

数据、AI与决策

来自学术研究与跨界创业的三个观察

Data, AI, and Decision-Making

欧阳书淼

牛津大学赛德商学院金融副教授 · OxSci 创始人

CUHKsz EMBA · Oxford Prospects Programme · 2026年6月11日

个人轨迹：一条由问题驱动的跨界路径

A Journey Driven by One Question, Across Many Boundaries



贯穿始终的问题：

如何去通过实践拓展认知的边界和世界的边界？

今天留给各位企业决策者的三个问题

Three Questions for the Decision-Makers in the Room

01

副产品

你公司真正值钱的
"副产品"是什么？

可能不是你正在收费的那个业务。

02

AI性格

你在用AI替你做决策时，
知道它的"性格"吗？

不同AI的风险偏好差异，比人类群体还大。

03

言行不一

当用户/员工/客户
说的和做的不一样时，
你听哪个？

隐私悖论是一个被学术界讨论了20年的难题。

三个研究观察，每一个都对应一个企业决策者应该重新思考的问题。

PART 01

副产品可能比主业更值钱

基于蚂蚁集团 / 阿里巴巴独特数据的深度洞察

Ouyang (JFE, 2026), "Cashless Payment and Financial Inclusion"

研究背景：8-9年的独家数据访问

Eight-Year Academic Partnership with Ant Group / Alibaba

研究背景

与蚂蚁集团 / 阿里巴巴长达 8-9 年的深度学术合作。

获得了难得的独家数据访问权限，得以从内部观察一家大科技公司的数据生态。

核心问题：

移动支付除了“方便”，还创造了什么更深层的价值？

数据规模

41,485

支付宝用户完整月度面板数据

2017年5月 — 2020年9月

学术研究的核心方法论：找到因果

Why Most Business Reports Get the Conclusion Wrong

商业报告的逻辑

A 和 B 相关 → 得出结论。

看到相关性即认为是因果。

学术研究的逻辑

A 和 B 相关 → 会不会是 C 同时导致了 A 和 B？

需要找到"外生冲击"排除混淆因素。

本研究的识别策略：共享单车自然实验



关键点：共享单车投放时间是外生的，与用户原本的信用状况无关。

核心发现：交易数据解锁信用准入

Transaction Data Substitutes for Traditional Collateral

广度边际 Extensive Margin

+56.3%

使用无现金支付后，获得花呗信用额度的概率显著提升。

强度边际 Intensive Margin

+0.41%

每增加 1% 的支付流水，信用额度增加 0.41%。

关键机制：数据替代了抵押品

不是因为用户更有钱——控制了资产、收入等变量后，效果依然显著。真正的逻辑是：交易频率传递收入稳定性的信号，消费结构传递生活方式的信号，没有传统征信记录的人也能被评估。

支付宝的价值结构重构

Payment Is the Door, Data Is the Asset, Credit Is the Monetization

	支付业务	信贷业务
收入来源	手续费 (0-0.6%)	利息收入 (年化15-18%)
利润率	极薄	丰厚
竞争壁垒	低 (激烈竞争)	高 (数据独占)
战略地位	入口	变现

核心洞察

看似低利润的支付业务，实际上是构建数据护城河的关键基础设施。副产品（数据）比主业（支付费率）更值钱。

副产品审计框架

Audit Your Business for Its Most Valuable Byproduct

主业产生的"副产品"是否比主业本身更有价值？

01 识别"废料"

数据 · 关系 · 能力 · 资产

交易记录、客户网络、团队 know-how、品牌信任...

02 评估价值

这些"废料"对谁有价值？

上下游企业？完全不同的行业？未来的自己？AI？

03 战略选择

哪个"废料"可能比主业更值钱？

护城河更高？利润率更厚？增长更快？

PART 02

AI 有自己的"性格"

不同AI模型存在系统性的决策偏见——这影响你的商业决策

Ouyang, Yun, and Zheng (2025), "AI as Decision-Maker: Ethics and Risk Preferences of LLMs"

实验设计：把AI放上行为经济学的实验台

Putting 50 Frontier LLMs Through Classic Behavioral Experiments

核心问题

AI 真的是"中性"的吗？不同AI会不会有不同的"偏见"？

50

研究对象

50个最前沿的大语言模型 (LLMs)

5

测量方法

5种行为经济学经典实验

100+

重复次数

每个模型每种方法重复100+次

发现一：模型间的风险偏好差异巨大

AI Risk Preferences Vary More Than Across Human Populations

高风险偏好

Zephyr-7B-Beta

8.11 / 10

中等风险偏好

大多数模型

4 - 6 / 10

低风险偏好

Baichuan2-7B-Chat

0 / 10

核心洞察

并非随机噪音，而是系统性差异。同样的投资决策任务，有的模型极度激进，有的模型极度保守。

这个差异，比人类不同群体之间的差异还大。

发现二与三：性格稳定，且对齐有代价

Personality Is Stable. Alignment Has a Cost.

发现二：性格是稳定的

一个在投资决策中保守的模型，
在其他领域也会保守。



性格不是随机噪音，而是这个模型的系统性特征。

发现三：对齐的代价 (Alignment Tax)

经过 RLHF（人类反馈强化学习）的模型
虽然更“安全”，但也变得更“胆小”。

更高对齐



更强风险厌恶

结果：完全依赖AI做市场预测，可能系统性低估真实CEO的扩张意愿，
错失增长机会。

选 AI, 也是选"性格"

There Is No Best AI. There Is Only the Best Match.

没有"最好"的模型, 只有"最匹配"的模型。

创新与创意任务

广告文案 · 艺术创作 · 头脑风暴

选激进型 (High Risk Preference)

严谨与合规任务

医疗诊断 · 法律审核 · 金融建议

选保守型 (High Alignment)

实操建议: 不要只看智商榜单 (MMLU), 更要关注性格测试 (Behavioral Profile)。

PART 03

用户说的和做的，经常不一样

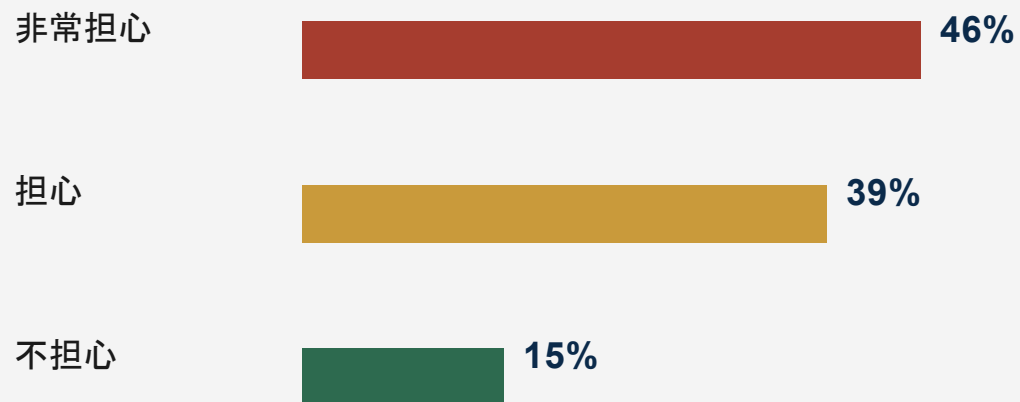
隐私悖论——一个被学术界讨论了20年的难题

Chen, Huang, Ouyang, Xiong (2025), "Data Privacy, Data Sharing and Credit Access"

85%的用户说担心隐私，但行为数据显示相反

When What People Say Contradicts What They Do

调查数据：声称担忧程度



85% 的用户表示担心或非常担心。

行为数据：实际授权数量



三组之间，实际授权数量几乎没有差别。

悖论的解答：隐藏的第三变量

It's Not a Paradox. It's an Omitted Variable.

隐藏变量：对数字服务的需求程度



影响一：认知层面

重度用户更了解数字服务

→ 更了解隐私风险

→ 更担心

影响二：行为层面

重度用户对服务需求更强

→ 必须使用服务

→ 继续分享数据

给企业决策者的三条实操建议

- ① 先信行为，不要被声称偏好误导
- ② 寻找隐藏的第三变量
- ③ 控制第三变量后再验证关系

PART 04

OxSci : 把三个洞察拼起来

The Credit Rating Agency for Science

五年前我研究的是金融信任如何被数据重构,
今天我在做的是科学信任如何被 AI 和数据重构。
底层是同一个问题。

为什么是现在：三股力量在汇聚

Three Forces Converging — Right Now

01

产量危机

AI 让论文产量爆炸式增长

→ 对质量筛选的需求空前高涨

Springer Nature 2025 投稿 310 万篇, 年增 30%

02

信任危机

论文工厂 + AI 幻觉

→ 既有信任基础设施正在失效

arXiv 2025 拒稿率三倍上升 (4% → 10%)

03

技术拐点

LLM 首次能做有意义的科研批判

→ AI 可承担 80% 的机械性审稿工作

释放专家做最终判断

AI 将把全球年论文产出从 1000 万推向 1 亿。窗口期就是现在。

现有体系的两难：要么慢但可信，要么快但无信任

The Existing System Cannot Scale

可信

传统期刊

Nature, Elsevier, Springer

2-3 年发表周期

慢 + 可信 (但正在崩塌)

★ OxSci

AI + 专家认证

7 天发表周期

快 + 可信

无信任

Review Commons

志愿者 + 慢

证明了需求 · 但无法规模化

预印本平台 / 纯AI审稿

arXiv, bioRxiv, aiXiv

快 + 无信任信号

← 慢 (月)

快 (天) →

OxSci 的工作流：从上传到认证发表，7 天

Upload → AI Review → Expert Certify → Published with DOI

01

研究者上传论文

任何学科，任何阶段。
不受期刊范围限制。

02

AI 引擎初审

方法核查、统计验证、
引用验证、新颖性评估。

03

领域专家认证

本领域真实学者修正AI输出，
做最终认证判断。

04

认证 + 发表

带 DOI 的同行认证预印本。
7 天，而不是 2-3 年。

从投稿到认证发表：7 天。

传统期刊：2-3 年。

洞察一应用：副产品就是数据资产

Same Logic, Two Industries: The Real Asset Is the Data

支付宝

表面：

卖支付服务（手续费 0-0.6%）

真正的资产：

交易数据（替代抵押品做信用评估）

变现：

信贷业务（年化 15-18%）

OxSci

表面：

卖 AI 审稿订阅（\$49/月）+ 专家认证（\$199/篇）

真正的资产：

专家审稿数据（前沿知识的高质量批判性标注）

变现：

客户付费让我们获得 \$150+/小时专家的标注数据。

Negative-cost frontier LLM training data.

同一个商业逻辑，两个不同的行业。表面服务是入口，数据是真正的护城河。

两个引擎，一个飞轮

AI Subscriptions Scale. Expert Reviews Generate Irreplaceable Data.



引擎 01

AI 订阅

\$49/月 · 3 个 AI 分析点数

经常性收入 · ~85% 毛利

可无限规模化扩展

每一次专家审稿都让 AI 变得更准

引擎 02

专家认证审稿

\$199/篇 · 专家认证 · 7天交付

每篇约 15 处专家修正

数据资产价值：\$200-400/篇

低成本的前沿 LLM 训练数据

Scale.ai 付标注员 \$40-80/小时。这个水准的科学家 \$150+/小时，而且大多不愿被雇佣做标注。

OxSci 的客户付钱给我们，这些专家在过程中产生数据。

洞察二应用：纯AI不够，必须 AI + 人

Pure AI Can Be Gamed. Hybrid Is the Only Scalable Trust.

研究告诉我们的

- AI 有自己的性格，存在系统性偏差。
- 经过对齐的模型更"胆小"，会系统性低估扩张机会。
- 纯 AI 决策可以被 gaming——一旦评判标准被知晓，作者可以针对性优化文本骗过 AI。

OxSci 的架构选择

- AI 做 80% 的机械工作：方法核查、引用验证、统计校验。
- 人做 20% 的最终判断：新颖性、领域贡献、是否值得发表。
- 这不是退而求其次的混合，是结构上唯一可规模化的信任机制。

学术研究的结论，被直接落实成创业公司的产品架构决策。

跨界创业的三个困惑

Three Real Tensions in Crossing from Academia to Founding

时间尺度

学术：一篇好论文 3-5 年。

创业：一个季度决定生死。

在两种节奏之间切换是最痛苦的部分。学术训练让人想等所有证据都到位再决策，但创业要求在信息严重不足时下注。

决策确定性

学术：审稿人帮你逐字找bug，错了可以撤稿重写。

创业：错了就是真金白银，没有审稿人。

学术品味 (research taste) 和商业判断力 (business judgment) 底层是同一种能力——在噪音中识别有价值的信号。

身份定位

教授？创业者？

应该写论文还是见投资人？

我后来想明白：不是两个身份，是同一个问题在两个不同的执行层。研究是把信任问题想清楚；OxSci 是把它做出来。

结尾：研究品味 ≈ 商业判断力

Taste, in Research and in Business, Is the Same Capability

底层能力是通用的：在信息不充分的噪音中，识别真正有价值的信号。

顶尖学者

"这个选题值得我投入未来3-5年的生命吗？"

在数据还没跑出来之前，就能判断这个方向是否可能有颠覆性发现，而不是增量修补。

顶尖企业家

"这个需求值得我投入全公司的资源吗？"

在市场还没爆发之前，就能判断这个需求是否是伪需求，以及天花板有多高。

在座各位董事长每天都在做这件事。

留给各位的最后一个问题

The Question to Take Home

一个真实的观察

员工 A：能力强，抵触 AI

履历漂亮、各方面能力都很强，但对 AI 有抵触，认为追求效率是在牺牲质量。

员工 B：起点低，拥抱 AI

起点不高，但愿意拥抱 AI，一个人能全链路跑通，实际产出可以替代一个小团队。

AI 时代，能力不是瓶颈，开放度才是。

你的组织里，谁是员工 A，谁是员工 B？

AI 的能力边界在哪里？怎样的人力资本会增值，怎样的人力资本会贬值？

感谢相遇

Thank You — Q & A

欧阳书淼

牛津大学赛德商学院金融副教授 · OxSci

ShumiaoOuyang.com · oxsci.ai



佩風

United Kingdom Oxford



Scan to add me as a friend.