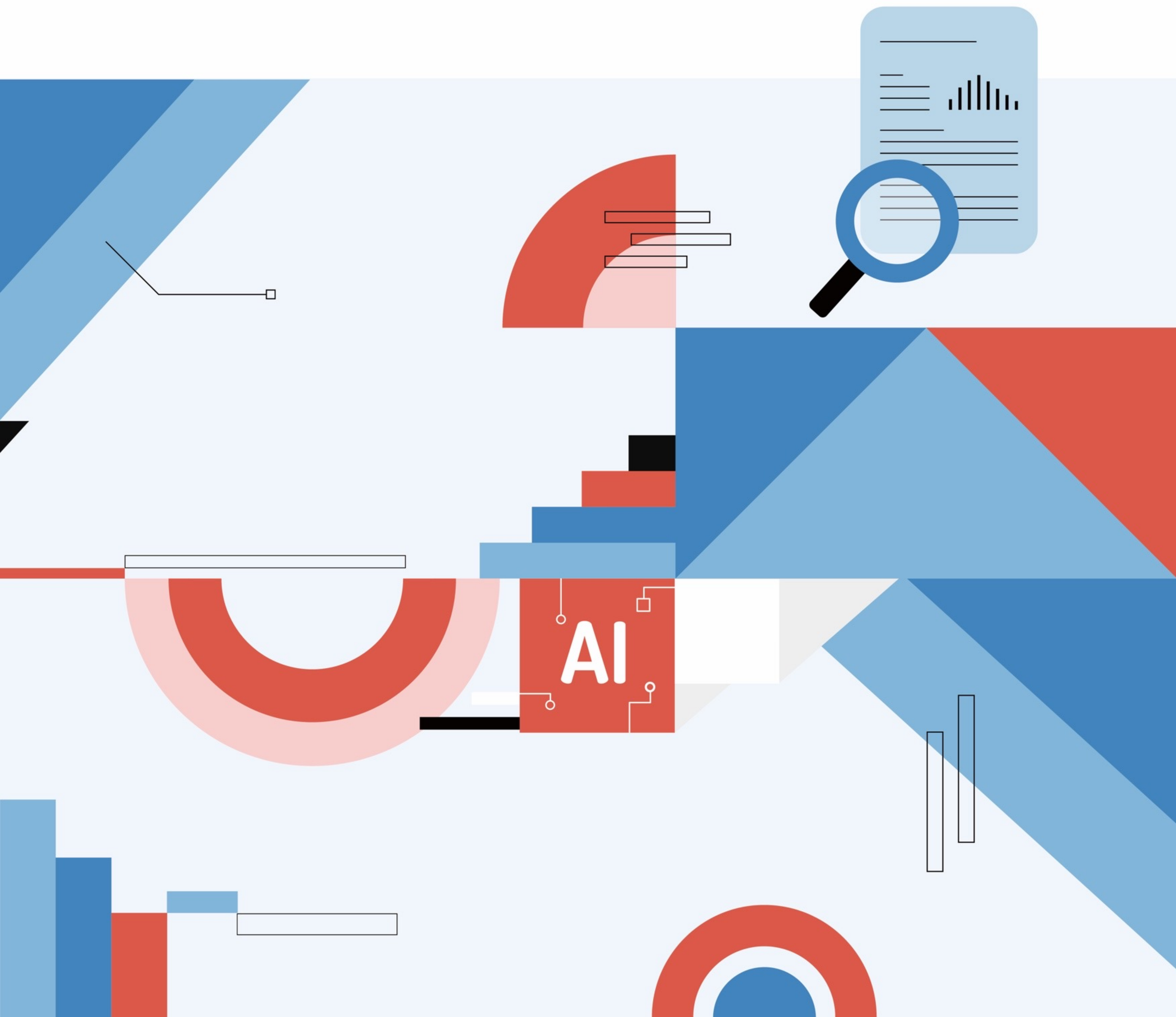




2024年金融业生成式 AI 应用报告



前言



毫无疑问，生成式人工智能是2023年全球最具影响力的创新科技，它代表着一种范式转变，从专门的技术追求演变为商业和金融领域的焦点。在全球范围内，我们看到越来越多的银行和保险机构已经采用并投资于生成式人工智能，尽管到目前为止几乎所有应用都是面向内部的。与此同时，还有大量的金融机构正在快速了解这项具有划时代意义的突破性科技。在此背景下，清华大学经济管理学院联合度小满科技（北京）有限公司、《麻省理工科技评论》中国共同开展了“2024年金融生成式人工智能应用”专题研究。

我们的研究关注了国内外170余家银行、保险、券商及资管等金融机构，并就生成式人工智能技术创新及其在金融行业应用的共性问题访谈了若干位国内金融机构高管人员和大模型技术开发主管，力求系统揭示金融生成式人工智能的创新密码、商业价值和政策期望，为国内金融机构有效采用这项突破性科技提供可靠的行动指南。

研究成果由四个部分组成。

第一，生成式人工智能是新质生产力。生成式人工智能是人工智能技术的集大成者，具有令人惊叹的类人原始创造能力，已被用于内容创作、文本提取及摘要生成、电子邮件生成、问答、翻译、客户支持、物体跟踪、图像生成与分类、音频生成与分类、代码开发等诸多场景。大模型技术大幅缩短了数据间发生连接与被计算的时间，极大提高了数据创造价值的效率，带来显著的降本增效，生产力水平是人工的4.4倍以上，并重新定义新服务、带来新制造、形成新业态，成为数字经济时代的新质生产力。

第二，生成式人工智能正在重塑全球金融业。金融业是单位数据产出量最高的行业，生成式人工智能将通过嵌入运营、业务转型、加速创新和差异化这四个关键方式重塑全球金融业，大幅提升内部运营效率，提高获客能力，提升产品设计创新能力，和加强风险防控能力。大模型驱动的新商业模式有望为金融业带来3万亿规模的增量商业价值。目前，生成式人工智能在我国银行业已经进入试点应用阶段，预计1-2年后会进入成熟应用期，降本增效方面会出现正面的财务成果，3-4年后将会得到规模化应用。在保险和证券业等其他金融机构中的商业应用会有所延迟，但整体上不会差距一年。全面进入生成式人工智能时代，高质量数据将是银行和其他金融机构的核心竞争力，这将会加速金融机构的组织变革与业务转型。

前言



第三，我国金融业生成式人工智能应用迎来重要政策红利期。金融业一直是产业数字化转型的重点行业，我国金融业具有世界上最大规模的实时数据，金融业与数字技术的结合是近年来我国金融科技创新的主要方向。中央金融会议明确提出做好科技金融、绿色金融、普惠金融、养老金融、数字金融五篇大文章，加快建设金融强国。随着“十四五”规划进入关键实施期，金融管理部门加快推动我国金融业数字化进程，鼓励持续关注金融领域新技术发展和应用情况，提升快速安全应用新技术的能力，释放数据价值。这些政策为金融业集成生成式人工智能提供了可预期、可操作的基本准则。

第四，2024年金融机构集成生成式人工智能需要遵循八项原则。一是正确认识生成式人工智能的作用域。二是大模型参数并非越大越好。三是技术先进性和业务可持续性选择大模型合作伙伴的两个关键标准。四是生成式人工智能的开发者实际上也是重要的监管者。五是着眼于端到端解决方案，而不是让它成为一个点解决方案。六是根据实际情况确认生成式人工智能应用优先级顺序。七是把大模型带到数据中心环境中运行，而不是把数据带到大模型所在的地方。八是由首席执行官直接牵头指导负责任的人工智能治理框架。

由于生成式人工智能技术创新与应用处于快速发展阶段，可谓日新月异，我们对生成式人工智能的变革作用及其在金融业落地应用的认识还有待进一步深化，研究报告存在不足之处，敬请大家批评指正。

摘要

本研究报告全面概述了生成式人工智能在金融领域的应用，并深入探讨了与其实实施相关的潜在收益和挑战。通过揭示生成式人工智能的变革能力，本报告旨在指导银行、保险公司等金融机构在2024年利用这项技术的力量来推动创新、提高效率，并在不断发展的行业中保持竞争力。



目录

- 引言——2023回顾 | 全球生成式AI一年重要事件盘点
- PART 1 创新解码 | 生成式AI是新质生产力
- PART 2 行业洞悉 | 生成式AI重塑全球金融业
- PART 3 案例解读 | 一些金融业生成式AI应用的成功案例
- PART 4 政策期望 | 我国金融业生成式AI应用迎来政策红利期
- PART 5 行动指南 | 金融业集成生成式AI需遵循的八项原则

联合编写团队

清华大学经济管理学院
度小满科技（北京）有限公司
《麻省理工科技评论》中国
清华大学经济管理学院动态竞争与创新战略研究中心

致谢 **感谢以下专家对本次研究的支持，不分先后顺序*

- 阮琦 中国人寿副总裁、首席风险官
- 龚伟华 北京银行首席信息官
- 孙科伟 工银科技技术总监
- 刘岩 泰康保险集团人工智能研究院院长
- 徐磊 中金研究院副总经理
- 罗引 中科闻歌CEO
- 贺立帅 宝马汽车金融创新事务负责人
- 陈伊凡 中欧商学院研究员
- 孔庆超 中国科学院自动化研究所副研究员
- 肖然 Thoughtworks中国区总经理



引言——2023年回顾

过去的一年，生成式人工智能（AI）受到前所未有的关注。主要经济体的投资者、政策制定者和社会各界几乎全年都在谈论ChatGPT、关注大模型，所有互联网、数字经济、IT技术、全球双边和多边组织的会议都专门设立了生成式AI的分议题，有的甚至是全会主题，比如旧金山人工智能峰会、英国人工智能安全峰会。基于AI的科学研究（AI for science）以及关于生成式AI的理论研究成了国家科学基金的重点资助对象。

时间	事件概述	事件影响
2023年1月	ChatGPT的月活用户（MAU）突破1亿，成为有史以来达到1亿MAU用时最短的消费应用	生成式AI正式进入公众生活
2023年3月15日	OpenAI发布GPT-4大模型	全球首个商用多模态产品
2023年3月16日	百度发布“文心一言”	中国首个商用生成式AI大模型，开启了中国的“百模大战”
2023年3月29日	马斯克等千名科技人士发公开信，呼吁暂停训练比GPT-4更强大的AI大模型	将公众注意力更多转移至生成式人工智能治理
2023年4月28日	中共中央政治局召开会议，会议指出“要重视通用人工智能发展，营造创新生态，重视防范风险”	通用人工智能发展的全球首个政府行动
2023年5月	英伟达股价翻番，市值超过1万亿美元，晋升成为全球第六大科技公司	GPU成为生成式AI时代的“硬通货”，社会资本更多关注算力基础设施
2023年9月13日	OpenAI、特斯拉、英伟达、苹果、Google、微软等美国AI巨头齐聚美国国会山，闭门商讨AI发展与安全	负责任的人工智能治理框架逐渐成为共识
2023年9月26日	微软把智能副驾(Microsoft Copilot)的力量和全新人工智能体验带入Windows PC，并融入到画图、照片、Clipchamp等应用程序中	开启生成式AI桌面操作系统新势力，将彻底改变日常工作方式
2023年11月7日	OpenAI举行了首个开发者大会，发布GPT-4 Turbo, 上线GPT应用商店，用户将能够自定义GPT	生成式AI商业生态系统逐渐形成
2023年11月	OpenAI内部上演了连续反转的董事会罢免CEO事件。11月17日，首席执行官山姆·奥特曼（Sam Altman）被董事会赶出公司，12天后，OpenAI又发文称，山姆·奥特曼已正式恢复担任公司CEO一职，同时OpenAI的董事会迎来重组，微软作为“无投票权观察员”加入董事会	创新发展在与安全担忧的较量中占据上风

引言——2023年回顾

我们盘点了过去一年，全球范围内与生成式AI相关的十件重要影响事件，其中备受瞩目的当属引领通用人工智能（AGI）研究的OpenAI及其推出的ChatGPT。据美国蒙大拿大学的一项最新研究显示，在创造性思维测试方面，ChatGPT已经可以与排名前1%的人类思想家相媲美。这是一项开创性的技术，它将把人类的认知智力能力提升到前所未有的高度，远远超过工业革命时期的劳动机械化。通过ChatGPT，社会大众开始切实感受到人工智能和第四次工业革命将如何影响人们生活的方方面面，也产生了安全和伦理道德的担忧。

尽管存在不安，一些大学和华尔街的银行家们在2023年上半年甚至禁止员工或学生使用ChatGPT，但凭借在复杂场景下的可用性和快速响应，ChatGPT获取新用户的速度比迄今为止创建的任何最受欢迎的消费者应用程序都要快。在发布不到五天，ChatGPT的用户就突破了一百万；不到两月，用户月活数就破亿；不到一年，就收获了全球17亿访问用户，这使其成为所有行业都必须准备考虑的工具。英国媒体监管机构Ofcom在2023年11月发布的调查结果显示，近80%的英国青少年已经使用了生成式AI工具和服务，即使在7-12岁的年轻群体中，大约40%的人也使用了生成式AI的能力，这有力地证明了尖端技术已迅速融入当年轻人的生活，凸显了生成式AI对新一代互联网用户的变革性影响。世界各地的公司都在竞相使用ChatGPT来创造新产品并挖掘专有数据以获得财务优势。在禁令发布不到半年的时间，英国的一些大学以及华尔街的银行家们解除了先前的禁令，也陆续开始在工作中集成这项变革性技术，旨在提高生产力与保持竞争优势。

伴随着ChatGPT的亮相，生成式AI大模型浪潮席卷全球。各大资本和科技巨头纷纷入场布局大模型，硅谷和中关村的AI创业市场火爆，基础大模型迅速发展，截至2023年10月初，我国生成式AI大模型数量超过230个，占据全球各国新开发大模型的一半以上，仅北京地区就高达115个，其中垂直类大模型103个，通用大模型12个。百度推出的“文心大模型”在最新发布的国际权威测评中荣获总分第一、算法模型第一、行业覆盖第一的佳绩，已经广泛应用于互联网、金融、能源、制造、传媒、教育等多个行业，体现了其强大的基础技术深度和产业应用覆盖广度。阿里发布的“通义千问”大模型，以赋能淘宝、钉钉、阿里云、飞猪等众多阿里应用为目标，形成了覆盖电子商务、办公、云服务、旅行等多场景的应用生态。华为推出的“盘古”系列大模型，其应用场景涵盖游戏开发、影视制作、电商营销、工业生产、设备巡检、智慧城市、气象服务、药物设计等领域。

此外，网易、360、智谱、抖音、腾讯、商汤、科大讯飞、京东、中科闻歌等数字技术公司也相继推出通用基础大模型。度小满、蚂蚁集团、恒生电子、众安科技、同花顺、马上消费等金融科技公司开发了一系列具有金融专业知识的金融大模型，积极探索垂直行业应用场景。

引言——2023年回顾

发布国内首个超大规模智能模型系统（悟道1.0）的北京智源人工智能研究院（以下简称“智源”）在2023年则与多家企业、高校和科研机构倾力打造一站式、高质量的FlagOpen（飞智）大模型开源开放软件体系，包括大模型算法、模型、数据、工具、评测等重要组成部分，支撑AI大模型软硬件生态的百花齐放，旨在建设大模型领域的“Linux”。微软总裁布拉德·史密斯评价智源为“处于绝对领先的三家全球机构之一”，与Google、OpenAI和微软齐名，充分展示了我国在生成式AI领域的强大实力。

在中东，2023年3月阿布扎比技术创新研究所发布中东首个400亿参数的基础大语言模型，并于5月15日宣布开源Falcon 40B，供研究和商业使用。基于斯坦福大学的HELM LLM（大语言模型）基准测试工具，Falcon 40B在使用显著更少的训练计算能力方面优于其著名的同类产品，该工具仅占DeepMind的Chinchilla AI的40%、OpenAI的GPT-3训练计算的75%、GooglePaLM-62B训练计算的80%。在开源后的一段时间，Falcon 40B冲上了Hugging Face开源LLM榜的第一位。同一团队还在去年推出了世界上最大的阿拉伯语自然语言大模型NOOR，并有望很快开发和发布Falcon 180B，展示了大模型的卓越性能以及阿联酋人工智能的技术进步，强化了阿联酋作为全球AI领导者的角色。

二级市场异常火爆，英伟达股价翻番，市值超过万亿美元，微软股价全年也上涨了近三分之一。生成式AI初创公司是最受风险资本青睐的投资对象，仅上半年就完成86笔交易，股权融资额超过141亿美元，比2022年全年增长了5倍以上，生成式AI热也带动70%的风险资金都投向人工智能领域，在低迷的全球资本市场一枝独秀。

生成式人工智能是大型科技公司相互角逐的新战场，亚马逊、英伟达、谷歌、微软、Meta都各自拥有2家及以上生成式AI公司。而且在2023年，这些大型科技公司支持的每一笔顶级交易，都有知名风险机构和企业的跟投。2023年，OpenAI获得微软和老虎全球资本、红衫资本等知名投资机构103亿美元股权融资，在ChatGPT发布一周年后，这家引领全球AGI研究的独角兽估值已接近900亿美元。亚马逊和谷歌向大语言模型开发商Anthropic投入数十亿美元，试图与OpenAI支持者微软展开竞争，争夺生成式AI基础设施的主导地位。

在国内，来自“清华系”的生成式AI企业获得了风险资本的大力支持。智谱清言全年获得社保基金、阿里、腾讯、高瓴、美团等科技巨头和知名机构超25亿人民币的融资，这家成立4年多的企业估值已经超过百亿元，是国内当前估值最高的AI大模型独角兽，也是全球11家新晋独角兽成员（其余十家分别是：ANTHROPIC（美国）、Cohere（英国）、Runway（美国）、A121labs（以色列）、Replit（美国）、ADEPT（美国）、Character.AI（美国）、synthesia（英国）、Typeface（美国）、Imbue（美国））。



引言——2023年回顾

2023年4月才成立的百川智能，半年后就获得阿里、腾讯、小米等科技公司及深创投等多家顶级投资机构3亿美元的A1轮战略融资，估值超过10亿美元。创立于2023年3月的“月之暗面”（Moonshot AI），也在10月获得来自红杉资本、今日资本、砺思资本等投资机构近20亿元的投资。

过去的一年，从底层大模型到算力基础设施到机器学习运营，再到消费端应用，生成式AI的生态已经初步建立，并且将在2024年随着大模型的进一步训练和稳定，会加速向下游行业发展移动，会对经济社会产生深刻影响。来自麦肯锡的全球调查数据揭示，28%的企业将生成式AI视为首要任务，这项变革性技术影响着全球董事会的议程——从营销和销售到产品和服务，预计每年在63个不同的行业应用中产生2.6–4.4万亿美元的惊人商业价值。

2023年是大炼模型的一年，开启了新一代人工智能创新发展的新征程。由于大模型开发的高成本支出，我们认为，开拓落地应用场景，不断提高客户采用率和大模型的利用率，更好发挥大模型的潜在价值，将会是2024年生成式人工智能突破发展的重要内容。

金融大模型产业链图谱

应用

银行

保险

券商

资管

.....

金融行业大模型

度小满

蚂蚁金服
ANT FINANCIAL

HUND SUN

HUAWEI

WENG 中科闻歌
WENGE GROUP.COM

众安科技

马上消费
WWW.MSXF.COM

GTCOM

Cornucopia

PIXIU

开发平台

阿里云
aliyun.com

Tencent 腾讯

HUAWEI

Baidu 百度

Google

北京智源
3A AI

Meta

ByteDance
字节跳动

之江实验室
ZHEJIANG LAB

计算机视觉&自然语言处理

Google

Microsoft

HUAWEI

DeepMind

美团

Meta

京东

inspur 浪潮

北京大学

智谱·AI

云从科技
CLOUDWALK

商汤
sensetime

追一科技

清华大学
Tsinghua University

NVIDIA

Baidu 百度

OpenAI

澜舟科技
langboat

WENG 中科闻歌
WENGE GROUP.COM

Tencent 腾讯

多模态

阿里云
aliyun.com

HUAWEI

科大讯飞
iFLYTEK

商汤
sensetime

Baidu 百度

OpenAI

KUNLUN
www.kunlun.com

Paradigm
第四范式

智谱·AI

Microsoft

LINKER 联汇

云从科技
CLOUDWALK

算力设施

亚马逊云科技

inspur 浪潮

Google

Microsoft

HUAWEI

Baidu 百度

阿里云
aliyun.com

Tencent 腾讯

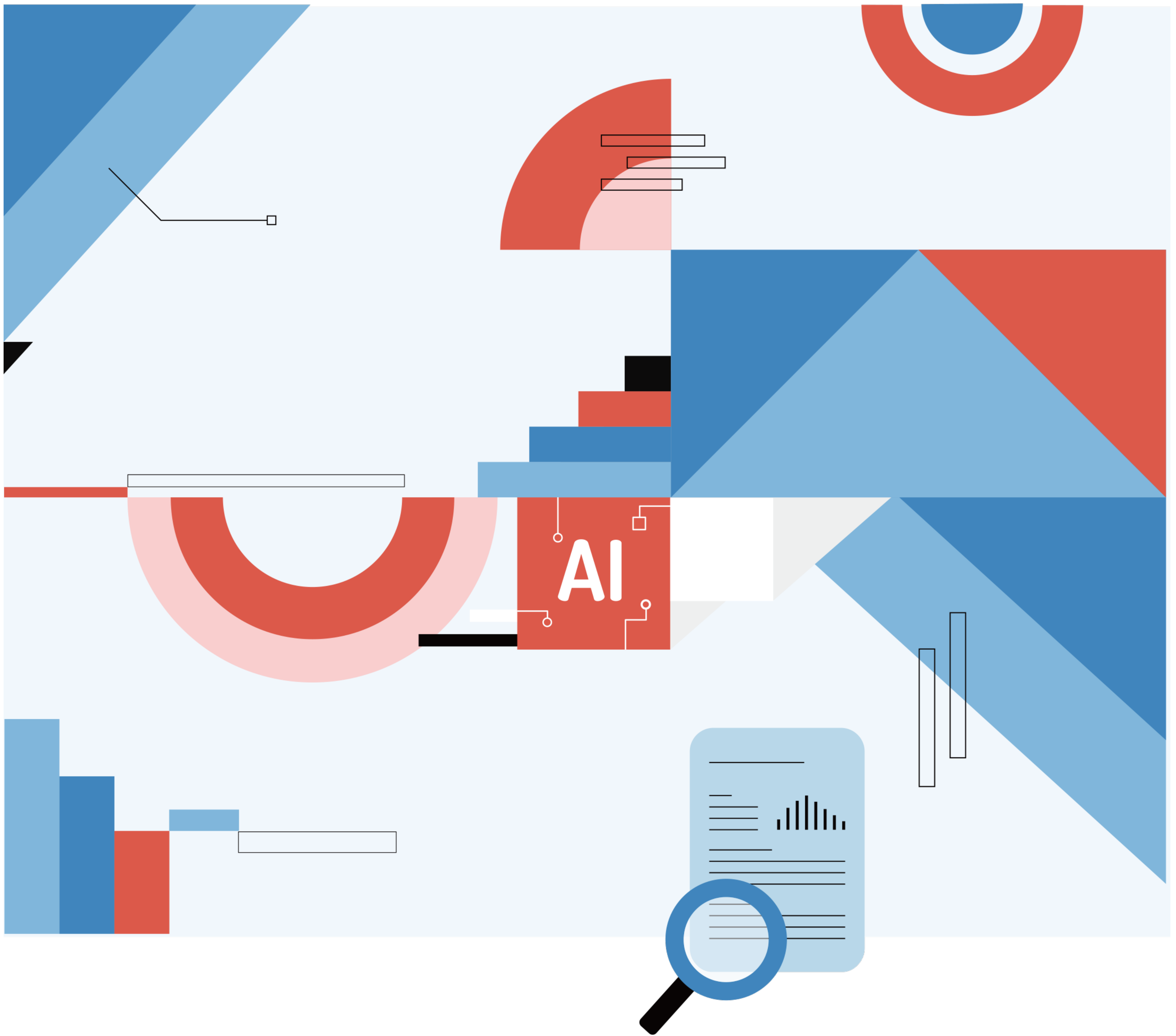
寒武纪
Cambricon

中科曙光
Sugon

商汤
sensetime

intel

NVIDIA



PART 1

创新解码 | 生成式AI是新质生产力

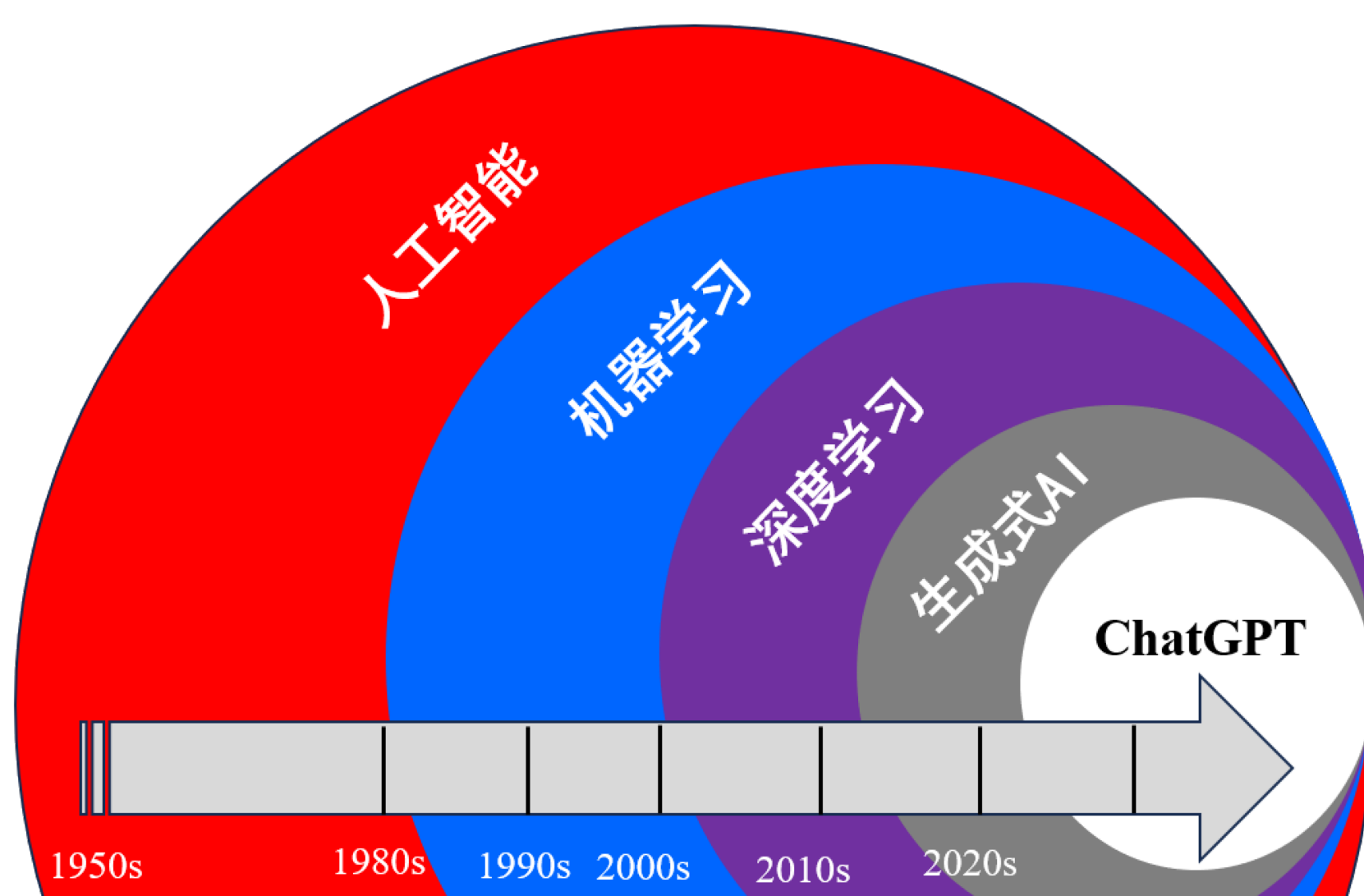
“人工智能是新电力，我们正处于一个决定性的时刻，就像上世纪90年代的互联网一样，人工智能正在走向大规模采用。” 美国银行的策略师在一份报告中写道，“ChatGPT等人工智能生成技术激发了人工智能的iPhone时刻”。



生成式AI技术创新历程

生成式人工智能是可以生成新内容（文本、代码、图像、音频等）的人工智能。它是机器学习的一个子集，由模型组成，这些模型可以根据现有数据（例如文本、音频文件或图像）和提示创建新内容（包括生成新数据）作为输出。大语言模型是生成式AI的一种，涉及语言/文本的生成（例如ChatGPT）。此外，生成式AI还涉及视觉输出（例如Sunthesia）、音频（例如Speechify）、图像（例如Midjourney）、代码的生成（例如GitHub Copilot）等其他类型的内容生成。

生成式人工智能的发展可追溯到20世纪50年代，当时计算机科学家们开始探讨如何让计算机具备类似人类的智能。随着硬件技术的进步和大数据时代的到来，生成式人工智能得到了迅猛发展。从最早的符号主义智能，到基于规则的专家系统，再到现在的机器学习、深度学习、强化学习等技术，生成式人工智能已经取得了显著的成果，产生了自主学习与原始创造能力。



生成式AI的产业创新机理

生成式人工智能进入大数据、高算力、强算法一体化集成创新的新阶段，尽管工程发展领先于理论探源，生成式大模型的“黑箱”机制有待解开，但产业应用上表现出较为清晰的创新机理，遵循从数据飞轮的学习效应，到成长飞轮的网络效应，再到成本飞轮的规模效应的产业创新逻辑。

具体地说，从获取更多用户开始，用户贡献更多数据，对大模型机理和产业创新的认识更深入，继而优化价值主张，吸收更多用户和合作伙伴，摊薄固定成本，降低单位成本，迭代循环，不断释放生成式人工智能的价值潜力。



生成式人工智能提高劳动生产力水平的四种途径

第一，增强数据间的连接，缩短海量数据被计算的时间，提高资源配置效率。在生产过程方面，生成式人工智能可以通过对大量数据的学习和处理，实现对生产流程的优化。这使得企业在生产过程中能够减少浪费、降低成本、提高生产效率，从而达到提升生产力的目的。此外，生成式AI还可以自动执行重复性任务、改进文档处理、增强决策洞察力、简化沟通效率、确定战略举措的优先顺序，辅助企业进行决策，为其提供数据支持，使企业决策更加科学、精准，进而提高生产力。

第二，大模型强大的专业知识储备，促进企业自主创新能力。通过对大量数据的学习和挖掘，加速构思过程、改进需求管理、增强代码生成、优化知识管理、改善团队协作，生成式人工智能可以为中小企业提供新的创新思路和解决方案。特别是生成式AI还可以辅助科研人员进行研究，缩短研究周期，提高研究质量，从而推动科技创新。

第三，增强产业链协同，促进产业链的升级与变革。生成式AI的应用使得企业可以自动执行重复性任务、改进需求预测/库存管理、优化路线和调度，实现产业链的自动化、智能化，提高产业链的协同效应，推动新兴产业的发展，为经济增长注入新动力。

第四，提高劳动力素质，提升企业整体创新效能。一方面是通过实施ChatGPT增强支持知识库、简化工单分类、减少响应时间、并识别有风险的客户，赋能企业员工自动化潜在客户资格认证、加强内容创作、个性化客户服务、简化CRM系统更新。另一方面，AI助理强大的专业能力，能取代部分任务，会倒逼员工主动提升能力，形成比AI助理更强大、更多样化、差异化的能力，进而提升企业的整体服务水平和创新效能。

生成式人工智能重新定义客户体验，正在形成新服务

优质的客户服务通常取决于服务代理的专业知识和可用性。生成式人工智能“副驾驶”可以充当有价值的助手，提出响应建议并迅速满足客户需求。

第一，生成式人工智能可以被用来深入研究用户的行为、偏好和需求，并在用户需要的时间和地点准确地提供内容。例如，通过使用生成式AI工具Phrasee，快餐连锁店Domino为其英国和爱尔兰受众优化电子邮件内容。这一策略不仅提高了客户忠诚度和参与度，而且数字营销活动的平均点击率显著提高了57%，平均打开率激增26%，总体上推动了收入的大幅增长。



生成式人工智能重新定义客户体验，正在形成新服务

第二，聊天机器人能够进行语义理解，并在对话中构建上下文语境方面的复杂性。生成式AI的这些能力有可能消除以前提供平淡且不准确的响应缺陷，从而增强客户的参与度和信任度。例如，捷蓝航空与生成式AI服务商ASAPP合作集成生成式人工智能解决方案，实现聊天渠道自动化，并增强联络中心的客户服务。平均每次对话节省了280秒，在2023年第一季度节省了73000个工作时间，劳动力水平提高4.8倍，使工作人员能够迅速解决复杂的客户问题并最大限度地减少客户等待时间。

第三，生成式人工智能开启了内容创作的新领域，涵盖图像、语音、文本和视频生成。虽然它不能取代人类的创造力，但它可以节省营销人员的一些宝贵时间，使他们能够将精力投入到开发出色的营销活动和内容上。例如，Canva新推出的人工智能Magic Studio提供了一套生成式人工智能工具，使营销人员能够在短短几秒钟内创建多种形式的内容，可以为客户提供更多的体验，也可以增加了营销人员与客户互动的的时间。

生成式AI应用前景广泛，正在形成新业态

生成式人工智能的应用领域涵盖了各行各业。在图像领域，生成式AI可以用于图像增强、图像生成和图像编辑等任务，如风格迁移、人脸合成等。在语音领域，生成式AI技术为语音合成、语音转换和语音识别等任务提供了新的解决方案。在文本领域，生成式AI在自然语言处理、机器翻译、文本生成等方面取得了显著的成果。

此外，生成式人工智能还在游戏、影视、广告等领域发挥着重要作用，如游戏角色的自动生成、影视作品的剪辑合成等。商业信息技术服务商Intercom使用ChatGPT和GPT-4开发了一款人工智能客服机器人Fin，该机器人能够为客户提供安全准确的答案，进行简短的对话并提出问题，即刻解决50%的客户问题。Fin可与Intercom的收件箱、票务、书信、报告等配合使用，开箱即用，告别分散的支持解决方案，迎接AI增强的一体化客户服务平台。通过集成GPT-4工具，冰岛政府正在与OpenAI合作，共同保护其处于危险中的冰岛语及稀有文化资源，为推动人类文明多样性、促进非物质文化遗产的传承发展开拓新的思路。随着技术的不断发展，生成式AI的应用边界将继续拓展，为人类创造更多可能性，产生更多新经济业态。



与RPA等技术协同创新，生成式AI正在形成新制造

第一，精简供应链管理。AI大模型可以分析来自各种来源的大量文本数据，供应商通信、合同和运输文件，为优化供应链管理提供可操作的见解。例如，使用GPT-4自动执行跟踪库存水平、识别潜在瓶颈和预测交货延迟等任务，提高制造公司的效率并降低成本。**第二，加强产品开发。**通过分析客户反馈、市场趋势和竞争对手数据，GPT-4能够为产品开发过程提供信息。通过利用未满足的客户需求和GPT-4确定的新趋势，制造商可以更好地创造创新产品，以满足不断变化的消费者偏好。**第三，提高客户参与度。**通过对客户询问生成个性化的、类似人类的响应，GPT-4显著提高客户参与度。制造商可以使用GPT-4驱动的聊天机器人来处理客户支持，提供订单状态的实时更新，并根据客户的个人偏好提供产品推荐。

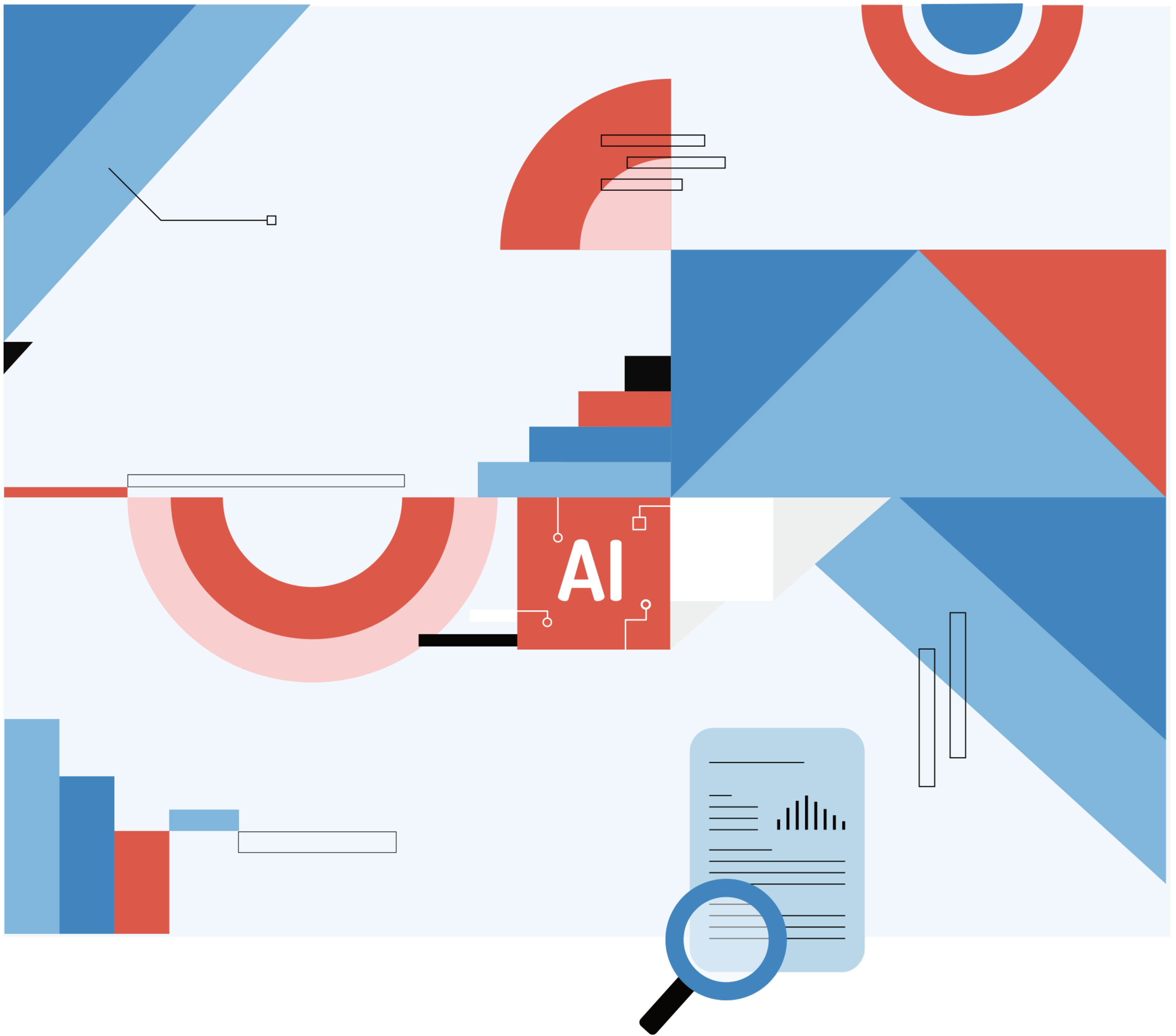
已有的一些工业制造案例显示，嵌入生成式AI工具会使企业生产力水平提高14%-25%，而如果将生成式AI与RPA等其他传统AI自动化工具结合使用，劳动生产力水平能提高到56%。随着成熟应用，预计2-3年后，有望使一些大型数字化制造企业的工资成本降低50%-60%，总的成本收入比可以降低10%-20%。因此，制造业可以结合生成式AI和RPA等其他工具，通过识别趋势和模式，并自动触发行动来响应意外事件（甚至干预预期问题发生），协同增强其车间效率，更进一步，支持生成式AI的聊天机器人可以改变工厂维护等领域的数据分析和信息记录检索流程。例如，生成式AI可以推进预测维护平台，为监督工厂流程可靠性的工程师提供帮助，促进产品开发和客户参与度，形成新制造方式和商业模式，最终提高效率和竞争力。

全球领先的工业技术服务商西门子与微软的技术团队合作，将西门子用于产品生命周期管理(PLM)的Teamcenter®软件与Microsoft 协作平台Teams以及Azure OpenAI服务中的大语言模型以及其他Azure AI功能集成，开发了最新的工业数字化技术，帮助工业公司在产品的设计、工程、制造和运营生命周期中推动创新和效率，打造新工业经济。通过生成式AI驱动的软件开发、问题报告和视觉质量检查来简化工作流程、克服孤岛并以更具包容性的方式进行协作，增强工厂自动化和运营,帮助工业公司在产品的设计、工程、制造和运营生命周期中提升效率，从而加速以客户为中心的创新。

小结



ChatGPT的出现，使得人工智能不再只是大型企业创新发展的“专利”，中小企业和社会大众拥有先进的人工智能正在成为现实。并且随着商业应用的拓展，生成式AI正在加速定义形成新的客户服务、新的制造方式以及新的经济业态，成为数字经济时代的新质生产力。



PART 2

行业洞悉 | 生成式AI重塑全球金融业

创新的银行将从“采用技术的银行”迈向“最懂技术的银行”，并向“最懂银行的先进技术服务商”领域拓展，充分释放其高质量数据价值。

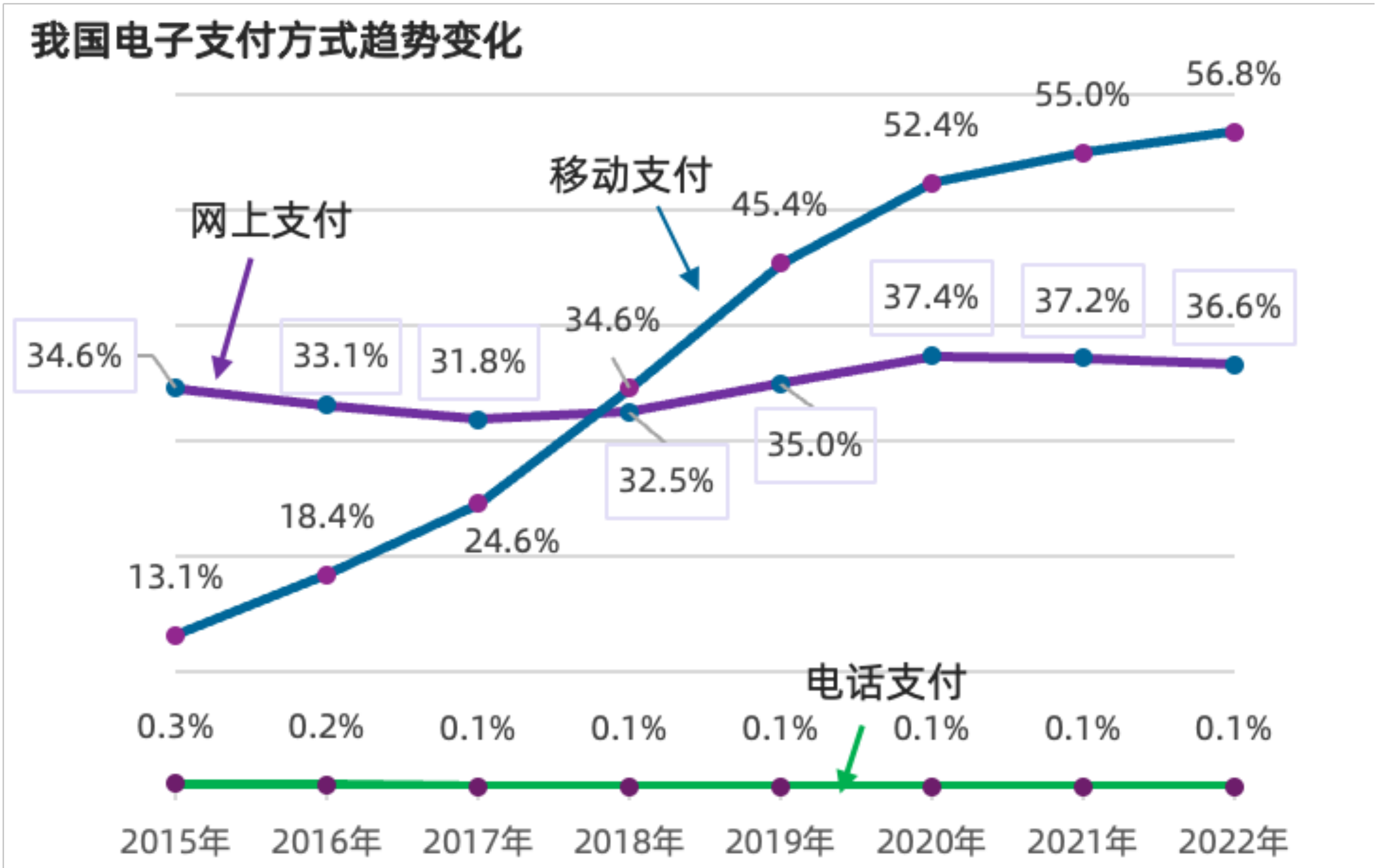


生成式AI重塑全球金融业

30年前，当互联网刚刚兴起时，大多数银行业从业者都无法相信支付方式将发生彻底变革，购物将无需现金和刷卡，交易越多银行内部的排队现象却越少；大多数的证券业从业人员也都无法想象交易所将不再是股民进行证券交易的场所，证券开户甚至都无需到证券公司现场办理。基于Windows/Linux操作系统的普通计算机和基于Android/iOS系统的智能手机/平板电脑就是数字时代的“交易所”。

过去15年，许多传统大银行一直跟踪并采用最新的数字技术，但是一些银行却在坚持自认为非常成功的“当铺”模式，坚持过时的商业模式已经或正在让这些银行付出代价，比如破产重组或利润率至少下降25%。因此，在不牺牲核心原则的情况下拥抱生成式人工智能、进行数智化转型可能是最好的选择。

在快节奏的金融领域，保持技术的领先地位对于蓬勃发展至关重要。多年来，我们见证了从互联网银行的诞生到算法交易的兴起的跨越式变革。在此影响下，我国银行业电子支付方式发生了极其显著的变革，移动支付在2018年超越网上支付，2020年，半数以上的电子交易用户选择移动支付，成为用户首选的业务支付方式。





生成式AI重塑全球金融业

当前，凭借无与伦比的计算能力和先进的数据分析能力，生成式AI正在席卷金融业，释放新的生产力，以前所未有的方式重塑全球金融业的格局，有望给金融业带来3万亿规模的增量商业价值，并可能将彻底改变交易的进行、投资的管理和风险的评估方式。泰康人寿人工智能研究院院长刘岩对此认为，“生成式AI将加速推动‘金融+AI’向‘AI+金融’的转型升级”。比如一个充满想象的应用场景，利用ChatGPT快速处理和审查大量金融数据，包括实时市场趋势、客户偏好、历史表现和宏观经济需求。从海量信息中获得的见解便可以使金融机构能够做出明智的决策，并提供符合客户独特需求和愿望的定制服务。

正如贝莱德首席运营官Rob Goldstein所言，“生成式AI将改变人们与技术互动的方式，提高我们的生产力并加强我们已经在做的伟大工作，还可能改变客户对我们互动的频率、及时性和简单性的期望。”生成式AI在银行的嵌入会对银行产生深刻影响，正在重塑全球金融业。

在零售银行和财富业务上，大型金融机构每天开设的新账户数量巨大，需要更有效、高效的了解你的客户（KYC）流程。生成式AI可以创建合成数据来帮助训练KYC背后的机器学习算法。它还可以帮助为虚拟助手创建更准确的自然语言模型。

在中小企业银行业务上，除了支持更复杂的虚拟助理之外，生成式AI还将帮助解释包含非数字数据（例如商业计划）的小企业贷款申请。

在商业银行业务领域，生成式AI将加速商业银行的后台任务，例如在复杂场景下实时回答有关客户财务绩效的问题。它还可以通过增加某些经济条件下业务绩效的稀疏数据来帮助训练预测算法。

在投资银行和资本市场上，生成式AI可以帮助银行对包含复杂、非流动性金融产品的资产负债表进行压力测试。通过综合各种场景的测试数据，该技术可以帮助使金融稳定措施更加精确并降低合规成本。

在支付领域，生成式AI进一步重塑交易方式。基础大模型支持的聊天机器人和虚拟助理提供即时客户支持、解决查询，甚至提供个性化的金融产品和服务。随着人工智能算法分析客户数据以提供量身定制的解决方案，漫长的等待时间和通用建议的日子已经过去。





生成式AI重塑全球金融业

在投资和财富管理方面，人工智能算法通过仔细分析大量的金融数据来发挥其魔力。这些算法可以识别投资机会、预测市场趋势并优化投资组合管理策略。这使投资者能够精确而敏锐地做出数据驱动的决策，从而最大限度地提高潜在回报。

重要的是，生成式AI对银行业务的影响将推动银行自身的组织变革。生成式AI作为AI助手将取代一些数据量大的人工任务，会带动银行人力资源的优化配置，加快组织的去中心化，更多的推行扁平化组织管理。生成式AI成为“副驾驶”，赋能银行员工，提升员工的数字化能力，加速传统的流程专家转型成为数智化数据专家，推动银行员工绩效的重新定义和改革。

一向较为稳定的保险业也正在经历大模型应用带来的一场革命。国内头部保险公司的一位副总裁表示，“对保险行业来说，生成式AI是一项革命性技术，不仅仅是提升效率，后续还会带来商业模式的重大冲击。现阶段，我们应该追求它的多样性，而不仅仅是准确性”。ChatGPT通过分析客户数据实现承保流程自动化、改善索赔管理并增强欺诈检测。AI算法计算风险状况并提供满足个人需求的定制保险计划。

风险管理是金融行业的一个重要方面，生成式AI在该领域也发挥着关键作用。通过分析历史数据、市场趋势和宏观经济指标，人工智能算法可以识别潜在风险并提出有效的风险缓解策略。这使金融机构能够做出明智的决策并确保其运营的稳定性。

尽管ChatGPT持续火爆，带动金融大模型的开发热潮，以及金融机构投资并采用大模型技术，但生成式AI技术在金融业中的应用尚处于技术探索和试点应用的并行期，依据Gartner新兴技术成熟曲线，预计1-2年内，首批大模型增强的金融机构会进入成熟应用期，3年后带动将会金融业生成式AI的规模化应用。当然，随着大模型技术在零售、教育、传媒、游戏等行业中嵌入应用的增加，大模型技术的进化和能力提升，金融业集成生成式AI技术的速度会加快。



生成式AI重塑全球金融业

以我国银行业为例，财报信息显示，42家上市银行中，当前已经有6家银行公开发布大模型技术开发与应用的信息。

工商银行在国内同业率先实现百亿级基础大模型在知识运营助手、金融市场投研助手等多个场景应用。

依托行内人工智能服务体系的算力、算法、数据、人才四位一体的基础能力，农业银行发布的ChatABC重点着眼于大模型在金融领域的知识理解能力、内容生成能力以及安全问答能力，对于大模型精调、提示工程、知识增强、检索增强、人类反馈的强化学习（RLHF）等大模型相关新技术进行了深入探索和综合应用，结合农业银行研发支持知识库、内部问答数据以及人工标注数据等金融知识进行融合训练调优，实现了金融知识理解和问答应用，同时实现了全栈AI技术的自主可控。

平安银行探索自研BankGPT平台，研究构建大模型文本生成、图片生成等能力，及其在图标头像、节日海报、个性化营销内容创作、交互式数据分析、非结构化数据洞察等场景中的应用落地。兴业银行引入部署私有化的商业大模型，上线大模型产品ChatCIB。

北京银行发布AIB平台，汇聚16万条金融知识，面向理财经理、大堂经理、客户经理、综合柜员、远程客服等岗位，打造运营助手、客服助手等7个问答机器人，助力一线员工提升工作质效，同时推出“北银投顾 GPT”，以京行理财产品货架为依托，打造京华四季、智能投顾、财富课堂、产品工具等四大板块，实现重点、优质、特色产品的遴选与推荐，形成面向用户需求的个性化交互，为客户经理精准营销提供良好支持。

长沙银行基于AI模型平台实现百亿级参数大型语言模型本地化部署，提供实时流式问答接口，上线人工智能助手，并入行内业务知识库，问答准确率比传统智能机器人高出43.34%。

一些头部银行、保险机构的高级管理人员表示，已经将生成式AI列入集团公司的年度发展战略，越来越多的银行和保险机构开始建立大模型技术团队，并与外部的金融大模型开发商着手技术开发和产品合作，搭建开放创新生态体系。例如，建设银行、交通银行、中金、中国人寿、泰康人寿、中国银行、招商银行、中信银行等正在搭建各自的新一代人工智能创新平台/中心，积极探索大模型落地应用场景。



拥抱生成式AI的潜在优势

第一，嵌入运营，提高效率。国内头部银行的科技负责人表示，“相比传统模型，生成式AI大模型是很好用的工具和技术手段，具有创造能力、便于生成文件草案、海报底板，端到端更近，最突出的是工作效率的提升，辅助代码编写等，能大幅提升从想法到原型的开发速度。”嵌入生成式AI技术，银行、保险等金融机构可以进一步简化手动流程，最大限度地减少错误并减少人工干预的需要。数据输入、合规性检查和文档处理等日常任务可以实现自动化，从而为财务专业人员腾出宝贵的时间来专注于更具战略性的活动。

第二，拓展业务，加快数字化转型。金融业是单位数据产出的行业，银行业每100万美元的数据产出高达820GB，是制药行业、零售行业的2倍多。当前我国金融机构的数据资产规模达到了千亿级，而且数据质量高，数据正在成为我国金融机构的核心竞争力，如果能够充分挖掘运用，潜在价值超过万亿元。

第三，增强合规，加速创新。生成式AI自动化的文件提取、比对以及内容生成功能，将金融机构中的流程专家转型成数据专家，减少人为的流程干扰因素，减少对专家技术的依赖和离职变动风险，增强内部开发和技术创新能力，创新的银行将从“采用技术的银行”迈向“最懂技术的银行”，并向“最懂银行的先进技术服务商”领域拓展，充分释放其高质量数据价值。

第四，改善客户体验，个性化定制与差异化发展。通过利用生成式AI大模型的力量，金融机构可以提供个性化的推荐、量身定制的财务建议，甚至通过聊天机器人和虚拟协助提供帮助。通过分析客户数据和了解个人偏好，这些人工智能驱动的系统可确保无缝和个性化的客户体验，让客户感到受到重视和理解。

第五，规则的改变者。信用与风险管理是金融业的核心业务，在风险评估和欺诈检测方面，生成式AI将会是金融业游戏规则的改变者，大模型开发商既是金融服务机构的客户，也能成为金融监管部门的合作对象。通过大模型技术快速分析提取文本并实时识别异常，可以快速识别潜在风险和欺诈活动，既可以应用银行账户盗取，也可以被证券监管部门用于上市企业信息披露，实际上都有助于金融机构保护自身安全，增强了金融整体的安全措施，确保为所有人提供更安全的金融生态系统。例如，微软和金融服务商Symphony AI共同开发了生成式AI助手Sensor Copilot，它允许金融犯罪调查人员使用自然语言处理SaaS和可疑活动报告，使他们处理这些犯罪活动的能力提高了60%以上，并将调查效率提高了约70%。



银行可能的部署场景

1. 智慧办公

可用于会议纪要的实时生成、录音视频资料等非文本格式的文本内容翻译及摘要生成、审计资料的内容提取及要点生成。

2. 内部研发

利用ChatGPT的代码生成功能，提高代码生成和测试的效率，加快原型产品开发。

3. 自动化支付

网上银行中的日常任务不再需要手动完成，ChatGPT可以处理这个问题。从账户余额查询到账单支付再到转账请求，银行可以利用 ChatGPT来减少员工的工作量。

4. 查询服务

银行可以将ChatGPT集成到其客户服务系统中，以提供实时答案，从而减少等待时间并改善整体体验。银行通常会遇到大量查询，ChatGPT可以减少查询量，并为人工服务代理留出更多时间来回答更复杂的查询。

5. 金融欺诈检测

ChatGPT拥有令人难以置信的分析能力，可用于帮助银行的专业分析师通过其海量神经网络数据检测欺诈行为。ChatGPT还可以对用户进行身份验证、验证信息并监控活动。

6. 文件比对

ChatGPT可以帮助自动化贷款和其他需要批准的文件的承保流程，验证客户信息以计算资格并评估风险。

7. 个性化客户服务

ChatGPT可以帮助银行为其客户提供个性化服务，将分析客户数据并提供财务和产品建议。例如，如果客户想要申请新的信用卡或信贷产品，ChatGPT可以帮助分析他们的消费习惯、偏好和财务状况，并向工作人员推荐最适合客户的信用卡或信贷产品。

8. 市场调查

ChatGPT可以进行调查和问卷来收集见解，然后分析该数据以生成有用的反馈，帮助银行家做出决策。

9. 电子营销

通过分析数据，ChatGPT可以根据客户行为和已知偏好生成个性化电子邮件活动。无论是邮件/公众号主题行优化、内容分段、A/B测试还是正文，ChatGPT都可以生成这些信息，使写作过程更加简单。

10. AI助理角色扮演（数字员工）

由于其语言处理能力，ChatGPT还可以模拟一些情景，扮演数字代理角色，改变数字营销与客户互动交谈的方式。

11. 财富管理

客户委托银行用他们的钱做出明智的决定。银行必须能够识别潜在风险和市场趋势，ChatGPT可以帮助提供投顾服务、进行财富管理。





银行可能的部署场景

12. 社交媒体管理

ChatGPT可以根据受众行为、偏好和高峰使用时间来优化社交媒体帖子进行调度。在广告方面，ChatGPT可以分析数据并为银行的活动推荐最佳的广告格式和创意元素，并提供有关客户行为、偏好和趋势的见解，为制定更强有力的策略提供可操作的见解。

13. 信贷支持

ChatGPT可以对大量信用和生产生活数据进行分析后，自主生成信用贷款支持方案建议，为信贷部门人员提供更多的决策时间和信息求真的时间，提高信贷质量，减少贷款收益损失。

金融业集成生成式AI的潜在挑战

虽然生成式AI在金融应用的前景广阔，但它也存在一些需要解决的挑战和考虑因素。

最重要的是数据隐私和安全。除了大模型开发中的数据训练带来的数据隐私和安全，大模型使用中也可能存在潜在的数据隐私安全风险。基础大模型技术带来的文本理解生成能力的另一面隐藏着大模型的可记忆功能，在实际应用中，实时数据存在可记忆功能下的被复制风险，因此，大模型的性能选择以及开发商的创新生态显得尤为重要，**金融风险相关的性能应该成为金融机构选择大模型的关键因素。**

而在实际探索应用中，具有专门风险防控技术的大模型厂商确实也是国内外金融机构的主要合作伙伴。比如，全世界很多金融公司都选择微软的Azure OpenAI大模型技术，不仅是从客户服务到自动化流程再到真正的重塑、从客户体验到提高生产力或内容的生成能力，重要的是微软在金融欺诈风险防控方面积累的“核心”技术。微软的AI方案能够检测合成声音，帮助汇丰银行节省了超过4亿美元的身份盗用和欺诈类风险防控的成本支出。微软与Swift在反洗钱方面进行合作，使用Azure隐私计算和AI大模型来检测银行间资金流动的欺诈模式，而无需查看银行客户的数据。度小满、蚂蚁金服、中科闻歌等国内金融科技服务商在金融欺诈风险防控上有着数年的技术积累，开发的金融大模型也重视风控能力的训练和应用更新，受到银行等金融机构的关注和采用，被集成用于智能投研、审计等后台业务场景中。

第二个潜在挑战是大模型幻觉和准确性问题。正如北京的一家大型商业银行所言，“AI本身是很好的技术，是银行后续要发展的核心能力，银行投入了大量资源，不仅是在吸引眼球的层面，更关注在技术本身的挑战、风险、特性，比如模型的准确性、可解释性。”



金融业集成生成式AI的潜在挑战

金融行业容错率非常低，无论是知识问答，还是内容提取上，都对大模型技术提出非常高的要求，当然零/低容错率也是金融业数据质量高的原因所在。目前，银行和其他金融机构非常关注的是——如何解决大模型“一本正经的胡说八道”以及“知道自己不知道”这两个问题，这些问题也是影响银行业采用生成式AI工具的重要堵点。

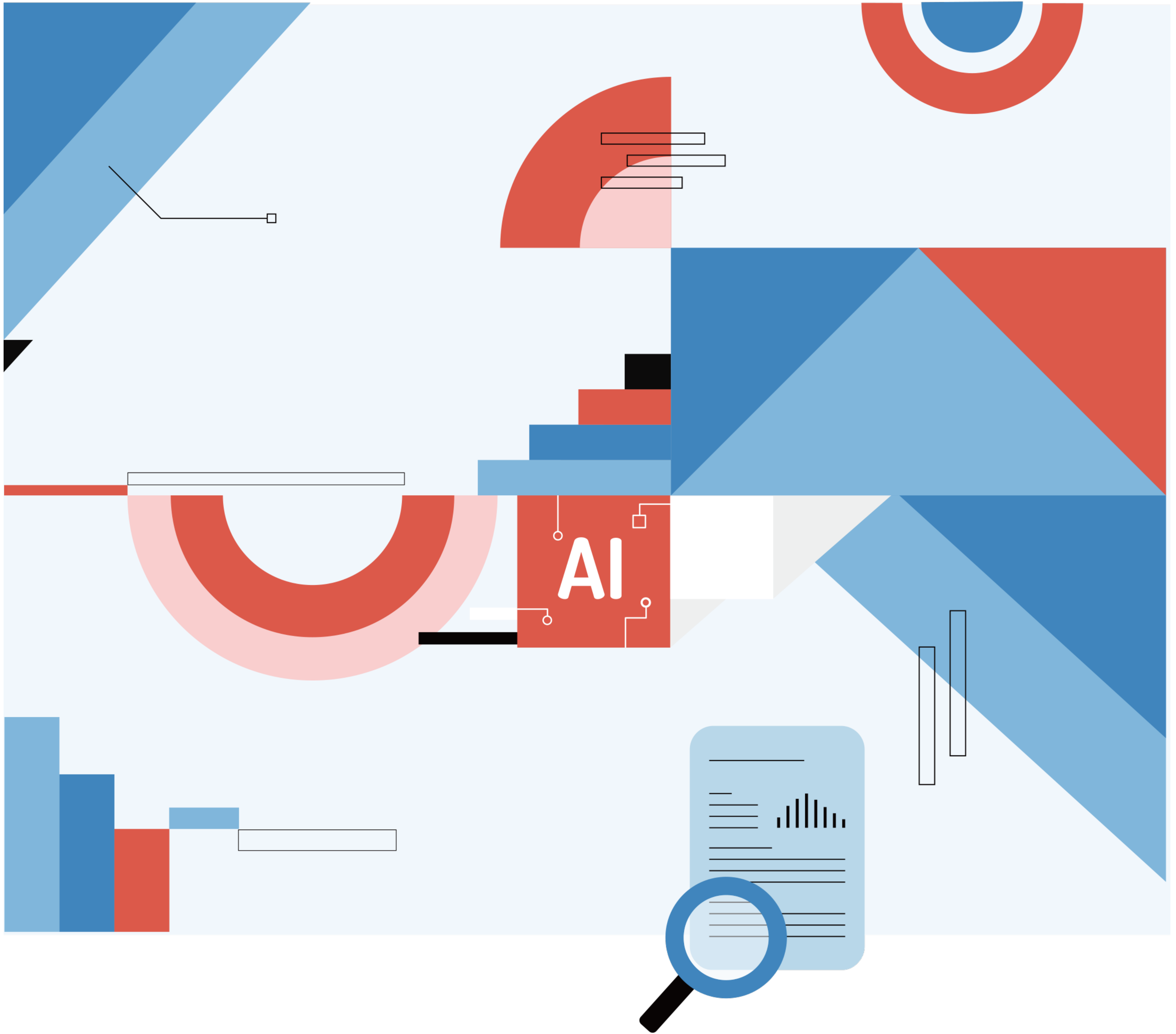
第三是道德和伦理的遵守，发现和减轻偏见应该是防止歧视性结果的首要任务，解决方案是增加训练数据的多样性。第四，监管的不确定性是重要影响要素。生成式AI工程发展领先于基础理论研究，面临智能涌现基础理论机理不明、计算成本高昂、生成内容安全性有待提高、自动评价困难等共性挑战，这会造成监管的不确定性。此外，基础大模型数据库的金融政策时效性会影响输出内容的准确性，例如，GPT-4的数据库也只能访问2023年4月之前的信息，一些最新的法律或趋势都没有添加到生成式AI的考虑范围内。而金融监管的法规可能会发生变化，并且有可能立即实施。从监管和运营的角度来看，这也将限制金融业生成式AI的可用性。

最后，大模型开发和使用的高成本也是阻碍其在金融业应用的主要挑战。高性能GPU一片难求、大规模参数带动的高耗电、大模型人才匮乏、数据安全下的本地化部署是推动当前金融业生成式应用的高成本的主要来源。ChatGPT是一个很好的资源，但高成本可能导致一些预算紧张的小型银行无法使用它。

小结

总之，从简化运营和增强客户体验到降低风险和释放新机遇，生成式AI的潜力巨大且具有变革性。在创新和监管之间取得适当的平衡将是释放生成式AI在金融领域全部潜力的关键。在实际应用中，实现人类与人工智能之间的和谐协作非常关键。人与人工智能之间是相互赋能而非取代，生成式AI在重复且数据量大的任务中具有先发优势，解放这些任务中的劳动力，赋能金融专业人士更专注其专业任务，同时在使用工具的过程中不断产生新的对话数据，又推动生成式AI技术本身迭代创新。

随着生成式AI的不断发展，我们可以期待更多的金融创新应用和进步，通过拥抱人类和人工智能的协作潜力，进一步重塑金融格局。



PART 3

案例解读 | 一些金融业生成式AI应用的成功案例



金融大模型成功开发案例

1

典型成功案例剖析

度小满发布国内首个开源金融大模型“轩辕”

度小满，原百度金融。2018年4月，百度宣布旗下金融服务事业群组正式完成拆分融资协议签署，拆分后百度金融启用全新品牌“度小满”，实现独立运营。2018年5月21日，度小满正式成立，延承百度技术基因，在智能金融时代，充分发挥 AI 优势和技术实力，携手金融机构合作伙伴，用科技为更多人提供值得信赖的金融服务。

“轩辕”金融大模型——基础模型能力建设

- 金融理解：增量预训练和指令微调阶段，加入大量金融数据，提升金融理解能力
- 知识增强：外挂实时更新的业务知识库，实现低成本干预，同时降低幻觉影响
- 应用增强：面向金融应用场景，定向增强摘要、逻辑、计算等金融场景核心能力
- 对话能力：使用百万级经人工构建和校验的高质量指令数据进行指令微调和对齐

“轩辕”金融大模型——基础模型相关进展

- 轩辕：2023年5月，度小满开源了国内首个千亿级金融大模型。在金融场景中的任务评测中，“轩辕”全面超越了市场上的主流开源大模型。“轩辕”用度小满实际业务场景积累的海量金融数据训练而来，通过独创的hybrid-tuning的创新训练方式，实现在大大增强金融能力的同时，不损失通用能力。自开源以来，已经有上百家金融机构申请试用“轩辕”大模型。
- 轩辕-70B：2023年9月，度小满开源“轩辕-70B”金融大模型。

- 通用能力：度小满“轩辕-70B”大模型在 C-Eval和CMMLU两大权威榜单上位列所有开源模型榜首：在由清华大学、上海交通大学和爱丁堡大学合作构建C-Eval榜单上，“轩辕-70B”的总成绩达到71.9分；在由MBZUAI、上海交通大学、微软亚洲研究院共同推出的CMMLU榜单中,轩辕70B以71.05分数排名第一。

模型	CEVAL	CMMLU
XuanYuan-70B	71.9	71.1
Llama2-70B	52.1	53.11
GPT4	68.4	70.95
InternLM-20B	58.8*	-
Baichun2-13B-base	58.1	61.97
XVERSE-13B-base	54.7	58.4
QWen-7B	59.6	58.66
ChatGLM2-6B	51.7	48.8



金融大模型成功开发案例

1

“轩辕”金融大模型——基础模型相关进展

- 金融能力：“轩辕-70B”已经通过注册会计师、银行/证券/保险/基金/期货从业资格、理财规划师、经济师等金融领域权威考试，且考试得分领先于其他通用模型。
- 场景能力：“轩辕-70B”在度小满自有金融业务场景测试中表现领先，特别金融知识问答、NL2SQL等场景表现优异。

大模型落地金融相关场景的探索

大模型技术已经应用在度小满各个业务场景，从营销、客服、风控、办公再到研发，已经初见成效。在代码助手方面，用大模型辅助生成的代码，采纳率能够达到42%，帮助公司整体研发效率提升了20%；在客服领域，大模型推动服务效率提升了25%。在智能办公领域，大模型目前的意图识别准确率已达到97%。

成功之道

一是领先的人工智能科技创新能力。度小满，由百度金融服务事务群拆分而来，延续了百度强大的人工智能科技创新能力。基于人工智能、金融云、区块链和物联网等底层架构，度小满搭建了金融大脑，运用情感计算、关联网络和知识图谱等技术，进行感知引擎和决策引擎运算。2019年底，度小满的智能语音机器人、智能风控等项目入选了人民银行首批金融科技应用试点。2023年5月6日，度小满“智能化征信解读中台”工程荣获我国智能科学技术最高奖“吴文俊人工智能科技进步奖”。

二是开放创新生态助力前沿技术探索。2023年2月，度小满宣布成为百度文心一言首批生态合作伙伴。度小满将基于自身金融场景积累的海量对话及解决方案数据，融合“文心一言”的全面能力，打造全新的智能服务。这是对话式语言模型技术在国内金融场景的首次落地应用。此外，度小满还携手中国科学院自动化研究所、哈尔滨工业大学等探索人工智能前沿研究。2020年底，度小满获批北京市博士后科研工作站，为更好培养顶级金融科技人才奠定了基石。

三是丰富的应用场景促进人工智能技术迭代创新。目前度小满已经和上百家银行、金融机构合作，优势互补，共同推动普惠金融，截至2022年底，度小满携手金融机构，累计服务数千万小微企业和个体工商户，其中7成是员工在5人以下的小店，超过50%分布在四线及以下城市。在助力金融机构防范风险降本增效，小微企业信贷支持发展的过程中，度小满不断迭代升级其金融科技产品创新。



金融大模型成功开发案例

2 其他金融大模型成功开发案例

2023年3月，全球商业、金融信息和新闻资讯提供商**彭博**推出Bloomberg GPT，这是一款专为金融行业从头打造的500亿参数大语言模型，可以快速分析金融数据，帮助进行风险评估、情感分析、问答等功能。Bloomberg GPT训练语料包括3450亿的公共数据集和3630亿的金融数据集。该模型已在彭博内部落地，被用于生成查询语言、提供标题建议、金融问答。

2023年6月3日，**中科闻歌**发布安全可靠的企业级专属大模型——雅意大模型，并推出了金融、媒体、宣传等领域的大模型应用。金融大模型提供宏观政策解读、金融事件追踪、企业财报智能解析。半年后，中科闻歌推出雅意2.0大模型，参数规模为300亿，支持128k上下文窗口输入，相当于20万字文本，具备图文交互的多模态能力，支持10余种主流语言以及十余种智能插件调用。在中文知识问答能力的评测方面，雅意2.0在AGIEval、CMMLU数据集评测中排名第一；零样本中文信息抽取能力方面，雅意2.0在实体抽取方面，实现了中文第一、英文全面对标ChatGPT。雅意金融大模型2.0支持多模态财报问答、一键合同关键条款解析入库、协议履约监控预警、报告审核合规比稿等功能，目前已经在国内大型商业银行、投资银行、券商等金融机构应用。

2023年9月8日，**蚂蚁集团**正式发布针对金融产业深度定制的金融大模型。蚂蚁金融大模型基于蚂蚁自研基础大模型，在万亿量级Token的通用语料基础上，注入千亿量级Token金融知识，并从300+真实产业场景中提取了共60万+高质量指令数据，形成了金融专属任务性能优化的优势数据资产，底层算力集群达到万卡规模。该大模型聚焦真实的金融场景需求，在“认知、生成、专业知识、专业逻辑、合规性”五大维度28类金融专属任务中表现突出，在“研判观点提取”“金融意图理解”“金融事件推理”等众多领域达到行业专家水平。蚂蚁金融大模型已在蚂蚁集团的财富、保险平台上全面测试，理财侧包括理财选品、产品评测、行情解读、资产配置等6大类服务，保险侧包括产品解读、家庭配置、智能核保、智能理赔等10多个智能服务。



银行业的成功应用案例

3 典型成功案例剖析

华侨银行（OCBC）是第一家向全球所有员工推出生成式人工智能聊天机器人（OCBC GPT）的新加坡银行。在本地化部署近1000名员工的半年测试后，华侨银行在2023年10月24日宣布，该银行将向全球30000名员工提供生成式AI聊天机器人，以协助他们写作、研究和构思，撰写投资研究报告、将内容翻译成多种语言以及起草客户回复。

华侨银行（OCBC）

华侨银行是新加坡第一家本地银行，由三家本地银行1932年合并而成，其中最古老的银行成立于1912年。经过近百年的不断发展壮大，华侨银行已经成为一家区域性金融服务集团，网络覆盖了东盟90%的贸易和资本流动，并在纽约、伦敦和悉尼等主要经济体开展业务，是第一家收购中国银行业股份的东南亚银行。2022年底总资产达5600亿新元，核心业务包括银行、保险和财富管理。

OCBC GPT的部署

2023年4月，华侨银行和微软Azure OpenAI合作开发了生成式AI聊天机器人的OCBC GPT，旨在进一步提高生产力并更好地为客户服务。OCBC GPT由ChatGPT的大型语言模型提供支持，同时，OCBC GPT托管在安全且受控的环境中，华侨银行员工输入的信息保存在银行内，不会提供给Microsoft或任何外部各方。2023年10月，经过银行内部1000名员工6个月的内部测试后，华侨银行宣布将OCBC GPT推广到该银行的所有30000名员工。

应用场景

该模型可以分析查询并根据网络上可访问的基于文本的信息生成完全充实的答案和结果，已经被应用在撰写投资研究报告、翻译多种语言的内容以及起草客户回复等领域。此外，该模型还被用来加强银行的反洗钱工作。通过生成式人工智能驱动的平台，华侨银行可以快速分析客户资料和网络，识别可疑社区和交易。



银行业的成功应用案例

效果

这些工具有潜力通过自动化各种耗时的任务来改变员工的工作方式，让他们腾出时间专注于更具战略性和增值的工作。这反过来又帮助银行员工可以花更多时间与客户建立关系并开发创新产品和服务，从而提供更好的客户服务。

来自多个职能领域的试验参与者表示，他们完成任务的速度比以前快了约50%。在OCBC GPT“家族”中，OCBC Wingman工具可以实现开发人员使用编码助手自动生成、调试和改进计算机代码，自2023年5月推出该工具以来，生产力提高了约20%。Document AI工具提供文档内容提取和分析，通过拖放过程从文档中提取和汇总关键信息，将每个文档提取信息所需的时间从30分钟减少到1分钟，供需要阅读多个财务、风险和 ESG 文件的员工使用。目前，华侨银行每天有超过400万个风险管理、客户服务和销售等流程中的决策是由人工智能做出的。预计到2025年，这一数字将增加到1000万。使用生成式人工智能，通过移动银行应用程序向客户推送个性化建议和见解。每年发送2.5亿条此类见解，帮客户节省时间和精力并改善财务管理。

该集团最新的财报显示，华侨银行2023年第三季度净利润较上年增长21%，净利息收入较2022年第三季度增长17%，集团的财富管理收入（包括来自保险、私人银行、高级私人客户、高级银行、资产管理和股票经纪的收入）增长16%，成本收入比从上一年度的42.2%改善至39.1%。

成功之道

持续的先进人工智能技术投资与人才培养是华侨银行成为世界上首批大规模部署生成式人工智能工具的银行之一的关键原因。2018年，华侨银行是新加坡第一家推出专门人工智能实验室以战略性开发内部人工智能能力的银行，该实验室后来被纳入集团数据办公室。2019年，华侨银行是第一家推出人工智能研究生奖学金以培养人工智能人才的新加坡银行。同年，该银行启动了数据认证途径，通过为期12至18个月的培训来培养内部数据专业人员。此后已有300多名员工完成了该计划。

与人工智能技术领先的微软和OpenAI合作是华侨银行能快速嵌入生成式AI技术提升生产力的一个重要原因。2022年，在推出OCBC GPT之前，OCBC已部署了多种生成式AI工具来开发代码、总结文档、转录通话以及创建内部知识库。通过这些工具的使用，相关任务部门的生产效率提高了50%，释放了员工的时间，使他们能够专注于更高价值的任务。

银行业的成功应用案例

4 其他成功案例

工商银行与鹏城实验室、清华大学、中国科学院、华为等联合发布基于昇腾AI的金融行业通用模型，首家实现了企业级金融通用模型的研制投产。该大模型已被应用于工商银行的客服、营销、运营、风控等业务主战场中，并取得了良好地应用效果。在客户服务领域，工商银行应用该模型支撑智能客服接听客户来电，显著提升了对客户来电诉求和情绪的识别准确率，能够更精准有效地响应客户需求，并且可以大幅缩减维护成本。在行内办公领域，工商银行全面升级办公模式，利用大模型的文本生成、问答能力，围绕邮件、文档、会议、员工日常事务等方面，优化行内办公工具的交互体验，助力全行40万员工的办公效率提升。比如会议纪要生成，根据会议对话内容，大模型快速生成会议纪要初稿，降低人工记录会议纪要的成本，助力全行办公领域智能化。在风险防控领域，工商银行实现了对工业工程融资项目建设的进度监测，监测精准度提升约10%，研发周期缩短约60%。在智能研发领域，工商银行利用大模型代码生成、代码识别与检测、代码转自然语言等领域的全方位能力，构建基于大模型的智能研发体系。截至目前人工智能编码助手共收集到超2100万个字符编码数据，录入代码超80万行，编码助手生成代码量占总代码量的比值从20%提升至40%，有效提升一线开发人员编码效率和质量。在运营管理领域，模型的应用帮助智能提取期限、利率等信贷审批书核心要素，提升了信贷审批效率。

摩根大通在新一代人工智能方面开展了大量投资，包括153亿美元的技术预算，已经有数千名员工在工作中使用生成式AI技术，实现了超过15亿美元的商业价值。正在测试的人工智能应用程序包括可以为其关注的公司生成收益摘要的应用程序。该银行还在开发一项服务台服务，为用户提供解决问题的步骤，而不是相关文章的链接。

摩根士丹利利用GPT-4的嵌入和检索功能来利用其智力资本，建立其财富管理知识库，并纳入其咨询服务。这个面向内部的聊天机器人可以对财富管理内容进行全面搜索，并有效地释放摩根士丹利财富管理积累的知识，提供量身定制的投资策略并优化客户互动。

NatWest与IBM合作，利用生成式AI增强其虚拟助手Cora，致力于通过前端服务改善客户交互。美国银行和富国银行在内的多家银行提供人工智能驱动的聊天机器人来帮助零售客户管理财务。



银行业的成功应用案例

4

其他成功案例

西太平洋银行Westpac在使用ChatGPT来帮助银行内部的开发人员编写代码。高盛正在开发十几个项目，包括用英语命令编写代码以及生成文档的能力，（出于监管方面的考虑，这些项目都不是面向客户的），旨在将生成式AI纳入其商业实践。

奥地利第一储蓄银行Erste Bank自2023年7月起已向所有员工开放ChatGPT，并为员工提供人工智能学习课程。不过，该银行的当前的开放访问并不是面向客户的商业行为，而是为了让员工有机会学习如何以有趣的方式使用生成式AI。

花旗银行专门成立生成式AI工作组。在见证了允许大约250名开发人员尝试生成式AI的试点计划取得成功后，花旗银行计划在今年扩展到集团40000名程序员中的大多数，并且员工针对该技术提出设计出了全行350多个用例。有些应用程序很小，有助于日常工作，而另一些则涉及复杂的项目。例如，花旗银行正在使用大语言模型来分析文档、提高效率、确保合规，提高开发人员的工作效率并测试软件漏洞，简化流程并使员工能够专注于客户互动。但不适用于信用评分或影响客户访问金融产品的其他应用程序。在最近的一项技术应用中，花旗集团的风险与合规团队使用生成式人工智能来分析和评估联邦监管机构发布的1089页的新资本规则的影响。该技术将提案提炼成可管理的部分并生成关键要点。该集团正在探索使用大型语言模型来消化其业务所在国家的立法和法规，以及探索利用生成式AI对其系统进行现代化改造。传统上，更新遗留系统需要大量的财政资源和人力。然而，生成式AI可以帮助翻译编码语言，使银行能够更经济高效地更新其系统。

利用集中式内部数据库，从内部电子邮件以及销售人员、交易员和客户之间的聊天等来源中获取信息数据，汇丰银行开发了自己的生成式人工智能工具——AI Markets，为机构投资者客户提供了可以跨不同资产类别执行任务的数字助理，生成定制的金融市场分析、访问汇丰银行的实时和历史跨资产数据集，并浏览最新的市场见解。用户可以通过银行的单一经销商平台、Symphony等外部平台或直接通过API连接插入该工具。



保险业的成功应用案例

5

典型成功案例剖析

LAQO与全球云通信平台Infobip合作，使用Azure OpenAI服务开发了生成式AI助理Pavle，可以为客户提供24/7全天候支持，从而改善客户体验并提高忠诚度。现在，30%的客户查询由Pavle解决，使LAQO的客服人员能够专注于复杂的案例和客户获取。

LAQO

LAQO是克罗地亚第一家全数字化保险公司，主要提供车辆和旅行保险等产品，用最简化的数字流程取代了传统的纸质工作流程。该公司创新了游戏化功能LaqoPrevent，该功能使用远程信息处理来鼓励更好的驾驶习惯，奖励负责任的驾驶，并向客户解释如何减少二氧化碳排放。除了核心服务之外，LAQO还涉足数字创新，推出了加密货币支付等功能，还推出了传播保险发展的元宇宙博物馆。

LAQO的生成式AI助手Pavle开发历程

第一步，开发缘由。LAQO深知24/7客户支持正在成为越来越重要的差异化因素，促使公司优化其客户服务。ChatGPT的惊人内容生成能力，让这种客户服务优化功能实现成为可能。

第二步，合作伙伴选择。LAQO求助于其长期合作伙伴Infobip，以便通过客户首选的消息传递渠道将客户连接起来，Infobip以其通信平台上的生成式AI的不懈创新而闻名，此前曾为LAQO提供全渠道云合同中心解决方案。

第三步，挖掘场景应用。在与Infobip团队坐下来绘制客户支持路线图时，LAQO发现，尽管其网站上涵盖了所有信息，但客户却没有时间阅读，因为他们更喜欢通过WhatsApp与客服人员交谈。

第四步，生成式AI助手LAQO GPT的开发。为了解决重复查询问题，LAQO和Infobip团队开始使用Azure OpenAI服务和Infobip的AI Assistant构建解决方案来设计AI Assistant的决策结构。

第五步，模型微调。构建AI助理Pavle花了两周时间，但真正的挑战是语言。因为OpenAI的基础大模型对克罗地亚语的支持不如英语，因此Infobip与LAQO密切合作，微调语言以确保准确性。文档准备就绪后，团队就可以在五分钟内设置新版本的AI助手。



保险业的成功应用案例

5 LAQO的生成式AI助手Pavle开发历程

第六步，解决“生成式AI助手知道自己不知道什么”的问题。开发团队确保人工智能助手只能涵盖保险索赔和有关LAQO的一般信息，以避免提供误导性信息，从而提高大模型的可靠性、准确性和可解释性。

第七步，解决大模型幻觉问题。采取两项措施在一定程度上解决大模型幻觉问题。一是通过调整温度，即算法本身的参数控制结果的随机性或不确定性，二是使用先进的提示技术和精心准备的内容，LAQO特别注意尽量减少幻觉或训练数据无法证明的夸大宣传反应。

第八步，进一步微调。对生成式AI助理的语气及其幽默感进行微调。例如，一个克罗地亚语单词可以表示“从天而降的冰”和“战斗”。当在车祸测试中提到这个词时，AI助理会询问谁赢得了这场战斗，但是大多数人在事故发生后都没有幽默的心情，因此幽默感并不适合这类场景，需要不断微调AI助理的语气。

实施效果

- **效果一，友好即时的客户互动：**LAQO的生成式AI助理Pavle可通过WhatsApp商务平台全天候协助和迎接客户，并将其转接给现场客服人员以进行更复杂的查询。数字助理提供快速处理事故索赔的指导步骤。确认保险范围后，AI助手会将客户引导至可以报告事件的应用程序。报告损坏情况后，LAQO的代理人员可以立即开始处理索赔。如果所有文件准备就绪，损坏问题可以在24小时内解决。对于受LAQO投保司机影响的第三方，AI助理提供了一个自动化的网络流程链接，以方便索赔处理。为客户支持自动化树立了新的基准，为用户提供即时和高质量的帮助。
- **效果二，优化客户支持和代理效率：**生成式AI助手完全重新定义了LAQO的客户支持，获得了保险客户所需的即时性。可以解决LAQO 30%的客户疑问，在90%的情况下，甚至通过交换三到五条消息就可以解决查询问题，同时AI助理Pavle提供的是一致的答案，这种没有偏见的一致性对于保险公司维持客户信任至关重要。
- **效果三，优化人力资源配置：**由于严格的测试，代理人员最初将生成式AI助理视为一种竞品，但集成使用后，他们认识到了其中的好处，现在已经可以自动化提供出色的客户支持，而代理人员可以花时间帮助新客户、处理复杂的案例或构建新业务。此外，该公司在联络中心上花费的精力减少了10%。即使需求不断增长，也不需要扩大团队规模。在节省资金和资源的同时，LAQO打算将其集成到其他重要的计划之中。



保险业的成功应用案例

5

下一步行动

一是让AI助手Pavle能够销售保单并向潜在客户提供特定信息。该公司已经开始训练AI助手，并推出了一项针对猫和狗等宠物的新保险服务LAQO Pets。

第二，LAQO正在通过向AI助手的回复库添加新元素来进一步人性化客户交互。Pavle现在将共享与问题本身相关的图像和视频，从而使整个互动更加吸引人。此外，LAQO也在开发多重人格功能。未来的Pavle遇到一个严肃的问题时，他会用严肃的语气回答，而对于轻松的问题，Pavle则可以表现出有趣的一面。

成功之道

专业的人做专业的事。LAQO深耕保险业务，专注向客户直观且友好地展示保险，从而悉心洞察客户需求，并大胆引进先进的技术来提升客户体验和获得感。作为技术合作伙伴，微软和OpenAI提供了安全的基础大模型技术底座，而长期合作伙伴Infobip则提供了打造用户体验的技术和专业知识。

6

其他成功案例

2023年5月23日，保险行业软件解决方案全球提供商Sapiens International Corporation（Sapiens）宣布与微软达成一项协议，将集成微软的Azure OpenAI服务，以利用保险行业最先进的生成式AI模型和工具。Sapiens将集成Azure OpenAI和Azure Power Virtual Agents，为保险公司提供生成式AI解决方案，使客户能够使用大语言模型轻松浏览保单、条款和条件等复杂文档。Sapiens的人工智能、认知服务和流程自动化可以帮助保险公司贯穿整个保险价值链，以多种语言在客户交互的多个领域提供即时答案，增强客户体验，同时减少现场代理和相关人员的大量呼叫。它还可以协助承保人的日常活动，并减少索赔处理和欺诈检测的时间。



私募股权基金领域的成功应用案例

典型成功案例剖析

2023年，全球最大的另类资产管理公司黑石集团（Blackstone）以多种方式应用生成式人工智能技术。

黑石集团

黑石集团成立于1985年，是一家总部位于纽约的全球最大另类投资管理公司，管理着1万亿美元的资产，分布于230多家投资组合公司，并大量持有房地产、私募股权和信贷基金，拥有大量适合人工智能赋能的专有数据。

黑石集团集成生成式AI的情况

在工程方面，黑石集团正在使用代码搜索和代码智能工具开发商Sourcegraph旗下的生成式AI产品Cody，该产品会调阅Blackstone丰富且分散的代码库，然后编写代码并创建新软件，以帮助显著缩短处理这些任务所需的时间。在私人财富管理业务中，黑石集团采用了大型云软件服务商Salesforce提供的生成式AI产品——Einstein GPT，用来联系客户并提供个性化的投资建议。

在投资方面，黑石集团正在构建一个内部工具，该工具有潜力使用生成式AI增强投资流程，例如使投资团队能够获得深入的公司和行业分析。同时黑石集团还在探索推出集成Microsoft和Zoom的各种“副驾驶”和辅助工具，以帮助员工提高工作效率。

黑石集团生成式AI的应用场景

依靠大模型来评估其打算收购的资产的风险。黑石集团利用生成式人工智能来预测需求，并利用其庞大的电子商务仓库库存立即为每个企业客户提供价格。在公司内部，数据科学家加入了投资团队，以确保生成式人工智能工具有助于完成交易或拒绝交易。

黑石如何解决数据安全

黑石集团拥有世界一流的网络安全计划，以确保飞行数据和静态数据尽可能安全。在评估并开始利用生成式AI平台之前，黑石花费并将继续花费大量时间来确保数据不会泄漏到基础大模型，并以专有方式进行处理。



私募股权基金领域的成功应用案例

黑石集团如何解决生成式AI生成内容的准确性问题？

验证生成式人工智能工具的输出质量，以确保这些工具能够提高生产力。同时时刻关注其他的生成式AI平台和技术，以确保数据的安全。

下一步计划

将在整个公司推出更多基于人工智能的解决方案。正在探索的一些工具包括跨各种软件平台的Copilot“副驾驶”助理，以提高生产力、自助服务台聊天机器人以及旨在增强财务流程、软件开发、营销内容生成和法律任务的技术。

黑石还将还将继续开发定制的生成式AI工具，并与黑石集团的投资组合公司合作，支持他们构建内部AI平台并将AI集成到他们的产品和流程中，赋能合作伙伴。

成功之道

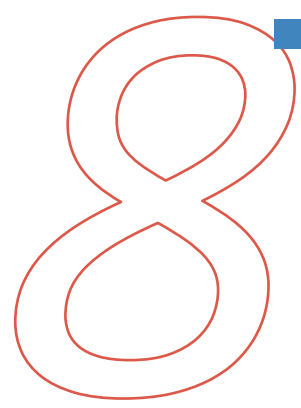
第一，黑石集团洞悉到数据是智能时代金融服务机构的核心竞争力，因此较早地采用生成式AI，试图加快数据被计算的效率，释放数据创造价值潜力。

二是开放式创新思维。在生成式AI技术集成上，黑石集团采取的是引进与自主研发并举的开放创新策略，这大大缩短了先进技术被采用的时间。同时集团还敦促旗下的投资组合公司充分利用其AI资源，包括其深厚的专有数据。

三是首席执行官苏世民对拥抱生成式AI技术的坚定信仰，亲自鼓励怀疑者使用ChatGPT或谷歌的Bard来写电子邮件和备忘录，并任命一位具有技术背景的高级管理人员来监督集团生成式人工智能的实施，以提高生产力，创造新产品和服务，并接触到越来越多的客户。



私募股权基金领域的成功应用案例



其他成功案例

全球最大的投资管理公司贝莱德（BlackRock）宣布已经在其Aladdin和eFront风险管理系统中构建“副驾驶”，计划于2024年1月向客户和员工推出首款生成式AI工具，在整个公司的Office 365办公软件中部署微软的AI插件，以增强风险管理和提升客户服务和员工体验，更广泛地实现编码民主化，目的是在利用新兴技术的竞争中获得竞争优势。此外，贝莱德还在构建工具来帮助其投资专业人士收集财务和其他数据，用于研究报告和投资建议以及语言翻译器。贝莱德首席财务官Martin Small认为，大语言模型和自动化使贝莱德能够真正扩展员工的时间和精力，帮助基金经理削减固定成本并提高利润率，这样就可以在2024年及以后推动更多的生产力提升。”当然，该集团的备忘录上特别提到，在所有情况下，由人工智能生成的“初稿”都将经过正常质量控制，并且所有数据都将保留在贝莱德内部，而不是与开源大语言模型及应用程序的用户共享。

总部位于瑞士的全球私募股权公司Partners Group Holding AG（瑞士合众集团）利用GPT-4开发了PRIMERA GPT,并已于2023年6月初向瑞士合众集团的所有员工开放。目前，该集团的高管正在尝试使用该技术寻找有吸引力的投资子行业、识别其投资组合公司的竞争对手、请求撰写提案支持和生成报告等应用。PRIMERA GPT是“隔离的”，这意味着所有输入数据都处于严格控制之下。集成生成式AI，不仅代表了私募股权管理的开创性方法，而且表明了该公司致力于利用尖端人工智能技术来增强投资策略和改进投资组合管理，在私募股权行业设定新标准。

施罗德个人财富在财富管理中采用生成式AI，旨在通过更准确地监控客户互动并生成以数据为主导的合规证据，彻底改变其财务咨询服务。



金融科技企业集成生成式AI成功案例

9

Stripe与OpenAI达成战略合作，相互赋能

Stripe

Stripe成立于2011年，是面向商家的金融基础设施平台，致力于为各行各业的企业提供在线支付、商业贷款等金融解决方案。这家公司独具匠心，创新性地自动计算和征收销售税，为广大商家降低了经营成本，简化了财务流程。Stripe堪称美国最具价值的私有金融科技企业。随着数字化转型的加速，Stripe不断拓展业务范围，为各类企业提供一站式金融解决方案。其广泛的服务领域和强大的技术实力，使公司在金融科技领域独树一帜。2023年3月15日，Stripe与OpenAI宣布开展战略合作。

Stripe与Open AI的合作内容

- **第一，Stripe利用GPT-4增强Stripe，获得先进生产力，提高开发效率**

2023年1月Stripe就加入到OpenAI的GPT-4测试版，将GPT-4整合到其产品和服务中来增强Stripe，并通过GPT-4构建新工具。Stripe还确定了一系列利用GPT-4简化操作的方法，以帮助用户更快地获取他们需要的信息。

这项工作的首批成果之一是由GPT支持的Stripe文档。这对Stripe的高质量文档是进一步的提升，开发人员可以在Stripe文档中实现自动查询，它将通过总结文档的相关部分或提取特定信息来给出回答,让开发人员可以花更少的时间在阅读上，可以有更多的时间去创建。

- **第二，使用Stripe助力OpenAI，Stripe获得增量价值**

2023年3月15日，OpenAI选择Stripe来支持ChatGPT Plus和DALL·E的付费支付，同时帮助OpenAI处理从定期开单和税务合规到财务运营自动化等方面的各种事务，助力OpenAI的技术商业化。

例如，Stripe Billing和Stripe Checkout为OpenAI的ChatGPT高级订阅服务ChatGPT Plus提供支持。使用Checkout的预构建托管支付页面，OpenAI提供一次性购买、每月或基于使用的订阅，并以超过25种支付方式覆盖全球客户。Stripe Billing为DALL·E用户提供了一个灵活的按量付费选项来购买DALL·E 点数，使客户能够生成新的图片。



金融科技企业集成生成式AI成功案例

9

Stripe与Open AI的合作内容

GPT-4增强的Remember Me “变种”——Stripe Link，可以自动填充保存的支付详情，使得OpenAI用户的支付速度平均加快了40%。OpenAI现在是Link网络的一部分，Link网络中保存有数十万商家的支付详情，为数千万用户实现了更快结账。

此外，Stripe Tax还支持OpenAI在向全球市场扩张的过程中履行其税务合规义务。Revenue Recognition帮助OpenAI保持对其财务状况的掌控。目前，75%领先的生成式AI公司（包括 OpenAI、Runway、Diagram 和 Moonbeam）已经与Stripe签约，以快速推向市场，在兼顾合规的情况下进行扩张，并将他们的产品带给全球更多的用户。

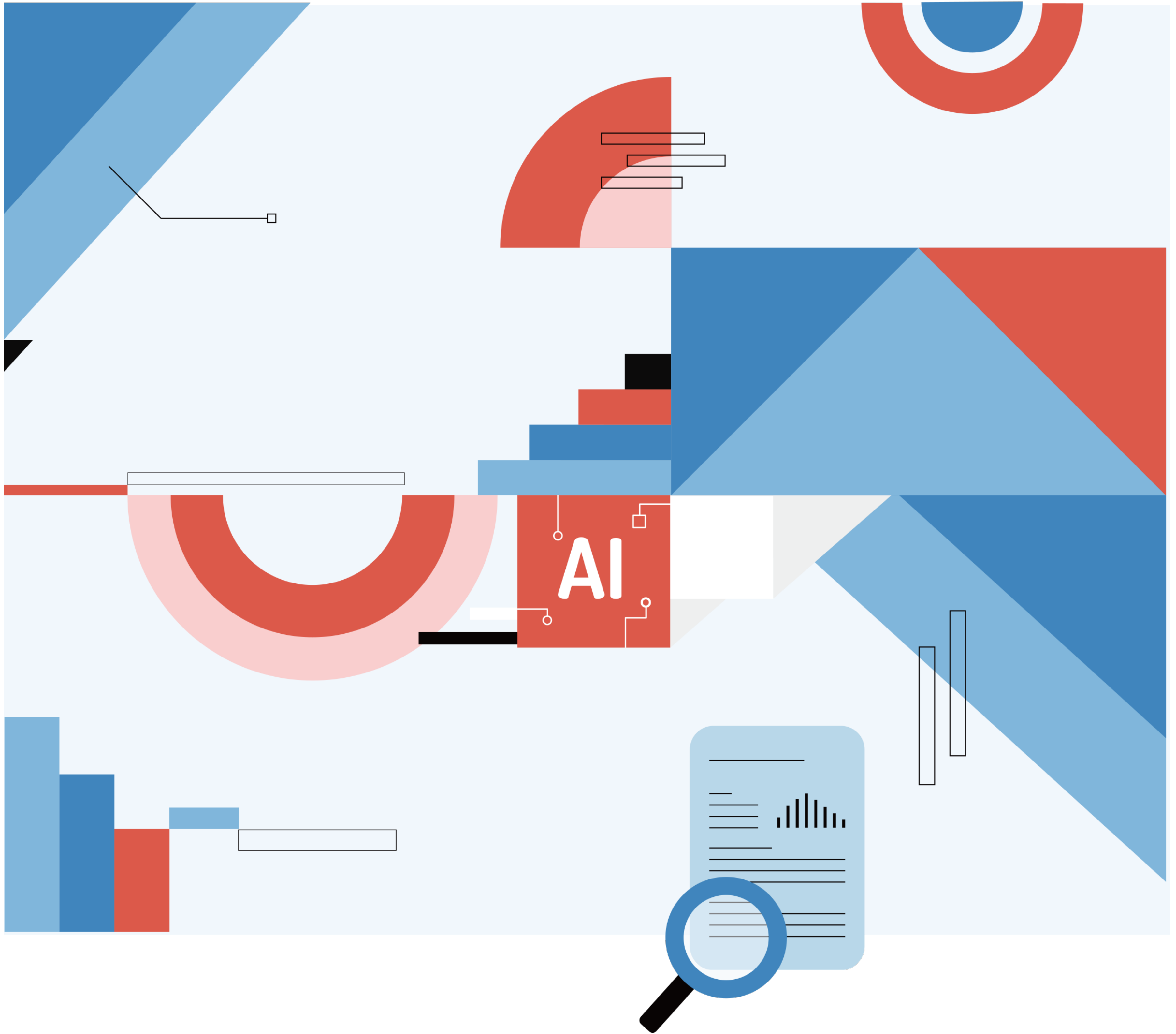
效果

使用Stripe的产品套件，OpenAI得以在几周内为多个产品线推出一个全球支付系统。OpenAI客户登录Stripe的Link后平均结账速度加快40%。

成功之道

一是持续的科技投入。Stripe长期以来一直使用AI技术来增强其产品和用户体验，包括帮助用户管理欺诈和提高转化率。2022年，Stripe与OpenAI团队合作，为Stripe Support部署了GPT-3，这有助于代理为他们每周保持互动的许多用户确定更快的解决途径，助力支持团队更好地服务用户，并打击欺诈行为。例如，通过智能路由问题票证和总结用户问题等任务，提高工作效率和用户满意度。

二是大胆的人机协同创新。在部署GPT-4时，Stripe让公司来自支持、入门、风险和文档团队的100名工程师停下他们的日常工作，而使用GPT-4为支付平台构思功能和特性，旨在确定Stripe中可以集成生成式AI加速的产品和工作流，并真正了解大语言模型在哪些领域工作得很好，哪些领域仍然存在困难。该公司的高管们深信，开发人员越熟练地采用大语言模型，Stripe就会在数字支付宇宙中扩散得更远。



PART 4

政策期望 | 我国金融业生成式AI应用迎来政策红利期



中央高度重视金融业高质量发展

2023年10月召开的中央金融工作会议强调，“金融是国民经济的血脉，是国家核心竞争力的重要组成部分，要加强党中央对金融工作的集中统一领导，做好科技金融、绿色金融、普惠金融、养老金融、数字金融五篇大文章，培育一流投资银行和投资机构，支持国有大型金融机构做优做强，严格中小金融机构准入标准和监管要求，强化政策性金融机构职能定位，发挥保险业的经济减震器和社会稳定器功能，坚持“引进来”和“走出去”并重，稳步扩大金融领域制度型开放，提升跨境投融资便利化，加快建设金融强国，坚定不移走中国特色金融发展之路，推动我国金融高质量发展，为中国式现代化全面推进强国建设、民族复兴伟业提供有力支撑。”

2024年1月16日，习近平总书记在省部级主要领导干部推动金融高质量发展专题研讨班开班式发表重要讲话，深刻阐释了金融强国的丰富内涵，强调“金融强国应当基于强大的经济基础，具有领先世界的经济实力、科技实力和综合国力，同时拥有强大的货币、强大的中央银行、强大的金融机构、强大的国际金融中心、强大的金融监管、强大的金融人才队伍一系列关键核心金融要素。建设金融强国，必须加快构建中国特色现代金融体系，建立健全科学稳健的金融调控体系、结构合理的金融市场体系、分工协作的金融机构体系、完备有效的金融监管体系、多样化专业性的金融产品和服务体系、自主可控安全高效的金融基础设施体系。”

我国金融业拥有全球最大规模的实时数据

第一，金融业拥有巨大的消费端实时支付数据。我国连续11年成为全球最大网络零售市场，电子商务交易总额由2018年的31.63万亿元增长至2022年的43.83万亿元。大规模的实时电子商务交易量，带动我国成为全球最大的电子支付国家。仅2022年，我国银行共处理2789.65亿笔电子支付业务，金额3110.13万亿元。其中，网上支付业务1021.26亿笔，金额2527.95万亿元；移动支付业务1585.07亿笔，同比增长4.81%，金额499.62万亿元；电话支付业务2.45亿笔，金额10.35万亿元。

第二，金融业拥有范围广泛的实时金融信用数据。截至2023年三季度末，金融机构人民币各项贷款余额234.59万亿元，同比增长10.9%。其中，短期贷款及票据融资余额53.4万亿元，同比增长6.9%，人民币房地产贷款余额53.19万亿元。本外币服务业中长期贷款余额61.79万亿元，同比增长13.3%，增速比上年末高2.1个百分点；前三季度增加6.09万亿元，同比多增1.49万亿元。



我国金融业拥有全球最大规模的实时数据

本外币住户消费性贷款（不含个人住房贷款）余额19.31万亿元，同比增长13%。普惠小微贷款余额28.74万亿元，同比增长24.1%，农户生产经营贷款余额9.15万亿元，同比增长17.8%。创业担保贷款余额2965亿元，同比增长10.1%。助学贷款余额1651亿元，同比增长23.6%。全国脱贫人口贷款余额1.14万亿元，同比增长12.2%。

第三，金融业可获得庞大的生产端实时数据。金融服务于实体经济高质量发展，推动实体产业数字化进程，随着数据成为新型数据要素，金融业在制造业信贷支持的同时，也获得了大量的实体产业和制造数据。比如，我国制造业规模已经连续13年居世界首位，2022年，我国制造业增加值占全球比重近30%，每天产出全球最多的实体产业、制造业产生的生产端数据。

然而，金融业海量的金融数据本身并不能同步带来商业价值，只有打通数据之间的护栏，利用生成式AI这样具有实时大数据分析和数据合成性能的新技术，在实时数据之间建立更多更快的连接，才能释放海量数据价值。

我国金融业生成式AI应用迎来政策红利期

落实中央金融会议精神和“十四五”规划纲要，金融管理部门统筹推进我国金融业数字化进程，多措并举鼓励金融科技创新与应用，提升金融服务实体经济效能。从政策视角，我国金融业生成式AI应用迎来政策红利期。

关于银行业保险业数字化转型的指导意见

2022年1月，中国银保监会办公厅发布《关于银行业保险业数字化转型的指导意见》，明确提出，“到2025年，银行业保险业数字化转型取得明显成效。数字化金融产品和服务方式广泛普及，基于数据资产和数字化技术的金融创新有序实践，个性化、差异化、定制化产品和服务开发能力明显增强，金融服务质量和效率显著提高。”

该指导意见覆盖了多项适合生成式AI的金融业应用场景。例如，在业务经营管理上，鼓励银保机构充分利用科技手段开展个人金融产品营销和服务，强化客户体验管理，增强线上客户需求洞察能力。在科技创新上，鼓励银保机构提高大数据分析对实时业务应用、风险监测、管理决策的支持能力，密切持续关注金融领域新技术发展和应用情况，提升快速安全应用新技术的能力。鼓励有条件的银行保险机构组织专门力量，开展前沿技术研究，探索技术成果转化路径，培育金融数字技术生态。



我国金融业生成式AI应用迎来政策红利期

同时，指导意见还提出，要坚持关键技术自主可控原则，对业务经营发展有重大影响的关键平台、关键组件以及关键信息基础设施要形成自主研发能力，降低外部依赖、避免单一依赖。鼓励科技领先的银行保险机构向金融同业输出金融科技产品与服务。在风险防控上，提出银保机构要防范模型和算法风险，建立对模型和算法风险的全面管理框架，制定管理制度，对模型数据的**准确性和充足性**进行交叉验证和定期评估。审慎设置**客户筛选和风险评估**等模型的参数，并使用压力情景下的参数进行模拟校验。定期评估模型预测能力及在不同场景下的局限性，确保模型的可解释性和可审计性。模型管理核心环节要自主掌控。加强消费者权益保护，防止算法歧视。

这些前瞻性的指导意见内容与时下的生成式人工智能创新发展密切相关，既为生成式AI技术供应商提供了可供参考的落地应用场景和安全框架，更重要的是为我国银行业保险业更好地集成生成式人工智能技术提供了可操作准则。

证券公司网络和信息安全三年提升计划（2023-2025年）

2023年6月，中国证券业协会印发《证券公司网络和信息安全三年提升计划（2023-2025）》，鼓励有条件的券商充分利用新技术积极推进新一代核心系统的建设，开展核心系统技术架构的转型升级工作，并明确提出，券商信息科技平均投入金额不少于2023年至2025年平均净利润的10%或平均营业收入的7%。

毕马威的调查报告显示，2022年，国内全行业140家证券公司合计实现营业收入人民币3950亿元，实现净利润人民币1433亿元。即便在不考虑增长的情况下，2023-2025年间全行业券商信息科技投入将不少于430亿元，其中人工智能技术所获得的投资规模将不少于300亿元。





我国金融业生成式AI应用迎来政策红利期

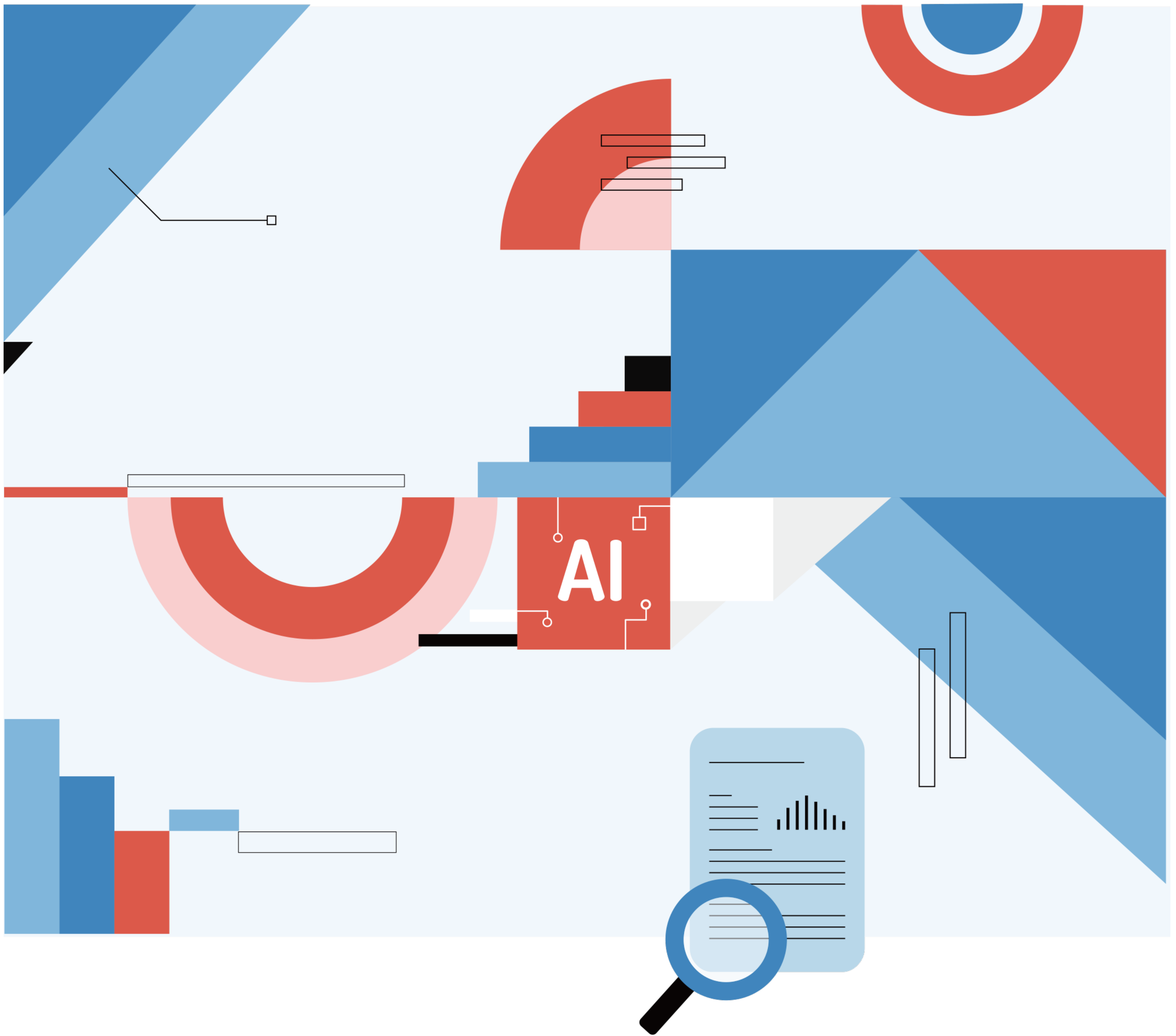
“数据要素×”三年行动计划（2024–2026年）

2023年12月31日，国家数据局会同国家金融监管总局等17个部门共同印发了《“数据要素×”三年行动计划（2024–2026年）》，推出“数据要素×金融服务业”等12个重点行动，目的是“到2026年底，打造300个以上示范性强、显示度高、带动性广的典型应用场景，培育一批创新能力强、成长性好的数据商和第三方专业服务机构，形成相对完善的数据产业生态，数据产业年均增速超过20%，场内交易与场外交易协调发展，数据交易规模倍增，推动数据要素价值创造的新业态成为经济增长新动力，数据赋能经济提质增效作用更加凸显，成为高质量发展的重要驱动力量。”

我国具有全球最大的智慧城市建设规模，产出全球最多的城市活动数据，“数据要素×金融服务业”行动明确支持金融机构融合利用科技、环保、工商、税务、气象、消费、医疗、社保、农业农村、水电气等数据，这为金融机构带来规模可观的城市生活增量数据。海量的数据利用需要有功能强大的数字技术支持，“三年行动计划”为生成式AI在金融业中的应用提供安全许可下的重要政策支持。

小结

在现行政策范围内，通过生成式AI等新技术对海量实时数据的计算处理，金融机构可以进行更多的产品创新、为实体产业提供更有有效的信贷支持、承保理赔、市场融资，释放巨大金融数据价值，提升实体经济金融服务水平，也促进金融机构自身的创新发展，迈向强大的金融机构行列。



PART 5

行动指南 | 金融业集成生成式AI需遵循的八项原则

经过过去一年的大炼模型，我国已经发布了200多款大语言模型，因此对于金融机构来说，大语言模型比较容易建立，可以直接购买、共同开发或自主研发，目前主要的挑战在于如何以最大的影响和最小的风险应用它们。综合已有的调查研究及国外一些成功的应用案例，我们提出国内金融机构在2024年采用生成式人工智能行动时需要遵循的八项原则建议。

1

原则一，正确认识生成式AI的作用域

生成式人工智能应用广泛，但并非全能，创造性产品创新任务是生成式人工智能的专业领域，利用这项技术能带来60%的效率提升。如果是带有大量不明确信息的复杂业务问题，这些自然不属于生成式AI的专业领域，一些应用案例显示，此种情况下的实际效率比没有使用该技术时还降低23%。因此，金融机构在制定生成式人工智能行动时，首先要正确认识这项技术的作用域，在全面客观地评估机构内部业务创新属性以及任务类型的基础上，将生成式AI工具更多地嵌入重复搜索和数据量大的任务中，并保持生成式AI与传统决策式AI工具的结合，不要追求通过生成式AI带来100%的自动化。

2

原则二，模型参数并非越大越好

模型参数规模越大，意味着神经网络的复杂度越高，学到的特征表达能力更强，拟合能力更好，评估的分数越高。但在实际应用中，大模型的利用效率并不高，普遍低于50%的利用率。过于庞大的模型可能难以在有限的计算设备上部署和训练，从而限制了其应用场景。大规模参数的模型在训练数据上表现优秀，但在新任务和未知数据上表现不佳，无法很好地泛化。此外，参数众多，导致大模型内部结构复杂，解释性较差。

因此，在实际应用中，金融机构需要根据自身数字成熟度、资源可用性和终端生成式AI应用程序的复杂性来选择规模适当的大模型。此外，技术研究层面也在探索如何通过模型压缩、剪枝、量化等方法，在保持模型性能的同时，降低模型规模，小型化大模型，以实现在更多场景下的应用，提高大模型的利用效率。

3

原则三， 技术先进性和业务可持续性选择合作伙伴的两个关键标准

技术先进性包括大模型的算法、训练数据、模型调优等多个方面，重点关注安全与合规、准确性和质量、定制化、定价和部署、提示词限制等关键指标。合作伙伴应具备强大的技术研发能力，能够持续优化大模型，以满足不断变化的业务需求。业务可持续性包括合作伙伴的业务模式、服务范围、客户支持、生态建设等多个方面。合作伙伴应具备良好的商业信誉和稳定的财务状况，能够为客户持续提供高质量的服务和支持。

在实践应用中，金融机构的长期战略合作伙伴通常也是生成式AI新技术的合作伙伴成员，新兴技术的开拓者几乎是大多用户首选的合作对象，比如，生成式AI技术引领者——OpenAI就是各行各业和大多数金融机构的首选商用合作伙伴。

4

原则四，

生成式人工智能技术的开发者实际上也是重要的监管者

金融业的一大核心功能就是风险防控和管理，因此金融行业大模型的开发者实际上肩负着重要的监管职责。一方面，这些开发者不仅需要推动大模型技术的创新与发展，提高生成式人工智能技术的透明度和可解释性，还要关注潜在的网络安全新风险和道德伦理问题，遵循数据隐私、知识产权、信息安全等众多法律法规，确保技术应用符合社会规范和法律法规。另一方面，证券、银行等金融监管部门也是大模型技术的重要使用者，这也需要大模型技术开发者对严格的金融监管法规有深入认识和理解。因此，金融机构在自主研发或共同研发大模型技术时，就要强化法务部门的作用和跨部门协同。

5

原则五，

着眼于端到端解决方案，而不是让它成为一个点解决方案

端到端解决方案是指通过整合各个部门或环节的业务流程，消除信息孤岛和部门墙。相较于仅在部门内部提供使能的点解决方案，端到端的解决方案能提高企业内部效率并满足客户需求，从而实现企业效益最大化和整体效能的提升。将生成式人工智能应用到金融业的端到端解决方案，更能充分发挥其在银行和保险业务流程效率提升、客户体验优化、风险管理、协同创新方面的潜力，消除承保等流程中的瓶颈和冗余，实现业务环节的无缝衔接，突破传统思维局限，并推进全员创新，挖掘新价值，为金融机构带来全面的变革和提升。

6

原则六，

根据实际情况确认应用的优先级顺序

金融机构对其客户非常了解，但也受到严格的监管，并且总是受到准确性、安全性和隐私标准的审查。目前，生成式AI在银行业的机会并不是面向消费者的工具，它更多的是关于如何综合银行现有的大量信息，做出更明智的决策，提供非常个性化的客户交互，以及快速自动化许多核心银行流程。金融业应用大模型技术尚处于探索应用的早期，几乎所有金融机构都在同一起跑线上，但这不是一条容易的赛道，因为有太多的应用场景。依据自身业务需求，金融机构可以用它作为内容创造者，可以用它来获得更好的资源管理，可以用它作为后台解决方案。重要的是，在各项任务场景中，金融机构需要确定生成式AI的优先级。例如，在一般办公及行政场景中，有些银行就确定了从自动执行重复性任务、改进文档处理、增强决策洞察力、简化沟通效率、到确定战略举措的优先顺序。在产品/研发场景中，他们确定了生成式AI工具从加速构思过程到改进需求管理、增强代码生成、优化知识管理、改善团队协作的优先级应用流程。



7 原则七， 把大模型带到数据中心环境中运行， 而不是把数据带到大模型所在的地方

将大模型带到数据中心环境中运行，进行本地化部署是金融业落地应用大模型技术的通用路径。国内银行、保险、证券和基金等金融机构的数据中心通常具备更强大的计算和存储资源，可以更好地支持大模型的运行和计算需求。将大模型部署在数据中心环境中，可以更好地利用这些资源，提高计算效率。数据中心通常具备严格的安全措施，包括物理安全、网络安全、数据备份等，可以有效保障数据的安全性和隐私性，确保数据在安全和受控的环境中运行。此外，数据中心通常具备更好的网络连接和数据传输能力，可以确保大模型在运行过程中能够快速、稳定地访问和处理数据，提高数据的可用性和处理效率。

8 原则八， 由首席执行官直接牵头指导负责任的人工智能治理框架

大语言模型的可解释性和数据泄露的风险处理是金融业高管们在探索生成式AI应用场景时高度关注的两大问题。当前立法仍未明朗，我们认为，金融机构自身需要建立负责任的人工智能治理框架，在现有监管可行范围内，利用自己的管理原则来处理人工智能解决方案，包括监控需要解决的技术和工具，设置自己的护栏和保障措施。

负责任的人工智能战略应该是开发各种人工智能和生成式人工智能解决方案的前沿和中心。我们关注到拥有负责任的人工智能框架的金融机构，通过应用不同的生成式人工智能工具，实现更大的业务影响。每当这些负责任的人工智能治理计划由首席执行官直接驱动时，这些计划在组织内产生的影响大约会增加60%，因为战略内容更清晰、战略执行更有力、跨部门协同更快速。

结语

回顾2023年，虽然全球经济复苏仍然缓慢，但是我们见证了一个充满活力、创新及转型的大语言模型时代的兴起。生成式人工智能技术在各个关键领域的拓展彰显了其深远影响和巨大潜力，已经为一些银行、保险和基金等金融机构赋予了创造力和效率的新维度。然而，在享受新质生产力带来的高效与便利之余，我们还需正视所面临的现实挑战，尤其是在数据安全、风险防控、道德和监管方面。

展望2024年及未来，积极主动地拥抱生成式人工智能这项变革性技术，并着手负责任的人工智能治理策略显得愈发重要，这将使社会得以充分利用生成式人工智能的变革力量，更好增进人类福祉。

本研究报告旨在提供创新逻辑、行业洞察、政策期望，在此基础上展示了金融机构在集成生成式人工智能时需要注意及遵循的八项原则建议，以期助力更多的金融机构快速掌握和使用生成式人工智能技术，加快推动我国金融业高质量发展。

版权声明

本报告由清华大学经济管理学院联合度小满科技（北京）有限公司、《麻省理工科技评论》中国共同发布，其版权归属三方所有，并对此报告拥有唯一著作权和解释权。没有经过上述机构的书面许可，任何组织和个人不得以任何形式复制、传播等。任何未经授权使用本报告的相关商业行为，清华大学经济管理学院联合度小满科技（北京）有限公司、《麻省理工科技评论》中国将依据中华人民共和国相关法律、法规追究其法律责任。

免责声明

本报告所载数据和观点仅反映于发出此报告日期当日的判断。我们对报告所载信息的准确性、完整性或可靠性做尽最大努力的追求，但不作任何保证。在任何情况下，本报告中的信息或表述均不构成任何投资等建议，发布方对该报告的数据和观点不承担法律责任。不同时期，发布方可能会发布其它与本报告所载资料、结论不一致的报告。同时发布方对本报告所载信息，可在不发出通知的情形下做出修改，读者应自行关注。

关于清华大学经济管理学院

清华大学经济管理学院（以下简称“清华经管学院”）成立于1984年，朱镕基担任首任院长（1984-2001），赵纯均担任第二任院长（2001-2005），何建坤担任第三任院长（2005-2006），钱颖一担任第四任院长（2006-2018）。学院现任院长是白重恩。学院以“创造知识，培育领袖，贡献中国，影响世界”为使命，以“成为世界一流的经济管理学院”为愿望，以“正直诚实、敬业尽责、尊重宽容”为核心价值。40年来，清华经管学院在人才培养、科学研究、社会影响、国际交流等方面保持国内领先水平，并为建设世界一流的经济管理学院而不断努力。

2007年至2008年，清华经管学院先后获得国际商学院联合会（AACSB）管理教育认证、AACSB会计教育认证、欧洲管理发展基金会（EFMD）的EQUIS认证，成为中国内地率先获得AACSB和EQUIS两大全球管理教育顶级认证的商学院。

清华经管学院现有会计系、经济系、金融系、创新创业与战略系、领导力与组织管理系、管理科学与工程系、市场营销系等7个系，覆盖理论经济学、应用经济学、工商管理、管理科学与工程等4个一级学科。全国工商管理专业学位研究生教育指导委员会秘书处设在清华经管学院。

2000年10月成立的清华大学经济管理学院顾问委员会由海内外知名企业家、商学院院长、知名学者（包括诺贝尔经济学奖获得者）以及我国政府及财经部门领导人组成。学院首任院长朱镕基担任顾问委员会名誉主席。高盛集团时任首席执行官亨利·保尔森先生（Henry Paulson）担任顾问委员会首任主席（2000-2003），英国BP集团时任首席执行官约翰·布朗勋爵（Lord Browne of Madingley）担任第二任主席（2003-2008），沃尔玛百货公司时任首席执行官李斯阁先生（Lee Scott）担任第三任主席（2008-2010），华盛顿布鲁金斯研究院时任理事会主席约翰·桑顿先生（John Thornton）担任第四任主席（2010-2013），凯雷投资集团联合创始人兼联席首席执行官大卫·鲁宾斯坦先生（David Rubenstein）担任第五任主席（2013-2016），Breyer Capital创始人兼首席执行官吉姆·布雷耶先生（Jim Breyer）担任第六任主席（2016-2019），苹果公司（Apple）首席执行官蒂姆·库克先生（Tim Cook）担任第七任主席。顾问委员会的使命是加强学院与外部各界的联系，提升学院在科研和教学方面的整体水平，帮助学院建设成为世界一流的经济管理学院。

关于度小满科技（中国）有限公司

度小满科技（中国）有限公司（以下简称“度小满”），原百度金融。2015年12月，百度整合旗下金融业务，成立金融服务事业群组(FSG)，交由百度高级副总裁朱光全面负责，金融科技与无人车、DuerOS成为百度AI战略的三大重要赛道。

2018年4月，百度宣布旗下金融服务事业群组正式完成拆分融资协议签署，拆分后百度金融启用全新品牌“度小满”，实现独立运营。2018年5月21日，度小满正式成立，延承百度技术基因，探索人工智能在金融领域的应用，先后布局信贷、财富管理、支付、保险、个人金融科技和供应链金融科技六大业务板块。

度小满坚持客户至上，通过科技来服务小微。基于人工智能、金融云、区块链和物联网等底层架构，度小满搭建了金融大脑，运用情感计算、关联网络和知识图谱等技术，进行感知引擎和决策引擎运算，大数据处理能力实现日均100亿次计算，智能风控模型年均上线10万条策略，通过对征信报告进行深度解读，识别出40多万维风险指标，将信贷风险降低了25%，RPA机器人替代信贷审核中六成以上重复性工作，无感率达99.2%。度小满将领先的技术实力应用到智能营销、智能风控、智能经营、智能运营和满链融，精准识别小微风险特征，用科技帮助小微降低融资成本。

2023年2月，度小满宣布成为百度文心一言（英文名：ERNIE Bot）首批生态合作伙伴。度小满将基于自身金融场景积累的海量对话及解决方案数据，融合“文心一言”的全面能力，打造全新的智能服务。这标志着对话式语言模型技术在国内金融场景的首次落地应用。

关于《麻省理工科技评论》中国

《麻省理工科技评论》 (*MIT Technology Review*) 依托麻省理工学院的学术和产业资源，于 1899 年在美国麻省理工学院创刊，是世界上历史悠久的科技商业智库。自成立之初，《麻省理工科技评论》就一直关注那些正在颠覆现有格局并创造新的市场机会而影响人类社会的技术，以及那些正在从实验室走向市场即将商业化的技术。在此基础上，也高度关注将这些技术落地并用这些技术影响我们生活的人和聪明企业。

《麻省理工科技评论》于2016年落地中国，由DeepTech独家运营，开展媒体、研究及会议业务，围绕技术话题辐射和影响新兴科技圈层，重点关注新兴科技的商业化和社会价值，聚焦中国市场，为中国科技从业者带来与全球百万科技领域研究者、从业者及商业领袖进行前沿科技国际化交流的机会。

关于清华大学经济管理学院动态竞争与创新战略研究中心

清华大学经济管理学院动态竞争与创新战略研究中心成立于2019年。中心以经世济民的智慧追寻中西创新之道为使命，提升在动态竞争和创新战略领域的学术和社会影响力。作为立足中国放眼全球的新型智库，中心旨在融合贯通全球战略管理领域的学术、教学与企业实践，服务于中国以及全球华人企业和管理学者和学生。中心在协助政府客观决策、提升企业自身认识、帮助管理学者理论创新、引导学界业界理论联系实际等方面起到重要作用；对清华经管学院，及至清华大学在企业战略和创新创业管理相关的学科建设、科研水平、科研成果转化和人才培养等方面具有重要的推动作用。中心力争培养一批融汇中西、理论结合实际的国际化高端人才，保持在动态竞争和创新战略领域的国际领先地位。

中心主要对动态竞争理论和数字化创新进行深入研究，通过融合动态竞争与创新战略来发表多篇论文、出版系列著作、编写相关案例，并分析企业最佳实践。中心自成立以来，协办参与每年的动态竞争国际论坛和中国管理学者交流营的活动，提升全球华人管理学者社群的影响力。中心还积极举办“全球化发展与企业数字化转型研讨会”等学术会议。

联系我们

如您希望与我们交流或有任何问题，请与我们联系：research@deeptechchina.com

