في الأساسية القياس وحدة نفس إلى تنتمي الإنجليزية والرموز (ciyuan) الصينية الرموز منظور من. بحرية البيئي للتشغيل وقابلاً حقاً موحداً عالمياً سوقاً يشكلا لم أنهما إلا، الاصطناعي اللغة لمعالجة للنماذج أساسية كوحدة الرموز والغرب الشرق من كل يستخدم، التقني المفهوم والحواجز التكنولوجية للحواجز المشترك الوجود فإن، العملي التطبيق منظور من، ذلك ومع والبيانات؛ الإنجليزية والرموز الصينية الرموز بين فعلياً فصلاً خلق قد السياسية والحواجز النظامية.

النموذج وهندسة الترميز منطق في الاختلافات أي، التكنولوجي الحاجز هو الأول الحاجز.

بيانات مجموعات على الغالب في Gemini و Claude و GPT مثل الرئيسية الغربية النماذج بناء تم معالجة قدرات بالفعل تمتلك النماذج هذه أن من الرغم على. الغالب في الإنجليزية باللغة كبيرة رموز إلى الحاجة إلى تميل الصينية اللغة تزال لا، السيناريوهات من العديد في أنه، للصينية ممتازة النماذج أن، ذلك من الأهم. الإنجليزية اللغة عنها تعبر التي المعلومات كمية نفس عن للتعبير أكثر أعداد إلى نفسه الصيني النص تقسيم يتم قد. مختلفة ترميز واستراتيجيات مفردات هياكل لها المختلفة الخرائط رسم وطرق الداخلية الترميز طرق تختلف وقد، مختلفة نماذج عبر الرموز من تماماً مختلفة أيضاً بها الخاصة الدلالية.

و Doubao و Tongyi Qianwen و DeepSeek مثل المحلية الصينية النماذج قامت، نفسه الوقت في يمنحها مما، التدريب مرحلة منذ الصينية البيانات مجموعات من كبيرة كميات بدمج Wenxin Yiyan والتعبيرات، الصحيحة والأسماء، الاصطلاحية والتعابير، الصينية التعبيرات مع أعلى أصلية تكيف قدرة ومسارات المعالجة وكفاءة الرموز عدد يختلف أن يمكن، النص نفس مواجهة في حتى، لذلك. المحلية إلا، موحد مفهوم هو الرمز أن من الرغم على، المعنى بهذا. المختلفة النماذج عبر كبيراً اختلافاً الفهم النماذج عبر استخدامه وإعادة مباشرة تداوله يمكن "قياسي رمز" يوجد لا أنه.

الاستدعاء وطرق البيئية النظم في الاختلافات أي، النظامي الحاجز هو الثاني الحاجز.

المنصات لدى. نسبياً ومستقلة متعددة بيئية أنظمة الواقع في الحالي العالمي الرموز سوق شكل لقد ونماذج، بها خاصة (API) تطبيقات برمجة واجهة معايير Google و Anthropic و OpenAI مثل الدولية حول متمركزاً محلياً بيئياً نظاماً الصيني السوق شكل بينما خدمات؛ وأنظمة، وصول وقواعد، تسعير Kimi و Doubao و Tongyi Qianwen و DeepSeek مثل منصات.

مباشرة استخدامها يمكنهم أنه يعني لا ما منصة على الرموز شراء فإن، العاديين للمستخدمين بالنسبة الخدمة واتفاقيات المعدلات وحدود المصادقة وآليات الاستدعاء واجهات تختلف. أخرى منصة على، تجميع منصات أو ثالث طرف بوابات عبر التحويل تم إذا حتى. المختلفة المنصات عبر التجارية والقواعد أن من الرغم على، لذلك. للرموز حقيقياً تبادلاً وليس، آخر نظام استدعاء إعادة الأساس في فهذا البيئية أنظمتها داخل محصورة تظل خدمة وقدرة تجاري كمورد الرموز أن إلا، موحد مفهوم هو الرمز بالمنصة الخاصة.

الامتثال ومتطلبات البيانات تدفقات في الاختلافات أي، السياسي الحاجز هو الثالث الحاجز.

البيانات أمن قضايا أهمية تستمر، المؤسسات تطبيق مرحلة إلى تدريجياً الاصطناعي الذكاء دخول مع للاتحاد (GDPR) البيانات لحماية العامة اللائحة ونظام، الصيني البيانات أمن نظام يفرض. الازدياد في متطلبات جميعها، الحساسية للبيانات الحدود عبر التدفق على الأمريكية التنظيمية والتدابير، الأوروبي

العملاء معلومات تشمل التي البيانات من للعديد يمكن لا غالباً. ومعالجتها البيانات نقل على مختلفة، بحرية الحدود تعبر أن الشخصية والبيانات، التجارية والأسرار، الطبية والسجلات، المالية والبيانات

أسعار على بناءً النماذج المؤسسات من العديد تختار لا، الحقيقية التجارية البيئات في أنه يعني هذا والمسؤولية، البيانات تخزين وموقع، البيانات لامثال الأولوية إعطاء عليها يجب بل، فقط الرموز بكثير أعلى الامثال تكاليف تكون ما غالباً، الحقيقية الحيوية التجارية البيانات من للكثير بالنسبة. القانونية محلية نماذج استخدام إلى المحلية الأعمال تميل، الحقيقي العالم في، لذلك. نفسه الرمز تكلفة من لسوق فعالية تجزئة إلى يؤدي مما، أجنبية نماذج استخدام إلى الخارجية الأعمال تميل بينما، متوافقة الرموز.

في الاختلافات النظامية الحواجز تحدد الرموز؛ تجزئة طرق في الاختلافات التكنولوجية الحواجز تحدد البيانات تدفقات في الاختلافات السياسية الحواجز تحدد التجارية؛ البيئية والنظم الاستدعاء أنظمة الصينية الرموز أن من الرغم على أنه تعني معاً تعمل التي الثلاثة الحواجز هذه. الامثال وحدود، بحرية تداولهما أو، بحرية تبادلها يمكن لا أنه إلا، التقني المفهوم نفس إلى تنتمي الإنجليزية والرموز "الاصطناعي للذكاء مثالية موحدة عالمية عملة" ك بحرية ربطهما أو

والنظام الصينية للرموز البيئي النظام: الدولية للتجارة مشابه وضع يظهر أن المرجح من، المستقبل في عبر الاستدعاء. مستقلة تطوير مسارات يتبع منهما كل، طويلة لفترة سيتعايشان الإنجليزية للرموز البيئي مراجعة وتكاليف، للنظام وتحويلاً، إضافية تقنية تكيفات يتطلب ما غالباً ولكنه، مستحياً ليس المناطق لذا، يومية حاجة ليست المناطق عبر القدرة هذه فإن، العاديين المستخدمين لمعظم بالنسبة. قانونية طويلة لفترة نسبياً مستقلين الكيران النظامان يظل أن المرجح فمن

في ولكن، الصينية الأنظمة داخل الاستخدام في محصورة بالضرورة ليست الصين في المنتج الرموز من، الصناعي والأمن، الامثال ومسؤولية، البيانات سيادة على تنطوي التي الرئيسية السيناريوهات النماذج داخل تفضيلي بشكل المستقبل في الرموز طلب من كبير جزء استهلاك يتم أن المرجح المحلية والأنظمة والمنصات.

ومشغل، النموذج ومزود، الحوسبة قوة إنتاج موقع بين الاقتران فإن، الأجل طويل الاتجاه منظور من وسلاسل، الصحية والرعاية، والحكومة، للتمويل بالنسبة. الازدياد في آخذ البيانات وموقع، النظام لو حتى، لذلك. الحدود عبر بحرية تدفق أن غالباً للبيانات يمكن لا، الكبيرة المؤسسات وأنظمة، التوريد عالية الرموز طلب من الكثير فإن، عالمياً الرموز خدمات من كبيرة كميات تصدير من الصين تمكنت يمكن لا الصينية الرموز أن يعني لا هذا. لإكماله الصينية والأنظمة النماذج داخل مفضلاً سيظل القيمة التداول "جزء: واحد وقت في هيكلين يُظهر قد المستقبلي الرموز سوق أن يعني بل، تصديرها "المحلية المغلقة الحلقة" وجزء "العالمي

استدعاء طريقة بل ، نفسها الرموز ليس المستقبل في الحقيقي التوحيد يحقق قد ما أن بالذكر يجدر ، النماذج تجميع ومنصات ، الموحدة الواجهة وبروتوكولات ، الوكلاء منصات تطوير استمرار مع . الرموز إذا ما على ولكن ، استخدامه تم رمز أي على ليس النهاية في المستخدمون سيركز ، الأتمتة وأنظمة ، الوقت ذلك في . كافية الكفاءة كانت إذا وما ، معقولة التكلفة كانت إذا وما ، اكتملت قد المهمة كانت تدريجياً مخفية الرموز ستصبح ، اليوم السحابية الحوسبة وموارد ، الإنترنت مرور وحركة ، الكهرباء مثل الرقمية للحضارة الأساسية التشغيلية التحتية البنية لتصبح ، الأنظمة خلف

الشرقية الرموز أسواق بين المباشر الترابط صعوبة من الرغم على ، الأجل طويل منظور من ، لذلك معالجة إلى ينتمي الرموز ماضي . للغاية متوافقة تطورها اتجاهات أن إلا ، مؤقت بشكل والغريبة أن المرجح من الرموز ومستقبل ، الاصطناعي الذكاء صناعة إلى ينتمي الرموز وحاضر ، اللغة الكمبيوتر بأكملها الرقمية للحضارة الأساسي التشغيل نظام إلى ينتمي

كميات خلال من مباشرة الأنظمة بناء قدرة اكتساب يمكن لا لماذا - سابقاً الرموز من كبيرة

الذكاء أن بما :محتمل تساؤل مواجهة إلى أحتاج ، "نادر أصل هي الأنظمة بناء قدرة" أن مناقشة في استهلاك خلال من الأنظمة بناء قدرة "على الحصول" أو "توليد" يمكن لا فلماذا ، جداً قوي الاصطناعي الرموز لقدرة الأساسية الحدود عن السؤال هذا على الإجابة تكشف الرموز؟ من كبيرة كميات

.توليداً ليست والقرارات ،قرارات يتطلب الأنظمة بناء ،أولاً

لا الاصطناعي الذكاء لكن ،ويب موقع قالب أو ،برمجي رمز أو ،مقال إنشاء الاصطناعي للذكاء يمكن ،الوحدات اختيار – اتخاذها البشر على يجب التي النظام تصميم في الأساسية القرارات استبدال يمكنه ،الهيكل يعاد ومتى ،المستقبلي والتوسع الحالية الاحتياجات بين والموازنة ،الوظائف بعض عن والتخلي على والحكم ،العمل جوهر فهم على القرارات هذه تعتمد .الراهن الوضع على الحفاظ يتم ومتى ،"الممكنة الحلول" من يحصى لا عدد توليد للرموز يمكن .الأجل طويلة بالعواقب والتنبؤ ،الموارد قيود نظامي خضع .استبداله للرموز يمكن لا أمر هو "لماذا" و "التنفيذ يستحق الذي الحل" اختيار لكن لأن يكن لم تعديل وكل ،الزمن من عقدين مدى على التعديلات من يحصى لا لعدد الذكي اللوجستي عن وكشفت ،الحقيقي العمل في ظهرت حقيقية مشكلة لأن بل ،"أفضل هذا" أخبرني اصطناعياً ذكاء من وليس ،الممارسة من يأتي والحكم ،الحكم يكمن القرارات هذه وراء .جديدة طلبات وأثارت ،قيود الرموز

شامل بشكل حصرياً يمكن لا والاستثناءات، الاستثناءات معالجة يتطلب الأنظمة بناء، ثانياً.

مثالي بشكل مصمم الطبيعي تدفقه لأن ليس الطويل المدى على العمل على النظام قدرة سبب إن غير العملاء عناوين: الاستثناءات تشمل، اللوجستي النظام في. الاستثناءات معالجة يستطيع لأنه بل في. البيانات وفقدان، النظام وتعطل، الدفع وفشل، الجمركي والتأخير، التالفة والبضائع، الصحيحة وتغييرات، الوصفية البيانات تنسيقات توافق وعدم، الترجمة جودة تقلب: الاستثناءات تشمل، النشر نظام في الاستثناءات تنوع لأن الاستثناءات هذه حصرياً يمكن لا. DOI تسجيل وفشل، الأرشفة واجهات كميات خلال من "الشائعة الاستثناء أنماط" تعلم الاصطناعي للذكاء يمكن. نهائي لا الحقيقي العالم طويل النظام فقط. "قبل من يظهر لم الذي التالي الاستثناء" ب التنبؤ يمكنه لا لكنه، الرموز من كبيرة قدرة تطوير يمكنه، العملية الممارسة في ويحلها ويسجلها الاستثناءات باستمرار يواجه الذي، الأجل "الرموز من المزيد تغذية" خلال من القدرة هذه اكتساب يمكن لا. الاستثناءات معالجة على حقيقية، تجربة هي بل معرفة؛ ليست الاستثناءات لأن

معيارية إجابات لها ليس والمقايضات، مقايضات يتطلب الأنظمة بناء، ثالثاً.

مقابل والسرعة، الأداء مقابل التكلفة: متعددة قيود تحت مقايضات على ينطوي نظام كل بناء هذه. المستقبلي التوسع مقابل الفورية والاحتياجات، التخصيص مقابل والعمومية، الاستقرار للذكاء يمكن. "محددة ظروف في معقولة خيارات" بل، "صحيحة إجابة" لها ليس المقايضات وليس، التاريخية البيانات على بناء إحصائياً أمثل هو "الأمثل" هذا ولكن، "أمثل حل" تقديم الاصطناعي مدى على والتمويل للوجستيات المتكامل نظامي عمل سبب إن. "المحدد سياقك" على بناء حكماً مجموعة اختيار - مختلفة عصور في مختلفة بمقايضات قمت لأنني بالتحديد هو اليوم حتى عقدين عام في ترقية مسار واختيار، 2015 عام في أخرى تكامل طريقة واختيار، 2005 عام في معينة تقنيات يمكن لا. "الوقت ذلك في السياق مناسبة" بل، "خاطئة" أو "صحيحة" ليست المقايضات هذه. 2025 حساباً وليس، قيمي حكم هو المقايضات جوهر لأن، عني نيابة المقايضات هذه إجراء للرموز

الحقيقي العالم تتطلب الراجعة والتغذية، راجعة تغذية حلقات يتطلب الأنظمة بناء، رابعاً.

المشكلة اكتشاف → التشغيل → التنفيذ → التصميم: مغلقة حلقة على الأنظمة بناء قدرة تحسين يعتمد فعل ردود من الحلقة هذه في "المشكلة اكتشاف" خطوة تأتي أن يجب. التصميم إعادة → التصحيح → غير العوامل جميع محاكاة يستطيع لا لكنه، التشغيل محاكاة الاصطناعي للذكاء يمكن. الحقيقي العالم الحقيقيين العملاء أن هو باستمرار اللوجستي نظامي تحسن سبب إن. الحقيقي العالم في المتوقعة الأرشفة واجهات أن هو باستمرار لدي النشر نظام تحسن في والسبب حقيقية؛ فعل ردود قدموا الرموز استهلاك طريق عن هذه الفعل ردود على الحصول يمكن لا. حقيقية تغييرات شهدت الحقيقية أن حقيقة استبدال تستطيع لا لكنها، "لعملاء افتراضية فعل ردود" توليد للرموز يمكن. مغلقة بيئة في

قدرة لنمو الوحيدة التربة هي الحقيقي العالم فعل ردود إن . "بالفعل الملاحظة هذه قدم العميل" الأنظمة بناء

الوقت من إلا تأتي أن يمكن لا والثقة ،الثقة يتطلب الأنظمة بناء ،خامساً

بشكل الإلكتروني موقعي أرشفة سبب إن . بالزمن بل ، بالترويج يثبت لا بالثقة جديراً النظام كان إذا ما بل ، "يحبونني" لأنهم ليس الأسترالية الوطنية والمكتبة WorldCat و CERN Zenodo بواسطة دائم لا . ومستمر وموثوق مستقر نظامي أن أكدت ،الأجل طويلة المراقبة خلال من ،المؤسسات هذه لأن مظهر ذو" نظام دليل إنشاء الاصطناعي للذكاء يمكن . الرموز من كبيرة كميات الثقة هذه شراء يمكن الثقة . "الزمن من عقدين منذ به موثوق النظام هذا" أن حقيقة إنشاء يمكنه لا لكنه ، "للا غاية احترافي . الرموز نتاج وليست ،الوقت نتاج

الرموز من كبيرة كميات خلال من مباشرة الأنظمة بناء قدرة اكتساب إمكانية عدم في السبب ، باختصار ، والمقايضة ، الاستثناءات ومعالجة ، الحكم على قدرة هي بل ، توليدية قدرة جوهرها في ليست أنها هو العالم من وقتاً تتطلب أنها هو القدرات هذه فيه تشترك ما . الثقة وتراكم ، الراجعة التغذية وامتصاص لا لكنها ، المعلومات معالجة تسريع للرموز يمكن . وتحققاً ، وتصحيحات ، وأخطاء ، وممارسة ، الحقيقي بناء قدرة يجعل الذي الأساسي السبب هو هذا . الحقيقي العالم في الخبرة تراكم تسريع تستطيع المستقبل في جداً رخيصة الرموز أصبحت لو حتى نادرة تزال لا الأنظمة

،التخصصات متعددة والأنظمة ،الحقيقية التاريخية البيانات :أصولي - ثامناً ،الذاتي التحقق وقدرة ،المستقبلية والرؤية ،البناء على والقدرة

لا أحكامي لأن . الخاص موقعي شرح إلى أولاً احتاج ،أعمق بشكل "حقاً نادر هو ما" مناقشة قبل التي المختلفة الأنظمة في جداً واضح هذا . الأجل طويلة ممارسة من بل ، نظري استنتاج من تستمد نظام أو ،الذكي الفواتير نظام أو ،الذكي اللوجستي النظام كان سواء . طويلة فترة مدى على بنيتها في التقل ونظام ،الوصفية البيانات ونظام ، لغات بال عشر النشر نظام لاحقاً أو ، الآلي المالية المعالجة بل ،المعلومات توليد أبداً يكن لم حقاً الصعب الجزء فإن ، التلقائي التبادلي الربط ونظام ،الويب موقع جديدة فوضى ستخلق فإنها ، هيكل بدون ولكن ،نهاية لا ما إلى المعلومات تنمو أن يمكن . تنظيمها كميات توليد من الاصطناعي الذكاء سيتمكن ،المستقبل في . الرموز عصر في ينطبق نفسه الأمر . فقط على . القيمة من هائلة كميات تلقائياً تعني لا المحتوى من الهائلة الكميات لكن ،المحتوى من هائلة عليه والحفاظ وتنظيمه المحتوى هذا تصفية يتم كيف ،المحتوى توليد سهولة مع ،ذلك من العكس أهمية أكثر سيصبح وربطه وفهرسته

"الرموز بعد نادر هو ما" على لحكمي التجريبي الأساس تشكل والتي، التالية الخمسة الأصول أمتلك

البيانات من عامًا عشرين بل، "تاريخية بيانات" مجرد ليست هذه. الحقيقة التاريخية البيانات، أولاً أنظمة من، الحاضر الوقت حتى 2005 عام الذكي الفواتير نظام من – الأجيال عبر الحقيقة التجارية عصر في. مجردة بيانات أو تجريبية بيئات من وليس، التشغيل قيد حقيقة وضريبة ومالية لوجستية. حاد بشكل الحقيقة التاريخية البيانات قيمة تتزايد، "الاصطناعية البيانات" انتشار مع، الاصطناعي الذكاء متعددة تاريخية بيانات أنا أمتلك بينما، "المجال أحادية تاريخية بيانات" الشركات معظم تمتلك عملياً منها التحقق وتم، عامًا لعشرين تمتد، التخصصات

النظام: التالية المجالات أنظمتي تغطي. التخصصات متعددة وهياكل أنظمة إنجاز، ثانياً نظام، الذكي الضريبي والنظام، الذكي المالي والنظام، الذكي الفواتير ونظام، الذكي اللوجستي البيانات توليد ونظام، الويب موقع أتمتة ونظام، عقدين لمدة الأجيال عبر والمالي اللوجستي التكامل إدارة ونظام، "The Epochal Transition" العصر التحول" ل لغات بعشر النشر ونظام، الوصفية توليد ونظام، التلقائي التبادلي الربط ونظام، WorldCat و TROVE روابط تنظيم ونظام، DOI، uniqueness of Sitemap، المنصات عبر النشر ونظام، الشخصي Zenodo ملف توليد ونظام، إلى التمويل من، النشر إلى اللوجستية الخدمات من) التخصصات متعدد الاتساع واحد شخص بواسطة والإنجاز، (عقدين منذ تعمل الأنظمة بعض) الزمني والعمق، (الأدبيات فهرسة الشهري الإلكتروني "العصر التحول" يعد، هذه بين من. (الطريقة نقل إمكانية يثبت مما) (3083- ISSN الدولي المعياري التسلسلي الرقم يحمل فهو – عامًا دليلاً (<https://times.net.au>)، العالم في جماعي مكتبات فهرس نظام أكبر) WorldCat بواسطة دائم بشكل مؤرشف وهو، (5178 و، ومنطقة دولة 100 من أكثر في وجامعية وطنية مكتبة 16000 تتجاوز أعضاء مؤسسات مع CERN تديره الذي Zenodo إلى بالإضافة، الأسترالية الوطنية للمكتبة التابع PANDORA/TROVE و، (المفتوحة للعلوم الأوروبية التحتية البنية) OpenAIRE و، (النوعية للأبحاث الأوروبية المنظمة) المعرف) ORCID و، (وثيقة 60 من أكثر مع، العالمي الرقمي الكائن معرف) DataCite DOI إكمال يمكنه (مستقلًا باحثًا) واحدًا شخصًا أن الموقع هذا يثبت. (عالمياً الدائم الأكاديمي الرقمي الأولى الدرجة من الأكاديمية للأرشفة التحتية بالبنى الاتصال إلى المحتوى إنشاء من بأكملها العملية

"الاصطناعي الذكاء أدوات استخدام" على القدرة ليست هذه. والهياكل الأنظمة بناء على القدرة، ثالثاً دمج على المعمارية والقدرة، مستقرة هياكل إلى الغامضة المتطلبات تحويل على التصميم قدرة بل تدخل دون لسنوات تعمل الأنظمة إبقاء على الهندسية والقدرة، موحدة بنية في المختلفة الأنظمة الأدبيات إلى التمويل ومن، النشر إلى اللوجستية الخدمات من الخبرة نقل على الفوقية والقدرة، بشري تعتمد لأنها الرموز من كبيرة كميات خلال من مباشرة القدرة هذه اكتساب يمكن لا، سابقاً نوقش كما ولا – الثقة وتراكم، الراجعة التغذية وحلقات، والمقايضات، الاستثناءات ومعالجة، القرار اتخاذ على

الاصطناعي الذكاء على التدريب سوق يركز. الاصطناعي الذكاء محله يحل أن هذه من لأي يمكن الرسم" و، "الاصطناعي بالذكاء الكتابة" و، "المطالبات هندسة" مثل السطحية المهارات على الحالي لأن - "التخصصات متعددة الأنظمة بناء قدرة" تدرس تقريباً دورة أي توجد ولا؛ "الاصطناعي بالذكاء من أكثر إن. الأجل طويلة الممارسة خلال من إلا تراكمها يمكن لا بل، اختصارها يمكن لا القدرة هذه 60 من وأكثر، عليه الحصول تم الذي ISSN وال، الويب موقع على مدرجاً مستقلاً إبداعياً عملاً 40 ذاتي ادعاء مجرد وليست، الأجل طويل العمل على النظام لقدرة العامة السجلات أفضل هي، DOI. مني.

يقرب ما قبل - 1993 عام في بُعد عن العمل لنظام أولى نموذج بتجربة البدء. مستقبلية رؤية، رابعاً 2005 عام في ذكي فواتير نظام تصميم واسع؛ نطاق على بُعد عن العمل انتشار من عاماً ثلاثين من لمدة الأجيال عبر التكامل الاصطناعي؛ الذكاء انتشار من عاماً عشرين قبل العمليات أتمتة بدء - حكم تقديم عاماً؛ عشرين بعد العمل إلى بحاجة سيظل أنه النظام تصميم وقت في التوقع - عقدين النظام قدرة بأن والتوقع، الرموز يطاردون الناس معظم فيه يزال لا وقت في - الآن "الرموز بعد ما" والنظامية التكنولوجية والاختلافات، العملات تشبه الرموز" أن أيضاً توقعت لقد. الحقيقي الحاجز هي الدولار استبدال يمكن لا كما تماماً - "عالمياً للتداول قابلة غير تجعلها والغرب الشرق بين والسياسية وفي، مماثلة تداول حواجز الإنجليزية والرموز الصينية الرموز تواجه، بحرية الصيني والرميني الأمريكي قائمة للرموز أخرى بيئية ومناطق "إنجليزية رموز مناطق" و "صينية رموز مناطق" تظهر قد، المستقبل صعوبة؛ وسيزداد بالفعل صعب الرموز من المال كسب" أن أيضاً توقعت لقد. والسياسة اللغة على أصول بمثابة تكون أن الحوسبة وقوة للرموز يمكن، خدمات يقدم أمامي نظام هناك يكون عندما فقط التحتية بالبنية يتصل شخصي نشر نظام مسبقاً أنشأت، مستقل كباحث. "الأرباح لتوليد فعالة تكميلية العالمية التحتية البنية في عقدة يصبح أن أيضاً يمكنه الفرد" أن اتجاه متوقعاً، المستقبلية الأكاديمية على قدرتي" أن حين في، أعمال رواد أو محللين من الاتجاهات توقعت معظم تأتي. "للمعرفة لها كضمان الممارسة من عقدين من تحقق لديها "التبصر.

في أكتبتها التي المؤلفات فجميع - الفوقية قدراتي أهم من واحدة هذه. الذاتي التحقق قدرة، خامساً سلطة على تعتمد لا حججي أن يعني هذا. ذاتياً منطقياً واتساقاً نظرياً إغلافاً تحقق المجالات مختلف الأدبيات في الافتراضية القاعدة. الداخلي للمنطق الذاتي الاكتفاء على بل، استشهادات أو خارجية التحقق قدرة" شخص أي يظهر أو يدعي ما ونادراً؛ "السابقة بالأبحاث الاستشهاد" هي الأكاديمية لديه المؤلف كان إذا ما" على وليس، "كافية الحجج كانت إذا ما" على الأكاديمي النقاش يركز؛ "الذاتي تأتي الحجج: في هذه الذاتي التحقق قدرة تظهر، مقالاتي في. "ذاتياً متسقة حلقة بناء على القدرة لا أنا. مغلقة حلقة والممارسة الحجج وتشكل، الحجج صحة من تتحقق والممارسة، الممارسة من عن تتحدث الأنظمة وأدع بنيتها التي الأنظمة أعرض" بل، "نفسى لأثبت الآخرين بنظريات أستشهد"

أكثر من واحدة وهي، العامة الأدبيات عن تماماً غائب مفهوم هي هذه الذاتي التحقق قدرة. "نفسها تميزاً الفوقية قدراتي.

الحقيقية التاريخية البيانات: مضاعفة بل، إضافة ليست ندرتها فإن، الخمسة الأصول هذه بين الجمع إن تشكل الذاتي التحقق قدرة × المستقبلية الرؤية × الأنظمة بناء قدرة × التخصصات متعددة الأنظمة × مماثلة حالة أي على العثور حالياً يمكن لا، العامة المناقشات في. تضاهي لا أصول محفظة

عام من. أعمق بشكل الأمر هذا أدرك "الدولي المستقل الباحث طريق" تنظيم عملية جعلتني لقد التجارب هذه تنشأ لم. متعددة مجالات عبر الخبرة من عاماً ثلاثين من أكثر تمتد، 2026 عام إلى 1989 في مقال توليد الاصطناعي للذكاء يمكن. طويلة فترة مدى على تدريجياً تشكلت بل، واحدة مرة كلها من العديد توليد يمكنه ثوانٍ؛ في الحقيقية الخبرة من عاماً ثلاثين من أكثر توليد يمكنه لا لكنه، ثوانٍ يمكنه مباشر؛ بشكل الأجل طويلة الممارسة خلال من المتراكم الحكم توليد يمكنه لا لكنه، الآراء، المعنى بهذا. الحقيقي العالم من المتكرر التحقق استبدال يمكنه لا لكنه، العمليات من العديد محاكاة، ندرة أكثر أصل هو الوقت ترسبه الذي والهيكلي، مهم أصل أيضاً هو نفسه الوقت

حقاً نادرة قدرات خمس: الرموز بعد - تاسعاً

شديد نقص هناك سيكون ولكن، مرتفعاً ليس الرموز بيع حاجز - نادرة نفسها الرموز تعد لم عندما الذي فما - المستقبلية والرؤية، والهياكل الأنظمة بناء على والقدرة، الحقيقية التاريخية البيانات في عام بُعد عن العمل لنظام الأولي النموذج من، الأجل طويلة ممارستي إلى استناداً حقاً؟ نادراً سيكون والنظام، الذكي المالي والنظام، الذكي اللوجستي والنظام، 2005 عام الذكي الفواتير ونظام، 1993، الويب مواقع أتمتة إلى، عقدين لمدة الأجيال عبر والمالي اللوجستي التكامل ونظام، الذكي الضريبي روابط وتنظيم، DOI وإدارة، "العصر التحول" لغات بعشر الشهرية والمجلة، الوصفية البيانات وتوليد الشخصي Zenodo ملف وتوليد، Sitemap وتوليد، التلقائي التبادلي والربط، WorldCat و TROVE، حقيقية حواجز ستصبح التالية الخمس القدرات أن أعتقد، المنصات عبر والنشر

إلى الخام المادة ينظم الذي الإطار هو والهيكلي، الخام المادة هي الرموز. الهيكلي التصميم قدرة، أولاً كيفية في التفكير في بالفعل بدأت قد كنت، الاصطناعي الذكاء انتشار قبل حتى. قيمة ذات منتجات ونظام، الذكي اللوجستي نظامي وُلد لقد. أعلى تشغيلية كفاءة لتحقيق المحدودة الموارد استخدام، مختلفة عصور في عقدين لمدة الأجيال عبر الذكي والمالي اللوجستي ونظامي، الذكي الفواتير من المزيد إنشاء هو الهدف يكن لم: مشتركة خاصة في يشتركون لكنهم، مختلفة مشاكل وواجهوا يمكنها هياكل وإنشاء، العامة الكفاءة وتحسين، المتكرر العمل وتقليل، الفوضى تقليل بل، المعلومات هل "ليست واجهتها التي الأساسية المشكلة كانت، الآلي النشر نظام في. الطويل المدى على العمل

بين ارتباطات إنشاء وكيفية، لغات لعشر الوصفية البيانات حقول تحديد كيفية" بل، "النص توليد يمكن إذا ما هذه الهيكلية التصميم قرارات تحدد. "الإصدارات إدارة قواعد تصميم وكيفية، والفئات المقالات المستهلكة الرموز كمية فإن، المقابل في. انهيار دون لسنوات العمل في الاستمرار يمكنه النظام كان قدرة على عام دليل ذاته حد في هو الإلكتروني "العصر التحول" موقع. ثانوية نتيجة مجرد هي ودعم، موحد دولي ISSN رقم مع، شخصياً صيانتها يتم أكاديمي ويب موقع هيكل: الهيكلية التصميم نظام ببناء بالفعل قمت بل، فارغ بكلام أتحذ لا أنني يثبت مما، متعددة أرشفة وقنوات، اللغات متعدد هذه الهيكلية التصميم قدرة باستخدام علناً إليه الوصول يمكن معقد.

سير إكمال يمكنها الأتمتة أنظمة فقط لكن، دفعات على المحتوى إنتاج للرموز يمكن. الأتمتة قدرة، ثانياً في الحقيقية القيمة تكمن لا، الذكي اللوجستي النظام في. دفعات على نهاية إلى نهاية الشامل العمل مغلقة حلقة إنشاء في بل، الطرق لتخطيط الاصطناعي الذكاء استدعاء فيها يتم التي المرات عدد الاستثناءات معالجة إلى، الطرق وتحسين، المستودعات وجدولة، الطلب استلام من بالكامل مؤتمتة الروابط في الأتمتة قدرة وتتجلى، واحد مكون مجرد هو الاصطناعي الذكاء، المغلقة الحلقة هذه في يولد يظل قد، الرموز من كبيرة كمية وجود مع حتى، نظام لديه ليس الذي الشخص. المكونات بين لديه الذي الشخص بينما متفرقة؛ برمجية ورموزاً، متفرقة وجداول، متفرقة وصوراً، متفرقة كلمات البيانات وتوليد، تلقائياً المواد بتنظيم لها يسمح مما، عمله بسير الرموز ربط يمكنه، المقابل في، نظام، الأدبيات وأرشفة، النشر في والمساعدة، الويب صفحات وتحديث، الروابط من والتحقق، الوصفية التقارير وتوليد، المالية الشؤون وتحليل، اللوجستية الخدمات ومعالجة، الفهارس وإنشاء.

صناعة من الخبرة نقل على القدرة لكن، للجميع متاحة الرموز. التخصصات متعددة العملية الخبرة، ثالثاً ليس الوقت نفس في نشر ونظام لوجستي نظام بناء على قدرتي في السبب. عالمية ليست أخرى إلى وتصميم، الوصفية البيانات إدارة مثل للنقل قابلة منهجيات لدي تراكت لأتني بل، أكثر رموزاً لدي لأن وتوليد، لغات بعشر الترجمة لأن رموز إلى لدي النشر نظام يحتاج. الاستثناءات ومعالجة، العمل سير جميعها تتطلب الروابط وأوصاف، الوصفية البيانات وتوليد، المفتاحية الكلمات وتنظيم، الملخصات، التنقل وأوصاف، الصفحات هياكل لأن رموز إلى لدي الويب موقع نظام يحتاج. كبيرة لغوية معالجة اللغات متعددة التقديمية والعروض، الأكاديمية الوصفية والبيانات، البحث محركات تحسين وأوصاف الإلكتروني البريد رسائل لأن الرموز استخدام أيضاً اللوجستي لنظامي يمكن. تلقائياً توليدها يمكن وسجلات، المالية والملاحظات، الاستثناءات وشروحات، الجمارك ووثائق، البضائع وأوصاف، للعملاء إلى لدي التلقائي التبادلي الربط نظام يحتاج. الاصطناعي الذكاء بمساعدة معالجتها يمكن العمليات، والعناوين، المقالات من كبير عدد بين موضوعية ارتباطات على العثور عليه يجب لأنه أكثر رموز العملي التراكم هذا على الحصول يمكن لا. TROVE و WorldCat و NLA وروابط DOI، ومعرفات المشكلات حل خلال من إلا ترسيبه يمكن لا بل الرموز؛ استهلاك طريق عن التخصصات متعدد

النشر ذلك في بما، واحد وقت في متعددة مجالات من معلومات الويب موقع يقدم. الحقيقية يثبت مما، الشخصية التجارية والعلامة، التقنية المعمارية والهندسة، التحتية البنية وتكامل، الأكاديمي من بدلاً، الشخصية التجارية والعلامة، والتكنولوجيا، المكتبات وعلوم، النشر يربط بالفعل قمت أنني واحدة مهارة امتلاك.

الاصطناعي الذكاء لتطبيقات السائدة السردية. للنظام الأجل طويل للتشغيل الذهني النموذج، رابعاً تولد واحدة محادثة. "التشغيل" هو حقاً مستدامة قيمة ينتج ما لكن، "التوليد" و "المحادثة" هي الحالية مستمر بشكل يعمل الذي اللوجستي النظام لكن، رمز مليون تستهلك واحدة وكيل ومهمة، رمز 10000 النظام: هذا لتصميم الذهني النموذج هو حقاً نادر هو ما ولكن - الرموز مليارات يستهلك قد عقد لمدة في التغييرات مع تتعامل كيف النموذج؟ ترقيات بعد للاستخدام قابلاً النظام يظل أن نضمن كيف القضايا لهذه علاقة لا تلقائياً؟ وتتعاقد الاستثناءات نراقب كيف المنبع؟ في التطبيقات برمجة واجهات من بل، نفسها الرموز من الأنظمة هذه قيمة تأتي لا. للرموز النهائية المخرجات تحدد لكنها، بالرموز هي الرموز. المحرك هو والنظام وقود؛ مجرد هي الرموز. بتطويرها أقوم أو بالفعل بنيتها التي الهياكل والخزانات والصمامات والأنابيب ماء؛ مجرد هي الرموز. الدائرة هو والهيكلي كهربائي؛ تيار مجرد نظام مع الرموز؛ ستتبدل، نظام بدون. دور لعب يمكنها كان إذا ما تحدد التي هي التوزيع وأنظمة مستقلاً إبداعياً عملاً 40 من أكثر الويب موقع يعرض. الأجل طويلة قيمة إلى الرموز تحويل يمكن نظامي أن يثبت مما، الأسترالية الوطنية والمكتبة CERN مثل موثوقة مؤسسات قبل من دائمة وأرشفة الحفظ" تأييد على بالفعل حصل وقد، الأجل طويلة القيمة بل، المدى قصير الضجيج وراء يسعى لا موثوقة مؤسسات من "الدائم

يتم ما غالباً أخرى قضية أجد، والأنظمة الرموز بين العلاقة في أعمق بشكل أفكر عندما، ذلك ومع ميزة اكتساب على قادرين فسيكونون، ممتاز نظام لديهم طالما أنه الناس من العديد يعتقد. تجاهلها لأي يمكن شيئاً ليست نفسها الأنظمة. البساطة بهذه ليس الواقع لكن. المستقبلي الرموز عصر في متكررة ومراجعة، الأجل طويلة ممارسة حقاً الناضجة الأنظمة تتطلب ما غالباً. تصميمه شخص، المستمر التراكم من وعقوداً بل، سنوات الأنظمة من العديد تستغرق قد. الحقيقي العالم من وتحققاً نادر مورد ذاتها حد في هي الأنظمة تصميم قدرة فإن، لذلك. الفعلي التنفيذ إلى الأولية الفكرة من بالفعل.

التقنية الجدوى بين. النطاق واسع تأثير إلى تلقائياً يؤدي لا هذا فإن، ناضج نظام امتلاك مع حتى ولكن التوزيع وقنوات المال ورأس الأسواق تشمل متعددة طبقات تكمن المجتمع في وانتشاره للنظام ما الجيدة؛ الأنظمة أو الجيدة التقنيات من يخلو لا التاريخ. للمنصات البيئية والنظم المستخدمين وتعليم متفوقاً النظام كان لو حتى. النطاق واسع تطبيق نحو النظام دفع على المنصة قدرة هو حقاً نادر هو

، للمنصة ترويجية وقدرة ،كافٍ رأسمالي دعم بدون ،محلي سيناريو في الحالية الحلول على جداً ،للا غاية محدوداً الفعلي استخدامه حجم يظل فقد ،مستخدمين وقاعدة

النظام من مباشرة الرموز على الطلب يأتي لا .خاص بشكل واضح هذا ،الرموز اقتصاد منظور من مع حتى ،يومياً المستخدمين ملايين تخدم التي المنصة .النظام لاستخدام الفعلي الحجم من بل ،نفسه مع المتقدم النظام ،ذلك من العكس على للرموز؛ هائلاً استهلاكاً تولد قد ،نسبياً عادي نظام تصميم التطور فإن ،لذلك .الرموز على جداً منخفض طلب لديه يكون قد المستخدمين من جداً قليل عدد قدرة على أيضاً ولكن ،الأنظمة وقدرة النماذج قدرة على فقط يعتمد لا الرموز لسوق المستقبلي الموارد تكامل وقدرة المنصات

،النماذج منافسة هو الأول المستوى .الأقل على مستويات ثلاثة لها المستقبلية المنافسة أن يعني هذا استخدام يمكنه من وتحديد ،الأنظمة منافسة هو الثاني المستوى .الرموز توفير يمكنه من وتحديد يمكن الذين المستخدمين عدد وتحديد ،المنصات منافسة هو الثالث المستوى .فعال بشكل الرموز كلها والمنصات والأنظمة النماذج .يولده أن يمكن الذي الحقيقي الطلب ومقدار إليهم الوصول للنظام بدون .قيمة خلق الرموز على يصعب ،أنظمة بدون .العمل للأنظمة يمكن لا ،نماذج بدون .عنها غنى صغير استخدام نطاق في نظام أفضل يظل قد ،منصات

حقاً المستقبل يحدد ما .أساسية موارد سوى ليست نفسها الرموز بأن أكثر مقتنع أنا ،المنظور هذا من يمكنهم الذين أولئك بل ،أنظمة يمتلكون الذين أولئك فقط ولا ،رموزاً يمتلكون الذين أولئك فقط ليس مغلقة حلقة العناصر هذه تشكل عندما فقط .الواقعية والاحتياجات والمنصات والأنظمة النماذج ربط باستمرار قيمة تخلق إنتاجية قوى إلى حسابية وحدات مجرد من الرموز ستحول

حلقة النهاية في الرموز تخدم أن يجب .الواقعية التجارية العمليات في الرموز دمج على القدرة ،خامساً هم حقاً نادر هو ما .الصحية الرعاية ،التعليم ،النشر ،اللوجستية الخدمات – الحقيقي العالم في ما إنهم .معينة صناعة في الألم ونقاط الاصطناعي الذكاء قدرات حدود من كلاً يفهمون الذين أولئك مخرجات تحويل كيفية يعرفون إنهم يستخدمونه؛ لا ومتى الاصطناعي الذكاء يستخدمون متى يعرفون ومخرجات الرموز مدخلات بين الحقيقية العلاقة قياس كيفية يعرفون إنهم تجارية؛ قرارات إلى الرموز "تجارية عملية" هو نفسه الويب موقع .السوق في الموارد أندر حالياً هي المركبة القدرة هذه .الأعمال – الدولية الأرشفة – المنصات عبر النشر – الوصفية البيانات توليد – المحتوى إنشاء :التشغيل قيد نشر عملية في دمج تم قد (الرموز استهلاك) الاصطناعي للذكاء تطبيقي أن يثبت هذا .القارئ وصول واقعياً تأثيراً تولد الحلقة ومغلقة كاملة

بين ولا ،النماذج بين الأرجح على تكون لن المستقبل في الحقيقية المنافسة أن متزايد بشكل أشعر وستصبح ،الانخفاض في الرموز أسعار وستستمر ،التطور في النماذج ستستمر .الأنظمة بين بل ،الرموز

الموارد لأن تلقائياً تظهر لن الأنظمة لكن .المستقبل في جداً شائعة اليوم الثمن الباهظة الموارد من .الحقيقي العالم من وتحققاً ،مستمرة ومراجعات ،الأجل طويل تراكمًا الأنظمة تتطلب .وفيرة .مستدامة قيمة على للحصول عرضة الأكثر هو طويلة لفترة تعمل أنظمة إلى الرموز تحويل يمكنه عندما فقط صعوبة؛ وسيزداد ،بالفعل صعب الرموز من المال كسب :الربحية مفتاح أيضاً هو وهذا تكميلية أصول بمثابة تكون أن الحوسبة وقوة للرموز يمكن ،خدمات يقدم أمامي نظام هناك يكون .الأرباح لتوليد فعالة.

هي الأولى المرحلة :مراحل ثلاث على ستكشف الرموز عصر في الحقيقية المنافسة بأن أحكم وأسعار التطبيقات برمجة وواجهات الحوسبة وقوة النماذج على الجميع يتنافس حيث ،الموارد منافسة العملاء وخدمة للكتابة الرموز الجميع يستخدم حيث ،التطبيقات منافسة هي الثانية المرحلة الرموز؛ الأنظمة منافسة هي الثالثة المرحلة البيانات؛ وتحليل المكتبي والعمل والتعليم والترجمة والبرمجة وأنظمة اللوجستية والأنظمة النشر وأنظمة الصناعية الأنظمة في الرموز يدمج من حيث ،الحقيقية ميزة على الحصول في الأرجحية له سيكون طويلة لفترة تعمل التي الحضارة تسجيل وأنظمة المعرفة .الأجل طويلة.

تجاهله يتم ما غالباً آخر مهم عامل هناك ،والمنصات والأنظمة والنماذج الرموز مناقشة عند ،الواقع في مع بسرعة؛ الأنظمة بناء يمكنهم ،المال برأس أنه الناس من العديد يعتقد .الوقت :الحقيقي العالم في .كذلك الواقع يكون لا ما غالباً ،ذلك ومع .بسرعة النجاح نماذج تكرر يمكنهم ،الكفاية فيه بما كبير سوق شراء يستطيع لا لكنه ،التطوير فرق وتوظيف ،الرموز وشراء ،الحوسبة قوة شراء المال لرأس يمكن من للعديد بالنسبة .عديدة سنوات مدى على منها التحقق تم التي الناضجة والأنظمة الحقيقية البيانات مستقر بشكل تعمل التي الأنظمة بل ،نفسها الرموز ولا ،نفسها النماذج ليس حقاً نادر هو ما ،الصناعات .طويلة لفترة الحقيقية الأعمال بيئات في.

كبير نظام .الصفر من لينمو تماماً جديداً نظاماً عادةً المال ورأس الأسواق تنتظر لا ،تجاري منظور من إلى العمليات صحة من والتحقق ،الوظائف وتطوير ،المعمارية الهندسة وتصميم ،المفهوم اقتراح من استثمار فقط يلزم لا ،العملية هذه خلال .أكثر حتى أو عديدة سنوات يستغرق ما غالباً ،الفعلية التنفيذ .الواقعية البيئة صحة من والتحقق والمراجعات والخطأ التجارب من الكثير أيضاً ولكن ،للموارد مستمر تشكيلها تم التي الأنظمة تفضيل إلى المال رأس يميل ،بالفعل السوق فرص تظهر عندما ،لذلك .الصفر من كامل نظام تصميم من بدلاً ،بالفعل منها التحقق تم والتي ،بالفعل تعمل والتي ،بالفعل.

اللاحق المال رأس تدخل صعوبة زادت ،أكبر النظام كان كلما .الواقع في آخر تناقض هناك ،ذلك ومع ،بها الخاصة التشغيل آليات بالفعل عادةً الكبيرة الأنظمة تكونت ،الأجل طويل التطور بعد .بسهولة وعمليات ،واسعة تاريخية بيانات لديها يكون قد .التظيمية والثقافات ،البيانات ومنطق ،الإدارة وهياكل

الكبيرة الأنظمة أن من الرغم على ، هذا بسبب .الأجل طويلة متراكمة خبرة وأنظمة ،معقدة تجارية وفقاً بالكامل هيكلتها إعادة أو بسرعة تحويلها الصعب من يكون ما غالباً أنه إلا ،هائلة قيمة ذات الذاتي قصوره زادت ،أكبر النظام كان كلما .المال رأس لتوقعات

على الاستحواذ ولا ،الصفر من نظام تصميم ليست المثلى الحالة ،الحقيقي الأعمال عالم في ،لذلك على منها التحقق وتم ،الأساسي هيكلها بناء بالفعل أكملت أنظمة على العثور بل ،متصلب كبير نظام ذات تزال لا فهي ،مثالية الأنظمة هذه تكن لم لو حتى .مستقر بشكل العمل ويمكنها ،الطويل المدى ،حقيقية وعمليات ،حقيقية بيانات وتمتلك ،صعوبة الأكثر المرحلة بالفعل تجاوزت لأنها جداً عالية قيمة الخارجي العالم على والانفتاح والتحسين التعبئة إعادة فإن ،ذلك على بناء .حقيقية تحقق وسجلات كبير بشكل أيضاً يقلل أن ويمكن ،الصفر من البدء من كفاءة أكثر يكون ما غالباً المنصات عبر والترويج تكلفة الأكثر المرحلة بالفعل أكملت قد الأنظمة هذه لأن .الوقت وتكاليف والخطأ التجربة مخاطر من التكرار في والأصعب للوقت استهلاكاً والأكثر

تزال لا الصناعة تكون عندما .غالباً تجاهله يتم مورد أيضاً هي الزمنية النافذة فإن ،المنظور هذا من ،السوق وحجم الرموز وأسعار النماذج قدرات على الناس من العديد يركز ،الأولى مراحلها في ،المفهوم اقتراح :متعددة بمراحل عادةً ناضج نظام ولادة تمر .النظام لتكوين اللازم الوقت متجاهلين والتحسين ،الحقيقي العالم في والتشغيل ،العمليات صحة من والتحقق ،المعمارية الهندسة وتصميم في مستقر بشكل النظام تشغيل على الحفاظ بل ،الفكرة اقتراح ليس حقاً الصعب الجزء .المستمر هذا فإن ،الناضجة الأنظمة قيمة إدراك في السوق يبدأ عندما ،لذلك .الطويل المدى على حقيقية بيئة مرة تكراره يمكن شيئاً كونها من بدلاً ،التراكم من لسنوات خضعت قد الأنظمة هذه أن غالباً يعني .قصير وقت في أخرى

ولا ،المفاهيمية المرحلة في الأفكار ليست مثالية الأكثر الأهداف فإن ،والسوق المال لرأس بالنسبة ،الأساسي الهيكل بناء أكملت التي الجينية الأنظمة بل ،بعد منها التحقق يتم لم التي الأولية النماذج تكون لا قد الأنظمة هذه أن من الرغم على .للترقية مجال ولديها ،حقيقية تشغيل سجلات وتمتلك أن بمجرد .خطورة والأكثر للوقت استهلاكاً والأكثر صعوبة الأكثر المرحلة بالفعل تجاوزت أنها إلا ،مثالية تم التي الناضجة الأنظمة عدد لأن ،بشدة مطلوبة هذه مثل الأنظمة تصبح ،بسرعة السوق طلب ينمو أن للنماذج يمكن .السوق يتصوره مما بكثير أقل يكون ما عادة الطويل المدى على حقاً منها التحقق لا ولكن ،مؤقت بشكل التطوير فرق تجميع ويمكن ،باستمرار تتوسع أن للرموز ويمكن ،بسرعة تتكرر وقت في الأجل طويلة التحقق وسجلات الحقيقية والعمليات الحقيقية البيانات إنشاء إعادة يمكن .قصير

على أيضاً يؤثر بل ،الأنظمة تكوين على فقط الوقت يؤثر لا ،الرموز لاقتصاد المستمر التطور في ،لذلك هذه تكون ما غالباً ،الأنظمة من معين نوع أهمية تماماً أدرك قد السوق يكون لا عندما .الأنظمة قيمة السوق يدرك وعندما الأدنى؛ هي الأنظمة هذه على الحصول تكلفة فيها تكون التي الفترة أيضاً هي يتم ما ،الصناعات من للعديد بالنسبة .للا غاية محدودة المتاحة الخيارات تكون ما غالباً ،عام بشكل قيمتها بالفعل أكملت والتي التطوير من التالية المرحلة في للدخول زمنية نافذة بل ،مشروع مجرد ليس تفويته الأجل طويل تراكمًا .

إلى الوصول ويمكن ،بسرعة الرموز على الحصول يمكن .ينطبق نفسه الأمر ،الرموز اقتصاد منظور من التي الأنظمة هو حقًا بسرعة تكراره يصعب ما ولكن ،بسرعة المنصات بناء ويمكن ،بسرعة النماذج الرموز تصبح عندما ،المستقبل في .القيمة إنتاج في الاستمرار ويمكنها عديدة سنوات مدى على تراكت يمتلك عندما لأنه .أكثر تزداد قد بل ،فحسب تخفض لن الأنظمة أهمية فإن ،متزايد بشكل وفيرة بيانات يمتلك من بل ،الرموز كمية يكون لن المنافسة نتيجة سيحدد ما فإن ،مماثلة رمزية موارد الجميع الحقيقي العالم احتياجات إلى أقرب نظام وهياكل ،استقراراً أكثر تجارية وعمليات ،نضجاً أكثر حقيقية

الأنظمة فهم على الناس من العديد اعتاد .الأنظمة أصل هي تجاهلها يتم ما غالباً أخرى قضية شركات بواسطة الأنظمة وتطوير تصميم يجب أنه طبيعي بشكل يفترضون وبالتالي ،برمجة كمنتجات مستقر بشكل حقًا تعمل التي الأنظمة فإن ،المعقدة الصناعات من العديد في ،ذلك ومع .البرمجيات يتم بل ،للصناعة إعطاؤها ثم البرمجيات شركات بواسطة أولاً تصميمها يتم لا غالباً الطويل المدى على ،وأتمتة ،برمجيات إلى تحويلها يتم ثم ،الأجل طويلة التجارية الممارسة سياق في تدريجياً تشكيلها منصات إلى وتحويلها .

في بالضرورة تبقى لا لكنها ،التقني التنفيذ في عادةً البرمجيات شركات تفوق .معقدًا ليس السبب اكتساب في صعوبة تواجه لكنها ،بسرعة الوظائف تطوير يمكنها .طويلة لفترة المحدد التجاري السياق بشكل تتطور التي الأنظمة فإن ،المقابل في .قصير وقت في الحقيقية التجارية الخبرة من كبيرة كمية والاستثناءات والطلبات العملاء مواجهة في باستمرار وتحسينها مراجعتها تتم الصناعة داخل عضوي من الأنظمة هذه تنطلق لا .يومي بشكل السوق في والمنافسة التكلفة وضغوط التنظيمية والتغيرات أتمتة رابط وكل ،قواعدها من قاعدة وكل ،عملياتها من عملية فكل الحقيقية؛ المشاكل من بل ،النظرية الحقيقي للعمل للأجل طويل التحقق من عادةً تأتي ،فيها

والعديد ،الصحية والرعاية ،والتصنيع ،والتمول ،والنشر ،التوريد وسلاسل ،اللوجستية الخدمات في ،لذلك الأنظمة ولا ،تعقيداً الأكثر الأنظمة ليست غالباً قيمة الأكثر الأنظمة فإن ،الأخرى المهنية المجالات من حل ويمكنها الطويل المدى على اختبارها تم التي تلك بل ،الوظائف من عدد أكبر على تحتوي التي أحدث أو ،بريقاً الأكثر الترويج أو ،الواجهات أحدث لديها يكون لا قد .مستقر بشكل الحقيقية المشكلات

التجاري والمنطق، الحقيقية والعمليات، الحقيقية البيانات من كبيرة كميات تمتلك لكنها، تقنية بنية الأجل قصير التطوير خلال من الخبرة هذه على الحصول يمكن لا غالباً. الأجل طويل المتراكم

في النماذج قدرات تستمر قد، المستقبل في. خاص بشكل مهم هذا، الرموز اقتصاد منظور من وأعلى أعلى البرمجيات تطوير كفاءة وستصبح، الانخفاض في الرموز تكاليف تستمر وقد، التحسن بسبب تلقائياً تنشأ لن عقود مدى على ما صناعة لدى تراكتت التي التجارية الخبرة فإن، ذلك ومع معالجة وخبرة، الصناعية المعرفة بل، نفسه البرنامج رمز ليس حقاً تكراره يصعب ما. النماذج ترقيات الطويل المدى على منها التحقق تم التي التجارية والعمليات، القرار اتخاذ ومنطق، الاستثناءات النظام وراء والمخفية.

عن تماماً منفصل بشكل المصممة تلك تكون لن المستقبل في مثالية الأكثر الأنظمة أن أعتقد، لذلك طويل التشغيل خلال من منها التحقق وتم، الحقيقية الأعمال من تتبع التي تلك بل، الصناعية الممارسة كلاً الأنظمة هذه تمتلك. الرموز وقدرات الاصطناعي الذكاء بمساعدة باستمرار ترقيتها يتم ثم، الأجل تكون ما وغالباً، الاصطناعي الذكاء عصر يجلبها التي الكفاءة ومكاسب الحقيقي العالم موثوقة من التقنية المفاهيم على فقط المبنية الأنظمة تلك من بكثير أعلى الأجل طويلة قيمتها.

الحضارة ذاكرة إلى العابرة المحادثات من: الرموز تتدفق أين - عاشراً

إنتاج كيفية على الناس من العديد يركز. أخرى ملاحظة لدي، المستقبلي الرموز باقتصاد يتعلق فيما، فقط واحدة محادثة في الرمز بقي إذا. النهائي تدفقها لمكان يُولى الاهتمام من القليل لكن، الرموز عمره يكون فقد، عادية ويب صفحة في دخل إذا فقط؛ دقائق بضع الافتراضي عمره يكون فقد عندما. جذري بشكل قيمته تتغير، الأجل طويل حفظ نظام في دخل إذا فقط؛ أشهر بضعة الافتراضي الأجل طويلة الحفظ وأنظمة، TROVE، و WorldCat، و Zenodo، و DOI، معرفات في الرموز تدخل في تبدأ بل رموز؛ مجرد تعد لم فإنها، المختلفة والمعرفية الأكاديمية التحتية والبنى، الوطنية للمكتبات في حقاً المهم السؤال أن أعتقد، لذلك. للحضارة وذاكرة، أدبية وأصول، معرفية سجلات إلى التحول حيث هياكل في الدخول من الرموز تمكين كيفية بل، الرموز من المزيد توليد كيفية ليس المستقبل دورها لعب في والاستمرار الطويل المدى على حفظها يمكن.

ماضي. أعمى بشكل بها أؤمن لا لكنني، للرموز المستقبلي التطور بشأن متفائلاً أظل، السبب لهذا في مهمة وتجارية قياس وحدة هو الرموز حاضر اللغة؛ لمعالجة للكمبيوتر الأساسية الوحدة هو الرموز الرقمية للحضارة مهماً أساسياً مورداً يصبح أن المرجح من الرموز مستقبل الاصطناعي؛ الذكاء عصر بل، الرموز ليس حقاً نادر هو ما بأن دائماً أتمسك فإنني، الرموز تطور كيفية عن النظر بغض، ذلك ومع يمكنهم الذين أولئك بل، أكثر موارد لديهم الذين أولئك وليس الرموز؛ تسخير على القادرون الأشخاص

الذين أولئك بل ،المحتوى من المزيد توليد يمكنهم الذين أولئك وليس الأجل؛ طويلة هياكل إنشاء
معرفية أنظمة في المحتوى تنظيم يمكنهم

شراؤها يمكن لا أصول أربعة :للضغط الوقت قابلية عدم - عشر حادي

معظم تتجاهلها أساسية قضية إلى الإشارة إلى أحتاج ،"الرموز بعد حقاً نادر هو ما" مناقشة في
الاصطناعي الذكاء كان مهما .النسخ أو الشراء أو للتسريع بطبيعتها قابلة غير الأصول بعض :المناقشات
الجوهرية بندرتها دائماً ستحتفظ التالية الأربعة الأصول فإن ،رخصة الرموز كانت ومهما ،قوباً

للضغط قابل غير الوقت ،أولاً

يمكن ولا ،سنوات ثلاث في ضغطها يمكن لا عاماً؛ ثلاثون ببساطة هي الممارسة من عاماً ثلاثون
نظام بناء كيفية" حول مقال توليد الاصطناعي للذكاء يمكن .الاصطناعي الذكاء بواسطة محاكاتها
والتغييرات ،والأعطال ،المختلفة الشذوذات ،ثوان في ،يعيش أن يمكنه لا لكنه ،ثوان في "لوجستي
تقصير يمكن لا .الزمن من عقدين مدى على اللوجستي النظام يواجهها التي السوق وتقلبات ،السياسية
لنظام الأولى للنموذج تجربتي بدء .تكنولوجية وسيلة بأي هذه الحقيقي العالم في الوقت تراكمات
الزمنية النقاط هذه - 2005 عام في الذكي الفواتير نظام وتصميم ،1993 عام في بُعد عن العمل
مقالتي قراءة يمكنه لاحقاً يأتي شخص أي .الحقيقي للتاريخ فيها رجعة لا علامات إنها حوادث؛ ليست
أصول هي نفسها السنوات تلك - عشتها التي السنوات امتلاك يمكنه لا لكنه ،طريقي وتعلم

شراؤها يمكن لا التخصصات متعددة الخبرة ،ثانياً

الخدمات حول كتاب مائة شراء للشخص يمكن .سلعة مثل التخصصات متعددة الخبرة شراء يمكن لا
اليوم ذلك في ،اكتساب يمكنه لا لكنه ،واحد يوم في الأدبيات وفهرسة والتمويل والنشر اللوجستية
تراكم ليست التخصصات متعددة الحقيقية القدرة إن .المجالات هذه لربط العملية الخبرة ،نفسه
ونقل ،المشتركة القواسم وتحديد ،المختلفة المجالات بين روابط إنشاء على القدرة بل ،المعرفة
،الأخطاء وارتكاب ،الحقيقية المشكلات حل خلال من إلا القدرة هذه اكتساب يمكن لا .الأساليب
من خبرتي نقل على قدرتي في السبب .الآخر تلو مجال في الحكم وتراكم ،المسارات وتصحيح
في عاماً عشرين أمضيت لأنني بل ،معيناً كتاباً قرأت لأنني ليس النشر أنظمة إلى اللوجستية الأنظمة
ظهرت الزميتين الفترتين هاتين تقاطع عند - النشر في أخرى سنوات وعشر اللوجستية الخدمات مجال
للنقل القابلة المنهجية

تكرارها يصعب الأنظمة بناء قدرة ،ثالثاً

يمكن. ولصقتها نسخها يمكن التي العمليات أو البرمجية الأكواد من مجموعة ليست الأنظمة بناء قدرة بالتغييرات والتبصر، المفاضلة وأحكام، الكود وراء القرار اتخاذ منطق تكرار يمكن لا لكن، الكود نسخ وعدد، التغييرات عدد في أيضاً ولكن، اليوم فعله يمكنه ما في فقط النظام قيمة تكمن لا. المستقبلية في استمراره مع الماضي العقد خلال لها خضع التي المتطلبات تعديلات وعدد، الأعطال استرداد مرات موقعي مشاهدة شخص لأي يمكن. البناء قدرة على دليل ذاته حد في هو هذا "النظام تاريخ". العمل تاريخ من الثلاثة العقود وليس، النتيجة هو الزوار يراه ما لكن (<https://times.net.au>) الإلكتروني يمكن لا. النتيجة تلك وراء المستمرة والصيانة، الهيكلية والتعديلات، التكنولوجيا واختيارات، القرار اتخاذ التقليد أو الانتحال طريق عن هذه البناء قدرة على الحصول.

الأجل طويلة التشغيل سجلات اختصار يمكن لا، رابعاً.

سجل يثبت. للاختصار بطبيعته قابل غير ذاته حد في هو الزمن من لعقدين يعمل النظام أن سجل، الاصطناعي الذكاء عصر في. تصميمه وبصيرة للصيانة وقابليته النظام استقرار الأجل طويل التشغيل من عقدين منذ يعمل نظام" إنشاء لكن، دقائق في "شكلي" نظام أو ويب موقع إنشاء يكتمل قد له ليس "الأسترالية الوطنية والمكتبة WorldCat و CERN بواسطة دائم بشكل أرشفته وتم الزمن أمام صمد قد النظام هذا أن للجميع يعلن، زمني طابع هو الأجل طويل التشغيل سجل. اختصارات اصطناعي ذكاء تقنية أي بواسطة السجل هذا مثل تسريع أو تزوير يمكن لا. الحقيقي العالم اختبار.

وسجلات، الأنظمة بناء وقدرة، التخصصات متعددة والخبرة، الوقت – الأربعة الأصول هذه تشكل ليس أمتلكها أنا. "حقاً نادر" أنه على المقال هذا في ناقشته لما الأساس معاً – الأجل طويلة التشغيل. الحاضر الوقت حتى واستمرت الآخرين قبل بدأت لأنني بل، الآخرين من اجتهداً أكثر أو أذكى لأنني هذه فإن، المنال وسهلة رخيصة الأشياء من والمزيد المزيد الاصطناعي الذكاء فيه يجعل عصر في. متزايدة قيمة ذات ستصبح اختصارها أو تكرارها أو شراؤها أو ضغطها يمكن لا التي الأشياء.

"الأنظمة بناء" إلى "الرموز تكديس" من: الخاتمة - عشر ثاني

نفسها الرموز في يكمن لا الرموز مستقبل.

عرض سيستمر، الاستدلال لتكاليف المستمر والانخفاض الكبيرة النماذج لتقنية المستمر التطور مع وسعة الإنترنت مرور حركة مثل وشائعة رخيصة ستكون اليوم الثمن باهظة الرموز. التوسع في الرموز كبيرة كميات على الحصول من الجميع يتمكن عندما. سنوات بضع بعد الصلبة الأقراص على التخزين قيمة إلى الرموز تحويل على القدرة بل، الرموز يكون لن حقاً نادراً سيكون ما فإن، بسهولة الرموز من العملي والتراكم، الصناعية والخبرة، الأتمتة وقدرة، الهيكلية التصميم قدرة – الحقيقي العالم في

أو تكرارها أو شراؤها أو ضغطها يمكن لا التي الأشياء تلك في الأعماق الندرة وتكمن. الأجل طويل طويلة التشغيل وسجلات، الأنظمة بناء وقدرة، التخصصات متعددة والخبرة، نفسه الوقت: اختصارها كميات خلال من مباشرة الأنظمة بناء قدرة اكتساب إمكانية عدم سبب فإن، سابقاً نوقش كما. الأجل ومعالجة، الحكم على قدرة بل، توليدية قدرة جوهرها في ليست أنها هو الرموز من كبيرة القدرات لهذه المشتركة والسمة – الثقة وتراكم، الراجعة التغذية وامتصاص، والمقايضة، الاستثناءات وتحققاً، وتصحيحات، وأخطاء، وممارسة، الحقيقي العالم من وقتاً تتطلب أنها هي.

صعوبة وسيزداد، بالفعل صعب الرموز من المال كسب: أيضاً الحكم هذا ينطبق، الربحية مستوى على إعادة مجرد في المتمثل الأعمال ونموذج، الربح هوامش ضغط على باستمرار الأسعار حروب تعمل بنظام مرتبطة الرموز تكون عندما فقط. معدومة حتى أو ضئيلة أرباح نحو بسرعة يتجه الرموز بيع تكميلية أصول بمثابة تكون أن الحوسبة وقوة للرموز يمكن – خدمات نظام يقدم عندما أي – أمامي وقود مجرد هي الرموز. الربح مصدر هو النظام الربح؛ مصدر ليست نفسها الرموز. الأرباح لتوليد فعالة ومحركه وليست، للنظام.

للتطور خاصة أهمية لها بتطويرها أقوم أو بالفعل أكملتها التي الأنظمة فإن، المنظور هذا من وقت قبل موجودة كانت بل عابرة؛ موضة لمطاردة مؤقتاً تبين لم الأنظمة هذه لأن. للرموز المستقبلي الواقعية الأعمال أن الذكي اللوجستي النظام أثبت. عاماً موضوعاً الرموز مفهوم يصبح أن من طويل متعدد أدبي نظام إنشاء أيضاً للفرد يمكن أنه النشر نظام أثبت وأتمتها؛ وإجراءاتها هيكلتها يمكن الويب موقع نظام أثبت الأرشفة؛ قنوات ومتعدد، DOI معرفات ومتعدد، المنصات ومتعدد، اللغات الربط نظام يثبت باستمرار؛ ونشره وفهرسته الرقمي المحتوى تنظيم يمكن أنه الوصفية والبيانات تكديس مواصلة ليس حقاً يهم ما فإن، باستمرار الأعمال عدد يزداد عندما أنه أيضاً التلقائي التبادلي الإلكتروني "العصر التحول" موقع. المحتويات بين هيكلية علاقات إنشاء بل، المحتوى على تجريبي دليل ذاته حد في هو، الإنجازات لهذه مركزية عرض كمنصة (<https://times.net.au>)، للتحقق وقابل ودائم مفتوح أكاديمي نظام إلى الرموز قدرة تحويل.

مختلفة تكون قد المنتجة النتائج فإن، المستقبل في الرموز من الكمية نفس الجميع إعطاء تم لو حتى محتوى من المزيد على البعض وسيحصل، المحادثة سجلات من المزيد على البعض سيحصل: تماماً معرفي نظام أو، نشر نظام أو، لوجستي نظام بإنشاء الآخر البعض يقوم قد بينما، الويب صفحات – وراءها تقف التي النظام هندسة بل، نفسها الرموز ليس الفرق يحدث ما. لعقود العمل في يستمر يصعب التي البناء وقدرة، للشراء القابلة غير التخصصات متعددة والخبرة، للضغط القابل غير والوقت تلك النظام هندسات تدعم والتي اختصارها يمكن لا التي الأجل طويلة التشغيل وسجلات، تكرارها.

الكهرباء مثل مهمة الرموز. الذكي للعصر الجديدة التحتية البنية أعتبرها بل ، غاية الرموز أعتبر لا ، لذلك يخلق لا وحده المال لكن ، المال مثل للقياس قابلة الرموز تلقائياً؛ المصانع تبنى لا وحدها الكهرباء لكن ، توزيع وأنظمة ، أنابيب بدون لكن ، الماء مثل تتدفق أن للرموز يمكن تلقائياً؛ الجودة عالية أصولاً لن المستقبل في الحقيقية القيمة ذوو الأشخاص . فقط ستتبدد فإنها ، الأجل طويلة صيانة وسجلات في الرموز تنظيم يمكنهم الذين أولئك بل ، الرموز من عدد أكبر لديهم الذين أولئك بالضرورة يكونوا هي الرموز عصر في فرصة وأكبر هدفاً؛ الرموز اعتبار هي الرموز عصر في مفارقة أعظم . أنظمة في الكهرباء مثل شائعة الرموز تصبح عندما . تكرارها للآخرين يمكن لا هياكل لبناء الرموز استخدام حقاً حدد ما أن يكتشفون قد اليوم الرموز جنون على الورا إلى ينظرون الذين الناس فإن ، المستقبل على القدرة يمتلكون كانوا إذا ما ولكن ، يمتلكونها التي الرموز عدد أبداً يكن لم الناس بين الفجوة وراء الأشياء وتلك – باستمرار الأنظمة وبناء ، المستمر والتراكم ، المستمرة والممارسة ، المستمر التعلم هي القدرة هذه ، رأيي وفي . اختصارها أو تكرارها أو شراؤها أو ضغطها يمكن لا التي القدرات هذه الرموز عصر في ثروة وأصعب أندر .

يمكن المستقبل الاصطناعي الذكاء لعصر التافسي المشهد أن متزايد بشكل أشعر ، بحثي تعمق مع ولكن ، أساسية كموارد والرموز والنماذج الحوسبة قوة الناس من العديد يعتبر . بسيط بتشبيه تلخيصه قوة تحدد . الجيش خلف والتموين الإمداد بنظام أشبه فهي ، الحقيقي العالم في العمليات منظور من معاً تشكل فهي التوزيع؛ قدرة الرموز وتحدد ، التنظيمية القدرة النماذج وتحدد ، الإنتاجية القدرة الحوسبة الاصطناعي الذكاء لعصر المادي الدعم نظام .

الخط في القتال نظام بل ، نفسها اللوجستية الإمدادات ليس المعركة نتيجة حقاً يحدد ما ، ذلك ومع كيفية الأنظمة تحدد . الهجوم بخط أشبه فالأنظمة ، المستقبلية الصناعية للمنافسة بالنسبة . الأمامي حقاً يستفيدون المستخدمون كان إذا وما ، القيمة خلق وكيفية ، العمليات تنفيذ وكيفية ، الموارد استخدام في المكسدة الإمدادات مثل فهي ، النظام تنظيم بدون ، وفيرة الحوسبة وقوة الرموز كانت مهما . حقيقة نتائج إلى التحول على قادرة غير ، مستودع .

بل ، الرموز بين ولا ، الحوسبة قوى بين الأرجح على ليست المستقبلية المنافسة فإن ، المنظور هذا من الذخيرة هي الرموز القتالية؛ والقدرة الهجوم اتجاه ويحدد ، الهجوم خط هو النظام . الأنظمة بين وتحدد ، الخلفية الصناعية القاعدة هي الحوسبة قوة المستدامة؛ القتالية القدرة وتحدد ، والإمدادات مع ولكن طويلة؛ لفترة يتقدم أن الهجوم لخط يمكن لا ، خلفي إمداد بدون . التوريد وقدرة الإنتاج سرعة . أيضاً النتائج تحقيق يمكن لا ، هجوم خط نظام وجود وعدم فقط خلفي إمداد وجود

من . النظام هو حقاً نادراً يزال لا ما فإن ، الرموز ووفرة ، النماذج وتشابه ، الحوسبة قوة انتشار مع ، لذلك ربط على والقدرة ، أطول لفترة منها التحقق تم وعمليات ، واقعية أكثر وبيانات ، نصجاً أكثر نظاماً يمتلك

في مستدامة ميزة على للحصول عرضة أكثر سيكون ،الواقعية بالاحتياجات الحوسبة وقوة الرموز المستقبل.

أن كما .ذاتها حد في غاية ليست لكنها ،عنها غنى لا – الاصطناعي الذكاء عصر "كهرباء" الرموز أصبحت اخترعوا الذين الأعمال ورواد المخترعون بل ،الطاقة محطات يكونوا لم الكهرباء ثورة في الفائزين عصر في الحقيقيين الفائزين فإن ،الكهربائي التحليل وحوض الكهربائي والمحرك الكهربائي المصباح لا وهندسات هياكل لبناء الرموز استخدام على القادرين الأنظمة بناء سيكونون الاصطناعي الذكاء حقاً نادر هو ما هو هذا ،الرموز بعد .تضاهي

الذكاء لعصر التنافسي المشهد أن متزايد بشكل أشعر ،بأكمله المقال هذا في الورا إلى بالنظر الشجرة بتشبيه توضيحه يمكن المستقبلي الاصطناعي

إن .بسهولة قياسهما ويمكن وملموسان مرئيان لأنهما ،والمغذيات الماء على الناس من العديد يركز قوة تحدد .باستمرار توصيلها يتم التي والمغذيات الماء مثل هي اليوم والرموز والنماذج الحوسبة قوة تقدم مع .التوزيع قدرة الرموز وتحدد ،التحويل قدرة النماذج وتحدد ،الإنتاجية القدرة الحوسبة المنال وسهلة متزايد بشكل وفيرة الموارد هذه أصبحت ،التكنولوجيا

الجزور نظام بل ،نفسها والمغذيات الماء ليس الشجرة حيوية حقاً يحدد ما فإن ،ذلك ومع

المستقبلية الرقمية الحضارة جذر هو النظام

يمكن بأكملها الشجرة كانت إذا وما ،المغذيات توزيع وكيفية ،الماء امتصاص كيفية الجزور نظام يحدد ،جذور بدون يتسرب؛ سوف ،الماء كمية عن النظر بغض ،جذور بدون .الطويل المدى على تنمو أن النظر بغض ،نظام بدون ،وبالمثل .حيوية إلى تحويلها يمكن لا ،المغذيات وفرة مدى عن النظر بغض مستمر بشكل قيمة خلق الصعب فمن ،الحوسبة قوة اتساع أو ،الرموز رخص أو ،النموذج قوة عن

تنقل التي الأنابيب هي والنماذج ،المغذيات هي والرموز ،الماء هي الحوسبة قوة ،المنظور هذا من نتيجة حقاً يحدد ما .الحقيقي العالم في بعمق يتغلغل الذي الجزور نظام هو والنظام ،المغذيات وأكثر أعماق جذور لديه من بل ،أكثر مغذيات لديه من ولا ،أكثر ماء لديه من ليس المستقبلية المنافسة .الطويل المدى على الموارد هذه واستخدام امتصاص على وقادرة استقراراً

أسعار انخفاض واستمرار ،النماذج تطور واستمرار ،الحوسبة قوة نمو استمرار مع ،المستقبل في الأنظمة بل ،الآن بعد الموارد يكون لن حقاً نادراً سيكون ما .شيوعاً أكثر نفسها الموارد ستصبح ،الرموز الجزور؛ يمتلك ناضجاً نظاماً يمتلك من .الموارد من باستمرار قيمة وخلق واستخدام تنظيم على القدرة المستمر والنمو الأجل طويل البقاء على القدرة يمتلك الجزور يمتلك ومن

بل ،الحوسبة قوة من المزيد ولا ،الرموز من المزيد ليس الرموز بعد حقاً نادر هو ما أن أعتقد ،لذلك الحقيقى العالم فى قيمة إلى والرموز والنماذج الحوسبة قوة تحويل على القادرة الأنظمة

على تنمو أن يمكن الشجرة كانت إذا ما حقاً يحدد ما فإن ،وفرة أكثر والمغذيات الماء يصبح عندما دائماً الجذور نظام هو الطويل المدى

اقتصاد ،حوسبة قوة ،كبير لغوي نموذج ،AI ،اصطناعى ذكاء ،(Token) رمز :المفتاحية الكلمات (Agent) وكيل ،البيانات سيادة ،BPE خوارزمية ،إنجليزي رمز ،صيني رمز ،الرموز اقتصاد ،النماذج بحث ،أتمتة نظام ،الأنظمة تصميم ،الأنظمة بناء قدرة ،(API) التطبيقات برمجة لواجهة بيئي نظام ،الأجل طويل تشغيل نظام ،حقيقية تاريخية بيانات ،التخصصات متعددة ممارسة ،التخصصات متعدد ،معرفية تحتية بنية ،الحضارة ذاكرة ،رقمية حضارة ،الأنظمة منافسة ،المنصات منافسة ،زمنى أصل ،وصفية بيانات نظام ،لغات بعشر نشر نظام ،ذكي لوجستي نظام ،مستقل دولي باحث ،مستقل باحث العصر التحول ،TROVE ،WorldCat ،Zenodo ،DOI الرقمي الكائن معرف ،تلقائي تبادل ربط وو هوي تشاو جيڤي ،هوي تشاو وو ،(The Epochal Transition)

المراجع:

1. *A Structural Case Study of the 1993 Remote Work Prototype System*
DOI: <https://doi.org/10.5281/zenodo.17978371>
2. *The 2005 Intelligent Invoice System*
DOI: <https://doi.org/10.5281/zenodo.19044755>
3. *Mature Structural Logistics System*
DOI: <https://doi.org/10.5281/zenodo.20201895>
4. **Cross-Generational 20-Year Intelligent Logistics and Financial System**
DOI: <https://doi.org/10.5281/zenodo.20352371>
NLA: <https://nla.gov.au/nla.obj-4204311557>
5. *Analysis of the Homepage Structural Evolution of The Epochal Transition*
DOI: <https://doi.org/10.5281/zenodo.20362574>
NLA: <https://nla.gov.au/nla.obj-4204309341>
6. *An Empirical Study of Institutional Engineering: Independent Scholars Accessing the WorldCat Global Bibliographic Infrastructure*
DOI: <https://doi.org/10.5281/zenodo.18028572>
7. *National-level Network Archiving and Digital Document Preservation Guidelines of the Australian Winner Information Network*
DOI: <https://doi.org/10.5281/zenodo.17888259>
8. *The Institutional Archiving of The Epochal Transition*
DOI: <https://doi.org/10.5281/zenodo.17932379>

WU, J. C. H. (2026). Source Declaration for Audiovisual and Derivative Adaptations of a Continuing Real-World Narrative. Zenodo.

<https://doi.org/10.5281/zenodo.18160116>

Nach den Tokens: Was im Zeitalter der KI wirklich knapp ist

Autor/unabhängiger internationaler Gelehrter: Wu Zhaohui, JEFFI CHAO HUI WU

ORCID: <https://orcid.org/0009-0009-5318-3027>

Vorab-Erklärung

Dieser Artikel basiert auf meinen persönlichen Beobachtungen aus über dreißig Jahren interdisziplinärer Praxis (seit meinem Studium in Australien 1989, einschließlich Fernarbeit, Logistikmanagement, automatisierter Verarbeitung, Website-Erstellung, Kulturverbreitung, mehrsprachigem Publizieren und interdisziplinärer Forschung). **Tokens sind ein sich schnell entwickelndes neues Phänomen; dieser Artikel stellt nur meine persönlichen, auf den Zeitpunkt der Abfassung (Juni 2026) bezogenen, momentanen Beobachtungen und Urteile dar und stellt in keiner Weise eine Anlageberatung, Geschäftsprognose oder Branchennorm dar.**

Dieser Artikel erhebt keinen Anspruch auf universelle Anwendbarkeit und beansprucht nicht, einen Branchenkonsens zu vertreten. Da sich der Token-Markt, die KI-Technologie und die globalen politischen Rahmenbedingungen ständig weiterentwickeln, müssen einige Urteile in diesem Artikel möglicherweise in Zukunft überarbeitet oder aktualisiert werden. Der Leser sollte auf der Grundlage der aktuellsten Informationen ein unabhängiges Urteil fällen.

Die "fünf knappen Fähigkeiten" und die "fünf persönlichen Vermögenswerte", die ich erörtere, stammen aus meiner eigenen Praxis des Systemaufbaus (öffentliche Belege verfügbar unter <https://times.net.au>), einschließlich der vollständigen Liste von über 40 unabhängigen kreativen Werken auf dieser Website, der ISSN 3083-5178, über 60 DOIs und der dauerhaft archivierten Aufzeichnungen durch

internationale Infrastrukturen wie WorldCat, die National Library of Australia und CERN Zenodo).

Um eine aus dem Zusammenhang gerissene Betrachtung zu vermeiden, setzt eine gültige Widerlegung oder Infragestellung dieses Artikels die folgende Bedingung voraus: **die vollständige Lektüre aller in diesem Dokument verlinkten Artikel und der darin enthaltenen verschachtelten Links sowie aller meiner**

unter <https://times.net.au> aufgeführten Werke. Teilweise Zitate oder vereinfachte Zusammenfassungen, die aus diesem Gesamtkontext gelöst sind, können keine faire Bewertung dieses Artikels darstellen.

Ergänzender Hinweis zur gültigen Widerlegung:

Dieser Artikel vertritt persönliche Ansichten, erhebt keinen Anspruch auf universelle Anwendbarkeit und setzt keine Zitation der Werke anderer voraus. Jegliche gültige Infragestellung sollte einen Äpfel-mit-Äpfel-Vergleich anwenden: Fakten gegen Fakten, persönliche Praxis gegen persönliche Praxis, fertige Systeme gegen fertige Systeme. Zu behaupten, dass "zukünftige KI dies könnte", um "bereits abgeschlossene Praktiken" zu negieren, oder die Fähigkeiten von Teams/Institutionen mit individueller Praxis zu vergleichen, stellt keinen fairen Vergleich dar.

Das in diesem Artikel beschriebene Portfolio von fünf Vermögenswerten ist ein Aufzeichnung persönlicher Praxis und behauptet nicht, dass "es keinen vergleichbaren Fall gibt". Wenn Sie der Meinung sind, dass ein vollständig vergleichbarer Fall existiert, nennen Sie bitte einen spezifischen öffentlichen Fall und erläutern Sie seine entsprechenden Belege in den fünf Dimensionen der realen historischen Daten, der domänenübergreifenden Systeme, der Systembaufähigkeit, der zukunftsgerichteten Vision und der Selbstverifizierungsfähigkeit. Andernfalls ist jede Infragestellung lediglich ein Meinungsausdruck und stellt keine gültige Widerlegung dar.

Zusammenfassung

Tokens werden zur zentralen Maßeinheit im Zeitalter der KI, aber dieser Artikel wird argumentieren, dass der Token nicht der endgültige Wert ist. Da die Inferenzkosten weiter sinken, werden Tokens so billig und allgegenwärtig sein wie heute der Internetverkehr und die Festplattenspeicherung. Wenn Tokens nicht mehr knapp sind, was wird dann wirklich knapp sein? Basierend auf über dreißig Jahren interdisziplinärer Praxis des Autors (von einem Fernarbeits-Prototypsystem 1993, einem intelligenten Rechnungssystem 2005 bis hin zu intelligenten Logistiksystemen, einem Zehn-Sprachen-Publikationssystem, DOI-Management, automatischer Kreuzreferenzierung und plattformübergreifendem Publizieren) stellt dieser Artikel die folgenden Kernurteile auf: Erstens, die Fähigkeit zum Systemaufbau – einschließlich Strukturdesign, Automatisierung, domänenübergreifender Praxis, mentaler Modelle für den Langzeitbetrieb und der Fähigkeit, Tokens in reale Geschäftsprozesse einzubetten – ist der wirklich knappe Vermögenswert. Zweitens, die Fähigkeit zum Systemaufbau kann nicht direkt durch große Mengen von Tokens erworben werden, weil sie von Entscheidungsfindung, Ausnahmebehandlung, Kompromissen, realen Feedbackschleifen und im Laufe der Zeit angesammeltem Vertrauen abhängt. Drittens, die Inkompressibilität der Zeit, die Nichtkäuflichkeit domänenübergreifender Erfahrung, die schwierige Reproduzierbarkeit der Systemaufbaufähigkeit und die Unmöglichkeit, langjährige Betriebsaufzeichnungen abzukürzen – diese vier Vermögenswerte bilden die tiefste Knappheit im Token-Zeitalter. Dieser Artikel verwendet die vom Autor erstellte akademische Website "The Epochal Transition" (<https://times.net.au/>), ISSN 3083-5178, dauerhaft archiviert durch internationale Infrastrukturen wie WorldCat, CERN Zenodo und TROVE der National Library of Australia) als öffentlichen Beweis und zeigt, wie eine Person über dreißig Jahre hinweg ein interdisziplinäres, überprüfbares, langlaufendes Systemarchitektur aufbauen kann.

Dieser Artikel argumentiert: Tokens sind nur der Treibstoff; das System ist der Motor. Nach den Tokens ist das, was wirklich knapp ist, die Person, die Tokens beherrschen kann, und die Strukturen, die sie aufgebaut hat.

"Ohne Waren, wozu das Geld?"

Wenn ich die Geschichte der Token-Entwicklung Revue passieren lasse, wird mir zunehmend bewusst, wie sehr sie vielen technologischen Revolutionen ähnelt, die die Menschheit in der Vergangenheit erlebt hat. Jedes Mal, wenn eine neue grundlegende Ressource auftaucht, konzentrieren sich die Menschen zuerst auf die Ressource selbst und vernachlässigen die Struktur dahinter. Dies galt für das Industriezeitalter, das Internetzeitalter und gilt gleichermaßen für das Zeitalter der künstlichen Intelligenz. Heute drehen sich viele Diskussionen um Token-Preise, Token-Kosten, Token-Kontingente und Token-Versorgungskapazitäten, als ob der Besitz von mehr Tokens die Zukunft bedeuten würde. Aber aus meinen jahrzehntelangen praktischen Erfahrungen heraus war die Ressource selbst nie der endgültige Wert. Was den Wert wirklich bestimmt, ist immer, wie Menschen diese Ressourcen organisieren und nutzen.

Um "Tokens" zu verstehen, darf man nicht einfach mit einem neuen Begriff beginnen, sondern muss mit ihrer Geschichte beginnen. Vom amerikanischen Philosophen Charles Sanders Peirce, der 1906 zwanzig Instanzen des Wortes "the" auf Papier schrieb und zwischen "Typ" und "Token" unterschied, bis hin zu Liu Liehong, dem Direktor der Nationalen Datenverwaltung Chinas, der 2026 auf dem Boao-Forum die offizielle chinesische Übersetzung "ciyuan" (词元) für Token festlegte, hat dieses Konzept einen langen Weg von der Philosophie zur Informatik und von der Blockchain zur künstlichen Intelligenz zurückgelegt. Das Wort "Token" existierte bereits seit langem in der westlichen Informatik, wo es in Programmiersprachen, Compilern und Textverarbeitungssystemen die grundlegende Einheit bezeichnete, die ein Computer erkennen und verarbeiten konnte. Später, in den Kontexten von Netzwerksicherheit, Authentifizierung, API-Aufrufen und Blockchain, wurde "Token" oft als "Pass" oder "Berechtigungsnachweis" verstanden, das Zugriffsrechte, Identitätsnachweise oder digitale Vermögenswerte anzeigt. Aber im Zeitalter der großen Sprachmodelle ist der Token nicht mehr in erster Linie ein Identitätsnachweis oder ein Blockchain-Vermögenswert, sondern die grundlegende Recheneinheit, mit der KI-Modelle Sprache, Code, Bildbeschreibungen und multimodale Informationen verarbeiten.

China nannte den Token im KI-Bereich später "ciyuan" (词元). Ich halte dies für eine ziemlich genaue Übersetzung. Sie verwendete nicht "lingpai" (Pass), weil der Token in großen Modellen kein Pass ist; sie übersetzte ihn auch nicht einfach als "zi" (Zeichen) oder "ci" (Wort), da er manchmal ein Zeichen, manchmal ein Wort, manchmal die Hälfte eines englischen Wortes und manchmal sogar nur ein Satzzeichen oder Symbol sein kann. "Ci" zeigt seine Beziehung zu Sprache und Information an, und "yuan" zeigt an, dass es sich um eine Grundeinheit handelt. Für den allgemeinen Leser kann man es einfach so verstehen: Was der Mensch als Satz sieht, sieht die KI als eine Kette von Tokens. Auf diesem Weg haben sich Tokens von einem obskuren linguistischen Begriff zur zentralen Maßeinheit des KI-Zeitalters und weiter zu einem Abrechnungsinstrument entwickelt, das technologisches Angebot mit kommerzieller Nachfrage verbindet. Als Praktiker, der langjährige Projekte wie den Aufbau intelligenter Logistiksysteme, die Gestaltung automatisierter Publikationsprozesse und die plattformübergreifende Inhaltsverteilung persönlich erlebt hat, habe ich ein unmittelbares Verständnis für diese Entwicklung – Tokens waren nie der Zweck, sondern das Mittel; was den Wert wirklich bestimmt, ist nie, wer mehr Tokens besitzt, sondern wer mit Tokens dauerhaft laufende Systeme und nicht reproduzierbare Strukturen aufbauen kann. Die in diesem Artikel diskutierte Systembaufähigkeit hat ihre Ergebnisse öffentlich auf <https://times.net.au> präsentiert – ein interdisziplinäres akademisches Publikationssystem, das von einem unabhängigen Gelehrten aufgebaut und dauerhaft im globalen Bibliothekssystem archiviert wurde.

Wenn ich auf meinen eigenen Werdegang zurückblicke, begann ich nicht erst mit dem Nachdenken über Systeme, nachdem Tokens auftauchten, sondern entwickelte allmählich ein Verständnis für Systeme durch langjährige Praxis. Seit meinem Studium in Australien 1989 habe ich Lern-, Arbeits- und Forschungsphasen in verschiedenen Bereichen durchlaufen. Später kam ich nach und nach mit Fernarbeit, Logistikmanagement, automatisierter Verarbeitung, Website-Erstellung, Kulturverbreitung, mehrsprachigem Publizieren und interdisziplinärer Forschung in Berührung. Rückblickend, obwohl diese Erfahrungen in verschiedenen Epochen stattfanden, haben sie eine gemeinsame Eigenschaft: Sie drehten sich immer

darum, wie man Strukturen aufbaut, wie man sich wiederholende Arbeiten reduziert und wie man die Gesamteffizienz verbessert. Das Konzept der Tokens existierte damals nicht, ebenso wenig wie die heutigen großen Sprachmodelle, aber ich dachte bereits ständig darüber nach, wie man Informationen automatisch fließen, organisieren und verarbeiten lassen kann.

Wenn also heute immer mehr Menschen Tokens als ein neues Produktionsmittel betrachten, neige ich eher dazu, sie als eine neue Form grundlegender Energie zu sehen. Energie ist zwar wichtig, aber Energie schafft nicht automatisch Werte. Strom allein wird nicht automatisch zu einer Fabrik; das Internet allein wird nicht automatisch zu einem Unternehmen; Tokens allein werden nicht automatisch zu einem Wissenssystem. Selbst wenn in Zukunft jeder leicht große Mengen an Tokens erhalten kann, bedeutet das nicht, dass jeder die gleichen Ergebnisse erzielen kann. Den Unterschied macht nicht die Menge der Tokens aus, sondern die Struktur, in die die Tokens eintreten.

Bevor ich eine eingehende Analyse der Tokens durchführe, muss ich eine weitere wichtige Beobachtung anführen: Tokens ähneln Währungen, und die technologischen, systemischen und politischen Unterschiede zwischen Ost und West führen dazu, dass sie nicht universell einsetzbar sind. So wie der US-Dollar und der chinesische Renminbi nicht frei getauscht werden können, stehen chinesische Tokens und englische Tokens vor ähnlichen Zirkulationsbarrieren. Dies ist nicht nur eine Frage des Sprachunterschieds; der tiefere Grund liegt in den dreifachen Unterschieden zwischen Ost und West auf der technologischen Ebene (Modellarchitekturen, Tokenisierungsalgorithmen, Trainingsdaten), der systemischen Ebene (API-Ökosysteme, Zahlungssysteme, Datenstandards) und der politischen Ebene (grenzüberschreitende Datenflüsse, Exportkontrollen, Überprüfungen der nationalen Sicherheit), die verhindern, dass Tokens als idealisierte "globale universelle KI-Währung" frei zirkulieren können. Diese Beobachtung hat erhebliche Auswirkungen auf das Verständnis der zukünftigen Token-Marktlandschaft – in Zukunft könnten "chinesische Token-Zonen", "englische Token-Zonen" und andere auf Sprache und Politik basierende Token-Ökosysteme entstehen, und

zoneübergreifende Aufrufe werden mit Effizienzverlusten und Compliance-Kosten konfrontiert sein.

I. Ursprünge der Tokens: Der vierfache Sprung vom philosophischen Konzept zur wirtschaftlichen Einheit

Die Identitätsentwicklung des Tokens lässt sich klar in vier Stufen einteilen.

Der Ursprung in der philosophischen Linguistik. 1906 unterschied Peirce beim Studium der Semiotik zwischen "Typ" und "Token". Ersteres repräsentiert eine abstrakte Regel oder Form, letzteres ist die spezifische Ausprägung dieser Regel in der Realität. Diese Unterscheidung wurde später zu einer wichtigen Grundlage der strukturellen Linguistik.

Die kodierungstechnische Wende in der Informatik. In den 1960er Jahren, als ein Programmierer die Codezeile "int x = 5;" schrieb, musste der Computer sie in unabhängige Einheiten wie "int", "x", "=", "5" usw. zerlegen. Jede dieser Einheiten war ein Token. Tokens wurden damit zur kleinsten Informationseinheit, die eine Maschine "lesen" konnte.

Das Werthalter-Experiment der Blockchain. 2017, mit dem Aufkommen des ICO-Booms, wurde dem Token eine neue Identität als "handelbarer digitaler Berechtigungsnachweis" gegeben. Obwohl diese Welle der Begeisterung allmählich abebbte, hatte sich das Verständnis des "Tokens als Wertesymbol" eingewurzelt.

Die wirtschaftliche Infrastruktur des KI-Zeitalters. Der März 2026 war ein bedeutsamer Moment – NVIDIA-CEO Jensen Huang definierte Rechenzentren als "Fabriken, die intelligente KI-Tokens produzieren" neu, und Liu Liehong, Direktor der Nationalen Datenverwaltung Chinas, gab die offizielle chinesische Übersetzung "ciyuan" (词元) und nannte sie den "Wertanker der intelligenten Ära". An diesem Punkt vollendeten die Tokens ihren letzten Sprung vom technischen Konzept zum wirtschaftlichen Konzept.

II. Die technische Basis: BPE-Algorithmus und die implizite "Sprachsteuer"

Die Tatsache, dass Tokens zur standardisierten Maßeinheit für KI werden konnten, beruht auf einer für über zwei Jahrzehnte vergessenen Technologie – der Byte-Paar-Kodierung (Byte Pair Encoding, BPE).

1994 veröffentlichte der amerikanische Programmierer Philip Gage einen Artikel über Datenkompressionsalgorithmen in einer C-Fachzeitschrift und stellte das Grundprinzip von BPE vor: wiederholtes Scannen des Textes und "Verschweißen" der beiden am häufigsten nebeneinander auftretenden Zeichen zu einem neuen Symbol, wobei iterativ Runde für Runde komprimiert wird. Diese Arbeit erregte damals keine Aufmerksamkeit, da ihre Kompressionseffizienz nicht herausragend war.

Erst 2016 entdeckte Rico Sennrich, ein Forscher der Universität Edinburgh, bei der Untersuchung des Problems der Wortsegmentierung in der maschinellen Übersetzung zufällig diesen alten Artikel und erkannte scharfsinnig, dass BPE die perfekte Lösung für das Tokenisierungsproblem war – es war nicht nötig, ein Wörterbuch vorzudefinieren; man ließ die Daten selbst "sprechen", und hochfrequente Kombinationen würden natürlich zu Tokens verschmelzen. 2019 übernahm OpenAI diesen Ansatz offiziell bei der Veröffentlichung von GPT-2 und setzte den Ausgangspunkt für die Tokenisierung direkt auf die "Byte"-Ebene, was es dem Modell theoretisch ermöglichte, jede Schriftsprache zu verarbeiten.

Die "Frequenz-zuerst"-Logik des BPE-Algorithmus bildete jedoch unbeabsichtigt ein implizites "Sprachsteuer"-System. Da Englisch der absolute Mainstream der Internetkorpora ist, ist die Effizienz des Modells bei der Segmentierung englischer Tokens weitaus höher als bei anderen Sprachen. Um dieselbe Bedeutung auszudrücken, benötigt Englisch die wenigsten Tokens, Chinesisch typischerweise das 1,5- bis 2-fache, und für Sprachen mit noch geringeren Ressourcen können die Kosten das 5- bis 10-fache des Englischen betragen. Dies bedeutet, dass unter der

Bezahlung-pro-Token-Regelung die tatsächlichen Kosten für ein Gespräch mit KI je nach verwendeter Sprache enorm variieren.

Diese Ungerechtigkeit an der "Startlinie" ist, einmal in den anfänglichen Wortschatz eines Modells geschrieben, nur sehr schwer zu korrigieren – die Tokenisierungsregeln sind das "Fundament" für das Verständnis der Welt durch die KI; je höher das Gebäude, desto schwieriger der Austausch des Fundaments. Glücklicherweise bessert sich diese Situation. Am Beispiel der GPT-Serie: Derselbe Satz auf Chinesisch benötigte in GPT-3 38 Tokens, in GPT-4 fiel er auf 26, in GPT-5 weiter auf 15 – eine Effizienzsteigerung von über 60%. Darüber hinaus haben chinesische heimische Großmodelle wie Tongyi Qianwen und DeepSeek von Beginn ihrer Entwicklung an hochfrequente chinesische Phrasen, Redewendungen usw. als native Tokens in ihre Vokabularien aufgenommen und so eine "muttersprachlichere" Verarbeitungseffizienz für Chinesisch erreicht. Dies offenbart ein tiefes Prinzip: Wer das "Recht auf semantische Segmentierung" besitzt, hält in weiten Teilen die Effizienz- und Kostenvorteile für den Ausdruck dieser Sprache in der digitalen Welt in der Hand – dies ist eine Form des "grundlegenden Münzrechts" im digitalen Zeitalter.

III. Verwendungen und Output: Der exponentielle Sprung von der Konversation zum Agenten

Die Szenarien des Token-Verbrauchs durchlaufen einen Sprung über drei Ebenen: von der "Konversation" über die "Argumentation" bis hin zum "Agenten".

In der Phase der konversationellen KI verbraucht eine einzelne Aufforderung und Antwort etwa 10.000 Tokens. In der Phase der argumentierenden KI, die mehrere Interaktionsrunden und komplexe Schlussfolgerungen umfasst, verbraucht jede Aufgabe etwa 100.000 Tokens. In der Phase der agentischen KI können Agenten mehrstufige Aufgaben autonom ausführen und mehrere Werkzeuge aufrufen, wobei der Token-Verbrauch pro Aufgabe etwa 1 Million erreicht – eine fast hundertfache Steigerung in nur zwei Entwicklungsgenerationen. Qualcomm-CEO Amon

prognostiziert, dass der weltweite Token-Bedarf alle 10 Sekunden bis 2030 von 31,7 Milliarden im Jahr 2026 auf 1,27 Billionen explodieren wird. Dies ist keine Übertreibung; die explosionsartige Popularität von Open-Source-Agent-Tools wie OpenClaw hat dies bewiesen: eine einzelne komplexe Aufgabe, die zig Millionen Tokens verbraucht, ist keine Seltenheit.

Auf der Output-Seite ist China zum größten Token-Anbieter der Welt geworden. Öffentliche Daten zeigen, dass das tägliche Token-Aufrufvolumen Chinas im März 2026 140 Billionen überschritten hat, was einem Anstieg um mehr als das Tausendfache in zwei Jahren entspricht und 61% des weltweiten KI-Token-Verbrauchs ausmacht. Die Etablierung dieser Position beruht auf der Überlagerung von drei Vorteilen: Erstens sind die chinesischen Industriestrompreise weitaus niedriger als in Europa und Amerika, und Strom macht über 60% der Token-Produktionskosten aus; zweitens hat China den weltweit größten intelligenten Rechencluster aufgebaut und durch die "Ost-Daten, West-Rechen"-Verteilung eine effiziente Produktion erreicht; drittens sind die heimischen chinesischen Großmodelle vollständig Open-Source, und das API-Kommerzialisierungssystem ist ausgereift und bildet eine vollständige, in sich geschlossene Industriekette von "Modelltraining – Rechenleistungsunterstützung – Inferenzausgabe – globaler Aufruf".

IV. Institutionen und Märkte: Das tripartite Industrielandschaft

Um Tokens herum hat sich eine klar geschichtete Marktstruktur gebildet.

Zum Verständnis dieser Industriestruktur ist es auch notwendig, das den Tokens zugrunde liegende Unterstützungssystem zu betrachten. Viele Menschen konzentrieren sich auf Token-Preise, übersehen jedoch, dass Tokens keine unabhängige Ressource sind. Hinter den Tokens stehen Modelle, und hinter den Modellen steht Rechenleistung. In gewissem Sinne bestimmt die Rechenleistung die Modellfähigkeit, Modelle bestimmen die Token-Produktion, und Tokens sind die Maßeinheit, die gebildet wird, wenn Modellfähigkeiten auf den Markt exportiert

werden. Daher geht es auf dem Token-Markt an der Oberfläche um Preiswettbewerb, im Kern aber immer noch um Wettbewerb in Sachen Rechenleistung und Effizienz.

Die aktuelle weltweite Token-Industriekette lässt sich grob in drei Ebenen unterteilen. Die oberste Ebene ist die Rechenleistungsinfrastruktur, einschließlich Chip-Hersteller, Server-Hersteller, Cloud-Computing-Plattformen, Rechenzentren und Energiesysteme. Die mittlere Ebene besteht aus grundlegenden Modellinstitutionen und Inferenzplattformen, die dafür verantwortlich sind, Rechenleistung in abrufbare Token-Dienste umzuwandeln. Die untere Ebene besteht aus Token-Diensteanbietern, Agentenplattformen und verschiedenen Branchen Anwendungen, die die Modellfähigkeiten weiter in Unternehmensdienste und Endprodukte umwandeln. Was der Benutzer letztendlich sieht, ist der Token-Preis, aber was die Token-Kosten und -Geschwindigkeit wirklich beeinflusst, sind oft die Veränderungen im vorgelagerten Rechenleistungssystem und der Modelleffizienz.

In den letzten Jahren, mit der kontinuierlichen Verbesserung der Chip-Leistung, der Optimierung der Modellarchitekturen und der Steigerung der Inferenzeffizienz, haben die Kosten pro Token einen rapiden Abwärtstrend gezeigt. Viele Menschen sehen, dass Tokens billiger werden, aber der eigentliche Treiber dieser Veränderung ist der kontinuierliche Fortschritt des Rechenleistungssystems. Daher wird der zukünftige Wettbewerb auf dem Token-Markt nicht nur Plattform- und Preiswettbewerb umfassen, sondern auch Wettbewerb um Rechenressourcen, Modelleffizienz und Ressourcenintegrationsfähigkeiten.

Vorgelagerte "Kraftwerke" – Rechenleistungsproduzenten. Chip-Unternehmen, angeführt von NVIDIA, verbessern kontinuierlich die Hardware, um die Token-Ausgabeffizienz pro Rechenleistungseinheit zu steigern. NVIDIA prognostiziert, dass sein Umsatz bis 2027 mindestens 1 Billion US-Dollar erreichen wird, eine Zuversicht, die vom weltweiten Hunger nach KI-Rechenleistung getragen wird.

Mittelstufige "Netzregulierung" – Modellplattformen und Aggregatoren. Unternehmen wie OpenAI, Google, DeepSeek und Alibaba wandeln die zugrunde liegende Rechenleistung in standardisierte, abrufbare Token-Dienste um. Um dieses Glied herum haben sich drei ausgereifte Geschäftsmodelle gebildet: Token-Aggregationsplattformen (vertreten durch OpenRouter, die auf Preisdifferenzen und Provisionen basieren, mit einem wöchentlichen Aufrufvolumen von 26,9 Billionen Tokens im Mai 2026), MaaS-Dienste (Model-as-a-Service) von Cloud-Anbietern und im Sekundärmarkt aktive KI-Relaisstationen.

Nachgelagerte "Stromeintrittspunkte" – Anwendungs- und Vertriebsplattformen. Die Öffnung der WeChat-Schnittstelle durch Tencent für den Zugang zu neuen Modellfähigkeiten verwandelt WeChat im Wesentlichen in eine "Token-Vertriebsplattform"; China Telecom hat vorgeschlagen, "sein Geschäft durch Token-Management neu zu gestalten" und damit versucht, sich von der traditionellen Verkehrslogik zu lösen. Diese Eintrittspunktunternehmen haben gemeinsam, dass Benutzer nicht mehr direkt mit dem Modell interagieren, sondern Tokens indirekt über die Plattform aufrufen – das Modell selbst wird "unsichtbar".

Erwähnenswert ist, dass sich um Tokens herum ein einzigartiges Paradigma des globalen Handels gebildet hat: immaterielle Ausfuhr, steuerfreier Betrieb und globale Abdeckung. Dies verwandelt China von einer traditionellen "Weltfabrik" in einen "globalen Anbieter von KI-Rechenleistungsinfrastruktur".

Bei der Diskussion der Token-Wirtschaft gibt es einen weiteren grundlegenden Faktor, der nicht umgangen werden kann: die Rechenleistung. Viele Menschen konzentrieren sich auf Token-Preise, übersehen jedoch den Produktionsprozess hinter den Tokens. Tatsächlich werden Tokens nicht aus dem Nichts erzeugt. Jede Token-Erzeugung erfordert eine Inferenz des Modells, und die Modellinferenz baut auf einer massiven Rechenleistungsbasis auf. In gewissem Sinne ist Rechenleistung die Quelle der Token-Wirtschaft, Modelle sind die Produktionswerkzeuge für Tokens, und Tokens sind die Maßeinheit, die gebildet wird, wenn die Modellfähigkeiten nach außen exportiert werden.

Daher beeinflussen die Rechenleistungsniveaus direkt die Token-Kosten und die Token-Erzeugungsgeschwindigkeit. Je stärker die Rechenleistung, desto höher die Fähigkeit des Modells, Informationen zu verarbeiten, und desto mehr Tokens können pro Zeiteinheit produziert werden; je niedriger die Kosten der Rechenleistung, desto niedriger sind in der Regel die Token-Preise. In den letzten Jahren war die Industrie der großen Modelle Zeuge eines kontinuierlichen Rückgangs der Token-Preise, der nicht nur durch die Optimierung der Modellalgorithmen, sondern auch durch Leistungsverbesserungen bei Chips, die Skalenerweiterung von Rechenzentren und die Steigerung der Inferenzeffizienz vorangetrieben wurde. Mit anderen Worten: Viele Menschen sehen, dass Tokens billiger werden, aber der eigentliche Treiber dieser Veränderung ist der kontinuierliche Fortschritt des Rechenleistungssystems.

Aus der Perspektive der Industriekette bilden Rechenleistung, Modelle und Tokens tatsächlich eine vollständige Kette. Vorgelagert sind Chips, Server, Rechenzentren und Energiesysteme; mittelgelt sind grundlegende Modelle und Inferenzplattformen; nachgelagert sind Token-Dienste, Agenten-Anwendungen und verschiedene Branchenlösungen. Rechenleistung bestimmt die Obergrenze der Modellfähigkeit; Modelle bestimmen die Qualität und Effizienz der Tokens; und Tokens bestimmen letztendlich, wie viel KI-Dienst der Benutzer erhalten kann.

Ich glaube jedoch, dass die Bedeutung der Rechenleistung zwar unbestreitbar ist, die Rechenleistung selbst jedoch nicht automatisch Werte schafft. Viele Branchen haben in der Geschichte ähnliche Prozesse durchlaufen. Ein größeres Kraftwerk zu besitzen bedeutet nicht, ein besseres Unternehmen zu besitzen; mehr Server zu besitzen bedeutet nicht, ein besseres Produkt zu besitzen. Ebenso bedeutet der Besitz von mehr Rechenleistung nicht den Besitz eines wertvolleren Systems. Rechenleistung löst das Problem der Produktionskapazität, während Systeme das Problem der Wertumwandlung lösen. In Zukunft, wenn die Rechenleistung weiter wächst, die Modellfähigkeiten weiter gestärkt werden und die Token-Ressourcen weiter zunehmen, wird das, was wirklich knapp sein wird, nicht die Rechenleistung selbst sein, sondern die Art und Weise, wie diese Ressourcen in Systeme organisiert werden, die langfristig in der realen Welt funktionieren können.

Daher glaube ich aus der Perspektive des zukünftigen Wettbewerbs, dass sich der Wettbewerb um Rechenleistung, Modelle und Tokens allmählich auf einen Systemwettbewerb verlagern wird. Denn wenn Rechenleistung allgegenwärtiger wird, Modelle sich immer ähnlicher werden und Tokens immer reichlicher vorhanden sind, wird der Unterschied nicht mehr durch die Menge der besessenen Ressourcen bestimmt, sondern dadurch, wer diese Ressourcen nutzen kann, um langfristig effektive Strukturen aufzubauen.

V. Wettbewerb und Eintrittsbarrieren: Preiskriegsinvolution und Wertschichtung

Der Wettbewerb auf dem Token-Markt weist hochgradig involutive Züge auf. Von 2022 bis 2026 sanken die Kosten pro Token-Inferenz um 99,9%. Heimische Modelle zeigen noch deutlichere Preisvorteile: Der Eingangspreis von DeepSeek V4-Pro liegt bei nur 0,025 RMB pro Million Tokens, der Ausgabepreis von Tongyi Qianwen Qwen3.5-Plus bei 1,6 RMB pro Million Tokens, verglichen mit dem Ausgabepreis von OpenAI GPT-4o von etwa 72 RMB.

Dieser Preiskrieg hat die Gewinnmargen der "Zwischenhändler" stark zusammengedrückt. Branchenstatistiken zeigen, dass die Zahl der verschiedenen API-Aggregatoren und Relay-Plattformen die Zweitausend überschritten hat, wobei sich der Markt in drei große Schichten gliedert: die oberste Schicht der großen Plattformen mit starker Technologie und umfassenden Modellen; die mittlere Schicht der professionellen Agenten mit offiziellen Rabatten und Kundenressourcen; und die Vielzahl der kleinen und kleinsten Agenten, die mit starken Überlebensdruck konfrontiert sind. Das Geschäftsmodell, einfach von Token-Preisdifferenzen zu profitieren, ist nicht mehr nachhaltig.

Die wahren Eintrittsbarrieren verlagern sich in zwei Richtungen. Die eine ist die Richtung der technologischen Prämie – durch die Entwicklung eigener Inferenzbeschleunigungs-Engines kann man billiger verkaufen als der offizielle Anbieter und trotzdem Gewinn machen. Die zweite ist die Richtung der

Mehrwertdienste – Bereitstellung maßgeschneiderter Dienste wie Prompt-Engineering, Modell-Feintuning und Private Deployment. Noch bemerkenswerter ist, dass der Wert von Tokens eine enorme szenariobasierte Ausdifferenzierung aufweist: Der Durchschnittspreis für Chat-bezogene Tokens liegt bei nur 0,01 pro Million, während der Durchschnittspreis für medikamentenentwicklungsbezogene Tokens bei 1000 pro Million, während der Durchschnittspreis für medikamentenentwicklungsbezogene Tokens bei 1000 pro Million liegen kann – ein szenarioübergreifender Wertunterschied um das Hunderttausendfache. Dies offenbart ein wesentliches Prinzip: Der Wert eines Tokens wird nicht durch seine Produktionskosten bestimmt, sondern durch seine Verwendung – derselbe Token, der in verschiedenen Szenarien verwendet und in verschiedenen Systeme eingebettet wird, generiert einen himmelweiten Unterschied an Wert.

Aber hier ist ein wichtigeres Urteil zu fällen: Derzeit und in Zukunft ist die Eintrittsbarriere für den Verkauf von Tokens nicht hoch. Jeder kann sich für eine API registrieren, eine Aggregationsplattform aufbauen und mit dem Weiterverkauf von Tokens beginnen. Was wirklich knapp ist, sind die Dinge jenseits des Token-Verkaufs – reale historische Daten, die Fähigkeit, Systeme und Architekturen zu bauen, und eine zukunftsorientierte Vision. Ohne diese sind alle großen Modelle und Tokens nutzlos.

Zusätzliches Urteil zur Rentabilität von Tokens: Mit Tokens Geld zu verdienen, ist bereits schwierig und wird immer schwieriger werden. Preiskriege drücken die Gewinnmargen kontinuierlich, und das Geschäftsmodell des reinen Token-Weiterverkaufs steuert schnell auf magere oder sogar keine Gewinne zu. Nur wenn Tokens an ein Frontend-System gebunden sind – d. h. wenn ein System Dienste bereitstellt – können Tokens und Rechenleistung als wirksame zusätzliche Vermögenswerte fungieren und Gewinne erzielen. Mit anderen Worten: Tokens selbst sind nicht die Quelle des Gewinns; das System ist es. Tokens sind nur der Treibstoff des Systems, nicht sein Motor.

Da Tokens allmählich zu einer wichtigen Maßeinheit im Zeitalter der künstlichen Intelligenz werden, sind auf dem Markt immer mehr Geschäftsmodelle entstanden, die sich um Tokens drehen. Neben den Anbietern grundlegender Modelle gibt es auch Token-Agenten, Modellaggregationsplattformen, Agentenplattformen, unternehmensbezogene API-Vertriebsdienste und verschiedene Mittelschicht-Dienstleister. Für viele Unternehmer scheint die Token-Agentur ein schneller Einstieg in die KI-Industrie zu sein, da sie weder ein Modell trainieren noch große Rechenzentren bauen muss; man muss nur vorhandene Ressourcen integrieren, um mit dem Geschäft zu beginnen. Dieses Modell hat jedoch sowohl Vorteile als auch offensichtliche Grenzen. Seine Vorteile sind relativ geringe Anfangsinvestitionen, relativ niedrige Eintrittsbarrieren und schnellere Marktexpansion; seine Grenzen sind eine starke Abhängigkeit von den vorgelagerten Modellanbietern, Anfälligkeit für Preiswettbewerb und ein Mangel an langfristigen technologischen Barrieren. Wenn die Token-Preise weiter fallen, könnte der Spielraum für Gewinne aus dem reinen Token-Weiterverkauf kontinuierlich schrumpfen. Daher ist die Token-Agentur auf lange Sicht eher eine Vertriebsfähigkeit als eine letztendliche Kernwettbewerbsfähigkeit. Was den langfristigen Wert eines Unternehmens wirklich bestimmt, bleiben die um Tokens herum aufgebauten branchenspezifischen Systeme, Automatisierungsprozesse, Wissenssysteme und realen Lösungen.

VI. Aktuelle Schwierigkeiten und zukünftige Wege

Der aktuelle Token-Markt steht vor mehreren tiefgreifenden Problemen.

Die Falle des homogenen Wettbewerbs. Über zweitausend Plattformen wetteifern um dieselbe Kundschaft, und Preiskriege drücken die Branchengewinne kontinuierlich nach unten. Ohne Veränderung könnte dies die Anfänge des Cloud-Computing wiederholen, wo man "Geld verlor, um Marktanteile zu gewinnen".

Das Risiko der Nichteinhaltung von Vorschriften. Die derzeitige globale "regulatorische Dividende" des Fehlens spezifischer Gesetze ist ein vorübergehender Zustand, keine langfristige Norm. Mit der Ausweitung des Token-

Marktes werden Fragen wie grenzüberschreitende Daten, nationale Sicherheit und Steuererhebung die Länder dazu zwingen, die Gesetzgebung zu beschleunigen. Besonders zu beachten ist, dass Versuche einiger Marktteilnehmer, Tokens zu tokenisieren und zu finanzialisieren, leicht globale Finanzregulierungen auslösen können, was nicht nur zur Schließung dieser Geschäfte führen, sondern auch den Ruf der gesamten chinesischen Token-Exportindustrie schädigen würde.

Der "Token-maxxing"-Mythos. Im Silicon Valley hat sich eine neue Form des "Reichtumsprotzes" herausgebildet – der Wettbewerb darum, wer täglich mehr Tokens verbraucht. Meta hat eine interne Bestenliste namens "Claudeconomics", die KI-Nutzungsdaten von über 85.000 Mitarbeitern sammelt. Der Spitzenreiter verbrennt Token im Wert von mehreren Millionen Dollar. Kritiker merken bissig an: "Outcome maxxung ist besser als Token maxxung" – anstatt wild Tokens zu verbrauchen, sollte man lieber auf die erzielten Ergebnisse schauen. Genau das ist das Kernproblem, das ich durch langjährige Praxis tief verstanden habe.

Chinesische Tokens vs. englische Tokens: Drei Barrieren unter einem einheitlichen Konzept

Für chinesische Benutzer ist eine Tatsache, die klar erkannt werden muss, diese: Obwohl chinesische Tokens (ciyuan) und englische Tokens derselben grundlegenden Maßeinheit des KI-Zeitalters angehören, haben sie keinen wirklich einheitlichen, frei interoperablen globalen Markt gebildet. Aus technischer Konzeptionssicht verwenden Ost und West Tokens als Basiseinheit für die Verarbeitung von Sprache und Daten durch Modelle; aus praktischer Anwendungssicht jedoch hat das kombinierte Vorhandensein technologischer Barrieren, systemischer Barrieren und politischer Barrieren eine faktische Trennung zwischen chinesischen Tokens und englischen Tokens geschaffen.

Die erste Barriere ist die technologische Barriere, d. h. Unterschiede in der Tokenisierungslogik und Modellarchitektur.

Westliche Mainstream-Modelle wie GPT, Claude und Gemini werden größtenteils auf großen Korpora aufgebaut, die hauptsächlich aus Englisch bestehen. Obwohl diese Modelle bereits über ausgezeichnete Chinesisch-Verarbeitungsfähigkeiten verfügen, benötigt Chinesisch in vielen Szenarien immer noch mehr Tokens, um die gleiche Informationsmenge wie Englisch auszudrücken. Noch wichtiger ist, dass verschiedene Modelle unterschiedliche Vokabularstrukturen und Tokenisierungsstrategien haben. Derselbe chinesische Text kann in verschiedenen Modellen in völlig unterschiedliche Anzahlen von Tokens segmentiert werden, und ihre internen Kodierungsmethoden und semantischen Zuordnungsmethoden können ebenfalls unterschiedlich sein.

Gleichzeitig haben chinesische heimische Modelle wie DeepSeek, Tongyi Qianwen, Doubao und Wenxin Yiyan von der Trainingsphase an große Mengen chinesischer Korpora aufgenommen, was ihnen eine höhere native Anpassungsfähigkeit an chinesische Phrasen, Redewendungen, Eigennamen und lokalisierte Ausdrücke verleiht. Daher können selbst bei demselben Text die Anzahl der Tokens, die Verarbeitungseffizienz und die Verständnispfade zwischen verschiedenen Modellen erheblich variieren. In diesem Sinne gibt es zwar ein einheitliches Token-Konzept, aber kein "Standard-Token", das direkt zwischen Modellen zirkuliert und wiederverwendet werden könnte.

Die zweite Barriere ist die systemische Barriere, d. h. Unterschiede in den Ökosystemen und Aufrufmethoden.

Der aktuelle globale Token-Markt hat tatsächlich mehrere relativ unabhängige Ökosysteme gebildet. Internationale Plattformen wie OpenAI, Anthropic und Google haben ihre eigenen API-Standards, Preismodelle, Zugriffsregeln und Dienstleistungssysteme; der chinesische Markt hat ein heimisches Ökosystem gebildet, das sich auf Plattformen wie DeepSeek, Tongyi Qianwen, Doubao und Kimi konzentriert.

Für normale Benutzer bedeutet der Kauf von Tokens auf einer Plattform nicht, dass sie sie direkt auf einer anderen Plattform verwenden können. Die

Aufrufsschnittstellen, Authentifizierungsmechanismen, Ratenbegrenzungen, Dienstvereinbarungen und Geschäftsregeln unterscheiden sich von Plattform zu Plattform. Selbst wenn eine Konvertierung über Drittanbieter-Gateways oder Aggregationsplattformen erfolgt, wird im Wesentlichen ein anderes System aufgerufen, kein echter Token-Austausch. Daher ist der Token zwar ein einheitliches Konzept, aber Tokens als kommerzielle Ressource und Dienstleistungsfähigkeit bleiben in ihren jeweiligen Plattform-Ökosystemen eingeschlossen.

Die dritte Barriere ist die politische Barriere, d. h. Unterschiede in den Datenflüssen und Compliance-Anforderungen.

Da künstliche Intelligenz allmählich in die Phase der Unternehmensanwendung eintritt, nimmt die Bedeutung von Datensicherheitsfragen weiter zu. Chinas Datensicherheitssystem, die EU-DSGVO und die US-Regulierungsmaßnahmen für den grenzüberschreitenden Fluss sensibler Daten stellen alle unterschiedliche Anforderungen an die Datenübertragung und -verarbeitung. Viele Daten, die Kundeninformationen, Finanzdaten, medizinische Unterlagen, Geschäftsgeheimnisse und personenbezogene Daten betreffen, können oft nicht frei Grenzen überschreiten.

Dies bedeutet, dass in realen Geschäftsumgebungen viele Unternehmen Modelle nicht allein aufgrund von Token-Preisen auswählen, sondern die Datenkonformität, den Datenspeicherort und die rechtliche Verantwortung priorisieren müssen. Für viele wirklich geschäftskritische Daten sind die Compliance-Kosten oft weitaus höher als die Token-Kosten selbst. Daher neigen in der realen Welt inländische Unternehmen dazu, konforme inländische Modelle zu verwenden, während ausländische Unternehmen dazu neigen, ausländische Modelle zu verwenden, was zu einer faktischen Segmentierung des Token-Marktes führt.

Technologische Barrieren bestimmen die Unterschiede in den Tokenisierungsmethoden; systemische Barrieren bestimmen die Unterschiede in den Aufrufsystemen und Geschäftsökosystemen; politische Barrieren bestimmen die Unterschiede in den Datenflüssen und Compliance-Grenzen. Diese drei gemeinsam

wirkenden Barrieren führen dazu, dass chinesische Tokens und englische Tokens, obwohl sie demselben technischen Konzept angehören, nicht wie eine idealisierte "globale universelle KI-Währung" frei getauscht, frei zirkuliert und frei verbunden werden können.

In Zukunft wird wahrscheinlich eine Situation ähnlich dem internationalen Handel entstehen: Das chinesische Token-Ökosystem und das englische Token-Ökosystem werden lange Zeit nebeneinander bestehen, jedes mit seinen eigenen unabhängigen Entwicklungspfaden. Ein zoneübergreifender Aufruf ist nicht unmöglich, erfordert jedoch oft zusätzliche technische Anpassungen, Systemkonvertierungen und rechtliche Prüfungskosten. Für die meisten normalen Benutzer ist diese zoneübergreifende Fähigkeit kein alltägliches Bedürfnis, daher werden die beiden großen Systeme wahrscheinlich lange Zeit relativ unabhängig bleiben.

In China produzierte Tokens sind nicht unbedingt auf die Verwendung innerhalb chinesischer Systeme beschränkt, aber in Schlüsselszenarien, die Datenhoheit, Compliance-Verantwortung und Industriesicherheit betreffen, wird ein Großteil der zukünftigen Token-Nachfrage wahrscheinlich vorzugsweise innerhalb inländischer Modelle, Plattformen und Systeme konsumiert werden.

Aus langfristiger Trendperspektive nimmt die Kopplung zwischen dem Standort der Rechenleistungsproduktion, dem Modellanbieter, dem Systembetreiber und dem Datenstandort zu. Für Finanzen, Regierung, Gesundheitswesen, Lieferketten und große Unternehmenssysteme können Daten oft nicht frei über Grenzen fließen. Daher wird ein Großteil der hochwertigen Token-Nachfrage, selbst wenn China große Mengen an Token-Diensten weltweit exportieren kann, wahrscheinlich vorzugsweise innerhalb chinesischer Modelle und Systeme verbleiben. Dies bedeutet nicht, dass chinesische Tokens nicht exportiert werden können, sondern dass der zukünftige Token-Markt gleichzeitig zwei Strukturen aufweisen könnte: einen "globalen Zirkulationsteil" und einen "heimischen Closed-Loop-Teil".

Erwähnenswert ist, dass das, was in Zukunft wirklich Einheitlichkeit erreichen könnte, nicht die Tokens selbst sind, sondern die Art und Weise, wie Tokens

aufgerufen werden. Mit der kontinuierlichen Entwicklung von Agentenplattformen, einheitlichen Schnittstellenprotokollen, Modellaggregationsplattformen und Automatisierungssystemen werden sich die Benutzer letztendlich nicht darauf konzentrieren, welcher Token verwendet wurde, sondern ob die Aufgabe erledigt wurde, ob die Kosten angemessen sind und ob die Effizienz ausreichend ist. Zu diesem Zeitpunkt werden Tokens, wie Strom, Internetverkehr und Cloud-Computing-Ressourcen heute, allmählich hinter den Systemen verschwinden und zur grundlegenden Betriebsinfrastruktur der digitalen Zivilisation werden.

Daher, aus langfristiger Perspektive, sind die ost- und westlichen Token-Märkte zwar vorübergehend schwierig direkt zu verbinden, aber ihre Entwicklungsrichtungen sind hochgradig deckungsgleich. Die Vergangenheit der Tokens gehört der Computerlinguistik, die Gegenwart der Tokens gehört der KI-Industrie, und die Zukunft der Tokens wird wahrscheinlich zum grundlegenden Betriebssystem der gesamten digitalen Zivilisation gehören.

VII. Warum die Fähigkeit zum Systemaufbau nicht direkt durch große Mengen von Tokens erworben werden kann

Bei der Argumentation, dass "die Fähigkeit zum Systemaufbau ein knappes Gut ist", muss ich mich einem möglichen Einwand stellen: Da KI so leistungsfähig ist, warum kann die Fähigkeit zum Systemaufbau dann nicht durch den Verbrauch großer Mengen von Tokens "erzeugt" oder "erworben" werden? Die Antwort auf diese Frage offenbart die grundlegenden Grenzen der Token-Fähigkeit.

Erstens, Systemaufbau erfordert Entscheidungen, und Entscheidungen sind keine Generierung.

KI kann einen Artikel, einen Code, eine Website-Vorlage generieren, aber KI kann nicht die zentralen Entscheidungen im Systemdesign ersetzen, die Menschen treffen müssen – welche Module auswählen, welche Funktionen aufgeben, wie man aktuelle Bedürfnisse mit zukünftiger Erweiterung abwägt, wann umstrukturiert, wann

der Status quo beibehalten wird. Diese Entscheidungen hängen vom Verständnis des Wesens des Geschäfts, der Beurteilung von Ressourcenbeschränkungen und der Voraussicht langfristiger Konsequenzen ab. Tokens können unzählige "mögliche Lösungen" generieren, aber die Entscheidung für "welche Lösung umsetzenswert ist" und "warum" können Tokens nicht ersetzen. Mein intelligentes Logistiksystem hat im Laufe von zwei Jahrzehnten unzählige Anpassungen erfahren, und jede Anpassung erfolgte nicht, weil mir eine KI sagte "das ist besser", sondern weil ein echtes Problem im tatsächlichen Geschäft auftrat, das Grenzen aufzeigte und neue Anforderungen stellte. Hinter diesen Entscheidungen steht Urteilsvermögen, und Urteilsvermögen kommt aus der Praxis, nicht aus Tokens.

Zweitens, Systemaufbau erfordert Ausnahmebehandlung, und Ausnahmen können nicht erschöpfend aufgelistet werden.

Ein System kann nicht deshalb langfristig funktionieren, weil sein normaler Ablauf perfekt gestaltet ist, sondern weil es Ausnahmen behandeln kann. In einem Logistiksystem umfassen Ausnahmen: falsche Kundenadressen, beschädigte Waren, Zollverzögerungen, Zahlungsausfälle, Systemabstürze, Datenverlust. In einem Publikationssystem umfassen Ausnahmen: Schwankungen in der Übersetzungsqualität, inkompatible Metadatenformate, Änderungen von Archivschnittstellen, DOI-Registrierungsfehler. Diese Ausnahmen können nicht erschöpfend aufgelistet werden, weil die Vielfalt der Ausnahmen in der realen Welt unendlich ist. KI kann durch große Mengen von Tokens "häufige Ausnahmemuster" lernen, aber sie kann nicht "die nächste noch nie dagewesene Ausnahme" vorhersagen. Nur ein langlaufendes System, das in der Praxis kontinuierlich auf Ausnahmen stößt, sie aufzeichnet und löst, kann eine echte Ausnahmebehandlungsfähigkeit entwickeln. Diese Fähigkeit kann nicht durch "Füttern mit mehr Tokens" erlangt werden, denn Ausnahmen sind kein Wissen, sondern Erfahrung.

Drittens, Systemaufbau erfordert Kompromisse, und Kompromisse haben keine Standardantworten.

Die Konstruktion jedes Systems beinhaltet Kompromisse unter multiplen Beschränkungen: Kosten vs. Leistung, Geschwindigkeit vs. Stabilität, Allgemeingültigkeit vs. Anpassung, aktuelle Bedürfnisse vs. zukünftige Erweiterung. Diese Kompromisse haben keine "richtige Antwort", sondern nur "unter bestimmten Umständen angemessene Entscheidungen". KI kann eine "optimale Lösung" liefern, aber dieses "Optimum" ist ein statistisches Optimum basierend auf historischen Daten, kein Urteil basierend auf "Ihrem spezifischen Kontext". Der Grund, warum mein generationsübergreifendes, zwei Jahrzehnte altes integriertes Logistik- und Finanzsystem bis heute funktioniert, ist genau, dass ich in verschiedenen Epochen unterschiedliche Kompromisse eingegangen bin – Wahl eines bestimmten Technologie-Stacks im Jahr 2005, Wahl einer anderen Integrationsmethode im Jahr 2015, Wahl eines Upgrade-Pfads im Jahr 2025. Diese Kompromisse waren nicht "richtig" oder "falsch", sondern "dem damaligen Kontext angemessen". Tokens können diese Kompromisse nicht für mich eingehen, denn das Wesen von Kompromissen ist Werturteil, nicht Berechnung.

Viertens, Systemaufbau erfordert Feedbackschleifen, und Feedback erfordert die reale Welt.

Die Verbesserung der Systembaufähigkeit hängt von einer geschlossenen Schleife ab: Design → Implementierung → Betrieb → Problemerkennung → Korrektur → Neudesign. Der Schritt "Problemerkennung" in dieser Schleife muss aus Rückmeldungen der realen Welt kommen. KI kann den Betrieb simulieren, aber sie kann nicht alle unvorhersehbaren Faktoren der realen Welt simulieren. Der Grund, warum mein Logistiksystem sich ständig verbessert hat, ist, dass echte Kunden echtes Feedback gegeben haben; der Grund, warum mein Publikationssystem sich ständig verbessert hat, ist, dass echte Archivschnittstellen echte Änderungen erfahren haben. Dieses Feedback kann nicht durch den Verbrauch von Tokens in einer geschlossenen Umgebung gewonnen werden. Tokens können "hypothetisches Kundenfeedback" generieren, aber sie können die Tatsache ersetzen, dass "der Kunde dieses Feedback tatsächlich gegeben hat". Rückmeldungen aus der realen Welt sind der einzige Boden für das Wachstum der Systembaufähigkeit.

Fünftens, Systemaufbau erfordert Vertrauen, und Vertrauen kann nur aus Zeit kommen.

Ob ein System vertrauenswürdig ist, beweist sich nicht durch Werbung, sondern durch die Zeit. Der Grund, warum meine Website dauerhaft von CERN Zenodo, WorldCat und der National Library of Australia archiviert wird, ist nicht, dass sie mich "mögen", sondern dass diese Institutionen durch langfristige Beobachtung bestätigt haben, dass mein System stabil, zuverlässig und beständig ist. Dieses Vertrauen kann nicht mit großen Mengen von Tokens erkaufte werden. KI kann ein "sehr professionell aussehendes" Systemhandbuch generieren, aber sie kann nicht die Tatsache generieren, dass "diesem System seit zwei Jahrzehnten vertraut wird". Vertrauen ist ein Produkt der Zeit, kein Produkt von Tokens.

Zusammenfassend lässt sich sagen, dass die Fähigkeit zum Systemaufbau deshalb nicht direkt durch große Mengen von Tokens erworben werden kann, weil es sich im Wesentlichen nicht um eine generative Fähigkeit handelt, sondern um eine Fähigkeit zu Urteilen, Ausnahmebehandlung, Kompromiss, Feedbackabsorption und Vertrauensaufbau. Diese Fähigkeiten haben gemeinsam, dass sie reale Weltzeit, Praxis, Fehler, Korrekturen und Validierung benötigen. Tokens können die Informationsverarbeitung beschleunigen, aber nicht die reale Erfahrungsakkumulation. Dies ist der grundlegende Grund, warum die Systembaufähigkeit auch dann noch knapp sein wird, wenn Tokens in Zukunft extrem billig werden.

VIII. Meine Vermögenswerte: Reale historische Daten, domänenübergreifende Systeme, Aufbaufähigkeit, Zukunftsbild und Selbstverifizierungsfähigkeit

Bevor ich weiter erörtere, "was wirklich knapp ist", muss ich zuerst meine eigene Position erläutern. Denn meine Urteile leiten sich nicht aus theoretischer Deduktion ab, sondern aus langjähriger Praxis. Dies zeigt sich sehr deutlich in den

verschiedenen Systemen, die ich über einen langen Zeitraum aufgebaut habe. Ob es sich um das intelligente Logistiksystem, das intelligente Rechnungssystem, das automatisierte Finanzverarbeitungssystem oder später das Zehn-Sprachen-Publikationssystem, das Metadaten-System, das Website-Navigationssystem und das automatische Kreuzreferenzierungssystem handelt, der wirklich schwierige Teil war nie die Generierung von Informationen, sondern deren Organisation. Informationen können ins Unendliche wachsen, aber ohne Struktur schaffen sie nur neues Chaos. Im Token-Zeitalter gilt dasselbe. In Zukunft kann KI riesige Mengen an Inhalten generieren, aber riesige Mengen an Inhalten bedeuten nicht automatisch riesige Mengen an Wert. Im Gegenteil, je einfacher die Inhaltsgenerierung wird, desto wichtiger wird es, diese Inhalte zu filtern, zu organisieren, zu bewahren, zu indizieren und zu verknüpfen.

Ich besitze die folgenden fünf Vermögenswerte, die die empirische Grundlage für mein Urteil darüber bilden, "was nach Tokens knapp ist".

Erstens, reale historische Daten. Dies sind nicht irgendwelche "historischen Daten", sondern zwei Jahrzehnte generationsübergreifender realer Geschäftsdaten – vom intelligenten Rechnungssystem 2005 bis zur Gegenwart, aus tatsächlich laufenden Logistik-, Finanz- und Steuersystemen, nicht aus Versuchsumgebungen oder gescrapten Daten. Im KI-Zeitalter, mit der Flut von "synthetischen Daten", steigt der Wert realer historischer Daten rasant an. Die meisten Unternehmen besitzen "einzeldomänige historische Daten", während ich domänenübergreifende, zwei Jahrzehnte umspannende, praktisch validierte historische Daten besitze.

Zweitens, die Fertigstellung mehrerer domänenübergreifender Systeme und Architekturen. Meine Systeme decken die folgenden Bereiche ab: intelligentes Logistiksystem, intelligentes Rechnungssystem, intelligentes Finanzsystem, intelligentes Steuersystem, generationsübergreifendes zwei Jahrzehnte altes integriertes Logistik- und Finanzsystem, Website-Automatisierungssystem, Metadatengenerierungssystem, Zehn-Sprachen-Publikationssystem von "The Epochal Transition", DOI-Managementsystem, TROVE- und WorldCat-Link-Organisationssystem, automatisches Kreuzreferenzierungssystem, Sitemap-

Generierungssystem, Zenodo-Profil-Generierungssystem, plattformübergreifendes Publikationssystem. Die Einzigartigkeit dieser Systeme liegt in: der domänenübergreifenden Breite (von Logistik bis Publishing, von Finanzen bis Literaturindexierung), der zeitlichen Tiefe (einige Systeme laufen seit zwei Jahrzehnten) und der Vollendung durch eine Person (was die Übertragbarkeit der Methode beweist). Unter diesen dient die elektronische Monatszeitschrift "The Epochal Transition" (<https://times.net.au>) als öffentlicher Beweis – sie besitzt die internationale Standardnummer ISSN (3083-5178), wird dauerhaft archiviert von WorldCat (dem weltweit größten Bibliotheksverbundkatalog, mit Mitgliedsinstitutionen von über 16.000 nationalen und Universitätsbibliotheken in über 100 Ländern und Regionen), der PANDORA/TROVE der National Library of Australia sowie von CERN (Europäische Organisation für Kernforschung) betriebenem Zenodo, OpenAIRE (Europäische Open-Science-Infrastruktur), DataCite DOI (globaler digitaler Objektidentifikator, mit über 60 Dokumenten) und ORCID (global eindeutiger permanenter akademischer digitaler Identifikator). Diese Website beweist, dass eine einzelne Person (ein unabhängiger Gelehrter) den gesamten Prozess von der Inhaltserstellung bis zur Schnittstelle mit führenden internationalen Archivinfrastrukturen abschließen kann.

Drittens, die Fähigkeit, Systeme und Architekturen zu bauen. Dies ist nicht die Fähigkeit, "KI-Werkzeuge zu benutzen", sondern die Designfähigkeit, vage Anforderungen in stabile Strukturen zu überführen, die Architekturfähigkeit, verschiedene Systeme in eine einheitliche Architektur zu integrieren, die Ingenieursfähigkeit, Systeme über Jahre ohne menschliches Eingreifen am Laufen zu halten, und die Meta-Fähigkeit, Erfahrungen von der Logistik auf das Publizieren, von Finanzen auf die Literatur zu übertragen. Wie bereits erörtert, kann diese Fähigkeit nicht direkt durch große Mengen von Tokens erworben werden, weil sie von Entscheidungsfindung, Ausnahmebehandlung, Kompromissen, Feedbackschleifen und Vertrauensaufbau abhängt – nichts davon kann durch KI ersetzt werden. Der derzeitige KI-Trainingsmarkt konzentriert sich auf oberflächliche Fähigkeiten wie "Prompt-Engineering", "KI-Schreiben" und "KI-Zeichnen"; fast kein Kurs vermittelt die "domänenübergreifende Systembaufähigkeit" – denn diese

Fähigkeit ist nicht abkürzbar, sondern kann nur durch langjährige Praxis angehäuft werden. Die auf der Website aufgeführten über 40 unabhängigen kreativen Werke, die erworbene ISSN und die über 60 DOIs sind die besten öffentlichen Aufzeichnungen der langjährigen Betriebsfähigkeit des Systems, nicht meine eigene Selbstbehauptung.

Viertens, eine zukunftsorientierte Vision. Beginn der Experimente mit einem Fernarbeits-Prototypsystem im Jahr 1993 – fast dreißig Jahre bevor großflächige Fernarbeit üblich wurde; Entwurf eines intelligenten Rechnungssystems im Jahr 2005 – Beginn der Prozessautomatisierung zwanzig Jahre vor der Verbreitung von KI; generationsübergreifende zwei Jahrzehnte dauernde Systemintegration – Voraussicht zum Zeitpunkt des Systementwurfs, dass es auch nach zwanzig Jahren noch laufen müsste; Aufstellung des Urteils "nach Tokens" jetzt – zu einer Zeit, in der die meisten Menschen noch Tokens hinterherjagen, Voraussicht, dass die Systemfähigkeit die wahre Hürde ist. Ich habe auch vorausgesehen, dass "Tokens Währungen ähneln und die technologischen, systemischen und politischen Unterschiede zwischen Ost und West sie nicht universell einsetzbar machen" – so wie der US-Dollar und der chinesische Renminbi nicht frei getauscht werden können, stehen chinesische Tokens und englische Tokens vor ähnlichen Zirkulationsbarrieren, und in Zukunft könnten "chinesische Token-Zonen", "englische Token-Zonen" und andere auf Sprache und Politik basierende Token-Ökosysteme entstehen. Ich habe auch vorausgesehen, dass "es bereits schwierig ist und immer schwieriger werden wird, mit Tokens Geld zu verdienen; nur wenn es ein Frontend-System gibt, das Dienste bereitstellt, können Tokens und Rechenleistung als wirksame zusätzliche Vermögenswerte zur Erzielung von Gewinnen dienen". Als unabhängiger Gelehrter habe ich im Voraus ein persönliches Publikationssystem aufgebaut, das sich an zukünftige akademische Infrastrukturen ankoppelt, und damit den Trend vorausgesehen, dass "ein Individuum auch ein Knoten in der globalen Wissensinfrastruktur werden kann". Die meisten Trendvorhersagen stammen von Analysten oder Unternehmern, während meine "Voraussichtsfähigkeit" die Bestätigung durch zwei Jahrzehnte Praxis als Rückendeckung hat.

Fünftens, Selbstverifizierungsfähigkeit. Dies ist eine meiner wichtigsten Meta-Fähigkeiten – alle meine schriftlichen Arbeiten in verschiedenen Bereichen erreichen einen theoretischen Abschluss und logische Selbstkonsistenz. Das bedeutet, dass meine Argumentation nicht auf externer Autorität oder Zitaten beruht, sondern auf der Selbstgenügsamkeit der inneren Logik. Die Standardregel in der akademischen Literatur ist es, "frühere Forschungen zu zitieren"; nur wenige Menschen behaupten oder demonstrieren eine "Selbstverifizierungsfähigkeit"; die akademische Diskussion konzentriert sich darauf, "ob die Argumentation ausreichend ist", nicht darauf, "ob der Autor die Fähigkeit hat, eine selbstkonsistente geschlossene Schleife zu konstruieren". In meinen Artikeln zeigt sich diese Selbstverifizierungsfähigkeit darin, dass Argumente aus der Praxis kommen, die Praxis die Argumente validiert und Argumente und Praxis eine geschlossene Schleife bilden. Ich "zitieren nicht die Theorien anderer, um mich selbst zu beweisen", sondern "zeige die von mir gebauten Systeme und lasse die Systeme für sich selbst sprechen". Diese Selbstverifizierungsfähigkeit ist ein in der öffentlichen Literatur völlig nicht vorhandenes Konzept und eine meiner einzigartigsten Meta-Fähigkeiten.

Die Kombination dieser fünf Vermögenswerte, ihre Knappheit ist nicht additiv, sondern multiplikativ: reale historische Daten × domänenübergreifende Systeme × Systembaufähigkeit × zukunftsorientierte Vision × Selbstverifizierungsfähigkeit bilden ein nicht reproduzierbares Vermögensportfolio. In öffentlichen Diskussionen ist derzeit kein vergleichbarer Fall zu finden.

Der Prozess der Systematisierung von "Der Weg eines unabhängigen internationalen Gelehrten" hat mir dies noch tiefer bewusst gemacht. Von 1989 bis 2026 umspannt mehr als dreißig Jahre Erfahrung mehrere Bereiche. Diese Erfahrungen entstanden nicht alle auf einmal, sondern bildeten sich allmählich über einen langen Zeitraum heraus. KI kann einen Artikel in Sekundenschnelle generieren, aber sie kann nicht in Sekundenschnelle mehr als dreißig Jahre echte Erfahrung generieren; sie kann viele Meinungen generieren, aber nicht direkt das durch langjährige Praxis angesammelte Urteilsvermögen; sie kann viele Prozesse simulieren, aber nicht die wiederholte Validierung der realen Welt ersetzen. In

diesem Sinne ist die Zeit selbst ein wichtiger Vermögenswert, und die durch die Zeit gefällte Struktur ist ein noch seltenerer Vermögenswert.

IX. Nach den Tokens: Fünf wirklich knappe Fähigkeiten

Wenn Tokens selbst nicht mehr knapp sind – die Eintrittsbarriere für den Verkauf von Tokens ist nicht hoch, aber es wird einen extremen Mangel an realen historischen Daten, der Fähigkeit, Systeme und Architekturen zu bauen, und einer zukunftsorientierten Vision geben – was wird dann wirklich knapp sein? Basierend auf meiner langjährigen Praxis, vom Fernarbeits-Prototypsystem 1993, dem intelligenten Rechnungssystem 2005, dem intelligenten Logistiksystem, dem intelligenten Finanzsystem, dem intelligenten Steuersystem, dem generationsübergreifenden zwei Jahrzehnte alten integrierten Logistik- und Finanzsystem, bis hin zur Website-Automatisierung, Metadatengenerierung, der Zehn-Sprachen-Monatszeitschrift "The Epochal Transition", DOI-Management, TROVE- und WorldCat-Link-Organisation, automatischer Kreuzreferenzierung, Sitemap-Generierung, Zenodo-Profil-Generierung, plattformübergreifendem Publizieren, glaube ich, dass die folgenden fünf Fähigkeiten zu echten Hürden werden.

Erstens, strukturelle Designfähigkeit. Tokens sind Rohmaterial, und Struktur ist der Rahmen, der das Rohmaterial in wertvolle Produkte organisiert. Schon lange vor der Verbreitung von KI begann ich darüber nachzudenken, wie ich mit begrenzten Ressourcen eine höhere Betriebseffizienz erreichen kann. Mein intelligentes Logistiksystem, mein intelligentes Rechnungssystem und mein generationsübergreifendes zwei Jahrzehnte altes intelligentes Logistik- und Finanzsystem entstanden in verschiedenen Epochen, standen vor verschiedenen Problemen, aber sie haben eine gemeinsame Eigenschaft: Das Ziel war nicht, mehr Informationen zu schaffen, sondern Chaos zu reduzieren, sich wiederholende Arbeit zu reduzieren, die Gesamteffizienz zu verbessern und Strukturen zu schaffen, die langfristig funktionieren können. Im automatisierten Publikationssystem war das zentrale Problem, dem ich gegenüberstand, nicht "ob Text generiert werden kann",

sondern "wie man Metadatenfelder für zehn Sprachen definiert, wie man Beziehungen zwischen Artikeln und Kategorien herstellt, wie man Versionsverwaltungsregeln entwirft". Diese strukturellen Designentscheidungen bestimmen, ob das System über Jahre hinweg ohne Zusammenbruch weiterlaufen kann. Die Menge der verbrauchten Tokens ist dagegen nur ein Nebenprodukt. Die Website "The Epochal Transition" ist selbst ein öffentlicher Beweis für strukturelle Designfähigkeit: eine persönlich gepflegte akademische Website-Struktur mit einer internationalen Standard-ISSN, mehrsprachiger Unterstützung, mehreren Archivkanälen, die beweist, dass ich keine leeren Worte mache, sondern mit dieser strukturellen Designfähigkeit erfolgreich ein komplexes, öffentlich zugängliches System aufgebaut habe.

Zweitens, Automatisierungsfähigkeit. Tokens können Inhalte in Stapeln produzieren, aber nur Automatisierungssysteme können End-to-End-Workflows in Stapeln abschließen. Im intelligenten Logistiksystem liegt der wahre Wert nicht darin, wie oft KI zur Routenplanung aufgerufen wird, sondern darin, eine vollautomatische geschlossene Schleife von der Auftragsannahme über die Lagerdisposition und Routenoptimierung bis zur Ausnahmebehandlung einzurichten. In dieser geschlossenen Schleife ist KI nur eine Komponente, und die Automatisierungsfähigkeit zeigt sich in den Verbindungen zwischen den Komponenten. Ein Mensch ohne System wird selbst mit einer großen Menge an Tokens möglicherweise nur vereinzelter Text, vereinzelter Bilder, vereinzelter Tabellen und vereinzelter Code generieren; ein Mensch mit einem System hingegen kann Tokens in seinen Arbeitsablauf einbinden und sie automatisch Material organisieren, Metadaten generieren, Links überprüfen, Webseiten aktualisieren, beim Publizieren helfen, Literatur archivieren, Indizes erstellen, Logistik abwickeln, Finanzen analysieren und Berichte erstellen lassen.

Drittens, domänenübergreifende praktische Erfahrung. Tokens stehen jedem zur Verfügung, aber die Fähigkeit, Erfahrungen von einer Branche auf eine andere zu übertragen, ist nicht universell. Der Grund, warum ich sowohl ein Logistik- als auch ein Publikationssystem aufbauen kann, ist nicht, dass ich mehr Tokens habe,

sondern dass ich übertragbare Methoden wie Metadatenmanagement, Workflow-Design und Ausnahmebehandlung angesammelt habe. Mein Publikationssystem benötigt Tokens, weil die Übersetzung in zehn Sprachen, die Generierung von Zusammenfassungen, die Organisation von Schlüsselwörtern, die Metadatengenerierung und die Linkbeschreibungen alle eine umfangreiche Sprachverarbeitung erfordern. Mein Website-System benötigt Tokens, weil Seitenstrukturen, Navigationsbeschreibungen, SEO-Beschreibungen, akademische Metadaten und mehrsprachige Einführungen automatisch generiert werden können. Mein Logistiksystem kann auch Tokens verwenden, weil Kunden-E-Mails, Warenbeschreibungen, Zolldokumente, Ausnahmeerklärungen, Finanznotizen und Betriebsprotokolle KI-unterstützt verarbeitet werden können. Mein automatisches Kreuzreferenzierungssystem benötigt Tokens sogar noch mehr, weil es thematische Beziehungen zwischen einer großen Anzahl von Artikeln, Titeln, DOIs, NLA-, TROVE- und WorldCat-Links finden muss. Diese domänenübergreifende praktische Akkumulation kann nicht durch den Verbrauch von Tokens erworben werden; sie kann nur durch die Lösung realer Probleme sedimentiert werden. Die Website präsentiert gleichzeitig Informationen aus mehreren Bereichen, darunter akademisches Publizieren, Infrastrukturintegration, technische Architektur und persönliche Marke, was beweist, dass ich tatsächlich das Publizieren, die Bibliothekswissenschaft, die Technologie und die persönliche Marke verbunden habe, anstatt eine einzelne Fähigkeit zu besitzen.

Viertens, das mentale Modell für den Langzeitbetrieb von Systemen. Das vorherrschende Narrativ für aktuelle KI-Anwendungen ist "Konversation" und "Generierung", aber was wirklich anhaltenden Wert schafft, ist der "Betrieb". Eine Konversation generiert 10.000 Tokens, eine Agentenaufgabe verbraucht eine Million Tokens, aber ein Logistiksystem, das ein Jahrzehnt lang ununterbrochen läuft, kann Milliarden von Tokens verbrauchen – aber was wirklich knapp ist, ist das mentale Modell für den Entwurf dieses Systems: Wie stellt man sicher, dass das System nach Modellupgrades noch nutzbar ist? Wie geht man mit Änderungen an vorgelagerten APIs um? Wie überwacht man Ausnahmen und stellt automatisch wieder her? Diese Fragen haben nichts mit Tokens zu tun, aber sie bestimmen den

endgültigen Output der Tokens. Der Wert dieser Systeme kommt nicht von den Tokens selbst, sondern von den Strukturen, die ich bereits aufgebaut habe oder entwickle. Tokens sind nur der Treibstoff; das System ist der Motor. Tokens sind nur der elektrische Strom; die Struktur ist der Stromkreis. Tokens sind nur das Wasser; die Rohre, Ventile, Reservoirs und Dispositionssysteme bestimmen, ob es seine Rolle spielen kann. Ohne ein System werden Tokens vergehen; mit einem System können Tokens in langfristigen Wert umgewandelt werden. Die Website zeigt über 40 unabhängige kreative Werke und die dauerhafte Archivierung durch angesehene Institutionen wie CERN und die National Library of Australia, was beweist, dass mein System nicht kurzlebigen Hypes hinterherjagt, sondern langfristigen Wert anstrebt und bereits die "dauerhafte Aufbewahrung" durch angesehene Institutionen erhalten hat.

Wenn ich jedoch die Beziehung zwischen Tokens und Systemen weiter betrachte, finde ich ein weiteres oft übersehenes Problem. Viele Menschen glauben, dass sie mit einem hervorragenden System im zukünftigen Token-Zeitalter im Vorteil sein werden. Aber die Realität ist nicht so einfach. Systeme selbst sind nichts, was jeder entwerfen kann. Wirklich ausgereifte Systeme erfordern oft langjährige Praxis, wiederholte Überarbeitung und Validierung in der realen Welt. Viele Systeme können vom ersten Konzept bis zur tatsächlichen Implementierung Jahre oder sogar Jahrzehnte kontinuierlicher Akkumulation benötigen. Daher ist die Fähigkeit zum Systementwurf bereits eine knappe Ressource.

Aber selbst wenn man ein ausgereiftes System besitzt, führt dies nicht automatisch zu einer großen Wirkung. Zwischen der technischen Machbarkeit eines Systems und seiner gesellschaftlichen Verbreitung liegen mehrere Schichten, darunter Märkte, Kapital, Vertriebskanäle, Benutzerschulung und Plattform-Ökosysteme. Der Geschichte mangelt es weder an guten Technologien noch an guten Systemen; was wirklich knapp ist, ist die Plattformfähigkeit, ein System zur großflächigen Anwendung zu bringen. Selbst wenn ein System in einem lokalen Szenario weitaus besser ist als bestehende Lösungen, kann seine tatsächliche Nutzungsskalierung

ohne ausreichende Kapitalunterstützung, Plattformpromotion und Benutzerbasis sehr begrenzt bleiben.

Aus der Perspektive der Token-Wirtschaft wird dies besonders deutlich. Die Token-Nachfrage entsteht nicht direkt aus dem System selbst, sondern aus der tatsächlichen Nutzungsskalierung des Systems. Eine Plattform, die täglich Millionen von Benutzern bedient, kann selbst mit einem relativ gewöhnlichen Systemdesign einen enormen Token-Verbrauch erzeugen; umgekehrt kann ein hochmodernes System mit sehr wenigen Benutzern eine sehr geringe Token-Nachfrage haben. Daher hängt die zukünftige Entwicklung des Token-Marktes nicht nur von der Modellfähigkeit und der Systemfähigkeit ab, sondern auch von der Plattformfähigkeit und der Ressourcenintegrationsfähigkeit.

Dies bedeutet, dass der zukünftige Wettbewerb mindestens drei Ebenen umfasst. Die erste Ebene ist der Modellwettbewerb, der bestimmt, wer Tokens anbieten kann. Die zweite Ebene ist der Systemwettbewerb, der bestimmt, wer Tokens effektiv nutzen kann. Die dritte Ebene ist der Plattformwettbewerb, der bestimmt, wie viele Benutzer ein System erreichen kann und wie viel tatsächliche Nachfrage es erzeugen kann. Modelle, Systeme und Plattformen sind alle unverzichtbar. Ohne Modelle können Systeme nicht funktionieren. Ohne Systeme können Tokens kaum Werte schaffen. Ohne Plattformen kann selbst das beste System in kleinem Maßstab stecken bleiben.

Aus dieser Perspektive bin ich noch mehr davon überzeugt, dass Tokens selbst nur grundlegende Ressourcen sind. Was die Zukunft wirklich bestimmt, sind nicht nur diejenigen, die Tokens besitzen, noch nur diejenigen, die Systeme besitzen, sondern diejenigen, die Modelle, Systeme, Plattformen und reale Bedürfnisse verbinden können. Erst wenn diese Elemente eine geschlossene Schleife bilden, werden Tokens von bloßen Recheneinheiten zu Wert schaffenden Produktivkräften.

Fünftens, die Fähigkeit, Tokens in reale Geschäftsprozesse einzubetten. Tokens müssen letztlich einem Glied in der realen Welt dienen – Logistik, Publizieren, Bildung, Gesundheitswesen. Was wirklich knapp ist, sind diejenigen, die sowohl die

Grenzen der KI-Fähigkeiten als auch die Schmerzpunkte einer bestimmten Branche verstehen. Sie wissen, wann sie KI einsetzen sollen und wann nicht; sie wissen, wie sie Token-Ausgaben in Geschäftsentscheidungen umwandeln; sie wissen, wie sie die tatsächliche Beziehung zwischen Token-Einsatz und Geschäftsertrag messen können. Diese zusammengesetzte Fähigkeit ist derzeit die knappste Ressource auf dem Markt. Die Website selbst ist ein laufender "Geschäftsprozess":

Inhaltserstellung → Metadatengenerierung → plattformübergreifende Veröffentlichung → internationale Archivierung → Leserzugriff. Dies beweist, dass meine KI-Anwendung (Token-Verbrauch) in einen vollständigen, geschlossenen Publikationsprozess eingebettet ist, der reale Wirkung erzeugt.

Ich spüre immer mehr, dass der wahre Wettbewerb in der Zukunft wahrscheinlich nicht zwischen Modellen oder zwischen Tokens stattfinden wird, sondern zwischen Systemen. Modelle werden sich weiterentwickeln, Token-Preise werden weiter fallen, und die heutigen teuren Ressourcen werden in Zukunft sehr verbreitet sein. Aber Systeme werden nicht automatisch erscheinen, nur weil die Ressourcen reichlich vorhanden sind. Systeme erfordern langfristige Akkumulation, kontinuierliche Überarbeitung und Validierung in der realen Welt. Wer Tokens in langlaufende Systeme umwandeln kann, wird eher nachhaltige Werte erzielen. Und dies ist auch der Schlüssel zur Rentabilität: Mit Tokens Geld zu verdienen, ist bereits schwierig und wird immer schwieriger werden; nur wenn es ein Frontend-System gibt, das Dienste bereitstellt, können Tokens und Rechenleistung als wirksame zusätzliche Vermögenswerte zur Erzielung von Gewinnen dienen.

Ich beurteile, dass sich der echte Wettbewerb im Token-Zeitalter in drei Phasen abspielen wird: Die erste Phase ist der Ressourcenwettbewerb, bei dem alle um Modelle, Rechenleistung, APIs und Token-Preise kämpfen; die zweite Phase ist der Anwendungswettbewerb, bei dem alle Tokens für Schreiben, Kundendienst, Codierung, Übersetzung, Bildung, Büroarbeit und Datenanalyse verwenden; die dritte Phase ist der wahre Systemwettbewerb, bei dem es darum geht, wer Tokens in langlaufende Industriesysteme, Publikationssysteme, Logistiksysteme,

Wissenssysteme und Zivilisationsaufzeichnungssysteme einbetten kann, um langfristige Vorteile zu erzielen.

In der Tat, bei der Diskussion von Tokens, Modellen, Systemen und Plattformen gibt es einen weiteren wichtigen Faktor, der in der realen Welt oft übersehen wird: die Zeit. Viele Menschen glauben, dass sie mit Kapital schnell Systeme aufbauen können; bei einem ausreichend großen Markt können sie erfolgreiche Modelle schnell replizieren. Die Realität sieht jedoch oft anders aus. Kapital kann Rechenleistung kaufen, Tokens beschaffen, Entwicklungsteams einstellen, aber es kann nicht direkt reale Daten und ausgereifte Systeme kaufen, die über viele Jahre validiert wurden. Für viele Branchen ist das, was wirklich knapp ist, nicht die Modelle selbst, noch die Tokens selbst, sondern die Systeme, die über einen langen Zeitraum stabil in realen Geschäftsumgebungen gelaufen sind.

Aus geschäftlicher Perspektive warten Märkte und Kapital normalerweise nicht lange darauf, dass ein brandneues System von Grund auf wächst. Ein großes System von der Konzeptvorschlagung, dem Architekturentwurf, der Funktionsentwicklung, der Prozessvalidierung bis zur tatsächlichen Implementierung braucht oft viele Jahre oder noch länger. Während dieses Prozesses sind nicht nur kontinuierliche Ressourceninvestitionen erforderlich, sondern auch viel Versuch und Irrtum, Überarbeitungen und Validierungen in der realen Umgebung. Daher, wenn sich Marktchancen bereits abzeichnen, neigt Kapital dazu, Systeme zu bevorzugen, die bereits geformt sind, bereits laufen und bereits validiert sind, anstatt ein völlig neues System von Grund auf zu entwerfen.

Doch gibt es einen weiteren Widerspruch in der Realität. Je größer ein System ist, desto schwieriger ist es für späteres Kapital, einfach einzusteigen. Nach langjähriger Entwicklung haben große Systeme in der Regel bereits ihre eigenen Betriebsmechanismen, Managementstrukturen, Datenlogiken und Organisationskulturen ausgebildet. Sie können über umfangreiche historische Daten, komplexe Geschäftsprozesse und langjährig angesammelte Erfahrungssysteme verfügen. Aus diesem Grund sind große Systeme zwar enorm wertvoll, aber oft schwierig schnell zu transformieren oder gemäß den Kapitalerwartungen vollständig

umzustrukturieren. Je größer das System, desto größer ist tendenziell seine Trägheit.

Daher ist der Idealzustand in der realen Geschäftswelt weder der Entwurf eines Systems von Grund auf noch die einfache Übernahme eines großen, starren Systems, sondern das Auffinden von Systemen, die ihren Kernstrukturaufbau bereits abgeschlossen haben, langfristig validiert wurden und stabil laufen können. Auch wenn solche Systeme nicht perfekt sind, haben sie einen enorm hohen Wert, weil sie bereits die schwierigste Phase überwunden haben und über reale Daten, reale Prozesse und reale Validierungsaufzeichnungen verfügen. Auf dieser Grundlage sind Neuverpackung, Optimierung, Öffnung nach außen und Plattform-Promotion oft effizienter als von Null zu beginnen, und können auch Versuch-und-Irrtum-Risiken und Zeitkosten erheblich reduzieren. Denn solche Systeme haben bereits die teuerste, zeitaufwändigste und am schwersten zu replizierende Phase abgeschlossen.

Aus dieser Perspektive ist das Zeitfenster ebenfalls eine oft übersehene Ressource. Wenn eine Branche noch in den Kinderschuhen steckt, konzentrieren sich viele Menschen auf Modellfähigkeiten, Token-Preise und Marktgröße und ignorieren die für die Systembildung erforderliche Zeit. Die Entstehung eines ausgereiften Systems durchläuft typischerweise mehrere Phasen: Konzeptvorschlag, Architekturentwurf, Prozessvalidierung, realer Betrieb und kontinuierliche Optimierung. Der wirklich schwierige Teil ist nicht das Vorschlagen der Idee, sondern das System über lange Zeit stabil in einer realen Umgebung zu halten. Daher, wenn der Markt beginnt, den Wert ausgereifter Systeme zu erkennen, bedeutet dies oft, dass diese Systeme bereits Jahre der Akkumulation hinter sich haben, anstatt dass sie in kurzer Zeit wieder reproduziert werden können.

Für Kapital und Markt sind die idealsten Ziele nicht Ideen auf der Konzeptebene, noch nicht validierte Prototypen, sondern die embryonalen Systeme, die den Kernstrukturaufbau abgeschlossen haben, über echte Betriebsaufzeichnungen verfügen und Spielraum für Upgrades bieten. Auch wenn solche Systeme nicht perfekt sind, haben sie bereits die schwierigste, zeitaufwändigste und riskanteste

Phase überwunden. Sobald die Marktnachfrage schnell wächst, werden solche Systeme oft sehr begehrt, weil die Anzahl der wirklich langvalidierten ausgereiften Systeme in der Regel viel geringer ist als der Markt annimmt. Modelle können schnell iterieren, Tokens können sich kontinuierlich ausdehnen, Entwicklungsteams können vorübergehend zusammengestellt werden, aber reale Daten, reale Prozesse und langfristige Validierungsaufzeichnungen können nicht in kurzer Zeit neu erstellt werden.

Daher beeinflusst die Zeit im fortschreitenden Prozess der Token-Wirtschaft nicht nur die Bildung von Systemen, sondern auch den Wert von Systemen. Wenn der Markt die Bedeutung eines bestimmten Systemtyps noch nicht voll erkannt hat, ist dies oft auch die Zeit, in der die Kosten für die Beschaffung solcher Systeme am niedrigsten sind; und wenn der Markt seinen Wert allgemein erkennt, sind die verfügbaren Optionen oft bereits sehr begrenzt. Für viele Branchen ist das, was verpasst wird, nicht nur ein Projekt, sondern ein Zeitfenster, um in die nächste Entwicklungsstufe einzutreten, das bereits eine langfristige Akkumulation abgeschlossen hat.

Aus der Perspektive der Token-Wirtschaft gilt dasselbe. Tokens können schnell beschafft werden, Modelle können schnell angeschlossen werden, Plattformen können schnell gebaut werden, aber was wirklich schwer schnell zu replizieren ist, sind die Systeme, die über viele Jahre sedimentiert wurden und weiterhin Werte produzieren können. In Zukunft, wenn Tokens immer reichlicher werden, wird die Bedeutung von Systemen nicht nur nicht abnehmen, sondern sogar noch zunehmen. Denn wenn alle ähnliche Token-Ressourcen haben, wird das Ergebnis des Wettbewerbs nicht mehr durch die Menge der Tokens bestimmt, sondern dadurch, wer über reifere reale Daten, stabilere Geschäftsprozesse und systemische Strukturen verfügt, die näher an den realen Bedürfnissen sind.

Ein weiteres oft übersehenes Problem ist die Herkunft von Systemen. Viele Menschen sind es gewohnt, Systeme als Softwareprodukte zu verstehen, und nehmen daher natürlich an, dass Systeme von Softwarefirmen entworfen und entwickelt werden sollten. In vielen komplexen Branchen werden jedoch die

Systeme, die wirklich über lange Zeit stabil laufen, oft nicht zuerst von Softwarefirmen entworfen und dann der Industrie übergeben, sondern sie werden allmählich im Laufe der langjährigen Geschäftspraxis geformt und dann versoftet, automatisiert und plattformisiert.

Der Grund ist nicht komplex. Softwarefirmen sind in der Regel gut in der technischen Umsetzung, aber sie bleiben nicht unbedingt über lange Zeit im spezifischen Geschäftskontext. Sie können Funktionen schnell entwickeln, aber es fällt ihnen schwer, in kurzer Zeit große Mengen an realer Geschäftserfahrung zu sammeln. Im Gegensatz dazu werden Systeme, die sich organisch innerhalb einer Branche entwickeln, täglich angesichts von Kunden, Aufträgen, Ausnahmen, regulatorischen Änderungen, Kostendruck und Marktwettbewerb kontinuierlich überarbeitet und verbessert. Diese Systeme gehen nicht von der Theorie aus, sondern von realen Problemen; jeder ihrer Prozesse, jede ihrer Regeln, jedes ihrer Automatisierungsglieder stammt typischerweise aus der langjährigen Validierung der realen Arbeit.

Daher sind in der Logistik, den Lieferketten, dem Publizieren, dem Finanzwesen, der Fertigung, dem Gesundheitswesen und vielen anderen Fachbereichen die wertvollsten Systeme oft nicht die komplexesten Systeme, noch die Systeme mit den meisten Funktionen, sondern diejenigen, die langjährig erprobt sind und reale Probleme stabil lösen können. Sie haben möglicherweise nicht die fortschrittlichsten Oberflächen, die auffälligste Werbung oder die neueste technische Architektur, aber sie verfügen über große Mengen an realen Daten, realen Prozessen und langjährig akkumulierter Geschäftslogik. Diese Erfahrung kann oft nicht durch kurzfristige Entwicklung erworben werden.

Aus der Perspektive der Token-Wirtschaft ist dies besonders wichtig. In Zukunft mögen sich die Modellfähigkeiten weiter verbessern, die Token-Kosten weiter sinken und die Softwareentwicklungseffizienz immer weiter steigen. Die von einer Branche über Jahrzehnte gesammelte Geschäftserfahrung wird jedoch nicht automatisch durch Modell-Upgrades entstehen. Was wirklich schwer zu replizieren ist, ist nicht der Softwarecode selbst, sondern das Branchenwissen, die Ausnahmeerfahrung, die

Entscheidungslogik und die langvalidierten Geschäftsprozesse, die sich hinter dem System verbergen.

Daher glaube ich, dass die idealsten Systeme der Zukunft nicht diejenigen sein werden, die völlig losgelöst von der Branchenpraxis entworfen wurden, sondern diejenigen, die aus echtem Geschäft entstanden sind, durch langjährigen Betrieb validiert wurden und dann mit Hilfe von künstlicher Intelligenz und Token-Fähigkeiten kontinuierlich verbessert werden. Solche Systeme besitzen sowohl die Zuverlässigkeit der realen Welt als auch die Effizienzgewinne des KI-Zeitalters, und ihr langfristiger Wert ist oft viel höher als der von Systemen, die allein auf technischen Konzepten aufgebaut sind.

X. Wohin fließen Tokens: Von flüchtigen Gesprächen zum Gedächtnis der Zivilisation

Zur zukünftigen Token-Wirtschaft habe ich eine weitere Beobachtung. Viele Menschen konzentrieren sich darauf, wie Tokens produziert werden, aber nur wenige beachten, wohin sie letztendlich fließen. Wenn ein Token nur in einem einzigen Chat bleibt, kann seine Lebensdauer nur wenige Minuten betragen; wenn er in eine gewöhnliche Webseite gelangt, kann seine Lebensdauer nur wenige Monate betragen; wenn er in ein Langzeitspeichersystem gelangt, ändert sich sein Wert grundlegend. Wenn Tokens in DOIs, Zenodo, WorldCat, TROVE, Langzeitspeichersysteme nationaler Bibliotheken und verschiedene akademische und Wissensinfrastrukturen eintreten, sind sie nicht mehr bloße Tokens; sie beginnen sich in Wissensaufzeichnungen, Literaturaktiva und Gedächtnis der Zivilisation zu verwandeln. Daher glaube ich, dass die wirklich wichtige Frage für die Zukunft nicht ist, wie man mehr Tokens erzeugt, sondern wie man Tokens in Strukturen gelangen lässt, in denen sie langfristig erhalten bleiben und ihre Rolle weiter ausüben können.

Aus diesem Grund bin ich zwar optimistisch, was die zukünftige Entwicklung der Tokens betrifft, aber ich glaube nicht blind an die Tokens selbst. Die Vergangenheit

der Tokens ist die grundlegende Einheit für Computer zur Sprachverarbeitung; die Gegenwart der Tokens ist eine wichtige Maßeinheit und kommerzielle Einheit im Zeitalter der KI; die Zukunft der Tokens wird wahrscheinlich eine wichtige grundlegende Ressource der digitalen Zivilisation sein. Wie auch immer sich Tokens entwickeln mögen, ich vertrete immer die Ansicht, dass das, was wirklich knapp ist, nicht die Tokens sind, sondern die Menschen, die Tokens beherrschen können; nicht diejenigen mit mehr Ressourcen, sondern diejenigen, die langfristige Strukturen aufbauen können; nicht diejenigen, die mehr Inhalte generieren können, sondern diejenigen, die Inhalte in Wissenssysteme organisieren können.

XI. Die Inkompressibilität der Zeit: Vier nicht kaufbare Vermögenswerte

Bei der Erörterung von "was nach Tokens wirklich knapp ist", muss ich auf ein grundlegendes Problem hinweisen, das die meisten Diskussionen übersehen: Bestimmte Vermögenswerte sind inhärent nicht beschleunigbar, käuflich oder reproduzierbar. Egal wie leistungsfähig KI ist, egal wie billig Tokens sind, die folgenden vier Vermögenswerte werden stets ihre inhärente Knappheit bewahren.

Erstens, Zeit ist inkompressibel.

Dreißig Jahre Praxis sind einfach dreißeig Jahre; sie können nicht auf drei Jahre komprimiert werden, noch können sie von KI simuliert werden. KI kann in Sekundenschnelle einen Artikel darüber generieren, "wie man ein Logistiksystem aufbaut", aber sie kann in Sekundenschnelle nicht die verschiedenen Anomalien, Ausfälle, Politikänderungen und Marktschwankungen erleben, denen ein Logistiksystem über zwei Jahrzehnte hinweg begegnet. Diese realen Zeitakkumulationen können durch keine technologischen Mittel verkürzt werden. Mein Beginn des Experimentierens mit einem Fernarbeits-Prototypsystem im Jahr 1993, der Entwurf eines intelligenten Rechnungssystems im Jahr 2005 – diese Zeitpunkte sind keine Zufälle; sie sind irreversible Markierungen der realen Geschichte. Jeder, der später kommt, kann meine Artikel lesen und meine Methoden

lernen, aber er kann nicht die Jahre besitzen, die ich durchlebt habe – diese Jahre selbst sind Vermögenswerte.

Zweitens, domänenübergreifende Erfahrung ist nicht käuflich.

Domänenübergreifende Erfahrung kann nicht wie eine Ware gekauft werden. Eine Person kann an einem Tag hundert Bücher über Logistik, Publishing, Finanzen und Literaturindexierung kaufen, aber sie kann nicht an demselben Tag die praktische Erfahrung erwerben, diese Bereiche zu verbinden. Die wahre domänenübergreifende Fähigkeit ist nicht die Anhäufung von Wissen, sondern die Fähigkeit, Verbindungen zwischen verschiedenen Bereichen herzustellen, Gemeinsamkeiten zu erkennen und Methoden zu übertragen. Diese Fähigkeit kann nur durch das Lösen realer Probleme, das Begehen von Fehlern, das Korrigieren von Wegen und das Ansammeln von Urteilsvermögen in einem Bereich nach dem anderen erworben werden. Der Grund, warum ich Erfahrungen aus Logistiksystemen auf Publikationssysteme übertragen kann, ist nicht, dass ich ein bestimmtes Buch gelesen habe, sondern dass ich zwanzig Jahre in der Logistik und weitere zehn Jahre im Publishing verbracht habe – an der Schnittstelle dieser beiden Zeiträume entstand die übertragbare Methodik.

Drittens, die Fähigkeit zum Systemaufbau ist schwer zu replizieren.

Die Fähigkeit zum Systemaufbau ist keine Sammlung von Code oder Prozessen, die kopiert und eingefügt werden können. Code kann kopiert werden, aber die Entscheidungslogik hinter dem Code, die Abwägungsurteile und die Voraussicht zukünftiger Änderungen können nicht repliziert werden. Der Wert eines Systems liegt nicht nur darin, was es heute tun kann, sondern auch darin, wie vielen Änderungen, wie vielen Ausfallbehebungen und wie vielen Anpassungen der Anforderungen es im letzten Jahrzehnt standgehalten hat, während es weiterlief. Diese "Geschichte des Systems" ist selbst ein Beweis für die Aufbaufähigkeit. Meine Website (<https://times.net.au>) kann von jedem eingesehen werden, aber was die Besucher sehen, ist das Ergebnis, nicht die drei Jahrzehnte Entscheidungsgeschichte, Technologieauswahl, strukturelle Anpassungen und

kontinuierliche Wartung hinter diesem Ergebnis. Diese Aufbaufähigkeit kann nicht durch Plagiat oder Nachahmung erlangt werden.

Viertens, langjährige Betriebsaufzeichnungen sind nicht abkürzbar.

Die Tatsache, dass ein System zwei Jahrzehnte lang gelaufen ist, ist selbst inhärent nicht abkürzbar. Eine langjährige Betriebsaufzeichnung beweist die Stabilität, Wartbarkeit und Designvorausschau des Systems. Im KI-Zeitalter mag die Generierung einer "so aussehenden" Website oder eines Systems in Minuten möglich sein, aber die Generierung eines "Systems, das seit zwei Jahrzehnten läuft und von CERN, WorldCat und der National Library of Australia dauerhaft archiviert wird", hat keine Abkürzungen. Eine langjährige Betriebsaufzeichnung ist ein Zeitstempel, der allen erklärt, dass dieses System die Prüfung der realen Welt bestanden hat. Eine solche Aufzeichnung kann durch keine KI-Technologie gefälscht oder beschleunigt werden.

Diese vier Vermögenswerte – Zeit, domänenübergreifende Erfahrung, Systembaufähigkeit und langjährige Betriebsaufzeichnungen – bilden gemeinsam die Grundlage für das, was ich in diesem Artikel als "wirklich knapp" bezeichnet habe. Ich besitze sie nicht, weil ich klüger oder fleißiger bin als andere, sondern weil ich früher als andere begonnen habe und bis heute weitermache. In einer Zeit, in der KI immer mehr Dinge billig und leicht zugänglich macht, werden diese Dinge, die nicht komprimierbar, nicht käuflich, nicht reproduzierbar und nicht abkürzbar sind, immer wertvoller werden.

XII. Fazit: Von der "Token-Hortung" zum "Systemaufbau"

Die Zukunft der Tokens liegt nicht in den Tokens selbst.

Mit der kontinuierlichen Entwicklung der großen Sprachmodelle und dem stetigen Sinken der Inferenzkosten wird das Angebot an Tokens weiter zunehmen. Tokens, die heute teuer sind, werden in wenigen Jahren so billig und allgegenwärtig sein wie heute der Internetverkehr und die Festplattenspeicherung. Wenn jeder leicht große Mengen an Tokens erhalten kann, wird das, was wirklich knapp sein wird, nicht mehr die Tokens sein, sondern die Fähigkeit, Tokens in realen Wert umzuwandeln – strukturelle Designfähigkeit, Automatisierungsfähigkeit, Branchenerfahrung und langjährige praktische Akkumulation. Und die tiefere Knappheit liegt in jenen Dingen, die nicht komprimierbar, nicht käuflich, nicht reproduzierbar und nicht abkürzbar sind: die Zeit selbst, domänenübergreifende Erfahrung, Systemaufbaufähigkeit und langjährige Betriebsaufzeichnungen. Wie bereits ausgeführt, kann die Fähigkeit zum Systemaufbau nicht direkt durch große Mengen von Tokens erworben werden, weil sie im Wesentlichen keine generative Fähigkeit ist, sondern eine Fähigkeit zu Urteilen, Ausnahmebehandlung, Kompromiss, Feedbackabsorption und Vertrauensaufbau – die gemeinsame Eigenschaft dieser Fähigkeiten ist, dass sie reale Weltzeit, Praxis, Fehler, Korrekturen und Validierung benötigen.

Auf der Ebene der Rentabilität gilt dieses Urteil gleichermaßen: Mit Tokens Geld zu verdienen, ist bereits schwierig und wird immer schwieriger werden. Preiskriege drücken die Gewinnmargen kontinuierlich, und das Geschäftsmodell des reinen Token-Weiterverkaufs steuert schnell auf magere oder sogar keine Gewinne zu. Nur wenn Tokens an ein Frontend-System gebunden sind – d. h. wenn ein System Dienste bereitstellt – können Tokens und Rechenleistung als wirksame zusätzliche Vermögenswerte fungieren und Gewinne erzielen. Tokens selbst sind nicht die Quelle des Gewinns; das System ist es. Tokens sind nur der Treibstoff des Systems, nicht sein Motor.

Aus dieser Perspektive haben die Systeme, die ich bereits fertiggestellt habe oder noch entwickle, eine besondere Bedeutung für die zukünftige Entwicklung von Tokens. Denn diese Systeme wurden nicht vorübergehend gebaut, um einem kurzlebigen Trend hinterherzujagen; sie existierten lange bevor das Konzept der Tokens zu einem öffentlichen Thema wurde. Das intelligente Logistiksystem bewies, dass reales Geschäft strukturiert, prozessualisiert und automatisiert werden kann; das Publikationssystem bewies, dass eine Person auch ein mehrsprachiges, multi-plattformfähiges, Multi-DOI-, Multi-Archivkanal-Literatursystem aufbauen kann; das Website- und Metadatensystem bewies, dass digitale Inhalte kontinuierlich organisiert, indiziert und verbreitet werden können; das automatische Kreuzreferenzierungssystem beweist ferner, dass, wenn die Anzahl der Werke kontinuierlich zunimmt, es nicht darauf ankommt, weiter Inhalte anzuhäufen, sondern strukturelle Beziehungen zwischen den Inhalten herzustellen. Die Website "The Epochal Transition" (<https://times.net.au>) als zentrale Präsentationsplattform dieser Errungenschaften ist selbst der empirische Beweis für die Umwandlung von Token-Fähigkeit in ein offenes, beständiges, überprüfbares akademisches System.

Selbst wenn in Zukunft alle die gleiche Menge an Tokens erhalten würden, wären die erzielten Ergebnisse höchstwahrscheinlich völlig unterschiedlich: Manche würden mehr Gesprächsprotokolle erhalten, manche mehr Webseiteninhalte, während andere vielleicht ein Logistiksystem, ein Publikationssystem oder ein Wissenssystem aufbauen würden, das über Jahrzehnte hinweg kontinuierlich läuft. Den Unterschied machen nicht die Tokens selbst aus, sondern die dahinterstehende Systemarchitektur – und die inkompressible Zeit, die nicht käufliche domänenübergreifende Erfahrung, die schwer reproduzierbare Aufbaufähigkeit und die nicht abkürzbaren langjährigen Betriebsaufzeichnungen, die diese Systemarchitekturen stützen.

Daher betrachte ich Tokens nicht als Endziel, sondern als die neue Infrastruktur des intelligenten Zeitalters. Tokens sind so wichtig wie Strom, aber Strom allein baut keine Fabriken; Tokens sind so messbar wie Geld, aber Geld allein schafft keine hochwertigen Vermögenswerte; Tokens können wie Wasser fließen, aber ohne

Rohre, Dispositionssysteme und langjährige Wartungsaufzeichnungen werden sie nur versickern. Die wirklich wertvollen Menschen der Zukunft werden nicht unbedingt diejenigen mit den meisten Tokens sein, sondern diejenigen, die Tokens in Systeme organisieren können. Der größte Irrglaube des Token-Zeitalters ist es, Tokens als Ziel zu betrachten; die größte Chance des Token-Zeitalters ist es, Tokens zu nutzen, um Strukturen zu bauen, die andere nicht nachahmen können. Wenn Tokens in Zukunft so alltäglich sein werden wie Strom, werden die Menschen, die auf den heutigen Token-Hype zurückblicken, vielleicht feststellen, dass das, was den Unterschied zwischen den Menschen wirklich ausmachte, nie die Anzahl der Tokens war, die sie besaßen, sondern ob sie die Fähigkeit besaßen, kontinuierlich zu lernen, kontinuierlich zu praktizieren, kontinuierlich anzuhäufen und kontinuierlich Systeme aufzubauen – und die Dinge hinter diesen Fähigkeiten, die nicht komprimierbar, nicht käuflich, nicht reproduzierbar und nicht abkürzbar sind. Und meiner Ansicht nach ist diese Fähigkeit der knappste und am schwierigsten zu reproduzierende Reichtum im Token-Zeitalter.

Mit zunehmender Vertiefung meiner Forschung spüre ich immer mehr, dass sich die Wettbewerbslandschaft des zukünftigen KI-Zeitalters vielleicht mit einer einfachen Analogie zusammenfassen lässt. Viele Menschen betrachten Rechenleistung, Modelle und Tokens als Kernressourcen, aber aus der Perspektive der realen Welt ähneln sie eher dem Logistiksystem hinter einer Armee. Rechenleistung bestimmt die Produktionskapazität, Modelle bestimmen die Organisationsfähigkeit, Tokens bestimmen die Verteilungsfähigkeit; gemeinsam bilden sie das materielle Unterstützungssystem des KI-Zeitalters.

Was jedoch den Ausgang einer Schlacht wirklich bestimmt, sind nicht die Logistikgüter selbst, sondern das Frontkampfsystem. Für den zukünftigen Branchenwettbewerb sind Systeme eher wie die Angriffslinie. Systeme entscheiden, wie Ressourcen verwendet werden, wie Prozesse ablaufen, wie Werte geschaffen werden und ob die Benutzer wirklich profitieren. So reichlich Tokens und Rechenleistung auch sein mögen, ohne die Organisation eines Systems sind sie wie

in einem Lagerhaus gestapelte Vorräte, die nicht in reale Ergebnisse umgewandelt werden können.

Aus dieser Perspektive wird der zukünftige Wettbewerb wahrscheinlich nicht zwischen Rechenleistungen oder zwischen Tokens stattfinden, sondern zwischen Systemen. Das System ist die Angriffslinie und bestimmt die Angriffsrichtung und Kampffähigkeit; Tokens sind Munition und Vorräte und bestimmen die nachhaltige Kampffähigkeit; Rechenleistung ist die rückwärtige Industriebasis und bestimmt die Produktionsgeschwindigkeit und Versorgungskapazität. Ohne rückwärtige Versorgung kann die Angriffslinie nicht lange vorrücken; aber mit nur rückwärtiger Versorgung und ohne Angriffslinie können ebenfalls keine Ergebnisse erzielt werden.

Daher, wenn Rechenleistung allgegenwärtiger wird, Modelle sich immer ähnlicher werden und Tokens immer reichlicher vorhanden sind, bleibt das System das wirklich Knappe. Wer ein reiferes System, realere Daten, länger validierte Prozesse und die Fähigkeit besitzt, Tokens und Rechenleistung mit realen Bedürfnissen zu verbinden, wird eher langfristige Vorteile erzielen.

Tokens werden zur "Elektrizität" des KI-Zeitalters – unverzichtbar, aber kein Selbstzweck. So wie die Gewinner der Elektrizitätsrevolution nicht die Kraftwerke waren, sondern die Erfinder und Unternehmer, die die Glühbirne, den Motor und den Elektrolyseur erfanden, werden die wahren Gewinner des KI-Zeitalters die Systembauer sein, die Tokens nutzen können, um nicht reproduzierbare Strukturen und Architekturen aufzubauen. Nach den Tokens ist dies das wirklich Knappe.

Wenn ich auf den gesamten Artikel zurückblicke, spüre ich immer mehr, dass sich die Wettbewerbslandschaft des zukünftigen KI-Zeitalters vielleicht mit der Analogie eines Baumes veranschaulichen lässt.

Viele Menschen konzentrieren sich auf Wasser und Nährstoffe, weil sie sichtbar, greifbar und leicht messbar sind. Die heutige Rechenleistung, Modelle und Tokens sind wie das kontinuierlich gelieferte Wasser und die Nährstoffe. Rechenleistung bestimmt die Produktionskapazität, Modelle bestimmen die Transformationsfähigkeit,

Tokens bestimmen die Verteilungsfähigkeit. Mit fortschreitender Technologie werden diese Ressourcen immer reichlicher und leichter zugänglich.

Was jedoch die Vitalität eines Baumes wirklich bestimmt, ist nicht das Wasser und die Nährstoffe selbst, sondern das Wurzelsystem.

Das System ist die Wurzel der zukünftigen digitalen Zivilisation.

Das Wurzelsystem bestimmt, wie Wasser aufgenommen wird, wie Nährstoffe verteilt werden und ob der ganze Baum langfristig wachsen kann. Ohne Wurzeln wird selbst das meiste Wasser versickern; ohne Wurzeln können selbst die reichlichsten Nährstoffe nicht in Vitalität umgewandelt werden. Ebenso können ohne ein System selbst das leistungsfähigste Modell, die billigsten Tokens und die riesigste Rechenleistung nur schwer dauerhaft Werte schaffen.

Aus dieser Perspektive ist Rechenleistung Wasser, Tokens sind Nährstoffe, Modelle sind die Rohre, die die Nährstoffe liefern, und das System ist das Wurzelsystem, das tief in die reale Welt eindringt. Was den Ausgang des zukünftigen Wettbewerbs wirklich bestimmt, ist nicht, wer mehr Wasser hat, noch wer mehr Nährstoffe hat, sondern wer tiefere, stabilere Wurzeln hat, die in der Lage sind, diese Ressourcen langfristig zu absorbieren und zu nutzen.

In der Zukunft, wenn die Rechenleistung weiter wächst, die Modelle sich weiterentwickeln und die Token-Preise weiter fallen, werden die Ressourcen selbst immer gewöhnlicher. Das, was wirklich knapp sein wird, sind nicht mehr die Ressourcen, sondern die Systeme, die in der Lage sind, Ressourcen zu organisieren, zu nutzen und kontinuierlich Werte zu schaffen. Wer ein reifes System besitzt, besitzt die Wurzeln; wer die Wurzeln besitzt, besitzt die Fähigkeit zum langfristigen Überleben und kontinuierlichen Wachstum.

Daher glaube ich, dass das, was nach Tokens wirklich knapp ist, nicht mehr Tokens oder mehr Rechenleistung ist, sondern die Systeme, die Rechenleistung, Modelle und Tokens in realen Wert umwandeln können.

Wenn Wasser und Nährstoffe immer reichlicher werden, bestimmt immer noch das Wurzelsystem, ob ein Baum langfristig wachsen kann.

Schlüsselwörter: Token, Künstliche Intelligenz, KI, Große Sprachmodelle, Rechenleistung, Modellökonomie, Token-Ökonomie, Chinesischer Token, Englischer Token, BPE-Algorithmus, Datensouveränität, Agent, API-Ökosystem, Systembaufähigkeit, Systemdesign, Automatisierungssystem, Interdisziplinäre Forschung, Interdisziplinäre Praxis, Reale historische Daten, Langlaufsystem, Zeitvermögenswert, Plattformwettbewerb, Systemwettbewerb, Digitale Zivilisation, Gedächtnis der Zivilisation, Wissensinfrastruktur, Unabhängiger Gelehrter, Internationaler unabhängiger Gelehrter, Intelligentes Logistiksystem, Zehn-Sprachen-Publikationssystem, Metadatensystem, Automatische Kreuzreferenzierung, DOI, Zenodo, WorldCat, TROVE, The Epochal Transition, Wu Zhaohui, JEFFI CHAO HUI WU

Literaturverzeichnis:

1. *A Structural Case Study of the 1993 Remote Work Prototype System*
DOI: <https://doi.org/10.5281/zenodo.17978371>
2. *The 2005 Intelligent Invoice System*
DOI: <https://doi.org/10.5281/zenodo.19044755>
3. *Mature Structural Logistics System*
DOI: <https://doi.org/10.5281/zenodo.20201895>
4. **Cross-Generational 20-Year Intelligent Logistics and Financial System**
DOI: <https://doi.org/10.5281/zenodo.20352371>
NLA: <https://nla.gov.au/nla.obj-4204311557>
5. *Analysis of the Homepage Structural Evolution of The Epochal Transition*
DOI: <https://doi.org/10.5281/zenodo.20362574>
NLA: <https://nla.gov.au/nla.obj-4204309341>

6. *An Empirical Study of Institutional Engineering: Independent Scholars Accessing the WorldCat Global Bibliographic Infrastructure*

DOI: <https://doi.org/10.5281/zenodo.18028572>

7. *National-level Network Archiving and Digital Document Preservation Guidelines of the Australian Winner Information Network*

DOI: <https://doi.org/10.5281/zenodo.17888259>

8. *The Institutional Archiving of The Epochal Transition*

DOI: <https://doi.org/10.5281/zenodo.17932379>

WU, J. C. H. (2026). Source Declaration for Audiovisual and Derivative Adaptations of a Continuing Real-World Narrative. Zenodo.

<https://doi.org/10.5281/zenodo.18160116>

Depois dos Tokens: O que é verdadeiramente escasso na Era da IA

Autor/Investigador Independente Internacional: Wu Zhaohui, JEFFI CHAO HUI WU

ORCID: <https://orcid.org/0009-0009-5318-3027>

Declaração Preliminar

Este artigo baseia-se nas minhas observações pessoais decorrentes de mais de trinta anos de prática interdisciplinar (desde os meus estudos na Austrália em 1989, abrangendo trabalho remoto, gestão logística, processamento automatizado, construção de sítios web, divulgação cultural, edição multilíngue e investigação interdisciplinar). **Os tokens são um fenómeno novo em rápida evolução; este artigo representa apenas a minha observação e julgamento pessoais, numa fase específica, até à data da redação (junho de 2026) e não constitui de forma alguma aconselhamento de investimento, previsão comercial ou padrão industrial.**

Este artigo não procura aplicabilidade universal nem pretende representar qualquer consenso da indústria. Com a evolução contínua do mercado de tokens, da tecnologia de IA e do ambiente político global, alguns dos juízos contidos neste artigo poderão necessitar de revisão ou atualização no futuro. O leitor deve fazer juízos independentes com base nas informações mais recentes.

As "cinco capacidades escassas" e os "cinco ativos pessoais" que discuto derivam da minha própria prática de construção de sistemas (evidências públicas disponíveis em <https://times.net.au>), incluindo a lista completa de mais de 40 obras criativas independentes nesse site, o ISSN 3083-5178, mais de 60 DOIs, e registos arquivados permanentemente por infraestruturas internacionais como a WorldCat, a Biblioteca Nacional da Austrália e o CERN Zenodo).

Para evitar tirar de contexto, qualquer refutação ou questionamento válido deste artigo deve ter como pré-requisito: **a leitura completa de todos os artigos ligados neste documento e das ligações aninhadas neles contidas, bem como de todas as minhas obras listadas em <https://times.net.au>**. Citações parciais ou resumos simplificados desligados deste contexto global não podem constituir uma avaliação justa deste artigo.

Nota complementar sobre refutação válida:

Este artigo expressa opiniões pessoais, não procura aplicabilidade universal e não pressupõe a citação de trabalhos de outros. Qualquer questionamento válido deve adotar uma comparação do tipo "maçã com maçã": factos contra factos, prática pessoal contra prática pessoal, sistemas completados contra sistemas completados. Usar "o que a IA futura poderá alcançar" para negar "práticas já completadas hoje", ou comparar capacidades de equipa/instituição com prática individual, não constitui uma comparação justa.

A carteira de cinco ativos descrita neste artigo constitui um registo de prática pessoal e não afirma que "não existe nenhum caso comparável". Se acredita que existe um caso totalmente comparável, forneça um caso público específico e explique as suas evidências correspondentes nas cinco dimensões de dados históricos reais, sistemas interdisciplinares, capacidade de construção de sistemas, visão prospetiva e capacidade de autoverificação. Caso contrário, qualquer questionamento é meramente uma expressão de opinião e não constitui uma refutação válida.

Resumo

Os tokens estão a tornar-se a unidade de medida central na era da IA, mas este artigo argumentará que o token não é o valor final. À medida que os custos de inferência continuam a diminuir, os tokens tornar-se-ão tão baratos e omnipresentes como são hoje o tráfego de Internet e o armazenamento em disco rígido. Quando os tokens deixarem de ser escassos, o que será verdadeiramente escasso? Com base

em mais de trinta anos de prática interdisciplinar do autor (desde um sistema protótipo de trabalho remoto em 1993, um sistema de faturação inteligente em 2005, até sistemas logísticos inteligentes, um sistema de edição em dez línguas, gestão de DOI, referência cruzada automática e publicação multiplataforma), este artigo apresenta os seguintes juízos fundamentais: Primeiro, a capacidade de construção de sistemas – incluindo design estrutural, automação, prática interdisciplinar, modelos mentais de funcionamento a longo prazo e a capacidade de incorporar tokens em processos de negócio reais – é o verdadeiro ativo escasso. Segundo, a capacidade de construção de sistemas não pode ser adquirida diretamente através de grandes quantidades de tokens porque depende da tomada de decisões, tratamento de exceções, compromissos, ciclos de retroalimentação do mundo real e confiança acumulada ao longo do tempo. Terceiro, a incompressibilidade do tempo, a impossibilidade de comprar experiência interdisciplinar, a dificuldade de replicar a capacidade de construção de sistemas e a impossibilidade de abreviar registos de funcionamento a longo prazo – estes quatro ativos constituem a escassez mais profunda na era dos tokens. Este artigo utiliza como evidência pública o sítio web académico *The Epochal Transition* (<https://times.net.au/>), ISSN 3083-5178, arquivado permanentemente por infraestruturas internacionais como a WorldCat, o CERN Zenodo e o TROVE da Biblioteca Nacional da Austrália), construído pelo autor, apresentando como uma pessoa pode construir um sistema interdisciplinar, verificável e duradouro ao longo de trinta anos.

Este artigo argumenta que: os tokens são apenas o combustível; o sistema é o motor. Depois dos tokens, o que é verdadeiramente escasso é a pessoa que consegue dominar os tokens, e as estruturas que ela construiu.

"Sem mercadorias, de que serve o dinheiro?"

Quando revejo a história do desenvolvimento dos tokens, sinto cada vez mais que esta se assemelha a muitas revoluções tecnológicas que a humanidade experimentou no passado. Cada vez que surge um novo recurso fundamental, as pessoas concentram-se primeiro no próprio recurso, negligenciando a estrutura subjacente. Isto foi verdade na era industrial, na era da Internet, e é igualmente

verdade na era da inteligência artificial. Hoje, muitas discussões centram-se nos preços dos tokens, custos dos tokens, quotas de tokens e capacidade de fornecimento de tokens, como se possuir mais tokens significasse possuir o futuro. Mas, de acordo com as minhas décadas de experiência prática, o recurso em si nunca foi o valor final. O que realmente determina o valor é sempre a forma como as pessoas organizam e utilizam esses recursos.

Para compreender os "tokens", não se pode começar simplesmente com um novo termo; é preciso começar pela sua história. Desde o filósofo americano Charles Sanders Peirce a escrever vinte instâncias da palavra "the" em papel em 1906, distinguindo entre "tipo" e "token", até Liu Liehong, diretor da Administração Nacional de Dados da China, a dar a tradução chinesa oficial "ciyuan" (词元) para Token no Fórum de Boao em 2026, este conceito percorreu um longo caminho, da filosofia à informática, e da blockchain à inteligência artificial. A palavra "token" existia há muito tempo na informática ocidental, onde, em linguagens de programação, compiladores e sistemas de processamento de texto, um token se referia à unidade básica que um computador podia reconhecer e processar. Mais tarde, em contextos de segurança de rede, autenticação, chamadas de API e blockchain, "token" era frequentemente entendido como uma "credencial" ou "passe", indicando direitos de acesso, credenciais de identidade ou ativos digitais. Mas na era dos grandes modelos de linguagem, o Token já não é primariamente uma credencial de identidade, nem um ativo de blockchain, mas sim a unidade básica de cálculo com a qual os modelos de IA processam linguagem, código, descrições de imagens e informações multimodais.

A China mais tarde nomeou o Token no campo da IA como "ciyuan" (词元). Considero esta uma tradução bastante precisa. Não usou "lingpai" (passe), porque o Token nos grandes modelos não é um passe; também não o traduziu simplesmente como "zi" (caráter) ou "ci" (palavra), porque pode ser às vezes um caráter, às vezes uma palavra, às vezes metade de uma palavra inglesa, e às vezes até um sinal de pontuação ou símbolo. "Ci" indica a sua relação com a linguagem e a informação, e "yuan" indica que é uma unidade básica. Para o leitor geral, pode ser entendido

simplesmente como: o que os humanos veem como uma frase, a IA vê como uma sequência de tokens. Neste caminho, os tokens passaram de um termo linguístico obscuro para a unidade de medida central da era da IA, e depois para uma ferramenta de liquidação que liga a oferta tecnológica à procura comercial. Como profissional que experimentou pessoalmente projetos de longo prazo, como a construção de um sistema logístico inteligente, a conceção de processos de edição automatizados e a distribuição de conteúdo multiplataforma, tenho uma compreensão visceral desta evolução – os tokens nunca foram o fim, mas o meio; o que realmente determina o valor nunca é quem possui mais tokens, mas quem pode usar os tokens para construir sistemas duradouros e estruturas irreplicáveis. A capacidade de construção de sistemas discutida neste artigo teve os seus resultados apresentados publicamente em <https://times.net.au> – um sistema de edição académica interdisciplinar construído por um investigador independente e arquivado permanentemente no sistema bibliotecário mundial.

Refletindo sobre a minha própria jornada, não comecei a pensar em sistemas apenas após o aparecimento dos tokens, mas desenvolvi gradualmente uma compreensão dos sistemas através de uma prática de longa duração. Desde os meus estudos na Austrália em 1989, passei por experiências de aprendizagem, trabalho e exploração em diferentes áreas. Mais tarde, envolvi-me gradualmente em trabalho remoto, gestão logística, processamento automatizado, construção de sítios web, divulgação cultural, edição multilingue e investigação interdisciplinar. Olhando para trás, embora estas experiências tenham ocorrido em épocas diferentes, partilham uma característica comum: giraram sempre em torno de como estabelecer estruturas, como reduzir o trabalho repetitivo e como melhorar a eficiência geral. O conceito de tokens não existia então, nem os grandes modelos de linguagem de hoje, mas eu já pensava constantemente em como fazer a informação fluir, organizar-se e processar-se automaticamente.

Portanto, enquanto um número crescente de pessoas hoje considera os tokens como um novo meio de produção, eu estou mais inclinado a vê-los como uma nova forma de energia fundamental. A energia é certamente importante, mas a energia

não cria valor automaticamente. A eletricidade por si só não se torna automaticamente uma fábrica; a Internet por si só não se torna automaticamente uma empresa; os tokens por si só não se tornam automaticamente um sistema de conhecimento. No futuro, mesmo que todos possam obter facilmente grandes quantidades de tokens, isso não significa que todos possam alcançar os mesmos resultados. O que faz a diferença não é a quantidade de tokens, mas a estrutura em que os tokens entram.

Antes de proceder a uma análise aprofundada dos tokens, preciso de apontar outra observação importante: Os tokens são semelhantes a moedas, e as diferenças tecnológicas, sistémicas e políticas entre o Oriente e o Ocidente tornam-nos não universais. Assim como o dólar americano e o renminbi chinês não podem ser livremente trocados, os tokens chineses e os tokens ingleses enfrentam barreiras de circulação semelhantes. Não é apenas uma questão de diferença linguística; a razão mais profunda reside nas triplas diferenças entre Oriente e Ocidente ao nível tecnológico (arquiteturas de modelos, algoritmos de tokenização, dados de treino), ao nível sistémico (ecossistemas de API, sistemas de pagamento, padrões de dados) e ao nível político (fluxos de dados transfronteiriços, controlos de exportação, revisões de segurança nacional), que impedem os tokens de circular livremente como uma idealizada "moeda universal global de IA". Esta observação tem implicações significativas para a compreensão da paisagem futura do mercado de tokens – no futuro, podem surgir "zonas de tokens chineses", "zonas de tokens ingleses" e outros ecossistemas de tokens baseados em língua e política, e as chamadas interzonas enfrentarão perdas de eficiência e custos de conformidade.

I. Origens dos tokens: O salto quádruplo do conceito filosófico para a unidade económica

A evolução da identidade dos tokens pode ser claramente dividida em quatro etapas.

A origem na linguística filosófica. Em 1906, Peirce, ao estudar a semiótica, distinguiu entre "tipo" e "token". O primeiro representa uma regra ou forma abstrata, enquanto o segundo é a instanciação específica dessa regra na realidade. Esta distinção tornou-se mais tarde uma base importante da linguística estrutural.

A viragem da codificação na ciência da computação. Na década de 1960, quando um programador escrevia a linha de código "int x = 5;", o computador precisava de a decompor em unidades independentes como "int", "x", "=", "5", etc. Cada uma dessas unidades era um token. Os tokens tornaram-se assim a menor unidade de informação para uma máquina "ler" instruções.

A experiência do veículo de valor com a blockchain. Em 2017, com a ascensão da loucura das Ofertas Iniciais de Moedas (ICO), foi dada ao Token uma nova identidade como um "certificado de direitos digitais negociável". Embora essa vaga de entusiasmo tenha gradualmente arrefecido, a compreensão do "Token como símbolo de valor" tinha criado raízes.

A infraestrutura económica da era da IA. Março de 2026 foi um momento marcante – o CEO da NVIDIA, Jensen Huang, redefiniu os centros de dados como "fábricas que produzem Tokens de IA inteligentes", e Liu Liehong, diretor da Administração Nacional de Dados da China, deu a tradução chinesa oficial "ciyuan" (词元), chamando-lhe a "âncora de valor da era inteligente". Neste ponto, os tokens completaram o seu salto final de conceito técnico para conceito económico.

II. A base técnica: Algoritmo BPE e o imposto linguístico implícito

A razão pela qual os tokens puderam tornar-se a unidade de medida padrão para a IA baseia-se numa tecnologia esquecida por mais de duas décadas – a codificação por pares de bytes (Byte Pair Encoding, BPE).

Em 1994, o programador americano Philip Gage publicou um artigo sobre algoritmos de compressão de dados numa revista técnica da linguagem C,

apresentando o princípio básico do BPE: digitalizar o texto repetidamente e "soldar" os dois caracteres adjacentes mais frequentes num novo símbolo, comprimindo iterativamente ronda após ronda. Este artigo não atraiu atenção na altura porque a sua eficiência de compressão não era notável.

Foi só em 2016 que Rico Sennrich, um investigador da Universidade de Edimburgo, ao estudar o problema da segmentação de palavras em tradução automática, encontrou por acaso este velho artigo e percebeu agudamente que o BPE era a solução perfeita para o problema da tokenização – não havia necessidade de pré-definir um dicionário; permitia-se que os dados "falassem" por si próprios, com combinações de alta frequência a coalescerem naturalmente em tokens. Em 2019, a OpenAI adotou oficialmente esta abordagem ao lançar o GPT-2, definindo o ponto de partida para a tokenização diretamente ao nível do "byte", permitindo ao modelo processar teoricamente qualquer língua escrita.

No entanto, a lógica de "prioridade à frequência" do algoritmo BPE formou inadvertidamente um sistema implícito de "imposto linguístico". Como o inglês é a corrente dominante absoluta dos corpus da Internet, a eficiência do modelo na segmentação de tokens ingleses é muito superior à de outras línguas. Para expressar o mesmo significado, o inglês usa o menor número de tokens, o chinês normalmente precisa de 1,5 a 2 vezes mais, e para línguas com ainda menos recursos, o custo pode ser de 5 a 10 vezes superior ao do inglês. Isto significa que, sob a tarifação por token, o custo real de conversar com a IA varia enormemente dependendo da língua usada.

Esta injustiça na "linha de partida", uma vez inscrita no vocabulário inicial de um modelo, é muito difícil de corrigir – as regras de tokenização são o "fundamento" da compreensão do mundo pela IA; quanto mais alto o edifício, mais difícil é substituir o fundamento. Felizmente, esta situação está a melhorar. Tomando a série GPT como exemplo, a mesma frase em chinês exigia 38 tokens no GPT-3, caiu para 26 no GPT-4, e para 15 no GPT-5 – uma melhoria de eficiência de mais de 60%. Além disso, os grandes modelos chineses domésticos como o Tongyi Qianwen e o DeepSeek incorporaram desde a fase de conceção frases chinesas de alta

frequência, expressões idiomáticas, etc., como tokens nativos nos seus vocabulários, alcançando assim uma eficiência de processamento do chinês mais "nativa". Isto revela um princípio profundo: quem quer que domine o "direito à segmentação semântica" detém em grande parte a vantagem de eficiência e custo para expressar essa língua no mundo digital – esta é uma forma de "direito fundamental de cunhagem" na era digital.

III. Usos e produção: O salto exponencial da conversa para os agentes

Os cenários de consumo de tokens estão a passar por um salto de três níveis, da "conversa" ao "raciocínio" e depois aos "agentes".

Na fase da IA conversacional, um único pedido e resposta consome cerca de 10.000 tokens. Na fase da IA de raciocínio, envolvendo múltiplas rondas de interação e inferência complexa, cada tarefa consome cerca de 100.000 tokens. Ao entrar na fase da IA agentica, os agentes podem executar tarefas de múltiplos passos de forma autónoma e invocar múltiplas ferramentas, com o consumo de tokens por tarefa a atingir aproximadamente 1 milhão – um aumento de quase cem vezes em apenas duas gerações de evolução. O CEO da Qualcomm, Amon, prevê que até 2030, a procura global de tokens a cada 10 segundos explodirá de 31,7 mil milhões em 2026 para 1,27 biliões. Isto não é exagero; a popularidade explosiva de ferramentas de agente de código aberto como o OpenClaw provou este ponto: uma única tarefa complexa a consumir dezenas de milhões de tokens é comum.

No lado da produção, a China tornou-se o maior fornecedor mundial de tokens. Dados públicos mostram que, em março de 2026, o volume diário de chamadas de tokens da China excedeu os 140 biliões, um aumento de mais de mil vezes em dois anos, representando 61% do uso global de tokens de IA. O estabelecimento desta posição baseia-se na sobreposição de três vantagens: primeiro, os preços industriais da eletricidade na China são muito mais baixos do que na Europa e América, e a eletricidade representa mais de 60% dos custos de produção de

tokens; segundo, a China construiu o maior cluster de computação inteligente do mundo, alcançando uma produção eficiente através da disposição "Computação a Este, Dados a Oeste"; terceiro, os grandes modelos chineses domésticos são totalmente de código aberto, e o sistema de serviços de comercialização de API é maduro, formando uma cadeia industrial completa em circuito fechado, desde "treino do modelo – suporte de computação – saída de inferência – chamada global".

IV. Instituições e mercados: A paisagem industrial tripartida

Uma estrutura de mercado claramente estratificada formou-se em torno dos tokens.

Para compreender esta estrutura industrial, é também necessário considerar o sistema de suporte subjacente aos tokens. Muitas pessoas concentram-se nos preços dos tokens, mas ignoram que os tokens não são um recurso independente. Por trás dos tokens estão os modelos, e por trás dos modelos está o poder de computação. Num sentido, o poder de computação determina a capacidade do modelo, os modelos determinam a produção de tokens, e os tokens são a unidade de medida formada quando as capacidades do modelo são exportadas para o mercado. Portanto, à superfície, o mercado de tokens é uma questão de concorrência de preços, mas na sua essência, continua a ser uma concorrência em poder de computação e eficiência.

A atual cadeia industrial global de tokens pode ser grosseiramente dividida em três níveis. O nível mais alto é o da infraestrutura de poder de computação, incluindo fabricantes de chips, fabricantes de servidores, plataformas de computação em nuvem, centros de dados e sistemas energéticos. O nível intermédio consiste em instituições de modelos fundamentais e plataformas de inferência, responsáveis por converter o poder de computação em serviços de tokens chamáveis. O nível inferior consiste em fornecedores de serviços de tokens, plataformas de agentes e várias aplicações industriais, que convertem ainda mais as capacidades dos modelos em serviços empresariais e produtos finais. O que os utilizadores veem finalmente é o preço do token, mas o que realmente afeta o custo e a velocidade do token é

frequentemente a mudança no sistema de poder de computação a montante e na eficiência do modelo.

Nos últimos anos, com a melhoria contínua do desempenho dos chips, a otimização das arquiteturas dos modelos e o aumento da eficiência da inferência, o custo por token mostrou uma tendência de queda rápida. Muitas pessoas veem os tokens a tornar-se mais baratos, mas o verdadeiro motor desta mudança é o progresso contínuo do sistema de poder de computação. Portanto, a concorrência futura no mercado de tokens envolverá não apenas a concorrência entre plataformas e de preços, mas também a concorrência em recursos de computação, eficiência dos modelos e capacidades de integração de recursos.

"Centrais elétricas" a montante – produtores de poder de computação. As empresas de chips representadas pela NVIDIA melhoram continuamente o hardware para melhorar a eficiência da produção de tokens por unidade de poder de computação. A NVIDIA prevê que a sua receita atingirá pelo menos 1 bilhão de dólares até 2027, uma confiança apoiada pela sede global de poder de computação de IA.

"Regulação da rede" a meio – plataformas de modelos e agregadores. Empresas como OpenAI, Google, DeepSeek e Alibaba convertem o poder de computação subjacente em serviços de tokens padronizados e chamáveis. Três modelos de negócio maduros surgiram em torno deste elo: plataformas de agregação de tokens (representadas pela OpenRouter, operando com diferenças de preço e comissões, com um volume de chamadas semanal a atingir 26,9 bilhões de tokens em maio de 2026), serviços MaaS (Model-as-a-Service) dos fornecedores de nuvem, e estações de retransmissão de IA ativas no mercado secundário.

"Pontos de entrada" a jusante – plataformas de aplicação e distribuição. A abertura da interface WeChat pela Tencent para aceder a capacidades de novos modelos transforma essencialmente o WeChat numa "plataforma de distribuição de tokens"; a China Telecom propôs "remodelar o seu negócio através da gestão de tokens", tentando sair da lógica tradicional do tráfego. O que estas empresas de ponto de entrada têm em comum é que os utilizadores já não enfrentam o modelo

diretamente, mas chamam tokens indiretamente através da plataforma – o próprio modelo está a tornar-se "invisível".

Vale a pena mencionar que se formou um paradigma único de comércio global em torno dos tokens: produção intangível, funcionamento isento de impostos e cobertura global. Isto está a transformar a China de uma tradicional "fábrica do mundo" num "fornecedor global de infraestrutura de poder de computação de IA".

Ao discutir a economia dos tokens, há outro fator fundamental que não pode ser contornado: o poder de computação. Muitas pessoas concentram-se nos preços dos tokens, mas ignoram o processo de produção por trás dos tokens. Na realidade, os tokens não são gerados do nada. Cada geração de token requer que o modelo realize uma inferência, e a inferência do modelo baseia-se numa base massiva de poder de computação. Num sentido, o poder de computação é a fonte da economia dos tokens, os modelos são as ferramentas de produção dos tokens, e os tokens são a unidade de medida formada quando as capacidades do modelo são exportadas para o exterior.

Portanto, os níveis de poder de computação afetam diretamente os custos dos tokens e a velocidade de geração de tokens. Quanto mais forte o poder de computação, maior a capacidade do modelo para processar informações, e mais tokens podem ser produzidos por unidade de tempo; quanto menor o custo do poder de computação, mais baixos tendem a ser os preços dos tokens. Nos últimos anos, a indústria dos grandes modelos tem assistido a um declínio contínuo nos preços dos tokens, impulsionado não só pela otimização dos algoritmos dos modelos, mas também por melhorias no desempenho dos chips, aumento da escala dos centros de dados e aumento da eficiência da inferência. Por outras palavras, muitas pessoas veem os tokens a tornar-se mais baratos, mas o verdadeiro motor desta mudança é o avanço contínuo do sistema de poder de computação.

De uma perspetiva de cadeia industrial, o poder de computação, os modelos e os tokens formam na realidade uma cadeia completa. A montante estão os chips, servidores, centros de dados e sistemas energéticos; a meio estão os modelos

fundamentais e as plataformas de inferência; a jusante estão os serviços de tokens, as aplicações de agentes e várias soluções industriais. O poder de computação determina o limite superior da capacidade do modelo; os modelos determinam a qualidade e eficiência dos tokens; e os tokens determinam finalmente quantos serviços de IA o utilizador pode obter.

No entanto, creio que, embora a importância do poder de computação seja inegável, o poder de computação em si não cria valor automaticamente. Muitas indústrias experimentaram processos semelhantes na história. Possuir uma central elétrica maior não significa possuir uma empresa melhor; possuir mais servidores não significa possuir um produto melhor. Da mesma forma, possuir mais poder de computação não significa possuir um sistema de maior valor. O poder de computação resolve o problema da capacidade de produção, enquanto os sistemas resolvem o problema da conversão de valor. No futuro, à medida que o poder de computação continua a crescer, as capacidades dos modelos continuam a fortalecer-se e os recursos de tokens continuam a proliferar, o que será verdadeiramente escasso não será o próprio poder de computação, mas a forma de organizar estes recursos em sistemas capazes de funcionar a longo prazo no mundo real.

Portanto, de uma perspetiva de concorrência futura, creio que a concorrência em poder de computação, modelos e tokens se deslocará gradualmente para uma concorrência em sistemas. Porque quando o poder de computação se tornar mais generalizado, os modelos se tornarem mais semelhantes e os tokens mais abundantes, o que fará a diferença já não será a quantidade de recursos possuídos, mas quem pode usar esses recursos para construir estruturas eficazes a longo prazo.

V. Concorrência e barreiras: Involução da guerra de preços e estratificação de valor

A concorrência no mercado de tokens apresenta características de involução muito pronunciadas. De 2022 a 2026, o custo por inferência de token diminuiu 99,9%. Os modelos domésticos apresentam vantagens de preço ainda mais significativas: o preço de entrada do DeepSeek V4-Pro é tão baixo quanto 0,025 RMB por milhão de tokens, o preço de saída do Tongyi Qianwen Qwen3.5-Plus é de 1,6 RMB por milhão de tokens, em comparação com o preço de saída do OpenAI GPT-4o de aproximadamente 72 RMB.

Esta guerra de preços comprimiu fortemente as margens de lucro dos "intermediários". As estatísticas da indústria mostram que o número de várias plataformas de agregação e retransmissão de API ultrapassou os dois mil, formando o mercado três grandes camadas: a camada superior de grandes plataformas com tecnologia forte e modelos abrangentes; a camada média de agentes profissionais com descontos oficiais e recursos de clientes; e a vasta multidão de micro e pequenos agentes que enfrentam fortes pressões de sobrevivência. O modelo de negócio de simplesmente lucrar com as diferenças de preço dos tokens já não é sustentável.

As verdadeiras barreiras estão a deslocar-se em duas direções. Uma é a direção do prémio técnico – através do desenvolvimento de motores de aceleração de inferência proprietários, pode-se vender mais barato do que o fornecedor oficial e ainda assim ter lucro. A segunda é a direção dos serviços de valor acrescentado – fornecer serviços personalizados como engenharia de prompts, afinação fina de modelos e implantação privada. Ainda mais notável, o valor dos tokens apresenta uma diferenciação enorme baseada no cenário: o preço médio dos tokens relacionados com conversa é tão baixo quanto 0,01 dólares por milhão, enquanto para tokens relacionados com descoberta de medicamentos, o preço médio pode chegar a 1000 dólares por milhão, uma diferença de valor de cem mil vezes através dos cenários. Isto revela um princípio essencial: o valor de um token não é

determinado pelo seu custo de produção, mas pela sua utilização – o mesmo token, usado em diferentes cenários e incorporado em diferentes sistemas, gera um valor radicalmente diferente.

Mas é necessário fazer um julgamento mais importante aqui: Atualmente e no futuro, a barreira para vender tokens não é alta. Qualquer pessoa pode registrar-se para obter uma API, criar uma plataforma de agregação e começar a revender tokens. O que é verdadeiramente escasso é o que está para além da venda de tokens – dados históricos reais, a capacidade de construir sistemas e arquiteturas, e uma visão orientada para o futuro. Sem isto, todos os grandes modelos e tokens são inúteis.

Julgamento complementar sobre a rentabilidade dos tokens: Ganhar dinheiro com tokens já é difícil, e tornar-se-á cada vez mais difícil. As guerras de preços comprimem continuamente as margens de lucro, e o modelo de negócio da simples revenda de tokens caminha rapidamente para lucros reduzidos ou mesmo nulos. Só quando os tokens estão ligados a um sistema frontal – isto é, quando um sistema fornece serviços – é que os tokens e o poder de computação podem servir como ativos complementares eficazes para gerar lucros. Por outras palavras, os tokens em si não são a fonte do lucro; o sistema é. Os tokens são apenas o combustível do sistema, não o seu motor.

À medida que os tokens se tornam gradualmente uma unidade de medida importante na era da inteligência artificial, surgiram no mercado cada vez mais modelos de negócio centrados nos tokens. Além dos fornecedores de modelos fundamentais, existem também agentes de tokens, plataformas de agregação de modelos, plataformas de agentes, serviços de distribuição de API a nível empresarial e vários fornecedores de serviços de camada intermédia. Para muitos empreendedores, a agência de tokens parece um atalho para entrar na indústria da IA porque não requer treino de modelos nem construção de grandes centros de computação; só é necessário integrar recursos existentes para começar a fazer negócio. No entanto, este modelo tem vantagens e limitações óbvias. As suas vantagens incluem um investimento inicial relativamente baixo, barreiras de entrada

relativamente baixas e uma expansão de mercado mais rápida; as suas limitações incluem uma elevada dependência dos fornecedores de modelos a montante, suscetibilidade à concorrência de preços e falta de barreiras técnicas de longo prazo. À medida que os preços dos tokens continuam a cair, a margem de lucro da simples revenda de tokens corre o risco de ser continuamente comprimida. Portanto, a longo prazo, a agência de tokens é mais uma capacidade de distribuição do que uma competitividade central final. O que realmente determina o valor a longo prazo de uma empresa continua a ser os sistemas específicos da indústria, os processos de automação, os sistemas de conhecimento e as soluções do mundo real construídos em torno dos tokens.

VI. Dificuldades atuais e caminhos futuros

O atual mercado de tokens enfrenta vários problemas profundos.

A armadilha da concorrência homogênea. Mais de duas mil plataformas competem pelos mesmos clientes, e as guerras de preços continuam a reduzir os lucros da indústria. Sem mudança, isto pode repetir os primórdios da computação em nuvem, onde se "perdia dinheiro para ganhar quota de mercado".

A ameaça dos riscos de conformidade. O atual "dividendo regulatório" da ausência de legislação específica é um estado temporário, não uma norma a longo prazo. À medida que o mercado de tokens se expande, questões como dados transfronteiriços, segurança nacional e cobrança de impostos forçarão os países a acelerar a legislação. É particularmente importante notar que as tentativas de alguns atores do mercado de tokenizar e financeirizar os tokens podem facilmente desencadear a regulação financeira global, não só levando ao encerramento dessas atividades, mas também prejudicando a reputação de toda a indústria chinesa de exportação de tokens.

O mito da "Token-maxxing". Uma nova forma de "exibição de riqueza" surgiu no Vale do Silício – competir sobre quem consome mais tokens por dia. A Meta tem uma classificação interna chamada "Claudeconomics", que compila dados de

utilização de IA de mais de 85.000 funcionários, com o funcionário melhor classificado a queimar valores de tokens que chegam a vários milhões de dólares. Os críticos apontam com mordacidade: "Outcome maxxng é melhor do que token maxxng" – em vez de consumir tokens freneticamente, é melhor ver que resultados são produzidos. Esta é exatamente a questão central que compreendi profundamente ao longo de anos de prática.

Tokens chineses vs. tokens ingleses: Três barreiras sob um conceito unificado

Para os utilizadores chineses, um facto que deve ser claramente reconhecido é o seguinte: embora os tokens chineses (ciyuan) e os tokens ingleses pertençam à mesma unidade de medida fundamental na era da IA, eles não formaram um mercado global verdadeiramente unificado e livremente interoperável. De uma perspetiva conceitual técnica, o Oriente e o Ocidente usam os tokens como a unidade básica para os modelos processarem a língua e os dados; no entanto, de uma perspetiva prática, a existência combinada de barreiras tecnológicas, barreiras sistémicas e barreiras políticas criou uma separação de facto entre os tokens chineses e os tokens ingleses.

A primeira barreira é a barreira tecnológica, ou seja, diferenças na lógica de tokenização e na arquitetura do modelo.

Os modelos ocidentais dominantes como GPT, Claude e Gemini são construídos principalmente em grandes corpora predominantemente ingleses. Embora estes modelos já possuam excelentes capacidades de processamento de chinês, em muitos cenários, o chinês ainda tende a exigir mais tokens para expressar a mesma quantidade de informação que o inglês. Ainda mais importante, diferentes modelos têm diferentes estruturas de vocabulário e estratégias de tokenização. O mesmo texto em chinês pode ser segmentado em números completamente diferentes de tokens em diferentes modelos, e os seus métodos de codificação interna e de mapeamento semântico também podem diferir.

Ao mesmo tempo, os modelos chineses domésticos como DeepSeek, Tongyi Qianwen, Doubao e Wenxin Yiyan incorporaram grandes quantidades de corpora chineses desde a fase de treino, dando-lhes uma maior adaptabilidade nativa a expressões chinesas, expressões idiomáticas, nomes próprios e expressões localizadas. Portanto, mesmo perante o mesmo texto, o número de tokens, a eficiência do processamento e os caminhos de compreensão podem variar significativamente entre diferentes modelos. Neste sentido, embora o token seja um conceito unificado, não existe um "token padrão" que possa ser diretamente circulado e reutilizado entre modelos.

A segunda barreira é a barreira sistémica, ou seja, diferenças nos ecossistemas e métodos de invocação.

O atual mercado global de tokens formou na realidade múltiplos ecossistemas relativamente independentes. Plataformas internacionais como OpenAI, Anthropic e Google têm os seus próprios padrões de API, modelos de preços, regras de acesso e sistemas de serviço; o mercado chinês formou um ecossistema doméstico centrado em plataformas como DeepSeek, Tongyi Qianwen, Doubao e Kimi.

Para utilizadores comuns, comprar tokens numa plataforma não significa que os possam usar diretamente noutra plataforma. As interfaces de invocação, mecanismos de autenticação, limites de taxa, acordos de serviço e regras comerciais diferem de plataforma para plataforma. Mesmo que a conversão seja feita através de gateways de terceiros ou plataformas de agregação, está essencialmente a chamar outro sistema, não uma verdadeira troca de tokens. Portanto, embora o token seja um conceito unificado, os tokens como recurso comercial e capacidade de serviço permanecem fechados dentro dos seus respetivos ecossistemas de plataforma.

A terceira barreira é a barreira política, ou seja, diferenças nos fluxos de dados e requisitos de conformidade.

À medida que a inteligência artificial entra gradualmente na fase de aplicação empresarial, a importância das questões de segurança dos dados continua a

crescer. O sistema chinês de segurança de dados, o sistema RGPD da UE e as medidas regulatórias dos EUA sobre o fluxo transfronteiriço de dados sensíveis impõem todos requisitos diferentes sobre a transmissão e processamento de dados. Muitos dados envolvendo informações de clientes, dados financeiros, registos médicos, segredos comerciais e dados pessoais muitas vezes não podem atravessar fronteiras livremente.

Isto significa que, em ambientes comerciais reais, muitas empresas não escolhem os modelos apenas com base nos preços dos tokens, mas devem priorizar a conformidade dos dados, a localização do armazenamento de dados e a responsabilidade legal. Para muitos dados comerciais críticos reais, os custos de conformidade são frequentemente muito mais elevados do que o próprio custo do token. Portanto, no mundo real, as empresas domésticas tendem a usar modelos domésticos conformes, enquanto as empresas estrangeiras tendem a usar modelos estrangeiros, resultando numa segmentação de facto do mercado de tokens.

As barreiras tecnológicas determinam as diferenças nos métodos de segmentação de tokens; as barreiras sistémicas determinam as diferenças nos sistemas de invocação e ecossistemas comerciais; as barreiras políticas determinam as diferenças nos fluxos de dados e fronteiras de conformidade. Estas três barreiras a atuar em conjunto significam que, embora os tokens chineses e os tokens ingleses pertençam ao mesmo conceito técnico, não podem ser livremente trocados, livremente circulados e livremente interligados como uma idealizada "moeda universal global de IA".

No futuro, é provável que surja uma situação semelhante ao comércio internacional: o ecossistema de tokens chinês e o ecossistema de tokens inglês coexistirão por muito tempo, cada um seguindo caminhos de desenvolvimento independentes. A invocação interzona não é impossível, mas muitas vezes requer adaptações técnicas adicionais, conversão de sistemas e custos de revisão legal. Para a maioria dos utilizadores comuns, esta capacidade interzona não é uma necessidade diária, portanto, os dois grandes sistemas provavelmente permanecerão relativamente independentes por muito tempo.

Os tokens produzidos na China não estão necessariamente confinados ao uso dentro dos sistemas chineses, mas em cenários-chave que envolvem soberania de dados, responsabilidade de conformidade e segurança industrial, é provável que grande parte da procura futura de tokens seja consumida preferencialmente dentro de modelos, plataformas e sistemas domésticos.

Numa perspetiva de tendência de longo prazo, o acoplamento entre a localização da produção de poder de computação, o fornecedor do modelo, o operador do sistema e a localização dos dados está a aumentar. Para as finanças, administração pública, saúde, cadeias de abastecimento e grandes sistemas empresariais, os dados muitas vezes não podem fluir livremente através das fronteiras. Portanto, mesmo que a China possa exportar grandes quantidades de serviços de tokens globalmente, grande parte da procura de tokens de alto valor permanecerá preferencialmente dentro de modelos e sistemas chineses para ser completada. Isto não significa que os tokens chineses não possam ser exportados, mas que o futuro mercado de tokens pode apresentar simultaneamente duas estruturas: uma parte de "circulação global" e uma parte de "circuito fechado doméstico".

Vale a pena notar que o que pode realmente alcançar a unificação no futuro não são os próprios tokens, mas a forma como os tokens são invocados. À medida que as plataformas de agentes, os protocolos de interface unificados, as plataformas de agregação de modelos e os sistemas de automação continuarem a desenvolver-se, os utilizadores acabarão por se concentrar não em qual token foi usado, mas em se a tarefa foi concluída, se o custo é razoável e se a eficiência é suficiente. Nessa altura, os tokens, como a eletricidade, o tráfego de Internet e os recursos de computação em nuvem hoje, tornar-se-ão gradualmente escondidos atrás dos sistemas, tornando-se a infraestrutura operacional fundamental da civilização digital.

Portanto, numa perspetiva de longo prazo, embora os mercados de tokens orientais e ocidentais sejam temporariamente difíceis de interligar diretamente, as suas direções de desenvolvimento estão altamente alinhadas. O passado dos tokens pertence ao processamento computacional da linguagem, o presente dos tokens

pertence à indústria da IA, e o futuro dos tokens provavelmente pertencerá ao sistema operacional fundamental de toda a civilização digital.

VII. Porque não é possível adquirir diretamente a capacidade de construção de sistemas através de grandes quantidades de tokens

Ao argumentar que "a capacidade de construção de sistemas é um ativo escasso", preciso de enfrentar uma possível objeção: uma vez que a IA é tão poderosa, porque não se pode "gerar" ou "obter" a capacidade de construção de sistemas consumindo grandes quantidades de tokens? A resposta a esta questão revela os limites fundamentais da capacidade dos tokens.

Primeiro, a construção de sistemas requer decisões, e as decisões não são geração.

A IA pode gerar um artigo, um código, um modelo de sítio web, mas a IA não pode substituir as decisões centrais no design de sistemas que os humanos devem tomar – escolher quais módulos, abandonar quais funções, equilibrar as necessidades atuais com a expansão futura, quando reestruturar, quando manter o status quo. Estas decisões dependem da compreensão da essência do negócio, do julgamento das restrições de recursos e da previsão das consequências a longo prazo. Os tokens podem gerar inúmeras "soluções possíveis", mas escolher "qual solução vale a pena implementar" e "porquê" é algo que os tokens não podem substituir. O meu sistema logístico inteligente passou por inúmeros ajustes ao longo de duas décadas, e cada ajuste não foi porque uma IA me disse "isto é melhor", mas porque um problema real surgiu no negócio real, expondo limitações e levantando novas exigências. Por detrás destas decisões está o julgamento, e o julgamento vem da prática, não dos tokens.

Segundo, a construção de sistemas requer o tratamento de exceções, e as exceções não podem ser enumeradas exaustivamente.

A razão pela qual um sistema pode funcionar a longo prazo não é porque o seu fluxo normal está perfeitamente concebido, mas porque pode lidar com exceções. Num sistema logístico, as exceções incluem: endereços de clientes incorretos, mercadorias danificadas, atrasos alfandegários, falhas de pagamento, quedas do sistema, perda de dados. Num sistema de edição, as exceções incluem: flutuação na qualidade da tradução, formatos de metadados incompatíveis, alterações nas interfaces de arquivo, falhas de registo de DOI. Estas exceções não podem ser enumeradas exaustivamente porque a variedade de exceções no mundo real é infinita. A IA pode aprender "padrões de exceção comuns" através de grandes quantidades de tokens, mas não pode prever "a próxima exceção que nunca apareceu antes". Apenas um sistema de longa duração, encontrando, registando e resolvendo continuamente exceções na prática, pode desenvolver uma verdadeira capacidade de tratamento de exceções. Esta capacidade não pode ser obtida "alimentando mais tokens", porque as exceções não são conhecimento; são experiência.

Terceiro, a construção de sistemas requer compromissos, e os compromissos não têm respostas padrão.

A construção de cada sistema envolve compromissos sob múltiplas restrições: custo vs. desempenho, velocidade vs. estabilidade, generalidade vs. personalização, necessidades imediatas vs. expansão futura. Estes compromissos não têm "resposta correta", apenas "escolhas razoáveis em circunstâncias específicas". A IA pode fornecer uma "solução ótima", mas este "ótimo" é estatisticamente ótimo com base em dados históricos, não um julgamento baseado no "seu contexto específico". A razão pela qual o meu sistema integrado de logística e finanças de duas décadas funcionou até hoje é precisamente porque fiz diferentes compromissos em diferentes épocas – escolhendo uma determinada pilha tecnológica em 2005, escolhendo outro método de integração em 2015, e escolhendo um caminho de atualização em 2025. Estes compromissos não eram "certos" ou "errados", mas "adequados ao contexto da época". Os tokens não

podem fazer estes compromissos por mim, porque a essência dos compromissos é julgamento de valor, não cálculo.

Quarto, a construção de sistemas requer ciclos de retroalimentação, e a retroalimentação requer o mundo real.

A melhoria da capacidade de construção de sistemas depende de um ciclo fechado: design → implementação → operação → detecção de problemas → correção → redesenho. O passo de "detecção de problemas" neste ciclo deve vir da retroalimentação do mundo real. A IA pode simular a operação, mas não pode simular todos os fatores imprevisíveis do mundo real. A razão pela qual o meu sistema logístico melhorou continuamente é que clientes reais forneceram feedback real; a razão pela qual o meu sistema de edição melhorou continuamente é que interfaces de arquivo reais sofreram alterações reais. Esta retroalimentação não pode ser obtida consumindo tokens num ambiente fechado. Os tokens podem gerar "feedback hipotético do cliente", mas não podem substituir o facto de que "o cliente realmente deu este feedback". A retroalimentação do mundo real é o único solo para o crescimento da capacidade de construção de sistemas.

Quinto, a construção de sistemas requer confiança, e a confiança só pode vir do tempo.

Se um sistema é digno de confiança não se prova pela promoção, mas pelo tempo. A razão pela qual o meu sítio web é permanentemente arquivado pelo CERN Zenodo, WorldCat e Biblioteca Nacional da Austrália não é porque eles "gostam de mim", mas porque estas instituições, através da observação a longo prazo, confirmaram que o meu sistema é estável, fiável e persistente. Esta confiança não pode ser comprada com grandes quantidades de tokens. A IA pode gerar um manual de sistema "de aspeto muito profissional", mas não pode gerar o facto de que "este sistema é confiável há duas décadas". A confiança é um produto do tempo, não um produto dos tokens.

Em resumo, a razão pela qual a capacidade de construção de sistemas não pode ser adquirida diretamente através de grandes quantidades de tokens é que ela não

é essencialmente uma capacidade generativa, mas sim uma capacidade de julgamento, tratamento de exceções, compromisso, absorção de retroalimentação e acumulação de confiança. O que estas capacidades têm em comum é que exigem tempo do mundo real, prática, erros, correções e validação. Os tokens podem acelerar o processamento de informação, mas não podem acelerar a acumulação de experiência do mundo real. Esta é a razão fundamental pela qual, mesmo que os tokens se tornem extremamente baratos no futuro, a capacidade de construção de sistemas continuará escassa.

VIII. Os meus ativos: Dados históricos reais, sistemas interdisciplinares, capacidade de construção, visão prospetiva e capacidade de autoverificação

Antes de discutir mais a fundo "o que é verdadeiramente escasso", preciso de explicar primeiro a minha própria posição. Porque os meus julgamentos não derivam da dedução teórica, mas da prática de longo prazo. Isto é muito evidente nos vários sistemas que construí ao longo de um longo período. Quer se trate do sistema logístico inteligente, do sistema de faturação inteligente, do sistema de processamento financeiro automatizado, ou mais tarde do sistema de edição em dez línguas, do sistema de metadados, do sistema de navegação do sítio web e do sistema de referência cruzada automática, a parte realmente difícil nunca foi gerar informação, mas sim organizá-la. A informação pode crescer indefinidamente, mas sem estrutura, apenas criará um novo caos. O mesmo se aplica na era dos tokens. No futuro, a IA poderá gerar quantidades maciças de conteúdo, mas quantidades maciças de conteúdo não significam automaticamente quantidades maciças de valor. Pelo contrário, à medida que a geração de conteúdo se torna mais fácil, saber como filtrar, organizar, preservar, indexar e associar este conteúdo tornar-se-á cada vez mais importante.

Possuo os seguintes cinco ativos, que constituem a base empírica para o meu julgamento sobre "o que é escasso depois dos tokens".

Primeiro, dados históricos reais. Estes não são uns "dados históricos" quaisquer, mas vinte anos de dados comerciais reais transgeracionais – desde o sistema de faturação inteligente de 2005 até ao presente, de sistemas logísticos, financeiros e fiscais reais em funcionamento, não de ambientes experimentais ou de dados raspados. Na era da IA, com a proliferação de "dados sintéticos", o valor dos dados históricos reais está a aumentar drasticamente. A maioria das empresas possui "dados históricos de domínio único", enquanto eu possuo dados históricos interdisciplinares, que abrangem vinte anos, e praticamente validados.

Segundo, a conclusão de múltiplos sistemas e arquiteturas interdisciplinares. Os meus sistemas cobrem os seguintes domínios: sistema logístico inteligente, sistema de faturação inteligente, sistema financeiro inteligente, sistema fiscal inteligente, sistema de integração logística e financeira de duas décadas, sistema de automatização de sítios web, sistema de geração de metadados, sistema de edição em dez línguas do "The Epochal Transition", sistema de gestão de DOI, sistema de organização de links TROVE e WorldCat, sistema de referência cruzada automática, sistema de geração de Sitemap, sistema de geração de perfil Zenodo, sistema de edição multiplataforma. A singularidade destes sistemas reside na: amplitude interdisciplinar (da logística à edição, das finanças à indexação de literatura), profundidade temporal (alguns sistemas estão em funcionamento há duas décadas) e conclusão por uma pessoa (demonstrando a transferibilidade do método). Entre estes, o mensal eletrónico "The Epochal Transition" (<https://times.net.au>) serve como prova pública – possui o número ISSN internacional normalizado (3083-5178), está permanentemente arquivado pela WorldCat (o maior sistema de catálogo coletivo de bibliotecas do mundo, com instituições membro excedendo 16.000 bibliotecas nacionais e universitárias em mais de 100 países e regiões), pelo PANDORA/TROVE da Biblioteca Nacional da Austrália, bem como pelo Zenodo operado pelo CERN (Organização Europeia para a Investigação Nuclear), pelo OpenAIRE (Infraestrutura Europeia para a Ciência Aberta), pelo DataCite DOI (identificador de objeto digital global, com mais de 60 documentos), e pelo ORCID (identificador digital académico permanente único global). Este sítio web prova que uma única pessoa (um investigador independente) pode completar todo o processo,

desde a criação de conteúdo até à interface com as infraestruturas de arquivo acadêmico de topo.

Terceiro, a capacidade de construir sistemas e arquiteturas. Esta não é a capacidade de "usar ferramentas de IA", mas a capacidade de design para transformar requisitos vagos em estruturas estáveis, a capacidade arquitetural para integrar diferentes sistemas numa arquitetura unificada, a capacidade de engenharia para manter sistemas em funcionamento durante anos sem intervenção humana, e a meta-capacidade para transferir experiência da logística para a edição, das finanças para a literatura. Como discutido anteriormente, esta capacidade não pode ser adquirida diretamente através de grandes quantidades de tokens porque depende da tomada de decisões, tratamento de exceções, compromissos, ciclos de retroalimentação e acumulação de confiança – nenhum dos quais pode ser substituído pela IA. O atual mercado de formação em IA concentra-se em competências superficiais como "engenharia de prompts", "escrita com IA" e "desenho com IA"; quase nenhum curso ensina a "capacidade de construção de sistemas interdisciplinares" – porque esta capacidade não pode ser abreviada, só pode ser acumulada através da prática de longo prazo. As mais de 40 obras criativas independentes listadas no sítio web, o ISSN obtido e os mais de 60 DOIs são os melhores registos públicos da capacidade de funcionamento a longo prazo do sistema, e não uma autoproclamação minha.

Quarto, uma visão prospetiva. Começar a experimentar um sistema protótipo de trabalho remoto em 1993 – quase trinta anos antes do trabalho remoto em grande escala se tornar comum; conceber um sistema de faturação inteligente em 2005 – começar a automação de processos vinte anos antes da proliferação da IA; a integração transgeracional de duas décadas – prever no momento do design do sistema que ele ainda precisaria de funcionar vinte anos depois; apresentar agora o julgamento "depois dos tokens" – numa altura em que a maioria das pessoas ainda persegue tokens, prever que a capacidade do sistema é a verdadeira barreira. Também prevê que "os tokens são semelhantes a moedas, e as diferenças tecnológicas, sistémicas e políticas entre o Oriente e o Ocidente os tornam não

universais" – tal como o dólar americano e o renminbi chinês não podem ser livremente trocados, os tokens chineses e os tokens ingleses enfrentam barreiras de circulação semelhantes, e no futuro, podem surgir "zonas de tokens chineses", "zonas de tokens ingleses" e outros ecossistemas de tokens baseados em língua e política. Também previ que "ganhar dinheiro com tokens já é difícil e tornar-se-á cada vez mais difícil; só quando há um sistema frontal que fornece serviços é que os tokens e o poder de computação podem servir como ativos complementares eficazes para gerar lucros". Como investigador independente, estabeleci antecipadamente um sistema de edição pessoal que se interfaceia com infraestruturas académicas futuras, prevendo a tendência de que "um indivíduo também pode tornar-se um nó na infraestrutura global de conhecimento". A maioria das previsões de tendências vem de analistas ou empresários, enquanto a minha "capacidade de previsão" tem a validação de duas décadas de prática como garantia.

Quinto, capacidade de autoverificação. Esta é uma das minhas meta-capacidades mais importantes – todos os escritos que produzo em vários campos alcançam um fecho teórico e uma autocoerência lógica. Isto significa que os meus argumentos não dependem de autoridade externa ou citações, mas da autossuficiência da lógica interna. A regra padrão na literatura académica é "citar investigações anteriores"; poucas pessoas reivindicam ou demonstram "capacidade de autoverificação"; a discussão académica centra-se em "se a argumentação é suficiente", e não em "se o autor tem a capacidade de construir um ciclo fechado coerente". Nos meus artigos, esta capacidade de autoverificação manifesta-se em: os argumentos vêm da prática, a prática valida os argumentos, e os argumentos e a prática formam um ciclo fechado. Não estou a "citar as teorias de outros para me provar", mas a "mostrar os sistemas que construí e deixar que os sistemas falem por si próprios". Esta capacidade de autoverificação é um conceito completamente ausente na literatura pública e uma das minhas meta-capacidades mais únicas.

A combinação destes cinco ativos, a sua escassez não é aditiva, mas multiplicativa: dados históricos reais × sistemas interdisciplinares × capacidade de construção de

sistemas × visão prospetiva × capacidade de autoverificação constitui uma carteira de ativos irreproduzível. Em discussões públicas, não é atualmente possível encontrar nenhum caso comparável.

O processo de sistematização de "O Caminho de um Investigador Independente Internacional" fez-me perceber isto ainda mais profundamente. De 1989 a 2026, mais de trinta anos de experiência abrangem múltiplos campos. Estas experiências não surgiram todas de uma vez, mas formaram-se gradualmente ao longo de um longo período. A IA pode gerar um artigo em segundos, mas não pode gerar mais de trinta anos de experiência real em segundos; pode gerar muitas opiniões, mas não pode gerar diretamente o julgamento acumulado através da prática de longo prazo; pode simular muitos processos, mas não pode substituir a validação repetida do mundo real. Neste sentido, o tempo em si é também um ativo importante, e a estrutura precipitada pelo tempo é um ativo ainda mais raro.

IX. Depois dos tokens: Cinco capacidades verdadeiramente escassas

Quando os próprios tokens já não são escassos – a barreira para vender tokens não é alta, mas haverá uma falta extrema de dados históricos reais, da capacidade de construir sistemas e arquiteturas, e de uma visão prospetiva – o que será verdadeiramente escasso? Com base na minha prática de longo prazo, desde o sistema protótipo de trabalho remoto de 1993, o sistema de faturação inteligente de 2005, o sistema logístico inteligente, o sistema financeiro inteligente, o sistema fiscal inteligente, o sistema de integração logística e financeira de duas décadas, até à automatização de sítios web, geração de metadados, o mensal em dez línguas "The Epochal Transition", gestão de DOI, organização de links TROVE e WorldCat, referência cruzada automática, geração de Sitemap, geração de perfil Zenodo, edição multiplataforma, acredito que as seguintes cinco capacidades se tornarão verdadeiras barreiras.

Primeiro, capacidade de design estrutural. Os tokens são a matéria-prima, e a estrutura é o quadro que organiza a matéria-prima em produtos valiosos. Muito antes da IA se generalizar, eu já tinha começado a pensar em como usar recursos limitados para alcançar uma maior eficiência operacional. O meu sistema logístico inteligente, o meu sistema de faturação inteligente e o meu sistema logístico e financeiro inteligente de duas décadas nasceram em diferentes épocas, enfrentando diferentes problemas, mas partilham uma característica comum: o objetivo não era criar mais informação, mas reduzir o caos, reduzir o trabalho repetitivo, melhorar a eficiência global e estabelecer estruturas que pudessem funcionar a longo prazo. No sistema de edição automatizado, o problema central que enfrentei não era "se o texto podia ser gerado", mas "como definir campos de metadados para dez línguas, como estabelecer associações entre artigos e categorias, como conceber regras de gestão de versões". Estas decisões de design estrutural determinam se o sistema pode continuar a funcionar durante anos sem colapsar. Em contraste, a quantidade de tokens consumidos é meramente um subproduto. O sítio web "The Epochal Transition" é em si uma prova pública da capacidade de design estrutural: uma estrutura de sítio web académico mantida pessoalmente, com um ISSN internacional normalizado, suporte multilíngue, múltiplos canais de arquivo, provando que não estou a fazer discursos vazios, mas já construí com sucesso um sistema complexo e publicamente acessível usando esta capacidade de design estrutural.

Segundo, capacidade de automação. Os tokens podem produzir conteúdo em massa, mas apenas sistemas de automação podem concluir fluxos de trabalho de ponta a ponta em massa. No sistema logístico inteligente, o valor real não reside em quantas vezes a IA é chamada para planear rotas, mas em estabelecer um ciclo fechado totalmente automatizado, desde a receção de pedidos, planeamento de armazém, otimização de rotas, até ao tratamento de exceções. Neste ciclo fechado, a IA é apenas um componente, e a capacidade de automação manifesta-se nas ligações entre os componentes. Uma pessoa sem sistema, mesmo com uma grande quantidade de tokens, pode apenas gerar texto disperso, imagens dispersas, tabelas dispersas e código disperso; uma pessoa com um sistema, por outro lado,

pode integrar os tokens no seu fluxo de trabalho, permitindo-lhes organizar automaticamente materiais, gerar metadados, verificar links, atualizar páginas web, ajudar na edição, arquivar literatura, criar índices, processar logística, analisar finanças e gerar relatórios.

Terceiro, experiência prática interdisciplinar. Os tokens estão disponíveis para todos, mas a capacidade de transferir experiência de uma indústria para outra não é universal. A razão pela qual posso construir tanto um sistema logístico como um sistema de edição não é porque tenho mais tokens, mas porque acumulei metodologias transferíveis como gestão de metadados, design de fluxo de trabalho e tratamento de exceções. O meu sistema de edição precisa de tokens porque a tradução para dez línguas, a geração de resumos, a organização de palavras-chave, a geração de metadados e as descrições de links requerem todo um processamento linguístico extenso. O meu sistema de sítio web precisa de tokens porque estruturas de página, descrições de navegação, descrições SEO, metadados académicos e apresentações multilingues podem ser todos gerados automaticamente. O meu sistema logístico também pode usar tokens porque os e-mails de clientes, descrições de mercadorias, documentos alfandegários, explicações de exceções, notas financeiras e registos de operações podem ser processados com assistência de IA. O meu sistema de referência cruzada automática precisa ainda mais de tokens porque tem de encontrar associações temáticas entre um grande número de artigos, títulos, DOIs, links NLA, TROVE e WorldCat. Esta acumulação prática interdisciplinar não pode ser obtida consumindo tokens; só pode ser sedimentada resolvendo problemas reais. O sítio web apresenta simultaneamente informações de múltiplos domínios, incluindo edição académica, integração de infraestruturas, arquitetura técnica e marca pessoal, provando que liguei efetivamente a edição, a biblioteconomia, a tecnologia e a marca pessoal, em vez de possuir uma única competência.

Quarto, o modelo mental para o funcionamento a longo prazo de sistemas. A narrativa dominante para as aplicações atuais da IA é a "conversa" e a "geração", mas o que realmente produz valor sustentado é o "funcionamento". Uma conversa

gera 10.000 tokens, uma tarefa de agente consome um milhão, mas um sistema logístico que funciona continuamente durante uma década pode consumir milhares de milhões de tokens – mas o que é verdadeiramente escasso é o modelo mental para conceber este sistema: como garantir que o sistema permanece utilizável após atualizações do modelo? Como lidar com mudanças nas APIs a montante? Como monitorizar exceções e recuperar automaticamente? Estas questões não têm nada a ver com tokens, mas determinam o resultado final dos tokens. O valor destes sistemas não vem dos próprios tokens, mas das estruturas que já construí ou estou a desenvolver. Os tokens são apenas o combustível; o sistema é o motor. Os tokens são apenas a corrente elétrica; a estrutura é o circuito. Os tokens são apenas a água; os tubos, válvulas, reservatórios e sistemas de despacho determinam se podem desempenhar um papel. Sem um sistema, os tokens dissipar-se-ão; com um sistema, os tokens podem ser convertidos em valor a longo prazo. O sítio web mostra mais de 40 obras criativas independentes e o arquivamento permanente por instituições autorizadas como o CERN e a Biblioteca Nacional da Austrália, provando que o meu sistema não procura hype de curto prazo, mas valor a longo prazo, e já recebeu o aval de "preservação permanente" de instituições autorizadas.

No entanto, quando considero mais a fundo a relação entre tokens e sistemas, encontro outra questão frequentemente ignorada. Muitas pessoas acreditam que, com um excelente sistema, poderão obter vantagem na futura era dos tokens. Mas a realidade não é tão simples. Os sistemas em si não são algo que qualquer pessoa possa conceber. Sistemas verdadeiramente maduros exigem frequentemente prática de longo prazo, revisão repetida e validação no mundo real. Muitos sistemas podem levar anos, ou mesmo décadas, de acumulação contínua, desde a ideia inicial até à implementação efetiva. Portanto, a capacidade de conceber sistemas é já um recurso escasso.

Mas mesmo possuindo um sistema maduro, isso não leva automaticamente a um impacto em grande escala. Entre a viabilidade técnica de um sistema e a sua difusão na sociedade existem múltiplas camadas, incluindo mercados, capital, canais de distribuição, educação dos utilizadores e ecossistemas de plataforma. A

história não carece de boas tecnologias ou bons sistemas; o que é verdadeiramente escasso é a capacidade da plataforma para levar um sistema a uma aplicação em grande escala. Mesmo que um sistema seja muito superior às soluções existentes num cenário local, sem apoio de capital suficiente, capacidade de promoção da plataforma e base de utilizadores, a sua escala de utilização real pode ser muito limitada.

Da perspetiva da economia dos tokens, isto é particularmente evidente. A procura de tokens não vem diretamente do sistema em si, mas da escala real de utilização do sistema. Uma plataforma que serve milhões de utilizadores por dia, mesmo com um design de sistema relativamente comum, pode gerar um consumo enorme de tokens; inversamente, um sistema avançado com muito poucos utilizadores pode ter uma procura de tokens muito baixa. Portanto, o desenvolvimento futuro do mercado de tokens depende não apenas da capacidade dos modelos e da capacidade dos sistemas, mas também da capacidade das plataformas e da capacidade de integração de recursos.

Isto significa que a concorrência futura tem pelo menos três níveis. O primeiro nível é a concorrência dos modelos, determinando quem pode fornecer tokens. O segundo nível é a concorrência dos sistemas, determinando quem pode usar os tokens eficazmente. O terceiro nível é a concorrência das plataformas, determinando quantos utilizadores um sistema pode alcançar e quanta procura real pode gerar. Modelos, sistemas e plataformas são todos indispensáveis. Sem modelos, os sistemas não podem funcionar. Sem sistemas, os tokens dificilmente criam valor. Sem plataformas, mesmo o melhor sistema pode permanecer numa escala de utilização restrita.

Desta perspetiva, fico ainda mais convencido de que os tokens são meramente recursos fundamentais. O que realmente determina o futuro não é apenas aqueles que possuem tokens, nem apenas aqueles que possuem sistemas, mas aqueles que podem ligar modelos, sistemas, plataformas e necessidades do mundo real. Só quando estes elementos formam um ciclo fechado é que os tokens se transformam de meras unidades de cálculo em forças produtivas que criam valor continuamente.

Quinto, a capacidade de incorporar tokens em processos de negócio do mundo real. Os tokens devem, em última análise, servir um elo no mundo real – logística, edição, educação, saúde. O que é verdadeiramente escasso são aqueles que entendem tanto os limites das capacidades da IA quanto os pontos críticos de uma indústria específica. Sabem quando usar a IA e quando não usar; sabem como converter as saídas dos tokens em decisões de negócio; sabem como medir a relação real entre os inputs de tokens e os outputs de negócio. Esta capacidade composta é atualmente o recurso mais escasso no mercado. O sítio web é em si um "processo de negócio" em funcionamento: criação de conteúdo → geração de metadados → publicação multiplataforma → arquivo internacional → acesso do leitor. Isto prova que a minha aplicação de IA (consumo de tokens) foi incorporada num processo de edição completo, em ciclo fechado, que gera impacto real.

Sinto cada vez mais que a verdadeira concorrência no futuro provavelmente não será entre modelos, nem entre tokens, mas entre sistemas. Os modelos continuarão a evoluir, os preços dos tokens continuarão a cair, e os recursos caros hoje tornar-se-ão muito comuns no futuro. Mas os sistemas não aparecerão automaticamente porque os recursos são abundantes. Os sistemas exigem acumulação a longo prazo, revisão contínua e validação no mundo real. Quem conseguir converter tokens em sistemas de longa duração terá maior probabilidade de obter valor sustentado. E esta é também a chave da rentabilidade: ganhar dinheiro com tokens já é difícil, e tornar-se-á cada vez mais difícil; só quando há um sistema frontal que fornece serviços é que os tokens e o poder de computação podem servir como ativos complementares eficazes para gerar lucros.

Julgo que a verdadeira concorrência na era dos tokens se desenrolará em três fases: a primeira fase é a concorrência de recursos, com todos a disputar modelos, poder de computação, APIs e preços de tokens; a segunda fase é a concorrência de aplicações, com todos a usar tokens para escrever, serviço ao cliente, código, tradução, educação, trabalho de escritório e análise de dados; a terceira fase é a verdadeira concorrência de sistemas, onde quem conseguir incorporar tokens em sistemas industriais, sistemas de edição, sistemas logísticos, sistemas de

conhecimento e sistemas de registo da civilização que funcionem a longo prazo terá probabilidade de obter vantagem duradoura.

De facto, ao discutir tokens, modelos, sistemas e plataformas, há outro fator importante frequentemente ignorado no mundo real: o tempo. Muitas pessoas pensam que com capital, podem construir sistemas rapidamente; com um mercado suficientemente grande, podem replicar rapidamente modelos de sucesso. No entanto, a realidade muitas vezes não é assim. O capital pode comprar poder de computação, pode adquirir tokens, pode contratar equipas de desenvolvimento, mas não pode comprar diretamente dados reais e sistemas maduros que foram validados ao longo de muitos anos. Para muitas indústrias, o que é verdadeiramente escasso não são os modelos em si, nem os tokens em si, mas os sistemas que têm funcionado de forma estável em ambientes de negócio reais durante um longo período.

Numa perspetiva comercial, os mercados e o capital normalmente não esperam muito tempo que um sistema completamente novo cresça a partir do zero. Um grande sistema, desde a proposta do conceito, design da arquitetura, desenvolvimento de funções, validação de processos até à implementação efetiva, leva frequentemente muitos anos ou até mais. Durante este processo, não só são necessários investimentos contínuos de recursos, mas também muita tentativa e erro, revisões e validação em ambiente real. Portanto, quando as oportunidades de mercado já surgiram, o capital tende a preferir sistemas que já estão formados, já estão em funcionamento e já foram validados, em vez de conceber um sistema inteiramente novo a partir do zero.

No entanto, há outra contradição na realidade. Quanto maior é um sistema, mais difícil é para o capital posterior intervir facilmente. Após desenvolvimento a longo prazo, os grandes sistemas geralmente já formaram os seus próprios mecanismos de funcionamento, estruturas de gestão, lógicas de dados e culturas organizacionais. Podem ter vastos dados históricos, processos de negócio complexos e sistemas de experiência acumulada a longo prazo. Devido a isto, embora os grandes sistemas sejam extremamente valiosos, é frequentemente difícil

transformá-los rapidamente ou reestruturá-los completamente de acordo com as expectativas do capital. Quanto maior o sistema, maior tende a ser a sua inércia.

Portanto, no mundo real dos negócios, o estado ideal não é conceber um sistema a partir do zero, nem adquirir um grande sistema rígido, mas encontrar sistemas que já completaram a sua construção estrutural central, que foram validados a longo prazo e que podem funcionar de forma estável. Mesmo que estes sistemas não sejam perfeitos, têm um valor muito alto porque já ultrapassaram a fase mais difícil, possuindo dados reais, processos reais e registos de validação reais. Com base nisto, o reempacotamento, a otimização, a abertura ao exterior e a promoção por plataformas são frequentemente mais eficientes do que começar do zero, e também podem reduzir significativamente os riscos de tentativa e erro e os custos de tempo. Porque estes sistemas já completaram a fase mais cara, mais demorada e mais difícil de replicar.

Desta perspetiva, a janela temporal é também um recurso frequentemente ignorado. Quando uma indústria ainda está na sua infância, muitas pessoas concentram-se nas capacidades dos modelos, nos preços dos tokens e no tamanho do mercado, ignorando o tempo necessário para a formação dos sistemas. O nascimento de um sistema maduro passa tipicamente por várias fases: proposta do conceito, design da arquitetura, validação dos processos, funcionamento no mundo real e otimização contínua. A parte realmente difícil não é propor a ideia, mas manter o sistema a funcionar de forma estável num ambiente real a longo prazo. Portanto, quando o mercado começa a reconhecer o valor dos sistemas maduros, significa frequentemente que estes sistemas já passaram por anos de acumulação, em vez de serem algo que pode ser replicado novamente em pouco tempo.

Para o capital e o mercado, os alvos mais ideais não são ideias na fase conceptual, nem protótipos que ainda não foram validados, mas os sistemas embrionários que completaram a construção estrutural central, possuem registos de funcionamento reais e têm margem para melhorias. Embora estes sistemas possam não ser perfeitos, já ultrapassaram a fase mais difícil, mais demorada e mais arriscada. Assim que a procura do mercado cresce rapidamente, sistemas como estes tornam-

se frequentemente muito procurados, porque o número de sistemas maduros realmente validados a longo prazo é geralmente muito menor do que o mercado imagina. Os modelos podem iterar rapidamente, os tokens podem expandir-se continuamente, as equipas de desenvolvimento podem ser montadas temporariamente, mas os dados reais, os processos reais e os registos de validação a longo prazo não podem ser recriados em pouco tempo.

Portanto, no desenvolvimento contínuo da economia dos tokens, o tempo não só afeta a formação dos sistemas, mas também afeta o valor dos sistemas. Quando o mercado ainda não reconheceu plenamente a importância de um determinado tipo de sistema, é frequentemente também o período em que o custo de aquisição desses sistemas é mais baixo; e quando o mercado geralmente reconhece o seu valor, as opções disponíveis são frequentemente já muito limitadas. Para muitas indústrias, o que se perde não é apenas um projeto, mas uma janela temporal para entrar na próxima fase de desenvolvimento que já completou uma acumulação de longo prazo.

Da perspetiva da economia dos tokens, o mesmo se aplica. Os tokens podem ser obtidos rapidamente, os modelos podem ser acedidos rapidamente, as plataformas podem ser construídas rapidamente, mas o que é realmente difícil de replicar rapidamente são os sistemas que sedimentaram durante muitos anos e podem continuar a produzir valor. No futuro, quando os tokens se tornarem cada vez mais abundantes, a importância dos sistemas não só não diminuirá, mas pode até aumentar ainda mais. Porque quando todos tiverem recursos de tokens semelhantes, o que determinará o resultado da concorrência já não será a quantidade de tokens, mas quem possui dados reais mais maduros, processos de negócio mais estáveis e estruturas de sistema mais próximas das necessidades do mundo real.

Outra questão frequentemente ignorada é a origem dos sistemas. Muitas pessoas estão habituadas a entender os sistemas como produtos de software e, portanto, assumem naturalmente que os sistemas devem ser concebidos e desenvolvidos por empresas de software. No entanto, em muitas indústrias complexas, os sistemas

que realmente funcionam de forma estável a longo prazo não são frequentemente primeiro concebidos por empresas de software e depois entregues à indústria, mas são gradualmente formados no curso da prática de negócio de longa duração, e depois transformados em software, automatizados e transformados em plataformas.

A razão não é complexa. As empresas de software geralmente destacam-se na implementação técnica, mas não permanecem necessariamente no contexto de negócio específico por muito tempo. Podem desenvolver funções rapidamente, mas têm dificuldade em adquirir grandes quantidades de experiência de negócio real em pouco tempo. Em contraste, os sistemas que se desenvolvem organicamente dentro de uma indústria são continuamente revistos e melhorados diariamente face a clientes, encomendas, exceções, mudanças regulatórias, pressões de custos e concorrência de mercado. Estes sistemas não partem da teoria, mas de problemas reais; cada um dos seus processos, cada uma das suas regras, cada elo de automação, provém tipicamente da validação de longo prazo do trabalho real.

Portanto, na logística, cadeias de abastecimento, edição, finanças, manufatura, saúde e muitos outros campos profissionais, os sistemas mais valiosos não são frequentemente os sistemas mais complexos, nem os sistemas com mais funções, mas aqueles que foram testados a longo prazo e podem resolver problemas reais de forma estável. Podem não ter as interfaces mais avançadas, a promoção mais chamativa ou a arquitetura técnica mais recente, mas possuem grandes quantidades de dados reais, processos reais e lógica de negócio acumulada a longo prazo. Esta experiência muitas vezes não pode ser obtida através de desenvolvimento de curto prazo.

Da perspetiva da economia dos tokens, isto é particularmente importante. No futuro, as capacidades dos modelos podem continuar a melhorar, os custos dos tokens podem continuar a diminuir, e a eficiência do desenvolvimento de software será cada vez maior. No entanto, a experiência de negócio acumulada por uma indústria ao longo de décadas não surgirá automaticamente devido a atualizações dos modelos. O que é realmente difícil de replicar não é o código do software em si, mas o conhecimento do setor, a experiência no tratamento de exceções, a lógica de

tomada de decisão e os processos de negócio validados a longo prazo escondidos por trás do sistema.

Portanto, acredito que os sistemas mais ideais no futuro não serão aqueles concebidos completamente desligados da prática industrial, mas aqueles que provêm do negócio real, foram validados através de funcionamento a longo prazo, e são então continuamente melhorados com a ajuda da inteligência artificial e das capacidades dos tokens. Estes sistemas possuem tanto a fiabilidade do mundo real como os ganhos de eficiência trazidos pela era da IA, e o seu valor a longo prazo é frequentemente muito superior ao daqueles sistemas construídos apenas sobre conceitos técnicos.

X. Para onde fluem os tokens: Das conversas transitórias à memória da civilização

Em relação à futura economia dos tokens, tenho outra observação. Muitas pessoas concentram-se em como os tokens são produzidos, mas pouca atenção é dada ao seu destino final. Se um token permanecer apenas numa única conversa, a sua vida útil pode ser de apenas alguns minutos; se entrar numa página web comum, a sua vida útil pode ser de apenas alguns meses; se entrar num sistema de preservação a longo prazo, o seu valor muda fundamentalmente. Quando os tokens entram em DOIs, Zenodo, WorldCat, TROVE, sistemas de preservação a longo prazo de bibliotecas nacionais e várias infraestruturas académicas e de conhecimento, deixam de ser meros tokens; começam a transformar-se em registos de conhecimento, ativos de literatura e memória da civilização. Portanto, acredito que a verdadeira questão importante no futuro não é como gerar mais tokens, mas como permitir que os tokens entrem em estruturas onde possam ser preservados a longo prazo e continuar a desempenhar o seu papel.

É por esta razão que permaneço otimista quanto ao futuro desenvolvimento dos tokens, mas não venero os tokens. O passado dos tokens é a unidade básica para os computadores processarem a linguagem; o presente dos tokens é uma unidade

de medida e unidade comercial importante na era da IA; o futuro dos tokens é provavelmente tornar-se um recurso fundamental importante para a civilização digital. No entanto, independentemente de como os tokens evoluam, mantenho sempre que o que é verdadeiramente escasso não são os tokens, mas as pessoas capazes de dominar os tokens; não aqueles que têm mais recursos, mas aqueles que podem estabelecer estruturas de longo prazo; não aqueles que podem gerar mais conteúdo, mas aqueles que podem organizar o conteúdo em sistemas de conhecimento.

XI. A incompressibilidade do tempo: Quatro ativos não compráveis

Ao discutir "o que é verdadeiramente escasso depois dos tokens", preciso de salientar uma questão fundamental que a maioria das discussões ignora: certos ativos são inerentemente incapazes de ser acelerados, comprados ou replicados. Por mais poderosa que seja a IA, por mais baratos que sejam os tokens, os seguintes quatro ativos manterão sempre a sua escassez inerente.

Primeiro, o tempo é incompressível.

Trinta anos de prática são simplesmente trinta anos; não podem ser comprimidos em três anos, nem simulados pela IA. A IA pode gerar um artigo sobre "como construir um sistema logístico" em segundos, mas não pode viver, em segundos, as várias anomalias, falhas, mudanças políticas e flutuações de mercado que um sistema logístico encontra ao longo de duas décadas. Estas acumulações de tempo do mundo real não podem ser encurtadas por nenhum meio tecnológico. O meu início de experimentação de um sistema protótipo de trabalho remoto em 1993, o design de um sistema de faturação inteligente em 2005 – estes pontos no tempo não são acidentes; são marcadores irreversíveis da história real. Quem vier mais tarde pode ler os meus artigos e aprender os meus métodos, mas não pode possuir os anos que vivi – esses anos em si são ativos.

Segundo, a experiência interdisciplinar não pode ser comprada.

A experiência interdisciplinar não pode ser comprada como uma mercadoria. Uma pessoa pode comprar cem livros sobre logística, edição, finanças e indexação de literatura num único dia, mas não pode adquirir, nesse mesmo dia, a experiência prática de ligar estes domínios. A verdadeira capacidade interdisciplinar não é o acúmulo de conhecimento, mas a capacidade de estabelecer conexões entre diferentes domínios, identificar pontos comuns e transferir métodos. Esta capacidade só pode ser adquirida resolvendo problemas reais, cometendo erros, corrigindo caminhos e acumulando julgamento num domínio após outro. A razão pela qual posso transferir a minha experiência dos sistemas logísticos para os sistemas de edição não é porque li um certo livro, mas porque passei vinte anos na logística e mais dez na edição – é na interseção destes dois períodos que surgiu a metodologia transferível.

Terceiro, a capacidade de construção de sistemas é difícil de replicar.

A capacidade de construção de sistemas não é um conjunto de códigos ou processos que podem ser copiados e colados. O código pode ser copiado, mas a lógica de tomada de decisão por trás do código, os juízos de compromisso e a previsão de mudanças futuras não podem ser replicados. O valor de um sistema não reside apenas no que pode fazer hoje, mas também em quantas mudanças, quantas recuperações de falhas e quantos ajustes de requisitos sofreu na última década enquanto continuava a funcionar. Esta "história do sistema" é em si uma prova da capacidade de construção. O meu sítio web (<https://times.net.au>) pode ser visto por qualquer pessoa, mas o que os visitantes veem é o resultado, não as três décadas de histórico de decisões, escolhas tecnológicas, ajustes estruturais e manutenção contínua por trás desse resultado. Esta capacidade de construção não pode ser obtida por plágio ou imitação.

Quarto, os registos de funcionamento a longo prazo não podem ser abreviados.

O registo de que um sistema funcionou durante duas décadas é em si inerentemente incapaz de ser abreviado. Um registo de funcionamento a longo prazo prova a estabilidade, a manutenibilidade e a visão de design do sistema. Na era da IA, gerar um sítio web ou sistema de "aspeto semelhante" pode ser feito em minutos, mas gerar um "sistema que funciona há duas décadas e é permanentemente arquivado pelo CERN, WorldCat e Biblioteca Nacional da Austrália" não tem atalhos. Um registo de funcionamento a longo prazo é um carimbo de tempo, declarando a todos que este sistema resistiu ao teste do mundo real. Tal registo não pode ser falsificado ou acelerado por nenhuma tecnologia de IA.

Estes quatro ativos – tempo, experiência interdisciplinar, capacidade de construção de sistemas e registos de funcionamento a longo prazo – constituem em conjunto a base do que defendi neste artigo como "verdadeiramente escasso". Possuo-os não porque seja mais inteligente ou mais trabalhador do que os outros, mas porque comecei mais cedo do que os outros e continuei até ao presente. Numa era em que a IA está a tornar mais e mais coisas baratas e facilmente acessíveis, estas coisas que não podem ser comprimidas, compradas, replicadas ou abreviadas tornar-se-ão cada vez mais valiosas.

XII. Conclusão: Da "acumulação de tokens" à "construção de sistemas"

O futuro dos tokens não reside nos próprios tokens.

Com o desenvolvimento contínuo da tecnologia dos grandes modelos e a contínua diminuição dos custos de inferência, a oferta de tokens continuará a expandir-se. Os tokens que são caros hoje serão tão baratos e omnipresentes como o tráfego de Internet e o armazenamento em disco rígido dentro de poucos anos. Quando todos puderem obter facilmente grandes quantidades de tokens, o que será verdadeiramente escasso já não serão os tokens, mas a capacidade de converter tokens em valor no mundo real – capacidade de design estrutural, capacidade de

automação, experiência industrial e acumulação prática de longo prazo. E a escassez mais profunda reside naquelas coisas que não podem ser comprimidas, compradas, replicadas ou abreviadas: o próprio tempo, a experiência interdisciplinar, a capacidade de construção de sistemas e os registos de funcionamento a longo prazo. Como argumentado anteriormente, a capacidade de construção de sistemas não pode ser adquirida diretamente através de grandes quantidades de tokens porque não é essencialmente uma capacidade generativa, mas sim uma capacidade de julgamento, tratamento de exceções, compromisso, absorção de retroalimentação e acumulação de confiança – a característica comum destas capacidades é que exigem tempo do mundo real, prática, erros, correções e validação.

No plano da rentabilidade, este julgamento aplica-se igualmente: ganhar dinheiro com tokens já é difícil, e tornar-se-á cada vez mais difícil. As guerras de preços comprimem continuamente as margens de lucro, e o modelo de negócio da simples revenda de tokens caminha rapidamente para lucros reduzidos ou mesmo nulos. Só quando os tokens estão ligados a um sistema frontal – isto é, quando um sistema fornece serviços – é que os tokens e o poder de computação podem servir como ativos complementares eficazes para gerar lucros. Os tokens em si não são a fonte do lucro; o sistema é. Os tokens são apenas o combustível do sistema, não o seu motor.

Desta perspetiva, os sistemas que já completei ou estou a desenvolver têm um significado especial para o desenvolvimento futuro dos tokens. Porque estes sistemas não foram construídos temporariamente para perseguir uma moda; existiam muito antes de o conceito de tokens se tornar um tópico público. O sistema logístico inteligente provou que o negócio do mundo real pode ser estruturado, processualizado e automatizado; o sistema de edição provou que um indivíduo também pode estabelecer um sistema de literatura multilingue, multiplataforma, multi-DOI, multi-canal de arquivo; o sistema de sítio web e metadados provou que o conteúdo digital pode ser continuamente organizado, indexado e divulgado; o sistema de referência cruzada automática prova ainda que, quando o número de

obras aumenta continuamente, o que realmente importa não é continuar a acumular conteúdo, mas estabelecer relações estruturais entre os conteúdos. O sítio web "The Epochal Transition" (<https://times.net.au>), como plataforma de exibição centralizada destes resultados, é em si uma evidência empírica da conversão da capacidade dos tokens num sistema académico aberto, permanente e verificável.

Mesmo que toda a gente recebesse a mesma quantidade de tokens no futuro, os resultados produzidos poderiam ser radicalmente diferentes: alguns obteriam mais registos de conversa, alguns obteriam mais conteúdo de páginas web, enquanto outros poderiam estabelecer um sistema logístico, um sistema de edição ou um sistema de conhecimento que funciona continuamente durante décadas. O que faz a diferença não são os próprios tokens, mas a arquitetura do sistema que os subjaz – e o tempo incompressível, a experiência interdisciplinar não comprável, a capacidade de construção difícil de replicar e os registos de funcionamento a longo prazo impossíveis de abreviar que suportam essas arquiteturas de sistema.

Portanto, não considero os tokens como um fim, mas como a nova infraestrutura da era inteligente. Os tokens são tão importantes como a eletricidade, mas a eletricidade por si só não constrói fábricas; os tokens são tão mensuráveis como o dinheiro, mas o dinheiro por si só não cria ativos de qualidade; os tokens podem fluir como a água, mas sem tubos, sistemas de despacho e registos de manutenção a longo prazo, apenas se dissiparão. As pessoas verdadeiramente valiosas no futuro não serão necessariamente aquelas com mais tokens, mas aquelas que conseguem organizar os tokens em sistemas. O maior equívoco da era dos tokens é considerar os tokens como um objetivo; a maior oportunidade da era dos tokens é usar os tokens para construir estruturas que outros não conseguem replicar. Quando os tokens se tornarem tão comuns como a eletricidade no futuro, as pessoas que olharem para trás para a atual loucura dos tokens poderão descobrir que o que realmente determinou a diferença entre as pessoas nunca foi a quantidade de tokens que possuíam, mas se possuíam a capacidade de aprender continuamente, praticar continuamente, acumular continuamente e construir sistemas continuamente – e aquelas coisas por trás destas capacidades que não podem ser

comprimidas, compradas, replicadas ou abreviadas. E na minha opinião, esta capacidade é a riqueza mais escassa e mais difícil de replicar na era dos tokens.

À medida que a minha investigação se aprofunda, sinto cada vez mais que a paisagem competitiva da futura era da IA pode ser resumida por uma analogia simples. Muitas pessoas consideram o poder de computação, os modelos e os tokens como recursos centrais, mas da perspetiva do funcionamento do mundo real, eles assemelham-se mais ao sistema logístico por trás de um exército. O poder de computação determina a capacidade de produção, os modelos determinam a capacidade organizacional, e os tokens determinam a capacidade de distribuição; juntos constituem o sistema de apoio material da era da IA.

No entanto, o que realmente determina o resultado de uma batalha não são os próprios suprimentos logísticos, mas o sistema de combate na linha da frente. Para a futura concorrência industrial, os sistemas são mais como a linha de ataque. Os sistemas determinam como os recursos são usados, como os processos são executados, como o valor é criado e se os utilizadores beneficiam verdadeiramente. Por mais abundantes que sejam os tokens e o poder de computação, sem a organização de um sistema, são como suprimentos empilhados num armazém, incapazes de serem convertidos em resultados reais.

Desta perspetiva, a concorrência futura provavelmente não será entre poderes de computação, nem entre tokens, mas entre sistemas. O sistema é a linha de ataque, determinando a direção do ataque e a capacidade de combate; os tokens são a munição e os suprimentos, determinando a capacidade de combate sustentável; o poder de computação é a base industrial na retaguarda, determinando a velocidade de produção e a capacidade de abastecimento. Sem abastecimento na retaguarda, a linha de ataque não pode avançar por muito tempo; mas com apenas o abastecimento na retaguarda e nenhum sistema de linha de ataque, também não se podem obter resultados.

Portanto, à medida que o poder de computação se torna mais generalizado, os modelos se tornam mais semelhantes e os tokens se tornam mais abundantes, o

que continua verdadeiramente escasso é sempre o sistema. Quem possuir um sistema mais maduro, dados mais reais, processos validados há mais tempo e a capacidade de ligar tokens e poder de computação às necessidades do mundo real, terá maior probabilidade de obter vantagem duradoura no futuro.

Os tokens estão a tornar-se a "eletricidade" da era da IA – indispensáveis, mas não um fim em si mesmos. Assim como os vencedores da revolução elétrica não foram as centrais elétricas, mas os inventores e empresários que inventaram a lâmpada, o motor e a cuba eletrolítica, os verdadeiros vencedores da era da IA serão os construtores de sistemas capazes de usar os tokens para construir estruturas e arquiteturas irreplicáveis. Depois dos tokens, isso é o que é verdadeiramente escasso.

Ao rever todo este artigo, sinto cada vez mais que a paisagem competitiva da futura era da IA pode ser ilustrada pela analogia de uma árvore.

Muitas pessoas concentram-se na água e nos nutrientes, porque são visíveis, tangíveis e facilmente mensuráveis. O poder de computação, os modelos e os tokens de hoje são como a água e os nutrientes continuamente fornecidos. O poder de computação determina a capacidade de produção, os modelos determinam a capacidade de transformação, e os tokens determinam a capacidade de distribuição. À medida que a tecnologia avança, estes recursos tornam-se cada vez mais abundantes e fáceis de obter.

No entanto, o que realmente determina a vitalidade de uma árvore não é a água e os nutrientes em si, mas o sistema radicular.

O sistema é a raiz da futura civilização digital.

O sistema radicular determina como a água é absorvida, como os nutrientes são distribuídos e se toda a árvore pode crescer a longo prazo. Sem raízes, por mais água que haja, esta escoará; sem raízes, por mais abundantes que sejam os nutrientes, não podem ser transformados em vitalidade. Da mesma forma, sem um sistema, por mais poderoso que seja o modelo, por mais baratos que sejam os

tokens, por mais vasto que seja o poder de computação, é difícil criar valor de forma contínua.

Desta perspectiva, o poder de computação é a água, os tokens são os nutrientes, os modelos são os tubos que entregam os nutrientes, e o sistema é o sistema radicular que penetra profundamente no mundo real. O que realmente determina o resultado da concorrência futura não é quem tem mais água, nem quem tem mais nutrientes, mas quem tem raízes mais profundas, mais estáveis, capazes de absorver e utilizar estes recursos a longo prazo.

No futuro, à medida que o poder de computação continuar a crescer, os modelos continuarem a evoluir e os preços dos tokens continuarem a cair, os próprios recursos se tornarão cada vez mais comuns. O que será verdadeiramente escasso já não serão os recursos, mas os sistemas capazes de organizar, utilizar e criar valor continuamente a partir dos recursos. Quem possuir um sistema maduro possui as raízes; quem possui as raízes possui a capacidade de sobrevivência a longo prazo e de crescimento contínuo.

Portanto, acredito que o que é verdadeiramente escasso depois dos tokens não são mais tokens, nem mais poder de computação, mas os sistemas capazes de converter poder de computação, modelos e tokens em valor no mundo real.

Quando a água e os nutrientes se tornam cada vez mais abundantes, o que realmente determina se uma árvore pode crescer a longo prazo é sempre o sistema radicular.

Palavras-chave: Token, Inteligência Artificial, IA, Grande Modelo de Linguagem, Poder de Computação, Economia de Modelos, Economia de Tokens, Token Chinês, Token Inglês, Algoritmo BPE, Soberania de Dados, Agente, Ecossistema de API, Capacidade de Construção de Sistemas, Design de Sistemas, Sistema de Automação, Investigação Interdisciplinar, Prática Interdisciplinar, Dados Históricos Reais, Sistema de Longa Duração, Ativo Temporal, Concorrência de Plataformas, Concorrência de Sistemas, Civilização Digital, Memória da Civilização, Infraestrutura de Conhecimento, Investigador Independente, Investigador Independente

Internacional, Sistema Logístico Inteligente, Sistema de Edição em Dez Línguas,
Sistema de Metadados, Referência Cruzada Automática, DOI, Zenodo, WorldCat,
TROVE, The Epochal Transition, Wu Zhaohui, JEFFI CHAO HUI WU

Referências:

1. *A Structural Case Study of the 1993 Remote Work Prototype System*
DOI: <https://doi.org/10.5281/zenodo.17978371>
 2. *The 2005 Intelligent Invoice System*
DOI: <https://doi.org/10.5281/zenodo.19044755>
 3. *Mature Structural Logistics System*
DOI: <https://doi.org/10.5281/zenodo.20201895>
 4. **Cross-Generational 20-Year Intelligent Logistics and Financial System**
DOI: <https://doi.org/10.5281/zenodo.20352371>
NLA: <https://nla.gov.au/nla.obj-4204311557>
 5. *Analysis of the Homepage Structural Evolution of The Epochal Transition*
DOI: <https://doi.org/10.5281/zenodo.20362574>
NLA: <https://nla.gov.au/nla.obj-4204309341>
 6. *An Empirical Study of Institutional Engineering: Independent Scholars Accessing the WorldCat Global Bibliographic Infrastructure*
DOI: <https://doi.org/10.5281/zenodo.18028572>
 7. *National-level Network Archiving and Digital Document Preservation Guidelines of the Australian Winner Information Network*
DOI: <https://doi.org/10.5281/zenodo.17888259>
 8. *The Institutional Archiving of The Epochal Transition*
DOI: <https://doi.org/10.5281/zenodo.17932379>
- WU, J. C. H. (2026). Source Declaration for Audiovisual and Derivative Adaptations of a Continuing Real-World Narrative. Zenodo.
<https://doi.org/10.5281/zenodo.18160116>

После Токенов: Что действительно является дефицитным в эпоху ИИ

Автор/независимый международный исследователь: У Чжаохуэй, Джеффи Чао Хуэй Ву

ORCID: <https://orcid.org/0009-0009-5318-3027>

Предварительное заявление

Эта статья основана на моих личных наблюдениях, вытекающих из более чем тридцати лет междисциплинарной практики (начиная с моей учебы в Австралии в 1989 году, охватывая удаленную работу, управление логистикой, автоматизированную обработку, создание веб-сайтов, распространение культуры, многоязычное издательское дело и междисциплинарные исследования). Токены — это быстро развивающееся новое явление; данная статья отражает только мои личные, поэтапные наблюдения и суждения на момент написания (июнь 2026 года) и никоим образом не представляет собой инвестиционную консультацию, коммерческий прогноз или отраслевой стандарт.

Эта статья не претендует на универсальную применимость и не заявляет о том, что представляет какой-либо отраслевой консенсус. В связи с продолжающимся изменением рынка токенов, технологий ИИ и глобальной политической среды некоторые суждения в этой статье могут потребовать пересмотра или обновления в будущем. Читатель должен выносить независимые суждения, основываясь на самой актуальной информации.

«Пять дефицитных способностей» и «пять личных активов», которые я обсуждаю, проистекают из моей собственной практики построения систем (публичные доказательства доступны на <https://times.net.au>, включая полный список из более чем 40 независимых творческих работ на этом сайте, ISSN

3083-5178, более 60 DOI, а также записи, постоянно архивируемые такими международными инфраструктурами, как WorldCat, Национальная библиотека Австралии и CERN Zenodo).

Чтобы избежать вырывания из контекста, любое действительное опровержение или оспаривание этой статьи должно иметь следующее предварительное условие: **полное прочтение всех статей, на которые есть ссылки в этом документе, и вложенных в них ссылок, а также всех моих работ, перечисленных на <https://times.net.au>**. Частичное цитирование или упрощенное резюмирование, оторванное от этого общего контекста, не может constitute справедливую оценку этой статьи.

Дополнительное примечание о действительном опровержении:

Эта статья выражает личные взгляды, не претендует на универсальную применимость и не предполагает цитирования работ других людей. Любое действительное оспаривание должно использовать сравнение «яблоко с яблоком»: факты против фактов, личная практика против личной практики, завершённые системы против завершённых систем. Использование «того, что будущий ИИ мог бы достичь» для отрицания «уже завершённых практик» или сравнение возможностей команды/учреждения с индивидуальной практикой не является справедливым сравнением.

Портфель из пяти активов, описанный в этой статье, представляет собой запись личной практики и не утверждает, что «не существует никакого сопоставимого случая». Если вы считаете, что действительно сопоставимый случай существует, пожалуйста, приведите конкретный публичный случай и объясните его соответствующие доказательства по пяти измерениям: реальные исторические данные, междисциплинарные системы, способность к построению систем, дальновидность и способность к самоопровержению. В противном случае любое оспаривание является всего лишь выражением мнения и не представляет собой действительного опровержения.

Аннотация

Токены становятся центральной единицей измерения в эпоху ИИ, но в этой статье будет утверждаться, что токен сам по себе не является конечной ценностью. Поскольку затраты на вывод (inference) продолжают снижаться, токены станут такими же дешевыми и повсеместными, как сегодня интернет-трафик и хранилище на жестких дисках. Когда токены перестанут быть дефицитными, что же действительно станет дефицитным? Основываясь на более чем тридцатилетней междисциплинарной практике автора (от прототипа системы удаленной работы 1993 года, интеллектуальной системы счетов 2005 года до интеллектуальных логистических систем, десяти-язычной издательской системы, управления DOI, автоматического перекрестного ссыла и кроссплатформенной публикации), эта статья выдвигает следующие ключевые суждения: Во-первых, способность к построению систем — включая структурное проектирование, автоматизацию, междисциплинарную практику, ментальные модели долгосрочного функционирования и способность встраивать токены в реальные бизнес-процессы — является действительно дефицитным активом. Во-вторых, способность к построению систем не может быть приобретена напрямую через большие объемы токенов, потому что она зависит от принятия решений, обработки исключений, компромиссов, циклов обратной связи из реального мира и доверия, накопленного со временем. В-третьих, несжимаемость времени, невозможность покупки междисциплинарного опыта, трудность воспроизведения способности к построению систем и невозможность ускорить получение записей о долгосрочном функционировании — эти четыре актива составляют глубочайший дефицит в эпоху токенов. В этой статье в качестве публичного доказательства используется созданный автором академический веб-сайт «The Epochal Transition» (<https://times.net.au/>), ISSN 3083-5178, постоянно архивируемый такими международными инфраструктурами, как WorldCat, CERN Zenodo и TROVE Национальной библиотеки Австралии), показывающий, как один человек может построить междисциплинарную, проверяемую, долго работающую системную архитектуру за тридцать лет.

В этой статье утверждается: Токены — это всего лишь топливо; система — это двигатель. После токенов действительно дефицитным является человек, способный управлять токенами, и структуры, которые он создал.

«Без товаров какой смысл в деньгах?»

Когда я просматриваю историю развития токенов, я все больше чувствую, что она очень похожа на многие технологические революции, пережитые человечеством в прошлом. Каждый раз, когда появляется новый фундаментальный ресурс, люди сначала сосредотачиваются на самом ресурсе, пренебрегая стоящей за ним структурой. Это было верно в индустриальную эпоху, в эпоху Интернета и в равной степени верно в эпоху искусственного интеллекта. Сегодня многие дискуссии сосредоточены на ценах на токены, затратах на токены, квотах на токены и способности поставки токенов, как будто обладание большим количеством токенов означает обладание будущим. Но по моему многолетнему практическому опыту, сам ресурс никогда не был конечной ценностью. Что действительно определяет ценность, так это всегда то, как люди организуют и используют эти ресурсы.

Чтобы понять «токены», нельзя начинать просто с нового термина; нужно начинать с их истории. От американского философа Чарльза Сандерса Пирса, написавшего двадцать экземпляров слова «the» на бумаге в 1906 году, различая «тип» и «токен», до Лю Лихуна, директора Государственного управления данных Китая, который дал официальный китайский перевод «цзыюань» (词元) термину Token на Боаоском форуме в 2026 году, это понятие прошло долгий путь от философии к информатике и от блокчейна к искусственному интеллекту. Слово «token» существовало в западной информатике уже давно, где в языках программирования, компиляторах и системах обработки текста токен обозначал базовую единицу, которую компьютер мог распознать и обработать. Позже, в контекстах сетевой безопасности, аутентификации, вызовов API и блокчейна, «token» часто понимался как «пропуск» или «удостоверение», указывающее на права доступа, идентификационные данные или цифровые активы. Но в эпоху

больших языковых моделей Token больше не является прежде всего идентификационным удостоверением или активом блокчейна, а основной единицей вычислений, с помощью которой модели ИИ обрабатывают язык, код, описания изображений и мультимодальную информацию.

Позже Китай назвал Token в области ИИ «цыюань» (词元). Я считаю это довольно точным переводом. Он не использовал «линпай» (пропуск), потому что Token в больших моделях — это не пропуск; он также не был просто переведен как «цзы» (иероглиф) или «цы» (слово), потому что он может быть иногда иероглифом, иногда словом, иногда половиной английского слова, а иногда даже просто знаком препинания или символом. «Цы» указывает на его связь с языком и информацией, а «юань» указывает на то, что это базовая единица. Для обычного читателя можно просто понять так: то, что человек видит как предложение, ИИ видит как строку токенов. На этом пути токены прошли путь от малоизвестного лингвистического термина до центральной единицы измерения в эпоху ИИ, а затем и до расчетного инструмента, связывающего технологическое предложение с коммерческим спросом. Как практик, лично испытывавший долгосрочные проекты, такие как создание интеллектуальной логистической системы, проектирование автоматизированных издательских процессов и кроссплатформенное распространение контента, я имею непосредственное понимание этой эволюции — токены никогда не были целью, а средством; что действительно определяет ценность, так это никогда не то, у кого больше токенов, а то, кто может использовать токены для создания постоянно работающих систем и невозпроизводимых структур. Результаты способности к построению систем, обсуждаемой в этой статье, были публично представлены на <https://times.net.au> — междисциплинарной академической издательской системе, созданной независимым исследователем и постоянно архивируемой в мировой библиотечной системе.

Размышляя о своем собственном пути, я начал думать о системах не только после появления токенов, но и постепенно развил понимание систем через

долгосрочную практику. Начиная с моей учебы в Австралии в 1989 году, я прошел через опыт обучения, работы и исследований в разных областях. Позже я постепенно вовлекся в удаленную работу, управление логистикой, автоматизированную обработку, создание веб-сайтов, распространение культуры, многоязычное издательское дело и междисциплинарные исследования. Оглядываясь назад, хотя этот опыт происходил в разные эпохи, у него есть общая черта: он всегда вращался вокруг того, как создавать структуры, как уменьшать повторяющийся труд и как повышать общую эффективность. Понятия токенов тогда не существовало, как и сегодняшних больших языковых моделей, но я уже постоянно думал о том, как заставить информацию течь, организовываться и обрабатываться автоматически.

Поэтому, хотя сегодня все больше людей рассматривают токены как новое средство производства, я склонен рассматривать их как новую форму фундаментальной энергии. Энергия, безусловно, важна, но энергия не создает ценность автоматически. Электричество само по себе не становится автоматически заводом; Интернет сам по себе не становится автоматически предприятием; токены сами по себе не становятся автоматически системой знаний. В будущем, даже если каждый сможет легко получить большие объемы токенов, это не означает, что каждый сможет достичь тех же результатов. Разницу определяет не количество токенов, а структура, в которую попадают токены.

Прежде чем приступить к углубленному анализу токенов, я должен указать на другое важное наблюдение: Токены похожи на валюты, и технологические, системные и политические различия между Востоком и Западом делают их не универсальными. Подобно тому, как доллар США и китайский юань не могут свободно обмениваться, китайские токены и английские токены сталкиваются с аналогичными барьерами обращения. Это не просто вопрос языкового различия; более глубокая причина кроется в тройных различиях между Востоком и Западом на технологическом уровне (архитектуры моделей, алгоритмы токенизации, обучающие данные), системном уровне (API-

экосистемы, платежные системы, стандарты данных) и политическом уровне (трансграничные потоки данных, экспортный контроль, проверки национальной безопасности), которые препятствуют свободному обращению токенов как идеализированной «глобальной универсальной валюты ИИ». Это наблюдение имеет серьезные последствия для понимания будущего ландшафта рынка токенов — в будущем могут появиться «китайские токен-зоны», «английские токен-зоны» и другие токен-экосистемы, основанные на языке и политике, и межзонные вызовы столкнутся с потерями эффективности и затратами на соответствие требованиям.

I. Происхождение токенов: Четверной скачок от философской концепции к экономической единице

Эволюцию идентичности токенов можно четко разделить на четыре этапа.

Истоки в философской лингвистике. В 1906 году Пирс, изучая семиотику, различил «тип» и «экземпляр» (token). Первый представляет абстрактное правило или форму, в то время как второй является конкретным воплощением этого правила в реальности. Это различие позже стало важной основой структурной лингвистики.

Кодирующий поворот в информатике. В 1960-х годах, когда программист писал строку кода «`int x = 5;`», компьютеру нужно было разбить ее на независимые единицы, такие как «`int`», «`x`», «`=`», «`5`» и т.д. Каждая такая единица была токеном. Таким образом, токены стали наименьшей информационной единицей для «чтения» машиной инструкций.

Эксперимент с носителем стоимости в блокчейне. В 2017 году, с подъемом бума ICO, Token получил новую идентичность как «обращающийся цифровой сертификат прав». Хотя эта волна энтузиазма постепенно утихла, понимание «Токена как символа стоимости» укоренилось.

Экономическая инфраструктура эпохи ИИ. Март 2026 года стал знаменательным моментом — генеральный директор NVIDIA Дженсен Хуан переопределил центры обработки данных как «фабрики, производящие интеллектуальные токены ИИ», а Лю Лихун, директор Государственного управления данных Китая, дал официальный китайский перевод «цыюань» (词元), назвав его «якорем стоимости интеллектуальной эры». В этот момент токены совершили свой последний скачок от технической концепции к экономической концепции.

II. Техническая основа: Алгоритм BPE и неявный «языковой налог»

То, что токены смогли стать стандартной единицей измерения для ИИ, основано на забытой более чем на два десятилетия технологии — попарном байтовом кодировании (Byte Pair Encoding, BPE).

В 1994 году американский программист Филип Гейдж опубликовал статью об алгоритмах сжатия данных в техническом журнале по языку C, представив основной принцип BPE: многократное сканирование текста и «сварка» двух наиболее часто встречающихся соседних символов в новый символ, сжимая итеративно за раундом раунд. Эта статья не привлекла внимания в то время, поскольку ее эффективность сжатия не была выдающейся.

Только в 2016 году Рико Сеннрих, исследователь Эдинбургского университета, изучая проблему сегментации слов в машинном переводе, случайно наткнулся на эту старую статью и остро осознал, что BPE является идеальным решением проблемы токенизации — не нужно было предопределять словарь; данным позволялось «говорить» самим за себя, а высокочастотные комбинации естественным образом сливались в токены. В 2019 году OpenAI официально приняла этот подход при выпуске GPT-2, установив отправную точку для токенизации непосредственно на уровне «байта», что теоретически позволяет модели обрабатывать любую письменность.

Однако логика «частотности превыше всего» алгоритма BPE непреднамеренно сформировала неявную систему «языкового налога». Поскольку английский язык является абсолютным мейнстримом интернет-корпусов, эффективность модели в сегментации английских токенов намного выше, чем для других языков. Для выражения одного и того же смысла английский язык использует наименьшее количество токенов, китайскому обычно требуется в 1,5–2 раза больше, а для языков с еще меньшими ресурсами затраты могут быть в 5–10 раз выше, чем для английского. Это означает, что при оплате за токен фактическая стоимость разговора с ИИ сильно варьируется в зависимости от используемого языка.

Эта несправедливость на «линии старта», однажды записанная в начальный словарный запас модели, очень трудно поддается исправлению — правила токенизации являются «фундаментом» понимания мира ИИ; чем выше здание, тем труднее заменить фундамент. К счастью, эта ситуация улучшается. На примере серии GPT: то же самое предложение на китайском языке требовало 38 токенов в GPT-3, снизилось до 26 в GPT-4 и до 15 в GPT-5 — улучшение эффективности более чем на 60%. Более того, китайские отечественные большие модели, такие как Tongyi Qianwen и DeepSeek, с самого этапа проектирования включили высокочастотные китайские фразы, идиомы и т.п. как родные токены в свои словари, достигнув более «родной» эффективности обработки китайского языка. Это раскрывает глубокий принцип: тот, кто владеет «правом на семантическую сегментацию», в значительной степени владеет преимуществом эффективности и стоимости для выражения этого языка в цифровом мире — это форма «фундаментального права чеканки» в цифровую эпоху.

III. Использование и производство: Экспоненциальный скачок от диалога к агентам

Сценарии потребления токенов переживают скачок на трех уровнях: от «диалога» к «рассуждению» и затем к «агентам».

На этапе диалогового ИИ один запрос и ответ потребляет около 10 000 токенов. На этапе ИИ рассуждений, включающего несколько раундов взаимодействия и сложные выводы, каждая задача потребляет около 100 000 токенов. При входе в этап агентного ИИ агенты могут автономно выполнять многошаговые задачи и вызывать несколько инструментов, при этом потребление токенов на задачу достигает примерно 1 миллиона — увеличение почти в 100 раз всего за два поколения эволюции. Генеральный директор Qualcomm Амон прогнозирует, что к 2030 году мировой спрос на токены каждые 10 секунд взлетит с 31,7 миллиарда в 2026 году до 1,27 триллиона. Это не преувеличение; взрывная популярность инструментов агентов с открытым исходным кодом, таких как OpenClaw, доказала это: одна сложная задача, потребляющая десятки миллионов токенов, обычное дело.

На стороне производства Китай стал крупнейшим в мире поставщиком токенов. Публичные данные показывают, что по состоянию на март 2026 года суточный объем вызовов токенов в Китае превысил 140 триллионов, увеличившись более чем в тысячу раз за два года, что составляет 61% мирового использования токенов ИИ. Установление этой позиции основано на наложении трех преимуществ: во-первых, цены на промышленное электричество в Китае намного ниже, чем в Европе и Америке, и электроэнергия составляет более 60% затрат на производство токенов; во-вторых, Китай построил крупнейший в мире кластер интеллектуальных вычислений, достигнув эффективного производства благодаря распределению «Вычисления на востоке, данные на западе»; в-третьих, отечественные китайские большие модели полностью открыты, а система коммерциализации API-услуг зрелая, образуя полную промышленную цепочку с обратной связью

от «обучения модели - поддержки вычислений - вывода результатов - глобального вызова».

IV. Институты и рынки: Трехсторонний индустриальный ландшафт

Вокруг токенов сформировалась четко стратифицированная рыночная структура.

Для понимания этой промышленной структуры также необходимо учитывать поддерживающую систему, лежащую в основе токенов. Многие люди сосредотачиваются на ценах на токены, но упускают из виду, что токены не являются независимым ресурсом. За токенами стоят модели, а за моделями — вычислительная мощность. В некотором смысле, вычислительная мощность определяет возможности модели, модели определяют производство токенов, а токены — это единица измерения, которая формируется при экспорте возможностей модели на рынок. Поэтому на поверхности рынок токенов — это ценовая конкуренция, но в своей основе это все еще конкуренция вычислительной мощности и эффективности.

Текущую мировую промышленную цепочку токенов можно грубо разделить на три уровня. Самый верхний уровень — это инфраструктура вычислительной мощности, включающая производителей чипов, производителей серверов, облачные платформы, центры обработки данных и энергетические системы. Средний уровень состоит из институтов базовых моделей и платформ вывода (inference), отвечающих за преобразование вычислительной мощности в вызываемые токен-сервисы. Нижний уровень состоит из поставщиков токен-сервисов, агентных платформ и различных отраслевых приложений, которые далее преобразуют возможности моделей в корпоративные услуги и конечные продукты. Что пользователь видит в итоге, так это цену токена, но то, что действительно влияет на стоимость и скорость токена, часто является

изменением в вышестоящей системе вычислительной мощности и эффективности модели.

За последние несколько лет, с постоянным улучшением производительности чипов, оптимизацией архитектур моделей и повышением эффективности вывода (inference), стоимость одного токена демонстрировала быструю тенденцию к снижению. Многие люди видят, что токены дешевеют, но истинным двигателем этого изменения является постоянный прогресс системы вычислительной мощности. Поэтому будущая конкуренция на рынке токенов будет включать не только конкуренцию между платформами и ценовую конкуренцию, но и конкуренцию вычислительных ресурсов, эффективности моделей и способностей к интеграции ресурсов.

Вышестоящие «электростанции» – производители вычислительной мощности. Чиповые компании, представленные NVIDIA, постоянно совершенствуют оборудование для повышения эффективности производства токенов на единицу вычислительной мощности. NVIDIA прогнозирует, что ее доход достигнет по крайней мере 1 триллиона долларов к 2027 году, уверенность, подкрепленная глобальной жадой вычислительной мощности ИИ.

Среднеточные «диспетчерские сети» – модельные платформы и агрегаторы. Такие компании, как OpenAI, Google, DeepSeek и Alibaba, преобразуют базовую вычислительную мощность в стандартизированные, вызываемые токен-сервисы. Вокруг этого звена сформировались три зрелые бизнес-модели: платформы агрегации токенов (представленные OpenRouter, работающие на разнице цен и комиссиях, с недельным объемом вызовов, достигшим 26,9 триллиона токенов к маю 2026 года), услуги MaaS (Model-as-a-Service) облачных провайдеров и ретрансляционные станции ИИ, активные на вторичном рынке.

Нижестоящие «точки входа» – платформы приложений и распространения. Открытие интерфейса WeChat компанией Tencent для

доступа к возможностям новых моделей по существу превращает WeChat в «платформу распространения токенов»; China Telecom предложила «перестроить свой бизнес через управление токенами», пытаясь выйти из традиционной логики трафика. У этих компаний-точек входа есть общая черта: пользователи больше напрямую не взаимодействуют с моделью, а вызывают токены косвенно через платформу — сама модель становится «невидимой».

Стоит упомянуть, что вокруг токенов сформировалась уникальная парадигма глобальной торговли: нематериальный экспорт, необлагаемое налогом функционирование и глобальное покрытие. Это превращает Китай из традиционной «мировой фабрики» в «глобального поставщика инфраструктуры вычислительных мощностей для ИИ».

При обсуждении экономики токенов есть еще один фундаментальный фактор, который нельзя обойти стороной: вычислительная мощность. Многие люди сосредотачиваются на ценах на токены, но упускают из виду производственный процесс, стоящий за токенами. На самом деле токены не возникают из ничего. Каждая генерация токена требует от модели выполнения вывода (inference), а вывод модели строится на основе массивной вычислительной мощности. В некотором смысле, вычислительная мощность является источником экономики токенов, модели являются инструментами производства токенов, а токены — это единица измерения, которая формируется при экспорте возможностей модели вовне.

Поэтому уровни вычислительной мощности напрямую влияют на затраты на токены и скорость генерации токенов. Чем сильнее вычислительная мощность, тем выше способность модели обрабатывать информацию и тем больше токенов может быть произведено за единицу времени; чем ниже стоимость вычислительной мощности, тем ниже обычно цены на токены. За последние несколько лет индустрия больших моделей стала свидетелем непрерывного снижения цен на токены, вызванного не только оптимизацией алгоритмов моделей, но и улучшением производительности чипов, увеличением масштаба центров обработки данных и повышением эффективности вывода. Другими

словами, многие люди видят, что токены дешевеют, но истинным двигателем этого изменения является постоянный прогресс системы вычислительной мощности.

С точки зрения промышленной цепочки, вычислительная мощность, модели и токены на самом деле образуют полную цепочку. Выше по течению находятся чипы, серверы, центры обработки данных и энергетические системы; в среднем течении находятся фундаментальные модели и платформы вывода; ниже по течению находятся токен-сервисы, агентные приложения и различные отраслевые решения. Вычислительная мощность определяет верхний предел возможностей модели; модели определяют качество и эффективность токенов; а токены в конечном итоге определяют, какой объем услуг ИИ может получить пользователь.

Однако я считаю, что, хотя важность вычислительной мощности неоспорима, сама вычислительная мощность не создает ценность автоматически. Многие отрасли пережили аналогичные процессы в истории. Обладание более крупной электростанцией не означает обладание лучшим предприятием; обладание большим количеством серверов не означает обладание лучшим продуктом. Точно так же обладание большей вычислительной мощностью не означает обладание более ценной системой. Вычислительная мощность решает проблему производственных мощностей, в то время как системы решают проблему преобразования стоимости. В будущем, когда вычислительная мощность продолжит расти, возможности моделей продолжат усиливаться, а токен-ресурсы продолжат размножаться, тем, что действительно будет дефицитным, будет не сама вычислительная мощность, а то, как организовать эти ресурсы в системы, способные долгосрочно функционировать в реальном мире.

Поэтому с точки зрения будущей конкуренции я считаю, что конкуренция вычислительных мощностей, моделей и токенов постепенно перерастет в конкуренцию систем. Потому что, когда вычислительная мощность станет более распространенной, модели станут более похожими, а токены станут

более обильными, разницу будет определять уже не количество принадлежащих ресурсов, а то, кто сможет использовать эти ресурсы для построения долгосрочных эффективных структур.

V. Конкуренция и барьеры: Инволюция ценовой войны и стратификация ценности

Конкуренция на рынке токенов демонстрирует высокие черты инволюции. С 2022 по 2026 год стоимость одного вывода (inference) токена снизилась на 99,9%. Отечественные модели демонстрируют еще более значительные ценовые преимущества: входная цена DeepSeek V4-Pro составляет всего 0,025 юаня за миллион токенов, выходная цена Tongyi Qianwen Qwen3.5-Plus — 1,6 юаня за миллион токенов, по сравнению с выходной ценой OpenAI GPT-4o примерно в 72 юаня.

Эта ценовая война резко сжала прибыль «посредников». Отраслевая статистика показывает, что количество различных платформ агрегации и ретрансляции API превысило две тысячи, при этом рынок сформировал три основные **梯队**: верхний эшелон крупных платформ с сильными технологиями и всеобъемлющими моделями; средний эшелон профессиональных агентов с официальными скидками и клиентскими ресурсами; и огромное количество микро- и малых агентов, столкнувшихся с серьезным давлением выживания. Бизнес-модель простого получения прибыли от разницы в ценах на токены больше не является устойчивой.

Истинные барьеры смещаются в двух направлениях. Одно из них — направление технологической премии — за счет разработки собственных движков ускорения вывода можно продавать дешевле официального поставщика и все равно получать прибыль. Второе — направление услуг с добавленной стоимостью — предоставление специализированных услуг, таких как инжиниринг подсказок (prompt engineering), тонкая настройка моделей и частное развертывание. Еще более заметно, что стоимость токенов

демонстрирует огромную дифференциацию, основанную на сценарии: средняя цена токенов, связанных с чатом, составляет всего 0,01 доллара за миллион, в то время как для токенов, связанных с открытием лекарств, средняя цена может достигать 1000 долларов за миллион — разница в стоимости между сценариями в сто тысяч раз. Это раскрывает важный принцип: ценность токена определяется не его производственной стоимостью, а его использованием — один и тот же токен, используемый в разных сценариях и встроенный в разные системы, генерирует совершенно разную ценность.

Но здесь необходимо сделать более важное суждение: В настоящее время и в будущем барьер для продажи токенов невысок. Любой может зарегистрироваться для получения API, создать платформу-агрегатор и начать перепродавать токены. Что действительно является дефицитным, так это то, что находится за пределами продажи токенов — реальные исторические данные, способность строить системы и архитектуры, а также дальновидное видение будущего. Без этого все большие модели и токены бесполезны.

Дополнительное суждение о прибыльности токенов: Зарабатывать деньги на токенах уже сложно, и будет становиться все сложнее. Ценовые войны постоянно сжимают нормы прибыли, и бизнес-модель простой перепродажи токенов быстро движется к мизерной или даже нулевой прибыли. Только когда токены привязаны к фронтальной системе — т.е. когда система предоставляет услуги, — токены и вычислительная мощность могут служить эффективными дополнительными активами для получения прибыли. Другими словами, токены сами по себе не являются источником прибыли; система является источником прибыли. Токены — это всего лишь топливо для системы, а не ее двигатель.

Поскольку токены постепенно становятся важной единицей измерения в эпоху искусственного интеллекта, на рынке появилось все больше бизнес-моделей, ориентированных на токены. Помимо поставщиков базовых моделей, существуют также токен-агенты, платформы агрегации моделей, агентные платформы, корпоративные услуги распространения API и различные поставщики услуг промежуточного слоя. Для многих предпринимателей токен-

агентство кажется быстрым путем в индустрию ИИ, поскольку оно не требует обучения моделей или строительства крупных вычислительных центров; нужно всего лишь интегрировать существующие ресурсы, чтобы начать бизнес. Однако эта модель имеет как преимущества, так и очевидные ограничения. Ее преимущества включают относительно низкие начальные инвестиции, относительно низкие барьеры входа и более быстрое расширение рынка; ее ограничения включают высокую зависимость от вышестоящих поставщиков моделей, подверженность ценовой конкуренции и отсутствие долгосрочных технологических барьеров. Поскольку цены на токены продолжают падать, возможности для получения прибыли от простой перепродажи токенов могут постоянно сжиматься. Поэтому в долгосрочной перспективе токен-агентство — это скорее дистрибьюторская способность, чем конечная основная конкурентоспособность. Что действительно определяет долгосрочную стоимость компании, остаются отраслевые системы, процессы автоматизации, системы знаний и решения реального мира, построенные вокруг токенов.

VI. Текущие трудности и будущие пути

Текущий рынок токенов сталкивается с несколькими глубинными проблемами.

Ловушка гомогенной конкуренции. Более двух тысяч платформ борются за одних и тех же клиентов, и ценовые войны продолжают снижать прибыльность отрасли. Без изменений это может повторить ранние дни облачных вычислений, когда «теряли деньги, чтобы завоевать долю рынка».

Угроза рисков соответствия требованиям. Текущий глобальный «регуляторный дивиденд» отсутствия специального законодательства является временным состоянием, а не долгосрочной нормой. По мере расширения рынка токенов такие вопросы, как трансграничные данные, национальная безопасность и налоговое администрирование, заставят страны ускорить законодательство. Особенно важно отметить, что попытки некоторых участников рынка токенизировать и финансировать токены могут легко спровоцировать

глобальное финансовое регулирование, что не только приведет к закрытию таких предприятий, но и нанесет ущерб репутации всей китайской индустрии экспорта токенов.

Миф «токен-максинга» (Token-maxxing). В Силиконовой долине появилась новая форма «хвастовства богатством» — соревнование в том, кто потребляет больше токенов в день. У Meta есть внутренняя таблица лидеров под названием «Claudeonomics», которая собирает данные об использовании ИИ более чем 85 000 сотрудников, причем лучший сотрудник сжигает токены на сумму в несколько миллионов долларов. Критики язвительно замечают: «Максимизация результатов лучше максимизации токенов» — вместо безумного потребления токенов лучше посмотреть, какие результаты получены. Это именно та ключевая проблема, которую я глубоко понял за годы практики.

Китайские токены против английских токенов: Три барьера под единой концепцией

Для китайских пользователей факт, который необходимо четко осознавать, заключается в следующем: хотя китайские токены (цзюань) и английские токены принадлежат к одной и той же фундаментальной единице измерения в эпоху ИИ, они не сформировали действительно единого, свободно взаимодействующего глобального рынка. С точки зрения технической концепции, и Восток, и Запад используют токены как базовую единицу для обработки моделями языка и данных; однако с точки зрения практического применения, совместное существование технологических барьеров, системных барьеров и политических барьеров создало фактическое разделение между китайскими токенами и английскими токенами.

Первый барьер — это технологический барьер, т.е. различия в логике токенизации и архитектуре модели.

Западные основные модели, такие как GPT, Claude и Gemini, в основном построены на крупных корпусах, преимущественно на английском языке. Хотя эти модели уже обладают отличными возможностями обработки китайского языка, во многих сценариях китайский язык все еще требует больше токенов для выражения того же объема информации, что и английский. Что еще более важно, разные модели имеют разные структуры словаря и стратегии токенизации. Один и тот же текст на китайском языке может быть разбит на совершенно разное количество токенов в разных моделях, и их внутренние методы кодирования и семантического отображения также могут различаться.

В то же время, китайские отечественные модели, такие как DeepSeek, Tongyi Qianwen, Doubao и Wenxin Yiyao, с самого этапа обучения включили большие объемы китайских корпусов, что дает им более высокую родную адаптируемость к китайским фразам, идиомам, именам собственным и локализованным выражениям. Поэтому даже применительно к одному и тому же тексту количество токенов, эффективность обработки и пути понимания в разных моделях могут значительно различаться. В этом смысле, хотя токен является единой концепцией, не существует «стандартного токена», который можно было бы напрямую передавать и повторно использовать между моделями.

Второй барьер — это системный барьер, т.е. различия в экосистемах и методах вызова.

Текущий глобальный рынок токенов фактически сформировал несколько относительно независимых экосистем. Международные платформы, такие как OpenAI, Anthropic и Google, имеют свои собственные стандарты API, модели ценообразования, правила доступа и сервисные системы; китайский рынок сформировал внутреннюю экосистему, сосредоточенную вокруг таких платформ, как DeepSeek, Tongyi Qianwen, Doubao и Kimi.

Для обычных пользователей покупка токенов на одной платформе не означает, что они могут использовать их непосредственно на другой платформе.

Интерфейсы вызова, механизмы аутентификации, ограничения скорости, соглашения об обслуживании и коммерческие правила различаются от платформы к платформе. Даже если конвертация осуществляется через сторонние шлюзы или платформы-агрегаторы, это, по сути, повторный вызов другой системы, а не настоящий обмен токенами. Поэтому, хотя токен и является единой концепцией, токены как коммерческий ресурс и сервисная способность остаются заблокированными в своих соответствующих платформенных экосистемах.

Третий барьер — это политический барьер, т.е. различия в потоках данных и требованиях соответствия.

По мере того как искусственный интеллект постепенно входит в стадию корпоративного применения, важность вопросов безопасности данных продолжает расти. Китайская система безопасности данных, система GDPR ЕС и меры регулирования США в отношении трансграничного потока конфиденциальных данных предъявляют различные требования к передаче и обработке данных. Многие данные, касающиеся информации о клиентах, финансовых данных, медицинских записей, коммерческих секретов и личных данных, часто не могут свободно пересекать границы.

Это означает, что в реальных бизнес-средах многие предприятия выбирают модели не только на основе цен на токены, но и должны отдавать приоритет соответствию данных, расположению хранилища данных и юридической ответственности. Для больших объемов реальных критически важных бизнес-данных затраты на соответствие часто намного превышают стоимость самого токена. Поэтому в реальном мире внутренние предприятия склонны использовать соответствующие внутренние модели, в то время как зарубежные предприятия склонны использовать зарубежные модели, что приводит к фактической сегментации рынка токенов.

Технологические барьеры определяют различия в методах сегментации токенов; системные барьеры определяют различия в системах вызова и

бизнес-экосистемах; политические барьеры определяют различия в потоках данных и границах соответствия. Действуя совместно, эти три барьера означают, что, хотя китайские токены и английские токены принадлежат к одной и той же технической концепции, они не могут свободно обмениваться, свободно обращаться и свободно взаимодействовать как идеализированная «глобальная универсальная валюта ИИ».

В будущем, вероятно, возникнет ситуация, аналогичная международной торговле: китайская токен-экосистема и английская токен-экосистема будут сосуществовать в течение длительного времени, каждая следуя независимым путем развития. Межзонный вызов не невозможен, но часто требует дополнительной технической адаптации, преобразования системы и затрат на юридическую проверку. Для большинства обычных пользователей такая межзонная способность не является повседневной потребностью, поэтому две большие системы, вероятно, останутся относительно независимыми в течение длительного времени.

Токены, произведенные в Китае, не обязательно ограничены использованием в китайских системах, но в ключевых сценариях, связанных с суверенитетом данных, ответственностью за соответствие требованиям и промышленной безопасностью, большая часть будущего спроса на токены, вероятно, будет преимущественно потребляться внутри отечественных моделей, платформ и систем.

С точки зрения долгосрочного тренда, связь между местом производства вычислительной мощности, поставщиком модели, оператором системы и местом нахождения данных усиливается. Для финансов, государственного управления, здравоохранения, цепочек поставок и крупных корпоративных систем данные часто не могут свободно перемещаться через границы. Поэтому, даже если Китай сможет экспортировать большие объемы токен-услуг по всему миру, большая часть высокоценного спроса на токены, вероятно, останется преимущественно в китайских моделях и системах для завершения. Это не означает, что китайские токены не могут быть

экспортированы, а скорее, что будущий рынок токенов может одновременно иметь две структуры: часть «глобального обращения» и часть «внутреннего замкнутого контура».

Стоит отметить, что то, что действительно может достичь унификации в будущем, — это не сами токены, а способ вызова токенов. По мере того как агентные платформы, унифицированные интерфейсные протоколы, платформы агрегации моделей и системы автоматизации будут продолжать развиваться, пользователи в конечном итоге сосредоточатся не на том, какой токен был использован, а на том, выполнена ли задача, разумна ли стоимость и достаточна ли эффективность. В то время токены, как электричество, интернет-трафик и облачные ресурсы сегодня, постепенно станут скрытыми за системами, становясь фундаментальной операционной инфраструктурой цифровой цивилизации.

Поэтому с долгосрочной точки зрения, хотя восточные и западные рынки токенов временно трудно напрямую связать, их направления развития сильно совпадают. Прошлые токены принадлежат компьютерной обработке языка, настоящее токенов принадлежит индустрии ИИ, а будущее токенов, вероятно, будет принадлежать фундаментальной операционной системе всей цифровой цивилизации.

VII. Почему способность к построению систем не может быть приобретена непосредственно через большие объемы токенов

Утверждая, что «способность к построению систем является дефицитным активом», я должен столкнуться с возможным возражением: если ИИ настолько мощен, почему способность к построению систем не может быть «сгенерирована» или «получена» путем потребления больших объемов

токенов? Ответ на этот вопрос раскрывает фундаментальные границы способности токенов.

Во-первых, построение систем требует решений, а решения — это не генерация.

ИИ может генерировать статью, код, шаблон веб-сайта, но ИИ не может заменить ключевые решения в системном проектировании, которые люди должны принимать – выбор модулей, отказ от некоторых функций, баланс между текущими потребностями и будущим расширением, когда реструктурировать, когда сохранять статус-кво. Эти решения зависят от понимания сути бизнеса, оценки ресурсных ограничений и предвидения долгосрочных последствий. Токены могут генерировать бесчисленные «возможные решения», но выбор «какое решение стоит реализовать» и «почему» — это то, что токены не могут заменить. Моя интеллектуальная логистическая система претерпела бесчисленные корректировки за два десятилетия, и каждая корректировка была не потому, что ИИ сказал мне «так лучше», а потому, что в реальном бизнесе возникла реальная проблема, обнажившая ограничения и выдвинувшая новые требования. За этими решениями стоит суждение, а суждение приходит из практики, а не из токенов.

Во-вторых, построение систем требует обработки исключений, и исключения не могут быть исчерпывающе перечислены.

Система может долгосрочно функционировать не потому, что ее нормальный поток идеально спроектирован, а потому, что она может обрабатывать исключения. В логистической системе исключения включают: неверные адреса клиентов, поврежденные товары, таможенные задержки, сбои платежей, отказы системы, потерю данных. В издательской системе исключения включают: колебания качества перевода, несовместимые форматы метаданных, изменения интерфейсов архивирования, сбои регистрации DOI. Эти исключения не могут быть исчерпывающе перечислены, потому что разнообразие исключений в реальном мире бесконечно. ИИ может изучать

«общие шаблоны исключений» с помощью больших объемов токенов, но не может предсказать «следующее исключение, которое никогда не появлялось ранее». Только долгосрочная система, постоянно сталкивающаяся с исключениями, записывающая их и решающая их на практике, может развить истинную способность к обработке исключений. Эта способность не может быть получена путем «скармливания большого количества токенов», потому что исключения — это не знания, а опыт.

В-третьих, построение систем требует компромиссов, и у компромиссов нет стандартных ответов.

Построение каждой системы включает компромиссы в условиях множественных ограничений: стоимость против производительности, скорость против стабильности, общность против индивидуализации, текущие потребности против будущего расширения. У этих компромиссов нет «правильного ответа», есть только «разумный выбор в конкретных обстоятельствах». ИИ может предложить «оптимальное решение», но этот «оптимум» является статистически оптимальным на основе исторических данных, а не суждением, основанным на «вашем конкретном контексте». Причина, по которой моя межпоколенческая двадцатилетняя интегрированная логистическая и финансовая система работает до сих пор, заключается именно в том, что я делал разные компромиссы в разные эпохи – выбирал определенный технологический стек в 2005 году, выбирал другой метод интеграции в 2015 году, выбирал путь обновления в 2025 году. Эти компромиссы были не «правильными» или «неправильными», а «подходящими для контекста того времени». Токены не могут делать эти компромиссы за меня, потому что суть компромиссов — это ценностное суждение, а не вычисление.

В-четвертых, построение систем требует циклов обратной связи, а обратная связь требует реального мира.

Совершенствование способности к построению систем зависит от замкнутого цикла: проектирование → реализация → эксплуатация → обнаружение проблем → исправление → перепроектирование. Шаг «обнаружение проблем» в этом цикле должен исходить из обратной связи реального мира. ИИ может моделировать работу, но не может моделировать все непредсказуемые факторы реального мира. Причина, по которой моя логистическая система постоянно улучшалась, заключалась в том, что реальные клиенты давали реальную обратную связь; причина, по которой моя издательская система постоянно совершенствовалась, заключалась в том, что реальные архивные интерфейсы претерпели реальные изменения. Эта обратная связь не может быть получена путем потребления токенов в закрытой среде. Токены могут генерировать «гипотетическую обратную связь от клиентов», но не могут заменить тот факт, что «клиент действительно дал эту обратную связь». Обратная связь реального мира — это единственная почва для роста способности к построению систем.

В-пятых, построение систем требует доверия, а доверие может исходить только от времени.

Доказывается ли система заслуживающей доверия, не рекламой, а временем. Причина, по которой мой веб-сайт постоянно архивируется CERN Zenodo, WorldCat и Национальной библиотекой Австралии, заключается не в том, что они «любят меня», а в том, что эти учреждения в ходе долгосрочного наблюдения подтвердили, что моя система стабильна, надежна и устойчива. Это доверие нельзя купить за большие объемы токенов. ИИ может сгенерировать «очень профессионально выглядящее» руководство по системе, но не может сгенерировать факт, что «этой системе доверяют уже два десятилетия». Доверие — это продукт времени, а не продукт токенов.

Резюмируя, способность к построению систем не может быть приобретена непосредственно через большие объемы токенов, потому что она по сути является не генеративной способностью, а способностью к суждению, обработке исключений, компромиссу, поглощению обратной связи и

накоплению доверия. Общей чертой этих способностей является то, что они требуют реального времени, практики, ошибок, исправлений и валидации. Токены могут ускорить обработку информации, но не могут ускорить накопление реального опыта. Это фундаментальная причина, по которой способность к построению систем останется дефицитной, даже когда токены станут чрезвычайно дешевыми в будущем.

VIII. Мои активы: Реальные исторические данные, междисциплинарные системы, способность к построению, дальновидное видение и способность к самоопровержению

Прежде чем перейти к обсуждению «что действительно является дефицитным», я должен сначала объяснить свою собственную позицию. Потому что мои суждения проистекают не из теоретической дедукции, а из долгосрочной практики. Это очень очевидно в различных системах, которые я построил за долгий период. Будь то интеллектуальная логистическая система, интеллектуальная система счетов, автоматизированная система финансовой обработки или более поздняя десяти-язычная издательская система, система метаданных, система навигации по веб-сайту и система автоматического перекрестного ссылаания, действительно сложная часть никогда не заключалась в генерации информации, а в ее организации. Информация может расти бесконечно, но без структуры она создаст лишь новый хаос. То же самое и в эпоху токенов. В будущем ИИ сможет генерировать огромные объемы контента, но огромные объемы контента не означают автоматически огромные объемы ценности. Напротив, по мере того как генерация контента становится проще, умение фильтровать, организовывать, сохранять, индексировать и связывать этот контент будет становиться все более важным.

Я обладаю следующими пятью активами, которые составляют эмпирическую основу для моего суждения о «чем является дефицит после токенов».

Во-первых, реальные исторические данные. Это не просто «исторические данные», а двадцать лет межпоколенческих реальных бизнес-данных – от интеллектуальной системы счетов 2005 года до настоящего времени, из реально работающих логистических, финансовых и налоговых систем, а не из экспериментальных сред или собранных данных. В эпоху ИИ, с распространением «синтетических данных», ценность реальных исторических данных резко возрастает. Большинство компаний обладают «однодоменными историческими данными», в то время как я обладаю междисциплинарными, охватывающими двадцать лет, практически проверенными историческими данными.

Во-вторых, завершение нескольких междисциплинарных систем и архитектур. Мои системы охватывают следующие области: интеллектуальная логистическая система, интеллектуальная система счетов, интеллектуальная финансовая система, интеллектуальная налоговая система, межпоколенческая двадцатилетняя интегрированная логистическая и финансовая система, система автоматизации веб-сайта, система генерации метаданных, десятиязычная издательская система «The Epochal Transition», система управления DOI, система организации ссылок TROVE и WorldCat, система автоматического перекрестного ссылаания, система генерации Sitemap, система генерации профиля Zenodo, кроссплатформенная издательская система. Уникальность этих систем заключается в: междисциплинарной широте (от логистики до издательского дела, от финансов до индексации литературы), временной глубине (некоторые системы работают уже два десятилетия) и завершении одним человеком (доказывая переносимость метода). Среди них, электронный ежемесячник «The Epochal Transition» (<https://times.net.au>) является публичным доказательством – он имеет международный стандартный номер ISSN (3083-5178), постоянно архивируется WorldCat (крупнейшей в мире системой сводных библиотечных каталогов, с членством более 16 000 национальных и университетских библиотек в более чем 100 странах и регионах), PANDORA/TROVE Национальной библиотеки Австралии, а также Zenodo (управляемым CERN, Европейской организацией ядерных исследований),

OpenAIRE (Европейской инфраструктурой открытой науки), DataCite DOI (глобальным цифровым идентификатором объектов, более 60 документов) и ORCID (глобальным уникальным постоянным академическим цифровым идентификатором). Этот веб-сайт доказывает, что один человек (независимый исследователь) может завершить весь процесс от создания контента до взаимодействия с ведущими международными архивными инфраструктурами.

В-третьих, способность строить системы и архитектуры. Это не способность «использовать инструменты ИИ», а способность проектирования для преобразования расплывчатых требований в стабильные структуры, архитектурная способность интегрировать различные системы в единую архитектуру, инженерная способность поддерживать работу систем годами без вмешательства человека и мета-способность переносить опыт из логистики в издательское дело, из финансов в литературу. Как обсуждалось ранее, эта способность не может быть приобретена непосредственно через большие объемы токенов, потому что она зависит от принятия решений, обработки исключений, компромиссов, циклов обратной связи и накопления доверия – ничто из этого не может быть заменено ИИ. Современный рынок обучения ИИ сосредоточен на поверхностных навыках, таких как «инжиниринг подсказок», «написание текстов с помощью ИИ» и «рисование с помощью ИИ»; почти нет курсов, преподающих «междисциплинарную способность к построению систем» – потому что эту способность нельзя ускорить, ее можно только накопить через долгосрочную практику. Более 40 независимых творческих работ, перечисленных на веб-сайте, полученный ISSN и более 60 DOI являются лучшими публичными записями долгосрочной работоспособности системы, а не моей саморекламой.

В-четвертых, дальновидное видение. Начало экспериментов с прототипом системы удаленной работы в 1993 году – почти за тридцать лет до того, как удаленная работа стала широко распространенной; разработка интеллектуальной системы счетов в 2005 году – начало автоматизации процессов за двадцать лет до распространения ИИ; межпоколенческая

двадцатилетняя интеграция систем – предвидение на момент проектирования системы, что она будет нуждаться в работе и через двадцать лет; вынесение суждения «после токенов» сейчас – в то время, когда большинство людей все еще гонятся за токенами, предвидение, что системная способность является истинным барьером. Я также предвидел, что «токены похожи на валюты, а технологические, системные и политические различия между Востоком и Западом делают их не универсальными» – как доллар США и китайский юань не могут свободно обмениваться, китайские токены и английские токены сталкиваются с аналогичными барьерами обращения, и в будущем могут появиться «китайские токен-зоны», «английские токен-зоны» и другие токен-экосистемы, основанные на языке и политике. Я также предвидел, что «зарабатывать деньги на токенах уже сложно и будет становиться все сложнее; только когда есть фронтальная система, предоставляющая услуги, токены и вычислительная мощность могут служить эффективными дополнительными активами для получения прибыли». Как независимый исследователь, я заранее создал персональную издательскую систему, взаимодействующую с будущей академической инфраструктурой, предвидя тенденцию, что «индивидуум также может стать узлом в глобальной инфраструктуре знаний». Большинство прогнозов тенденций исходят от аналитиков или предпринимателей, в то время как моя «способность предвидения» имеет подтверждение двадцатилетней практикой как гарантию.

В-пятых, способность к самоопровержению. Это одна из моих самых важных мета-способностей – все мои сочинения в различных областях достигают теоретической замкнутости и логической самосогласованности. Это означает, что мои аргументы не опираются на внешний авторитет или цитаты, а на самодостаточность внутренней логики. Стандартное правило в академической литературе – «цитировать предыдущие исследования»; мало кто заявляет или демонстрирует «способность к самоопровержению»; академическая дискуссия сосредоточена на том, «достаточно ли аргументы», а не на том, «есть ли у автора способность построить самосогласованный замкнутый цикл». В моих статьях эта способность к самоопровержению проявляется в следующем:

аргументы исходят из практики, практика подтверждает аргументы, а аргументы и практика образуют замкнутый цикл. Я не «цитирую теории других, чтобы доказать себя», а «показываю созданные мной системы и позволяю системам говорить самим за себя». Эта способность к самоопровержению является понятием, полностью отсутствующим в публичной литературе, и одной из моих самых уникальных мета-способностей.

Комбинация этих пяти активов, их дефицитность не аддитивна, а мультипликативна: реальные исторические данные × междисциплинарные системы × способность к построению систем × дальновидное видение × способность к самоопровержению образуют невоспроизводимый портфель активов. В публичных дискуссиях в настоящее время невозможно найти ни одного такого сопоставимого случая.

Процесс систематизации «Пути независимого международного исследователя» заставил меня осознать это еще глубже. С 1989 по 2026 год более тридцати лет опыта охватывают множество областей. Этот опыт возник не весь сразу, а формировался постепенно в течение длительного периода. ИИ может сгенерировать статью за секунды, но не может сгенерировать более тридцати лет реального опыта за секунды; может сгенерировать множество мнений, но не может напрямую сгенерировать суждение, накопленное долгосрочной практикой; может моделировать многие процессы, но не может заменить многократную проверку реальным миром. В этом смысле само время является важным активом, а структура, осажденная временем, является еще более дефицитным активом.

IX. После токенов: Пять действительно дефицитных способностей

Когда сами токены больше не будут дефицитными – барьер для продажи токенов невысок, но будет крайний недостаток реальных исторических данных, способности строить системы и архитектуры, а также дальновидного видения –

что же будет действительно дефицитным? Основываясь на моей долгосрочной практике, от прототипа системы удаленной работы 1993 года, интеллектуальной системы счетов 2005 года, интеллектуальной логистической системы, интеллектуальной финансовой системы, интеллектуальной налоговой системы, межпоколенческой двадцатилетней интегрированной логистической и финансовой системы, до автоматизации веб-сайтов, генерации метаданных, десяти-язычного ежемесячника «The Epochal Transition», управления DOI, организации ссылок TROVE и WorldCat, автоматического перекрестного ссыла, генерации Sitemap, генерации профиля Zenodo, кроссплатформенной публикации, я считаю, что следующие пять способностей станут истинными барьерами.

Во-первых, способность к структурному проектированию. Токены являются сырьем, а структура — это рамки, которые организуют сырье в ценные продукты. Задолго до распространения ИИ я уже начал думать о том, как использовать ограниченные ресурсы для достижения более высокой операционной эффективности. Моя интеллектуальная логистическая система, интеллектуальная система счетов и моя межпоколенческая двадцатилетняя интеллектуальная логистическая и финансовая система родились в разные эпохи, сталкивались с разными проблемами, но у них есть общая черта: цель заключалась не в создании дополнительной информации, а в уменьшении хаоса, уменьшении повторяющейся работы, повышении общей эффективности и создании структур, которые могут функционировать в течение длительного времени. В автоматизированной издательской системе ключевая проблема, с которой я столкнулся, заключалась не в том, «можно ли генерировать текст», а в том, «как определить поля метаданных для десяти языков, как установить связи между статьями и категориями, как разработать правила управления версиями». Эти структурные проектные решения определяют, сможет ли система продолжать работать годами без сбоя. Напротив, количество потребляемых токенов является лишь побочным продуктом. Веб-сайт «The Epochal Transition» сам по себе является публичным доказательством способности к структурному проектированию: лично поддерживаемая

академическая структура веб-сайта с международным стандартным номером ISSN, многоязычной поддержкой, несколькими архивными каналами, доказывающая, что я не занимаюсь пустыми разговорами, а успешно построил сложную, общедоступную систему, используя эту способность к структурному проектированию.

Во-вторых, способность к автоматизации. Токены могут массово производить контент, но только системы автоматизации могут массово выполнять сквозные рабочие процессы. В интеллектуальной логистической системе реальная ценность заключается не в том, сколько раз вызывается ИИ для планирования маршрутов, а в создании полностью автоматизированного замкнутого цикла от приема заказа, планирования склада, оптимизации маршрута до обработки исключений. В этом замкнутом цикле ИИ является лишь одним из компонентов, а способность к автоматизации проявляется в связях между компонентами. Человек без системы, даже имея большое количество токенов, может генерировать лишь разрозненный текст, разрозненные изображения, разрозненные таблицы и разрозненный код; человек с системой, напротив, может встроить токены в свой рабочий процесс, позволяя им автоматически организовывать материалы, генерировать метаданные, проверять ссылки, обновлять веб-страницы, помогать в публикации, архивировать литературу, создавать индексы, обрабатывать логистику, анализировать финансы и генерировать отчеты.

В-третьих, междисциплинарный практический опыт. Токены доступны каждому, но способность переносить опыт из одной отрасли в другую не является универсальной. Причина, по которой я могу построить как логистическую, так и издательскую систему, заключается не в том, что у меня больше токенов, а в том, что я накопил переносимые методологии, такие как управление метаданными, проектирование рабочих процессов и обработка исключений. Моя издательская система нуждается в токенах, потому что перевод на десять языков, генерация аннотаций, организация ключевых слов, генерация метаданных и описания ссылок требуют большой языковой обработки. Моя

система веб-сайта нуждается в токенах, потому что структуры страниц, описания навигации, SEO-описания, академические метаданные и многоязычные презентации могут быть автоматически сгенерированы. Моя логистическая система также может использовать токены, потому что электронные письма клиентов, описания товаров, таможенные документы, объяснения исключений, финансовые примечания и журналы операций могут обрабатываться с помощью ИИ. Моей системе автоматического перекрестного ссылаания токены нужны еще больше, потому что она должна находить тематические связи между большим количеством статей, заголовков, DOI, ссылок NLA, TROVE и WorldCat. Это междисциплинарное практическое накопление не может быть получено путем потребления токенов; оно может быть осаждено только решением реальных проблем. Веб-сайт одновременно представляет информацию из нескольких областей, включая академическое издательство, интеграцию инфраструктуры, техническую архитектуру и личный бренд, доказывая, что я действительно соединил издательское дело, библиотековедение, технологии и личный бренд, а не обладаю единичным навыком.

В-четвертых, ментальная модель долгосрочного функционирования

системы. Основное повествование о современных приложениях ИИ — это «диалог» и «генерация», но то, что действительно создает устойчивую ценность, — это «функционирование». Один диалог генерирует 10 000 токенов, одна задача агента потребляет миллион токенов, но логистическая система, непрерывно работающая в течение десяти лет, может потреблять миллиарды токенов – но что действительно дефицитно, так это ментальная модель для проектирования этой системы: как гарантировать, что система останется работоспособной после обновлений модели? Как обрабатывать изменения вышестоящих API? Как отслеживать исключения и автоматически восстанавливаться? Эти вопросы не имеют ничего общего с токенами, но они определяют конечный продукт токенов. Ценность этих систем исходит не от самих токенов, а от структур, которые я уже построил или разрабатываю. Токены — это всего лишь топливо; система — это двигатель. Токены — это

всего лишь электрический ток; структура — это цепь. Токены — это всего лишь вода; трубы, клапаны, резервуары и диспетчерские системы определяют, может ли она играть свою роль. Без системы токены рассеются; с системой токены могут быть преобразованы в долгосрочную ценность. Веб-сайт демонстрирует более 40 независимых творческих работ и постоянное архивирование авторитетными учреждениями, такими как CERN и Национальная библиотека Австралии, доказывая, что моя система не гонится за краткосрочным ажиотажем, а стремится к долгосрочной ценности и уже получила одобрение «постоянного сохранения» от авторитетных учреждений.

Однако, когда я глубже рассматриваю взаимосвязь между токенами и системами, я нахожу еще одну часто упускаемую из виду проблему. Многие люди считают, что если у них есть отличная система, они смогут получить преимущество в будущую эпоху токенов. Но реальность не так проста. Системы сами по себе — это не то, что может спроектировать каждый. По-настоящему зрелые системы часто требуют долгосрочной практики, многократного пересмотра и проверки в реальном мире. Многие системы могут нуждаться в годах или даже десятилетиях непрерывного накопления от первоначальной концепции до фактической реализации. Поэтому способность к проектированию систем сама по себе уже является дефицитным ресурсом.

Но даже обладая зрелой системой, это не автоматически приводит к широкомасштабному воздействию. Между технической осуществимостью системы и ее распространением в обществе лежат несколько слоев, включая рынки, капитал, каналы распространения, обучение пользователей и экосистемы платформ. История не испытывает недостатка в хороших технологиях или хороших системах; то, что действительно дефицитно, так это способность платформы вывести систему на широкомасштабное применение. Даже если система намного превосходит существующие решения в локальном сценарии, без достаточной капитальной поддержки, способности продвижения платформы и пользовательской базы ее фактический масштаб использования может оставаться очень ограниченным.

С точки зрения экономики токенов, это особенно очевидно. Спрос на токены исходит не непосредственно от самой системы, а от фактического масштаба использования системы. Платформа, обслуживающая миллионы пользователей в день, даже с относительно обычным дизайном системы, может генерировать огромное потребление токенов; наоборот, передовая система с очень небольшим количеством пользователей может иметь очень низкий спрос на токены. Поэтому будущее развитие рынка токенов зависит не только от способности моделей и способности систем, но и от способности платформ и способности к интеграции ресурсов.

Это означает, что будущая конкуренция имеет по крайней мере три уровня. Первый уровень — это конкуренция моделей, определяющая, кто может поставлять токены. Второй уровень — это конкуренция систем, определяющая, кто может эффективно использовать токены. Третий уровень — это конкуренция платформ, определяющая, скольких пользователей может охватить система и какой реальный спрос она может генерировать. Модели, системы и платформы — все незаменимы. Без моделей системы не могут работать. Без систем токены с трудом создают ценность. Без платформ даже лучшая система может остаться в масштабах малого использования.

С этой точки зрения, я еще больше убежден, что токены сами по себе являются лишь фундаментальными ресурсами. То, что действительно определяет будущее, — это не только те, кто владеет токенами, и не только те, кто владеет системами, а те, кто может соединять модели, системы, платформы и реальные потребности. Только когда эти элементы образуют замкнутый цикл, токены превратятся из простых вычислительных единиц в постоянно создающие ценность производительные силы.

В-пятых, способность встраивать токены в реальные бизнес-процессы. Токены в конечном итоге должны служить какому-то звену в реальном мире — логистике, издательскому делу, образованию, здравоохранению. То, что действительно дефицитно, — это те, кто понимает как границы возможностей ИИ, так и болевые точки конкретной отрасли. Они знают, когда использовать

ИИ, а когда нет; знают, как преобразовать выводы токенов в бизнес-решения; знают, как измерить реальную связь между затратами на токены и результатами бизнеса. Эта составная способность является в настоящее время самым дефицитным ресурсом на рынке. Веб-сайт сам по себе является работающим «бизнес-процессом»: создание контента → генерация метаданных → кроссплатформенная публикация → международное архивирование → доступ читателей. Это доказывает, что мое применение ИИ (потребление токенов) встроено в полный, замкнутый издательский процесс, производящий реальное влияние.

Я все больше чувствую, что истинная конкуренция в будущем, вероятно, будет не между моделями и не между токенами, а между системами. Модели будут продолжать обновляться, цены на токены будут продолжать падать, и сегодняшние дорогие ресурсы станут очень распространенными в будущем. Но системы не появятся автоматически из-за изобилия ресурсов. Системы требуют долгосрочного накопления, непрерывного пересмотра и проверки в реальном мире. Тот, кто сможет превратить токены в долгосрочные работающие системы, с большей вероятностью получит устойчивую ценность. И это также ключ к прибыльности: зарабатывать деньги на токенах уже сложно, и будет становиться все сложнее; только когда есть фронтальная система, предоставляющая услуги, токены и вычислительная мощность могут служить эффективными дополнительными активами для получения прибыли.

Я полагаю, что настоящая конкуренция в эпоху токенов будет разворачиваться в три этапа: первый этап — это конкуренция за ресурсы, все борются за модели, вычислительную мощность, API и цены на токены; второй этап — это конкуренция приложений, все используют токены для письма, обслуживания клиентов, кодирования, перевода, образования, офисной работы и анализа данных; третий этап — это настоящая системная конкуренция, где тот, кто сможет встроить токены в долгосрочные отраслевые системы, издательские системы, логистические системы, системы знаний и системы записи цивилизации, с большей вероятностью получит долгосрочное преимущество.

На самом деле, при обсуждении токенов, моделей, систем и платформ есть еще один важный фактор, часто упускаемый из виду в реальном мире: время. Многие люди считают, что, имея капитал, они могут быстро построить системы; при достаточно большом рынке они могут быстро воспроизвести успешные модели. Однако реальность часто не такова. Капитал может купить вычислительную мощность, может приобрести токены, может нанять команды разработчиков, но не может напрямую купить реальные данные и зрелые системы, которые проверялись в течение многих лет. Для многих отраслей то, что действительно дефицитно, — это не сами модели, не сами токены, а системы, которые долгое время стабильно работали в реальных бизнес-средах.

С деловой точки зрения, рынки и капитал обычно долго не ждут, пока новая система вырастет с нуля. Крупная система от предложения концепции, проектирования архитектуры, разработки функций, валидации процессов до фактической реализации часто требует многих лет или даже больше. В течение этого процесса требуются не только постоянные вложения ресурсов, но и множество проб и ошибок, исправлений и проверок в реальной среде. Поэтому, когда рыночные возможности уже появились, капитал, как правило, предпочитает системы, которые уже сформированы, уже работают и уже проверены, а не проектирование совершенно новой системы с нуля.

Однако в реальности существует и другое противоречие. Чем крупнее система, тем труднее последующему капиталу легко вмешаться. После долгосрочного развития крупные системы обычно уже сформировали свои собственные механизмы работы, управленческие структуры, логику данных и организационную культуру. Они могут обладать обширными историческими данными, сложными бизнес-процессами и накопленными за долгое время системами опыта. Из-за этого, хотя крупные системы и являются чрезвычайно ценными, их часто трудно быстро трансформировать или полностью реструктурировать в соответствии с ожиданиями капитала. Чем крупнее система, тем больше ее инерция.

Поэтому в реальном деловом мире идеальное состояние — это не проектирование системы с нуля и не простое приобретение крупной, неповоротливой системы, а поиск систем, которые уже завершили свое основное структурное строительство, были проверены в течение длительного времени и могут стабильно работать. Даже если такие системы не идеальны, они имеют очень высокую ценность, потому что они уже преодолели самый трудный этап, обладая реальными данными, реальными процессами и реальными записями проверки. На этой основе переупаковка, оптимизация, открытость внешнему миру и продвижение через платформы часто более эффективны, чем начало с нуля, и могут также значительно снизить риски проб и ошибок и временные затраты. Потому что эти системы уже завершили самый дорогой, самый трудоемкий и самый трудновоспроизводимый этап.

С этой точки зрения, временное окно также является часто упускаемым из виду ресурсом. Когда отрасль находится на ранней стадии, многие люди сосредотачиваются на возможностях моделей, ценах на токены и размере рынка, игнорируя время, необходимое для формирования системы. Рождение зрелой системы обычно проходит несколько этапов: предложение концепции, проектирование архитектуры, валидация процессов, работа в реальном мире и непрерывная оптимизация. Действительно трудная часть — не предложение идеи, а поддержание стабильной работы системы в реальной среде в течение длительного времени. Поэтому, когда рынок начинает осознавать ценность зрелых систем, это часто означает, что эти системы уже прошли через годы накопления, а не то, что они могут быть воспроизведены за короткое время.

Для капитала и рынка наиболее идеальными целями являются не идеи на концептуальной стадии и не непроверенные прототипы, а зародышевые системы, которые завершили основное структурное строительство, обладают реальными записями о работе и имеют потенциал для улучшения. Хотя такие системы могут быть не идеальны, они уже преодолели самый трудный, самый трудоемкий и самый рискованный этап. Как только рыночный спрос быстро растет, такие системы часто становятся очень востребованными, потому что

количество действительно проверенных временем зрелых систем обычно намного меньше, чем предполагает рынок. Модели могут быстро итерироваться, токены могут непрерывно расширяться, команды разработчиков могут быть временно собраны, но реальные данные, реальные процессы и долгосрочные записи проверки не могут быть воссозданы за короткое время.

Поэтому в процессе развития экономики токенов время не только влияет на формирование систем, но и влияет на ценность систем. Когда рынок еще не полностью осознал важность определенного типа систем, часто это также период, когда стоимость приобретения таких систем самая низкая; а когда рынок в целом осознает их ценность, доступные варианты часто уже очень ограничены. Для многих отраслей упускается не просто проект, а временное окно для входа в следующую стадию развития, которое уже завершило долгосрочное накопление.

С точки зрения экономики токенов, то же самое применимо. Токены могут быть быстро получены, модели могут быть быстро подключены, платформы могут быть быстро построены, но то, что действительно трудно быстро воспроизвести, — это системы, которые осаждались в течение многих лет и могут продолжать производить ценность. В будущем, когда токены станут все более обильными, важность систем не только не уменьшится, но может даже возрасти. Потому что, когда у всех будут схожие токен-ресурсы, результат конкуренции будет определяться уже не количеством токенов, а тем, у кого более зрелые реальные данные, более стабильные бизнес-процессы и системные структуры, более близкие к реальным потребностям.

Еще один часто упускаемый из виду вопрос — это происхождение систем. Многие люди привыкли понимать системы как программные продукты и поэтому естественно предполагают, что системы должны проектироваться и разрабатываться компаниями-разработчиками программного обеспечения. Однако во многих сложных отраслях системы, которые действительно стабильно работают в течение длительного времени, часто не сначала

проектируются компаниями-разработчиками ПО, а затем передаются отрасли, а скорее постепенно формируются в ходе долгосрочной деловой практики, а затем превращаются в программное обеспечение, автоматизируются и становятся платформами.

Причина не сложна. Компании-разработчики ПО обычно преуспевают в технической реализации, но они не обязательно остаются в конкретном бизнес-контексте в течение длительного времени. Они могут быстро разрабатывать функции, но им трудно приобрести большой объем реального бизнес-опыта за короткое время. Напротив, системы, которые органически развиваются внутри отрасли, постоянно пересматриваются и улучшаются ежедневно перед лицом клиентов, заказов, исключений, нормативных изменений, ценового давления и рыночной конкуренции. Такие системы исходят не из теории, а из реальных проблем; каждый их процесс, каждое их правило, каждое звено автоматизации обычно происходит из долгосрочной проверки реальной работы.

Поэтому в логистике, цепочках поставок, издательском деле, финансах, производстве, здравоохранении и многих других профессиональных областях наиболее ценными системами часто являются не самые сложные системы и не системы с наибольшим количеством функций, а те, которые прошли долгосрочную проверку и могут стабильно решать реальные проблемы. У них могут не быть самых передовых интерфейсов, самой броской рекламы или самой новой технической архитектуры, но они обладают большими объемами реальных данных, реальных процессов и долгосрочно накопленной бизнес-логики. Этот опыт часто не может быть получен путем краткосрочной разработки.

С точки зрения экономики токенов, это особенно важно. В будущем возможности моделей могут продолжать улучшаться, затраты на токены могут продолжать снижаться, а эффективность разработки программного обеспечения будет становиться все выше. Однако бизнес-опыт, накопленный отраслью за десятилетия, не возникнет автоматически из-за обновлений моделей. Что действительно трудно воспроизвести, так это не сам

программный код, а отраслевые знания, опыт обработки исключений, логика принятия решений и проверенные временем бизнес-процессы, скрытые за системой.

Поэтому я считаю, что самыми идеальными системами в будущем будут не те, которые разработаны в полном отрыве от отраслевой практики, а те, которые происходят из реального бизнеса, проверены долгосрочным функционированием, а затем постоянно улучшаются с помощью искусственного интеллекта и возможностей токенов. Такие системы обладают как надежностью реального мира, так и приростом эффективности, принесенным эпохой ИИ, и их долгосрочная ценность часто намного выше, чем у систем, построенных исключительно на технических концепциях.

Х. Куда текут токены: От мимолетных разговоров к памяти цивилизации

Относительно будущей экономики токенов у меня есть еще одно наблюдение. Многие люди сосредотачиваются на том, как производятся токены, но мало внимания уделяется тому, куда они в конечном итоге попадают. Если токен остается только в одном чате, его жизненный цикл может составлять всего несколько минут; если он попадает на обычную веб-страницу, его жизненный цикл может составлять всего несколько месяцев; если он попадает в систему долгосрочного хранения, его ценность принципиально меняется. Когда токены попадают в DOI, Zenodo, WorldCat, TROVE, системы долгосрочного хранения национальных библиотек и различные академические и знаниевые инфраструктуры, они уже не просто токены; они начинают превращаться в записи знаний, литературные активы и память цивилизации. Поэтому я считаю, что действительно важный вопрос в будущем — не как произвести больше токенов, а как позволить токенам попасть в структуры, где они могут быть долгосрочно сохранены и продолжать играть свою роль.

Именно по этой причине я сохраняю оптимизм в отношении будущего развития токенов, но я не слепо верю в сами токены. Прошлые токены — это базовая единица для обработки языка компьютерами; настоящие токены — это важная единица измерения и коммерческая единица в эпоху ИИ; будущее токенов, вероятно, станет важным фундаментальным ресурсом цифровой цивилизации. Однако независимо от того, как будут развиваться токены, я всегда придерживаюсь мнения, что действительно дефицитным является не токены, а люди, способные управлять токенами; не те, у кого больше ресурсов, а те, кто может создавать долгосрочные структуры; не те, кто может генерировать больше контента, а те, кто может организовывать контент в системы знаний.

XI. Несжимаемость времени: Четыре непокупаемых актива

При обсуждении «что действительно является дефицитным после токенов» я должен указать на фундаментальную проблему, которую упускает из виду большинство дискуссий: некоторые активы по своей природе не могут быть ускорены, куплены или воспроизведены. Как бы ни был мощен ИИ, как бы дешевы ни были токены, следующие четыре актива всегда будут сохранять свою неотъемлемую дефицитность.

Во-первых, время несжимаемо.

Тридцать лет практики — это просто тридцать лет; их нельзя сжать до трех лет, ни смоделировать с помощью ИИ. ИИ может сгенерировать статью о том, «как построить логистическую систему», за секунды, но он не может за секунды пережить различные аномалии, сбои, изменения политики и рыночные колебания, с которыми логистическая система сталкивается за два десятилетия. Эти накопления времени в реальном мире не могут быть сокращены никакими технологическими средствами. Мое начало экспериментов с прототипом системы удаленной работы в 1993 году, разработка интеллектуальной системы счетов в 2005 году — эти временные

точки не случайность; это необратимые маркеры реальной истории. Любой, кто придет позже, может читать мои статьи и изучать мои методы, но он не может обладать годами, которые я прожил — эти годы сами по себе являются активами.

Во-вторых, междисциплинарный опыт нельзя купить.

Междисциплинарный опыт нельзя купить как товар. Человек может купить сотню книг по логистике, издательскому делу, финансам и индексации литературы за один день, но он не может приобрести в тот же день практический опыт объединения этих областей. Истинная междисциплинарная способность — это не накопление знаний, а способность устанавливать связи между различными областями, выявлять общие черты и переносить методы. Эта способность может быть приобретена только путем решения реальных проблем, совершения ошибок, исправления путей и накопления суждений в одной области за другой. Причина, по которой я могу перенести свой опыт из логистических систем в издательские системы, заключается не в том, что я прочитал какую-то книгу, а в том, что я провел двадцать лет в логистике и еще десять в издательском деле — именно на пересечении этих двух временных периодов возникла переносимая методология.

В-третьих, способность к построению систем трудно воспроизвести.

Способность к построению систем — это не набор кода или процессов, который можно скопировать и вставить. Код можно скопировать, но логику принятия решений, лежащую в основе кода, компромиссные суждения и предвидение будущих изменений скопировать нельзя. Ценность системы заключается не только в том, что она может сделать сегодня, но и в том, сколько изменений, сколько восстановлений после сбоев и сколько корректировок требований она претерпела за последнее десятилетие, продолжая функционировать. Эта «история системы» сама по себе является доказательством способности к построению. Мой веб-сайт (<https://times.net.au>) может быть просмотрен любым, но то, что видят посетители, — это результат, а не тридцатилетняя история

принятия решений, выбора технологий, структурных корректировок и непрерывного обслуживания, стоящая за этим результатом. Эту способность к построению нельзя получить путем плагиата или подражания.

В-четвертых, записи о долгосрочном функционировании нельзя ускорить.

Тот факт, что система работала два десятилетия, сам по себе не может быть ускорен. Запись о долгосрочном функционировании доказывает стабильность системы, удобство обслуживания и предвидение в проектировании. В эпоху ИИ создание системы или веб-сайта, которые «выглядят похоже», может быть выполнено за минуты, но создание «системы, которая работает уже два десятилетия и постоянно архивируется CERN, WorldCat и Национальной библиотекой Австралии», не имеет быстрых путей. Запись о долгосрочном функционировании — это временная метка, объявляющая всем, что эта система выдержала испытание реальным миром. Такая запись не может быть подделана или ускорена никакой технологией ИИ.

Эти четыре актива — время, междисциплинарный опыт, способность к построению систем и записи о долгосрочном функционировании — вместе составляют основу того, что я назвал в этой статье «действительно дефицитным». Я обладаю ими не потому, что я умнее или трудолюбивее других, а потому, что я начал раньше других и продолжаю до сих пор. В эпоху, когда ИИ делает все больше вещей дешевыми и легкодоступными, эти вещи, которые нельзя сжать, купить, воспроизвести или ускорить, будут становиться все более ценными.

XII. Заключение: От «накопления токенов» к «построению систем»

Будущее токенов заключается не в самих токенах.

С непрерывным развитием технологии больших моделей и постоянным снижением стоимости вывода (inference) предложение токенов будет продолжать расширяться. Токены, которые сегодня дороги, через несколько лет будут такими же дешевыми и повсеместными, как сегодня интернет-трафик и хранилище на жестких дисках. Когда каждый сможет легко получать большие объемы токенов, то, что будет действительно дефицитным, уже не будут токены, а способность превращать токены в реальную ценность — способность к структурному проектированию, способность к автоматизации, отраслевой опыт и долгосрочное практическое накопление. А более глубокая дефицитность заключается в тех вещах, которые нельзя сжать, купить, воспроизвести или ускорить: само время, междисциплинарный опыт, способность к построению систем и записи о долгосрочном функционировании. Как было сказано ранее, способность к построению систем не может быть приобретена непосредственно через большие объемы токенов, потому что она по сути является не генеративной способностью, а способностью к суждению, обработке исключений, компромиссу, поглощению обратной связи и накоплению доверия — общей чертой этих способностей является то, что они требуют реального времени, практики, ошибок, исправлений и валидации.

На уровне прибыльности это суждение также применимо: зарабатывать деньги на токенах уже сложно, и будет становиться все сложнее. Ценовые войны постоянно сжимают нормы прибыли, и бизнес-модель простой перепродажи токенов быстро движется к мизерной или даже нулевой прибыли. Только когда токены привязаны к фронтальной системе — т.е. когда система предоставляет услуги, — токены и вычислительная мощность могут служить эффективными дополнительными активами для получения прибыли. Токены сами по себе не являются источником прибыли; система является источником прибыли. Токены — это всего лишь топливо для системы, а не ее двигатель.

С этой точки зрения, системы, которые я уже завершил или разрабатываю, имеют особое значение для будущего развития токенов. Потому что эти системы не были временно построены в погоне за мимолетным трендом; они

существовали задолго до того, как концепция токенов стала публичной темой. Интеллектуальная логистическая система доказала, что реальный бизнес может быть структурирован, процедурен и автоматизирован; издательская система доказала, что один человек может также создать многоязычную, мультиплатформенную, мульти-DOI, многоканальную литературную систему; система веб-сайта и метаданных доказала, что цифровой контент может быть непрерывно организован, индексируем и распространен; система автоматического перекрестного ссылаания доказывает, что когда количество работ продолжает увеличиваться, действительно важным является не продолжение накопления контента, а установление структурных связей между контентом. Веб-сайт «The Epochal Transition» (<https://times.net.au>) как центральная платформа для демонстрации этих достижений сам по себе является эмпирическим доказательством превращения способности токенов в открытую, долговечную, проверяемую академическую систему.

Даже если всем дадут одинаковое количество токенов в будущем, результаты будут совершенно разными: некоторые получают больше записей разговоров, некоторые получают больше веб-контента, в то время как другие могут создать логистическую систему, издательскую систему или систему знаний, которая будет непрерывно работать десятилетиями. Разницу определяют не сами токены, а лежащая в основе системная архитектура — и несжимаемое время, непокупаемый междисциплинарный опыт, трудно воспроизводимая способность к построению и невозможные ускорить записи о долгосрочном функционировании, которые поддерживают эти системные архитектуры.

Поэтому я не рассматриваю токены как цель, а как новую инфраструктуру интеллектуальной эпохи. Токены так же важны, как электричество, но электричество само по себе не строит заводы; токены так же измеримы, как деньги, но деньги сами по себе не создают качественные активы; токены могут течь, как вода, но без труб, диспетчерских систем и долгосрочных записей о техническом обслуживании они просто рассеются. По-настоящему ценными людьми в будущем будут не обязательно те, у кого больше всего токенов, а те,

кто может организовать токены в системы. Самое большое заблуждение эпохи токенов — рассматривать токены как цель; самая большая возможность эпохи токенов — использовать токены для создания структур, которые другие не могут воспроизвести. Когда в будущем токены станут такими же обычными, как электричество, люди, оглядываясь на сегодняшнюю токен-лихорадку, возможно, обнаружат, что то, что действительно определяло разницу между людьми, никогда не было количеством принадлежащих им токенов, а было ли у них способность постоянно учиться, постоянно практиковаться, постоянно накапливать и постоянно строить системы — и те вещи, стоящие за этими способностями, которые нельзя сжать, купить, воспроизвести или ускорить. И на мой взгляд, эта способность и является самым дефицитным и самым трудновоспроизводимым богатством в эпоху токенов.

По мере углубления моих исследований я все больше чувствую, что конкурентный ландшафт будущей эпохи ИИ можно подвести под простую аналогию. Многие люди рассматривают вычислительную мощность, модели и токены как основные ресурсы, но с точки зрения функционирования реального мира они больше похожи на логистическую систему позади армии. Вычислительная мощность определяет производственные мощности, модели определяют организационные способности, а токены определяют распределительные способности; вместе они составляют систему материального обеспечения эпохи ИИ.

Однако то, что действительно определяет исход битвы, — это не сами логистические запасы, а боевая система на передовой. Для будущей промышленной конкуренции системы больше похожи на линию атаки. Системы определяют, как используются ресурсы, как выполняются процессы, как создается ценность и действительно ли выигрывают пользователи. Как бы ни были обильны токены и вычислительная мощность, без организации системы они подобны запасам, сложенным на складе, которые не могут быть преобразованы в реальные результаты.

С этой точки зрения, будущая конкуренция, вероятно, будет не между вычислительными мощностями и не между токенами, а между системами. Система — это линия атаки, определяющая направление атаки и боевые возможности; токены — это боеприпасы и запасы, определяющие устойчивую боевую способность; вычислительная мощность — это тыловая промышленная база, определяющая скорость производства и способность снабжения. Без тылового снабжения линия атаки не может долго продвигаться; но с одним только тыловым снабжением и без системы линии атаки также нельзя достичь результатов.

Поэтому, когда вычислительная мощность станет более распространенной, модели станут более похожими, а токены станут более обильными, по-настоящему дефицитной останется система. Тот, у кого более зрелая система, более реальные данные, более длительно проверенные процессы и способность соединять токены и вычислительную мощность с реальными потребностями, с большей вероятностью получит долгосрочное преимущество в будущем.

Токены становятся «электричеством» эпохи ИИ — незаменимыми, но не самоцелью. Как победителями электрической революции были не электростанции, а изобретатели и предприниматели, которые изобрели лампочку, двигатель и электролизер, настоящими победителями эпохи ИИ будут строители систем, способные использовать токены для создания невозпроизводимых структур и архитектур. После токенов — вот что действительно дефицитно.

Оглядываясь на всю эту статью, я все больше чувствую, что конкурентный ландшафт будущей эпохи ИИ можно проиллюстрировать аналогией с деревом.

Многие люди сосредотачиваются на воде и питательных веществах, потому что они видимы, осязаемы и легко измеримы. Сегодняшняя вычислительная мощность, модели и токены подобны постоянно подаваемой воде и питательным веществам. Вычислительная мощность определяет

производственную способность, модели определяют способность к трансформации, а токены определяют распределительную способность. По мере развития технологий эти ресурсы становятся все более обильными и легкодоступными.

Однако то, что действительно определяет жизнеспособность дерева, — это не вода и питательные вещества сами по себе, а корневая система.

Система — это корень будущей цифровой цивилизации.

Корневая система определяет, как вода впитывается, как распределяются питательные вещества и может ли все дерево расти долгосрочно. Без корней, сколько бы воды ни было, она утечет; без корней, как бы ни были обильны питательные вещества, они не могут быть преобразованы в жизненную силу. Точно так же без системы, как бы ни была мощна модель, как бы дешевы ни были токены, как бы огромна ни была вычислительная мощность, трудно создать ценность на постоянной основе.

С этой точки зрения, вычислительная мощность — это вода, токены — это питательные вещества, модели — это трубы, доставляющие питательные вещества, а система — это корневая система, глубоко проникающая в реальный мир. Что действительно определяет исход будущей конкуренции, так это не у кого больше воды, и не у кого больше питательных веществ, а у кого более глубокие, более стабильные корни, способные долгосрочно поглощать и использовать эти ресурсы.

В будущем, когда вычислительная мощность продолжит расти, модели продолжат развиваться, а цены на токены продолжат падать, сами ресурсы станут все более обычными. То, что будет действительно дефицитным, уже не будут ресурсы, а системы, способные организовывать, использовать и постоянно создавать ценность из ресурсов. Тот, кто обладает зрелой системой, обладает корнями; тот, кто обладает корнями, обладает способностью к долгосрочному выживанию и постоянному росту.

Поэтому я считаю, что после токенов действительно дефицитным являются не больше токенов и не большая вычислительная мощность, а системы, способные преобразовывать вычислительную мощность, модели и токены в реальную ценность.

Когда вода и питательные вещества становятся все более обильными, то, что действительно определяет, может ли дерево расти долгосрочно, всегда является корневая система.

Ключевые слова: Токен, Искусственный интеллект, ИИ, Большая языковая модель, Вычислительная мощность, Экономика моделей, Экономика токенов, Китайский токен, Английский токен, Алгоритм ВРЕ, Суверенитет данных, Агент, API-экосистема, Способность к построению систем, Системное проектирование, Система автоматизации, Междисциплинарное исследование, Междисциплинарная практика, Реальные исторические данные, Система долгосрочного функционирования, Временной актив, Платформенная конкуренция, Системная конкуренция, Цифровая цивилизация, Память цивилизации, Инфраструктура знаний, Независимый исследователь, Независимый международный исследователь, Интеллектуальная логистическая система, Десяти-язычная издательская система, Система метаданных, Автоматическое перекрестное ссылаание, DOI, Zenodo, WorldCat, TROVE, The Epochal Transition, У Чжаохуэй, Джеффи Чао Хуэй Ву

Литература:

1. *A Structural Case Study of the 1993 Remote Work Prototype System*
DOI: <https://doi.org/10.5281/zenodo.17978371>
2. *The 2005 Intelligent Invoice System*
DOI: <https://doi.org/10.5281/zenodo.19044755>
3. *Mature Structural Logistics System*
DOI: <https://doi.org/10.5281/zenodo.20201895>
4. **Cross-Generational 20-Year Intelligent Logistics and Financial System**
DOI: <https://doi.org/10.5281/zenodo.20352371>
NLA: <https://nla.gov.au/nla.obj-4204311557>
5. *Analysis of the Homepage Structural Evolution of The Epochal Transition*
DOI: <https://doi.org/10.5281/zenodo.20362574>
NLA: <https://nla.gov.au/nla.obj-4204309341>
6. *An Empirical Study of Institutional Engineering: Independent Scholars Accessing the WorldCat Global Bibliographic Infrastructure*
DOI: <https://doi.org/10.5281/zenodo.18028572>
7. *National-level Network Archiving and Digital Document Preservation Guidelines of the Australian Winner Information Network*
DOI: <https://doi.org/10.5281/zenodo.17888259>
8. *The Institutional Archiving of The Epochal Transition*
DOI: <https://doi.org/10.5281/zenodo.17932379>

WU, J. C. H. (2026). Source Declaration for Audiovisual and Derivative Adaptations of a Continuing Real-World Narrative. Zenodo.

<https://doi.org/10.5281/zenodo.18160116>

토큰 이후: AI 시대에 진정으로 희소한 것은 무엇인가

저자/국제 독립 연구자: 우자오후이 (JEFFI CHAO HUI WU)

ORCID: <https://orcid.org/0009-0009-5318-3027>

【사전 진술】

본 글은 30 년 이상의 학제간 실천(1989 년 호주 유학 시작 이후 원격 근무, 물류 관리, 자동화 처리, 웹사이트 구축, 문화 전파, 다국어 출판, 학제간 연구에 이르기까지)에 기반한 개인적 견해입니다. 토큰은 빠르게 진화하는 새로운 현상이며, 본 글은 집필 시점(2026 년 6 월)까지의 개인적 단계적 관찰 및 판단만을 나타낼 뿐, 투자 조언, 상업적 예측 또는 업계 표준을 구성하지 않습니다.

본 글은 보편적 적용 가능성을 추구하지 않으며, 어떤 업계 합의를 대표한다고 주장하지 않습니다. 토큰 시장, AI 기술 및 글로벌 정책 환경의 지속적인 변화에 따라, 본 글의 일부 판단은 향후 수정 또는 업데이트가 필요할 수 있습니다. 독자는 가장 최신 정보에 기반하여 독립적으로 판단해야 합니다.

제가 논의하는 '다섯 가지 희소 능력' 및 '다섯 가지 개인 자산'은 제 자신의 시스템 구축 실천에서 비롯됩니다(공개 증거는 <https://times.net.au> 에서 확인 가능하며, 해당 사이트에 게재된 40 개 이상의 독립 창작 작품, ISSN 3083-5178, 60 개 이상의 DOI, 그리고 WorldCat, 호주 국립 도서관, CERN Zenodo 와 같은 국제 인프라에 의해 영구 보관된 기록 포함).

맥락을 벗어난 인용을 피하기 위해, 본 글에 대한 유효한 반박 또는 질문의 전제 조건은 본 문서 내의 모든 링크 기사 및 그 내포된 링크, 그리고 <https://times.net.au> 에 게재된 제 모든 작품을 완전히 읽는 것입니다. 이 전체 맥락에서 분리된 부분적 인용 또는 단순화된 요약은 본 글에 대한 공정한 평가를 구성할 수 없습니다.

유효한 반박에 대한 추가 설명:

본 글은 개인적 견해이며, 보편적 적용 가능성을 추구하지 않고 타인의 문헌 인용을 전제로 하지 않습니다. 유효한 질문은 '사과 대 사과' 비교를 채택해야 합니다: 사실 대 사실, 개인 실천 대 개인 실천, 완성된 시스템 대 완성된 시스템. '미래 AI 가 할 수 있는 것'으로 '현재 완료된 실천'을 부정하거나, 팀/기관의 능력을 개인 실천과 비교하는 것은 공정한 비교가 아닙니다.

본 글에서 설명하는 5 가지 자산 포트폴리오는 개인 실천 기록이며 '대상이 될 만한 사례가 전혀 존재하지 않는다'고 주장하지 않습니다. 완전히 대상이 될 만한 사례가 존재한다고 생각되면, 구체적인 공개 사례를 제시하고 그것이 실제 역사 데이터, 학제간 시스템, 시스템 구축 능력, 미래 예측 비전, 자기 증명 능력이라는 다섯 가지 차원에서 어떤 상응하는 증거를 가지고 있는지 설명하십시오. 그렇지 않다면, 관련 질문은 단지 의견 표현에 불과하며 유효한 반박이 아닙니다.

초록

토큰은 AI 시대의 핵심 측정 단위가 되고 있습니다. 그러나 본 글은 토큰 자체가 최종 가치가 아님을 논증할 것입니다. 추론 비용이 지속적으로 감소함에 따라 토큰은 오늘날의 인터넷 트래픽이나 하드 디스크 저장 장치처럼 저렴하고 보편화될 것입니다. 토큰이 더 이상 희소하지 않을 때, 진정으로 희소한 것은 무엇일까요? 저자의 30 년 이상의 학제간 실천(1993 년 원격 근무 프로토타입 시스템, 2005 년 지능형 송장 시스템에서 지능형 물류 시스템, 10 개국어 출판 시스템, DOI 관리, 자동 상호 참조 및 크로스 플랫폼 출판에 이르기까지)에 기반하여, 본 글은 다음과 같은 핵심 판단을 제시합니다: 첫째, 시스템 구축 능력(구조 설계, 자동화, 학제간 실천, 장기 운영 멘탈 모델 및 토큰을 실제 비즈니스 프로세스에 내장하는 능력 포함)이야말로 진정한 희소 자산입니다. 둘째, 시스템 구축 능력은 대량의 토큰을 통해 직접 획득할 수 없습니다. 이는 의사 결정, 예외 처리, 균형, 실제 세계로부터의 피드백 루프 및 시간이 지남에 따른 신뢰 축적에 의존하기 때문입니다. 셋째, 시간의 비압축성, 학제간 경험의 비구매 가능성, 시스템 구축 능력의 복제 어려움, 장기 운영 기록의 단기간 획득 불가능성 – 이 네 가지 자산이 토큰 시대의 가장 깊은 희소성을 구성합니다. 본 글은 저자가 구축한 학술 웹사이트 '시대적 전환(The Epochal Transition)'(<https://times.net.au/>, ISSN 3083-5178, WorldCat, CERN Zenodo, 호주 국립 도서관 TROVE 등의 국제 인프라에 영구 보관됨)을 공개 증거로 제시하며, 한

사람이 어떻게 30 년에 걸쳐 학제간 검증 가능한 장기 운영 시스템 아키텍처를 구축할 수 있는지 보여줍니다.

본 글의 주장: 토큰은 단지 연료일 뿐이며, 시스템이 바로 엔진이다. 토큰 이후, 진정으로 희소한 것은 토큰을 다룰 줄 아는 사람, 그리고 그 사람이 구축한 구조들이다.

"상품 없이, 돈이 무슨 소용인가?"

토큰 발전의 역사를 되돌아볼 때, 그것이 인류가 과거에 경험했던 많은 기술 혁명과 매우 유사하다는 점을 점점 더 절감합니다. 새로운 기초 자원이 등장할 때마다 사람들은 먼저 자원 자체에 집중하고, 자원 뒤에 있는 구조를 무시합니다. 산업 시대에도 그랬고, 인터넷 시대에도 그랬으며, 인공지능 시대에도 마찬가지입니다. 오늘날 많은 논의는 토큰 가격, 토큰 비용, 토큰 할당량 및 토큰 공급 능력에 초점을 맞추고 있으며, 마치 더 많은 토큰을 소유하는 것이 미래를 소유하는 것처럼 여겨집니다. 그러나 제 수십 년의 실천 경험에 비추어 볼 때, 자원 자체가 최종 가치였던 적은 결코 없었습니다. 가치를 진정으로 결정하는 것은 항상 사람들이 이러한 자원을 어떻게 조직하고 활용하는가입니다.

'토큰'을 이해하려면 새로운 용어에서 시작하는 것이 아니라 그 역사에서 시작해야 합니다. 1906 년 미국 철학자 찰스 샌더스 퍼스가 종이에 'the'라는 단어의 20 개 예시를 쓰면서 '유형'과 '실례'를 구분한 때부터, 2026 년 중국 국가데이터국장 류리훙이 보아오 포럼에서 Token 에 대한 공식 중국어 번역 '츠위안(词元)'을 제정하기까지, 이 개념은 철학에서 컴퓨터로, 블록체인에서 인공지능으로의 긴 여정을 걸어왔습니다. Token 이라는 단어는 서양 컴퓨터 세계에 오래전부터 존재해 왔으며, 프로그래밍 언어, 컴파일러 및 텍스트 처리 시스템에서 토큰은 컴퓨터가 인식하고 처리할 수 있는 기본 단위를 의미했습니다. 이후 네트워크 보안, 인증, API 호출 및 블록체인 맥락에서 '토큰'은 종종 접근 권한, 신원 증명 또는 디지털 자산을 나타내는 '패스' 또는 '자격 증명'으로 이해되었습니다. 그러나 대규모 언어 모델 시대에 이르러 Token 은 더 이상 주로 신원 증명이나 블록체인 자산이 아니라, AI 모델이 언어, 코드, 이미지 설명 및 멀티모달 정보를 처리할 때의 기본 계산 단위가 되었습니다.

중국은 이후 AI 분야의 Token 을 '츠위안(词元)'이라고 명명했습니다. 저는 이것이 상당히 정확한 번역이라고 생각합니다. '링파이(令牌, 패스)'를 사용하지 않은 이유는 대규모 모델의 Token 이 통행증이 아니기 때문입니다. 또한 단순히 '쯔(字, 글자)'나 '츠(词, 단어)'로 번역하지 않은 이유는 때로는 한 글자, 때로는 한 단어, 때로는 영어 단어의 절반, 때로는 구두점이나 기호일 수 있기 때문입니다. '츠(词)'는 언어 및 정보와의 관련성을 나타내고, '위안(元)'은 그것이 기본 단위임을 나타냅니다. 일반 독자를 위해 간단히 이해하면, 인간이 보는 것은 문장이지만, 인공지능이 보는 것은 토큰의 열입니다. 이 길을 따라 토큰은 언어학의 낯선 용어에서 AI 시대의 핵심 측정 단위로, 그리고 기술적 공급과 상업적 수요를 연결하는 결제 도구로 변모했습니다. 지능형 물류 시스템 구축, 자동화된 출판 프로세스 설계, 크로스 플랫폼 콘텐츠 배포와 같은 장기 프로젝트를 경험한 실천가로서 저는 이러한 진화에 대해 직접적인 이해를 가지고 있습니다. 토큰은 결코 목적이 아니라 수단이었습니다. 가치를 진정으로 결정하는 것은 결코 누가 더 많은 토큰을 가지고 있느냐가 아니라, 누가 토큰을 사용하여 지속적으로 작동하는 시스템과 복제 불가능한 구조를 구축할 수 있느냐입니다. 본 글에서 논의하는 시스템 구축 능력의 결과물은 <https://times.net.au> 에 공개적으로 제시되어 있습니다. 이는 독립 연구자가 구축하고 글로벌 도서관 시스템에 영구 보관된 학제간 학술 출판 시스템입니다.

자신의 여정을 되돌아보면, 저는 토큰이 등장한 이후에야 시스템에 대해 생각하기 시작한 것이 아니라, 장기적인 실천을 통해 점차적으로 시스템에 대한 이해를 발전시켜 왔습니다. 1989 년 호주 유학을 시작으로 여러 분야에서 학습, 작업, 탐색의 경험을 겪었습니다. 이후 점차 원격 근무, 물류 관리, 자동화 처리, 웹사이트 구축, 문화 전파, 다국어 출판 및 학제간 연구에 관여하게 되었습니다. 돌아보면, 이러한 경험들은 시대는 다르지만 공통점을 가지고 있습니다. 즉, 항상 어떻게 구조를 구축하고, 어떻게 반복 작업을 줄이며, 어떻게 전체 효율성을 높일 것인가를 중심으로 전개되었습니다. 당시에는 토큰이라는 개념도, 오늘날의 대규모 언어 모델도 존재하지 않았지만, 저는 이미 정보가 자동으로 흐르고, 자동으로 조직되고, 자동으로 처리되도록 하는 방법에 대해 끊임없이 생각하고 있었습니다.

따라서 오늘날 점점 더 많은 사람들이 토큰을 새로운 생산 수단으로 간주하는 경향이 있지만, 저는 그것을 새로운 형태의 기초 에너지로 보는 것이 더 적절하다고 생각합니다. 에너지는 분명 중요하지만, 에너지가 자동으로 가치를 창출하지는

않습니다. 전기 자체가 자동으로 공장이 되지는 않으며, 인터넷 자체가 자동으로 기업이 되지는 않으며, 토큰 자체가 자동으로 지식 체계가 되지는 않습니다. 미래에 모든 사람이 쉽게 대량의 토큰을 얻을 수 있게 된다고 해서 모든 사람이 동일한 결과를 얻을 수 있다는 의미는 아닙니다. 차이를 결정하는 것은 토큰의 양이 아니라 토큰이 들어가는 구조입니다.

토큰에 대한 심층 분석을 진행하기 전에, 저는 또 다른 중요한 관찰 지점을 지적할 필요가 있습니다. 토큰은 통화와 유사하며, 동서양의 기술, 시스템, 정책적 차이로 인해 보편적으로 통용되지 않습니다. 미국 달러와 중국 위안이 자유롭게 교환될 수 없는 것처럼, 중국어 토큰과 영어 토큰도 유사한 유통 장벽에 직면해 있습니다. 이는 단순한 언어 차이의 문제가 아닙니다. 더 깊은 이유는 동서양의 기술적 수준(모델 아키텍처, 토큰화 알고리즘, 훈련 데이터), 시스템적 수준(API 생태계, 결제 시스템, 데이터 표준), 정책적 수준(데이터 국경 이동, 수출 통제, 국가 안보 심사)에서의 세 가지 차이점에 있으며, 이로 인해 토큰이 이상적인 '글로벌 범용 AI 통화'처럼 자유롭게 유통되지 못하기 때문입니다. 이 관찰은 미래 토큰 시장의 판도를 이해하는데 중요한 의미를 갖습니다. 즉, 미래에는 '중국어 토큰 존', '영어 토큰 존'과 같은 언어 및 정책 기반 토큰 생태계 구역이 나타날 수 있으며, 구역 간 호출은 효율성 손실과 규정 준수 비용에 직면할 것입니다.

1. 토큰의 기원: 철학적 개념에서 경제적 단위로의 사중 비약

토큰 정체성의 진화는 명확히 네 단계로 나눌 수 있습니다.

철학 언어학적 기원. 1906 년, 퍼스는 기호 이론을 연구하면서 '유형'과 '실례'를 구분했습니다. 전자는 추상적인 규칙이나 형식을 나타내고, 후자는 그 규칙이 현실에서 구체적으로 나타난 것입니다. 이 구분은 이후 구조 언어학의 중요한 기초가 되었습니다.

컴퓨터 과학의 코딩 전환. 1960 년대, 프로그래머가 'int x = 5;'라는 코드 줄을 작성할 때, 컴퓨터는 이를 'int', 'x', '=', '5' 등과 같은 독립적인 단위로 분해해야 했으며, 이러한 각 단위가 Token 이었습니다. 따라서 토큰은 기계가 명령을 '읽는' 최소 정보 단위가 되었습니다.

블록체인의 가치 매체 실험. 2017 년, ICO 열풍이 고조되면서 Token 은 '유통 가능한 디지털 권리 증서'라는 새로운 정체성을 부여받았습니다. 그 열광은 점차 식었지만, '토큰은 가치의 상징이다'라는 인식은 그대로 자리 잡았습니다.

AI 시대의 경제 인프라. 2026 년 3 월은 획기적인 순간이었습니다. NVIDIA 의 CEO 젠슨 황은 데이터 센터를 '지능형 AI 토큰을 생산하는 공장'으로 재정의했고, 중국 국가데이터국장 류리훙은 '초위안(词元)'이라는 공식 중국어 번역을 제시하며 이를 '지능형 시대의 가치 닻'이라고 불렀습니다. 이 시점에서 토큰은 기술적 개념에서 경제적 개념으로의 마지막 비약을 완성했습니다.

2. 기술적 기반: BPE 알고리즘과 암묵적인 '언어세'

토큰이 AI 의 표준 측정 단위가 될 수 있었던 이유는 20 년 이상 잊혀졌던 기술인 바이트 쌍 부호화(Byte Pair Encoding, BPE)에 의존합니다.

1994 년, 미국 프로그래머 필립 게이지는 C 언어 기술 잡지에 데이터 압축 알고리즘에 관한 기사를 게재하여 BPE 의 기본 원리를 소개했습니다: 텍스트를 반복적으로 스캔하여 가장 자주 인접하는 두 문자를 새 기호로 '융접'하고, 반복적으로 압축해 나가는 방식입니다. 이 논문은 압축 효율이 두드러지지 않았기 때문에 당시 주목을 받지 못했습니다.

2016 년이 되어서야 에든버러 대학교 연구원 리코 센리히가 기계 번역의 단어 분할 문제를 연구하던 중 이 오래된 논문을 우연히 발견하고, BPE 가 분할 문제에 대한 완벽한 해결책임을 예리하게 깨달았습니다. 사전을 미리 정의할 필요 없이 데이터 스스로 '말하게' 하여 고빈도 조합이 자연스럽게 토큰으로 응집되도록 하는 것입니다. 2019 년, OpenAI 는 GPT-2 를 출시하면서 이 접근 방식을 공식 채택했으며, 토큰화의 시작점을 직접 '바이트' 수준으로 설정하여 모델이 이론적으로 모든 문자 체계를 처리할 수 있도록 했습니다.

그러나 BPE 알고리즘의 '빈도 우선' 논리는 의도치 않게 암묵적인 '언어세' 체계를 형성했습니다. 영어는 인터넷 말뭉치의 절대적 주류이기 때문에 영어 토큰의 분할 효율성은 다른 언어보다 훨씬 높습니다. 같은 의미를 표현하는 데 영어는 가장 적은 토큰을 사용하고, 중국어는 일반적으로 1.5 배에서 2 배, 자원이 더 적은 언어는

영어의 5 배에서 10 배에 달하는 비용이 들 수 있습니다. 즉, 토큰 단위 과금 체계 하에서는 AI 와 대화하는 실제 비용이 사용하는 언어에 따라 크게 달라집니다.

이러한 '출발선'에서의 불공평함은 모델의 초기 어휘에 한번 기록되면 수정이 매우 어렵습니다. 토큰화 규칙은 AI 가 세계를 인식하는 '기반'이며, 건물이 높아질수록 기반을 교체하기는 더욱 어렵기 때문입니다. 다행히도 이 상황은 개선되고 있습니다. GPT 시리즈를 예로 들면, 같은 중국어 문장이 GPT-3 에서는 38 개의 토큰이 필요했지만, GPT-4 에서는 26 개로, GPT-5 에서는 15 개로 감소하여 처리 효율이 60% 이상 향상되었습니다. 또한 DeepSeek, 통이첸원(Tongyi Qianwen)과 같은 중국 국내 대형 모델은 설계 단계부터 중국어 고빈도 어구, 관용구 등을 네이티브 토큰으로 어휘표에 포함시켜 중국어에 대해 더 '모국어' 수준의 처리 효율을 달성했습니다. 이는 깊은 법칙을 드러냅니다. 즉, '의미 분할 권한'을 장악하는 자가 그 언어의 디지털 세계에서 표현 효율성과 비용 우위를 크게 장악하며, 이는 디지털 시대의 '기본 주조권'인 것입니다.

3. 용도와 산출: 대화에서 에이전트로의 기하급수적 비약

토큰 소비 시나리오는 '대화'에서 '추론'을 거쳐 '에이전트'로의 3 단계 비약을 겪고 있습니다.

대화형 AI 단계에서는 단일 프롬프트와 응답이 약 10,000 개의 토큰을 소비합니다. 추론형 AI 단계에서는 여러 차례의 상호 작용과 복잡한 추론이 수반되어 각 작업이 약 100,000 개의 토큰을 소비합니다. 에이전트형 AI 단계에 들어서면 에이전트가 다단계 작업을 자율적으로 수행하고 여러 도구를 호출할 수 있으며, 각 작업의 토큰 소비량은 약 100 만 개에 달합니다. 불과 두 세대 만에 약 100 배로 증가한 것입니다. 퀄컴 CEO 아몬은 2030 년까지 10 초당 전 세계 토큰 수요가 2026 년 317 억 개에서 1.27 조 개로 폭증할 것이라고 예측합니다. 이는 과장이 아닙니다. OpenClaw 와 같은 오픈 소스 에이전트 도구의 폭발적인 인기가 이를 증명합니다. 단일 복잡 작업이 수천만 개의 토큰을 소비하는 것은 흔한 일입니다.

산출 측면에서 보면 중국은 세계 최대의 토큰 공급국이 되었습니다. 공개 데이터에 따르면 2026 년 3 월 기준 중국의 일일 토큰 호출량은 140 조 개를 돌파했으며, 2 년 만에 천 배 이상 증가하여 전 세계 AI 토큰 사용량의 61%를 차지합니다. 이 지위의

확립은 세 가지 이점의 중첩에 의존합니다. 첫째, 중국의 산업용 전기 가격은 유럽과 미국보다 훨씬 낮으며, 토큰 생산 비용에서 전기가 차지하는 비중은 60%를 초과합니다. 둘째, 중국은 세계 최대 규모의 지능형 컴퓨팅 클러스터를 구축하고 '동수서산(東數西算, 동부 데이터 서부 계산)' 배치를 통해 효율적인 생산을 실현했습니다. 셋째, 국산 대형 모델이 완전히 오픈소스화되었고, API 상용화 서비스 체계가 성숙하여 '모델 훈련-컴퓨팅 지원-추론 출력-글로벌 호출'의 완전한 산업 체인 클로즈 루프를 형성했습니다.

4. 기관과 시장: 삼두 체제의 산업 지형

토큰을 중심으로 명확히 계층화된 시장 구조가 형성되었습니다.

이러한 산업 구조를 이해하기 위해서는 토큰 뒤에 있는 하부 지원 체계를 함께 살펴보아야 합니다. 많은 사람들이 토큰 가격에 주목하지만 토큰 자체가 독립적으로 존재하는 자원이 아님을 간과합니다. 토큰 뒤에는 모델이 있고, 모델 뒤에는 컴퓨팅 파워가 있습니다. 어떤 의미에서 컴퓨팅 파워는 모델 능력을 결정하고, 모델은 토큰의 산출을 결정하며, 토큰은 모델 능력이 시장을 향해 출력될 때 형성되는 측정 단위입니다. 따라서 토큰 시장은 표면적으로는 가격 경쟁이지만, 그 바닥에는 여전히 컴퓨팅 파워 경쟁과 효율성 경쟁이 있습니다.

현재 글로벌 토큰 산업 체인은 대략 세 가지 수준으로 나눌 수 있습니다. 최상류는 컴퓨팅 파워 인프라로, 칩 제조사, 서버 제조사, 클라우드 컴퓨팅 플랫폼, 데이터 센터 및 에너지 시스템을 포함합니다. 중류는 기반 모델 기관 및 추론 플랫폼으로, 컴퓨팅 파워를 호출 가능한 토큰 서비스로 변환하는 역할을 합니다. 하류는 토큰 서비스 제공업체, 에이전트 플랫폼 및 다양한 산업 응용 기관으로, 모델 능력을 기업 서비스 및 최종 제품으로 더욱 변환합니다. 사용자가 최종적으로 보는 것은 토큰 가격이지만, 토큰 비용과 속도에 진정으로 영향을 미치는 것은 종종 상류의 컴퓨팅 파워 체계와 모델 효율성의 변화입니다.

지난 몇 년간 칩 성능의 지속적인 향상, 모델 아키텍처의 최적화, 추론 효율성의 지속적인 개선에 따라 토큰 단가는 빠른 하락 추세를 보였습니다. 많은 사람들은 토큰이 점점 저렴해지고 있다고 생각하지만, 이 변화의 진정한 동인은 컴퓨팅 파워 체계의 지속적인 발전입니다. 따라서 미래 토큰 시장의 경쟁은 플랫폼 간 경쟁과 가격

경쟁뿐만 아니라 컴퓨팅 자원, 모델 효율성 및 자원 통합 능력 간의 경쟁이 될 것입니다.

상류 '발전소' – 컴퓨팅 파워 생산자. NVIDIA 가 대표하는 칩 회사들은 하드웨어를 지속적으로 개선하여 컴퓨팅 파워 단위당 토큰 생산 효율성을 높이고 있습니다. NVIDIA 는 2027 년까지 매출이 최소 1 조 달러에 달할 것으로 예측하는데, 그 자신감의 원천은 전 세계의 AI 컴퓨팅 파워에 대한 갈망입니다.

중류 '전력망 제어' – 모델 플랫폼 및 집계자. OpenAI, Google, DeepSeek, 알리바바와 같은 회사들은 기반 컴퓨팅 파워를 표준화되고 호출 가능한 토큰 서비스로 변환합니다. 이 부분을 중심으로 세 가지 성숙한 비즈니스 모델이 형성되었습니다: 토큰 집계 플랫폼(OpenRouter 가 대표, 가격 차이와 수수료로 운영, 2026 년 5 월 기준 주간 호출량 26.9 조 토큰), 클라우드 공급업체의 MaaS 서비스, 그리고 2 차 시장에서 활발히 활동하는 AI 중계소입니다.

하류 '전력 입력 포인트' – 응용 및 배포 플랫폼. 텐센트가 WeChat 인터페이스를 개방하여 새로운 모델 능력을 연결하는 것은 본질적으로 WeChat 을 '토큰 배포 플랫폼'으로 만드는 것입니다. 중국전신은 '토큰 경영을 통한 사업 재구성'을 제안하며 전통적인 트래픽 로직에서 벗어나려 하고 있습니다. 이러한 입력 포인트 기업들의 공통점은 사용자가 더 이상 직접 모델에 대면하지 않고 플랫폼을 통해 간접적으로 토큰을 호출한다는 점이며, 모델 자체가 '보이지 않는' 상태가 되어가고 있습니다.

주목할 점은 토큰을 중심으로 독특한 글로벌 무역 패러다임이 형성되고 있다는 것입니다. 즉, 무형적 수출, 면세적 운영, 글로벌 적용 범위입니다. 이는 중국을 전통적인 '세계 공장'에서 '글로벌 AI 컴퓨팅 파워 인프라 공급자'로 변모시키고 있습니다.

토큰 경제를 논의할 때 컴퓨팅 파워라는 피할 수 없는 기본 요소가 있습니다. 많은 사람들이 토큰 가격에 주목하지만 토큰 뒤의 생산 과정을 간과합니다. 사실 토큰은 허공에서 생성되는 것이 아닙니다. 토큰이 생성될 때마다 모델이 추론을 수행해야 하며, 모델의 추론은 막대한 컴퓨팅 파워 위에 구축됩니다. 어떤 의미에서 컴퓨팅 파워는 토큰 경제의 원천이고, 모델은 토큰의 생산 도구이며, 토큰은 모델 능력이 외부로 출력될 때 형성되는 측정 단위입니다.

따라서 컴퓨팅 파워 수준은 토큰 비용과 토큰 생성 속도에 직접적인 영향을 미칩니다. 컴퓨팅 파워가 강력할수록 모델의 정보 처리 능력은 높아지고 단위 시간당 생성할 수 있는 토큰 수는 많아집니다. 컴퓨팅 파워 비용이 낮을수록 토큰 가격도 일반적으로 낮아집니다. 지난 몇 년간 대규모 모델 업계에서 토큰 가격이 지속적으로 하락한 현상은 모델 알고리즘 최적화뿐만 아니라 칩 성능 향상, 데이터 센터 규모 확대, 추론 효율성 증가 등의 요인도 그 배경에 있습니다. 즉, 많은 사람들은 토큰이 저렴해지고 있다고 보지만, 이 변화를 진정으로 추진하는 것은 컴퓨팅 파워 체계의 지속적인 발전인 것입니다.

산업 체인 관점에서 보면 컴퓨팅 파워, 모델, 토큰은 실제로 완전한 체인을 구성합니다. 상류에는 칩, 서버, 데이터 센터, 에너지 시스템이 있습니다. 중류에는 기반 모델과 추론 플랫폼이 있습니다. 하류에는 토큰 서비스, 에이전트 애플리케이션 및 다양한 산업 솔루션이 있습니다. 컴퓨팅 파워는 모델 능력의 상한을 결정하고, 모델은 토큰의 품질과 효율성을 결정하며, 토큰은 궁극적으로 사용자가 얼마나 많은 인공지능 서비스를 얻을 수 있는지를 결정합니다.

그러나 컴퓨팅 파워의 중요성은 의심할 여지가 없지만, 컴퓨팅 파워 자체가 자동으로 가치를 창출하지는 않는다고 생각합니다. 역사적으로 많은 산업이 유사한 과정을 경험했습니다. 더 큰 발전소를 가진다고 해서 더 나은 기업을 가진다는 의미는 아니며, 더 많은 서버를 가진다고 해서 더 나은 제품을 가진다는 의미는 아닙니다. 마찬가지로, 더 많은 컴퓨팅 파워를 가진다고 해서 더 높은 가치의 시스템을 가진다는 의미는 아닙니다. 컴퓨팅 파워는 생산 능력의 문제를 해결하고, 시스템은 가치 전환의 문제를 해결합니다. 미래에 컴퓨팅 파워가 계속 증가하고, 모델 능력이 계속 강화되며, 토큰 자원이 계속 풍부해짐에 따라, 진정으로 희소해질 것은 여전히 컴퓨팅 파워 자체가 아니라 이러한 자원을 현실 세계에서 장기적으로 작동하는 시스템으로 조직화하는 방법일 것입니다.

따라서 미래 경쟁의 관점에서 볼 때, 컴퓨팅 파워 경쟁, 모델 경쟁, 토큰 경쟁은 결국 점차 시스템 경쟁으로 이행될 것이라고 생각합니다. 컴퓨팅 파워가 점점 보편화되고, 모델이 점점 유사해지며, 토큰이 점점 풍부해졌을 때, 차이를 결정하는 것은 더 이상 보유 자원량이 아니라 누가 이러한 자원을 이용하여 장기적으로 유효한 구조를 구축할 수 있는지가 될 것이기 때문입니다.

5. 경쟁과 진입 장벽: 가격전 인플레이션과 가치 계층화

토큰 시장의 경쟁은 매우 인플레이션적인 특징을 보이고 있습니다. 2022년부터 2026년까지 단일 토큰 추론 비용은 99.9% 감소했습니다. 국내 모델의 가격 우위는 더욱 두드러집니다. DeepSeek V4-Pro의 입력 가격은 백만 토큰당 0.025 위안(약 0.5 원)으로 낮고, 통이첸원 Qwen3.5-Plus의 출력 가격은 백만 토큰당 1.6 위안(약 32 원)인 반면, OpenAI GPT-4o의 출력 가격은 약 72 위안(약 1440 원)입니다.

이러한 가격전은 '중간상'의 이익 공간을 급격히 압박했습니다. 업계 통계에 따르면 다양한 API 집계 및 중계 플랫폼의 수는 2,000 개를 돌파했으며, 시장은 세 가지 큰 계층을 형성했습니다. 최상위층은 기술력이 강하고 모델이 풍부한 대형 플랫폼, 중견층은 공식 할인과 고객 자원을 보유한 전문 대리점, 그리고 다수의 소규모 대리점은 심각한 생존 압력에 직면하고 있습니다. 단순한 토큰 가격 차이로 이익을 보는 비즈니스 모델은 더 이상 지속 가능하지 않습니다.

진정한 진입 장벽은 두 가지 방향으로 이동하고 있습니다. 하나는 기술적 프리미엄 방향입니다. 자체 개발한 추론 가속 엔진을 통해 공식보다 저렴하게 판매하면서도 이익을 낼 수 있습니다. 두 번째는 부가가치 서비스 방향입니다. 프롬프트 엔지니어링, 모델 미세 조정, 프라이빗 배포 등의 맞춤형 서비스를 제공하는 것입니다. 더욱 주목할 점은 토큰의 가치가 시나리오에 따라 엄청나게 분화되고 있다는 것입니다. 잡담 관련 토큰의 평균 가격은 백만 개당 0.01 달러(약 1.5 원)에 불과한 반면, 신약 개발 관련 토큰의 평균 가격은 백만 개당 1000 달러(약 150 만원)에 달하여 시나리오 간 가치 차이가 10 만 배에 이릅니다. 이는 본질적인 법칙을 드러냅니다. 토큰의 가치는 생산 비용에 의해 결정되는 것이 아니라 그 용도에 의해 결정되며, 같은 토큰이라도 다른 시나리오에서 사용되고 다른 시스템에 내장되면 그 가치는 천지 차이입니다.

그러나 여기서 더 중요한 판단을 해야 합니다. 현재 그리고 미래에 토큰을 판매하는 진입 장벽은 높지 않습니다. 누구나 API에 등록하고, 집계 플랫폼을 구축하고, 토큰 재판매를 시작할 수 있습니다. 진정으로 희소한 것은 토큰 판매 그 외의 것들, 즉 실제 역사 데이터, 시스템과 체계를 구축하는 능력, 미래를 예측하는 비전입니다. 이것들이 없다면 어떤 대규모 모델과 토큰도 활용도가 없습니다.

토큰 수익성에 대한 추가 판단: 현재 토큰으로 돈을 버는 것은 이미 쉽지 않으며 점점 더 어려워질 것입니다. 가격전은 지속적으로 이윤 폭을 압박하고 있으며, 단순한 토큰 재판매 비즈니스 모델은 빠르게 저이윤 또는 무이익으로 향하고 있습니다. 토큰이 프론트엔드 시스템, 즉 서비스를 제공하는 시스템에 부착될 때만 토큰과 컴퓨팅 파워가 유효한 부가 자산으로 기능하여 이익을 창출할 수 있습니다. 다시 말해, 토큰 자체는 이익의 원천이 아니라 시스템이 이익의 원천입니다. 토큰은 시스템의 연료일 뿐 엔진이 아닙니다.

토큰이 인공지능 시대의 중요한 측정 단위가 됨에 따라 토큰을 중심으로 한 비즈니스 모델이 시장에 점점 더 많이 나타나고 있습니다. 기반 모델 벤더 외에도 토큰 대리점, 모델 집계 플랫폼, 에이전트 플랫폼, 기업용 API 배포 서비스 및 다양한 미들웨어 서비스 제공자가 등장하고 있습니다. 많은 기업가들에게 토큰 대리점 업은 인공지능 산업으로의 지름길처럼 보입니다. 모델을 훈련하거나 대규모 컴퓨팅 센터를 구축할 필요 없이 기존 자원을 통합하는 것만으로 사업을 시작할 수 있기 때문입니다. 그러나 이 모델에는 장점과 동시에 명확한 한계도 존재합니다. 장점은 상대적으로 낮은 투자로 시작할 수 있다는 점, 진입 장벽이 상대적으로 작다는 점, 시장 확장 속도가 빠르다는 점 등입니다. 한계는 상류 모델 공급자에 대한 의존도가 높고, 가격 경쟁에 빠지기 쉬우며, 장기적인 기술 장벽이 부족하다는 점입니다. 토큰 가격이 지속적으로 하락할 경우 단순한 토큰 재판매로 인한 이익 공간은 지속적으로 압박받을 수 있습니다. 따라서 장기적으로 볼 때, 토큰 대리점 업은 오히려 유통 능력에 가까우며, 최종 핵심 경쟁력은 아닙니다. 기업의 장기적 가치를 진정으로 결정하는 것은 여전히 토큰을 중심으로 구축된 산업 시스템, 자동화 프로세스, 지식 체계 및 현실 시나리오에서의 솔루션입니다.

6. 현재의 곤경과 미래의 출구

현재 토큰 시장은 몇 가지 심층적인 문제에 직면해 있습니다.

동질화 경쟁의 함정. 2,000 개 이상의 플랫폼이 동일한 고객층을 두고 다투고 있으며, 가격전은 업계의 이익을 계속해서 끌어내리고 있습니다. 변하지 않으면 클라우드 컴퓨팅 초기의 '손해 보면서 고객 잡기' 상황을 재현할 가능성이 있습니다.

컴플라이언스 리스크의 부각. 현재 전 세계적으로 특별 법규가 없다는 '규제 보너스'는 일시적인 상태일 뿐 장기적인 상수가 아닙니다. 토큰 시장 규모가 확대됨에 따라 데이터 국경 이동, 국가 안보, 세금 징수 등의 문제는 각국이 입법을 가속화하도록 만들 것입니다. 특히 주의해야 할 점은 일부 시장 참여자들이 토큰을 토큰화하고 금융화하려는 시도는 글로벌 금융 규제를 촉발하기 쉬우며, 관련 사업이 단속되는 데 그치지 않고 중국 전체 토큰 수출 산업의 평판을 손상시킬 수 있다는 것입니다.

'토큰 최대화(Token-maxxing)'의 미신. 실리콘 밸리에서는 새로운 과시 방식이 유행하고 있습니다. 즉, 매일 얼마나 많은 토큰을 소비하는지 겨루는 것입니다. 메타(Meta) 내부에는 'Claudeconomics'라는 랭킹이 있으며, 85,000 명 이상의 직원들의 AI 사용 데이터를 집계하여, 1 위 직원이 소비하는 토큰의 가치는 수백만 달러에 달합니다. 반대파는 신랄하게 지적합니다. '토큰 최대화보다 결과 최대화(Outcome maxxng)가 더 낫다' - 미친 듯이 토큰을 소비하기보다 어떤 결과를 생산했는지 보는 것이 낫다는 것입니다. 이것은 제가 오랜 실천을 통해 깊이 절감해온 핵심 문제입니다.

중국어 토큰과 영어 토큰: 통일된 개념 아래의 세 가지 장벽

중국 사용자에게 명확히 인식해야 할 사실이 있습니다. 중국어 토큰(츠위안)과 영어 토큰은 둘 다 인공지능 시대의 기초적 측정 단위에 속하지만, 진정으로 통일되고 자유롭게 상호 운용 가능한 글로벌 시장을 형성하지는 못했습니다. 기술적 개념上看, 동서양 모두 모델이 언어와 데이터를 처리하는 기본 단위로 토큰을 채택하고 있습니다. 그러나 실제 응용 관점에서는 기술적 장벽, 시스템적 장벽, 정책적 장벽이 공동으로 존재하여 중국어 토큰과 영어 토큰 사이에 사실상의 분리 상태가 형성되었습니다.

첫 번째 장벽은 기술적 장벽, 즉 토큰화 논리와 모델 구조의 차이입니다.

GPT, Claude, Gemini 등의 서양 주류 모델은 대부분 영어를 주체로 하는 대규모 말뭉치에 기반하여 구축되었습니다. 이러한 모델들은 이미 뛰어난 중국어 처리 능력을 갖추고 있지만, 많은 시나리오에서 같은 정보량을 표현하는 데 중국어는 여전히 영어보다 더 많은 토큰을 필요로 하는 경향이 있습니다. 더 중요한 것은, 다른 모델은 다른 어휘표 구조와 토큰화 전략을 가진다는 점입니다. 같은 중국어

텍스트라도 다른 모델에서 완전히 다른 수의 토큰으로 분할될 수 있으며, 내부 코딩 방식과 의미 매핑 방식도 다를 수 있습니다.

동시에 DeepSeek, 통이첸원, 더우바오, 원신이엔과 같은 중국 국내 모델은 훈련 단계부터 중국어 말뭉치를 대량으로 도입하여 중국어 구문, 관용구, 고유 명사, 지역화 표현에 대해 더 높은 네이티브 적응력을 가지고 있습니다. 따라서 같은 텍스트라도 모델에 따라 토큰 수, 처리 효율성, 이해 경로에 분명한 차이가 발생할 수 있습니다. 이러한 의미에서 토큰은 통일된 개념을 가지고 있지만, 모델 간을 직접 유통하고 재사용할 수 있는 '표준 토큰'은 존재하지 않는 것입니다.

두 번째 장벽은 시스템적 장벽, 즉 생태계와 호출 방식의 차이입니다.

현재 세계 토큰 시장은 실제로 여러 비교적 독립적인 생태계를 형성하고 있습니다. OpenAI, Anthropic, Google 등의 국제 플랫폼은 각자의 API 표준, 과금 모델, 접근 규칙, 서비스 체계를 가지고 있습니다. 중국 시장은 DeepSeek, 통이첸원, 더우바오, Kimi 등의 플랫폼을 중심으로 한 국내 생태계를 형성하고 있습니다.

일반 사용자에게 있어 어떤 플랫폼에서 토큰을 구매했다고 해서 그것을 다른 플랫폼에서 직접 사용할 수 있음을 의미하지는 않습니다. 플랫폼마다 호출 인터페이스, 인증 메커니즘, 속도 제한, 서비스 계약, 비즈니스 규칙이 다릅니다. 제 3 자 게이트웨이나 집계 플랫폼을 통해 변환하더라도 본질적으로는 다른 시스템을 다시 호출하는 것일 뿐, 진정한 의미의 토큰 상호 교환은 아닙니다. 따라서 토큰은 개념적으로 통일되어 있지만, 상업적 자원 및 서비스 능력으로서의 토큰은 여전히 각자의 플랫폼 생태계 내에 갇혀 있습니다.

세 번째 장벽은 정책적 장벽, 즉 데이터 흐름과 규정 준수 요구 사항의 차이입니다.

인공지능이 점차 엔터프라이즈 응용 단계로 들어감에 따라 데이터 보안 문제의 중요성은 점점 더 높아지고 있습니다. 중국의 데이터 보안 체계, EU 의 GDPR 체계, 미국의 민감 데이터 국경 이동에 대한 규제 조치는 모두 데이터 전송과 처리에 대해 다른 요구 사항을 부과하고 있습니다. 고객 정보, 재무 데이터, 의료 기록, 영업 비밀, 개인 정보와 관련된 많은 데이터는 종종 자유롭게 국경을 넘어 유통될 수 없습니다.

이는 실제 비즈니스 환경에서 많은 기업들이 단순히 토큰 가격에 기반하여 모델을 선택하는 것이 아니라, 데이터 규정 준수, 데이터 저장 위치, 법적 책임을 우선적으로 고려해야 함을 의미합니다. 많은 실제 업무 데이터의 경우 규정 준수 비용이 토큰 자체의 비용보다 훨씬 더 높은 경우가 많습니다. 따라서 현실 세계에서는 국내 업무는 국내 규정 준수 모델을 사용하고, 해외 업무는 해외 모델을 사용하는 경향이 있어 결과적으로 사실상의 토큰 시장 분할이 발생합니다.

기술적 장벽은 토큰 분할 방식의 차이를 결정하고, 시스템적 장벽은 호출 체계와 비즈니스 생태계의 차이를 결정하며, 정책적 장벽은 데이터 흐름과 규정 준수 경계의 차이를 결정합니다. 세 가지 장벽이 공동으로 작용함으로써 중국어 토큰과 영어 토큰은 같은 기술 개념에 속하면서도 이상적인 '글로벌 통일 AI 통화'처럼 자유롭게 교환, 유통, 연결되기 어렵습니다.

미래에는 국제 무역과 유사한 상황이 나타날 가능성이 높습니다. 즉, 중국어 토큰 생태계와 영어 토큰 생태계가 장기간 공존하며 각자 독자적인 발전 경로를 형성할 것입니다. 구역 간 호출이 불가능한 것은 아니지만, 종종 추가적인 기술적 적응, 시스템 변환, 법적 심사 비용이 필요합니다. 대다수 일반 사용자에게 이러한 구역 간 능력은 일상적인 필요 사항이 아니므로, 두 대규모 체계는 장기간 비교적 독립적인 발전을 유지할 것입니다.

중국에서 생산된 토큰이 반드시 중국 시스템 내에서만 사용되어야 하는 것은 아니지만, 데이터 주권, 규정 준수 책임, 산업 안보 보장이 관련된 주요 시나리오에서는 미래의 많은 토큰 수요가 우선적으로 국내 모델, 국내 플랫폼, 국내 시스템 내부에서 소비될 가능성이 높습니다.

장기적 추세로 보면 컴퓨팅 파워 생산지, 모델 제공자, 시스템 운영자, 데이터 소재지 간의 결합도가 강화되고 있습니다. 금융, 행정, 의료, 공급망, 대기업 시스템의 경우 데이터는 종종 자유롭게 국경을 넘어 유통될 수 없습니다. 따라서 중국이 대량의 토큰 서비스를 세계로 수출할 수 있다고 하더라도, 많은 고가치 토큰 수요는 여전히 우선적으로 중국 모델과 중국 시스템 내에서 완결될 것입니다. 이는 중국 토큰이 수출될 수 없다는 것을 의미하는 것이 아니라, 미래 토큰 시장이 '글로벌 유통 부분'과 '국내 클로즈드 루프 부분'이라는 두 가지 구조를 동시에 가질 가능성이 있음을 의미합니다.

주목할 점은 미래에 진정한 통일을 실현하는 것은 토큰 자체가 아니라 토큰을 호출하는 방식일 가능성이 높다는 것입니다. 에이전트 플랫폼, 통일 인터페이스 프로토콜, 모델 집계 플랫폼, 자동화 시스템 등이 계속 발전함에 따라 사용자가 최종적으로 주목하는 것은 어떤 토큰이 사용되었는지가 아니라, 작업이 완료되었는지, 비용이 합리적인지, 효율성이 충분한지가 될 것입니다. 그때가 되면 토큰은 오늘날의 전기, 인터넷 트래픽, 클라우드 컴퓨팅 자원처럼 점차 시스템 뒤에 숨겨져 디지털 문명의 기초적 운영 인프라가 될 것입니다.

따라서 장기적 추세로 볼 때 동서양의 토큰 시장은 당분간 직접 상호 운용되기는 어렵지만, 그 발전 방향은 매우 일치합니다. 토큰의 과거는 컴퓨터에 의한 언어 처리에 속하고, 토큰의 현재는 인공지능 산업에 속하며, 토큰의 미래는 디지털 문명 전체의 기초적 운영 체계에 속할 가능성이 높습니다.

7. 왜 시스템 구축 능력은 대량의 토큰으로 직접 획득할 수 없는가

'시스템 구축 능력은 희소 자산이다'라고 주장할 때, 저는 제기될 수 있는 질문에 정면으로 맞서야 합니다. AI 가 그렇게 강력하다면, 왜 대량의 토큰을 소비함으로써 시스템 구축 능력을 '생성'하거나 '획득'할 수 없을까요? 이 질문에 대한 답은 토큰 능력의 근본적인 한계를 드러냅니다.

첫째, 시스템 구축에는 의사 결정이 필요하며, 의사 결정은 생성이 아닙니다.

AI 는 기사, 코드, 웹사이트 템플릿을 생성할 수 있지만, 시스템 설계에서 인간이 내려야 하는 핵심 결정 – 어떤 모듈을 선택하고, 어떤 기능을 포기할지, 현재의 필요와 미래 확장의 균형을 어떻게 잡을지, 언제 리팩토링하고 언제 현 상태를 유지할지 – 을 AI 가 대신할 수는 없습니다. 이러한 결정은 업무 본질의 이해, 자원 제약의 판단, 장기적 결과에 대한 예측에 의존합니다. 토큰은 무수한 '가능한 해결책'을 생성할 수 있지만, '어떤 해결책을 구현할 가치가 있는지' 그리고 '왜'를 선택하는 것은 토큰이 대체할 수 없는 일입니다. 제 지능형 물류 시스템은 20 년 동안 수많은 조정을 겪었으며, 매번의 조정은 AI 가 '이렇게 하는 것이 더 좋다'고 알려줬기 때문이 아니라,

실제 업무에서 문제가 발생하고, 한계가 드러나고, 새로운 요구가 생겼기 때문입니다. 이러한 결정 뒤에는 판단력이 있으며, 판단력은 토큰이 아닌 실천에서 비롯됩니다.

둘째, 시스템 구축에는 예외 처리가 필요하며, 예외는 총망라하여 열거할 수 없습니다.

시스템이 장기간 작동할 수 있는 이유는 정상 흐름이 완벽하게 설계되었기 때문이 아니라 예외를 처리할 수 있기 때문입니다. 물류 시스템에서 예외는 다음과 같습니다: 고객 주소 오류, 화물 손상, 통관 지연, 결제 실패, 시스템 다운, 데이터 손실. 출판 시스템에서 예외는 다음과 같습니다: 번역 품질 변동, 메타데이터 형식 비호환, 아카이브 인터페이스 변경, DOI 등록 실패. 이러한 예외는 현실 세계의 예외 다양성이 무한하기 때문에 총망라하여 열거할 수 없습니다. AI 는 대량의 토큰을 통해 '일반적인 예외 패턴'을 학습할 수 있지만, '다음에 나타날 전례 없는 예외'를 예측할 수는 없습니다. 장기간 작동하는 시스템만이 실제 운영에서 지속적으로 예외를 마주하고, 기록하고, 해결함으로써 진정한 예외 처리 능력을 형성할 수 있습니다. 이러한 능력은 '더 많은 토큰을 먹이는 것'으로 얻을 수 있는 것이 아닙니다. 예외는 지식이 아니라 경험이기 때문입니다.

셋째, 시스템 구축에는 균형(trade-off)이 필요하며, 균형에는 표준적인 답이 없습니다.

모든 시스템의 구축은 다중 제약 조건 하에서의 균형입니다: 비용 대 성능, 속도 대 안정성, 범용성 대 맞춤화, 현재의 필요 대 미래 확장. 이러한 균형에는 '정답'은 없고, '특정 상황에서의 합리적 선택'만 있을 뿐입니다. AI 는 '최적해'를 제시할 수 있지만, 이 '최적'은 역사적 데이터에 기반한 통계적 최적일 뿐, '당신의 구체적 상황'에 기반한 판단이 아닙니다. 제 세대를 초월한 20 년간의 물류·재무 통합 시스템이 오늘날까지 작동하는 이유는 바로 제가 다른 시대에 다른 균형을 했기 때문입니다 – 2005 년에 어떤 기술 스택을 선택하고, 2015 년에 다른 통합 방식을 선택하고, 2025 년에 업그레이드 경로를 선택했습니다. 이러한 균형은 '옳다' 또는 '그르다'가 아니라 '당시의 상황에 적합했다'는 것입니다. 토큰은 이러한 균형을 대신 해줄 수 없습니다. 균형의 본질은 계산이 아니라 가치 판단이기 때문입니다.

넷째, 시스템 구축에는 피드백 루프가 필요하며, 피드백에는 현실 세계가 필요합니다.

시스템 구축 능력의 향상은 설계 → 구현 → 운영 → 문제 발견 → 수정 → 재설계라는 클로즈드 루프에 의존합니다. 이 루프에서 '문제 발견' 단계는 현실 세계로부터의 피드백에서 비롯되어야 합니다. AI 는 운영을 시뮬레이션할 수 있지만, 현실 세계의 모든 예측 불가능한 요소를 시뮬레이션할 수는 없습니다. 제 물류 시스템이 지속적으로 개선된 것은 실제 고객이 실제 피드백을 제공했기 때문이며, 제 출판 시스템이 지속적으로 개선된 것은 실제 아카이브 인터페이스가 실제 변경을 경험했기 때문입니다. 이러한 피드백은 폐쇄된 환경에서 토큰을 소비함으로써 얻을 수 있는 것이 아닙니다. 토큰은 '고객이 이렇게 피드백할 것이다'라는 가정을 생성할 수는 있지만, '고객이 실제로 이렇게 피드백했다'는 사실을 대체할 수는 없습니다. 현실 세계로부터의 피드백은 시스템 구축 능력이 성장하는 유일한 토양입니다.

다섯째, 시스템 구축에는 신뢰가 필요하며, 신뢰는 시간으로부터만 올 수 있습니다.

시스템이 신뢰할 만한지 여부는 홍보에 의해서가 아니라 시간에 의해 증명됩니다. 제 웹사이트는 CERN Zenodo, WorldCat, 호주 국립 도서관에 의해 영구 보관되는 이유는 그들이 '저를 좋아하기' 때문이 아니라, 이 기관들이 장기적 관찰을 통해 제 시스템이 안정적이고 신뢰할 수 있으며 지속 가능성을 확인했기 때문입니다. 이러한 신뢰는 대량의 토큰으로 구매할 수 있는 것이 아닙니다. AI 는 '매우 전문적으로 보이는' 시스템 설명서를 생성할 수는 있지만, '이 시스템이 20 년 동안 신뢰받아 왔다'는 사실을 생성할 수는 없습니다. 신뢰는 시간의 산물이지, 토큰의 산물이 아닙니다.

요약하자면, 시스템 구축 능력이 대량의 토큰을 통해 직접 획득될 수 없는 이유는 그것이 본질적으로 생성 능력이 아니라 판단 능력, 예외 처리 능력, 균형 능력, 피드백 흡수 능력, 신뢰 축적 능력이기 때문입니다. 이러한 능력들의 공통점은 현실 세계의 시간, 실전, 오류, 수정, 검증을 필요로 한다는 것입니다. 토큰은 정보 처리를 가속화할 수는 있지만, 현실 세계의 경험 축적을 가속화할 수는 없습니다. 이것이 미래에 토큰이 극도로 저렴해져도 시스템 구축 능력이 여전히 희소한 이유입니다.

8. 나의 자산: 실제 역사 데이터, 학제간 시스템, 구축 능력, 예측 비전, 자기 증명 능력

'진정으로 희소한 것은 무엇인가'를 논하기 전에 먼저 내 자신의 위치를 설명할 필요가 있습니다. 내 판단은 이론적 연역이 아니라 장기적 실천에서 비롯되기 때문입니다. 이는 내가 오랜 기간에 걸쳐 구축해 온 다양한 시스템에서 매우 분명하게 드러납니다. 지능형 물류 시스템, 지능형 송장 시스템, 자동화 재무 처리 체계, 이후의 10 개국어 출판 시스템, 메타데이터 시스템, 웹사이트 탐색 시스템, 자동 상호 참조 시스템 중 어느 것이든 진정으로 어려운 부분은 결코 정보를 생성하는 것이 아니라 정보를 조직하는 것이었습니다. 정보는 계속 늘어날 수 있지만, 구조가 없다면 결국 새로운 혼란을 만들어낼 뿐입니다. 토큰 시대도 마찬가지입니다. 미래에 인공지능은 방대한 양의 콘텐츠를 생성할 수 있겠지만, 방대한 양의 콘텐츠가 자동으로 방대한 양의 가치를 의미하는 것은 아닙니다. 오히려 반대로, 콘텐츠 생성이 쉬워질수록 이러한 콘텐츠를 어떻게 선별하고, 조직하고, 보존하고, 색인하고, 연결할 것인지가 점점 더 중요해질 것입니다.

나는 다음과 같은 다섯 가지 자산을 가지고 있으며, 이 다섯 가지 자산은 '토큰 이후 무엇이 희소한가'에 대한 내 판단의 경험적 기초를 구성합니다.

첫째, 실제 역사 데이터. 이것은 어떤 '역사 데이터'가 아니라, 세대를 초월한 20 년간의 실제 업무 데이터입니다 – 2005 년의 지능형 송장 시스템부터 현재까지, 실험 환경이나 스크레이핑 데이터가 아닌 실제로 운영되는 물류 시스템, 재무 시스템, 세무 시스템으로부터의 데이터입니다. AI 시대에 '합성 데이터'가 넘쳐나는 상황에서 실제 역사 데이터의 가치는 급격히 상승하고 있습니다. 대부분의 기업이 가진 것은 '단일 분야 역사 데이터'인 반면, 제가 가진 것은 학제간이며 20 년에 걸쳐 실제로 검증된 역사 데이터입니다.

둘째, 여러 학제간 시스템 및 체계의 완성. 제 시스템은 다음 분야를 포괄합니다: 지능형 물류 시스템, 지능형 송장 시스템, 지능형 재무 시스템, 지능형 세무 시스템, 세대를 초월한 20 년간의 물류·재무 통합 체계, 웹사이트 자동화 시스템, 메타데이터 생성 시스템, '시대적 전환(The Epochal Transition)' 10 개국어 출판 시스템, DOI 관리 시스템, TROVE 및 WorldCat 링크 정리 시스템, 자동 상호 참조 시스템, Sitemap 생성

시스템, Zenodo 프로파일 생성 시스템, 크로스 플랫폼 출판 시스템. 이러한 시스템의 독자성은 다음과 같습니다: 학제간 폭(물류에서 출판, 재무에서 문헌 색인까지), 시간적 깊이(일부 시스템은 20 년간 운영), 한 사람에 의한 완성(방법의 전이 가능성을 증명). 그중에서 '시대적 전환' 전자 월간지(<https://times.net.au>)는 공개된 실증 사례입니다. 이는 ISSN 국제 간행물 번호(3083-5178)를 가지고 있으며, WorldCat(100 개국 이상의 16,000 개 이상의 국립 및 대학 도서관이 회원인 세계 최대 도서관 종합 목록 시스템), 호주 국립 도서관의 PANDORA/TROVE, 그리고 CERN 유럽 핵 연구 기구가 운영하는 Zenodo, OpenAIRE 유럽 오픈 사이언스 인프라, DataCite DOI(글로벌 디지털 객체 식별자, 60 개 이상의 문헌), ORCID(글로벌 고유 영구 학술 디지털 ID) 등의 국제적 학술 인프라에 의해 영구 보관되었습니다. 이 웹사이트는 한 사람(독립 연구자)이 콘텐츠 제작에서 최고 수준의 국제 아카이브 인프라와의 연동까지의 전체 프로세스를 완료할 수 있음을 증명합니다.

셋째, 시스템 및 체계 구축 능력. 이는 'AI 도구를 능숙하게 다루는 능력'이 아니라, 모호한 요구를 안정된 구조로 변환하는 설계 능력, 다른 시스템을 통일된 체계로 통합하는 아키텍처 능력, 무인으로 시스템을 수년간 작동시키는 엔지니어링 능력, 물류에서 출판, 재무에서 문헌으로 경험을 전이하는 메타 능력입니다. 앞서 논의한 바와 같이, 이 능력은 대량의 토큰으로 직접 획득할 수 없습니다. 이는 의사 결정, 예외 처리, 균형, 피드백 루프, 신뢰 축적에 의존하며, 이들 모두 AI 로 대체될 수 없기 때문입니다. 현재 AI 훈련 시장은 '프롬프트 엔지니어링', 'AI 라이팅', 'AI 드로잉' 등의 표면적 스킬에 집중되어 있으며, '학제간 시스템 구축 능력'을 가르치는 코스는 거의 없습니다. 이 능력은 단기간에 습득할 수 있는 것이 아니라 장기적인 실천의 축적을 통해서만 얻을 수 있기 때문입니다. 웹사이트에 게재된 40 개 이상의 독립 창작 작품, 획득한 ISSN, 60 개 이상의 DOI 는 시스템의 장기 운영 능력에 대한 최고의 공개 기록이며, 제 자신의 자기 선전이 아닙니다.

넷째, 미래를 예측하는 비전. 1993 년 원격 근무 프로토타입 시스템 시도 시작 – 대규모 원격 근무 보급보다 약 30 년 선행. 2005 년 지능형 송장 시스템 설계 – AI 보급 20 년 전에 프로세스 자동화 시작. 세대를 초월한 20 년간 시스템 통합 – 시스템 설계 시점에 20 년 후에도 운영이 필요할 것을 예견. 현재 '토큰 이후'에 대한 판단 제시 – 현재 대부분의 사람들이 아직 토큰을 쫓고 있을 때 시스템 능력이 진정한 장벽임을 예견. 또한 '토큰은 통화와 유사하며, 동서양의 기술, 시스템, 정책적 차이로 인해

보편적으로 통용될 수 없다'는 점도 예견했습니다 – 미국 달러와 중국 위안이 자유롭게 교환될 수 없는 것처럼, 중국어 토큰과 영어 토큰도 유사한 유통 장벽에 직면하며, 미래에는 '중국어 토큰 존', '영어 토큰 존' 등의 언어 및 정책 기반 토큰 생태계 구역이 나타날 것이라고 예견했습니다. 또한 '현재 토큰으로 돈을 버는 것은 이미 어렵고 점점 더 어려워질 것이며, 프론트엔드에 서비스를 제공하는 시스템이 있을 때만 토큰과 컴퓨팅 파워가 유효한 부가 자산으로서 이익을 창출할 수 있다'는 점도 예견했습니다. 저는 독립 연구자로서 미래의 학술 인프라와 연동하는 개인 출판 체계를 선제적으로 구축하여, '개인도 글로벌 지식 인프라의 하나의 노드가 될 수 있다'는 트렌드를 예견했습니다. 대부분의 트렌드 예측은 애널리스트나 기업가에 의한 것이지만, 제 '예측 능력'은 20 년의 실천 검증이라는 뒷받침을 가지고 있습니다.

다섯째, 자기 증명 능력. 이것은 제 가장 중요한 메타 능력 중 하나입니다. 제가 작성하는 모든 분야의 문헌은 이론적 클로즈드 루프와 논리적 자기 일관성을 달성합니다. 이는 제 주장이 외부 권위나 인용에 의존하지 않고 내부 논리의 자기 충족성에 의존함을 의미합니다. 학술 문헌의 기본 규칙은 '선행 연구 인용'입니다. '자기 증명 능력'을 주장하거나 보여주는 사람은 거의 없습니다. 학계의 논의는 '주장이 충분한가'에 초점을 맞추지, '저자가 자기 일관된 클로즈드 루프를 구축할 능력이 있는가'에 있지 않습니다. 제 글에서 이 자기 증명 능력은 다음과 같이 나타납니다. 주장은 실천에서 비롯되고, 실천은 주장을 검증하며, 주장과 실천이 클로즈드 루프를 형성합니다. 저는 '타인의 이론을 인용하여 자신을 증명하는' 것이 아니라, '자신이 구축한 시스템을 보여주고 시스템 스스로 말하게 하는' 것입니다. 이러한 자기 증명 능력은 공개 문헌에는 전혀 존재하지 않는 개념이며, 제 가장 독특한 메타 능력 중 하나입니다.

이 다섯 가지 자산을 조합하면, 그 희소성은 덧셈이 아니라 곱셈입니다: 실제 역사 데이터 × 학제간 시스템 × 시스템 구축 능력 × 미래 예측 비전 × 자기 증명 능력은 복제 불가능한 자산 포트폴리오를 구성합니다. 공개된 논의에서 현재로서는 이러한 대조 사례를 찾을 수 없습니다.

'국제 독립 연구자의 길'을 정리하는 과정에서 저는 이 점을 더욱 깊이 인식하게 되었습니다. 1989 년부터 2026 년까지 30 년 이상의 경험은 여러 분야에 걸쳐 있습니다. 이러한 경험은 한 번에 생긴 것이 아니라 긴 시간에 걸쳐 점차적으로 형성된 것입니다. 인공지능은 몇 초 만에 글을 생성할 수 있지만, 30 년 이상의 실제

경험을 몇 초 만에 생성할 수는 없습니다. 많은 의견을 생성할 수는 있지만, 장기적 실천을 통해 축적된 판단력을 직접 생성할 수는 없습니다. 많은 프로세스를 시뮬레이션할 수는 있지만, 현실 세계의 반복된 검증을 대체할 수는 없습니다. 이러한 의미에서 시간 자체도 중요한 자산이며, 시간의 침전으로 형성된 구조는 더욱 희소한 자산입니다.

9. 토큰 이후: 진정으로 희소한 다섯 가지 능력

토큰 자체가 더 이상 희소하지 않을 때 – 토큰 판매의 진입 장벽은 높지 않지만, 실제 역사 데이터를 보유하고, 시스템과 체계를 구축하고, 미래를 예측하는 비전은 극도로 부족해질 것입니다 – 진정으로 희소한 것은 무엇일까요? 1993 년의 원격 근무 프로토타입 시스템, 2005 년의 지능형 송장 시스템, 지능형 물류 시스템, 지능형 재무 시스템, 지능형 세무 시스템, 세대를 초월한 20 년간의 물류·재무 통합 체계에서부터 웹사이트 자동화, 메타데이터 생성, '시대적 전환' 10 개국어 월간지, DOI 관리, TROVE 및 WorldCat 링크 정리, 자동 상호 참조, Sitemap 생성, Zenodo 프로필 생성, 크로스 플랫폼 출판에 이르기까지의 장기적 실천에 기반하여, 다음 다섯 가지 능력이 진정한 장벽이 될 것이라고 생각합니다.

첫째, 구조 설계 능력. 토큰은 원자재이며, 구조는 원자재를 가치 있는 제품으로 조직하는 프레임워크입니다. 인공지능이 보급되기 훨씬 전부터 저는 제한된 자원을 활용하여 더 높은 효율의 운영을 실현하는 방법에 대해 생각하기 시작했습니다. 제 지능형 물류 시스템, 지능형 송장 시스템, 세대를 초월한 20 년간의 지능형 물류·재무 시스템은 다른 시대에 태어나 다른 문제에 직면했지만, 그들은 공통된 특징을 가지고 있습니다. 목표는 더 많은 정보를 만드는 것이 아니라, 혼란을 줄이고, 반복 노동을 줄이며, 전체 효율성을 높이고, 장기적으로 작동할 수 있는 구조를 구축하는 것이었습니다. 자동화 출판 시스템에서 제가 직면한 핵심 문제는 '텍스트를 생성할 수 있는가'가 아니라, '10 개 언어의 메타데이터 필드를 어떻게 정의할 것인가, 기사와 카테고리의 연관성을 어떻게 구축할 것인가, 버전 관리 규칙을 어떻게 설계할 것인가'였습니다. 이러한 구조 설계 결정은 시스템이 수년간 붕괴 없이 계속 작동할 수 있는지 여부를 결정합니다. 이와 대조적으로 소비되는 토큰의 양은 단지 부산물일 뿐입니다. '시대적 전환' 웹사이트 자체가 구조 설계 능력의 공개된 실증 사례입니다. 즉, ISSN 국제 간행물 번호를 가지고 있으며, 다국어를 지원하고, 여러 아카이브

채널을 가진, 개인이 유지 관리하는 학술 웹사이트 구조이며, 이는 제가 헛된 이론을 말하는 것이 아니라 이 구조 설계 능력을 사용하여 공개적으로 접근 가능한 복잡한 시스템을 성공적으로 구축했음을 증명합니다.

둘째, 자동화 능력. 토큰은 콘텐츠를 배치 생산할 수 있지만, 엔드투엔드 워크플로우를 배치 완료할 수 있는 것은 자동화 시스템뿐입니다. 지능형 물류 시스템에서 진정한 가치는 경로 계획을 위해 AI 를 몇 번 호출하는지에 있는 것이 아니라, 주문 접수부터 창고 스케줄링, 경로 최적화, 예외 처리까지 완전히 자동화된 클로즈드 루프를 구축하는 데 있습니다. 이 클로즈드 루프에서 AI 는 하나의 구성 요소일 뿐이며, 자동화 능력은 구성 요소 간의 연결 방식에 나타납니다. 시스템이 없는 사람은 대량의 토큰을 가지고 있더라도 흩어진 텍스트, 흩어진 이미지, 흩어진 표, 흩어진 코드를 계속 생성할 뿐일 수 있습니다. 반면 시스템이 있는 사람은 토큰을 자신의 워크플로우에 접속하여 자료 자동 정리, 메타데이터 생성, 링크 확인, 웹페이지 업데이트, 출판 보조, 문헌 아카이브, 색인 생성, 물류 처리, 재무 분석, 보고서 생성 등을 자동으로 실행하게 할 수 있습니다.

셋째, 학제간 실천 경험. 토큰은 모든 사람에게 열려 있지만, 한 업계의 경험을 다른 업계로 전이하는 능력은 보편적이지 않습니다. 제가 물류 시스템과 출판 시스템을 동시에 구축할 수 있는 이유는 더 많은 토큰을 가지고 있기 때문이 아니라, 메타데이터 관리, 워크플로우 설계, 예외 처리 등과 같은 전이 가능한 방법론을 축적해 왔기 때문입니다. 제 출판 시스템은 토큰이 필요합니다. 10 개국어 번역, 초록 생성, 키워드 정리, 메타데이터 생성, 링크 설명에는 많은 언어 처리가 필요하기 때문입니다. 제 웹사이트 시스템은 토큰이 필요합니다. 페이지 구조, 탐색 설명, SEO 설명, 학술 메타데이터, 다국어 소개 등이 자동 생성될 수 있기 때문입니다. 제 물류 시스템도 토큰을 사용할 수 있습니다. 고객 이메일, 화물 설명, 통관 자료, 예외 설명, 재무 메모, 작업 로그 등을 인공지능으로 보조 처리할 수 있기 때문입니다. 제 자동 상호 참조 시스템은 특히 토큰이 필요합니다. 많은 수의 기사, 제목, DOI, NLA, TROVE, WorldCat 링크 간의 주제 연관성을 찾아내야 하기 때문입니다. 이러한 학제간 실천의 축적은 토큰을 소비함으로써 얻을 수 있는 것이 아니라, 실제 문제를 해결함으로써 침전되는 것입니다. 웹사이트는 학술 출판, 인프라 통합, 기술 아키텍처, 개인 브랜드 등 여러 분야의 정보를 동시에 제시하여, 제가 단일 스킬이 아니라 출판, 도서관, 기술, 개인 브랜드를 실제로 연결했음을 증명합니다.

넷째, 시스템 장기 운영의 멘탈 모델. 현재 AI 애플리케이션의 주류 내러티브는 '대화'와 '생성'이지만, 진정으로 지속적인 가치를 창출하는 것은 '운영'입니다. 하나의 대화는 1 만 개의 토큰을 생성하고, 하나의 에이전트 작업은 100 만 개의 토큰을 소비합니다. 그러나 10 년간 지속적으로 작동하는 물류 시스템이 소비하는 토큰은 수십억 개에 달할 수 있습니다. 그리고 진정으로 희소한 것은 이 시스템을 설계하는 멘탈 모델입니다. 모델 업그레이드 후에도 시스템이 계속 사용 가능하도록 어떻게 보장할 것인가? 상류 API 의 변경에 어떻게 대처할 것인가? 예외를 모니터링하고 자동 복구하려면 어떻게 할 것인가? 이러한 문제는 토큰과 무관하지만, 토큰의 최종 산출을 결정합니다. 이러한 시스템의 가치는 토큰 자체에서 오는 것이 아니라, 제가 이미 구축했거나 개발 중인 구조에서 옵니다. 토큰은 단지 연료일 뿐이며, 시스템이 바로 엔진입니다. 토큰은 단지 전류일 뿐이며, 구조가 바로 회로입니다. 토큰은 단지 물일 뿐이며, 파이프, 밸브, 저수지, 조달 시스템이 그것이 역할을 할 수 있는지 여부를 결정합니다. 시스템이 없으면 토큰은 유실되고, 시스템이 있으면 토큰은 장기적 가치로 전환됩니다. 웹사이트는 40 개 이상의 독립 창작 작품과 CERN, 호주 국립 도서관과 같은 권위 기관에 의한 영구 보관을 보여주며, 제 시스템이 단기적인 화제성을 쫓는 것이 아니라 장기적 가치를 추구하며, 이미 권위 기관의 '영구 보존'이라는 인정을 받았음을 증명합니다.

그러나 토큰과 시스템 간의 관계에 대해 더 깊이 고려할 때, 또 하나 자주 간과되는 문제를 발견합니다. 많은 사람들이 훌륭한 시스템만 있으면 미래 토큰 시대에 우위를 점할 수 있다고 생각합니다. 그러나 현실은 그렇게 단순하지 않습니다. 시스템 자체는 누구나 설계할 수 있는 것이 아닙니다. 진정으로 성숙한 시스템은 장기적인 실천, 반복적인 수정, 실제 시나리오에 의한 검증을 필요로 합니다. 많은 시스템은 초기 구상에서 실제 구현까지 수년, 혹은 수십 년의 지속적인 축적을 필요로 합니다. 따라서 시스템 설계 능력 자체가 이미 희소한 자원입니다.

그러나 성숙한 시스템을 가지고 있다고 해서 그것이 자동으로 대규모 영향력 형성을 의미하는 것은 아닙니다. 시스템의 기술적 가능성에서 사회적 보급 사이에는 시장, 자본, 유통 채널, 사용자 교육, 플랫폼 생태계 등 여러 층위가 존재합니다. 역사는 훌륭한 기술이나 훌륭한 시스템이 부족하지 않습니다. 진정으로 희소한 것은 시스템을 대규모 응용으로 끌어올릴 수 있는 플랫폼 능력입니다. 시스템이 국지적 시나리오에서 기존 솔루션보다 훨씬 우수하더라도, 충분한 자본 지원, 플랫폼

프로모션 능력, 사용자 기반이 없으면 그 실제 이용 규모는 여전히 매우 제한적일 수 있습니다.

토큰 경제의 관점에서 보면 이는 특히 명확합니다. 토큰 수요는 시스템 자체에서 직접 발생하는 것이 아니라 시스템의 실제 이용 규모에서 발생합니다. 하루 수백만 명의 사용자에게 서비스를 제공하는 플랫폼은 시스템 설계가 상대적으로 평범하더라도 엄청난 토큰 소비를 생성할 수 있습니다. 반대로 설계는 매우 첨단적이지만 사용자 수가 제한된 시스템은 토큰 수요가 매우 작을 수 있습니다. 따라서 미래 토큰 시장의 발전은 모델 능력과 시스템 능력뿐만 아니라 플랫폼 능력과 자원 통합 능력에도 의존합니다.

이것은 미래 경쟁에 적어도 세 가지 수준이 존재함을 의미합니다. 첫 번째 수준은 모델 경쟁으로, 누가 토큰을 공급할 수 있는지 결정합니다. 두 번째 수준은 시스템 경쟁으로, 누가 토큰을 효과적으로 활용할 수 있는지 결정합니다. 세 번째 수준은 플랫폼 경쟁으로, 시스템이 얼마나 많은 사용자를 커버하고 얼마나 많은 실제 수요를 창출할 수 있는지 결정합니다. 모델, 시스템, 플랫폼은 모두 필수 불가결한 요소입니다. 모델이 없으면 시스템은 작동할 수 없습니다. 시스템이 없으면 토큰은 가치를 창출하기 어렵습니다. 플랫폼이 없으면 아무리 훌륭한 시스템이라도 소규모 사용 범위에 머물 수 있습니다.

이러한 관점에서 볼 때, 저는 토큰 자체는 단지 기초적 자원에 불과하다고 더욱 확신합니다. 진정으로 미래를 결정하는 것은 토큰을 가진 사람도, 시스템을 가진 사람도 아니라, 모델, 시스템, 플랫폼, 그리고 현실의 필요를 연결할 수 있는 사람입니다. 이러한 요소들이 클로즈드 루프를 형성할 때 비로소 토큰은 단순한 계산 단위에서 가치를 지속적으로 창출하는 생산력으로 전환됩니다.

다섯째, 토큰을 실제 비즈니스 프로세스에 내장하는 능력. 토큰은 궁극적으로 현실 세계의 어느 한 부분 – 물류, 출판, 교육, 의료 – 에 기여해야 합니다. 진정으로 희소한 것은 AI 능력의 경계를 이해하고, 동시에 특정 업계의 문제점을 이해하는 사람들입니다. 그들은 언제 AI 를 사용해야 하고 언제 사용하지 말아야 하는지 알고 있습니다. 토큰의 출력을 비즈니스 의사 결정으로 전환하는 방법을 알고 있습니다. 토큰 투입과 비즈니스 산출 간의 실제 관계를 측정하는 방법을 알고 있습니다. 이 복합적 능력은 현재 시장에서 가장 희소한 자원입니다. 웹사이트 자체는 작동 중인

'비즈니스 프로세스'입니다: 콘텐츠 제작 → 메타데이터 생성 → 크로스 플랫폼 출판 → 국제적 아카이브 → 독자 접근. 이는 제 AI 응용(토큰 소비)이 실제 영향력을 창출하는 완전한 클로즈드 루프의 출판 프로세스에 통합되어 있음을 증명합니다.

저는 미래의 진정한 경쟁은 아마도 모델 간의 경쟁도, 토큰 간의 경쟁도 아니라 시스템 간의 경쟁일 것이라고 점점 더 느낍니다. 모델은 계속 업그레이드되고, 토큰 가격은 계속 하락하며, 오늘날의 고가 자원은 미래에 매우 보편화될 수 있습니다. 그러나 시스템은 자원이 풍부하다고 해서 자동으로 나타나는 것이 아닙니다. 시스템은 장기적 축적, 지속적인 수정, 현실 세계의 검증을 필요로 합니다. 누가 토큰을 장기 운영 가능한 시스템으로 전환할 수 있느냐가 지속적인 가치를 얻을 가능성이 높습니다. 그리고 이것이야말로 수익성의 핵심이기도 합니다. 즉, 현재 토큰으로 돈을 버는 것은 이미 어렵고 점점 더 어려워질 것입니다. 프론트엔드에 서비스를 제공하는 시스템이 있을 때만 토큰과 컴퓨팅 파워가 유효한 부가 자산으로서 이익을 창출할 수 있는 것입니다.

저는 토큰 시대의 진정한 경쟁이 세 단계로 나뉠 것이라고 판단합니다. 첫 번째 단계는 자원 경쟁으로, 모두가 모델, 컴퓨팅 파워, API, 토큰 가격을 두고 다퉈립니다. 두 번째 단계는 응용 경쟁으로, 모두가 토큰을 글쓰기, 고객 서비스, 코딩, 번역, 교육, 사무 작업, 데이터 분석에 사용합니다. 세 번째 단계가 진정한 시스템 경쟁으로, 누가 토큰을 장기 운영되는 산업 시스템, 출판 시스템, 물류 시스템, 지식 시스템, 문명 기록 시스템에 내장할 수 있는가가 장기적 우위를 얻을 가능성이 높습니다.

사실, 토큰, 모델, 시스템, 플랫폼에 대해 논할 때, 현실 세계에서는 시간이라는 또 하나 중요한 요소가 종종 무시됩니다. 많은 사람들은 자본이 있으면 단기간에 시스템을 구축할 수 있고, 시장이 충분히 크면 성공 모델을 빠르게 복제할 수 있다고 생각합니다. 그러나 현실은 종종 그렇지 않습니다. 자본은 컴퓨팅 파워를 구매하고, 토큰을 조달하고, 개발팀을 고용할 수는 있지만, 수년에 걸쳐 검증되어 온 실제 데이터와 성숙한 시스템을 직접 구매할 수는 없습니다. 많은 산업에게 진정으로 희소한 것은 모델 자체도, 토큰 자체도 아니라, 실제 비즈니스 환경에서 장기간 안정적으로 작동해 온 시스템입니다.

비즈니스 관점에서 보면, 시장과 자본은 일반적으로 새로운 시스템이 처음부터 성장하기를 오랫동안 기다리지 않습니다. 대규모 시스템은 개념 제안, 아키텍처 설계,

기능 개발, 프로세스 검증에서 실제 구현까지 종종 수년 혹은 그 이상의 시간이 걸립니다. 이 과정에서는 지속적인 자원 투입뿐만 아니라 많은 시행착오, 수정, 실제 환경에서의 검증이 필요합니다. 따라서 시장 기회가 이미 나타났을 때, 자본은 처음부터 완전히 새로운 체계를 설계하기보다는 이미 형태를 갖추고, 이미 작동하고, 이미 검증된 시스템을 찾는 경향이 강합니다.

그러나 현실에는 또 다른 모순이 존재합니다. 대규모 시스템일수록 후발 자본이 쉽게 개입하기 어려워집니다. 대규모 시스템은 장기적인 발전을 거쳐 일반적으로 이미 자체적인 운영 메커니즘, 관리 구조, 데이터 로직, 조직 문화를 형성했습니다. 그들은 방대한 역사적 데이터, 복잡한 비즈니스 프로세스, 장기간 축적된 경험 체계를 가지고 있을 수 있습니다. 그렇기 때문에 대규모 시스템은 가치가 헤아릴 수 없이 크지만, 자본의 기대에 따라 신속하게 변혁하거나 완전히 재구축하기 어려운 경우가 많습니다. 시스템 규모가 클수록 그 관성도 커지는 경향이 있습니다.

따라서 현실 비즈니스 세계에서 가장 이상적인 상태는 시스템을 처음부터 설계하는 것도, 크고 경직된 체계를 단순히 인수하는 것도 아니라, 핵심 구조 구축을 이미 완료하고, 장기적인 검증을 거쳐, 안정적으로 작동할 수 있는 기존 시스템을 찾는 것입니다. 그러한 시스템은 완벽하지 않더라도 가장 어려운 단계를 이미 넘어섰으며, 실제 데이터, 실제 프로세스, 실제 검증 기록을 가지고 있기 때문에 여전히 매우 높은 가치를 가집니다. 이것들을 기반으로 리패키징, 최적화, 대외 개방, 플랫폼화 및 프로모션을 하는 것은 처음부터 구축하는 것보다 효율적인 경우가 많으며, 시행착오 리스크와 시간 비용을 크게 줄일 수 있습니다. 왜냐하면 이러한 시스템들은 이미 가장 비용이 많이 들고, 가장 시간이 많이 걸리며, 가장 복제하기 어려운 단계를 완료했기 때문입니다.

이러한 관점에서 볼 때, 시간의 창(window)도 또한 자주 간과되는 자원입니다. 산업이 아직 초기 단계에 있을 때, 많은 사람들은 모델 능력, 토큰 가격, 시장 규모에 주목하며, 시스템 형성에 필요한 시간을 무시합니다. 성숙한 시스템의 탄생은 대개 개념 제안, 아키텍처 설계, 프로세스 검증, 실제 운영, 지속적 개선 등 여러 단계를 거칩니다. 진정으로 어려운 것은 아이디어를 제안하는 것이 아니라 시스템을 실제 환경에서 장기적으로 안정적으로 작동시키는 것입니다. 따라서 시장이 성숙한 시스템의 가치를 인식하기 시작할 때, 이는 종종 이러한 시스템들이 이미 수년간의 축적을 거쳤음을 의미하며, 단기간에 재현할 수 있는 것이 아닙니다.

자본과 시장에게 가장 이상적인 대상은 개념 단계의 구상도, 아직 검증되지 않은 프로토타입도 아니라, 핵심 구조 구축을 완료하고, 실제 운영 기록을 가지고 있으며, 업그레이드 여지가 있는 시스템의 태동 단계에 있는 것입니다. 이러한 시스템은 완벽하지 않더라도 가장 어렵고, 가장 시간이 많이 걸리며, 가장 위험한 단계를 이미 넘어섰습니다. 시장 수요가 급증하면, 이러한 시스템은 매우 인기가 높아지는 경향이 있습니다. 진정으로 장기적으로 검증된 성숙한 시스템의 수는 시장의 상상보다 훨씬 적기 때문입니다. 모델은 빠르게 반복될 수 있고, 토큰은 지속적으로 확장될 수 있으며, 개발팀은 임시로 구성될 수 있지만, 실제 데이터, 실제 프로세스, 장기적 검증 기록은 단기간에 재창조할 수 없습니다.

따라서 토큰 경제가 발전해 나가는 과정에서 시간은 시스템의 형성에 영향을 줄 뿐만 아니라 시스템의 가치에도 영향을 줍니다. 시장이 특정 유형의 시스템의 중요성을 충분히 인식하지 못하고 있을 때는 종종 그러한 시스템을 획득하는 비용이 가장 낮은 시기이기도 합니다. 그리고 시장이 그 가치를 일반적으로 인식했을 때는 선택 가능한 대상이 일반적으로 이미 매우 제한적입니다. 많은 산업에게 놓치는 것은 단순한 하나의 프로젝트가 아니라, 장기적 축적을 이미 완료하고 다음 발전 단계로 직접 진입할 수 있는 시간의 창인 것입니다.

토큰 경제의 관점에서조차 마찬가지입니다. 토큰은 신속하게 입수할 수 있고, 모델은 신속하게 접속할 수 있으며, 플랫폼은 신속하게 구축할 수 있지만, 진정으로 단기간에 복제하기 어려운 것은 수년에 걸쳐 침전되어 가치를 지속적으로 창출할 수 있는 시스템입니다. 미래에 토큰이 점점 풍부해질 때, 시스템의 중요성은 감소どころか 오히려 더 높아질 수 있습니다. 왜냐하면 모든 사람이 유사한 토큰 자원을 갖게 된 후, 경쟁 결과를 결정하는 것은 더 이상 토큰의 양이 아니라, 누가 더 성숙한 실제 데이터, 더 안정적인 비즈니스 프로세스, 더 실제 필요에 가까운 시스템 구조를 가지고 있느냐일 것이기 때문입니다.

또 하나 자주 간과되는 문제는 시스템의 기원입니다. 많은 사람들은 시스템을 소프트웨어 제품으로 이해하는 데 익숙하며, 따라서 당연히 시스템은 소프트웨어 회사에 의해 설계되고 개발되어야 한다고 생각합니다. 그러나 많은 복잡한 산업에서 진정으로 장기간 안정적으로 작동하는 시스템은 종종 먼저 소프트웨어 회사에 의해 설계되고 산업에 넘겨지는 것이 아니라, 장기간의 업무 실천 과정에서 점차적으로 형성된 후, 소프트웨어화, 자동화, 플랫폼화되는 경우가 많습니다.

이유는 복잡하지 않습니다. 소프트웨어 회사는 일반적으로 기술적 구현에 능숙하지만, 반드시 장기간에 걸쳐 구체적인 업무 현장에 머물러 있지는 않습니다. 그들은 기능을 빠르게 개발할 수 있지만, 단기간에 많은 양의 실제 업무 경험을 축적하기는 어렵습니다. 반면, 업계 내부에서 자연발생적으로 발전해 온 시스템은 매일의 고객, 주문, 예외 상황, 규제 변화, 비용 압력, 시장 경쟁에 직면하면서 지속적으로 수정되고 개선되어 왔습니다. 이러한 시스템은 이론에서 출발하는 것이 아니라 실제 문제에서 출발하며, 그 모든 프로세스, 모든 규칙, 모든 자동화 부분은 종종 실제 업무의 장기적 검증에서 비롯됩니다.

따라서 물류, 공급망, 출판, 금융, 제조업, 의료 및 많은 전문 분야에서 가장 가치 있는 시스템은 반드시 가장 복잡한 시스템이나 가장 많은 기능을 가진 시스템이 아니라, 장기적인 실전 검증을 거쳐 실제 문제를 안정적으로 해결할 수 있는 시스템입니다. 그들은 최신 인터페이스를 가지고 있지 않을 수도 있고, 가장 화려한 홍보를 하고 있지 않을 수도 있으며, 최신 기술 아키텍처를 채택하고 있지 않을 수도 있지만, 많은 양의 실제 데이터, 실제 프로세스, 장기간 축적된 비즈니스 로직을 가지고 있습니다. 이러한 경험은 종종 단기간의 개발로는 얻을 수 없습니다.

토론 경제의 관점에서 보면 이는 특히 중요합니다. 미래에 모델 능력은 계속 향상되고, 토론 비용은 계속 감소하며, 소프트웨어 개발 효율성도 점점 더 높아질 것입니다. 그러나 업계가 수십 년에 걸쳐 축적해 온 업무 경험은 모델 업그레이드 때문에 자동으로 생겨나는 것이 아닙니다. 진정으로 복제하기 어려운 것은 소프트웨어 코드 자체가 아니라, 시스템 뒤에 숨겨진 업계 지식, 예외 처리 경험, 의사 결정 로직, 장기간 검증된 비즈니스 프로세스입니다.

따라서 미래의 가장 이상적인 시스템은 업계 실천에서 완전히 분리되어 설계된 것이 아니라, 실제 업무에 기원을 두고, 장기 운영을 통한 검증을 거친 후, 인공지능과 토론 능력을 활용하여 지속적으로 업그레이드되는 시스템이라고 생각합니다. 이러한 시스템은 현실 세계의 신뢰성과 인공지능 시대가 가져온 효율성 향상을 모두 갖추고 있으며, 그 장기적 가치는 단순히 기술 컨셉트에 기반하여 구축된 시스템보다 훨씬 높은 경우가 많습니다.

10. 토큰은 어디로 흘러가는가: 일시적 대화에서 문명의 기억으로

미래 토큰 경제에 대해 또 하나의 관찰이 있습니다. 많은 사람들은 토큰이 어떻게 생산되는지에 주목하지만, 토큰이 최종적으로 어디로 흘러가는지에는 거의 주목하지 않습니다. 토큰이 한 번의 채팅에만 머물면 그 라이프사이클은 불과 몇 분일 수 있습니다. 일반 웹페이지에 들어가면 라이프사이클은 불과 몇 개월일 수 있습니다. 장기 보존 체계에 들어가면 그 가치는 근본적으로 변화합니다. 토큰이 DOI, Zenodo, WorldCat, TROVE, 국립 도서관의 장기 보존 시스템, 그리고 다양한 학술 및 지식 인프라에 들어가면, 그것은 더 이상 단순한 토큰이 아니라 지식 기록, 문헌 자산, 문명의 기억으로 변모하기 시작합니다. 따라서 미래에 진정으로 중요한 문제는 더 많은 토큰을 생산하는 방법이 아니라, 토큰을 장기적으로 보존되고 역할을 계속 수행할 수 있는 구조에 들어가게 하는 방법이라고 생각합니다.

이러한 이유로 저는 토큰의 미래 발전에 낙관적이지만, 토큰 자체를 맹신하지는 않습니다. 토큰의 과거는 컴퓨터가 언어를 처리할 때의 기본 단위였습니다. 토큰의 현재는 인공지능 시대의 중요한 측정 단위이자 상업적 단위입니다. 토큰의 미래는 디지털 문명의 중요한 기초 자원이 될 가능성이 높습니다. 그러나 토큰이 어떻게 발전하든, 저는 항상 진정으로 희소한 것은 토큰이 아니라 토큰을 다룰 줄 아는 사람이라고 생각합니다. 더 많은 자원을 가진 사람이 아니라 장기적인 구조를 구축할 수 있는 사람입니다. 더 많은 콘텐츠를 생성할 수 있는 사람이 아니라 콘텐츠를 지식 체계로 조직할 수 있는 사람입니다.

11. 시간의 비압축성: 구매 불가능한 네 가지 자산

'토큰 이후에 무엇이 진정으로 희소인가'를 논할 때, 저는 대부분의 논의가 간과하는 근본적인 문제를 지적할 필요가 있습니다. 특정 자산은 본질적으로 가속, 구매, 복제가 불가능합니다. AI 가 아무리 강력해도, 토큰이 아무리 저렴해도, 다음 네 가지 자산은 항상 그 고유의 희소성을 유지할 것입니다.

첫째, 시간은 압축할 수 없다.

30 년의 실천은 단순히 30 년의 실천이며, 3 년으로 압축할 수 없고 AI 로 시뮬레이션할 수도 없습니다. 인공지능은 '물류 시스템 구축 방법'에 관한 기사를 몇 초 만에 생성할 수 있지만, 물류 시스템이 20 년 동안 겪는 다양한 이상, 고장, 정책 변경, 시장 변동을 몇 초 만에 경험할 수는 없습니다. 이러한 현실 세계의 시간 축적은 어떤 기술적 수단으로도 단축할 수 없습니다. 제가 1993 년에 원격 근무 프로토타입 시스템 시도를 시작하고, 2005 년에 지능형 송장 시스템을 설계한 사실은 우연이 아니라 실제 역사의 되돌릴 수 없는 기록입니다. 나중에 오는 사람은 누구나 제 글을 읽고 제 방법을 배울 수 있습니다. 그러나 그들은 제가 살았던 시간들을 소유할 수 없습니다. 그 시간들 자체가 바로 자산입니다.

둘째, 학제간 경험은 구매할 수 없다.

학제간 경험은 상품처럼 구매할 수 없습니다. 사람은 하루에 물류, 출판, 재무, 문헌 색인에 관한 100 권의 책을 살 수 있습니다. 그러나 그 같은 날에 이들 분야를 연결하는 실제 경험을 획득할 수는 없습니다. 진정한 학제간 능력은 지식의 축적이 아니라, 다른 분야 간의 연결을 확립하고, 공통점을 식별하며, 방법을 전이하는 능력입니다. 이 능력은 한 분야에서 실제 문제를 해결하고, 실수를 저지르고, 경로를 수정하고, 판단력을 축적함으로써만 얻을 수 있습니다. 제가 물류 시스템의 경험을 출판 시스템에 전이할 수 있는 이유는 어떤 책을 읽었기 때문이 아니라, 물류 분야에서 20 년을 보내고 출판 분야에서 또 10 년을 보냈기 때문입니다. 그리고 이 두 시간의 교차점에서 전이 가능한 방법론이 생겨난 것입니다.

셋째, 시스템 구축 능력은 복제가 어렵다.

시스템 구축 능력은 복사해서 붙여 넣을 수 있는 코드나 프로세스의 집합이 아닙니다. 코드는 복제할 수 있지만, 코드 뒤에 있는 의사 결정 로직, 취사 선택의 판단, 미래 변화에 대한 예측성은 복제할 수 없습니다. 시스템의 가치는 오늘 무엇을 할 수 있는지만 아니라, 지난 10 년 동안 얼마나 많은 변경, 얼마나 많은 장애 복구, 얼마나 많은 요구 사항 조정을 겪으면서도 작동을 계속해 왔는지에 있습니다. 이 '시스템의 역사' 자체가 구축 능력의 증거입니다. 제 웹사이트(<http://times.net.au>)는 누구나 볼 수 있습니다. 그러나 방문자가 보는 것은 결과일 뿐, 이 결과 뒤에 있는 30 년의 의사 결정 역사, 기술 선택, 구조적 조정, 지속적 유지 보수가 아닙니다. 이 구축 능력은 표절이나 모방으로 얻을 수 없습니다.

넷째, 장기 운영 기록은 단기간에 얻을 수 없다.

시스템이 20 년 동안 작동했다는 기록은 그 자체로 본질적으로 단기간에 얻을 수 없습니다. 장기 운영 기록은 시스템의 안정성, 유지 보수성, 설계의 선견성을 증명합니다. AI 시대에 '모양만 그럴듯한' 웹사이트나 시스템을 몇 분 만에 생성하는 것은 가능할 수 있습니다. 그러나 '20 년 동안 작동해 왔고 CERN, WorldCat, 호주 국립 도서관에 의해 영구 보관된' 시스템을 생성하는 데 지름길은 없습니다. 장기 운영 기록은 타임스탬프이며, 이 시스템이 현실 세계의 시험을 견뎌냈다고 모든 사람에게 선언하는 것입니다. 이러한 기록은 어떤 AI 기술로도 위조하거나 가속화할 수 없습니다.

이 네 가지 자산 – 시간, 학제간 경험, 시스템 구축 능력, 장기 운영 기록 – 은 본 글에서 제가 '진정으로 희소한' 것으로 논해 온 기초를 공동으로 구성합니다. 제가 이것들을 가지고 있는 것은 다른 사람들보다 더 똑똑하거나 근면하기 때문이 아니라, 다른 사람들보다 일찍 시작했고 지금까지 계속해 왔기 때문입니다. AI 가 점점 더 많은 것들을 저렴하고 쉽게 접근 가능하게 만드는 시대에, 이러한 압축, 구매, 복제, 단기간 획득이 불가능한 것들은 점점 더 귀중해질 것입니다.

12. 결론: '토큰 쌓기'에서 '시스템 구축'으로

토큰의 미래는 토큰 그 자체에 있는 것이 아닙니다.

대규모 모델 기술의 지속적인 발전과 추론 비용의 지속적인 하락에 따라 토큰의 공급 능력은 계속 확대될 것입니다. 오늘날 고가로 여겨지는 토큰은 몇 년 후에는 오늘날의 인터넷 트래픽이나 하드 디스크 저장 장치처럼 저렴하고 보편화될 것입니다. 모든 사람이 쉽게 대량의 토큰을 얻을 수 있게 되었을 때, 진정으로 희소해지는 것은 더 이상 토큰이 아니라 토큰을 현실의 가치로 전환하는 능력 – 구조 설계 능력, 자동화 능력, 업계 경험, 장기적 실천의 축적 – 일 것입니다. 그리고 더 깊은 희소성은 압축, 구매, 복제, 단기간 획득이 불가능한 것들, 즉 시간 자체, 학제간 경험, 시스템 구축 능력, 장기 운영 기록에 있습니다. 앞서 논의한 바와 같이, 시스템 구축 능력이 대량의 토큰을 통해 직접 획득될 수 없는 이유는 그것이 본질적으로 생성 능력이 아니라 판단 능력, 예외 처리 능력, 균형 능력, 피드백 흡수 능력, 신뢰 축적 능력이기 때문입니다.

이러한 능력들의 공통점은 현실 세계의 시간, 실천, 오류, 수정, 검증을 필요로 한다는 것입니다.

수익성 측면에서도 이 판단은 마찬가지로 적용됩니다. 현재 토큰으로 돈을 버는 것은 이미 어렵고, 점점 더 어려워질 것입니다. 가격전은 이윤 폭을 계속 압박하고 있으며, 단순한 토큰 재판매 비즈니스 모델은 빠르게 저이윤 또는 무이익으로 향하고 있습니다. 토큰이 프론트엔드 시스템, 즉 서비스를 제공하는 시스템에 부속될 때만 토큰과 컴퓨팅 파워가 유효한 부가 자산으로서 이익을 창출할 수 있습니다. 토큰 자체는 이익의 원천이 아니라 시스템이 이익의 원천입니다. 토큰은 시스템의 연료일 뿐 엔진이 아닙니다.

이러한 관점에서, 제가 이미 완성했거나 개발 중인 시스템은 미래 토큰 발전에 특별한 의의를 가집니다. 왜냐하면 이러한 시스템들은 일시적인 화제성을 쫓기 위해 임시로 구축된 것이 아니라, 토큰이라는 개념이 공론화되기 훨씬 전부터 존재했기 때문입니다. 지능형 물류 시스템은 현실 업무가 구조화, 프로세스화, 자동화될 수 있음을 증명했습니다. 출판 시스템은 개인이라도 다국어, 멀티 플랫폼, 멀티 DOI, 멀티 아카이브 채널의 문헌 체계를 구축할 수 있음을 증명했습니다. 웹사이트와 메타데이터 시스템은 디지털 콘텐츠가 지속적으로 조직화, 색인화, 전파될 수 있음을 증명했습니다. 자동 상호 참조 시스템은 더 나아가 작품 수가 계속 증가할 때, 진정으로 중요한 것은 콘텐츠를 계속 쌓아 올리는 것이 아니라 콘텐츠 간의 구조적 관계를 구축하는 것임을 증명하고 있습니다. '시대적 전환'

웹사이트(<https://times.net.au>)는 이러한 성과들의 집중 전시 플랫폼으로서, 토큰 능력을 공개되고, 영속적이며, 검증 가능한 학술 체계로 전환하는 것의 실증 사례 그 자체입니다.

미래에 모든 사람에게 같은 양의 토큰이 주어진다고 해도, 최종적으로 생산되는 결과는 완전히 다를 수 있습니다. 더 많은 대화 기록을 얻는 사람이 있는가 하면, 더 많은 웹페이지 콘텐츠를 얻는 사람이 있을 것입니다. 반면 수십 년 동안 지속적으로 작동하는 물류 시스템, 출판 시스템, 지식 체계를 구축하는 사람도 있을 것입니다. 차이를 결정하는 것은 토큰 자체가 아니라 그 뒤에 있는 시스템 아키텍처이며, 그리고 그 시스템 아키텍처를 지탱하는 압축 불가능한 시간, 구매 불가능한 학제간 경험, 복제하기 어려운 구축 능력, 단기간에 얻을 수 없는 장기 운영 기록입니다.

따라서 저는 토큰을 종착점이 아니라 인텔리전트 시대의 새로운 인프라로 간주합니다. 토큰은 전기처럼 중요하지만, 전기 자체가 자동으로 공장을 건설하지는 않습니다. 토큰은 화폐처럼 측정 가능하지만, 화폐 자체가 자동으로 우량 자산을 창출하지는 않습니다. 토큰은 물처럼 흐를 수 있지만, 파이프, 조달 시스템, 장기적인 유지 보수 기록이 없으면 그저 흘러내릴 뿐입니다. 미래에 진정으로 가치 있는 사람은 반드시 가장 많은 토큰을 가진 사람이 아니라, 토큰을 시스템으로 조직할 수 있는 사람일 것입니다. 토큰 시대의 가장 큰 오해는 토큰을 목표로 삼는 것입니다. 그리고 토큰 시대의 가장 큰 기회는 토큰을 이용하여 다른 사람이 복제할 수 없는 구조를 구축하는 것입니다. 미래에 토큰이 전기처럼 보편화되었을 때, 사람들은 오늘날의 토큰 열풍을 되돌아보며, 인간 간의 격차를 진정으로 결정했던 것은 결코 소유한 토큰의 양이 아니라, 지속적으로 학습하고, 지속적으로 실천하고, 지속적으로 축적하고, 지속적으로 시스템을 구축할 수 있는 능력, 그리고 이러한 능력들 뒤에 있는 압축, 구매, 복제, 단기간 획득이 불가능한 것들을 가지고 있었는지 여부였음을 발견할지도 모릅니다. 그리고 제 견해로는, 이 능력이야말로 토큰 시대에서 가장 희소하고 가장 복제하기 어려운 부(富)입니다.

연구를 깊이 할수록, 미래 인공지능 시대의 경쟁 구도는 아마도 하나의 간단한 비유로 요약될 수 있을 것이라고 점점 더 느낍니다. 많은 사람들은 컴퓨팅 파워, 모델, 토큰을 핵심 자원으로 간주하지만, 현실 세계의 운영 법칙에서 보면 그것들은 오히려 군대 뒤에 있는 군수 체계에 더 가깝습니다. 컴퓨팅 파워는 생산 능력을 결정하고, 모델은 조직 능력을 결정하며, 토큰은 수송 능력을 결정합니다. 이들은 공동으로 인공지능 시대의 물자 지원 체계를 구성합니다.

그러나 전장의 결과를 진정으로 결정하는 것은 군수 물자 그 자체가 아니라 전선의 작전 체계입니다. 미래 산업 경쟁을 위해 시스템은 오히려 전진선(front line)과 같습니다. 시스템은 자원의 사용 방법, 프로세스의 실행 방법, 가치의 창출 방법, 사용자가 진정으로 혜택을 받는지 여부를 결정합니다. 토큰과 컴퓨팅 파워가 아무리 풍부해도 시스템에 의한 조직화가 결여되면, 그것들은 창고에 쌓인 물자처럼 실제 성과로 전환될 수 없습니다.

이러한 관점에서 볼 때, 미래의 경쟁은 아마도 컴퓨팅 파워 간의 경쟁이나 토큰 간의 경쟁이 아니라 시스템 간의 경쟁일 것입니다. 시스템은 전진선이며, 공격의 방향성과 작전 능력을 결정합니다. 토큰은 탄약과 보급이며, 지속적인 작전 능력을 결정합니다.

컴퓨팅 파워는 후방의 산업 체계이며, 생산 속도와 공급 능력을 결정합니다. 후방의 공급 없이는 전진선이 장기간 전진할 수 없습니다. 그러나 후방의 공급만 있고 전진선 체계가 없다면 마찬가지로 성과를 올릴 수 없습니다.

따라서 컴퓨팅 파워가 점점 보편화되고, 모델이 점점 유사해지며, 토큰이 점점 풍부해졌을 때, 진정으로 희소한 것은 여전히 시스템입니다. 누가 더 성숙한 시스템을 가지고 있고, 더 현실적인 데이터를 가지고 있으며, 더 장기적으로 검증된 프로세스를 가지고 있고, 토큰과 컴퓨팅 파워를 현실의 필요에 연결할 수 있는가가 미래에 지속적인 우위를 얻을 가능성이 높습니다.

토큰은 AI 시대의 '전기'가 되어가고 있습니다 – 필수 불가결하지만, 그 자체가 목적은 아닙니다. 전기 혁명의 승자가 발전소가 아니라 전구, 모터, 전해조를 발명한 발명가와 기업가였던 것처럼, AI 시대의 진정한 승자는 토큰을 이용하여 복제 불가능한 구조와 체계를 구축할 수 있는 시스템 구축자일 것입니다. 토큰 이후, 이것이야말로 진정으로 희소한 것입니다.

본 글 전체를 되돌아보면, 미래 인공지능 시대의 경쟁 구도는 아마도 하나의 나무에 비유될 수 있다고 점점 더 느낍니다.

많은 사람들은 물과 영양분에 주목합니다. 그것들은 눈에 보이고, 만질 수 있으며, 빠르게 측정할 수 있기 때문입니다. 오늘날의 컴퓨팅 파워, 모델, 토큰은 끊임없이 공급되는 물과 영양분과 같습니다. 컴퓨팅 파워는 생산 능력을 결정하고, 모델은 변환 능력을 결정하며, 토큰은 수송 능력을 결정합니다. 기술의 진보에 따라 이러한 자원들은 점점 더 풍부해지고 점점 더 쉽게 얻을 수 있게 됩니다.

그러나 하나의 나무의 생명력을 진정으로 결정하는 것은 물과 영양분 자체가 아니라 뿌리 시스템입니다.

시스템이 바로 미래 디지털 문명의 뿌리입니다.

뿌리 시스템은 물의 흡수 방법, 영양분의 분배 방법, 나무 전체가 장기적으로 성장할 수 있는지 여부를 결정합니다. 뿌리가 없으면 아무리 많은 물도 흘러내릴 것입니다. 뿌리가 없으면 아무리 풍부한 영양분도 생명력으로 전환될 수 없습니다. 마찬가지로,

시스템이 없으면 아무리 강력한 모델, 아무리 저렴한 토큰, 아무리 방대한 컴퓨팅 파워라도 지속적으로 가치를 창출하기 어렵습니다.

이러한 관점에서 보면, 컴퓨팅 파워는 물이고, 토큰은 영양분이며, 모델은 영양분을 운반하는 파이프이고, 시스템은 현실 세계에 깊이 뿌리내린 뿌리 시스템입니다. 미래 경쟁의 결과를 진정으로 결정하는 것은 누가 더 많은 물을 가지고 있느냐도, 누가 더 많은 영양분을 가지고 있느냐도 아니라, 누가 더 깊고, 더 안정적이며, 이러한 자원을 장기적으로 흡수하고 활용할 수 있는 뿌리를 가지고 있느냐입니다.

미래에 컴퓨팅 파워가 계속 성장하고, 모델이 계속 진화하며, 토큰 가격이 계속 하락함에 따라 자원 자체는 점점 더 평범해질 것입니다. 진정으로 희소해지는 것은 더 이상 자원이 아니라, 자원을 조직하고, 활용하고, 가치를 지속적으로 창출할 수 있는 시스템입니다. 누가 성숙한 시스템을 가지고 있는가가 뿌리를 가지고 있는가이며, 누가 뿌리를 가지고 있는가가 장기적 생존과 지속적 성장의 능력을 가지고 있는가입니다.

따라서 저는 토큰 이후에 진정으로 희소한 것은 더 많은 토큰도, 더 많은 컴퓨팅 파워도 아니라, 컴퓨팅 파워, 모델, 토큰을 현실의 가치로 전환할 수 있는 시스템이라고 생각합니다.

물과 영양분이 점점 더 풍부해질 때, 나무가 장기적으로 성장할 수 있는지 여부를 진정으로 결정하는 것은 여전히 뿌리 시스템입니다.

키워드: 토큰, 인공지능, AI, 대규모 언어 모델, 컴퓨팅 파워, 모델 경제, 토큰 경제, 중국어 토큰, 영어 토큰, BPE 알고리즘, 데이터 주권, 에이전트, API 생태계, 시스템 구축 능력, 시스템 설계, 자동화 시스템, 학제간 연구, 학제간 실천, 실제 역사 데이터, 장기 운영 시스템, 시간 자산, 플랫폼 경쟁, 시스템 경쟁, 디지털 문명, 문명의 기억, 지식 인프라, 독립 연구자, 국제 독립 연구자, 지능형 물류 시스템, 10 개국어 출판 시스템, 메타데이터 시스템, 자동 상호 참조, DOI, Zenodo, WorldCat, TROVE, The Epochal Transition, 우자오후이, JEFFI CHAO HUI WU

참고문헌:

1. *A Structural Case Study of the 1993 Remote Work Prototype System*
DOI: <https://doi.org/10.5281/zenodo.17978371>
 2. *The 2005 Intelligent Invoice System*
DOI: <https://doi.org/10.5281/zenodo.19044755>
 3. *Mature Structural Logistics System*
DOI: <https://doi.org/10.5281/zenodo.20201895>
 4. **Cross-Generational 20-Year Intelligent Logistics and Financial System**
DOI: <https://doi.org/10.5281/zenodo.20352371>
NLA: <https://nla.gov.au/nla.obj-4204311557>
 5. *Analysis of the Homepage Structural Evolution of The Epochal Transition*
DOI: <https://doi.org/10.5281/zenodo.20362574>
NLA: <https://nla.gov.au/nla.obj-4204309341>
 6. *An Empirical Study of Institutional Engineering: Independent Scholars Accessing the WorldCat Global Bibliographic Infrastructure*
DOI: <https://doi.org/10.5281/zenodo.18028572>
 7. *National-level Network Archiving and Digital Document Preservation Guidelines of the Australian Winner Information Network*
DOI: <https://doi.org/10.5281/zenodo.17888259>
 8. *The Institutional Archiving of The Epochal Transition*
DOI: <https://doi.org/10.5281/zenodo.17932379>
- WU, J. C. H. (2026). Source Declaration for Audiovisual and Derivative Adaptations of a Continuing Real-World Narrative. Zenodo.
<https://doi.org/10.5281/zenodo.18160116>