

ChatGPT 开启 AI 新纪元，AIGC 投资 框架梳理

核心观点：

- **ChatGPT 指明了 NLP 生成领域的技术方向，从技术本源上加速 AIGC 发展。**我们认为，ChatGPT 对 AIGC 发展的意义在于：一方面，GPT 作为 NLP 生成领域模型的突破，将迅速解决 AI 文本生成、AI 代码生成等领域的痛点；另一方面，AI 文本生成、AI 代码生成作为 AI 音视频、游戏等其他领域的技术基础，其突破发展也将加速 AIGC 在音视频、游戏等场景中的渗透（比如生成 AI 绘画提示词，或生成调用计算引擎等）。全球巨头争相追赶，与 ChatGPT 水平尚存一定差距。国外公司中，谷歌发布的聊天机器人 Bard 具有与 ChatGPT 接近的技术水平，大约相差半年左右。国内公司中，百度、华为、字节跳动、阿里、腾讯等巨头均在大模型方向布局，整体发展水平与 ChatGPT 相差大概一到两年左右，约接近 GPT-3 的水平。
- **AIGC 多场景应用处于爆发前夜，千亿市场打开。**随着大模型算法突破以及算力成本下降，AIGC 发展痛点逐步解除，亟待爆发；同时又有 ChatGPT 助力，文本、代码、图像生成有望率先成熟，视频、游戏将紧随其后。我们预测，到 2025 年 AIGC 在网络文学领域、文本分析领域、绘画及图片领域、数字音乐领域技术将相对成熟，因此渗透率分别为 70%/60%/60%/50%；网络视频领域、游戏领域由于技术原因渗透率相对较低，分别约为 30%/25%。基于上述假设，我们分别对 AIGC 细分领域做市场规模预测，预计 2025 年中国 AIGC 市场规模有望达到 1600 亿。
- **AIGC 产业链：基础层最先受益，中间层巨头占优。**AIGC 产业链可大致分为基础层、中间层和应用层三层架构。鉴于大模型训练需要巨大的算力规模，基础层中的芯片、服务器等硬件提供商将最先受益。而对于中间层来说，技术积累以及资金实力都是关键竞争力，因此我们认为科技巨头更具有竞争优势。应用层中，可类比移动互联网时代，会爆发出很多杀手级应用，用户体验和模式创新都将成为“杀手锏”。
- **GPT-3 训练所需总成本：模型的算力需求几乎与参数量呈同比增长。**单一 V100 GPU 芯片进行一次 GPT-3 13B 模型的训练，大约需要 2144 天；进行一次 GPT-3 175B 模型的训练，大约需要 29120 天。我们测算 GPT-3 13B 训练总成本约为： $\$3.06 / H * 24 H/D * 365 D/Y * 26Y = \0.7 Million；同理，GPT-3 175B 训练总成本约为 \$9.5 Million。在不计 RAM、CPU、SSD 驱动器、电源等其他情况下，GPT-3 运行所需芯片成本至少在 20 万美元以上。
- **投资建议：**ChatGPT 开启 AI 新纪元，AIGC 千亿市场处于爆发前夜。产业链基础层中，我们重点推荐全球 AI 服务器龙头浪潮信息（000977.SZ），关注中科曙光（603019.SH）、中国长城（000066.SZ）；中间层中，推荐人工智能技术积累深厚的拓尔思（300229.SZ）、科大讯飞（002230.SZ）；应用层中，推荐彩讯股份（300634.SZ）、嘉和美康（688246.SH）、金山办公（688111.SH）、同花顺（300033.SZ），关注万兴科技（300624.SZ）。
- **风险提示：**行业竞争加剧的风险；产业发展进度不达预期的风险；技术风险；政策风险。

计算机行业

推荐 维持评级

分析师

吴砚靖

☎：（8610）66568589

✉：wuyanqing@chinastock.com.cn

执业证书编码：S0130519070001

邹文倩

☎：（8610）86359293

✉：zouwenqian@chinastock.com.cn

执业证书编码：S0130519060003

行业数据

2023-2-17



资料来源：Wind，中国银河证券研究院整理

目 录

一、ChatGPT 开启 AI 新纪元，指明 NLP 生成领域的技术方向	3
（一）ChatGPT 通过引入 RLHF 技术解决生成模型的核心问题，大大提高类人成熟度	3
（二）ChatGPT 商业化规划及仍需讨论的问题	4
（三）全球巨头争相追赶，与 ChatGPT 水平尚存一定差距	6
二、AIGC 多场景应用处于爆发前夜，千亿市场打开	6
（一）AIGC 发展痛点逐步解除，亟待爆发	6
（二）ChatGPT 助力，文本、代码、图像生成有望率先成熟，视频、游戏紧随其后	8
（三）AIGC 海外公司布局梳理	12
（四）AIGC 细分市场预测：2025 年有望百倍增长	14
三、AIGC 产业链：基础层最先受益，中间层巨头占优	16
（一）AIGC 产业链结构及巨头布局	16
（二）GPT 模型所需算力测算	18
四、A 股相关公司介绍	20
五、投资建议与风险提示	23

一、ChatGPT 开启 AI 新纪元，指明 NLP 生成领域的技术方向

ChatGPT 指明了 NLP 生成领域的技术方向。预训练语言模型对于 NLP 理解领域以及生成领域的问题都是相当有效的，在过去几年里，发展出的主要模型有 Bert 和 GPT。从文本对话的过程来看，理解到生成之间的连接是一个关键问题。然而，Bert 在 NLP 理解领域获得了巨大的成功，但是在 NLP 生成领域却表现不佳，这是由于 Bert 训练时所采用的语言模型所决定的。Bert 采用的是 DAE 语言模型，只学习到了词的上下文表征的能力，即理解语言的能力，但没有学习到如何组织语言的能力。而 chatGPT 成功解决了这个问题，可以说 ChatGPT 指明了 NLP 生成领域的技术方向。

ChatGPT 从技术本源上加速 AIGC 发展。AIGC 的底层技术主要有 NLP、GAN、扩散模型等，其中的代表就是语言生成中的 chatGPT 和图像生成中的扩散模型（Diffusion Model）。2022 年 8 月 StabilityAI 宣布开源 Stable Diffusion 模型，AI 图像生成成本迅速下降，并且能获得惊人的高质量绘画作品，达到学习几年、十几年的画师同等水平。而 2022 年 11 月 OpenAI 发布的 chatGPT 虽然是自然语言生成领域的产品，但是可以生成 AI 绘画提示词，或生成调用计算引擎，从而进一步加速 AI 绘画、AI 游戏、AI 音视频领域的应用。

我们认为，ChatGPT 对 AIGC 发展的意义在于：一方面，GPT 作为 NLP 生成领域模型的突破，将迅速解决 AI 文本生成、AI 代码生成等领域的痛点；另一方面，AI 文本生成、AI 代码生成作为 AI 音视频、游戏等其他领域的技术基础，其突破发展也将加速 AIGC 在音视频、游戏等场景中的渗透（比如生成 AI 绘画提示词，或生成调用计算引擎）。

（一）ChatGPT 通过引入 RLHF 技术解决生成模型的核心问题，大大提高类人成熟度

ChatGPT 是在 GPT 3.5 大语言模型（LLM,即 Large Language Model）的基础上，加入“基于人类反馈的强化学习（RLHF, Reinforcement Learning from Human Feedback）”来不断微调（Fine-tune）预训练语言模型，使得 LLM 模型学会理解不同类型的命令指令，并通过多重标准合理判断基于给定的 prompt 输入指令，输出的是否为优质信息（这些标准包括：富含信息、内容丰富、对用户有帮助、无害、不包含歧视信息等）。因此，ChatGPT 使得人机对话更加人性化，更富有逻辑性，大大提高了 AI 的类人成熟度。

ChatGPT 的技术发展路径：从 GPT-1 到 InstructGPT，数据量与参数量不断增加，模型精度和语言能力提升。1) 2018 年，在自然语言处理领域（NLP）刚兴起时，OpenAI 就推出了初代 GPT，它的运行逻辑是：先通过无标签数据学习生成语言模型，并能够运用于一些与有监督任务无关的 NLP 任务中。此后再根据特定的下游任务进行有监督的微调，提高其泛化能力。2) 2019 年推出的 GPT-2 扩展了网络参数和数据集，进行多任务学习，可以在数据量足够丰富且模型容量足够大时，通过训练语言模型就能够完成有监督学习的任务。3) 对比 GPT-2，2020 年推出的 GPT-3 最显著的特征是庞大的数据量和参数投入，整体训练过程耗资 1200 万美元，投入数据量达上万亿，模型参数量达到 1750 亿。GPT-3 延续了前两代 GPT 的技术架构，但改变了“大规模数据集预训练+下游数据标注微调”的方式，采用情境学习（in-context learning）来提高模型对话输出的性能。GPT-3 的规模和语言能力几乎是最强大的。它能在不做微调的情况

下，在一些传统的 NLP 任务中表现得更好，包括实现闭卷问答、模式解析、纯语言建模、机器翻译等；在新的领域，GPT-3 将 NLP 的应用扩展到缺乏足够训练数据的领域，例如在开发程序代码、文章生成和信息检索领域取得了实质性的进展。然而，GPT-3 在推理和理解能力上还有较长的路要走，在自然语言推理（NLI）任务中表现不佳。

ChatGPT 在 InstructGPT 模型上做改进，实现更好地遵循用户意图、更少的虚假信息。 OpenAI 在 2022 年 3 月发布了 InstructGPT。该语言模型通过“基于人类反馈的强化学习（RLHF）和监督学习”来提高输出质量。虽然 InstructGPT 的参数数量仅为 13 亿左右，相比于 GPT-3 缩小了 100 倍以上；但在遵循指令方面，能够更好地遵循用户意图，将有害的、不真实或者有偏差的信息输出最小化。在优化的模型上，ChatGPT 基于 InstructGPT 进一步改进，在模型结构和训练流程上遵循上述方式，但收集和标注数据的方式上发生了变化，通过采取监督学习的方式让模型学习人类排序的方式。

表 1: GPT 四代的对比

模型	发布时间	参数量	预训练数据量	预估成本	消耗资源
GPT-1	2018 年 6 月	1.17 亿	约 5GB		在 8 个 GPU 上训练一个月
GPT-2	2019 年 2 月	15 亿	40GB		在 256 个 Google Cloud TPU v3 上训练一周
GPT-3	2020 年 5 月	1750 亿	45TB	训练一次 460 万美元，总成本 1200 万美元	在 355 个 GPU 上训练一年
ChatGPT	2022 年 12 月	13-1750 亿	数十 GB	训练一次的成本高达 1 千万美元	根据参数量而定

资料来源：人民数字，品玩，中国银河证券研究院

（二）ChatGPT 商业化规划及仍需讨论的问题

ChatGPT 成为史上最快达到 1 亿月活跃用户的应用。 ChatGPT 在 1 月达到 1 亿月活跃用户，平均每天有 1300 多万访客。目前 OpenAI 已推出每月 20 美元（约合 135 人民币）的 ChatGPT 用户个人订阅计划，即使在软件的使用高峰时段，订阅用户也可以优先使用 AI 聊天机器人。

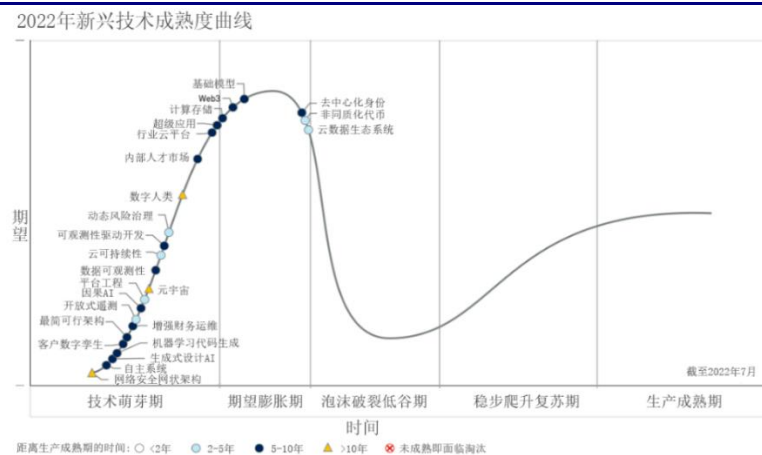
表 2: 各大软件月活数突破一亿所用时间

软件名	ChatGPT	TikTok	Instagram	Facebook	Twitter
所用时间	两个月	九个月	两年半	四年半	五年

资料来源：新浪，中国银河证券研究院

在 Gartner 发布的 2022 年新兴技术成熟度曲线中，生成性 AI 处于技术萌芽期并加速发展中。Gartner 将生成性 AI 列为 2022 年五大影响力技术之一，预测到 2025 年生成性 AI 所创造的数据可占到所有已生产数据的 10%。

图 1: Gartner 发布的 2022 年新兴技术成熟度曲线



资料来源: Gartner, 中国银河证券研究院

ChatGPT 的能力目前几乎可以涵盖各个自然语言交互领域,例如聊天机器人、对话系统、智能客服、信息检索、主题建模、文本生成和总结、NLP 作为服务的翻译、转录、总结等等,未来应用领域将面向蓝海。例如,在聊天机器人领域,目前 ChatGPT 已经能基本满足用户提供个性化需求和信息服务;在需要智能客服的电商、金融、医疗、教育、政务等垂直领域等,ChatGPT 能够结合行业特点和需求,构建自动应答系统,为客户提供快速、准确的问题解答。除此外,在传媒、娱乐、设计和影视领域,ChatGPT 能够协助完成一些较低层次的任务,包括文稿生成、采访助手、摘要总结等,或将提高行业的运行效率。

ChatGPT 存在的缺陷之一: 目前 ChatGPT 在精准性、真实性、重复率和依赖性等问题上尚待改善。第一、由于技术实现的问题,ChatGPT 会不可避免地写出一些似是而非、或者荒谬的答案,这将导致植入虚假数据和误导用户的风险。ChatGPT 依然没有完全克服大型语言模型(LLM)的这一常见缺点;第二、在较长的会话中,由于训练数据的偏差和过度修正,ChatGPT 会过度强调某些短语或者句子,导致重复性高的问题;第三、ChatGPT 的强大能力依赖语料库、数据量的抓取和复杂的训练过程。ChatGPT 模型依赖于大规模离线语料进行训练,往往不能充分接受并采用在线提供的即时信息,难以理解对话中提及的因果关系,也无法基于已有信息进行推测,这距离人类举一反三的能力相差较远。

ChatGPT 存在的缺陷之二: 其应用过程中安全性、伦理约束、权责界定以及知识产权等相关法律问题尚未解决。一方面,由于 RLFH 并不能完全避免 ChatGPT 训练库中学习到的不道德或有偏见的回答,也会导致在模糊提示或引导回答的过程中让 ChatGPT 输出一些有害信息,导致输出结果的安全性降低,有可能产生权责无法界定的法律问题。另一方面,在创造性、创作伦理和知识产权等方面并未形成有效界定。在数据挖掘、大规模计算、统计、多线程工作等数据处理分析领域,人工智能有着人类不可比拟的优势,但是以“创新和感知”为基础的创造过程是机器学习和模型难以训练的。目前 ChatGPT 能够在用户的引导下快速生成小说、诗歌、散文、编程等需要创造力的内容,或许将对创作者和以版权为基础的行业造成冲击。文本生成的过程是基于数据库内容的学习,这是否会构成对被抓取作品的侵权,ChatGPT 生成的文本内容是否具有著作权,是否属于该用户等一系列问题的答案尚不明确。

（三）全球巨头争相追赶，与 ChatGPT 水平尚存一定差距

国外公司中，谷歌发布的聊天机器人 Bard 具有与 ChatGPT 接近的技术水平，可能相差半年左右。2023 年 2 月 6 日，Google CEO 发布了与 ChatGPT 对标的产品 Bard。Bard 是基于由 Google 于两年前就开始的 LaMDA（Language Model for Dialogue Applications，用于对话应用的语言模型）模型的轻量级版本。LaMDA 和 GPT-3 本质上是非常类似的模型，都是使用 Transformer 架构的大语言模型配以 RLHF，已公开的 LaMDA 最大的版本是 1370 亿参数的模型，比 GPT-3 略少。从谷歌 22 年发布的 LaMDA 的论文来看，LaMDA 关注三个关键目标：质量、安全和凭据，从这三个关键目标来看的话，Google 相比于 OpenAI 还是谨慎了很多，这可能也是 Bard 迟迟未推出的原因。

国内公司中，百度、华为、字节跳动、阿里、腾讯等巨头均在大模型方向布局，整体发展水平与 ChatGPT 相差大概一到两年左右，约达到接近 GPT-3 的水平。其中，百度进展比较突出，百度的文心一言应用是百度基于文心大模型技术推出的生成式对话产品，将于 2023 年 3 月完成内测，面向公众开放。文心大模型的参数量已经达到 2600 亿，已经超过 GPT-3 水平。从 GPT 模型发展三要素算力、数据、模型来看，国内巨头与 ChatGPT 的主要差距体现在模型结构的设计上，由于 GPT3.5 缺少相关论文，也没有开源，国内在代码的训练及推理上仍缺少积累。

二、AIGC 多场景应用处于爆发前夜，千亿市场打开

（一）AIGC 发展痛点逐步解除，亟待爆发

AIGC 并非新兴概念，只是过往受限于优质训练数据不足、高额的算力成本或是模型算法难题，AIGC 发展较慢，仅限于实验范围。在人工智能尚不具备完整表达和生成内容能力的时期，人工智能生成内容仅能用于简单的 AI 问答和聊天机器人领域，受限于模型算法，生成内容没有创新，也较为呆板。但随着算力成本下降、数据形态和数量更加丰富，在不同领域 AIGC 都得到了飞速发展。

1. AI 文本生成领域：痛点在于上下文关联、自然语言生成的能力

文本生成领域是 AIGC 最早触及的领域。除了较早进行探索的智能客服和聊天机器人领域，早期文本生成主要应用于辅助写作或者结构化写作。此前的文本生成一直面临的巨大挑战在于：（1）上下文关联能力；（2）自然语言生成能力；（3）逻辑能力。由于传统的聊天机器人采用“基于模板”或者“基于规则”的算法，开发人员通过预设一些较为固定的规则和模板后，AI 根据用户输入的内容在已有模板中进行认知识别和匹配，然后生成回复。这种方法效率高、训练难度较低，但是训练后的模型缺乏普适性和灵活度，且难以支持复杂程度较高的会话任务，也不能进行上文记忆。因此人们在与传统聊天机器人对话时会感觉到“答非所问”、“车轮式对话”。

对于通用性内容的文本生成来说，语言流畅性、思维深度和原创性是评价文章内容的标准。人工智能输出文字的本质在于计算机语言对于人类自然语言的模拟，创造来源于人工智能基于逻辑的计算和判断，此前人工智能无法生成流畅文本的重要原因在于训练数据的质量和数

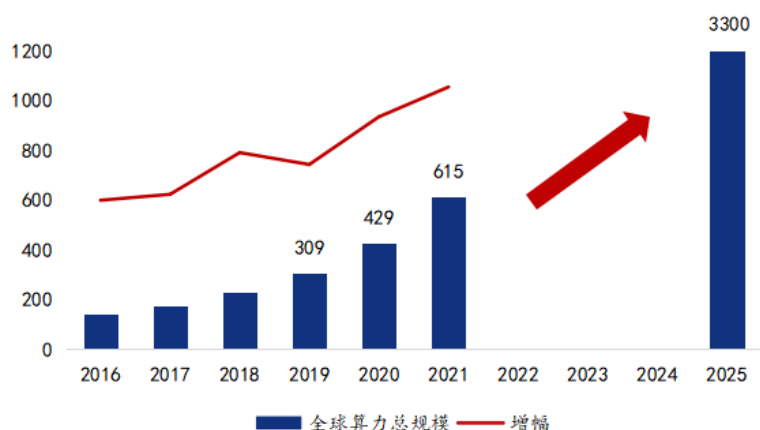
量都较小。GPT 模型的“有监督微调”和“对齐”的技术架构提高了模型对数据的理解能力，也相应提高了结果匹配精度，在语句搭配的选择上更具优势，因此生成文本的可读性有了大幅提升。

2. AI 绘画领域：痛点在于高额的算力成本

AI 学习绘画的实质过程是通过已有画作的训练数据，通过模型和参数迭代输出有规律的像素组合。由于图片所包含的信息量远高于文字包含的信息量，像素点位置、色彩、数量等共同组成庞大的参数组，因此深度学习该参数组需要足量的算力。目前 AI 绘画根据文字提示输出图片需要经历两个过程：理解文字提示所包含的信息，再匹配生成相应的图像。此前一直难以发展的原因在于：（1）绘画细节不足、分辨率低；（2）参数量过大导致渲染时间过长；（3）算力和内存不足。

此后，新的深度学习模型 CLIP（Contrastive Language-Image Pre-Training），通过优化采样的方式将完成了“文字-图片”匹配的跨越，而计算机视觉模型 Diffusion 的模型降维优化降低了对显卡的需求，加上算力成本的逐渐降低，AI 绘画应用迅速崛起。

图 2：全球算力保持快速稳定增长，平均增速超过 40%（单位：EFlops）

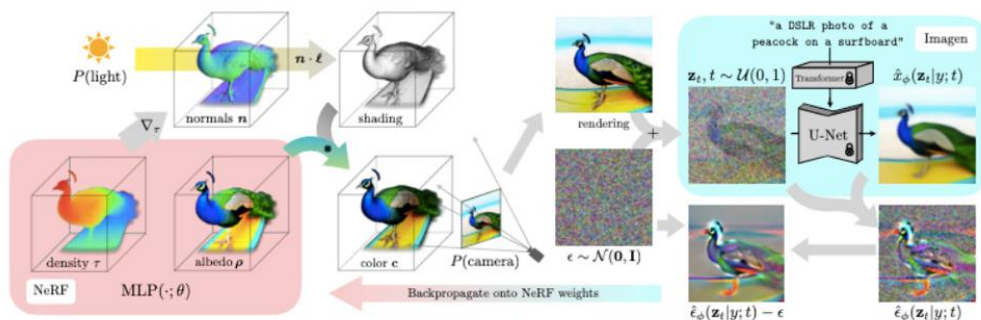


资料来源：中国信通院，中国银河证券研究院

3. AI 音视频、游戏等其他领域：需要 AI 文本生成和 AI 绘画提供技术基础辅助

在 AI 文本生成和 AI 绘画有所发展的技术基础上，音频、视频、2/3D 建模、游戏等等领域才逐步被 AI 所渗透。以音频领域为例，未来 AIGC 将在语音合成和音乐合成领域有所进展。但由于一段音频包括的数据量相较于文字和图片都更大，也更抽象，因此目前还难以找到很好的训练方式。同理，以交互和实时反馈为基础的游戏领域也是如此。游戏制作过程中涉及的每一个轮次在未来都将有相应的生成式 AI 模型发展，从 2/3D 建模、对话生成、对话语音封装等等，需要前期在文字、图像和音频领域有坚实的技术积累。

图 3: 3D 模型生成: 从文本-图像-3D 模型



直接从文本生成3D的过程所需的标注数据过多，因此Dreamfusion生成3D的过程：

先使用一个预训练2D Diffusion模型基于文字提示(prompt)生成二维图像后，再通过梯度下降法生成3D模型，并不断迭代。

资料来源: DREAMFUSION 论文, 中国银河证券研究院

(二) ChatGPT 助力，文本、代码、图像生成有望率先成熟，视频、游戏紧随其后

大模型辅以新架构，推动 AIGC 进入杀手级应用涌现阶段。在 2015 年前，人工智能发展以小模型为主导地位，能够较好地适应在垂直领域的分析任务，但是在通用交互相关的任务上，由于表达和理解能力欠缺，无法完成生成类的任务。但随着计算机深度学习模型的进展，人们开发出新型神经网络架构 transformer、扩散模型 diffusion 等新技术，同时算力成本降低，开发人员才逐渐开始探索并训练大语言模型（Large Language Model，即大模型）。

我们认为，ChatGPT 对 AIGC 发展的意义在于：一方面，GPT 作为 NLP 生成领域模型的突破，将迅速解决 AI 文本生成、AI 代码生成等领域的痛点；另一方面，AI 文本生成、AI 代码生成作为 AI 音视频、游戏等其他领域的技术基础，其突破发展也将加速 AIGC 在音视频、游戏等场景中的渗透（比如生成 AI 绘画提示词，或生成调用计算引擎）。

图 4: AIGC 在不同场景的发展进度预测

	2020年以前	2020年	2022年	预计2023年	预计2025年	预计2030年
文本领域	垃圾邮件检测 翻译 基础问答回复	基础文案撰写 生成初稿	完成长篇写作 完善初稿	对专业文本内容 垂直微调	终稿水平优于人类 平均水平	终稿水平优于专业 作家水平
代码生成	单行代码补全	多行代码生成	更长篇幅 更高的准确度	更多语言种类 更符合垂直行业	直接从文本到产品 (终稿)	实现从文本到产品 终稿，效果优于全 职开发人员成果
图像生成			美术作品 商标 摄影作品	产品设计、建筑 等实物模型	产品设计、建筑 等终稿	终稿优于专业艺术 家、设计师、摄影师
视频、3D和 游戏领域			3D、视频模型的 初次尝试	视频和3D文件的 初稿	视频和3D文件的 二稿	AI Roblox 实现电子游戏和电 影个性化

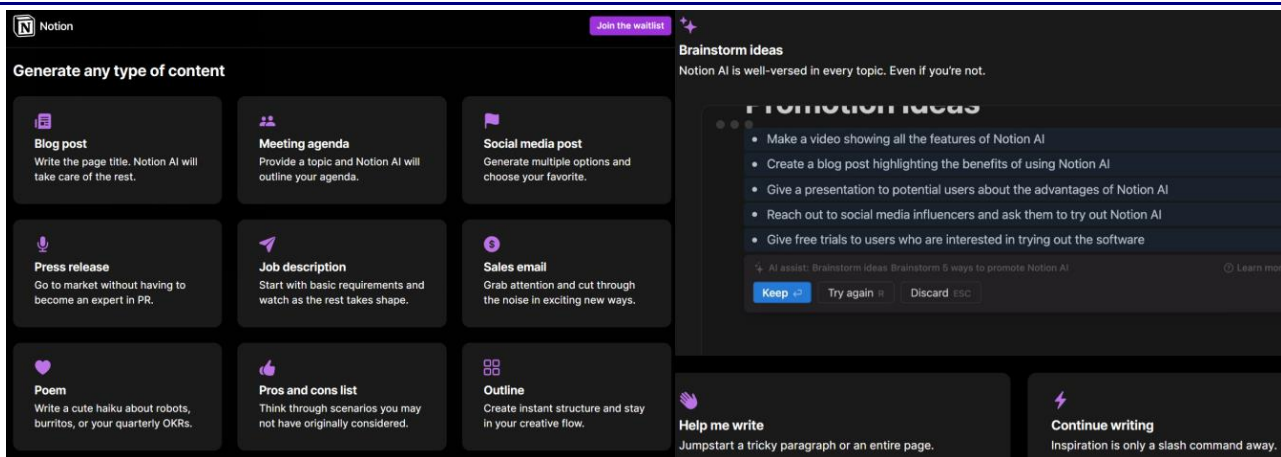
大模型应用进展：● 初次尝试 ● 基本完成 ● 准备就绪

资料来源：红杉资本官网，中国银河证券研究院

ChatGPT 成功的关键是人类反馈强化与奖励机制，其应用层的良好效果，为 AIGC 领域提供了普适性的模型优化思路。目前内容生成领域还处于用户创作（UGC）或者 AI 辅助用户创作（AIUGC）为主的阶段，但 ChatGPT 的底层技术已被逐渐迁移向以下领域：

（1）在文本生成领域，AI 在文章润色、拼写检查修改等辅助性写作领域的的能力将更加稳定，除此外，自然语言生成能力的提升赋予 AI 更好地完成结构化程度较高的文本初稿，例如电子邮件、新闻等领域，甚至在一些非结构化的文章，例如社交媒体、广告营销、文学等领域的创作也能提供帮助；理解归纳能力的提升将推动大量办公辅助类工具的落地，例如会议纪要、文档翻译、笔记管理等，提高整体办公效率。

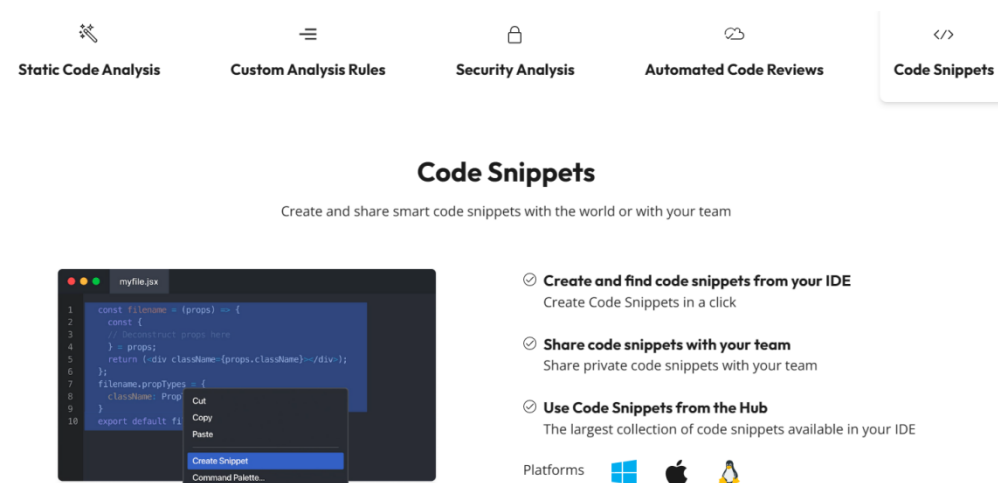
图 5：NOTION 推出较为全面的 AI 产品



资料来源：Notion 官网，中国银河证券研究院

（2）在代码开发领域，基于上下文理解能力，AI 进行代码注释、代码补全或代码检查的相关应用将有可能在短期内落地。虽然与文本生成的底层技术相类似，但是代码生成对 AI 工具的稳定性、以及自然语言和代码的相互转换的精准性提出了更高的要求。目前已有海外公司研发出稳定的 AI 辅助工具，可以实时显示代码示例并提供错误反馈。

图 6: Codiga 除静态代码分析外, 推出 AI 代码检查和漏洞自动修复, 也可进行组内代码协作



资料来源: Codiga 官网, 中国银河证券研究院

(3) 在智能问答领域, 目前 AI 文本生成的技术已相对成熟, 理解能力和表达能力的提升突破了现有客服“回答千篇一律、答非所问”的情况。未来, 从底层模型的微调和强化将使得模型分化出不同的技能树, 或将在“问题解决、服务效率和使用体验”等三个方面全面提升用户的使用体验。

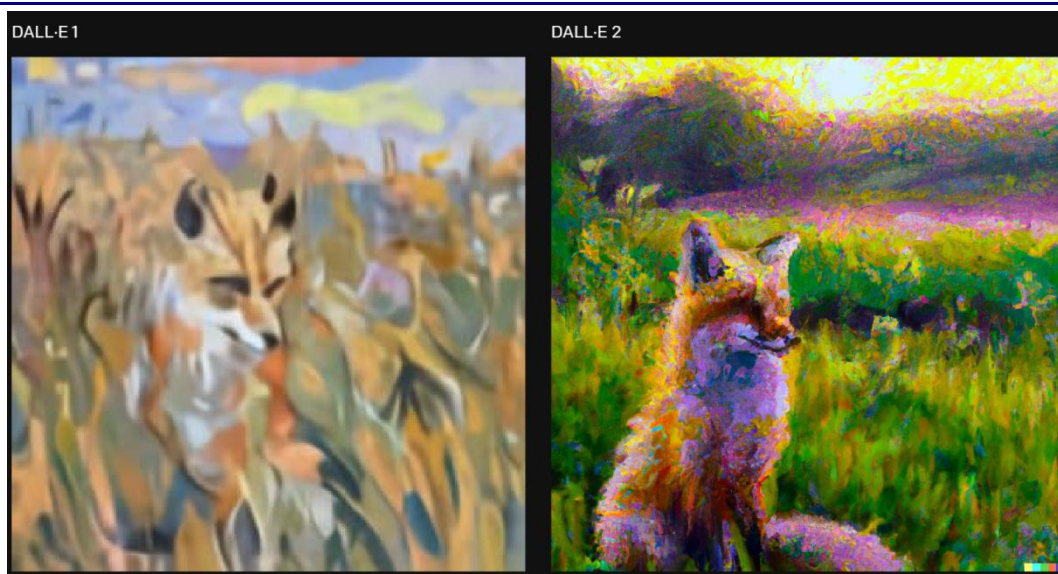
图 7: 聊天机器人 BlenderBot 的进步



资料来源: 澎湃网, 中国银河证券研究院

(4) 在图像生成领域, 目前 GPT 模型的图像生成能力相对弱于 Diffusion 模型, 但是在利用 GPT 生成文字提示的基础上, AI 可以根据扩散模型生成更优质的 AI 绘画作品。未来, 传统图片编辑能力将大幅提升, 抠图、去除特定主体等技术将更加优化, AI 按照文字提示或者示意图生成图片的能力将进一步提升。

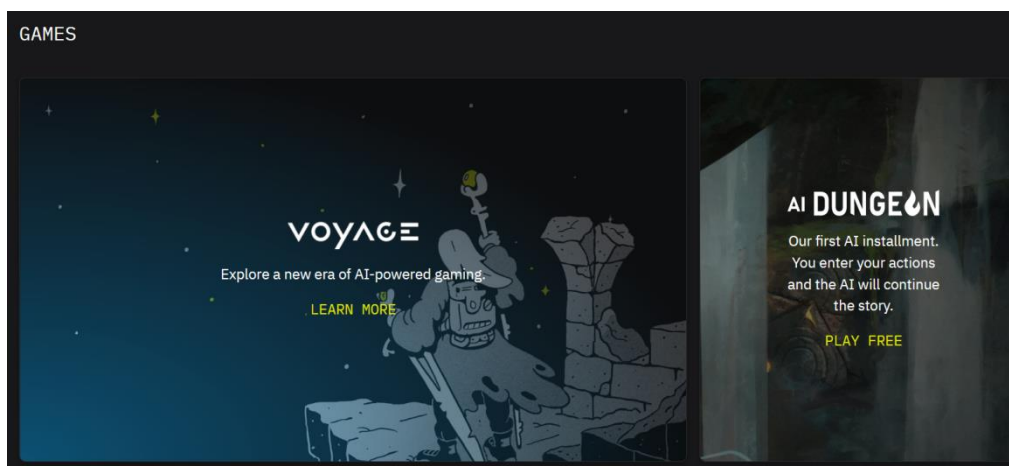
图 8: 相同算法下, DALL-E 算力提升和模型优化带来的画质不同



资料来源: DALL-E 官网, 中国银河证券研究院

(5) 在视频、游戏和 3D 领域, 所需底层技术更多, 技术能力更加交叉, 需要以目前已有模型作为基础技术实现。以视频领域为例, 视频是音频和图像的结合, 语音合成目前已有探索, 但是耗时和相似度还有所欠缺, 要实现完全自然的人类语音的合成仍然需要时间的积累。目前对 AI 视频生成的探索中, 应用场景、素材选择的限制仍然较多, 在未来音频和图像领域发展到更加成熟的阶段后, 视频生成或许将拥有更大的自由度。

图 9: AI 生成的文字冒险游戏



资料来源: Latitude 官网, 中国银河证券研究院

(三) AIGC 海外公司布局梳理

目前海外 AIGC 行业正呈现百花争艳的状态。根据 Leonis Capital 统计，截至 2022 年 12 月，海外对于生成式 AI 的风险投资从 2020 年约五亿美金增长到超过 22 亿美金，增长幅度超过 400%。目前文本和绘画领域的技术成熟度较高，音视频领域、代码生成和更为复杂的多模态生成领域目前底层技术原理已经基本具备，但还需要静候发展。

表 3: AI 文本领域部分海外公司应用汇总

AIGC 细分领域	应用场景	应用名称	公司名称	备注
AI 文本生成	开发相应模型作为 AI 文字底层协议	-	OpenAI	使用 GPT3 模型
		Copy.ai	Copy.ai	使用 GPT3 模型
		Jasper.ai	Jasper.ai	使用 GPT3 模型
	文案生成 ai 工具， 有较为全面的应用场景	OthersideAI	OthersideAI	使用 GPT3 模型
		Notion AI	Notion	
		Anyword	Anyword	
	针对社交媒体博文、邮件、营销等 应用场景的文案生成的 ai 工具	Copysmith	Copysmith	
		Ghostwriter	Replit	
		Regie.ai	Regie.ai	
		Rytr	Rytr.me	
		Smartwriter.ai	Smartwriter.ai	
		Contenda	Contenda	
		LEX	LEX	
	写作辅助 ai 工具，重写或改述生成更符合语境的文本	Quillbot	Quillbot	
		Wordtune	AI21 Labs	
AI 交互式 文本生成	聊天机器人	ChatGPT	OPENAI	使用 GPT3 模型
		Amazon Lex	Amazon	
		Bard	Google	LaMDA 对话应用语言模型
		Dasha	Dasha	使用 GPT3 模型，使用 TTS 技术
		PolyAI	PolyAI	基于 ML 的 ConveRT 技术
		Rasa	Rasa	
		Senseforth.ai	Senseforth	

资料来源：中国银河证券研究院整理

表 4: AI 绘画领域部分海外公司应用汇总

AIGC 细分领域	应用场景	应用名称	公司名称	备注
AI 绘画	AI 绘图底层协议， 推出的 stable AI 能生成图片、音视频等	Stable AI	Stability AI	开源 Stable Diffusion 模型
		DALL E	OpenAI	使用 Stable Diffusion 模型
	将描述性的文字或文字提示 自动转化为图片的 AI 绘画平台	Disco diffusion	Accomplice	使用 Stable Diffusion 模型
		Midjourney	Midjourney	使用 Stable Diffusion 模型
		Photosonic AI	Writesonic	使用 Stable Diffusion 模型

	Craiyon	Craiyon	是 DALL-E mini demo
	DeepAI	DeepAI	
	DeepDream	Google	
	Imagen	Google	
	make-a-scene	Meta	
	NightCafe	NightCafe	
	NUWA	微软	
	Parti	Google	
	Rosebud.ai	Rosebud	

资料来源：中国银河证券研究院整理

表 5: AI 代码生成领域部分海外公司应用汇总

AIGC 细分领域	应用场景	应用名称	公司名称
AI 代码生成	根据代码上下文和语法完成全功能和整行代码补全	Amazon CodeWhisperer	Amazon
		Cogram	Cogram
		GitHub Copilot	微软与 OpenAI 共同推出
		MutableAI	MutableAI
	根据上传代码自动生成程序注释	Tabnine	Tabnine
		Mintlify	Mintlify
	基于上下文进行自动代码检查和代码分析	Codiga	Codiga
		Kite	Kite

资料来源：中国银河证券研究院整理

表 6: AI 音视频领域部分海外公司应用汇总

AIGC 细分领域	应用场景	应用名称	公司名称	备注
AI 音频	由 AI 生成每周节目内容的数字播客，可进行语音合成	Podcast.ai	Podcast.ai	Play.ht 的语言模型
		Play.ht	Play.ht	peregrine 语音模型
	将文本生成语音的 ai 工具	Resemble AI	Resemble	
		WellSaid.ai	WellSaid.ai	
		Mubert API	Mubert	
AI 视频	实时视频剪辑、协作 ai 工具	Runway.AI	Runway	
		Designs.AI	Designs.AI	
		Hour One	Hour One	
		Make-A-Video	Meta	
	根据模板、已生成的虚拟形象和输入文本即可生成完整的视频	Imagen Video	Google	
		Phenaki	Google	
		Rephrase.ai	Rephrase	
		Synthesia	Synthesia	
		Wombo.AI	WOMBO	

拼凑式视频生成 ai,	Gliacloud	GliaCloud
输入文本后自动生成个性化的视频内容,	Tavus	Tavus
区分其中文字和标题		

资料来源：中国银河证券研究院整理

表 7: AI 游戏领域部分海外公司应用汇总

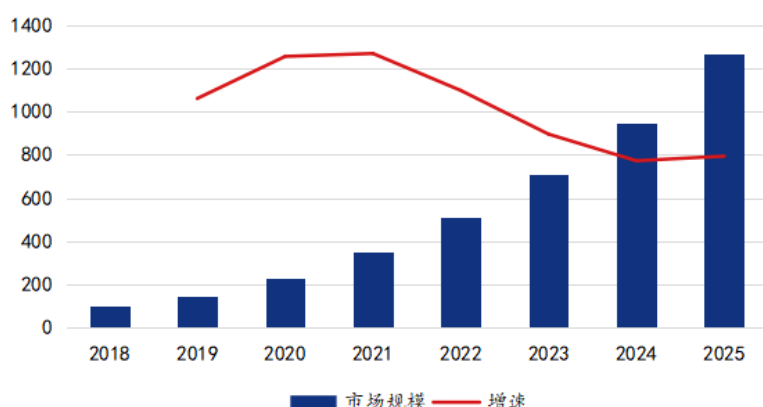
AIGC 细分领域	应用场景	应用名称	公司名称
AI 游戏领域	Ai 文字冒险游戏, 根据玩家交互生成故事	AI Dungeon	Latitude
	文本、ai 角色和语音	Charisma.ai	Charisma
	以 AI 创建的角色为特色的扮演游戏	Arrowmancer	Spellbrush

资料来源：中国银河证券研究院整理

(四) AIGC 细分市场规模预测：2025 年有望百倍增长

目前, AIGC 所覆盖的领域基本属于 AI 软件与服务。根据 Statista 的预测数据, 全球 AI 软件的市场规模将在 2025 年超过 1200 亿美元, 2021 年到 2025 年的 CAGR 为 38.01%。在 AIGC 发展的推动下, 未来人工智能市场规模将以更快的速度增长。

图 10: 全球 AI 软件的市场规模 (单位: 亿美元)

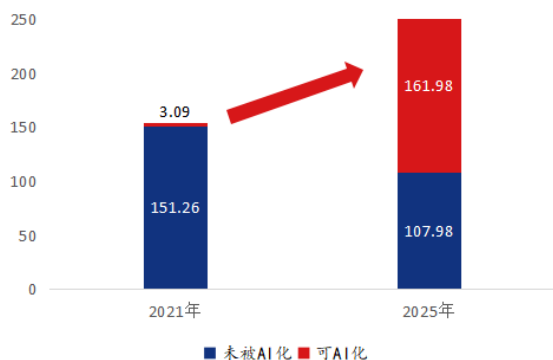


资料来源: Statista, 中国银河证券研究院

虽然目前 AIGC 已经涉及文字、图像、代码、语音、视频、3D、游戏等多种内容形式的生成, 但是每个领域的技术成熟度、底层技术难度均有差距, 因此未来 AIGC 的渗透率将有所差别。2021 年 AI 制作的比例仅占市场的 1%-2% 左右, 随着技术发展, 我们预测, 到 2025 年 AIGC 在网络文学领域、文本分析领域、绘画及图片领域、数字音乐领域技术将相对成熟, 因此渗透率分别为 70%/60%/60%/50%, 网络视频领域、游戏领域由于技术原因渗透率相对较低, 分别约为 30%/25%。

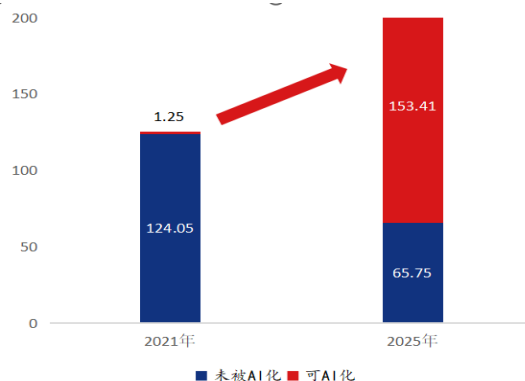
基于上述假设, 我们分别对 AIGC 细分领域针对中国市场做 2025 年的市场规模预测, 预计 2025 年中国 AIGC 市场规模有望达到 1600 亿。

图 11: 文本分析领域的市场规模 (单位: 亿元)



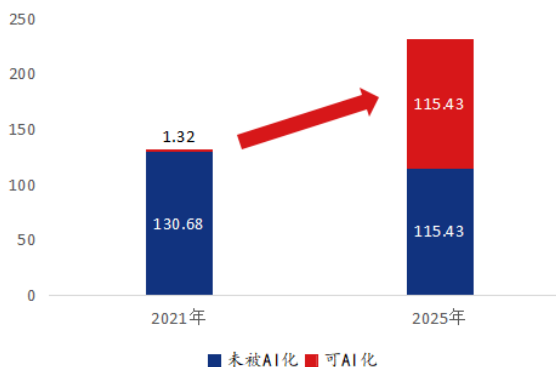
资料来源: IMARC Services, 中国银河证券研究院

图 12: 网络文学领域的市场规模 (单位: 亿元)



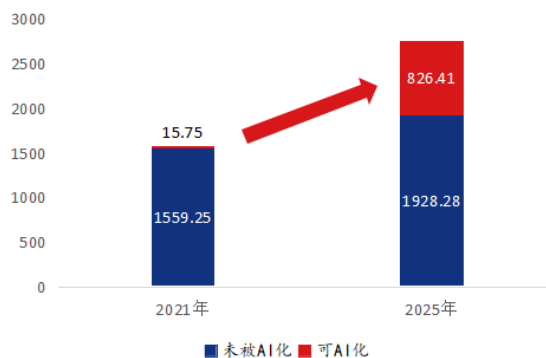
资料来源: 华经产业研究院, 中国银河证券研究院

图 13: 数字音乐领域的市场规模 (单位: 亿元)



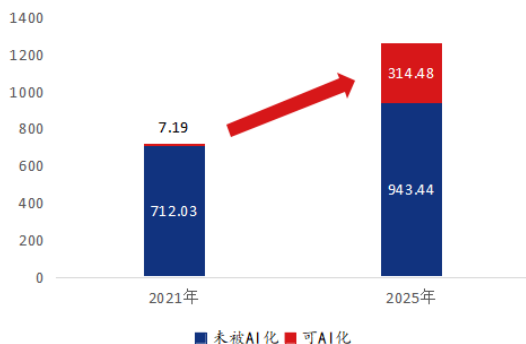
资料来源: 艾媒咨询, 中国银河证券研究院

图 14: 网络视频领域的市场规模 (单位: 亿元)



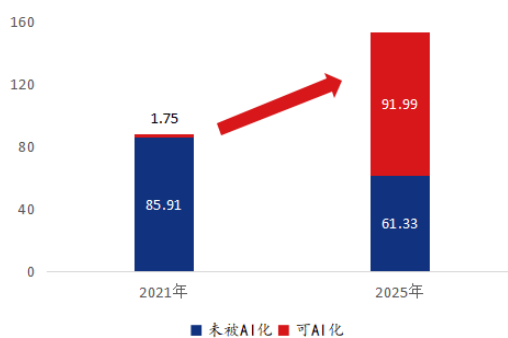
资料来源: 前瞻产业研究院, CNNIC, 国银河证券研究院

图 15: 游戏领域的市场规模 (单位: 亿元)



资料来源: 国音数协游戏工委, 伽马数据, 中国银河证券研究院

图 16: 绘画及图片领域的市场规模 (单位: 亿元)



资料来源: 中国产业研究院, 国银河证券研究院

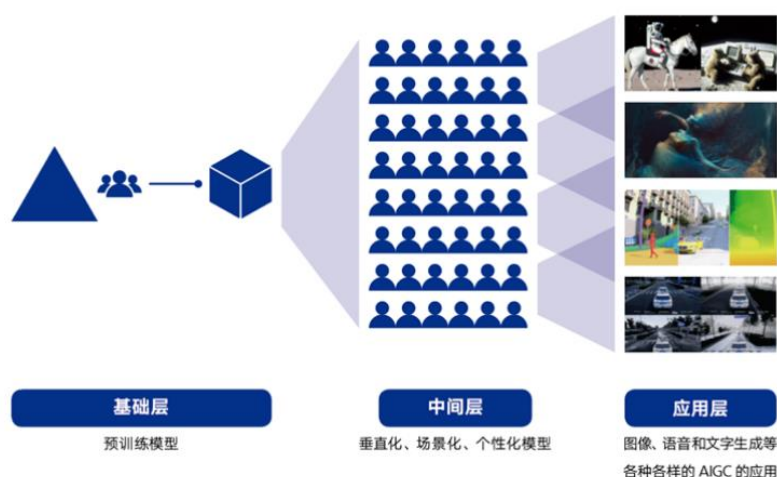
三、AIGC 产业链：基础层最先受益，中间层巨头占优

（一）AIGC 产业链结构及巨头布局

AIGC 产业链可大致分为上中下三层架构。**第一层为基础层**，以预训练模型为基础搭建的 AIGC 技术基础设施层，包括芯片、服务器等硬件提供商以及大模型技术提供商。**第二层为中间层**，通过基础层的模型，进行特定训练，形成垂直化、场景化、定制化的应用工具。**第三层为应用层**，即面向 C 端用户提供图像、语音、文字生成等各种 AIGC 的应用。

鉴于大模型训练需要巨大的算力规模，基础层中的芯片、服务器等硬件提供商将最先受益。而对于中间层来说，技术积累以及资金实力都是关键竞争力，因此我们认为科技巨头更具有竞争优势。应用层中，可类比移动互联网时代，会爆发出很多杀手级应用，用户体验和模式创新都将成为“杀手锏”。

图 17: AIGC 产业链



资料来源：腾讯研究院、中国银河证券研究院

表 8: AIGC 产业链巨头布局

产业链	公司	布局
基础层及中间层	英伟达	推出的深度学习框架、计算机视觉库、自然语言处理库等工具可以帮助开发者更快更好地开发出更先进的 AI 应用。此外，英伟达还推出了一系列的 AI 开发服务，包括 AI 开发咨询、AI 开发培训、AI 开发实施等，以帮助开发者更好地掌握 AI 开发技术。去年年底，英伟达发布了 Magic3D，根据文本描述自动生成 3D 模型，正式宣布加入 AIGC 赛道。
	高通	研发了一系列解决方案，包括基于深度学习的自然语言处理（NLP）、计算机视觉（CV）和机器学习（ML）等技术。这些解决方案可以帮助开发者更快地开发出更加智能的应用程序，从而提高用户体验。公司最新的 Snapdragon 865 处理器支持 AIGC，并且支持多种 AI 框架，包括 TensorFlow、Caffe2、PyTorch 等。此外，高通还推出了 Snapdragon Neural Processing Engine（NPE），它可以支持多种 AI 模型，并且可以提供更快的 AI 性能。

中间层及应用层	AMD	在 AIGC 方面取得了巨大的进展，其 Radeon Instinct 系列 GPU 已经成为 AI 计算的首选。它们拥有更高的计算性能，更低的功耗，更快的训练时间，以及更多的 AI 模型支持，今年年初，代号为 Instinct MI300 的芯片，也是 AMD 首款数据中心/HPC 级的 APU 正式发布，号称能将 ChatGPT、DALL·E 等大模型的训练时间从几个月缩短至几周。此外，公司还推出了 MIVision AI 开发套件，为 AI 开发者提供了一个完整的解决方案，以支持他们的 AI 应用程序。此外，公司还推出了一款名为“ROCm”的开源 AI 开发平台，可以帮助开发人员更轻松地创建和部署 AI 应用程序。
	微软	近日公司推出了由 ChatGPT 支持的最新版本 Bing（必应）搜索引擎和 Edge 浏览器，新版本的 Bing（搜索）会采取类似 ChatGPT 的方式，回答具有大量上下文的问题，此外，必应中的新聊天机器人可以帮助进行起草和翻译邮件、完善查询功能、重写代码等。再早几日，微软宣布旗下所有产品将全线整合 ChatGPT，其中包括了 Bing 搜索引擎、Word、PPT、Excel 等 Office 软件、Azure 云服务、Teams 聊天程序等，预计在 3 月，ChatGPT 将内置于 Bing 搜索。
	谷歌	公司于近日公布了智能聊天机器人 Bard，这是一款基于应用程序语言模型(LaMDA)的实验性人工智能程序，其主要利用来自网络的广泛的知识与信息，与大型语言模型的强大创造力相结合，输出高质量的答案，公司 CEO 表示该项技术将很快集成到谷歌的搜索引擎中。在去年年底，谷歌还一次性发布了多项 AIGC 技术成果，包括文本图像模型 Imagen 以及 AI 写作协作工具 LaMDA Wordcraft 等。
	百度	拥有强大的 AI 技术团队以及大量的研究成果，掌握着深度学习、自然语言处理、计算机视觉等技术，不断推出新的产品和服务，以满足用户的需求。到目前百度所推出的 AIGC 产品中，产业级搜索系统“文心百中”与 AI 作画平台“文心一格”均使用了百度所开发的大模型 ERNIE，据百度官方表公布，百度正在开发类似 chatGPT 的相关技术与产品，并将其命名为“文心一言”，英文名为 ERNIE Bot，预计在今年三月份完成内测后向公众开放。
	Meta	公司旗下人工智能实验室 Meta AI，发布了从文本生成视频的 AI 系统“Make-A-Video”，即根据输入的自然语言文本生成一段 5 秒钟左右的短视频，并在此基础上，拓展到从图像生成视频和从视频生成视频。此外，近日公司发表的一篇论文引起了广泛关注，该论文介绍了一种会使用工具的语言模型 Toolformer，也即是在生成文本的过程中，遇到特定的任务，Toolformer 会直接调用所需工具的 API 来解决面前的任务。
	阿里	推出了基于深度学习的自动文本生成技术，基于自然语言处理的自动文档生成技术，基于机器学习的自动图像生成技术，以及基于深度学习的自动视频生成技术，这些都属于 AIGC 相关技术。近年来，阿里研发了通义-AliceMind 语言大模型，可以实现 AI 阅读理解。此外，公司还推出了会议 AI 助理“听悟”、智能对话机器人“小蜜”等产品，阿里版 ChatGPT 目前正在研发中，最终命名尚未确定，据曝光信息来看，阿里很可能将 AI 大模型技术与钉钉生产力工具深度结合。
	腾讯	腾讯混元 AI 大模型团队推出了万亿级别中文 NLP 预训练模型 HunYuan-NLP-1T，目前该模型已落地于腾讯广告、搜索、对话等产品，例如在广告制作上，可以利用人工智能将广告文案生成广告视频，可以大幅降低广告视频的制作成本。此外，腾讯还推出了“梦幻写手”，可以协助用户生成稿件，腾讯 AI 开放平台，可以为客户提供语音识别、知识图谱等工具。对于 ChatGPT 与 AIGC 相关方面的研发，公司表示将基于此前的技术储备，有序进行各项研究。
	字节跳动	正力求开发出更加智能的内容生成系统，可以根据用户的需求和偏好，自动生成高质量的内容。此外，研究人员还在开发更加智能的自然语言处理技术，可以更好地理解用户的意图，从而提供更加准确的内容生成结果。目前，字节跳动的人工智能实验室（AI Lab）正在开展类似 ChatGPT 和 AIGC 的相关研发，据预计，其研发的相关产品，将主要集中在 VR 内容生成上。

资料来源：新浪、网易、IT之家，中国银河证券研究院

(二) GPT 模型所需算力测算

GPT-3 训练所需算力：

根据 OPENAI 公布的论文附录，GPT-3 175B 共有 1746 亿个参数，训练该模型需要经过 3.14×10^{23} 个浮点运算，总算力为 3640 PF-days。换言之，在每秒进行 1 万亿次浮点运算的计算速度下，GPT-3 175B 需要耗时 364 天。

图 18：三类语言模型训练所用的总算力、参数(Params)、训练数据量(Tokens)等

Model	Total train compute (PF-days)	Total train compute (flops)	Params (M)	Training tokens (billions)	Flops per param per token	Mult for bwd pass	Fwd-pass flops per active param per token	Frac of params active for each token
T5-Small	2.08E+00	1.80E+20	60	1,000	3	3	1	0.5
T5-Base	7.64E+00	6.60E+20	220	1,000	3	3	1	0.5
T5-Large	2.67E+01	2.31E+21	770	1,000	3	3	1	0.5
T5-3B	1.04E+02	9.00E+21	3,000	1,000	3	3	1	0.5
T5-11B	3.82E+02	3.30E+22	11,000	1,000	3	3	1	0.5
BERT-Base	1.89E+00	1.64E+20	109	250	6	3	2	1.0
BERT-Large	6.16E+00	5.33E+20	355	250	6	3	2	1.0
RoBERTa-Base	1.74E+01	1.50E+21	125	2,000	6	3	2	1.0
RoBERTa-Large	4.93E+01	4.26E+21	355	2,000	6	3	2	1.0
GPT-3 Small	2.60E+00	2.25E+20	125	300	6	3	2	1.0
GPT-3 Medium	7.42E+00	6.41E+20	356	300	6	3	2	1.0
GPT-3 Large	1.58E+01	1.37E+21	760	300	6	3	2	1.0
GPT-3 XL	2.75E+01	2.38E+21	1,320	300	6	3	2	1.0
GPT-3 2.7B	5.52E+01	4.77E+21	2,650	300	6	3	2	1.0
GPT-3 6.7B	1.39E+02	1.20E+22	6,660	300	6	3	2	1.0
GPT-3 13B	2.68E+02	2.31E+22	12,850	300	6	3	2	1.0
GPT-3 175B	3.64E+03	3.14E+23	174,600	300	6	3	2	1.0

资料来源：Language Models are Few-Shot Learners (2020)，OPENAI，中国银河证券研究院

论文中表明，GPT-3 的模型训练是在微软提供的 V100 GPU 上完成的。由于微软与 OPENAI 并未公布人才合作共建的超级计算机的硬件细节，未知其 V100 的运算成本效率，因此 V100 GPU 型号细节参考 NVIDIA 官网公布的 TESLA V100。其采用 NVIDIA Volta 架构，带有 16GB 和 32GB 两种配置，单个 GPU 即可提供高达 100 个 CPU 的性能，训练吞吐量比使用 CPU 快 32 倍，推理吞吐量比使用 CPU 快 24 倍，配备 V100 GPU 的单台服务器可以取代数百台仅配备通用 CPU 的服务器来处理传统的高性能计算和人工智能工作负载。

图 19: Tesla V100 的参数

NVLink 版本 V100		PCIe 版本 V100
NVIDIA GPU 加速 带来高性能	双精度	双精度
	7.8 TeraFLOPS	7 TeraFLOPS
	单精度	单精度
	15.7 TeraFLOPS	14 TeraFLOPS
	深度学习	深度学习
	125 TeraFLOPS	112 TeraFLOPS
双向互联带宽		PCIe
	NVLink	32 GB/s
	300 GB/s	
GPU 内存 显存带宽		容量
		32/16 GB HBM2
		带宽
		900 GB/s

资料来源: NVIDIA, 中国银河证券研究院

GPT-3 训练所需总成本:

模型的算力需求几乎与参数量呈同比增长。单一 V100 GPU 芯片进行一次 GPT-3 13B 模型的训练, 大约需要 2144 天; 进行一次 GPT-3 175B 模型的训练, 大约需要 29120 天。由于 NVIDIA 官网公布的最快的 V100 深度学习性能达到 125 TFlops, 按照 GPT-3 13B 模型的算力需求 268 PF-days, 单个 V100 GPU 需要进行: $268/0.125=2144$ 天的运算; 同理, 完成一次 GPT-3 175B 运算的算力需求为 3640 PF-days, 单个 V100 GPU 需要进行 29120 天的计算。GPT-3 13B 发展到 GPT-3 175B, 参数量扩大 13.73 倍, 算力需求扩大 13.58 倍, 两者几乎同比增长。

若以一个容量为 28 Tflops 的 V100 GPU 为例, 完成 GPT-3 13B 和 GPT-3 175B 模型理论上分别需要 26 个 GPU 年和 355 个 GPU 年的运算。根据 Cloud GPU Comparison 公布的数据, Microsoft Azure 的 V100 型号使用成本为 \$3.06/H, 我们测算 GPT-3 13B 训练总成本约为: $\$3.06/H * 24 H/D * 365 D/Y * 26Y = \0.7 Million ; 同理, GPT-3 175B 训练总成本约为 \$9.5 Million。

GPT-3 运行所需芯片成本:

除此外, 根据 OpenAI 的白皮书, GPT-3 使用每个参数 16 位的半精度浮点变量, 这说明 1750 亿参数至少需要有 $175*4=700$ GB 的空间才能存储 (每个参数需要 4 个字节)。若以此计算, 相当于至少 22 块 32GB 内存的 Tesla V100 GPU, 每块约 9000 美元, 则不计 RAM、CPU、SSD 驱动器、电源等其他情况下, 芯片成本在至少 20 万美元以上。

四、A 股相关公司介绍

1、浪潮信息——全球 AI 服务器龙头

技术发展：浪潮信息人工智能服务器的市占率位居全球第一，在中国市场份额连续五年保持 50% 以上。公司行业竞争力强劲，致力于十四五期间成为全球服务器龙头公司。根据 IDC 数据显示，2022 年，公司在全球服务器行业市占率排名第三，市占率分别为 9.9%；在国内服务器行业，公司更是当之无愧的龙头厂商，2020 年市占率 35.6%，高于排名第二第三的华为和新华三总和。

业务布局：2 月 10 日，浪潮信息推出 AI 算力服务产品，将基于国内领先智算中心的算力基础设施，为国内客户提供先进、开放、高性能、高质量的 AI 算力资源和配套服务，支撑 AI 模型构建、训练和推理的业务全流程，赋能生成式 AI 产业创新。目前，AI 算力服务相关信息已在元脑生态 AISTore 平台上线，包括有多种算力、数据、算法等可选产品。浪潮信息此次推出的 AI 算力服务产品，包括支持国际和国内领先的 4 种 AI 算力产品、经“源”清洗验证过的语言或多模态的大模型数据集的 2 种数据产品、由浪潮“源”专家团队提供的语言或多模态的大模型训练支持的 2 种算法产品以及 AI 训练推理资源管理调度的 1 种平台产品。

2、彩讯股份——信创+国资云+AIGC 三驾马车

技术发展：公司在智能问答、图像识别、语音语义识别等产品方向均有所涉猎，此外，公司还在电信、金融等行业的智能文件识别、智能客服机器人系统、5G 消息 chatbot 系统等方面都有着深厚的技术储备，其中一部分成果已经实现落地应用。与 AI 的深度融合，可以让彩讯邮件系统更加智能地实现邮件摘要自动生成，邮件撰写智能等功能。5G 消息未来有机会成为 ChatGPT 入口。彩讯 5G 消息 CSP 解决方案可定制接入客户自有 chatbot 或提供垂直行业 chatbot 一体化解决方案，通过 5G 消息号“变身”为智能客服助手，提供更智能的交互体验。

业务布局：公司力求将云计算、人工智能以及大数据等新兴技术应用到传统产业链，促进各行各业的数字化转型升级。智慧渠道、协同办公、云和大数据为公司的三大业务线，主要为行业客户提供产业互联网应用平台开发和运营服务。目前是信创产业发展的风口，公司在此趋势下积极布局，自研的彩讯 Richmail（Richmail 信创安全增强电子邮件系统解决方案）未来有望在党政、金融等领域多点发力。公司作为中国移动供货商，云业务深度受益移动云业务加速，预计维持较高增速。

3、嘉和美康——电子病历龙头，AI 智慧医疗先行者

技术发展：公司在国内医疗大数据和人工智能方面具有丰富的技术与经验储备，智慧医疗类产品在全国市场排名领先。据 IDC 报告，公司在 2021 年中国医疗大数据解决方案市场中排名第二。凭借深厚的临床信息化建设经验和上千家医院核心信息系统的数据处理优势，公司对数据处理技术进行二次创新转化，自主研发了智能医学数据中台，对大规模多源异构医疗数据进行深度处理和分析，对不同维度数据进行聚合、关联，形成适用于不同场景的数据模型，打造出拥有自主知识产权的覆盖智能诊前服务、临床决策支持、病历内涵质控及大数据科研等

不同应用场景的医疗 AI 系统，实现了数据驱动下的患者服务、临床辅助、科研反哺、质控护航的多位一体医疗服务模式，完成了对“诊前-诊中-诊后”全量医疗数据的全流程闭环应用。

业务布局：临床辅助决策支持系统（CDSS）、大数据科研分析平台、AI 病历内涵质控系统、智能预问诊系统、智能分诊系统是公司智慧医疗产品体系的五大核心产品。公司智慧医疗产品体系以人工智能技术为核心驱动，以医院内真实诊疗数据为基础，以能够提供高度集约化、标准化数据整合、治理及服务的智能医学数据中台为支撑，形成的覆盖临床诊疗、科研支持、医务管理、患者服务等多个应用场景的智慧医疗全生态产品矩阵，能够实现“诊前-诊中-诊后”智慧医疗服务闭环，为医疗机构、科研院校、卫生主管部门等提供数字化、智能化、精细化的智慧医疗创新服务。

智慧医疗产品在智慧医院建设中比重不断增加，随着智慧医院建设要求不断提高，以大数据和人工智能等新兴技术为基础的智慧医疗产品，场景丰富度和成熟度得到根本提升，在医疗过程中发挥的作用不断凸显。公司智慧医疗产品也成为主营业务中新的增长点，2022 上半年，公司智慧医疗产品增速明显，软件产品销售收入 3,107.72 万元，同比增长 128.09%。

4、科大讯飞——人工智能行业龙头企业

技术发展：AI 语音技术为公司之本，公司在 AI 语音市场份额第一，由此延展至人工智能多个领域。目前公司已进入 AI2.0 战略时代，处于 AI 技术的红利兑现期，目标是实现 AI 应用规模化落地。在此时期内，公司将在于教育、消费者、开放平台、智慧城市、医疗、汽车等重点行业规模化推广人工智能解决方案。科大讯飞主导建设了认知智能全国重点实验室，使得自身核心技术始终保持在世界前沿水平，并且面向认知智能领域前后共开源了 6 大类、超过 40 个通用领域的系列中文预训练语言模型，是业界最为流行的中文预训练模型系列之一。

业务布局：公司在 2019 年启动 AI2.0 战略，以“开放平台+赛道”的模式，在智慧教育、消费者业务、智慧医疗、智慧城市等应用场景布局。在平台上，公司依靠讯飞人工智能开放平台，为客户提供 AI 解决方案，围绕自身平台构建产业生态；在赛道上，公司为各行各业提供产品服务以及解决方案，主要有智慧教育、智慧医疗、智慧城市、消费者业务、智慧汽车、智慧金融等。

5、拓尔思——专注于 NLP(语义智能)领域的行业领军者

技术发展：在人工智能领域，拓尔思是国内最早研发自然语言处理（NLP）的公司之一，在知识图谱、NLP、OCR 等领域都有着自主可控的底层技术，也是行业领军者。此外，公司的语义智能技术还包括了知识图谱、智能搜索等领域，并且在相关领域处于领先地位。如今公司将以预训练大模型、In-Context Learning、instruction tuning 等技术为基础，把文本生成、交互式生成、跨模态生成、小样本学习等功能作为研发重点，力求在问答式 AI、智能创作、搜索引擎等领域实现成功应用。

业务布局：公司先前主要业务为向各级党政机关和企事业单位等提供人工智能和大数据

工具软件平台以及解决方案，现在正逐步从提供软件产品向提供云平台服务转型升级，以软件产品+云服务的商业模式成为目前公司的核心竞争力所在。

6、金山办公——推动 OA 与 AIGC 高效融合的办公软件龙头

技术发展：公司在办公软件具备 30 余年的技术积累，且具有办公领域专业研究的优势，对核心技术具有自主知识产权。目前关键技术研究重点针对数据云存储、智能数据处理、安全云文档以及数据协同共享等多个领域。由于办公领域与 AIGC 天然的结合点，能大幅提高工作效率，实现“软件产品向价值工具”的转变。目前公司在研项目包括智能协作创新和产业化公共服务平台、智能办公软件集成等，目前已完成国产软硬件深度适配基本技术的研究实现、产品架构设计及第一阶段功能定制开发等工作。

业务布局：公司一直探索人工智能和大数据分析技术的发展，并应用于用户文档创作和在线协同中，以提供“多屏、云、内容、AI、协作”的办公生活方式，依托 AI 构建产品云服务体系。2022 年 11 月，金山办公与华为签订战略合作备忘录，进一步在鸿蒙生态领域展开深度合作，逐步将 WPS 及其他旗下产品适配华为 1+8+N 设备，以期实现设备协同和生态融合的理念，未来将推动“跨设备、跨用户、跨服务”的办公形态。

7、同花顺——AI+互联网金融信息服务提供商

技术发展：公司在人工智能技术方面具有丰富的经验积累，2009 年落地的“i 问财”部门是最早为股票信息服务平台，逐步完善智能问答功能，并从金融垂直领域扩展到多领域生成答案。2015 年公司自主研发的自然语言系统中，语音识别表现处于第一梯队，此后上线的语音系产品包括智能外呼、智能客服、语音质检等，其智能客服机器人已应用于多家大中型证券公司及基金公司。同花顺长期保持高研发投入，不断加大对机器学习、自然语言理解、语音识别、人机交互、数字人等关键技术研发，正在推出更多 ALL IN AI 的产品。

业务布局：公司从 2009 年开始布局人工智能领域，先后推出自主研发的股票信息检索服务、智能问答、语音识别等技术，目前已上线 AI 开放平台，可面向客户提供数字虚拟人、智能金融问答、智能语音、智能客服机器人、智能质检机器人、会议转写系统、智慧政务平台、智能医疗辅助系统等 40 余项人工智能产品及服务，应用于证券、基金、银行、保险、运营商、高校、企业、科研院所、政府部门等多个行业。目前公司已在金融领域以外推出智能电梯产品、AI 医疗内镜助手等，未来将推动 AI 和更多垂直领域的结合。

8、万兴科技——全球领先的新生代数字创意赋能者

技术发展：公司正积极探索 3D 模型设计与渲染技术、动捕 AI 技术和基于 GAN 网络形象生成等技术。截至目前，公司已在多个产品中实现虚拟人、文生图、AI 智能抠像、AI 智能降噪、AI 音乐重组、AI 换脸等功能。

业务布局：公司已布局虚拟数字人、虚拟场景、虚拟直播等业务，并于近日在全球元宇宙

大会论坛上宣布将布局 AIGC 赛道,同时,公司旗下首款 AI 绘画软件“万兴爱画”也已经正式开启公测。未来,公司将在进一步挖掘消费类软件海外市场的同时,大力开拓国内市场,提高自身在国内市场的占有率,因为目前中国 IT 市场正处于蓬勃发展时期,公司需要抓住这一机遇,并且中国是全球重要市场,实现全球化自然离不开开拓国内市场。

9、云从科技——提供人机协同操作系统和行业解决方案的人工智能企业

技术发展: 依靠在人脸识别领域的技术积累,云从科技首先从人脸识别切入 AI 行业,之后相继在图像识别、语音识别、自然语言处理、机器学习等 AI 领域取得突破和成果,公司有着自主研发的跨镜追踪(ReID)、3D 结构光人脸识别、双层异构深度神经网络和对抗性神经网络技术等人工智能技术,均处于业界领先水平。截至目前,公司已经基本形成了人工智能感知、认知、决策的技术闭环,全链自主可控。

业务布局: 公司从人脸识别技术出发,通过感知、认知、决策的全链人工智能技术,将人工智能解决方案在智慧金融、智慧治理、智慧出行、智慧商业等行业进行实践。此外,公司以自身研发的通用操作系统为基础,实现人工智能技术在特定场景下的应用,如火眼人脸大数据、活体检测软件等产品,最终形成软硬件相结合的人工智能解决方案。

五、投资建议与风险提示

ChatGPT 开启 AI 新纪元, AIGC 千亿市场处于爆发前夜。产业链基础层中,我们重点推荐全球 AI 服务器龙头浪潮信息(000977.SZ), 关注中科曙光(603019.SH)、中国长城(000066.SZ); 中间层中, 推荐人工智能技术积累深厚的拓尔思(300229.SZ)、科大讯飞(002230.SZ); 应用层中, 推荐彩讯股份(300634.SZ)、嘉和美康(688246.SH)、金山办公(688111.SH)、同花顺(300033.SZ), 关注万兴科技(300624.SZ)。

风险提示:

行业竞争加剧的风险; 产业发展进度不达预期的风险; 技术风险; 政策风险。

插图目录

图 1: Gartner 发布的 2022 年新兴技术成熟度曲线	5
图 2: 全球算力保持快速稳定增长, 平均增速超过 40% (单位: EFlops)	7
图 3: 3D 模型生成: 从文本-图像-3D 模型	8
图 4: AIGC 在不同场景的发展进度预测	8
图 5: NOTION 推出较为全面的 AI 产品	9
图 6: Codiga 除静态代码分析外, 推出 AI 代码检查和漏洞自动修复, 也可进行组内代码协作	10
图 7: 聊天机器人 BlenderBot 的进步	10
图 8: 相同算法下, DALL-E 算力提升和模型优化带来的画质不同	11
图 9: AI 生成的文字冒险游戏	11
图 10: 全球 AI 软件的市场规模 (单位: 亿美元)	14
图 11: 文本分析领域的市场规模 (单位: 亿元)	15
图 12: 其他文本生成领域的市场规模 (单位: 亿元)	15
图 13: 互联网音频领域的市场规模 (单位: 亿元)	15
图 14: 网络视频领域的市场规模 (单位: 亿元)	15
图 15: 游戏领域的市场规模 (单位: 亿元)	15
图 16: 图片领域的市场规模 (单位: 亿元)	15
图 17: AIGC 产业链	16
图 18: 三类语言模型训练所用的总算力、参数(Params)、训练数据量(Tokens)等	18
图 19: Tesla V100 的参数	19

表 格 目 录

表 1: GPT 四代的对比	4
表 2: 各大软件月活数突破一亿所用时间	4
表 3: AI 文本领域部分海外公司应用汇总	12
表 4: AI 绘画领域部分海外公司应用汇总	12
表 5: AI 代码生成领域部分海外公司应用汇总	13
表 6: AI 音视频领域部分海外公司应用汇总	13
表 7: AI 游戏领域部分海外公司应用汇总	14
表 8: AIGC 产业链巨头布局	16

分析师承诺及简介

本人承诺，以勤勉的执业态度，独立、客观地出具本报告，本报告清晰准确地反映本人的研究观点。本人薪酬的任何部分过去不曾与、现在不与、未来也将不会与本报告的具体推荐或观点直接或间接相关。

吴砚靖 TMT/科创板研究负责人

北京大学软件项目管理硕士，10 年证券分析从业经验，历任中银国际证券首席分析师，国内大型知名 PE 机构研究部执行总经理。具备一二级市场经验，长期专注科技公司研究。

邹文倩 计算机/科创板团队分析师

复旦大学金融硕士，复旦大学理学学士；2016 年加入中国银河证券研究院；2016 年新财富入围团队成员。

评级标准

行业评级体系

未来 6-12 个月，行业指数（或分析师团队所覆盖公司组成的行业指数）相对于基准指数（交易所指数或市场中主要的指数）

推荐：行业指数超越基准指数平均回报 20% 及以上。

谨慎推荐：行业指数超越基准指数平均回报。

中性：行业指数与基准指数平均回报相当。

回避：行业指数低于基准指数平均回报 10% 及以上。

公司评级体系

推荐：指未来 6-12 个月，公司股价超越分析师（或分析师团队）所覆盖股票平均回报 20% 及以上。

谨慎推荐：指未来 6-12 个月，公司股价超越分析师（或分析师团队）所覆盖股票平均回报 10% - 20%。

中性：指未来 6-12 个月，公司股价与分析师（或分析师团队）所覆盖股票平均回报相当。

回避：指未来 6-12 个月，公司股价低于分析师（或分析师团队）所覆盖股票平均回报 10% 及以上。

免责声明

报告由中国银河证券股份有限公司（以下简称银河证券）向其客户提供。银河证券无需因接收人收到本报告而视其为客户。若您并非银河证券客户中的专业投资者，为保证服务质量、控制投资风险，应首先联系银河证券机构销售部门或客户经理，完成投资者适当性匹配，并充分了解该项服务的性质、特点、使用的注意事项以及若不当使用可能带来的风险或损失。

本报告所载的全部内容只提供给客户做参考之用，并不构成对客户的投资咨询建议，并非作为买卖、认购证券或其它金融工具的邀请或保证。客户不应单纯依靠本报告而取代自我独立判断。银河证券认为本报告资料来源是可靠的，所载内容及观点客观公正，但不担保其准确性或完整性。本报告所载内容反映的是银河证券在最初发表本报告日期当日的判断，银河证券可发出其它与本报告所载内容不一致或有不同结论的报告，但银河证券没有义务和责任去及时更新本报告涉及的内容并通知客户。银河证券不对因客户使用本报告而导致的损失负任何责任。

本报告可能附带其它网站的地址或超级链接，对于可能涉及的银河证券网站以外的地址或超级链接，银河证券不对其内容负责。链接网站的内容不构成本报告的任何部分，客户需自行承担浏览这些网站的费用或风险。

银河证券在法律允许的情况下可参与、投资或持有本报告涉及的证券或进行证券交易，或向本报告涉及的公司提供或争取提供包括投资银行业务在内的服务或业务支持。银河证券可能与本报告涉及的公司之间存在业务关系，并无需事先或在获得业务关系后通知客户。

银河证券已具备中国证监会批复的证券投资咨询业务资格。除非另有说明，所有本报告的版权属于银河证券。未经银河证券书面授权许可，任何机构或个人不得以任何形式转发、转载、翻版或传播本报告。特提醒公众投资者慎重使用未经授权刊载或者转发的本公司证券研究报告。

本报告版权归银河证券所有并保留最终解释权。

联系

中国银河证券股份有限公司 研究院

深圳市福田区金田路 3088 号中洲大厦 20 层

上海浦东新区富城路 99 号震旦大厦 31 层

北京市丰台区西营街 8 号院 1 号楼青海金融大厦

公司网址：www.chinastock.com.cn

机构请致电：

深广地区：苏一耘 0755-83479312 suyiyun_yj@chinastock.com.cn

程曦 0755-83471683 chengxi_yj@chinastock.com.cn

上海地区：何婷婷 021-20252612 hetingting@chinastock.com.cn

陆韵如 021-60387901 luyunru_yj@chinastock.com.cn

北京地区：唐嫚玲 010-80927722 tangmanling_bj@chinastock.com.cn