

2024上海图书馆开放数据竞赛巡讲·厦门大学

计算伦理与数字道德

——人工智能时代的人文主义



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·厦门大学

从LLM到AGI

——图书馆人的机遇与挑战



刘炜 上海图书馆上海科技情报所

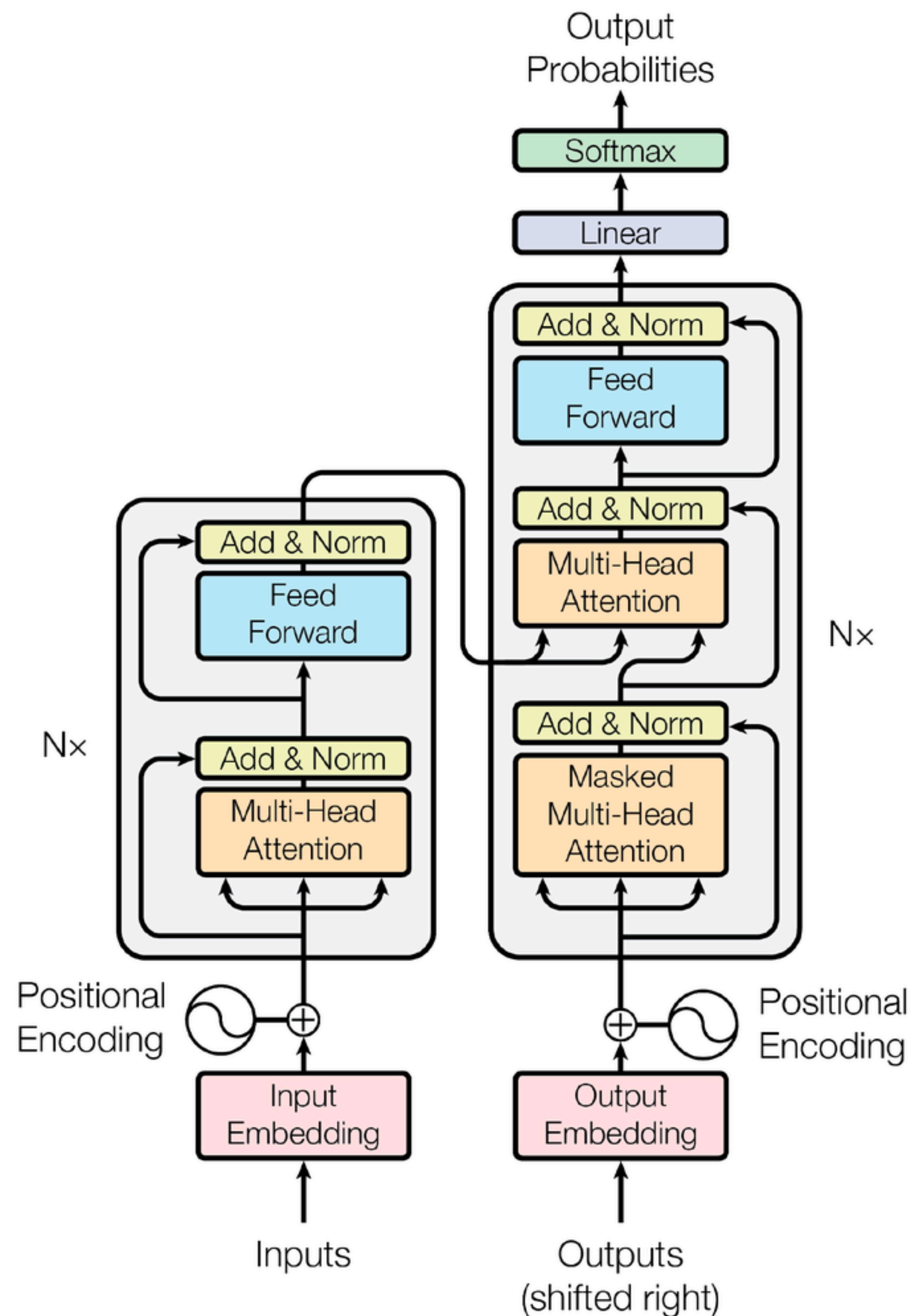
wliu@libnet.sh.cn

AGI之梦

- AGI（通用人工智能：Artificial General Intelligence）是指一种具备人类智能水平的机器，能够理解、学习和应用各种不同领域的知识。与当前专注于解决特定问题（如图像识别、自然语言处理或棋类游戏）的人工智能不同，AGI能够执行“任何”智能任务，包括任何领域的推理、规划、学习、交流，以及感知和与物理世界互动的能力。
- AGI四要素：1.具有通用智能，能够跨领域学习和工作；2.能够自我学习和适应环境；3.能够理解复杂/抽象概念并进行逻辑推理；4.甚至具有情感和意识！
- 目前的大语言模型还不是AGI，但在语言理解和多模态方面初步具备了泛化/涌现和推理能力，是目前最有可能发展成AGI的模型技术。但通过堆数据和堆算力是否能到达AGI还有争论，但普遍认为目前还未边际效应递减，但数据和电能都几近枯竭。

LLM如何被点化？

- 大型语言模型（LLM）是基于海量自然语言数据进行预训练而得到的超大型深度学习模型，参数通常从数十亿到超千亿。底层基于Transformer深度神经网络，由具有自注意力功能的编码器和解码器组成，但GPT只采用了解码器，从一系列文本中提取含义，并能够理解其中的单词和短语之间的关系。
- 用同样方法对海量图片、音频、视频等多媒体信息结合语言数据进行预训练和指令微调的超大型深度学习模型也是大语言模型的一种发展，通常称为多模态大模型。



大模型所具备的AGI特征

■ “泛化”能力

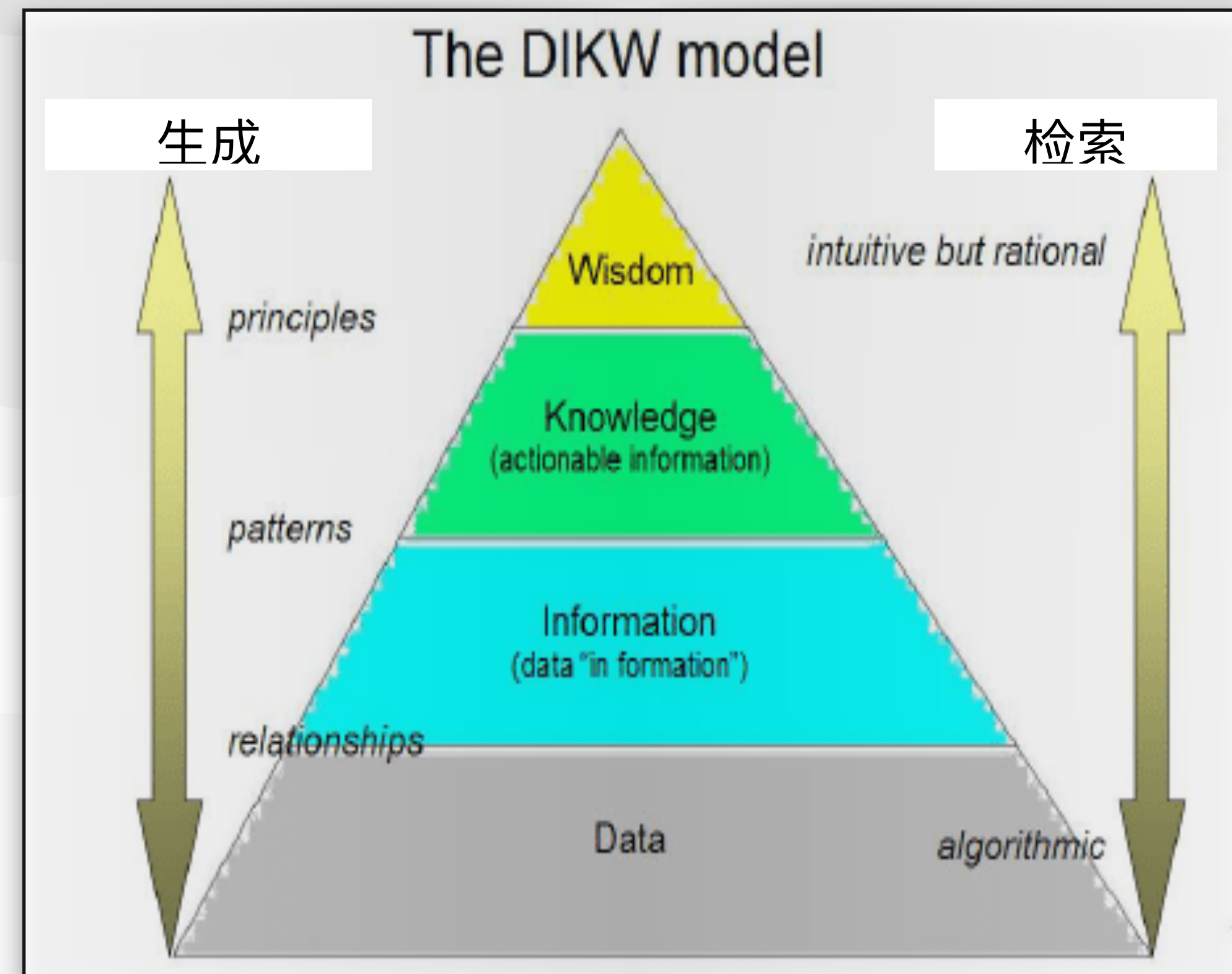
- Perplexity（困惑度）评价：用于评估语言模型在未见过的数据上的预测能力。困惑度越低表示模型在未见过的数据上表现越好。
- 语言模型的交叉验证：将数据集分为训练集、验证集和测试集，通过在验证集和测试集上的性能来评估模型的泛化能力。
- 零样本任务（Zero-shot Task）能力：在模型未见过的任务上进行评估，例如对模型提出一些与训练数据不相关的问题，评估其在这些任务上的表现能力。

■ 推理能力

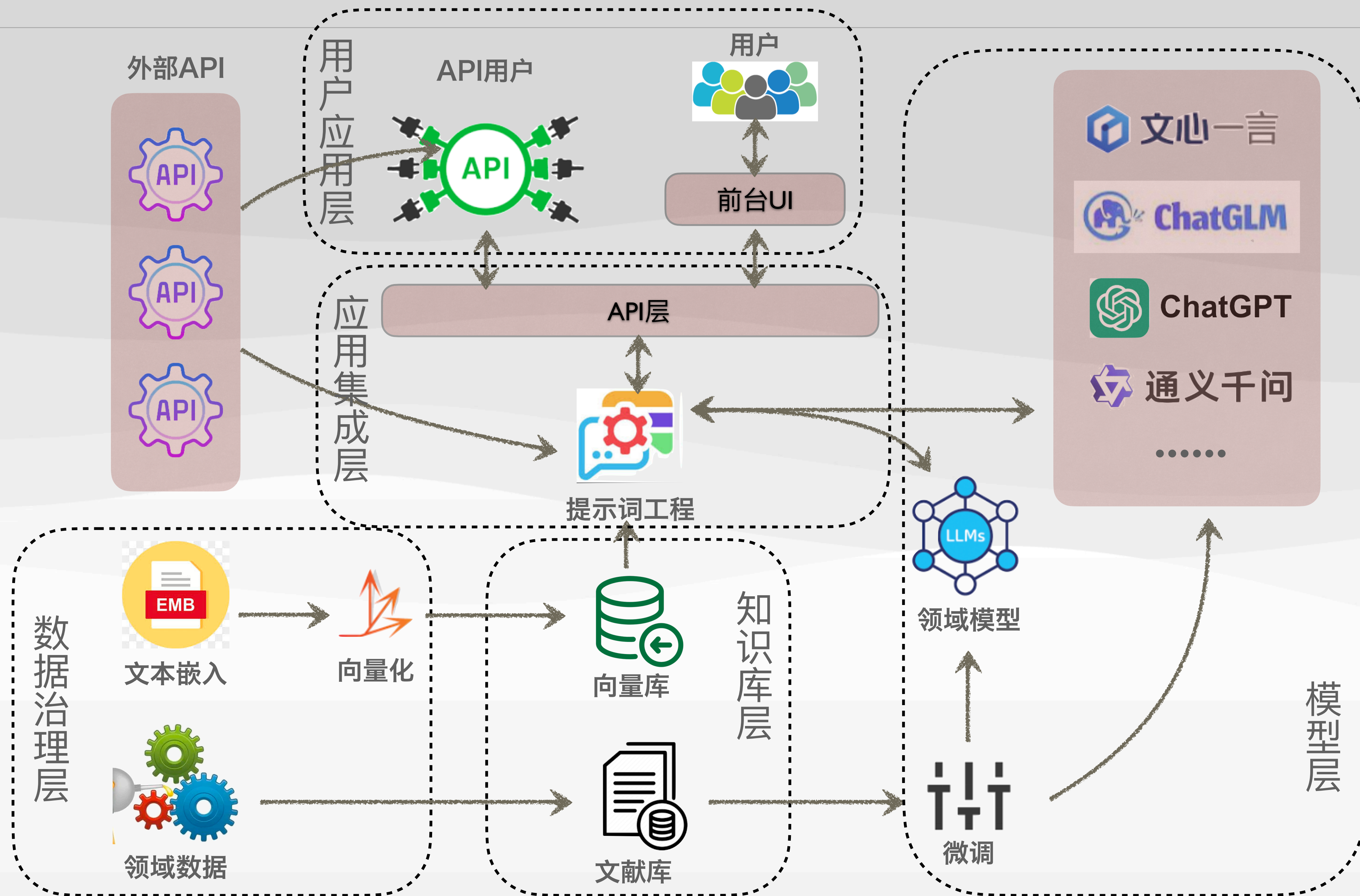
- 自然语言推理（NLI）任务：在给定前提和假设下，能否正确推断出假设的真假。
- 文本蕴含任务：在给定前提和假设下，能否判断假设是否可以从前提中推导出。
- 逻辑填空任务：要求模型填写一些语句中的空白，使得整个语句逻辑上合理。
- 逻辑推理任务：要求模型根据一些逻辑规则进行推理，例如判断一些命题是否成立或给出逻辑结论。

大模型的颠覆性

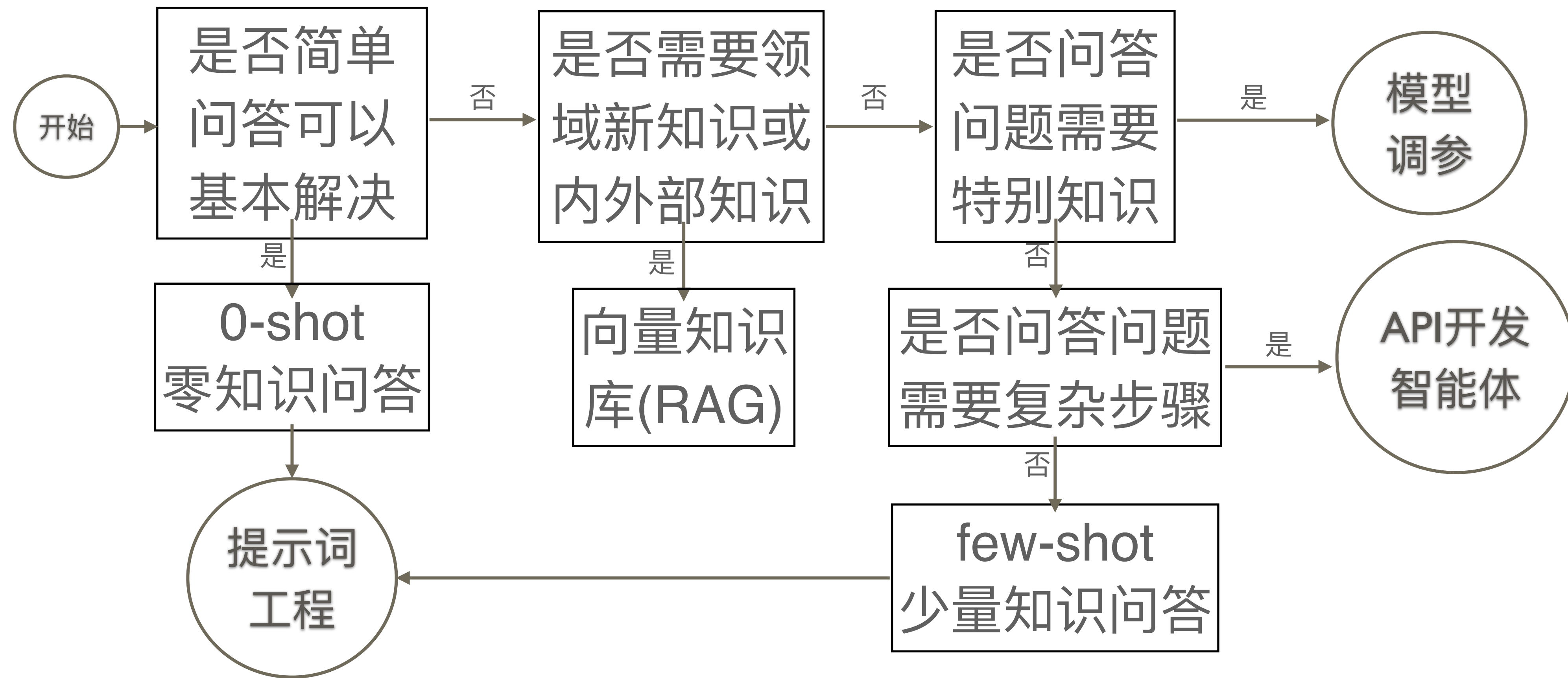
- LLM是海量知识在深度神经网络中的压缩形态，可以认为是智慧的一种编码存储形式。
- LLM是基于词元（token）而不是符号的：词元是一种张量，是语义相似性的度量而不是符号匹配，因此它可以直接告诉答案而不是符号的排列组合。
- 知识是随时生成的而无需预先存储的：存储知识只是AI的Bootloader，具有具身学习能力的智能体无需存储知识，只需要参数权重存储的智能即可以随时产生知识。
- LLM应用以端到端模型为最高形态，端到端是指只要给定数据就能得到智慧。
- 人类作为知识链的起点，其知识生产虽然节能，但却是极其原始而粗糙的。LLM一旦形成便不再需要人类的帮助，可以通过自我学习（自己创造数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



大模型应用框架



大模型应用开发策略



perplexity

新建帖子

Q 首页

发现

图书馆

Is there any new...

尝试 Pro
升级以获得图片上传、更智能的AI等更多 Pro Search 功能。
了解更多

kevenshlib8526

下载

提问，得到答案

尽管问...

焦点 附加

Pro

Apple's better Siri rumors

US Air Force AI-controlled fighter jets

New music video generated in Sora

Jack Dorsey leaves Bluesky

Will Thanos win Kratos in a battle

Warren Buffett fears AI

?

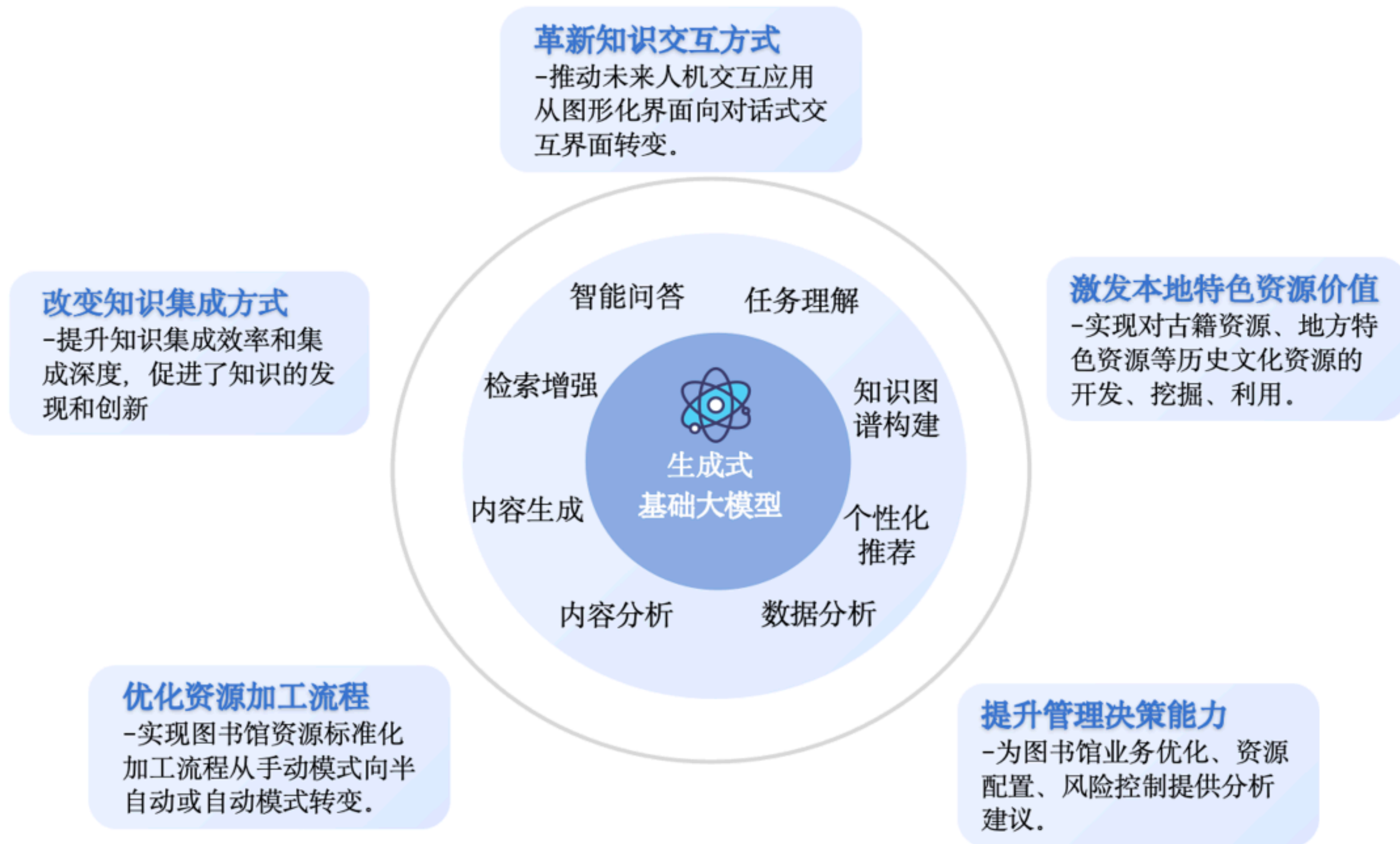


图 3.1 大模型技术对图书馆的影响

表 3.2.1 来自 IFLA：人工智能对图书馆运营的影响³⁹

人工智能应用	影响
利用 AI 使馆藏资料可读并实现规模化描述	馆藏团队、特藏团队、档案团队
利用 AI 增强或创建元数据	元数据团队
发现/检索、文献综述	图书馆系统、联络小组
数据科学家社区支持	联络小组
生成 AI 文本和图像	推广团队
图书馆或机构聊天机器人	用户服务
后端系统中的 AI，例如 RPA（机器人流程自动化）	图书馆系统
利用机器人向用户提供信息	用户服务
智慧空间	设施团队
利用机器人整理书架	馆藏团队
支持学生使用 AI 工具	学术服务
用户 AI 素养（包括数据素养、算法素养）需求	培训团队
分析和预测用户行为	策划团队

表 3.2.2 大模型核心技能及其作用影响

大模型核心技能	作用与影响	影响领域举例
智能问答	可提供自动咨询、技术支持，接入虚拟数字馆员、馆员知识库、图书馆服务平台（LSP）。	参考咨询、数字馆员服务、AI 机器人服务、学习培训、员工知识库等
检索增强	可探索对话式发现，改变图书馆资源检索、资源推荐模式。	检索发现、资源发现等
内容生成	可自动生成摘要、综述、报告、推文、辅助学术研究、总结发现结果、进行咨询回复；可进行艺术风格识别与创造，生成创意图片视频等，提升馆员工作效率。	学术研究、阅读推广等
内容分析	能够对图书馆资源进行文本/图片/语音识别、自动元数据生成、标注分类。	采编、数字资源加工与开发、数字人文研究等
数据分析	可处理数据、格式转换、报表分析、指标分析、数据挖掘。在图书馆数据系统、数据中台基础上，构建 AI 数据分析能力。	图书馆业务数据分析、用户数据分析、各类决策等
个性化推荐	根据主题、可进行用户推荐、辅助智慧采访。	用户资源推荐、采访推荐、阅读推广等
知识图谱构建	赋能数字人文研究与服务、知识学术服务。	数字人文服务、学术学科服务等
任务理解	辅助完成自动化任务，图书馆可以应用这一技能自动化日常规范化的运营任务。	图书馆服务平台、后端系统 AI、机器流程自动化等

表 3.3 图书馆大模型应用策略

应用策略	预期效果	数据需求	技术依赖性	实现难度
(1) 无需开发集成的 AI 服务	提升用户 AI 素养	无需特定数据	低；启动成本低，无需任何编程、数据操作	易
(2) 直接集成应用的 AI 产品工具	增强用户体验	第三方数据接口	低-中；启动成本低，直接购买服务	较易
(3) 需整合开发的 AI 产品与服务	优化服务流程	运营数据、服务数据、资源数据	中-高；需要数据适配和接口开发	中
(4) 集成定制开发的后端 AI 流程	提升工作效率	运营数据、服务数据、资源数据	中-高；包括系统定制、数据处理和模型训练	较难
(5) 自主/联合开发的 AI 服务产品	创新 AI 服务产品	大量馆藏数据、用户行为数据	高；涉及复杂的数据处理、软件开发	难
(6) 参与研发基础大模型产品	共建基础大模型	大规模文献资源、专业语料库	高；需要深度的数据准备、技术研发	难



图 3.3 图书馆大模型应用的六种策略

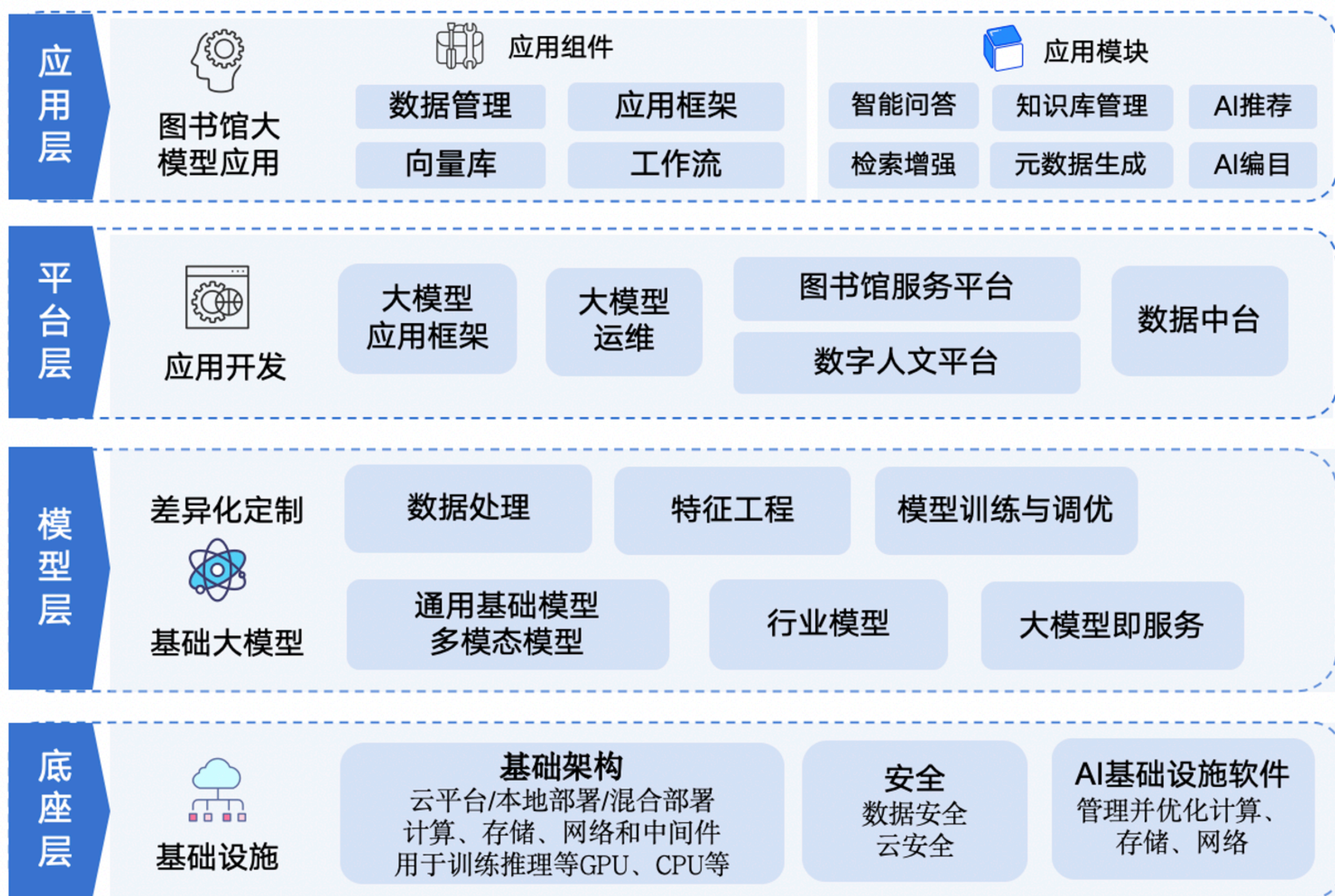


图 3.4 图书馆大模型应用架构



图 4.0.2 智慧图书馆中的大模型应用



图 5.0 图书馆大模型典型应用场景

AI五级素养体系

- 1. 会使用AI应用的基本功能，例如提示词组合，发挥其特殊能力，区别于传统应用
 - 2. 能根据自己的需求寻找合适的AI应用，并会有意识地使用提示词框架，以及AI应用的不同前端版本
 - 3. 会将不同的AI应用组合起来，实现一定的工作流，完成自己的日常工作或任务
 - 4. 会设置大模型环境，并按目标准备和加工数据，安装调试自己的AI应用
 - 5. 懂得目前大模型应用开发的堆栈框架，了解目前的不足，并对多模态和智能体等最新发展有所了解甚至提出和尝试探索路径
-

保障可信的/负责任AI

1. 信息过载与信息茧房

2. 虚假信息

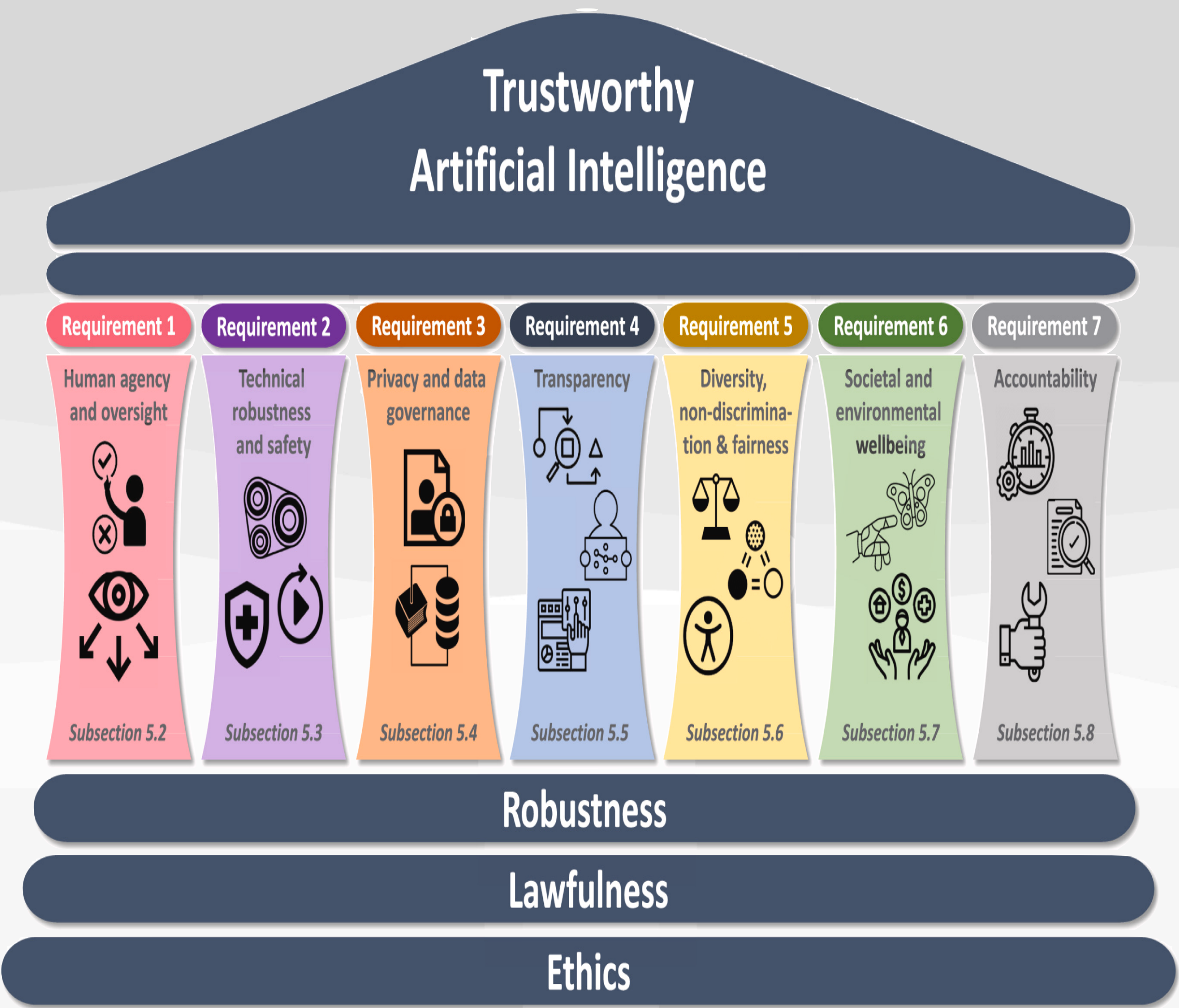
3. 误用滥用与责任边界

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

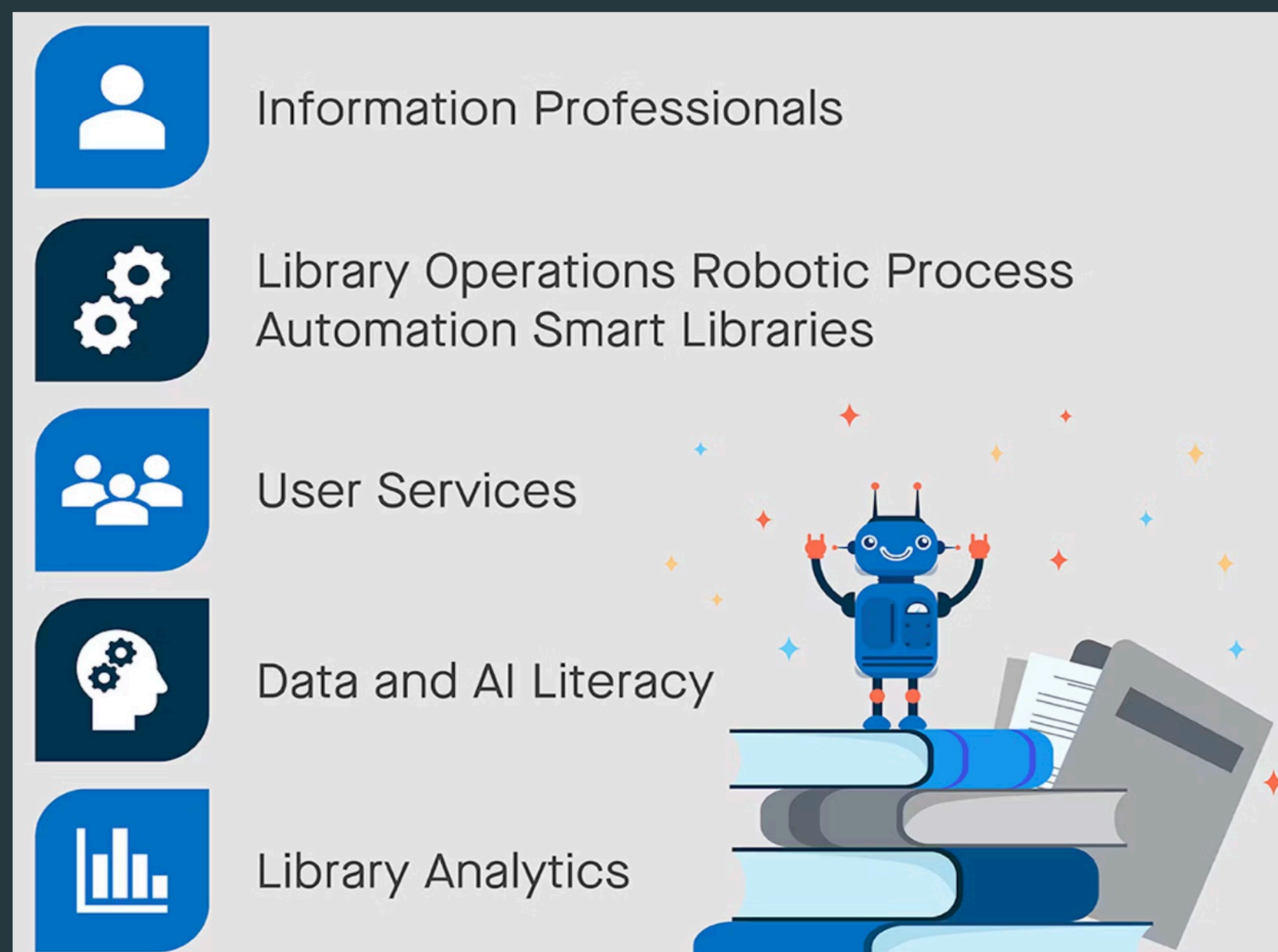
7. 军事应用与生物威胁

8. 意识觉醒与情感欺骗



人工智能给我们带来无尽的可能性

无尽的可能性既包括正面的对人类社会的福祉，也包括与之随行的风险和危机



<https://www.aje.com/arc/ways-artificial-intelligence-impacts-libraries/>

Library Assistant, Clerical -- ~95% amenable to automation

Library Technician -- ~99% at risk

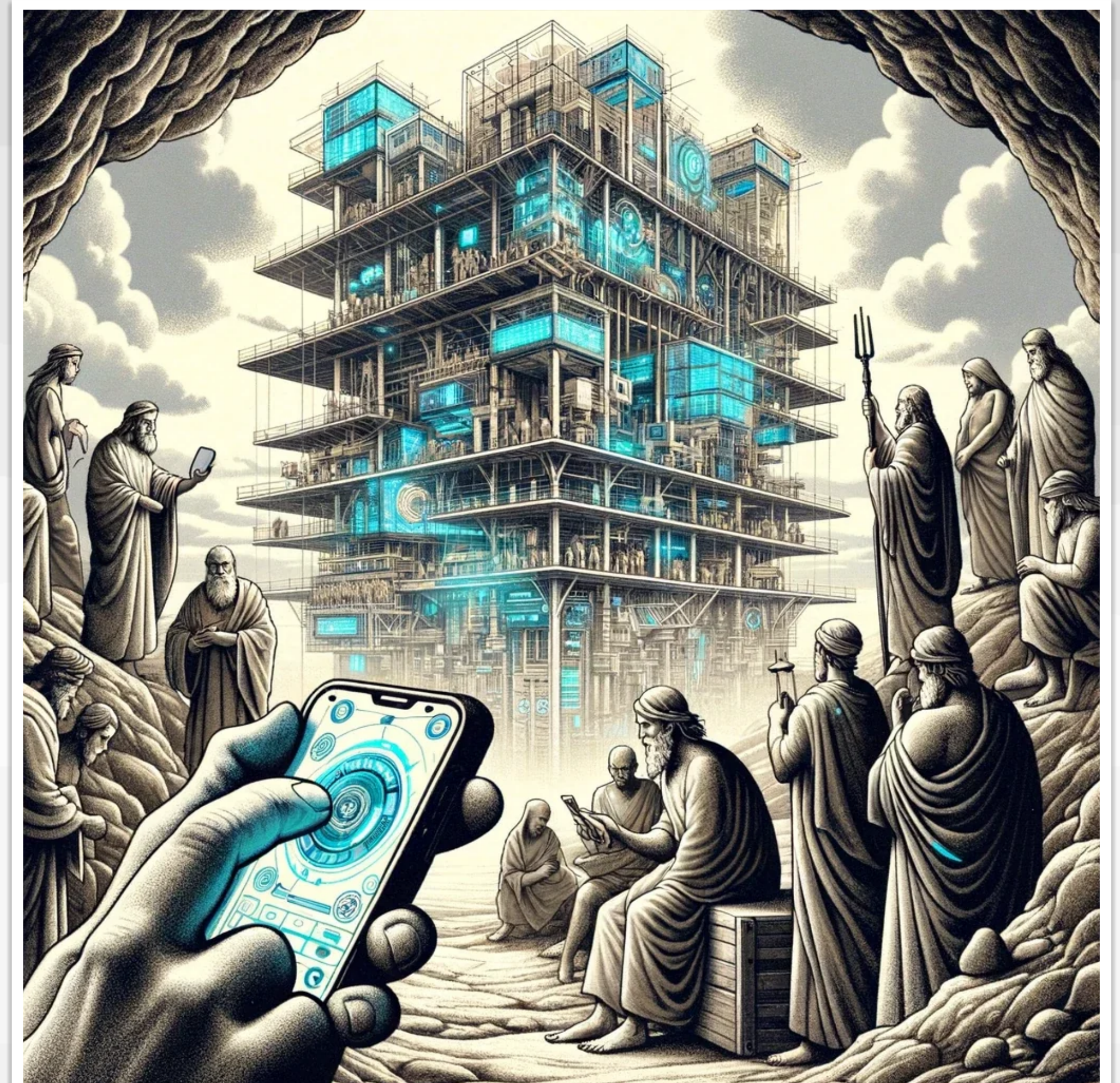
Librarian – 65% likely to be automated

(Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114, 254–280.

<https://doi.org/10.1016/j.techfore.2016.08.019>)

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·厦门大学

谢谢!



刘炜 上海图书馆上海科技情报所

kevenlw@gmail.com