滨湖金融小镇
BINHU FINANCIAL TOWN

# 金融科技行业
## 信息汇编

滨湖金融小镇

BINHU FINANCIAL TOWN

# 前言

人工智能已成为新一代信息时代的核心技术，广泛应用于多个领域，为数字经济的发展和产业数字化转型提供了底层支撑，并在各种应用场景中发挥着至关重要的作用。其中，最常见的应用场景包括推荐系统、预测分析等。过去十年来，人工智能技术在持续提高和改进，并不断冲击着人类的认知。

自2021年起，生成式人工智能连续两年入选Gartner《Hype Cycle for Artificial Intelligence》，被认为是未来重要的AI技术趋势。2022年以来，生成式AI产品不断涌现，生成内容模态多样，引起广泛的关注和讨论。2022年11月，OpenAI发布ChatGPT，定义为一个基于自然语言大模型的通用对话系统，仅用2个月就创造了APP用户过亿的新记录。此前，APP用户破亿最快的记录是字节跳动TikTok的9个月，每个创造用户过亿时间记录的APP都成为了一个时代的符号，ChatGPT的的发布同样具有划时代的意义。

作为一个现象级技术产品，ChatGPT在人工智能生成内容（AIGC）领域的表现无疑是革命性的，将对文字乃至多模态的AIGC应用具有里程碑式的重要意义，甚至可能对整个社会结构、企业生存甚至大国之间的博弈产生冲击。目前，我国人工智能产业规模已超4000亿元，各行各业都已开展人工智能布局，百度、字节跳动、阿里、华为、腾讯纷纷布局ChatGPT产品。其中，商业银行业是AI的最好最快商业级应用场景，互联网数据中心（IDC）的报告显示，我国90%的银行已经试水人工智能的应用，其中招商银行已率先引入“ChatGPT”智能对话机器人。

本期金融科技行业汇编梳理了我国ChatGPT在金融产业的发展情况，按照“支撑层（芯片、服务器和云服务）-嫁接层（大模型和大数据服务）-金融应用层（持牌金融机构、金融科技公司和监管层）”将产业链分为三层，对技术趋势和国内代表性公司也进行了归纳汇总。

# 目录
CONTENTS



22
2023/04/18 总第贰贰期

# 一、ChatGPT的发展历程

## 发展历史

**2017年6月，6500万参数的Transformer。**2017年6月，谷歌大脑团队（Google Brain）在神经信息处理系统大会（NeurIPS，该会议为机器学习与人工智能领域的顶级学术会议）发表了一篇名为“Attention is all you need”《自我注意力是你所需要的全部》的论文。作者在文中首次提出了基于自我注意力机制（self-attention）的变换器（transformer）模型，并首次将其用于理解人类的语言，即自然语言处理。

**2018年6月，1.17亿参数的GPT-1。**2018年6月，在谷歌的 Transformer 模型诞生一周年时，OpenAI公司发表了论文“Improving Language Understanding by Generative Pre-training”《用生成式预训练提高模型的语言理解力》，推出了具有1.17亿个参数的GPT-1（Generative Pre-training Transformers，生成式预训练变换器）模型。最终训练所得的模型在问答、文本相似性评估、语义蕴含判定、以及文本分类这四种语言场景，都取得了比基础Transformer模型更优的结果，成为了新的业内第一。

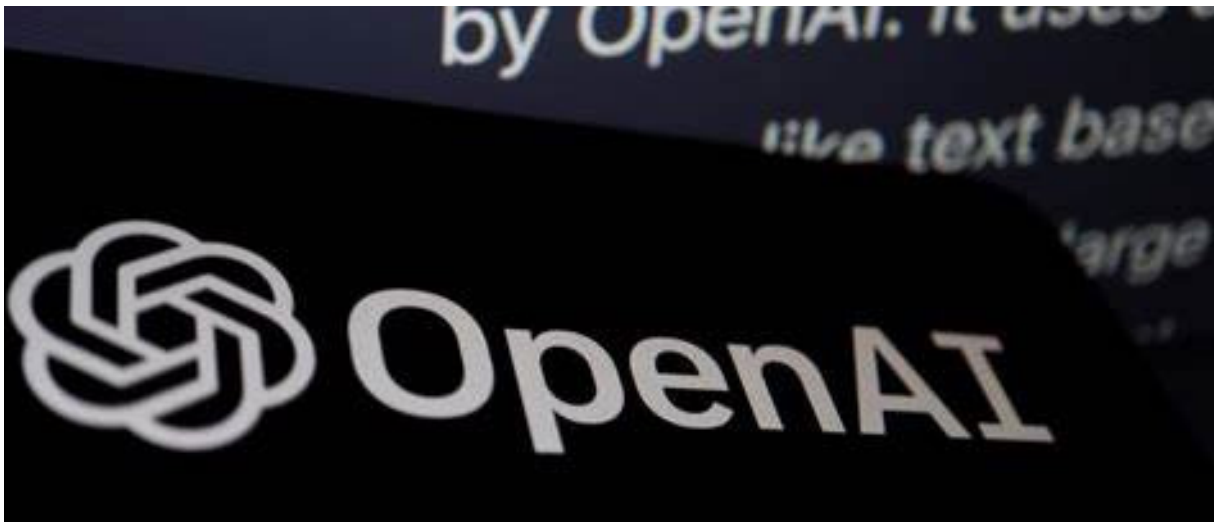
**2019年2月，15亿参数的GPT-2。**2019年2月，OpenAI推出了GPT-2，同时，他们发表了介绍这个模型的论文“Language Models are Unsupervised Multitask Learners”《语言模型是无监督的多任务学习者》。相比于GPT-1，GPT-2并没有对原有的网络进行过多的结构创新与设计，只使用了更多的网络参数与更大的数据集。在性能方面，除了理解能力外，GPT-2 在文本内容生成方面表现出了强大的天赋，并在多个特定的语言建模任务上实现了那时的最佳性能。

**2020年5月，1750亿参数的GPT-3。**2020年5月，OpenAI发布了GPT-3，这是一个比GPT-1和GPT-2强大得多的系统。同时发表了论文“Language Models are Few-Shot Learner”《小样本学习者的语言模型》。GPT-3的神经网络是在超过45TB的文本上进行训练的，数据相当于整个维基百科英文版的160倍。并且，GPT-3有1750亿参数。GPT-3作为一个无监督模型（现在经常被称为自监督模型），几乎可以完成自然语言处理的绝大部分任务，例如面向问题的搜索、阅读理解、语义推断、机器翻译、文章生成和自动问答等等。

**2022年3月，13亿参数的InstructGPT。**2022年3月，OpenAI发布了InstructGPT。并发表论文“Training language models to follow instructions with human feedback”《结合人类反馈信息来训练语言模型使其能理解指令》。InstructGPT的目标是生成清晰、简洁且易于遵循的自然语言文本。InstructGPT模型基于GPT-3模型并进行了进一步的微调，开发人员通过结合监督学习+从人类反馈中获得的强化学习，来提高GPT-3的输出质量。通过训练，获得了更真实、更无害，而且更好地遵循用户意图的语言模型 InstructGPT。

**2022年11月，约20亿参数的ChatGPT。**2022年11月30日，OpenAI公司在社交网络上向世界宣布他们最新的大型语言预训练模型（LLM）：ChatGPT。ChatGPT是OpenAI对GPT-3.5微调后开发出来的对话机器人。可以说，ChatGPT模型与InstructGPT模型是姐妹模型，都是使用RLHF（从人类反馈中强化学习）训练的。不同之处在于数据是如何设置用于训练（以及收集）的。从功能来看，ChatGPT可以用更接近人类的思考方式参与用户的查询过程，可以根据上下文和语境，提供恰当的回答，并模拟多种人类情绪和语气。

**2023年3月，1000亿参数的GPT-4。**2023年3月14日，OpenAI公司在社交网络上向世界宣布他们最新的大型语言预训练模型：GPT-4。GPT-4 是一个大型多模态模型（接受图像和文本输入，发出文本输出），虽然在许多实际场景中不如人类，但在各种专业和学术基准测试中表现出人类水平的性能，例如通过模拟律师考试并获得了约排名前10%的成绩。在日常交流中，GPT-3.5 和 GPT-4 之间的区别可能很小，当任务的复杂性达到足够的阈值时，差异就会出现——GPT-4 比 GPT-3.5 更可靠、更有创意，并且能够处理更细微的指令。



# 二、ChatGPT的发展现状和应用前景

## 国内外ChatGPT技术的发展现状

ChatGPT技术扩散堪比工业革命。微软首席执行官萨蒂亚·纳德拉表示，现阶段人工智能领域的发展已经可以用“指数级”来形容，ChatGPT等生成式人工智能技术扩散堪比工业革命。微软将ChatGPT视为新一代技术革命，将ChatGPT整合进Bing搜索引擎、Office全家桶、Azure云服务、Teams程序等产品中。而ChatGPT技术的核心AI大模型，将成为继移动互联网之后，未来最大的技术平台；而以聊天机器人为界面，加上图像、音乐、文本等多模态模型的发展，将诞生世界级大型企业。

ChatGPT是生成式AI的一种形式，Gartner《2022年度重要战略技术趋势》将生成式AI放在了第一位，并预测到2025年，生成式AI将占到所有生成数据的10%（目前还不足1%）。红杉资本预测，生成式AI有潜力产生数万亿美元的经济价值。全球范围来看，大模型的应用已经不局限于NLP领域，计算机视觉、多模态等领域的大模型开始涌现。目前大模型包括三类：自然语言处理（NLP）模型，如 Open AI的 ChatGPT 模型，Google的LaMDA；计算机视觉（CV）模型，如微软Florence；多模态模型，如Open AI的 GPT-4 模型，Google 的 Parti。

回顾ChatGPT的迭代，OpenAI至少进行了三次技术路线的“自我革命”。从2018年GPT-1的推出到今年的GPT-4，OpenAI用了近5年。人工智能在当下的中国展现出全面开花的热闹，资金雄厚、人才济济的互联网大厂被视作最能与OpenAI比肩的选手。

ChatGPT引发的大模型热潮汹涌，各大企业AI大模型落地的消息接踵而至。3月16日，百度率先发布大语言模型“文心一言”，百度在会上用PPT展示了文心一言在文学创作、商业文案创作、数理推算、中文理解、多模态生成五个使用场景中的综合能力。发布会后一小时内，排队申请文心一言企业版API调用服务测试的企业用户已达3万多家，申请产品测试网页多次被挤爆,文心大模型在市场格局中处于第一梯队,IDC评估结果显示，百度文心大模型处于第一梯队，产品能力、生态能力达到L4水平，应用能力达到L3水平。

4月7日，阿里云自研大模型“通义千问”开始邀请用户测试。通义千问是一个超大规模的语言模型，功能包括多轮对话、文案创作、逻辑推理、多模态理解、多语言支持，能够帮助用户续写小说，编写邮件等。现阶段该模型主

要定向邀请企业用户进行体验测试。目前，通义千问仅支持自然语言处理，尚不支持文生图等跨模态功能，4月8日，京东集团副总裁何晓冬称将在今年发布新一代大模型“ChatJD”，同日，华为云介绍了“盘古”大模型的进展和应用。4月9日，基于360GPT大模型开发的人工智能产品矩阵“360智脑”率先落地搜索场景。更多的AI大模型正在飞奔而来的路上。腾讯、字节跳动、同花顺等企业的AI大模型产品近日将会陆续亮相。

## 国内类ChatGPT大模型

| 一、已公布具体数据 | | | |
|---|---|---|---|
| 模型名称 | 自研单位 | 参数规模 | 训练数据来源 |
| 文心 | 百度 | 1000亿 | 百度自有 |
| MOSS | 复旦大学 | 200亿 | 公开的中英文数据 |
| HunYuan | 腾讯 | 10000亿 | —— |
| 通义 | 阿里 | —— | 达摩院内部数据 |
| 盘古 | 华为 | 10000亿 | 华为语料库 |
| ChatJD | 京东 | 千亿级 | —— |
| 二、已公布名称即将发布的大模型 | | | |
| 模型名称 | 自研单位 | 模型名称 | 自研单位 |
| 天工 | 昆仑万维 | 360智脑 | 360 |
| 子曰 | 网易 | 商量 | 商汤 |
| 星火 | 科大讯飞 | | |

*数据来源：公开数据整理*

大模型的竞争很可能像九十年代PC操作系统的竞争一样，具有“垄断性”的倾向和趋势，其本质还是在于大模型和操作系统一样，都是一个技术新时代的“基础设施”。如同需要搭乘操作系统的软件一样，所有的人工智能产品，尤其是生成式人工智能，乃至未来可能的通用型人工智能，都需要依靠背后的人工智能大模型才能完成训练、输出等一系列动作。目前国产大模型与美国的国际顶尖大模型相比仍然有一定的差距，涵盖数据训练、算法等方面。但恰恰是落后的时候，要直面差距、接受批评、迎头赶上。自2020年起，中国的大模型数量骤增，仅2020年到2021年，中国大模型数量就从2个增至21个，和美国量级同等，大幅领先于其他国家。未来，不管是政府还是资本方面，都应给予大模型研发相关的企业和机构更多的支持和宽容，共同推进中国人工智能大模型的发展。

# 以ChatGPT为基础的金融应用场景

以ChatGPT为基础、加载专业金融数据库的应用程序，兼顾自然语言处理能力和严格金融专业知识，将为金融行业带来更高效、更准确的信息处理和决策分析能力，同时也将为金融机构提供更高质量的客户服务和风险管理能力。

## ChatGPT的金融场景

| 能力 | 金融运用场景 | | |
|---|---|---|---|
| 文本理解 | 市场行情分析 | 风险监控 | 客户分析 |
| 对话 | 客服、投顾 | 营销 | 金融知识科普 |
| 文本生成 | 公告撰写 | 产业报告编写 | |
| 程序编写 | 交易程序 | 简化管理流程 | |

*数据来源：公开数据整理*

**1.智能投顾与客服。**ChatGPT加持下的智能投顾，所提供的建议与针对性营销更加契合客户的特定需求和偏好，回应语句更加顺畅、合理且能应变的情况更多。**一方面，ChatGPT能更简易的说明金融知识。**金融产品通常具有很多技术术语和复杂的流程，这使得客户很难理解和使用这些产品。ChatGPT的对话方式更趋近于人类，可以帮助客户更轻松地理解和使用这些产品。**另一方面，ChatGPT更适应于应对多样化的情绪交流。**随着金融机构客户数量的增加，客户服务团队可能难以承受大量的询问和投诉，经过特殊设计的模型更能察觉人工不易察觉的情绪反馈，从而响应针对性内容，用户体验会更优。海通证券正在探索ChatGPT在证券业账户全景分析、智能投顾服务等场景的应用。

**2.市场分析与风险识别。**金融行业涉及的数据量非常大，包括市场趋势、客户投资组合、历史交易记录等等，行业分析人员一般要收集财务报告、市场数据和新闻文章，使用此信息创建财务模型以预测未来绩效并评估潜在风险。经过专业数据库调整的ChatGPT可以对这些信息并行处理和分析，简化流程帮助金融分析师更准确地评估公司的财务状况、盈利能力、偿债能力等指标，从而预测其财务状况及未来的发展趋势。

**3.撰写金融报告及论文。**ChatGPT具备强大的指令学习能力，其能够理解的任务指令不仅包括回答问题，还包括信息检索、文章写作、问题求解、程序设计、作曲等等。同时，ChatGPT能够精准捕捉上下文所确定的代词所指，在多轮对话中准确进行意图识别。ChatGPT撰写的金融业论文已具有相当效应：根据相关文献[ Dowling, M., & Lucey, B. (2023). ChatGPT for (finance) research: The Bananarama conjecture. Finance Research Letters, 103662.]，在32个具有由经验丰富的作者和审稿人组成的团队审核下，最终版本生成的金融论文有较高概率通过审核刊印；财通证券研究团队运用ChatGPT撰写超过6000字的医美行业研究报告。

**4.自动化交易编程。**GPT模型的预训练数据中包含大量源代码，因此，它在基础代码编程效率上有较大优势。ChatGPT可以自动化生成交易策略和执行交易，通过ChatGPT，机构可以更快速、准确地执行交易，提高交易效率和收益率。明泽投资在量化交易上的底层代码许多已由ChatGPT完成，中泰证券尝试将相关技术用在辅助系统建设代码编写方面。

**5.金融知识科普。**金融机构可以利用其在金融领域的海量知识储备，运用在金融消费者教育领域。ChatGPT通过模型和基础专业知识的训练，可为金融用户提供生动、有趣、及时且准确率高的金融科普知识，提升普通消费者的金融素养。

**6.参与日常管理。**金融机构有许多日常事务性工作，需要人员投入大量精力处理，诸如人事、会务、活动等工作，而引入ChatGPT能够大量节省交流沟通、方案写作、PPT制作等方面的成本，使员工能投入更重要的工作。

在金融机构中，商业银行业是AI的最好最快商业级应用场景。商业银行应用人工智能在算法风险、隐私保护、信息安全等方面仍然面临较大的挑战。在进行数据分析、采集、存储、使用时一定要注意客户信息隐私的安全。如能合理利用，以ChatGPT为代表的人工智能可广泛应用于产品、流程、营销、运营和风控等多方面。

**1.应用于智能客服管理。**语音机器人苹果Siri、百度小度、微软小冰、天猫精灵已泛应用于居民生活中；商业银行也实现了各种各样的AI应用场景，如对账智能外呼、信用卡催收智能外呼、个人客户回访智能外呼等，ChatGPT强大的内容定义和智能交互能力将升级银行智能机器人外呼服务，使交互更加智能与精准。

**2.应用于营销服务。**ChatGPT能够解决线上线下协同营销过程中的断点问题，帮助商业银行根据客户画像开展千人千面的服务与营销。ChatGPT可帮助员工搭建一个全能型数字化服务平台，随时随地获取产品的最新数据、客户喜好、用户需求，结合客户的实时表情行为分析，为客户做好个性化的推荐服务。

**3.应用于产品全生命周期管理。**商业银行可运用用户画像、推荐引擎、大数据等人工智能技术整合所有产品数据，以大数据和深度学习分析模型进行产品分析、数据积累与挖掘，不断推进产品改进，提升用户体验；对产品的引入及退出进行全方位管控，引导产品快速更新与迭代，发挥最优价值。

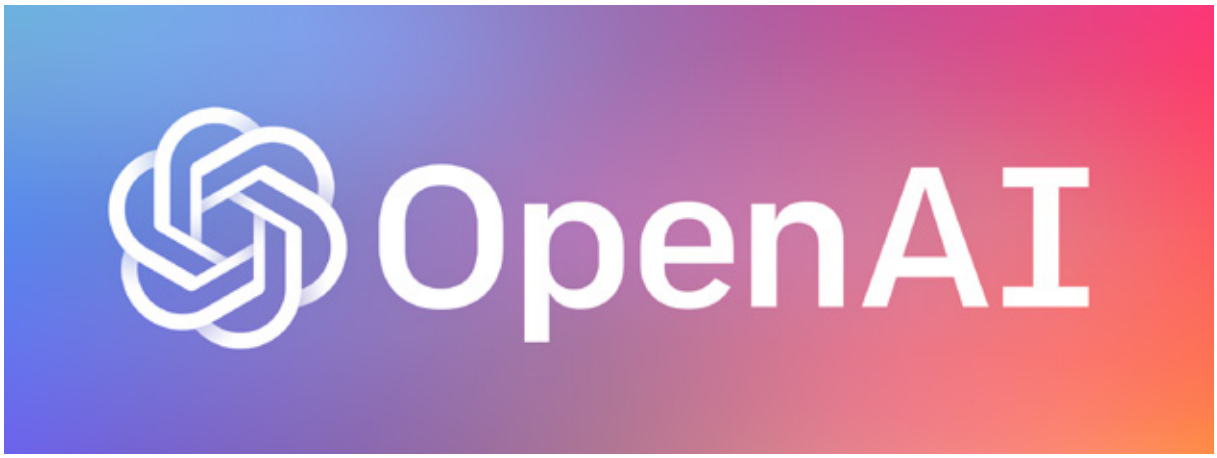
**4.应用于RPA场景拓展。**RPA是流程自动化机器人（Robotic Process Automation）的简称。RPA通过模拟人在电脑中的一系列操作，实现业务流程中重复性、规范性、跨系统操作的数据采集、传输、加工等事项自动处理，帮助商业银行快速处理大量重复性工作。但RPA主要处理结构化数据，大量非结构化数据、复杂业务场景仍然无法处理。ChatGPT以强大的自然语言生成、逻辑推理、多样性、灵活性、可扩展性等特点将推动RPA机器人能够理解人类的语言并作出判断和决策，可帮助商业银行拓展更多RPA应用场景。运用ChatGPT可生成千人千面的定制化理财、运营和风险、可疑交易报告。

**5.应用于流程优化。**将商业银行网点智能设备的形态、功能、服务状态、服务年限、运行速度、库存状态等数据输入至AI模型；实时采集客户在银行网点的办理状态和时间等数据，智能推测设备是否运行正常。在客户出现操作异常时，如客户在使用智能设备过程中出现超长办理、无法提交等现象，及时反馈至后台处理中心，并通过ChatGPT自动记录异常信息，自学习分析原因，快速生成功能优化需求，帮助商业银行实现敏捷开发。

**6.应用于网点竞争力提升。**运用ChatGPT等人工智能技术对网点的资产状况、客户特征、员工特性以及周边商圈情况进行智能化的感知，将网点的实时地理位置、附近人均收入水平、周边商圈数量、政府机构、企业单位、居民住宅、学校医院数、同业网点数、人口密度、人口流动性、停车位、交通便利度、位置重要性进行画像。运用多种组合模型进行虚拟银行形态配置，并在真实环境中试运行；并通过改变某一变量或几个变量产生的网点竞争力提升，测算出最佳的银行网点运营模式并形成网点竞争力报告。

**7.应用于数字化运营推广。**招商银行已尝试应用ChatGPT生成的文案推介信用卡。商业银行可运用ChatGPT海量的文本数据资源和分析能力，快速产出同业调研分析报告，进行业务产品的迭代优化，升级用户体验。还可运用ChatGPT辅助撰写商业计划书、功能需求文档、产品宣传文案等，这将大幅提升商业银行的数字化运营能力。

**8.应用于商业银行反赌反诈。**当前，电信网络已成为发案最多、上升最快、涉及面最广、人民群众反映最强烈的犯罪类型。银行传统的风险客户识别、电话回访、账户管控等反赌反诈工作需耗费大量人力物力，但仍存在管控不够精准的现象。银行运用叠加ChatGPT技术的智能外呼机器人可实现差别化智能外呼、个性化客户识别，将极大减少商业银行反赌反诈工作压力，提升防控的精准性和有效性。利用数据的不断丰富和完善，人工智能自学习精准预测可能发生涉赌涉诈的账户，帮助银行采取有效措施进行管控，在减少涉案账户的同时减轻客户投诉压力。



**国内已有许多运营主体从事相关工作，主要包括客服、营销、服务数字化转型等领域。**

## 浙大网新

浙大网新科技股份有限公司成立于2001年，以浙江大学综合应用学科为依托的信息技术咨询和服务集团，行业数智化专家。金融科技上，浙大网新聚焦传统金融产品重塑、全方位金融总包、智能金融知识图谱、智能金融理财、互联网金融服务、企业信用画像等领域帮助客户重塑传统金融产品与服务，长年服务于美国道富银行、中国外汇交易所、上海清算所等重量级客户。代表性产品有BlueMorphoIDE（闪蝶）、网新恒天金融数据服务平台、微宽投研系统、恒信互联网融资交易系统、IN-Talk智能语义问答系统等。

## 云从科技

云从科技集团股份有限公司成立于2015年，是第一家在科创板成功上市的人工智能平台公司，致力于助推人工智能产业化进程和各行业智慧化转型升级。一方面，打造了人机协同操作系统，为客户提供信息化、数字化和智能化的人工智能服务；另一方面，基于人机协同操作系统，赋能智慧金融、智慧治理、智慧出行、智慧商业等应用场景。金融科技上，云从科技先后完成中国银行生物识别算法项目、打造“5G智能+智慧网点”、山东省城商行联盟人脸识别平台等项目，持有集成生物识别分析平台（IBIS）、灵云数据智能风控平台等金融科技产品。

## 凌志软件

凌志软件股份有限公司成立于2003年，是一家聚焦金融科技的软件公司。为客户提供咨询、设计、开发、维护等全方位软件服务，业务范围涵盖了证券、保险、银行、信托、资产管理等金融领域。公司致力于云计算、大数据、人工智能等新兴技术在金融行业的应用，在金融科技领域拥有较强的竞争优势。金融科技上，凌志软件具有O2O客户智能精准营销服务、大投行业务数字化转型解决方案、面向机构服务的综合金融服务等多项解决方案。服务国泰君安、华泰证券、中信证券、招商证券、平安资产管理、富国基金等多家金融客户。

## 华胜天成

北京华胜天成科技股份有限公司成立于1998年，是一家面向全球客户提供云计算解决方案和数字化服务的公司。华胜天成按照“云数为轴、四轮驱动”战略方向，以“新基建”和“信创”为两大重点工程，积极打造行业数字化方案、产品增值服务、产业数字化运营、专业运维外包的核心竞争力。金融科技上，华胜天成主打智慧银行解决方案，实现服务于银行数据业务的、基于鲲鹏的、全栈安可的国产化替代换战略，形成双方合作的案例。借助华胜天成在移动金融的应用建设经验，推动华为移动金融安可终端项目的实施，拓展华胜天成在手机终端市场与华为在金融行业的市场合作。华胜天成自主可控中间件产品，已完成与鲲鹏云服务、华为云Stack鲲鹏混合云解决方案兼容性认证。

## 信雅达

信雅达科技股份有限公司成立于1996年，是一家面向行业客户提供完整解决方案、产品及服务的金融科技企业。近年来，公司持续聚焦数智化新技术与金融业务的深度融合，持续输出面向未来的新一代数智化端到端大型解决方案及运营服务，致力于以数字金融技术链接金融行业客户与政企客户。金融科技上，信雅达业务覆盖数字化运营、数字化营销、数字化风控、数字化监管、数字化产业金融等领域，同时生产支付终端、信息加密设备等产品。客户涵盖众多国家级、地方级银行。

# 三、国内ChatGPT相关产业链发展现状

随着国内厂商陆续布局ChatGPT类似产品，GPT大模型预训练、调优及日常运营将引入大量算力需求，且AI模型日益复杂、计算数据量快速增长、人工智能应用场景不断深化，AI相关产业将迎来跨越式增长。从ChatGPT的金融产业链构成上，由支撑层（硬件）嫁接层（软件）和应用层共同组建完成。其中：

**·芯片是ChatGPT的大脑，是决定AI计算能力的核心因素。**一方面，专用的AI芯片在构造上与传统PC芯片不同，具有专门服务AI的架构和设计，尤其是满足AI训练和推理时产生的大量并行计算。另一方面，由于ChatGPT产业的高度技术集成度，关键技术及资源都集中在头部一至两家机构中。虽然GPU为现在主流算力芯片，但为达成算力最大化，异构芯片很可能会脱颖而出。

**·服务器是ChatGPT的身体，是影响芯片效能发挥的支撑措施。**服务器由CPU、GPU、内存、主板、散热、电源等组件构成，是服务AI训练和推理的物理平台，出于成本、能耗等因素考量，部分AI应用厂商会选择云服务开展业务。

**·插件是ChatGPT的眼耳，是弥补弱点、扩展使用场景的必须步骤。**原始的ChatGPT接口只具备广泛且浅层的文字处理能力，缺乏严谨产业知识、仅能通过问答交流、无法联动外在部件等缺陷，使其无法应对各行各业的现实需求。只有发挥AI平台公司在数据、技术和场景等方面的积累，为ChatGPT加装特定插件，才能实现效用最大化。

**·应用是ChatGPT的灵魂，金融场景是其重要应用领域。**金融科技产业是金融业务与现代科技的“最后一公里”，只有在ChatGPT接口的基础上，融合嫁接具体的金融业务，才能真正触达用户。


# 支撑层——芯片、服务器及云服务提供商

## 1.芯片设计与制造

芯片是以半导体为原材料，把集成电路进行设计、制造、封测后，所得到的实体产品。目前，按照工艺制程可以分：5nm芯片、7nm芯片、14nm芯片、28nm芯片等，国际最先进的制程是台积电和三星的3nm（良品率仅有10-20%），国内最先进的制程是中芯国际的14nm。

**ChatGPT的运算芯片在架构和性能上有别于传统芯片。**ChatGPT先后经过四次迭代，参数量与训练文本量呈指数级增长：GTP-1（约5GB训练文本，1.17亿参数量）、GPT-2（约40GB训练文本，15亿参数量）、GTP-3（约45TB训练文本，1750亿参数量）以及刚刚发布的GPT-4（1000万亿）。因此，高效智能的AI专业运算芯片，是支撑ChatGPT模型在海量的训练数据中学习、实现高质量的生成输出的必要组成部分。特性上：**一要具有较强的分布式计算能力。**包括数据并行、模型并行、流水并行等分布式计算方案，计算效率尤其关键。**二要带有大容量高带宽的内存方案。**在每个AI芯片内部有效提升数据处理能力和算力利用率，结合HBM以及CXL等新型存储技术进一提升本地存储能力和算力利用率。**三要支持更高的单芯片计算负载。**以降低整体系统复杂度，并降低TCO成本。

## AI计算芯片分类表

| 分 类 | | 国内外代表企业 | |
|---|---|---|---|
| 通用类芯片 | GPU | 英伟达、AMD | 国芯科技 |
| | FPGA | Altera、Xilinx | 复旦微电、安路科技、紫光同创 |
| 定制类芯片 | 半定制 | | 深鉴科技、百度 |
| | 全定制（ASIC） | 谷歌TPU | 寒武纪 |
| 类脑计算芯片 | | IBM、WestWellLab | |

*数据来源：公开数据整理*

受限于设备、技术专利等因素，中国芯片厂商在AI计算芯片这一细分领域和国外巨头的差距远远比其他领域要大，根本原因就是技术门槛非常高，核心技术只掌握在及其少数的公司手上。国内已初步构建出以下芯片产业格局：

**·寒武纪**

寒武纪是一家专注于人工智能芯片产品的研发与生产商。提供云边端一体、软硬件协同、训练推理融合、具备统一生态的系列化智能芯片产品和平台化基础系统软件。其代表性产品为思元AI训练设备（包括智能加速卡、智能边缘计算模组）、智能处理器Cambricon系列及MagicMind开发平台。

**·国芯科技**

杭州国芯科技股份有限公司成立于2001年，专注于数字电视及物联网人工智能领域的芯片设计和系统方案开发。深耕人工智能领域，率先推出多款面向物联网的人工智能芯片，拥有自主研发的神经网络处理器、指令集及编译器等核心技术。芯片产业上，一方面，该公司正在积极布局可用于人工智能应用的CPU和芯片设计技术，正在研发的募投项目——基于C*CoreCPU核的SoC芯片设计平台设计及产业化项目中包括“开发面向边缘计算与人工智能应用的SoC芯片设计平台”内容，基于在研的RISC-V指令带有AI功能的CRV4AI/CRV7AI两大CPU核，面向边缘计算与人工智能应用，设计开发两类SoC芯片设计平台——基于CRV4AICPU核的SoC芯片设计平台和基于CRV7AICPU核的SoC芯片设计平台，分别用于边缘计算领域的嵌入式人工智能SoC芯片以及高端控制和数据处理密集应用的人工智能主控芯片。目前该项目在研发建设过程中。另一方面，公司参股企业南京智绘微正在开发完全自主指令的GPU芯片，其新一代GPUIDM929已基于14nm工艺完成设计并进入流片阶段。公司和南京智绘微已签订合作协议，在该芯片的研发投入、后端设计、流片验证和市场应用推广方面展开合作。

**·复旦微电**

上海复旦微电子集团股份有限公司成立于1998年，是国内从事超大规模集成电路的设计、开发、生产（测试）和提供系统解决方案的专业公司，也是国内成立最早、首家上市的股份制集成电路设计企业。芯片产业上，该公司常年耕耘FPGA芯片的研发。从2004年至今，该公司已陆续推出百万门级、千万门级和亿门级FPGA产品，主要应用于高速通信、信号处理、图像处理、工业控制等应用领域。公司的PSoC芯片兼具了SOC的灵活性和通用性、FPGA的硬件可编程性和专用AI加速核或GPU的高效性，针对人工智能的不同应用领域，可将各种算力需求和控制逻辑用最合适的资源组合实现，可以低成本、高效能的实现人工智能应用的快速部署、应用的动态重构和快速升级。

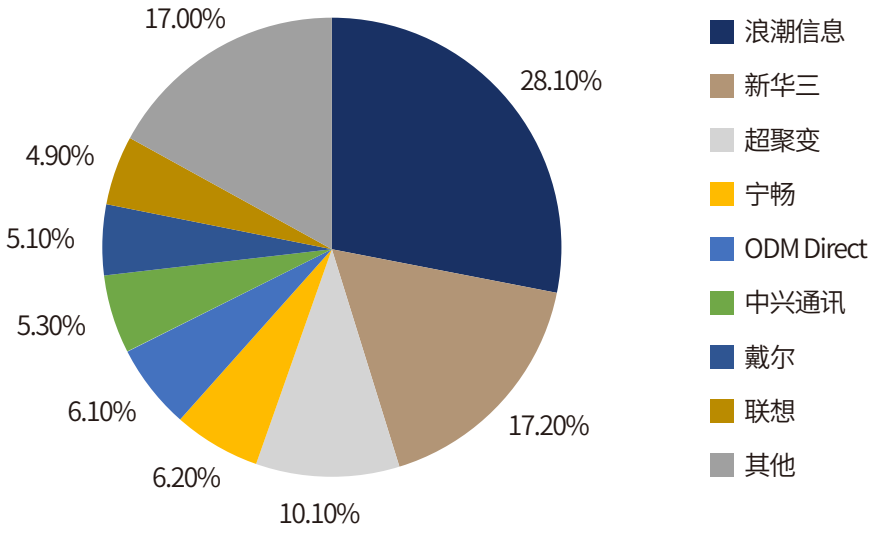
## 2.服务器及云服务商

服务器是整合芯片、内存、主板、散热等组件并经过科学优化的高性能计算机，按架构分为标准服务器和异构服务器。AI专用服务器多采用1-2块CPU+多块GPU的架构组成，按照GPU数量可以分为四路、八路和十六路服务器，其中八路AI服务器最常见。

**能提供更大吞吐量的异构服务器成为ChatGPT的首选服务器。**据OpenAI公开数据推测，每进行一次GPT-3175B模型的预训练需要的算力约3640PFlop/s-day，若以浪潮信息目前算力最强的AI服务器NF5688M6进行计算，在预训练期限分别为3、5、10天的假设下，单一厂商需采购的AI服务器数量分别为243、146、73台。服务器搭载2颗第三代IntelXeon可扩展处理器+8颗英伟达A800GPU，最低芯片采购成本约为0.7亿元。

我国相关厂商在服务器领域拥有较高市场占有度，尤其在新兴市场的AI服务器，浪潮信息、华为和新华三等都占有较高份额。

## 2022年全球服务器市场占比




**·同方股份**

同方股份有限公司是1997年由清华大学出资成立的高科技企业。形成了核技术应用、智慧能源、数字信息、科技创新主干产业集群，技术、产品和服务已遍及五大洲一百余个国家和地区。云服务方面，同方有云是中国市场重要的云端服务和运维提供商，在上海、深圳、广州、长沙、武汉等多地设立分支机构，从基础环境建设、硬件设备、云平台操作系统，到存储、大数据、云安全产品与方案，为企业提供公有云、私有云、混合云建设一站式服务，打造以开源技术为核心的系列产品，包括UOS私有云平台、UDS分布式存储系统、UCS容器管理平台、UOS超融合一体机等。

**·城地香江**

上海城地香江数据科技股份有限公司成立于1997年，主营业务逐步由工程建设拓展为云基础设施产业、云数据服务产业和岩土工程产业三大板块。云基础设施服务产业，依托在通信基础设施领域深耕20多年经验，拥有机电工程总承包一级、拥有增值电信业务经营许可证（云服务牌照）、ABB开关柜授权和ISO9001、泰尔认证等。为三大运营服务商、阿里巴巴、拼多多、快手、华为等提供相关产品及服务。在北京、上海、广州、深圳、南京等地建有多个高等级数据中心，并顺利通过CQCA级数据中心认证。

**·环旭电子**

环旭电子股份有限公司成立于1976年，现已成为全球电子设计制造领导厂商，在SiP模块领域居行业领先地位，同时向国内外知名品牌厂商提供设计、生产制造、微小化、行业软硬件解决方案以及物料采购、物流与维修服务等全方位D(MS)2服务。服务器方面，环旭电子提供L10服务器系统设计服务，包括服务器主板、固件BIOS和

BMC、子卡（背板、附加卡等）、外壳和散热设计以及系统集成，设计在台湾工厂进行管理，在深圳和昆山工厂量产。具有生产单/双路服务器主板和系统、刀片服务器等产品的能力。

**·云赛智联**

云赛智联股份有限公司是上海仪电（集团）有限公司旗下的上市公司，是一家以云计算与大数据、行业解决方案及智能化产品为核心业务的专业化信息技术服务企业。云服务方面，云赛智联运营宝山云计算中心和徐汇数据中心两个DCaaS，均以国际T3+标准规划、设计，总面积15200平方米，共有机柜1950个。

## ● 嫁接层——人工智能平台公司

2022年3月，OpenAI开始逐步开放内部插件，并允许第三方在GPT模型内创建插件。一方面，插件拓展了GPT模型自身能力的边界，为模型添加“眼睛和耳朵”。以ChatGPT模型为基础，接入第三方插件使得大模型得以联网，既弥补了原模型在信息真实度、时效性等方面的不足，也拓展了用户的使用场景，更大程度上解放生产力。另一方面，随着第三方插件逐渐丰富，最终将形成一个基于AI的全新系统生态，甚至会改变生产关系。ChatGPT模型开创性地发现了RLHF+transformer+精细标注数据这一方式，通过工程化落地的方式使普通人感受到AI的强大与便利。ChatGPT的引入与广泛应用，就像IOS和安卓重构了手机、互联网等产业，将对人们生活产生极大影响。

**丰富完善的AI产业集群对拓展ChatGPT使用场景有极大地促进作用。**经过多年发展，人工智能产业已深度服务赋能政务、交通、金融、教育、工业等方方面面，积累了多样的数据、场景和客群。市场上AI公司的富足程度，将极大影响ChatGPT服务社会发展的触及面。

中国人工智能市场规模在2016年-2020年持续增长，市场规模从2016年的154亿元增长至2020年的1280亿元，年复合增长率达到69.79%。随着新基建产业愈发受到国家重视，人工智能产业未来将持续增长，预计2022年将达2729亿元。

**·星环科技**

星环信息科技（上海）有限公司成立于2013年6月，致力于打造企业级大数据基础软件，围绕数据的集成、存储、治理、建模、分析、挖掘和流通等数据全生命周期提供基础软件与服务，构建明日数据世界。公司以上海为总部，以北京、南京、广州、新加坡为区域总部，在郑州、成都、重庆、济南设有支持中心，同时在深圳、西安等地设有办事机构，并在加拿大设有海外分支机构。AI产品上，星环科技建立了多个产品系列：一站式大数据基础平台TDH、分布式分析型数据库ArgoDB及交易型数据库KunDB、基于容器的智能数据云平台TDC、大数据开发工具TDS、智能分析工具Sophon和超融合大数据一体机TxDataAppliance等，并拥有多项专利技术。目前公司产品已经在十几个行业应用落地，拥有超过一千家终端用户。

**·海天瑞声**

北京海天瑞声科技股份有限公司成立于2005年，是一家深耕于智能语音、计算机视觉、自然语言理解等领域的全球AI训练数据服务商。具有ISO9001质量、ISOIEC27701、ISOIEC27001管理体系认证。AI产品上，海天瑞声在语音识别、语音合成、自然语言处理、计算机视觉处理等数据集领域均有较强技术力，运用于自动驾驶、人脸姿态、虚拟主播等热门场景。同时，在严格遵循数据安全法、个人信息保护法、GDPR等法规的基础上，在全球进行190种语言、方言的采集，多场景图像、视频采集，多行业领域文本语料制作。相关产品在智能金融领域中，运用于智能营销、身份识别&金融助手、智能风控和供应链金融。

**·思必驰**

思必驰科技股份有限公司成立于2007年，是国内专业的对话式人工智能平台公司，拥有全链路的智能语音语言技术，自主研发了新一代人机交互平台(DUI)，和人工智能芯片(TH1520)。截至2021年底，思必驰拥有各类已授权知识产权900余件，其中已授权专利近400项，软件著作权近300项，并拥有中英文综合语音技术。

**·开普云**

开普云信息科技股份有限公司成立于2000年，是国内领先的基于AI大模型的行业数字化服务提供商。累计服务包括中办、国办、全国人大、中纪委、国家电网等2000多家政府、媒体和企业。在数智政务、数智内容、数智安全、数智司法等多维平台和人工智能核心技术引擎支持下，开普云发展出新一代数融平台、低代码创设平台和内容安全检测平台。在政务服务场景中积累的政务知识图谱可以纠正目前AIGC模型中普遍存在的事实性错误，提高AIGC的内容质量。

**·智洋创新**

智洋创新科技股份有限公司成立于2006年，是一家聚焦行业数字化转型的人工智能企业。公司集研发、生产、销售、服务于一体，成功应用于电力、水利、轨道交通及应急管理等业务领域。AI产品上，智洋创新在机器视觉、传输链路、边缘计算等领域均有产品。其中推出产品AI魔方，可通过AI魔方定义的开发接口，结合所在行业的功能需求进行二次开发，快速推出适用于该行业的人工智能应用。

# 免责声明

《金融科技行业信息汇编》是合肥滨湖金融小镇管理有限公司推出的专题分析类的非盈利报告。内容聚焦于国内外金融行业的热点领域——金融科技，并结合对信息的简要分析和评述，发出“滨湖金融小镇”的见解和声音。旨在服务于地方金融发展的需要，为集团公司、各子公司和相关专业人士提供参考。

《金融科技行业信息汇编》基于公开渠道和专业数据库资料搜集整理而成，但本公司对这些信息的准确性和完整性不作任何保证。信息汇编中的内容和意见仅供参考，在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议。合肥滨湖金融小镇管理有限公司不对使用《金融科技行业信息汇编》及其内容所引发的任何直接或间接损失负任何责任。



合肥滨湖金融小镇管理有限公司