

McKinsey
& Company

麦肯锡 中国金融业CEO季刊

全球洞见 中国实践

捕捉生成式AI新机遇



2023年秋季刊





麦肯锡 中国金融业CEO季刊

总 编： 倪以理 曲向军

编辑委员会： 韩 峰 容觉生
方浩翔 方溪源
胡艺蓉 曾文静
李静瑶 宋 戈
邱外山 王喆宸
蒋子翔 贺 珏
鲁志娟

特别顾问： Nick Leung
Joydeep Sengupta
Alex Sawaya

我们谨此向全球金融咨询业务部门的众多同事表示诚挚感谢。他们是本书8篇文章作者或内容贡献者：Michael Chui、Alex Singla、Alex Sukharevsky、Bryce Hall、Lareina Yee、Eric Hazan、Shivani Gupta、Abhisek Jena、Begum Ortaoglu、Barr Seitz、张力、Roger Roberts、Tanya Rodchenko、Delphine Zurkiya、Stephanie Brauckmann、Anusha Dhasarathy、Martin Harrysson、Klemens Hjartar、Alharith Hussin、Naufal Khan、Sam Nie、Chandrasekhar Panda、Henning Soller、Nikhil Srinidhi、Asin Tavakoli、Niels Van der Wildt、Anna Wiesinger、陳香輝 (Jenny Chan)、Glenn Leibowitz (王磊智)、刘猛、Kai Shen、Naveen Sastry、Enno De Boer、Ray Zhou、Diane Henry (排名不分先后) 等。

麦肯锡公司2023年版权所有。

2023年秋季刊

麦肯锡中国金融业CEO季刊

刊首语

我们置身于技术革新时代。在 2023 年众多科技趋势和投资热点中，生成式人工智能（下称GenAI）无疑是年度明星技术。伴随着ChatGPT的横空出世，GenAI成为各行各业关注和热议的话题。全球科技巨头和AI厂商纷纷下场，唯恐错过此番科技盛宴。行业翘楚和媒体将GenAI浪潮类比昔日的移动互联网机遇，认为它将对全球经济和各个行业带来深远影响，企业也将迎来重大变革机遇。

与传统AI相比，GenAI拥有四大核心优势——自动化和效率提升、个性化和定制化、创造性和创新能力、以及解释性和透明度。麦肯锡研究显示，人工智能（AI）整体将为全球经济带来高达25.6万亿美元的正经济影响，而其中来自GenAI的贡献将高达7.9万亿美元¹。这其中，零售和消费品行业、银行业、制药和医疗产品行业的价值潜力最大。

以银行业为例，当前四类最主流的应用方式包括：任务和流程自动化、用户互动、内容创作、以及代码编写加速，合计贡献占GenAI总效益的75%。而从企业经营和业务部门的视角来看，其对营销与销售、客户运营与产品研发三大部门的影响最大。不过值得注意的是，金融机构在应用GenAI时需关注模型幻觉、恶意使用、信息泄露等风险。

由于GenAI蓬勃发展，其产业规模也在高速增长，投资者纷纷入局。2022年GenAI市场收入为400亿美元²，预计2027年及2032年将分别达到3,990亿美元和13,040亿美元¹，2022~2032年复合增长率达42%¹。中国市场，2022年规模约人民币660亿元³，2020~2025年复合增速将达到84%³，2025年中国GenAI市场将占全球市场规模（2,170亿美元）的14%。可见，GenAI不仅会为全球经济创造巨大价值，其本身也蕴藏巨大投资机会。

¹ 基于2022年全球经济结构的预估

² 数据来源：Bloomberg Intelligence

³ 数据来源：《中国AI数字展望2021~2025》，前瞻产业研究院和中关村大数据产业联盟

本季刊围绕生成式AI主题，通过4大章节共8篇文章，全面深入分析了GenAI对各主要行业的影响、价值链投资机会、中国GenAI市场现状和未来趋势以及企业如何布局GenAI，从而真正挖掘其价值。

第一章，我们分析了GenAI对全球经济的深刻影响及其在各行业的广泛应用，并从企业领导者角度出发，提炼出《生成式AI：CEO必读指南》以及《生成式AI引发科技变革：CIO和CTO参考指南》。

第二章，我们以投资者视角深入探讨在GenAI产业链的六大环节上——专用硬件、云平台、基础模型、模型中心和MLOps、应用及服务——分别存在哪些重要投资机会，并聚焦价值链前五个环节中的明星独角兽企业。

第三章，全面分析了中国GenAI市场发展动态，包括市场规模、技术栈、商业化应用和风险管理这些大家普遍关心的话题，为读者呈现一个全景式的中国GenAI市场概览。

第四章，就如何布局GenAI从而挖掘其价值，我们提出实用策略和建议，并盘点了市场上三类服务商的特点和各自优势，帮助企业管理者更好筛选合适的合作伙伴。

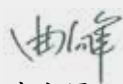
我们深知，对许多企业而言，如何着手部署GenAI仍是一个待解的难题，独立探索GenAI能力也会遇到诸多难题，因此，本期季刊深入浅出从不同方面探讨GenAI的影响和行业应用潜力，希望能帮助企业捕捉这一历史性新机遇，谱写第二增长曲线。

祝您开卷愉快！



倪以理

全球资深董事合伙人
麦肯锡中国区主席



曲向军

全球资深董事合伙人
麦肯锡中国区金融机构咨询业务负责人

目录

第一部分

未来已来，生成式AI驱动全球生产力革新与商业模式重塑

008 GenAI 将为全球各行业创造巨大价值

GenAI 技术在提升行业生产效率和促进产品创新方面具有显著优势，预计未来将颠覆全球各行各业的现有格局。在我们分析的 63 个用例中，GenAI 将在不同行业产生 2.6 万亿至 4.4 万亿美元的价值。其中，零售和消费品行业、银行业、制药和医疗产品行业的价值潜力最大。

040 2023 年人工智能发展现状：生成式 AI 的突破之年

麦肯锡关于生成式 AI 现状的最新年度调查，综述生成式 AI 工具在 2023 年的爆炸式增长及其影响，包括 GenAI 的普遍应用，AI 领域领先企业如何抢先布局，对人工智能相关人才的需求变化以及 GenAI 普及率和影响范围。

060 生成式 AI：CEO 必读指南

本文旨在帮助 CEO 及其团队思考生成式 AI 的价值创造场景以及如何开始应用。包括 CEO 的 GenAI 入门指南，4 个旨在提高组织效能的案例解析，以及 CEO 如何发挥关键作用利用 GenAI 带领企业走向成功。

084 生成式 AI 引发科技变革：CIO 和 CTO 参考指南

本文通过与数十位科技企业领袖的深度对话和对 50 多家企业的 GenAI 举措分析，为科技企业领袖列出了成功部署 GenAI 的九大举措，用以创造价值、协调技术和数据、扩展解决方案并管理 GenAI 风险。



第二部分



洞悉先机，生成式AI浪潮下的全球产业机遇与投资价值

104 投资机构篇：挖掘生成式 AI 产业新机遇

本文深入分析了GenAI 价值链的六大环节——专用硬件、云平台、基础模型、模型中心和 MLOps、应用和服务，帮助投资机构了解其中蕴藏的大量投资机遇。

第三部分



赋能百业，中国企业积极布局生成式 AI“新赛道”

126 中国生成式 AI 行业发展现状及未来趋势

本文深入浅出地分析了中国 GenAI 行业的发展现状和趋势，详细剖析中国在多个细分领域存在的机遇与挑战，帮助企业探索适合自身发展的商业模式，抓住 GenAI 技术红利。

第四部分



行稳致远，企业规模化部署生成式 AI 的“道”与“术”

146 从战略到规模化实施，“四步走”全面落地 GenAI

企业在考虑引入 GenAI 时应以速度为先，遵循“四步走”方法论，高效实现 GenAI 部署和规模化应用：首先快速梳理和构建战略；然后确定用例图谱和转型路径；其次试点优先用例，为规模化建立基础；最后推进全方位的运营模式转型以规模化推广 GenAI。



167 让 GenAI 应用真正为企业创造商业价值

企业可以从 GenAI 应用中获得至少三方面价值，一是减少手工作业并缩短冗长的业务流程，二是协助团队进行高级分析以及更加准确地提炼洞察，三是加速活动 / 产品的设计与交付过程。

第一部分

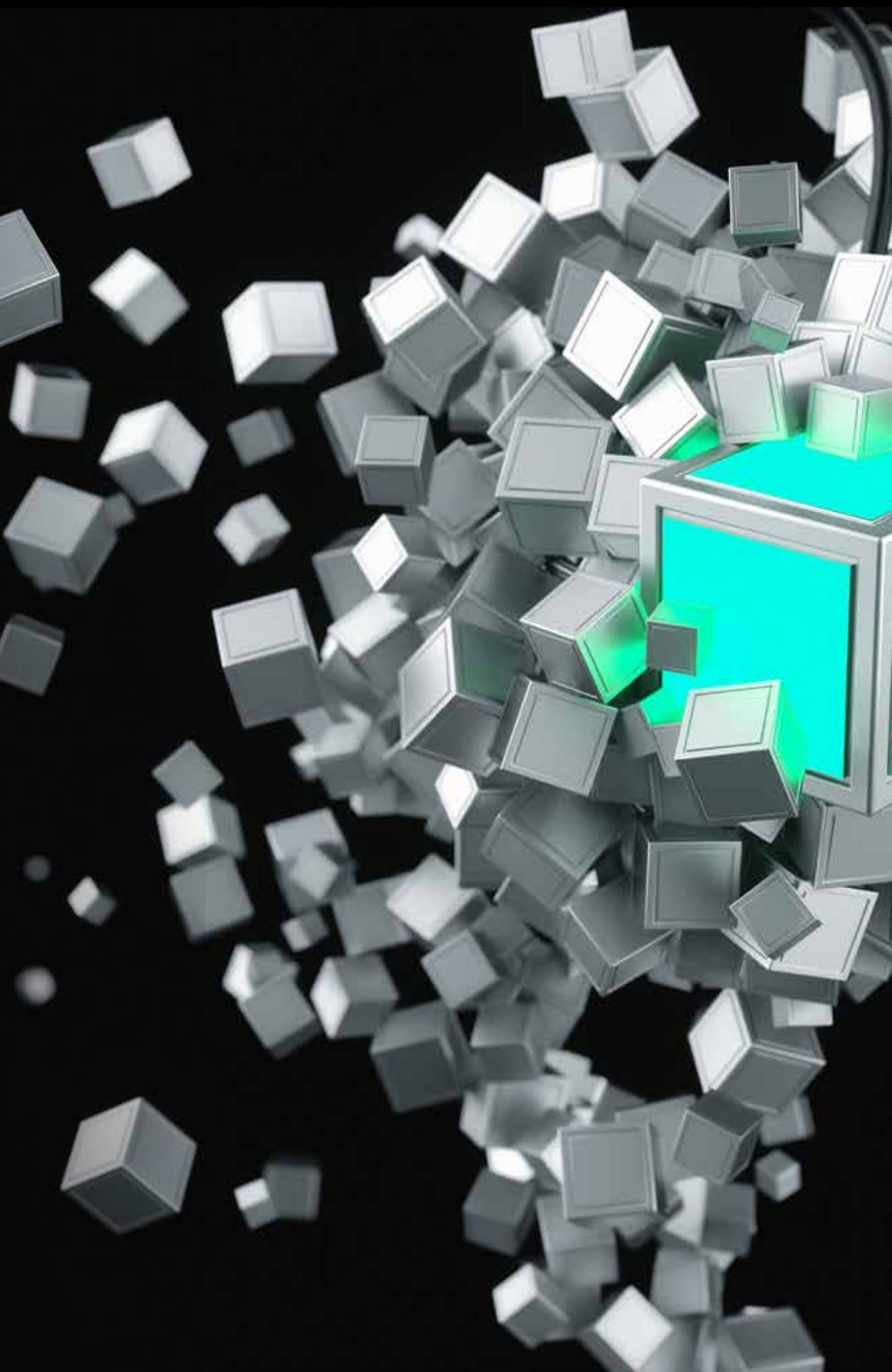
未来已来，生成式AI推动全球生产力革新与商业模式重塑

8 GenAI 将为全球各行业创造巨大价值

40 2023 年人工智能发展现状：生成式 AI 的突破之年

60 生成式 AI：CEO 必读指南

84 生成式 AI 引发科技变革：CIO 和 CTO 参考指南





GenAI将为全球各行业 创造巨大价值

GenAI将在未来发展中发挥极其重要的作用。通过提高生产效率、推动创新能力和改变行业竞争格局，GenAI将为全球经济带来巨大价值。各行各业的领先企业已经开始积极应用GenAI，并已取得显著成果。

Michael Chui, Alex Singla, Alex Sukharevsky, Bryce Hall
和Lareina Yee

数字时代人工智能正成为推动科技进步和社会变革的强大力量。智能手机技术、自动驾驶功能、乃至零售商多样的数字化工具，无一不在展示着人工智能的威力——但曾几何时，AI技术进步几乎难被公众察觉。谷歌DeepMind开发的AlphaGo在2016年击败围棋世界冠军选手时，只得到短暂关注，随后即从公众视野消失。

然而今年，ChatGPT以前所未见的方式吸引了世界关注，开启了人们的想象之门。这要归功于其广泛的实用性：这一工具具备理解自然语言并创造内容的“超能力”，几乎任何人都可以使用它。全球各类市场参与者于是开始研究并关注生成式人工智能（以下简称GenAI）对商业和社会的影响。

未来，预计这项前沿技术将为各行各业创造高达7.9万亿美元的价值。本章将深入探讨GenAI在不同领域的应用以及它所带来的积极影响。

第一部分：GenAI是技术催化剂，助力产业进步

人工智能一步一个脚印发展至今，生成式人工智能实现突破，展现出巨大潜力。

对于GenAI这个概念，人们可能感到相对新鲜。事实上，人工智能的发展经历了若干阶段，从“规则型人工智能”（rule-based AI）到“判别式人工智能”（discriminative AI），最终演变到“生成式人工智能”（generative AI，简称GenAI）。这些阶段代表了人工智能在算法、学习方法和应用领域上的不断演进和创新。

- **规则型人工智能：**上世纪60年代到80年代，AI发展处于初始阶段，这一时期的AI主要基于事先定义的规则和逻辑，通过逐步推理和匹配规则来解决问题。使用这种方法的代表是专家系统，1970年代开发的MYCIN系统便是一例，它使用了大量规则来诊

断细菌感染。然而，这些规则型系统的局限性在于其缺乏通用性和灵活性，无法适应复杂的现实问题。

- **判别式人工智能：**进入1990年代，AI技术开始关注从输入数据中学习特定模式和规律，以进行分类、识别和预测。尽管近年来深度神经网络（DNN）的出现使得判别式AI在图像识别、语音识别、自然语言处理和推荐系统等领域取得了突破性成果，但在这一发展阶段，人工智能仍缺乏生成新数据的能力。
- **生成式人工智能：**这个阶段的起点可追溯到2014年，当时生成对抗网络（GAN）的提出引发了人们对GenAI的广泛关注。GAN由生成器和判别器组成，通过对抗训练的方式生成逼真的数据样本。GAN在文本生成、图像生成和音频生成等领域取得了重大突破，使得AI能够以更具创造性的方式执行任务。

GenAI是传统AI（规则型AI以及判别式AI）的进化产物，与传统AI相比，GenAI具有以下四大核心优势：

- **自动化和效率提升：**传统AI需要经过繁琐的手工特征工程和模型调整，而GenAI则能够自动从大量数据中学习，生成高质量的输出。这使得GenAI能够自动化许多重复性和繁琐的任务，从而提高工作效率并减少人工错误。与传统AI相比，GenAI能够更快构建模型、处理数据、生成结果，可助力企业更快做出决策、提供服务。
- **个性化和定制化：**传统AI往往是基于统计分析和大规模数据的结果，缺乏对个体差异的精确处理。而GenAI通过学习每个个体的数据和反馈，能够根据个人需求和上下文生成个性化输出。这使得GenAI能够为用户提供更加定制化的体验和解决方案，满足不同用户需求，提高用户满意度和忠诚度。
- **创造性和创新能力：**传统AI主要是基于规则和已有数据进行预测和决策，局限于已知模式和规律。而GenAI具备生成新的、未

知内容和想法的能力，能够从学习到的数据中创造出全新解决方案。通过GenAI的创造性和创新能力，企业可以发现新洞察，探索新业务模式，不断创新产品和服务。

- **可解释性和透明度：**传统AI模型往往是黑盒模型，其决策过程难以被解释和理解。而GenAI在生成输出时能够提供一定程度的解释，具备一定的可解释性，可以揭示其决策逻辑和原因。这对于金融、医疗等需要可解释性的领域尤为重要，有助于建立信任、满足监管要求，也让人们更容易接受和采纳GenAI的决策。

持续创新也将带来新的挑战。训练和使用GenAI系统所需的计算能力有可能成为开发瓶颈，如何让GenAI成为“更具责任感的AI”也增加了GenAI技术的综合成本。从2017年到2022年，针对GenAI的投资以74%的年复合增长率狂飙（同期，AI总体投资的年增长率为29%）。从地理角度来看，对GenAI的投资主要来自北美科技巨头和风险投资公司。从2020年到2022年，总部位于美国的GenAI相关公司筹集总计约80亿美元资金，占同期此类公司总融资的75%。

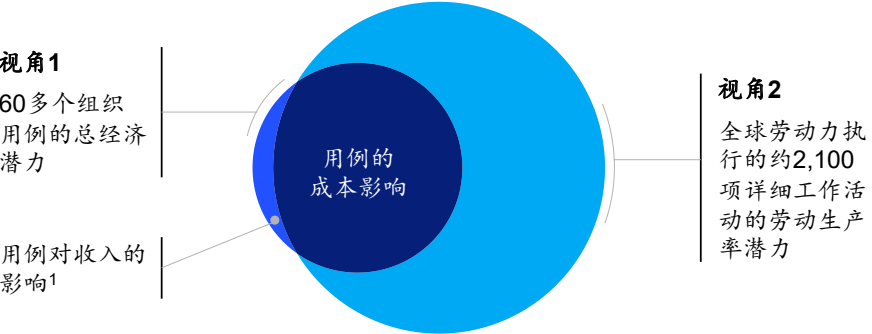
GenAI以其在全球生产活动中重塑知识工作的潜力震惊了世界。在销售和营销、客户运营和软件开发等职能部门，GenAI已显露出变革工作方式和提高绩效的能力。在这个过程中，它可以跨越多行业释放数万亿美元价值。

GenAI的兴起将推动行业格局改变，带来可观的经济增长。

不难看出，GenAI技术在提升行业生产效率和促进产品创新方面具有显著优势，预期未来将颠覆全球各行各业的现有格局。我们预测，人工智能整体将为全球经济带来高达25.6万亿美元的正面经济影响，而其中来自GenAI的贡献高达7.9万亿美元。这既包括GenAI用例本身为企业带来的直接收入增加及成本优化，也包括了GenAI推动全行业生产效率提升所带来的经济价值。

了解GenAI为经济和社会创造价值的潜力，将有助于企业制定关键决策。我们采用两个互为补充的视角，以确定当前的GenAI技术可以在哪些方面提供最大价值以及价值规模（见图1）。

图1: GenAI的潜在影响可以通过两个视角进行评估



1. 定量分析，收入影响被重新定义为相应支出的生产率增加，以保持与成本影响的可比性，而不是假设任何特定市场的额外增长。

资料来源：麦肯锡分析

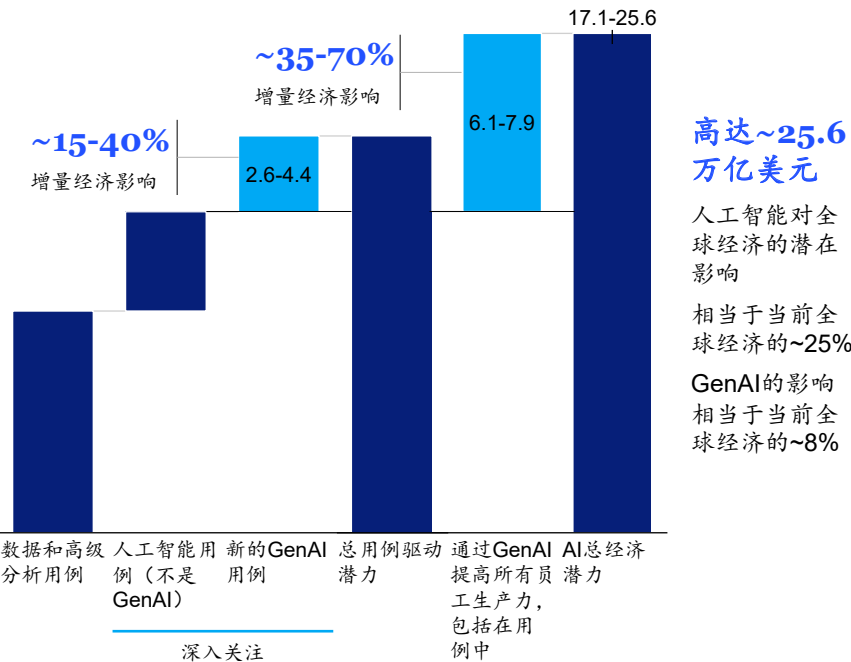
第一个视角扫描企业能够应用的GenAI用例。我们将“用例”定义为，将GenAI技术定向应用于特定业务并产生可衡量的价值。例如，营销中的典型用例是应用GenAI来生成创意内容，例如个性化的会员电子邮件等，其可衡量的价值包括降低成本以及通过大规模提升内容质量来提升转化率。我们确定了63个生成式人工智能用例，涵盖16个业务功能，当跨行业应用时，每年可提供2.6万亿至4.4万亿美元的总经济效益。

第二个视角分析GenAI对大约850个职业生产活动的潜在影响，以补充第一个视角。我们通过建模，预估GenAI在哪个时间点可以执行构成全球经济活动的2,100多个“详细生产活动”，例如“与相关人员沟通运营计划”。凭借分析结果，我们能够预估GenAI的能力可能会给全球劳动生产效率带来的影响。

两个分析视角中的部分影响相互重叠（我们暂且归纳为生产效率提升的结果），剔除这种重叠后，生成式人工智能的总经济效益（包括我们探索的主要用例以及当该技术应用于知识工作者的活动时可能实现的生产效率提高）每年达6.1万亿美元至7.9万亿美元（见图2）。

图2: 人工智能技术对全球经济的潜在影响在17.1-25.6万亿美元之间，相当于~25%的生产率增长

人工智能对全球经济的潜在影响，万亿美元



资料来源：麦肯锡全球研究院

GenAI产业自身也拥有巨大市场价值及发展潜力

GenAI当下已进入高速成长期，它的市场规模在未来将会迅速增长，投资者也纷纷入场，潜在企业及个人用户都对其功能抱有期待、兴趣浓厚，这些无不说明，GenAI行业的春天已经到来。

彭博数据显示，2022年GenAI行业的全球市场收入达400亿美元，并且以平均每年40%的增长率迅速上升。预计到2032年，这一数字将增至1.3万亿美元。目前，中国市场规模约是全球总量的1/10，预计到2025年将达到2,000亿元人民币¹，平均年增速超过60%。这一趋势表明，在为全球经济带来巨大贡献的同时，GenAI自身也拥有巨大市场潜力。

巨大增长潜力自然吸引到了全球投资者的目光。这些投资者看中GenAI的创新潜力和商业机会，将资金注入到这些初创企业中，支持它们的研发、扩展和市场推广。截至2023年第一季度，全球有近400家GenAI行业初创企业获得了私募股权或风险投资注资。这些初创企业各有所长，涵盖GenAI价值链各个环节，包括基础模型、行业模型及不同模态的具体应用（如文字生成、代码生成、视频生成等）。行业估值从2022年至今已经翻了一番，并且还呈不断增长趋势。

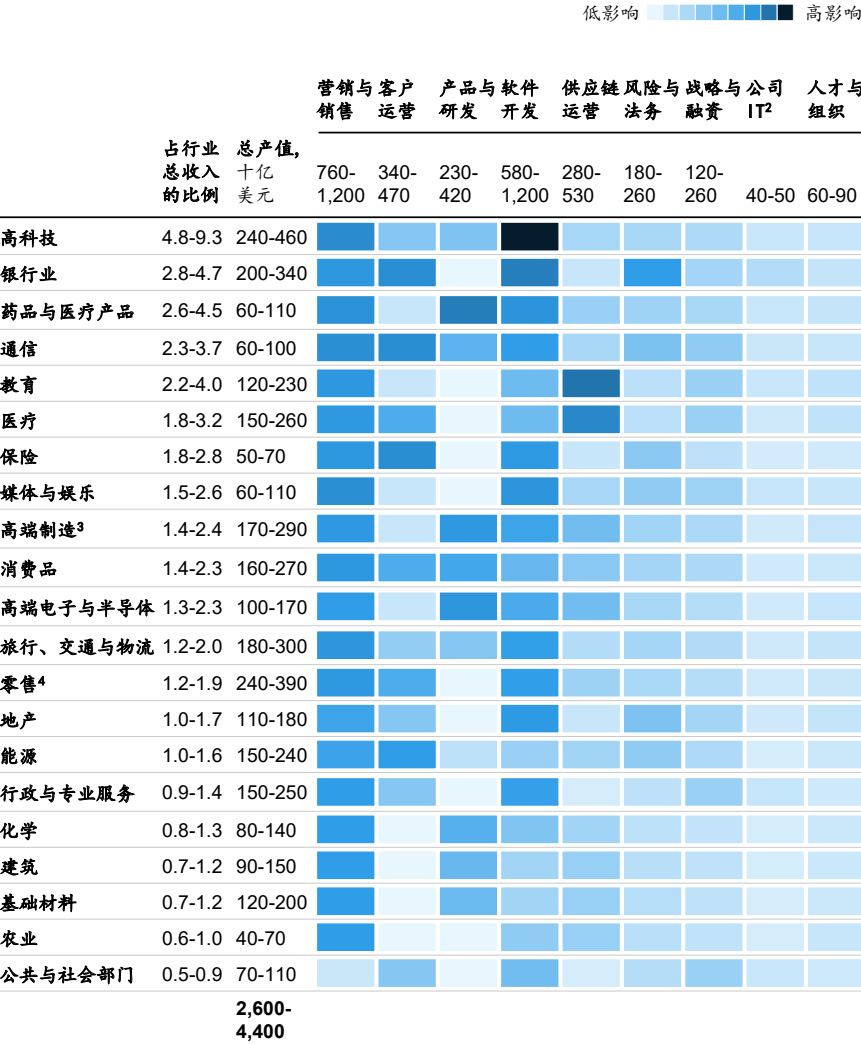
这些初创企业所推出产品和服务正在全球市场引起广泛关注和兴趣。在2022年12月，全球对于“生成式人工智能（Generative AI）”这一关键词的搜索量高达年初的8倍。OpenAI基于GenAI技术研发的聊天应用ChatGPT最为人所熟知，该应用一经推出就引起了前所未有的广泛关注，使用者数量激增，其升级版ChatGPT Plus在全球内突破性实现了大规模个人用户付费订阅的商业模式。

第二部分：GenAI在不同职能领域与行业都有广泛应用

我们将从职能领域及行业两个维度来探讨GenAI对全球产业的影响。从职能领域来看，GenAI对营销与销售、软件开发、客户运营与产品研发三大部门影响最大，其中在营销与销售领域，GenAI的价值达7,600亿至1.2万亿美元。从行业来看，GenAI产生价值最大的三个行业为高科技、零售及银行业。其中在高科技行业，GenAI的价值将达2,400亿至4,600亿美元。从所占行业收入比例来看，高科技、银行业和药品及医疗产品行业位居前三，其中在高科技行业，GenAI所带来的价值占行业收入的比例达4.8%~9.3%（见图3）。

图3: GenAI用例在不同行业和部门中具有不同规模的影响

GenAI在不同行业和部门中的产值¹



注：由于四舍五入，数字之和可能不等于100%

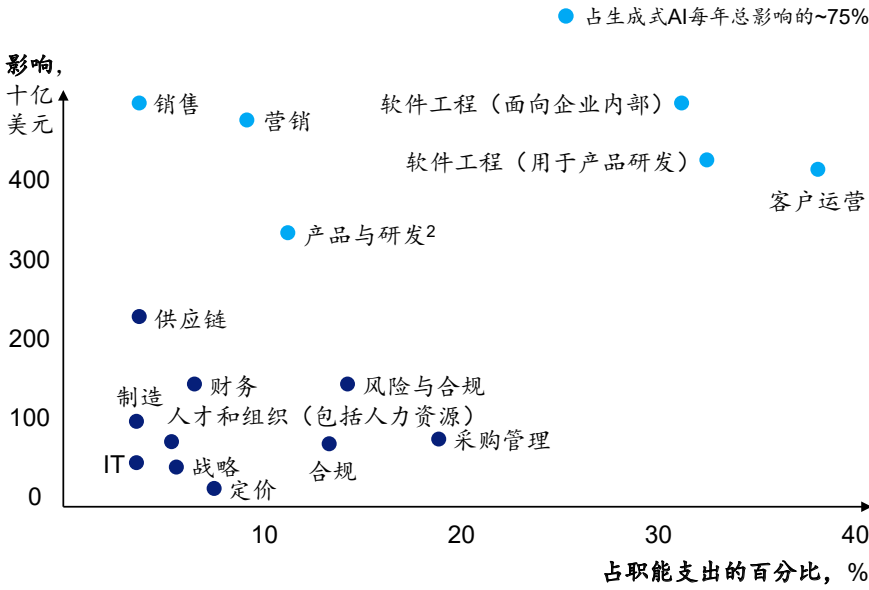
- 1. 不包括实施成本（例如培训、许可证）
- 2. 不包括软件工程
- 3. 包括航空航天、国防和汽车制造
- 4. 包括汽车零售

资料来源：比较行业服务(CIS)、IHS Markit；牛津经济杂志；麦肯锡公司和业务职能数据库；麦肯锡制造和供应链 360；麦肯锡销售导航；麦肯锡数据库Ignite；麦肯锡分析

按职能领域划分的GenAI价值潜力

GenAI能为绝大多数职能领域创造价值，不过，以价值占该职能领域支出的比例以及所创造价值量衡量，我们发现，GenAI的价值潜力在各职能的分布并不均衡。其在客户运营、营销和销售、软件工程和研发四大领域所创造的价值占到GenAI用例年度总价值的75%（见图4）。

图4：在少数职能中使用GenAI可以实现用例的大部分价值¹



1. 注意：用例价值按行业平均计算
2. 不含软件工程

资料来源： IHS Markit；牛津经济研究院；麦肯锡企业和业务职能数据库；麦肯锡制造和供应链360；麦肯锡销售导航；Ignite，麦肯锡数据库；麦肯锡分析

下面我们将以客户运营、营销和销售、软件工程和研发四大领域为例，说明GenAI如何在不同领域产生价值。

客户运营

GenAI通过自助服务可大大提升客户体验和客服效率，并能提升人工客服的服务技能。研究发现，在一家拥有5,000名客服人员的公司中，应用生成式AI使每小时的问题解决率提高了14%，处理问题所花时间缩短了9%，与经理对话的请求数量减少了25%。

我们预计，在客户运营领域应用GenAI将可以带来相当于该领域成本30%~45%的价值。

GenAI对客户运营的影响



自动化客服

客户与类似人类的聊天机器人进行交互，该机器人对复杂的查询提供即时、个性化的响应，确保无论客户语言或位置如何，都能获得一致的品牌声音

辅助人工互动

人工客服使用人工智能开发的呼叫脚本，并在电话交谈期间接收实时帮助和响应建议，即时访问相关客户数据以进行量身定制的实时信息传递



摘要与建议生成

人工客服将收到简洁的对话摘要，用以记录客户投诉和所采取行动

人工客服将使用由AI生成的自动化、个性化的见解，包括量身定制的后续消息或个性化的指导建议

以下是GenAI在不同场景下的用例：

- 快速响应客户咨询：GenAI驱动的聊天机器人可以对复杂的客户查询提供个性化的即时响应，大大减少人工客服团队需要处理的客户任务量。我们预计生成式人工智能可将人工服务的任务量减少50%，同时还能提升客户体验。
- 通过大量数据训练，优化客户服务能力：GenAI可以基于过往客户互动数据，快速理解不同客户群体的需求，从而做到从一开始便能更有针对性地解答客户疑问、提供优质服务，这有助于提升客户转化率。
- 捕捉客户数据，提升客户价值：GenAI能够快速捕获及分析客户行为数据，据此识别客户可能感兴趣的产品并进行精准推送，从而提升客户转化与复购率。

营销和销售

GenAI在营销和销售领域产生的价值可分别占各自领域全球支出的5%~15%和3%~5%。我们将以如下应用场景为例，探讨GenAI如何在营销与销售领域创造价值。

- 高效创建个性化内容，显著增加客户转化率、留存率和客户价值。
 - 营销内容个性化：企业应用GenAI，可向不同画像客户群发送更为个性化的营销信息。例如，群发的营销邮件可根据实际需要被自动翻译成多种语言，还可面向不同文化背景和地理区域的受众，有针对性地使用差异化图像，采用更适合受众的风格。
 - 产品推送千人千面：GenAI可依据文本、图像和语音格式的信息输入及对客户资料的深刻理解，生成个性化的产品描述，增强产品介绍网页对客户吸引力，提升转化率和销售额。

GenAI对营销和销售的影响



战略制定

销售和营销人员从非结构化数据源（例如，社交媒体、新闻、研究报告、产品信息和客户反馈）中收集市场趋势和客户信息，并起草营销和销售内容

建立认知

客户会看到根据其画像量身定制的广告系列

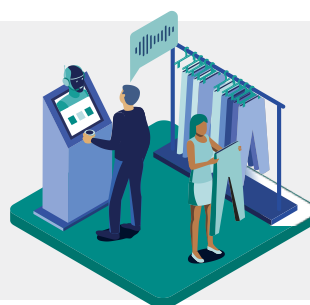


购买决策

客户可以访问全面的产品信息和动态建议，例如进行“试穿”

交易转化

由GenAI支持的虚拟销售代表模拟人类品质（例如同理心、个性化沟通和自然语言处理），以与客户建立信任和融洽关系



客户留存

更有可能通过定制的消息和奖励来留存，他们可以与人工智能驱动的聊天机器人进行交互，这些聊天机器人主动管理关系，减少人工客服需要处理的对话

- 提升搜索引擎优化 (SEO) 结果。

GenAI可帮助优化网页标题、图像标签和URL等数字营销要素的搜索引擎结果,支持SEO数字内容的创建,帮助营销人员将内容准确分发给目标客户,实现精准营销,降低广告成本。

- 提升销售效率。

基于客户数据库信息, GenAI可以帮助销售人员快速理解特定客户需求,使之能在初期接触时更有针对性地解答客户疑问、提供优质服务,从而能有效提升客户转化率。此外, GenAI能够基于客户与产品信息创建对话脚本,推动向上销售和交叉销售。GenAI还可以自动跟进与挖掘潜在客户价值,促成客户与人工销售代理的直接互动。而且在销售团队的客户外展中, GenAI亦可提供更多个性化的信息。

产品研发

GenAI能够快速生成候选设计、提升效率,还可以改进设计本身。所创造价值占整体研发成本的比例达10%~15%。GenAI在产品研发的应用示例如下:

- 快速生成候选产品设计,加速研发进程。

GenAI模型可以基于过往相关产品的特征生成初步设计(如候选药物分子、电路等),从而缩短产品研发到上市所需时间。但其局限性在于,产品设计模型具有较强的行业特征,因而GenAI在不同行业中的应用互通性较低。

- 优化产品设计。

GenAI可以通过基于对类似产品数据的学习与分析,协助设计师更有效地选择和使用材料,从而可降低包装、物流和生产等成本。

- 提升产品测试效率、改进质量。

GenAI的深度学习能力能有效优化产品模拟与测试阶段的系统虚拟仿真能力,帮助缩短复杂系统的测试时间。

GenAI对产品研发的影响



早期研究分析

研究人员使用GenAI来协助市场报告、产品或解决方案的起草

虚拟设计

研究人员使用GenAI生成基于提示的草稿和设计，使他们能够使用更多设计选项快速迭代

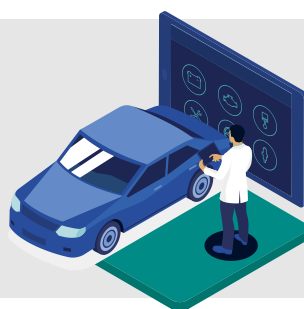


虚拟模拟

如果与新的深度学习生成式设计技术相结合，研究人员可以加速和优化虚拟仿真阶段

产品测试

研究人员优化测试用例以实现更高效的测试，从而减少物理构建和测试所需的时间



软件工程

GenAI为软件工程开辟了新的可能性，在收到软件工程师以自然语言提出的代码描述后，GenAI即可协助生成代码。根据麦肯锡分析，GenAI对软件工程生产效率的直接影响可能占当前该领域年度支出的20%~45%。

研究发现，软件开发人员使用Microsoft GitHub Copilot后，完成任务的速度比不使用该工具的软件开发人员快56%²。同时，麦肯锡内部对软件工程团队的实证研究发现，接受过GenAI工具培训的员工表示自己拥有更好的工作体验，幸福感、沉浸感和成就感均有所提升。

GenAI在软件工程领域的应用如下：

- 自动生成代码。

GenAI可以协助软件工程师生成初始代码草稿、进行代码更正和重构及根本原因分析并生成新的系统设计。通过加速代码生成，软件工程师可集中更多时间用于架构设计，促进软件功能迭代与优化。

- 赋能软件创新。

GenAI能够基于对现存代码和算法的学习与分析，生成创新性代码与算法解决方案，在全自动模式下创造出新的特性、功能、甚至新的应用程序，赋能软件创新。

GenAI对软件开发的影响



启动和规划

软件工程师和产品经理使用GenAI来协助分析、清理和标记大量数据，例如用户反馈、市场趋势和现有系统日志

系统设计

工程师使用GenAI来创建多个IT架构设计并迭代潜在的配置，从而加速系统设计并缩短上市时间



代码撰写

工程师可以借助AI进行代码编写，通过协助生成草稿、快速查找错误及充当易于导航的知识库来缩短开发时间

测试

通过采用可以增强测试功能的算法来提升测试质量，并可以自动生成测试用例和测试数据



维护

工程师使用有关系统日志、用户反馈和性能数据的AI洞察来帮助诊断问题，提供修复建议，并预测其他高优先级的改进领域

按行业划分的GenAI价值潜力

在我们分析的63个用例中，GenAI将在不同行业产生2.6万亿至4.4万亿美元的价值。

麦肯锡分析显示，GenAI可以通过提升营销、客户互动等，为零售业贡献大约3,100亿美元价值。在银行业，GenAI可通过承担客服、代码生成等任务，进一步提升员工工作效率。在生命科学行业，GenAI有望为药物发现和开发做出重大贡献（见图5）。

图5: GenAI部署于部分行业的用例中时，可以提供巨大的价值

职能性价值驱动因素关键用例的部分示例（非穷尽）

		该职能对行业的价值潜力 高 </				
--	--	---	--	--	--	--

资料来源：麦肯锡分析

下面我们将分别探讨GenAI在零售和消费品行业、银行业、制药和医疗产品行业中的应用。

零售和消费品行业

GenAI将改变零售和消费品公司的游戏规则，为零售和消费品行业创造4,000亿至6,600亿美元的价值，价值潜力占行业总收入的1.2%~2.0%。

GenAI可通过以下方式为零售和消费品行业创造价值：

1. 重塑客户互动模式

GenAI为零售和消费品公司提供了许多交叉销售和追加销售的机会，还能帮助企业捕获关键信息用于改进产品。此外，GenAI在助力企业扩大客户群、增加收入机会、提升整体营销投资回报率方面也大有可为。

如今，无论是选购服装、化妆品，还是购物体验本身，消费者越来越多地寻求方方面面的个性化体验，而GenAI可以改善这种体验。聊天机器人可以通过个性化营销活动来增强客户价值管理，从而提高客户的满意度和品牌忠诚度。

2. 加速关键领域价值创造

GenAI能助力营销和销售文案写作，帮助打开思路、提出创意营销理念，减少消费者研究过程中的人工投入，并加速内容分析和创建。文字和视觉效果的改进可以扩大产品知名度、提高销售转化率。

3. 快速解决客户问题，加深了解客户需求

电商时代，企业与消费者的有效互动格外重要。零售商可将现有的人工智能工具与生成式人工智能相结合，以增强聊天机器人能力，使之更接近人工客服。GenAI聊天机器人可以直接响应客户的各种要求，如查询、跟踪或取消订单、要求打折和追加销售等。将此类任务交由GenAI进行自动化处理，有助于将人工客服解放出来去处理更复杂的客户问题。

4. 助力设计创新

GenAI设计工具发展极快,可加速数字化设计过程,从而加快产品的迭代进程。设计师使用GenAI工具可以从头开始生成包装设计,也可以对现有设计进行修改。



银行业

GenAI将为银行业带来2000亿至3400亿美元的价值,相当于行业年收入的2.8%~4.7%。此外,生成式人工智能工具还可帮助各银行增强客户满意度,改善决策和员工体验,并能更好地监控欺诈和风险,从而有效降低风险。

银行业的如下特征也为GenAI在银行业大显身手提供了舞台:

- 庞杂的IT架构。几十年来,银行一直在投资技术,因而积累了大量的遗留系统,IT架构各自孤立且繁复庞杂。
- 庞大的客户服务团队。银行业依赖大量业务服务代表,如呼叫中心客服人员、财富管理顾问等。
- 严格的监管环境。作为一个受到严格监管的行业,银行业有大量的风险、合规和法律需求。
- 繁重的文件工作。GenAI的影响可以跨越整个组织,协助所有员工编写电子邮件、创建业务演示文稿和完成其他任务。

GenAI可通过以下方式为银行业创造价值：

1. 作为虚拟客服，降本增效。

接受过政策、研究和客户互动等专有知识训练的GenAI机器人可以提供全天候的支持。这种面向客户的聊天机器人可以评估用户请求，并根据主题、难度和客户类型等特征选择最佳服务专家来解决这些问题。通过GenAI，专业服务人员可以快速访问所有相关信息，例如产品指南和政策，以即时满足客户请求。

例如，摩根士丹利正在使用GPT-4构建人工智能助手，该工具旨在帮助该行数万名财富经理从庞大的内部知识库中快速找到信息、整理出答案。该模型结合了搜索和内容创建功能，财富经理可以随时利用它为任何客户查找信息，提供定制内容。

2. 加速代码生成，更快交付软件

GenAI工具可通过四大应用加速代码生成和软件交付。

首先，GenAI工具可以通过所输入代码或自然语言根据上下文起草代码，帮助开发人员更快地编写代码并减少摩擦，同时启用自动翻译以及无代码和低代码工具。

其次，这些工具可以自动生成代码测试、确定其优先级、并运行和审查不同的代码测试，从而加速测试过程、扩大覆盖范围并提高有效性。

第三，生成式人工智能的自然语言翻译能力可以优化遗留框架的集成和迁移。

最后，这些工具可以检查代码以识别计算中的缺陷和低效环节，输入更稳健有效的代码。

3. 大规模制作个性化内容

生成式AI工具可以利用现有文档和数据集简化内容生成。这些工具可以根据客户的档案和历史,创建个性化的营销和销售内容,还能提供多种替代方案用于A/B测试。此外,生成式人工智能可以自动生成模型文档,识别缺失的文档并扫描相关法规更新,为内容变化创建提醒。

生命科学行业

我们的分析发现, GenAI将在生命科学行业产生600亿至1,100亿美元的价值,相当于该行业年收入的2.6%~4.5%。

GenAI在生命科学领域的用例覆盖从药物发现、药物实验到患者治疗、商业化推广的价值链全流程。GenAI可借助图像或自然语言等模型为研发人员提供更具有潜力的药物分子,从而缩短药物前期研发和实验时间。同时GenAI还能为医疗机构和患者提供更精准的治疗方案和更高效的药物。

- 在药物发现与早期开发阶段, GenAI为研发团队提供更精准的药物研发工具,提升新药研发效率,缩短研发所需时间。具体而言, GenAI能够通过建立更准确的筛选标准,提升目标分子识别精度,从而缩减每个阶段的分子实验规模,快速推进药物研发进程。同时, GenAI还能够基于患者病程大数据信息,为新药的适应症搜寻提供更科学的建议。
- 在药物试验阶段, GenAI基于受训的患者数据,为研发人员提供不同患者群体所需的特定试验设计信息,从而有助于试验人员更科学地进行试验设计。同时GenAI还能被用来对患者进行数字化预试验,方便研发人员根据预试验的临床结果快速迭代试验设计。在试验过程中, GenAI则以聊天机器人的形式,为试验患者解答相关问题并记录相关数据,提升患者管理的自动化程度。试验结束时, GenAI能自动生成相关报告以及洞察分析,大

大节省试验结果分析总结用时。

- 在患者治疗阶段，对研发机构或药企而言，GenAI能够大规模筛选和分析药物的真实世界数据，并评估药物对不同患者的实际效果，为研发提供反馈。对医疗专业人员而言，GenAI能够基于患者病史和电子病历数据产生对于临床决策的建议，同时搜索最新文献信息，帮助医疗专业人员了解最新科研结果。
- 在商业化阶段，GenAI能以聊天机器人的形式实时回复销售代表或医生的药物相关问题，并基于群体偏好自动生成患者教育内容和医生互动内容，提升药品推广效率。

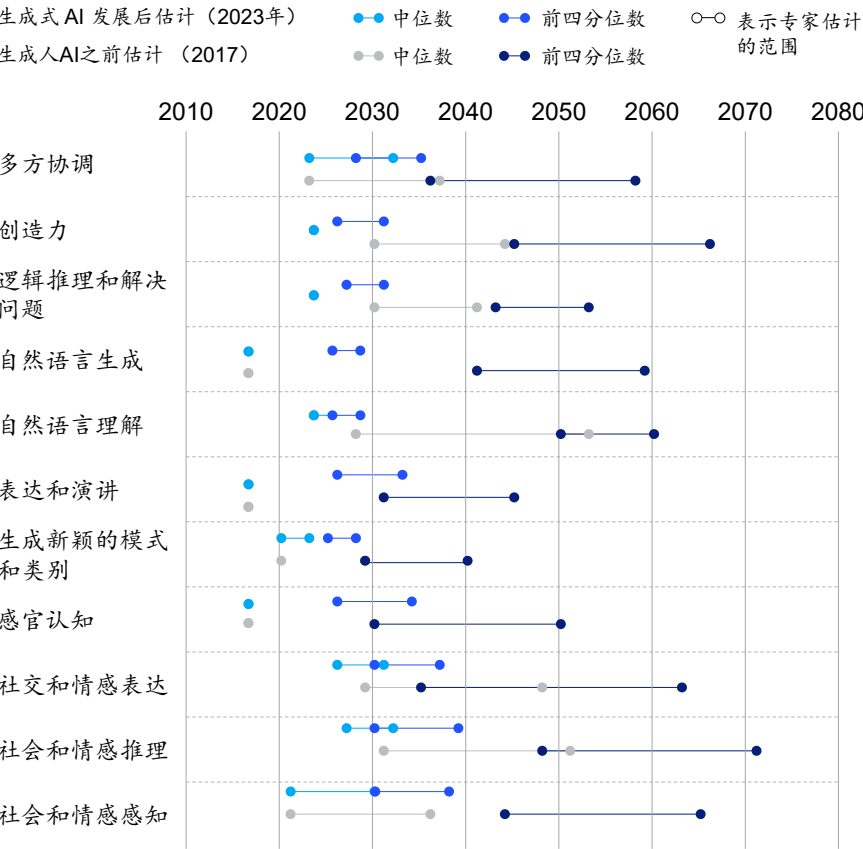


第三部分: GenAI将给全球产业的效率、创新及竞争格局带来巨大影响

GenAI的发展是科技进步的又一次飞跃。在GenAI的推动下，科技在诸多工作上有望实现相当于人类中位数水平的表现，且发展速度在加快，科技的各项能力表现达到人类工作表现前25%水准的时间比先前预估已有所提前（见图6）。例如，麦肯锡全球研究院之前的分析认为，科技在自然语言理解方面最早会在2027年到达人类中位数水准，但根据最新分析，这一时间已提前到2023年。

图6: 科技能力，科技可实现的人类表现水平

职能性价值驱动因素关键用例的部分示例（非穷尽）



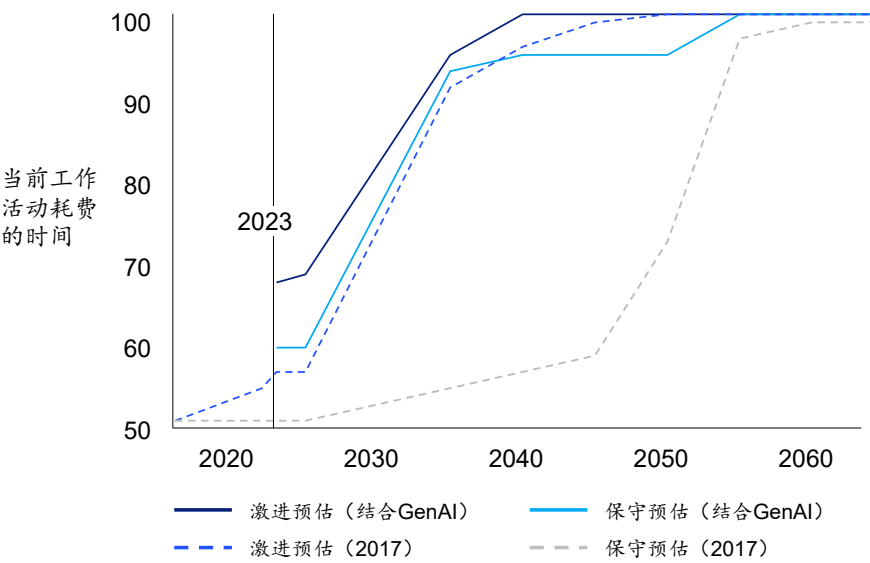
资料来源：麦肯锡分析

整体而言，将GenAI与现有技术结合，可通过三种方式对全球产业产生巨大影响：提高企业生产效率、推动企业产品创新以及改变行业竞争格局。

提高生产效率

生成式AI的出现，可大幅提升各项工作的自动化比例。其自然语言能力的加速发展，可推动技术潜力飞速跃升。据麦肯锡分析，若在现有技术中使用GenAI，则理论上可实现自动化的工作量（按小时数衡量）百分比将从50%左右增至60%~70%。按照最乐观情景估计，最快到2040年，企业即可实现100%的手工作业替代（见图7）。

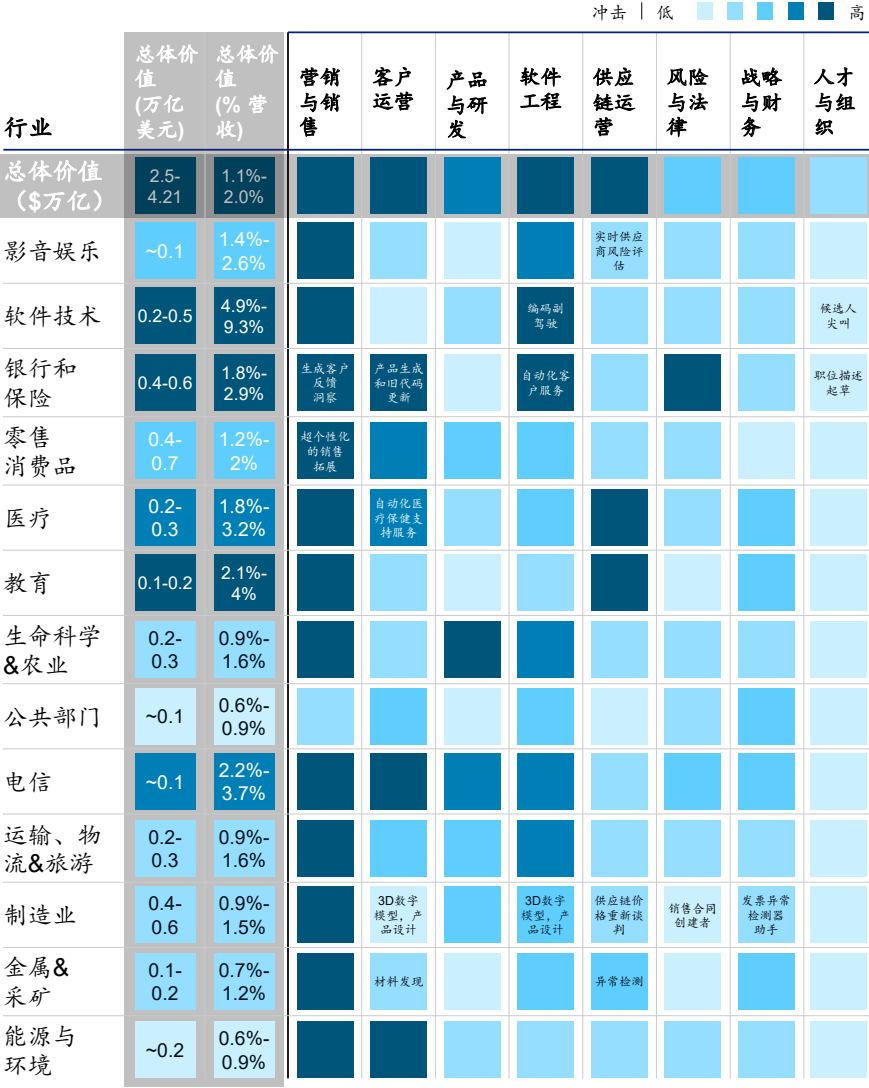
图7: 生成式人工智能的出现推动了技术自动化的潜力
按不同预估划分的技术自动化潜力，%



资料来源：麦肯锡分析

GenAI主要通过优化现有流程和任务的方式提升企业生产效率；流程优化的最终目标是提高日常工作效率、降低成本、提升生产能力。GenAI能够通过自动化和智能化的方式减少人力投入、优化资源配置、加速工作流程。根据麦肯锡估算，仅从提高企业生产效率的角度衡量，GenAI就将为世界经济带来高达2.5万亿至4.2万亿美元的贡献，相当于整体营收增长1.1%~2.0%（见图8）。大多数行业都能从中受益。从各职能领域来看，营销和客户运营的生产效率提升幅度最大，这也是GenAI能在实践中最早产生价值的两个领域。而且，大部分此类GenAI应用，目前市场上已有成熟工具，企业可直接使用，资源投入较低。

图8：效率提升：GenAI将影响各个职能部门，对营销、客户运营和工程的影响最大

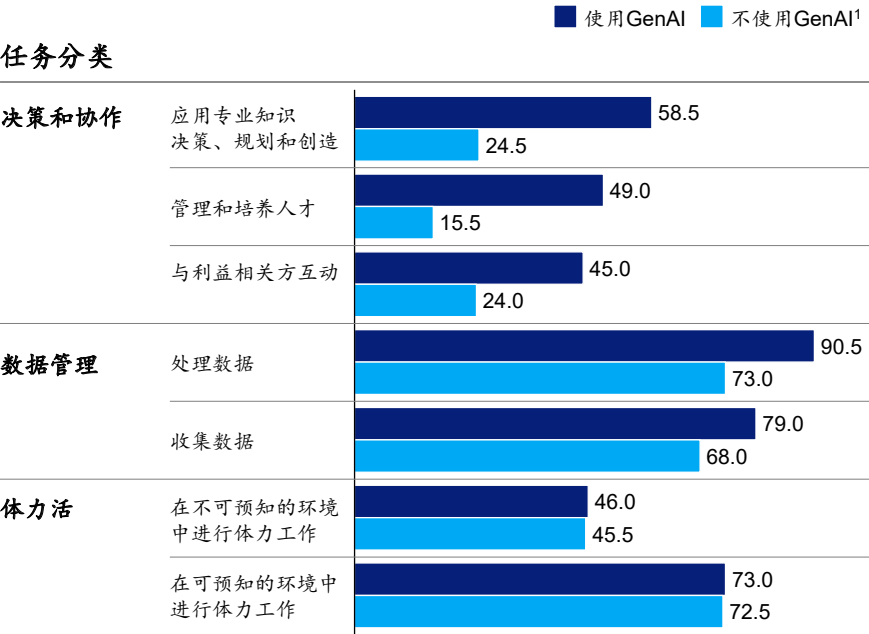


资料来源：麦肯锡分析

推动产品创新

GenAI不仅可助力企业提高生产效率，还能够推动企业的产品创新。这是因为，以完成认知类任务为目标而设计的GenAI能够影响知识类工作，特别是涉及决策和协作的任务。而在以往的技术变革中，这类任务的自动化潜力最低（见图9）。在我们的分析中，若将GenAI应用于专识类任务，则实现自动化的技术潜力比值会比不使用GenAI有明显提升，高出34个百分点。

图9: GenAI可对协作和专识应用类任务等早先自动化潜力较
低的领域产生最大影响



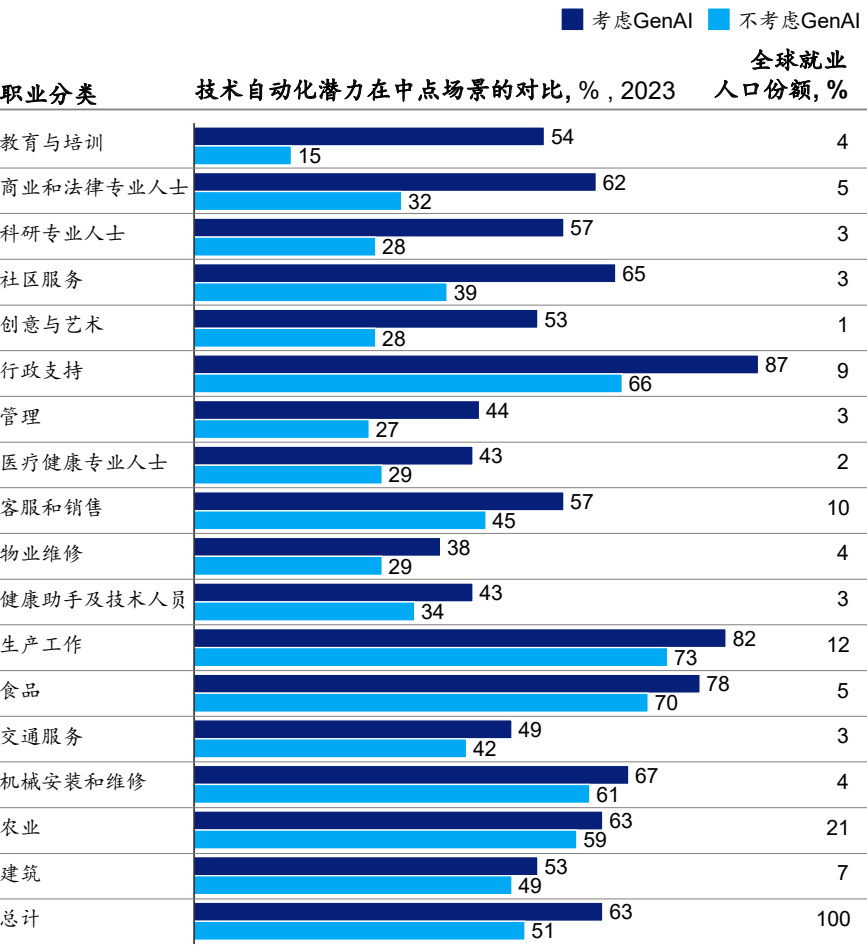
注：由于四舍五入，数字可能不能加总
资料来源：麦肯锡全球研究院分析

那些对创新要求较高的工作（其中40%的工作需要具备至少中等水平的人类自然语言理解能力才能完成），在GenAI技术的助力下，也能实现自动化潜力的显著提升。这意味着，诸多涉及沟通、文档和人际互动的工作都有可能被GenAI自动化，这将加速教育、科研、创意和软件开发等职业转型，而这一前景预测是在GenAI诞生之前未曾有过的（见图10）。

若要有效推动产品创新，则需要所应用的技术具备创造全新产品、服务或改进现有产品的能力，能够制造出可带来显著竞争优势的应用、或是能整合积淀公司数据的产品。GenAI具有学习、分析和生成能力，可通过对海量数据的深入学习和不断迭代，识别数据中的模式和趋势，从而引领产品和服务创新，为企业提供新颖的解决方案和设计。

图10: GenAI带动的技术进步将加速教育、科研、创意等职业转型

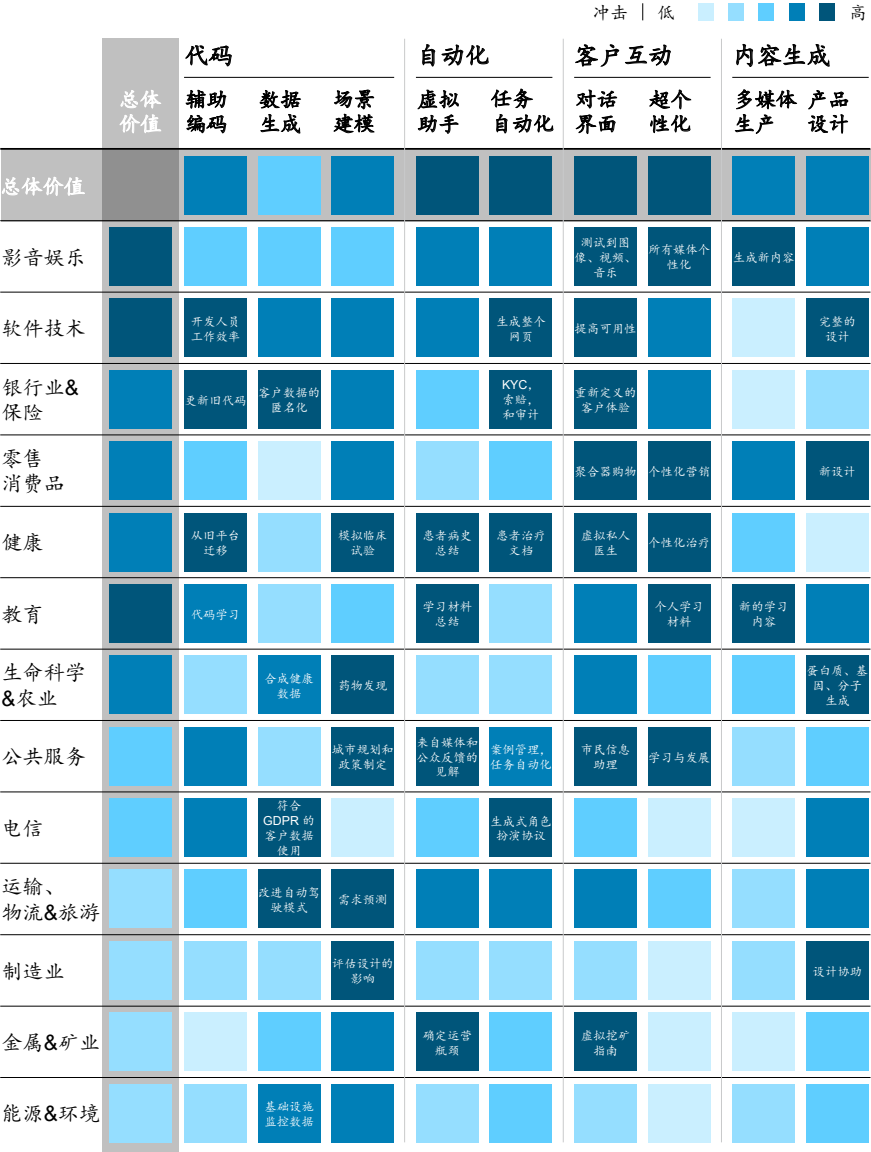
生成式AI对技术自动化潜力在重点场景的潜在影响，2023



注：由于四舍五入，数字可能不能加总
包括来自 47 个国家的数据，约占全球就业人数的 80%
资料来源：麦肯锡全球研究院分析

能够从GenAI应用的这一能力中受益最深的行业当属影音娱乐、软件开发与教育（见图11）。例如，GenAI能协助影音娱乐企业生成突破想象力的多媒体内容，软件开发企业可利用GenAI设计全新的产品形态，学校则能够通过GenAI针对每个学生特点设计作业及课后辅导。

图11: 产品创新：我们预计软件和技术以及媒体和娱乐领域的影响最大



1. 每个行业都包含行业特定的软件 2.与行业无关的软件 3.包括航空航天和国防、汽车和装配、化学品、半导体、基础材料 4.包括石油和天然气、电力

资料来源：麦肯锡分析

不过，产品创新类的应用必须基于GenAI基础模型和企业数据进行定制化开发，因而更为复杂，没有现成解决方案可用，企业需要投入更多资源才能从GenAI技术上获得价值。

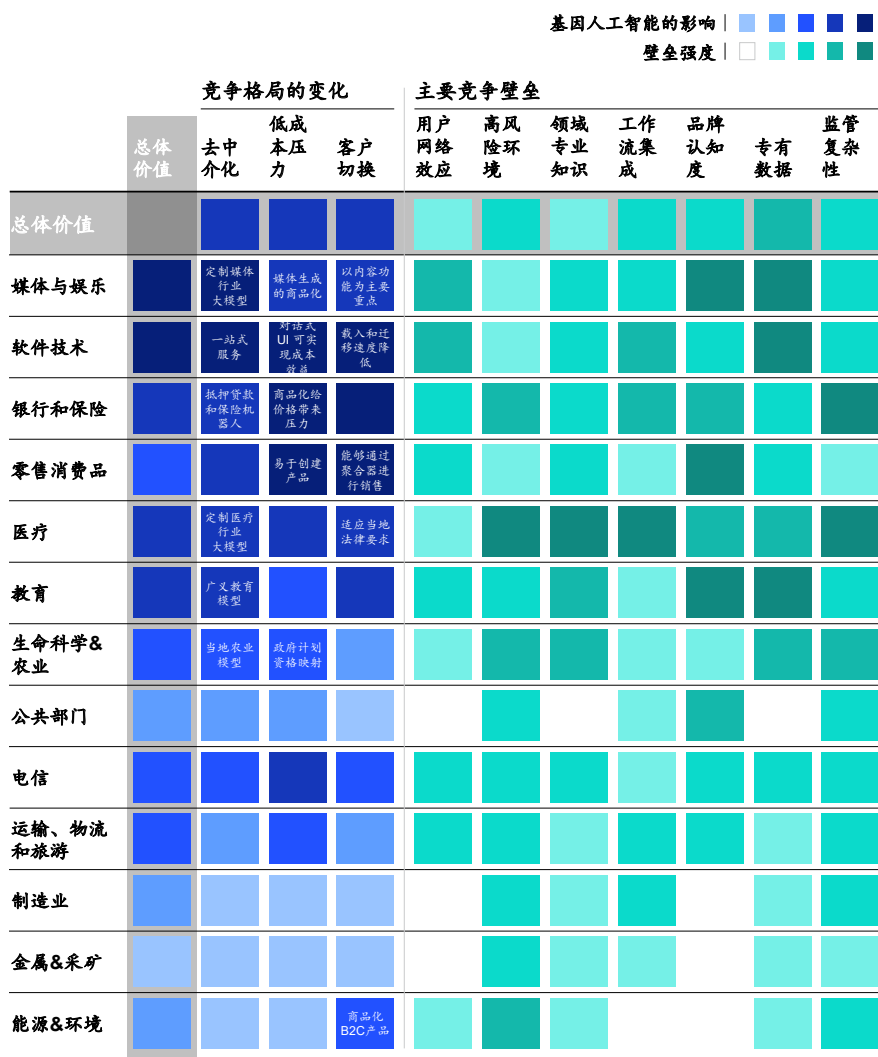
改变行业竞争格局

GenAI的应用不仅能提高企业生产效率和创新能力，还将改变行业的竞争格局。新技术能有效降低产品和服务成本，降低客户切换产品的前期投入，这会对行业现有参与者和竞争者产生巨大冲击。企业可通过以下三种方式显著提升竞争力：

- 变革客服形态。企业可借助GenAI全面升级与客户的交互方式。例如企业可启用基于GenAI的虚拟助手，打造全新的一站式解决方案，通过削减客户获取信息的中间环节，简化价值链。
- 严格管控成本。企业可通过彻底改革内部操作，引入新的生产模式，从而实现更高效的运作，构建竞争对手无法企及的成本优势。
- 降低客户迁移成本。利用GenAI，企业可大大降低客户从竞争对手迁移的成本，提升客户的迁移意愿和能力，从而改变原有行业的竞争格局。例如企业可大幅降低重构代码的耗时和成本，为客户带来全新的自动化体验。

整体来看，未来软件开发、传媒行业受到的冲击可能最大。以软件开发行业为例，GenAI将通过整合开发环境价值链、降低软件开发专业门槛以及提升开发环境兼容度，驱动行业格局转型。身处这些行业的企业若想拔得头筹，则需提早行动，投入大量资源构建新的竞争优势。而竞争壁垒较高的医药健康等行业，短期内受冲击程度可能较小（见图12）。

图12：竞争变化：我们预计对转换成本、专有内容和成本优势的影响最大



第四部分：企业和社会应用GenAI时的注意事项

事物皆有两面性，这项技术在带来巨大价值潜力的同时，也会带来新的重大挑战。鉴于GenAI推广应用速度之快，相关各方应当立即行动起来，在拥抱机会的同时，做好防范新风险、迎接新挑战的准备。

如今部分风险已经显现，比如，人们会对GenAI系统所产出的内容产生如下担忧：它们是否会因训练数据中存在“抄袭”而侵犯知识产权？模型给出的答案是否准确无误、是否具有可解释性？由GenAI生成的内容是否公平、无偏见，是否会有伤人的偏见？经济方面也面临挑战。未来十年，将有大量工作因GenAI而发生变化。

当前企业和社会的任务是妥善处理好这项技术的积极潜力与潜在破坏力。我们的热情既要用来挖掘新技术的潜力，也不要忘记积极面对可能出现的新挑战。为此，我们需要解决以下一些关键问题。

公司和商业领袖需要快速行动，抓住生成式人工智能可能带来的潜在价值，同时管理好相关风险：

- 思考GenAI及其他AI技术未来几年如何改变员工职业和所需技能，研究并制定招聘与再培训计划及人力资源的其他策略，帮助员工顺利转型。
- 制定策略确保GenAI不会被部署在可能危害社会的负面场景中。

从业者、消费者和其他公民需要关注GenAI技术发展，理解这项技术如何影响自身生活生计：

- 思考如何从自身出发，平衡GenAI带来的便利与其对经济环境和工作的负面影响。
- 在GenAI融入普通人生活的过程中，积极发声、参与决策。

结语

不难看到，作为一项前沿技术，GenAI将在未来发展中发挥极其重要的作用。通过提高生产效率、推动创新能力和改变行业竞争格局，GenAI将为全球经济带来巨大价值。各行各业的领先企业已经开始积极应用GenAI，并已取得显著成果。随着技术的进一步发展创新，GenAI将持续为全球带来更多价值和机遇，创造更美好的未来。

¹“数据来源于《中国AI数字商业展望2021~2025》，前瞻产业研究院整理。其GenAI应用规模统计口径为应用GenAI技术生成的数字内容的市场规模，统计方式为数字内容市场规模乘以GenAI渗透率，其规模的推导，综合考虑了国家商务局，Gartner，第三方调研机构，专家访谈等多个数据源。

²数据来源：Peter Cihon等人所发布论文《The impact of AI on developer productivity: Evidence from GitHub Copilot》缘政治韧性》(How to build geopolitical resilience amid a fragmenting global order)，麦肯锡公司，2022年9月。

Michael Chui是麦肯锡公司及麦肯锡全球研究院全球董事合伙人，常驻湾区分公司；

Eric Hazan是麦肯锡全球资深董事合伙人，常驻巴黎分公司；

Roger Roberts是麦肯锡全球董事合伙人，常驻湾区分公司；

Alex Singla是麦肯锡全球资深董事合伙人，常驻芝加哥分公司；

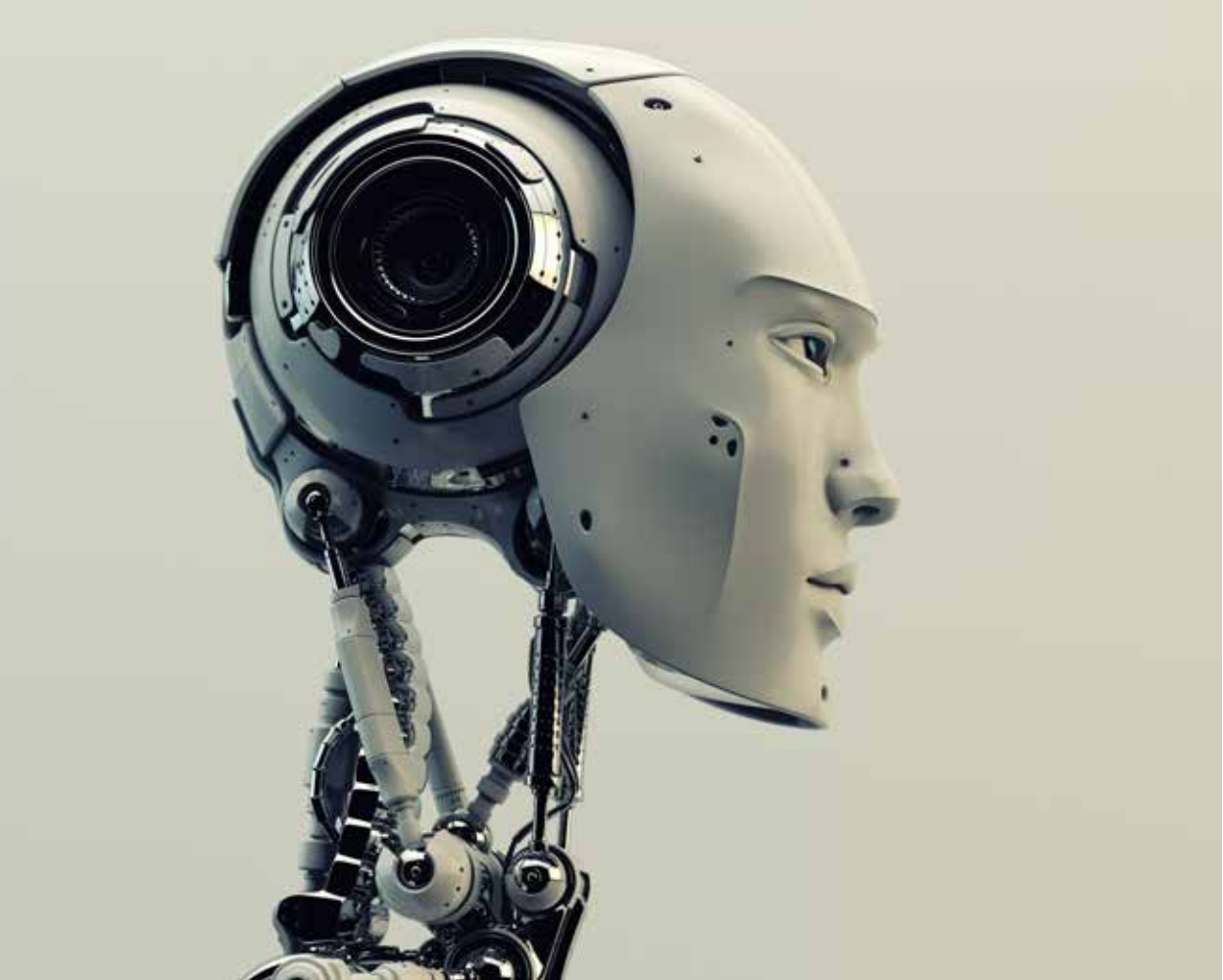
Kate Smaje和**Alex Sukharevsky**均是麦肯锡全球资深董事合伙人，常驻伦敦分公司；

Lareina Yee是麦肯锡全球资深董事合伙人，常驻湾区分公司；

Rodney Zimmel是麦肯锡全球资深董事合伙人，常驻纽约分公司。

感谢曲向军、韩峰、胡艺蓉、李静瑶、宋戈、王喆宸和邱外山对此文的贡献。

麦肯锡公司2023年版权所有。



2023年人工智能发展现状： 生成式AI的突破之年

如今，生成式AI工具遍地开花，各组织均在快速部署；麦肯锡调查的受访者们预计，该技术将对自己所在行业及就业产生重大影响。

Michael Chui, Eric Hazan, Lareina Yee, Bryce Hall, Alex Singla
和Alexander Sukharevsky

麦肯锡针对人工智能发展现状的最新年度全球调研结果证实，生成式人工智能（简称GenAI）工具已出现爆炸式增长。许多此类工具至今推出尚不满一年，但已有1/3的受访者表示，其所在组织会在至少一项业务职能中经常使用GenAI。随着这些最新进展，人工智能已经从一个技术话题上升为企业领导的关注焦点：近1/4受访高管表示，他们会在工作中使用GenAI工具；而在已应用人工智能的企业中，有超过1/4的受访者表示GenAI已被列入董事会议程。此外，40%的受访者表示，其所在组织将会因GenAI的最新进展而增加对人工智能的整体投入。调查结果表明，GenAI相关风险管理仍处于早期阶段：即便是针对受访者眼中最常见的不准确问题，也只有不到一半受访者表示，所在企业正在采取措施降低风险。

原本已在自身系统内嵌入人工智能能力的组织，也在积极探索GenAI的潜力；而那些最看好传统AI价值潜力的组织（我们称之为人工智能高绩效组织）在GenAI工具应用上同样当仁不让，走在了其他企业前面¹。

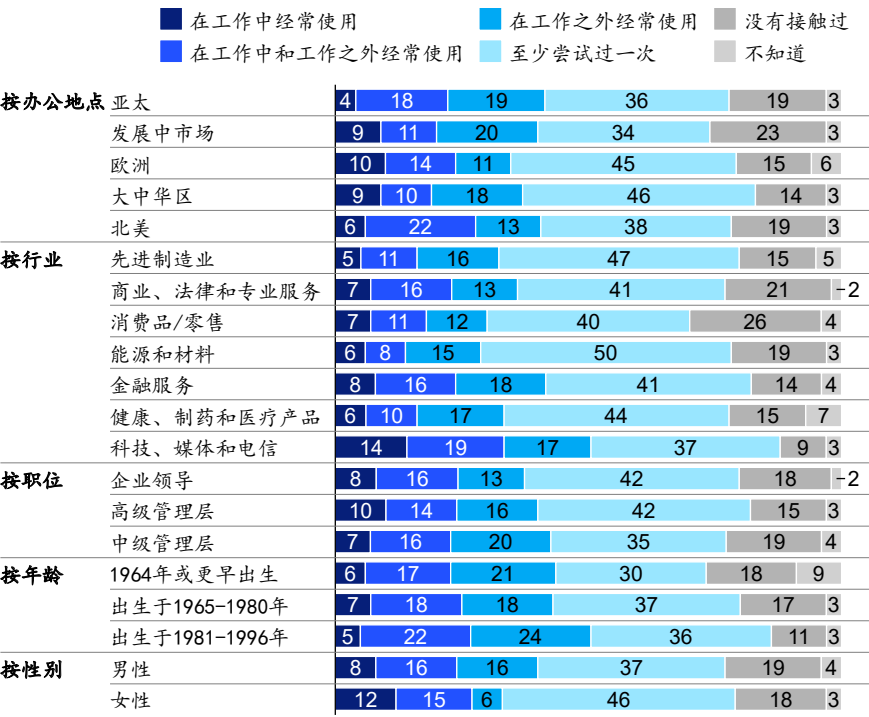
受访者们预测，GenAI将对业务产生巨大影响，组织的员工状况也将发生较大改变。他们预计某些领域会减员，而且为应对人才需求变化，企业需要开展大规模技能再培训。不过，虽然GenAI应用有可能促使企业增加对其他人工智能工具的采用，但我们发现，组织对其他技术工具的采用几乎并无实质性增加。2022年以来，采用人工智能工具的组织占比一直保持稳定，而且人工智能应用也仅集中在少数几项业务职能上。

尽管仍在早期阶段，但GenAI的应用已相当普遍

这项调查于2023年4月中旬进行，调查结果表明，尽管GenAI进入公众视野并不久，但对这些工具的尝试性应用已经相当普遍，受访者

们预计这些新能力将改变其所在行业（见图1）。GenAI已引起商界人士的兴趣：受访者不分地区、行业和资历等级，人人都在工作中和工作之外尝试使用GenAI。79%的受访者表示，他们至少在工作中或工作之外接触过GenAI。22%的受访者表示，他们会在工作中经常使用GenAI。不同资历受访者的GenAI使用率非常相似，而从行业和地区来看，技术行业和北美地区受访者的GenAI使用率最高。

图1:各地区、行业和资历级别的受访者表示，他们已经在使用生成式人工智能工具。



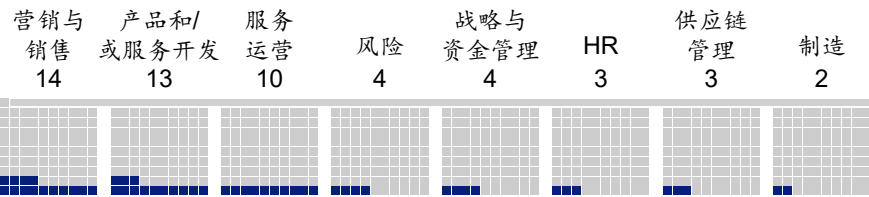
备注：由于四舍五入，各项数据加总后不一定等于100%。亚太地区受访人数 = 164；欧洲 = 515；北美 = 392；大中华区（包括香港和台湾）= 337；发展中国市场（包括印度、拉丁美洲、中东和北非）= 276。先进制造业（包括汽车和装配、航空航天和国防、先进电子和半导体）受访人数 = 96；商业、法律和专业服务 = 215；消费品和零售业 = 128；能源和材料 = 96；金融服务 = 248；健康、制药和医疗产品 = 130；科技、媒体和电信 = 244。企业领导受访人数 = 541；高级管理层 = 437；中层管理层 = 339。1964年或更早出生的受访者人数 = 143；1965年-1980年出生的受访者 = 268；1981年-1996年出生的受访者 = 80。不是所有受访者都提供了年龄信息。男性受访者人数 = 1025；女性受访者人数 = 156。调查样本还包括选择“非二元性别”或“其他”的受访者，但数量不多，不具有统计意义。

资料来源：麦肯锡全球人工智能调查，1684名受访者涵盖组织各个层级，2023年4月11日至21日进行

组织层面上，GenAI的应用现在也十分普及。1/3的受访者表示，他们所在的组织至少在一个职能中经常使用GenAI，在已经应用AI的组织中，有60%都在使用GenAI。更重要的是，在已采用人工智能的组织中，有40%都表示由于GenAI的巨大潜力，他们的公司预期将增大对人工智能的投入，28%的企业表示GenAI的应用已被列入董事会议程。从应用AI技术的业务职能来看，新工具与人工智能的整体应用状况相同，应用最多的职能部门都是营销和销售、产品服务开发以及客户服务和后台支持等服务运营部门（见图2）。这表明企业正将这些新工具部署到价值潜力最大的地方。在我们之前的研究中，上述3个职能领域，再加上软件工程，可实现GenAI用例全年总价值的75%左右。

图2：生成式AI工具最常见的用途是营销和销售、产品服务开发以及服务运营。

表示其组织在相应职能中经常使用GenAI的受访者占比（%）¹



各职能的GenAI常见用例，受访者占比%

营销和销售	产品和/或服务开发	服务运营
起草文本初稿	识别客户需求趋势	采用聊天机器人（如用于客户服务）
9	7	6
个性化营销	起草技术文件	预测服务趋势或异常情况
8	5	5
总结文本文档	设计新产品	起草文件初稿
8	4	5

1. 问题仅针对表示其企业已至少在一个职能中采纳人工智能技术的受访者。显示数据已经过调整，以代表所有受访者。
资料来源：麦肯锡全球人工智能调查，1684名受访者涵盖组织各个层级，2023年4月11日至21日进行

麦肯锡评论



Alex Singla

全球资深董事合伙人, QuantumBlack, AI by McKinsey
全球负责人

有关GenAI讨论愈发成就在几个月前, 高管之间的对话还相当初级, 主要是理解它是什么, 看看哪些是炒作, 哪些是现实。现在, 短短6个月过后, 商界领袖们已经在讨论更加复杂的问题了。从调查结果中我们看到, 有近1/3的公司在至少一个业务职能中使用GenAI。这凸显出企业对“GenAI在商界可行”这一看法的理解和接受程度。

接下来问题将是, 企业该如何采取下一步行动, 以及GenAI的发展轨迹是否会与我们在更广泛的人工智能领域观察到的模式相同, 即采用率稳定在50%左右。我们从数据中看到, 在GenAI发展潜力的鼓舞下, 已经使用人工智能的企业中有近一半企业计划增加对人工智能的投入, 部分原因是他们认识到, 企业需要具备更广泛的能力才能充分利用GenAI。

若要迈出下一步, 将GenAI从实验性工具转变为商业引擎, 运用它为企业实现丰厚的投资回报, 则要先解决一大批问题。这些问题包括: 确定GenAI在组织中有哪些具体机会、治理和运营模式应该是什么、如何最好地管理第三方(例如云和大语言模型提供商)、需要具备哪些条件才能有效管理各种风险, 并理解该技术对人员和技术栈的影响, 清楚如何在实现短期收益与奠定大规模应用所需的长期基础之间找到平衡。这些问题都很复杂, 但它们是释放真正重要价值的关键。

在当前这一早期阶段，人们对GenAI未来影响的预期很高：3/4的受访者预计GenAI将在未来3年内对其所在行业的竞争性质产生重大或颠覆性影响。技术和金融服务行业的受访者预期GenAI将产生颠覆性影响的比例最高。我们之前的研究表明，虽然所有行业都会受到一定程度的冲击，但受影响程度会有所不同²。最依赖知识工作的行业可能会受到较大影响，也因此有望收获较多价值。我们的估算显示，毫无意外，科技公司受到GenAI的影响最大——价值潜力相当于行业全球收入的9%；但除此之外，银行业（价值潜力可达5%）、制药和医疗产品（增值同样可达5%）和教育（可达4%）等基于知识的行业，也会受到重大影响。相比之下，航空航天、汽车和先进电子等制造业受到的颠覆性影响可能较小。这与之之前对制造业影响最大的技术浪潮形成鲜明对比，出现这一局面是因为，GenAI的优势在于基于语言的活动，而非那些需要体力劳动的活动。

调查结果显示，许多组织尚未着手解决来自GenAI的潜在风险

根据调查，几乎没有企业已为GenAI的广泛使用做好了充分准备——或者说，企业并没有准备好应对这些工具可能带来的商业风险（见图3）。在已采用人工智能的企业受访者中，只有21%的人表示，企业已制定相关政策，规范员工工作中的GenAI技术使用行为。在被问及GenAI的具体应用风险问题时，针对最常被提及的GenAI的风险（即不准确问题），很少有受访者表示他们的公司已有风险防范措施到位。受访者提及不准确问题的频率高于网络安全和监管合规，后两项是之前调查显示的整体人工智能技术最常见风险。只有32%的受访者表示正在着手应对不准确问题，这一比例低于网络安全风险的应对比例（38%）。有趣的是，网络安全的应对比例明显低于去年调查中的相关数字，去年提及正采取措施应对人工智能相关网络安全风险的受访者比例为51%。总体而言，正如我们前几年所看到的那样，大多数受访者表示他们的组织并未着手解决人工智能相关风险。

麦肯锡评论

**Alexander Sukharevsky**

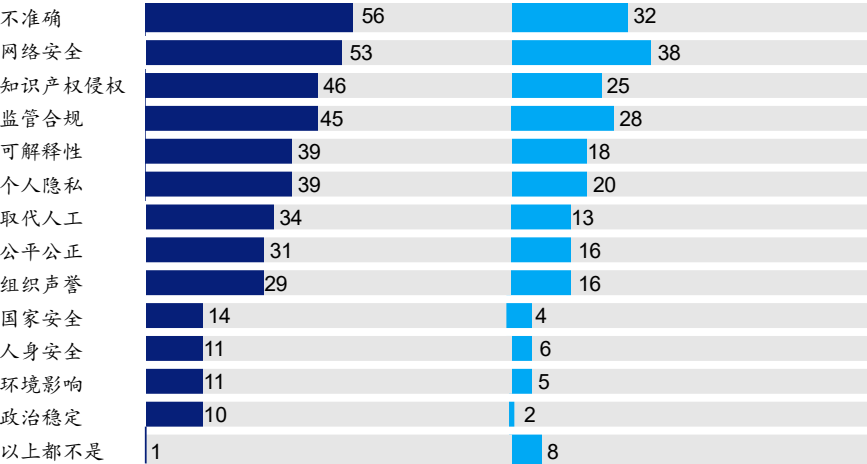
全球资深董事合伙人, QuantumBlack, AI by McKinsey
全球负责人

大家都知道GenAI有风险。但同时,普遍存在的焦虑和害怕使企业领导难以有效应对这些风险。我们最新的调查结果显示,只有略多于20%的公司制定了生成式AI相关风险防范政策。这些政策往往侧重于保护公司的数据、知识和其他知识产权等专有信息。这些当然很重要,但我们发现,许多风险可以通过改变反映既定政策的业务技术架构来解决。

然而,真正的问题在于企业对风险的看法过于狭隘。企业还需要关注一系列重大风险,包括社会风险、人道主义风险、可持续性风险。事实上,与一些人想象的世界末日场景相比,GenAI那些意想不到的后果更有可能给世界带来问题和麻烦。以最具建设性的方式接触GenAI的企业正在尝试应用这项技术,同时制定结构化流程来识别和解决更广泛的风险。这些企业会部署测试版用户,并设置专门团队,预判GenAI应用可能的脱轨之处,从而更好地预测某些不良后果。这些企业还与业内最优秀、最具创造力的人才合作,确定什么是对组织和整个社会而言都最有益的结果。对于可能出现的新风险以及新机遇进行全盘、结构化的深入思考,理解其本质,这对于促成GenAI以负责任的方式实现增长、有效提升生产效率至关重要。

图3:不准确、网络安全和知识产权侵权是生成式AI应用中最常被提及的风险。

企业认为重要且正在着手解决的生成式AI相关风险，
受访者占比（%）¹



1. 仅对表示其所在企业已在至少一个职能采纳AI技术的受访者提问。上述两组受访者人数均为913。
资料来源：麦肯锡全球人工智能调查，1684名受访者涵盖组织各个层级，2023年4月11日至21日进行

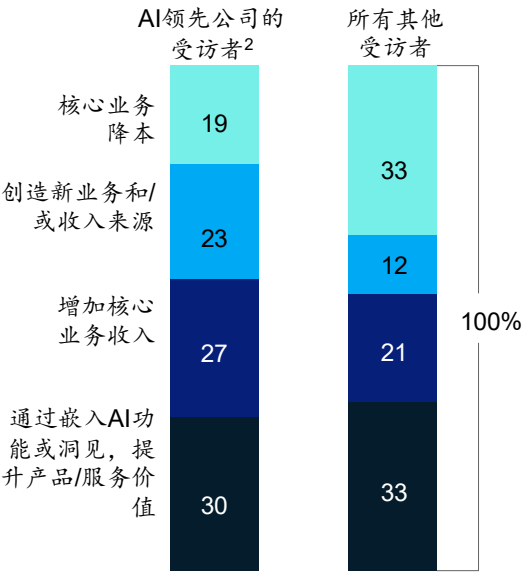
AI领域的领先企业已在GenAI应用上 抢先一步

调查结果显示，人工智能高绩效企业（即据受访者披露，2022年至少20%的息税前利润源自人工智能应用的组织）正全力投入包括GenAI和较传统人工智能在内的所有人工智能能力。这些早先就已从人工智能中获得巨大价值的组织，在GenAI应用方面也已经比其他组织领先一步，这些企业会在更多职能领域应用GenAI，特别是在产品和服务开发、及风险和供应链管理方面。放眼所有人工智能能力（包括较传统的机器学习、机器人流程自动化和聊天机器人），人工智能高绩效企业比其他企业在产品和服务开发上都有较多应用；这些企业利用人工智能来优化产品开发周期，为现有产品添加新功能，以及创建新的基于人工智能的产品等。此外，这些企业在风险模拟和人力资源领域（例如绩效管理、组织设计和劳动力部署优化）也会更多使用人工智能。

还有一个不同于其他企业的方面是：高绩效企业的人工智能活动更少以降本为目标，而其他组织会将降本视为首要任务（见图4）。就来自人工智能高绩效企业的受访者而言，他们表示GenAI的首要目标是创造全新业务或收入来源的可能性是其他人的两倍，而且他们最有可能提到基于人工智能的新功能提高了现有产品的价值。

图4:来自人工智能高绩效企业的受访者将降本视为生成式AI首要目标的比例较低。

组织GenAI活动的首要目标，
受访者占比（%）¹



备注：由于四舍五入，各项数据加总后不一定等于100%。
1. 仅对表示其所在企业已至少在一个职能采纳AI技术的受访者提问。
2. 表示其组织2022年至少有20%的息税前利润源自AI应用的受访者。人工智能高绩效企业受访者人数 = 45；所有其他受访者人数 = 712。
资料来源：麦肯锡全球人工智能调查，1684名受访者涵盖组织各个层级，2023年4月11日至21日进行

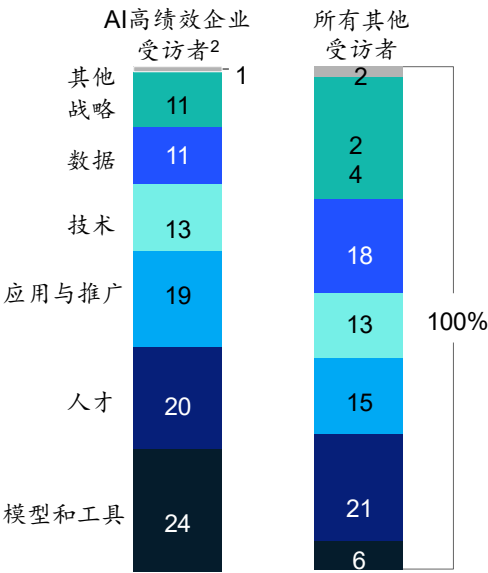
正如我们前几年所见，这些高绩效组织在人工智能方面的投资远多于其他组织：来自人工智能高绩效组织的受访者更有可能（可能性是其他受访者的五倍以上）表示，他们将超过20%的数字预算用于人工智能。这些组织内部也更广泛地使用人工智能能力。来自高绩效企业的受访者比其他人更有可能表示，他们的组织已在4个或

更多业务职能中采用了人工智能，并且他们在工作流程中嵌入的人工智能能力数也较多。例如，来自高绩效企业的受访者更常表示，除了GenAI和相关的自然语言能力之外，他们还将知识图谱嵌入到至少一种产品或业务职能流程中。

虽然人工智能高绩效企业也面临获取人工智能价值方面的挑战，但调查结果表明，他们所面临的困难反映出其在人工智能应用方面相对成熟，而其他企业面对的则是人工智能采用方面更基础的战略性问题（见图5）。人工智能高绩效企业的受访者最常提到的最大挑战在于模型和工具（如在生产中监控模型性能、以及需要随着时间推移重新训练模型）。相比之下，其他受访者提到的则是战略问题，例如设定明确的人工智能愿景并与业务价值挂钩，或寻求足够的资源支持。

图5:模型和工具是领先企业面临的最大AI挑战，其他企业则更多被战略难题所困。

实现人工智能潜力价值的最大挑战，
在受访者中占比（%）¹



注：由于四舍五入，各项数据加总后不一定等于100%。
1. 该问题仅针对表明自己所在组织某一职能领域已采用人工智能技术的受访者。
2. 表示其组织2022年至少有20%的息税前利润源自AI应用的受访者。AI高绩效企业受访者人数 = 49；其他受访者人数 = 792。
资料来源：麦肯锡全球人工智能调查的1684名受访者涵盖组织各个层级，2023年4月11日至21日进行



麦肯锡评论

Bryce Hall

全球副董事合伙人

过去6年，我们每年开展一次全球人工智能研究，每次都得出一个相同的结论——领先企业通常会从更广阔的视角出发，权衡考量成功必备的要素。领先企业格外善于聚焦价值，并有针对性地调整组织架构，成功捕捉价值。观察领先企业运用GenAI技术的方式，可以清晰地看到上述规律。

例如，就战略而言，我们分析中的领先企业会梳理各个业务领域，寻找人工智能创造高价值的机会。值得注意的是，他们不仅仅是在应用GenAI时这么做。虽然炙手可热的GenAI新技术让人心动，但其实AI应用中有远超过半数的潜在价值来自非GenAI技术。领先企业往往会秉持审慎态度，根据潜在价值全面评估人工智能领域的机会。

这样的方法适用于所有能力相关领域。以技术和数据为例，领先企业会精准识别捕捉目标价值所需的能力，包括运用公司和行业数据训练大型语言模型的能力。目前存在两条路线，一是直接采用现有的人工智能服务（即所谓的“接受者”方式），二是通过诸如调整并训练模型、使其能够运用内部自有数据的方式，培养自身能力，打造竞争优势（即所谓的“创造者”方式）。领先企业会对这两者的效果和效率进行评估和测试。

调研结果进一步验证了我们的结论：在诸如机器学习运维（MLOps）法等人工智能技术的应用方面，即使是人工智能高绩效企业也尚未达到最佳实践水平，不过它们更有潜力做到这一点。例如，来自人工智能高绩效企业的受访者中，仅35%表示其所在组织会尽量选择整合现有模块，而非重新创造；不过，这已远高于19%这一其他企业受访者回复的比例。

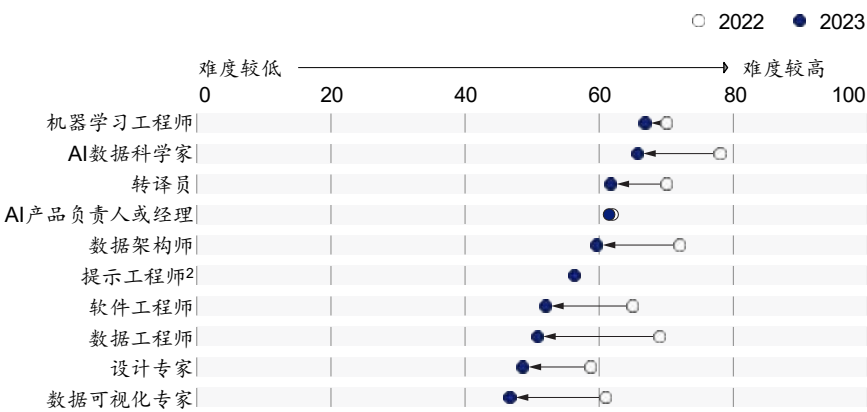
目前，为落地一些更具转型意义的AI用例，并尽可能保证过程中的安全，亟需大量的专业MLOps技术和实践经验。其中包括实时模型运行，即通过检测系统和即时预警，快速解决问题，保证GenAI系统始终处于管控之下。虽然领先企业在这方面表现突出，但仍有进步空间：来自AI高绩效企业的受访者中，有1/4表示所在企业会全面监控所有系统，并已配备即时预警机制，而其他企业中，仅12%能够做到这一点。

人工智能相关人才需求已发生变化，人工智能技术或将给劳动力带来巨大影响

我们的最新调研结果表明，企业为支持人工智能目标而需要的人才岗位已发生变化（见图6）。过去一年，应用人工智能技术的企业招募最多的岗位是数据工程师、机器学习工程师和AI数据科学家，这些都是上一次调研中受访者普遍提及的岗位。不过，招募AI相关软件工程师（去年最热门的招聘岗位）的受访企业占比大大缩小，比例由上一次调研的39%降至28%。随着GenAI技术的应用，对相关技能的需求也在增加，提示工程等新兴领域的岗位悄然兴起，采用人工智能技术的企业中，7%的受访者表示在过去一年里招募了提示工程领域的人才。

图6:人工智能相关岗位招聘仍然很难，不过受访者反映许多岗位的招聘难度在2022年后有所下降。

表示其所在组织面临人工智能相关岗位招聘难题的受访者占比（%）¹



1. 该问题仅针对表明自己所在组织某一职能领域已采用人工智能技术、且过去12个月招聘相关岗位的受访者。未展示回复“简单”、“不简单也不难”或“不知道”的受访者。
2. 2022年末询问受访者。
资料来源：麦肯锡全球人工智能调查的1684名受访者涵盖组织各个层级，2023年4月11日至21日进行

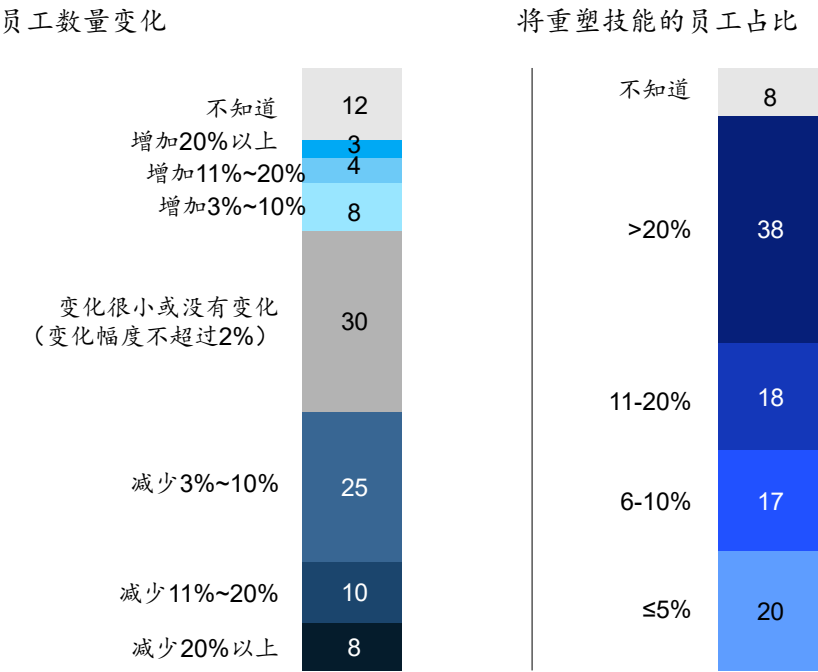
研究结果还显示，招募人工智能相关人才依然很难，不过，过去一年的情况多少有所好转，这与2022年底到2023年上半年科技公司的裁员潮可能有关。同时，反映很难招到AI数据科学家、数据工程师和数据可视化专家的受访企业较上一次调查有所减少，而机器学习工程师和AI产品负责人的招募工作难度仍然较大，与去年基本一致。

受访者预测，未来3年，人工智能技术将重塑许多工作岗位（见图7）。整体而言，受访者预计会有更多员工转换技能，而非离职。已应用AI企业的近4成受访者预计，企业将有超过20%的成员经历技能重塑，而8%的受访者认为组织规模将缩减20%以上。

具体到各职能部门可能受到的GenAI冲击，服务运营是唯一一个大多数受访者均认为员工数会减少的职能部门（见图8）。这与我们近期的研究成果基本一致，即虽然GenAI技术会进一步提高我们所估算的可自动化工作内容的占比（由50%升至60%~70%），但这并不代表相关岗位的全部工作可以完全实现自动化。

图7:受访者预计人工智能将对企业的员工队伍有重大影响

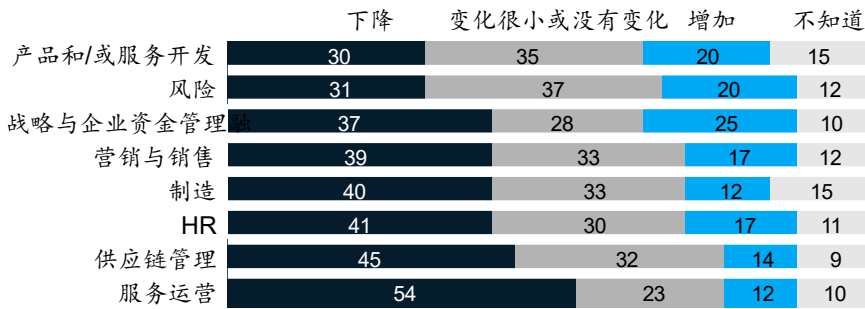
对未来3年应用人工智能技术对企业员工队伍所产生影响的预期，
受访者占比（%）¹



注：由于四舍五入，各项数据加总后不一定等于100%
1. 仅对表示其所在企业已至少在一个职能采纳AI技术的受访者提问，受访者人数 = 913
资料来源：麦肯锡全球人工智能调查的1684名受访者涵盖组织各个层级，2023年4月11日至21日进行

图8:服务运营是唯一一个大多数受访者都认为会因生成式AI而缩减规模的职能部门。

未来3年生成式AI对员工数量的影响，按业务职能划分，
受访者占比（%）¹

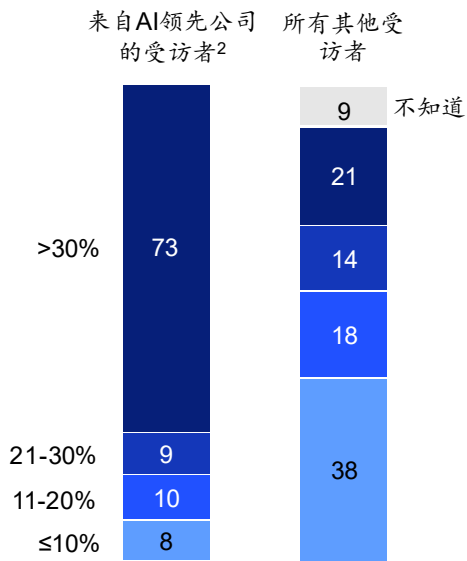


注：由于四舍五入，各项数据加总后不一定等于100%
1. 该问题仅针对表明自己所在组织某一职能领域已采用人工智能技术的受访者
资料来源：麦肯锡全球人工智能调查的1684名受访者涵盖组织各个层级，2023年4月11日至21日进行

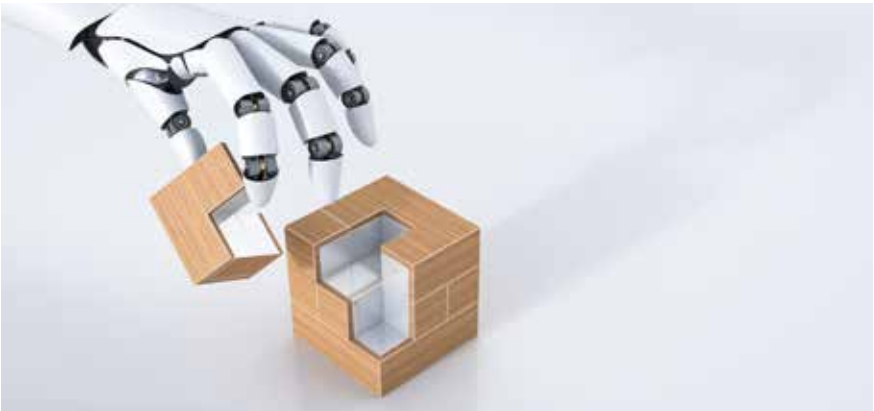
AI领先公司重塑员工技能的力度预计将超过其他企业（见图9）。在表示所在企业未来3年会重塑30%以上员工技能的受访者中，来自AI领先公司的受访者人数是其他企业的3倍。

图9:来自AI领先公司的受访者所预计的企业重塑员工技能的力度大于其他公司。

预计应用人工智能后公司未来3年需要重塑技能的员工比例，受访者占比（%）¹



1. 该问题仅针对表明自己所在组织某一职能领域已采用人工智能技术的受访者
2. 表示其组织2022年至少有20%的息税前利润源自AI应用的受访者。来自AI领先企业的受访者人数 = 50；所有其他受访者人数 = 863
资料来源：麦肯锡全球人工智能调查的1684名受访者涵盖组织各个层级，2023年4月11日至21日进行





麦肯锡评论

Lareina Yee

麦肯锡全球资深董事合伙人；麦肯锡技术委员会主席

GenAI技术初试锋芒，企业就已经在预见该技术将对人才队伍产生的巨大影响，包括开放新的工作机会，改变工作方式，乃至引入提示工程等全新的岗位类型。GenAI技术可用于辅助几乎所有工作岗位，这既是它的优势，也是它最大的挑战。

传统AI技术虽然重要，但影响到的员工范围较小，仅针对在机器学习、数据科学或机器人领域拥有深厚技能积累的人才，这一点与GenAI技术存在显著差异。传统AI技术人才的专业化程度较高，因此在市场中总是一才难求。我们的调研结果也凸显出相关岗位的招聘难度。而反观GenAI技术，虽然开发大型语言模型、训练生成式模型等工作，对从业者技能的要求依然很高，但几乎人人都可以使用这一技术，用户不需要任何数据科学专业背景或机器学习专业知识，就能有效地利用GenAI技术完成工作。这有些像是从主机电脑到个人电脑的发展，前者只有技术专家才能操作，而个人电脑人人皆可掌握。这是一次颠覆性转变，技术成了人人可用的强大工具。

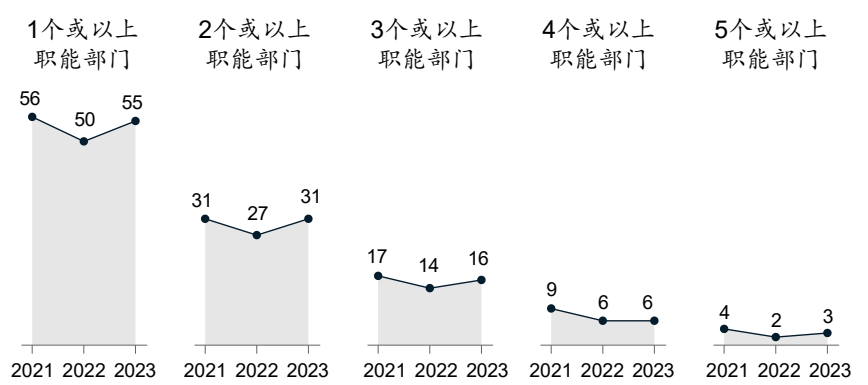
以GenAI为工具的观点在我们的调查中也有所体现。大多数情况下，企业都将GenAI技术看作是强化人工工作的工具，而非人工的替代品。目前，我们看到很多公司都在不断探索GenAI技术，重点关注最有可能有效提升营收或生产效率的实用领域，如利用GenAI工具更新遗留代码，或缩短科学研究和探索的周期。上述列举的种种只是GenAI技术的冰山一角，预计未来相关应用的落地速度会显著提升。

虽然GenAI技术备受瞩目，但普及率和影响范围依然有限

虽然GenAI工具的应用正在快速普及，但调研数据表明，企业整体的AI技术普及率并未因为这些新工具而显著提升（见图10）。采用人工智能技术的公司比例依然止步不前，目前为止，55%的受访企业表示已采用相关技术。表示所在企业已在1个以上职能部门采用人工智能技术的受访者比例不足1/3，这表明AI应用范围依然较为有限。根据受访者的回答，应用人工智能技术最多的职能部门是产品和服务开发及服务运营，这与此前4次调研的结果一致。仅23%的受访者称，去年息税前利润中有至少5%来自人工智能技术的应用，比例与上一次调查基本相同，表明价值挖掘潜力依然较大。

图10:不足1/3的受访者表示，所在企业已经1个以上职能部门采用人工智能技术——该数字在2021年后几乎没有变化。

受访者所在组织应用人工智能技术的业务部门数，受访者占比（%）¹



1. In 2021, n = 1843; in 2022, n = 1492; in 2023, n = 1684
资料来源：麦肯锡全球人工智能调查的1684名受访者涵盖组织各个层级，2023年4月11日至21日进行



麦肯锡评论

Michael Chui

麦肯锡全球研究院董事合伙人

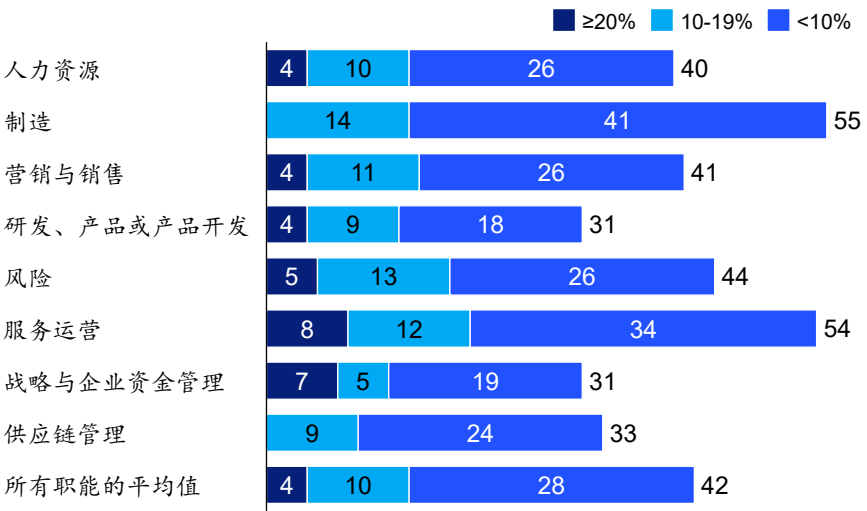
我们总在强调GenAI技术的重要性——当然，有鉴于该技术的革命性潜力，如此强调并不过分——但这次调研结果提醒我们，在更广阔的人工智能领域，还有更多价值等待挖掘。事实上，部分研究成果表明，非生成式AI的潜在价值有可能甚至比GenAI更大。在提升预测准确率、优化物流网络、生成下一次购物推荐等领域应用更广泛的AI技术，有望为企业创造更多价值。

尽管受访者反馈的整体人工智能普及率稳定维持在55%左右，但有超过2/3的受访者表示自己所在公司已计划加大对人工智能领域的投资力度。我们不断看到AI领域的佼佼者着手构建基础、培养能力、为价值创造铺路。这件事可以这样理解，在挖掘人工智能价值潜力方面，“富者愈富”此言不虚。企业对GenAI技术的浓厚兴趣能否成为推动人工智能广泛普及的契机，让我们拭目以待吧。

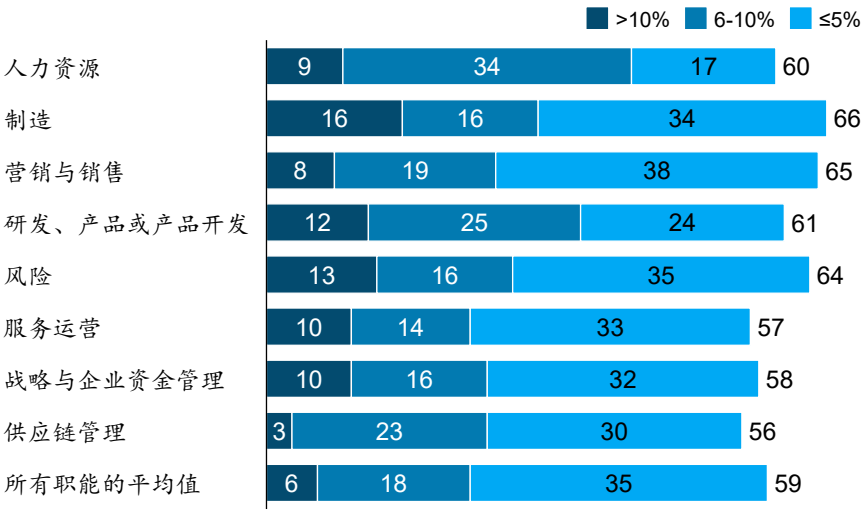
**超过2/3的受访者预测
未来3年所在组织将加大对
人工智能领域的投资力度。**

图11:企业继续在采用人工智能技术的职能部门看到长期效益。

2022年AI技术应用的降本成效，受访者占比（%）¹



2022年AI用例落地后营收增加，受访者占比（%）²



1. 该问题仅针对表明自己所在组织某一职能领域已采用人工智能技术的受访者。不展示回答“成本上涨”、“无变化”、“不适用”或“不知道”的受访者。

2. 该问题仅针对表明自己所在组织某一职能领域已采用人工智能技术的受访者。不展示回答“营收下降”、“无变化”、“不适用”或“不知道”的受访者

资料来源：麦肯锡全球人工智能调查，1684名受访者涵盖组织各个层级，2023年4月11日至21日进行

企业继续在采用人工智能技术的业务领域中看到长期效益，并计划在未来加大相关领域的投资力度（见图11）。多数受访者表示，人工智能技术在相关业务部门的应用带来了营收增长。超过2/3的受访者预测，未来3年，所在企业将增加在人工智能领域的投入。

本次研究相关信息

本次调研于2023年4月11日至21日在线上开展，共收到1684份问卷回复，全面覆盖各地区、行业、公司规模、专业职能和工作年限。其中，913名受访者表示所在公司至少已在1个职能部门采用AI技术，并被问及有关其组织中人工智能应用的问题。为针对回复率差异做出调整，我们根据受访者国家对全球GDP的贡献值对数据进行了加权处理。

¹ 我们将人工智能高绩效企业定义为，至少20%的息税前利润源自人工智能应用的企业（根据受访者提供数据）。

² 《The economic potential of generative AI: The next productivity frontier》，麦肯锡，2023年6月14日。

Michael Chui是麦肯锡全球研究院董事合伙人，常驻湾区分公司；

Lareina Yee是麦肯锡全球资深董事合伙人，常驻湾区分公司；

Bryce Hall是麦肯锡全球资深董事合伙人、QuantumBlack全球负责人，常驻芝加哥分公司；

Alexander Sukharevsky是麦肯锡全球资深董事合伙人、QuantumBlack全球负责人，常驻伦敦分公司。

鸣谢Shivani Gupta、Abhisek Jena、Begum Ortaoglu、Barr Seitz和张力对本次调研工作的贡献。

麦肯锡公司2023年版权所有。



生成式AI: CEO必读指南

生成式AI (Generative AI) 在飞速演进，CEO们也在探索其商业价值及潜在风险。为此，我们提供一份生成式AI核心概要，供广大CEO们参考。

Michael Chui、Roger Roberts、Tanya Rodchenko、Alex Singla 和Alex Sukharevsky、Lareina Yee和Delphine Zurkiya共同撰写，谨代表Mckinsey Digital旗下科技委员会 (McKinsey Technology Council) 和QuantumBlack, AI by McKinsey的观点。

自ChatGPT、Bard、Claude、Midjourney以及其他内容生成工具问世以来,人们对生成式AI抱有很高期待。各企业CEO自然也在思考:这究竟是科技炒作,还是颠覆行业格局的机遇?如果是后者,那生成式AI能给自身业务带来什么价值?

ChatGPT的大众版仅两个月就吸引到1亿用户。它以史无前例的方式推动了AI的普及,已成为迄今增长最快的应用程序。无与伦比的易用性让生成式AI有别于以往所有AI技术。用户不需要专修机器学习就可以开展交互、获取价值——只要会提问,几乎人人都能用。就像个人电脑或iPhone等其他突破性技术一样,一款生成式AI平台可以衍生出许多应用程序,适用于各个年龄段和教育水平的用户群体,人们无论身处何地,能够上网即可使用。

而实现这一切,依靠的是驱动生成式AI聊天机器人的基础大模型,它们是经由大量非结构化、无标签数据(如文本、音频等各类形式)训练的庞大神经网络。基础大模型可处理各种各样的任务。相比之下,以往的AI模型通常适用范围更“窄”,往往只能执行一项任务,如预测客户流失率等。而一个基础大模型则既能为一份2万字的量子计算技术报告生成内容摘要,又能为园艺公司起草市场进入策略,还能根据冰箱里的10种食材给出5张不同的食谱。不过,在其丰富功能的背后,目前还存在结果不够准确的短板,这也让人们再度关注起AI的风险管理问题。

在监管得当的情况下,生成式AI不仅可以为企业开辟新用例,还可以加速、扩展或改进现有用例。以电销场景为例,经过专门训练的AI模型可以帮助销售人员发现追加销售机会,但截至目前,这些模型通常还只能根据通话前收集的人口统计信息和购买规律等静态客户数据来判断追加销售的可能性。生成式AI工具则可根据实际对话内容,利用内部客户数据、外部市场趋势和社交媒体影响者数据,实时为销售人员提供追加销售建议。同时,生成式AI还可以为销售人员撰写销售话稿,供其根据具体情况进行调整。

上述例子只展示了AI技术对人类工作潜在影响的一个侧面，而实际上，几乎所有知识工作者都有可能因使用生成式AI而获益。尽管生成式AI最终可能会让部分工作自动化，但其价值将更多来自于被嵌入日常工具（如电子邮件或文字处理软件）后知识工作者对它的使用。这类升级后的工具可以大幅提高生产力。

CEO们想知道是否应立即采取行动，以及如果采取行动，该从何开始。有些人可能从中看到了机遇，希望通过重塑人与生成式AI应用程序协同工作的方式，在竞争中弯道超车。其他人则可能希望谨慎行事，在进行大规模投资之前先尝试几个用例，增进对生成式AI的理解。企业也需要评估自身是否具备必要的技术专识、技术及数据架构、运营模式以及风险管理流程，这些是更进一步部署生成式AI时所需要的。

本文旨在帮助CEO及其团队思考生成式AI的价值创造场景以及如何开始应用。首先，我们总结了生成式AI的入门指南，以帮助CEO更好地了解AI日新月异的发展现状和可行技术选择。第二部分将通过4个旨在提高组织效能的案例，探讨企业如何应用生成式AI。这些案例来自我们对早期采用者的观察，并介绍了在技术、成本和运营模式要求等方面的各种选择。最后，我们将探讨CEO如何发挥关键作用，利用生成式AI带领企业走向成功。

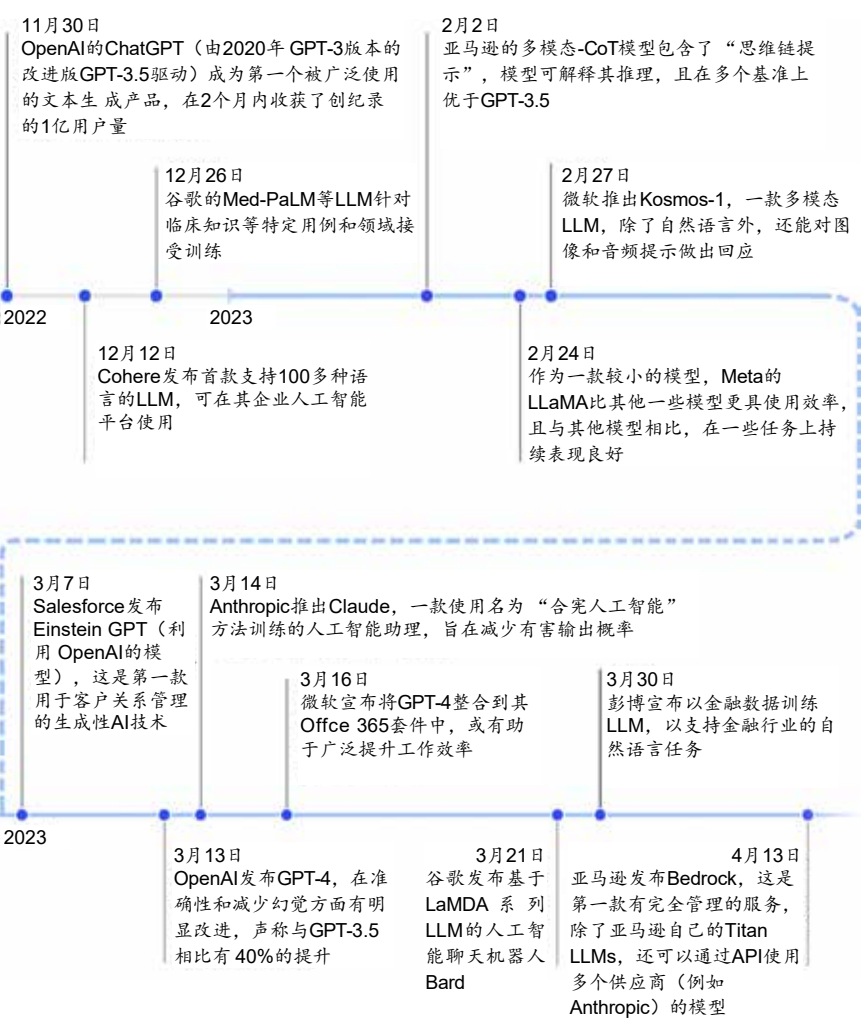
人们对生成式AI的期待显而易见，企业高管自然希望借此东风运筹帷幄，有计划地快速推进。我们希望本文能让商业领袖更全面地了解生成式AI未来潜力。

生成式AI入门指南

生成式AI技术飞速发展(见图1)。发布周期之短、初创公司数量之众、与现有软件的整合之快,皆不同凡响。在本节,我们将探讨生成式AI应用的广度,并简要介绍该技术,包括阐明其与传统AI的区别。

图1: 生成式人工智能技术飞速演进

在ChatGPT推出后的几个月里,主要大型语言模型(LLM)的发展时间线



资料来源:麦肯锡分析

不仅是聊天机器人

生成式AI可用于工作自动化、辅助强化与加速推进。从本文宗旨出发,我们将着重阐述生成式AI如何辅助并强化人类工作,对其替代人类工作的潜力不作探讨。

ChatGPT等文本生成式聊天机器人备受关注,不过,生成式AI同样可以在图像、视频、声音以及计算机代码等更广泛的内容领域提供助力。在企业,生成式AI也可以发挥广泛功能,比如分类、编辑、总结、回答问题和起草内容。针对上述每项活动,各个业务职能和工作流程的具体工作方式转变都有可能创造价值,部分示例如下:

分类

- 反欺诈分析师可以将交易描述和客户文件输入到生成式AI工具中,要求其识别欺诈交易。
- 客户服务经理可以使用生成式AI工具根据客户满意度对客户通话音频文件进行分类。

编辑

- 撰稿人可以利用生成式AI纠正语法,并使文字风格与顾客的牌调性相匹配。
- 平面设计师可以利用生成式AI从图像中移除过时标志。

总结

- 制片助理可以用数小时的活动录像创作精彩视频集锦。
- 业务分析师可以创建维恩图(Venn diagram)总结高管发言要点。

回答问题

- 制造企业员工可以向基于生成式AI的“虚拟专家”咨询有关操作流程的技术问题。
- 消费者可以向聊天机器人询问新家具的组装方式。

起草内容

- 软件开发者可以让生成式AI完成整段代码, 或者提供建议以补全现有代码的未完成语句。
- 营销经理可以使用生成式AI起草不同版本的营销活动信息。

随着技术的发展和成熟, 这类生成式AI可以更好地融入企业工作流程, 实现任务自动化并直接执行特定操作(例如, 在会议结束时自动发送纪要)。该领域已经有一些工具面市。

生成式AI与其他AI的不同之处

顾名思义, 生成式AI和过往AI技术或分析工具的主要区别在于, 该技术能够生成新内容, 所生成的新内容通常以“非结构化”形式(如书面文本或图像)呈现, 而非以表格形式排列(见侧边栏《术语表》中的生成式AI相关术语)。

其底层技术是一类被称为基础大模型的人工神经网络, 其灵感来自于人类大脑中数十亿相互连接的神经元。人工神经网络需要通过深度学习加以训练, “深度”即指神经网络中的层数之众、之深。深度学习技术推动了AI领域的众多新进展。

而某些特质使得基础大模型区别于过往的深度学习模型。首先, 训练基础大模型可以使用体量庞大、类型多样的非结构化数据。例如, 一类被称为大型语言模型的基础大模型可以通过互联网上公开可用且涵盖各类主题的大量文本进行训练。其他深度学习模型虽然也可以处理大量非结构化数据, 但训练所用的数据集通常更具体。例如, 为了让模型识别照片中的某些物体, 需要使用一组特定图像对其进行训练。

事实上, 其他深度学习模型往往只能执行一项此类任务。例如, 它们要么对照片中的物体分类, 要么执行预测等其他功能。相比之下, 基础大模型可以同时实现上述功能, 并且还能够生成内容。上述能力的积累是通过从所摄取的广泛训练数据中学习规律和关系实现的, 比如, 通过规律和关系学习, 基础大模型能够预测句子中的

术语表

应用程序接口 (API) 是一种通过编程访问 (通常是外部的) 模型、数据集或其他软件的方式。

AI是指软件有能力执行过去需要人类智能才能完成的任务。

深度学习 (deep learning) 是机器学习的分支, 使用由相互连接的多层“神经元”组成的深度神经网络, 这些连接具有可训练的参数或权重。它在学习图像、文本和音频等非结构化数据方面尤为有效。

微调 (fine-tuning) 是指调试预训练模型以使其更好地处理特定任务的过程。这需要在相对较短的时间内, 通过标记的数据集进行训练, 这个数据集比最初训练模型的数据集小得多。这一额外训练使模型能够学习并适应较小数据集中的细微差异、术语和特定规律。

基础大模型 (foundation model, FM) 是基于大量非结构化、无标签数据训练的深度学习模型, 可以直接用于广泛的任务, 也可以通过微调适应特定任务。GPT-4、PaLM、DALL·E2 和 Stable Diffusion便属于这类模型。

生成式AI (Generative AI) 通常指使用基础大模型构建的AI, 具有以往AI所没有的能力, 比如生成内容的能力。基础大模型也可用于非生成性目的 (例如, 根据通话记录将用户的情绪分类为负面或正面), 这类用例的结果相较早期模型有明显改进。为方便起见, 本文在提到生成式AI时, 包括所有基础大模型用例。

图形处理器 (graphics processing units, GPU) 是计算机芯片, 最初为制作计算机图形 (如视频游戏) 而开发, 同样可支持深

度学习应用。相比之下,传统的机器学习和其他分析工具通常在被称为计算机“处理器”的中央处理器(CPU)上运行。

大型语言模型(large language model, LLM) 是一类基础大模型,可处理大量非结构化文本,学习单词或词组(称为token)之间的关系。这使得大型语言模型能够生成自然语言文本,执行总结或提取知识点等任务。GPT-4 (ChatGPT的基础大模型) 和LaMDA (Bard的基础大模型) 均是大型语言模型。

机器学习(machine learning, ML) 是AI的一个分支,在该领域中,模型接受训练、学习大量样例数据点之后获得能力。机器学习算法通过处理数据和经验(而非接收明确的编程指令)来发现规律并学习如何做出预测和推荐。算法也会自我调试,能对新的数据和经验做出更有效的反应。

MLOps指的是扩展和维持AI和机器学习的工程模式和实践,包括一整套覆盖整个机器学习生命周期的实践(数据管理、开发、部署和实时运营)。目前,这些实践很多都由辅助软件(任务标准化、简化或自动化工具)支持或优化。

提示工程(prompt engineering) 是指设计、改进和优化输入提示以引导生成式AI 模型产生所需(即准确)输出的过程。

结构化数据(structured data) 是指以表格、数据库或电子表格等形式呈现的数据,能够有效地用于训练某些机器学习模型。

非结构化数据(unstructured data) 指缺乏统一格式或结构的数据(例如文本、图像和音频文件),通常需要使用更先进的技术以生成见解。

下一个单词。这就是为什么ChatGPT能够回答不同主题的问题、而DALL-E 2和Stable Diffusion能够根据描述生成图像。

鉴于基础大模型的多功能性，企业可以使用同一模型实现多个业务用例，这是早期深度学习模型难以实现的。一款纳入了公司产品信息的基础大模型可能同时用于回答客户问题和协助工程师开发新版产品。因此，企业可以搭建应用并更快实现收益。

然而，基础大模型当前的运行方式决定了它们并不适用于所有类型的应用。例如，大型语言模型容易产生“幻觉”，用看似可信、实则错误的论断回答问题（见侧边栏《负责任地使用生成式AI》）。此外，基础大模型并不总能给出作答的基本推理或依据来源。这意味着在错误可能造成损害、或需要对回答进行解释的情况下，企业对应用无人监督的生成式AI需慎之又慎。生成式AI目前也不适用于直接分析大量的表格数据或解决高级数值优化问题。研究人员正在努力克服这些限制。

生成式AI生态系统正在兴起

基础大模型是生成式AI的“大脑”，而正在兴起的整个价值链将支持该技术的训练和使用（见图2）¹。专用硬件提供了训练模型所需的庞大算力。云平台则提升了对这类硬件的利用。MLOps和模型中心供应商提供企业所需工具、技术和实践，让企业能够调试使用基础大模型并将其部署到终端用户应用之中。许多公司正在进入市场，主打依托基础大模型、能够执行特定任务的应用程序，例如帮助某公司处理客户服务问题。

鉴于训练模型需要庞大的计算资源而完善模型需要大量人力，因而首款基础大模型的开发需要巨大的投资。这导致相关开发工作被部分科技巨头、有充足投资支持的初创公司以及一些开源合作研究机构（例如BigScience）主导。然而，针对有效处理若干任务的小规模模型的研究、以及提高训练效率的工作，均方兴未艾。这最终

可能为更多企业打开市场。部分初创公司已经成功开发了自己的模型,例如, Cohere、Anthropic和AI21实验室都建立并训练了自己的大型语言模型。

图2: 支持生成式AI系统的价值链正在迅速发展

生成式AI价值链

服务

围绕如何利用生成式AI(如,训练、反馈和强化学习)的专业知识而提供的服务

应用

使用基础模型的B2B或B2C产品,基本直接使用或根据特定用例进行微调

模型中心和MLOps

管护、托管、微调或管理基础模型的工具(例如,应用程序和基础模型之间的虚拟门面)

基础模型

用于建立生成式AI应用的核心模型

云平台

提供计算机硬件访问的平台

专用硬件

为训练和运行模型而优化的加速器芯片(例如,图形处理器 GPU)

资料来源:麦肯锡分析

负责任地使用生成式AI

生成式AI伴生了各种风险。企业CEO需要一开始从团队和流程设计上就做好风险防控。这不仅是为了满足不断变化的监管要求,也是为了保护业务并赢得消费者的数字信任(我们将在后文中提供关于这一点的方法建议)¹。

公平性:不完美的训练数据或开发模型工程师的决策瑕疵,可能让模型产生算法偏向。

¹ Jim Boehm、Liz Grennan、Alex Singla和Kate Smaje 2022年9月12日《为什么数字信任真正重要》(“Why digital trust truly matters”)。

知识产权 (IP)：训练数据和模型输出可能带来重大的知识产权风险, 包括侵犯版权、商标权、专利权或其他合法受到保护的材料权利。即便所使用的生成式AI工具来自供应商, 企业也需要了解训练过程中使用了哪些数据以及这些数据在工具输出中的使用方式。

隐私性: 如果用户的输入信息以某种可识别个人身份的形式出现在模型输出中, 则可能引发隐私问题。生成式AI也可能被用于创作和传播虚假信息、深度伪造和仇恨言论等恶意内容。

安全性: 生成式AI有可能被不法分子用来加剧网络攻击的复杂程度和侵害速度, 也可被操纵用于制造恶意输出。例如, 通过名为“提示注入” (prompt injection) 的技术, 第三方可以给模型提供新的指令, 诱导模型产出模型作者和终端用户用意之外的输出。

可解释性: 生成式AI依赖拥有数十亿参数的神经网络, 人们因而很难解释某个答案从何而来。

可靠性: 模型对相同的提示会产生不同的回答, 使用户难以评估输出的准确性和可靠性。

组织影响: 生成式AI可能会对劳动力产生重大影响, 对某些特定群体和社区的负面影响可能尤为巨大。

社会和环境的影响: 基础大模型的开发和训练可能会危害社会和环境, 包括增加碳排放 (例如, 训练一个大型语言模型可能会排放约315吨二氧化碳)²。

² Ananya Ganesh、Andrew McCallum和Emma Strubell 2019年6月5日《深度学习在自然语言处理中的能源和政策考虑》 (“Energy and policy considerations for deep learning in NLP”) 《计算语言学协会第57届年会论文集》 (Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics)。



将生成式AI应用于工作

CEO们应将探索生成式AI列入工作议程，不能仅视之为“可选项”。生成式AI可通过广泛用例创造价值。起步的经济和技术要求并非高不可攀，而无所作为则可能导致迅速被竞争对手甩开。每一名CEO都应同管理团队一起思考竞争领域与方法。一些CEO可能会认定，生成式AI将为企业带来变革性机遇，全面重塑从研发到营销、从销售到客户运营等各个领域。还有些CEO则可能选择从小处着手，再逐渐扩大规模。一旦做出决策，AI专家便可根据场景需要，通过相应技术路径执行战略。

生成式AI在企业中的很多应用（尽管未必是全部价值）将来自员工对现有软件中新嵌入功能的使用。电子邮件系统可以给出邮件初稿；生产应用将根据描述生成演示文稿的初稿；财务软件可对财务报告中的要点给出文字描述；客户关系管理系统则可以提供客户互动建议。这些功能可以提高所有知识工作者的生产效率。

然而，在某些用例中，生成式AI或可带来更具变革性的影响。接下来，我们将介绍不同行业企业利用生成式AI重塑组织工作方式的4个范例²，范围涵盖了从资源需求极少到资源投入很高的不同情况。（这些例子的快速比较和更多技术细节，请参见图3。）

图3：生成式人工智能对组织的要求从低到高不等，具体取决于用例

低 1 ————— 6 高

用例	技术路径	成本	技术人才	专有数据	流程调整
改变软件工程工作	使用软件即服务(SaaS)工具	许多SaaS工具提供固定费率的订阅服务，每名用户每月10至30美元；一些产品则按使用情况定价 1	不太需要技术人才——可能负责选择合适的解决方案和轻度的整合工作 1	模型基本是直接使用的，因此不需要专有数据 1	流程基本保持不变，但工作人员应系统地检查模型结果的准确性和适当性 1
帮助客户经理跟上公共信息和数据的步伐	通过模型API构建软件层	需要前期投资开发用户界面、整合解决方案并建立后处理层 API使用和软件维护的持续成本 3	需要软件开发、产品管理和数据库集成能力，因此需要至少1名数据科学家、机器学习工程师、数据工程师、设计师和前端开发人员 2	模型基本是直接使用的，因此不需要专有数据 1	可能需要一些流程以实现对提示和结果的存储，可能也需要一些防护机制，出于风险或成本考虑而限制使用 2
解放客户支持代表时间进行更有价值的工作	内部对开源模式进行微调	数据清理和标记以及模型微调导致人力资本成本增加，初始成本比基于API构建高出约2倍 模型维护和云计算的持续成本更高 4	经验丰富的数据科学和工程团队，拥有机器学习运营(MLOps)知识和资源，可以检查或创建所需的标签数据 3	需要专有、带标签的数据集以微调模型，尽管在某些情况下该数据及可能相对较小 3	需设置流程以管理分流、将问题上报人工处理以及对模型安全做定期评估 4
加快研究者识别相关细胞特征的速度，助力药物发现	从零开始训练基础模型	前期人力资本和技术基础设施成本导致初始成本比基于API构建高出约10-20倍 模型维护和云计算的持续成本与上一条类似 6	需要大型数据科学和工程团队，具备博士水平的学科知识、MLOps最佳实践以及数据和基础设施管理技能 6	基础模型可以通过大量公开数据训练，但长期的差异性优势来自于增加自有的标记或未标记数据（更易于收集） 3	包括以上所有流程，在对外部数据进行训练时，需要开展彻底的法律审查，以防止发生知识产权问题 4

资料来源：麦肯锡分析



改变软件工程师工作

第一个范例的复杂度相对较低,使用现成的生成式AI解决方案,不需要内部定制,因而可以立即提高生产力。

软件工程师绝大部分工作内容是编写代码。这个过程劳神费心,需要大量的试错以及对私域和公域文件的研究。某公司由于缺少足够的熟练软件工程师,功能和错误修复请求被大量积压。

为了提高工程师的工作效率,该公司使用了一款基于AI的代码补全产品,该产品被整合到工程师的编码软件之中。工程师可以使用自然语言撰写代码描述,而AI提供若干满足该描述要求的代码块变体。工程师可以从做出选择,进行必要改进,然后点击插入代码。

我们的研究表明,这类工具可以让开发人员生成代码的速度提升高达50%。它还可以帮助调试纠错,提高开发产品质量。不过,生成式AI目前并不能取代熟练的软件工程师。事实上,较有经验的工程师从这些工具中获益最多,工作效率提升幅度最大,该产品对缺乏经验的开发者则效果一般,有时甚至有负面影响。其中一项已知风险在于,AI生成的代码可能包含漏洞或其他错误,因此必须有软件工程师亲身参与,以确保代码的质量和安全性(见本文最后一节,了解减轻风险的方法)。

这一现成的生成式AI编码工具成本相对较低，上市所需时间短，因为该产品已经可用，不需要进行大量的内部开发。具体成本因软件供应商而异，使用费为每户每月10到30美元不等。在选择工具时，企业一定要向供应商了解许可和知识产权问题，以确保生成的代码不会违规。

这款新工具需要由一支跨职能的小团队负责支持，他们主要负责选择软件供应商并监控性能，其中也包括检查知识产权和安全问题。工具实施只需工作流程和政策方面的变更。由于该工具完全是现成的软件即服务 (SaaS) 类型，额外的计算和存储成本极低或为零。



帮助客户经理及时了解公共信息和数据

有些企业可能决定利用基础大模型（通过API或开放模型）构建自己的生成式AI应用，而非使用现成工具。其投资要求高于上一范例，但这样做有助于以更加定制化的方法满足公司的具体环境和需求。

在本例中，一家大型对公银行希望利用生成式AI提高客户经理的工作效率。客户经理为及时了解客户情况和当前重点，需要花大量时间阅读企业年报和业绩发布会记录等篇幅庞大的文件。这一工作让客户经理能够为客户提供契合其特定需求的服务。

该银行决定通过API接通基础大模型,构建解决方案。该解决方案可以快速查阅文件并为客户经理的提问总结答案。银行围绕基础大模型构建了额外软件层,以优化用户体验、实现工具与公司系统的整合并进行风险与合规控制。由于一些大型语言模型会产生“幻觉”误导,因而尤其需要对模型输出进行检查,正如金融机构会检查初级分析师的工作输出。客户经理也要接受培训,学会用适当的提问方式获得最精准的回复(也称“提示工程”)。相应的工具输出和信息来源的简化验证流程也已落实到位。

在本例中,生成式AI可以加快客户经理的分析过程(从几天缩短到几小时),提高工作满意度,并有机会捕捉到客户经理可能会忽视的想法。

开发成本主要集中在用户界面的构建和集成工作,这一工作需要数据科学家、机器学习工程师或数据工程师、设计师和前端开发人员投入时间。运营支出包括软件维护以及使用API的费用。具体成本取决于模型选择和第三方供应商费用、团队规模以及达到最简可行产品所需的时间。

减少客户服务用时,让客服代表有时间去做更有价值的工作

更复杂一些的应用是对基础大模型进行微调。在本例中,一家公司采用了针对对话场景进行优化的基础大模型,并使用自有的高质量客户聊天记录及行业特有的问答对基础大模型进行微调。公司所处行业会使用专业术语(例如法律、医学、房地产和金融)。客户服务速度是竞争优势的重要来源。

这家公司的客户服务代表每天要处理数百个来电咨询,响应时间有时过长,导致用户不满意。该公司决定引入生成式AI客服机器人来处理大部分客户需求。其目标是以符合公司品牌和用户偏好的方式回应客户。微调和测试基础大模型的一个环节是确保回复与公司设

定的领域术语、品牌承诺和风格基调保持一致；这需要进行持续监控以评估系统在包括顾客满意度等多个维度上的表现。

该公司创建了多阶段的产品路线图，以最大程度减少潜在的模型错误。第一阶段，公司对聊天机器人进行了内部试点。员工可以给模型的建议“点赞”或“拍砖”，而模型能够从这些输入中学习。下一步，模型“旁听”客户服务对话并提供建议。当技术经过充分测试后，第二阶段开始，模型直接面向客户使用，并保持一名真人员工参与。最终，当公司领导者对技术有十足信心时，便可以实现大面积自动化。

在本例中，生成式AI解放了客服代表，使其能够专注于价值更高和更复杂的客户咨询工作，既提高了员工效率和工作满意度，也提升了服务水平和客户满意度。该机器人可以访问所有的客户内部数据，并能“记住”先前的对话（包括电话通话），明显领先于现有的客户聊天机器人。

为实现效益，该用例需要在软件、云基础设施和技术人才方面进行较大投资，并在风险和运营方面进行更高水平的内部协调。一般而言，微调基础大模型的成本是借助API建立一个或多个软件层的两到三倍。云计算所需的人才和第三方成本（若微调自托管模型）或API费用（若通过第三方API微调）是主要的成本增量。为实施解决方案，公司需要数据运营和MLOps专家的帮助，也需要如产品管理、设计、法务和客户服务专家等其他职能部门的输入。





加速药物发现

当没有合适的基础大模型可用,公司需要从头建立模型时,就会出现最复杂且定制化程度最高的生成式AI用例。这一情形可能会出现在专业性较强的行业,或者在所处理数据集与现有基础大模型训练所用数据集大不相同的情况下,接下来介绍的医药行业用例正属此类。从头训练基础大模型会伴生技术、工程和资源方面的巨大挑战。使用性能更高的模型所带来的额外投资回报应当超过相应的财务和人力资本成本。

在本例中,一家制药公司的药物发现研究人员必须根据显微镜图像决定后续实验。研究者们拥有百万量级的图像数据集,其中包含了大量与药物发现有关、人眼难以解读的细胞特征视觉信息。这些图像用以评估潜在的候选治疗方法。

该公司决定创建一款工具,帮助科学家了解药物化学与显微镜图像记录之间的关系,以加速研发工作。由于这种多模态模型仍处于起步阶段,该公司决定训练自有模型。为了建立这一模型,团队成员同时使用了训练图像类基础大模型的真实世界图像和公司内部的庞大显微镜图像数据集。

训练后的模型能够预测可能导致有利结果的候选药物，并能提高精准识别相关细胞特征的能力，为药物发现增加价值。这让药物发现过程更高效、更有效，不仅缩短了价值实现时间，还减少了不准确、具有误导性或最终失败的分析次数。

一般而言，从零开始训练模型的成本是围绕模型API创建软件所需成本的10到20倍。更大规模的团队（包括博士水平的机器学习专家）以及更高的计算和存储支出是主要的成本增量。训练基础大模型的预计成本因所需的模型性能水平和建模复杂性而存在较大差异。上述因素影响到所需的数据集大小、团队组成和计算资源。在本用例中，工程团队和持续的云服务费用耗费了大部分成本。

该公司发现需要对技术基础设施和流程进行重大升级，包括访问多个GPU实例以训练模型，获取各类工具以在多个系统间调配训练，以及部署MLOps最佳实践来限制成本和项目持续时间。另外，要对数据进行收集、整合（确保不同数据集的格式和分辨率一致）和清理（过滤低质量数据，删除重复数据，并确保数据分配符合预期用途），这涉及大量处理工作。由于基础大模型从零开始训练，因此最终模型需要经过严格测试，以确保输出准确、使用安全。

上述案例对CEO的关键启示

上述用例为CEO们应用生成式AI带来一系列启示：

为工作和职场带来切实效益的变革性用例已然存在。从制药业、银行业到零售业，众多企业正在建立一系列用例，以捕捉价值创造潜力。具体入手点可大可小，取决于企业的目标抱负。

应用生成式AI的成本差异很大，影响因素包括用例和软件所需数据、云基础设施、技术专识和风险缓解措施。无论用例如何，企业必须考虑风险问题，在此，某些用例会需要更多资源投入。

快速起步有其优势，率先建立起基本的业务逻辑将帮助公司更好应用生成式AI。

起步时需考虑的因素

CEO在推动企业关注生成式AI方面发挥着重要作用。在结尾部分,我们将探讨CEO在踏上征途时需要熟记的策略,其中有很多与过往技术浪潮兴起时企业高管应当做出的反应一致。然而,生成式AI也带来了独有的挑战,这包括其超越以往技术变革的空前发展速度及随之而来的应对难度。

跨职能部署生成式AI

过去,许多组织以孤立试验的方式启动对传统AI的探索。然而,鉴于生成式AI独特的风险考量以及基础大模型支撑全组织、多用例的能力,企业应当以更加精细和协调的方法加以管理。例如,使用专有数据进行微调以反映企业品牌特质的模型,可以在多个用例(如生成个性化的营销活动和产品描述)和多个业务职能(如产品开发和营销)中加以部署。

为此,我们建议组建由公司领导组成的跨职能小组(例如,代表数据科学、工程、法务、网络安全、营销、设计和其他业务职能)。这样的跨职能小组不仅可以发现并优先处理价值最高的用例,还能保证整个组织的执行协调一致且安全。

端对端领域重塑,而非仅专注于用例生成式AI是可以改变组织运作方式的强大工具,对价值链中的特定业务领域(例如,零售商的营销或制造商的运营)具有特别的影响。生成式AI的部署十分便利,企业因而很容易将应用局限于全盘业务下的零星用例。因此,坚持全盘视角、按领域划分用例群至关重要,从这一视角出发进行规划能为全体业务职能带来最大的变革潜力。随着生成式AI与其他传统AI应用协同并进、此前无法实现的工作方式逐渐诞生,企业正在不断重新定义其目标状态。

实现满载的技术栈

现代化的数据和技术堆栈几乎是任何生成式AI策略成功的基础。CEO们应当向首席技术官了解以确定公司在计算资源、数据系统、工具和模型访问（通过模型中心的开源方式或通过API的商业模式）方面是否具备所需的技术能力。

例如，生成式AI的命脉是流畅获取具体业务背景或问题的相应数据。无法有效协调或随时使用自身数据的公司，将无法微调生成式AI以探索更多潜在的变革性用途。设计可扩展的数据架构同样重要，这包括数据治理和安全流程。根据不同的用例情况，现有的计算和工具配置基础设施（可以通过云供应商采购或在内部建立）可能也需要升级。明确的数据和基础设施战略应立足于从生成式AI中获得的业务价值和竞争优势。上述因素都至关重要。

打造“灯塔”

CEO们需要避免在规划阶段止步不前。新的模型和应用正迅速被开发和发布。例如，GPT-4在2023年3月发布，此前ChatGPT（GPT-3.5）发布于2022年11月，而GPT-3则是在2020年。在商业世界，时间尤为重要，而生成式AI技术的快节奏特性要求企业迅速行动以把握优势。CEO们可以采取下述几种方式稳步推进。

尽管生成式AI仍处于早期阶段，但应尽快在企业内部展示其对运营模式的重要影响，这可以通过“灯塔方法”来实现。例如，一种推进方式是打造“虚拟专家”，让一线员工能够利用专有的知识源，为客户提供最相关的内容。这一方法能提高生产力、激发热情，并使企业能够在向客户扩展应用之前，在内部对生成式AI展开测试。

和其他技术创新浪潮一样，“概念验证疲劳”将会出现，许多公司会陷入“试点炼狱”的困境。然而，鼓励概念验证仍然是快速测试和完善有价值业务用例、以便日后向邻近用例扩展的不二选择。通过聚

焦产生有益结果的早期成功实例,企业可以积攒势头,以此为基础扩大规模,充分发挥生成式AI的多功能性。这一方法可以帮助企业推动更广泛的AI采用,营造创新文化,从而保持竞争优势。如上所述,企业需要组建跨职能领导团队,确保有计划地协调推动概念验证。

平衡风险与价值创造

正如此前详细介绍的4个范例所示,商业领袖必须在生成式AI所涉及的价值创造机会与风险之间取得平衡。我们近期的全球AI调查显示,尽管已有超过一半的组织采用传统AI技术,但大多数组织并未对相关的大部分风险采取应对措施³。生成式AI再次引发了人们对许多同类风险的关注,比如AI可能会让隐藏在训练数据中的偏向被固化;同时生成式AI还带来了一些新风险,比如产生幻觉的倾向。

因此,跨职能领导团队不仅要为生成式AI的使用建立总体道德原则和指导方针,还要对每个潜在用例所伴生的风险有全面了解。

重要的是,要寻找与组织的整体风险容忍度相一致的初始用例,并且设置相应结构以减轻相应风险。例如,零售组织可以优先考虑价值稍低但风险也较小的用例,比如创建营销内容初稿和其他需要人工参与的任务;同时,可能搁置价值更高、但风险更大的用例,比如自动起草和发送高度个性化营销电邮的工具。这种以风险为考量的做法能够让企业建立必要的控制机制,妥善管理生成式AI并保持合规。

CEO与其团队还要密切关注生成式AI监管的最新动态,包括与消费者数据保护和知识产权相关的规定,以保护公司远离法律纠纷。正如目前对AI和数据的监管,各国对生成式AI所采取的监管方式可能各不相同。企业需要调整工作方法以校准流程管理、文化和人才管理方式,确保能够规模化地应对快速发展的监管环境和生成式AI的风险。

应用生态系统方法建立合作伙伴关系

商业领袖应注重建立和维护平衡得当的联盟网络。企业的收购和联盟战略应继续聚焦建立合作伙伴生态系统，以适应不同的场景并解决生成式AI对不同技术栈层面的需求，同时要注意避免锁定供应商。

与正确的公司合作可以加速推进执行。企业不需要自己建立所有的应用或基础大模型，而是可以与生成式AI供应商和专家合作，更快地采取行动。例如，企业可以与模型供应商合作，为特定行业定制模型，或与提供可扩展云计算等能力的基础设施供应商合作。

企业可以借助他人的专业知识，让最新的生成式AI技术迅速地为我所用。然而，生成式AI模型只是冰山一角，价值创造还需要包括多个其他要素。

聚焦所需的人才和技能

为有效应用生成式AI创造商业价值，公司需要打造技术能力并提升现有员工的技能水平。这需要领导层共同努力，根据企业的优先用例确定所需能力，这可能不限于技术方面，也包括工程、数据、设计、风险、产品和其他业务职能的人才组合。

正如上文用例所示，技术和人才需求因具体实施的性质（从最简单的使用现成解决方案、到最复杂的从零建立基础大模型）而大不相同。例如，为了建立生成式模型，企业可能需要博士水平的机器学习专家；而如果要利用现有模型和SaaS产品开发生成式AI工具，一名数据工程师和一名软件工程师或许足以领导此工作。

除了雇用合适的人才，企业还需要培训现有员工。生成式AI应用程序使用基于提示的对话式用户界面，因而使用简便，但用户仍然需要优化提示输入，了解技术限制，并懂得在何时何地可以合理地把应用融入工作流程。领导层应提供使用生成式AI工具的明确指导方针，

并开展持续的教育和培训,让员工了解其风险。培养自发研究和实验的文化也可以激发员工创新流程和产品,从而有效整合AI工具。

多年来,企业一直在探索AI之路、追求宏远的目标,许多公司已经收获了新的收入来源、改进了产品、提升了运营效率。其中许多成功都源于AI技术,它们仍然是处理特定任务的最优工具,企业应继续保持这方面的努力。然而,生成式AI带来了又一次实现重大飞跃与无限可能的机会。虽然该技术的运营和风险框架还未完全建立,但商业领袖们深知应当启动生成式AI旅程。但从何开始,又如何开始?每家公司的答案不尽相同,公司内部也众说纷纭。有些企业放手一搏,而其他公司则从小规模试验起步。最佳方法将取决于公司的目标抱负和风险偏好。但无论目标如何,关键是要迈出第一步,边做边学。

¹ 更多信息请参阅麦肯锡2023年4月26日《在生成式AI价值链中探索机遇》

² 这些范例来自我们客户工作和公开案例的汇总整理,而非反映特定公司的具体事例。

³ 2022年12月6日麦肯锡《2022年AI现状及5年回顾》

Michael Chui是麦肯锡全球研究院董事合伙人,常驻湾区分公司;

Roger Roberts是麦肯锡全球董事合伙人,常驻湾区分公司;

Tanya Rodchenko是麦肯锡全球副董事合伙人,常驻湾区分公司;

Lareina Yee是麦肯锡技术委员会主席、麦肯锡全球资深董事合伙人,均常驻湾区分公司;

Alex Singla是麦肯锡全球资深董事合伙人,常驻芝加哥分公司;

Alex Sukharevsky是全球资深董事合伙人兼Quantumblack, AI by McKinsey全球负责人,常驻伦敦分公司;

Delphine Zurkiya是麦肯锡全球资深董事合伙人,常驻波士顿分公司。



生成式AI引发科技变革： CIO和CTO参考指南

企业CIO和CTO可以采取九大举措，利用生成式AI重塑自身业务与科技。

Michael Chui, Eric Hazan, Lareina Yee, Bryce Hall, Alex Singla
和Alexander Sukharevsky

几乎每一天，都会有媒体报导生成式AI催生的新业务革命。其热度之高也不难理解——麦肯锡研究发现，生成式AI每年创造的价值高达2.6万亿~4.4万亿美元。¹

要实现上述价值，首席信息官（CIO）和首席技术官（CTO）的作用至关重要。但我们不应忘记，类似故事也曾多次上演。每当新技术出现，例如互联网、移动技术和社交媒体，各类试验和试点项目便纷至沓来，而事实又反复证明，要实现显著商业价值往往困难重重。过往变革的经验教训如今依然适用，特别是在新技术从试点向规模化过渡的阶段。生成式AI的爆发式增长为企业CIO和CTO创造了独特机会，他们应在过往经验教训基础上，为企业领导层指引道路，真正将生成式AI的潜力转化为可持续的业务价值。

在与数十位科技企业领袖深度对话并分析了50多家企业（包括麦肯锡自身）的生成式AI举措后，我们为科技企业领袖列出了九大举措，用以创造价值、协调技术和数据、扩展解决方案并管理生成式AI的风险（参见边栏“关键术语表”）：

1. **迅速行动，确定公司对采用生成式AI的态度**，与员工切实沟通，为他们提供可靠的技术获取途径。
2. **重构业务，围绕生产力、增长和新业务模式制定价值创造用例**。开发“财务人工智能”（FinAI）能力，估算生成式AI的真实成本和回报。
3. **重塑技术职能**，聚焦在软件开发中快速构建生成式AI能力，加速降低技术负债，并大幅减少IT运营中的人工作业。
4. **充分利用现有服务或改编生成式AI开源模型**，开发专属能力（独立构建和运营生成式AI模型可能需要投入数千万到数亿美元，至少短期内如此）。
5. **升级企业技术架构，集成管理生成式AI模型**，合理编排其彼此之间，以及与现有人工智能和机器学习（ML）模型、应用程序和数据源的协作运行方式。

6. **开发数据架构**, 处理结构化和非结构化数据源, **获取高质量数据**。
7. **创建中央跨职能生成式AI平台团队**, 按需向产品和应用程序团队提供已批准的模型。
8. 通过投资提升关键岗位(软件开发人员、数据工程师、MLOps工程师和安全专家)以及更多非技术员工的技能。然而, 由于生成式AI对不同职能的影响各异, 需**根据具体岗位及员工的熟练程度定制技能提升计划**。
9. **评估最新风险情况, 制定持续性风险缓释措施**, 解决模型、数据和策略问题。

关键术语表

生成式AI属于人工智能技术, 通过机器学习(ML)技术对大量(公开)数据进行训练、习得规律并创建新内容(文本、代码、图像、视频)。

基础模型(FM)是基于大量非结构化、无标签数据训练的深度学习模型, 可以直接用于广泛的任务, 也可以通过微调适应特定任务。GPT-4、PaLM 2、DALL·E2和Stable Diffusion便属于这类模型。

大型语言模型(LLM)是一类基础大模型, 可处理大量非结构化文本, 学习单词或词组(称为token)之间的关系。这使得大语言模型能够生成自然语言文本, 执行总结或提取知识点等任务。Cohere Command属于大语言模型; LaMDA则是谷歌Bard背后的大语言模型。

微调(Fine-tuning)是指调试预训练模型以使其更好地处理特定任务的过程。这需要在相对较短的时间内, 通过标记的数

据集进行训练，这个数据集比最初训练模型的数据集要小得多。这一额外训练使模型能够学习并适应较小数据集中的细微差异、术语和特定规律。

提示工程 (Prompt engineering) 是指设计、改进和优化输入提示以引导生成式AI模型产生所需（即准确）输出的过程。

1、明确公司对生成式AI应用的态度

随着生成式AI的应用日益普及，部分企业CIO和CTO为防范风险，禁止员工使用公开App。然而企业可能因此错失创新机遇，员工甚至会认为这样会阻碍其习得重要的新技能。

相反，CIO和CTO应该与风险部门负责人一起努力，思考并兼顾企业培养生成式AI技能的重要意义以及降低风险的实际需要。这就需要企业明确对应用生成式AI的态度，围绕企业能够接受的风险水平，就如何将生成式AI纳入企业整体战略达成共识。这样一来，企业就能迅速确定全局战略和指导方针。

在明确政策之后，领导者应将其传达给全公司，由CIO和CTO为员工提供可靠的新技术获取途径，以及好用易懂的指导方针。有些企业已经就生成式AI展开全面沟通，为具体用户群提供广泛的访问权限，并通过弹出式窗口，在用户向模型输入内部数据时发出警告，还创建了指导页面，在用户每次访问公开生成式AI服务时给出提示。



2、识别价值创造用例，通过提高生产力、促进增长和打造新业务模式创造价值

CIO和CTO应帮助企业避免陷入狂热的“用例陷阱”。最有益的做法是与CEO、CFO和其他业务领导合作，仔细思考生成式AI如何挑战现有业务模式，如何为新业务模式打开大门、创造新价值来源。在深入了解技术可能性后，CIO和CTO应全盘扫描生成式AI能够为企业带来最大价值的机会和问题以及低价值领域。因为有时生成式AI并非最佳选择。

例如，麦肯锡的研究表明，生成式AI可以将某些营销用例（例如，通过分析非结构化和抽象数据以了解客户偏好）的生产率提高约10%，并将客户支持（例如，通过智能机器人）的生产率提高达40%²。CIO和CTO能够帮助企业树立大局观，按领域（如客户旅程或业务流程）或用例类型（如创意内容创建或虚拟坐席）实现最优的用例分群，从而让生成式AI发挥最大价值。然而，单纯寻找机会并非最重要的战略工作，现有生成式AI用例比比皆是。CIO和CTO的工作重点是根据企业初始人才和能力限制，进行可行性和资源估算，帮助业务部门确定生成式AI的优先级排序。

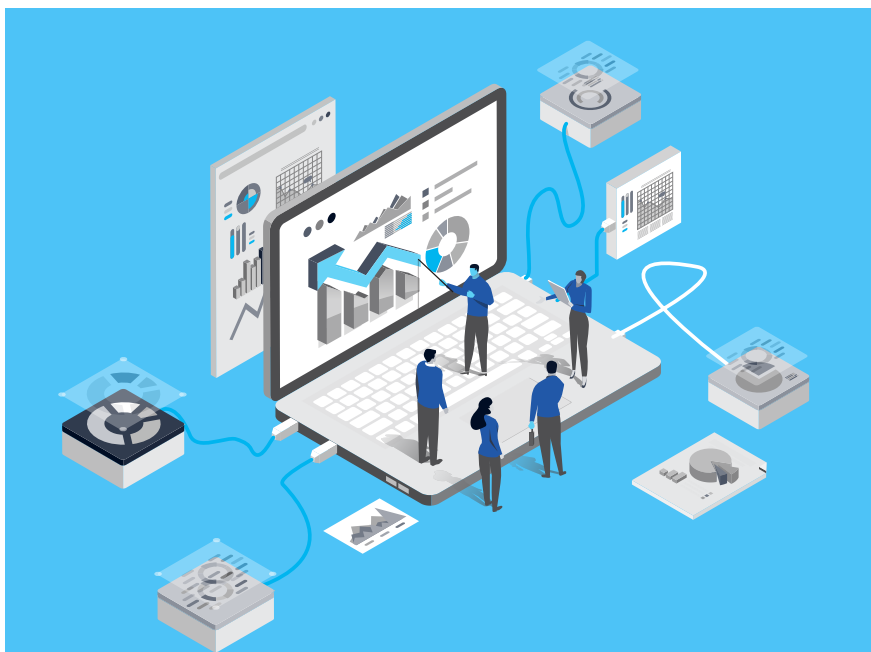
为提供这样深入的咨询建议，技术领导必须与业务部门合作，共同开发FinAI能力，估算生成式AI项目的真实成本和回报。成本计算将非常复杂，因为单位经济效益必须考虑各模型和供应商成本、模型间的相互作用（例如一次查询可能需要多个模型输入，每个模型都产生费用）、持续使用费以及人工监督成本。

3、重塑技术职能

生成式AI有可能彻底改变技术职能部门的工作方式。CIO和CTO需要全面审视生成式AI对所有技术领域的潜在影响，但关键是迅速行动，积累经验和专业知识。

初期工作可以集中在以下三个领域：

- **软件开发：**麦肯锡研究表明，生成式AI的编码支持能力可将代码开发速度提高35%至45%，将代码重构速度提高20%至30%，将代码文档编写速度提高45%至50%³。生成式AI还可帮助实现测试流程自动化，模拟边缘场景（例如，模拟异常情况以进行预警），让团队在发布前开发出更有韧性的软件，并使新开发人员更快熟悉业务（例如，通过向生成式AI提出有关代码库的问题）。发挥上述优势需要企业开展广泛培训（更多内容请参见第8项举措），并通过DevSecOps实践，实现开发管线集成和部署的自动化，从而管理激增的代码量。
- **技术负债：**技术负债可能占到技术预算的20%至40%，并大大减缓开发速度⁴。CIO和CTO应审查其技术债的资产负债表，确定代码重构、代码翻译和测试用例自动生成等生成式AI功能如何帮助企业快速减少技术负债。
- **IT 运营 (ITOps)：**CIO和CTO需要重新审视ITOps生产率，确定如何利用生成式AI加快流程。生成式AI以下方面尤为有益：通过自助服务坐席自动实现密码重置、状态查询或基本诊断等任务；通过优化路径选择，加快问题分流和解决；浮现主题或优先级等有用情境，并生成推荐响应；通过分析大量日志流提高可观察性，识别真正需要关注的事件；开发文档，如标准操作程序、事件后记或性能报告。



4、充分利用现有服务或改编生成式AI 开源模型

当涉及制定战略以发展生成式AI能力时，经典的“租用、外购或自建”决策可进行相应调整。基本原则不变：对于能够让企业获得独家优势的领域，公司应投资培养生成式AI能力，而对于大宗商品类，则可利用现成服务。

CIO和CTO可将不同选项分为三种模式进行思考：

- **取用者 (Taker)** ——通过聊天界面或应用程序接口 (API) 使用公开模型，很少或无需定制。典型例子包括用于生成代码的现成解决方案（如，GitHub Copilot），或协助设计师生成和编辑图像的解决方案（如，Adobe Firefy）。这一模式对工程和基础设施要求最低，通常能最快投入使用。这些模型本质上类似于大宗商品，通过提示 (prompt) 向公开模型传送数据。

- **塑造者 (Shaper)** ——将模型与内部数据和系统整合，生成更定制化的结果。其中一个例子是将生成式AI工具与客户关系管理 (CRM) 及财务系统关联，将客户的过往销售和参与记录纳入其中，由此支持销售交易。另一种是利用公司内部文件和聊天记录对模型进行微调，协助客户支持专员的工作。对于希望扩展生成式AI能力、开发更多专属功能，或满足更高安全性或合规性需求的公司而言，塑造者模式更为适用。

这一模式通过两种常见方法将数据与生成式AI模型集成。一种是“让模型找数据”，即模型托管在企业基础设施上，可以是在企业本地，或者在云环境中。例如，Cohere在客户的云基础设施上部署基础模型，从而减少数据传输需要。另一种方法是“让数据找模型”，企业可以将其数据汇总，并在云基础设施上部署大模型副本。这两种方法都能实现访问基础模型的目的，最终的选择取决于企业在何处放置其工作负载。

- **创造者 (Maker)** ——建立自己的基础模型，针对极具个性的业务案例。打造基础模型既昂贵又复杂，需要大量数据、深厚的专业知识和强大的计算能力。该方案需要一次性大量投入以构建和训练模型，有时可达数千万甚至数亿美元。成本取决于多种因素，如训练基础设施、模型架构选择、模型参数数量、数据规模和专家资源等。

每种模式都有成本，需要技术领导斟酌 (图1)。虽然技术随时间的推移不断进步，更高效的模型训练方法、更低的图形处理器 (GPU) 计算成本等在不断拉低成本，但创造者模式固有的复杂度意味着短期内很少会有组织问津。大多数企业会寻求一方面通过取用者模式快速获取服务，同时利用塑造者模式在基础模型上构建专属能力。

图1: 每种模式都有对应的成本，需要技术领导斟酌

类型	示例用例	总拥有成本(估算)
取用者 (Taker)	• 面向软件开发人员的现成编码助手 • 通用的客户服务聊天机器人，仅具有提示工程和文本聊天	• ~ 50 万至 200 万美元，一次性费用 — 现成的编码助手：~50 万美元用于集成。费用包括一个 6 人的团队工作 3 到 4 个月 — 通用的客服聊天机器人：~200 万美元用于在第三方模型 API 之上构建插件层。费用包括一个 8 人团队工作 9 个月 • ~ 50 万美元每年，重复性费用 — 模型推断：现成的编码助手：每年 ~20 万美元，每日 1,000 个用通用的客服聊天机器人：~每年 20 万美元，假设每天 1,000 个客户聊天和每次聊天 10,000 个代币 — 插件层维护：每年 ~20 万美元，假设开发成本的 10%
塑造者 (Shaper)	• 根据特定行业知识和聊天记录进行微调的客服聊天机器人	• ~ 200 万到 1000 万美元，一次性费用，除非模型需进一步微调 — 数据和模型管道构建：~50 万美元。成本包括 5 到 6 名机器学习工程师和数据工程师，他们工作 16 到 20 周，收集和标记数据并执行数据 ETL¹ — 模型微调²：~每次训练运行 10 万到 600 万美元³ — 低成本情景：成本包括计算（例如算力），和 2 名数据科学家工作 2 个月的成本 — 高成本情景：基于公共开源模型的计算和微调成本 — 插件层构建：~100 万美元到 300 万美元，成本包括一个 6 到 8 人的团队工作 6 到 12 个月 • ~ 50 万美元每年，重复性费用 — 模型推理：每年重复性，~50 万美元。假设每天有 1,000 次音频和文本聊天 — 插件层维护：~50 万美元。假设 MLOps 平台每年花费 100,000 到 250,000 美元⁴，1 名机器学习工程师花费 50% 到 100% 的时间监视模型性能 — 插件层维护：每年 ~30 万美元，占开发成本 10%
创造者 (Maker)	• 经模型，用于协助患者诊断过培训的基础	• 500 万到 2 亿美元，一次性，除非模型经过微调或重新训练 — 模型开发：~50 万美元。成本包括 4 名数据科学家花费 3 到 4 个月的时间利用现有研究进行模型设计、开发和评估 — 数据和模型管道：~50 万到 100 万美元。成本包括 6 到 8 名机器学习和数据工程师工作 ~12 周来收集数据并执行数据 ETL — 模型训练⁵：每次训练运行 ~400 万美元到 2 亿美元。成本包括 4 到 6 名数据科学家工作 3 到 6 个月的计算和人工成本 — 插件层建筑：~100 万到 300 万美元。费用包括一个 6 到 8 人的团队，工作 6 到 12 个月 • ~100 万至 500 万美元，每年重复性费用 — 模型推断：~10 万到 100 万美元每年每 1,000 个用户 假设每位医生每天看 20 到 25 名患者，患者每次就诊 6 到 25 分钟 — 模型维护：~每年重复 100 万美元到 400 万美元。假设 MLOps 平台每年花费 250,000 美元⁴和 3 到 5 名机器学习工程师来监视模型性能 — 插件层维护：每年 ~30 万美元，占开发成本 10%

1 Extract, transform, and load (ETL), 即提取、转换和加载

2 模型在由 ~100,000 页行业特定文档和来自 ~1, 000 个客户代表的 5 年聊天记录 (~480 亿个代币) 组成的数据

集上进行微调。较低的终端成本包括在开源模型（例如 LLaMA）上重新训练的 1% 参数和闭源模型的上限参数。

聊天机器人可以通过文本和音频访问

3 每次训练运行后，都会根据超参数、数据集和模型架构的使用对模型进行优化。模型可以在需要时定期刷新（例如，使用新数据）

4 Gilad Shaham, "Build or buy your MLOps platform: Main considerations," LinkedIn, 2021/11/3

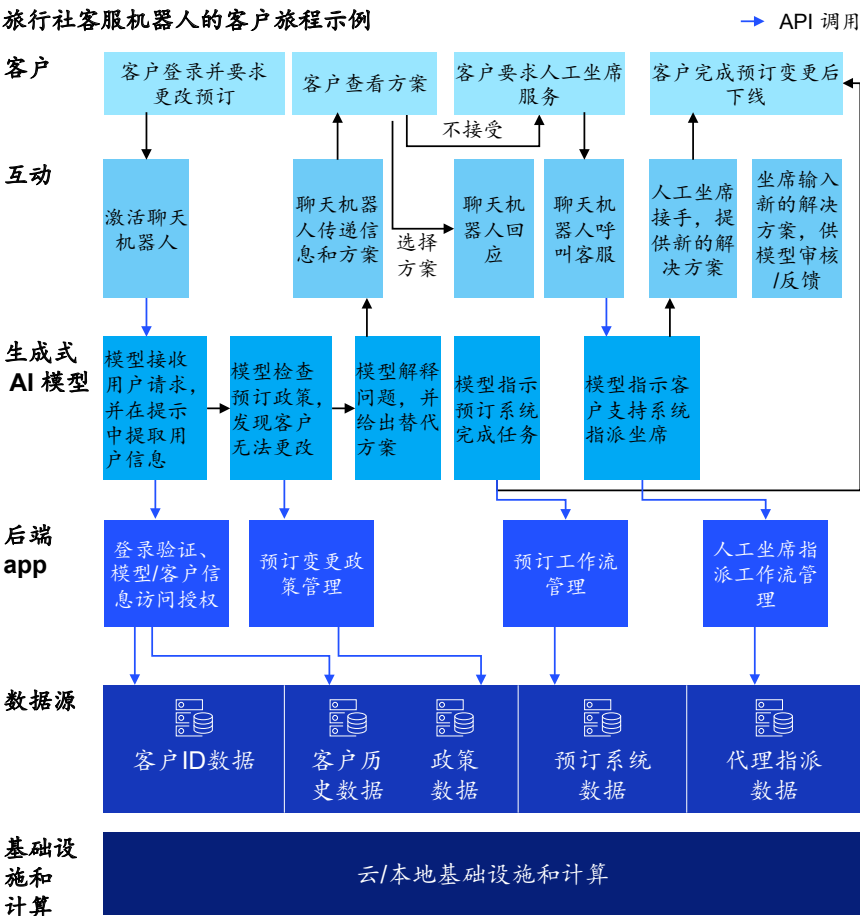
5 模型在 650 亿到 1 万亿个参数和 1.2 到 2.4 万亿个代币的数据集上进行训练。该工具可以通过文本和音频访问

资料来源：麦肯锡分析

5、升级企业技术架构，集成管理生成式AI模型

企业将使用规模、复杂度和能力各不相同的多种生成式AI模型。这些模型必须能够与彼此、以及企业现有系统或应用程序协同工作，才能产生价值。因此，为生成式AI单独构建技术堆栈只会增加复杂度。我们可以看看消费者向某旅行社客服人员询问解决预订问题的事例(图2)。在与客户互动时，生成式AI模型需要访问多个应用程序和数据源。

图2: 在关键触点集成生成式AI，实现客户旅程的量身定制



资料来源：麦肯锡分析

对于“取用者”模式，没有必要进行如此复杂的协调工作。然而，企业如果希望作为取用者或塑造者拓展生成式AI的优势，其CIO和CTO就需要升级技术架构。首要目标是将生成式AI模型集成到内部系统和企业应用中，并为各种数据源建立工作流管线。归根结底，企业技术架构的成熟度决定了其整合和扩展生成式AI的能力。

LangChain和LlamaIndex等最新的集成和编排框架大大减少了将不同生成式AI模型与其他应用和数据源关联的工作量。此外还出现了几种集成模式，让模型在响应用户查询时调用API（例如GPT-4可以调用函数），并通过外部数据集提供情境数据，协助用户查询，这一技术被称为检索式增强生成方法（Retrieval Augmented Generation）。技术负责人需要为组织确定参考架构和标准集成模式（如，标准API格式和参数，以识别调用API的用户和模型）。

要有效整合生成式AI，技术架构中必须包含五大要素（图3）：

情境管理和缓存为模型提供来自企业数据源的相关信息。只有在正确的时间访问相关数据，模型才能理解情境，产生令人信服的输出结果。缓存可存储常见问题的处理结果，以便更快、以更低成本做出响应。

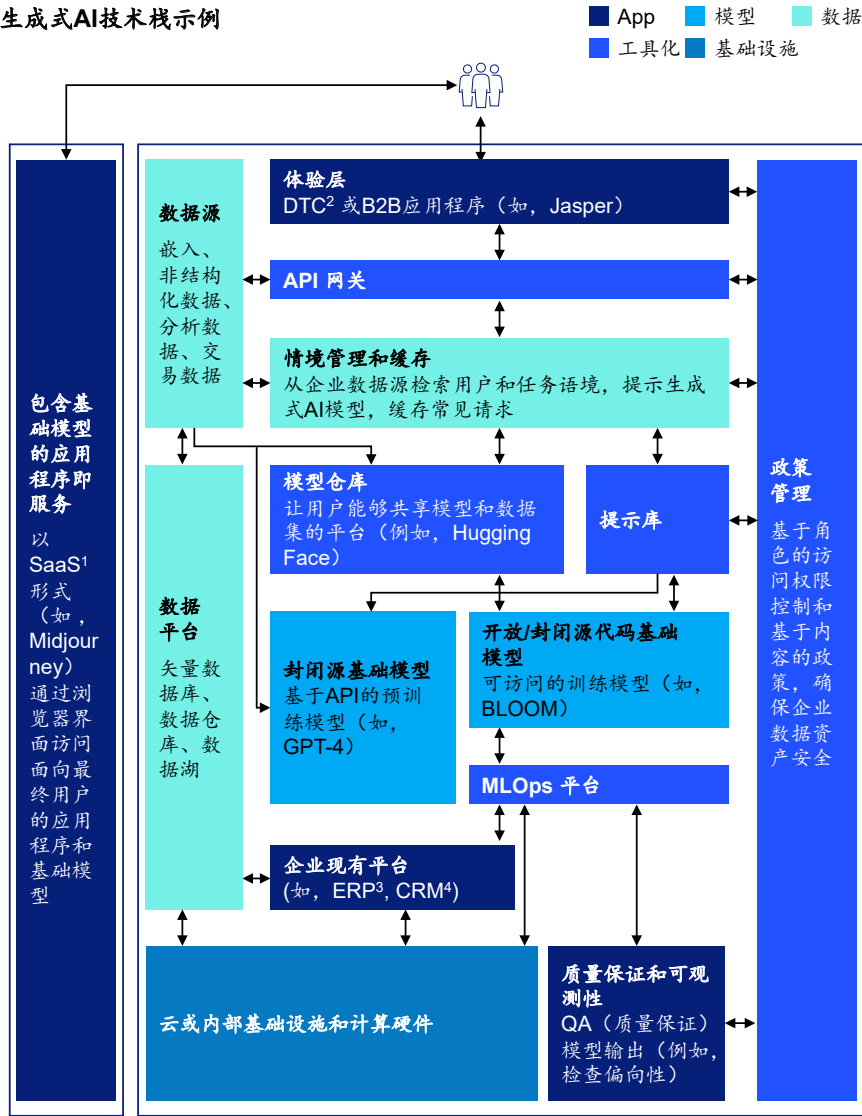
策略管理确保对企业数据资产的合理访问。这一控制机制可确保人力资源部门的生成式AI模型不会被组织的其他部门访问，例如包含员工薪酬详情的模型。

模型仓库汇集了经过训练和批准的模型，可按需开通服务，并充当模型检查点、权重和参数的存储库。

提示库汇集了对生成式AI模型做出的优化指令，包括模型更新时的新版提示。

MLOps平台，包括升级的MLOps能力，以应对复杂的生成式AI模型。例如，MLOps管道需要有工具（instrumentation），用于衡量特定任务性能，如测评模型检索正确知识的能力。

图3：生成式AI技术栈正在形成



- 1. 软件即服务
- 2. 直接面向客户
- 3. 企业资源规划
- 4. 客户关系管理

资料来源：麦肯锡分析

在发展架构的过程中, CIO和CTO需要管理由生成式AI供应商和工具组成的快速发展的生态系统。云供应商提供多种多样的大规模硬件和基础模型, 以及不断丰富的成套服务。同时, MLOps和模型仓库供应商提供了工具、技术和实践, 用于调整某个基础模型并将其部署到生产中。其他公司则提供建立在基础模型之上、用户可直接访问的应用程序, 用以执行特定任务。CIO和CTO需要评估如何装配和整合这些不同的技术能力, 从而有效部署和运行生成式AI模型。

6、开发数据架构, 获取高质量数据

企业能否通过生成式AI模型创造并扩大价值, 例如降低成本、改进数据、保护知识等, 将取决于对自身数据的利用。要取得优势, 就必须建立良好的数据架构, 将生成式AI模型与内部数据源关联, 后者可提供必要情境或帮助调整模型, 从而产生更有效的输出。

为此, CIO、CTO和首席数据官必须协力开展以下工作:

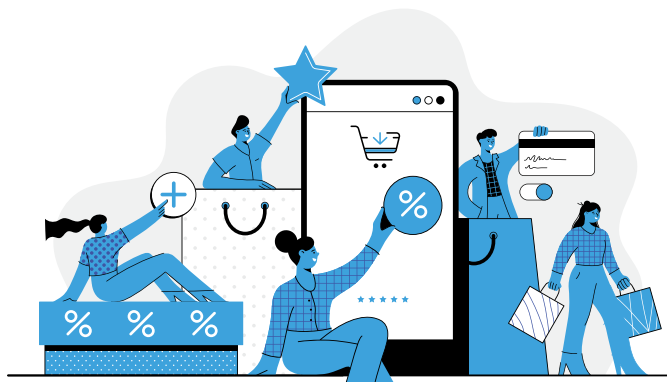
- 分类和组织数据, 便于生成式AI模型使用。技术负责人需要开发综合数据架构, 囊括结构化和非结构化数据源。这需要先制定标准和指导原则, 优化数据以便生成式AI使用——例如, 用合成样本增强训练数据, 提升多样性和规模; 将不同类型的媒体转换为标准数据格式; 添加元数据, 提高可追溯性和数据质量; 以及更新数据。
- 确保现有基础设施或云服务能够支持存储和处理大量数据, 以供生成式AI应用程序使用。
- 优先开发数据管道, 将生成式AI模型与提供“语境理解”的相关数据源关联。新兴方法包括使用矢量数据库来存储和检索嵌入向量 (embeddings, 即特殊格式的知识), 作为生成式AI模型的输入; 以及语境学习方法, 如“小样本提示”, 即为模型提供回答范例。

7、创建跨职能生成式AI平台团队

大多数科技企业都在迈向某种产品和平台运营模式。企业CIO和CTO需要将生成式AI能力整合到该运营模式中，发展现有基础设施，快速扩展生成式AI的应用规模。第一步是建立生成式AI平台团队，其核心任务是开发和维护平台服务。通过这个平台，可以按需为产品和应用程序团队开通经批准的生成式AI模型。平台团队还制定协议，规定生成式AI模型与内部系统、企业应用程序和工具的集成方式，并开发和实施标准化方法（如负责任的AI框架）来管理风险。

企业CIO和CTO需要确保平台团队成员具备相应技能。团队需要一名高级技术领导担任总经理。其他主要角色包括：软件工程师，负责将生成式AI模型集成到现有系统、应用程序和工具；数据工程师，负责建立管道，将模型与各种记录系统和数据源关联；数据科学家，负责选择模型和设计提示；MLOps工程师，负责管理多模型和多版本的部署和监控；ML工程师，负责利用新数据源对模型进行调整；风险专家，负责管理数据泄漏、访问权限、输出准确性和偏向性等安全问题。平台团队的具体组成取决于全企业的用例使用情况。在某些情况下（比如要创建面向客户的聊天机器人），企业需要强有力的产品管理和用户体验（UX）人才和资源。

现实而言，平台团队最初需要处理的优先用例范围较窄，而随着可重复使用能力的培养，以及对能力效果的进一步了解，他们的工作范围会逐渐扩大。技术负责人应与业务负责人密切合作，评估哪些业务案例需要资助和支持。





8、根据具体岗位和熟练程度定制技能提升计划

生成式AI有可能大幅提高员工生产力，并增进其能力。但是，AI对不同岗位和技能水平的助益并不均等，这就要求领导者重新思考如何培养员工切实需要的技能。

例如，我们对生成式AI工具GitHub Copilot的最新实证研究发现，该工具可帮助软件工程师将代码编写速度提高35%至45%⁵。但对于不同水平的人员，其影响存在较大差异：资深开发人员的速度提高了50%到80%，而初级开发人员的速度却下降了7%到10%。这是因为生成式AI工具的输出结果需要工程师对代码进行点评、验证和改进，而缺乏经验的软件工程师很难做到这一点。相反，在客服等技术含量较低的岗位上，生成式AI对低技能水平员工的帮助非常明显，一项研究显示，生成式AI让生产率提高了14%，员工流失率也有所下降⁶。

上述差异凸显出技术领导需要与首席人力资源官 (CHRO) 合作，重新设计人才管理战略，打造一支未来之师。招募一流的生成式AI核心人才将至关重要，鉴于这些人才与日俱增的稀缺性和战略意义，技术领导应建立人才留存机制，如设计有竞争力的薪酬、提供参与企业重要战略工作的机会等。

然而，技术领导不应止步于人才招聘。几乎所有现存岗位都会受到生成式AI的影响，因此更应聚焦对不同角色、专业水平和业务目标的技能要求，并相应提升员工能力。以软件开发人员为例：新手的培训重点在于加快其成长为一流代码审查员和代码生成成员的速度。与写作和编辑的区别类似，代码审查也需要具备不同的技能组合。软件工程师需要懂得什么是好代码；审查生成式AI生成的代码功能、复杂度、质量和可读性；扫描漏洞，同时确保自己不会给代码造成质量或安全问题。此外，软件开发人员在编写代码时还需要学会换位思考，更好理解用户意图，由此创建提示和定义情境数据，帮助生成式AI工具给出更好答案。

除了培训技术人才，企业CIO和CTO还可以为非技术人才的生成式AI技能培养发挥重要作用。

在学习如何使用生成式AI工具写邮件或管理基本任务之外，企业员工还要习惯于运用类种技术提高绩效和产出。CIO和CTO可以帮助调整企业学院模式，提供这类培训和相应认证。

由于新手工程师的价值不断降低，企业应加快从传统人才金字塔结构（初级水平的员工人数最多）向钻石型结构（大部分技术人员经验丰富）转变。实际上，这意味着尽快强化初级员工技能，同时减少内容简单的人工岗位（如单元测试编写）。



9、评估最新风险格局，制定持续性风险缓释措施

生成式AI带来一系列新的道德问题和风险，包括“幻觉”（即生成式AI模型会根据最高概率的响应给出错误回答）；意外泄露个人信息；模型使用的大型数据集存在固有偏向性；以及涉及知识产权（IP）的高度不确定性。企业CIO和CTO需要熟悉道德、人道和合规问题，不仅遵守法律条文（因国家而异），还要本着负责任的精神管理企业声誉。

要应对这一新局面，需要详细审查企业的网络实践，并更新软件开发流程，以便在模型开发启动之前评估风险并确定缓解措施。这样既能减少问题，又能确保流程不会放缓。此外，针对AI幻觉行之有效的风险缓解措施包括：在模型生成回答时，调整模型的创造性水平（业界称为“温度参数”）；使用相关内部数据增强模型，提供更丰富的情境；使用信息库，对可生成的内容设置安全护栏；使用“审核”模型检查输出；做出明确的免责声明。早期生成式AI用例应侧重于出错成本较低的领域，以帮助组织克服必然经历的挫折并积累经验教训。

保护数据隐私的关键是建立和执行敏感数据标记规范，在不同域中设置数据访问限制（如人力资源的薪资数据），当数据被外部使用时增加额外防护手段，并采取隐私保护措施。例如，为降低访问权限控制风险，一些组织建立了策略管理层，一旦模型接到提示，就会根据角色对访问做出相应限制。为降低知识产权风险，企业CIO和CTO应坚持要求基础模型供应商对所用数据集的IP情况（数据源、许可和所有权）保持透明。

-
- ¹ 《生成式AI的经济潜力：下一个生产力前沿》（The economic potential of generative AI: The next productivity frontier），麦肯锡，2023年6月14日
 - ² 《生成式AI的经济潜力：下一个生产力前沿》The economic potential of generative AI: The next productivity frontier），麦肯锡，2023年6月14日
 - ³ Begum Karaci Deniz, Martin Harrysson, Alharith Hussin, Shivam Srivastava 所著《生成式AI解锁开发者生产力》（Unleashing developer productivity with generative AI），麦肯锡，2023年6月27日
 - ⁴ Vishal Dalal, Krish Krishnakanthan, Björn Münstermann, Rob Patenge 所著《科技债：收回技术权益》（Tech debt: Reclaiming tech equity），麦肯锡，2020年10月6日
 - ⁵ 《生成式AI解锁开发者生产力》（Unleashing developer productivity with generative AI），2023年6月27日
 - ⁶ Erik Brynjolfsson, Danielle Li, Lindsey R. Raymond 所著《生成式AI进行时》（Generative AI at work），美国国家经济研究局（NBER）工作论文，编号31161，2023年4月

Aamer Baig是麦肯锡全球资深董事合伙人，常驻芝加哥分公司；

Sven Blumberg是麦肯锡全球资深董事合伙人，常驻杜塞尔多夫分公司；

Eva Li是麦肯锡顾问，常驻旧金山湾区分公司；

Megha Sinha是麦肯锡全球董事合伙人，常驻旧金山湾区分公司；

Douglas Merrill是麦肯锡全球董事合伙人，常驻南加州分公司；

Adi Pradhan是麦肯锡全球副董事合伙人，常驻多伦多分公司；

Stephen Xu是麦肯锡全球副董事合伙人，常驻多伦多分公司；

Alexander Sukharevsky是麦肯锡全球资深董事合伙人，常驻伦敦分公司。

作者感谢Stephanie Brauckmann、Anusha Dhasarathy、Martin Harrysson、Klemens Hjartar、Alharith Hussin、Naufal Khan、Sam Nie、Chandrasekhar Panda、Henning Soller、Nikhil Srinidhi、Asin Tavakoli、Niels Van der Wildt和Anna Wiesinger为本文所做的贡献。

麦肯锡公司2023年版权所有。

The background is a solid dark blue color. Scattered throughout the upper and middle portions of the page are numerous 3D cubes of various sizes. Some cubes are dark blue, while others have a lighter blue or cyan glow, particularly on their top and front faces. The cubes appear to be floating or falling, creating a sense of depth and movement.

第二部分

洞悉先机，生成式AI浪潮下的全球产业机遇与投资价值





投资机构篇：挖掘生成式AI产业新机遇

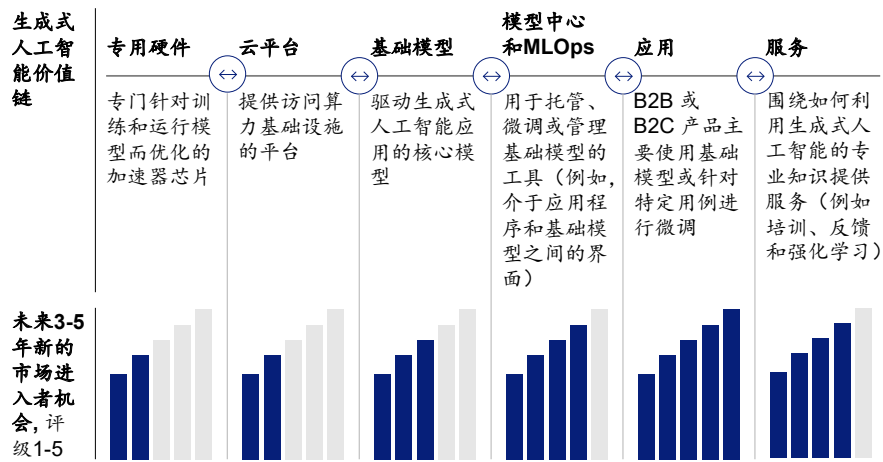
综观GenAI价值链的六大环节，各个环节都蕴藏着大量机遇，有待投资公司去发掘。

曲向军，韩峰，方溪源，胡艺蓉，李静瑶和邱外山

引言

生成式人工智能 (Generative AI, 以下简称GenAI) 价值链由六大环节组成, 分别是专用硬件、云平台、基础模型、模型中心和MLOps、应用和服务。相较于传统AI价值链, GenAI新增了基础模型这个关键环节, 且整体系统较传统AI系统复杂许多, 应以全新视角来观察。伴随着GenAI的蓬勃发展, 价值链上的各个环节都蕴藏着大量机遇, 但我们在深入研究后发现各环节的市场机会存在明显差异 (见图1)。由于部分环节对资源投入、专业知识的门槛要求较高, 且先发者凭借自身优势已形成一定行业壁垒, 这些都对新进入者和小型企业展业构成了较大阻力。

图1: GenAI价值链六大环节的差异化市场机会



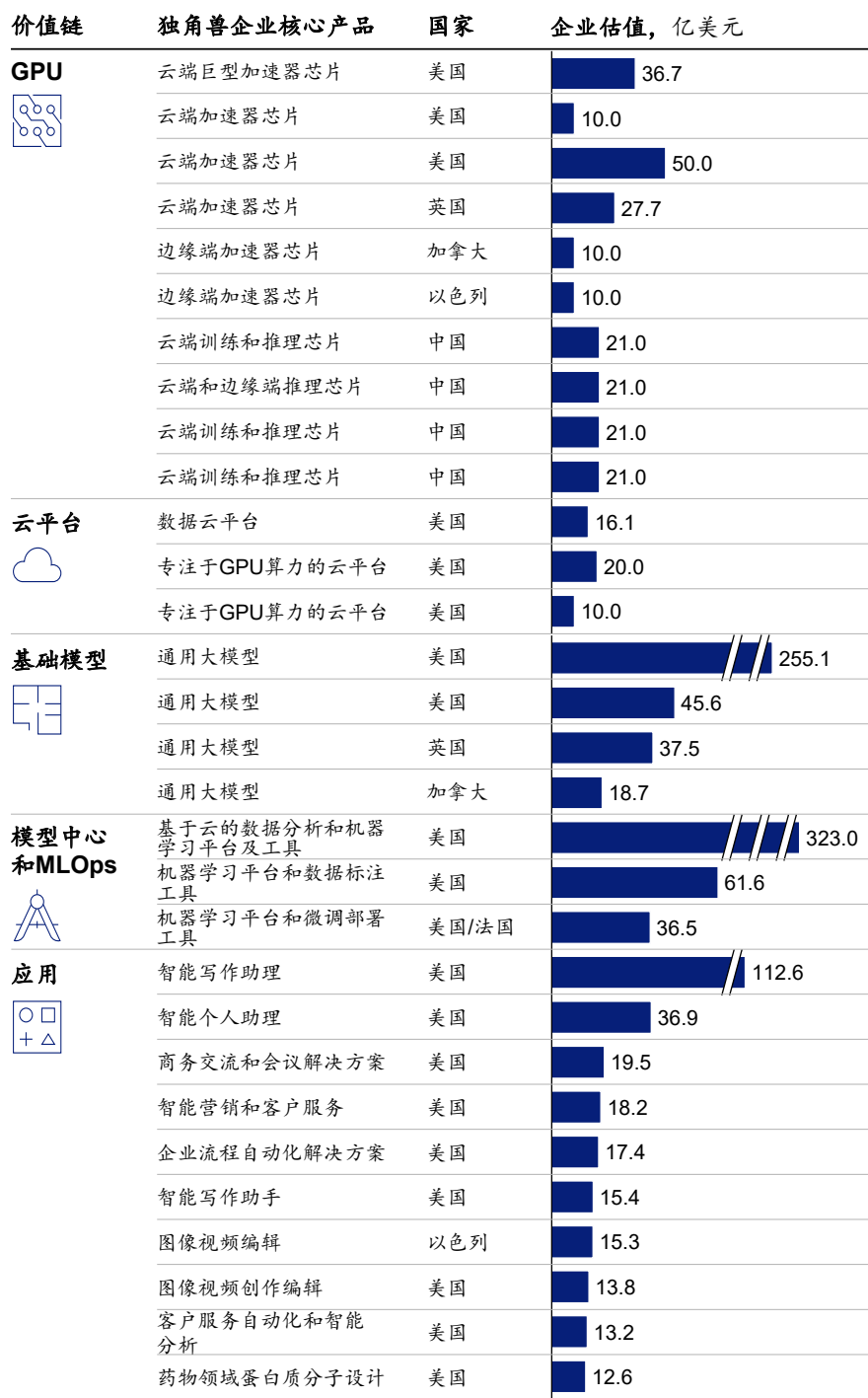
资料来源:麦肯锡分析

我们预计, 价值链上游环节将继续由科技巨头主导, 而应用环节将成为GenAI价值链上未来增长最快的市场, 为一众科创企业和新进入者提供大量价值创造机会。

为了进一步分析生成式人工智能的商业价值, 本文将聚焦GenAI价值链六大环节中的独角兽企业, 深入浅出地为读者剖析这些成功企业如何在各自的领域和市场中探索与实践最新的人工智能技术。

在深入研究各赛道排名前十的独角兽企业后，我们发现应用赛道的高价值成长企业数量最多，而基础模型赛道最有可能诞生超级独角兽（见图2）。

图2：GenAI价值链六大环节的差异化市场机会



资料来源：Pitchbook（PE，VC及并购项目数据平台），麦肯锡研究

GenAI价值链市场机会剖析： 硬件和云平台呈现巨头垄断格局， 科创企业和初创公司可瞄准垂直 行业或细分领域的应用环节，抢 占市场份额

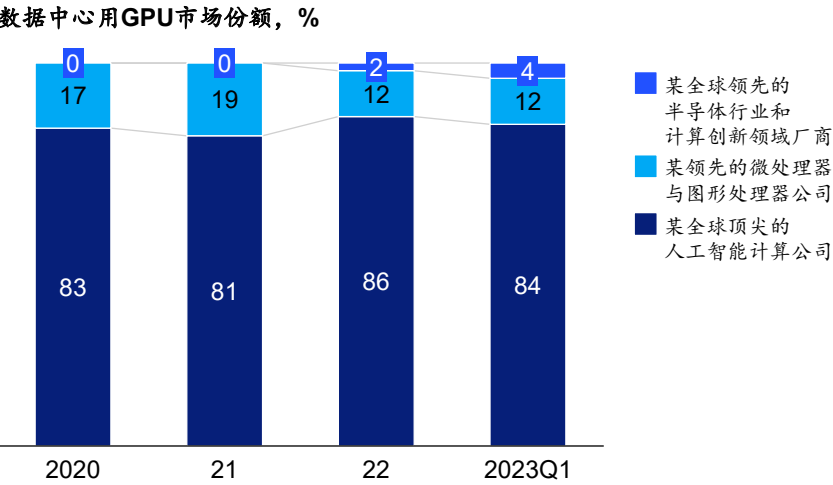
一、专用硬件：市场壁垒较高，已形成巨头垄断

GenAI模型的训练、微调和推理过程都依赖算力基础设施的支持，尤其训练耗费最大。算力硬件的核心是以GPU和TPU为代表的计算芯片，即能够并行处理数十亿个参数数据的“加速器”芯片。

尽管全球市场上存在部分小型生产厂商，但就GPU和TPU的设计和生 产而言，其本质上是高度集中的，全球三家头部专用GPU厂商占据了约99%的市场份额。某市占率第一的全球领先厂商自1999年推出全球首款GPU后，通过持续迭代创新，至今已成为全球计算专用GPU设计市场的主导者（见图3），其CUDA软件平台已经成为训练芯片设计行业的标准；此外某全球顶尖的互联网科技公司独占TPU设计市场，而某领先的半导体制造公司则生产几乎所有的AI加速器芯片。这些都使新进入者面临高昂的启动成本，包括科技研发、标准适配和产业链合作，拥有相关专业知识和技能的从业人员也成为稀缺资源。

我们认为，全球头部的专用GPU厂商未来将继续通过自身长期投入而构建起的强大行业壁垒，在全球训练用芯片市场居主导；在推理用芯片市场，由于客户有可能因为性价比而更换其它厂商，因此独角兽芯片企业仍有参与机会，代表企业包括美国的某主打4纳米AI加速器芯片的初创公司，以及中国某AI神经网络解决方案厂商等。

图3：全球数据中心GPU市场份额（%）



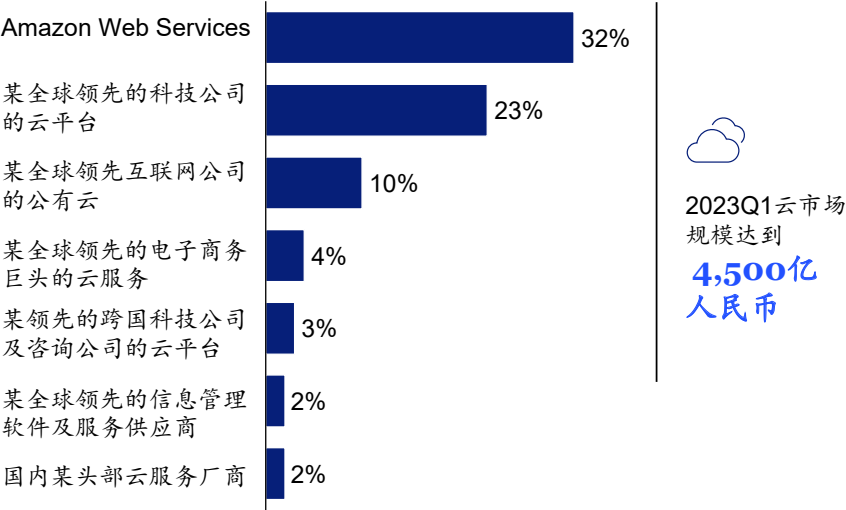
资料来源：Jon Peddie Research

二、云平台：大型云厂商一枝独秀，包揽80%市场份额

当前以GPU和TPU为代表的算力硬件价格昂贵且稀缺，使得大多数企业在本地图部署和维护GenAI模型的投入产出比很低，且执行上较为困难。因此目前大部分训练、微调和运行大型模型的工作皆在云端，企业通过云厂商提供的服务即可轻松获得算力，并根据实际需求管理相关支出。

全球云市场呈现明显的头部集中态势，排名前八的大厂占据约80%的市场份额（见图4），头部厂商凭借规模优势、产品体系和运营经验构建起难以逾越的市场壁垒。我们预计，新入局的玩家将难以在短期内抢占市场份额，初创企业如果未能取得大型企业客户的支持（例如部分希望减少对大型云平台依赖的客户），将难以实现自身竞争目标。

图4：2023Q1全球云基础设施服务市场竞争格局



资料来源: Statista

三、基础模型：GenAI价值链的核心环节，通用大模型赛道头部集中趋势明显，行业大模型赛道仍有市场空白

（一）基础模型即大型深度学习模型，是生成式人工智能的核心环节

通过大量数据训练，基础模型能够回应提示、生成与训练数据相符但具有新颖性的各式模态的内容。根据输出内容的模态，基础模型主要分为文本模型、图像模型、音频模型、视频模型等四大类，此外还有3D模型、多模态模型等。基础模型能够支持广泛的任务，在开发完成的基础模型之上，可利用其内容生成能力，进一步构建各式应用。例如，以全球领先AI研究实验室开发的文字生成的自回归语言模型，使用深度学习生成人类可以理解的自然语言，可以用于解决下游各种类型的生成任务。以此类深度学习应用为代表的基础模型，可以产出媲美人类写作水平的高质量文本；从全球备受瞩目的ChatGPT，到SaaS内容生成工具，例如某AI文案生成工具，以及某AI驱动的内容生成器。此类AI工具利用人工智能的力量来创建高质量的文章、电子邮件、博客文章和社交媒体帖子，与聘请文案相比，参与率更高，成本更低。此类应用，都是基于上述OpenAI基础模型的具体应用。

(二) 开发基础模型的壁垒很高, 依赖海量数据、专业知识以及高昂投入

就海量数据而言, 全球领先的AI厂商旗下的自回归语言模型, 其训练数据集包含约13万亿token, 相当于1.8PB的海量文本数据(根据OpenAI官网, 1000个token大约对应750个英文单词)。其数据集来源包括从维基百科、政府网站、社交媒体和图书中爬取的公共数据, 以及来自大型数据库的私有数据(例如OpenAI与付费图库Shutterstock达成合作, 用后者提供的专有图像训练其图像模型); 就跨领域专业知识而言, 所涉及的专业领域包括数据准备、选择匹配目标的模型架构、训练模型, 以及微调模型改进输出(涉及到标记模型输出的质量, 并将结果反馈给模型以供其进一步学习); 就投入而言, 由于训练过程需要不断重复并全程辅以海量计算资源支持, 类似的基础模型训练至少需要花费数百万美元、耗时数月甚至更久。据估算, 某顶尖AI研究实验室仅训练上一代自回归语言模型, 就已投入400万~1,200万美元。该实验室训练近期发布的新一代模型的投入可能高达上亿美元。

(三) 较高的壁垒导致通用大模型市场由科技巨头和独角兽企业主导

GenAI模型的竞争是技术、数据和资金的综合较量。由于开源大模型难以计算具体收入, 闭源厂商也较少单独披露大模型相关的财务数据, 导致无法统计全球各大厂商的市场份额。但我们可以看到, **目前市场上文本模型的数量最多, 在全球8个主要的大语言模型中, 有4个来自科技巨头, 4个来自AI独角兽企业(见图5)。**与之类似, 国内知名的GenAI大模型也基本出自互联网巨头、AI独角兽或顶尖科研院校。无论是海外还是国内市场, 各大玩家均拥有顶尖的AI技术人才, 科技巨头的优势在于其雄厚的资本实力和研发能力, 独角兽企业则拥有更敏捷的组织 and 更高效创新机制。随着文本模型日趋成熟, 基于其它模态模型的赛道也将逐渐兴起; 鉴于其它模态模型的开发成本显著高于文本模型, 预计通用大模型赛道未来仍是以科技巨头和独角兽企业为主的战场。

图5：海外GenAI大模型（按模态区分）的主要玩

大模型 ■ 开源 ■ 非开源, 可调用APIs² ■ 非公开 ■ PaaS³ ■ MLOps³

	文字	代码	图像	音频/音乐	视频	3D	蛋白质结构/DNA序列	PaaS/ MLOps ³
某全球领先的科技公司	Orca-13B	Phi-1		VALL-E	GODIVA	RODIN Diffusion	MoLeR	AzureAI/ Azure OpenAI
某全球领先的AI研究实验室	GPT-4	GitHub Copilot ¹	DALL-E2	Jukebox		Point-E		
Meta	LLaMA2	CodeCopilot ² / InCoder	Make-a-scene	AudioGen	Make-a-video	Builder Bot	ESMFold	
某全球顶尖的互联网科技公司	PaLM2		Imagen	MusicLM	Imagen Video	Dream Fusion	AlphaFold2	PubSub / VertexAI/ CloudRun
英国某全球化AI研究型企业	StableLM		Stable Diffusion 2	Dance Diffusion			LibreFold	
Amazon/ AWS	Titan	Code Whisperer		Deep composer				Bedrock ⁴ / SageMaker ⁵
Apple						GAUDI		
某全球顶尖的GPU厂商	MT-NLG		Edify		Edify	Edify	MegaMolB ART	
某企业级AI平台独角兽	Family of LLMs							
美国某AI初创公司	Claude2							
美国某AI非营利性研究机构	Jurassic-2							

注意：产品列表仅供参考，并不代表麦肯锡公司的背书

1. 由 GitHub 和 OpenAI 开发
2. 基于 LLaMA2 构建，供内部使用LLaMA2 获取 Python 和其他语言的代码建议
- 3.从技术上讲，此类别不是LLM，而是一种解决方案。平台即服务（PaaS）是指第三方提供商向用户提供硬件和软件工具的云计算模型。机器学习操作（MLOps）在生产环境中可靠、高效地部署和维护机器学习模型
4. Bedrock 通过 API 提供对来自亚马逊/AWS 和知名 AI 初创公司的预训练基础模型的访问
5. SageMaker 是一项完全托管的服务，可帮助数据科学家和开发人员大规模构建、训练和部署机器学习模型

资料来源：麦肯锡分析

(四) 在行业大模型及专项任务模型领域，仍存在市场空白

这两类模型大多是在通用大模型的基础上，通过输入专业数据和进一步训练产生的。目前，市场上已发布的行业大模型和专项任务模型并不多，较具代表性的包括某全球商业、金融信息和财经资讯的领先提供商开发的金融大模型等，意味着可能会有更多的垂类企

业和初创企业入局。我们认为，那些资本实力雄厚、拥有丰富行业或任务数据积淀的企业，将有较大机会从竞争中胜出。此外，部分大型企业可能出于提高数据安全和隐私级别的考虑，希望将大型语言模型 (LLMs) 纳入自身IT运营系统中，市场上已有部分玩家 (如某领先的生成式AI独角兽公司) 针对此类需求，提供围绕LLMs的专项服务。

(五) 基础模型赛道有望诞生可媲美巨头厂商的超级独角兽

基础模型需要长期和大量资金的投入，因其开发成本高昂，故难以呈现百花齐放的市场竞争格局，反而较容易出现“赢家通吃”的规模效应。正如许多互联网超级巨头均诞生于平台领域，我们认为，当下的基础模型市场极有可能正在孕育GenAI领域未来的超级独角兽。



四、模型中心和MLOps：科技巨头与独立厂商形成差异化竞争

（一）模型中心和MLOps承担在基础模型之上构建应用的重任

具体而言，主要包括两项工作：一是建立模型仓库，提供存储和访问基础模型的空间；二是开发专门的MLOps工具，用于微调和部署基础模型的应用。例如，引入额外训练数据以及构建允许应用与模型交互的API，这也是模型中心和MLOps的价值所在。

（二）该赛道的业务模式与模型开放程度有直接关系

对于源代码不开放的闭源模型，模型厂商通常会提供端到端的模型中心和MLOps服务，即通过许可协议以API的形式提供对模型的访问，以及额外的MLOps功能，以使用户在不同的应用中微调和部署大模型。

对于任何人均可自由使用和调整代码的开源模型，一众独立厂商正如雨后春笋般兴起。鉴于开源社区已经围绕大模型打造了一个完整的生态圈，独立厂商可以仅提供模型聚合器能力，为开发团队提供访问不同基础模型的权限；也可提供对模型端到端的MLOps功能（如使用专有数据微调基础模型，并将其部署在应用中），代表厂商包括某全球领先的聊天机器人厂商和某全球领先的数据开发&分析平台等。对于希望利用生成式人工智能技术但缺乏相关人才和开发工具的企业而言，端到端的产品模式能更好地满足其需求。此类独立厂商能否赢得市场，将取决于其产品功能、使用体验（如开发者社区口碑），以及是否具备合适的商业模式（见图6）。

图6：模型中心和MLOps领域知名独立厂商优势

技术维度	技术和开发者		GitHub星数, 千	技术介绍	优势
模型训练和微调	分布式训练加速框架	某大规模并行AI训练系统	30	在DeepSpeed的数据并行、流水并行、一维模型并行的基础上，自研了序列并行等技术，将原有的训练并行维度从3维提高到5-6维	使用方便且训练并行效率高：相比Megatron-LM仅需一半的计算资源即可复现GPT3
		某聊天机器人厂商开发的分布式训练工具	5	在多GPU/TPU训练场景中自动进行硬件资源分配并优化混合精度策略（如FP16和32），同时集成了多种分布式训练技术的python库	支持一站式的分布式训练
	训练微调技术	某聊天机器人厂商旗下的参数高效微调工具	7	集成了各种微调技术（e.g., LoRA, 提示微调等）的python库，能节省微调内存并提高微调后模型性能	最少仅需微调0.01%的参数就能实现模型性能的提高
	训练微调平台	某自动机器学习的人工智能(AI)平台	2	提供了可视化界面让使用者以零代码的方式进行大模型微调，集成了多种最新的微调技术（如LoRA和低精度微调）	支持实时反馈模型微调的表现，从而支持快速迭代微调实验
模型部署	部署服务	某机器学习分布式框架	26	通过Batch inference技术并行化对输入数据批量推断从而加速模型推理	非常适合需要多模型组合的场景,可以在Python代码中构建由多个ML模型和业务逻辑组成的复杂推断服务
模型应用开发	编排	某基于语言模型开发应用程序的框架	49	将大模型连接到外部数据源，并通过Agent的方式让大模型与真实世界进行交互	集成了136种的外部数据源 (Arxiv, Wiki) 和工具库 (Qdrant)
模型管理平台	框架	某全球领先的数据开发&分析平台	15	支持大模型训练的表现跟踪、版本控制、部署管理等	对模型生命周期管理进行一站式管理

资料来源:麦肯锡分析

五、应用：终端应用领域孵化出最多GenAI独角兽企业

终端应用构建于基础模型之上，是大模型能够完成指定任务的关键环节，随着GenAI不断展现出巨大的发展潜力和广阔的应用前景，应用市场赛道可谓百花齐放，既包括开发全新应用的新玩家，也包括向当前应用添加创新功能的老玩家，以及诸多希望在应用领域弯道超车、打造自身竞争优势的科创企业。

（一）我们预期在短期内，为垂直行业和特定功能开发、基于微调模型所建立的应用能够脱颖而出，在GenAI早期阶段创造更多价值

首先，微调模型门槛不高。企业只需引入额外相关数据或按需调整模型参数，便能为特定用例产出优质内容。虽然训练基础模型需要大量数据，往往费用昂贵又耗时，但对基础模型进行微调所需的数据量较少，成本较低且可在较短时间内完成，大幅降低了参与门槛。

其次，基于专有数据的微调能够实现更高质量的输出。具体而言，企业可基于以下三个领域的专有数据，对模型进行微调：

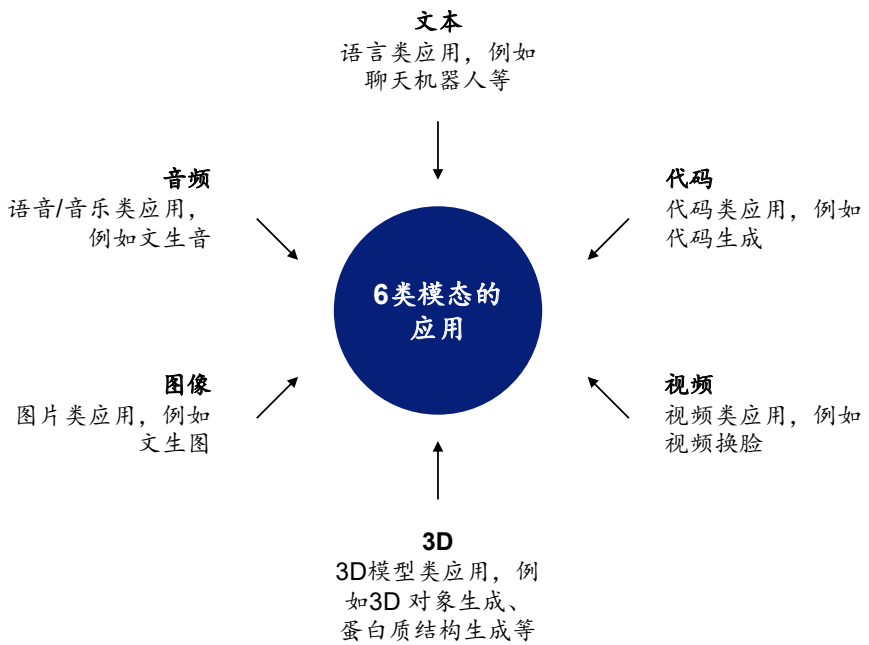
- **洞察垂直行业或客户的具体需求并总结数据。**例如，专注于法律领域的GenAI应用的开发者将法律数据集输入到OpenAI的GPT-3，并测试不同的提示词，以确保微调后的模型能够生成远胜于ChatGPT的法律文件。
- **利用日常商业运营产生的专有数据。**例如，某软件开发商为某银行开发一款基于GenAI的客服机器人，可将该银行呼叫中心的对话数据纳入其中，使模型随着专有数据的增加而不断迭代优化，持续提升客户体验。

- 从终端用户的评价和反馈系统中获得专有数据。例如，终端用户的打分评价系统（如好评/差评等），OpenAI就专门邀请用户对其输出质量进行评价，以反哺模型并持续训练ChatGPT。根据OpenAI的相关报告，评价反馈数据非常有助于改进底层模型。

(二) GenAI应用可分为六大主要模态，市场格局差异化明显

按照输出内容，GenAI应用可分为六大主要模态，即文本、代码、音频、图像、视频以及3D（见图7），基本涵盖所有人类在真实世界中感知与表达信息的主要方式，也体现出GenAI适合于理解性和创造性工作的巨大潜能。每种模态下所衍生的具体应用场景，都蕴藏着一个充满无限憧憬、前景广阔的市场。例如，文本类应用中的总结提炼法律法规文件、代码类应用中的合成训练数据集，以及视频类应用中的长视频自动转换短视频等（见图8）。

图7： 初创企业拥有最大机会的应用层存在6类模态的应用



资料来源:麦肯锡分析

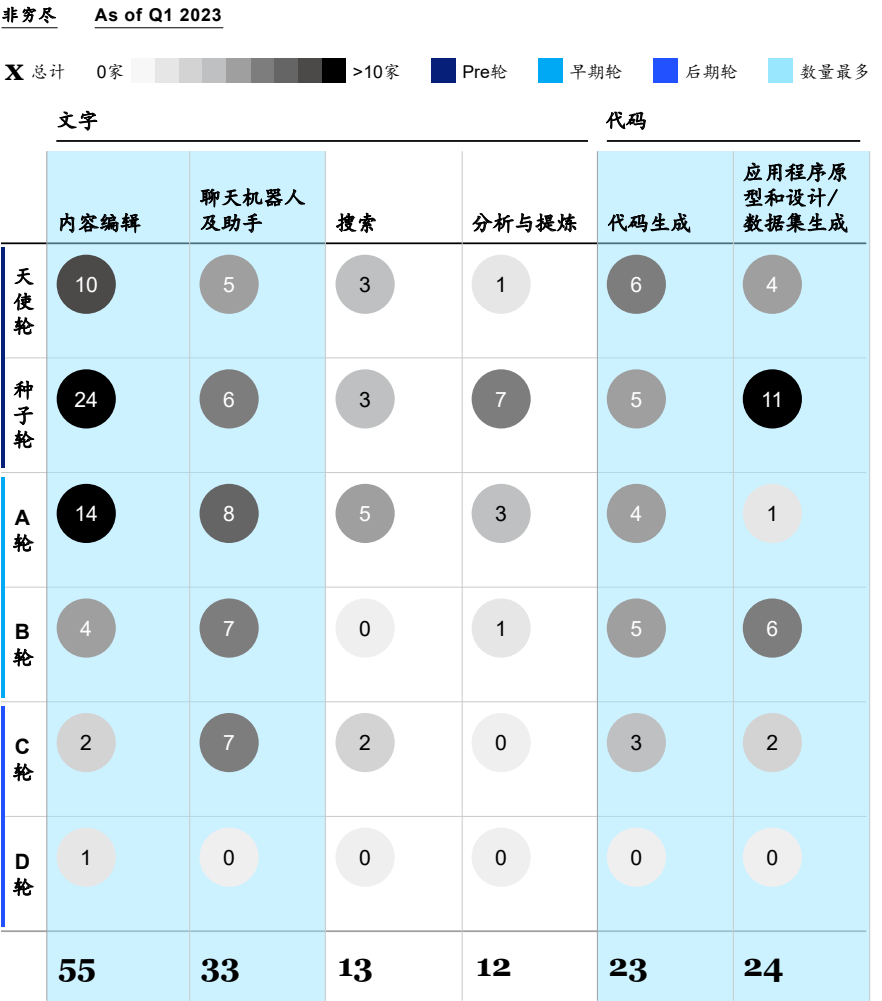
图8：GenAI应用六大主要模态的典型用例

模态	应用	用例示例	
文本 	内容创作	- 营销：创建个性化的电子邮件和帖子 - 人才招聘：起草面试问题、职位描述	
	聊天机器人或助手	客户服务：使用聊天机器人提高网络娱乐的转化率	
	搜索	- 进行更自然的网络搜索 - 利用企业知识增强内部搜索工具	
	分析和综合	- 销售：分析客户互动以提取见解 - 风险和法律：总结合规文件	
代码 	代码生成	IT：通过自动化代码建议加速程序开发和提升质量	
	应用原型和设计	IT：快速生成交互界面设计	
	数据集生成	生成数据集以提高AI模型训练质量	
图像 	库存图片生成器	营销和销售：生成独特的媒体图片	
	图像编辑器	营销和销售：快速个性化内容	
音频 	文本生成语音	培训：创建教育用配音	
	声音创作	娱乐：制作自定义声音而不侵犯版权	
	音频编辑	娱乐：播客后期制作免去重复录制	
3D 	三维对象生成	- 电子游戏：绘制场景、角色 3D建模：为建筑设计创建室内设计模型和虚拟舞台	
	产品设计与发现	- 制造：优化材料设计 - 药物发现：加速R&D过程	
视频 	视频创作	- 娱乐：生成适配各线上平台的短视频素材 - 培训或学习：使用AI头像创建视频课程或企业宣讲视频	
	视频编辑	- 电子商务：为通用视频添加个性化元素 - 娱乐：后期制作去除背景图像和背景噪音	
	翻译和语音调整	- 视频配音：使用AI语音或原声语音转换新语言 - 现场翻译：用于公司会议、视频会议 - 语音克隆：复刻演员声音或改变效果	
	换脸和脸部调整	- 虚拟效果：实现快速老化、年轻化、化妆、假发和修复 - 换脸和深度仿真视觉效果	
		- 后期制作中的口型同步或“视觉”配音：快速编辑实现多语言发布 - 视频会议：实时凝视校正	

资料来源:麦肯锡分析

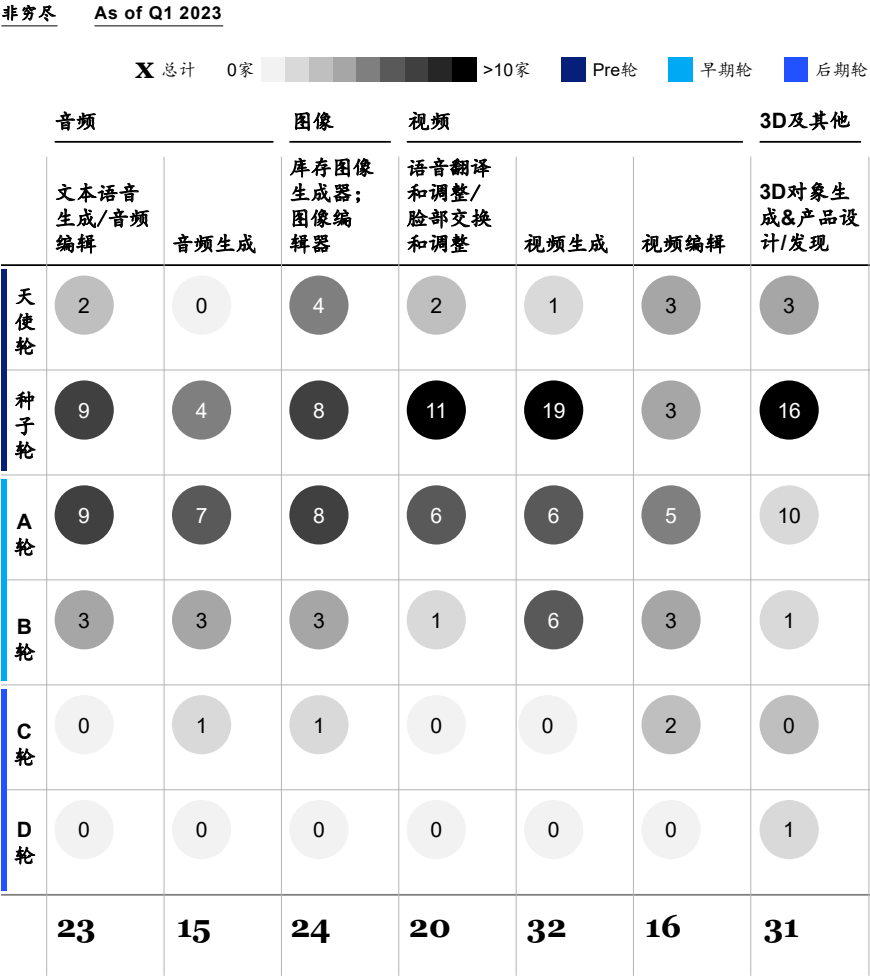
为展现GenAI应用六大模态下细分赛道的发展现状，我们盘点了321家处于不同融资阶段的海外知名科创企业。其中，**文本、视频和代码类初创企业数量最多**，占比分别为35%、14%及22%。就企业成熟度而言，**文本和代码模态下的应用相对更成熟**，不仅获得融资的企业数量最多，且约10%的科创企业已进入C轮+融资阶段；相比之下，视频类应用企业仍处于早期阶段，仅有不到3%的企业进入C轮+融资阶段（见图9）。从估值来看，约70%的独角兽企业集中在文本赛道，其中成立于2009年的拼写检查和智能建议工具目前估值最高，达到112亿美元；而聚焦于其余模态应用的独角兽，在估值和数量上都远低于文本模态赛道。

图9：应用概览（1/2）全球范围内，终端用户应用程序在文本生成模式上最成熟，代码生成在国外应用也较多



资料来源: Pitchbook (PE, VC及并购项目数据平台), Antler (新加坡base的VC平台), Crunchbase (投融资信息平台), Gain.pro (聚焦一级市场的数据平台)

图10：应用概览（2/2）全球范围内，终端用户应用程序在文本生成模式上最成熟，代码生成在国外应用也较多



资料来源: Pitchbook (PE, VC及并购项目数据平台), Antler (新加坡base的VC平台), Crunchbase (投融资信息平台), Gain.pro (聚焦一级市场的数据平台)

(三) GenAI价值链中,应用赛道的高价值成长企业数量最多,约一半独角兽企业来自应用赛道

尽管如此,我们更愿意把现阶段视作GenAI应用的婴儿期,因为大型厂商刚开始尝试将GenAI与现有产品结合,一众初创企业正着手筹备其首款产品。就各个模态的发展程度而言,文本模态应用正处于蓬勃发展阶段,其余几个模态应用中则有为数不多的几个崭露头角者,我们预计未来在这些领域将会涌现更多创新者和挑战者,以不同方式颠覆现有的行业竞争格局,并持续创造价值。

接下来,我们将以4家知名独角兽企业为例,深入剖析GenAI创业公司取得成功的关键要素,包括技术实力、数据积累,以及自身产品能力。

1. 独角兽公司A成立于2022年,仅发布过一款个人智能助理应用Pi。这家只有35名员工的独角兽公司,已获得包括位于硅谷的全球领先科技公司、顶尖GPU厂商、以及一众硅谷大咖的投资,其成功秘诀在于:
 - 明星创始团队的加成,三位联合创始人分别是总部位于英国的领先AI研究型企业的联合创始人以及该机构的前资深科学家,和硅谷某全球领先的职场社交平台联合创始人;
 - 公司拥有强大的技术实力,在顶尖GPU厂商的支持下,正着手构建世界最大的AI集群,以开发目前市场上最复杂的大语言模型,并对标OpenAI;
 - 其推出的智能个人助理PI,产品定位别具特色,主打随时随地的陪伴,以诙谐有趣、富有同理心、高度仿真的对话模式为特点,与当下炙手可热的ChatGPT形成差异化竞争。
2. 独角兽公司B成立于2013年,是一家专注于开发照片和视频编辑软件的公司。该公司于2019年获得来自某国际知名投行的C轮融资后,跻身独角兽企业行列,其成功秘诀在于:

- 公司拥有深厚的技术实力，创始团队来自以色列顶尖高校的计算机视觉和图形学专业，掌握核心算法技术；
 - 旗下产品提供一流的用户体验，核心产品注重提供简洁易用的用户界面和流畅的编辑体验，使用户数量得以快速增长；
 - 由于始终专注于开发移动app，随着移动社交媒体和短视频平台的井喷式增长，其照片和视频编辑软件也受到全球多国用户的欢迎。
3. 独角兽公司C利用GenAI技术提供新型生物制药解决方案，其生成式人工智能平台将生物学研究与大量行业数据相结合，缩短了发现和设计新型蛋白质药物的实验周期。其成功秘诀在于：
- 采用先进的GenAI技术，能够高效模拟和预测药物的生物效应，从而加速药物研发进程；
 - 具备较强的行业数据生成能力，其专利技术能够多重合成遗传细胞并生成数十亿蛋白质，为训练GenAI模型提供了大规模、高质量的数据集；
 - 核心团队拥有足够的行业深度，成员具备生物医药行业资深背景，对药物研发流程和生物医药理论有深入理解。
4. 独角兽公司D是一家成立于2016年的云编程平台创业公司，短短几年时间便拥有超过2000万开发者用户，在2023年初获得B+轮融资后，跻身独角兽企业行业。其成功秘诀在于：
- 优秀的产品设计对用户增长至关重要：公司产品支持50多种开发语言，可帮助初级开发者避开复杂的环境部署，直接上手编程。对于有经验的开发者，则可协助用户构建和测试各种应用程序；

- 依靠海量代码数据训练编码大模型，例如，公司于2022年上线AI辅助编程工具，并于今年发布27亿参数的编码大模型，可支持20种编程语言，版本性能甚至优于OpenAI同类模型；
- 在硅谷顶尖科技互联网巨头等重量级合作伙伴的加持下，独角兽公司D的技术实力得以进一步增强，从而获得资本市场青睐。今年年初，科技巨头将旗下人工智能大语言模型与独角兽D的产品相结合，进一步提升了其代码能力。

六、服务：整体解决方案领域被大厂垄断，但垂类领域仍有参与机会

与传统人工智能类似，未来GenAI相关的专属服务将如雨后春笋般出现，并以专业集成商、垂类服务商、咨询公司等三类机构为主，以帮助企业洞察最新商业机会，以及缩小在复杂技术和能力方面的差距。

- 我们预计现有AI服务提供商将不断发展自身能力，以覆盖新的市场机会（如某全球领先的跨国科技及咨询公司开发的电脑问答系统等）；
- 在依托大模型产品提供增值服务的整体解决方案领域，将由现有科技大厂垄断；
- 垂类市场仍有较多空间，中小玩家可凭借特定领域或行业的专业知识和数据积累进入市场，代表厂商包括总部位于美国的AI策略咨询和解决方案提供商、一家位于印度的AI培训机构，以及一家位于乌克兰的端到端AI产品开发和咨询服务机构等。

值得注意的是，服务环节尚未出现独角兽级别的企业，市场仍处于一个未被充分挖掘的状态，蕴藏的大量商机尚待耕耘和发掘。在可预见的未来，随着越来越多企业客户意识到生成式人工智能技术的创新价值和实践难点，并逐渐开始寻求专业培训和咨询服务，

相关领域的专业机构将迎来新一轮发展契机。然而，这一切都有待GenAI相关服务市场的进一步成熟和发展，以及服务供应商在技术实力、行业数据和解决方案等方面的不断积累。

结语

综观GenAI价值链的六大环节，各个环节都蕴藏着大量机遇，有待投资公司去发掘。鉴于价值链上游环节（如专用硬件、云平台、基础模型等）市场壁垒较高，未来将继续由科技大厂主导，呈现巨头垄断格局。我们预计GenAI价值链下游环节将提供较多市场机会，其中应用环节将成为GenAI价值链上增长最快的市场，科创企业和初创公司可瞄准垂直行业或细分领域的应用环节，抢占市场份额。此外，在行业大模型及专项任务模型领域，以及GenAI相关服务的垂类领域，仍存在一定市场空白，中小玩家可结合自身优势有选择性地进入相关市场。随着GenAI应用逐渐成熟和规模化，未来会涌现出优秀企业，投资行业应该紧跟市场趋势，及时捕捉优质投资机会。

¹ 数据来源：Bloomberg Intelligence

² 数据来源：Gartner

³ 数据来源：《中国AI数字商业产业展望2021~2025》，前瞻产业研究院、中关村大数据产业联盟联合发布

⁴ 数据来源：Canalys《2022年第一季度中国云计算市场报告》

⁵ 数据来源：2023年《中国人工智能大模型地图研究报告》，中关村论坛

⁶ 据来源：“2023年网站所用语言统计”，互联网研究网站W3Techs

曲向军是麦肯锡全球资深董事合伙人，中国区金融机构咨询业务负责人，常驻香港分公司；

韩峰是麦肯锡全球董事合伙人，常驻深圳分公司；

胡艺蓉是麦肯锡全球副董事合伙人，常驻上海分公司；

王喆宸是麦肯锡咨询顾问，常驻上海分公司。

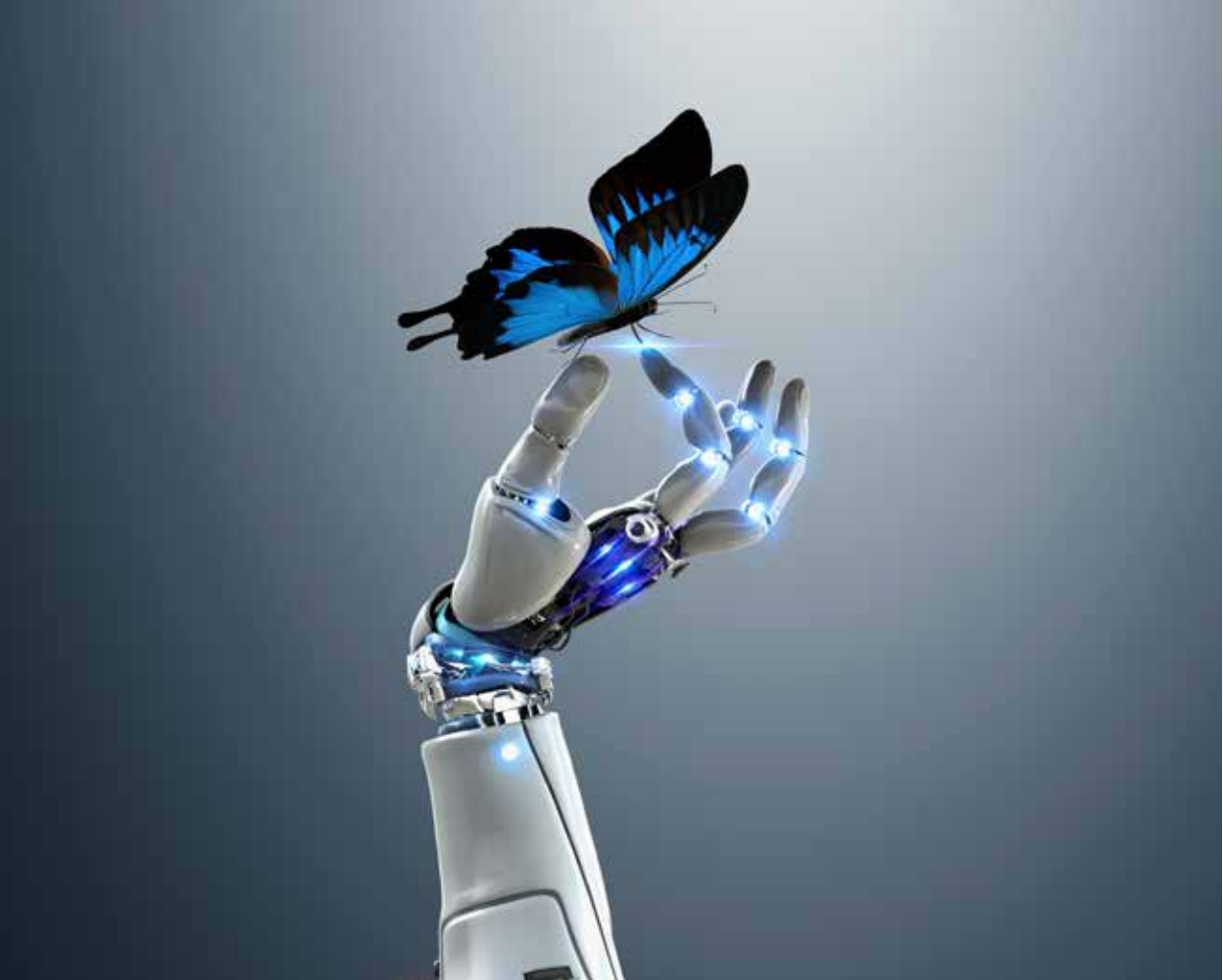


第三部分

赋能百业，中国企业积极 布局生成式AI“新赛道”

126 中国生成式 AI 行业发
展现状及未来趋势





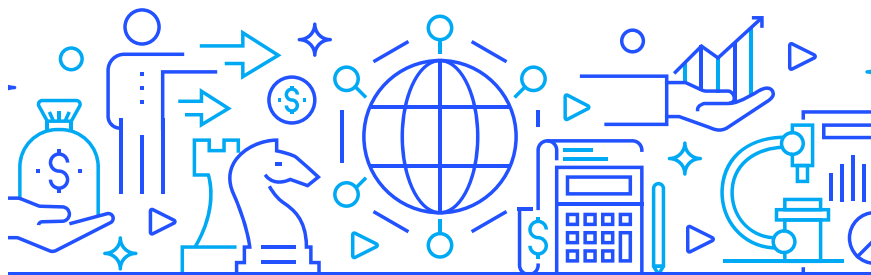
中国生成式AI行业发展 现状及未来趋势

本文深入浅出地分析中国GenAI行业的发展现状和趋势,详细剖析中国多个细分领域存在的机遇与挑战,帮助企业探索适合自身发展的商业模式,抓住GenAI技术带来的红利。

曲向军 韩峰 胡艺蓉和王喆宸

在本章中, 我们将从市场规模、技术栈、商业化应用和风险管理四大维度, 深入探讨中国生成式AI (简称GenAI) 行业现状和发展趋势, 核心观察包括:

- **市场规模:** 全球人工智能市场正处于高速增长阶段, 预计2032年将达到1.3万亿美元, 2022~2035年全球市场规模增量主要来源于训练侧硬件、广告应用和软件。与此同时, 预计国内GenAI市场规模将在2025年突破人民币2,000亿元。
- **技术栈:** 硬件方面, 国内大模型在算力和中文训练资源上存在瓶颈, 但国产化进程不断加速。在模型方面, 国内通用大模型数量可观, 行业大模型有望成为中国GenAI的制胜赛道。在应用方面, 目前国内GenAI在垂直应用领域的布局主要聚焦文本、图像和音视频生成, 海外同行更偏向于布局代码生成等领域。
- **商业化应用:** 国内GenAI企业利用垂直场景优势, 将GenAI优先应用于商业化基础设施较完善的新兴行业 (如电商、传媒、娱乐和游戏等), 随着未来SaaS市场不断成熟与企业付费意愿提升, 预计国内GenAI应用将进一步拓展到传统行业, 并探索出更完善的商业模式。
- **风险管理:** 考虑到中文语料的复杂性和用户规模, 包括模型幻觉、恶意使用、信息泄露在内的三大风险值得重点关注。随着国家互联网信息办公室出台《生成式人工智能服务管理暂行办法》, 并针对数据训练、内容标识、产权保护等提出具体监管要求, 目前国内GenAI的法律监管框架已初步建立。

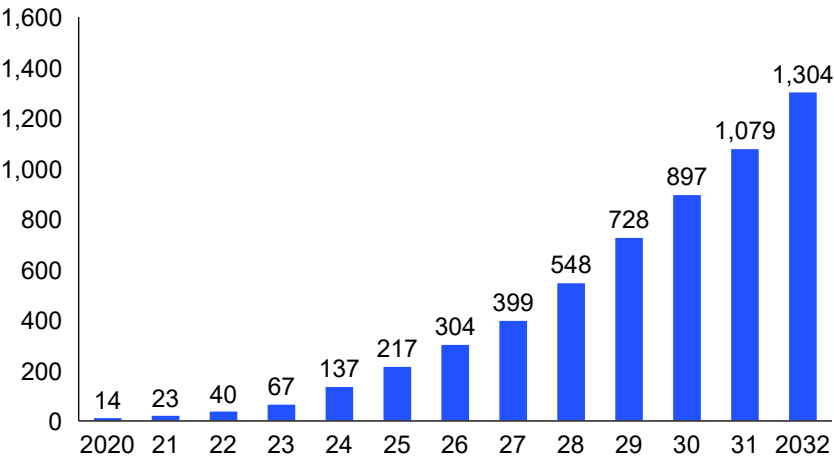


一、市场规模

2022年全球GenAI市场整体收入为400亿美元，预计2027年及2032年将分别达到3990亿美元和1.3万亿美元（见图1）¹，2022~2032年复合增长率高达42%¹。到2025年，GenAI产生的数据将占有所有生成数据的10%²（2021年仅为1%²）。

图1：2020~2032年全球生成式人工智能应用规模

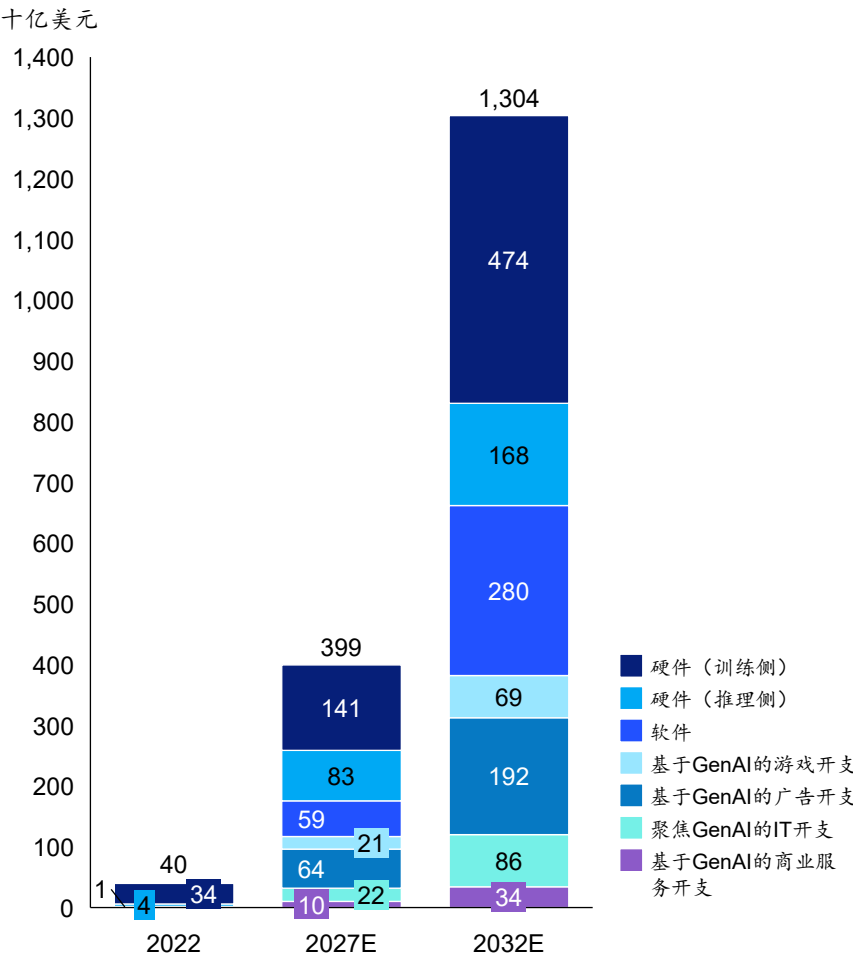
2020-2032年全球生成式人工智能应用规模，十亿美元



资料来源：Bloomberg Intelligence

预计2022~2035年全球市场规模增量主要来源于训练侧硬件、广告应用和软件。其中，在训练侧硬件方面的增量约4,440亿美元，其中仅GenAI基础设施服务（GenAI as a service）一项，增量就有2,448亿美元，年复合增长率将达到60%；在广告应用方面，相关年复合增长率将达到125%，增量达1924亿美元；而软件方面，在约2,800亿美元的增量中，GenAI助理软件增量可观，达890亿美元，年复合增长率有望达到70%（见图2）。

图2：2022`2032年全球生成式人工智能细分行业市场规模



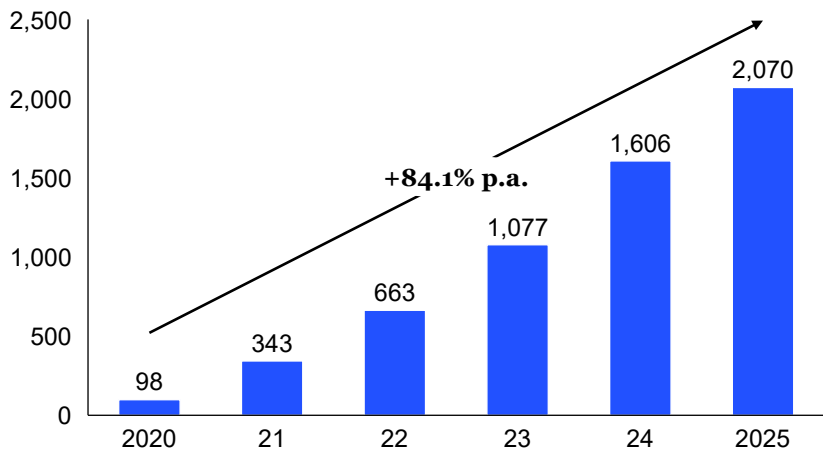
资料来源：Bloomberg Intelligence

2022年中国GenAI市场规模约人民币660亿元³，预计2020~2025年复合增速将达84%³，到2025年将占全球市场规模（2,170亿美元）的13%（见图3）。

与此同时，近两年国内GenAI融资市场热度持续上升，前十大融资事件总规模达8.7亿美元。其中，某领先的端到端AI制药公司、光某大模型初创公司等头部企业获得逾2亿美元的单笔融资，多轮融资累计规模分别超3亿美元、2.5亿美元（见图4）。

图3：2020~2025年中国生成式人工智能应用规模

2020-2025年中国生成式人工智能应用规模， 亿元



资料来源：《中国AI数字商业产业展望2021-2025》，前瞻产业研究院、中关村大数据产业联盟联合发布

图4：海外生成式AI赛道典型融资案例

企业	轮次	融资时间	产品类型	融资规模， 百万美金
某AI制药公司	C轮	2022.10	AI制药	210
某AI大模型公司	A轮	2022.10	AI大模型	209
某GenAI与语音交互为核心的AI公司	D轮	2023.1	AI生成语音	131
某AI制药公司	D轮	2022.11	AI制药	92
某AI非结构化数据处理和分析公司	C轮	2022.12	AI非结构化数据处理和分析软件	53
某大模型创业公司	Pre-A轮	2022.1	AI大模型	46
某AI非结构化数据处理和分析公司	B轮	2023.2	AI非结构化数据处理和分析软件	37
某GenAI与语音交互为核心的AI公司	C轮	2023.3	AI生成语音	35
某AI制药公司	B轮	2022.11	AI制药	33
某AI和多模态通用搜索技术的公司	A轮	2021.1	AI神经搜索技术	26

资料来源: Pitchbook数据库，数据截至2023年7月



二、技术栈

在本部分中，我们将结合GenAI价值链的六大环节（包括专用硬件、云平台、基础模型、模型库和工具、应用、服务）探讨中国GenAI技术栈的发展现状。

（一）专用硬件：基础设施建设资源受限，但国产化进程不断加速

由于无法获得高端GPU（图形处理器）芯片A100与H100，国内大模型研发机构在算力上遭遇瓶颈。为应对上述困境，中国本土芯片研发厂商加强科研，在算力技术上实现重要突破。

目前，中国市场可以获得的A800 GPU芯片，传输速度仅为顶尖产品A100的70%。由于AI技术发展高度依赖先进GPU，需借助芯片进行大量模型训练和扩展，部分国内科研机构 and 科技企业因此受到较大影响，尤其在目前大模型参数量快速膨胀、算力需求显著增加的情况下。

为了应对算力瓶颈，国内领先算力芯片公司力求在产品端实现突破。例如，国内某量产商业人工智能芯片厂商，在头部大型云服务提供商旗下一款知识增强大语言模型产品中，实现了对部分业务场景的部署。某专注于通用GPU芯片及解决方案研发及销售的高科技公司和另一计算机图形芯片设计公司则希望开发出接近高端芯片性能的产品。全球领先的ICT（信息与通信）基础设施和智能终端提供商旗下的人工智能处理器则被用于内部大模型开发训练。目前，国产GPU的算力规模大多在1000TOPS以内，上述某计算机图形芯片设计公司已能达到2000TOPS，但距离国际领先的H100芯片高达4000TOPS的算力仍有一定差距（见图5）。

图5：国产AI算力芯片公司产品对比

全球顶尖GPU厂商						国内顶尖GPU厂商					
公司名称	产品类型	产品	算力	主频	制程						
某全球领先的GPU厂商	训练+推理+整机	A100	1248TOPS	/	7nm						
		H100	3958TOPS	/	4nm						
某量产商业AI芯片厂商	训练+推理+整机	训推一体 MLU370	256TOPS	/	7nm						
		训练侧 MLU290	512TOPS								
某国产服务器CPU芯片龙头	训练	训练侧 DCU8100	/	1.5GHz	7nm						
某通用GPU芯片厂商	训练+推理	推理侧 MXN100	/	/	7nm						
某通用GPU芯片及超级算力系统提供商	由训练切入推理	训练侧 BI	295TOPS	/	7nm						
某计算机图形芯片设计公司	训练	训练侧 BR100	2000TOPS	/	7nm						
某AI神经网络解决方案厂商	训练+推理+整机	推理侧 i20	256TOPS	1.5GHz	/						
		训练侧 T21	256TOPS								
某AI芯片公司	训练	训推一体昆仑芯 2 代	256TOPS	/	7nm						
某互联网巨头半旗下的半导体厂商	推理	推理侧含光 800	820TOPS	/	12nm						
某通用AI GPU制造初创企业	训练	云端芯片 Compass	/	/	/						
某云端AI计算平台公司	由推理切入训练	推理侧 Goldwasser UL/	32-64TOPS/	/	/						
		训练侧 Goldwasser XL	512TOPS								

资料来源：各公司官网; 民生证券研究院

(二) 云平台：全球竞争格局趋同，未来国内GenAI厂商将依赖云平台完成模型训练与调整

在云平台领域，国内和海外市场均呈现头部厂商占据绝对主导地位的垄断格局。2022年，国内四家头部厂商占据国内79%⁴的市场份额；在海外市场，截至2023年一季度，前三大厂商占据全球65%的市场份额。

由于国内云计算行业发展尚未成熟，目前以定制化程度较高、毛利率低的IaaS为主，相关支出占比70%，PaaS和SaaS仅占30%；而全球云计算市场则恰好相反，IaaS支出占比28%，PaaS和SaaS占比72%。

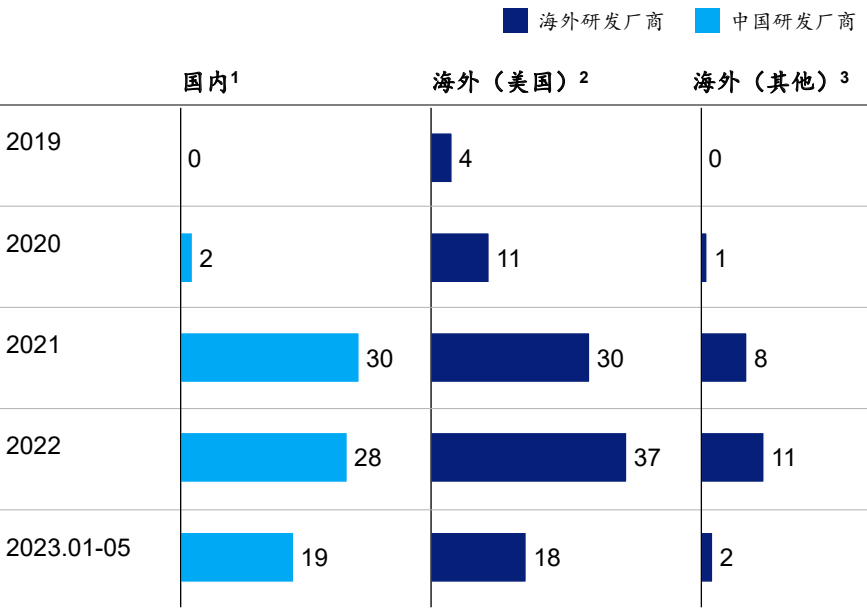
展望未来，我们预计海内外头部厂商的客户群体将持续分化，国内GenAI厂商将更多依赖国内云平台完成模型训练与调整。

(三) 基础模型：通用大模型数量可观，行业大模型可能是国产GenAI的出路

在通用大模型方面，截至2023年5月，中国已发布79个10亿参数级以上的大模型⁵，在大模型发布数量上位居全球第二，仅次于美国。在全球模型发布数量前十的GenAI大模型厂商中，中国研发机构及厂商占据4席（见图6）。

图6：全球生成式AI厂商大模型发布数量

根据2019.01至2023.05期间发布的大模型数量



注：大模型市场处于一个快速变化阶段，统计数据据发布日或已发生重大变化

1. 中国国内大模型厂商包括国内某AI研发机构，国内某顶尖高校，国内Top3互联网厂商等

2. 海外（美国）的大模型厂商主要包括硅谷众多全球领先的互联网厂商，元宇宙公司，顶尖AI研究型实验室，跨国科技公司，和AI非营利性研究机构等

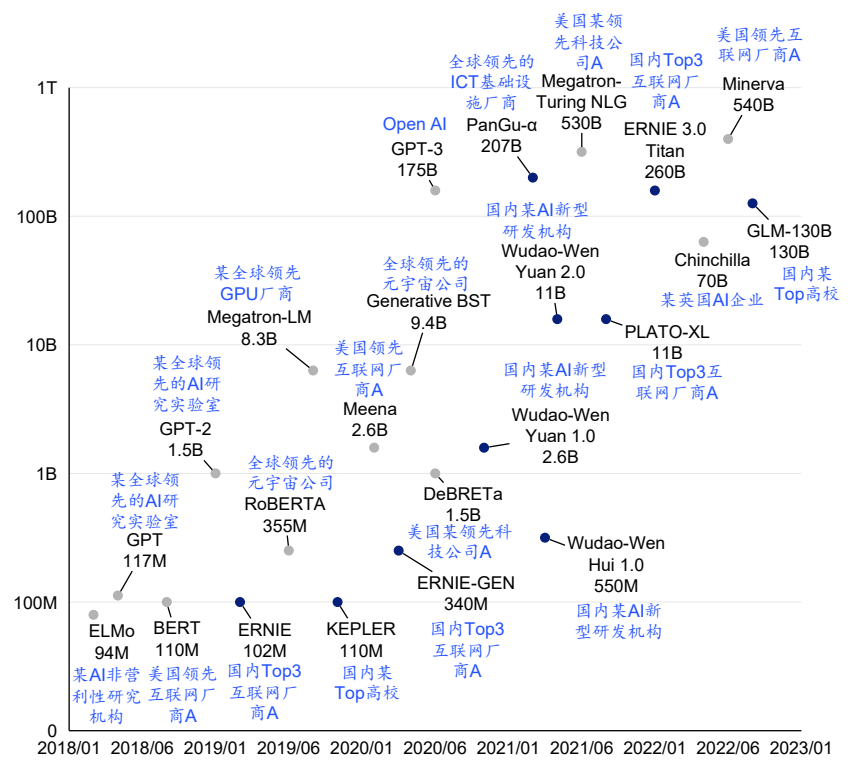
3. 海外（其他）的大模型发布厂商主要来自英国和欧盟

资料来源：公开资料收集，包括厂商公布和投行券商行研报告（例如，Leonis Capital等）

通用大模型需要海量数据作训练用途，但中文网站数量从全球来看占比仅为1.4%⁶（英文网站占比达54%⁷），可用作训练的公开中文语料库（含文本、图片、视频等素材）往往数量有限、质量不均。与此同时，国内用户在网站和移动端App产生的海量数据，基于保护用户隐私的原因，大部分也未能应用在大模型训练上，对模型训练的效率 and 精度都造成不利影响。

尽管在训练数据上受到诸多限制，中国通用大模型技术仍然不断追赶国际领先水平，参数规模紧跟国际领先水准并实现快速提升（见图7）。

图7：2018~2023年全球大模型厂商发布的主要模型



资料来源: Leonis Capital 研究报告

相较于适用于多领域、多任务的通用大模型，行业大模型更依赖于开发者对垂直场景的理解和海量行业数据支持，在算力受限的背景下，行业大模型更有可能成为中国GenAI厂商的制胜赛道。

就行业大模型而言，国内市场正呈现差异化竞争格局：某头部互联网大厂的行业大模型根植于游戏、广告、内容创作等多媒体领域，成为提升广告和游戏场景创作效率的生产力工具；国内前三的超大规模云服务商开发的产业级知识增强大模型则凭借其对中文语言的强理解能力，赋能汽车行业售后服务与知识库构建、医疗行业单据识别、社科行业版权保护与词条管理等；某全球领先的ICT（信息与通信）基础设施和智能终端提供商旗下的包含NLP、CV、多模态和科学计算的中文大模型则在工业物流、新药研发、气象预测等领域得到广泛应用。

（四）模型库和工具：国内外围绕开源模型，涌现出大模型工具独立厂商

如第二章所述，海外市场围绕可自由调整代码的开源模型，已涌现出一批独立厂商，可实现模型训练和微调、模型部署、模型应用开发等细分技术功能。随着国内GenAI市场的不断发展，各种开发和维护工具等也将逐渐成熟。



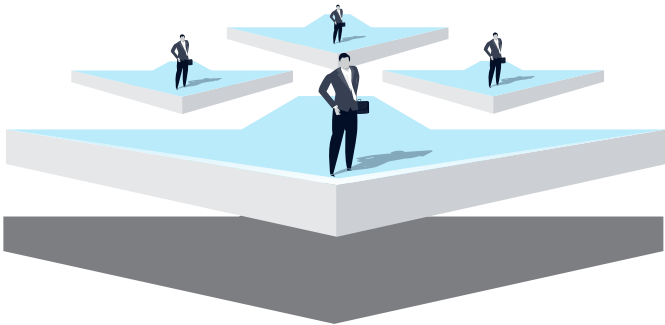
(五) 应用: 中国创业公司尚处于早期阶段, 聚焦领域相对集中

中国GenAI领域的创业公司融资轮次靠前, 集中在天使轮到Pre-A轮之间, 融资金额大多小于1亿元人民币, 正处于方兴未艾的发展初期, 行业潜力巨大。其中, 某全球领先的跨国科技公司旗下的人工智能聊天机器人作为准独角兽企业, 于2022年11月完成10亿元A+轮融资, 投后估值约10亿美元(见图8)。

图8: 国内生成式AI赛道典型融资案例

产品类型	融资时间	成立时间	轮次	金额
AI心理咨询	2021.12	2021	天使	1,000万美元
AI生成视频	2022.3	2021	Pre-A	~1,000万美元
AI生成图片	2022.5	2020	A	800万美元
AI生成图片	2022.8	2020	天使	-
AI知识智能	2022.9	2019	B	数亿元
AI虚拟人	2022.11	2020	A+	10亿元
AI角色生成	2022.11	2022	天使+	-
AI虚拟人	2022.12	2022	天使	数千万元
AI数字诊疗	2023.2	2021	Pre-A	-
AI生成3D视频	2023.2	2022	天使	数千万元
AI生成图片	2023.2	2020	A	~1,000万美元
AIGC空间	2023.3	2022	Pre-A	数千万元

资料来源: 汉能资本



反观海外GenAI行业，科创企业规模体量相对较大，涌现出多家细分应用领域独角兽企业，其中某总部位于伦敦的全球化AI研究型企业估值高达38亿美元（见图9）。

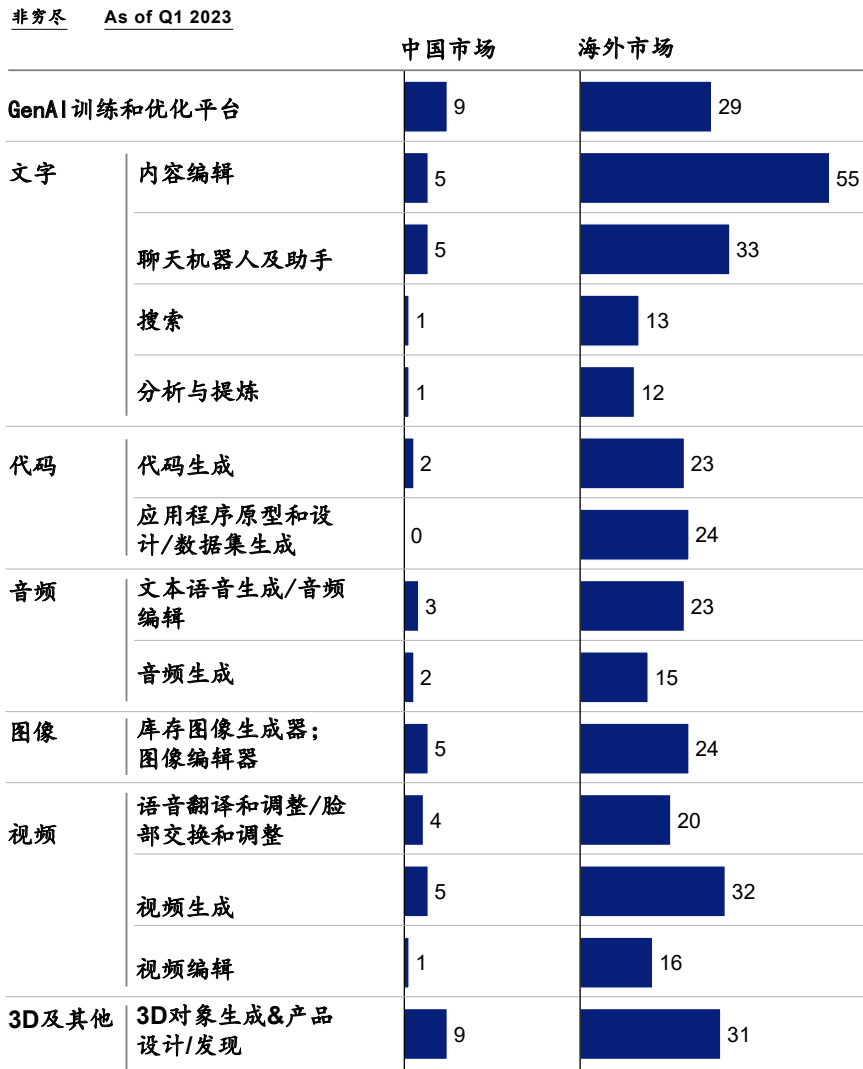
图9：海外生成式AI赛道典型融资案例

产品类型	融资时间	估值，美金
AI生成图像	2022.10	38亿
AI生成图像	2022.10	15亿
AI生成视频	2022.12	14亿
AI生成程序	2023.2	约10亿
AI虚拟人	2023.3	约10亿
AI生成音频	2022.11	5.5亿
AI生成文本	2023.1	约3亿
AI生成语音	2022.1	约2亿
AI办公软件	2022.11	1.1亿
AI生成文本	2021.1	约1亿1

资料来源：汉能资本

中国GenAI垂直应用领域主要聚焦文本、图像和音视频生成，海外市场则在上述领域以外出现了大量基于GenAI的开发平台、数据分析平台及代码编写平台（见图10），其原因在于海外以早期程序语言（如COBOL）编写的系统较多，许多企业面临较高编程人力成本，因此对编程辅助软件有较高需求。

图10：中国和全球截止至2023年Q1的按模态和应用赛道分的生成式AI初创公司融资数量



资料来源: Pitchbook (PE, VC及并购项目数据平台), Antler (新加坡base的VC平台), Crunchbase (投融资信息平台), Gain.pro (聚焦一级市场的数据平台)

同时, 目前中国GenAI企业目标客群多为终端用户, 而海外市场存在大规模企业客户群体, 随着中国SaaS市场成熟以及企业付费意愿提升, 预计国内文本生成和图像生成初创企业将快速拓展至企业客户市场。

（六）服务：潜力尚未充分挖掘

由于海内外GenAI行业尚处于快速发展期，基于大模型产品的服务市场潜力还未被充分挖掘，服务市场有待进一步发展。

三、商业化应用

中国GenAI行业的商业化应用呈现出两大特点：一是行业分布集中，主要集中在商业化发展较成熟的优势产业；二是大部分企业尚处于积极探索自身商业模式的早期阶段。

（一）中国GenAI应用集中于商业化发展成熟的优势产业

中国GenAI企业利用国内垂直场景优势，将GenAI优先应用于商业化基础设施较完善的新兴行业中；而欧美同行则利用当地成熟的SaaS市场，将GenAI应用到高科技、通信及各种传统行业（如医疗、教育等），应用范围更加广泛。

在国内，GenAI应用发展最迅速的领域包括电商、传媒、娱乐和游戏，尤其是数字虚拟人和电商视频营销，而大部分传统行业（如金融、能源、教育等）仍处于小规模试点阶段。GenAI应用得以在互联网电商平台蓬勃发展，原因在于中国拥有优质的电商和供应链生态以及庞大的消费者群体，为GenAI应用落地提供了契机。

典型的行业应用案例包括视频行业的某生成式AI的3D短视频内容厂商、电商行业的某领先GenAI视频大模型研发企业以及游戏行业的某互联网平台旗下的游戏与AI研究和应用机构等。上述企业中，生成式AI的3D短视频内容厂商能够基于文本生成3D视频内容，并通过抖音、快手、B站等短视频平台进行推广，极大提高了短视频内

容生产者的创作效率；GenAI视频大模型研发企业则深耕电商行业，以AI生成虚拟人主播，赋能各行业、各品牌电商；游戏与AI研究和应用机构则利用人工智能打造AI陪玩、AI竞技机器人，并通过自然语言技术赋予非玩家角色人设，使其能够在游戏中与玩家互动对话。

（二）中国市场GenAI商业模式尚在探索中

目前，大部分中国市场的GenAI创业公司刚刚完成标准化产品输出，开始进入初级商业化探索阶段，市场上主流的商业模式包括云资源售卖、模型API调用、SaaS收费、素材收费等。由于国内企业对软件的付费意愿不高、市场有待进一步培育，以及企业对SaaS部署方式存在数据安全方面的顾虑，大规模应用GenAI的商业模式仍有待探索。在欧美市场，SaaS付费订阅模式已基本成熟，一众海外GenAI企业纷纷构建起可持续的SaaS商业模式。



四、风险管理

GenAI赋能各行各业的同时，也带来了诸如公平性受损、侵犯知识产权、信息泄露、恶意使用、安全威胁、模型幻觉及第三方风险等负面影响和挑战。其中，模型幻觉、恶意使用、信息泄露三大风险值得重点关注（见图11）。

图11: 使用GenAI 有 8 个关键风险类别

风险类别	潜在风险描述	生成式 AI 对风险格局的影响
公平性 受损	算法偏差;算法决策背后缺乏问责制;将生成的内容错误地表示为人为创建的内容	生成式 AI 可能会由于开发/部署阶段的不完善训练数据或工程决策而导致算法偏差 生成式 AI 可以生成无法明确识别为 AI 生成的内容, 从而导致用户混淆或欺骗
知识产权 侵权	侵犯受版权保护、商标、专利或其他受法律保护的材料	生成式人工智能和基础模型通常利用基于互联网的数据, 导致潜在的知识产权侵权事件 (例如侵犯版权、抄袭)
信息泄露	未经授权使用/披露个人信息或敏感信息, 违反隐私要求或期望	生成式人工智能可能会通过 (可能无意的) 使用用于模型训练的个人或其他敏感信息来加剧隐私问题, 从而产生潜在的敏感结果。
恶意使用	人工智能生成的恶意内容发布 (例如虚假信息、欺诈)	可以利用生成式 AI 来创建和传播恶意内容 (例如, 虚假、仇恨言论、广告活动)
安全威胁	生成式 AI 系统中可能被破坏或利用的漏洞, 从而导致安全性或合规性问题	生成式 AI 驱动的应用程序可能会受到安全漏洞和操纵的影响 (例如, 用户或不良行为者可能会通过混淆、有效负载拆分、虚拟化绕过安全过滤器)
模型幻觉	无法正确解释或解释模型输出 (即“黑盒算法”问题);对可能导致不良结果的不准确性进行建模	生成式人工智能对具有众多参数 (例如数万亿) 的神经网络的依赖, 在解释结果如何生成的能力方面带来了挑战。 基础模型可能会产生事实不正确或过时的答案
环境、社会、社会及管治	不符合环境、社会及管治标准;模型开发和使用的社会或环境后果造成的声誉损害	基础模型的培训和部署可能会增加碳排放并超出 ESG 承诺或预期 生成式人工智能可能会扰乱或取代员工, 带来新的挑战 (例如, 声誉风险、人才短缺)
第三方风险	与使用或集成第三方 AI 工具相关的风险 (例如, 缺乏控制、依赖、锁定)	组织通常使用或集成外部生成AI模型和工具, 而不是在内部开发它们, 从而带来与模型结果和依赖关系相关的新型第三方风险。

关键风险可能进一步导致监管、法律、声誉和商业后果

资料来源: 麦肯锡分析

- **模型幻觉:** 模型输出结果不符合事实。模型幻觉普遍存在于国内外所有的GenAI模型和应用, 但因为国内大模型的研发经验和技術实力尚有提升空间, 且中文语料相比英文更为复杂, 让模型幻觉问题变得更加棘手。

- **恶意使用:** 不当使用模型输出可能会带来欺诈、虚假信息传播等负面影响, 中国GenAI应用丰富、用户群体庞大, 所面临的恶意使用风险较为突出, 但由于国内在实名认证, 内容合规, 和打击虚假信息等方面落实到位, 有利于追溯恶意使用源头, 从而防范和控制相关风险。
- **信息泄露:** 用于模型训练的隐私或敏感信息可能会被GenAI应用作为输出内容, 同时人工智能模型存储系统内的个人信息也可能面临网络攻击和泄漏, 因此需要在内部部署数据、又处于强监管下的重点行业和企业, 需要主动防范信息泄露风险。
- **安全威胁:** 不仅负责任的AI (RAI-Responsible AI) 受到监管和企业关注, 安全的AI (SAI-Securing AI) 也逐渐进入企业和公众视野, 特别是AI系统全生命周期安全保护。有效的AI威胁管理直接影响到防范信息泄露和恶意使用。近期全国信息安全标准化技术委员会起草了相关评估规范和标准, 为算法提供者给与了机器学习算法生命周期安全指导意见⁷, 也可为监管评估提供参考。

为规范GenAI的应用, 中国早在2022年已经开始布局AI相关监管, 例如2022年3月发布了互联网信息服务算法推荐管理规定, 2022年11月发布了互联网信息服务深度合成管理规定。2023年7月, 国家网信办联合国家发展改革委、教育部、科技部等7部门发布《生成式人工智能服务管理暂行办法》, 对模型训练过程中可能涉及的数据歧视、知识产权保护、个人隐私及肖像权保护等作出明确规定, 行业监管框架进一步完善。

放眼海外, 美国国家标准与技术研究院于2023年1月发布了AI风险管理框架, 着力为AI开发, 治理, 和运营提供风险控制最佳实践。欧盟则于2023年6月通过《人工智能法案》授权草案, 对不同等级的人工智能所承担的责任和义务进行划分, 通过风险定级来实现分层治理, 并着力推动跨产业的人工智能立法, 以进一步防范人工智能可能带来的多重风险。

展望未来，中国GenAI相关的法律框架有望不断完善和细化，预计监管机构将通过建立分行业、分场景、分应用的差异化监管体系来兼顾各方权利与责任。

结语

当下全球人工智能市场正处于高速增长阶段，本文从市场规模、技术栈、商业化应用和风险管理四大维度，对国内外GenAI行业的发展现状和趋势进行了深入浅出的分析。尽管目前国内外市场在诸如大模型开发、应用布局、商业模式等多个领域都存在较大差异，我们看到中国GenAI行业正在不断追赶国际领先水平，预计未来中国GenAI相关技术和应用将逐渐成熟，并进一步探索出适合自身发展的商业模式。

¹ 数据来源: Bloomberg Intelligence
² 数据来源: Gartner
³ 数据来源: 《中国AI数字商业产业展望2021~2025》，前瞻产业研究院、中关村大数据产业联盟联合发布
⁴ 数据来源: Canalys 《2022年第一季度中国云计算市场报告》
⁵ 数据来源: 2023年《中国人工智能大模型地图研究报告》，中关村论坛
⁶ 数据来源: “2023年网站所用语言统计”，互联网研究网站W3Techs
⁷ 信息安全技术-机器学习算法安全评估规范，预计2024年3月实施

曲向军是麦肯锡全球资深董事合伙人，中国区金融机构咨询业务负责人，常驻香港分公司；
韩峰是麦肯锡全球董事合伙人，常驻深圳分公司；
胡艺蓉是麦肯锡全球副董事合伙人，常驻上海分公司；
王喆宸是麦肯锡咨询顾问，常驻上海分公司。

感谢Ray Zhou对此文的贡献。

麦肯锡公司2023年版权所有。



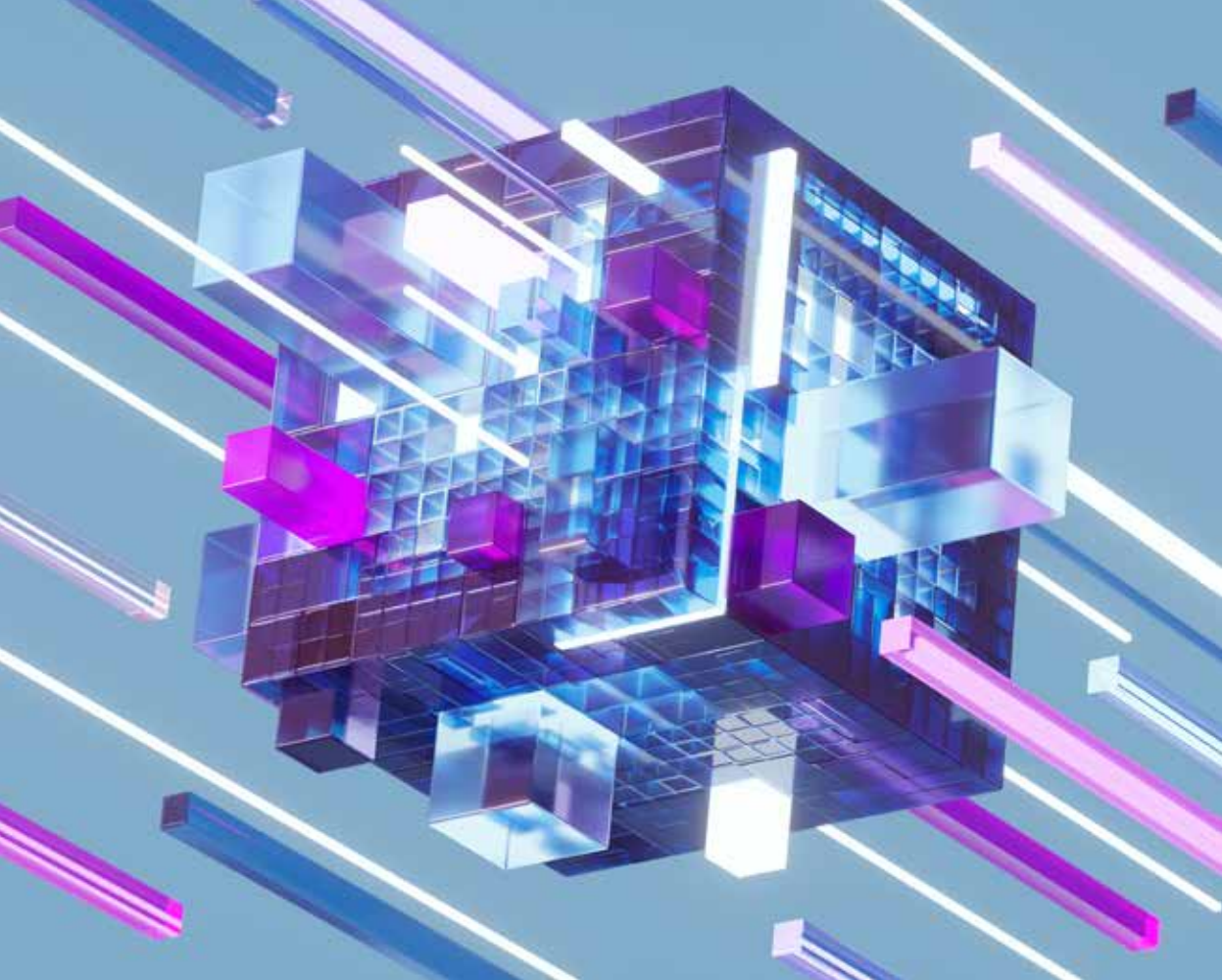
第四部分

行稳致远，企业规模化部署生成式AI的“道”与“术”

146 从战略到规模化实施，“四步走”全面落地
GenAI

167 让 GenAI 应用真正为企业创造商业价值





从战略到规模化实施， “四步走”全面落地GenAI

本章深入浅出地分析了企业规模化落地GenAI应用的流程，提出企业在引入GenAI时可遵循“四步走”方法论。

曲向军，韩峰，胡艺蓉和蒋子翔

引言

生成式AI (GenAI) 在不同行业和领域引发大量讨论,与上一波由DeepMind AlphaGo引发的全球人工智能热潮不同,这次除了投资圈和大企业外,资源相对较少的中小型企业也能通过诸如ChatGPT和Microsoft 365 Copilot等工具为员工和客户赋能,快速提升生产力和服务效率,实现弯道超车。

传统AI在企业落地的一大难点是数据质量差,经常被戏谑为“garbage in, garbage out”,导致许多AI项目回报周期长甚至遥遥无期。很多企业在前期评估后的结论是:若要享受AI带来的各种成果,需要先对数据治理进行大量投入,或直接判定企业尚未具备实施AI的条件。令人欣喜的是,最新的大语言模型(LLM)及图像生成等多模态大模型都已经涵盖了大量数据(如GPT4的训练数据集就达到13万亿token),这使企业无需对IT系统或数据清洗进行大量前期投入,即可借助GenAI达到既定目标。

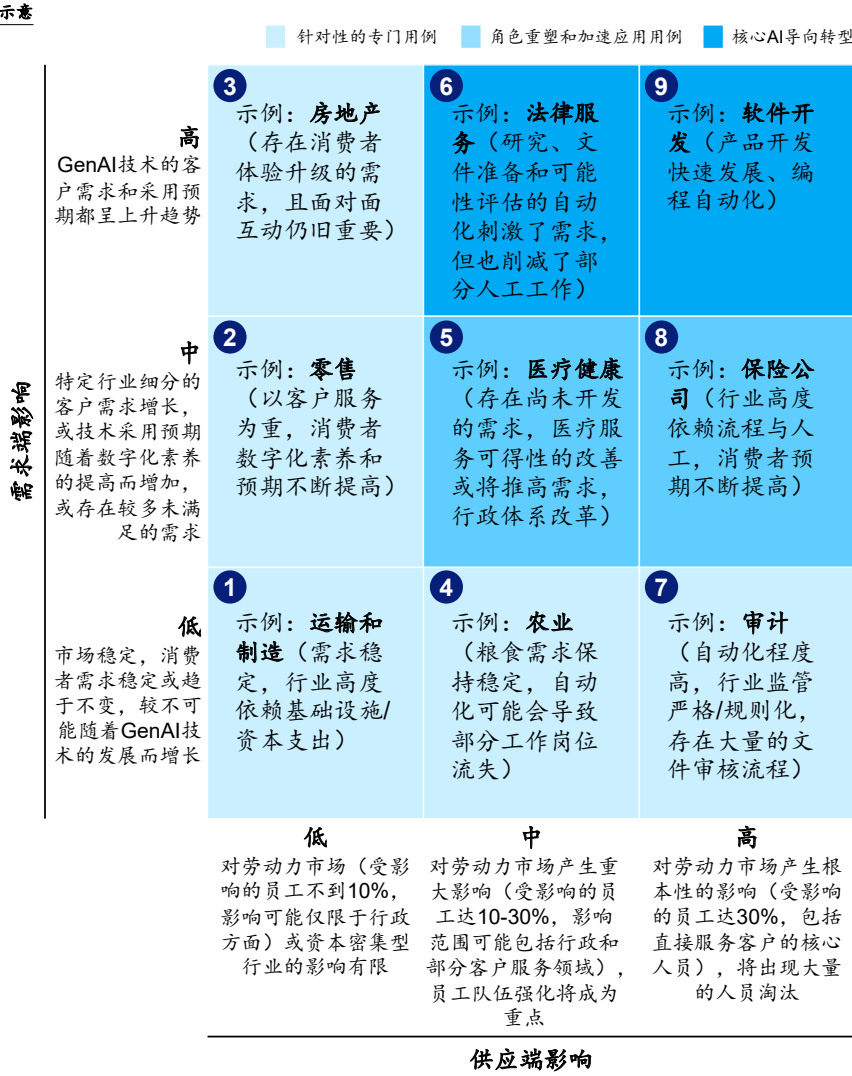
当下,全球人工智能技术正以空前速度持续创新和迭代升级,麦肯锡相关研究和项目经验表明,企业在考虑引入GenAI时应以速度为先,遵循“四步走”方法论,高效实现GenAI部署和规模化应用。

第一步:构建GenAI“内、外、快、慢”四大战略

企业在制定GenAI实施战略时,应同时兼顾**外部因素(如市场、股东等)**和**企业内部因素**。就**外部因素而言**,企业应从自身战略出发,站在宏观高度思考GenAI对行业生态的影响,包括但不限于直接竞争对手、上下游价值链转型等。例如,某全球领先GPU厂商股价从去年10月至今暴涨逾110%,使价值链上下游企业的估值也得到明显提升。此外,企业在制定相关战略时,应一并考虑与股东、投资者沟通时的话术和投资故事。

就内部因素而言，建议企业不能仅关注自动化对成本效能等方面的提升，更重要的是考虑如何借助GenAI重塑企业运营及文化（如鼓励创新等），并利用相关技术赋能企业人才。例如对软件开发供应商而言，GenAI带来的自动化代码编程能力可节约30%以上的人力资源（见图1），且在减省成本以外，还能提升现有系统的开发功能，实现产品快速迭代并优化用户体验，加速淘汰未能适应新技术趋势的竞争对手。

图1：因不同行业在供需分析矩阵中的位置差异，企业应找准自身卡位



资料来源：麦肯锡分析

在充分考虑对内、对外战略的基础上，企业应该同时制定“快、慢”两个实施赛道。所谓“慢赛道”是指GenAI端到端的体系化转型，企业应对所有GenAI战略选项进行快速梳理和全景扫描，并将其分为解决现有业务痛点（防守）以及驱动增长和新业务发展（进攻）两大类型，然后再根据所处行业特点、企业禀赋和供需分析，构建起最适合自身发展的GenAI战略。

同时，企业应兼顾“快赛道”的速赢举措，通过快速实现价值为企业高层提供“定心丸”，与慢赛道相辅相成。企业应尽快识别适合自身行业和定位的GenAI用例，重点是要通过用例或POC试点，短期内确定企业内部推动变革和GenAI落地的人才，以及外部技术合作方（如软件厂商、咨询公司、技术提供商、学术机构等）。企业应积极推动与同业、异业生态圈的各类合作，借力打力，充分把握算法升级迭代带来的重要机遇。

在试点成功落地并得到初步反馈后，企业应持续调整和完善现有GenAI战略，以动态适应不断变化的技术和市场趋势。



第二步：基于负责任的AI，制定转型路径图，聚焦用例“先动起来”

确立初版GenAI战略后，企业应在“负责任的AI (Responsible AI)”框架下，快速制定初版用例图谱和转型路径图。

(一) 负责任的AI (Responsible AI, 简称RAI)

负责任的AI是一个整体治理框架，尽管如何运用AI技术属于企业内部经营决策，但鉴于AI自动化执行和智能决策有可能对法律法规、社会伦理、产品和运营安全性产生巨大而深远的影响，企业（特别是金融业）在部署AI前应充分考虑各方面因素，并制定相应的RAI框架，为可能出现的AI风险提前准备预案和应对措施。

目前在全球范围内，RAI并没有统一标准，企业需要按自身战略和实际情况制定最适合的RAI框架，我们认为框架中至少应该包括：

- 明确AI模型“不适用”的场景，例如欧盟通过的《人工智能法案》对AI在招聘、绩效评估等方面的应用设有限制；
- 明确AI模型在透明度、可追踪性、公平性和稳定性等方面的要求；
- 明确AI用例在投产前所必需通过的基础测试和审批流程；
- 明确AI模型和应用在保护个人及企业数据隐私安全方面的具体要求。

全球各大AI厂商都制定了适合企业自身情况的RAI框架，以微软为例，其对于“负责任的AI”的定义包括六大要素：公平性、可靠性和安全性、隐私和安全、包容性、透明度、问责¹。此外，合规性和可靠性也是负责任的AI框架的重要内涵。例如，中国生成式人工智能服务管理办法强调AI模型生成的内容合规性。而在美国NIST AI风险框架和欧洲AI法案²提到内容可靠性。

(二) 用例图谱和转型路径图

就制定用例图谱而言,企业应首先明确需要优先进行GenAI部署的业务领域,为各业务分别设计用例,并形成GenAI整体转型路径图。具体而言,主要包括以下七大方面:

首先,为了将一个概念或者想法转变成实际行动,转型团队需要得到**高管和业务负责人的承诺**。这就涉及到对业务领域的定义,以及明确谁是关键利益相关者。团队应该在获得高层的充分承诺和支持后,再着手推进下一步行动。

接下来,团队需要**明确大模型用例**的主要目标。这一步骤可能涉及自然语言处理的相关任务(如生成摘要或分类信息、处理某些特定场景等),其主要目的是厘清大模型将用于哪些领域以及可能产生的效果。

然后,就是明确如何实现这一目标,即**确定技术方案**。这包括大模型从哪里获取数据、这些数据具体包括什么内容以及如何安全有效管理相关数据等。企业在这个阶段需要考虑的因素很多,包括数据安全性和风险管理等。

第四步是**确定用例优先级**。这涉及评估用例潜在效果(如收入、成本节约、客户体验等),并根据评估结果确定每个用例的优先级。此外,企业需要将用户需求和数据敏感性等因素一并纳入考量。

一旦确定了上述内容,就可以**开始设计和实施用例**了。在设计过程中,企业应确保所选择的用例符合企业自身使命、道德规范和整体业务战略,以确保企业在整个实施过程中角色定位明确,以及所选择的基础架构能够真正满足企业的实际需求。

在此基础上,企业可进一步设计**具体用户界面和体验**。根据用户在不同业务上的需求,输出形式可以多样化,如用户界面、用户交互、看板、驾驶舱、商业智能报告等。同时,企业需要建立一个高效评价机制,持续评估和优化模型输出。

最后，就是**制定转型路线图**。企业应根据前期战略和用例图谱，制定出详细计划，明确每个阶段的工作内容、时间节点和责任主体，以保证企业在推进转型过程中，始终保持清晰目标和方向，避免出现混乱。

第三步：评估用例试点成效，优化整体落地蓝图和路线图

在快速形成初版GenAI转型路径图及用例图谱后，企业需要根据用例的重要性与潜在效益进行优先级排序，以“先动起来”的方式逐步推进试点工作，为全公司规模化应用打下坚实基础。对于初次实施AI技术的企业，我们建议应从概念验证（Proof of Concept）试点开始，逐步开展工作。

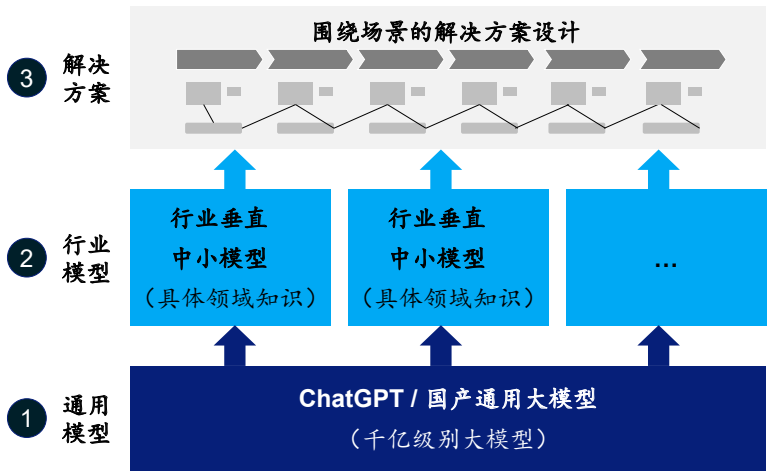
在实施试点前，为了评估某个用例的优势和弱点，企业可基于两个维度快速分析：**潜在成效**和**可行性**。潜在成效主要考察用例在创造业务价值方面的能力，此外企业还需要考虑单个用例在不同部门间的可扩展性，即这个用例能否应用于多个场景和部门。最后，企业还需要评估具体用例对人才、组织、业务流程和合作伙伴的价值，以了解其对整个企业生态的影响。

在**可行性**方面，企业需要评估具体用例的实施潜力，即在实际操作中能否成功执行。此外，企业还需要考虑落地的复杂程度，即实施这个用例需要花费多大的精力，同时估算实施用例所需要的实际工作量，以及可能遇到的内外部阻碍等。我们建议企业可以选取**潜在成效和可行性较高的2~3个用例进行POC试点**。

概念验证 (POC) 试点

在实际操作过程中,企业可从以下三个方面着手落地POC试点,具体包括:基础模型选择、行业模型选择以及场景解决方案设计(见图2)。

图2: GenAI POC核心设计点



对于大中型客户而言，落地大模型所需要做出的三个关键设计：

- 1 通用模型选择：**
 - 通过通用模型，用聊天对话的交互方式，调用千亿级别大模型的生成能力
 - ChatGPT是个交互入口，底层是OpenAI已经训练好的海量隐形知识库
 - 当前中国企业有Azure ChatGPT接口和国产大模型两种选择
- 2 行业模型选择：**
 - 通过行业中小模型，以相对可控的算力资源，基于具体领域的情况和知识，训练出企业需要的模型
 - 可选择多种部署及训练形式，其中“调用微调”需要更多的客户数据但可实现更大程度的客制化和安全性
- 3 场景解决方案设计：**
 - 基于具体使用场景的流程，进行解决方案设计及相关系统集成
 - 如银行应用场景，为满足合规性需求，进行本地私有化部署是必选项，且需要通过API集成CRM、营销系统、BI系统等
 - 由于模型训练及系统集成所需巨大资源，成本使用效率及落地价值成为解决方案设计的核心

资料来源：麦肯锡分析

1、基础模型选择：基于现有大模型或自研大模型进行开发

在基础模型方面，企业有两种选择：一是基于现有大型人工智能模型开发应用，二是从头开始研发自身模型，具体如何选择取决于企业的实际需求和资源。

使用现有大模型是一个经济、高效且实用的方式，企业可以将预先训练好的大模型转化为各种应用。这些模型主要分开源和闭源两种类型，**开源模型**可以在开发者社区中找到，可根据企业需求定制化开发；**闭源商业模型**则由人工智能模型厂商开发，按照企业具体用途和使用量收费。在**海外与国产商业模型**选择上，企业可以根据自身情况选择使用海外领先大模型接口或国内开发的大模型接口。

选择**自行研发大模型**的企业，需要拥有强大的大模型开发技术能力，并仔细评估投入和回报是否符合预期。开发自研大模型往往需要投入海量数据与高昂资金，例如OpenAI仅训练GPT-3就花费了约400万~1,200万美元，其在规模更大的GPT-4上的投入则可能高达数千万至上亿美元。

2、行业模型微调：针对不同大模型进行微调

企业在开发GenAI应用时，往往需要嵌入所处行业的特定场景，现有通用大模型并不能完全满足其专业化需求。因此，通常做法是在通用大模型的基础上进行微调（fine-tune），以得到更适合垂域场景的行业基础模型。

一般来说，行业大模型主要有以下两个优势：一是**技术门槛相对较低**。尽管基础模型的训练需要海量数据，成本高且耗时，但在通用模型基础上进行微调则只需要较少的数据，成本相对更低，而且可以在较短时间内完成，从而大幅降低了进入门槛。通过加入额外的行业数据或按需调整模型参数，企业就能为特定应用场景生成所需内容。另一方面，这种技术具有**很强实用性**。企业可以利用相对

可控的计算资源，结合行业现状和相关数据，训练出符合自身需求的模型。

行业模型微调包括多种部署和训练形式，允许企业根据实际需要选择。例如企业可采取“调用+微调”方式，尽管这种方式需要输入更多客户数据，但可以最大程度满足定制化和安全性方面的要求，使企业更好应对不断变化的市场环境。

3、场景解决方案设计：集成现有系统和功能，提供基于场景的客制化解决方案

在这一环节，企业首要任务是根据具体使用环境进行**流程设计**，这对后续解决方案的成功执行至关重要。以银行为例，为了满足各种合规要求，一种可能的解决方案是选择本地私有化部署，意味着银行需要把所有数据和计算资源都保存在本地，以确保客户数据隐私和安全。因此，在设计流程时，企业需考虑如何让各种业务流程与本地私有化部署结合，在合规和业务需求间找到平衡。

接下来，企业需要考虑的是与**相关系统的集成**。仍以银行为例，为实现由AI大模型驱动的业务流程自动化和智能化，企业需要通过API集成CRM、营销系统、商业智能系统等。只有这样，客户关系管理、营销策略以及商业数据分析等重要功能才能与核心业务流程无缝连接，为银行业务运营提供全面支持。

另外，由于AI大模型训练和系统集成需要大量资源，因此在设计解决方案过程中，**成本效益和投资回报率 (ROI)** 是企业需要特别关注的要点。企业在设计过程中，不仅要关注技术层面的实施，还需要充分考虑到成本效益和投资回报率，以确保解决方案具有实际应用价值。这就需要企业进行详细的成本分析，同时优化资源使用效率，以确保解决方案能顺利落地。



第四步：规模化推广GenAI，需要企业进行全方位的运营模式转型，并将AI内嵌到业务的每个环节

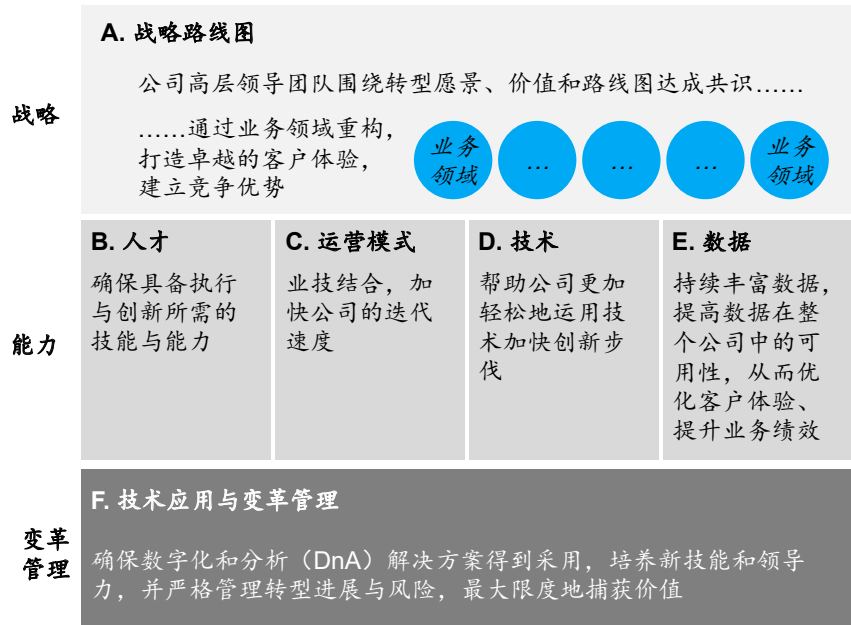
（一）规模化实施GenAI应用的六大要素

在规模化实施GenAI应用时，一个成功的运营模式应涵盖六大方面（见图3）：战略路线图、人才、运营模式、技术、数据以及技术应用与变革管理。

首先，领导层和员工的理解和接受度至关重要。企业想要把人工智能深度融入日常运营中，就需要通过培训和沟通来提高员工对新技术的理解和接受度。当员工能够充分利用人工智能的优势时，团队工作效率和质量将得到全面提升。

此外，企业还需要建立一支专门的人工智能专家团队，并通过持续培训提升员工能力，使所有员工都能将人工智能工具应用到日常工作中，这意味着企业需要在组织、人才等方面进行投入和改变。

图3: GenAI规模化实施所需能力模块



资料来源：麦肯锡分析

在运营模式上，企业需要将人工智能融入现有业务流程，并根据新技术带来的影响，重塑业务流程，这可能会涉及资源分配和考核体系优化。同时，企业需要构建强大的技术能力，以满足部署人工智能大模型的各种要求（如端到端基础技术能力、模型运维能力等）。

数据也是企业实施人工智能应用时必须考虑的重要因素。企业需要建立高效存储和处理大量复杂数据的技术能力，同时还需要培养员工将内部数据与人工智能相结合的技能，以提升GenAI应用落地的实际效能。

最后，为了确保人工智能解决方案顺利落地，并有效管理转型进程和风险，企业需要全面启动变革管理，以更好支持大规模技术能力建设和文化变革，帮助内部员工逐渐适应新的运营环境。

接下来,我们将围绕人才、运营模式、技术、数据、技术应用与变革管理展开分析,战略路线图由于已在上文中体现,故不再重复叙述。

1、组织人才: 设立专门的GenAI专家团队或部门, 以开展持续的能力建设和培训

转型重点: 企业在引入GenAI技术时,需要建立一支专门团队或部门,推动新技术研发、应用和推广。相关团队需要深入研究GenAI技术的最新发展,确保所部属的应用与企业需求高度契合。同时,团队成员应具备丰富行业经验和强大技术实力,只有这样才能更深入理解企业需求,并针对行业实际情况提供定制化的GenAI解决方案。

其中,团队成员的技术实力(尤其是在应对技术挑战和需求变化方面的能力),对企业而言至关重要。他们需要具备一系列复杂技能,例如熟练运用提示词工程(Prompt engineering)技术,掌握如何优化和部署GenAI模型,同时还需要熟悉多种深度学习和GenAI框架(如TensorFlow、PyTorch等),以及Llama-index等应用开发框架。此外,诸如高性能计算、分布式计算和大数据处理等技能也是必需的。

针对GenAI带来的技术变革,企业应当投入资源提升关键岗位的技能,帮助员工掌握GenAI相关技能及工具运用,以提升生产效率和工作满意度。

全球GenAI技术的快速发展,促使企业从上到下、从GenAI专职团队成员到每名员工都必须**持续学习**,因此企业需要主动提供足够的培训和交流机会,帮助GenAI团队跟上技术发展步伐,同时培训普通员工获得相关技能,为进一步扩展GenAI用例的应用范围打下坚实基础。部分大型企业甚至可以考虑打造内部的AI人才学院,为不同岗位的员工提供定制化课程和培养计划,以充分挖掘内部人才的潜力。

潜在阻碍:精通基础模型、GenAI技术并且熟悉业务的复合型人才数量有限,例如同时具备生成式AI中间件和应用开发经验的人才在市场上极少。

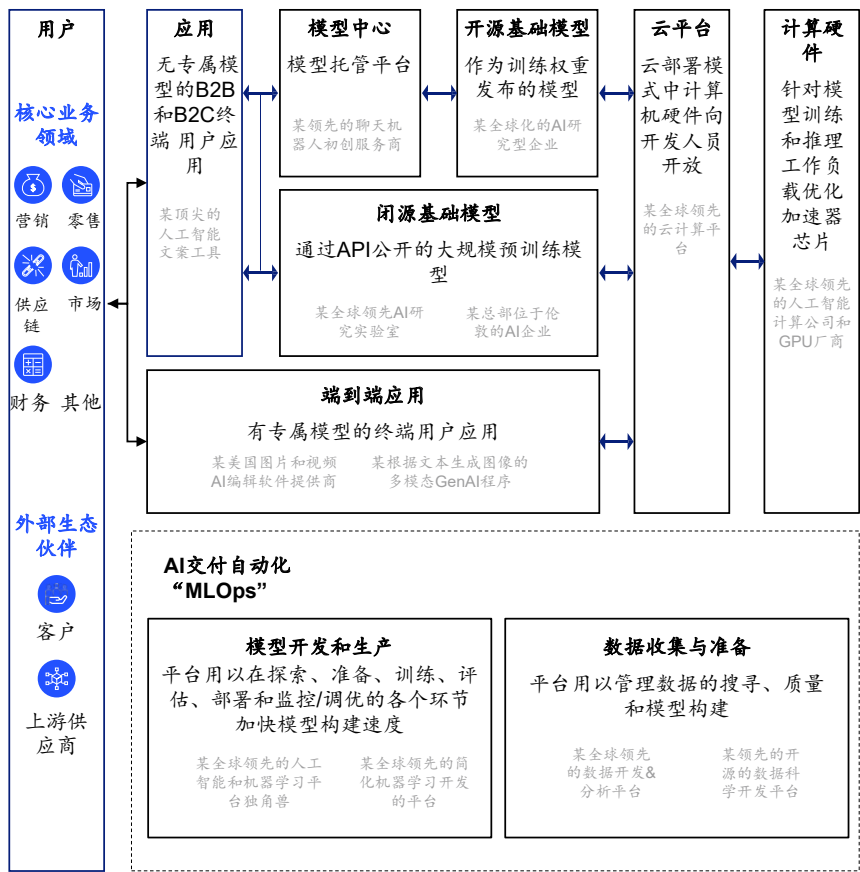
2、运营模式: 将GenAI融入现有业务流程,并根据GenAI的应用效果进行流程重塑,并优化配套资源分配和考核体系

转型重点:大部分企业在GenAI应用上采用平台运营模式,即在现有基础设施上构建及整合GenAI能力,快速和规模化应用GenAI。这就需要创建一个跨职能的GenAI中央平台团队,负责微调和维护企业GenAI模型及平台能力,根据业务需要向各团队提供经过审核的模型能力,并且通过标准化管理防范风险。

在转型过程中,企业首要任务是基于前期规划的GenAI用例,设计和推行**新业务模式**,这涉及到重新审视和调整业务流程及组织架构,以确保充分发挥GenAI技术优势。企业可根据GenAI技术的潜在效益,在关键业务领域对现有流程和职能全面革新,明确岗位角色和职责,以适应GenAI技术的快速发展。



图4：GenAI大模型技术栈地图



资料来源：麦肯锡分析

在这一过程中，**优化资源配置和调整考核体系**同样重要。企业应根据GenAI用例实施优先级合理分配资源，这可能涉及将资源分配到不同部门和项目，并根据实际情况进行调整。此外，为了鼓励员工积极参与新技术推广和应用，企业需要优化内部绩效指标和考核体系。

潜在阻碍：当企业尝试对传统运营模式进行革新，或者因实施自动化技术而缩减人员规模时，可能会严重打击士气，并在各层级和部门中产生巨大阻力。因此，企业应积极管理和应对内部阻力，确保转型顺利推行。

3、技术栈: 组建强大且能满足GenAI大模型部署的现代技术栈, 包括搭建端到端的GenAI技术能力及模型运维能力

转型重点: GenAI大模型的核心技术栈包括多模态大模型、数据库、MLOps工具、计算硬件等关键组成部分(见图4), 企业在转型过程中应重点关注以下三个子模块:

首先是**多模态大模型**, 包括通用基础模型、适用于特定领域的行业模型以及解决特定问题的专有模型。其次, 是存储和访问基础模型的**数据库**, 它们包括大模型专用的矢量数据库(如某向量相似性搜索引擎, 和某云原生开源向量数据库)、非结构化数据库(如某基于分布式文件存储的开源数据库系统和专门用于存储和导航关系的图数据库)以及结构化数据库(如SQL、Oracle等传统关系型数据库, 及Hive等分布式数据库)。

最后, 企业还需要专门的**MLOps工具**和技术来微调基础模型, 并将其部署到应用中。上述MLOps工具和技术主要包括: 分布式训练加速框架(如某开源深度学习训练优化库等)、大模型微调技术(如LoRA、P-tuning等)、GenAI通用模型框架(如PyTorch、TensorFlow等)以及模型部署工具(如Ray Studio、NVIDIA NeMo等)。此外, 企业还需要各种应用开发工具(如LangChain、Llama-index等)以及端到端模型管理平台(如H2O LLM Studio、MLFlow等)。

潜在阻碍: 企业可能因为各种原因, 不能获得所需的基础架构(例如云)和工具/API(如某全球顶尖的AI研究实验室和某企业级GenAI平台独角兽等)。此外, 由于大模型私有化部署成本极高, 成本和产出比也是需要考虑的因素之一。总而言之, 企业需要充分评估所有潜在问题, 并根据自身情况寻找合适的解决方案。

4、数据能力：为全面支撑GenAI应用，企业应建立高效存储、访问和处理海量多模态数据的能力，并打造自身数据护城河

转型重点：对于企业来说，**数据中台的架构和设计**非常重要。首先，企业需要一个能够处理大量数据的系统，不仅要存储海量数据，还要进行高速读写操作以及数据清洗和转换等。在引入新技术应用时，该系统应具备高效处理非常规数据、存储文本向量信息的能力，这是非常关键的一环。

此外，**数据治理体系的设计与实施**也是不可或缺的一部分。企业需要一个治理体系来全面管理、控制和保护数据，确保数据完整性、准确性、一致性和安全性。具体来说，数据分类、数据质量管理、数据安全和隐私保护，以及数据备份和恢复等都是该体系的重要组成部分。

同时，**数据伦理、安全和隐私问题**也不能忽视。企业需要遵循道德规范，公正和透明地使用数据，并确保数据保密性和完整性。这需要企业主动采取一些必要措施，如数据加密、访问控制、身份验证和数据备份等。

另一方面，企业需要**统一管理和维护核心数据资产**，以确保数据一致性、准确性和完整性，这对企业决策和分析至关重要。同时，全面管理和利用数据资产，也是确保数据资产价值和实用性的重要手段。

在当前大模型算法和应用快速发展的背景下，专有、特色和**高质量数据**变得越来越重要。因此，企业在构建完善的数据基础设施的基础上，还需要持续收集、存储和管理高质量的结构化/非结构化数据资产。

潜在阻碍: 企业可能会遇到诸如难以获取高质量的非结构化数据（如图像、PDF报告等）、国内企业数据出境和上云的安全问题，以及与消费者数据（如客户交易数据）相关的知识产权和隐私安全等问题，对此企业需要积极寻求解决办法。

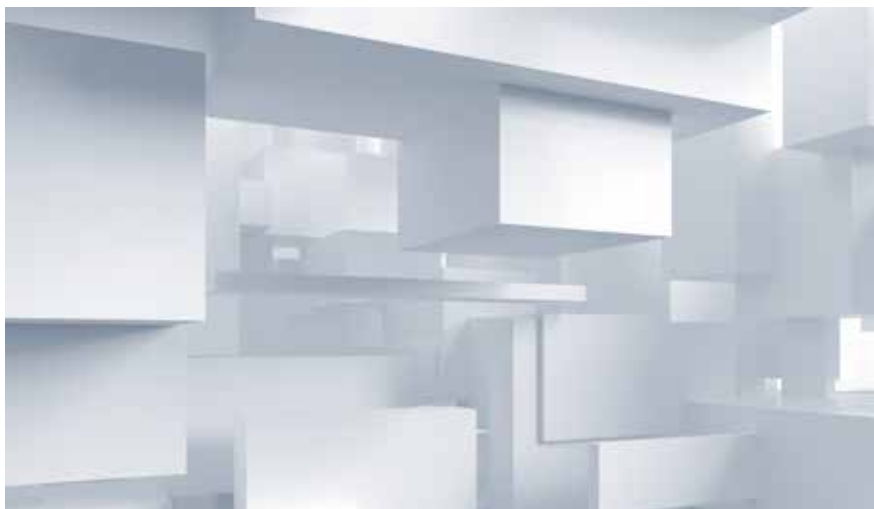
5、技术应用与变革管理：为确保GenAI解决方案全面落地、持续监督转型进程和管控风险，企业需全面启动变革管理，以支持大规模技术能力建设和文化变革

转型重点: 企业需要制定中长期计划，并持续培训，帮助员工理解并掌握新技术，提升工作效率和质量。这不仅涉及到员工技能提升，更是一场深刻的内部文化变革。

与此同时，理解和管控新技术的潜在风险同样重要。企业需要制定行之有效的应对策略，建立完备的**风险管理机制**，以确保在新技术应用过程中及时应对各种潜在风险和挑战。

对于新技术运用，企业还需进行严格审计，确保其**符合道德和法律规定**。在审计过程中，一旦发现问题，必须及时整改，以确保企业技术实践合规、可持续。其次，建立良好**沟通渠道和反馈机制**。及时了解员工对新技术的反馈和建议，帮助企业持续优化和改进新技术应用，从而更好推广新技术。最后，构建一个开放**生态系统**，让企业更好利用内外部人才、技术和合作伙伴，持续创新和发展。通过与合作伙伴共同研发和推广新技术，企业可实现产业链协同发展，从而提升自身竞争力。

潜在阻碍: 整个组织对新技术的理解可能还不够透彻，领导团队可能对新技术信心不足和支持不够，这些都需要企业认真对待并着手解决。



(二) 落地GenAI应用的两种布局路线： 开发POC应用、自研行业大模型

路线一、开发POC应用

在该路线下，企业借助基础LLM模型和各种辅助工具（如直接调用头部科技厂商的应用程序开发托管平台和人工智能开发平台的API），开发基于特定场景的自然语言交互应用，以解决实际问题并提升效率体验。这类应用往往先从1~2个业务领域的POC试点开始，例如营销客服机器人、个人办公助理和虚拟秘书（Copilot）、IT代码开发和运维助理等，进而实现全面推广和流程重塑。

典型案例包括某全球领先的客户关系管理（CRM）软件厂商发布的GPT产品，可以自动执行销售任务（例如自动发送电子邮件、安排会议、下发销售线索等），并自动总结笔记和沟通过程中的关键点；某领先GenAI应用企业主要帮助企业和个人用户写营销文案和AI绘画，使用GenAI生成文本和图像内容，通过自然语言应用在各种渠道（如社交媒体、广告、博客文章、电子邮件、网站等）进行数字化营销推广，并集成了SEO功能。

该路线为企业提供了**快速试错和验证GenAI技术**的契机，能帮助企业形成先发优势和构建竞争壁垒，避免错失最佳布局时机。同时，**成本相对可控**（一般仅为API接口费用及实施费用，年均费用百万人民币左右），已成为当前各行业领军企业的首选。

路线二、自研行业大模型

在该路线下，根据企业所处行业、卡位、禀赋和战略的不同，又可细分为两条子路径：首先，企业可以选择**对通用大模型进行深度微调并自行开发应用层**。这适用于对数据敏感度高且对输出质量有较高要求的行业（如法律、医疗和金融），其优点在于可以学习和应用专有数据，以及进行私有化部署；劣势则包括若企业选择使用开源模型，整体效果可能不如封闭大模型，而且应用开发框架将受限于现有的训练和部署框架。例如，某国际领先投行就使用大量文档和历史记录对其自研大模型进行微调，以协助财富顾问的日常工作，以及支持投行内部财富管理业务的发展。

另外，企业也可以选择**自主开发行业大模型和端到端应用**。这种方法是从零开始研发特定行业模型，然后将前端应用层进行封装以实现商用。其优点在于可以完全掌握大模型的数据、算法、知识产权、能力、模型资产等；劣势则包括可能无法与市场上领先的封闭大模型竞争，而且初期投入大、回报周期长，需要企业高层具备较大决心和耐性。

由于大模型技术是GenAI赛道的核心价值所在，实力雄厚的“大厂”通常会选择自主研发行业大模型，或利用自身数据优势对大模型深度微调。然而，该路线对企业大模型技术栈、人才和技术能力，以及推动技术转型的决心要求较高且成本费用高昂。例如，构建一个百亿以上参数的行业大模型至少需要采购8~10块市占率第一的GPU厂商的数据中心平台级引擎硬件，成本约为**2,500万~3,500万元**，考虑到后期开发、GPU训练能耗及运维成本，**整体费用可高达5,000万~7,000万元**。因此，企业在选择该路线时应充分考量投入产出比和长期回报等指标，综合考量后再做决定。

结语

本章深入浅出地分析了企业规模化落地GenAI应用的流程，提出企业在引入GenAI时可遵循“四步走”方法论：首先快速梳理GenAI战略选项，构建初版GenAI战略；其次确定GenAI用例图谱和转型路径图，并贯彻“负责任的AI”原则；然后对优先用例进行试点，为规模化建立基础；最后，企业需要推进全方位的运营模式转型，以规模化推广GenAI。

在企业首次落地GenAI用例时，我们建议快速完成第1、2步，并基于后续POC试点快速迭代，通过“先动起来”的策略，不断打磨和优化适合企业自身的GenAI战略及用例图谱。中国企业高效推进GenAI部署和规模化应用的征程刚刚开始，衷心希望一众领先企业能够把握AI浪潮下的产业发展机遇，不断前行。

¹ 资料来源于微软
² 特别是第四条第一点 (https://www.gov.cn/zhengce/zhengceku/202307/content_6891752.htm)

曲向军是麦肯锡全球资深董事合伙人，中国区金融机构咨询业务负责人，常驻香港分公司；
韩峰是麦肯锡全球董事合伙人，常驻深圳分公司；
胡艺蓉是麦肯锡全球副董事合伙人，常驻上海分公司；
蒋子翔是麦肯锡咨询顾问，常驻上海分公司。

感谢Ray Zhou对此文的贡献。

麦肯锡公司2023年版权所有。



让GenAI应用真正为企业创造商业价值

企业可以从GenAI应用中获得至少三方面价值，一是减少手工作业并缩短冗长的业务流程，二是协助团队进行高级分析以及更加准确地提炼洞察，三是加速活动/产品的设计与交付过程。

曲向军, 韩峰, 胡艺蓉和李静瑶

一、三类企业级GenAI服务提供商

按照部署GenAI应用的五大要素（即算法和技术栈、数据与基础设施、战略路线图、人才与运营模式、风险与治理），企业级GenAI服务提供商可划分为三类不同的市场参与者，服务内容涵盖从技术架构到规模化实施等多个范畴（见图1）。我们看到，这三类服务商在GenAI价值链上扮演着不同角色，能够为企业带来差异化价值。

图1：企业级GenAI服务提供商可划分为三大类别

服务范围

专注程度

		GenAI技术提供商	系统集成服务商	价值守护者
技术架构	算法和技术栈	开发和拥有尖端的核心模型 (FM, LLM等)	根据客户端规范和功能要求部署第三方模型	结合技术和业务专业知识定制模型，保障实际业务价值
	数据与基础设施	与基础设施提供商密切合作，确保无缝集成	构建和维护模型和基础架构之间的集成	根据最适合具体情况的解决方案，建议应构建的技术生态系统
规模化推动因素	战略路线图			确保用例路线图、行业宏观趋势和战略目标之间的一致性
	AI/ML人才与运营模式			构建组织人才支撑能力并设计运营模式
	风险与治理	解决模型的偏见、公平性和透明度问题	通过整体方法识别风险并定义防范举措	

资料来源：麦肯锡分析

1. **基础设施及平台提供商**: 主要提供基础平台和通用模型, 在GenAI价值链上的云平台 and 基础模型环节较为常见。它们为企业提供GenAI的核心技术能力, 使企业开发的应用能够顺利调用最新的大型深度学习模型。
2. **系统集成服务商**: 根据客户需求搭建行业和细分领域模型, 以及企业技术架构与基础模型之间的链接, 主要存在于GenAI价值链上的云平台、模型库与工具环节。它们在企业大型科技项目上拥有丰富经验, 可提供基础架构集成大模型能力的综合解决方案。
3. **价值守护者**: 凭借以业务为导向的方法论和端到端的专业知识, 协助企业落地GenAI应用并确保新技术能为机构带来真正商业价值, 主要存在于GenAI价值链服务环节。

它们通过构建战略愿景、增强企业内部能力和管理相关风险, 确保GenAI应用规模化实施后能为企业带来预期中的商业价值。此外, 部分公司也可协助企业定制各种行业和细分领域模型, 并提供相应的专业建议和技术指导, 例如, 如何构建适配于企业发展愿景的技术生态系统等。

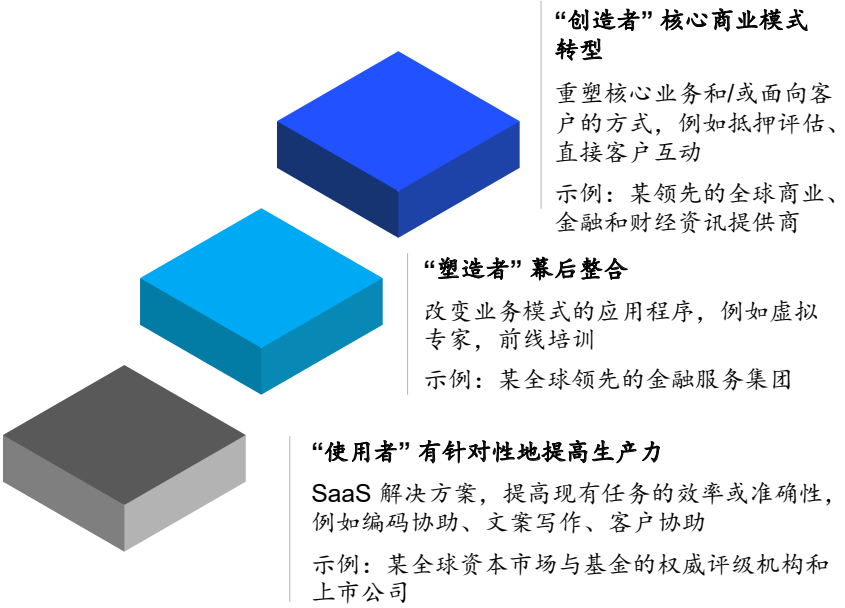


二、企业GenAI应用成熟度的三个阶段与相应解决方案

企业GenAI应用成熟度可分为三个阶段，处于初始阶段的“使用者”主要聚焦于内部效率提升，中级阶段的“塑造者”则着手优化展业模式，而处于高级阶段的“创造者”则开始重塑商业模式（见图2）。针对在GenAI应用成熟度上处于不同阶段的各类企业，三类企业级GenAI服务提供商能为客户提供不同支持，其中价值守护者可提供四类支持，覆盖从战略咨询到现状诊断、再到用例交付以及规模化推广各个环节的精准洞察、方案设计与能力建设。

图2：企业规模化部署GenAI应用的三个阶段

部分领先企业已成功跨越第三阶段，利用GenAI重塑商业模式



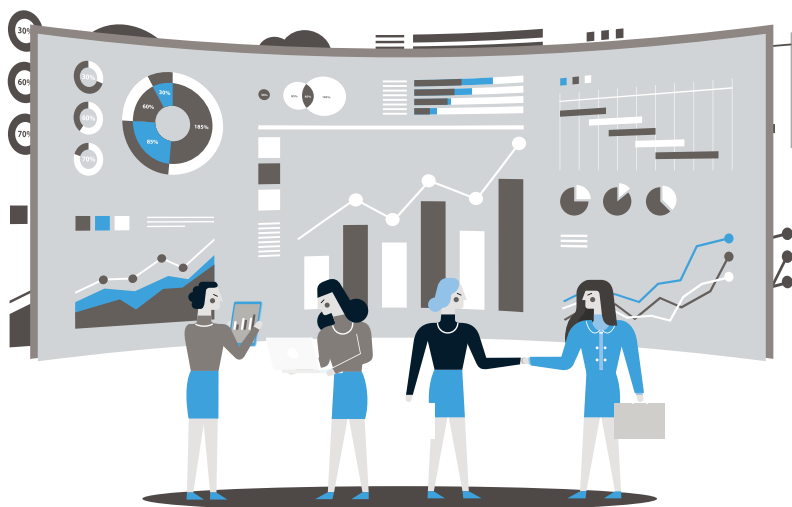
资料来源：South State Correspondent Division; Jasper AI; Bloomberg

基于企业GenAI应用成熟度的不同阶段，上述四类支持措施可进一步演化为三类具有针对性的举措，即知识导入、战略规划，以及共建与实施保障，为企业部署GenAI应用提供端到端保障并加速落地，确保新技术为机构带来真正的商业价值。

- **知识导入：**以大师课、专题工作坊等形式，为企业决策层演示GenAI大模型能力、业内领先实践案例及具体执行，帮助企业管理层加深认知，思考如何将新技术与公司实际情况相结合，广泛应用于各个业务领域和内部流程。

例如，某领先公司就成立了专门的GenAI和基础模型研究中心，通过持续投入优秀人才和相关资源，不断跟踪GenAI技术的最新发展趋势，协助企业量身定制不同阶段的GenAI战略。此外，一众海外领先公司还通过收购数据平台类服务商、与大型科技企业建立战略合作伙伴关系等方式将行业前沿技术为己所用，持续打造自身科技生态圈，以确保GenAI应用能真正为企业带来所需的商业价值，包括业务流程重塑、内部效率提升等。

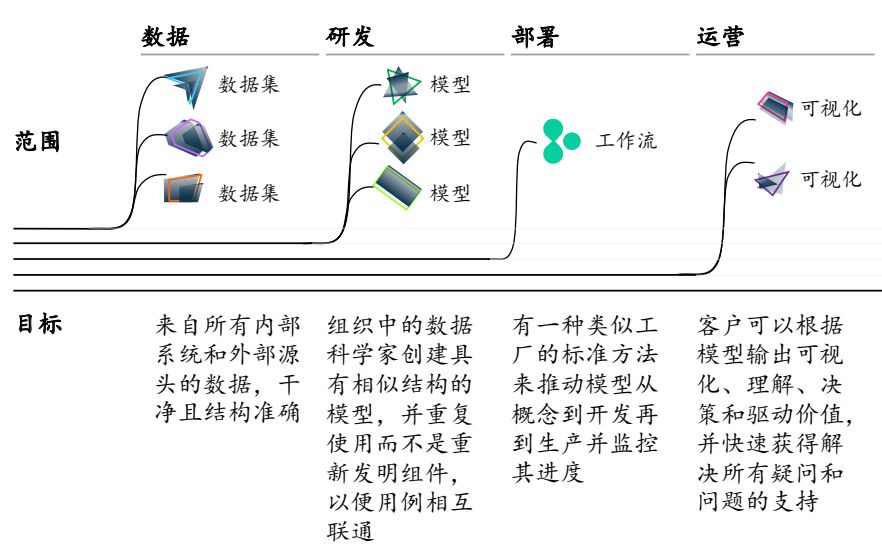
- **规划路径：**快速诊断企业部署GenAI的相关能力，形成整体实施战略及路线图。具体包括以下三个步骤：首先，结合企业现状识别GenAI赋能或重塑业务模式的价值创造机会点；其次，创建用例图谱并确定用例优先级；最后，制定具体的GenAI试点路径图。



- **共建与实施保障:** 企业可通过“四步走”，将GenAI技术优先应用于具有较高适用性的任务中，在用例顺利落地并产出价值（如业务重塑或降本增效）后再进行规模化推广。首先，企业应该明确关键评价指标（KPI/OKR），主要用于判断用例实施是否复合预期以及优化GenAI应用的具体设计方案。其次，着手设计合适的技术方案，探寻与大模型厂商的最佳合作模式，创建模型，用例归属部门设计GenAI应用的具体产品方案。然后以敏捷方式进行MVP测试，并持续迭代产品和技术方案。最后，规划在组织内实施用例的方法，优化现有运营模式和组织人才机制，并落实GenAI相关的新兴技术风险治理，以推动用例落地，取得实际应用成果。

部分领先公司已开发出规模化部署GenAI的产品组件，能够与企业现有技术栈实现良好兼容（见图3），既能作为行业分析产品的基座，也能为首次部署和驱动GenAI提供底层技术栈。该套组件可实现四大关键目标，包括清洗和整合内外部数据，搭建可扩展和重复使用的GenAI模型，工厂化的模型开发和监控方法，并以具备较高透明度的卓越性能辅助企业快速决策。

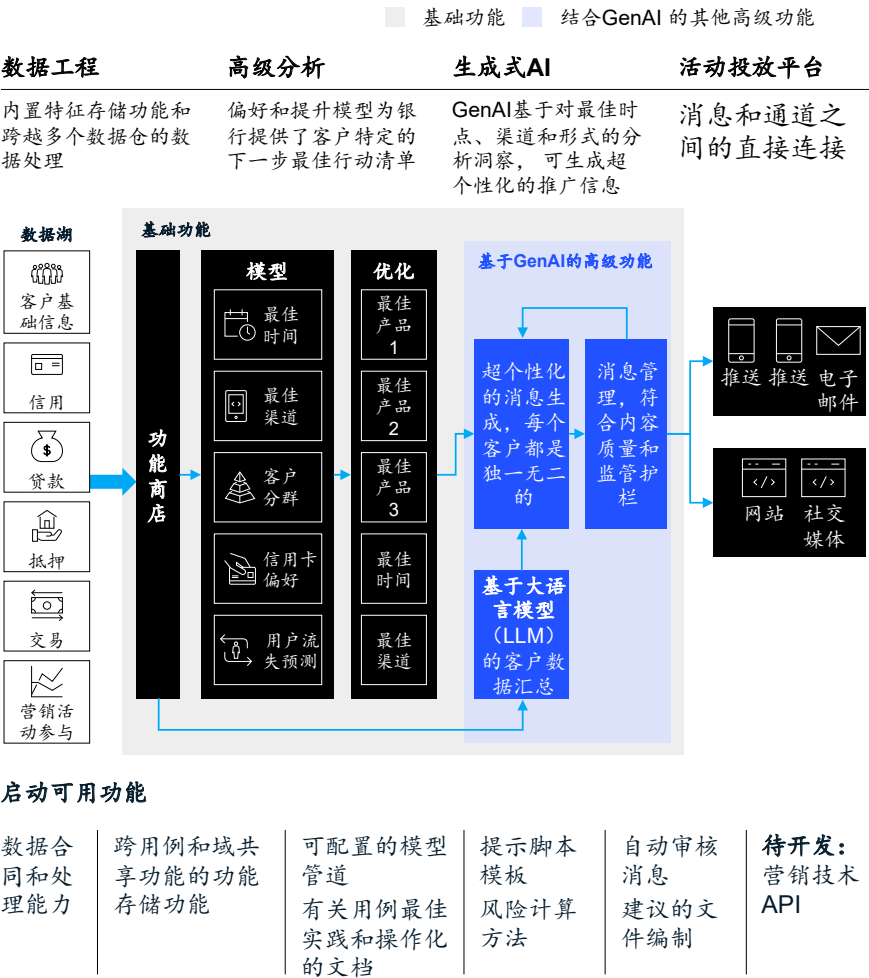
图3：企业规模化部署GenAI应用示意图



资料来源：麦肯锡分析

又如，某国际领先公司自主研发的内容生成AI，能够基于企业历史数据生成客户档案，自动分析和决策针对不同客户的活动投放内容，并生成个性化的推广信息（每个客户的推广信息均由高度相关的文本和图像组成），同时确保推广内容通过预先设定的风险管理机制评估，最后通过直连的投放平台完成一系列投放操作，将过去耗时数周的数字化营销活动缩短到几分钟，大幅提升了市场营销部门的工作效率（见图4）。

图4: GenAI协助企业高效完成数字化营销活动示意图





三、GenAI应用如何为企业带来真正商业价值

GenAI应用和技术可帮助企业大幅缩短冗长的业务流程、协助团队进行高效分析和提炼洞见，并加速活动/产品设计与交付过程，持续重塑企业现有商业模式。我们认为，其中较具代表性、能迅速为企业带来价值的GenAI应用包括以下几种：

1. **虚拟专家类型应用：**某全球商业银行利用GenAI技术，构建了用于提取ESG知识和信息的产品原型。过去，该银行一线客户经理需要为每个企业客户撰写和制作文本量极大的可持续发展报告，非常耗费人力和时间。在GenAI技术的帮助下，技术人员打造了一个基于GPT通用大模型的分析洞察提取引擎，只需简单四个步骤（包括数据源准备、提示词工程、推理处置和答案校验），就可从上百页的文档中提取信息，并自动填入每份可持续发展调查问卷，且预填答案正确率达到90%以上。银行客户经理只需对调查问卷进行简单复核便能轻松完成报告，节省了逾90%的工作量和时间。

2. **客户互动类应用:** 某大型通信公司为提升其呼叫中心的客户满意度, 利用GenAI技术对被标记为"优秀"和"欠佳"的通话记录进行分析, GenAI可自动拆解通话结构(例如开场语、客户问题识别等), 并分析电话客服是否在通话过程中展示出应有的专业能力和服务态度(如同理心, 特定客服词汇的使用等), 最后根据分析结果自动为客服人员创建所需培训(即针对被标识为能力欠佳的范畴进行专项培训)。在GenAI应用的帮助下, 该企业成功打造出一个可持续监测电话客服工作表现、不断优化服务水平的教练引擎。
3. **内容创作类应用:** 某消费品公司利用GenAI技术, 帮助设计师和工程师设计产品概念、制作原型、撰写产品故事和营销内容, 该GenAI应用还能根据消费者反馈, 自动对产品概念和营销内容进行微调, 此举大幅降低了企业的人力成本并缩短产品上市周期。在短短半年时间, 该公司便成功推出了一系列创新产品, 并利用GenAI创造了新奇有趣的用户体验, 受到当地消费者的追捧并迅速占领市场, 产品出口业务也呈现指数级增长。
4. **跨领域类应用:** 某知名生物医药企业计划在未来数年内, 对分布在不同地点的几十个药厂进行维修和翻新工程, 而每个药厂的翻新工程均需要近百名工程师参与, 并经历千余次的文档更新和流程验证。在GenAI技术的帮助下, 工程师们通过运用自动生成文档、流程验证以及多语种翻译等自动化功能, 大幅减少了流程中的错误/返工数量, 消除了因语言不通而引起的误解, 使项目管理工作变得更轻松和高效。截至目前, GenAI技术已帮助项目人员减省约30%的文档工作量, 释放超过500名工程师人力, 总周转时间减少约10%, 实现了明显的效率提升和质量优化。

概而言之，企业可以从GenAI应用中获得至少三方面价值，一是减少手工作业并缩短冗长的业务流程，二是协助团队进行高级分析以及更加准确地提炼洞察，三是加速活动/产品的设计与交付过程。

结语

展望未来，我们预计国内领先企业将通过规模化实施GenAI应用，快速提高自身生产力、整合现有业务模式和流程，并最终实现核心商业模式转型，达到既定战略目标。潮平两岸阔，风正一帆悬，当下全球人工智能技术处于高速发展阶段，是国内企业布局GenAI技术和实现规模化部署的重要机遇，祝愿一众领先企业能够借助新兴技术的东风，不断乘风破浪，创造更多商业价值。

曲向军是麦肯锡全球资深董事合伙人，中国区金融机构咨询业务负责人，常驻香港分公司；
韩峰是麦肯锡全球董事合伙人，常驻深圳分公司；
胡艺蓉是麦肯锡全球副董事合伙人，常驻上海分公司；
李静瑶是麦肯锡资深咨询顾问，常驻上海分公司。

麦肯锡公司2023年版权所有。

关于麦肯锡中国区金融咨询业务

麦肯锡中国区金融机构咨询业务服务于中国领先的银行、证券和信托公司、财富与资产管理公司、保险公司等金融机构，涉及转型与创新、互联网金融、全球化、战略、企业金融、销售与营销、运营、数字化与技术、风险与资本、领导力培养与文化转型等多个领域。我们在本地区共有约300多位专注于金融领域的咨询顾问，均在该领域拥有深厚的行业经验。与此同时，麦肯锡遍布全球的金融机构专业咨询顾问、研究人员及分析人员共同构成了庞大的资源体系，为我们中国本土的客户提供充分的支持与服务。

关于作者



曲向军
全球资深董事合伙人
香港分公司



韩峰
全球董事合伙人
深圳分公司



方浩翔
蓝跃领导人
香港分公司



方溪源
全球董事合伙人
香港分公司



容觉生
全球董事合伙人
香港分公司



胡艺蓉
全球副董事合伙人
上海分公司



曾文静
资深项目经理
上海分公司



李静瑶
资深咨询顾问
上海分公司



宋戈
资深咨询顾问
上海分公司



邱外山
资深咨询顾问
上海分公司



王喆宸
咨询顾问
上海分公司



蒋子翔
前咨询顾问

关于麦肯锡

麦肯锡是一家全球管理咨询公司，致力于帮助各类组织实现可持续且包容性增长。我们与私营、公共和社会部门的各类客户广泛合作，解决复杂问题，并为客户的所有利益相关方带来积极变化。我们将果敢的战略与变革性技术相结合，帮助组织实现更具可持续性的创新、持久的业绩改善，并打造立足当下、制胜未来的卓越员工团队。在中国，我们在北京、上海、深圳、香港、台北和成都开设了六家分公司，拥有超过一千名全球合伙人、咨询师和业务支持专家。了解更多信息，请访问 www.mckinsey.com.cn

上海

麦肯锡上海分公司
上海市湖滨路168号
企业天地3号楼21楼
邮编: 200021
电话: (86-21) 6385-8888
传真: (86-21) 6385-2000

北京

麦肯锡北京分公司
北京市朝阳区光华路1号
嘉里中心南楼19楼
邮编: 100020
电话: (86-10) 6561-3366
传真: (86-10) 8529-8038

香港

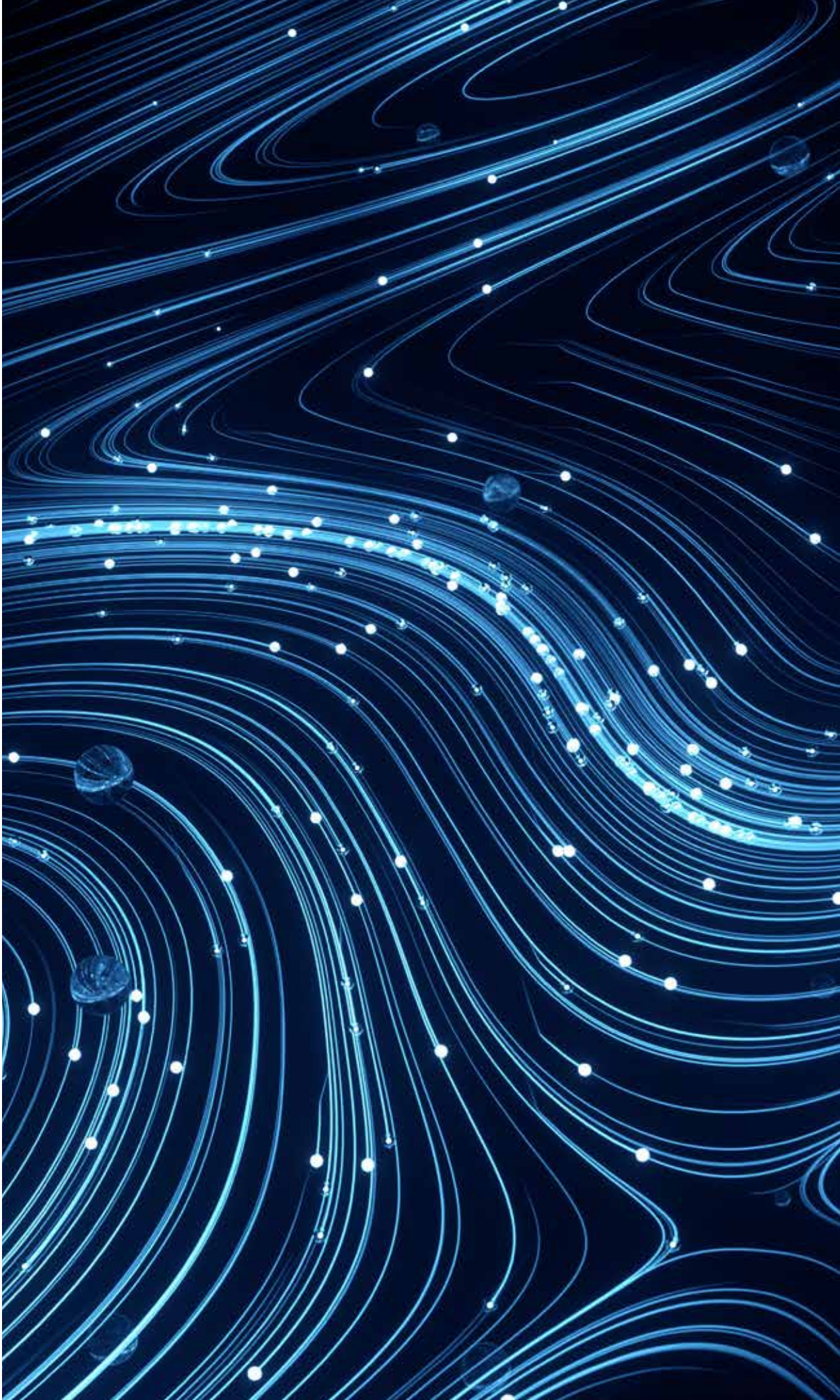
麦肯锡香港分公司
香港中环花园道3号
中国工商银行大厦40楼
电话: (852) 2868-1188
传真: (852) 2845-9985

台北

麦肯锡台北分公司
台北市信义路五段七号47楼110
电话: (886-2) 8758-6700
传真: (886-2) 8758-7700

深圳

麦肯锡深圳分公司
深圳市福田区中心四路
嘉里建设广场第三座13楼26室
邮编: 518000
电话: (86-755) 33973300



捕捉生成式AI新机遇
2023年秋季刊
麦肯锡公司版权所有©
麦肯锡中国区新媒体设计出品
McKinsey.com.cn

