

2024上海图书馆开放数据竞赛巡讲·中山大学

计算伦理与数字道德

——AGI时代的人文主义和数字人文



刘炜 上海图书馆上海科技情报所

wliu@libnet.sh.cn

2024上海图书馆开放数据竞赛巡讲·中山大学

GEN·AI与信息资源管理：机遇与挑战



刘炜 上海图书馆上海科技情报所

wliu@libnet.sh.cn

GENAI发展态势：上游成型，下游洗牌

- 英伟达点燃十万卡集群，AGI就在地平线
- 模型能力未见极限，应用架构与生态逐渐成型，国内外冰火两重天
- Apple Intelligence揭幕AI进入大众消费，应用即将爆发，AI大隐于形
- 素养普及任重道远

LLM继续飞奔

- GPT(5)还要18个月即能达到人类博士水平（AGI） —Mira

Claude 3.5 Sonnet benchmarks

	Claude 3.5 Sonnet	Claude 3 Opus	GPT-4o	Gemini 1.5 Pro	Llama-400b (early snapshot)
Graduate level reasoning <i>GPQA, Diamond</i>	59.4%* 0-shot CoT	50.4% 0-shot CoT	53.6% 0-shot CoT	—	—
Undergraduate level knowledge <i>MMLU</i>	88.7%** 5-shot	86.8% 5-shot	—	85.9% 5-shot	86.1% 5-shot
	88.3% 0-shot CoT	85.7% 0-shot CoT	88.7% 0-shot CoT	—	—
Code <i>HumanEval</i>	92.0% 0-shot	84.9% 0-shot	90.2% 0-shot	84.1% 0-shot	84.1% 0-shot
Multilingual math <i>MGSM</i>	91.6% 0-shot CoT	90.7% 0-shot CoT	90.5% 0-shot CoT	87.5% 8-shot	—
Reasoning over text <i>DROP, F1 score</i>	87.1 3-shot	83.1 3-shot	83.4 3-shot	74.9 Variable shots	83.5 3-shot Pre-trained model
Mixed evaluations <i>BIG-Bench-Hard</i>	93.1% 3-shot CoT	86.8% 3-shot CoT	—	89.2% 3-shot CoT	85.3% 3-shot CoT Pre-trained model
Math problem-solving <i>MATH</i>	71.1% 0-shot CoT	60.1% 0-shot CoT	76.6% 0-shot CoT	67.7% 4-shot	57.8% 4-shot CoT
Grade school math <i>GSM8K</i>	96.4% 0-shot CoT	95.0% 0-shot CoT	—	90.8% 11-shot	94.1% 8-shot CoT

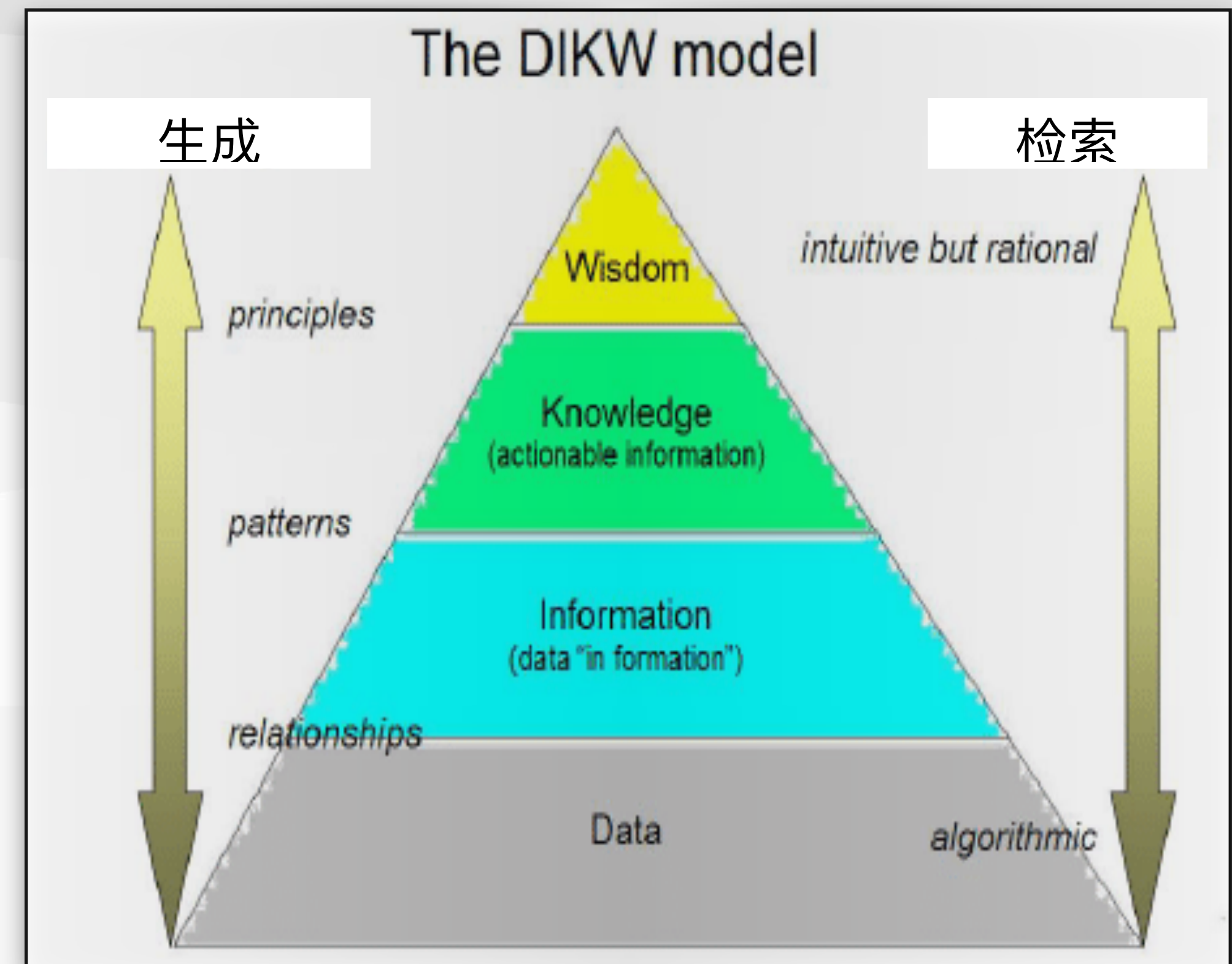
* Claude 3.5 Sonnet scores 67.2% on 5-shot CoT GPQA with maj@32

** Claude 3.5 Sonnet scores 90.4% on MMLU with 5-shot CoT prompting



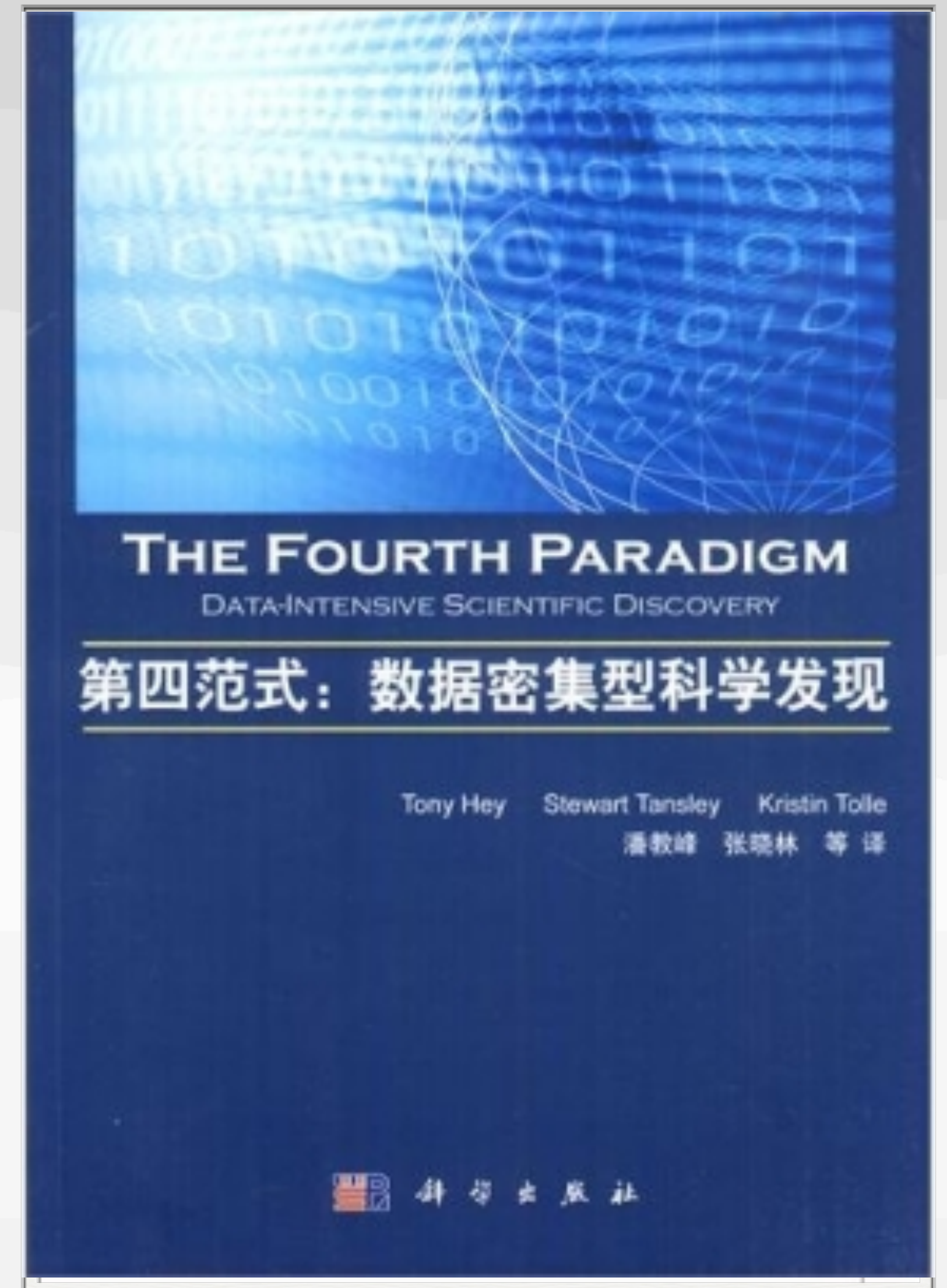
A REAL BREAKTHROUGH

- LLM是海量知识在深度神经网络中的压缩形态，可以认为是智慧的一种编码存储形式。
- LLM是基于词元（token）而不是符号的：词元是一种张量，是语义相似性的度量而不是符号匹配，因此它可以直接告诉答案而不是符号的排列组合。
- 知识是随时生成的而无需预先存储的：存储知识只是AI的Bootloader，具有具身学习能力的智能体无需存储知识，只需要参数权重存储的智能即可以随时产生知识。
- LLM应用以端到端模型为最高形态，端到端是指只要给定数据就能得到智慧。
- 人类作为知识链的起点，其知识生产虽然节能，但却是极其原始而粗糙的。LLM一旦形成便不再需要人类的帮助，可以通过自我学习（自己创造数据进行学习）而得到，并在应用中不断迭代（数据飞轮）。
- LLM短期赋能传统的知识工作，长期将会颠覆整个知识产业模式。



科学研究的范式理论

- 科学研究的范式（paradigm）理论来自于托马斯库恩，是指在特定历史时期内，科学家们共同遵循的一组基本信仰、方法和实践。
- 科学范式通常包括：1. 基本假设；2. 理论体系（相互联系的概念、原则和法则）；3. 方法论（观察、实验、数据收集和分析等的过程和规范要求）；4. 研究范围和问题；5. 实践和应用原则；6. 社会认可和共识（学术圈及利益相关者）。
- 范式概念常被泛化使用，指一项事物发展过程中的关键和本质要素，例如共同的指导原则、方法论、核心能力等。范式可能随着新的理论方法的出现而发生变革更替，新范式取代旧范式才会带来进步。



科学研究的五个范式

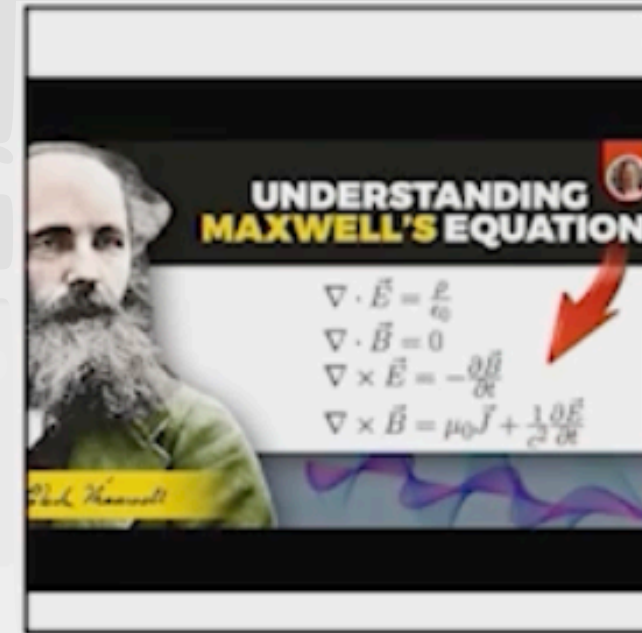
Empiricism



Pisa Tower
~1589
Free falling

伽利略、哥白尼及同时代的开普勒创建的实验观察模式，以经验主义和人的深度思考为主

Theory



United E&M
~1861
Maxwell's equation

以牛顿微积分和经典力学为代表的模型推演和理论精准预测，研究方法以演绎法为主

Computation



ENIAC
1946
First HPC

20世纪中计算机诞生，借此对复杂现象模拟仿真可推演出大量复杂现象，该范式通过可行的理论搜集数据，通过仿真发现未知的结论，

eScience



Human Genome
2003
Data Deluge

通过海量数据的收集代替人类传统的经验观察过程，借助机器的高算力代替人类的归纳推理，从而实现远超经验范式的理论归纳能力。

AI4Science



Deep Learning
Present
AI acceleration

通过“机器猜想”的应用于“科学智能”，通过不同的“算法思维”和“应用场景”的对撞，得到不同领域的“专业知识”，推导出未知结论

第一范式：经验科学

第二范式：理论科学

第三范式：计算科学

第四范式：数据密集型科学

第五范式：AI4S

GEN·AI造就第五范式

- 主要特征：模型驱动的自主研究，从辅助流程到辅助决策（真正融入过程的情报服务）；
- 规模范围：全量数据、全域知识、全程智能；数据不再是过去的数据；
- 智能体：更高效且更智能的数据收集、模拟、处理和分析；
- 多模态：各种形态的知识交流和多模态跨学科研究与融合；
- 随时随地随问题：深度挖掘数据潜在信息知识智慧；
- 行为方式：虚拟实验、自动实验，自动流程并优化验证。

AI4DH：数字人文研究的第五范式

- 数据的深度挖掘和分析
- 人机交互体验得到提升
- 跨学科协作与知识整合
- 数据可视化与多维理解
- 研究自动化与能力增强
- 更需注重伦理道德影响

智慧图书馆建设

- 没有大模型应用的图书馆不是智慧图书馆
- 不了解大模型的图书馆员很可能被淘汰

智慧图书馆大模型创新与应用白皮书
(征求意见稿 Ver0.9)



上海图书馆(上海科学技术情报研究所)
上海人工智能研究院
智慧图书馆技术应用联盟（云瀚联盟）

2023 年 9 月

· 图书馆领域大模型创新应用需求
调研报告
(征求意见稿 V0.2)

上海图书馆（上海科学技术情报研究所）
智慧图书馆技术应用联盟（筹）
2024 年 4 月 30 日

应用三部曲

- 正在普及：翻译、摘要、讨论、答疑、写作、归纳、整理、总结、朗读、转译...
 - 迎接挑战：UGC/AIGC、版权、界限模糊
 - 创新产业：PKE、二创，新生态
-

保障可信的/负责任AI

1. 信息过载与信息茧房

2. 虚假信息

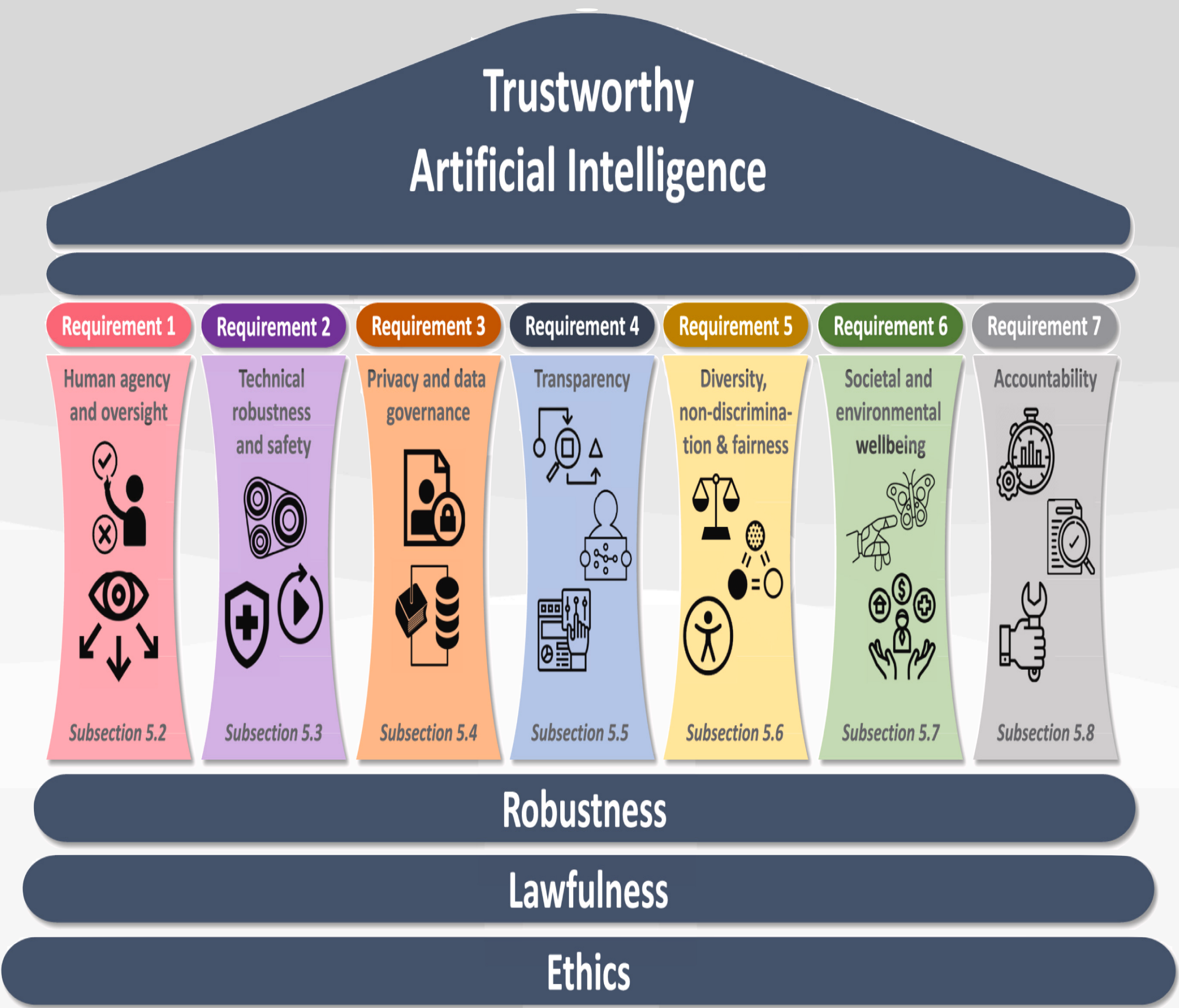
3. 误用滥用与责任边界

4. 技术素养与失业问题
5. 侵犯隐私与信息泄漏

6. 侵犯版权与诱导犯罪

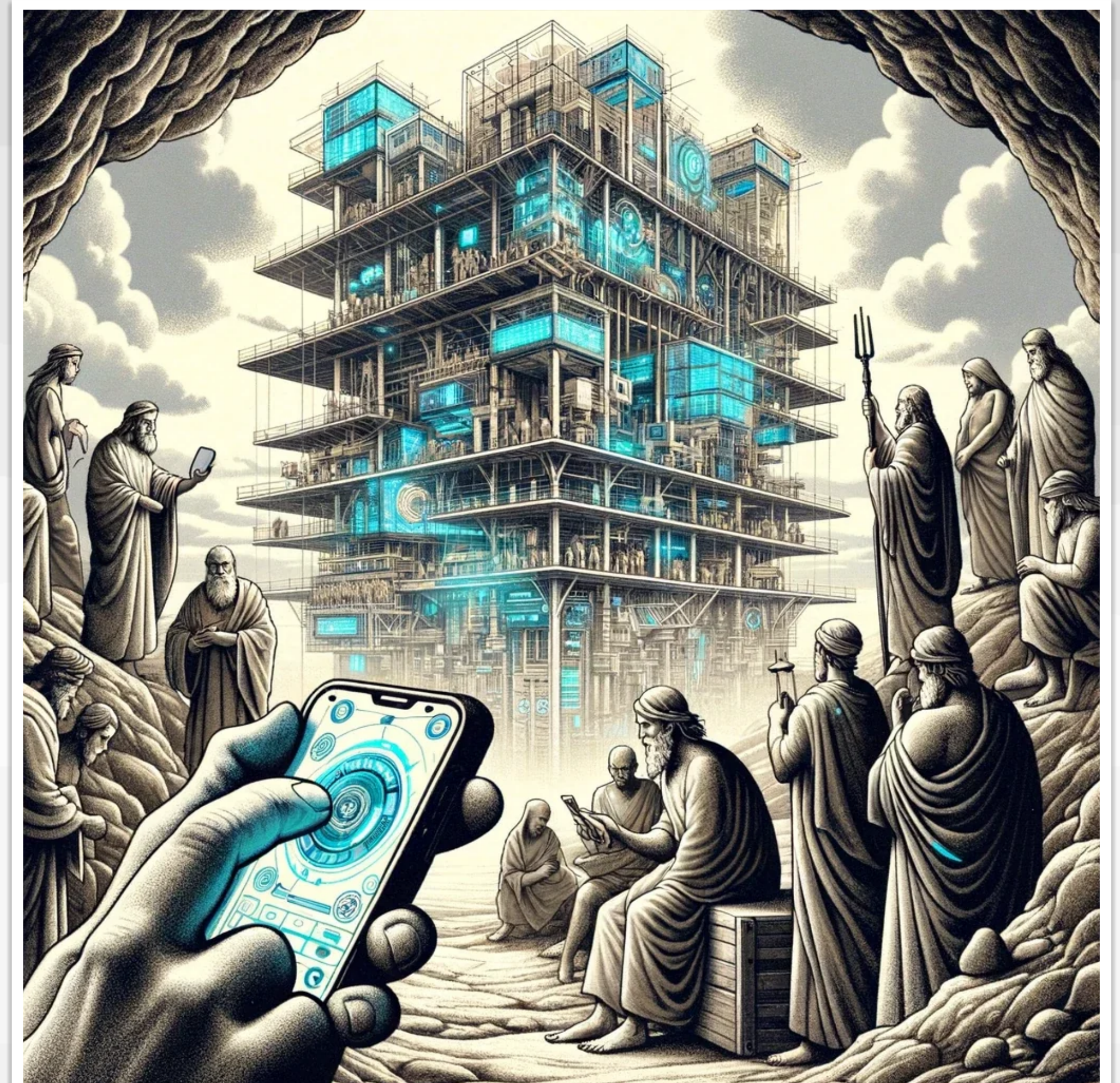
7. 军事应用与生物威胁

8. 意识觉醒与情感欺骗



捍卫人文主义

图书馆是黑暗森林中的灯塔，
图书馆员是AI时代的领航员。



2024上海图书馆开放数据竞赛巡讲·中山大学

谢谢!



刘炜 上海图书馆上海科技情报所

wliu@libnet.sh.cn