



向 AI 而行 共筑新质生产力

——行业大模型调研报告



腾讯研究院
Tencent
Research Institute

出品

2024年5月

专家推荐

人工智能大模型正在催生新一轮技术创新与产业变革，也将为工业、金融、广电等行业数字化转型和高质量发展带来新动能。当前市场以基础大模型为主，通识能力强，但缺少行业专业知识。如何将大模型融入千行百业，是下一阶段的发展重点。

很高兴地看到，腾讯研究院自 2023 年 8 月就启动了此方面探索，开展了覆盖十余个行业、上百位专家的访谈，在行业大模型概念厘清、应用进展、实现方式、安全治理、未来趋势等方面进行了体系化研究，为业界提供了有益借鉴。

大模型上云落地，能够加速形成新质生产力，推动数字经济的提质换档。期待与更多同仁共同探讨、共同努力。

——中国工程院院士、原副院长 邬贺铨

工业是国家之本、经济之基。“AI+ 制造”能够提高生产效率、提升产品质量、实现个性化定制、优化资源配置、促进绿色环保、实现人才结构优化等，进而加快推进中国工业的五个转型升级，实现工业的数字转型与智能化升级，促进新质生产力形成，使我国由制造大国向制造强国大步迈进。

腾讯的行业大模型报告，通过一手的大量调研，阐释了行业大模型的必要性和发展路径。报告从应用场景入手，给出了落地的具体实现方式，揭示了大模型在降本提效、业务创新和体验增强方面的价值，并给出了 Agent、多模态、端侧和云智一体等技术和生态展望，对行业具有很好的借鉴参考价值。

——中国工程院院士 李伯虎

AI 大模型的发展正在深刻改变着各行各业的格局。然而，随着大模型能力的持续增强，确保其真正为人类社会创造价值而非带来威胁，成为当前 AI 发展的重要课题。

行业大模型作为 AI 向善和可持续发展的重要维度，其安全、可信、负责任的发展日益离不开有效的治理和安全措施。在这方面，这份报告提出的见解为我们构建负责任的人工智能生态提供了重要参考。

希望业内外同仁能够在此基础上继续深入讨论，以达成更广泛的共识，共同推进人工智能的可持续发展之路。

——香港中文大学（深圳）教授、前海国际事务研究院院长 郑永年

广告是市场经济的传感器、晴雨表和风向标。在智能化浪潮的推动下，广告行业正迎来前所未有的变革。大模型在广告创意和生产、个性化推荐、效果分析等方面，已展现出巨大潜力。它正在重塑着广告行业的未来，为品牌与消费者之间建立更加精准和深入的链接提供了可能。

大模型是效率、体验和创造力的倍增器，腾讯的这份报告为我们展现了大模型在各行业应用的广阔前景、主要挑战和落地路径。相信在“人工智能+”战略的推动下，各行业将催生大量产品形态、生产力、生产方式、产业组织等多层面的创新，实现更高质量的发展。

——中国广告协会会长、国际广告协会全球副主席 张国华

大模型的出现使得 AI 再次受到资本市场的广泛关注，其核心逻辑在于大模型实现提质增效创造价值的同时，也提供了实现通用人工智能的可能选择。在算力、算法、数据、场景等四个决定大模型发展的关键要素中，数据和场景是我们相对优势的领域。而要更好地利用场景，成功实现商业落地，高质量的行业大模型必不可少。

高质量的行业大模型，必须具备专业化、低成本、高可靠解决行业痛点问题的能力。同时，基于云侧和端侧模型的完美结合，实现较高的执行效率。目前，大模型大规模的行业应用尚在探索，成体系的研究不多。腾讯研究院的报告基于大量一线访谈和调研，进行了全面系统的分析梳理，提供了理性、全面、有价值的研究参考。作为投资人，我很乐于并期待看到更多类似的高质量研究，以此推动大模型向通用人工智能发展，实现人类走向智能社会的美好愿景。

——中国互联网投资基金管理有限公司总经理 李筱强

在大模型应用上，Hugging Face 与 OpenAI 走向了不同的方向。OpenAI 正在搭建人工智能的封闭系统，Hugging Face 是希望每个人都可以访问生成式 AI 模型，包括各类企业和所有普通开发者。我们坚信，开源大模型会成为新的选项，开源社区也将大有可为。腾讯的行业大模型报告，与 Hugging Face 的理念不谋而合，大量的一手调研数据表示，我们不管是从底层能力还是应用场景方面，都应该引导一种百花齐放的局面。一方面，大公司在算力、数据、资金等方面拥有天然的竞争优势，但是场景理解不足；另一方面，因为通用大模型对算力和数据量的高需求，许多企业无力承担。我特别推荐不管是技术供给还是需求方的人士，都可以通过阅读这份报告，了解供需两股力量的匹配情况，共同推进生成式 AI 模型的能力开放，让普通开发者，也能够通过不同的技术组合，将生成式 AI 应用到产品开发和工作流程中。

—— Hugging Face 高级工程师 王铁震

这几年通用大模型技术快速发展，展现出越来越强的智能，但很多传统行业应用推进得其实并不快，因为企业有大量的具体问题解决不好。对企业而言，大模型应用需要综合考虑行业专业性、数据安全、持续迭代和综合成本等多种因素。针对这些现实情况，我去年就提出要重点发展行业大模型。基于行业大模型，构建自己的专属模型，也许是企业更优的选项，尤其大部分企业的算力、数据等资源有限。当时支持行业大模型的人并不多，但到今年我们看到更多的企业已经开始尝试并取得进展。

我们研究院做的这份报告基于一线的大量实践反馈，做了系统深入的归纳总结，让大家能比较充分了解行业大模型发展的真实情况，厘清其中的关键争议和困惑问题，比如行业大模型与通用大模型的关系等。我们始终认为，人工智能发展的根本目标是落地于行业、服务于人。能真正解决用户需求、距离场景和数据更近的企业，将拥有大模型的未来。

——腾讯集团高级执行副总裁、云与智慧产业事业群 CEO 汤道生

应用场景是大模型的“磨刀石”。腾讯拥抱大模型的一个重要方式，就是积极推动腾讯自己的各类业务与自研的混元大模型结合，同时在文化消费、生产提效、科技普惠等多个领域探索落地场景，发掘大模型的真实应用价值。这也是腾讯多年来形成的务实精神的体现，把自己做试验田，内部打磨成熟再向外推广应用，通过“由内而外”的路线给用户提供更好用的产品。

随着腾讯大模型的不断成熟，我们将会连接和服务更多的行业应用场景。哪些行业和场景需求强，哪些应用实现方式更适合，如何衡量应用的价值和成效……这些都是需要我们持续思考和解答的问题。这份行业大模型研究，提供了一份不错的预研参考。未来，腾讯将继续专注于自主可控的大模型技术发展路线，不断迭代技术，积极落地新应用，助力“人工智能+”稳步迈进。

——腾讯集团副总裁 蒋杰

序

我们正处在一个智力变革的路口。

很少有一项技术像大模型这样，全球热度持续了快两年，还在不断涌现新热点。这反映出大模型带来了巨大的可能性，我自己有一个判断叫“智力平替”。从文生文、文生图到文生视频，大模型不断解锁新能力，逐步替代传统的智力劳动，使得原本需要专业团队长时间才能完成的工作，现在仅需一个简单的提示或草图即可实现。结果就是，人类的智力成果可以云端的方式、通过大模型提供给每个人，一个人带领 AI 团队就能完成复杂的任务，这可能会成为未来世界的常态。

然而，在大模型技术持续快速迭代的同时，我们也越来越发现一个现实问题：大模型在具体行业中的应用，推进好像没那么快。很多企业在大模型是一边期待、一边焦虑，想尽快用起来，又不太知道怎么用，即使是有试用的，效果也不明显，存在不少困惑和挑战。这反映出，目前大模型技术的能力与行业应用需求间还存在差距，要用一个通用的大模型来直接满足各行各业的特定需求，至少现在还有一定距离。那么，应该怎么办？

行业大模型应运而生。坦白说，一开始我有疑虑，觉得这是一个伪概念，业界对此也有不少讨论和争议。后来我们团队坚持要进行研究、用事实说话，我也支持。在我看来，大模型发展到今天，研究已经不能只停留在眺望未来的视角，更需要贴近实践，用深度的研究推动创新探索和落地。这份行业大模型调研报告，就这样最终成型了。

报告做了大量的一线实践调研，访谈了超过百位的各界专家，系统归纳解答了一系列共性问题，包括为什么要发展行业大模型、行业大模型是什么、哪些行业和场景发展快、如何衡量成功、如何负责任地应用等。通过这项研究，我认为行业大模型，至少现阶段是客观存在的，通常基于通用大模型调整而来，是很多企业提高模型专业性的有效选择，也是国家落地“人工智能+”的最后一公里。而决定大模型影响大小的核心因素，就在“知识密度”，密度越高的行业或场景，受影响会越大，比如教育行业、研发和设计环节等等。

当然，这项研究只是起了个头，希望能对大家有所启发和帮助。也要特别感谢所有给予我们指导、接受我们访谈的各界专家和朋友们，希望未来我们继续携手合作，推动大模型深入各行各业，一起创造和见证一个以人为本的人机共生时代的来临。

——腾讯研究院院长 司晓

目录

CONTENTS

发展背景	01
一 大模型引发的智能革命	02
参数规模大	03
泛化能力强	04
支持多模态	05
二 发展行业大模型的必要性	05
大模型存在“不可能三角”问题	05
行业对大模型有内生需求	07
行业大模型是落地“人工智能+”的最后一公里	07
三 行业大模型的特点	09
行业大模型既有模型、也含应用	09
行业大模型大多生长于通用大模型之上	09
行业大模型的本质是解决方案	11

应用进展

一	行业大模型应用阶段划分	13
二	行业大模型应用场景分析	14
	不同行业的大模型应用快慢有别，需求和数据是两个关键因素	17
	大模型在垂直场景的应用渗透，呈现“微笑曲线”特征	18
	不同行业对大模型能力有“三大共性需求”	20
三	终极之问：行业大模型怎样算成功？	22
	避免两个误区：片面追求技术性能或短期收益	23
	评估三类价值：降本提效、业务创新和体验增强	24
	构建一个模式：高质量数据飞轮	37

实现方式

一	引导：提示工程	40
二	外挂：检索增强生成	42
三	优化：精调	45
四	原生：预训练	48
五	典型案例：多种方式组合	49
	商业案例：金融大模型 + 个股异动写稿	49
	科研案例：scBERT 单细胞基因数据分析大模型	52

安全与治理 54

- 一 行业大模型治理的三项原则 55
- 二 价值对齐是行业大模型可信应用的基础保障 56
- 三 行业大模型更需重视全生命周期的安全保障机制 57
- 四 AI 沙盒在行业中有较成熟探索，可成为实现监管目标的可行方案 59
- 五 合成数据有望为行业大模型增加数据来源 60
- 六 行业大模型是走向 AI 绿色可持续发展的路径之一 61

未来展望 62

- 一 "人工智能+" 促进行业大模型应用提速 63
- 二 多模态大模型为数实融合打开新空间 64
- 三 AI Agent 有望为各行业注入生产力 65
- 四 行业大模型将出现更多的端侧形态 67
- 五 云智能一体支持行业大模型加速落地 68

编写团队 70



发展背景



大模型引发的智能革命

"日益强大的人工智能系统将从根本上改变科技产品的工作方式、创造新的类别并重组整个行业。从头开始是好的。"

"Increasingly powerful AI systems will fundamentally change how tech products work, create new categories, and restructure entire industries. starting with a cleanslate is good."

——山姆·奥特曼 (sam Altman), openAI首席执行官

2022年11月30日，openAI发布大语言模型（LLM, Large Language Model）chatGPT，用户数5天过百万、2个月过亿的成绩，一度刷新纪录成为史上增长最快的应用。chatGPT的全民爆火揭开了人工智能（AI, Artificial Intelligence)大模型时代的序幕，也预示着AI迈向通用人工智能（AGI, Artificial General Intelligence)的新一轮冲刺。

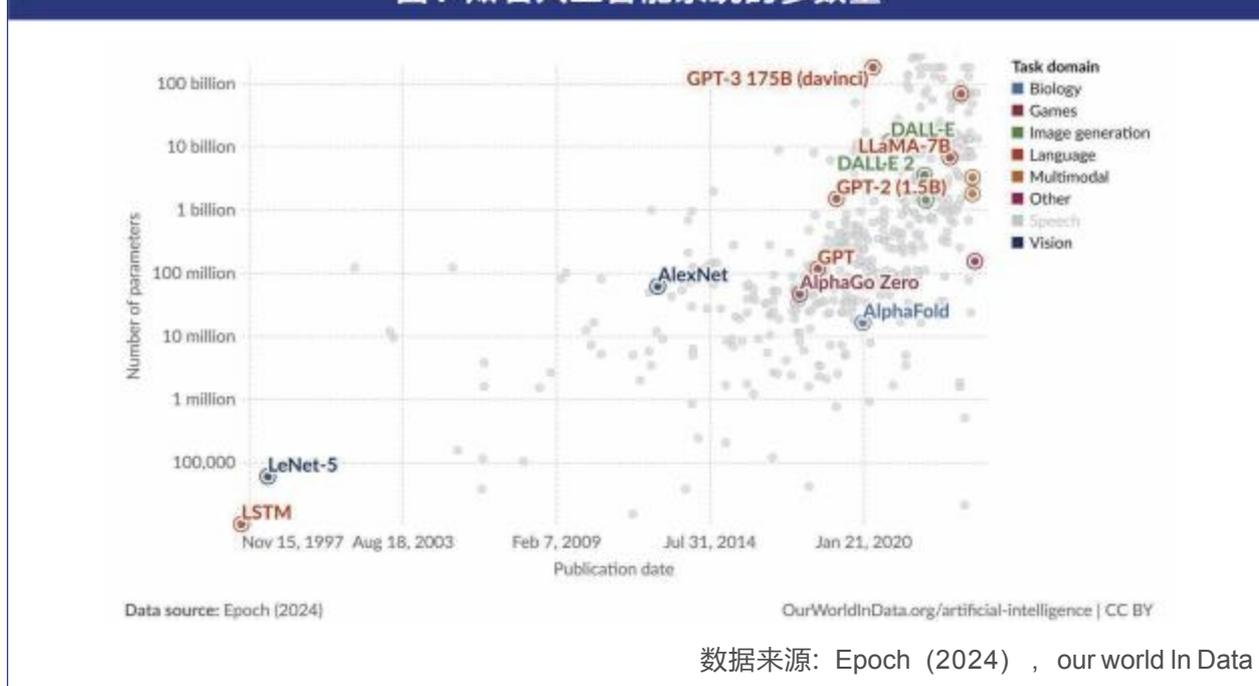
目前业界对大模型并没有形成明确统一的定义，狭义上可指代大语言模型，基于Transformer技术框架；广义上则包含了语言、声音、图像、视频等多模态大模型，技术框架也涵盖stable Diffusion等。学界如李飞飞等人工智能学者指出，这些模型可被称为基础模型（FM, Foundation Model）。

在大模型出现之前,人工智能通常需要针对特定的任务和场景设计专门的算法模型,能够执行的也是训练数据范围内的单一任务。大模型的突破,关键在于展现出了类人的通用智能"涌现"能力,能够学习多个领域知识、处理多种任务,因此也被称为通用大模型。具体而言,大模型具备以下特点:

1. 参数规模大

大模型的参数规模远大于传统深度学习模型。大模型发展呈现"规模定律"(scaling Law)特征,即:模型的性能与模型的规模、数据集大小和训练用的计算量之间存在幂律关系,性能会随着这三个因素的指数增加而线性提高,通俗而言就是"大力出奇迹"。不过"大"并没有一个绝对的标准,而是一个相对概念。传统模型参数量通常在数万至数亿之间,大模型的参数量则至少在亿级,并已发展到过万亿级的规模。如OpenAI的GPT-1到GPT-3,参数量从1.1亿大幅拉升到1750亿,GPT-4非官方估计约达1.8万亿。

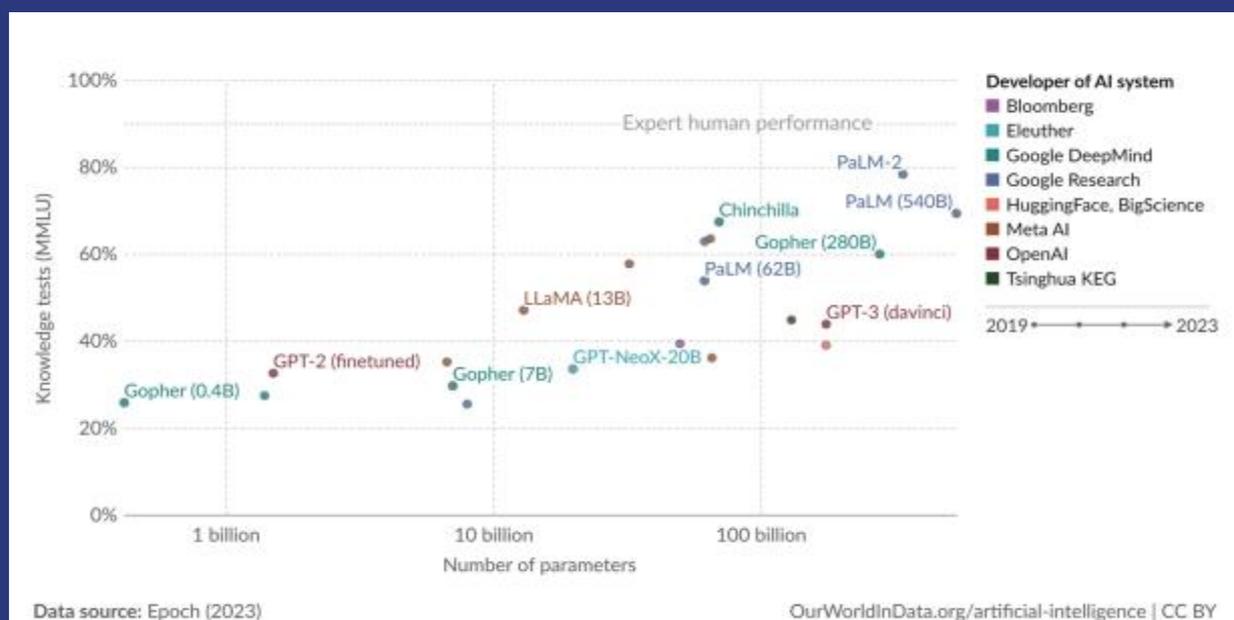
图：知名人工智能系统的参数量



2. 泛化能力强

大模型能够有效处理多种未见过的数据或新任务。基于注意力机制 (Attention) ，通过在大规模、多样化的无标注数据集上进行预训练，大模型能够学习掌握丰富的通用知识和方法,从而在广泛的场景和任务中使用，例如文本生成、自然语言理解、翻译、数学推导、逻辑推理和多轮对话等。大模型不需要、或者仅需少量特定任务的数据样本，即可显著提高在新任务上的表现能力。如openAI曾用 GPT-4参加了多种人类基准考试，结果显示其在多项考试中成绩都超过了大部分人类（80%以上），包括法学、经济学、历史、数学、阅读和写作等。

图：代表性语言模型参数量与性能（MMLU 评估结果）对比

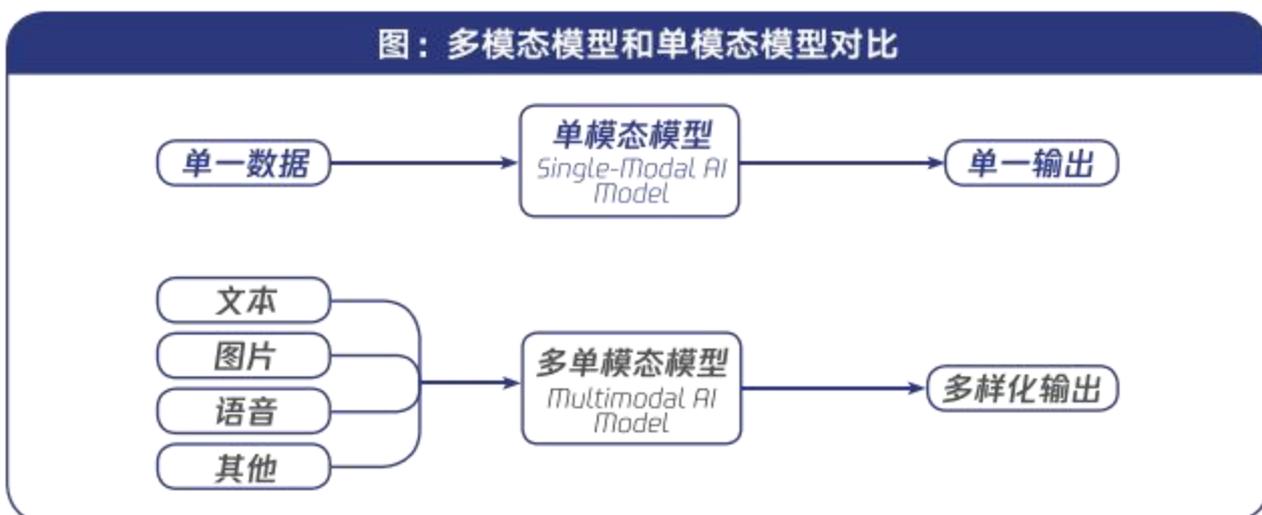


数据来源: Epoch (2023) , our world In Data

注: MMLU (Massive Multi task Language understanding) ，大规模多任务语言理解,是一项用于评估大模型在各种语言理解任务上能力的综合测评，涵盖历史、文学、科学、数学等广泛主题和领域

3. 支持多模态

大模型可以实现多种模态数据的高效处理。传统深度学习模型大多只能处理单一数据类型（文本、语音或图像），大模型则可以通过扩展编 / 解码器、交叉注意力（Cross-Attention）、迁移学习（Transfer learning）等方式，实现跨模态数据的关联理解、检索和生成。多模态大模型（LMMs, Large Multimodal Models）能够提供更加全面的认知能力和丰富的交互体验，拓宽AI处理复杂任务的应用范围，成为业界探索迈向通用人工智能的重要路径之一。典型如OpenAI的Sora模型推出，掀起了全球多模态大模型的发展新热潮。



发展行业大模型的必要性

1. 大模型存在“不可能三角”问题

Scaling Law 驱动通用大模型性能不断提升，同时也产生了“不可能三角”的问题：

大模型在专业性、泛化性和经济性三方面很难兼得。

专业性

大模型处理特定领域问题或任务的准确性与效率。大模型专业性要求越高，越需要针对特定领域数据进行训练，这可能造成模型过拟合而降低泛化能力。此外，增加的数据收集和训练也会增加成本、降低经济性。

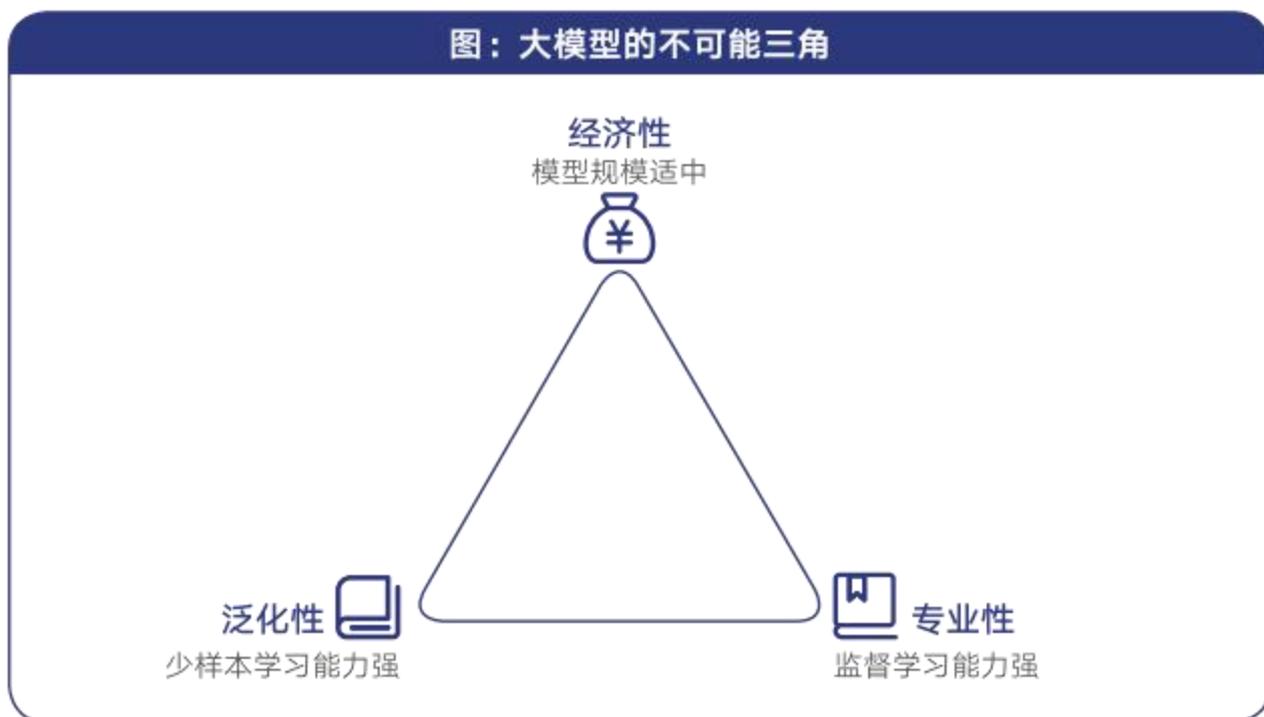
泛化性

大模型处理训练数据集之外新样本的表现能力。大模型泛化性要求越高，越需要多样化的大规模训练数据集、模型参数量也越多，这意味着模型训练和使用成本的增加、经济性降低，同时可能降低模型对特定问题处理的专业能力。

经济性

大模型训练和应用的投入产出比。大模型经济性要求越高，越需要消耗更少的算力

图：大模型的不可能三角



资源与成本满足性能需要，然而降低资源消耗基本上需要采用更小的模型或更少的参数，这又会降低模型的性能表现。

通用大模型以发展通识能力为主要目标，更侧重泛化性，在专业性和经济性方面很难充分满足具体行业/机构的特定需求，存在“有幻觉、成本高”等情况。

2. 行业对大模型有内生需求

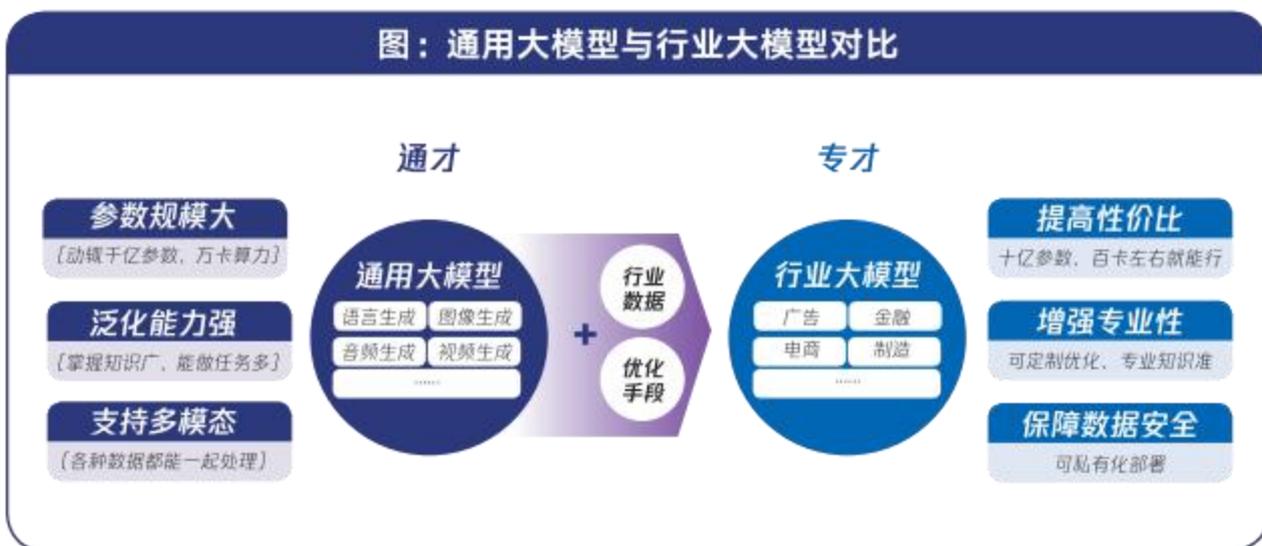
具体行业中机构从自身需求出发，采用大模型还有两个关键考量因素：竞争和安全。

将数据转化为竞争力是核心驱动：为有效提升竞争力，机构会尽力寻找性能最佳的模型，并利用行业专业数据或私有数据的独特资源，对模型进行定制调整和优化。目前GPT-4等市场头部通用大模型多为闭源，通过网页、APP应用服务大众用户，或以API标准化调用的方式服务开发者，可定制化调整的空间较小。

保障安全可控是底线要求：大模型不仅涉及机构私有数据的调用，还会与机构的业务、流程结合（如 copilot方式），这使得大模型使用越深、越需要重视安全可控。通用大模型通常基于公有云提供服务，会造成机构对私有数据、敏感数据安全的担忧。例如，2023年三星电子引入chat GPT不到20天，就引发代码上传、会议记录等三起内部资料泄露事件。此外，通用大模型本身的数据集、参数等对外多是黑盒，机构无法掌握技术细节，也会影响对模型的信任度。

3. 行业大模型是落地“人工智能+”的最后一公里

如上可见，通用大模型与行业/机构具体需求间存在差距，行业大模型成为弥合技术与需求间差距的必然产物，能有效支持各行各业加速落地大模型应用：



一是高性价比，行业大模型能在较小参数量模型的基础上，通过相对低成本的再训练或精调，达到较好的性能效果。十亿~百亿级参数量的行业大模型是目前主流选择，相比通用大模型动辄千亿级以上的参数量能明显节省开发成本；

二是可专业定制，行业大模型可基于开源模型开发，能对模型结构、参数等按需调整，更好地适配个性化的应用需要。通过模型即服务（MaaS, Model-as-a-Service）方式，机构可以从平台对接的多种模型中快速选择合适的使用，包括厂商已开发的行业大模型初始版本；

三是数据安全可控，行业大模型可采用私有化部署方式，机构能更放心地利用私有数据提升应用效果，减少数据安全的疑虑。

“通用大模型缺乏细分领域的专业知识，所以需要训练行业大模型，然后紧密结合业务系统实现可落地的智能应用。而且很多企业的私有数据是不允许对外泄露的，比如金融、医疗的客户信息和交互行为，

所以还需要提供大模型的私有化训练和本地化部署。 ”

——王超辉,北京仁科互动网络技术有限公司销售易智能产品负责人



行业大模型的特点

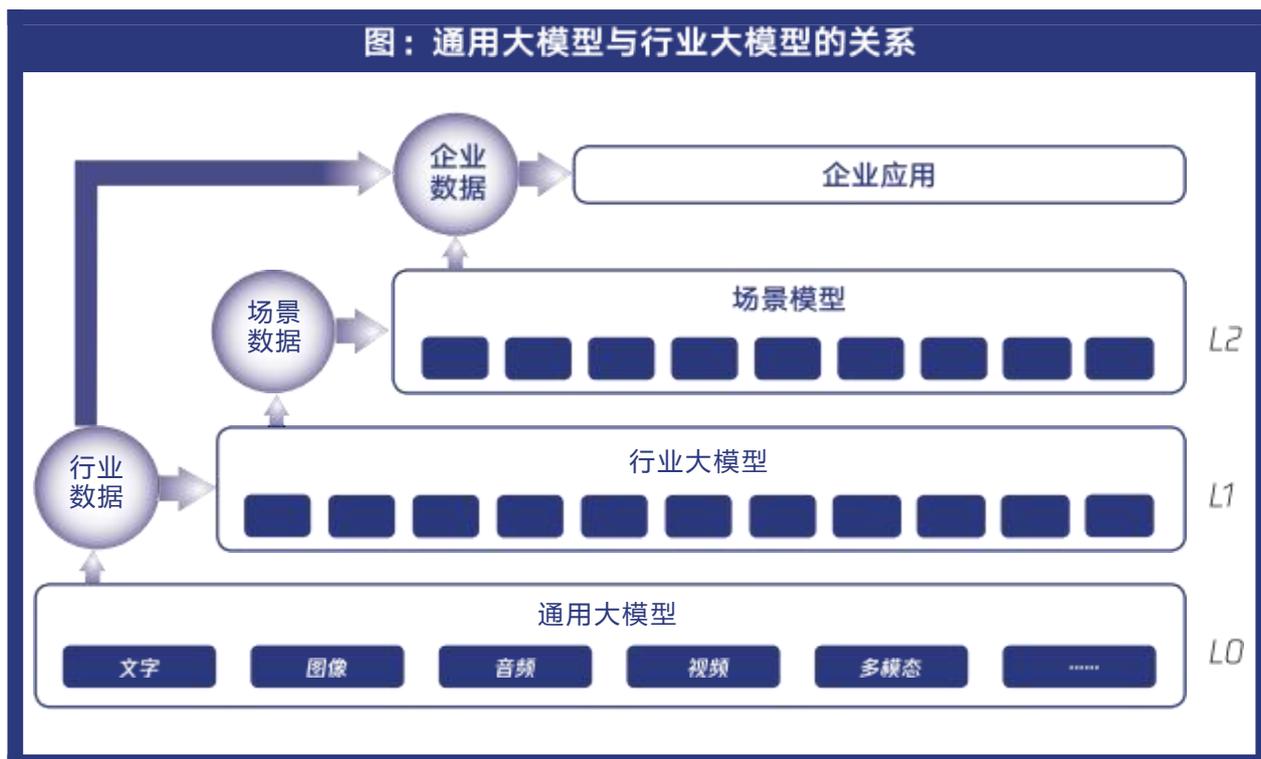
1.行业大模型既有模型, 也含应用

行业大模型是与通用大模型相对的概念。通用大模型侧重发展通识能力, 行业大模型则侧重发展专业能力。从行业实践看, 行业大模型不仅指开发一个行业专用的模型本身, 更多还包括基于通用大模型调整和开发的行业应用。因此, 广义上行业大模型可以归纳为: 利用大模型技术, 针对特定数据和任务进行训练或优化, 形成具备专用知识与能力的大模型及应用。此外, 国际上更多用垂直模型 (vertical Model) 或垂直人工智能 (vertical AI) 来表示, 国内还有用垂类模型、领域模型、专属模型等称谓, 都可以认为与行业大模型等同或包含在内。

2.行业大模型大多生长于通用大模型之上

行业大模型大多在通用大模型基础上构建。通用大模型具备丰富的知识和强大的泛化能力, 不仅能为行业大模型提供广泛的知识基础并提升交互体验, 还能显著节约从头训练模型所需的大量数据和算力资源, 大幅提升行业大模型开发及应用的效率和效果。通过对通用大模型进行提示工程、检索增强生成、精调、继续预训练/后训练等方式, 模型能够更好地处理特定数据或任务, 从而生成行业大模型的版本

(模型有变)或具备行业大模型的功能(模型不变),具体实现方式见报告第三章。今天市场上很多行业大模型,如金融、法律、教育、传媒、文旅等,大多是在Llama、SD、GLM、Baichuan等国内外主流开源大模型基础上构建。



"模型需要进一步的分级,如L0-L1-L2等从基础模型、垂直模型到某个领域专家,来逐步实现专业性的提升。所谓的垂直领域模型,更多是一个特定训练数据的加入,比如占到10-15%的预训练占比,从而输出行业大模型。"

——国内某大模型独角兽技术负责人

3. 行业大模型的本质是解决方案

行业大模型通常需要针对特定数据和任务进行定制开发或调整。行业大模型面向的是 B 端客户，每个客户都有独特的业务、数据、流程等，需要用大模型解决的具体问题也普遍存在个性化需求。因此，厂商提供的行业大模型不仅是产品和工具，更需要有定制服务与支持，甚至需要客户参与共建。可以这样理解，行业大模型中的产品通常是“毛坯房”，客户需要根据自身用途进行“装修”才能满足需要。例如，腾讯云 2023 年中就依托 TI 平台打造了行业大模型精选商店，推出了 MaaS 服务，为客户提供一站式的行业大模型解决方案，包括模型预训练、模型精调、智能应用开发等。

图：腾讯云 MaaS 服务体系



“我们认为，客户更需要有行业针对性的行业大模型，再加上企业自己的数据做训练或精调，才能打造出实用性高的智能服务。企业所需要的是在实际场景中真正解决了某个问题。”

——汤道生，腾讯集团高级执行副总裁、云与智慧产业事业群 CEO



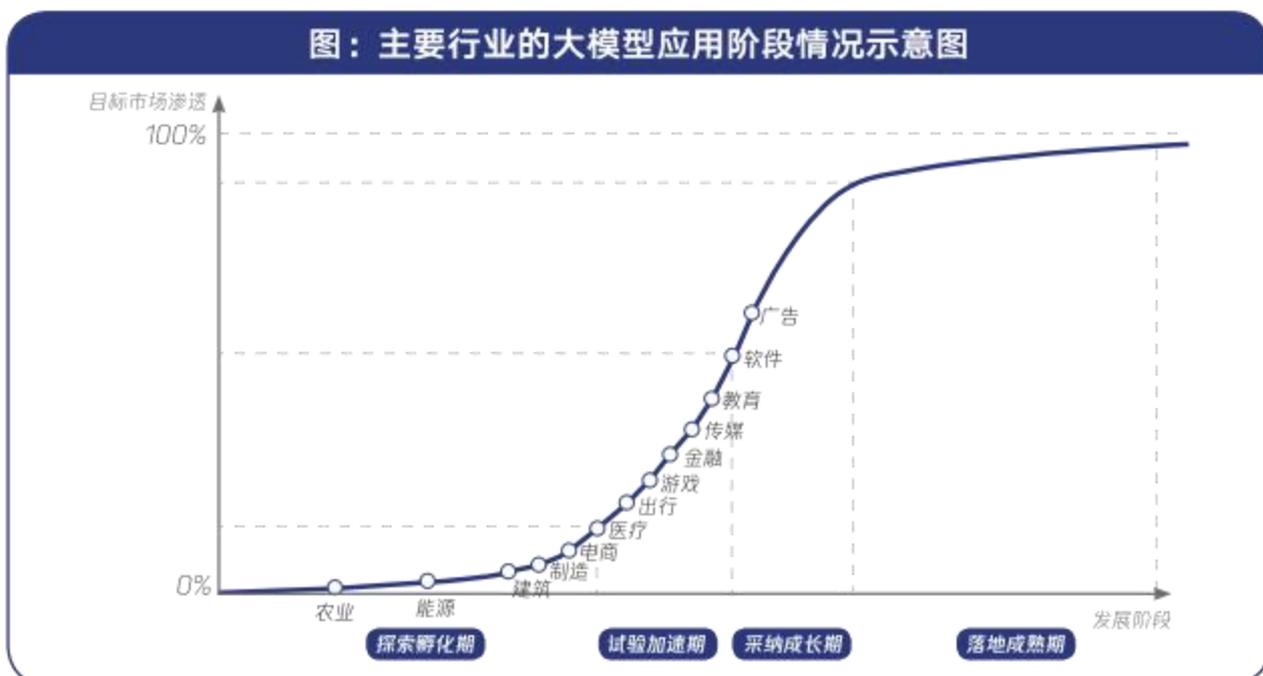
应用进展

在调研和横向比较各行业应用大模型技术的过程中，我们注意到不同行业的技术落地进度各异。这种差异主要由大模型技术的成熟度、行业的数字化水平、投入产出比、行业对专业性和准确性的要求以及安全可控等因素决定。

行业大模型应用阶段划分

参考埃弗雷特·罗杰斯(Everett M. Rogers)《创新的扩散》(Diffusion of Innovations)一书中对创新阶段的界定，我们从技术发展和市场渗透两个维度构建视图，结合一线调研数据，综合评估、定位各行业在 2024 年初的位置，以便对不同行业采纳大模型的进程做比较。结果显示：**目前行业在大模型技术的采用上主要集中在两个阶段：探索孵化期与试验加速期。**部分行业已经步入采纳成长期，尚未有行业达到落地成熟期。

图：主要行业的大模型应用阶段情况示意图



注 1：表中数据由专家综合打分得出，为 2023 年 12 月 31 日截面数据，所属阶段将随着时间迁移发展变化。

注 2：目标市场渗透指各行业从业人员，在工作过程中常态化应用大模型技术，占总体行业人员的百分比。

常态化应用：指任一工作流程月均使用工作天数在 50% 以上。

注 3：软件泛指以软件（包括 APP）为主要产品和服务的行业，包括互联网服务商和信息技术服务商。

阶段一： 探索孵化期

农业和能源等行业为代表。行业中尝试采用大模型的机构数量还较少，但仍有一些头部或创新意识强的机构积极探索。机构推进市场应用的关键在于证明技术的可行性和实用性，并能解决行业特有的挑战，面临较高风险和不确定性，同时有机会引领市场。如电力行业，国家电网公司正在积极尝试，但由于安全性等关键指标要求高，推进较谨慎、速度相对慢。

阶段二： 试验加速期

教育、金融、游戏与出行为代表。行业普遍具备相对良好的数据基础，探索应用大模型的机构数量快速增长，开始在特定应用场景产生经济价值。机构关注技术如何解决实际问题，如金融量化策略的赢率、游戏设计的降本增效等。成功案例是此阶段的风向标，实用效益能够吸引更多参与者加入。

阶段三： 采纳成长期

广告与软件行业为代表。行业中的主流机构，已普遍采纳并使用大模型。由于与大模型基础能力的高度匹配，目前在广告以及软件行业〔包括各类互联网应用〕，文案生成、文生图、代码生成与数据分析等能力，已经在不少机构被大量使用。继续扩大市场的关键在于进一步优化技术应用，提升用户体验和效率，同时降低成本。

阶段四： 落地成熟期

目前还没有行业达成。此阶段意味着大模型技术应用基本成熟，绝大多数机构已在主要生产运营场景中使用，并与供应商之间建立了稳定的商业合作关系。目前大模型技术本身还远未到成熟阶段，行业应用的成熟需要更长的时间。大模型的稳定性、可解释性、插件调用的可靠性等等，都是行业应用能步入成熟期的必要前提。



行业大模型应用场景分析

调研发现，多个行业都已经开始探索大模型技术在各生产环节的应用落地，具体涉及研发/设计、生产/制造、市场/销售、客户服务、经营管理等环节，典型场景总结如下：

表：行业大模型典型应用场景调研总结

第一产业：农林牧渔业

行业	研发/设计	生产/制造	市场/销售	客户服务	经营管理
农业	-生物育种	-智慧种植/养殖决策辅助	-农产品电商平台销售对话	-农业技术服务智能对话助手	-农业市场信息分析辅助

第二产业：制造业/建筑业/电力/能源

行业	研发/设计	生产/制造	市场/销售	客户服务	经营管理
建筑	-户型和总图设计生成 -意向图生成等 -开店选址助手 -技术资料对话	-建筑工程技术资料对话辅助施工	-住房销售广告素材生成	-物业智能客服/数字社区	-企业助手 -经营数据问答 -工单自动分类 -设备设施管理
制造	-工业产品外观设计	-工业代码生成 -工业机器人控制 -车间和设备管理	-市场需求预测和库存管理 -个性化营销 -产品营销内容生成	-产品智能客服 -机器人业务处理	-管理软件辅助 -工业知识问答和企业助手
能源	-新能源场景规划辅助	-智能巡检 -生产环节调度辅助 -知识图谱辅助运营	-售电公司智能销售	-售电与电网服务助手	-负荷预测及电力质量分析

第三产业：生产性服务业

行业	研发/设计	市场/销售	客户服务	经营管理
广告	-广告创意生成 -设计图生成	-广告精准投放	-广告主自助内容生成平台	-客户触达数据分析
金融	-风险管理和信贷评价 -资料汇总和投资分析	-理财产品销售 -差异化的广告投放	-金融智能客服	-数仓的自动调度 -营销合规监管 -银行催收合规监管
软件	-软件创意设计生成 -辅助芯片设计 - copl lot编程辅助	-销售线索自动评价	-应用智能客服	-数据中台智能分析

续上表：行业大模型典型应用场景调研总结

第三产业：消费者服务业

行业	研发/设计	市场/销售	客户服务	经营管理
电商	- 广告创意设计生成 - Copilot 编程辅助	- AI 驱动的数字人辅助销售 - 明星数字分身	- 电商智能客服	- 商家经营数据分析
传媒	- 采访提纲辅助 - 分镜头脚本生成 - 媒体内容生成与翻译 - AI 驱动数字人播报	- 广告投放素材生成	- 辅助用户运营 - 内容智能审核 - 账号安全	- 传媒运营数据分析
游戏	- 美术创意设计生成 - 大模型驱动的 NPC 对话 - 多语种翻译 - Copilot 编程辅助	- 产品宣传素材生成	- 社区活动运营支持 - 社区智能客服 - 陪伴玩家的虚拟人	- 数据中台智能分析
医疗	- 对话式药物研发助手 - DNA 测序识别基因与疾病关联	- 产品营销素材生成	- 临床辅助诊断 - 基于最新文献提供诊断依据 - 患者日常保健智能客服	- 病例结构化与检索辅助 - 医院资源管理调度 - 患者满意度文本分析
政务	- 政务数据库	- 政务服务宣传素材生成	- 政务服务智能助手 - 热线电话预处理	- 辅助公文写作 - 辅助数据报告制作
教育	- 辅助科研论文生成 - 辅助教学内容生成	- 机构/课程营销素材生成 - 机构/课程智能销售 Bot	- 虚拟老师/教练 - 个性化学习辅助 - 心理咨询与辅助	- 辅助精准测评
出行	- 路况图像视频等分析 - 出行系统语音交互	- 产品营销素材生成	- 车辆智能助手 - API 调用车机应用的乘坐服务	- 交通运营数据分析

来源：腾讯研究院行业大模型百人百问专家访谈整理

1.不同行业的大模型应用快慢有别，需求和数据是两个关键因素

各行业横向对比，大模型已经在多个行业落地，从进展快慢角度可归为三大类：

1)数字原生行业是大模型应用的先行者

互联网、游戏等数字原生行业,由于数字化程度高、数据积累丰富、技术接受能力强,成为大模型落地较快的行业。这些行业的大模型应用场景广泛,覆盖了营销、客服、内容生成等诸多环节,目前已积累了相对丰富和成熟的实践。例如,游戏行业利用大模型进行 NPC 对话生成、游戏素材创作、玩家行为分析等,提升研发效率和游戏体验;互联网TO C领域,利用大模型优化内容生成、信息搜索、智能助手等关键应用进展较快。

2) 生产性服务业成为传统行业结合大模型的示范区

金融、广告、软件等生产性服务业,因其产品和业务的非实物属性,在客户服务和数据处理等方面有强需求,比较适配当前大模型技术的能力,因此推进也较快,为其他传统行业的大模型应用探索提供了示范。例如,软件行业积极利用大模型代码生成等能力,优化从需求分析、开发、测试到运维的全流程,降低人力成本、提高应用开发和服务质量;金融机构则利用大模型增强服务的广度和精度,实现营销、风控、投研等环节的赋能提效。这些行业的大模型实践正加速走向成熟,并向场景纵深探索。

3)重资产行业在大模型的应用上处于局部探索阶段

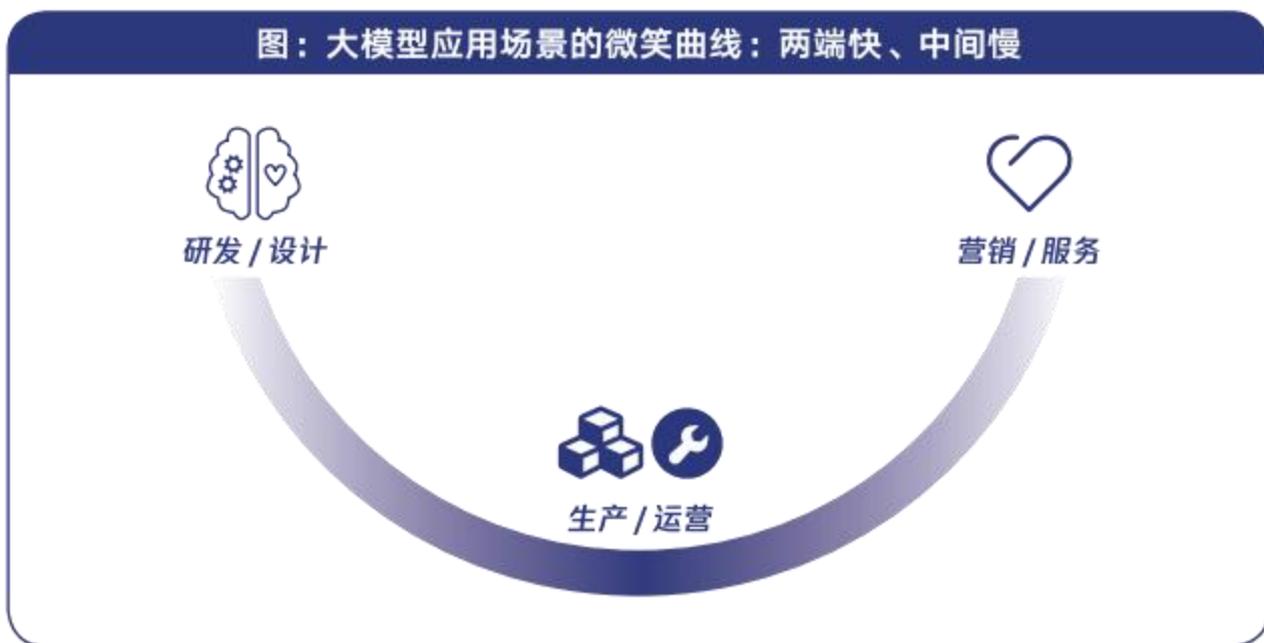
能源/ 电力、建筑、制造业等行业,大模型应用推进相对慢,主要受限于线下生产

流程的复杂性和高度的专业化。这些行业的核心在生产运营，需要在通用大模型能力的基础上，进一步深度整合行业专业知识，同时避免幻觉问题确保准确性和安全，因此需要更长期、渐进的过程。建筑行业为例，大模型在前端设计、渲染等环节已初显成效，但在建造、施工管理、安全监控等核心环节，仍需要大量行业知识的采集和专门训练；制造业则需要将大模型与工业互联网、数字孪生等基础设施及专业数据深度结合，才能在工艺优化、质量管控、设备维护等核心领域发挥更大价值。

整体而言，可以看出影响行业应用大模型速度有两个关键因素：一是数据可得性，高质量数据越容易获取、进展越快，如数字原生行业；二是需求适配度，行业的核心业务与大模型创意生成和交互能力越匹配、进展越快，如生产性服务业。

2. 大模型在垂直场景的应用渗透，呈现“微笑曲线”特征

深入行业环节纵向来看，当前大模型技术渗透呈现出类似产业微笑曲线的特征，即：在产业链高附加价值的两端（研发 / 设计和营销 / 服务），大模型应用落地较快；



而在低附加价值的中部（生产、组装等），大模型应用进程较慢。究其原因，大模型技术所带来的“智力即服务”的范式变化，特别适配微笑曲线两端、知识密集型和服务密集型领域，对人的能力提升乃至部分替代的效果显著，可以认为掀起了脑力劳动大规模工业化、自动化的新篇章。

1)营销/服务环节进展最快,跨行业通用性强是重要原因

在产业链下游的营销与服务环节，基于机构自有知识库的内容生成与智能对话，能使营销和服务活动的效率和体验得到显著提升，成为各行业尝试应用大模型的普遍先行领域。营销和服务大多直接面向 C端用户，跨行业通用性强，能够充分利用通用大模型的基础能力和通用的营销、服务知识，快速开发和调试出适配机构需要的应用。如营销方面，各行业都能基于广告大模型进行素材生成和精准投放；客服方面，各行业也能在通用的智能客服机器人快速搭载自有知识库，为用户提供专业的个性化服务。

2)研发/设计环节结合最深, 高质量专业数据集决定进程

在产业链上游的设计研发环节，大模型对海量知识的高效学习、推理和生成能力，不仅能大幅提升文案、影像、代码等内容创意的生成效率，还适用于生物、环境、材料等涉及海量科研数据处理的科学计算领域。是否具备高质量的专业数据集，决定了不同行业、领域在这个环节进程的快慢。文案、影像、代码等有大量基于互联网的开放、开源数据集，因此这些领域进展最快；有高质量、大规模开放数据集的科研领域进展也较快，典型像 Deep Mind开发的Alpha Fold,能够仅凭氨基酸预测蛋白质的 3D结构，大幅提升了蛋白质研究进程，其成功的背后有赖于采用了 PDB、uni prot等提供的开放数据集进行预训练；工业研发/设计方面，芯片、汽车等领域也已出现用大模型辅助设计生成的应用，但这些领域多涉及强商业竞争，

高质量的开放数据集很难获取,需要更多投入,实用进展相对慢。

3) 生产/制造环节进展相对慢,对人的辅助增强是目前主要结合点

处于产业链中间的生产制造环节,往往涉及对机器等各类实体的操作,需要人与设备、工艺、系统的适配,环节多、流程复杂,对安全性、准确性和稳定性要求高。目前大模型的能力主要体现在自然语言和图像的处理上,并不直接适用于生产制造环节复杂的数值计算、时序分析和实时决策等场景,往往需要针对性采集专业数据集进行专门的模型训练和开发,因此结合的进展相对慢。从目前行业实践看,大模型在生产制造领域的应用侧重在对人的辅助增强,以copilot(机器人助手)为主要形态,结合工业软件在工业仿真、生产监控、故障排查等环节辅助人提升处理能力。例如西门子与微软合作推出了siemens Industrial copilot,通过自然语言交互,维修人员可以获得详细的维修指导,工程师可以快速访问仿真工具,还能够快速生成、优化和调试复杂的自动化代码,几分钟即可完成以前耗时数周的任务,大幅提高生产力和效率。sight Machine推出的Factory copilot,让用户通过自然语言交互就能自动整理工厂运营相关的各类数据,快速生成易于理解的报告、邮件和图表,还能引导用户进行根因分析加快问题诊断。

3.不同行业对大模型能力有“三大共性需求”

虽然不同行业与大模型结合的进展和侧重点存在差异,但对大模型的当前优势和发展方向也存在共性的认识与需求,总体看有三个方面:

1) 内容生成与创意设计

主要运用大模型展现出的生成能力,包括文本生成、图像生成及代码、表格等泛文

本生成能力,结合特定行业、场景的数据,支持内容生成和创意设计。例如各行业都普遍需要营销和广告,不论是创意阶段的文案制作、还是市场阶段的广告素材生成,结合行业特定数据的广告大模型应用都能发挥更好的作用。在数字化的浪潮下,很多行业也需要增加代码开发、数据分析等工作,大模型辅助代码生成、数据分析和图表生成等,成为增长中的跨行业共性需求。

2)信息提炼与专业辅助

主要运用大模型的摘要、规划等能力,针对特定行业、场景的数据,辅助人进行专业知识的提炼、分析和加工。结合检索增强生成等技术,许多行业通过对话机器人实现此类助手型应用,覆盖研发设计、生产制造、营销服务等多环节。例如金融行业,结合金融领域专有知识形成的金融大模型,已能有效支持前后台工作,成为金融投资决策、风险评估等的关键支撑;医疗行业,科研人员可以通过对话式药物研发助手,便捷查询专业信息,助力新药研发,甚至辅助DNA测序识别疾病与基因的关联等。

3)任务调度与智能交互

行业对大模型的需求,更多还体现对其代理能力(Agent)的期待,希望大模型能与其他应用,甚至与现实世界的机器和设备等连通,在更广泛的范围协助进行任务调度和问题解决。这涉及实时数据处理、自动化控制、环境感知和决策支持等,对模型的响应速度、准确度和自适应性提出了更高要求,需要大模型插件生态、大模型与小模型的结合、多模态大模型等顺利发展。例如电力行业,希望通过大模型的任务调度和交互能力,有效整合传感器等数据,优化能源分配和消耗,提高能源领域的运行效率;交通行业,涉及复杂的现实环境交互,对安全性和准确性要求高,可能需要开发具备多模态能力的原生交通大模型才能满足需求。



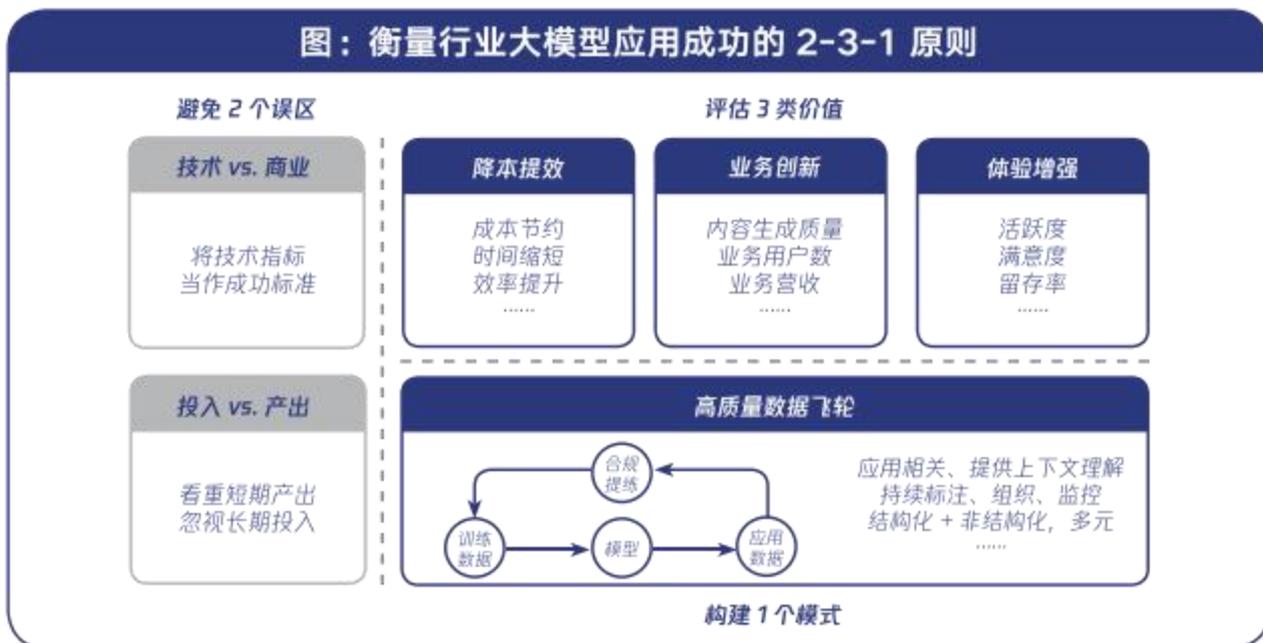
终极之问：行业大模型怎样算成功？

随着大模型的发展及在行业中应用的推进，越来越多的行业机构开始关心一个核心问题：行业大模型做成什么样才算成功？

这个问题也是当前业界普遍面临的挑战。大模型整体还处于发展早期阶段，一方面技术快速迭代蕴含了巨大的创新价值，另一方面在规模定律驱动下算力等投入还呈现指数级增长态势。许多行业机构几乎不知道从何入手，更不用说有充分的应用经验来衡量成功，然而如果缺乏衡量的标准和办法，又很难充分投入技术创新和应用，容易陷入决策困境。

基于本次研究对多方的实践调研，结合国际前沿的相关探索，我们尝试总结构建出当前衡量行业大模型应用成功的 2-3-1 原则：避免两个误区、评估三类价值、构

图：衡量行业大模型应用成功的 2-3-1 原则



建一个模式。之所以是原则，是因为对于具体的行业、机构而言，“成功”是一个相对概念、没有绝对标准。

1.避免两个误区：片面追求技术性能或短期收益

误区 1：将技术指标当作证明大模型成功的标准。

在大模型技术快速迭代的形势下，一些机构会将注意力集中在技术性能上，通过呈现指标数值的增长来反映大模型的成功，包括精度、召回率和提升率等，但这些指标远远不够。最大的问题是，技术指标的高低无法直接反映出大模型价值的大小，例如召回率从 50%提升到60%，算不算成功其实很难评判。更重要的应该是关注业务指标，例如用户数、使用量、收入等，这些指标能直观反映价值，作为评判成功的标准更具说服力。可将技术指标和业务指标建立联系，用业务发展来牵引技术开发和优化，也更容易论证技术性能提升所带来的价值。

误区 2：过度看重投资收益中的短期产出部分而忽视长期投入。

行业更关注大模型的实用性，越来越强调投入产出比，这本身没有问题。但一些机构会将大模型与成熟业务类似，要求明确的投入产出，甚至短期内实现正向盈利，这并不利于大模型应用的发展。大模型还在快速迭代，有大量的不确定性，比如技术突变、泡沫等，短期财务上很难有较好的表现，长期也很难评估准确。更合理的方式是将大模型作为研发或孵化项目，不强求短期的财务指标绝对值达成，转而关注业务、技术等指标的相对提升，采取广泛的组合投资策略长期持续投入才可能真正见效。

"人工智能投资的最佳回报率将来自广泛的人工智能投资组合，并以广泛且不断发展的投资理论为指导与战略优先事项保持一致，有助于根据业务影响分配资源。 遵循投资组合管理计划来确定大多数人工智

能用例的组织,达到人工智能实施成熟水平的可能性高出2.4倍。 "

——Gartner 《快速解答:人工智能投资的真实回报是什么?》

2.评估三类价值：降本提效、业务创新和体验增强

从国际头部企业的研究和实践经验看，目前评估大模型成功的方法整体呈现多维特征。例如，微软推出的《为人工智能成功奠定基础:领导者指南》列出了五个类别，包括商业策略、技术策略、AI策略、组织与文化、AI治理。谷歌在《新一代人工智能的 KPI》中归纳了三个主要领域,模型质量、系统质量和业务影响。

虽然各家评估方法和维度有所差异，但可以发现存在一个基本共性：**商业价值**（往往体现为业务需求及指标)是衡量大模型成功的最终指向。例如，微软认为商业策略是人工智能成功的第一步，业务需求目标主要有提高生产力、效率最大化、改善业务成果。谷歌认为企业可以从人工智能部署中获得多种价值，建议用于跟踪业务影响的指标有采用率、使用频率、会话长度、放弃率、用户满意度等，商业价值改进指标则列举了客服成本减少、营销简化时间节省、零售每次访问收入提升等示例。

"41% 的成熟人工智能组织使用与客户成功相关的业务指标。 "

——Gartner 《调查分析:人工智能优先策略带来回报增加》

从我们调研的实际情况看也类似，国内行业大模型应用实践的企业，在谈及大模型价值或成效的时候，大多也认为对商业、业务的影响是最关注的问题。整体上可以归纳为三类价值

1) 降本提效

降本提效主要是衡量行业大模型对机构本身运营管理影响，也是目前比较显性的成效表现。核心在于大模型能辅助增强人员能力，提升自动化水平进而简化流程，最终减少人力等成本、提升组织运营和管理效率。典型场景包括自动化开发、业务流程优化、智能化决策支持等，衡量成效的指标如大模型支持业务的成本节约、时间节约、效率提升、辅助决策的准确率等。

自动化开发

大模型能自动生成代码、提供编程建议，减少开发者的重复性工作，让他们更专注于复杂的系统设计和问题解决。同时，这也降低了编程门槛，使非专业人士能参与编程开发，促进技术的普及和创新。比如，大模型提供的代码示例和建议，帮助开发者避免常见错误、提高效率。大模型能理解编程语言的语法和结构，甚至项目的具体上下文，为开发者提供定制化支持。这种自动化不仅加快了开发流程，也能提高产品质量。

(··)

【广告】

"跨境卖家每年投入在商品拍摄上的费用几乎会占到整体GTV的1.5%，着实是一笔成本不小的开支，我们发现以往跨境卖家每请1个外籍模特的费用基本在2000元/小时，而用AgIG C模特商拍功能，可以将每张商拍图的成本降低

98%，同时生成时间缩短至 1分钟/张，大大降低买家营销成本，提高生产效率。”

——秦鹏，易点天下企业技术中心总经理



【软件】

“ICT行业之前很多公司都是靠外包，把需求和设计聊清楚，代码交给别人。这些原来可以交给别人去做的工作，后面可能都可以通过大模型来做。”

——茹炳晟，腾讯 Tech Lead



【游戏】

“大模型显著提升游戏行业效率，尤其在内容创作和发行方面。在生产端，stable Diffusion 等工具助力2D美术创作，同时辅助代码开发和游戏设计，如借助GPT辅助关卡设计。”

——程伟光，武汉多比特公司 AIGC事业部副总裁



【金融】

“在金融领域，代码助手为技术团队开辟了敏捷开发的新路径。传统瀑布式开发模式已不符合当下的开发节奏，月珀的加入不仅助力技术团队高效编写原型和代码模块，而且能够和人工集成、联调相结合，最大限度地降低劳动力密集且价值相对较低的技术环节成本。”

——柳晓光，毕马威中国金融业科技咨询服务合伙人

业务流程优化

大模型能通过智能助手等方式，嵌入业务流程环节中，提高环节间的交互效率，并通过自动化方式简化、优化流程，提升业务运营效率。例如游戏行业，大模型能使初步概念快速转化为可视化内容，降低策划与美术之间的沟通成本，提高游戏开发效率。能源行业，像电网运维检修，大模型结合向量数据库加速了知识图谱的创建，支持数据处理和决策流程提效，提升作业规范 and 安全性。金融行业，大模型已被用于自动化交易、风险管理和合规监控中，辅助识别欺诈行为等风险，同时提高交易的处理速度和精确度。协同办公领域也在融入大模型，如视频会议结合大模型能自动生成会议记录和摘要，形成行动计划并自动分发任务信息等。



【金融】

"大模型技术在金融领域深化客户交互与营销管理，通过 chat bot和数字人提高服务的个性化与效率。这些 AI工具不仅改善了客户体验，还加速了获客和展业过程，尤其在投资和保险方案制定中展现出精准服务能力。同时在办公协同方面，大模型自动化生成会议记录和摘要，提升了金融机构内部的工作流程效率。"

——柳晓光，毕马威中国金融业科技咨询服务合伙人



【能源】

"在运维检修等关键场景中，需要促进作业人员的安全、规范执行。传统方法如结合知识图谱和监控摄像头，比较复杂、效果有限。大模型和向量数据库的引入，能加速知识图谱创建过程，提高应用效果，简化电网的数据处理和决策流程。"

——杨劲松，北京未来式智能科技有限公司创始人兼CEO



【软件】

"IT技术的发展, 最开始做的是流程化, 如通过RP9 去帮助企业提效, 把重复劳动替换掉; 然后我们增加了智能化, 最开始落地在聊天机器人, 帮企业提供各种数字员工, 通过智能化的方式进一步提效; 再然后是 月1重塑 workflow, 也就是让客户原来的工作流程体系逐渐被RI替代。"

——李佳芮, 句子互动创始人&CEO



【游戏】

"传统方式上, 策划向美术传达概念草图时面临巨大挑战, 因难以通过言语让对方完全理解内心的视觉想法。 引入基于大模型的RIG C工具后, 策划可将初步概□快速转化为一系列 2D或3D 内容, 供和美术一起讨论, 有效避免意见分歧导致的内耗,极大简化了美术风格定调和选型过程, 显著提高游戏制作的效率。"

——陈俊文, 腾讯游戏天美工作室游戏A1负责人

智能化决策支持

大模型还能在运营分析和管理的上发挥作用。例如,在供应链管理中, 大模型能辅助分析预测需求, 支持库存优化,提高物流效率并降低成本。交通领域,大模型能整合人、车、路、环境等多元的数据,提供智能管控方案, 改善城市交通的流量管理。金融行业,大模型能辅助高效检索信息,提高了财富管理顾问等决策和服务效率。在数据分析方面,大模型能辅助数据探索、自动执行数据清洗和预处理等任务,节省分析时间、提升分析结果的时效性。广告行业, 以大模型为底座,通过投放的实时数据反馈来驱动创意素材的生成与预算分配中各项指标的设定,最终能建立起清晰的增长模型与投放策略。

((·))

【广告】

"通过广告营销全链路AI化,可以打通素材创意、投放管理、预算分配与效果数据四大核心环节。其中,AI预估模型,将传统拉新、留存、活跃、转化等运营场景与AI大模型相结合,可精准预估RO/. 9RP U、LTV LT等指标,帮助运营人员了解业务回收趋势与回本周期,提前进行营销策略的优化与调整。"

——秦鹏,易点天下企业技术中心总经理

/A

【交通】

"大模型在干线交通管理中,能将人员、车辆、道路、环境等各种因素纳入考量,通过计算得出高精度的最佳控制方案。在路网管理层面,传统的交通调整方案通常依赖于规范和标准、依赖于人的主观判断,因而充满了随机性与不确定性。相比之下,大模型能够在检测到城市道路或高速公路上的交通流异常情况时,依据交通流的动态数据,并结合无人机的地面和空中巡查等,对交通规划进行更为及时、精确的调整。"

——刘志远,东南大学交通学院教授、博导、副院长

羊

【金融】

"T75WT7利用open RI的技术来访问、处理和合成内容,以全球公司、行业、资产类别、资本市场和地区的洞察形式,吸收T SW 自身广泛的智力资本。 [该产品) 专门为、并且由摩根士丹利设计,通过适当的控制,财务顾问及其团队将利用内部功能提出问题并思考大量的内容和数据,以易于理解的格式提供答案,专门从 刀5WT的内容中生成,并带有链接源文件,帮助财务顾问更好地为客户服务。"

——摩根士丹利财富管理公司 [IT5WTJ



【建筑】

"在设计阶段, RIG C大模型促进了各方之间的交流以及需求的准确传递,从而实现了在设计流程全程的伴随式支持。基于这些观察不难看出,大模型不仅可以在项目初期提供有效的设计支持,满足规范标准和工程经验的要求,还在进一步支持设计师创新和优化设计、促进团队之间精准沟通等方面发挥作用。"

——刘济璃,北京市建筑设计研究院数字总监



【游戏】

"从目前大模型对数据分析日常工作的辅助来看,比较适合想法的产生。比如我们给简单的指令,它可以协助进行数据探索、数据清洗,帮助发现问题、诊断问题。"

——陈峭霖,腾讯游戏海外首席数据科学家

2) 业务创新

业务创新主要是衡量行业大模型对业务供给能力的影响,也是很多机构最希望实现的目标。核心在于大模型的生成能力能扩大内容供给,与应用场景的结合还可能创造新功能或业务,帮助提高业务的竞争力、扩大业务的市场空间。典型场景包括丰富内容创意、优化业务功能、开发新业务等,衡量成效的指标如大模型支持内容生成的速度和质量、业务的用户数、业务的营收等。

降低创意门槛

大模型能快速生成内容,从而降低内容创意门槛,尤其在广告、建筑、规划和设计领域表现显著。广告营销行业最典型,技术进步特别是文生图像和视频大模型如

Midjourney、stable Diffusion、sora和pika等不断迭代升级，最直接冲击的就是广告。这些模型使得创意不再是高不可攀的资源，普通人也能借助AI工具将其想法具象化，极大丰富广告创意的供给。又如建筑和规划行业，大模型可以贯穿从创意启动到审图阶段，辅助将创意逐步转化为合理的设计。

"文生图对我们影响比较大，与设计师的生产息息相关。尤其商业图像很严肃，用户付费要求比较高。之前是人工较多，服务对象较少，设计师水平差异比较大，成本比较高。现在能够通过AIGC工具服务更多的人。"

——卓也，稿定（厦门）科技有限公司视觉算法负责人

(·)

【广告】

"大模型应用显著提高了创意生产的效率，一方面降低营销人员与设计师间的沟通成本，通过自然语言直接生成概念图。另一方面，为创意人员提供新素材方向和灵感，建立素材库，助力创新。"

——王天夫，北京大学现代广告研究所所长



【建筑】

"建筑和规划领域，大模型的重要性可归纳为两个阶段：创意启动阶段，它提供了基于实际空间图纸的辅助工具；审图阶段，它以其规范性和逻辑推理能力为支撑。未来，行业特征可能将不再遵循传统的分类，而更倾向于分为规范逻辑型和发散想象型两种。"

——徐跃家，北京建筑大学建筑与城市规划学院副教授

丰富创意供给

一是扩大供给数量，由于创意生成门槛的降低，普通人也可以通过AI工具，将其创意想法进行呈现，并进行各种再加工和应用，AIGC+UGC的混合方式大幅度提升创意的供应量；二是丰富供给类型，目前文生图只是AIGC的1.0版本，随着sora等新技术的出现，“扔一本小说生成一本绘本或者一部电影”的想象正在走向现实。未来，文生文与文生图、文生视频、游戏引擎等结合，可能创造更多内容形态和服务方式。



【游戏】

“在没有AI的时代，中级原画师需要约一周时间才能完成一张高质量的场景图。然而借助 stable Diffusion，美术人员在简单培训后，一个上午就能产出百张图，从中挑选两张即可满足需求。此外，还可以进一步利用如control net之类的工具进行优化提效。”

——程伟光，武汉多比特公司AIGC事业部副总裁



【广告】

“互联网广告的作业流程包括策略设定、内容管理、创意生产、执行发布和数据分析五个连续迭代优化的环节。其中，执行发布和数据分析环节的人工智能应用已相对成熟，形成有效的“智能小循环”。然而，策略设定、内容管理和创意生产三个环节仍主要依赖人工，构成效率瓶颈。引入大模型到创意生产环节，可显著提高创意生产的规模、试错速度和优化效率，使得人工智能能够贯穿并赋能整个广告作业流程，实现整体流程效率的提升。”

——陈刚，北京大学新闻与传播学院院长



【软件】

"今天的互联网大家理解更多是生产和消费的连接,人和人的连接。通过新的AI供给重塑互联网,目前比较明确的热点在于内容行业,特点是生产和消费都在线上可以完成,比如 character.ai, Artifact和pika就是探索Gen AI时代的内容平台,提供了一种创造即消费,主被动一体的新体验。"

——杜頓康 博士, 金沙江创投合伙人

创新应用场景

大模型也有利于激发用户的需求,创造出全新的应用场景。例如开发者可以利用AI生成环境、角色、对话等,提供以AI生成内容为核心的新业务和个性化服务,在互联网、游戏、建筑等领域都有新尝试。

"基于大语言模型推出 月I咨询师的文字问答服务,好评率已高于真人, AI的营收也比之前翻了约一倍。可能相比真人,输出更稳定, 用户隐私安全也更放心。"

——任永亮, 测测 CEO



【软件】

"9I原生的企业会越来越多,因为成本足够低,可能性空间会很大。比如做照片的、视频的、数字人的。甚至是建个城市模型,可以迅速生成、任意组装和探索,戴上VR头显,就能看到很多可能性的空间。以前这种产品成本很高,有了AI能力之后,成本变低。月I的能力会极大降低元宇宙的成本,同时大大扩展范畴。"

——茹炳最, 腾讯Tech Lead

"我认为未来在虚拟空间搭建方面有很大的发展潜力。利用人工智能技术和低成本建设,我们可以创造出更加开阔和多样化的虚拟环境。举个例子,借助增强现实技术,可以通过显示屏展示虚拟场景,为老旧小区改造带来新的可能性。"

——朱玮, 国匠城 CEO



【建筑】

"大模型能够输入: 文本、图像、草图、三维重建模型、视频等数据,按照不同行业、不同业务场景的需求,按需自动化或半自动化生成高信息、高可计算性的三维场景,支持城市级数字底板生成与更新的3D大模型建设。同时具备泛化为通用模型的潜力,可延伸到影视、游戏、娱乐等行业应用。"

——张宇, 飞渡人工智能研究院院长

3) 体验增强

体验增强主要是衡量行业大模型对用户使用的影响,体现对用户的价值。大模型能提供自然语言交互能力,并随着向多模态、具身智能方向发展,为用户提供更加自然、丰富的体验,从而创造价值增值。典型场景包括交互方式变革、个性化服务、虚拟陪伴等,衡量成效的指标如大模型支持业务的用户使用量、活跃度、满意度、问题解决率、留存率等。

交互方式变革

大模型可能改变现有应用的交互方式,以更自然的对话方式,让用户使用应用和调用功能。例如,文档结合大模型,用户可以通过描述需求,让应用直接生成文字、图形等内容。更有想象空间的是与AR/VR等新交互技术和设备的结合,像苹果的vision pro,可能推动新一轮应用形态和体验的变革。



【软件】

"以 chat GPT为代表的 AI大语言模型应用将复杂交互做了根本性的颠覆，因为只有人与人的对话是最顺畅的，类 chat GPT应用以拟人化的预训练大语言模型提供了接近于人类交流的舒适度与流畅度。大模型应用于政府领域，使市民与数字政府之间的交互会更流畅和便捷，可以视为数字政府的专家级陪伴或助手。"

——吴炳坤,众数信科 CEO

"换位想象一下，如果所有的交互变成自然语言的交互方式呢？比如我要创建一个任务，原来在 5a35产品上要点击 60次才能完成，现在我一句话就能完成，这个一定会有助于 5aa5 的落地。"

——李佳芮, 句子互动创始人&CEO

个性化服务

大模型能够提供个性化服务。典型像教育领域，大模型可以根据每个学生的特点和兴趣，辅助提供千人千面的个性化教学内容和服。同时，在教学过程中，对话式学习也将提升学生学习过程的交互性，提高每个学生的学习兴趣。医疗领域，大模型可以为医生和患者提供助手服务，辅助实现大规模的精准医疗服务实施。



【教育】

"随着人工智能的全面深化应用，教育的精准化程度和个性化水平快速提升，教育方式不再是一成不变的，而是根据学生发展状态不断调整。无论学生处于何种状态，都会定制一个最适合他的学习方案，让他可以按照自己的进度进行学习。"

——曹培杰，中国教育科学研究院未来教育研究所副所长



【医疗】

"人类医生处理类似问题时并不会直接给出答案，而是追问症状，进一步排除可能性较低的疾病。腾讯医疗大模型就很好地实现了这一点，通过多轮提问推导出患者的症结所在。"

——郑治枫，腾讯杰出科学家、腾讯天衍实验室主任

虚拟陪伴服务

在产品服务优化环节，将大模型的能力集成到产品中，成为消费电子、汽车等领域产品智能化能力提升的探索焦点。例如智能音箱vifa chat Mini内置了chat GPT，在保持了专业声学标准的基础上，提升了自然语言交互体验，可用于情感支持、学习陪伴、工作助手等。



【互联网】

"在大模型没火之前，当时做得比较吃力，效果远不如现在。现在做聆心陪伴，大模型的对话能力很重要，不然用户无法持续聊下去。用户在不做假设的情况下聊天，体验不好的时候直接退出。我们用户的体验和对话的轮次息息相关，现在平均对话是60轮左右，单一客户是150轮左右。之前是单一模型和单一角色，现在可以生成各种模式。"

——郑叔亮，聆心智能 CEO

**【教育】**

"大模型有一个特别重要的特点，它可以模拟身份，它有角色感...大模型在未来教育中会成为一种陪伴的角色，甚至是不是可以用具身智能的方式，一个人形机器人陪伴用户成长，然后形成对一个人终身的陪伴。当然这种陪伴不仅仅是教育，因为大模型的能力是智能体的概念 (Agent)，某种程度上会帮助我们处理更多的事情，工作上的、专业上的事情等。"

——王绍兰，北京智谱华章科技有限公司总裁

**【出行】**

"结合车辆感知数据、语音输入和用户行为分析等因素，腾讯智能座舱大模型能够提供场景化分析和智能决策，带给用户更自然的对话体验、生成式的交互界面、多样化的场景编排和更个性化的出行服务建议等，革新座舱智能体验。"

——孙廷，腾讯智慧出行智能座舱产品总经理

3. 构建一个模式：高质量数据飞轮

数据是大模型能够运行并创造价值的核心能源。对具体的行业机构而言，通过大模型生成和扩展价值的核心能力，核心取决于如何充分利用自己特有的数据。在数据问题上往往有一个误区：数据量越大越好，有大量的数据就能训练出好的模型。然而实际上相比规模，数据质量对于大模型的性能更为重要，尤其是对专业性和准确性要求高的行业大模型。

行业大模型的构建，需要一开始就把高质量的数据环境纳入其中，通过体系化的数据治理设计，优先开发数据管道，让大模型能够与应用相关的机构自有数据源建立

连接, 以支持后续能持续不断获得有效的数据, 形成数据飞轮。高质量的数据环境, 不是企业任意数据拿过来就可以, 而是需要应用相关的、能提供上下文理解的数据, 重点投资于持续标记、组织和监控这些数据, 比如行业专家的问答内容。数据架构本身, 还需要涵盖结构化和非结构化的数据源, 支持多样化的数据处理。

"与行业客户合作中, 遇到最核心的问题是数据准备方面, 也是当前最大的成本项。客户往往不清楚要准备哪些数据、数据要整理成什么格式等。工业为例, 数据量是巨大的、一天几个 T, 但对于当前的大语言模型来说, 一些简单重复的时序数据, 没有太大意义, 需要重新规划训练思路。"

——国内某大模型独角兽公司技术负责人

可以建立一些指标和标准来评估和指引数据工作, 并根据实际的训练和应用不断优化完善。如数据的相关性, 评估数据符合模型应用的需要程度, 缺乏相关性的数据很可能降低模型性能。数据的可持续性, 评估数据是否容易持续获取, 能通过业务的用户使用及反馈直接获取的有效数据, 更容易支持数据飞轮的构建。总之, 高质量数据的飞轮是行业大模型成功的重要基础。



实现方式

行业大模型的构建和应用中，由于需求和目标不同，技术实现复杂性差异也较大。通过调研总结，目前机构在使用大模型适配行业应用过程中，从易到难主要有提示工程、检索增强生成、精调、预训练四类方式。

在机构的具体实践中，通常不会只用一种方式，而会组合使用，以实现最佳效果。例如，一个高质量的智能问答系统，会综合使用提示工程、检索增强生成和精调等方式。

图：行业大模型构建方式对比

	提示工程	检索增强生成	精调	预训练
适用场景	快速探索应用，如对话系统、文案创作等 优化提示词，可以显著提升生成内容的质量	需要引用大量外部知识，如问答系统、专业咨询等 生成内容准确性高，保障自有数据所有权	希望通用大模型在行业应用场景表现更好 具有较好的行业泛化能力	通用大模型缺乏目标任务相关知识和能力 专业性高，能准确理解并执行特定任务
代表技术	零样本提示 Zero-shot 少样本提示 Few-shot 思维链提示 CoT	文本嵌入 Text Embedding 稠密段落检索 DPR 向量数据库 Vector Database 重排序 Re-ranking	有监督微调 SFT 低秩调整 LORA 适配器层 Adapter Layers	无监督学习 UL 自监督学习 SSL 人类反馈强化学习 RLHF
实现特点	大模型不用调整 非常轻量化的技术开发 依赖通用大模型自身知识 一般与其他方式联合使用	大模型不用调整 难度不大，性价比比较高 利用外部知识库 目前使用广泛	会局部调整大模型 较复杂，高质量数据集是关键 将行业知识部分内化到大模型 目前使用较广泛	要全面调整或构建大模型 投入大、周期长 大量学习和掌握行业专业知识 目前较少使用

引导：提示工程

提示工程（Prompt Engineering）指通过针对性地设计提示词（prompt），来引

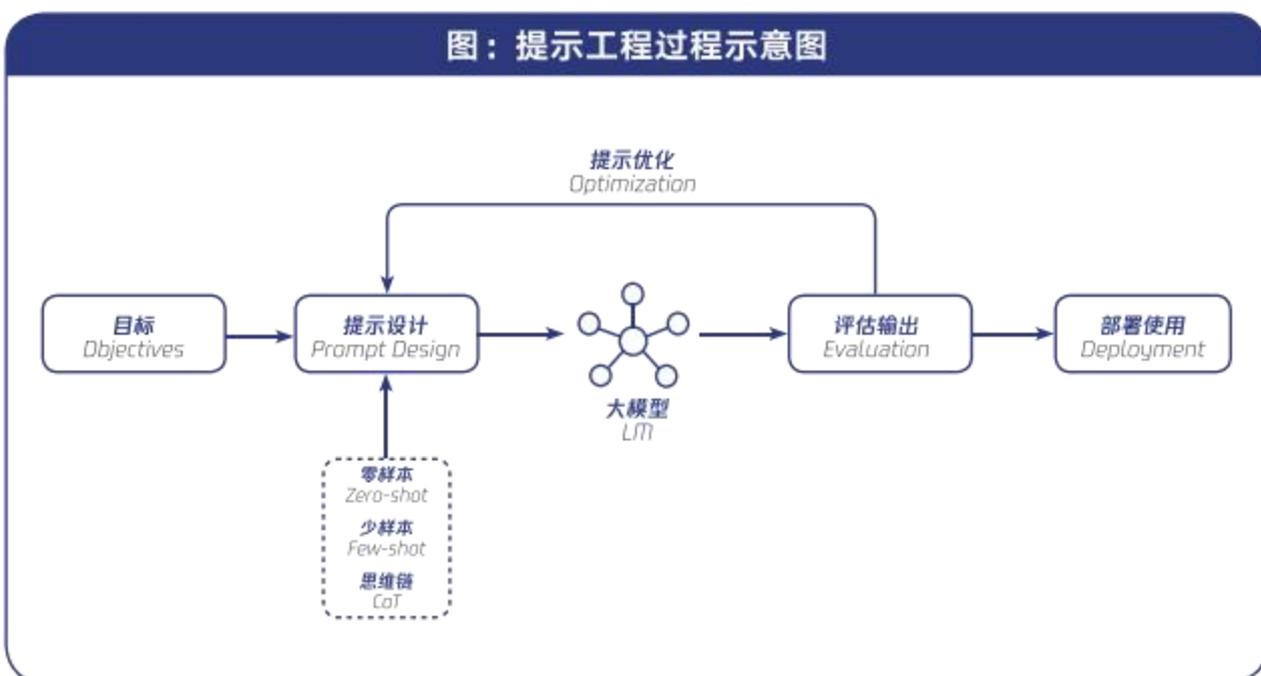
导大模型产生特定应用场景所需的输出。

提示工程上手相对简单，不需要批量采集与构建数据集，更不需要调整或训练模型本身，很多企业刚接触大模型时会采用这种方式来快速探索应用。通用大模型的能力虽然强大，较少的输入也可以生成内容，但随意输入可能产生无效或错误的输出。通过系统设计提示词，规范模型的输入输出方式，企业能够快速得到更准确和实用的结果。

“在行业大模型实践中，企业刚开始在不熟悉大模型的情况下，一般会先从提示词工程来开始使用大模型，特别是一些小的、简单的场景。”

——曹士圯，LangChain Partner Network 成员

图：提示工程过程示意图



提示工程成为持续优化大模型应用的基本方法。通过构建提示库并不断更新,企业的大模型应用开发人员能够在不同场景中重复使用这些提示词,再将用户的开放式输入封装到提示词中传给模型,使模型输出更相关、更准确的内容,避免用户反复试验从而提升体验。

任务的复杂度决定了提示工程的技术方式选择。简单任务可以用零样本提示、少样本提示的方式,不提供或少量提供示例给模型,让模型能快速输出结果,比如对某个文本进行正面或负面的评判。复杂任务则大多需要拆解为若干步骤、提供更多示例,采取思维链提示等方式,让模型能逐步推理输出更精准的结果,例如对一个复杂的工程问题进行数学求解。

提示工程的效果高度依赖通用大模型本身的能力。如果通用大模型训练时包含了行业应用相关的数据,提示工程就能有效引导模型输出更符合行业需要的结果;但若通用大模型本身内含的行业应用数据较少,提示工程的作用就会比较有限。



外挂：检索增强生成

检索增强生成 (RAG, Retrieval-Augmented Generation)指在不改变大模型本身的基础上,通过外挂知识库等方式,为模型提供特定领域的的数据信息输入,实现对该领域更准确的信息检索和生成。

RAG能有效帮助企业快速利用大模型处理私有数据,已经成为企业部署行业大模型应用的主流选择,特别适用于数据资源基础较好的企业、需要准确引用特定领域

知识的场景，如客服问答、内容查询与推荐等。主要优点有：

提高模型应用的专业准确性

让模型能基于特定数据生成内容、降低幻觉；

满足企业自有数据所有权保障的需要

模型本身只会查找和调用外挂的数据,不会吸收数据并训练成模型内含的知识；

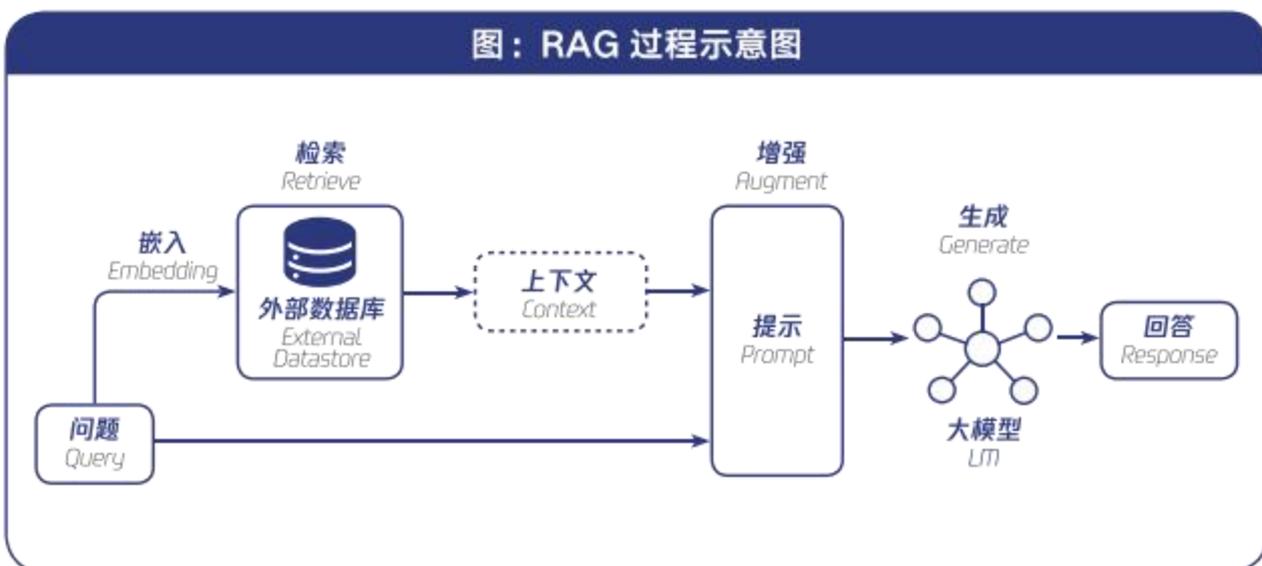
具备较高的性价比

底层大模型本身不做调整，不用投入大量算力等资源做精调或预训练，能够更快速开发和部署应用。

"农业行业来说，向用户反馈的知识准确性尤为重要。惠农网的91惠小农通过RAG思路，大幅提高了农技月1问答结果的丰度和准确性，而农技知识库本身来自以往农技专家在惠农网平台的沉淀，AI通过匹配识别问题和多个答案，聚合生成后对用户提问请求进行解答。"

——周取辉，湖南惠农科技有限公司 CTO

RAG 的能力核心是有效结合了检索和生成两种方法。基本思路是把私有数据进行切片，向量化后通过向量检索进行召回，再作为上下文输入到通用大模型，模型再进行分析 and 回答。具体应用时，当用户提出一个问题或请求，RAG 首先检索私有数据，找到与问题相关的信息。这些信息接着被整合到原始问题中，作为额外的上下文信息和原始问题一起输入大模型。大模型接到这个增强的提示后，将其与自己内部知识进行综合，最后生成更准确的内容。



向量化成为 RAG 提升私有数据调用效率的普遍手段。通过将各种数据统一转化成向量，能更高效地处理各类非结构化数据，进行相似性搜索，从而在大规模数据集中快速找到最相似向量，特别适合大模型检索和调用各种数据的需要。

“我们将企业知识库转成向量，然后保存到向量数据库中，用检索增强生成 (RAG) 的方式和大模型结合起来，从而让大模型掌握企业专业领域的知识，响应业务数据的实时变化，降低大模型的幻觉，增强智能产品的可信度和稳定性。”

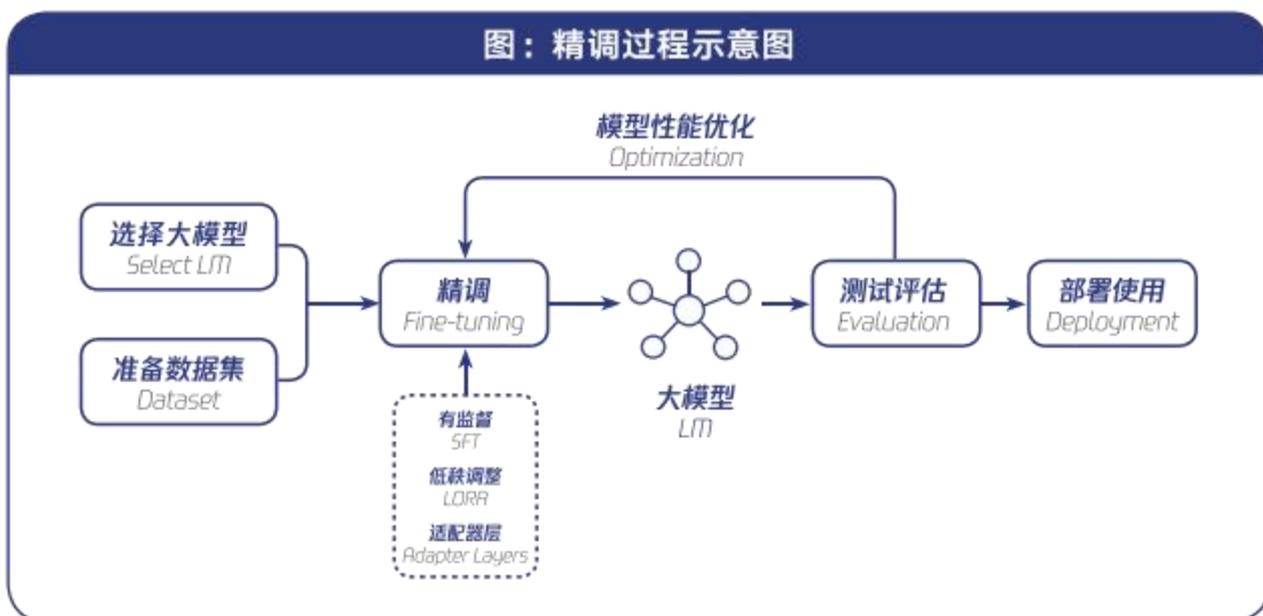
——王超辉，北京仁科互动网络技术有限公司销售易智能产品负责人



优化：精调

精调（FT，Fine-tuning）也常称为“微调”，是在已经预训练好的大模型基础上，基于特定数据集进一步调整大模型的部分参数，使模型能更好地适应业务场景、准确高效地完成特定任务。精调也是目前较为常用的行业大模型构建方法。

图：精调过程示意图



“目前采用开源模型微调方式，主流模型都在试用。重点在开发多轮对话功能，精调采用 SFT(Supervised Fine-tuning, 有监督微调)，大概用几万条业务专家精调的场景数据，还需要加上更多伦理、合规数据。”

——任永亮，测测 CEO

精调适用于特定领域对大模型有更高性能要求的场景。在行业应用中，当通用大模型不能准确理解或生成专业的内容时，可以通过精调的方式，提升大模型在理解行业特定术语和正确应用行业知识的能力，并确保大模型的输出符合特定的业务规则或逻辑。例如，在零售的智能客服场景中，大模型需要理解商品的知识，并按照公司的故障排查流程来提问和回应。

"例如我们在帮助很多假发垂直品类客户解决一系列繁琐的棚拍问题，我们的91GC数字营销创作平台可以自动生成不同年龄、肤色的数字人试戴素材。那在大模型训练过程中，我们结合自身垂直模型，训练不断微调各种数据，比如拍摄视角、发丝成色、柔顺度、光泽度等细节，可以更加逼真地满足客户展示需求。"

—秦鹏，易点天下企业技术中心总经理

精调会将行业知识内化到大模型的参数中。精调后的大模型不仅保有通用知识，还能较为准确地理解和使用行业知识，能更好地适应行业内的多样化场景，提供更加贴合实际需求的解决方案。例如，用医疗领域数据精调的医疗大模型，能够更准确解读专业的医学文献和病历报告，达到为医生提供辅助诊断的要求。

精调是对大模型定制优化和成本投入的折中选择。精调往往涉及大模型权重参数或模型结构的调整，并且需要多次迭代才能达到性能要求，因此相对提示工程、RAG等不改变模型本身的方式，会需要较长时间和较多的计算资源。当然，与从头预训练大模型相比，精调还是一种更为经济高效的方法，因为通常只用对模型做局部调整，所需训练的数据相对少。

高质量数据集是决定精调后模型性能的关键。数据集需要与业务场景密切相关,并且数据标注需要高度精准。高质量数据集既会来自企业内部数据提取也会来自外部数据的采集,均需进行专门的数据标注处理。这些数据需要具备代表性、多样性和准确性,并符合数据隐私等法规要求。只有当足够的高质量数据被用于训练时,精调才能真正发挥作用。

"行业数据的质量比数量更为重要,在做全量精调的过程中,加入的高质量行业数据占比大约在 10-15%;在SFT精调的实践经验中,高质量的问答对,最少要达到一万条,才会有比较好的效果,部分项目甚至高达百万条。"

——某国内大模型独角兽Z公司技术负责人

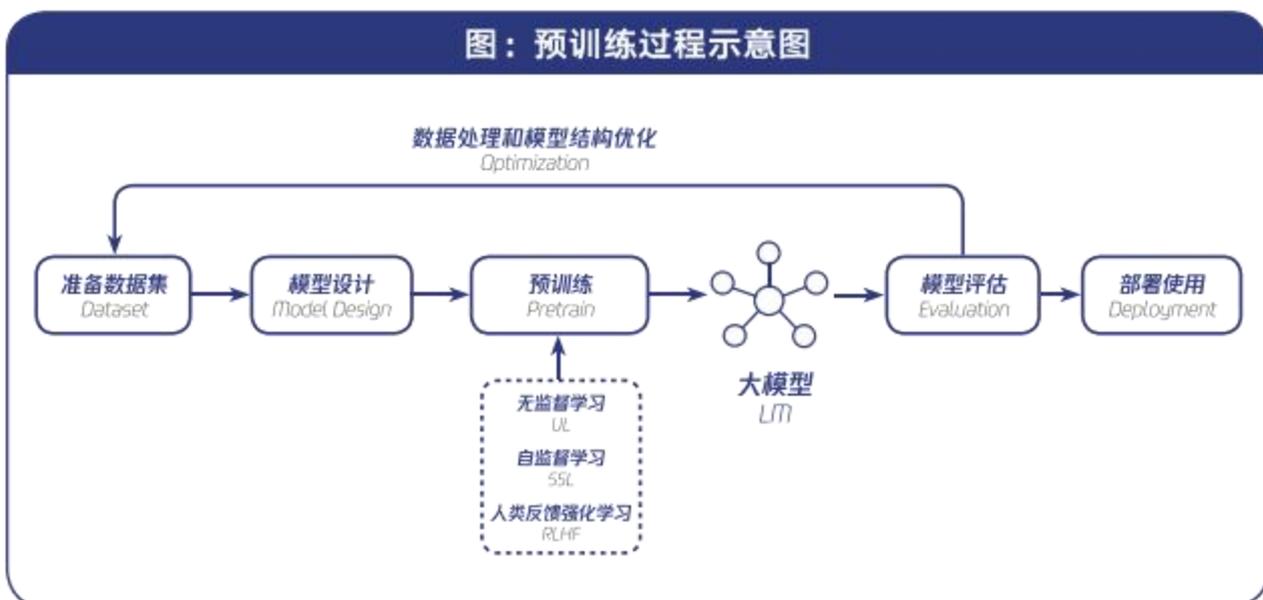
精调策略也直接影响着大模型最终的性能。精调分为全量精调和局部精调。局部精调的方法更为高效、在实际中也比全量精调使用更多,常见形式有:有监督精调(SFT, supervised Fine-tuning),在特定任务的标注数据上调整模型;低秩调整(LORA, LOW-Rank Adaptation),通过低秩矩阵更新减少所需学习参数量;适配器层(Adapter Layers)技术则在模型中加入小型网络层,专注训练这些层以适应新任务。精调策略的选择可根据具体任务需求、数据的可用性以及计算资源的限制来综合考虑。

四

原生：预训练

通过提示词工程、检索增强生成、精调三种方式都无法达到需求标准时，还可以选择预训练的方式，构建一个专门为特定行业定制的大模型。

图：预训练过程示意图



预训练行业大模型适用于与现有大模型差异较大的场景。预训练方式要求搜集并标注大量行业特定数据，涵盖文本、图像、交互记录，以及特殊格式数据（如基因序列）；在训练过程上，模型通常会采用从底层参数开始训练，或者基于已经具备一定能力的通用模型进行后训练（post-training，也称为二次增训），目的是使大模型更好地理解特定领域的术语、知识和工作流程，提高大模型在行业应用中的性能和准确性，确保其在该领域的专业性和效率。例如谷歌的蛋白质生成模型 AlphaFold2，是特定于生物信息学的大模型，其预训练涉及了对大量实验室测定的蛋白质结构数

据的深入分析和学习,使得模型能够捕捉到蛋白质序列与其空间结构之间的复杂关系,从而精准地理解和预测蛋白质的复杂三维结构。

预训练的方式一般投入成本较大,当前较少采用。预训练方式不仅需要大量的计算资源和长期的训练过程,还需要行业专家的密切协作和深度介入。此外,从头预训练还涉及复杂的数据处理和模型架构设计工作,以及在训练过程中不断地调优和验证。因此,只有少数企业和科研机构有能力采用这种高投入、高风险,而潜在回报同样高的方式。未来随着技术的进步和成本的降低,预训练行业大模型可能增加。

预训练行业大模型的技术流程与通用大模型相似,但更注重行业特性。在数据集准备上,从一开始就会加入行业特性的数据;在模型构建技术与流程上,和通用大模型预训练类似,会涉及模型架构设计、预训练任务挑选、大量数据处理、大规模无监督或自监督学习等。例如,使用自监督学习(SSL, self-supervised Learning)技术,通过从数据本身生成标签来学习数据的内在结构和特征,无需人工标注的数据;以及基于人类反馈的强化学习(RLHF, Reinforcement Learning from Human Feedback)技术,通过引入人类专家的主观反馈来引导模型的学习过程,以产生更高质量的输出。

五

典型案例：多种方式组合

1. 商业案例：金融大模型+个股异动写稿

行业大模型是在通用大模型基础上增强了模型的垂类领域知识,能更好地适配和解

决垂类任务或问题,前文提到的预训练、指令微调、RAG、prompt工程都是可增强垂类知识的方法,但效果与投入成正比。方案选型上需要综合考量该垂类的知识丰富度、复杂度和投入产出比。

腾讯金融大模型是在混元通用大模型基础上,在预训练阶段重点加入300B token以上的金融领域语料进行二次增训,使模型对金融知识体系有完整的吸收与理解,之后再通过几十万条高质量的金融指令数据进行了SFT和RLHF。

1) 为什么从预训练开始?

金融作为包含金融市场、产品、机构、风险、法规、量化等的综合学科,知识体系丰富且复杂,指令微调无法涵盖所有金融知识和任务,仅靠对通用大模型做RAG和prompt很难使模型具备较好的金融理解能力。好比招聘金融从业人员,其他条件一样时,还是会更倾向于金融专业背景的人。

2) 为什么需要针对金融领域做SFT和RLHF?

金融场景需要不断适应符合监管政策和合规要求,在产品推荐、投后服务等场景只有通过针对金融领域特定的SFT和RLHF,才能满足金融细分场景的针对性要求、专业风格和偏好需要等,此外从安全合规角度甚至对于一些用户问题需要拒绝回复。

3) 为什么还需要RAG?

通用大模型的创意往往来自它的幻觉率,但金融场景不同于写作等创意场景,对数字的实时性、准确性有非常高的要求,需要通过RAG提供更实时的信息,同时以RAG检索到的信息生成回复可大幅降低幻觉率,仅依赖基座大模型容易导致一招

致命的风险。

总之,在专业、合规、实时和准确等要求下,腾讯金融大模型选择从预训练阶段进行二次增训开始,经过 SFT和 RLHF,在实际推理阶段又整合内部金融知识库辅助 LLM生成,实现在金融任务上效果远超通用大模型。

4)金融场景应用有没有成功案例?

在股票投资服务场景中我们发现,股民当所关注的个股出现较大幅涨跌时,总希望了解背后的原因来辅助决策。腾讯自选股针对这个需求,每日挑选重点个股,进行涨跌原因解读和稿件推送,受到用户欢迎。之前采用人工编辑的方式有个问题,每日稿件生产量不足 10篇,尤其高质量稿件生产耗时长。利用金融大模型,我们有效实现了效率和质量的双提升,稿件生产快、覆盖个股全,同时质量还能达到人工编辑的好,编辑仅需对最终输出进行复核保障安全合规要求。在个股行情触发涨跌幅时,可 5分钟内生成稿件,效率提升5倍以上。

5)有什么需要注意的关键点?

一是大模型无法一步到位,我们原本将个股行情变化、与同时期发生的个股事件资讯全部给到大模型,希望大模型直接生成最后原因解读的稿件,但结果生成质量差、原因也不对口。最后我们发现和 COT 的经验一样,需要 step by step拆解问题,让大模型分步骤来解决;

二是大模型无法一招吃遍天下,需要与金融场景沉淀的"小模型"打配合、互相辅助,之前在业务场景中已使用的舆情正负面模型、事件分类模型等,是基于业务场景特性构建、根据业务指标优化的,用以辅助大模型输出能产生事半功倍的效果。

2. 科研案例：SC BERT单细胞基因数据分析大模型

SC BERT模型是腾讯AI Lab2022年发表的科研成果，首次将transformer运用到单细胞转录组测序数据分析领域，为生命科学和精准医疗领域提供了高质量的AI辅助分析方法。

1) 为什么选择预训练方式?

随着单细胞测序技术的成熟和广泛应用,基础研究和精准医疗领域产生了大量的单细胞测序数据,亟须进行精准的注释和分析。这些数据存在超高维度、高噪声和批次效应等特点,给数据分析带来了较大挑战。以往的注释方法,主要包括基于人工挑选基因注释、基于参考表达谱匹配、筛选基因及降维训练分类器等,存在引入人工偏差、损失全局信息、容易过拟合和泛化性差难迁移等问题。全球已经有较大量的公开单细胞数据集,虽然是很宝贵的研究资源,但其中数据质量参差不齐、缺乏精细标注,未被充分利用。

近年来,基于transformer框架和BERT范式的预训练大语言模型,已经在NLP领域取得了前所未有的突破。但是由于单细胞数据格式与自然语言差异大,难以直接将单细胞数据转换为Transformer对应的embedding输入方式让模型处理。

针对以上错综复杂的行业痛点,腾讯AI Lab针对单细胞测序数据格式,设计了基因表达谱的embedding作为Transformer的输入,充分利用大规模、多样化的公开数据集进行自监督预训练(包括不同来源、组织、细胞类型和测序平台等),有效地对超高维度、高噪声的数据进行编码表征,以数据驱动的端到端方式建立模型,使得模型不受批次效应影响并且具有较强的泛化能力。

2)怎么做的预训练和精调?

AI Lab研究人员以当时世界上最大的公开人类单细胞转录组数据集pang lao为基础，首先进行标准化和对齐,然后针对单细胞转录组数据独特的组成格式，分别设计基因嵌入式表示和表达谱嵌入式表示，处理好的数据作为预训练数据集。

SC BERT模型是基于 BERT架构，采用预训练+精调方式构建的行业大模型。首先在大规模无细胞类型标签的单细胞数据上，基于 MLM (Mask language model) 的方式进行自监督预训练，根据上下文进行回填,从而学习单细胞表达谱的模式。预训练后的模型加入细胞类型注释分类器，随后基于少量精标注在细胞类型注释任务上进行精调，从而实现对不同疾病类型、组织和细胞类型的精准注释。

3)效果怎么样?

该方法是单细胞测序领域首创性研究，开创了单细胞预训练大语言模型的研究方向。模型建立后进行了充分的精准测试，分别在 9 个独立数据集（涵盖 17种人体主要器官，超过 50种细胞类型，超过 500,000 个细胞）对应的细胞类型注释任务上验证效果，并且可扩展至新类的发现、新标记基因的发现、基因相互作用和基因表征等应用场景。研究成果发表于人工智能顶级期刊《Nature Machine Intelligence》，获得学术界和工业界的广泛关注和使用时，在同期发表期刊论文（435,968）中受关注度高居前5%（AI metrics统计）。



安全与治理

大模型大发展同时带来诸多安全和治理新问题, 引发全球各界广泛关注。 诸如 AI 幻觉、可解释性、安全攻击、价值对齐 (value alignment) 、版权保护等, 各国政府都在积极探索发展和监管上的平衡应对之策。 如我国已出台 《生成式人工智能服务管理暂行办法》, 明确提出坚持发展和安全并重、 促进创新和依法治理相结合的原则, 鼓励生成式人工智能技术在各行业、 各领域的创新应用, 同时保障技术应用的安全、可靠、可控。 欧盟在制定 《欧盟人工智能法》过程中, 增加了针对大模型的专门规定, 要求通用目的AI模型 (即大模型) 的提供者履行相关的透明度义务。 美国对于AI治理以软性监管措施和行业自律为主, 例如美国 NIST制定了自愿性的AI风险管理标准、 美国白宫 AI行政命令提出了 "红队测试指南" 、主流AI企业作出了自愿性的AI安全承诺等。 在此形势下, 行业大模型的有效治理也势在必行。

行业大模型是发挥我国在AI领域应用场景优势的重要载体, 但其安全和治理问题可能也更复杂, 不仅面临通用大模型的共性问题, 诸如可解释性、 网络安全、 价值对齐等, 也会面临特定行业的特殊监管要求, 如医疗、金融、交通、教育等强监管行业内存在的强制性标准等。 因此, 需要通过监管和治理创新, 疏通行业大模型在数据获取、算力供给、应用拓展等方面面临的堵点和难点以及相关监管障碍, 尽量给予充分的创新、试错空间, 避免超前治理、多方叠加治理等影响创新应用及时落地。



行业大模型治理的三项原则

针对行业大模型的特点, 我们结合各方实践和共识, 提出三项治理原则, 旨在确保行业大模型的应用不仅能在技术上先进、安全可靠, 而且在法律和伦理上也能得到充分的考量和保障。

一是可信可控原则：行业大模型的开发和应用，需要建立在可信可控的通用大模型之上。应优先选择提供充分安全保障、技术透明度和可控性等的大模型产品和服务，同时考虑到大模型作为行业应用的基础，需要确保技术底座的稳固性和通用性，以有效支持后续的行业定制和功能扩展。

二是数据安全和知识产权保护原则：在大模型的预训练、检索增强和精调等各个阶段，需要使用安全、合规的优质行业数据。确保数据来源于合法渠道，并且符合知识产权保护的要求。鉴于高质量共性知识通常受到知识产权保护，行业大模型需要与掌握这些关键数据的机构建立合作关系，共同解决数据合规问题，避免法律纠纷和诉讼风险。

三是顶层设计与行业协同原则：行业大模型的落地实施可能需要行业主管部门进行顶层设计和统筹规划，以确保行业共性知识库的合规性，并推动建立一个可扩展、可协同的行业大模型技术架构。这种架构应当促进产业链各环节之间的有效沟通和协作，使得各环节能够独立构建和应用其领域模型和企业模型，同时实现跨领域的协同和整合，从而提升整个行业的竞争力和创新能力。



价值对齐是行业大模型可信应用的基础保障

随着大模型的能力日益强大，如何让其行为和目的跟人类的价值观、偏好、伦理原则、真实意图之间实现协调一致，成为越来越重要的议题。从通用大模型的情况来看，目前业界主要是从人机价值对齐的方向来对AI大模型进行安全治理，并在技术上取得了可观的效果，在很大程度上确保大模型部署和使用中的安全与信任。

open AI、Anthropic等美国主流AI企业, 正把大模型价值对齐做成产品的一项市场竞争力。例如, open AI采取人类反馈强化学习 (RLHF)对齐, Anthropic采取AI反馈强化学习 (RLAIF) 对齐。此外, 业界还在探索训练数据质量把控、 对抗测试 (红队测试)、 模型评估、可解释AI方法、 伦理审查、 第三方服务等多元化手段, 共同确保负责任AI的发展。

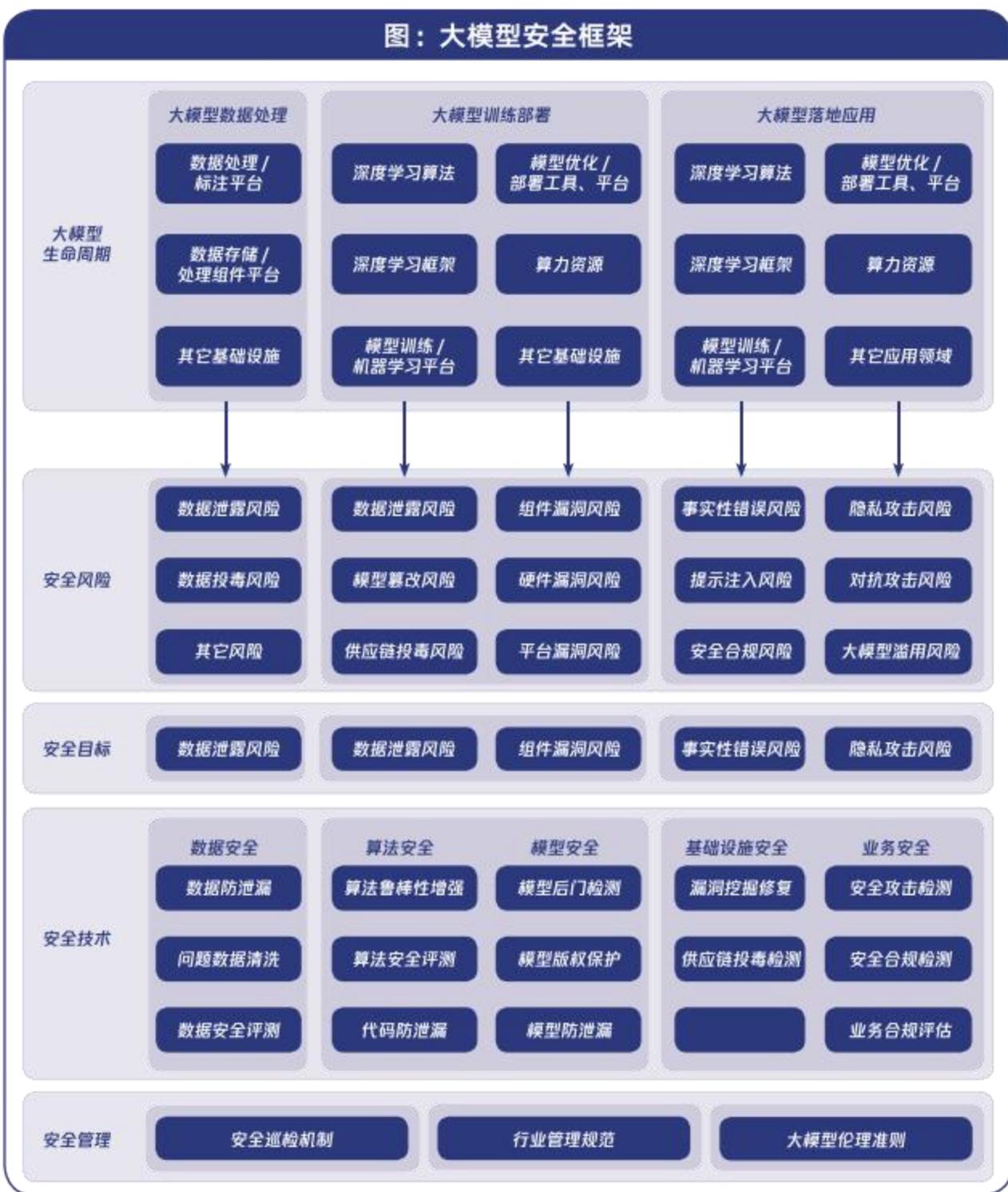
行业大模型对准确和安全的要求往往更高, 价值对齐更重要。 一方面, 选择做好价值对齐的通用大模型作为基础模型进行再开发, 可以大幅减少风险。 另一方面, 对于医疗、教育等领域面向公众提供公共服务的行业大模型, 再开发所采用的数据集、开发后形成的模型成果, 也应再采取类似方法进行价值对齐, 以确保模型的有效性、安全性、真实性和无害性。 为了支持对细分领域的行业大模型进行价值对齐, 可以在考虑具体行业的特殊性基础上, 形成专门的价值对齐指南、 标准和技术规范。



行业大模型更需重视全生命周期的安全保障机制

大模型的数据处理流程相比传统模型复杂, 许多不可控因素带来不容忽视的安全隐患问题, 包括数据隐私安全、 推理安全、 内容安全、 版权保护等, 需要体系化的安全机制, 对行业大模型而言更是如此。 腾讯提出的大模型安全框架, 针对全生命周期安全风险提供原则和方法指引, 包括内生安全及应用安全, 协助系统解决大模型生产研发流程中涉及的基础设施、 数据、 模型、 算法等方面的安全风险, 以及不合理应用可能引发决策失控风险等。 其中, 对于大模型生产研发流程中的内生安全问题, 应该由大模型研发团队及安全团队共同关注, 通过安全评估等机制来保障大模型的应用达到安全目标。 总体而言, 把安全措施落实到具体的研发、 训练、 测试、

部署发布等全流程环节，是行业共识提升大模型安全的方向，可采取包括 prompt 安全测评、红蓝对抗演习、源代码安全防护、基础设施漏洞安全防护等方式。



四

RI沙盒在行业中有较成熟探索 可成为实现监管目标的可行方案

"沙盒" (sandbox)作为源自计算机安全领域的一个技术概念,旨在提供一个安全的隔离测试环境,使软件应用、程序和代码可以在真实环境中运行,保障测试的准确和安全。在AI领域,沙盒环境同样提供了一个安全、可控的测试场所,让研究者和开发者可以自由地测试AI模型和算法,同时避免潜在的风险和不可预见的副作用。AI沙盒在模型安全性测试、性能评估、伦理和法律合规性检验、创新实验等方面,已有较多应用。AI沙盒也可以跟监管结合起来,作为一种创新性的监管方式,帮助实现监管目的。例如,在自动驾驶汽车领域,我国已采取试行汽车安全沙盒监管制度,支持测试前沿技术或新功能新模式。

已有多个国家和地区积极试点AI监管沙盒,如欧盟在其人工智能法案中引入该机制、新加坡特别推出了面向中小企业的 GenAI沙盒等。整体上, AI开发者可以向监管部门申请开展沙盒测试,与主管部门商定沙盒计划,包括可以在沙盒中受监督的现实世界条件下进行测试和应用。而监管部门将对AI企业提供指导、监管预期、监督和指导,以识别、控制潜在风险,支持AI监管沙盒的有效运作。这一机制尤其对AI领域的中小微企业和初创企业较为有利,可以显著降低合规成本。对行业大模型而言,尤其在传统强监管的领域,该机制可以提供一个合规的落地方式。行业主管部门可以针对各行业、领域特点,推出专项的AI监管沙盒计划,鼓励符合资格的企业加入,提供真实环境加快行业大模型测试、验证、试点和先行应用。

合成数据有望为行业大模型增加数据来源

合成数据指在数字世界中创造的、由计算机模拟技术或算法创建或生成的自标注信息,是可替代真实世界数据用以训练、测试、验证AI模型的数据。目前业界普遍判断,合成数据将在未来人工智能发展中扮演核心作用,成为通用人工智能的关键要素。当前,合成数据在交通、医疗、工业、金融等产业领域已经有先行探索,并取得积极成效。

合成数据有几个方面的优点:

- 1)促进创新,通过提供难以获取或受限数据的合成版本,推动科研和商业创新;
- 2)降低成本,相比于收集真实世界数据的高昂成本,提供了更经济的选择;
- 3)保护隐私,能够在保护个人隐私的同时,使得敏感数据得到有效地利用。

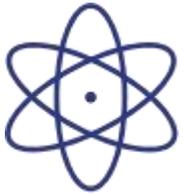
行业大模型大多需要特定领域的专业数据,数据稀缺和隐私安全问题更突出,合成数据可以有效解决数据供给、数据安全利用等问题,进一步夯实行业大模型发展的数据基础,助力行业大模型应用普及。为此,需要营造良好的制度和政策环境,积极扶持、鼓励合成数据技术、产业和相关应用的创新发展。可为合成数据的生成和使用制定明确的隐私保护标准和指南,推动合成数据商业化利用。在公共数据、行业数据开放共享方面,也可以探索通过合成数据方式,实现涉及个人信息、商业秘密、知识产权等数据的安全开放和共享。

六

行业大模型是走向 R1 绿色可持续发展的路径之一

大模型“大力出奇迹”的方式，训练和部署需要大量算力支持，导致大量的能源消耗和碳排放，其环境影响不容忽视。open AI首席执行官山姆·奥特曼承认人工智能行业正面临能源危机，因为下一代生成式AI系统将消耗大量能源，可能超出能源系统的承载能力。大模型的算力应用系统还需要大量淡水用于冷却处理器和发电，数据中心的水资源消耗量也已经显著增加。对此，研究界呼吁优先考虑节能问题，需要构建更高效的模型和重新设计数据中心等。

行业大模型本身就是对通用大模型的“降本增效”，参数量相对大幅减少，算力资源消耗也明显降低，这也意味着降低了能源消耗和碳排放。未来，行业大模型在发展应用过程中，可以将算力及能耗节约作为重点优化方向之一，加强环境影响治理，将绿色可持续打造成行业大模型的一个特性。可以通过试点示范等方式，激励更多行业大模型优化算法、节能减排。



未来展望



"人工智能 +" 促进行业大模型应用提速

2024年政府工作报告明确提出,要"深化大数据、人工智能等研发应用,开展'人工智能+'行动,打造具有国际竞争力的数字产业集群"。国家发改委也提出,将推进"人工智能+"行动,推动人工智能技术与经济社会各领域深度融合,支撑各行业应用创新,赋能百业智能化转型升级,提高生产效率,激发创新活力,重塑产业生态,培育经济发展新动能,形成更广泛的以人工智能为创新要素的经济社会发展新形态。

可以看到,在国家推进新质生产力的背景下,人工智能已经成为经济发展的核心动力之一。未来,国家将加大力度推进通用大模型和行业大模型应用,引导更多行业领域开放应用场景,加强典型应用落地的示范,让人工智能更好赋能千行百业。

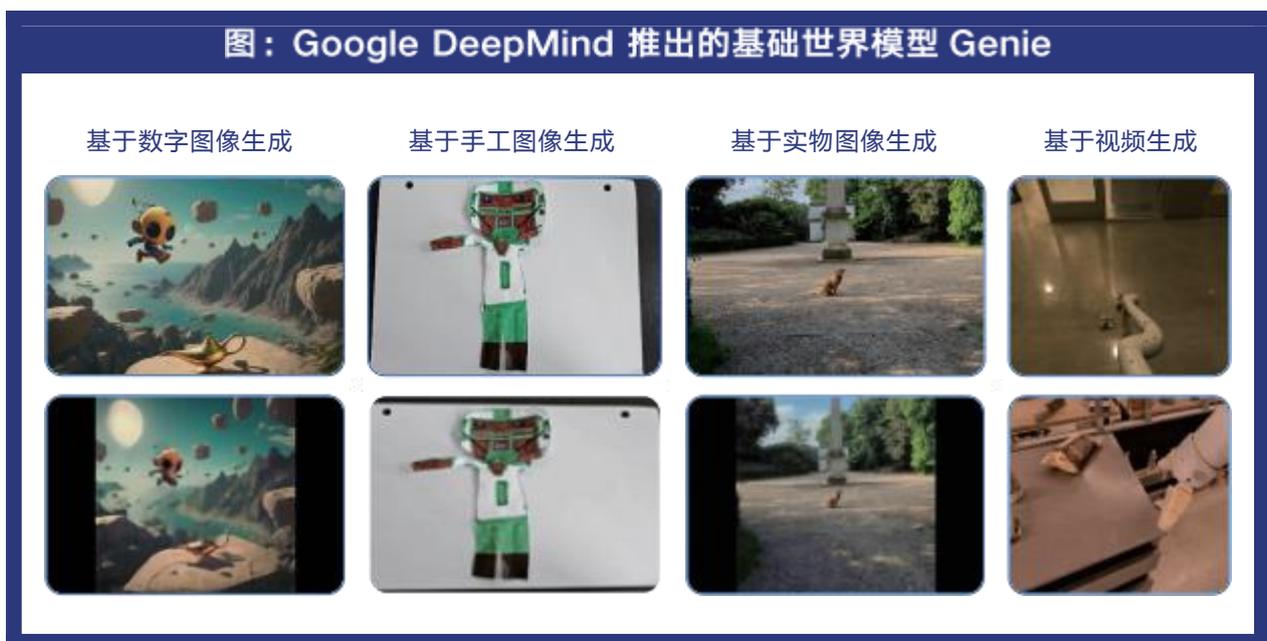
实际上,近年各地已经陆续出台了一系列鼓励措施,在国家政策指引下,预计各地推进产业落地会进一步加速。北京、上海、杭州、广东、重庆等十多个省市,已经发布相关政策积极推动通用人工智能技术在各类场景的应用,涉及金融、医疗、教育、制造、城市治理、交通、农业农村等多个领域。在这个过程中,行业数据集的建设、行业大模型训练和推理算力的配备、交叉型人才的培养等,都是决定行业大模型应用创新与落地快慢的重要因素。



多模态大模型为数实融合打开新空间

多模态大模型向世界模型演进，有望驱动各行业落地加速。多模态是人类世界的本来面貌，大模型技术正在从文本、图像、视频（2D和3D），到声、光、电，甚至分子、原子等各类模态的生成，并实现跨模态的迁移。谷歌发布的 Gemini，主打原生多模态大模型，无缝跨越文本、图像、视频、音频和代码。open AI发布的文生视频模型 sora，可以直接输出长达 60秒的高质量视频，尤其是sora对物理规律的模仿，已经具备了一定程度的世界模拟器能力，有望向世界模型进化。Google Deep Mind推出的基础世界模型Genie，基于互联网视频训练，可以从合成图像、照片甚至草图生成无数种可玩（动作可控）的世界。

图：Google DeepMind 推出的基础世界模型 Genie



多模态技术的持续进步,在丰富用户多维和沉浸式体验、提高多模态数据处理效率、理解复杂的现实世界场景、创新各种新产品形态和新服务形式等多方面,将产生巨大价值。结合行业知识,多模态大模型有望应用于千人千面的互动广告、交通态势感知、制造业产品研发设计、农业生产检测和优质育种等众多场景。

多模态大模型跟VR/AR、全真互联网的结合,也是未来值得期待的方向,有望强化数字世界和现实世界的深度连接。随着多模态大模型对物理规律认知的加强,大模型不仅能复刻现实世界,还可以把数字世界中的决策推演到物理世界,实现大模型从“发现——理解——决策——行动”的闭环,进而探索新的行业知识,优化产业研发、生产、销售和服务等全流程,成为各行业更能干的智能助手。



RIRgent有望为各行业注入生产力

以大模型为“大脑”的AI Agent (人工智能体)持续演进,有望逐步成为各行各业不可或缺的新型生产力。chat GPT发布后不久,open AI即推出开源项目Auto GPT,开创了新一代AI Agent的先河。与传统的人工智能应用方式相比, AI Agent具备通过独立思考、调用工具去逐步完成给定目标的能力。这些高度自主、能够迭代优化和独立运作的 AI Agent好比数字员工,预示着一个全新的生产力时代的来临。

尽管目前技术尚未成熟,但已有一些与工作流程深度耦合的 AI Agent涌现。生产力领域, Microsoft 365 copilot及其个人版copilot pro的推出,代表了AI Agent在提高工作效率和办公自动化方面的巨大潜力; 社交领域, Meta AI等产品将进一

步推动内容创作和社交互动的智能化,为用户带来更加丰富多样的体验;工业领域, Lev atas等工业视觉检查机器人的应用,显示了AI Agent在识别、判断和维修设备方面的高度自主性;营销领域, sales GPT等平台的出现,展示了AI Agent在感知情绪、个性化推广和客户服务中的潜力.....



在这一系列的进展背后,一个明显的趋势是AI正从简单的工具 (Tool), 进化为复杂的助手 (copilot)乃至代理 (Agent)。这一转变的拐点在于AI技术在深度学习和自主决策能力的突破。一旦AI Agent能够准确理解复杂的任务需求、自主选择最合适的解决方案,并有效控制任务进度,就能在各行各业实现智能化转型,推动生产力的指数级增长。未来, AI Agent可能集成到组织机构的各个层面,与人类员工、其他数字化系统及基础设施形成一个互联互通、高效运作的生态系统。这种深度整合不仅优化了组织结构和提升了运营效率,也将为企业创造前所未有的价值。

四

行业大模型将出现更多的端侧形态

大模型正在向端侧转移，AI推理将在手机、PC、耳机、音箱、XR、汽车，以及一系列可穿戴式新型终端上运行。端侧大模型具有一些独特优势，如本地数据处理效率更高，节省云端服务器带宽和算力成本，对用户数据更好的隐私保护，开启更多交互新方式、新体验等。其中，大模型在端侧的应用会让自然语言对话、姿态理解等人机新交互方式成为可能，将为用户带来流畅、便捷和贴心的全新使用体验，成为促进用户消费升级的主要动力。

作为新技术应用的风向标，手机和汽车行业有望优先落地端侧大模型技术。目前行业内已推出小米MiLM，VIVO蓝心大模型、OPPO安第斯、荣耀魔方等端侧大模型，三星也推出了由谷歌Gemini驱动的新智能手机。

未来，手机端侧大模型应用有2种可能性：

1) 手机AI助手，在操作系统层面提供服务，能跨越多个应用程序并提供全面支持。用户可以通过语音、文本或其他输入方式与全局AI助手交互，以执行各种任务，如搜索信息、管理日程、发送消息、控制智能家居设备等，甚至可以调用各类手机APP的接口。用户不用打开APP，就可以用语音交互完成各种服务，如点外卖、查天气、打车、购物等。类似苹果的sir i或安卓的Google Assistant，内置在手机操作系统中。

2) APP内AI助手，增强现有APP的语音交互和智能化能力。由各APP开发团队

提供，支持大模型在单个应用内对用户提供的支持。比如办公软件或社交软件内部的AI助手,每个助手通常针对其所在应用的特定功能和用例进行优化。

汽车行业的大模型落地，首先是智能座舱，问界、小鹏、蔚来、长城都自研或联合研发了座舱大模型，为用户提供更为贴心的交互体验。未来，用户可以直接跟座舱数字人对话，就能实现用车指导、导航、娱乐、服务信息查询、聊天陪伴等全面贴心的服务。其次是自动驾驶，以BEV+Transformer架构为核心，以重感知+轻地图的方案为主的城市NOA(Navigate on Autopilot领航辅助驾驶)将进入量产阶段。目前已有小鹏XNGP、理想AD Max等推出。加上2023年11月《关于开展智能网联汽车准入和上路通行试点工作的通知》正式落地，L3/L4的高阶智驾获得政策支持，商业化发展将迎来新阶段。

五

云智能一体支持行业大模型加速落地

云计算通过提供强大的基础设施、灵活的资源管理、全面的服务支持、更强的成本优化、开放的生态系统、专业化的工具和保障安全性等多方面的支持,将显著加快各行业大模型从研究探索到实际应用的落地过程。

首先，云计算提供了稳健的硬件基础设施，不仅包括提供CPU、GPU、AI专用加速芯片等异构算力，还具备万卡级以上并行计算的高速互连网络能力。在高性能芯片紧缺的背景下，这对于大模型的训练和推理至关重要。企业和研究机构无需投入巨额资金建设自己的数据中心，即可获得强大算力。其次,长期以来，行业落地的效果、成本和弹性往往不可兼得，而云+大模型的方式，有望带来新解法。云具有

极强的规模优势，从而带来用户成本的显著降低；而且随着最终用户的快速增长，推理服务方面的弹性扩容也更为必要，相比私有化部署，公有云提供了更为便捷可靠的方式，可以满足高吞吐量，高并发量等动态需求。第三，云计算服务商还提供了一系列专业化的服务和工具，如支持千亿向量搜索的向量数据库、专有机器学习平台、自动化的模型调优和超参数优化工具、数据加密、访问控制和合规性管理工具等，帮助行业用户更快搭建大模型应用。

图：腾讯云大模型高性能算力底座方案



随着 AI 应用的深入，模型的规模和复杂性将不断增加，MaaS 将日益成为行业用户云上用智的主流方式。用户或直接调用云的大模型 API，或借助全生命周期的大模型训练工具，生成适用于自身场景的大模型，并托管在云上，为最终用户提供高质量智能服务。

编写团队

顾问指导:

汤道生 腾讯集团高级执行副总裁、云与智慧产业事业群 CEO

蒋杰 腾讯集团副总裁

张正友 腾讯首席科学家、腾讯 AI Lab 及腾讯 Robotics X 实验室主任

湛炜标 腾讯金融科技副总裁

司晓 腾讯研究院院长

冯宏声 腾讯研究院高级顾问

项目策划:

刘琼 杨乐 张钦坤 杜晓宇 曹颖 田小军 吴朋阳 李瑞龙

编写人员:

刘琼 吴朋阳 李瑞龙 王强 袁媛 袁晓辉 曹建峰 王鹏 白惠天

李南 王枢 陈楚仪 牛福莲 闫德利 余一

主要访谈机构/专家 (排名不分先后) :

邱泽奇 北京大学社会学系教授

汪琼 北京大学教育学院教授

张文佳 北京大学城市规划与设计学院副院长、研究员

黄璞 北京大学政府管理学院教授、博导

王天夫 北京大学现代广告研究所所长

张昕 清华大学建筑学院院长聘教授

- 郭庆来 清华大学电机系教授
- 高峰 清华能源互联网研究院副院长
- 沈阳 清华大学新闻学院新媒体研究中心主任
- 王天夫 清华大学社会学系教授
- 安孟瑶 清华大学临床医学院博士后研究员
- 聂再清 清华大学国强教授、智能产业研究院首席研究员
- 童星 深圳清华研究院智慧能源研发中心副主任
- 陈洪 北京邮电大学文化和旅游研究基地负责人、北京邮电大学交互技术与体验系统文化和旅游部重点实验室副主任
- 仇沪毅 北京邮电大学文化和旅游研究基地讲师、博士
- 谢新水 首都师范大学管理学院院长、教授
- 郑跃平 中山大学政治与公共事务管理学院副教授、博导
- 郑磊 复旦大学国际关系与公共事务学院教授
- 叶宇 同济大学建筑与城市规划学院副教授，博导
- 刘超 同济大学建筑与城市规划学院助理教授
- 李兴华 同济大学中国交通研究院院长
- 線凯 北京交通发展研究院研发中心主任
- 曹剑东 交通运输部科学研究院信息中心总工
- 曹培杰 中国教育科学研究院未来教育研究所副所长
- 许世卫 农业农村部农业监测预警技术重点实验室主任
- 刘桂才 农业农村部信息中心原总工程师
- 谷晓峰 中国农业科学院生物技术研究所副所长
- 钟俊浩 上海市人工智能行业协会秘书长
- 王逸浩 上海市人工智能行业协会研究部部长
- 王宇 上海市人工智能实验室产业合作和知识产权部负责人

编写团队

- 何宛余 小库科技CEO
- 李濮实 中国金茂信息技术中心技术经理
- 李一帆 品览科技董事长
- 刘 刚 广联达科技股份有限公司副总裁、研究院院长
- 刘济璃 北京市建筑设计研究院数字总监
- 茅明睿 城市象限CEO
- 莫修权 清华大学建筑设计研究院有限公司副院长
- 宋银灏 广联达数字建筑研究院特聘首席专家
- 孙 亮 中国金茂信息技术中心副总经理
- 吴若飒 上格云CEO
- 张英楠 上海建工四建集团有限公司建筑人工智能研究室主任
- 张 宇 飞渡人工智能研究院院长
- 朱 玮 国匠城CEO
- 胡 嵩 北京一亩田新农网络科技有限公司 CTI
- 申 斌 湖南惠农科技有限公司 CEO
- 周取辉 湖南惠农科技有限公司 CTO
- 曹士坦 Langchain partner Network成员
- 杜頔康 金沙江创投合伙人
- 何 润 致趣百川联合创始人 &CEO
- 李佳芮 句子互动创始人兼CEO
- 程伟光 多比特AI平台事业部副总裁
- 柳晓光 毕马威中国金融业科技咨询服务合伙人
- 时培听 寄云科技创始人兼总裁
- 智 振 中工互联董事长
- 闫 洋 花房集团战略总监
- 徐 赫 赛文交通网创始人

- 王绍兰 北京智谱华章科技有限公司总裁
- 刘丁泉 北京智谱华章科技有限公司工业智能部 CTO
- 朱 红 国家电网首席A 科学家
- 杨劲松 未来式智能科技创始人与CEO
- 王 敏 七猫投放策略负责人
- 王雪纯 七猫创意团队负责人
- 张殿鹏 省广集团云与全媒介事业群数智中心总经理
- 叶 劲 省广集团云与全媒介事业群品效运营中心总经理
- 肖 龙 省广集团云与全媒介事业群品效运营中心资深创意总监
- 陈 竹 中信证券首席医疗健康行业分析师
- 曹艳林 中国医学科学院\北京协和医学院医学信息研究所医疗卫生法制研究室主任
- 乔 木 北京水木分子生物科技有限公司联合创始人、首席技术官
- 王少程 北京清华长庚医院内分泌科副主任医师
- 吴炳坤 众数信科 CEO
- 唐 鹏 电子政务理事会副秘书长、数旗智酷创始人
- 吴海波 we shop总经理
- 卓 也 稿定(厦门)科技有限公司视觉算法负责人
- 郑叔亮 北京聆心智能科技有限公司CEO
- 秦 鹏 易点天下企业技术中心总经理
- 任永亮 cece Group (测测集团) CEO
- 王超辉 北京仁科互动网络技术有限公司销售易智能产品负责人
上海稀宇科技有限公司
- 杜西库 腾讯金融科技副总裁
- 常 越 腾讯广告市场部副总经理
- 陈俊文 腾讯 IEG 天美 J5 工作室游戏 AI 技术专家

编写团队

- 邓民文 腾讯AI Lab智能计算中心总监
- 殷俊 腾讯游戏 CROS 研发效能部数字内容技术中心总监
- 贺志强 腾讯 IEG 天玑智趣工作室总监
- 林智超 腾讯 IEG天美T1工作室程序负责人
- 曹木勇 腾讯 IEG 研发效能部 Gvoice 研发组组长
- 廖诗颢 腾讯游戏魔方工作室群 AI 技术负责人
- 陈峭霖 腾讯游戏海外首席数据科学家
- 茹炳晟 腾讯微信支付Tech Lead
- 孙星 腾讯优图专家研究员
- 汪凯锋 腾讯云智能商业化中心金融业务负责人
- 何秀强 腾讯金融科技智能平台中心高级总监
- 周梦 腾讯金融科技数据科学专家
- 李峰 腾讯金融云首席数据科学家
- 施宇 腾讯广告商业 AI 部智能生成中心负责人
- 孙珏 腾讯智慧出行智能座舱产品总经理
- 沈沛 腾讯智慧出行战略总经理
- 张潇恺 腾讯智慧出行战略经理
- 陈梦珂 腾讯智慧交通副总经理
- 丁捷 腾讯智慧交通行业应用专家
- 姚建华 腾讯AI Lab AI医疗首席科学家
- 杨帆 腾讯AI Lab高级研究员
- 黄蕾 腾讯智慧农业行业总监
- 郭红英 腾讯云解决方案总监
- 金翔 腾讯云智能产品部智慧能源专家
- 朱季峰 腾讯朱雀实验室AI安全专家
- 唐梦云 腾讯朱雀实验室高级研究员



腾讯研究院

Tencent
Research Institute

向RI而行 共筑新质生产力

行业大模型调研报告



— 腾讯研究院公众号 —



— 腾讯研究院AI速递 —

腾讯研究院

腾讯公司设立的智库机构，旨在依托腾讯公司多元的产品与服务，围绕互联网发展的焦点问题，以科技向善为指南，通过开放合作的研究平台，汇集各界智慧，共同推动数字社会健康、有序的发展。