

# 2024年中国金融大模型 产业发展洞察报告

部门：TMT金融组      署名：孙石琦

# CONTENTS

# 目录

---

**01** 中国金融大模型产业发展背景介绍  
Background

---

**02** 金融大模型结构特征与优势能力分析  
Product Mix and Advantages

---

**03** 中国金融大模型产业卓越者评选  
Outstanding Enterprises

---

**04** 中国金融大模型产业未来趋势预判  
Industry Development Trends

---

**05** 中国金融大模型产业专家之声  
Industrial Experts' Insights

# 01 / 中国金融大模型产业 发展背景介绍

Background

# 技术崛起：大模型迭代历程回顾

## 多技术聚积融合推动大模型能力逐步完善，实现从“专用智能→通用智能→专业化智能”的模式转变

随着生成算法、预训练模型、多模态数据分析等AI技术的聚集融合，AIGC技术的实践效用迎来了行业级大爆发。在内部模型迭代与多技术协同的作用下，以“大规模”+“预训练”为核心的大模型产品，逐步由Transformer基础架构走向集图像识别、语义理解、视觉感知于一体的多模态、全应用模型体系，以产品化、业务化模式实现AIGC能力的进一步释放。与此同时，通用大模型技术的成熟推动了新一轮行业生产力变革，凭借其在场景泛化、技术涌现、强计算能力等方面的优势，大模型为多行业的发展带来了工业革命级生产能力的提升。随着专业化应用进程的推进，通用大模型在解决特定场景的专业化问题时暴露出部分局限性问题，行业深度信息的缺失导致其很难在专业领域提供高价值的服务，此外，通用大模型对于数据、算力的较高要求也提升了其在细分领域落地应用的门槛，相比之下，在专业度及成本方面更具优势的垂类大模型或将成为未来大模型技术落地应用的核心方式，预计未来5年，服务于特定行业的垂类大模型将实现超80%的专业领域覆盖，大模型能力将以专业化定制产品的方式为更多垂类专业领域赋能。

### 大模型技术迭代与产品模式应用分析



注释：各类模型的应用阶段描述仅为该阶段行业内最主流模型应用，并不代表该阶段仅存在一种或两种模型形式以及该模型形式仅存在于该段时期；

来源：专家访谈，公开资料，艾瑞咨询研究院自主研究及绘制

# 能力适配：大模型与金融数智化高度适配 iResearch 艾瑞咨询

大模型技术特征与金融业务数智化需求具有较高适配度，如何低运营成本的获取高价值大模型服务是金融行业大模型应用关注的核心问题

## 大模型技术的产业化应用——业务需求与技术能力的匹配

### 行业需求

### 技术能力



#### 金融行业数智化转型发展需求

##### ■ 金融行业的数据要素应用需求

- 金融行业属于数据、信息密集型产业，对于多渠道信息汇总及数据处理能力有较高要求，随着近年金融数据升级成为“**第五生产要素**”，数据在金融场景的应用得到了空前的重视。

##### ■ 基于经验的复杂决策制定

- 金融行业的众多业务场景对于从业人员的专业能力提出了基于专业知识储备与过往经验积累的双重要求，以满足金融业务对于**精准快速的有效判断**以及**高质量代码、文案产出的需求**。



#### 大模型的深度学习与智慧涌现应用

##### ■ 大模型的复杂信息处理能力

- 大模型是拥有**巨大参数量和计算能力**的深度学习模型，在自然语言的处理、计算机视觉应用、知识图谱问答等环节具有强大的数据处理与分析能力，结合深度神经网络结构，实现非结构化数据及复杂语句的理解与表达。

##### ■ 大模型的涌现能力

- 随着模型规模的提升，模型的性能将在突破某个临界阈值后显著提升到高于随机的水平，为用户提供**基于原始数据与经验的创造性解决方案**，包括上下文学习能力、知识容量扩展、泛化应用能力及复杂推理能力等。



#### 金融行业大模型技术应用

##### ■ 大模型技术为金融行业提供全新生产力

- 大模型在复杂信息处理与智慧涌现方面的特征与金融业务的数智化转型需求高度匹配，业务场景中面临的海量数据分析、交易决策制定、个性化投资建议等环节对于金融从业者的专业知识储备与经验判断提出了较高要求，而大模型的应用能够**优化数据处理模式，降低人力错误**，并在关键决策的判断上给予金融从业者系统性支持，以技术能力重塑金融业务流程。

##### ■ 通用大模型在金融领域的应用存在弊端

- 大模型的学习能力与泛化能力决定了其可以服务于多个领域，但在专业要求高的领域通用大模型技术**很难提供高价值服务**；此外大模型产品的高**运维成本**也将成为业务场景实际应用的阻碍。

来源：专家访谈，公开资料，艾瑞咨询研究院自主研究及绘制

# 产业定制：大模型技术的产业化应用

在投入提升与政策扶植的双重作用下，以大模型技术为底座、结合专业化金融能力的金融大模型产品应运而生

近年，在金融科技发展规划与金融机构数字化转型一系列政策的指导之下，国内金融行业的数智化转型颇具成效，越来越多的金融机构意识到数字化转型为自身业务发展带来的助力，行业内对于金融科技的资金投入逐年攀升，以金融机构作为主要服务对象的技术厂商迎来了需求端市场蓬勃发展的上升期，为前沿技术的产业化定制创造了优渥的市场环境；此外，金融作为监管合规要求较高的信息密集型行业，其监管体系的完善与监管要求的提升推动了机构应用前沿技术增强合规能力的需求，大模型作为有能力优化业务流程环节、完善合规应用的技术应用，其产业化及垂类应用的价值不容小觑。

## 大模型技术的产业化应用——金融大模型

### 科技应用成为金融行业发展的主旋律

- **政策的指导：**建设数字中国的战略号召与层层递进的**数字金融系列政策**，为金融机构的数字化转型提供了外在科技氛围营造与内在数字化发展指导；
- **机构的投入：**2024年中国金融机构**科技投入预计超4000亿元人民币**，后续将以**约12%**的增速保持增长。

### 金融行业全面进入超级监管时代

- **监管体制改革：**国内金融监管体系改革，设立了“一行一局一会”的**监管格局**，明确了中国金融“双峰”**监管框架**；
- **法律制度健全：**制定《网络安全法》《数据安全法》《个人信息保护法》在金融领域的**配套制度**，建设**高质量、多层次的国内金融监管体系**。



- 金融大模型将大模型技术与金融行业数据及场景信息相结合，应用大模型在**技术涌现、多任务学习、强计算高存储**等方面的优势服务信息密集型产业，通过**预训练、金融语料微调及专业数据强化对齐**等方式，实现大模型技术在金融行业的产业化落地。
- 大模型的产业化应用将在**智能风控、精准营销、财富顾问、投资规划**等多个业务场景中为金融机构提供全新生产力支持，解决场景中可能面临的**金融信息过载、复杂金融需求拆解、个性化投资建议**等需求，为机构与用户提供专业化、高价值服务。

数据来源：金融机构科技投入规模数据来自艾瑞咨询《2024年金融科技行业发展洞察报告》“2019-2027年中国金融科技投入情况”  
来源：专家访谈，公开资料，艾瑞咨询研究院自主研究及绘制

# 02 / 金融大模型结构特征 与优势能力分析

Product Mix and Advantages

# 金融大模型结构特征 与优势能力分析

## □ 金融大模型结构特征分析

- 金融大模型核心结构总览
- 基座支持：通用大模型
- 功能增强：大模型工具链
- 业务适配：大模型垂类场景训练

## □ 金融大模型优势能力分析

# 金融大模型核心结构总览

由通用大模型基座、专用模型工具链、私域业务场景训练三部分组成的完整金融大模型产品构建体系

## 金融大模型核心结构总览 ——以通用大模型为底座的金融大模型产业链分析



### 金融大模型核心组件结构分析：

- 金融大模型产品的核心构建流程主要由以上三个环节构成，首先是作为技术底座支持的通用大模型为金融大模型产品的构建提供底层技术能力、算法、基础参数等底层模型逻辑方面的支持，在此基础之上，以解决特定领域应用能力短板问题的模型工具链系列产品将作为模型层与应用层的桥梁，为垂类模型的构建提供应用层面的增强，此外，私域信息库的模型训练将全面提升大模型产品的业务适配能力，为机构提供客制化的模型服务

# 1.基座支持：通用大模型

## 国内通用大模型市场迎来“百模大战”时代，泛化能力的应用与专业能力的增强将成为通用大模型服务金融行业的两大主流方式

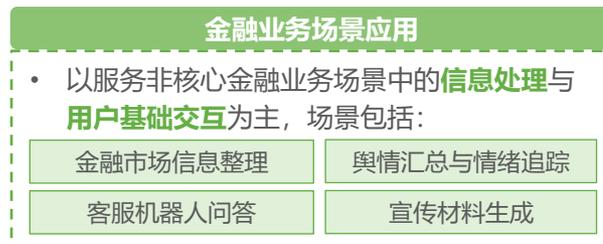
随着深度学习与生成式大模型的融合应用，以卷积神经网络、对抗生成网络、ResNet为代表的深度学习模型能够在自然语言处理、计算机视觉处理等方面实现较大突破，后续随着Transformer结构的应用，以Transformer和注意力机制作为组件、参数在十亿级别以上的自监督学习模型营运而生。伴随着近年大模型技术在实践应用场景的发展，目前通用大模型市场已进入“百模大战”的全新时代。从整体来看，以GPT4-Turbo、LaMDA 为代表的国际市场头部通用大模型，在模型架构能力、参数范畴、应用交互使用体验方面具有相对领先的优势，但随着近年国内大模型产业的爆发式增长，国内市场也出现了“文心大模型”、“通义千问2.0”等立足于自身大模型产品架构基础与生态体系建设的国产通用大模型产品，考虑到金融行业对于信创建设与数据保护有较高要求，国产通用大模型体系的建设对于大模型技术在国内金融市场的应用至关重要。在实际应用过程中，通用大模型的核心使用渠道主要有两种，一种是应用其泛化学习能力，将其他行业获取的信息与模型能力迁移至金融领域，解决非核心金融业务场景中的信息处理及交互需求，另一种方式是将其作为金融行业垂类应用模型的基础底座，在工具链增强、专有业务生态训练下成为具备更高专业化金融业务解决能力的金融大模型产品，满足金融行业对于大模型产品在专业术语的概念及理解、场景针对性解决方案以及高精度、合规性的需求。

### 通用大模型产品在金融行业的两大核心应用 ——金融大模型产品基座支持与大模型泛在化能力应用

#### 通用大模型泛在化能力应用：



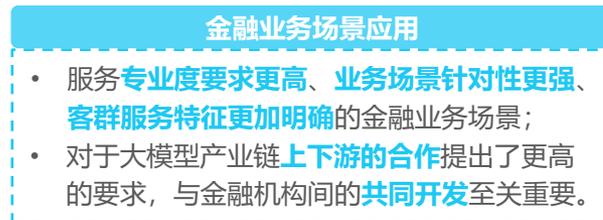
以泛在学习能力  
直接应用于金融领域



#### 通用大模型专业能力增强：



+大模型功能的增加  
+私域数据集训练



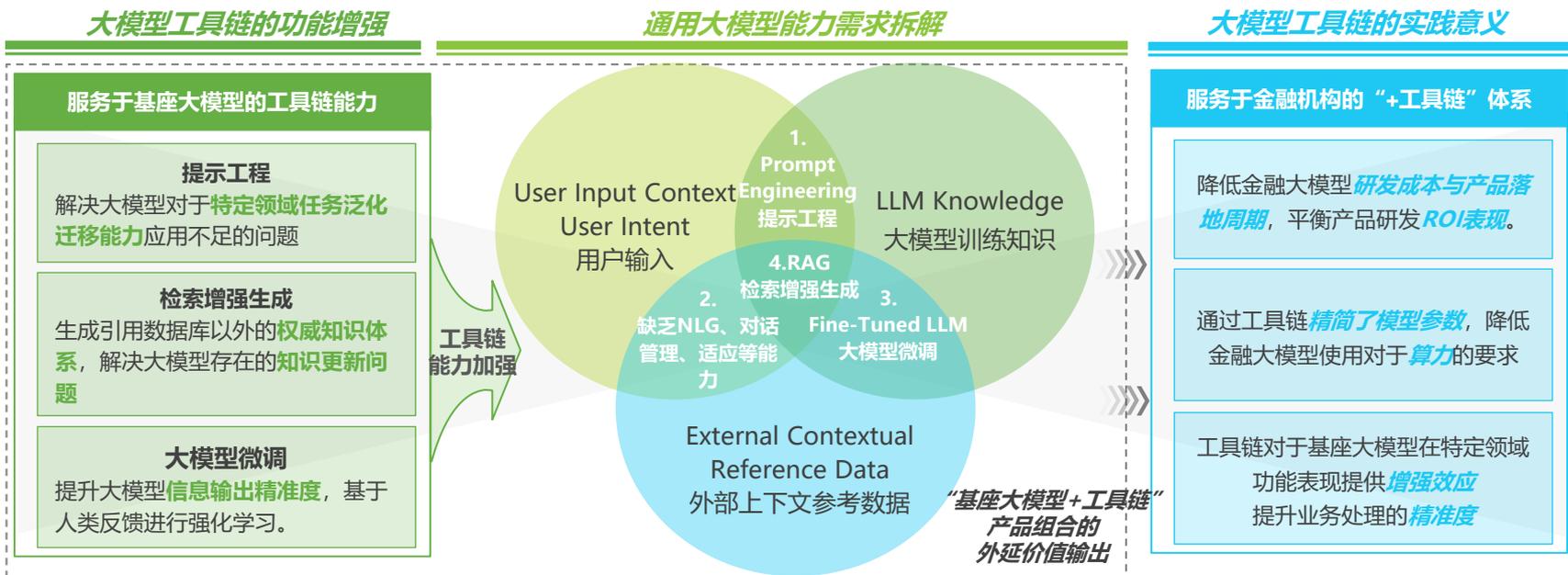
来源：专家访谈，公开资料，艾瑞咨询研究院自主研究及绘制

## 2.功能增强：大模型工具链

### 工具链作为通用大模型与垂类大模型间的桥梁，以便捷灵活的工具模块助力金融大模型建设降本增效

大模型的工具链产品是指在金融大模型产品研发的过程中，为了优化通用大模型在金融领域的应用表现，在产品开发、预训练、精调、压缩、推理、部署等环节提供高性能、灵活易用的工具模块，在简化复杂的训练流程、降低大模型技术使用门槛与落地周期的同时，提高了模型使用的精准度与可靠性，增强了金融大模型产品在投资回报方面的表现。目前在国内市场，大模型工具链的使用与通用大模型之间存在深度的绑定关系，工具链的核心技术提供商以大模型厂商为主，随着通用大模型技术能力的迭代，预计未来更多的工具链产品将逐步融合进大模型内部体系，以更高水平的通用大模型产品的形式为开发者提供完整一体化的AGI服务，但对于专业度壁垒较高的行业，例如金融行业，大模型厂商在技术层面的能力补足难以真正满足金融机构的应用需求，基于长期服务金融机构的技术积累与行业Know-how能力将成为工具链功能增强的重点，因此针对特定行业的专业化服务仍需依托于有较多行业深耕能力的专精厂商负责。

#### 大模型工具链的能力增强表现及“+工具链”的实践意义分析



来源：专家访谈，公开资料，艾瑞咨询研究院自主研究及绘制

# 3.业务适配：大模型垂类场景训练

大模型在特定信息域的定制化训练将全面提升其垂类业务场景的应用表现，数据信息的专业性与群体特征的捕捉能力将成为业务适配的关键

大模型在垂类业务场景应用的效果与其训练应用的文本数据息息相关，信息输入的质量将很大程度的决定了大模型在特定业务场景的应用表现，特别是在专业壁垒较高的金融行业，金融大模型对于金融市场运营逻辑的理解、专业术语的解读以及宏观市场环境的洞察大多依赖于大模型厂商在模型训练期间的信息输入，因此大模型厂商在金融领域的专业积累能力将极大的影响其大模型产品的应用效果，能否为模型训练提供专业化的场景信息、专业知识及行业know-how能力将成为判断其金融大模型可用性的关键。除了在大模型厂商专业知识库的场景训练之外，由金融机构提供的私域场景信息也将极大的提升金融大模型的建模精准度与问题解决能力，在合规分享的前提下，结合客户方私域信息的金融大模型产品将在模型结构与参数丰富度方面表现更为优秀，通过丰富大模型垂类场景训练的方式，为金融机构提供与业务场景及客群特征契合度更高的定制化大模型产品。

## 金融大模型垂类场景训练分析——基于业务场景与客群信息的定制化训练



来源：专家访谈，公开资料，艾瑞咨询研究院自主研究及绘制

# 金融大模型结构特征 与优势能力分析

## □ 金融大模型结构特征分析

## □ 金融大模型优势能力分析

- 产品构建期：降低成本、保留差异化竞争空间、与底层模型同步迭代
- 模型使用期：模型精准、结果可靠、产品结构轻量化
- 迭代维护期：应变能力强、维护及迭代成本低

# 阶段一：产品构建期

以通用大模型为底座的金融大模型构建体系，在具备成本控制与同频迭代优势的同时保留了产品的差异化、个性化竞争空间

## 完整构建流程之下的金融大模型产品优势能力分析 ——产品构建阶段

### 1 缩短产品落地周期，降低成本

- 从产品建设周期的角度来看，以通用大模型为基础的金融大模型产品的构建及产品落地周期要**明显短于“从0开始”搭建的垂类大模型产品**。
- 通过基座模型的支持，金融大模型产品的开发流程实现**将模型训练成本前置于通用模型阶段**，降低金融大模型产品的研发成本。



### 2 保障基础能力，留存产品差异化竞争空间

- 通用大模型作为技术基座**保障了金融大模型产品在基础技术能力以及通用业务场景的表现**。
- 在此之上，基于**不同厂商行业经验积累形成的大模型工具链与场景训练**，将充分体现不同厂商对于**金融行业的理解**以及各家金融大模型产品的**差异化竞争表现**。完整构建流程下的金融大模型产品将在保持较高基础技术能力的同时，提供差异化、客制化服务。

### 3 乘基座大模型迭代东风，与时俱进

- 建设于通用大模型基础之上的金融大模型产品，将**跟随其底座大模型的升级迭代，实现“被动”的产品升级**，其底层的技术能力、模型参数与逻辑、前沿科技的应用以及模型的使用方式与形态将**对齐大模型行业领先标准，与时俱进**。
- 对于金融大模型产品厂商来说，由于其底层的技术模块将实现一定程度的“自动化”更新，厂商可以将工作的重心放在**提升其工具链的使用效果与业务场景训练的专业度上**，更多的聚焦于大模型产品的垂类应用，实现大模型产品在金融行业深度探索。
- 以上的产品构建模式对于合作的底层通用大模型厂商的技术能力与未来发展提出了较高的要求，金融大模型厂商需**在产品构建初期明确靠谱的模型技术底座**，以减少后续产品应用、迭代的隐患。

# 阶段二：模型使用期

经由模型结构优化、专用功能增强与业务能力提升的金融大模型产品在资源需求、生成效率与问题解决专业度方面具有较强优势

相比于传统的通用大模型，金融大模型产品在使用阶段的核心优势主要来自于其结构特征与行业经验积累，通过完整垂类模型构建体系下的金融大模型产品，在模型算法能力、参数精准度以及产品结构模式方面具备更优化、更精准、更轻量化的特征，在其作用下，金融大模型产品的使用对于外部资源的需求程度相对较低，精简的模型产品不再受限于庞大的算力需求，在降低金融行业大模型产品使用门槛的同时也减少了机构的模型使用成本，结合金融大模型在金融场景下的问题解决能力与专业表现，轻量化的金融大模型产品将更有能力为金融机构提供更具性价比的专业化AGI服务，产品本身拥有较高的实用价值与推广应用条件。

## 完整构建流程之下的金融大模型产品优势能力分析

### ——模型使用阶段



与通用大模型相比，金融大模型产品结构特征：

|      |                  |
|------|------------------|
| 核心特征 | 模型算法优化，指向性强      |
|      | 模型参数精准，生成结果信息含量高 |
|      | 产品结构更加轻量化        |

### 金融大模型产品使用阶段优势能力分析：

#### ● 模型使用过程中的资源需求

- **信息需求：**金融大模型的使用具有更强的专业导向性，信息需求以专业场景下的表现为主，对于其他场景下的信息需求较少。
- **算力需求：**经过模型算法的优化与参数的精简，轻量化的金融大模型产品在使用过程中所需的算力需求相对较低。

#### ● 模型使用效率

- **缩短生成路径，提升使用效率：**经过模型结构优化，金融大模型产品在解决金融行业专业问题时的处理路径更加直接清晰，模型提供的结果生成效率更高。

#### ● 专业问题解决能力

- **专业场景的理解力：**在专业的金融行业公域、私域信息库训练的加持下，金融大模型产品对于金融业务场景的理解更加深刻。
- **解决方案提供能力：**金融大模型将为客户提供专业度更高、参考价值更强的金融问题解决方案，并且有能力服务专业要求水平更高的金融业务场景。

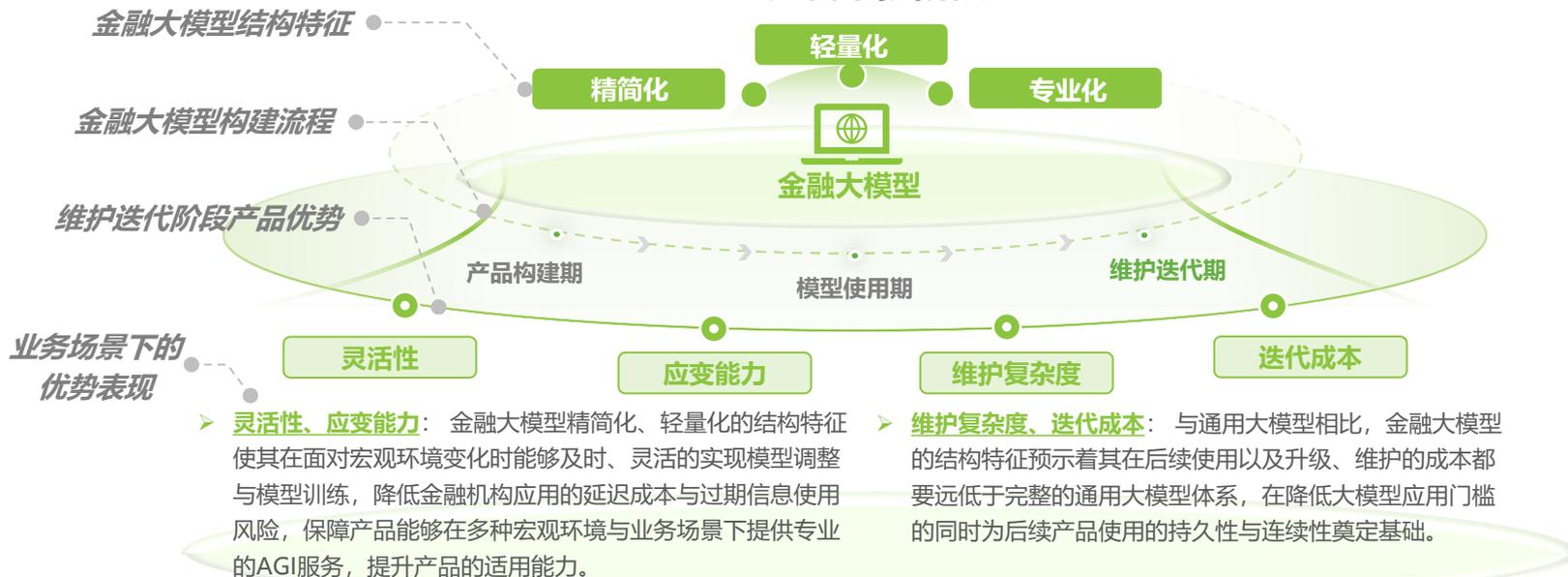
来源：专家访谈，公开资料，艾瑞咨询研究院自主研究及绘制

# 阶段三：迭代维护期

“小而精”的结构特征在保障金融大模型产品灵活度的同时，降低了产品后续使用的维护迭代门槛，提升产品的普适性

金融作为政策导向性与宏观影响性较强的服务行业，其自身机构的运营、决策与发展同外部的政策引导与市场环境息息相关，对于服务于该领域的大模型产品来说，需要有能及时根据外部信息的变化对模型结构进行调整，为金融机构提供及时、准确、专业的业务解决方案，以适配金融行业高效性、实时性需求。与传统通用大模型在金融领域的直接应用相比，垂类金融大模型“小而精”的结构特征使其在面对外部环境变化时拥有足够的灵活性与应变能力，模型调整与重新训练的反应更加敏捷，能够快速将市场变化反应在模型结构中，为金融机构提供最前沿的业务解决方案。此外，金融大模型产品的结构特征优势也在产品维护难度与成本方面得到充分的展现，为产品后续的迭代升级与长期应用奠定基础。

## 完整构建流程之下的金融大模型产品优势能力分析 ——迭代维护阶段



# 03 / 中国金融大模型产业 卓越者评选

Outstanding Enterprises

# 中国金融大模型产业 卓越者评选

## □ 中国金融科技行业卓越服务商评选

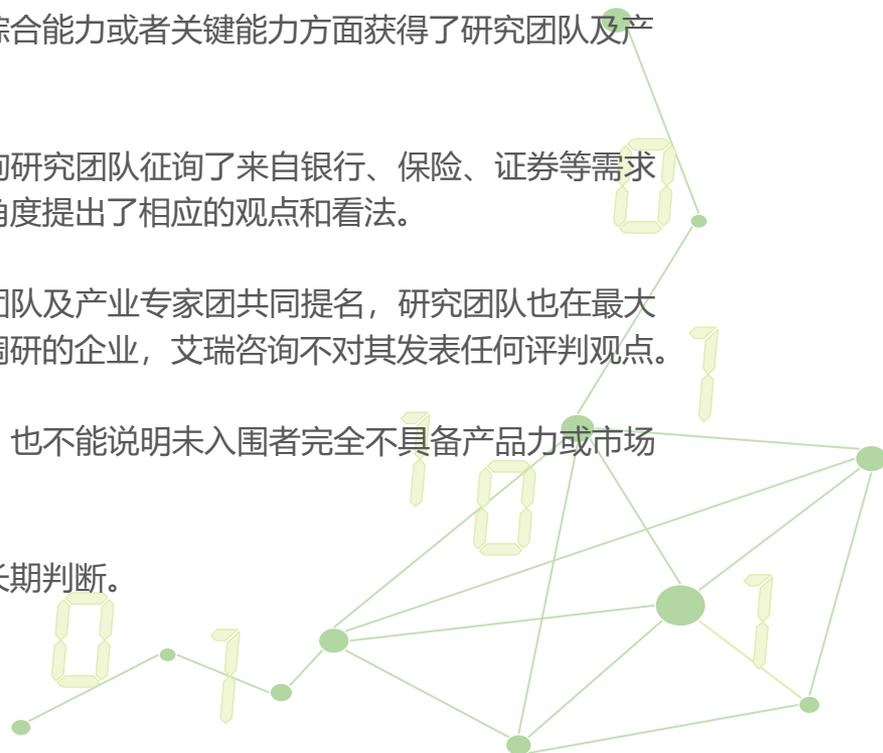
- “卓越者”评选说明
  - 入围说明
  - 评估指标介绍
- 入围厂商名单

# 卓越者入围说明

## Qualification Statement for Outstanding Enterprises

@iResearch: “中国金融大模型产业卓越服务厂商评选” 艾瑞咨询研究团队&产业专家团队

- 所有的入围者，均为接受过研究团队调研的企业，其在综合能力或者关键能力方面获得了研究团队及产业专家团的一致肯定。
- “中国金融大模型产业卓越服务厂商”的选定，艾瑞咨询研究团队征询了来自银行、保险、证券等需求端专家的意见，外部专家团站在技术水平及服务能力的角度提出了相应的观点和看法。
- “中国金融大模型产业卓越服务厂商”由艾瑞咨询研究团队及产业专家团共同提名，研究团队也在最大程度上实现了对厂商的触达。对未展开调研或者未接受调研的企业，艾瑞咨询不对其发表任何评判观点。
- “中国金融大模型产业卓越服务厂商”不代表企业排名，也不能说明未入围者完全不具备产品力或市场力的优势。
- 本次调研仅为对企业的当下能力判断，不代表对企业的长期判断。



# iResearch-卓越者 评估指标介绍

- 评测说明：评测模型主要包括服务厂商的行业深耕程度、产品与技术能力、服务能力及生态能力四个核心维度，最终评测结果的产出来自于对以上四个核心维度的综合考量，为了更全面的量化各服务厂商在核心维度下的具体表现，将核心维度进一步拆分至二级、三级评价指标，并给予其不同的评测重要程度等级，具体指标如下图所示：

## iResearch - 卓越者（2024，中国金融大模型产业服务厂商）核心评估指标



示例



一级指标  
二级指标  
三级指标

### 打分标准说明：

- 各项指标评测结果均来自于内部及外部评审团队，根据参选企业提供的资料及调研、走访结果，结合团队专家自身对行业的认知及经验积累，综合评断得出；
- 具体卓越者榜单情况见下页。

来源：专家访谈，公开资料，艾瑞咨询研究院自主研究及绘制

# iResearch-卓越者 入围厂商

## 中国金融大模型产业卓越服务厂商TOP20榜单

### iResearch – 中国金融大模型产业卓越服务厂商TOP20榜单



#### 金融业务数字化创新服务商

东方财富

度小满

马上消费  
WWW.MSXF.COM

QiFU  
奇富科技

同花顺



#### 核心技术应用服务商

百度智能云

百融云创  
Bairong, Inc.

HUAWEI 华为云

科大讯飞  
IFLYTEK

蚂蚁集团  
ANT GROUP

腾讯云



#### 金融科技IT服务商

HUNDSUN 恒生电子

恒生聚源  
IILDATA

幻方量化  
HIGH-FLYER QUANT

JDT 京东科技

SOFTSTONE  
软通动力

HOPERUN  
Information Technology  
润和软件

DCITS  
神州信息

TRANSWARP  
星环科技

中关村科金

# 04 / 中国金融大模型产业 未来趋势预判

Industry Development Trends

## 金融机构对于技术产品合规性、实践能力与稳定性的要求将指引供给端产业迭代方向；分工协作、强强结合或将成为未来产业发展的主流模式

金融行业属于政策导向型产业，在前沿技术产品的选择上对于产品的合规性、实践效果与表现稳定度方面均有较高要求，而对于金融大模型产品来说，作为其底层技术能力的生成式AI已被认为走到了技术发展曲线的膨胀阶段，如何将市场对于技术能力的期待转化为大模型的产品力将成为大模型产业发展面临的重要课题。结合金融行业应用的特殊性，未来金融大模型产业的发展将极大程度取决于需求端对于产品的认可程度，即金融机构对于大模型产品的使用意愿如何。考虑到金融机构对于新技术产品选择的审慎程度，未来中短期内大模型产品在金融领域的表现效果将极大影响机构对于该项产品的认可程度，进而影响后续产业整体的发展。对于金融大模型产品的供应方来说，如何在合规的背景下保障产品的技术能力与专业度、为机构提供更优质的AI使用体验是目前短中期发展的关键，就目前国内金融大模型市场的发展情况来看，单独一家机构很难实现从技术底座、能力增强到金融知识训练每个环节均具备行业领先优势，“多环节、多机构”的合作模式或将成为优质金融大模型产品的构建方式，由分别的技术厂商提供底层技术、能力增强与专业训练，充分发挥金融实践积累与技术创新迭代的协同作用，驱动国内金融大模型产业链的整体发展。

### 国内金融大模型产业发展趋势前瞻 ——以需求端为导向、业务与科技深度结合的跨领域产业发展

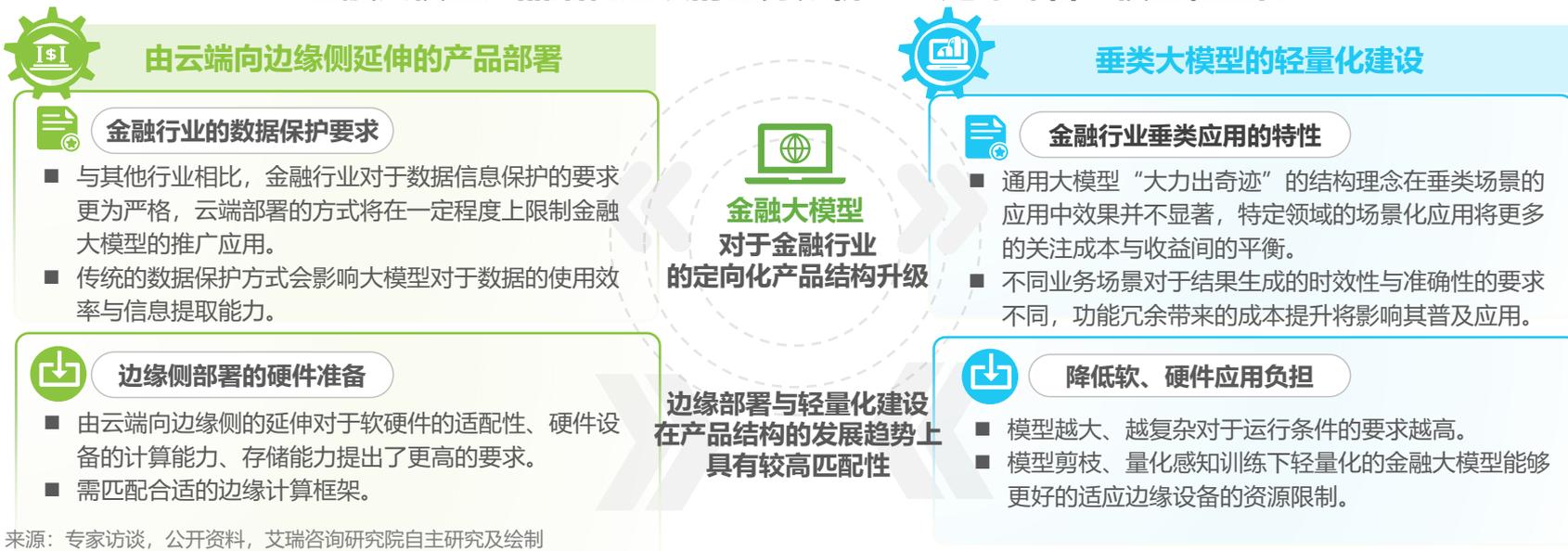


# 产品结构建设

## 数据合规应用与成本、成效间的权衡将成为影响垂类大模型实践效果的重要因素，边缘部署与轻量化建设或将成为金融大模型产品构建新趋势

金融行业对于数据信息保护有较高要求，云端部署意味着金融机构需要通过将业务数据信息上传至云端的方式来完成模型训练、获取输出结果，这对于金融机构自身的信息安全与合规应用来说是不小的挑战，市场整体的接受度与认可度也将由于该部署形式受到较大限制，多数涉及私域数据的专业化业务场景将很难接受云端部署的大模型应用，而基础的金融场景也难以发挥金融大模型在涌现性与复杂信息处理的能力，导致大模型的实践价值大打折扣。因此，大模型产品要想在金融行业实现高价值、大规模应用，应提升产品部署方式的多元化，由云端逐步延伸至边缘侧，增强产品部署的私密性，以此降低大模型产品应用过程中面临的信息安全问题，在合规的环境下满足金融机构的个性化定制需求。于此同时，与国际市场通用大模型的“规模军备竞赛”不同，庞大的参数对于垂类大模型来说过于冗余，特别是对于结果生成时效性要求较强的场景，参数量过大的大模型产品并不是好的选择，并且模型规模的扩张也会显著提高使用成本以及软、硬件的支持负担。因此，如何压缩模型、减少模型的计算复杂度与内存占比将成为未来行业发展关注的重点问题，通过合理的剪枝、低秩分解，将大模型产品轻量化，以精准的运营为客户提供稳定可靠的运算结果。

### 金融大模型产品结构建设的趋势分析——边缘部署与模型轻量化



来源：专家访谈，公开资料，艾瑞咨询研究院自主研究及绘制

# 技术能力迭代

## 以大模型能力为基础，构建具备环境感知、自主理解、决策制定与执行行动力的AI Agent智能体，推动大模型能力的落地应用

AI协同应用的技术演进路径预示着AI Agent将成为大模型技术迭代与落地应用的重要方向，在“动态决策”的大模型框架之下，AI Agent将以“超级拟人化”形态，接收外部指令、信息并进行深度理解，凭借其CoT (Chain of Thought)能力，通过计划、行动与观察相结合的方式弥补推理与实践的差距，实现有效地意图识别与决策制定。此外，大模型作为Agent的“大脑”将为其全自动决策能力提供规划、记忆、执行与工具四项能力补充，促使其有能力处理更复杂、更多元的决策任务，自主进行任务规划、环境感知、选择行动、观察纠错等一系列以任务目标为导向的行动，提升大模型技术的实际问题解决能力与决策结果的准确性，推动智能体的实践应用进程。

### AI协同应用的技术演进路径与智能体组成架构分析展示

#### ➤ AI协同应用的模式演进路径:



#### ➤ AI Agent智能体结构组成与总体架构剖析:

AI Agent 智能体=LLM+规划能力+记忆能力+工具使用



# 未来场景应用

加深技术产品间的协同应用，实现由非决策场景向决策场景的过渡，承担更重要的金融投资角色；技术、业务与合规方面的应用困境仍需重视

目前已有多项包括传统AI技术产品在内的前沿科技应用正服务于金融业务的多个环节，其服务模式与应用效果已达到较为成熟的水平。对于金融大模型来说，未来的服务模式并不是完全取代之前的模型或其他技术产品，而是应用大模型在语义理解、信息分析、内容生成等方面的优势，作为补充能力实现对于原有实践效果的提升。此外，金融大模型的应用也将逐步由简单的信息处理与图文生成拓展至与投资决策相关的核心业务板块，充分发挥大模型产品在智慧涌现与复杂信息处理方面的能力，为投研、投顾场景提供更全面的技术赋能。与此同时，金融大模型的信息处理能力也将在客户运营与挖掘中得到更为充分的应用，以高效低成本的方式触达长尾客群，实现私域用户需求的深度挖掘。在走向未来场景应用的过程中，大模型的技术能力、金融行业深耕程度与监管合规要求都将很大程度影响其的实践效果，如何在合规的前提下保障产品技术水平与业务能力将成为未来金融大模型产品之争的关键。

## 金融大模型产品的未来场景应用分析——服务模式、场景范畴与应用困境的讨论



来源：专家访谈，公开资料，艾瑞咨询研究院自主研究及绘制

# 05 / 中国金融大模型产业 专家之声

Industrial Experts' Insights

# iResearch – 中国金融大模型产业 专家之声

## 以“1+V+n”的方式构建AI基础模型的支撑体系， 与战略伙伴一同推动大模型在垂直领域的生态建设



车忠良

软通动力

集团高级副总裁

@iResearch 金融科技行业研究团队

“大模型在应用于某些特定的业务场景时需要依赖企业内部和细分领域的专业知识，这部分内容的迭代速度比较快，但受限于大模型知识来源与基础语料的更新频率，大模型自身的迭代周期相对较长，且迭代成本也比较昂贵。相比之下，专属于特定领域应用的小模型在时效性较强的细分专业领域更加适用。对于金融行业的大模型应用，我们主张的是在基座大模型的基础之上，与私域知识向量V及精确计量的n个小模型来进行配套，以“1+V+n”的方式构建AI基础模型的支撑体系，依托于我们与客户的共同研究，专注于“V+n”层面的构建。目前大模型广泛应用的两大场景分别是软件工程领域的自动编码与办公助手领域，后续随着大模型在专有领域生态建设的探索，未来将覆盖包括智能内控、智能投顾、监管合规等更多金融业务环节，实现业务场景的智能化变革。在国内大模型产业发展的基础之上，通过对行业内的垂类应用及场景普及做技术方向的迭代，并与大模型战略伙伴一道，共同打造适配国家金融机构发展的、完整的数智化信创解决方案。”

# iResearch – 中国金融大模型产业 专家之声



吴震操

恒生聚源

CEO

@iResearch 金融科技行业研究团队

**金融大模型建设需要厂商在金融领域有足够的经验积累，除了金融场景应用的数据，还包括金融业务的know-how能力**

“金融大模型产品的核心竞争点是算法、算力与数据。现阶段国内厂商在算法能力上各擅胜场，算力和公开语料也可通过外购等方式弥补，而领域数据将成为国内大模型市场差异化竞争的关键。金融大模型产品的建设需要厂商在金融领域有足够的经验积累，除了金融场景应用的数据，还包括金融业务的know-how能力。通过领域数据和机构内部数据对基座大模型进行训练、精调和强化学习，全面提升金融机构数智化应用的覆盖场景和应用大模型的文本处理和推理能力。用Agent+数据串联多个业务场景应用，重塑“由数据到知识，由知识到决策”的生成路径，为金融机构提供基于内外部数据生成的市场和业务洞察。

未来，金融大模型将作为知识管理和业务流程的引擎，实现更自然地指令控制、信息加工、内容生成以及业务流转等。对于技术厂商来说，应更多地与数据供应商和金融机构合作，探索如何更好的发挥基座模型在金融领域的效能，以更强的处理能力，更快的速度、更高的性能，打造更懂金融行业的大模型产品。”

# iResearch – 中国金融大模型产业 专家之声

## 以丰富的经验积累为金融机构提供安全可控的技术赋能， 以务实的态度携手客户实现从+AI迈向AI+的进阶



李智伟

中关村科金

CTO

@iResearch 金融科技行业研究团队

“ 目前国内大模型产业已进入快速迭代发展的全新阶段，技术与场景的匹配度越来越高，大模型在认知能力、抽取分析能力、生成能力、决策路径理解能力方面已越来越聪明，也更加人性化。尤其在金融行业，随着技术的成熟与应用快速渗透，大模型在满足金融业务主要场景需求时表现出其独特的价值和潜力。当前，正是金融大模型市场蓬勃发展的黄金时期，预示着后续的应用趋势将更加广泛而深刻。

加速大模型垂类应用落地，科技企业的垂类行业经验尤为关键。尤其是在金融行业，深厚的知识积淀和深刻的业务洞察是科技企业的核心优势。中关村科金拥有丰富的金融项目经验与技术专利，为大模型能力快速融入现有产品体系奠定坚实基础。同时，我们深耕模型工具链研发，并强化与大模型生态伙伴的协作，构建从模型选型、调优到应用部署的闭环产业链，确保高效实现从理论到实践的全面落地。

作为科技企业，中关村科金要成为金融行业数智化转型的赋能者，运用大模型技术助力金融机构做好五篇大文章，并推动全行业从+AI向AI+进阶。

”

# iResearch – 中国金融大模型产业 专家之声

## 模型底座与场景生态是金融大模型建设的两大优势 聚焦深度与精准度是金融大模型的下一步迭代方向



王丰辉

腾讯金融云  
副总经理

@iResearch 金融科技行业研究团队

“我们做大模型有两个优势，首先混元大模型是从零开始搭建的，这有助于解决大模型的风险和安全问题，也符合目前国内金融行业推进的融合创新导向。其次，混元大模型已在腾讯内部对接了600多个应用，众多的应用生态为大模型迭代提供了更丰富的场景，再先进的技术也要落到具体场景中才能产生实质价值。

代码助手与知识引擎是大模型在金融行业落地效果较好的两款产品，可用于辅助程序员代码编写，解决客服坐席、客户经理等知识储备不足的问题。目前大模型优先解决的是知识外包，随着迭代的推进，大模型逐步由泛化的广度向某领域的深度及准确度聚焦，再往后将会涉及到智能决策层面，由知识领域拓展到现实世界的认识。

对于后续规划，基座大模型肯定会关注Agent及多模态，金融大模型我们会深入金融场景，在落地层面积极推动大模型与金融业务的融合，与我们的金融客户一起，共同推动金融行业的提质增效。

”

BUSINESS  
COOPERATION

# 业务合作

## 联系我们



400 - 026 - 2099



ask@iresearch.com.cn



www.idigital.com.cn

www.iresearch.com.cn

官 网



微 信 公 众 号



新 浪 微 博



企 业 微 信



## LEGAL STATEMENT

# 法律声明

### 版权声明

本报告为艾瑞数智旗下品牌艾瑞咨询制作，其版权归属艾瑞咨询，没有经过艾瑞咨询的书面许可，任何组织和个人不得以任何形式复制、传播或输出中华人民共和国境外。任何未经授权使用本报告的相关商业行为都将违反《中华人民共和国著作权法》和其他法律法规以及有关国际公约的规定。

### 免责条款

本报告中行业数据及相关市场预测主要为公司研究员采用桌面研究、行业访谈、市场调查及其他研究方法，部分文字和数据采集于公开信息，并且结合艾瑞监测产品数据，通过艾瑞统计预测模型估算获得；企业数据主要为访谈获得，艾瑞咨询对该等信息的准确性、完整性或可靠性作尽最大努力的追求，但不作任何保证。在任何情况下，本报告中的信息或所表述的观点均不构成任何建议。

本报告中发布的调研数据采用样本调研方法，其数据结果受到样本的影响。由于调研方法及样本的限制，调查资料收集范围的限制，该数据仅代表调研时间和人群的基本状况，仅服务于当前的调研目的，为市场和客户提供基本参考。受研究方法和数据获取资源的限制，本报告只提供给用户作为市场参考资料，本公司对该报告的数据和观点不承担法律责任。



# THANKS

艾瑞咨询为商业决策赋能