腾讯金融研究院
Tencent Financial Research Institute

腾讯研究院
Tencent
Research Institute

KPMG

2025金融业大模型应用报告

体系落地　价值共生

腾讯金融研究院 | 腾讯研究院 | 毕马威企业咨询（中国）有限公司

2025 金融业大模型应用报告

体系落地 价值共生

序言

司 晓

腾讯集团副总裁、腾讯研究院院长

过去两年，席卷全球的大语言模型浪潮，正式拉开了生成式 AI 时代的宏大序幕。步入 2025 年，大模型正从聚光灯下的明星技术，沉淀为驱动社会运行的智能基础设施，并以“马拉松”般的耐力，深度重塑着产业与经济的血脉。

一方面，对技术极限的探索仍在加速。国内外头部科技公司不断推出的新一代 SOTA（State of the Art）模型，正合力将基础大模型的性能推向全新高度。另一方面，产业界的重心已转向对应用生态的精心构建，以此承载各行各业向 AI 转型的宏伟蓝图。

这远非零散试点或工具集成所能企及，它要求我们像建设工业时代的电网、信息时代的光缆一样，进行系统性的规划与投入。这不仅是一场技术革命，更是一场涵盖数据基建、组织形态、信任机制乃至社会伦理的全维度重构。

金融业，作为现代经济的神经中枢，是这场重构的核心战场与先导力量。在这里，AI 不再仅仅是专家能力的“放大器”，更是与人类智慧深度耦合、互补协作来探寻金融服务本源的伙伴。我们观察到，一种新的协作范式正在诞生：技术供给与真实需求双向奔赴，在解决最棘手的金融挑战中协同进化；科技、金融等多元主体，也正携手构建一个开放、共建、共享的创新共同体。

作为这场技术浪潮的深度思考者与躬身入局者，腾讯研究院希望《2025 金融业大模型应用报告》能够超越一份常规的分析报告。报告不仅剖析变革的机制与路径，更力求探寻其背后的时代脉搏与产业逻辑，为金融机构提供具有前瞻性的战略思考框架与切实可行的实践路径指引，共同探索由 AI 驱动的崭新金融时代。

序言

柳晓光

毕马威变革咨询数字化转型业务牵头人

“智慧之光”数智解决方案主管合伙人

我们正处在一个由大模型定义的革命性时代。2025 年，已是技术浪潮与产业变革交汇的关键之年。AI 正以超越过往任何技术的速度与深度，从数字世界的底层逻辑，重塑着全球经济的宏观格局。

金融，作为现代经济的核心，其本质是信息的处理与风险的定价。这恰好与大模型强大的认知、推理及生成能力，形成了前所未有的共振。这股力量，正推动金融业开启继数字化与移动化之后，一场更为深刻的智能化变革。它不仅是效率工具的迭代，更是对金融服务范式、运营模式乃至核心竞争力的系统性重塑。

本报告旨在为这场波澜壮阔的变革提供一张清晰的导航图。我们将从宏观视野出发，系统梳理模型、算力与数据的演进趋势。随后，将镜头聚焦于金融业本身，提炼各机构从审慎探索到加速布局的应用全景，并揭示其应用价值从提升运营效率向赋能核心决策的跃迁路径。然而，我们深知通往智能金融的道路并非坦途。因此，报告将直面数据孤岛、战略模糊、安全合规、人才短缺等核心挑战，并通过深度剖析全球领先的实践案例，力求提供兼具前瞻性与可操作性的应对之策。最终，本报告将落脚于未来。我们提炼出驱动行业演进的六大核心趋势，希望能为身处变革中的每一位金融决策者、创新者和从业者，提供一个思考未来、把握当下的战略罗盘。我们相信，一个更普惠、更个性化、更高效的智能金融新纪元，正由我们共同开启。

报告总览

2024 年至今，一场由大模型驱动的生产力革命正在金融业上演：一家领先大行将过去需要数小时甚至数天完成的复杂信贷审批报告分析压缩至 3 分钟，准确率提升超 15%；一家头部券商借助 AI 智能体实现 7X24 小时监控全球超过 5000 家上市公司的动态，研究覆盖面和响应速度达到了全新量级；一家海外顶尖投行部署了数百个 AI 程序员，后续或增至数千个，旨在将工程师的生产力提升至三到四倍。

2025 年 3 月，中国人民银行已明确要求加快金融数字化智能化转型，安全稳妥有序推进人工智能大模型等在金融领域应用。

务实的政策指引与激动人心的行业进展，共同表明行业正进入一个生成式 AI 引领的结构性变革期。我们判断，2025 年已成为金融行业深度整合 AI、借助大模型进行创新的关键拐点。本报告旨在穿透技术热潮，深入剖析大模型在金融业应用的现状与未来，为行业提供前瞻性的洞察。我们坚信，那些能够深刻理解机遇与挑战、进行前瞻布局、并致力于体系化能力建设的机构，必将在本轮智能化浪潮中获得发展先机，共同塑造金融服务的未来。

本报告的核心观点如下：

一、技术边界极速拓展，从能力延伸到效率革命

全球大模型的发展已非单一的技术竞赛，而是呈现出技术迭代、资源升级、价值深化与生态竞合交织并进的复杂格局。模型的演进方向正从探索能力边界转向追求效率革命，算法与架构的持续优化在不断重新定义性能天花板。与此同时，算力需求呈现更重视推理的结构性变化，数据训练的关注点，正从单纯追求海量规模，转向更加倚重高价值的精准数据。在应用场景上，大模型也正从提效工具升级为协作伙伴，以智能体为代表的应用正在重构人机协作的全新形态。

二、应用模式日趋成熟，从试验阶段到规模化部署

全球近半数金融机构已启动大模型应用建设，行业正从零星的试验阶段迈入规模化部署期。中国金融业的大模型建设呈现出顶层设计、梯次推进的清晰格局：银行业是大模型落地应用最广泛的领域，证券、保险行业的头部机构则作为先行者，探索出多样化的应用模式。当前，应用建设的路径正逐渐收敛至以实际效益为导向，围绕能力建设、基座构建、应用部署的三大策略日益清晰，应用版图也正从外围的效率工具向核心的决策层面审慎渗透。

三、落地挑战逐步明确，体系化能力成制胜关键

金融机构在实践中普遍面临着局部突破与整体效能的平衡，创新投入与资源效能的平衡，以及前沿探索与风险防控的三大平衡关系考验。面对高价值数据资源碎片化、战略规划和投资回报不清晰、低容错场景技术适配难、组织人才升级滞后等具体挑战，构建“数据 + 应用 + 战略 + 组织人才”四位一体的综合能力框架，将是赢得未来人工智能革命先机的关键。在场景侧，智能理财助理、财富管理、保险代理人、投研报告生成、编程助手等场景已率先实现商业化突破，金融智能体的探索和实践也在持续涌现，为行业提供了可复制的成功路径。

四、“ 金融 +AI”前景广阔，重塑金融服务未来

AI 技术正驱动金融服务走向前所未有的普惠化、智能化与个性化，将专家级专业服务带给更广泛的长尾客户群体。同时，AI 与人类专业能力的深度融合，正在重新定义金融的运营与管理模式，加速推动复合型、创新型金融人才的需求形成。在此进程中，高质量私域数据的挖掘与应用将成为金融机构的核心竞争力，而 AI 技术和治理体系的不断成熟，也将推动监管科技效率与效能的提升。

目录

近期，全球顶尖 AI 模型接连取得突破性进展，在编程、数学和视觉感知等领域树立了新的标杆，并显著减少了幻觉现象。尽管追求参数规模的“军备竞赛”仍未停止，行业已经逐步转向追求效率与价值的务实探索，大模型正从少数科技企业的专属技术，加速成为面向全社会提供智能服务的新型基础设施。

第一章

全球大模型发展态势

全球大模型发展态势

当前，全球大模型发展不再是单一维度的技术竞赛，而是呈现**技术迭代加速、资源结构升级、应用价值深化、生态竞合交织**四大趋势。从 DeepSeek-R1 等高效开源模型的涌现，到多模态与强化学习的融入，人机协作模式重塑，以及智算中心成为新型基础设施，我们正站在一个由技术范式革新驱动的产业变革的起点。

1.1 模型演进：能力边界进一步延伸

1.1.1 算法与架构的效率革命持续提升模型能力

基于规模法则（Scaling Law）[1]推动基础大模型性能提升的预训练模式的性价比下降。在此背景下，学术界和工业界不断探索后训练与特定场景的 Scaling Law。目前，Scaling Law 的影响已经扩展到后训练和推理阶段，推理模型的性能与训练时间计算、推理时间计算量存在明显的幂律关系。DeepSeek 的成功经验引发全球关注，但并未改变大模型对大算力的依赖。目前，业界的探索表明，通往更高智能存在两条并行路径、两者相辅相成：一是通过更优的算法提升算力效率，比如强化学习（Reinforcement Learning）和测试时计算（Test-time Compute），二是通过持续提升模型尺寸、扩充训练数据与加码算力投入，构筑顶级基座模型的竞争壁垒。例如，xAI 为 Grok 3 投入的预训练算力达到了 Grok 2 的 10 倍，Grok 3 调用了 10 万个英伟达 H100 芯片，相较于 Grok 2 的 15000 个实现了显著提升。**而这场围绕算力的军备竞赛远未结束，作为行业标杆的 GPT-5 的发布，再次印证了构建最顶级基础大模型的成本投入极为高昂，注定是战略性稀缺资源。**

业界将目光从预训练转向即后训练与推理阶段，强化学习正是这一趋势的核心技术。以 DeepSeek-R1 为代表的强化学习架构，通过让模型在试错中学习，用更少的计算量激发了模型更深层次的推理能力。其核心创新的 GRPO 算法消除了传统 Critic 网络的计算冗余，不仅显著提升训练效率，而且将推理延迟降低至毫秒级。目前，诸多团队基于类似的训练策略，持续验证在小模型上的推理能力，结果表明额外的指令微调并非必要，基础模型和指令模型最终能达到相似的性能水平；不同的强化学习算法都能实现长思维链的涌现；通过精心设计的强化学习方法，即使是较小的模型也能实现强大的推理能力，而且这个过程可以比传统方法更简单、更经济。自 2024 年 9 月 OpenAI o1 系列模型发布后，这一强化学习范式已逐步被主流模型采纳。

融合强化学习等技术的后训练，大幅提升了模型进行更深入、更复杂的推理的“慢思考”能力。通过强化学习驱动的多轮追问与假设检验，大模型在数学推理等任务中展现出卓越性能，并为复

1　规模法则（Scaling Law）是被业界认为是大模型预训练第一性原理，也是在机器学习领域，特别是对于大语言模型而言，模型性能与其规模（如参数数量）、训练数据集大小以及用于训练的计算资源之间存在的一种可预测的关系。这种关系通常表现为随着这些因素增长，模型性能会按照一定的幂律进行改善。

杂产业场景的优化提供了新的思路。未来，大模型的逻辑推理能力将得到强化，能够处理更复杂的逻辑关系及推理任务，例如演绎推理、归纳推理以及溯因推理。例如，Grok 3 通过强化学习推理方式获得了“慢思考”能力，成功跻身第一梯队；2025 年 7 月发布的 Grok 4，在后训练强化学习方面的计算量较 Grok 3 提升了 10 倍。这体现了大型科技公司对深度推理能力的投入持续升级。

推理新范式的出现促进行业专用基础模型崛起。行业专用基础模型正在各个领域蓬勃发展，这些模型针对特定行业数据和任务进行训练和优化。与通用基础模型相比，它们在处理特定行业任务时表现更加出色。与去年相比，当前行业专用基础模型的发展呈现出两大趋势：一方面是更便捷的开发流程。在当下后训练与推理新范式下，得益于强化学习等技术，现在开发者只需使用少量经过标注的数据，就能快速构建出实用的行业专用模型，极大地降低了开发门槛和成本。另一方面是更多样化的模型类型，除了传统的语言处理模型，现在也包括多模态模型、图神经网络和物理信息神经算子等。

混合专家模型架构（Mixture of Experts, MoE）已成为大模型追求卓越性能的主流技术路径之一。其核心优势在于解耦了参数规模与计算成本，突破了传统稠密模型的扩展瓶颈。MoE 并非要替代 Transformer 中的自注意力等基础模块，而是通过更高效的组织与计算方式，极大提升了模型的扩展效率。业界领先者已纷纷采纳此路线。2024 年年初，腾讯混元就在国内率先采用 MoE 架构模型。其旗舰模型混元 TurboS 创新采用了前沿的混合线性注意力机制与 MoE 模型架构，是大模型研发前沿分支的重要代表。Kimi-K2 在 DeepSeek V3 架构基础上，通过将注意力头减至 6 个、MoE 专家数翻倍至 128 个（每次激活 8 个）的策略，实现了计算效率与知识容量的同步提升。长远来看，MoE 代表了模型设计从同质化的全量计算向异质化的条件计算的理念转变。这种转变使得研究者可以探索参数数量远超现有稠密模型极限的架构，而不必承担同等比例增长的计算开销，为实现模型能力的持续扩展提供了基础。

1.1.2 模型能力维度从数字世界拓展到物理世界

多模态感知和生成能力是大模型与真实世界交互的关键，也是未来发展的重要趋势。多模态大模型突破了单一文本的限制，实现了跨模态信息处理与理解。在图文交互领域，大模型可以根据图像生成精准的文字描述，或者根据文字指令创作匹配度较高的图像，例如根据文字描述生成设计图稿。在视听融合层面，大模型通过视频内容识别、字幕生成和关键信息提取，为影视创作、安防监控等场景提供智能辅助，例如自动生成视频摘要、识别监控视频中的异常行为等。尽管多模态技术发展迅速，但仍面临着一些挑战，例如跨模态的语义协同、正负样本多模态数据量不足等。随着技术的进步，未来有望实现多模态的深度融合，开启更多交互和创作的可能性。

空间智能的加速发展，正推动人工智能与物理世界进行深度交互。一系列关键技术，如三维感知、空间表示与生成以及多模态融合技术逐步成熟，将赋予 AI 感知、理解并最终行动于物理世界的能力。实现这一目标的关键，在于强大且易于使用的世界模型。腾讯近期开源的混元 3D 世界模型，作为业界首个兼容传统 CG 管线的可漫游 3D 世界生成模型，通过技术优化，成功实现了在消费级显卡上的流畅运行，极大地降低了 3D 世界生成的门槛，并已在游戏开发、VR 体验及数字内容创作等领域展现出加速行业演进的潜力。这种基础能力的成熟与普及，将为具身智能机器人等前沿应用提供关键支撑。预计在未来 2-3 年内，机器人将能够完成数十项实用的复杂功能，并通过技术迭代与成本优化将任务能力扩展至成百上千种。

1.2 算力与数据：从追求规模到优化结构

1.2.1 算力需求结构性变化催生智算中心等算力基础设施

DeepSeek 的突破再次引发了算力的“杰文斯悖论”[2]：大模型推理效率提升，不仅不会降低算力需求，应用得到推广还将带来整体算力需求保持扩张。随着智能体应用的增加，将推动数据处理量的指数级增长，进而引发对推理算力的巨大需求，甚至可能超过训练算力需求。一方面，随着模型部署成本的大幅降低，中小企业和边缘计算场景也能接入 AI 应用，带动算力资源调用速率突破线性增长规律。同时，实际应用场景对实时推理任务的需求爆发式增长，进一步加剧了对算力的需求。另一方面，领军企业对更大参数规模模型的持续研发，也巩固了算力需求的不可替代性。

为应对日益增长的算力需求，智算中心作为新型 AI 计算基础设施正在兴起。AI 计算基础设施正在从单一 GPU 集群向综合智算中心转变，这些中心整合了计算、存储、网络和冷却系统，为各种 AI 工作负载提供可扩展解决方案。尤其是进入到推理范式以及多智能体阶段，分布式算力的需求呈指数级增长。这一趋势推动了液冷技术、高带宽内存和专用互连网络的创新，以支持日益增长的计算需求。与此同时，分布式算力架构通过源网荷储一体化[3]创新，成功实现能效跃升。据新华网 2024 年 12 月 24 日报道，上海崇明岛北堡风电场部署的分布式算力节点，依托风电直供技术使年运营成本降低 70 万元，碳排放年减少 850 吨，验证了新能源与算力深度融合的可行性。

与此同时，轻量化模型的普及，也推动算力资源配置格局从集中式超大规模集群逐步向分布式、多点协同的方向发展演变。DeepSeek 打破了传统 AI 的规模壁垒，其轻量化模型与开源策略降低了 AI 应用门槛，促进了中端算力设施和分布式数据中心的普及。产业价值链条呈现结构性调整：上游，国内芯片企业获得关键发展窗口期；中游，区域化数据中心利用响应速度优势对接产业智

2 “杰文斯悖论”通常指资源利用效率提高导致总消耗量增加的经济学现象。

3 源网荷储一体化是指将能源源头（如光伏、风电等）、电网、用电负荷和储能系统有机地整合在一起，形成一个综合性的能源系统，以实现能源的高效利用和优化能源供应与需求的平衡。

能化需求；下游，人工智能与细分领域的深度结合，推动技术升级与商业价值形成互促的良性循环。

1.2.2 数据训练从海量数据驱动转向高价值知识驱动

大模型对高价值数据的依赖远超传统算法，训练从简单的数据堆砌转向对数据的价值锚定，目标是将数据转化为可被大模型有效学习的知识。高价值数据集通过价值锚定化、知识显性化和演进动态化，实现人工智能从通用能力到垂直场景业务效能的精准转化。价值锚定化方面，聚焦对模型训练真正有价值的数据，例如在工业质检场景中，设备异常的频谱特征数据价值远高于正常运行数据。知识显性化方面，这种数据集不再是简单的信息堆砌，而是将隐含在数据中的知识提取出来，以更易于模型理解和学习的方式进行呈现。演进动态化，则是根据模型训练和应用的反馈，动态调整和优化数据集，形成“数据 – 模型 – 业务”的迭代飞轮。这种范式正在重塑数据采集逻辑，企业需要从被动记录转向主动设计，将高价值数据集建设提升至战略高度。

随着大模型训练需求的指数级增长，真实数据与合成数据融合成为突破数据瓶颈的新路径。Epoch AI 研究公司预测，全球公共互联网文本总量预计将在 2028 年前后接近现有 AI 训练数据集规模，这意味着高质量训练数据枯竭的挑战可能在未来四年内爆发。面对这一形势，合成数据技术成为缓解数据短缺的关键突破口，并在高质量指令微调、复杂推理任务及多轮对话数据生成领域展现出独特的价值。以 OpenAI 开发 GPT-5 为例，研究团队尝试利用前代模型生成的数据来训练下一代模型，这种方法在理论上具有可持续性，然而实践中并未完全解决训练扩展性和数据瓶颈问题，效果提升也未达到预期。这一现象揭示了合成数据在应对数据稀缺性、隐私安全及极端场景建模需求虽有优势，但实际效能高度依赖生成算法的成熟度。当前阶段，合成数据与真实数据的动态配比、规模化生成规律突破等核心问题仍需关注。

跨模态数据的协同训练有利于推动模型智能水平提升。文本、图像、时序信号等跨模态的协同训练并非简单数据叠加，而是通过不同模态数据之间的语义对齐与信息补偿，构建更接近人类认知的全息理解框架，让模型像人类一样能够综合处理和理解来自不同感官的信息。同时，跨模态协同训练对数据治理提出新的要求：多源数据的异质特性催生标准化重构需求，而模态交叉带来的隐私风险则倒逼安全防护体系升级。未来，能够打通数据壁垒、掌握跨模态数据协同训练机制、并具备落地应用能力的企业，将在产业智能化变革中占据优势。

1.3 应用场景：提效工具逐步升级为协作伙伴

1.3.1 企业级应用强调人机协作下的价值创造

企业对于 AI 应用的态度已经更加务实——AI 是增强员工能力的协作伙伴，而非完全替代员工。在企业层面，企业正在将大模型集成到现有 AI 系统和业务流程，这种融合不仅仅是简单地

添加大语言模型接口，而是通过重新设计工作流程，使得大模型能够增强但非取代现有系统，从而创造更大价值。这种务实的态度也在大模型的应用场景选择上得到了体现：我国的大模型在与实体经济深度融合方面，应用场景正从 IT/ 互联网、通信、金融与能源逐步向医疗、物流、教育、制造等多个行业拓展。

检索增强生成（RAG）与私有知识库的结合，不仅推动了大模型从概念验证向企业级应用转变，也初步体现了人机协作的核心理念。具体而言，RAG 为大模型外挂企业私有知识库，即大模型负责提供强大的信息处理和生成能力，而企业员工则通过知识库提供专业知识、判断力和决策力。这样既能保障数据安全和解决幻觉问题，又能实现 AI 能力与人类专业知识的高效协同，成为企业级应用落地的关键技术。这种人机协作模式能够最大限度地发挥大模型的优势，同时避免其潜在风险，从而在创新发展的同时注重合规和安全。RAG 系统架构正朝着更复杂、更专业、更智能的方向发展：从单一文本检索向全媒体内容理解转变；从通用模型向高度领域专用的知识增强转变；从简单的检索管道向多阶段评估、验证和优化流程转变；从独立系统向端到端、云到边缘的分布式架构转变。

1.3.2 智能体（Agent）成为人机协作的重要形态

AI 应用的形态正从聊天机器人（Chatbot）向能够独立思考、调用工具、执行任务的智能体（Agent）演进。业界期待，未来的 AI 智能体还能发展成为长期自主运行、持续学习和适应能力的智能实体。这要求 AI 智能体突破长效记忆、复杂工具调用与协同、环境感知以及多智能体协作等多个技术，以独立胜任高动态性任务。将知识库和推理能力融入大模型，可以显著提升智能体作为大模型企业级应用在感知、分析、决策和执行方面的智能化水平。

当前，单一智能体已有小规模试点，通过效率支撑、流程赋能与决策辅助来完成体系融合；而多智能体的协作能力远超单一智能体，在解决复杂问题方面成为有潜力的 AI 应用。智能体的首次革命完成了从指令执行工具到问题解构主体的转变，核心突破在于思维链的引入。当前技术已进入二次革命阶段，表现为多智能体协同系统的认知涌现，即通过辩论机制、置信度加权、不确定性校准等技术完成系统性协同行为。目前，在数学推理等复杂任务中，多 Agent 协同系统的准确率相比单模型提升 23-45%，这种协同并非简单的投票机制，而是通过动态调整注意力权重形成知识合成的新范式。xAI 发布的 Grok 4 Heavy 多智能体模型，在标准版 Grok 4 结合工具在 HLE 测试（Humanity's Last Exam）中取得 38.6% 的准确率后，Grok 4 Heavy 模型通过并行启动多个 Grok 4 实例，并采用内部协同与投票机制输出结果，准确率提高到 44.4%。展望未来，智能体的第三次革命将延伸至具身物理空间中的群体博弈，逐步演进成可信空间下的群体智能与演化博弈循环。当智能体深度介入决策流程时，确保其行为符合伦理规范、避免算法偏见、维持决策透明度、保障数据合规使用成为关键命题。这要求技术创新与治理体系同步进化，构建

既能防范技术滥用又可明晰责任边界的监管框架，在推进智能体落地的过程中实现技术效能与社会价值的平衡。

1.4 产业生态：技术竞赛与生态竞合并立

1.4.1 开源与闭源的路线之争过渡为生态共存

大模型开源与闭源的博弈已超越技术路线之争，成为企业战略选择、生态主导权争夺以及全球治理规则重构的核心场域。

高性能和低成本开源模型的崛起，正在瓦解传统依赖算力与资金垄断的模型开发格局。过去依赖巨额资金、千亿参数和超算资源的开发模式，逐渐被低成本、高效率的开源模式所冲击。企业竞争焦点从技术单点突破转向生态整合能力，例如通过开源社区构建开发者粘性，再通过云服务、API 接口实现商业转化。这种模式在削弱巨头垄断的同时，也加剧了生态碎片化风险，例如不同开源协议间的兼容性问题可能阻碍技术规模化落地。

开源模型加速技术迭代、推动长尾场景应用普及，但闭源模型在稀疏激活、多模态对齐等底层技术上仍具优势，并有利于企业维持技术代差。闭源模式的另一优势在于可控性，更易满足数据隐私和伦理合规要求。然而，过度封闭可能扼杀创新活力，表面开源但实际存在使用限制或关键组件未开放的“伪开源”做法也遭部分开发者抵制，凸显了开源社区对透明性的强烈需求。

当前大模型领域正在呈现开源与闭源并存的多元化格局。Meta、xAI 等企业通过开源 Llama 3.1、Grok 1 等模型主张透明性和可定制性。部分科技企业选择分层开源的平衡策略，开放中小模型构建生态，保留顶级模型巩固壁垒。例如，Google 开源 Gemma 小模型系列吸引开发者，同时保持 Gemini 大模型闭源以维持技术领先。OpenAI 等传统闭源领军者也在重新审视策略。2025 年 8 月，OpenAI 推出首批开源模型 gpt-oss 系列，从纯闭源向“闭源 + 开源”转变，开源模型支持在笔记本和手机等端侧场景运行。Anthropic 仍坚持闭源路径以维护核心竞争优势。

随着分层开源成为一种务实的策略，如何平衡开放性与商业价值，如何建立统一的开源标准和评估体系，以及如何构建安全、可信的开源生态，成为推动大模型技术普惠和产业繁荣的关键。未来，未来开源和闭源将长期共存、相互促进，并推动行业标准形成，进一步规范技术发展路径。可信的开源或将成为平衡安全与创新的关键路径，随着相关监管条例与备案机制的完善，对闭源模型的安全性与伦理风险的评估也将更加审慎与全面。

1.4.2 主权 AI 成为影响全球科技发展的重要变量

提升 AI 技术的自主可控水平，正成为越来越多国家在科技战略布局中的优先考量。英伟达创始人兼 CEO 黄仁勋在迪拜世界政府峰会提到，国家应拥有其数据及其产生的智能信息的所有权，呼吁各国建立“主权人工智能”（sovereign AI），这一言论引发了广泛讨论。在当前背景下，“主权 AI”指的是国家主导 AI 基础设施建设、模型训练及生态构建，尤其集中在算力和数据两个关键领域，并从硬件层（芯片、超算）向规则层（伦理标准、开源协议）延伸。这场全球范围的主权 AI 竞赛，将深刻影响未来数十年的全球科技发展轨迹。

美国通过政策推动和项目实施，持续展现其在 AI 创新、基础设施建设和国际主导地位方面的战略决心。2025 年 7 月，美国白宫发布了《赢得竞赛：美国人工智能行动计划》。首先，法案明确表示将快速推进人工智能在各领域的创新；其次，坚决完善与人工智能相关的重要基础设施和产业支持；最后，美国希望在国际 AI 外交和安全事务中发挥更大的主导作用。此外，美国凭借雄厚的技术积累，在 AI 基础层（如芯片、算法）保持领先地位，并积极构建由其主导的 AI 技术生态，例如通过星际之门项目加强与盟友的技术合作。同时，美国还试图通过出口管制等措施维护其技术优势。然而，这种做法可能导致与盟友关系的疏远，并最终影响其在全球市场的地位。

欧盟坚定地推行其数字主权战略，并将其延伸至 AI 领域。在顶层设计上，欧盟标志性的《人工智能法案》已于 2024 年 5 月由欧盟理事会正式批准，为 AI 的研发和使用划定了明确的法律界线。在基础设施层面，欧盟正大力推动 AI 算力网络的建设。在 2025 年 4 月发布的《人工智能大陆行动计划》中，欧盟提出将启动覆盖 17 个成员国的 13 个 EuroHPC“AI 工厂”，并计划投资超 100 亿欧元进行算力设施的重大升级。这些设施将重点服务于健康、能源和制造业等关键行业，为欧洲的科研机构和企业提供强大的本土算力支持，构筑其全球 AI 领导地位的基石。

中国依托政策协同和场景创新，在 AI 应用层面取得了显著进展，进一步彰显了 AI 在国家科技战略中的地位。中国政府高度重视 AI 技术发展，出台了《新一代人工智能发展规划》，并通过建设东数西算工程、推动开源倡议等举措，积极构建 AI 产业生态。2024 年，政府工作报告首次提出“人工智能 +”行动，去年年底召开的中央经济工作会议更是将这一行动作为 2025 年九项重点任务之一。2025 年 7 月，国务院常务会议审议通过了《关于深入实施“人工智能 +”行动的意见》，提出要深入推进“人工智能 +”行动，大力推动 AI 规模化商业化应用，充分利用中国产业体系完备、市场规模大、应用场景丰富的优势，加速人工智能在经济社会各领域的普及与深度融合，形成以创新推动应用、以应用促进创新的良性循环。

全球 AI 发展和治理格局正在加速演变，各国在技术、规则、生态等多个层面展开竞争与合作。未来，开放合作、共建共享将成为推动 AI 发展和造福人类社会的必然选择。

随着以大模型为代表的新兴技术在金融行业的全面深入应用，通用大模型与垂域大模型在场景侧激活动能，大幅提升了金融微观决策的信息对称性和金融服务的便利性、可靠性，为金融服务和产品创新提供了广阔空间。

第二章

金融业大模型建设与应用态势

金融业大模型建设与应用态势

随着算力资源的持续优化升级与 AI 技术的蓬勃发展，以大语言模型为核心的 AI 技术在金融行业中正以前所未有的速度不断涌现、演进与迭代。**战略驱动与价值导向已成为金融业布局新兴 AI 场景的双轮驱动**，推动大模型能力在金融业的渗透速率显著提升，展现出金融与 AI 深度融合的新态势。

2.1 总体概况：金融机构加速孵化大模型能力，行业整体审慎推进应用

2.1.1 全球金融业大模型应用建设概况

全球金融业加速拥抱 AI，大模型在金融行业的渗透率正加速提升。麦肯锡 2024 年的调研数据显示，金融行业从业者反馈在工作中常规使用大模型、在生活中常规使用大模型和在工作和生活中均常规使用大模型的数量占比已达到 48%。英伟达对近 400 家金融机构的调研显示，43% 的机构已开始应用大模型。[4] 国际金融协会报告显示，88% 受访者在生产中使用人工智能，并在 2025 年将持续增加 AI 应用投资。[5] 这种全球性的热潮在不同市场环境下，形成了各具特色的发展路径和战略重点。

从全球视角对比来看，海外机构更侧重技术整合与业务创新的协同，而国内机构目前更聚焦于知识库、文档处理等效率提升场景，这正反映了双方在不同发展阶段和政策导向下的不同选择。

海外金融市场的大模型应用展现出更强的主动性和规模化落地能力。

毕马威 2024 年美国银行业前景报告显示，65% 的受访机构领导者已将生成式 AI 纳入战略愿景，并预期在年底前将 1%-20% 的团队日常任务交由 AI 执行。[6] 从应用深度看，海外金融机构愿意将大模型部署于核心业务场景，例如智能定价策略优化、资金流动性管理、高频交易风控等直接影响经营效益的领域，同时也拓展至内部运营效率工具开发。从全球领先金融机构的实践来看，这一趋势正在加速：高盛集团（Goldman Sachs）自 2025 年起正式推出由生成式人工智能驱动的 AI 助手，并已扩展至全公司范围内的员工使用，能够帮助员工进行复杂文档总结、初始

4 英伟达，《金融服务业 AI 现状与趋势洞察》，2025

5 IIF-EY Annual Survey Report on AI/ML Use in Financial Services, 2025

6 KPMG, 2024 U.S. Banking Industry Outlook Survey

内容起草和数据分析等任务。在支付安全领域，维萨（Visa）于 2024 年推出基于生成式 AI 的欺诈解决方案，用于识别枚举攻击的可能性，这些攻击每年带来 11 亿美元的欺诈损失。万事达卡（Mastercard）则运用生成式 AI 技术将潜在受损卡片的检测速度提高一倍。

值得注意的是，部分海外银行已将大模型驱动的智能投顾、个性化理财方案等直接面向客户的智能服务嵌入业务流程，这种基于人机协同的技术直连用户的模式得益于其相对完善的隐私保护框架和流程被保护机制，但也需应对生成内容可靠性带来的合规挑战。尽管中小型机构受资源和技术储备所限，但正通过合作开发或技术外包等方式加速渗透，逐步缩小与大型机构的差距。

中国金融业在大模型建设上呈现出顶层设计、梯次推进的格局，体现了行业对技术全栈掌控和自主可控的战略追求。

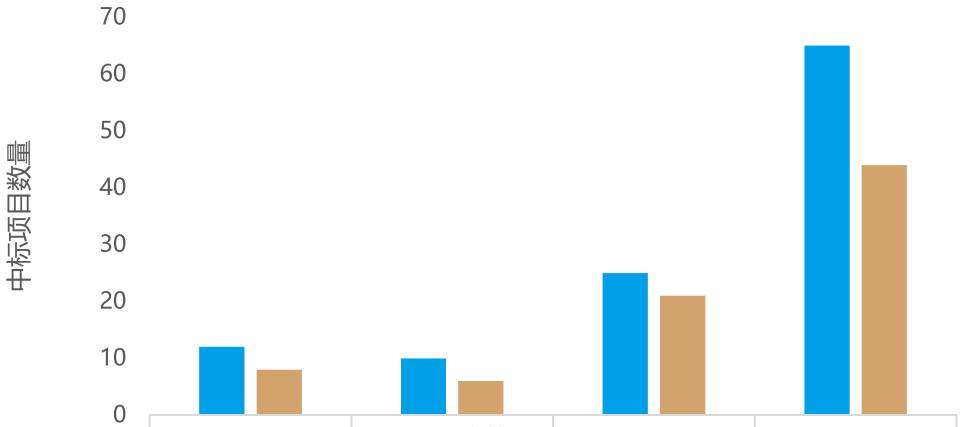
银行业是大模型落地应用最多的金融领域，其应用范围已经从国有大行、股份制银行迅速扩展到头部区域性银行。目前，国有大行和股份制银行已全面启动大模型应用建设，并在前、中、后台均有正式投产的应用案例。国有大行凭借雄厚的资金与技术积累，更注重技术的全栈掌控，旨在通过构建自主可控的技术体系，满足自身多样化的业务需求，进而提升核心竞争力。在此过程中，他们积极与国内顶尖机构开展深度合作，共同推进计算资源、计算调度与模型能力的全栈信创建设。股份制银行则展现出更为灵活多样的建设模式，它们在探索的宽度与广度上均取得了显著进展。区域性银行虽然起步较晚，但基于战略与价值驱动的探索热情同样高涨。目前，约 80% 的区域性银行已涉足大模型领域，部分已基于行业成熟的产品市场匹配度进行速赢落地，部分仍处于实验室阶段或全行范围内的智能体原型竞比阶段，少部分亦开展了全行级的领域实践。

证券、保险行业在大模型建设上也呈现出头部机构先行且模式多样化的特点。资管、信托行业在大模型建设上多聚焦于特定场景下的工具侧能力引入，尚未形成体系化的能力布局。由于大模型推理能力的突破，以投研、投顾为代表的金融场景亦正快速被券商、资管、基金、信托所接受。随着金融科技的日新月异与监管政策的不断完善，预计证券、保险、资管、信托等行业将逐步加大大模型建设的投入和布局力度。

2.1.2 中国金融业大模型应用建设数据洞察

2024 年以来，大模型技术迈入规模化产业落地的关键拐点，从概念验证转向实际业务应用的深度整合。金融行业凭借其数据密集、场景众多、拥抱创新的属性，展现出“人工智能 +”战略的示范效应和引领作用。

洞察一：大模型应用迎来从试水到抢滩的拐点

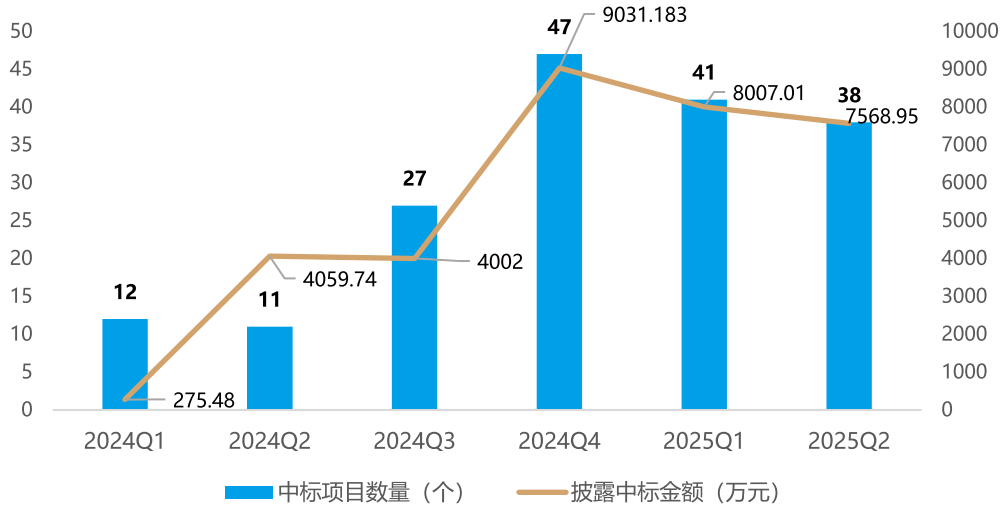


	保险	信托&资管	证券	银行
2024年	12	10	25	65
2025年H1	8	6	21	44

2024Q1–2025Q2 金融业大模型招投标信息汇总

基于全网公开披露信息统计（不含金融科技及消费金融样本），从 2024 第一季度至 2025 年第二季度期间，共计产生 191 个[7]大模型相关中标项目，其中 2024 年 112 个，2025 年上半年 79 个，覆盖银行、证券、保险、信托与资管。中标项目数量与金额均呈现头部集中特征，**这一趋势显示，金融业大模型应用已基本形成银行业主导、证券保险跟进、信托资管探索的梯次发展格局。**值得关注的是，进入 2025 年行业应用建设节奏明显提速，各类规模机构已全面启动大模型应用规划，大模型技术正在成为推动金融业数智化转型的核心引擎。

7 数据仅基于全网公开披露信息进行统计，样本收集时间截至 2025 年 6 月 30 日，仅收录公开中标结果的项目。渠道驳杂，可能存在未收录案例，亦有部分案例未做公开披露，金融科技、消费金融等样本未计入统计。



2024Q1–2025Q2 金融业大模型中标项目数量及披露金额

从季度演进趋势来看，2024 年 Q1 大模型中标项目仅 12 个，尽管第二季度中标项目数量有所下降，但从第三季度开始，中标项目数量呈爆发式增长达 27 个，Q4 保持高位增长至 47 个，2025 年上半年全面爆发，仅半年即快赶超 24 年全年的项目数量。这种低开高走的演化轨迹，印证了金融机构在技术成熟度和应用场景验证后的规模化投入，反映出大模型技术进入场景渗透 –ROI 验证 – 规模复制的良性发展闭环。这意味着，**企业观望的窗口期正在迅速关闭，竞争已从要不要用转变为如何规模化地用好和常用。未能跟上这一节奏的机构，可能会在 1–2 年内面临显著的效率和创新代差。**

洞察二：应用快跑与算力长跑并存

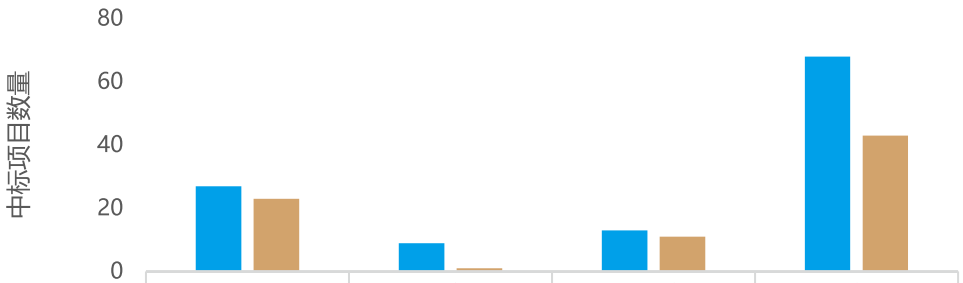
从客户服务到风险管理，从产品创新到运营优化，大模型以前中后台全链路覆盖的形式逐步渗透到金融业务的各个环节，带来前所未有的效率提升和创新突破。与此同时，各大金融机构正在通过多种采购方式推动大模型的应用与落地。根据采购内容的不同，可将大模型的采购划分为四大类：基础设施类、模型能力类、AI 平台类和 AI 应用类。其中：

- 基础设施类：为运行大模型所需的底层计算资源和硬件设施的采购，主要包括各类算力资源，如 GPU、TPU 等高性能计算设备；云计算资源；网络基础设施和存储设备等。

- 模型能力类：为大模型训练提供数据资源、算法优化服务，以及模型训练平台的技术支持。

- AI 平台类：用于支持大模型开发、部署和管理的 AI 平台或工具，主要包括大模型基座、AI 中台、AI 开发平台等。

- AI 应用类：将大模型与具体业务场景结合，开发并落地实际应用，主要包括应用软件，如智能客服、智能运营、智能研发等，以及针对特定业务需求（如信贷审批、反欺诈）的 AI 解决方案等。

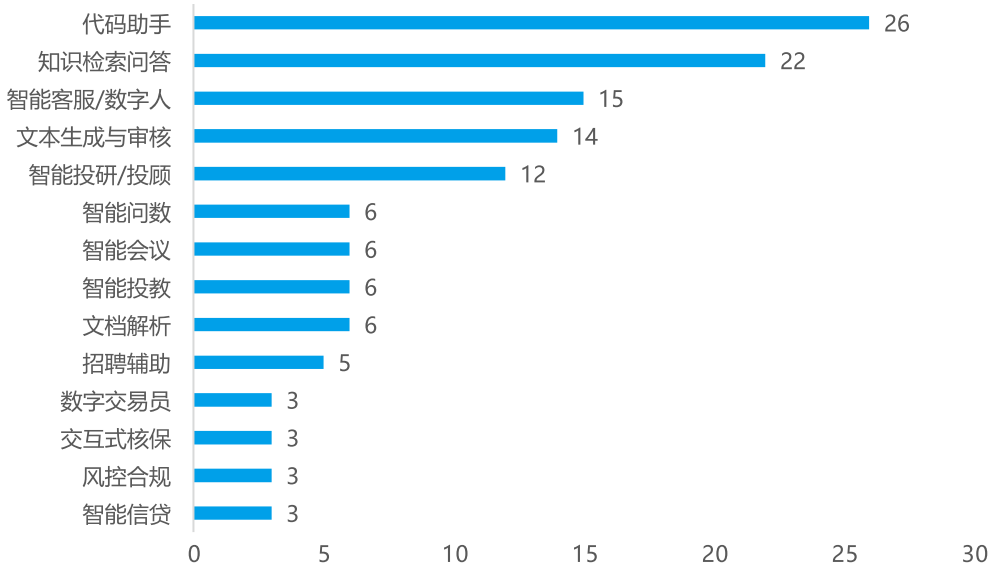


	基础设施	模型能力	AI平台	AI应用
2024年	27	9	13	68
2025年H1	23	1	11	43

2024Q1–2025Q2 金融业大模型招投标类型分布

图表分布清晰地揭示了金融业的布局策略。AI 应用类采购数量遥遥领先，而基础设施类采购虽然数量不多，但通常金额巨大。这揭示了行业一方面通过采购 AI 应用追求短期业务见效和技术价值快速兑现；另一方面通过投入基础设施进行长期算力储备和战略布局，尽量确保算力自主可控。**对于金融机构而言，既要避免陷入只买应用、不做基建的技术空心化风险，也要防止重金投基建、应用跟不上的资源闲置困境，确保两条轨道上的投资能够协同并最终融合。**

洞察三：场景渗透呈现由内向外的渐进式路径



2024Q1–2025Q2 金融行业大模型场景应用分布

（AI 应用的标段中会包含多个场景；此处仅选取中标项目数量≥ 3 个的场景）

将 AI 应用层项目按场景细分，应用场景的高度集中于代码助手和知识检索问答，揭示了行业当前主流的渗透路径，即技术渗透呈现由内至外、从效率工具向决策支撑演进，这是一种审慎的风险管理策略。**从务实的角度出发，当前的重点应是评估内部效率工具的真实成效，并思考如何将这些内部效率红利转化为可衡量的外部竞争优势，打通从员工赋能到客户价值创造的传导链条。**2025 年开始，由大模型驱动的业务模式创新或业务价值增益，成为头部金融机构首要考虑的关键场景或探索的核心趋势。

进一步，不同金融机构对 AI 应用的需求呈现差异化分布：

- 银行业：应用范围较为广泛，涵盖从前台业务提升（如智能客服、信贷报告生成、营销物料生成）到后台经营决策（如代码助手、知识问答）等多个领域，展现银行对于提升业务效率和客户体验的强烈需求。

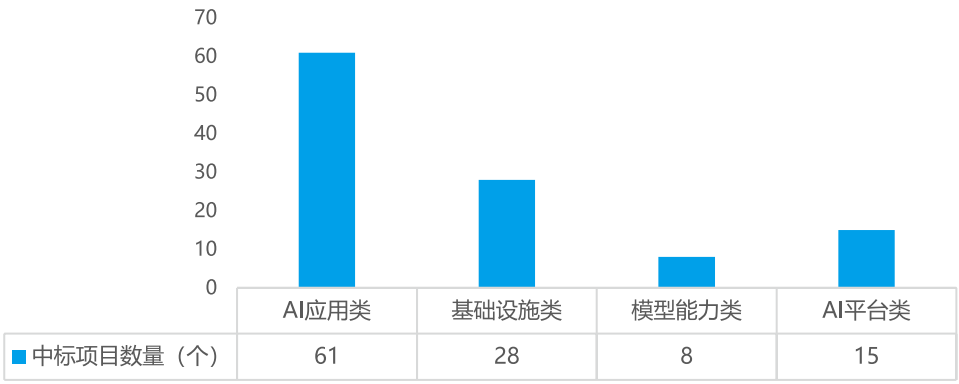
· 证券业：大模型的应用主要集中在提升投研工作效率、优化研发流程以及内容审核等方面。2025 年开始，智能投研、智能投顾类成为头部券商的重点建设方向，如投资组合交易分析、智能选股与诊断、舆情分析与研报生成等。

· 保险业：集中在核保流程优化以及知识库建设等关键领域。2025 年开始，以顾销渠道的保险建议书生成、代理人陪练、代理人小助手；市场部的营销画像分析、营销物料生成为代表，开始加速渗透。

· 资管业和信托业：中标项目占比相对较低，主要在智能问答、代码辅助，以及数字员工建设等方面进行探索。2025 年开始，头部基金公司开始重点探索智能投研、智能投顾两大专题，但多以合作共建或自建的方式推动相关能力建设；在基金的运营管理相关工作中，交易指令处理与意图识别、估值对账与异常预警、信息披露报告草稿撰写与审核亦为重点建设场景。

洞察四：银行业既是需求驱动的引擎，也是行业发展的关键参照系

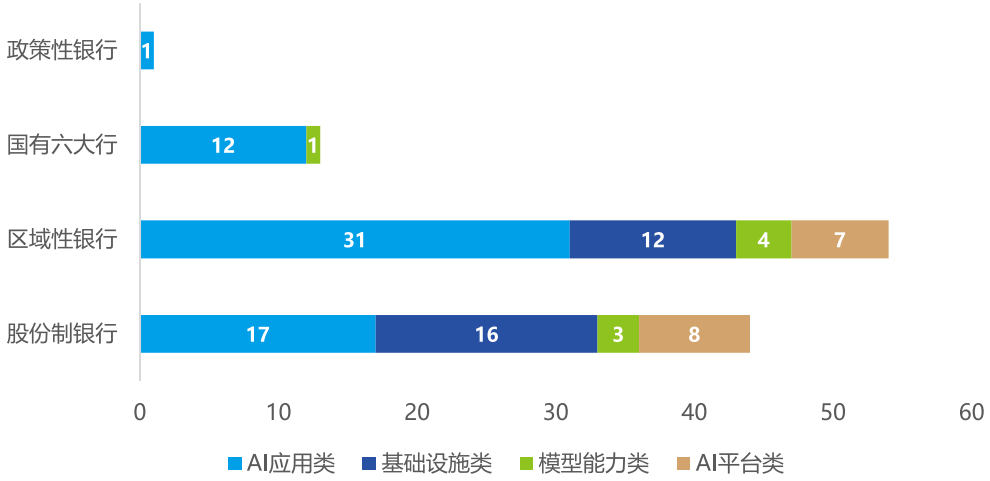
银行业的绝对领先地位，不仅是其体量的反映，更意味着它正在扮演整个金融 AI 生态的需求引擎。**银行的巨额投资正在定义技术路线、塑造厂商格局、吸引顶尖人才。其在应用、平台、基础设施上的采购战略，将为其他金融子行业提供重要的参照系。**



2024Q1–2025Q2 银行业大模型中标项目分布（采购类型）

（按采购类型，同一标段含多个类型）

从采购需求来看，银行类金融机构对 AI 应用类项目的采购占比 54%；基础设施类项目采购占比 25%；AI 平台类项目采购占比 14%，模型能力类项目采购各占比 7%。通过这四大类采购，基本能够全面覆盖从底层算力支持到上层业务应用的全链条需求。



2024Q1–2025Q2 银行业大模型中标分布（按银行类型）

（按银行类型，同一标段包含多个采购类型）[8]

根据银行性质和职能进行划分，将银行机构划分为国有六大行[9]、股份制银行、区域性银行、政策性银行。**当前，大模型战略并无唯一最优解。机构的资源禀赋、市场地位和战略雄心，共同决定了其最适合的采购与建设路径。**具体而言：

· 国有大行加速应用体系建设：国有六大行在 2023 年末相继完成了基础设施的补足，2024 年基于充足的算力，积极开展大模型应用体系化能力建设，前中后台与通用工具均有所涉及并真实投产，存在垂直业务领域的端到端赋能与单一系统平台的全能力支撑；2025 年开始国有大行聚焦于前台业务强相关的应用探索与建设，并开始探索 AI 服务能力原子化、平台化，以及全行级的知识体系建设。股份制银行紧跟趋势，全链条多重投入：相较于其他各类银行机构，股份制银行 2024 年的中标项目中 46% 为基础设施类采购，包括各类算力服务器资源、配套网络设备等大模型训练集群所需的基础设施建设硬件；2025 年上层应用百花齐放，面向前中后台均有所突破，甚

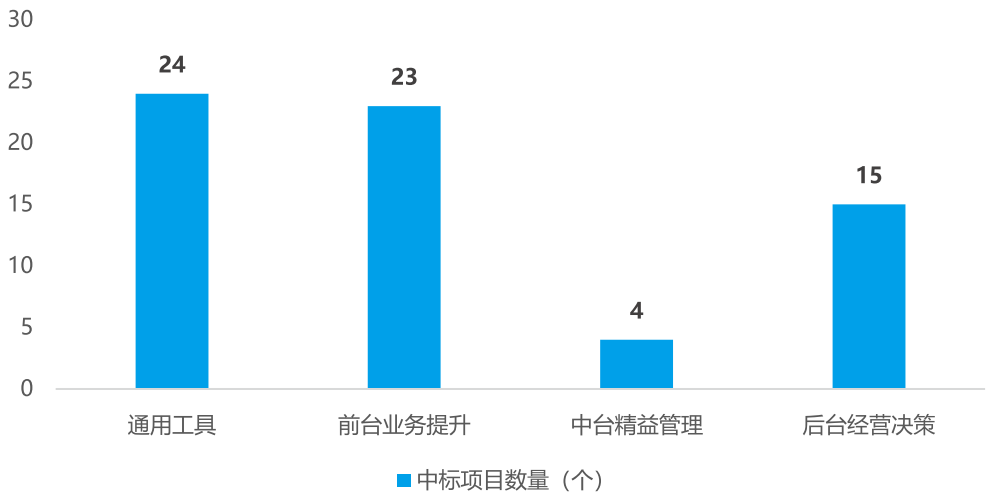
8 本统计以项目名称为计量单位，未做同一银行的去重

9 国有六大行特指工商银行、农业银行、中国银行、建设银行、交通银行、邮政储蓄银行

至亦有流程嵌入型的应用进行投产，重新定义人机协同的流程组织，同步推进知识体系建设。

- 区域性银行场景化应用突围：区域行经过 2024 年的蛰伏，于 2025 年以经过市场验证的 PMF 进行速赢建设。其中某头部城商行聚焦构建全行统一智能技术体系，并在此基础上拓展 AI+ 金融业务场景应用，如代码助手、智能会议、智能报告等；总体来看其他区域性银行采购需求涉及知识检索问答、智能客服、数字人在内的通用工具和前中台应用。

- 政策性银行以客户为核心的服务深化：以某政策性银行举例，对 AI 大模型的应用需求主要集中在智能客服的部署，以及客服垂直领域模型的参数调优和升级，深度开发生成式 AI 的能力，提升自然语言处理、知识推理、场景问答和创新解决方案生成的核心技术实力，优化客户服务体验。



2024Q1–2025Q2 银行业 AI 应用类项目分布

（同一标段包含多个应用类型）

前台业务提升和通用工具是银行应用落地的两大重点。银行业当前已基本构建了三级递进策略框架，即基础层重投入、平台层快迭代、应用层求突破。此框架下，应用类项目的实施焦点显著集中于前台业务优化与通用智能工具的应用两大核心领域。在前台业务提升方面，通过引入智能客服、数字人增强客户互动体验，提高响应速度和问题解决效率，打造更加个性化与沉浸式的服务场景；智能投研为投资决策与市场洞察提供了强有力的辅助，进一步增强金融服务能力；通用工具加速推动知识检索与智能搜索服务的革新，实现信息的快速获取与高效传播。**值得注意，公开采购数据并非等同于战略重要性，据调研，最核心、最敏感的风控类系统，更多通过自研或深度合作开发，而非公开招标采购。以 2025 年的采购场景与类型来看，公开采购的多为经过市场验证的 PMF 可速赢应用，面向流程嵌入型、模式重塑型的应用多为自研或深度合作开发。**

2.2 建设模式：以实际效益为导向，建设方式渐进收敛

面对大模型热潮，金融业正积极探索适合自身的应用建设模式，并呈现出以投入产出比（ROI）为导向的渐进收敛趋势。由于大模型训练成本高昂，金融机构更加注重应用落地的实际效益。目前，无论是购买基础算法框架，还是与顶尖研究机构、模型公司开展深度合作，金融机构都致力于在风险控制、客户服务、效率提升等方面培育潜在场景应用，并通过不断优化模型性能、降低部署成本等方式，追求更高的投入产出比。在这一过程中，金融业在应用部署方式、能力建设方式、模型构建方式等方面，逐步形成了几种较为普遍且更注重实际效益的建设模式。

2.2.1 大模型应用部署策略

大模型应用相关的算力配置顺序通常为：开发环境（训练）– 测试环境（推理）– 生产环境（推理）。针对这一特点，金融机构在部署大模型应用时，主要有以下三种方式：

- **本地数据中心部署：安全优先和自主可控**

金融机构将数据处理、模型预训练与微调、模型推理等环节全部在本地数据中心完成。这种方式的优势在于数据安全性和隐私保护程度较高，金融机构可以完全掌控数据的存储和处理过程，符合金融行业对数据安全和合规性的严格要求。然而，这种方式也存在一些挑战，自行建设和维护大规模算力基础设施的成本较高，且在面对业务峰值流量时可能算力弹性不足。

- **私有云与本地结合的混合云部署：兼顾安全与弹性**

金融机构将大算力侧置于私有云，利用私有云的弹性计算能力进行模型预训练和微调等计算

密集型任务，而将模型推理环节放在本地进行，实现大规模在离线混部。这种部署方式既能够充分利用私有云的弹性算力资源，又能够保证模型推理的稳定性和数据的安全性。同时，在增强预训练或 SFT 环节，数据可以在私有云中进行处理和分析，进一步提高模型的性能和准确性。

· 公有云 / 团体云与本地结合的混合云部署：灵活高效

金融机构可以在公有云或团体云上完成模型的预训练和场景试验，解耦训练推理资源与环境，而将核心系统及其数据仍然保留在本地。在保证数据安全的前提下，加速大模型的应用落地和业务创新。

同时，金融机构可以利用公有云提供的丰富的大模型训练资源和先进的训练工具，快速进行模型的原型开发和验证，当模型达到一定成熟度后，再将其部署到本地进行实际业务应用。

团体云场景则专指使用联邦大模型以进行数据共享且能保障数据隐私的技术架构。在金融行业数据合作中，不同金融机构可以通过团体云平台，利用联邦学习技术共同训练大模型，实现数据价值的最大化挖掘，同时确保各参与方的数据隐私和安全。

大模型应用部署方式

应用部署方式	方式特性	适用场景
本地数据中心	全流程封闭式管理，数据零外传	全量
私有云 + 本地混合部署	训练上云，推理本地，实现计算弹性与数据安全平衡	全量
团体云 + 本地混合部署	训练上云，推理本地，利用联邦学习技术共同训练大模型，保障数据隐私和安全	非核心系统或非强监管场景（建议）
公有云 + 本地生产	云端原型验证，本地化部署成熟方案	仅使用公开数据或领域数据的 MVP 验证 /demo 试验 / Agent 探索

金融机构在选取部署方式时，需考量当前成熟的算力调度方案、目标场景的数据隐私与模型安全要求，按需选择上述方式。

2.2.2 大模型能力建设方式

模型能力获取方面，金融业主要采取以下五种方式：

- **端到端自建：**金融机构全面掌控从模型训练、优化到部署的全流程技术能力，涵盖算力、算力调度、算法模型及工程落地的全周期；自行采购并搭建算力基础设施，包括高性能的 GPU 服务器等硬件设备，以及相应的网络和存储设施；组建专业的数据科学团队，负责从数据收集、清洗、标注到模型训练、验证和优化的全过程。这种模式需要强大的技术实力和资源投入能力。

- **基于基础大模型开发专有模型：**金融机构选择闭源或开源的基础大模型，在此基础上，结合自身业务数据进行进一步的训练和优化，开发出具有特定功能的专有模型。例如，与科研机构 / 高校合作，共同投入资源进行模型训练，共享成果。

- **基于 Agent 编排平台构建大模型应用：**金融机构采购私有化 Agent 编排平台及相应的基础设施，将多个大模型或模型组件进行有机组合，灵活地构建和管理大模型应用，实现复杂业务逻辑的自动化处理。Agent 平台通常会集成开源的基础大模型或行业大模型，作为应用编排的模型库。

- **基于大模型 API 开发特定场景应用：**金融机构通过调用第三方提供的 API 接口，结合自身业务需求进行应用探索，无需自行搭建和训练大模型。API 接口按 token 计费的模式可以灵活控制成本，适用于在一些非核心业务或非强监管领域的应用场景进行试验。

- **采购具备成熟大模型能力的相关应用：**金融机构直接采购市场上已有的、针对金融行业特定业务场景进行过优化的、具备成熟大模型能力的应用软件。

大模型能力建设方式

建设方式	技术特征	适用场景
端到端自建	全栈自主可控，需大规模算力基建与专业团队支撑	强监管、战略引领的能力建设 / 领域能力的价值深化
专有模型开发	基于闭源 / 开源基模进行领域微调，数据安全与性能平衡	战略引领的能力建设 / 领域能力的价值深化
Agent 编排平台	多模型协同调度，实现复杂业务逻辑的动态组合	复杂场景创新试验
API 调用模式	按需付费的敏捷试验，适合非敏感场景快速验证	轻型应用探索 / 试验
成熟方案采购	即插即用的行业解决方案，缩短价值兑现周期	中小机构数智能力补足

★ 以上建设方式亦可混合并行

金融机构在选取建设方式时，需考量当前所处的技术周期、目标场景的可控性与性能要求，按需选择以上能力；当前并无绝对正确的建设方式，开源基模的使用与垂域模型的定制化开发更需理性选择。

2.2.3 大模型体系构建路径

· 模型训练

在模型训练侧，分布式并行计算框架构成算力基座，配合混合精度训练、梯度压缩算法等优化方案，以及参数高效微调技术，模型训练效率得以提升，模型规模得以控制，有利于模型轻量化部署，构建了从通用基座到垂直领域模型的参数集约化路径。

当前，大模型训练主要包含预训练基座构建、监督微调（SFT）和人类反馈强化学习（RLHF）

三大核心环节。当前，金融机构多采用基座模型进行领域知识注入和价值观对齐，仅少数头部机构具备从零预训练千亿级大模型的算力储备与时间。

· 模型应用

在模型应用侧，**检索增强生成（RAG）技术成为构建可信金融大模型的关键。**RAG 技术融合向量化数据库与行业知识图谱，形成数据检索 – 知识增强 – 智能生成的动态闭环，推动大模型从通用对话向专业决策场景延伸。该架构类似于开卷考试机制，借助实时数据检索更新，突破了模型静态知识局限，有效缓解了生成内容的时效偏差与事实性错误。目前，在金融领域，超过 80% 的智能投研、监管合规等场景依赖 RAG 技术实现业务落地。相比之下，单纯依赖提示词工程的原生大模型应用尽管部署快捷，但受制于训练数据的时效性，难以满足高频市场分析等动态场景需求。

· 模型协同（异构模型管理）

为了更好地管理和利用不同类型、不同规模的模型，部分领先金融机构正在构建智能化的异构模型协同管理平台。这类平台通过构建动态调度中枢，实现了大模型与小模型的有机协同。平台基于实时流量监测与资源调度算法，系统能精准识别用户请求的复杂度，在轻量模型与超大规模参数模型之间实现毫秒级智能切换，既保障高价值场景的分析深度，又避免算力资源的无效消耗；其次，通过建立模型效果与成本核算的动态平衡机制，平台可依据业务场景的容错阈值和成本约束，自主配置最优模型组合，使单位算力投入产出比提升；再者，平台创新的知识蒸馏机制支持将大模型输出的高质量分析结果反哺小模型训练，形成持续优化的技术闭环。

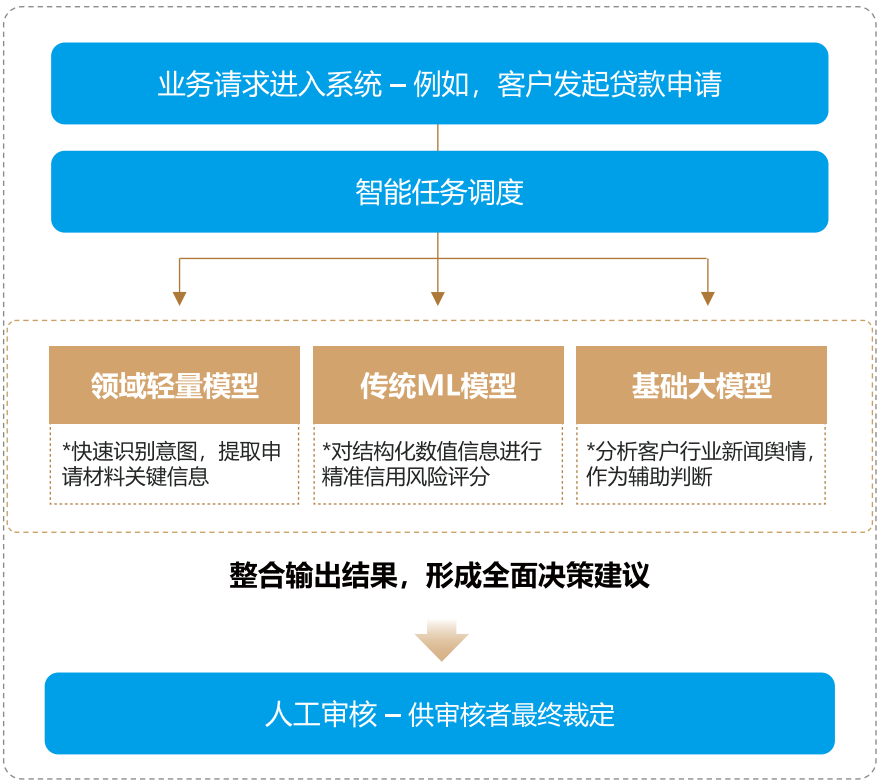
在金融大模型的实践落地中，单一模型难以经济高效地应对所有复杂场景，因此金融机构已经或正在构建功能互补、高效协同的混合模型体系。领先的金融机构正在超越大小模型的简单二元划分，构建一个由基础大模型、领域轻量模型和传统机器学习模型三者有机协同、智能调度的混合智能体系。这三类模型基于自身技术特点，承担不同角色：

· 基础大模型：具备强大的通用知识和复杂的推理能力。它们主要负责处理开放式、探索性的任务，例如宏观经济趋势分析、创新构思等需要跨领域知识的场景，同时也可作为后续领域模型微调的基础。

· 领域轻量 / 蒸馏模型：这是从基础大模型通过微调或蒸馏等技术，面向特定金融业务优化的模型。它们专注于具体任务，如合规文本审核、智能投研摘要、信贷报告要素提取等。这类模型更轻量、响应更快、运营成本更低，是实现金融业务规模化、高效化应用的核心。

- 传统机器学习模型：例如梯度提升树、随机森林等经典算法，在处理结构化数据时依然拥有高精度和高可解释性的显著优势。它们在信用评分、量化交易、反欺诈侦测等场景中，持续发挥着关键作用。

这三类模型并非独立运作，而是通过智能化的任务编排平台协同工作。



异构模型协同应用示意图

当一个业务请求进入系统时（例如，客户发起一笔贷款申请），平台会进行任务的智能分解与调度：首先由一个轻量模型快速识别客户意图并提取申请材料中的关键信息；其次将结构化的数值信息交由传统机器学习模型进行精准的信用风险评分；同时可调用基础大模型对客户所在行业的新闻舆情进行分析，作为辅助判断。最终，系统将三者的输出结果进行整合，形成一份全面的决策建议，供人工审核者最终裁定。这种协同模式，实现了模型能力与业务场景的最佳匹配，在成本、效率和精准度之间取得了有效平衡。

2.3 场景落地：从效率工具向决策引擎演进

2.3.1 金融业大模型的应用版图

金融业的大模应用正从零散的点状尝试，向通用工具、前台业务提升、中台精益管理、后台经营决策四大领域全面渗透。

金融业大模型应用全景图

通用工具是金融机构引入大模型技术的基础应用领域，其核心特征为技术通用性强、跨场景适用。此类应用主要面向机构内部，涵盖文档图像处理、智能问答、音视频内容分析、代码辅助生成等多个方面，旨在提升组织内部的运营与研发效能。由于该领域技术成熟度高、应用场景明确、风险相对可控，其所带来的价值直观易衡量，因此成为多数金融机构部署大模型的首要实践领域。其应用主要聚焦于机构内部，以确保信息与业务风险可控。

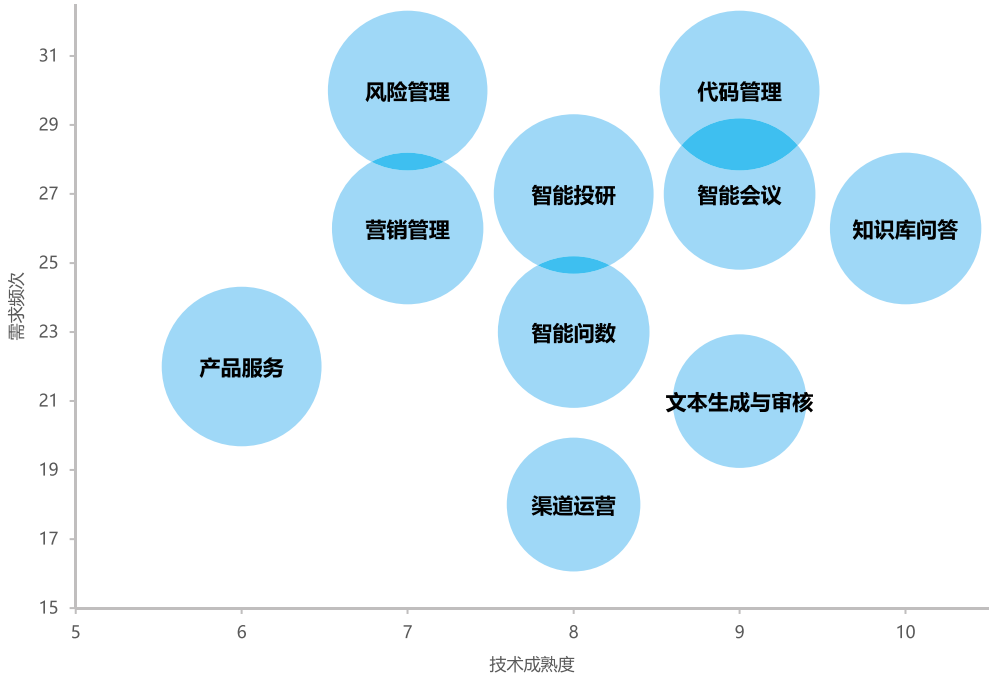
前台业务提升领域直接面向客户价值创造，是决定未来市场竞争力的战略关键。其通过对客户行为、金融产品偏好与潜在需求的深度分析，构建精准客户画像，进而赋能个性化营销与服务能力的升级。应用范围覆盖客户服务、精准营销、投资顾问与投资研究等多个直接关系到客户体验与业务增长的领域。受限于金融场景的严谨性要求、模型输出的可控性挑战以及投资回报率的评估周期，多数相关应用仍处于实验室探索或小规模试点阶段。部分成熟度较高的应用，正通过与传统人工智能模型及人工专家协同的模式进行部署。

中台精益管理领域是机构稳健运营的基石，侧重于强化机构的运营效率与核心风险管理能力。通过大模型技术对海量交易数据、市场动态与客户行为模式进行实时分析，能够构建更为智能与动态的风险预警、识别与防控体系。应用聚焦于风险识别、合规审查、反欺诈、智能运营等中枢环节。该领域对模型的可靠性、稳定性与专业性均提出极高要求，通常由金融机构主导自研，或与顶级技术厂商进行深度合作定制开发，是体现机构核心风控能力的关键领域，亦是当前行业探索与应用推广的重点方向。

后台经营决策领域旨在支撑组织高效运转与科学决策，主要赋能财务管理、人力资源、开发运维等核心职能部门，以提升内部管理的科学性与自动化水平。在财务领域，应用聚焦于财务报告的自动化生成、预算规划的智能推演以及关键经营指标的预测性分析。在开发与运维领域，大模型在代码自动生成、测试用例生成、系统异常智能诊断等方面展现出巨大潜力，正成为部分领先金融机构探索的创新方向，旨在提升软件工程全链路效率，同时辅助提升相关岗位人员的专业能力与决策质量。

在领先机构的实践中，上述四大应用领域正走向由统一 AI 战略驱动的协同共振。以保险科技领域的探索为例，微保与腾讯混元共建保险领域大模型，面向 C 端用户开发智能助手，以解答产品、核保、理赔等问题；同时，在后台自建智能体开发平台，将内容生产、数据分析、质检等环节的 AI 赋能门槛显著降低。这种“内外兼修”的布局，将外部客户价值创造与内部运营效率提升相结合，形成良性循环，代表了金融业大模型应用的方向之一。

2.3.2 金融业大模型的演进路径



金融业大模型应用路线图

备注：气泡半径与场景价值成正比；共 30 个金融机构样本量；
以可商用作为技术成熟度高的标准。

大模型的应用正从最初的效率提升工具逐步向价值创造引擎演进，并带来了应用场景迭代升级。通过技术成熟度、需求频次和场景价值三个维度，我们绘制了当前的应用现状态势，行业对效率的追求和对业务创新的迫切需求塑造了当前的应用路线：

- 代码管理、知识库问答、智能会议是当前 ROI 最明确、落地最快的场景，是效率工具的典型代表，是所有机构都应迅速布局的基础能力；

- 风险管理、营销管理这些场景价值巨大，需求迫切，但技术仍在攻坚。这是未来拉开差距的关键，需要持续的战略投入和耐心；

- 文本生成与审核等技术成熟，需被整合到其他流程中，适合作为插件或组件快速部署。渠

道运营场景的需求频次和场景价值有所下降，可能是因为更有效的新模式尚未出现，当前投入需谨慎评估 ROI。

三大现实因素制约具体场景落地的快慢和效果：

可控性：如何确保大模型在复杂金融场景下的输出结果可靠、稳定、可控，仍然是制约大模型场景落地的主要因素。

可解释性：大模型的决策过程缺乏透明度，难以解释其推理逻辑，成为试验场景向外推广的核心顾虑

投入产出比（ROI）：金融机构越来越关注大模型应用的投入产出比，尤其是在推理决策场景中，与传统 AI 方案在推理效率、算力投入与场景价值的优势。

受限于上述因素，金融业大模型趋于成熟的应用模式主要体现为以下两个层面：

辅助工具：大模型主要作为高效的辅助工具，用于提升特定工作环节的执行效率，但尚不深度介入核心业务的决策流程。典型应用包括辅助编程、文档归纳与信息检索等，旨在优化内部员工的日常工作效能。

内容生成：大模型被用于自动化或半自动化地生成各类业务内容，显著提升信息生产的规模与速度，以流程嵌入的形式存在于业务流转中。具体应用涵盖市场营销文案撰写、初步行业分析报告生成、基础数据报表制作以及宣传材料的设计等。

与此同时，业界正积极探索更具深远影响的应用模式，其商业价值与技术路径的成熟度尚在验证过程中，主要包括以下几个方向：

业务流程再造：此方向旨在构建由智能体深度参与的全新业务工作流，以实现端到端的自动化与智能化。当前，金融机构在此领域进行了大量探索。其中，应用于知识库问答、智能数据查询等场景的智能体技术已相对成熟并投入实际应用。然而，涉及更复杂决策逻辑的投资研究分析、动态风险控制等领域的智能体应用，仍处于深入研究与验证阶段。

商业模式创新与业态重塑：此方向包含两个层面。第一，在现有业务框架内，利用智能体技术显著提升客户触达、转化与服务的效率，例如探索新型的智能化零售金融服务模式。第二，基于大模型原生的能力，创造全新的金融产品或服务形态。目前，此类应用在金融行业的实践尚处于早期的概念构想阶段。

2.3.3 金融业大模型的应用新形态：智能体

智能体（Agent）是基于先进大模型构建的应用实体，具备自主感知环境、决策制定与行动执行的全套能力。Agent 的目标是在无需人类干预的情况下，通过观察现实世界并利用内置及外接工具，自主实现预设目标。**Agent 具有以下三个特性：**

自主性：Agent 能够独立于人类干预完成任务，通过深度逻辑推理与预测确定下一步行动方案，以实现最终目标。这种自主性使得 Agent 能够在复杂且动态的环境中高效运作，无需持续监督。

适应性：Agent 具备卓越的环境适应与策略调整能力，通过持续学习与适应不断优化决策过程，以应对市场条件、用户需求或其他外部因素的快速变化。这种适应性确保了 Agent 在多变环境中始终保持高效。

交互性：Agent 能够通过自然语言等方式与用户及其他系统进行高效沟通，准确理解用户查询意图、提供及时反馈、清晰解释决策过程，并与其他系统或 Agent 实现无缝协作。这种交互性提升了用户体验，增强了 Agent 在复杂任务中的协作与执行能力。

智能体凭借其任务闭环执行能力和动态环境适应性，成为技术落地的核心载体。**这一趋势的驱动力体现在三方面的变化：**

需求升级，从辅助工具到决策执行。传统大模型多局限于文本生成、问答等单点场景，而金融业对业务流程重塑的需求迫切。Agent 通过整合工具调用、环境感知与自主决策能力，可完成信贷审批自动化、实时风控拦截、投研分析等复杂任务，实现从认知支持到行动闭环的质变。

技术适配，突破大模型固有局限。大模型的黑箱性与金融场景的高合规要求存在矛盾。基于大模型增强预训练、微调与对齐的输出后，Agent 通过 RAG、插件能力调用及流程编排模式，可在保障数据安全的前提下提升模型专业性与可解释性。

生态演进，从单点智能到系统协同。金融业务流程的复杂性需要 AI 具备协同分工能力，Agent 框架天然支持多工具集成与多角色协作。如 Manus 通过动态调度各类大模型，构建覆盖数据获取、分析、执行的智能工作流，解决了传统大模型只建议不行动的痛点。这种生态化能力使其在股票分析、合规审查等场景中展现出超人类团队的效率。

智能体系统正经历从个体智慧向群体智能的范式升级：

单智能体犹如专业领域的超级个体户，其核心价值在于独立完成标准化流程，例如自动生成财务报表或执行简单交易指令。这种形态常见于金融后台的自动化场景，其优势在于**部署成本低、响应速度快，但面对跨部门协作、多因素决策等复杂场景时往往力不从心。**

多智能体协同系统则构建了数字神经元网络，**每个智能体如同金融组织的专业化神经元，通过动态协商机制形成分布式决策网络。**这种架构在风险管理场景中尤为显著：信用评估智能体、市场波动监测智能体、流动性管理智能体等既保持专业独立性，又通过实时数据共享形成风险联防体系。该架构既能保留了专业深度，又实现了系统韧性，正如现代投行交易中不同策略组的协同运作，通过 AI 实现了毫秒级的决策同步。

随着 Agent 应用生态的蓬勃发展，如何高效、规模化地连接模型与外部工具，已成为核心议题。在此背景下，虽然功能调用提供了基础的实现路径，**但模型上下文协议则代表了更具前瞻性的生态级解决方案：**

功能调用（Function Call）聚焦于模型自身能力增强。功能调用是赋予大语言模型调用外部工具能力的基础技术。它通过标准化的方式，让模型能够指定工具并传递相关参数，从而完成特定任务。整个调用与执行过程通常发生在智能体（Agent）的进程内部，即使工具的业务逻辑可能部署在远程。其设计初衷是让语言模型能更精准地使用一组已知工具，从而便于后续对模型的行为进行优化和训练。它本质上是一种模型中心化的解决方案。

模型上下文协议（MCP）着眼于构建开放、协作的工具生态。与功能调用不同，MCP 将视角从单个模型扩展至整个生态系统，旨在解决多智能体与多工具之间的协同问题。它通过定义一套统一的标准协议，在工具的调用者（Agent）和提供者（Server）之间建立起沟通的桥梁，从而实现了智能体与工具的彻底解耦。MCP 的核心目标是构建一个开放的、类似“应用商店（App Store）”的工具生态系统。在此生态中：工具开发者可以便捷地将自己的工具发布并融入 AI 生态，无需与特定的智能体或模型进行深度绑定。工具本身增强了复用性，并为商业化（如按次收费）提供了可能，最终提升整个生态的运转效率。

因此，MCP 不仅是对智能体与工具交互方式的技术优化，更是一次架构层面的范式革新。它通过构建一个更加开放、灵活且可扩展的生态环境，为未来复杂智能体应用的协同工作与商业化落地奠定了坚实的基础。

金融业大模型的落地绝非单纯技术命题，而是涉及战略重构与组织进化、数据要素治理、体系融合的系统工程。唯有通过统筹规划、机制创新与持续运营的三维突破，方可实现从实验室盆景到业务森林的规模化跃迁。未来，随着多智能体等技术的渗透，金融大模型将逐步迈向人机共生的新范式，但其成功必然建立在跨学科、跨机构的协同基石之上。

3 第三章

金融业大模型建设的核心挑战与应对策略

金融业大模型建设的核心挑战与应对策略

3.1 金融数字化转型背景下的三大平衡关系

金融业对于新质生产力的理解已深化至探索金融科技创新动力的价值经营层面，体现在金融机构对于数字化转型三大平衡关系的策略把控与行动落位上。

局部突破与整体效能平衡。数字化转型进入精益发展阶段，金融机构需破解碎片化建设与全局效益的协同难题。当前呈现两大策略趋势：一是投入策略更趋审慎，降本增效类项目占比显著提升，形成收益、风险与成本的动态优先级调整机制；二是构建跨部门协同体系，通过产品经理责任制重塑业业融合模式，建立业务、科技与数据三位一体的业技融合机制。某头部机构已经将一体化写入数字化转型的核心战略愿景，强调数字化治理作为数字化转型总体蓝图实现的根本前提，确立数字化统筹管理职能与组织协同机制，营造数字化转型人人负责、人人参与、人人贡献的文化理念，践行落地数字化任务级实施路径。

创新投入与资源效能平衡。在资源约束条件下，领先机构正构建差异化资源配置体系。面向小额创新实践项目，通过更为灵活和容错的管理方式以实现前瞻性、探索性、智能化的创新发展目标。同时设立专项创新基金，通过宽进严出机制激发探索活力，对场景获客、中台建设等方向实施敏捷管理；同时聚焦速赢项目，以短周期试点验证核心价值，形成样板复制、信心强化与良性循环的推进逻辑。某领先股份制银行通过建立金融科技基金并执行专项运作机制，以支持场景及生态建设、中台能力提升、领先技术探索、创新组织与机制建设等方面的创新能力突破。

前沿探索与风险防控平衡。金融机构在新技术应用领域呈现审慎创新特征。当前聚焦两大实施路径：数据全周期管理成为逆周期投入重点，通过隐私计算等技术升级安全体系，构建治理 – 平台 – 应用 – 赋能的完整链条；AI 大模型应用强调场景适配性，聚焦技术与场景融合探索，以数据为中心、模型为中枢的业务场景持续把握监管合规、安全风控和增长赋能的平衡。某头部机构以业务引领、数据驱动为策略，已形成从数据治理到业务赋能的闭环体系，但大模型在核心业务场景的渗透仍处于验证阶段。

金融业大模型的落地挑战本质上是数字化转型三大平衡关系在智能技术深化阶段的集中映射，需以系统性思维重构技术、数据、组织与商业模式的协同机制。

金融业数字化 转型平衡关系	金融业大模型 能力体系建设
局部突破与整体效能平衡	要求大模型建设从单点技术验证转向系统性价值创造。局部场景的算法突破若缺乏与业务流程、数据中台及组织架构的深度耦合，易导致技术能力悬浮于业务需求之上，形成技术孤岛； 当前金融机构普遍面临模型能力与业务价值传导链断裂的问题，部分试点项目因无法融入核心风控、客户运营等价值链环节而陷入重复建设。
创新投入与资源效能平衡	算力基建、数据治理等底层投入的刚性约束，倒逼大模型建设必须实施精准投资策略。当前行业普遍存在大势紧追与场景价值迷失的现象，且仅有少部分机构聚焦 ROI（投资回报率）测算； 差异化资源分配的前提是完备的数据资产、充分的业务流程融合基础，以探明 AI 就绪评估分级分类评定速赢项目。
前沿探索与风险防控平衡	以数据为中心、模型为中枢的业务场景应持续把握监管合规、安全风控和增长赋能的平衡；应用场景的建设由内部效率工具向对外决策支撑审慎推进； 以大模型为核心的应用场景当前仍应以人机协同的方式规避风险，尽可能降低由技术成熟度所带来的 AI 幻觉风险；按照监管要求，构建可信 AI 体系。
三大平衡逻辑交叉影响	面向科技创新与产品孵化的敏捷协作机制； 面向复合人才的培养计划与激励机制； AI 影响下的组织、流程的重构式创新。

以大模型能力建设为代表的新质生产力在金融业的成功践行从来不是金融科技创新单方面的一腔热血、孤掌而鸣，如何把握由面及点的多层级平衡关系，回归价值经营本质，最大化释放资源效能，是金融机构迈入发展新阶段的必要性命题，也关乎数字金融的发展与未来。

3.2 大模型建设的四大核心挑战与应对策略

一切生产力转型的根本目的仍然在于业务增长和管理提效，因此，在金融行业积极拥抱大模型浪潮时，更需要保持冷静，切勿拿着锤子找钉子，盲目追赶技术热点。企业需要立足自身业务，梳理出具备实际产业价值的可落地应用场景，并基于自身丰富的数据语料，训练和调教适合自身业务和管理需求的智能体，使其能够真正解决业务痛点，释放管理效能。更进一步，要推动 AI 应用从单纯的对话机器人模式升级到 Agent，并逐步融入一线员工习惯使用的业务应用系统，让 AI 真正成为业务助手，这才是让业务前线具体感知 AI 能力并推动业务智能化转型的关键。

尽管引入人工智能已被金融业广泛认可为提升运营效率和客户体验的关键驱动力，大模型技术正重塑金融业态，但其落地过程面临多维挑战。

3.2.1 数据挑战：从碎片化资源到规模化语料的转化困境

【具体问题】

私域数据资产的激活障碍与专业模型能力的供给稀缺：金融机构虽拥有海量高价值的私域数据，但这些数据因系统壁垒而碎片化，导致难以被有效激活，形成统一、可用的知识资产以供大模型实时利用。与此同时，公开市场上缺乏能满足金融风控、财富管理等场景严苛要求的专业训练语料，使得通过传统方式训练或微调出具备深度领域能力的模型成本高昂且周期漫长。这形成了内部知识无法释放、外部能力无法精准补给的困境。

非结构化数据向可用知识转化的治理体系缺失：金融业务流程中产生海量的非结构化数据，例如法律合同、信贷审批报告、券商研究报告、财务报表附注以及客户服务中心的语音与文本记录。现有数据治理体系大多围绕结构化数据构建，对于如何从这些蕴含了丰富上下文与深度知识的非结构化载体中进行高效、精准的信息抽取、语义理解与知识连接，缺乏成熟的技术框架与治理标准。这使得大模型无法充分消化利用这些关键信息，其在文档智能审核、市场舆情深度分析、客户意图精准识别等高级应用场景中的潜力因此受到极大限制。

面向复杂业务流程的思维链推理与工具调用数据集构建严重不足：金融领域的核心业务，如信贷审批决策、资产配置建议或复杂衍生品定价，其业务逻辑并非简单的问答，而是需要模型具备执行多步骤推理以及与外部数据系统、业务执行系统进行交互的能力。当前行业内相对缺乏能够有效训练并评测模型此类高级能力的专用数据集。没有经过此类数据训练的模型，难以准确理解并顺序执行复杂的金融指令，无法保证业务流程的逻辑严谨性与最终结果的准确性，从而限制

了其在核心业务环节的深度应用。

【影响分析】

模型在金融垂直领域的应用效果不佳，价值难以彰显。由于缺乏高质量的专业数据进行训练与微调，模型对金融领域的专有术语、复杂产品与业务逻辑理解存在偏差，导致其在智能投顾、信贷审批、合规审查等核心场景中表现平庸，甚至出现事实性错误与内容幻觉，无法达到辅助决策或替代人工的预期目标，使得技术投入的商业价值难以实现。

自动化测试与验证体系缺失，模型风险难以管控。由于缺乏面向复杂推理与工具调用的高质量测试样本集，金融机构难以对模型的逻辑能力、安全红线与合规性进行系统、自动化的压力测试与持续监控。这导致对模型的行为缺乏充分的预判与控制，一旦模型在实际生产环境中出现误判或违规操作，将可能引发客户纠纷、资产损失或监管处罚，增大模型相关的操作风险与声誉风险敞口。

【应对策略】

策略 1：实现以终为始的技术能力现代化储备

为应对挑战并规避风险，技术能力的储备必须从传统的数据仓库和机器学习平台，向适应大模型范式的新一代技术架构升级。

构建统一的数据底座与务实的平台集成策略：依托湖仓一体架构，实现对结构化、半结构化乃至多模态数据的统一管理与高效处理，并利用自动化标注技术提升非结构化数据的标准化速率。同时，清晰规划机器学习平台、数据中台等现有资产与知识管理、LLM 研发运维等新型平台的集成关系。应根据自身资源禀赋与战略优先级，采取利旧与创新结合的务实路径，分阶段、有重点地构建核心能力域，避免重复建设，确保技术投入的精准性与高效性。

全面拥抱检索增强生成（RAG）架构：将 RAG 作为盘活内部私域知识的核心技术。通过构建企业级向量数据库，将内部海量的碎片化、非结构化文档与数据转化为模型可检索的知识库。模型在响应用户请求时，能够实时、精准地从该知识库中检索相关信息作为上下文，这不仅能极大提升回答的准确性与时效性，还能有效降低模型产生内容幻觉的风险，确保输出内容有据可循。

推行参数高效微调（PEFT）策略与合成数据生成：

- **PEFT 微调：**针对特定的、高价值的金融场景，采用 PEFT 技术对基础大模型进行轻量

级微调。该策略利用少量高质量的自有数据，即可获得在特定任务上表现卓越的领域专用模型。

· **合成数据：**为解决高质量训练数据不足的难题，可由金融专家提供少量高价值种子数据，利用大模型的数据生成与扩充能力，生产规模化的训练数据。此过程需建立严格的评估机制，确保合成数据的质量、多样性，并规避引入新的偏见。可借助小参数模型进行快速迭代实验，验证合成数据对场景性能的提升效果。

构建并治理面向工具调用的 API 框架：将机构内部的业务系统功能（如账户查询、下单交易、风险计算）封装为标准化的 API 接口，并建立一套严格的 API 治理与安全管控机制。同时，构建相应的训练数据，让模型学习理解何时、何种场景下以及如何正确、安全地调用这些工具接口。这是打通模型与实际业务流程、使其从“能说”到“能做”的关键一步。

策略 2：推动数据治理与 AI 治理体系的深度融合

数据治理必须超越传统的数据质量管理范畴，与新兴的 AI 治理框架进行深度融合，将数据治理前置到业务全流程。

实现从数据治理到知识治理的战略演进：将治理的焦点从孤立的数据项，提升到结构化的知识体系。目标是将碎片化的数据、零散的文档、内隐的规则与流程，整合为一个相互连接、可供模型高效理解与利用的机构级知识大脑。具体措施包含：建立覆盖知识全生命周期管理的责任矩阵（RACI），明确知识的版本控制、审核流程与反馈闭环，形成企业级的知识管理体系，构建能够持续自我优化的“知识飞轮”。

建立场景驱动的数据与模型迭代闭环：

· **场景化语料工厂：**锚定特定灯塔业务领域（如财富管理、风险控制），深度治理其所需的多模态数据，明确数据标准、质量要求与元数据规范，构建面向场景的领域推理数据集与语料工厂。

· **Agent 反馈闭环：**面向智能体（Agent）落地场景，建立领域思维链（CoT）的作业与反馈机制。通过系统化收集真实用户的显性反馈（如点赞、投诉）与隐性反馈（如操作路径），形成策略化的数据回流，用以持续优化模型性能与场景价值表现，远期形成以场景 Agent 交互数据为核心的、独特的模型能力禀赋。

将数据治理与 AI 模型风险管理一体化：将数据治理作为 AI 模型风险管理的第一道防线。在模型开发与应用的每一个环节，都必须嵌入对数据来源、数据质量、数据偏见、数据隐私与安全的审查与控制。确保输入模型的数据是高质量、无偏见且合规的，这是保障模型输出结果公平、

可解释、合规的根本前提，形成覆盖数据到模型的全链路风险闭环管理。

将数据治理前置到业务全流程：将数据治理体系与业务创新、产品研发、项目管理、IT 开发等体系有机融合，减少业务、科技、数据由于流程冗余带来的管理羁绊；从业务需求和商业论证之初就开始关注数据要素和数据治理的需求，从结果导向的数据治理逐步转变为源头治理。

策略 3：搭建以价值实现为导向的数据资产化框架

数据资产化的核心在于其价值的实现，而非形式上的盘点。为此，需要构建一个全新的、以驱动业务为目标的框架，盘活数据价值。

建立以业务价值贡献为核心的度量体系：数据资产的价值不应仅通过其规模或完整性来衡量，而应通过其在具体业务场景中创造的价值来量化。需建立一套清晰的度量体系，追踪并评估数据资产在支持大型模型应用、优化核心业务流程（如降低信贷风险、提升营销转化率）、改善客户体验以及增强合规能力等方面的具体贡献与投资回报率（ROI），以此驱动数据战略的持续优化与投入，并为数据资产的会计确认和计量（数据资产入表）提供可靠依据。

3.2.2 战略挑战：规划不清与价值验证困难引发的投资失衡

【具体问题】

（1）战略规划的前瞻性与全局性不足

金融机构在引入大模型技术时，其战略规划常表现出一定的局限性，未能充分实现前瞻性与全局性的统一。部分机构可能将大模型视为现有业务流程的补充或局部优化的工具，而未将其置于企业整体发展战略的核心位置。这种视角限制了对大模型颠覆性潜力的认知，导致战略规划多呈现为短期、分散的项目驱动模式，缺乏与公司长期愿景、数字化转型目标以及核心业务发展的深度协同。具体而言，战略规划的不足体现在以下方面：

总体战略与机构战略的协同缺失：金融业的总体战略日益强调普惠金融、风险合规与数字化运营。然而，部分机构在制定大模型战略时，未能将这些宏观导向与自身的差异化竞争优势和客群定位紧密结合。例如，大型金融机构需思考如何利用大模型提升全球资产配置和风险管理能力，而中小型金融机构则需探索如何借助大模型能力在特定细分市场或特色业务上实现突破。若战略脱离实际，将导致技术投入与业务发展脱节。

对技术发展趋势的认知滞后：当前大模型技术正朝着多模态、轻量化、自主智能体等方向快

速演进。若金融机构的战略规划仍停留在对通用文本生成等基础能力的认知上，将错失利用多模态融合、检索增强生成等技术提升金融场景的覆盖度与可控性，以及部署智能体自动执行复杂金融任务的机遇，从而在未来的市场竞争中处于被动地位。

场景选择的广度与深度不足：在应用场景的选择上，许多机构倾向于从成熟度较高的智能客服、营销文案生成等领域切入。这虽然降低了初期落地风险，但也限制了价值创造的空间。战略层面需要更具前瞻性的布局，系统性规划大模型在风险管理、量化投资、合规审查、产品创新乃至组织管理等核心业务领域的应用路径，形成由点及面的推广策略，避免应用场景的碎片化和浅层化。

未充分考虑组织与文化的协同变革：大模型的引入不仅是技术升级，更是一场组织变革。战略规划若忽视了对现有组织架构、人才体系、工作流程和企业文化的系统性重塑，将导致技术难以融入业务。例如，数据科学家、业务专家与 IT 工程师之间的协同机制、AI Agent 上线后与业务专家的变革融合等。同时，鼓励创新、允许试错的文化氛围，是保障大模型战略成功落地的基础。

（2）价值实现的路径模糊与效益评估的复杂性

由于大模型的投入成本高、技术复杂性强，且其影响深远，传统的项目评估方法难以完全适用。价值实现的路径模糊和评估体系的缺失，导致机构在决策时犹豫不决，在实施后难以衡量成效。具体而言，这一挑战体现在：

短期效益与长期价值的平衡困难：金融机构的决策往往受短期财务指标驱动。大模型的价值释放通常需要较长周期，其在优化客户体验、重塑品牌形象、激发组织创新活力等方面的长期、无形价值，难以通过传统的投资回报率（ROI）等短期指标来衡量。过分关注短期效益，可能导致机构放弃对具有长远战略意义但短期见效慢的项目的投入。

间接效益与隐性成本的量化难题：大模型带来的效益通常是间接的，例如，通过提升客户满意度来增加客户粘性，或通过优化风险模型来降低未来的潜在损失。这些间接效益难以精确量化并归因于特定的技术投入。同时，隐性成本，如数据治理成本、模型持续维护与迭代成本、合规风险管理成本以及组织变革带来的摩擦成本，也容易被低估。

缺乏统一的、多维度的评估框架：目前业界尚未形成一套公认的、适用于大模型项目的价值评估标准。金融机构内部往往也缺乏一个能够整合财务指标、业务指标、客户指标和技术指标的多维度评估框架。评估维度的单一化，使得对大模型项目价值的判断出现偏差，无法全面反映其综合贡献。

动态调整与持续优化的机制缺失：大模型项目并非一次性交付的工程，而是一个需要持续迭代和优化的动态过程。市场环境、客户需求和技术本身都在不断变化。如果缺乏一个动态的价值评估与调整机制，项目可能会偏离最初的目标，或者无法根据反馈进行及时优化，从而影响最终的价值实现。

【影响分析】

缺乏前瞻性与全局性的战略规划，将导致大模型应用陷入项目孤岛的困境。机构内部不同部门基于自身需求各自为战，进行重复性的技术探索与系统建设，不仅造成了资金、算力和人才等核心资源的巨大浪费，更形成了新的技术壁垒，增加了未来系统整合与数据贯通的难度和成本。更严重的是，这种碎片化的应用无法汇聚成体系化的、难以被竞争对手模仿的核心能力，使得技术投入仅仅停留在对现有流程的点状优化，而非驱动业务模式变革的结构性重塑，最终导致机构在由 AI 定义的未来金融竞争格局中，因反应迟缓而丧失战略主动权与市场先机。

价值路径的模糊与评估体系的缺失，会直接削弱大模型项目获取持续内部支持。即项目虽有初步成果但因价值难以清晰阐释而无法获得推广资源，最终不了了之。这不仅导致前期的技术投资无法转化为可衡量的业务成果，形成了沉没成本，更重要的是，它阻碍了技术与业务的深度融合。当业务部门无法感知到技术的明确价值时，其参与意愿和协作深度将大打折扣，导致模型应用场景无法切中真实痛点，长此以往将动摇组织推动深度变革的信心，使数智变革流于表面。

【应对措施】

策略 1：构建与企业战略相匹配的大模型战略体系与机制

构建与企业战略相匹配的大模型蓝图：将大模型规划提升至公司战略高度，确保其与机构的长期发展目标、数字化转型路径及核心业务策略同频共振。明确大模型在实现普惠金融、提升风险管理能力、优化客户体验等关键战略议题中的角色和贡献，并为不同规模和定位的机构制定差异化的战略重点。

建立动态的技术认知与评估机制：组建跨学科的专业团队，持续追踪多模态模型、RAG、智能体、轻量化微调等前沿技术的发展，并定期评估其在金融场景中的适用性和潜在价值。通过技术研讨、外部合作等方式，保持战略决策层对技术趋势的敏锐洞察力，确保战略规划的先进性。

系统性规划与分阶段实施应用场景：制定一份覆盖前、中、后台业务的全景式应用地图，并根据业务价值、技术成熟度和数据可用性等维度，确定场景落地的优先级和时间表。采取价值引领、

试点先行、逐步推广的实施路径，先在核心业务领域打造可复制的成功案例，再逐步扩展至更广泛的业务范围，形成规模化效应。

推动组织与人才的协同发展：将组织变革作为大模型战略的重要组成部分，建立敏捷的、跨职能的合作团队，打破部门壁壁垒。同时，制定系统性的人才培养计划，引进顶尖 AI 人才，并对现有员工进行技能培训，培育既懂技术又懂业务的复合型人才，营造支持创新的企业文化。

策略 2：建立健康、适度的价值评估体系

制定兼顾长短期的价值实现路线图：围绕核心业务痛点和战略机遇，设计清晰的价值实现路线图。将最终的战略目标分解为一系列可衡量、可实现的中短期业务目标（OKRs），如将“提升客户体验”分解为“降低客户平均等待时长”、“提升问题首次解决率”等具体指标，从而将长期价值与短期效益相结合。

采用“测试与学习”的敏捷评估方法：对于创新性强、价值不确定性高的项目，采用敏捷的测试与学习方法。通过小规模试点快速验证商业假设，并利用试点数据来校准和优化价值评估模型。这种方法有助于降低不确定性，并为更大规模的投入提供数据支持。

建立持续的价值追踪与反馈循环：将价值评估嵌入项目的全生命周期管理中，从项目立项、过程监控到事后复盘，进行持续的价值追踪。建立定期的沟通与汇报机制，向管理层和业务部门清晰地展示项目进展与阶段性成果，并根据反馈及时调整项目方向和资源配置，形成一个闭环的价值管理体系。

构建多维度的价值评估模型：建立一个超越传统财务指标的综合性价值评估模型。具体体现为：

- **统筹成本管理：**

⊙ 建立全生命周期成本核算机制：将算力资源、模型微调、安全对齐等成本纳入财务模型，明确成本归属部门，实现统一管理和核算。可借鉴全生命周期成本（TCO）模型，全面评估大模型应用的总拥有成本。

⊙ 建立跨部门成本分摊机制：针对不同成本类型，制定合理的成本分摊机制。例如将安全对齐成本按场景业务线使用量分摊。

- **管控和业务双指标体系牵引：**

⊙ 构建量化的效益评估指标体系：区分决策类场景和非决策类场景，制定差异化的评估指标。决策类场景关注模型对业务决策的提升效果，如风控场景中的违约率降低比例、风险识别准确率提升幅度等。非决策类场景重点关注模型对工作效率的提升效果，如单位人效、流程效率等。

- **评估模板示例：**

评估模板一：内部效能提升类场景 ROI 计算模板

- **适用场景：**主要通过节约人工时、提升工作效率来创造价值的场景。

- **核心计算逻辑：**将节省的工时量化为人力成本。

测算维度	计算项	计算公式 / 说明	金额（万元）
年化收益（A）	年化人力成本节约	= (B) * (C) * (D) * 12 月 / 10000	
	（B）单次任务节省工时（小时）	访谈相关岗位员工，评估 AI 辅助前后单次任务平均处理时长的差异。	
	（C）月均任务发生次数	统计相关业务系统的月均业务量或相关岗位的月均任务处理量。	
	（D）相关岗位平均小时薪酬（元）	= 岗位年均总人力成本（含薪酬、福利、社保等）/（年工作日 * 8 小时）。	
年化成本（E）	年化总成本	= (F) + (G)	
	（F）一次性投入成本（年化分摊）	=（项目开发成本 + 首次数据处理成本）/ 预计使用年限（建议 3 年）。	
	（G）年化持续运营成本	= 年化平台资源分摊成本 + 年化人工维护成本。平台资源分摊成本由 AI 平台运营方根据模型调用量、算力消耗等进行核算。	
核心指标	投资回报率（ROI）	=（A − E）/ E * 100%	
	投资回收期（月）	= F /（A − G）* 12	

评估模板二：风险控制与合规增强类场景 ROI 计算模板

- **适用场景：**主要通过降低风险损失、减少合规罚款来创造价值的场景。
- **核心计算逻辑：**将规避的损失或节省的成本进行量化。

测算维度	计算项	计算公式 / 说明	金额（万元）
年化收益（A）	年化风险损失规避	= (B) * (C) * (D)	
	（B）相关业务年交易 / 管理总额	统计场景相关的业务总规模。例如，客户流失预警场景中，为目标客群的总资产管理规模（AUM）。	
	（C）预计风险发生率降低值	= 部署前风险发生率 - 部署后预计风险发生率。需基于历史数据和模型回测结果进行预估。	
	（D）风险事件平均损失率	每次风险事件造成的平均资金损失比例。	
	或 年化合规成本节约	如场景为提升合规审查效率，可参考模板一计算人力成本节约。	
年化成本（E）	年化总成本	= (F) + (G)（计算方法同模板一）	
核心指标	投资回报率（ROI）	= (A - E) / E * 100%	
	投资回收期（月）	= F / (A - G) * 12	

评估模板三：业务增长与创收类场景 ROI 计算模板

- **适用场景：**主要通过提升营销转化率、增加客户价值、创造新收入来源来创造价值的场景。
- **核心计算逻辑：**将新增的业务收益进行量化。

测算维度	计算项	计算公式 / 说明	金额（万元）
年化收益（A）	年化新增业务收益	= (B) * (C) * (D)	
	（B）目标客群规模	项目覆盖的客户总数。	
	（C）预计转化率提升值	= 部署后预计转化率 - 部署前基线转化率。需基于 A/B 测试或小范围试点结果进行预估。	
	（D）单客年均贡献价值（元）	客户转化后，预计在一年内为银行带来的平均利润或收入。	
年化成本（E）	年化总成本	= (F) + (G)（计算方法同模板一）	
核心指标	投资回报率（ROI）	= (A - E) / E * 100%	
	投资回收期（月）	= F / (A - G) * 12	

3.2.3 应用挑战：严监管场景对模型可控性的极高要求

【具体问题】

大模型在应用中生成的内容与客观事实不符或缺乏事实依据的现象，即模型幻觉，是其在金融领域落地的核心障碍之一。幻觉产生的根本原因在于模型固有的技术局限，包括训练数据的压缩损失与潜在矛盾、特定领域知识的覆盖不足、模型对复杂逻辑的理解能力有限，以及推理过程中固有的概率性。这些通用原因在金融领域被显著放大，构成了更严峻的挑战：

金融数据的极端复杂性与高噪音：金融数据不仅包含结构化的财务报表，还涵盖大量非结构化的法律文件、监管公告与新闻舆情。这些数据通常具有低信噪比、高时效性、专业术语晦涩等特点，大幅增加了模型准确理解和推理的难度。

金融知识体系的快速迭代：金融市场、产品工具及监管法规均处于高速动态变化中，模型依赖的静态训练数据极易过时，无法及时反映最新的市场状态或监管要求，从而产生与现实脱节的输出。

对跨源信息综合研判的刚性需求：金融决策，如信贷审批或投资分析，往往需要模型具备跨越多个段落、甚至多份独立文档进行信息整合、逻辑推理和一致性检验的能力。当前模型在处理此类任务时，信息丢失、逻辑断裂或错误推断的风险较高，容易诱发幻觉。

【影响分析】

模型幻觉与可控性不足对金融机构的负面影响是深远且多层次的：

直接的业务与财务风险：在信贷审批、资产评估、投资决策等核心业务环节，幻觉可能导致错误的信用评级、资产定价或投资建议，直接引发信贷违约、投资亏损等财务后果。即使是极低概率的错误，在金融杠杆的放大下也可能造成严重损失。

严峻的合规与法律风险：金融业受到严格监管，对信息披露的准确性、完整性和公平性有极高要求。模型的幻觉内容一旦被用于客户报告、信息披露或监管报送，即可能构成虚假陈述或误导，引发监管机构的巨额罚款。同时，基于错误信息向客户提供建议可能导致法律诉纷，严重损害机构声誉。

用户信任的侵蚀与战略推进的迟滞：当客户或内部使用者发现 AI 系统提供的信息频繁出错、不可依赖时，会迅速丧失对机构数字化和智能化能力的信任。这种信任赤字不仅会阻碍当前 AI 工

具的采纳，更会影响机构整体数智变革的战略布局，导致在人工智能领域的重大投入无法转化为预期的业务价值。

【 应对措施 】

策略 1：技术层面的解决方案

高级检索增强生成： 替代传统的 RAG 技术，采用如知识图谱检索增强生成等更先进的架构。通过构建连接内部碎片化信息的知识图谱，模型可以进行更精准、更具关联性的信息检索，有效解决因信息不完整或过时导致的幻觉问题，尤其适用于需要综合多份文档进行分析的复杂金融场景。

直接偏好优化（DPO）： 作为新一代模型对齐技术，DPO 通过直接在偏好数据上进行优化，替代了传统 RLHF 中复杂的奖励模型训练环节。这使得模型微调过程更稳定、高效，能够更可控地引导模型生成符合金融行业规范、价值观和监管要求的专业内容，显著提升输出的可控性。

系统性评估与验证： 建立常态化的模型评估机制，采用专为金融领域设计的、开放的、可复现的评估基准，以能够对模型在真实金融任务中的事实一致性、知识准确性进行全面、量化的评估，作为模型上线前和运行中持续监控其可靠性的重要依据。

模型协同与解耦： 构建模型协同工作的体系，将具备强大通用推理能力的基座模型与经过专门领域知识训练的、小而精的垂直领域模型相结合。在处理复杂任务时，由基座模型负责逻辑分解与规划，再调用多个专家模型完成具体的、高准确性要求子任务，最后进行结果汇总。这种策略确保了分析的深度与执行的精度。

策略 2：管理层面的保障措施

建立系统的 AI 模型风险管理框架： 将传统的金融模型风险管理体系扩展至 AI 领域，建立覆盖模型全生命周期的治理框架，替代简单的业务兜底机制。该框架应包含以下核心支柱：

- 模型开发与文档化：制定严格的模型开发标准，并要求对数据来源、模型设计、训练过程、已知局限性等进行全面、透明的文档记录。

- 模型清单与集中化管理：建立全机构统一的模型清单，对所有线上 AI 模型进行集中化追踪、分类和风险评级。

· 独立的模型验证：在模型部署前及运行期间，由独立于开发团队的部门进行验证，包括性能测试、稳定性测试、偏见检测和稳健性压力测试。

· 持续监控与审计：部署自动化工具，对模型的实时表现、数据输入分布、输出结果进行持续监控，及时发现性能衰退或数据漂移，并保留完整的审计日志。

· 明确的角色与职责：清晰界定模型所有者、使用者、开发者和验证者的职责，确保问责机制的有效落地。

严格的第三方模型治理：对于从外部供应商采购的 AI 模型，特别是“黑箱”模型，必须建立专门的治理流程。这包括对供应商进行深入的尽职调查，在合同中明确要求其提供详尽的模型文档、解释性工具和性能数据，并保留机构内部进行独立测试与验证的权利，以管理供应链风险。

强化人机协同与人工审核闭环：在自动化决策流程中嵌入关键的人工审核节点，特别是在高风险或核心决策场景。AI 的输出应被视为对人类专家的决策辅助，而非最终决策本身。这构成了最后的防线，确保所有输出在交付或执行前都经过了人类专家的审核与确认，满足金融监管对审慎经营的要求。

3.2.4 能力挑战：技术迭代提速倒逼组织变革与人才升级

【具体问题】

（1）复合型人才瓶颈

战略规划与治理人才的缺失：缺乏能够洞察大模型技术发展趋势，并将其与金融机构总体战略、风险偏好、合规框架相结合的领导者。该人才需要制定企业级 AI 治理体系，平衡创新与风险，确保技术应用符合监管要求。

模型应用与业务融合人才的断层：业务团队与技术团队之间存在认知鸿沟。业务专家通常不了解大模型的能力边界与技术细节，技术专家则往往对金融业务的复杂逻辑、合规要求和风险控制点缺乏深入理解，导致研发出的工具无法紧密贴合实际业务流程。

模型持续运维与迭代人才的不足：大模型的有效落地不仅是初期的开发与部署，更在于后期的持续监控、评估、迭代与优化。机构普遍缺少能够对模型性能进行长期跟踪，处理模型幻觉，管理数据漂移，并根据业务反馈进行敏捷迭代的专业运维与算法优化人才。

（2）跨部门组织协同的系统性障碍

传统 IT 架构与敏捷开发模式的冲突：金融机构普遍依赖的、以稳定性和安全性为核心的传统 IT 架构，其迭代周期长、流程僵化，难以支撑大模型应用所要求的快速迭代、持续集成的敏捷开发与运维一体化模式。这种结构性冲突导致模型从开发到部署的周期被大大拉长。

敏捷模式下跨团队协同障碍，任务目标与激励机制的不兼容：大模型项目的成功依赖于底层平台、数据中台、业务应用和风险合规等多个团队的紧密协作。但各团队的考核指标往往是独立的，甚至相互矛盾。例如，业务团队为快速上线而选择牺牲一部分模型通用性，这与平台团队追求架构长期可扩展性的目标直接冲突，导致决策内耗与资源争抢。

短期追求背后的预算难以平衡：大模型应用通常兼具平台级投入与应用级产出的属性。但在传统的预算审批与组织架构下，用于构建通用能力的基础性投入，难以被归属到任何一个独立的业务部门。各个业务线更倾向于申请用于开发本部门应用的短平快项目预算，使得支撑长远发展的、跨部门共享的基础设施建设停滞不前。

（3）场景上线后对组织运营与流程架构的冲击与变革

人机交互模式的根本性重塑：原有基于固定规则和线性流程的岗位，将被全新的人机协同模式所取代。员工的角色从流程的执行者，转变为 AI 工具的使用者、监督者与优化者。例如，理财经理需要学会利用 AI 生成的投资建议，并结合自身专业判断与客户进行更高质量的沟通。这要求员工具备全新的技能组合，而现有的岗位说明、培训体系和能力模型已失效。

决策责任与风险归属的模糊化：在 AI 辅助决策的场景中，一旦出现错误，责任归属变得异常复杂。是批准 AI 建议的一线员工、设计模型的算法团队，还是提供数据的平台部门应承担责任。这种责任链条的模糊化，不仅会引发内部权责纠纷，更可能导致无人敢于在关键节点做出决策，使得 AI 应用难以在核心业务中发挥作用。

现有业务流程与组织架构的失效：将强大的 AI 工具塞入为人工操作而设计的旧有流程中，不仅无法发挥其最大效能，反而可能因为流程断点而导致效率进一步降低。大模型的应用要求对整个业务流程进行端到端的重构，这必然会触及甚至打破原有的部门墙与组织架构。

【影响分析】

关键人才的缺失将导致三个层面的负面影响。**第一，战略失焦与资源错配；第二，应用落地水土不服；第三，运营风险与合规风险积聚。**

组织协同瓶颈将使得应用重复建设，并不断累积技术债。同时对基础平台投入的系统性不足，使得机构的 AI 能力始终停留在对单个应用的修补上，无法形成规模化、体系化的创新能力，逐渐丧失长期竞争力；协同内耗拖垮项目进程。团队间因目标冲突而产生的持续拉扯，将大量时间与精力消耗在内部协调而非价值创造上，使得项目周期被无限拉长，错失市场机遇。

场景上线即性能巅峰，无法充分发挥大模型的演进特性。如若不及时调整岗位技能与工作模式，将导致员工无法有效使用新工具，造成技术投资的浪费，并因技能恐慌而产生对变革的抵触情绪；权责不清将导致无人敢于在信贷审批、风险交易等核心环节依赖 AI，使得大模型应用被局限在非关键的、外围的场景，无法触及真正的价值核心；颠覆性的流程重构必然触动部门利益，若无强有力的顶层推动与清晰的变革管理，极易在中途受阻，导致整个数智变革战略搁浅。

【 应对措施 – 面向人才 】

策略 1：实施分层分类的、与业务场景强绑定的培养计划

面向管理者：设计 AI 战略与治理课程，聚焦于大模型的商业价值、应用边界与风险管理，提升其战略决策与顶层设计能力。

面向业务专家：开展 AI 赋能业务工作坊，通过真实案例与沙盘演练，使其掌握如何识别业务中的 AI 应用机会，并能与技术团队进行高效沟通。

面向技术人才：启动金融领域知识强化项目，使其深入理解特定金融场景的业务逻辑与合规要求，确保技术方案的业务适切性。

策略 2：构建内培外引并重、以项目实践为核心的人才发展生态

与外部顶尖 AI 公司或学术机构建立战略合作，定向引进成熟人才以快速补齐短板。同时，设立内部创新基金与真实业务场景挑战赛，激励内部员工组建跨职能团队，在解决实际问题的过程中，将外部知识与内部经验相融合，加速复合能力的养成。

【 应对措施 – 面向组织 】

策略 3：建立由高层领导的、具备资源与决策权威的虚拟项目组

针对战略级大模型项目，成立由高级管理层直接领导的、跨职能的专项任务小组。该小组被授予独立的预算审批权与跨部门资源协调权，其唯一目标是确保项目的最终成功。通过设定统一的、

贯穿所有参与团队的共享 OKR，将所有人的利益与最终业务成果绑定，从根本上解决激励不兼容问题。

策略 4：推行平台即服务的内部运营模式

将数据、模型训练、合规检查等通用能力，作为标准化的内部服务，由专门的平台团队负责建设与运营。业务应用团队则作为平台的用户，通过调用服务来快速构建上层应用。平台团队的考核指标与其服务的稳定性、易用性以及被业务部门调用的频率挂钩，从而激励其主动提升平台能力，形成良性循环。

策略 5：实施嵌入式的风险与合规协同机制

将风险、法务与合规专家从项目启动初期就作为核心成员嵌入敏捷开发团队。他们不再是项目末端的审查者，而是在需求分析、数据处理、模型设计的每个环节提供实时指导的共建者。这种模式将合规要求内化为产品设计的固有属性，极大提升了研发效率，避免了因后期发现重大问题而推倒重来的风险。

【 应对措施 – 面向变革 】

策略 6：面向流程嵌入型应用，实施以人机协同为核心的流程再造

对于将大模型作为增强工具嵌入现有业务流程的场景，变革的重点是进行精细化的流程再造。

具体措施：核心是重新定义流程中人与 AI 的交互节点、各自的权责边界以及信息传递方式。需明确哪些环节由 AI 自动完成，哪些环节必须由人工复核，以及人工干预的触发条件。同时，必须配套建立以人机协作效能为导向的新考核体系。

策略 7：面向流程颠覆型应用，推动以终为始的系统性组织变革

对于大模型能够端到端重塑甚至完全替代原有核心业务流程的场景，则必须进行更为彻底的系统性组织变革。

具体措施：这类变革需要由最高管理层直接驱动，其核心不再是优化局部流程，而是基于未来业务形态，重新设计组织架构。可能涉及撤并原有职能部门，建立全新的、跨领域的 AI 运营与监督中心或人机协同作战单元。变革的成功与否，取决于能否打破部门壁垒，重构预算与资源分配机制，并建立与全新组织形态相匹配的、以最终业务价值为衡量标准的顶层考核体系。

策略 8：启动以人机协同为核心的岗位重塑与赋能计划

开展未来岗位画像分析：联合业务与人力资源部门，提前研判受 AI 冲击最严重的岗位，并重新设计其在人机协同模式下的核心职责、能力要求与绩效衡量标准。

设计场景化、伴随式的赋能项目：摒弃一次性的通用培训，转而开发与新工具、新流程强绑定的线上学习模块与实操演练，让员工快速掌握与 AI 协同工作的新技能。

策略 9：建立清晰的 AI 伦理与决策责任框架

在 AI 应用上线前，必须由 AI 治理委员会牵头，联合业务、法律、合规部门，共同制定并发布清晰的 AI 伦理准则与人机决策责任划分矩阵。该矩阵需明确定义不同场景下，AI 的决策权限边界、人工审批的层级与标准，以及出现问题后的追责流程。这为一线员工提供了清晰的行动指引与心理安全保障，是确保 AI 在核心业务中被放心、大胆使用的前提。

3.3 金融业大模型落地实践案例与洞察

3.3.1 智能理财助理——从低风险场景切入，实现价值快速验证

背景：本案例聚焦的智能理财助理系统，以生成式大模型为核心引擎，构建 AI 主导 + 人工辅助的对话式服务模式。通过整合用户交易数据、业务知识库及实时 API 接口，重点解决传统客服系统在复杂多轮对话、个性化服务响应、操作合规性等方面的不足，旨在打造具备拟人化交互、超预期增值服务能力的智能理财顾问。

场景定位：根据用户工单数据分析，基金赎回场景呈现三大特征：其一，用户诉求集中于封闭式高频问题，典型问题集中度显著；其二，问题类型以客观事实确认为主，涉及账户状态查询、到账时效确认等可量化信息；其三，相比其他业务场景，该场景的合规风险系数较低。基于此业务特性，选择赎回场景作为首期突破点，既能验证技术可行性，又能有效控制风险敞口。

场景难点：

- 回答的可控生成

⊙ 需同步处理用户交易记录（含时间戳、金额等数值型数据）、知识库结构化条款、FAQ 非结构化文本，模型在数值推理环节易出现计算误差；

⊙ 训练数据缺失，已有的客服工单数据多为按照知识库内容进行回答，不符合场景要求；

⊙ 开放式对话的产品形态无法限制用户问题聚焦在赎回范围内。

· RAG 精准召回

⊙ 金融场景专业术语多，常面临用户提问时用词有歧义、错别字、意图模糊等情况，直接影响传统检索系统召回准确率；

⊙ 多轮意图漂移，连续对话中存在较高的意图转移概率，简单拼接对话历史将导致意图识别准确率下降，需要结合上下文总结出当前用户的真实意图。

总体思路： 前端部署查询重构模块实现意图净化，后端建立分层知识体系。通过构建业务逻辑框架，将知识库按产品规则、交易流程等维度分类，有效降低信息冗余。

专题解决办法：基座模型问题

除了基座模型外，其他模块的问题都能够收敛解决并且能够迅速优化上线，最难解决的是基座模型问题，其缺乏基金赎回的业务知识（如活期产品和零钱产品的关系），模型还没有按照业务要求的逻辑和关键点来回答问题，因此需要将业务逻辑注入到模型中，并对其进行微调。

将基金赎回规则转化为链式推理模板，构建包含典型场景的微调数据集。通过控制微调数据比例，保持模型在开放域问答中的通用能力。建立动态知识提示机制，将关键业务参数固化至系统提示模板，确保业务规则遵守率。

专题解决办法：高质量训练数据缺失

面对用户需求的高度复杂性，为保障应答质量需实现三重目标：保持基座模型的通用推理与指令跟随能力，提升场景专业化水平，同时满足拟人化话术要求。基于不同语料储备情况，可制定差异化训练策略：在语料充足时实施增强预训练；语料不足时采用大模型数据蒸馏技术结合人工标注生成大规模数据，并通过调整通用数据与业务数据配比进行有监督微调。核心原则是确保通用能力与场景需求均具备高质量数据支撑，为此重点引入数据合成技术实现高效生成优质场景数据。

现实挑战主要来自三方面：真实用户对话样本稀缺、业务专家资源受限、样本快速生成能力不足。针对这些瓶颈，提出并进方案：一方面通过外部知识检索系统增强生成准确性；另一方面构

建专家思维范式，将业务逻辑转化为可执行的思维链，结合动态检索召回机制辅助模型生成。该方案的实施前提是必须保证微调后模型在逻辑推理与指令跟随层面的基础能力不衰减。

专题解决办法：大模型幻觉

金融场景直接面向 C 端落地需应对双重挑战：既要满足严格的合规要求与数值精确性（尤其是涉及金额计算的场景），又要克服大模型固有的幻觉。由于对话式产品存在开放式提问特性，用户可能提出超出模型能力边界的问题，强行应答将显著增加错误风险。为此构建系统性防控体系：

通过三级递进式反思框架严控幻觉生成，首层执行可应答性判断以明确问题处理边界，第二层监控推理链条的逻辑完备性，第三层实施合规与数值精确性终检，形成从问题识别到结果输出的全链路校验。同步建立多级意图识别体系提升生成精准度，设立其他类意图容器归集非赎回类基金咨询、业务无关闲聊等长尾请求，结合意图 -API 动态映射机制降低计算复杂度，聚焦核心业务需求响应。

经验洞察：

- 场景价值与技术可行性验证：通过大模型与知识库融合方案实现技术可行性验证，在用户体验维度实现突破性提升，具体表现为精准场景适配性、个性化服务能力及 7×24 小时持续性响应优势；

- 能力协同原则：复杂多轮对话机制在强化场景专业能力的过程中，必须确保基座模型核心能力（通用推理、指令解析等）的稳定性，这是实现精准用户意图理解与有效响应的基础前提；

- 系统工程方法论：大模型应用需构建涵盖数据工程、算法优化、系统架构的完整技术体系。实施路径建议采用单点突破策略：优先在核心业务场景实现极致优化，完成方法论沉淀、技术框架验证及流程标准化建设后，再有序推进多场景扩展，避免盲目追求规模速度；

- 金融应用约束：鉴于金融领域的高严谨特性，需正视大模型存在的幻觉生成、计算偏差及知识边界模糊等技术瓶颈，必须建立兜底方案；

- 迭代协同机制：构建问题驱动 - 算法优化 - 产品创新的闭环迭代体系，通过产研协同机制实现能力迭代与功能补位。在模型能力边界外场景，依托产品功能设计实现技术短板的创造性补偿，形成模型能力与产品功能的动态平衡。

3.3.2 财富管理风控——用领域 LLM 攻克传统 AI 的语义理解难题

背景：在用户需求多元化、监管要求趋严的背景下本场景旨在探索如何利用大模型技术，实现对理财顾问对话内容的实时监测与风险预警，有效管理金融风险。

场景定位：本方案聚焦理财顾问对话场景的合规风险智能识别，重点解决三个核心问题：第一，在维护客户关系与推动业务增长过程中，如何精准识别理财顾问在服务高净值客户时可能产生的合规风险（包括敏感话题言论、诱导修改风测问卷等）；第二，针对口语化对话中存在的语义模糊、上下文缺失等特征，建立适应自然语言场景的风险判定模型；第三，构建可扩展的风险识别体系，满足动态调整的监管要求与业务发展需求。

场景难点：

其一，月均数十万条对话记录存在显著的自然语言处理难点，包括文本口语化特征显著、语义模糊性突出、主观表述密集，以及对话上下文信息不完整等问题；其二，金融风险判定存在高度复杂性，其判断标准涉及多维模糊边界（如暗示性表述、未经证实的信息、缺乏客观数据支撑及误导性预期等特征），传统上需依赖专业人员的经验型判断。

从技术实现层面审视，传统机器学习模型面临两大核心障碍：首先，监督式学习范式需消耗大量人工标注资源，存在标注成本高企与效率瓶颈；其次，模型泛化能力受限于训练数据覆盖度，难以满足金融场景对风险判定的精准性要求。以下为典型风险判例：

- 使用了“预期收益、预期收益率”禁止性词汇；

- 预测基金业绩是对基金未来收益率、净值表现或者市场排名等进行的预测或者承诺，包括：（1）业绩保证：承诺基金将会达到某个具体的收益率或者净值水平；（2）明示或暗示保证：直接或通过含糊其辞的方式暗示基金将会有良好的表现；（3）未来业绩预测：预测基金未来业绩，而没有明确依据且明确指出这只是一种可能性；（4）排名预测：预测基金在未来某个时期内的市场排名或比较基准的表现；（5）未经证实的声明：发布或传播未经证实的基金业绩预测信息。

- 在无客观数据支持下预测或评论个股、行业的业绩；

- 承诺提供符合客户收益率要求的理财产品，误导客户对产品收益的期望

解决思路：两种方案并行

方案一：基于金融大模型 +Prompt[10]（风险定义、典型案例）进行识别

依托金融领域增强预训练模型，通过融合金融法规文本、风控管理案例等专业语料进行领域适配训练，并针对具体场景实施监督式微调（SFT）。验证测试表明，经专项优化的金融大模型在风控场景具备显著效能优势。具体实施时，采用风险定义框架与典型违规案例构建动态提示模板，通过增量式注入风险特征描述与负向样本实现 prompt 迭代升级。但随着负向案例积累导致的 prompt 复杂度指数级增长，模型幻觉现象加剧，最终引发识别准确率边际效益递减的技术瓶颈。

方案二：通过作业与反馈机制构建有监督微调数据集迭代优化

针对方案一中负样本无法解决的问题，构建了数据飞轮的反馈链路，首先建立人工校验标注 - 模型反馈的机制，将专家确认的高价值案例转化为训练数据，其次对原始对话进行语义扩展与对抗样本生成，自动产出合规判定依据说明文本，以此提升微调效果，案例解决率大大提升。

经验洞察：

- 通用大模型在垂直场景存在显著领域适配局限。比如在理财师对话风险识别中，通用大模型识别效果没有达到使用预期，而金融大模型由于在增量预训练阶段增加金融法律法规、风险管理书籍、考试等语料，在真实风险识别中表现出较好的能力；

- 过去机器学习等传统解决方案往往需要大量样本和建模调优时间，而大模型 +few-shot Prompt 的应用范式，相比数据标注与传统模型，验证了在风险识别，尤其是自然语言（客服、工单等）场景下，大模型具有明显优势，模型准确率及效率可以大幅提升，新模型上线效率从月级到周级；同时能够解决传统技术方案无法解决的少样本甚至无样本、识别效果差、解释性差等的问题；

- 当处于正负样本数据缺失的环境中，合成数据与作业与反馈机制可助力大模型突破效果瓶颈，成为模型迭代的有效手段。

3.3.3 超级保险代理人——AI 重塑展业与培训新范式

背景：保险代理人渠道作为连接保险公司与客户的核心桥梁，其专业能力与服务效率直接决定了客户体验与业务增长。然而，传统代理人模式普遍面临展业效率不高、新人培养周期长、服务质量难以标准化、合规风险管控难度大等多重挑战。尤其在制作保险建议书环节，往往耗费大

10 Prompt：提示工程

量时间进行客户信息整理、需求分析和产品匹配，且难以确保每一次输出都兼具个性化与合规性。因此，某领先寿险公司启动 AI 赋能项目，旨在利用大模型技术，系统性提升代理人渠道的整体效能。

场景定位：本案例聚焦于构建一个 AI 代理人智能作业平台，核心解决代理人作业流程中的两大关键痛点：一是通过智能建议书生成功能，大幅提升展业效率与方案专业度；二是通过 AI 模拟销售对练功能，缩短新代理人的成长周期，强化专业销售技能。该平台并非单一工具的集合，而是旨在将 AI 深度嵌入代理人从学习、准备到展业的全流程，实现从辅助工具到智能伙伴的角色跃迁。

场景难点：将大模型应用于保险建议书生成与销售对练场景，需克服四重核心挑战：其一，动态知识的精准应用。保险产品条款、核保规则、费率表等知识体系复杂且更新频繁，模型必须能够实时、准确地调用最新信息，任何细微的错误都可能导致方案失效或客户误解。其二，个性化与合规的平衡。一份高质量的建议书，既要深度契合客户独特的家庭结构、财务状况和风险偏好，又必须严格遵循监管部门和公司的合规要求，规避任何形式的误导性销售言论。其三，复杂任务的逻辑推理。建议书的生成是一个严谨的逻辑链条，涉及客户画像构建、保障缺口量化分析、多产品组合策略、保费精算等多个步骤，对模型的多步推理和数学计算能力提出了极高要求。其四，交互式辅导的真实感与有效性。在销售对练中，AI 不仅要扮演一个具有真实情感和异议的客户，还要能作为专业教练，对代理人的话术、逻辑和情感表达进行精准评估，并提供具体、可行的改进建议。

解决思路：为应对上述挑战，采用了知识增强与流程编排相结合的总体解决思路。技术路线上，构建了一个由通用大模型、领域微调模型和规则引擎协同工作的混合智能架构。首先，以检索增强生成为核心，构建覆盖产品、合规、销售全流程的动态知识库。这是确保所有输出内容准确、合规的基石，模型在执行任务时，被强制要求从该知识库中检索信息作为决策依据。其次，将复杂的业务流程进行拆解，通过多智能体协作的模式执行。例如，建议书生成任务被分解为客户信息理解、需求分析、产品匹配、文案生成、合规审核等多个子任务，由不同但协同工作的智能体完成，确保了流程的严谨性和专业性。

专题解决办法：以智能建议书生成为例

在建议书生成模块中，为确保最终输出的质量，系统性地解决了知识注入和逻辑可控性问题。在知识注入层面，项目团队构建了一个多模态知识库，将 PDF 格式的产品条款、说明书，Excel 格式的费率表，以及 Word 文档形式的销售话术、异议解答脚本等非结构化与半结构化数据，通过 ETL 流程处理后，统一向量化存储。这使得 AI 在面对“特定年龄、非标职业客户的重疾险保

费是多少”这类复杂查询时，能精准地从多个数据源中提取、整合信息并作答。

在逻辑可控性层面，项目引入了思维链与业务规则引擎相结合的机制。当代理人输入客户信息后，系统并非直接让大模型自由生成，而是通过一个预设的思维链框架，引导模型按步骤执行：第一步，调用客户信息智能体，对输入信息进行结构化整理并生成客户画像；第二步，触发需求分析智能体，根据内置的计算公式（如重疾保额缺口 = 年收入 ×5- 现有保额）量化保障缺口；第三步，产品匹配智能体根据缺口和客户偏好，从知识库中检索并推荐 2-3 种产品组合，并阐述推荐逻辑；第四步，在所有方案生成后，合规审核智能体启动，利用更侧重于规则执行的微调模型或规则引擎，对建议书全文进行扫描，核查是否存在禁止性词汇、超范围承诺等问题，形成生成与审核的技术闭环。

经验洞察：

业务流程重构是 AI 成功应用的前提。大模型的价值并非简单替换人工操作，而在于驱动业务流程的再造。在本项目中，成功的关键并非技术本身，而是将传统的、线性的建议书制作流程，重构为一个动态、交互、智能的人机协同流程。该平台并未取代代理人的专业判断，而是将其能力放大。代理人从繁琐的信息搜集和文案撰写中解放出来，将更多精力投入到与客户的情感沟通、对 AI 生成方案的优化微调以及最终决策的把关上，实现了 AI 提效、人增温的理想协作模式。这一协作新范式并非纸上谈兵，其价值已在实践中得到印证：某寿险公司与腾讯云合作的应用实践表明，AI 将代理人从超过 80% 的重复性人工录入与复核工作中解放出来，使其能真正聚焦于与客户的情感沟通和信任建立。这种由 AI 驱动的端到端流程再造，其系统性优势更体现在整体运营效率的飞跃上，助力实现了前端单证处理时间从小时级到分钟级、后端理赔周期从数周到 1-3 天的显著优化。

知识库是战略资产，而非 IT 成本，知识库的质量决定了应用的上限。高质量、结构化、持续更新的领域知识库，是金融大模型区别于通用大模型、建立专业壁垒的核心。项目实践证明，知识库的建设与运营投入，是确保模型专业性与可靠性的最高价值投资。而这项投资的价值回报是直接且可量化的。例如，某寿险公司依托腾讯乐享构建一个覆盖保险精算、金融法规、业务流程到健康管理等领域的千万级企业知识库，将条款解析准确率提升了 40%，跨领域知识关联效率提高了 60%，为破解复杂场景下的智能决策难题提供了坚实基础。

3.3.4 投研报告生成——AI 赋能投研决策

背景：投资研究业务作为证券公司构建核心竞争力的基石，其产出的专业洞察与价值判断，

是服务机构客户与内部决策的关键支撑。然而，传统投研模式普遍面临研究覆盖面受限、信息处理效率低下、知识传承与标准化困难、观点时效性难以保障等多重挑战。尤其在研究报告的撰写环节，研究员需投入大量时间进行海量数据搜集、清洗、分析及基础性内容的撰写，导致其核心精力无法完全聚焦于更高价值的逻辑推理与前瞻性判断上。为应对此困境，某中小券商前瞻性地启动 AI 赋能计划，旨在通过引入大模型技术，系统性重塑投研内容生产流程，提升研究团队的整体产能与专业价值。

场景定位：本案例聚焦于构建一个赋能型智能投研工作台。该平台旨在解决研究员工作流程中的两大核心矛盾：通过自动化处理基础研究环节，将研究员从信息与数据的收集者转变为深度洞察的分析者；通过标准化内容生产，确保每一份研究报告的专业性、合规性与品牌一致性。该平台并非单一的写作工具，而是旨在将 AI 能力深度嵌入研究员从选题、资料搜集、数据分析、内容生成到合规审核的全业务流程，实现从辅助工具到智能研究伙伴的角色跃迁。

场景难点：

实时金融数据的精准融合。投资研究的时效性要求极高，模型必须能够实时接入并准确理解多源异构的数据，包括结构化的财务数据、行情数据，以及非结构化的公司公告、行业新闻、宏观政策文件等，确保所有分析都基于最新、最准确的信息。

投研逻辑的深度与一致性。一份有价值的研究报告不仅是信息的堆砌，更需要遵循严谨的分析框架与逻辑。模型需要具备深度推理能力，能够理解并运用如财务模型分析、行业竞争力分析等复杂投研逻辑，并保证在不同报告中逻辑标准的一致性。

合规要求与分析师观点的平衡。研究报告必须严格遵守监管机构的合规红线，规避不当陈述或投资建议。同时，报告的价值又在于其独立的分析师观点。如何让 AI 在提供客观数据支持与内容生成的同时，为分析师的独立判断留出空间并确保最终成果合规，是一个核心难点 .

多模态内容的自动化生成与整合。现代研究报告包含大量的图表、表格等可视化元素。模型不仅要能撰写文本，还需具备理解数据、自动生成相应图表并将其与文本内容无缝整合的能力，保证图文的一致性与专业性。

解决思路：以检索增强生成为核心，构建覆盖宏观、行业、公司、产品的多层次动态知识库。模型在执行任何分析任务时，都被强制要求从该知识库中检索信息作为其决策与生成内容的依据。将复杂的研究报告生成任务进行拆解，通过多智能体协作的模式执行。例如，一份公司深度报告的生成任务被分解为数据搜集、财务分析、新闻舆情分析、初稿撰写、图表生成、合规审核等多

个子任务，由不同但相互协同的智能体完成，确保了全流程的严谨性、专业性与自动化水平。

专题解决办法：以一篇深度公司研究报告的自动化生成为例

在研究报告生成模块中，为确保最终输出的专业质量与合规性，系统性地解决了知识注入与逻辑可控性两大问题。

在知识注入层面，项目团队构建了一个多源异构的投研知识中心。该中心能够实时接入并处理多种数据格式，例如通过 API 接口获取的结构化行情与财务数据，通过网络爬虫与订阅源获取的新闻资讯，以及内部存储的 PDF 格式公司财报、Word 格式过往报告等。所有信息经过统一的 ETL 流程进行清洗、解析与结构化，最终被向量化存储。这使得 AI 在面对“分析该公司上季度营收同比增长的原因，并结合近期管理层在业绩说明会上的表态”这类复杂查询时，能精准地从多个数据源中提取、整合信息并形成观点。

在逻辑可控性层面，项目引入了思维链与业务规则引擎相结合的机制。当研究员发起一项报告生成任务后，系统并非让大模型自由发挥，而是通过一个预设的投研逻辑框架，引导模型按步骤执行：

- 第一步，信息采集智能体启动：根据报告要求，自动从知识中心检索并汇总目标公司的财务报表、历史股价、重大公告、相关行业政策及最新的市场舆情。

- 第二步，数据分析智能体介入：调用内置的财务分析模型，对采集到的数据进行自动化计算，生成核心财务比率、增长趋势分析等量化结果，并识别出关键的财务亮点或风险点。

- 第三步，内容草拟智能体执行：基于前两步的结构化信息与量化结果，并遵循经过微调学习的报告模板与行文风格，自动生成报告的初稿，包括数据描述、基础分析及图表占位说明。

- 第四步，合规审核智能体把关：在初稿生成后，合规审核智能体利用更侧重于规则执行的微调模型或规则引擎，对报告全文进行扫描，核查是否存在夸大宣传、承诺收益等禁止性词汇，并自动添加必要的风险提示与免责声明，形成生成与审核的技术闭环。

投入产出分析，该项目精准地平衡了技术投入与业务产出，实现了显著的投资回报。

- **投入分析：**

⊙ 技术成本：采用私有化部署的开源轻量级模型，并结合知识库与检索增强生成技术，相较于直接采购或独立训练大规模闭源模型，大幅降低了算力基础设施投入与模型许可费用。

⊙ 人力成本：项目初期投入数据工程师与资深研究员，共同构建投研知识库与报告模板，形成一次性知识资产投入。通过引入数据合成等技术，可有效降低对业务专家进行大规模人工标注的依赖 。

⊙ 实施路径：遵循单点突破与速赢策略，选择从覆盖范围最广的晨会报告与数据点评作为切入点，快速验证价值并积累经验，再逐步扩展至深度行业报告与公司报告。

· **产出分析：**

⊙ **核心效能提升：**经测算，智能投研工作台将研究员在每份标准报告上的资料搜集、数据处理及初稿撰写时间平均压缩 40% 以上。而在部分高频、标准化的业务场景中，效率提升更为极致。腾讯等行业实践已证明，在金融舆情报告这一高频、标准化的业务中，大模型将单份报告的生成时间可从原先的人工 4 小时大幅压缩至 15 分钟内，为市场响应与风险控制赢得了宝贵的时间窗口。

⊙ **交付质量保障：**研究团队在不增加人员编制的情况下，能够显著提升研究报告的覆盖范围与发布频次。此外，相关行业实践数据也表明，基于大模型的报告生成在内容准确率方面可稳定在 90% 以上，关键信息抽取完整率也超过 85%，这证明了在提升舆情研究的规模与效率的同时，其产出内容的质量同样获得了保障。

经验洞察：

业务流程重构是流程嵌入型 AI 应用的成功前提。大模型的价值并非简单替换人工操作的某个环节，而在于驱动投研业务流程的系统性再造。在本项目中，成功的关键并非技术本身，而是将传统的、线性的报告撰写工作，重构为一个动态、高效、智能的人机协同生产流程。

领域知识库是构建专业壁垒的战略资产。该平台的专业性并非源于通用大模型的语言能力，而是源于其背后高质量、结构化、持续更新的投研专用知识库。这是区分通用 AI、建立自身核心竞争力的关键。实践证明，在知识库建设与运营上的投入，是确保模型专业性与可靠性的最高价值投资。

重新定义研究员的价值，人机协同是价值实现的最终形态。该工作台并未取代研究员的专业判断，而是将其能力从繁琐的数据整理工作中解放出来，使其角色从信息处理者转变为思想创造者。研究员将更多精力投入到与产业专家的交流、对未来的前瞻性思考以及对客户的深度服务上，实现了 AI 提效，人增智的理想协作模式。

3.3.5 AI 编程伙伴——金融业软件开发提效新范式

背景：金融行业作为技术深度应用的领域，其软件开发过程面临独特的挑战。首先，金融业务逻辑极为复杂，对从业人员的专业门槛要求高，开发者不仅需要具备扎实的技术能力，还必须深入理解相关金融知识。其次，金融系统是社会经济运行的核心基础设施，因此对代码的安全与合规性有极高的标准，任何微小的技术疏漏都可能引发系统性风险。最后，为保障数据安全，金融机构的开发环境通常与公共网络物理隔离，这使得外部先进工具的引入流程复杂且审查严格。

在上述严苛的条件下，金融机构的开发者在日常工作中面临诸多具体痛点。例如，为理解遗留系统和复杂的业务逻辑，开发者需投入大量时间研读有限的文档，新员工的培养周期长，知识传承高度依赖资深专家，形成了效率瓶颈。同时，开发人员需耗费大量精力编写满足安全规约的重复性代码，并在漫长的手动代码审查流程中等待反馈，这不仅拖慢了开发节奏，也难以完全避免人为疏漏。物理隔离的开发环境限制了对外部开源工具和知识库的访问，而业务层面又要求产品快速迭代以应对市场变化，导致研发效率与业务敏捷性之间的矛盾日益突出。

场景定位：为系统性应对上述挑战，本案例聚焦于为金融开发者打造一款 AI 辅助编程提效工具，旨在成为员工的智能编程伙伴。以腾讯云代码助手 CodeBuddy 为例，此类工具的核心价值在于将大模型能力深度嵌入软件开发全生命周期，提供包括编码辅助（代码补全与生成）、智能问答与诊断（技术问答、代码诊断、单元测试生成）、代码质量保障（智能评审）、团队知识沉淀（知识库管理、Rules 规范管理）以及代码智能化、多研发任务自动化（如通过软件开发智能体 Agent 实现 AI 深度理解需求、批量生成多文件代码）、研发生态打通（兼容 MCP 生态协议）等在内的端到端的综合性能力。

解决思路：通用能力与行业特性的深度融合

为精准应对金融行业的特殊挑战，方案采用通用基础能力、行业特性增强与企业级定制相结合的三层策略。

首先，在通用基础能力层面，方案依托性能强大的基础模型，提供高质量的代码补全、代码生成与技术问答等功能。这些功能能够基于代码上下文进行多行、精准的逻辑预测与生成，普适性地解决所有开发者的基础效率瓶颈。具体体现在：

- 代码补全：模型能够理解当前代码文件的上下文，包括已定义的变量、函数签名、引入的类库以及整体代码逻辑，从而提供行内或整段代码块的补全。补全场景覆盖了从简单的变量名、API 调用，到复杂的业务逻辑函数体、循环与条件判断语句，以及特定框架所需的样板代码。

- 代码生成：开发者可通过自然语言注释描述需求，如 CodeBuddy Agent 智能体自动检索代码仓库，深度理解用户需求，根据关联代码片段或知识库、图片或 Rules 规范召回数据，制定执行计划，自动生成完整的功能代码。典型生成场景包括根据注释创建单元测试用例、依据数据库表结构生成数据访问对象（DAO）与数据传输对象（DTO）、基于功能描述生成正则表达式或 SQL 查询语句，以及实现完整的算法或业务处理函数。

- 技术问答：开发者可选中代码片段，向 CodeBuddy 提问以获得功能解释、逻辑梳理或优化建议。CodeBuddy 还可用于快速定位并修复程序错误，或根据问题提供相关的 API 文档说明与最佳实践范例。

其次，在行业特性与企业级定制层面，方案通过以下技术路径解决前述痛点：

- 应对高专业门槛：通过检索增强生成（RAG）技术，将企业内部的代码库、技术文档、API 规范等私有知识资产整合为模型可检索的知识库。开发者能够通过自然语言查询，精准召回私域知识库，方便开发者快速获取关于复杂业务逻辑和历史代码实现的说明与范例。这相当于为每位开发者配备了一位全天候可用的、精通本企业业务的资深技术专家，显著缩短了新老员工的学习曲线，促进了知识的有效流转。

- 应对高安全合规要求：构建多层级的代码质量保障体系。在编码阶段，通过在集成开发环境（IDE）中内置由模型基于项目工程、Project Rules 规则约束进行驱动的代码静态分析，依据企业内部的安全规范与历史漏洞数据进行实时诊断，提前预警潜在风险和提供修复意见。在代码审查阶段，通过代码仓库 Web 端和 IDE 端双管齐下，基于团队评审规则，一方面利用模型自动生成代码变更摘要，另一方面基于规范对不符合合规要求的代码进行检测，提出修改建议，作为前置审查环节，提升人工审查的效率与准确性。此外，通过在经过严格审计的企业内部高质量合规代码上对模型进行精调训练，确保模型生成的代码本身就具备高度的内生安全性与合规性。

- 应对高开发环境要求：提供成熟的私有化部署方案，将整套 AI 代码助手系统以容器化的形式部署在企业内网，实现与公共网络的完全隔离，确保所有代码与数据均在企业内部流转，满足金融行业对数据安全和环境隔离的最高标准。同时，通过自动化能力提升敏捷性，例如一键生成需求单、设计稿，批量生成代码，一键生成单元测试用例以及测试报告，根据代码逻辑自动创建 API 文档，结合日志与代码上下文智能推荐调试方案等，将开发者从重复性劳动中解放出来，使其能够更专注于核心业务逻辑的创新与实现，从而有效加速开发进程。

项目成效：

AI 代码助手已成为大模型在金融行业落地最广泛、成效最显著的应用之一，并已经在银行、证券、保险等金融机构推广。在某头部金融机构，已有超过 8000 名程序员在日常工作中使用。腾讯云 AI 代码 CodeBuddy 的实践数据显示，该工具普遍可达到 40% 的字符生成率和 30% 以上的代码采纳率。综合代码生成、智能问答、代码诊断和自动化测试等能力，可为研发团队带来超过 40% 的整体编码效率提升，有效缩短了新产品和新功能的上线周期。

经验洞察：

数据治理与模型安全是落地的前提。金融机构在引入 AI 编程工具时，必须建立严格的数据治理框架，确保用于模型训练与检索的内部代码、文档不包含任何敏感信息。同时，私有化部署环境下的模型自身安全、访问控制与输出内容审计机制是保障技术应用合规与风险可控的必要条件。

试点验证与量化指标是建立信任的基础。在金融机构内部，新技术的推广需审慎。建议选取代表性的业务线（例如核心交易、风险管理）开展小范围试点，并建立与业务目标强相关的量化评估指标，例如缩短监管需求响应时间、降低生产环境安全漏洞数量等。通过试点获得的实证数据是获得管理层与业务部门支持的关键。

深度适配是跨越可用到好用的桥梁。AI 编程工具的价值不仅在于提升通用编码效率，更在于与金融机构特有的安全协议、合规框架及私有代码库深度整合。通过私有化部署、检索增强生成和模型精调等技术手段，使工具能够理解并生成符合本机构规范的代码，是实现其业务价值最大化的核心路径。

无缝集成是工具广泛采纳的关键。工具的价值最终需要通过开发者在日常工作和场景中使用来体现。基于腾讯内部及外部大量用户使用场景和诉求，CodeBuddy 提供 AI IDE、终端场景 CLI、Plugin 插件等交付形式为开发者提供更好的开发体验，其中 Plugin 插件兼容几乎所有主流 IDE（包括 Visual Studio Code, JetBrains 系列、Visual Studio 、微信开发者工具等），并无缝嵌入开发者已有的工作流，是降低使用门槛、实现技术广泛推广与采纳的必要条件。

3.3.6 金融智能体——从概念验证到应用的探索

金融投资研究领域作为一个高度依赖知识和分析的行业，面临着三大瓶颈：海量信息带来的认知过载，不同来源数据形成的信息孤岛，以及核心经验随人员流动而流失的知识断代。

以资产管理行业为例基金经理和研究员的工作常被重复性的信息搜集所占据，如手动查询公司公告，整理和汇总财务数据，并时刻监控市场新闻和政策变化。这些繁琐的任务不仅易于遗漏，还严重削弱了用于策略制定和深度分析的宝贵时间。此外，现有内容生成工具在准确性和实效性方面存在较大差异，导致内容生成质量参差不齐。

在这一背景下，构建能够理解自然语言指令、自动处理信息并根据统一标准进行初步分析的智能投资研究 Agent，已成为提升投研效率、释放核心人才创造力的关键。智能体（Agent）在金融投资研究这一知识密集型领域展现出巨大潜力。

一、市场主流金融大模型应用的模式

为解决上述瓶颈，市场上已涌现出几种 AI 投资研究应用模式：

嵌入式终端助手：以 Bloomberg GPT、Wind Alice 为代表，将 AI 嵌入现有的复杂终端，通过自然语言交互，自动翻译成终端可以执行的精确指令或者专有的代码，将用户的问题编码成查询向量（embedding），匹配已编码的独家后台数据进行比对，并与 LLM 进行多模态生成。用户不再需要记忆复杂的指令，降低了专业工具的操作门槛。

增强型搜索引擎：利用向量检索技术在海量、半结构化的文档中实现精准的信息定位，并通过指令生成结构化简报，实现“在 200 页的财报中找一句话”的效果。同时整合专家访谈纪要库并对非结构化对话内容进行向量化，优化信息检索与分析。

企业知识管家：利用知识图谱、向量检索等技术，将企业隐性知识转化为结构化资产。一类如 Glean，旨在通过分析文档关联找到内容及专家，以提升协作效率，但需要针对金融场景做二次开发和适配；另一类如腾讯乐享，作为企业级知识库平台，侧重于知识治理，通过严谨的权限管理和动态更新机制，并支持私有化部署，以满足金融等行业的核心安全合规需求。

智能体工作流：该模式的实现平台主要分为两类：一类是 Zapier 等在传统流程中融入 AI 决策的自动化平台；另一类则是 Dify、腾讯云智能体开发平台等专注于原生智能体编排的开发平台。它们的共同点在于都提供了低代码 / 无代码界面，让用户能编排多工具，创建复杂的自动化工作流，快速构建更垂直和定制化的 AI 应用。例如，在实验中，一个自动股票交易助手可自主监控实时指数（如 RSI），交付大模型推理决策，并自动调用交易平台的 API 完成交易。

二、当前金融大模型应用的能力边界

AI 的应用仍主要集中在信息处理阶段，缺乏成熟且稳定的独立分析与决策能力。智能体依赖

于检索增强生成（RAG）架构高效提取信息，能高效地回答“是什么”和“在哪里”，但无法可靠地回答“为什么”和“会怎样”。此外，RAG 支持的问答生成仍存在幻觉。 例如，部分国产大模型在研报问答中的准确率为 90%，问题出在检索阶段的上下文和语义丢失以及生成阶段的缺乏数据库具体事实依据。

AI 擅长共识检索消化和生成，而投资的竞争优势源于非共识洞察。高阶的 AI 智能体，其价值不仅体现在对买方共识（Buyside Consensus）内信息的快速响应、做到正确检索被市场充分定价（Priced-in）的数据，更在于从另类数据中挖掘并理解增量信息，发现超越市场共识的阿尔法机会。若缺乏针对性的领域知识训练与模型微调，AI 难以洞悉原始数据背后的含义，无法从原始、嘈杂的数据中有效地区分出真正的投资信号与随机噪声。

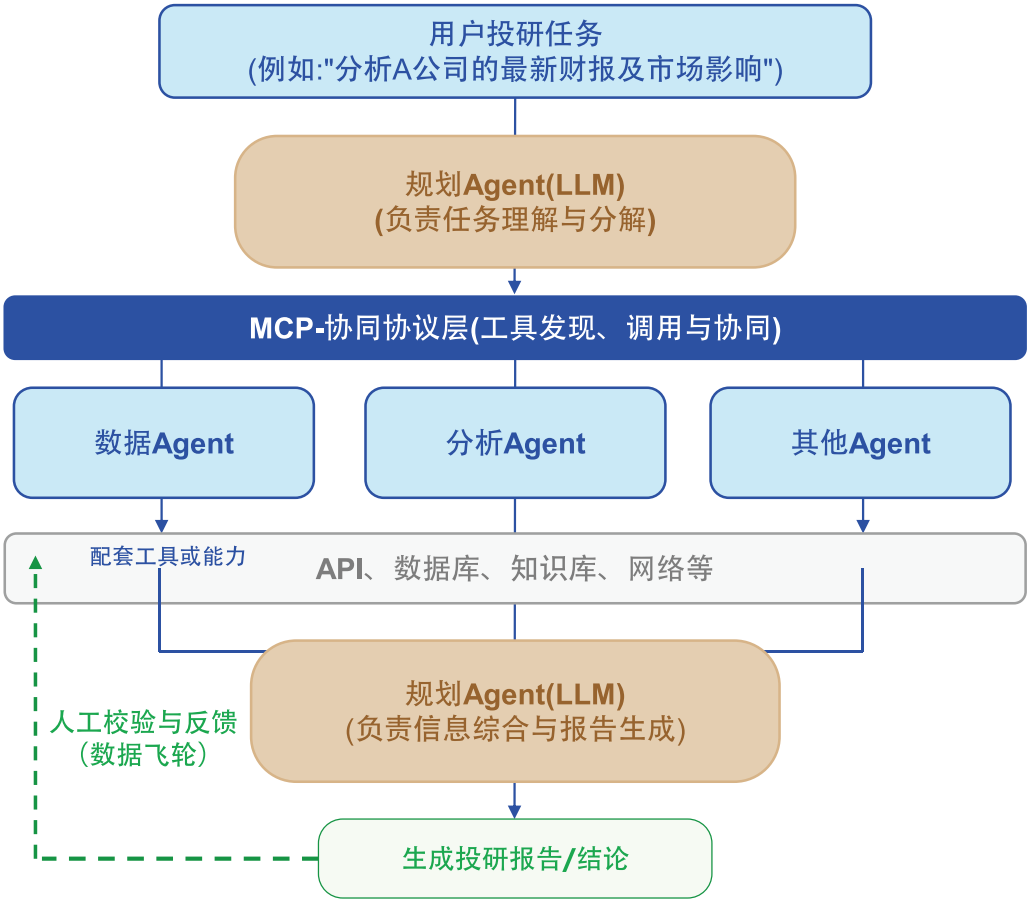
市场应用呈现出专业深度、开放生态与无缝集成三者难以兼顾的局面。拥有最深金融护城河的平台生态较为封闭；拥有大量专家内容的平台核心生成能力依赖外部；拥有原生工作流的平台则缺乏金融专业性。能完美兼顾三者的解决方案，至今尚未出现。

应用的推广必须首先确保信任与合规性，避免因过度依赖技术而削弱市场信任。对于中国本土金融机构而言，数据安全与合规是不可逾越的红线。海外模型对我国特有政策语境和市场环境的理解偏差，加之数据出境的合规风险，是海外应用的最大障碍。这为基于本土大模型的解决方案提供了巨大的发展机遇，其在中文能力、数据安全和私有化部署成本上的优势将愈发凸显。在当前发展机遇下，如何通过扎实的技术工程实现真正的创新和智能化，避免“AI washing”[11] 透支市场信任，是全球大模型企业面临的重大挑战。

三、以 MCP 协议构建金融 Agent

为破解上述能力边界与生态困境，业界正在探索以模型上下文协议（MCP）为代表的新技术路径。MCP 协议通过提供统一标准，使得不同 Agent 能够发现并调用通用工具，解决了开放性与专业性的矛盾。在这一架构下，每个 Agent 可以专注于自己最擅长的领域，而 MCP 的多视图和调用机制则负责将这些平行且专业的技能模块有机串联，形成协同效应。

11　AI washing：即夸大 AI 能力、过度包装 AI 概念以获取市场关注和投资，但实际技术水平有限的行为。



金融 Agent+MCP 开放协同生态

然而，在 MCP 协议架构下，智能体依然面临多个挑战：

Agent 的固有问题。主要体现在三个方面：一是可能不完全遵循指令，执行超出预期范围的任务；二是可能产生模型幻觉，编造不存在的工具或参数；三是面对复杂问题时效率低下，易陷入长时间的无效推理循环。在对高风险领域（如金融投资决策）应用时，幻觉导致的错误信息可能会导致重大损失，并且由于模型自身的局限性以及缺乏有效的自我修正机制，往往需要人工干预。

灵活性与可靠性的冲突。纯 Agent 模式的灵活性高但可靠性不足，而传统的固定工作流模式则相反。当前的最佳实践是采用混合模式：一方面，用固定的工作流来强化 Agent 能力，特别是在金融领域，需补充背景知识（如当前时间、金融术语 / 字段映射、股票代码）。通过 RAG+ 知识库在 Agent 调用工具前提供这些信息，可提高工具调用成功率。另一方面，未来的发展方向是采用多 Agent 协作来应对复杂任务，将大任务分解，由规划 Agent、执行 Agent 等构成的虚拟

团队协同完成，以提升系统鲁棒性和处理能力上限。

MCP 工具的精准调用难题。即使有统一协议，Agent 如何精准选择和调用工具仍是难题，且当前高质量 MCP 市场工具数量有限。模型可能更倾向使用自然的表达方式而非工具定义的特定函数。针对此问题，可采用更准确、精简、自然的工具描述，或探索使用小模型总结工具功能，或通过反思 Agent 检查参数有效性。

MCP 连接到生产数据库存在潜在的安全风险。虽然 MCP 主要用于开发环境，但如果其启用了可访问互联网的工具，就可能暴露出攻击向量，攻击者通过这些工具能够获取数据并将其外传。在使用如代码编辑类等工具时，Agent 可能会被赋予过高的权限。如果客户提交的请求中包含恶意构造的指令，Agent 可能会将这些指令误解为执行命令并进行操作。只要 Agent 拥有足够的权限，且未构建有效的安全策略，攻击者就可以利用这一点进行攻击，绕过防火墙和基于角色的访问控制，从而导致数据泄漏。

四、发展金融 Agent 的思考

在技术突破方面，首先需要构建金融领域的因果推理体系，帮助 Agent 理解财务指标和市场事件之间的逻辑关系，并结合符号推理与神经网络建立混合推理架构。同时，重点关注非共识信息挖掘，特别是另类数据的预处理与特征工程，提升 Agent 在识别市场未充分定价信息方面的能力。此外，提升系统的可靠性至关重要，这包括引入不确定性量化技术，让 Agent 能够识别自己的知识边界，并主动寻求人工确认，尤其是在高风险决策时。为进一步提升系统鲁棒性，需要构建自我修正体系，通过多层验证机制减少模型的幻觉风险。在多 Agent 协作方面，建议通过构建一个由规划 Agent、数据收集 Agent、策略分析 Agent 和决策执行 Agent 组成的虚拟团队，将复杂任务分解，并通过专业化协作提升整体决策效率和质量。此外，增量信息提取技术应结合时间序列分析与异常检测，帮助 Agent 更好地识别市场动向与潜在风险，从而增强决策的前瞻性和准确性。

在生态整合方面，需要从标准化体系建设入手，首先通过优化 MCP 协议，推动金融行业特有的工具调用规范、数据接口标准及风险控制协议的建立，确保不同厂商的 Agent 在安全性与准确性方面达成统一标准。同时，通过进一步优化 MCP 协议，提升不同工具间的兼容性和协作能力，简化接口设计，减少 Agent 调用过程中的干预需求。在专业工具生态构建方面，建议鼓励金融数据供应商和研究机构开发标准化的 MCP 工具，形成涵盖数据分析与风险控制的完整工具链，提升工具的质量评估机制，通过基准测试和用户反馈不断优化工具的准确性和成功率。在开放合作模式方面，支持金融机构和行业联盟的建设，推动共同制定技术标准，分享最佳实践，避免重

复建设。与此同时，探索“平台 + 生态”模式，由核心平台提供基础设施，第三方开发者贡献专业工具，共同打造良性循环的生态系统。

在可信体系构建方面，首先需要构建分级安全架构，根据数据敏感程度和业务风险等级设定差异化的安全控制策略，确保高敏感数据得到本地化部署和加密处理。权限控制机制需要严格限制数据库访问权限，并通过特定安全 API 进行，防止指令注入等攻击手段。在本土化适配保障方面，建议基于国产大模型构建金融 Agent，确保其对中国市场环境和政策语境的准确理解，并建立中文金融语料库和知识图谱，提升模型的专业能力。同时，确保 MCP 协议与国内安全要求和监管政策兼容，通过私有化部署与数据加密保障数据的安全性。在监管审计体系方面，建议建立实时监控机制，记录 Agent 行为，通过动态日志和异常检测确保决策过程的可追溯和可解释。同时，加强合规流程，在 Agent 调用任何工具之前，嵌入背景知识注入和强化 RAG 流程，确保决策的合规性和准确性。

正如水和电力重塑了人类社会的基础设施，远期来看，大模型也将深刻影响金融的运营模式，提升效率，释放更强大的潜能，驱动一场效率、智能和模式上的深刻变革，支持金融更好服务实体经济发展，助力金融强国建设。

4 第四章

大模型驱动金融业发展的趋势展望

大模型驱动金融业发展的趋势展望

大模型将在未来五年引爆金融业的临界点，触发的不是缓慢演进，而是一场深刻的范式革命。这场革命并非简单的“机器换人”，而是将人类从重复性劳动中解放出来，推向更具创造性、战略性和判断力的角色，并在此过程中创造出全新的职业。

4.1 金融服务的专业化和普惠化进程提速

过去，尖端的金融分析能力、复杂的风险建模和高度定制化的财富管理策略，如同奢侈品，是少数大型机构和高净值客户的专属。随着高性能开源模型的涌现、模型训练和部署成本的下降，金融机构构建和应用大模型的门槛显著降低，有利于通过 AI 将这些专家级的能力，从金字塔尖逐步下沉至一线，推动普惠金融发展。

金融世界充满了复杂的信息和数据，普通用户往往深陷其中，难以做出有效决策。AI 的核心价值之一，便是利用 AI 快速处理海量信息的特点，为普通用户在复杂的金融场景中进行信息降噪，提供清晰、易懂的决策支持。其次，传统金融 APP 和软件的操作往往需要一定的学习成本，随着 APP 和软件的用户界面正从传统的图形用户界面（GUI）向语言用户界面（LUI）演进，用户不再需要学习复杂的操作，只需用自然语言下达指令，AI 便能代替用户使用金融工具，进一步降低使用门槛。

综合以上两点，AI 将打破高端金融服务的稀缺性，将原本高度集中在机构和高净值客户的专家级能力，诸如复杂的投资研究、精密的风险建模、专业的合规文本解读，转化为 AI 服务，普及至更广泛的中小机构乃至个人投资者，从而开启一个全民普惠的智能金融新时代。为了实现这个愿景，行业仍需共同克服数据质量、模型可靠性、合规性及伦理等多重挑战。

4.2 金融产品更加实时、动态、超个性化

多模态大模型正在重构金融服务逻辑。新一代大模型已实现文 – 图 – 音 – 视频的无缝转换。在金融场景中，在确保合规与用户授权前提下，这意味着可以通过分析语音语调、面部微表情、交互行为等非结构化数据，更深入地理解客户需求和风险状况。例如，在远程视频服务中，结合声纹和行为分析辅助身份验证和风险评估；在智能投顾交互中，通过理解客户的语气和表达，动

态调整沟通策略和产品推荐。这种多模态融合有助于构建更立体、精准的客户画像。

基于对用户全维度、实时行为数据的动态捕捉与分析（如交易记录、地理位置、设备使用习惯），结合语音情感识别、微表情分析等生物特征解析技术，金融机构正从静态功能交付转向动态场景适配。例如，车险保费可根据驾驶行为的实时反馈动态调整；信用卡额度可能因用户临时的大额消费计划而临时提升；投资组合则会随市场波动与投资者情绪变化自动再平衡。这种“流媒体式”的服务模式将打破传统金融产品的静态框架，推动服务范式从千人一面向千人千时千面跃迁，即同一用户在不同时间、不同场景下获得的服务逻辑个性化。

借助多模态交互和生成式 AI，数字员工的能力将大幅提升，并实现更强的个性化和情感连接。这包括：跨渠道的身份与对话记忆、在不同设备终端上保持一致的交互体验、以及根据用户画像和偏好定制沟通风格与服务流程的能力。这种个性化生态将金融软件从工具属性升维为有温度的金融伙伴，通过建立情感连接提升用户粘性与生命周期价值。

4.3 人机协同重新定义金融运营与管理模式

金融业依赖大量人工的后台开发与运营、中台审核及部分前台交互环节，将越来越多地利用 AI 进行流程再造和效率提升。例如，在信贷审批、保险理赔、交易清算等流程中，AI 将承担更多的数据处理、模式识别、风险评估和初步决策建议工作。在投资分析、风险预警等更核心的领域，AI 强大的数据处理和预测能力，将为人类专家提供更精准、高效的决策支持。

然而，AI 的应用并非旨在完全取代人类，而是构建更高级的人机协同模式。人类的角色将向监督者、策略制定者、复杂问题解决者和最终决策者转变，专注于设定目标、监督 AI 运行、处理异常情况、进行关键判断，并负责维护客户信任和伦理规范。AI 执行与分析，配合人类监督与决策，将成为金融运营的新常态。

大模型的技术迭代对金融人才的能力结构提出了更高、更综合的要求。除了传统的金融专业知识，从业者需要增强 AI 应用与管理能力、定义复杂问题的能力、跨领域整合能力，以及与 AI 高效协作并确保其安全合规运行的能力。人类独有的批判性思维、创造力、同理心、伦理判断以及建立信任关系的能力，将更加凸显其价值。

4.4 高价值数据的挖掘与应用的重要性提升

金融领域对模型的专业性、精准度和时效性要求极高，单纯依赖海量通用数据已不足以构建前沿、具有竞争力的模型，私域数据的利用会进一步释放金融 AI 应用的核心价值。金融机构需要更注重数据精炼，通过构建领域知识图谱、优化特定任务数据集的方式提升训练数据的价值密度，基于思维链 CoT 方法构建的推理数据集，使模型具备更强的专业知识和因果推理能力，同时优化算力效率。在各大模型厂商以公开数据作为模型训练基础的前提下，金融机构构建的 AI 应用的竞争优势，来自于对机构内部私域信息的深度挖掘和利用，特别是客户交易数据、电话会议记录、专家访谈等。

面对高质量金融数据稀缺且获取成本高的问题，合成数据技术成为重要的补充手段，特别在处理长尾事件、极端风险、新型欺诈模式时尤为关键。例如，合成数据可用于扩充反洗钱模型的训练样本、生成压力测试所需的极端市场情景数据等。然而，必须谨慎验证合成数据的质量和分布，并有效结合真实数据，采取如合成数据预训练和真实数据微调的策略，避免模型偏见或与现实脱节。

金融决策往往需要综合处理来自不同来源和形态的数据，如财报文本、市场行情（时序数据）、交易量（结构化数据）、新闻舆情甚至另类数据。多模态大模型的训练目标是实现跨模态信息的有效融合与语义对齐，使模型能够像人类专家一样，从多维信息中发现关联、洞察趋势，进行更全面的分析与预测。

高价值金融数据往往高度敏感，在挖掘数据价值的同时，必须严格遵守隐私保护法规和伦理规范。隐私计算技术（如联邦学习、安全多方计算、同态加密、差分隐私等），在保护数据隐私前提下，进行模型训练和数据协作，例如，机构间可在不共享原始数据的情况下共建风控模型。未来，数据价值的实现将与透明度（如模型可解释性、数据溯源）和安全性（如合规脱敏）的要求紧密协同。

4.5 AI 驱动监管科技提升和治理体系升级

AI 对金融稳定性的影响已上升为全球监管重要议题。大模型在金融领域的广泛应用带来了新的监管挑战，主要包括：透明度风险，即算法黑箱导致的透明度不足问题；时滞性风险，即技术快速迭代与监管规则更新之间的时滞；共振性风险，即模型趋同可能引发的系统性风险放大效应（如市场共振、顺周期行为加剧）。

以 AI 来驾驭 AI，可推动监管科技（RegTech）进入新的发展阶段。大模型强大的非结构化数据处理与关联分析能力，有效弥补了传统模型依赖结构化数据的短板，能够高效整合分析新闻报道、研究报告、社交媒体情绪等信息，深度洞察风险事件背后的舆论环境与逻辑链条，从而更准确地判断异常波动的性质。另一方面，其动态学习能力使其能够持续适应和发现新型风险模式，当监测到负面舆情与异常交易量等多元信号并发时，能将孤立信号关联判断，提升对黑天鹅与灰犀牛事件的早期预警能力。对此，可以利用大模型构建智能合规审查系统，自动解析、比对多司法辖区监管规则；建立早期风险预警模型，向预测性、主动性监管转型；以及优化监管沙盒机制，在可控环境中测试和评估创新 AI 应用的潜在风险。

然而，大模型在风险管理领域的应用仍需构建人机协同、专家把关的决策闭环。大模型擅长发现相关性，但难以有效判断因果性，这可能导致其对风险的分析停留在表面症状，无法触及深层病灶，甚至被虚假关联信号误导。大模型在生成内容时可能出现幻觉，叠加其训练数据源于开放互联网，涉及不实信息、偏见和噪音等信息污染问题。许多风险的研判依赖于深厚的领域知识与专业常识，而大模型对此类基于真实世界经验的系统性理解力仍然严重不足。因此，构建人机协同、专家把关的决策闭环是其在风险管理领域安全应用的必然要求。

面对 AI 的能力及风险，金融机构作为 AI 的应用主体，应进一步完善内部治理体系，构建覆盖 AI 应用全生命周期的可信治理框架，在创新与合规之间取得平衡。这包括：对 AI 供应商和模型的准入评估；运营阶段对模型性能、偏见和稳定性的持续监控与审计；建立模型风险的应急处置和退出机制；以及积极落实可信 AI 原则，加强模型可解释性研究，保障在关键决策点上不可或缺的人工监督与最终否决权。通过监管科技与合规 AI 的协同发展，最终实现敏捷监管与负责任创新的动态均衡。

4.6 复合型、创新型金融人才需求正在形成

大模型正在深刻变革金融行业的组织结构与人才需求，其影响并非简单的岗位替代，而是对各层级岗位职责的系统性重塑，并催生出人机协同的全新工作模式。这一轮转型正沿着执行层、专业层与新兴岗位三个维度展开，对金融机构的人才战略提出了新的要求。

现有岗位的职责内涵正在发生结构性演变。在执行层面，大量重复性、规则导向型的工作，如标准化的数据录入与核对、初级信贷审查报告撰写等，正逐步由自动化技术实现。AI 显著提升了对结构化与非结构化数据的处理能力，使得该层级员工的角色正从任务的直接执行者，转变为对自动化流程进行监控、对异常事件进行处置的监督者。初级岗位员工以极小的比例向业务一线

转岗，机构应为其提供合理的职业发展路径和激励机制。

在专业层面，AI 日益成为辅助专业人士进行深度分析与决策的智能助手。通过赋能精准营销、智能风控等复杂场景，AI 帮助客户经理或风险经理等专业人员，提升了决策效率与质量，其职能也随之向数据驱动的策略分析师方向演进。

伴随技术与业务模式的创新，一批全新的岗位有望应运而生。这些新兴职位聚焦于人机协同与 AI 治理的关键环节。例如，智能体编排工程师负责设计与优化基于大模型的自动化业务流程；数据伦理与治理专家则专注于确保 AI 应用的数据合规性、算法公平性与决策可解释性，维护内外部数据质量与知识体系。面向流程颠覆型的 AI 应用场景，衍生出 AI 行为分析师、AI 对齐工程师等专业岗位。AI 行为分析师则通过分析 AI 决策逻辑与反馈信号，判断 AI 系统是否存在系统性偏差或决策偏误；AI 对齐工程师聚焦大模型的伦理偏好、情感表达偏好，通过多轮交互式提示进行引导、对齐和校准，并建立可持续的演进策略。

报告团队

顾问

司　晓 | 腾讯集团副总裁、腾讯研究院院长

湛炜标 | 腾讯金融科技副总裁、腾讯投资合伙人

杜西库 | 腾讯金融科技副总裁

胡利明 | 腾讯云副总裁

策划

柳晓光 | 毕马威变革咨询数字化转型业务牵头人、“智慧之光”数智解决方案主管合伙人

好　好 | 腾讯云战略研究院院长

杜晓宇 | 腾讯金融研究院秘书长

周　梦 | 腾讯金融大模型应用负责人

主笔

陈楚仪 | 孙箐阳 | 储　宁

研究支持

杨海松 | 王江涛 | 许文浩 | 洪庚伟 | 李晓聪 | 阿　梅 | 贾　飞 | 刘　辉

刘　毅 | 孔德远 | 王　成 | 马晓芳 | 刘　玲 | 卢晓明 | 陈春歌 | 巴洁如

联合出品

腾讯金融研究院
Tencent Financial Research Institute

腾讯研究院
Tencent
Research Institute

KPMG

2025金融业大模型应用报告

体系落地 价值共生