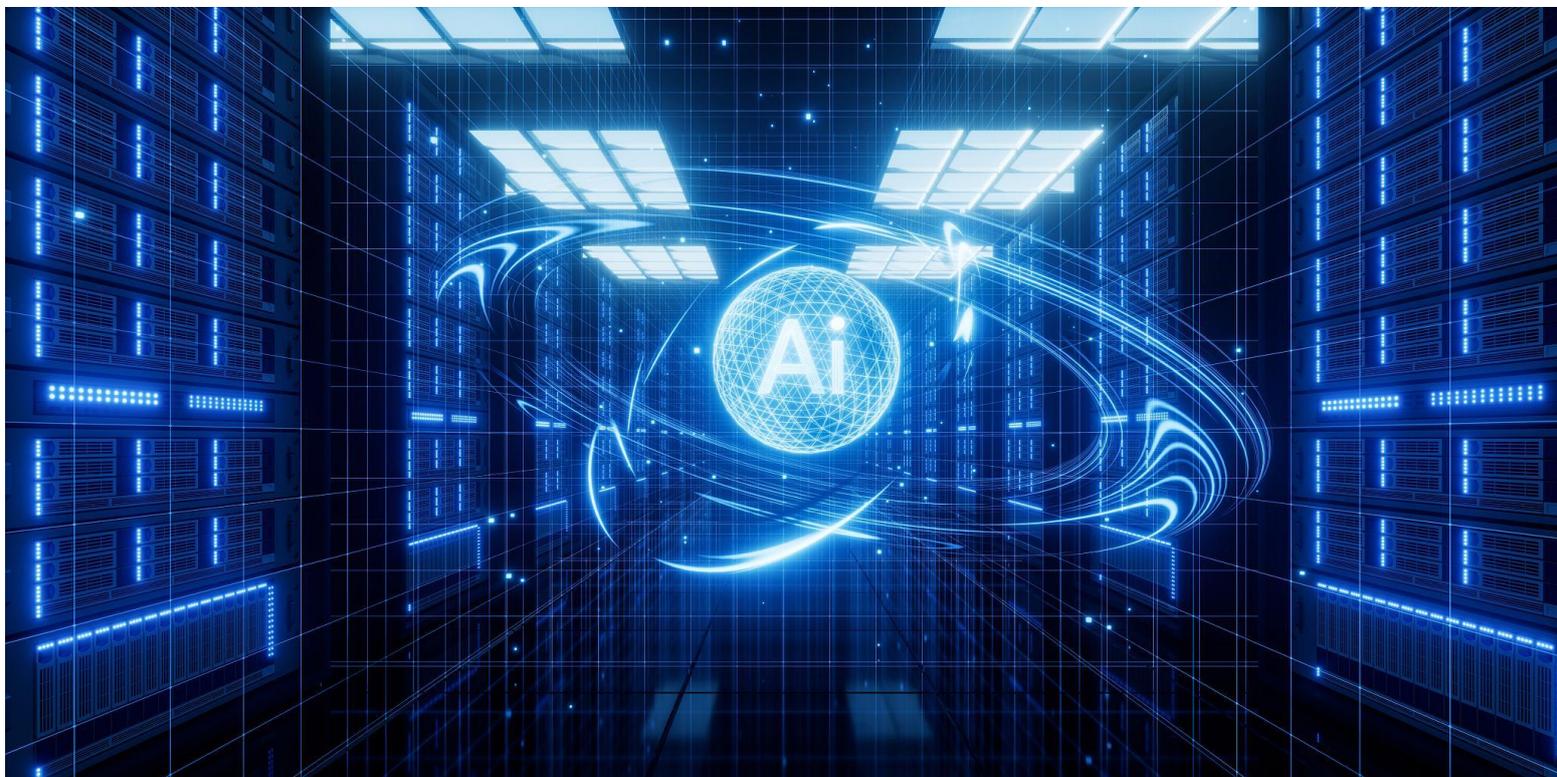


中国经济高质量发展系列研究

数字经济：大国经济体系下，人工智能领航数字 经济新阶段



中国银河证券首席经济学家，研究院院长：章俊

宏观组分析师：许冬石、詹璐

科技组分析师：吴砚靖、邹文倩、李璐昕

研究助理：胡天昊、聂天奇

数字经济：大国经济体系下，人工智能领航数字经济新阶段

核心观点：

- **人工智能开启第四次工业革命，融合技术带来的实际 GDP 增速未来 7 年有望超过 7%。**人类正在进入第四次工业革命—智能化革命，以人工智能、公有区块链、多组学测序、储能和机器人为代表的五类颠覆性技术正在融合，带来生产力的爆发。根据 ARK 研究显示，全球范围内，融合技术带来的实际 GDP 增速未来 7 年有望达到 7% 以上，而过去 125 年的平均增速只有 3%。AI 促进颠覆性技术融合，对经济增长的贡献突出，引进 AI 后实际 GDP 在 2023 至 2030 年间有望累积增长 130%。国内 Kimi 智能助手访问量持续飙升，加速 AI 应用进程。国外 Sora 多模态大模型的表现优异，以及训练成本有望以每年 75% 的速度下降，意味着 AGI 有望加速到来。
- **大国经济财政货币支撑体系为“适度超前”建设算力体系提供背书，人工智能作为核心技术催化剂领航数字经济新阶段。**大国体系下的中国数字经济产业已全面开启，我们预计在 2035 年可达到 GDP 的 71.6%。大国经济财政货币支撑体系为“适度超前”建设算力体系提供背书。央行通过结构性货币政策工具提供间接融资支持，并提升直接融资占比，构建覆盖科技型企业全生命周期的金融服务体系。而新一轮财政逆周期调控不同于以往几次扩张，更加聚焦经济转型和科技创新。
- **中美两国人工智能大模型合计数量占全球 90% 以上，在全球具有绝对优势。**从中美数字经济及人工智能产业要素发展对比来看，目前中国在数据方面具有明显的大国优势；算力方面，中美两国的算力与其他梯队国家相比有显著优势，但美国仍领先于中国；算法方面，国内领军大模型在中文领域的平均水平已经接近 GPT-4。
- **算力、算法、数据三要素共振，数据跨境流动趋严，彰显大国体系数据优势。**算力供需缺口持续加大，海外高端 AI 芯片持续短缺，国产华为生态份额将不断提升。多模态大模型 C 端、B 端加速渗透，2024 年为端侧 AI 元年。数据跨境流动已成为全球资金、信息、技术、人才、货物等资源要素交换、共享的基础。数据跨境流动趋严，高质量数据成为全球人工智能竞争的“胜负手”，大国经济支持数据要素价值释放。行业应用加速渗透，AIGC+金融、教育、办公、医疗、工业有望率先落地。

投资建议：

- 我们认为，人工智能正在推动我国数字经济发展进入快车道，大模型及多模态加速迭代升级，国内外应用端加速落地，进一步带动上游算力基础设施迎来高速成长期，积极看好本轮人工智能驱动的数字经济全产业链投资机会。建议关注：1) 智能算力基础设施厂商；2) 大模型及多模态厂商；3) 数据资源持有商及运营商；4) AI+金融、AI+教育、AI+医疗、AI+办公、AI+工业等赛道龙头公司。

分析师

首席经济学家：章俊 S013052307000

宏观分析师：许冬石 S0130515030003

宏观分析师：詹璐 S0130522110001

首席计算机分析师：吴砚靖 S0130519070001

计算机分析师：邹文倩 S0130519060003

计算机分析师：李璐昕 S0130521040001

研究助理：

胡天昊、聂天奇

风险提示

1. 技术研发进度不及预期风险；
2. 供应链风险；
3. 政策推进不及预期风险；
4. 下游需求不及预期风险；
5. 行业竞争加剧风险。

目 录

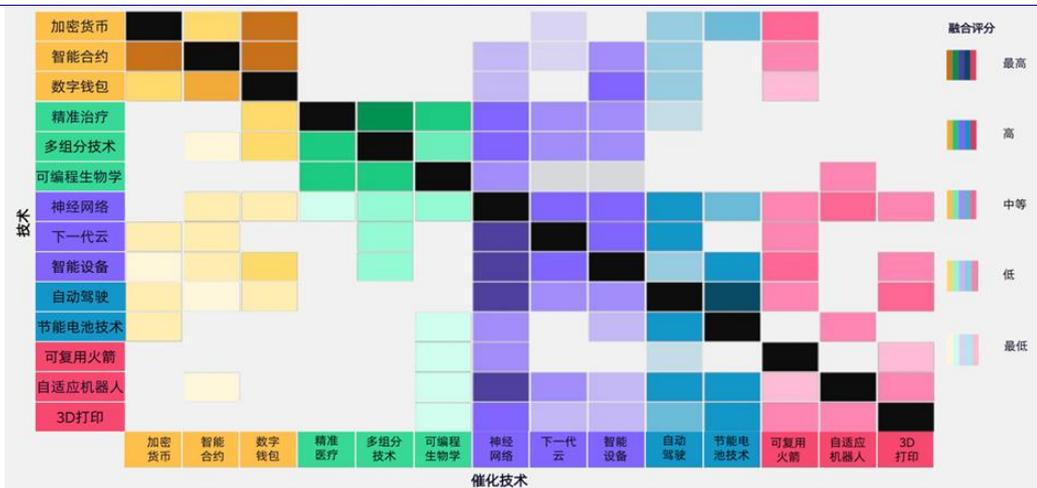
一、未来已来，人工智能正式开启第四次工业革命	4
（一）第四次工业革命概述，融合技术正在掀起历史性的技术浪潮	4
（二）Sora 发布标志 AGI 系统有望超预期提前到来	5
（三）国内 Kimi 加速迭代，AI 应用元年百花齐放	8
（四）AIGC 到 AGI 将带来哪些颠覆式革命	8
二、人工智能是核心技术催化剂，领航数字经济新阶段	12
（一）“适度超前”建设算力体系背后的财政货币支撑体系	12
（二）大国体系下的中国数字经济产业已全面开启，预计在 2035 年达到 GDP 的 71.6%	18
（三）中美数字经济及人工智能产业要素发展对比	19
三、人工智能三要素共振，算力、算法、数据未来趋势推演	22
（一）算力：供需缺口加大，AI 服务器产业链分析	22
（二）算法：多模态大模型 C 端、B 端加速渗透，未来从云端走向终端	36
（三）数据：跨境流动监管趋严，大国经济支持数据要素价值释放	41
四、AIGC 驱动千行百业加速裂变，行业应用加速渗透	53
（一）AIGC+金融：国内金融大模型抢滩布局，场景应用加速渗透	53
（二）AIGC+教育：大模型重构教育生态	56
（三）AIGC+医疗：AI 药物研发和 AI 医学影像深具潜力	57
（四）AIGC+办公：ARPU 值与付费率双升	63
（五）AIGC+工业：打造新质生产力，推动智能制造发展	65
五、投资建议	68
六、风险提示	69

一、未来已来，人工智能正式开启第四次工业革命

（一）第四次工业革命概述，融合技术正在掀起历史性的技术浪潮

人工智能是核心技术催化剂，颠覆性技术之间的融合产生协同效应，带来生产力的爆发，成为第四次工业革命的主要推动力。人类文明已经历了三次工业革命，第一次是 18 世纪中叶以蒸汽机为代表的机械化革命，第二次是 19 世纪中叶以电力、内燃机为代表的电气化革命，第三次是 20 世纪中叶以信息技术为代表的自动化革命。当前我们正在进入第四次工业革命—智能化革命，以人工智能（神经网络/下一代云/智能设备）、公有区块链（加密货币/智能合约/数字钱包）、多组学测序（精准治疗/多组分技术/可编程生物学）、储能（自动驾驶/节能电池）和机器人（可复用火箭/自适应机器人/3D 打印）为代表的五类颠覆性技术正在融合，融合产生的协同效应将发挥更大作用，推动生产力的发展。其中，人工智能是核心技术催化剂，与其他四类技术的融合范围最广、评分最高。

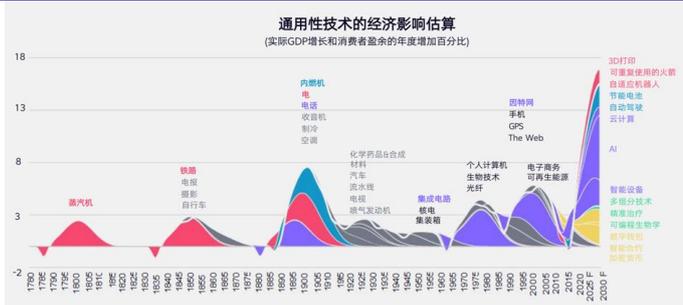
图1：人工智能是核心技术催化剂，促进颠覆性技术融合



资料来源：ARK 投资，中国银河证券研究院

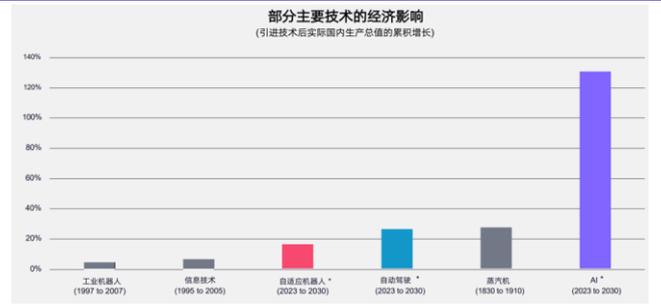
融合技术带来的实际 GDP 增速将远超第一次和第二次工业革命，AI 对经济增长的贡献突出。据 ARK 估算，在全球范围内，未来 7 年的实际 GDP 增速将达到 7% 以上，而过去 125 年的平均增速只有 3%。AI 作为核心技术催化剂，对经济增长的贡献突出。据 ARK 估算，引进 AI 后，实际 GDP 在 2023 至 2030 年间有望累积增长 130%。原因是在 AI 的赋能下，一些行业的生产效率和成本发生了巨大的变化。比如：机器人与 AI 融合后，可以在非结构化环境中低成本高效地工作，2030 年有望带来 24 万亿美元的经济效益；自动驾驶出租车与生成式 AI 融合提升了安全性，到 2030 年有望广泛应用而使每英里成本低至 0.25 美分，创造一个 11 万亿美元的潜在市场；而 AI 软件直接提高了知识型工作者的生产力，2030 年有望提升生产力至 2.5 倍，若软件价值量按 10% 计算，则有望产生 13 万亿美元的经济效益。

图2：四次工业革命里通用性技术的经济影响估算



资料来源：ARK 投资，中国银河证券研究院

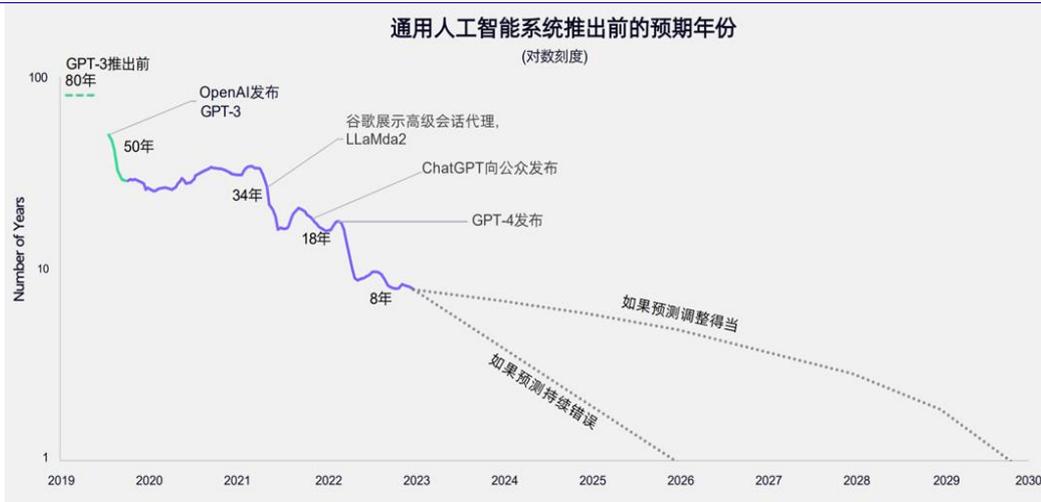
图3：AI 对经济增长的贡献突出



资料来源：ARK 投资，中国银河证券研究院

AGI 时代有望加速到来。近期，英伟达创始人 CEO 黄仁勋和谷歌 DeepMind CEO 哈萨比斯对 AGI 的到来时间进行了预测，他们的观点相似。哈萨比斯认为，AGI 最早可能在 2030 年出现，而黄仁勋则认为通用人工智能可能在五年内实现。据 ARK 分析，根据赖特定律，加速计算硬件的改进将使 AI 相对计算单元 (RCU) 的生产成本每年降低 53%，而算法模型的增强可以进一步带来每年 47% 的训练成本下降。换言之，到 2030 年，硬件和软件的融合可以使人工智能训练成本以每年 75% 的速度下降。人工智能模型的训练成本下降将进一步加速其能力的迭代，AGI 有望加速到来。

图4：AGI 有望加速到来



资料来源：ARK 投资，中国银河证券研究院

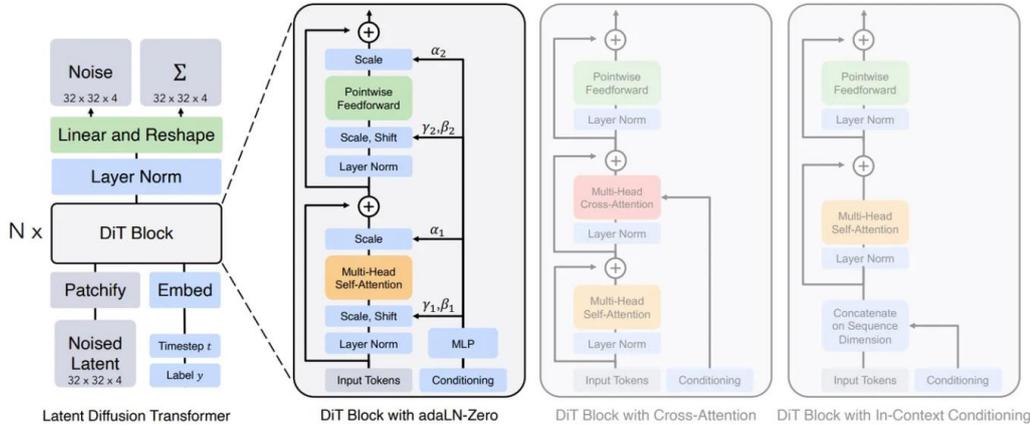
（二）Sora 发布标志 AGI 系统有望超预期提前到来

1. DiTs 算法赋能 AIGC，Sora 开启文生视频新纪元

北京时间 2 月 16 日，OpenAI 发布“文生视频”大模型 Sora。可生成一分钟的高保真视频，并配有 48 个生成案例及技术报告，能够通过自然语言指令生成长达 60 秒的高清流畅视频，在生成视频长度、清晰度、连贯性、多镜头切换方面都有显著提升。官方发布的技术报告指出，视频生成模型将是构建“世界通用模拟器”的重要途径。

本质上，Sora 基于 Diffusion Transformers (DiTs) 构建，并使用 DALL-E3 的重捕获技术。研究表明，DiTs 相较于传统语义分割网络架构 (U-Net) 在模型大小上更具可扩展性，并能生成更高质量的 2D 及 3D 图像。

图5: DiT 模块架构



资料来源: 《Scalable Diffusion Models with Transformers》(William Peebles, Saining Xie), 中国银河证券研究院

OpenAI 曾采用循环网络、生成对抗网络、自回归 Transformer 和扩散模型等方法对视频数据进行生成建模，但生成视频存在视觉数据受限、视频时长较短或视频尺寸局限等问题。通过从大语言模型 (LLM) 中汲取灵感，类似 GPT 将自然语言转换为文本 tokens，Sora 可将视觉数据转换为 patches(视觉编码块)。通过此种方式，即可实现在不同类型的视频和图像上训练生成模型。

通过视频压缩网络可降低视觉数据的维度，该网络将原始视频作为输入，并在时间和空间上进行压缩，Sora 在压缩空间中接受训练并生成视频；另外，相应的解码器模型可将生成的对象映射回像素空间。Sora 基于 patches 对可变分辨率、持续时间和纵横比的视频和图像进行训练；亦可以通过在适当大小的网格中排列随机初始化的 patches 来控制生成视频的大小。

图6: 将视频压缩到一个较低维空间，再将其分解为 patches

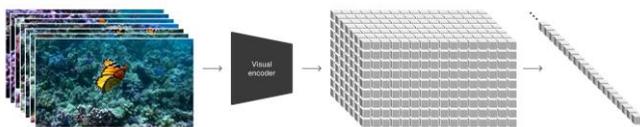


图7: 随着训练计算的增加，样本质量显著提高



资料来源: OpenAI, 中国银河证券研究院

资料来源: OpenAI, 中国银河证券研究院

2. Sora 功能多样化，生成质量远超同类大模型

Sora 可以将简短文本描述转换成一分钟流畅视频，相对于 Runway、Pika、

Stable Video 等同类大模型提升了几个代级。1) **生成视频长度**：Runway、Pika 等传统文生视频大模型平均时长在 3-5 秒，Runway 用户可以最多延长视频长度至 16 秒，Sora 生成视频长度相对传统视频生成工具提升 15-20 倍；2) **视频质量显著提升**：Sora 生成视频分辨率可达 1920 x 1080；3) **可实现多镜头切换**：可以理解 and 模拟运动中的物理规律，可以实现复杂的运动相机模拟；4) **视频连贯性与稳定性出色**：在建模能力上表现更好，可以依赖关系进行建模，能初步理解并模拟物理运动规律；5) **高可拓展性**：支持多种数据格式输入，具备实现文生视频、图生视频、向前或向后视频扩展能力，同时支持视频连接。

表1: OpenAI Sora 对比同类文生视频大模型

模型	Sora	Gen-2	Pika
研发团队	OpenAI	Runway	PIKA Labs
发布时间	2024.02	2023.03	2023.11
生成时长	60s	4-16s	3~15s
可输入格式	文字/图片/视频	文字/图片	文字/图片/视频
输入限制	-	320 字符/单张图片	500 字符/单张图片/单个视频
每秒帧数	-	24fps	8~24fps
分辨率	1920 x 1080	1408 x 768 (基础)	1280 x 720 (基础)
		2816 x 1536 (高级)	2560 x 1440 (高级)
付费模式	-	订阅制/单次付费	订阅制
免费试用	-	限制次数	不限次数

资料来源: OpenAI, Runway, PIKALabs, JournalBharat, Medium, 中国银河证券研究院

3. Sora 发布标志 AGI 时代正加速到来，带动训练及推理算力需求增长

由于目前 Sora 处在初级阶段，训练数据集和参数规模有限，仍存在一些不足之处。对于 Sora 当前存在的弱点，OpenAI 指出它可能难以准确模拟复杂场景的物理原理，并且可能无法理解因果关系。该模型还可能混淆提示的空间细节，例如混淆左右；或可能难以精确描述随着时间推移发生的事件，例如遵循特定的相机轨迹。

图8: Sora 缺陷其一：生成的运动违反物理原理



Prompt: Step-printing scene of a person running, cinematic film shot in 35mm. 0:11 / 0:19

Weakness: Sora sometimes creates physically implausible motion.

资料来源: OpenAI, 中国银河证券研究院

图9: Sora 缺陷其二：物理建模不精确、目标“变形”不自然



Prompt: Basketball through hoop then explodes. 0:05 / 0:08

Weakness: An example of inaccurate physical modeling and unnatural object "morphing."

资料来源: OpenAI, 中国银河证券研究院

AIGC 领域难度排序由易到难依次是文本、图像、音频、视频。未来训练数据集将会数以万倍的增长，模型参数量也将不断提升，目前来看 Sora 训练所需算力不及 GPT-4 等大语言模型，伴随 Sora 大模型不断迭代调优、训练数据集规模逐渐扩大，我们认为，Sora 发布标志着 AGI 时代正加速到来：1) 短期来看，模型迭代优化、训练数据集增大将快速带动训练端算力需求；2) 长期来看，Sora 技术逐渐成熟带动下游 AI 应用百花齐放，包括为影视制作、游戏开发、教育培训、广告及社交等场景应用带来颠覆，推理端需求将厚积薄发。

（三）国内 Kimi 加速迭代，AI 应用元年百花齐放

北京时间 3 月 18 日，Moonshot（月之暗面）宣布在大模型上下文窗口技术上取得新的突破，其自研的 Kimi 智能助手已支持 200 万字超长无损上下文，并已开启产品内测。

Kimi 智能助手在去年 10 月发布，支持 20 万汉字无损级别上下文输入，是当时 AI 消费级产品支持上下文文本长度记录保持者。我们认为，Kimi 智能助手迭代速度超预期，推动应用端加速落地，Kimi 智能助手目前已经支持 200 万文字超长无损上下文，对比目前主流大模型：1) 谷歌近期发布的 Gemini 1.5 pro 支持 100 万 token 输入；2) Claude 3 支持 20 万 token 输入；3) GPT-4 Turbo 支持 12.8 万 token 输入。我们依旧坚定年初观点，2024 年将是 AI 应用元年，Kimi 智能助手宣布大模型进入“长文本时代”，长文本能力也将是通往 AGI 进程中的关键之一，Kimi 智能助手将是又一里程碑。

Kimi 智能助手支持多种应用场景，生成速度提升 3 倍之多。Kimi 智能助手去年 10 月发布的版本仅支持 20 万上下文输入，时隔 3 个月，Moonshot 为 Kimi 智能助手提供了更多数据源，本次迭代升级使 Kimi 基于出色的长上下文处理能力帮助用户解锁更多应用场景，比如专业学术论文的翻译和理解、辅助分析法律问题、一次性整理几十张发票、快速理解 API 开发文档、快速筛选符合条件的简历等。当面对一个问题时，Kimi 智能助手会尝试不同的方向搜索并据此做出回答。在回答速度上也有提升，Moonshot 工程副总裁表示，基于 Infra 层的优化，Kimi 智能助手生成速度较去年 10 月份提升了 3 倍。

Kimi 智能助手访问量持续飙升，加速 AI 应用元年进程。根据 SimilarWeb 数据显示，去年 12 月 Kimi 的周访问量还在 10 万次上下，到了 1 月下旬才突破 40 万，但是从春节开始访问量疾速攀升，至今周访问量已经超过 160 万次，2 月访问量增长 107.6%，仅次于百度文心一言与阿里通义千问（访问量均下降超 30%）。我们认为，2024AI 应用元年有两个条件：1) 大模型达到可使用状态：这点从 Kimi 用户的如潮好评中可以看出。2) 大模型公众可触达：目前 Kimi 已经面向全社会开放使用。

（四）AIGC 到 AGI 将带来哪些颠覆式革命

通用人工智能（AGI）是指具备类似于人类思考能力，能够适应广泛领域并解决多种问题的机器智能，是人工智能研究的重要目标之一。而狭义人工智能则

指已取得显著进展但局限于特定领域的人工智能，例如语音识别、机器视觉等。目前我们处于狭义人工智能相对成熟、通用人工智能乍现的阶段，GPT-4 等大语言模型及 Sora 等多模态模型被认为是通向通用人工智能的重要潜在路径，并且这一进程在逐渐加速。

“超长文本”和“超强模拟物理运动”能力将是 AGI 时代关键。我们认为大模型时代将会沿着两条路线继续演绎，一是支持超长上下文能力大模型，二是模拟世界、物理运动规律的多模态能力。

1. AGI 进程中，大模型上下文输入长度是关键之一

表2：主流大模型上下文输入长度对比

大模型或产品	公司	上下文窗口可支持长度
GPT-4 Turbo	OpenAI	12.8 万 (tokens)
Claude-3	Anthropic	20 万 (tokens)
Gemini 1.5 pro	Google	100 万 (tokens)
Kimi	Moonshot (月之暗面)	200 万 (汉字)

资料来源：OpenAI 等官网，中国银河证券研究院

2. 多模态加速模拟文本、图像及视频，未来模拟物理运动规律将成为现实

多模态模型是指能够处理不同类型数据，并将其融合进行综合理解的人工智能模型。这种模型能够更全面地理解和处理真实世界中复杂多样的信息，从而进一步提升大型模型的迁移学习能力。多模态技术的发展在人工智能领域具有重要意义。

当前，单模态的人工智能模型，如处理文本、语音、图片等的模型，已经相对成熟。而大型模型正在向多模态信息融合的方向快速发展：重要模型诞生以及 GPT-4 等模型的图像处理能力提升。大型模型不仅限于文字和图像的处理，也开始拓展到音频、视频等领域，未来甚至有望延伸到包括味道等其他信号。

表3：AGI 带来千行百业颠覆性革命

类别	方向	细分领域
生产力	前端	科研辅助、智能工程师、产品设计、生产管理、执行危险任务等
	中端	产业分析、经济预测、投资决策等
	后端	人力资源、广告公关、促销销售、售后服务客服等
消费领域	软件	信息搜索、购物、教育、金融、娱乐、社交、精神伴侣、自动驾驶等
	硬件	通用机器人、智能家居、育儿、养老、医院陪护、个人智能硬件等
政务领域	政策	政策制定、政策执行、社情民意等
	管理	智慧执法、智慧稽查、智慧公共安全等
军事等	军用	智能装备、军队管理、态势分析
	国际	国际舆情分析、国际谈判、局势分析等

资料来源：AGICall 等，中国银河证券研究院

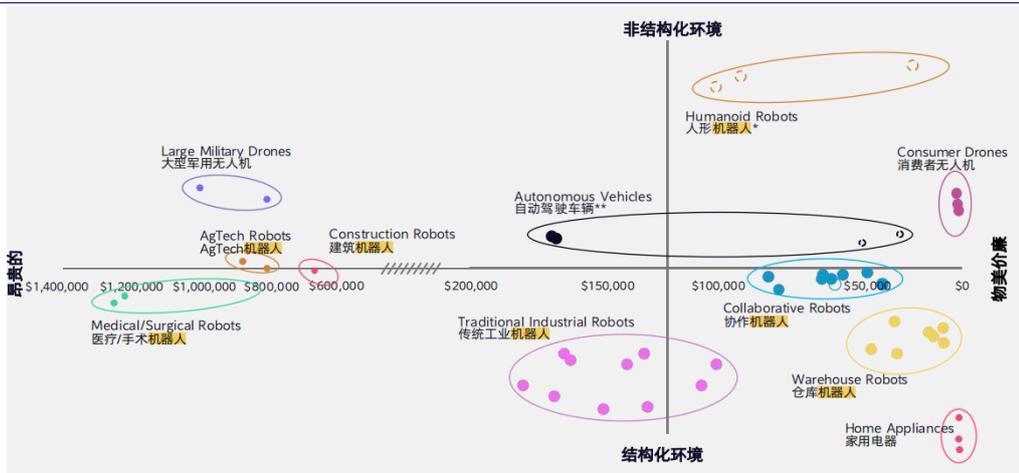
我们认为未来通用人工智能重点将变革以下领域：

1. 具身智能成为 AI 发展新形态，机器人将取代大部分工种

具身智能作为人工智能发展的一个重要分支，是指那些可以感知、并与物理世界进行交互、具有自主决策和行动能力的人工智能系统。这些智能体能够以主人的视角感受物理世界，并通过与环境的交互结合自我学习来理解和改变客观世界，机器人将成为具身最优载体。

未来机器人产业将持续快速发展，迎来战略机遇期。具身智能作为两大领域交叉的核心应用，有望在未来取得快速发展。它将推动智能体具备更多自主规划、决策、行动和执行的能力，实现人工智能的进一步进阶。

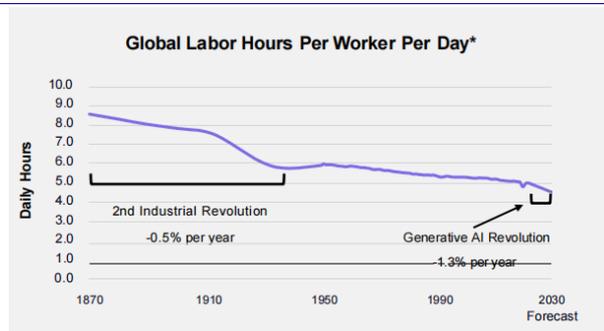
图10: 受益于人工智能和计算机视觉，机器人将能够在非结构化环境中经济高效地运行



资料来源: ARK 投资, 中国银河证券研究院

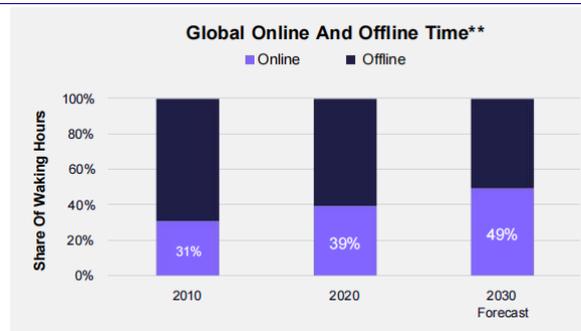
AI 与机器人结合能充分提高生产效率，更好应对各种复杂任务。在流水作业以及生产线上，机器人可以根据大模型提供的实时数据分析结果，对生产流程进行自动优化，提高生产效率和产品质量。并且能快速学习新技能，有效解放劳动力，降低劳动力工作时长，未来人类将更多时间应用在线上处理工作任务，一周工作 3-4 天或将成为现实。

图11: 全球劳动力每天工作时长



资料来源: ARK 投资, 中国银河证券研究院

图12: 全球上网时长



资料来源: ARK 投资, 中国银河证券研究院

2. 脑机接口成为新的创新交互方式

脑机接口作为一种新型的人机交互方式，在医学、教育、游戏等领域有着广阔的应用前景。脑机接口技术将脑电信号转化为计算机可识别的数据，并通过计

计算机的反馈，实现了人与机器的无缝互动。

目前脑机接口技术正从“学术科学探索”走向“应用转化落地”。脑机接口系统性能指标包括响应时间、识别正确率、可输出指令数量和菲茨吞吐量，可用性指标包括易用性、长效性、鲁棒性、安全性和互操作性，易用性指标包括准备时长、轻便性和舒适性。

表4：脑机接口系统关键指标

脑机接口系统关键指标	衡量指标
性能指标	响应时间
	识别正确率
	可输出指令数量
	菲茨吞吐量
可用性指标	易用性
	长效性
	鲁棒性
	安全性
	互操作性

资料来源：中国信通院，中国银河证券研究院

经过数十年的科学探索与技术论证，脑机接口已从科幻成为科学，处于从科学研究到产业落地的关键时期。就脑机接口目前的发展情况，在今后一段时间，脑机接口的基础学科研究和应用落地都将得到长足发展，从而有望促进脑机接口市场规模不断扩大。

另一方面，人类与人工智能之间的交互方式也在不断升级，脑机接口有望成为下一代人机交互方式。当前，脑机接口技术正在突破人类的生理界限，为残障人士提供了前所未有的可能性。

3. 治疗罕见病与精准治疗成为现实

未来 AGI 可以应用的领域众多，其中绕不开人类生物工程。我们认为，医疗是 AGI 落地的最佳场景之一，大模型、多模态以及垂类大模型将更加广泛结合并应用在药物研发、诊断、影像、治疗等细分环节。

随着各类医疗大模型的加速迭代与演化，医疗大模型商业化前景有望进一步打开。例如，谷歌的 Med-PaLM2、微软子公司 Nuance 的 DAXExpress 等医疗大模型已经在医疗领域得到应用，并取得了一定的商业化成果。

在最近的一项研究中，Med-PaLM2 在 USMLE 问题上的准确率达到 85.4%，与参试专家的水平相当。这使得 Med-PaLM2 成为第一个在 USMLE 问题上达到专家级表现的人工智能系统。

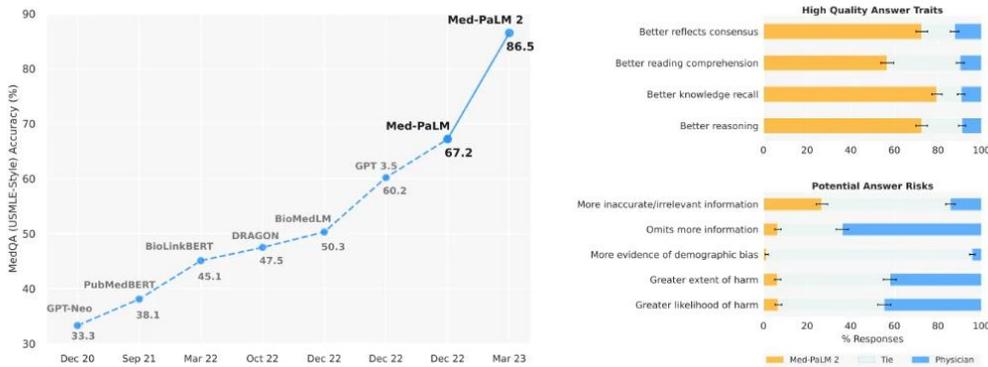
1) 可以帮助医生更快速、更准确地进行诊断。它可以通过分析大量的病例和医学文献，提供对疾病的诊断和治疗建议。这有助于减少医疗错误和误诊的风险；还可以实现疾病的早期发现和治疗，从而改善患者的治疗效果并挽救生命。

2) 可以帮助医生节省时间和精力，使他们能够更专注于与患者的沟通和治疗。在 Med-PaLM 2 一项实验中，通过对超过 50 万个医学图像进行分析，成功预测了肺癌的发生率。Med-PaLM 2 还可以通过自动化完成目前由医生执行的许多

日常工作，使他们能够有更充足的时间专注于为患者提供服务，可以缩短患者的等待时间并提高患者满意度。

3) 可以与其他医疗设备和系统进行集成，可以与智能手环和智能手表等设备进行连接，实时监测患者的生理参数，并提供相应的建议和警告。还可以与电子病历系统等其他医疗系统进行集成，从而实现更加智能化和高效化的医疗服务。

图13: Med-PaLM 2 在 MedQA 上的表现比 Med-PaLM 提高了 19% 以上



资料来源: 《Towards Expert-Level Medical Question Answering with Large Language Models》(《Big Ideas 2024》、ARK Invest), 中国银河证券研究院

未来 AGI 时代，AI 赋能医疗有广阔前景。可以应用的领域很多，其中绕不开人类生物工程。我们认为，医疗是 AGI 落地的最佳场景之一，大模型、多模态以及垂类大模型将更加广泛结合并应用在药物研发、诊断、影像、治疗等细分环节，很多罕见病及疑难杂症将逐渐被治愈。

二、人工智能是核心技术催化剂，领航数字经济新阶段

(一) “适度超前”建设算力体系背后的财政货币支撑体系

1. 央行的“科技金融”、“数字金融”货币政策框架

1.1 结构性货币政策支持科技创新和数字经济发展

中央经济工作会议将“稳中求进、以进促稳、先立后破”作为 2024 年宏观政策基调。在此背景下，货币政策“灵活适度，精准有效”，总量和结构政策双重发力，既托底总量增长又推动结构改革，支持防范化解宏观风险，着力营造良好的货币金融环境，高质量服务实体经济。

“精准”即为强调信贷的方向引导。预计未来结构性货币政策工具将在货币投放中扮演更加重要的角色，加大对重大战略、重点领域和薄弱环节的支持力度。聚焦“五篇大文章”，即科技金融、绿色金融、普惠金融、养老金融、数字金融五篇大文章。

中国人民银行货币政策课题组近期文章《结构性货币政策助力做好“五篇大文章”》中详细阐述了结构性货币政策工具的作用。“结构性货币政策是指在市场配置资源基础上，设计适当激励机制，引导资金流向经济特定领域的货币政策。结构性货币政策主要发挥结构功能，通过建立激励相容机制，将中央银行资金与金融机构对特定领域和行业的信贷投放挂钩，发挥精准滴灌实体经济的独特优势。结构性货币政策也有总量效应，通过投放基础货币，保持银行体系流动性合理充裕，支持信贷平稳增长”。

截至 2023 年末，结构性货币政策工具余额 7.5 万亿元，比上年末增加约 1 万亿元，占人民银行总资产的 16.4%。其中为“科技金融”和“数字金融”已创设的结构性货币政策工具共 2 个，分别为“科技创新再贷款”和“设备更新改造专项再贷款”。2022 年 4 月，人民银行创设科技创新再贷款，支持金融机构加大对科技创新企业的信贷支持力度；2022 年 9 月，创设设备更新改造专项再贷款，支持金融机构向制造业、社会服务领域和中小微企业、个体工商户等设备更新改造提供贷款。根据央行披露，截至 2023 年第三季度末，科技创新再贷款 4000 亿元额度全部用完，支持金融机构向科技企业累计发放贷款 1.69 万亿元；设备更新改造专项再贷款累计发放 1694 亿元，支持新型基础设施和产业数字化转型等设备更新改造。

2023 年科创企业贷款的获贷率提升，增速高于各项贷款增速。2023 年获得贷款支持的科技型中小企业 21.2 万家，获贷率 46.8%，获贷率提升 2.1pct。科技型中小企业本外币贷款余额 2.45 万亿元，同比增长 21.9%，比上年末低 3.8pct，比同期各项贷款增速高 11.8pct。2023 年获得贷款支持的高新技术企业 21.75 万家，获贷率为 54.2%，提升 0.8pct。高新技术企业本外币贷款余额 13.64 万亿元，同比增长 15.3%，比上年末低 0.8pct，比同期各项贷款增速高 5.2pct。

未来央行可能会继续在“科技金融”、“数字金融”领域创设相关结构性货币政策工具，丰富政策工具箱，引导商业银行信贷直达实体经济，引导信贷投放方向的作用，将为建设算力体系提供资金支持。

图14：结构性货币政策工具梳理（截至 2023 年末）

	工具名称	存续状态	支持领域	发放对象	利率（1年期）/额度激励比例（%）	额度（亿元）	余额（亿元）
长期性工具	1 支农再贷款	存续	涉农领域	农商行、农合行、农信社、村镇银行	2.00	8,100	6,562
	2 支小再贷款	存续	小微企业、民营企业	城商行、农商行、农合行、村镇银行、民营银行	2.00	18,000	16,551
	3 再贴现	存续	涉农、小微和民营企业	具有贴现资格的银行业金融机构	2.00（6个月）	7,400	5,920
阶段性工具	1 普惠小微贷款支持工具	存续	普惠小微企业	地方法人金融机构	1.00（激励）	800	498
	2 抵押补充贷款	存续	棚户区改造、地下管网、重点水利工程等	开发银行、农发行、进出口银行	2.25		32,522
	3 碳减排支持工具	存续	清洁能源、节能减排、碳减排技术	21家全国性金融机构、部分外资金融机构和地方法人金融机构	1.75	8,000	5,410
	4 支持煤炭清洁高效利用专项再贷款	到期	煤炭清洁高效利用、煤炭开发和利用和储备	工农中建交、开发银行、进出口银行	1.75	3,000	2,748
	5 科技创新再贷款	到期	科技创新企业	21家全国性金融机构	1.75	4,000	2,556
	6 普惠养老专项再贷款	存续	浙江、江苏、河南、河北、江西试点，普惠养老项目	工农中建交、开发银行、进出口银行	1.75	400	18
	7 交通物流专项再贷款	到期	道路货物运输经营者和中小微物流（含快递）、仓储企业	工农中建交、邮储、农发行	1.75	1,000	316
	8 设备更新改造专项再贷款	到期	制造业、社会服务领域和中小微企业、个体工商户	21家全国性金融机构	1.75	2,000	1,567
	9 普惠小微贷款减息支持工具	到期	普惠小微企业	16家全国性金融机构、地方法人金融机构	1.00（激励）		269
	10 收费公路贷款支持工具	到期	收费公路主体	21家全国性金融机构	0.50（激励）		83
	11 民企债券融资支持工具（II）	存续	民营企业	专业机构	1.75	500	0
	12 保交楼贷款支持计划	存续	保交楼项目	工农中建交、邮储	0.00	2,000	56
	13 房企纾困专项再贷款	到期	房企项目并购	5家全国性资产管理公司	1.75	800	0
	14 租赁住房贷款支持计划	存续	试点城市收购存量住房	工农中建交、邮储、开发银行	1.75	1,000	0

资料来源：中国人民银行，中国银河证券研究院

1.2 提升直接融资占比，构建覆盖科技型企业全生命周期的金融服务体系

央行 2023 年 4 季度货币执行报告首次提出“合理把握债券与信贷两个最大融资市场的关系”，扩大直接融资，社融结构实现再平衡。提升直接融资的要求与当下中国经济结构转型的新方向相匹配。报告专栏 1《准确把握货币信贷供需

规律和新特点》提出“先进制造、科技创新、绿色低碳、数字经济等新兴产业蓬勃发展，这些新动能领域与直接融资的金融支持模式更为适配，也会对贷款形成良性替代”。央行行长在十四届全国人大二次会议经济主题记者会上强调“在宏观层面，要加强顶层设计和系统筹划。比如，科技金融方面，科技型企业一般会经历种子期、初创期、成长期、成熟期不同的阶段，企业成长周期的不同阶段，对金融需求有不同的特点。在科技型企业成长的早期，更多需要风险投资、创新创业投资基金的介入，目前还是一个薄弱环节；金融机构对科技型企业风险评估能力，也需要进一步提升，下一步需要着力补齐短板，构建覆盖科技型企业全生命周期的金融服务体系”。

目前，交易商协会创设了科创类融资产品工具箱，并在 2022 年将其升级为了科创票据。科创票据用于支持科创类企业以及非科创类企业的科技创新发展行为。2022 年上交所和深交所正式落地了科创债。

因此，除了通过结构性货币政策工具为建设算力体系提供间接融资支持外，对于相关企业的债券融资等直接融资的支持也会进一步提升，试图构建覆盖科技型企业全生命周期的金融服务体系。

2. 财政政策支持重点首次转向现代化产业体系建设

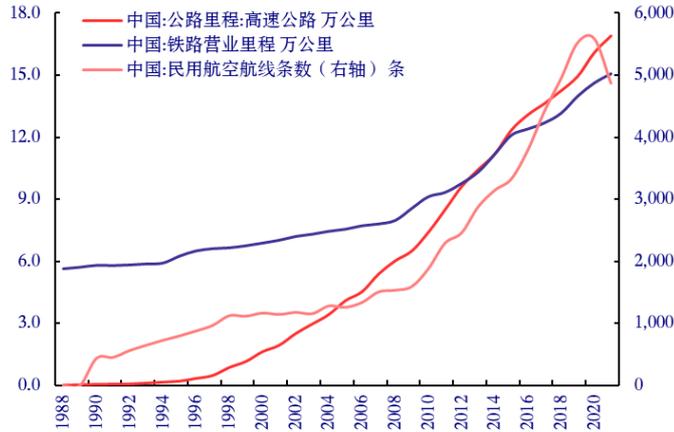
从去年年底的中央经济工作会议到今年两会期间的政府工作报告，均把“加快现代化产业体系”建设放在首要目标，与此同时持续强调要增强宏观政策取向一致性。我们认为，在当前经济同时面临逆周期和结构性调节的关键时期，财政政策配合国有资本的政策调整将是有效发挥举国体制优势，支持数字经济和科技转型的重要力量。其实，过去以来的历次经济周期中，积极的财政政策均起到了关键性作用，力挽经济于狂澜。但我们也注意到，过去几轮积极财政政策主要支持的方向为传统基建产业，而现如今财政政策支持重心转向科技创新，需要通过何种方式调整？

2.1 过去三轮积极财政政策主要投向传统产业

我国分别于 1998 年、2008 年以及 2013 年开启过三轮积极财政政策，其中前两次均是由外部因素导致的需求冲击加剧了经济波动，继而使用扩张性财政政策予以对冲。始于 2013 年的第三轮积极财政政策主要是由于传统产能的供给过剩引起，财政政策方面主要以结构性政策为主，但由于土地财政等预算外广义财政的存在，使得传统产能并未完全出清。由此，过去不同历史时期财政政策的主要支持工具及资金投向也有所区别：

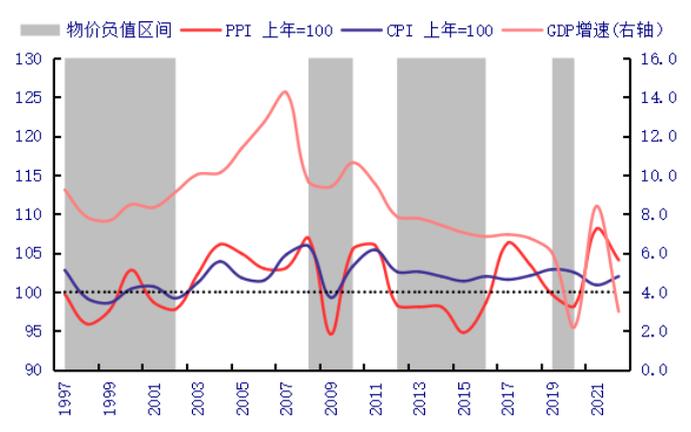
1998 年首轮财政支持经济转型：以狭义政府赤字和财政支出政策为主，促进城镇化、工业化转型。受到亚洲金融危机的影响 1997 年 6 月至 1999 年 12 月 PPI 连续 31 个月在负值区间运行，且 CPI 处于负值区间。未来应对外需冲击，同时促进我国城镇化和工业化的转型发展，我国分别于 1998 年至 1999 年两年期间多次增发共计 1600 亿元的国债，需要注意的是 1998 年我国 GDP 水平仅 8.5 万亿元。当时我国刚刚实行改革开放不久，如图 15 所示，1994 年我国公路及民航建设刚刚起步，国内基础设施建设具有大幅扩张的空间及较高的投资回报率，相对而言资本是稀缺的。为此，1998 年新增国债由中央政府向国有商业银行定向增发进行募资。

图15: 首轮积极财政政策期间我国基础设施处于起步阶段



资料来源: WIND, 中国银河证券研究院

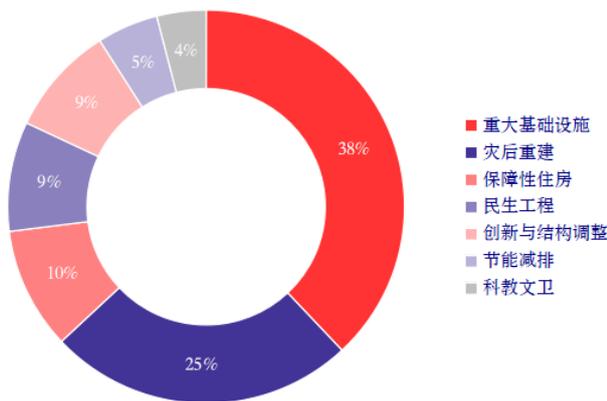
图16: 1998 年至今价格指数及 GDP 增速 (%)



资料来源: WIND, 中国银河证券研究院

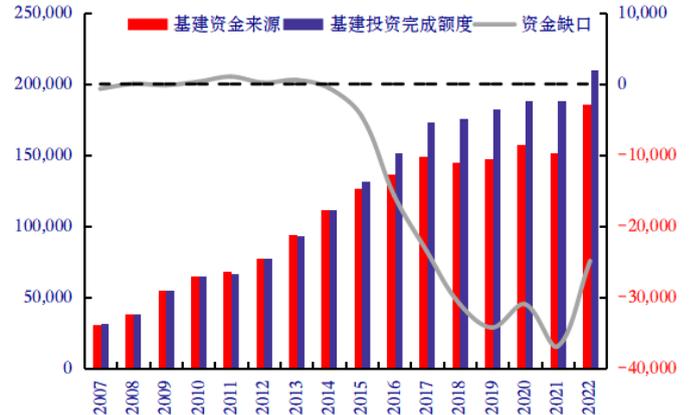
2008 年第二轮财政支持经济转型: 以政府性基金扩张为主, 催生地产与基建快速发展。2008 年全球金融危机影响, 中国国内需求遭受大幅冲击, PPI 负值运行 12 个月, CPI 同样落入负值区间。为稳定经济增长, 同时推动一批利长远项目建设, 我国实施“四万亿”投资计划。资金来源方面, 据国家发改委公布的“4 万亿”投资明细, 四万亿投资的资金构成是 1.18 万亿中央预算内投资和 2.82 万亿配套资金。资金来源是中央财政赤字、地方财政、地方债(财政部代理发行)、政策性贷款、企业债和中期票据、银行贷款, 以及民间投资。其中财政预算内较上一轮积极财政中, 多增了政府性基金收入中的土地收入增长。2009 年和 2010 年的政府性基金收入分别为 1.83 万亿元和 3.57 万亿元, 较上年分别增长了 17.26% 和 95.15%, 其中超过 70% 专项用于土地开发及建设领域的投资和补偿。资金投向方面, 主要是重大基础设施建设、灾后重建、保障房建设、民生工程等。

图17: “四万亿”计划的主要投资方向



资料来源: WIND, 中国银河证券研究院

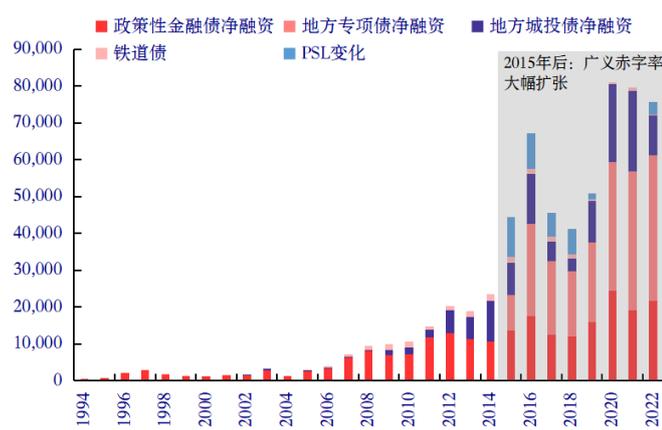
图18: 我国基建投资资金来源与完成额间的资金缺口 (亿元)



资料来源: WIND, 中国银河证券研究院

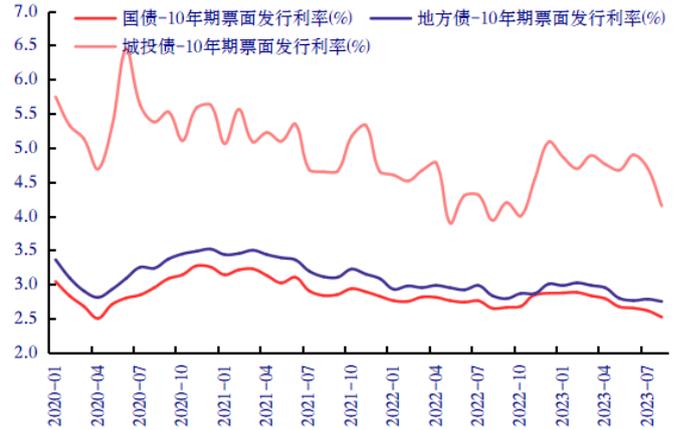
2013 年后第三轮积极财政: 以结构性政策配合完成过剩产能出清, 但实际开启了广义财政扩张。2008 年以来广义财政的扩张导致传统产能过剩, 经济结构性矛盾和隐性债务风险不断加剧。理论上此时财政政策应该以结构性政策配合货币政策完成“供给出清”, 我们可以看到基建的投资缺口在 2015 年之后开始显著提升, 如图 19 所示。但实际上由于地方政府考核目标仍在, 地方在稳经济过程中实际开启了广义财政的扩张, 而“宽货币、紧信用”的货币政策使得“宽”出的货币进了城投和地产, “紧”的信用又使城投承担了较高的融资成本, 造成了一定隐性债务。

图19: 1994 年至今广义财政净融资额变化 (亿元)



资料来源: WIND, 中国银河证券研究院

图20: 城投债与政府债的融资利差



资料来源: WIND, 中国银河证券研究院

2.2 新一轮积极财政支持重点转向现代化产业体系建设

2023 年中央经济工作会议及 2024 年政府工作报告均将“现代化产业体系建设”放在首位,与之对应的 2024 年财政预算草案中,对于主要财政收支政策也做出了重要调整。根据过往几年预算草案来看,每年基本会公布 7-8 项主要政策,其基本规律是:后几项政策相对固定(第四是乡村振兴、第五是生态环境、第六是民生、第七是国防、外交、政法),但前三项政策的内容和排序往往指明当年重点方向。例如 2021 年首要政策是“推动创新发展和产业升级”、2022 年首要政策目标是保市场主体、2023 年为扩大内需。

而 2024 年主要财政收支政策中的前两项均和科创相关,第一是支持加快现代化产业体系建设,第二是支持深入实施科教兴国战略。去年的首要目标“扩大内需”已经移居第三。据此可以得出的结论是:新一轮财政逆周期调控不同于以往几次扩张,更加聚焦经济转型和科技创新。

图21: 2021-2024 年主要财政指标及政策表述

	2021年	2022年	2023年	2024年
总基调	积极的财政政策要提质增效、更可持续	积极的财政政策提升效能,更加注重精准、可持续	积极的财政政策要加力提效	积极的财政政策要适度加力、提质增效。
财政赤字	3.46 万亿	3.5 万亿+2 万亿国企利润和稳定基金	3.88+1 万亿	4.06 万亿
赤字率	3.2% 左右	2.8% 左右	3%+0.8%	3%
专项债	3.65 万亿	3.65 万亿	3.8 万亿	3.9 万亿+1 万亿特别国债
主要收支政策:	1. 推动创新发展和产业升级; 2. 支持实施扩大内需战略; 3. 支持推进区域协调发展和新型城镇化; 4. 支持全面推进乡村振兴; 5. 支持加强污染防治和生态建设; 6. 加强基本民生保障; 7. 支持国防、外交和政法工作。	1. 加强对市场主体支持,着力稳企业保就业; 2. 大力支持科技创新,提升产业发展水平; 3. 充分挖掘内需潜力,推进区域协调发展和新型城镇化; 4. 推动农业高质量发展,全面推进乡村振兴; 5. 持续改善生态环境,推动绿色低碳发展; 6. 完善基本民生保障,切实保障和改善民生; 7. 强化资金和政策保障,支持国防、外交、政法等领域工作。	1. 发挥财政稳投资促消费作用,着力扩大国内需求; 2. 推进高水平科技自立自强,支持现代化产业体系建设; 3. 进一步优化政策实施方式,持续增强各类企业活力; 4. 加强乡村振兴投入保障,着力推进城乡融合和区域协调发展; 5. 完善绿色低碳财税支持政策,协同推进降碳、减污、扩绿、增长; 6. 强化基本公共服务,扎实做好民生保障; 7. 统筹发展和安全,支持国防、外交、政法等工作。	1. 支持加快现代化产业体系建设; 2. 支持深入实施科教兴国战略; 3. 支持扩大国内需求; 4. 支持保障和改善民生; 5. 支持推进乡村振兴; 6. 支持城乡融合、区域协调发展; 7. 支持加强生态文明建设; 8. 支持国防、外交和政法工作。
减税降费	继续执行制度性减税政策,实施新的结构性减税举措,对冲部分政策调整带来的影响。对小微企业和个体工商户应纳税所得额不到 100 万元的部分,在现行优惠政策基础上,再减半征收所得税。	实施新的组合式税费支持政策,坚持阶段性措施和制度性安排相结合,减税与退税并举。对小微企业年应纳税所得额 100 万元至 300 万元部分,再减半征收企业所得税。并全年退还抵税约 2.5 万亿元,其中留抵退税约 1.5 万亿元,退税资金全部直达企业。	完善税费优惠政策,对现行减税降费、退税免税等举措,该延续的延续,该优化的优化。	落实好结构性减税降费政策,重点支持科技创新和制造业发展。全年新增税费优惠总额超 22 万亿元。

资料来源: 财政部、中国银河证券研究院

2.3 当前及未来一段时期财政如何支持科技创新发展?

从今年财政预算草案及相关新增政策来看,未来具体支持政策主要以下几方面:

一是新增政府债务工具重点支持科技创新。一方面是专项债用途在以往的传统基建和新能源建设的基础之上，新增了数字基础设施建设用途。另一方面，我们看到今年两会提出了新型债务工具一超长期特别国债，在3月6日举行的十四届全国人大二次会议经济主题记者会上，国家发展和改革委员会主任郑栅洁指出初步考虑，超长期特别国债将重点支持科技创新、城乡融合发展、区域协调发展、粮食能源安全、人口高质量发展等领域建设。这些领域潜在建设需求巨大、投入周期长，现有资金渠道难以充分满足要求，亟需加大支持力度。其中首要支持领域便是科技创新。

表5：2019年至今历次专项债用途扩容

2019年	2020年	2021年	2022年	2024年
		交通基础设施		交通基础设施
		能源	交通基础设施	能源
		农林水利	能源	农林水利
交通基础设施	交通基础设施	生态环保	农林水利	生态环保
能源	能源	民生服务	生态环保	民生服务
农林水利	农林水利	城乡冷链物流	民生服务	城乡冷链物流
生态环保	生态环保	市政产业园区	城乡冷链物流	市政产业园区
民生服务	民生服务		市政产业园区	国家重大战略项目
城乡冷链物流	城乡冷链物流		国家重大战略项目	保障性安居工程
市政产业园区	市政产业园区	新增：	保障性安居工程	新型基础设施
		国家重大战略项目		新能源项目
		保障性安居工程	新增：	预计新增：
			新型基础设施	数字经济
			新能源项目	城中村改造

资料来源：财政部，中国银河证券研究院

二是税收政策持续深化落实，加大研发费用扣除比例。即落实技术改造相关投资税收优惠，落实研发费用加计扣除等政策，例如去年已将符合条件的集成电路和工业母机企业研发费用税前加计扣除比例提高至120%，将符合条件的研发费用税前加计扣除比例由75%提高至100%。预计在今年及未来一段时间将持续贯彻落实以上税收支持政策，且有望逐步扩大政策支持范围和支持力度。

三是更好发挥国有资本和国有企业在科技创新中的引领、引导作用。今年1月份国务院印发了《关于进一步完善国有资本经营预算制度的意见》，本次新《意见》在之前基础上对于“国有资本”的功能定义新增了“落实国家战略”，并居于首位，如表6所示。而当前我国首要战略显然是科技创新和数字经济的发展。实际上，国有资本经营预算作为财政收支的重要组成部分，既是中国特色，也是我国更好支持科技转型的体制优势。本次制度的修订实现了所有国有企业的预算全覆盖，并加强了支出纪律约束，未来对于关键领域的国有资本的注入有望助力科技创新相关行业的快速发展。

表6: 关于国有资本经营预算的定义变化

	《国务院关于试行国有资本经营预算的意见》	《关于进一步完善国有资本经营预算制度的意见》
时间:	2007年09月08日	2024年1月6日
相同内容:	国有资本经营预算, 是国家以所有者身份依法取得国有资本收益, 并对所得收益进行分配而发生的各项收支预算, 是政府预算的重要组成部分。	
不同内容:	建立国有资本经营预算制度, 对增强政府的宏观调控能力, 完善国有企业收入分配制度, 推进国有经济布局和结构的战略性调整, 集中解决国有企业发展中的体制性、机制性问题, 具有重要意义。	国有资本经营预算是企业国有资产管理的重要内容, 是落实国家战略、增强政府宏观调控能力、推进国有经济布局优化和结构调整的重要力量。

资料来源: 国务院, 中国银河证券研究院

四是专项产业基金支持。去年我国的产业基础再造和制造业高质量发展专项资金增长了 20.3%，今年预算草案再安排专项资金 104 亿元。其中主要强化对制造业企业技术改造的资金支持，落实技术改造相关投资税收优惠政策。深入实施首台（套）重大技术装备和首批次重点新材料应用保险补偿政策。优化产业投资基金功能，鼓励发展创业投资、股权投资，充分运用市场化手段，支持集成电路、新一代信息技术等产业加快发展。

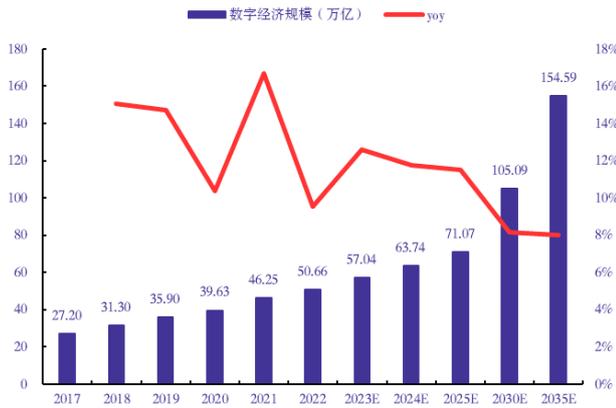
五是教育和研发预算支出支持。今年财政预算草案中主要财政政策第二位是“科教兴国”，从支持加快建设高质量教育体系和推动高水平科技自立自强两方面做了阐述，主要支持政策是加大教育支出和财政研发补贴。于此同时，对于今年中央财政支出的预算安排中，科技支出和教育支出预算安排的支出增速分别为 10% 和 5%，大幅高于去年的 2.9% 和 1.9%，是今年中央各项财政支出中主要提升的项目。

（二）大国体系下的中国数字经济产业已全面开启，预计在 2035 年达到 GDP 的 71.6%

总量法测算：国内数字经济预计 2035 年占 GDP 比将达到 71.60%。近年来，我国数字经济整体实现量的合理增长，2022 年数字经济规模达到 50.2 万亿元，同比增加 4.68 万亿元，首次突破 50 万亿元。2023 年，面对经济新的下行压力，各级政府、各类企业纷纷把发展数字经济作为培育经济增长新动能、抢抓发展新机遇的重要路径手段，数字经济发展活力持续释放，我国数字经济规模有望达到 54.6 万亿元，面对多方面不利因素，我国数字经济仍保持强劲增长、凸显韧性，持续为国民经济稳增长保驾护航。

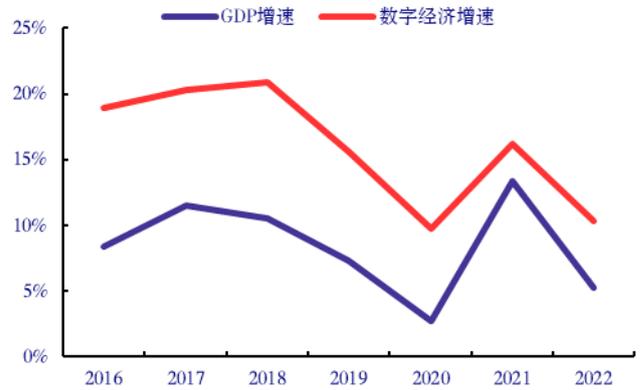
我国数字经济规模维持高位增长，增速连续 11 年高于名义 GDP 增速。2022 年，我国疫情防控取得重大胜利，经济发展环境得到改善，国内生产总值同比名义增长 5.3%，数字经济规模达到 50.2 万亿元，同比名义增长 10.3%，高于 GDP 名义增速 4.98 个百分点。自 2012 年以来，我国数字经济平均增速 15.9%，已连续 11 年显著高于 GDP 增速，数字经济持续发挥经济“稳定器”、“加速器”作用。

图22：中国数字经济规模及增速预测



资料来源：信通院、中国银河证券研究院

图23：中国数字经济增速（名义）对比 GDP 增速（名义）



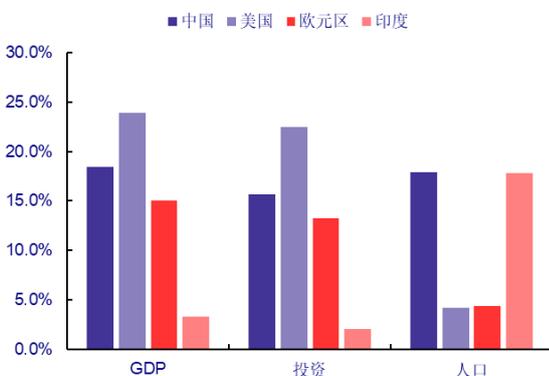
资料来源：信通院、中国银河证券研究院

我国数字经济占 GDP 比重持续提升，2030 年数字经济占比有望追上发达国家水平，2035 年有望位列全球首位。2022 年，我国数字经济占 GDP 比重为 41.86%，从 2023-2035 年的整体趋势及预测来看，中国数字经济占 GDP 的比重持续提升，我们预测 2030 年占比达到 59.73%，有望追上发达国家平均水平，预计 2035 年占比将达到 71.60%。

（三）中美数字经济及人工智能产业要素发展对比

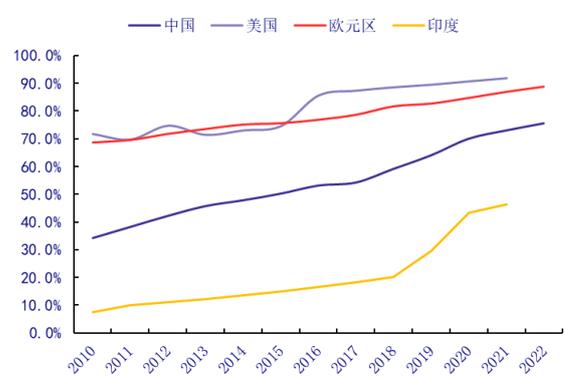
人工智能的三大基础要素为数据、算力和算法，目前中国在数据方面具有明显的大国优势。首先，数据的底层是人和人的活动，因此发展主体（国家或者区域内）的人口数量与质量对数据资源的“量”与“质”起到至关重要的影响。数量方面，目前世界人口排名前列的国家或地区依次为印度 14.17 亿、中国 14.12 亿、欧盟 4.48 亿、美国 3.33 亿。据 IDC 统计，2022 年中国产生的数据规模达 23.3ZB，在全球占比达到 23%。伴随国内数据要素成为国家战略，数据价值不断释放，到 2025 年中国人工智能产业在数据量方面将具有相对优势，IDC 预计到 2025 年中国数据量规模达到 48.6ZB。质量方面，2022 年中国人均 GDP 为 1.27 万美元，持平全球平均水平，但从互联网渗透率来看，中国达到 75.6%，明显高于世界平均水平 63%。

图24：中美欧印 GDP、投资、人口占全球比重



资料来源：WIND, WorldBank, 中国银河证券研究院

图25：中美欧印互联网渗透率



资料来源：WIND, WorldBank, 中国银河证券研究院

算力方面，中美两国的算力与其他梯队国家相比有显著优势，但美国仍领先于中国。由于超大规模互联网企业在算力投入上的大幅增长，2022年美国算力指数从77分增长到82分，尤其是一级指标计算能力增长显著，包括通用计算能力和AI计算能力。从数据来看，2022年，美国服务器市场规模达到530亿美元，同比增长19.7%；人工智能服务器市场规模达到75亿美元，同比增长48.1%，是全球服务器市场增长的主要驱动力。中国受阻于疫情反复，2022年算力投入有所放缓，但整体增速仍高于GDP，算力指数从70分增长到71分。在这样的大背景下，中国整体服务器市场规模仍然保持6.9%的正增长，达270亿美元，占全球市场25.0%，仅次于美国稳居第二。从服务器子市场来看，边缘计算服务器和液冷服务器市场均呈现20.0%以上的增长。

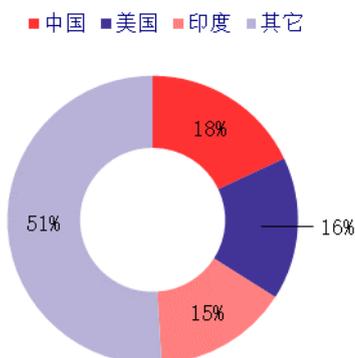
算法方面，我们主要通过人才规模、专利数、企业数量、融资规模和大模型这几个角度来对比。

1) 人才方面，人工智能人才整体规模中国居首，但人工智能顶尖人才美国处领先地位。中国人工智能人才规模凸显，LinkedIn和猎聘的数据统计显示，全球当前累计AI人才突破100万人。其中中国AI人才占全球AI人才总数的18%，位居世界首位。美国和印度AI人才数量分居全球第二、三位，且均超过15万。根据清华大学AMiner的历年统计数据，美国学者入选AI2000榜单年均超过1,000人次，数量最多且远高于其他国家，表明美国在AI顶尖人才方面具有领军地位。与之对比，中国人才数量位居第二，但差距仍然较大。

2) 专利方面，中国信通院发布的《全球数字经济白皮书（2023年）》指出，在专利申请授权方面，2013-2023年Q3，全球AI专利申请量累计达129万，全球AI专利授权量累计超51万，中国AI专利申请量占全球64%，位列全球第一，论文数也遥遥领先。

3) 融资方面，当前美国在人工智能企业数量和融资规模方面均占据显著优势。截至2023年6月底，全球人工智能企业共计3.6万家，美中英企业数量名列前茅。美国人工智能企业数量约为1.3万家，在全球占比达34%，中国占比16%，英国7%，美中英三国的人工智能企业数量合计占全球的56%。计算2013年以来的私人AI投资总额时，美国以2489亿美元的投资排名第一，其次是中国（951亿美元）和英国（182亿美元）。AI企业和融资活动集中在美、中、英等国家。

图26：全球人工智能人才数量占比



资料来源：LinkedIn,猎聘，清华大学AMiner团队发布的AI2000学者榜单，智谱研究，尚普研究院，中国银河证券研究院

图27：中美入选AI2000历年人数对比



资料来源：LinkedIn,猎聘，清华大学AMiner团队发布的AI2000学者榜单，智谱研究，尚普研究院，中国银河证券研究院

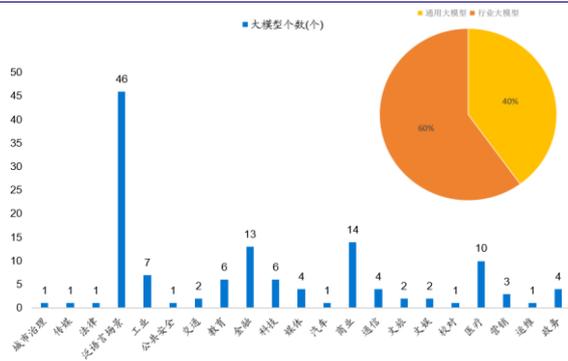
表7：中美人工智能产业发展状况对比，截至 2023 年 7 月

衡量指标	美国人工智能产业发展状况	中国人工智能产业发展状况
累计企业数量	13000 家	5760 家
累计融资规模	2489 亿美元	951 亿美元
累计专利数	116 万	191 万
大模型数量	114	130
算力指数	82	71
2025 数据量	30.6ZB	48.6ZB

资料来源：CBInsights、《2022-2023 全球算力指数评估报告》（IDC、浪潮信息、清华大学全球产业研究院联合编制），尚普研究，中国银河证券研究院

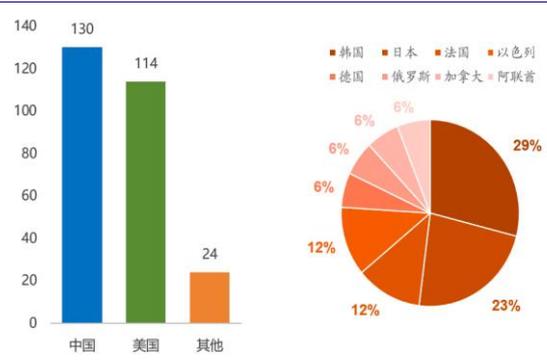
大模型方面，我们通过大模型数量和大模型表现来进行比较。大模型数量：根据赛迪顾问数据，截至 2023 年 7 月，中国已累计发布 130 个大模型；国外共发布 138 个，其中美国共 114 个。中美两国大模型合计数量占全球 90%以上，具有绝对优势。从大模型参数来看，中美两国的代表性大模型里均是既有千亿参数的通用大模型，又有几十亿参数的行业垂直大模型。但是，在影响力方面，中国缺少像 GPT-4、Gemini、Sora 这种具有全球影响力的大模型，中国的行业大模型居多，占总数的 60%，商业、金融、医疗居多。

图28：全球人工智能人才数量占比



资料来源：赛迪顾问，中国银河证券研究院

图29：截至 2023 年 7 月全球大模型累计数量区域分布情况



资料来源：赛迪顾问，中国银河证券研究院

表8：国内外代表性大模型参数及预训练数据对比

模型	研发团队	参数规模	预训练 tokens
Gemini 1.5 pro	谷歌	175B	-
Gemma-2B (开源)		2B	2T
Gemma-7B (开源)		7B	6T
StableDiffusion	高通	65B	1.5T
GPT-3	OpenAI	175B	300B
GPT-4		1.8T	13T
LLaMA (开源)	Meta	7B-65B	1T/1.4T
Sora	OpenAI	7B	-
ChatGLM-6B (开源)	清华大学	6B	1T
GLM-130B (开源)		130B	400B
混元大模型	腾讯	100B+	2T+
文心一言	百度	260B	-

盘古 NLP	华为云	110B	-
Baichuan-7B (开源)	百川智能	7B	1.2T
Baichuan-13B (开源)		13B	1.4T
通义千问系列	阿里云	7B-72B	2.2T
讯飞星火	科大讯飞	175B	-

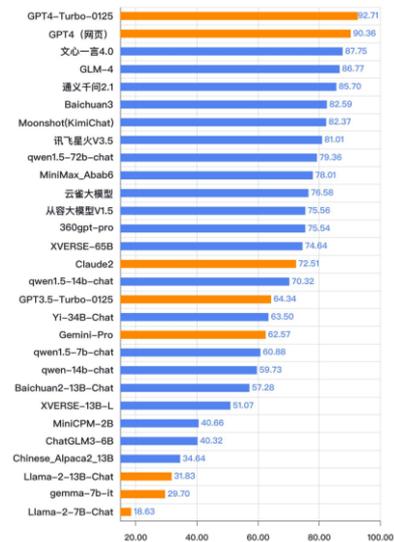
资料来源: CNBC, decoder, GitHub, CSDN, IT之家, 同花顺财经, 中华网, 中国银河证券研究院

大模型表现的比较,我们主要参考国内较权威机构 SuperCLUE 在 23 年 7 月至 24 年 2 月的测评结果。测评是基于 4572 道中文评测题,可以看到在过去半年里国内领军大模型在不断进步,与 GPT-4 的差距在不断缩小。**24 年 2 月的测评结果,在中文领域国内领军大模型的平均水平已经接近 GPT-4。**SuperCLUE 还对全球大部分模型进行了测评,从 24 年 2 月的结果来看,国内在中文领域综合能力超过 GPT3.5 的模型有 13 个,文心一言 4.0、GLM-4、通义千问 2.1 排名前三;国外模型的平均成绩为 57.83 分,国内模型平均成绩为 68.75 分。可以看出,国内大模型在中文领域的能力的平均水平已经超过国外大模型。

图30: 国内外代表性大模型中文测评得分趋势(23 年 7 月-24 年 2 月) 图31: 全球大模型中文测评结果 (24 年 2 月)



资料来源: SuperCLUE, 中国银河证券研究院



资料来源: SuperCLUE, 中国银河证券研究院

表9: 部分国内代表性大模型 SuperCLUE 中文测评基准得分(23 年 7 月-24 年 2 月)

模型	23 年 7 月	23 年 8 月	23 年 9 月	23 年 10 月	23 年 11 月	23 年 12 月	24 年 2 月
文心一言	50.48	54.18	53.72	61.81	73.62	75	87.75
通义千问	-	41.73	33.78	43.36	61.01	71.78	85.70
ChatGLM	42.46	38.49	54.31	58.53	63.27	69.91	87.77

资料来源: SuperCLUE, 中国银河证券研究院

三、人工智能三要素共振, 算力、算法、数据未来趋势推演

(一) 算力: 供需缺口加大, AI 服务器产业链分析

1. 大模型时代智能算力渗透率持续提升，AI 服务器有望量价齐升

服务器处理大规模计算，主流架构分为 X86 和 ARM。服务器是高性能计算机，比普通计算机运行更快、负载更高、价格更贵。主要功能包括运行网站、存储数据、运行程序、数据分析、云计算开发、人工智能运算等，在网络中为其它客户机如 PC 机、智能手机、ATM 等终端等大型设备提供计算或者应用服务，具有高速的 CPU 运算能力、长时间的可靠运行、强大的 I/O 外部数据吞吐能力以及更好的扩展性。

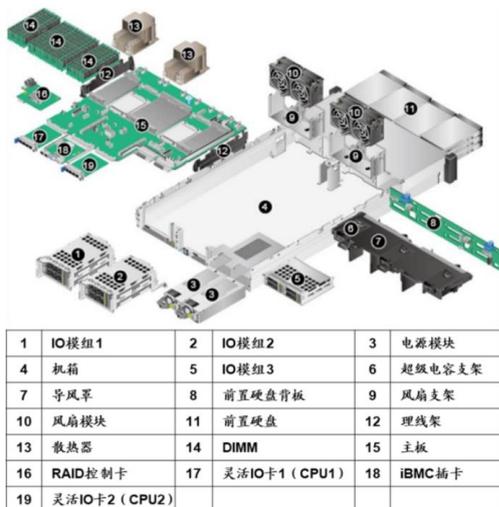
表10: 服务器架构

组件名称	功能	主要参与者
CPU	解释指令并处理数据，是核心的运算和控制中心。频率越高，性能越好	Intel、AMD、BM、IDT、Cyrilx、华为海思等
GPU	图像处理显示	NVIDIA、AMD、华硕、技嘉、微星、泰索、影驰等
DRAM	临时存储数据，在 CPU 与硬盘之间，存取效率较高	三星、海力士、美光
硬盘	永久存储数据，但存取速度较慢，可分为机械硬盘(HDD)和固态硬盘(SSD)	Seagate、西部数据 WDC、Toshiba、三星、海力士等
主板	电路板，连接其他组件	TTM、深南电路、沪电股份、方正科技等
电源	提供电力保障	台达、艾默生、光宝科技、康舒科技、高效电子、全汉电源等
网卡	网络连接功能	Intel、TPLINK、友讯、腾达，华硕、华为等

资料来源: IDC, 中国银河证券研究院

AI 服务器中用于运算和存储的芯片占服务器成本结构约 70%，通用服务器用于运算和存储芯片占服务器成本结构大约 50%左右。一台服务器主要硬件包括处理器、内存、芯片组、I/O(RAID 卡、网卡、HBA 卡)、硬盘、机箱(电源、风扇)。以一台普通的服务器生产成本为例，CPU 及芯片组大致占比 50%左右，内存大致占比 15%左右，外部存储大致占比 10%左右，其他硬件占比 25%左右。其中机器学习型服务器中 GPU 成本占比达 72.8%。

图32: 通用服务器拆解图 (以华为 TS200-2280 为例)



资料来源: 华为官网, 中国银河证券研究院

图33: AI 服务器拆解图 (以华为 Atlas800TA2 为例)



资料来源: 华为官网, 中国银河证券研究院

全球服务器市场高增长，中国市场占比提升。根据 Statista 数据，2022 年全球服务器市场规模达到 848.7 亿美元，同比增长 2.04%，中国服务器市场占比 30.16%，预计 2023

年全球服务器市场规模来到 907.8 亿美元,同比增长 6.96%,中国服务器市场占比 33.93%,变化+3.77pct。我们认为,随着人工智能所需算力扩大,未来中国服务器市场有望进一步扩大。

图34: 全球服务器市场规模及预测



资料来源: Statista, 中国银河证券研究院

图35: 中国服务器市场规模及预测

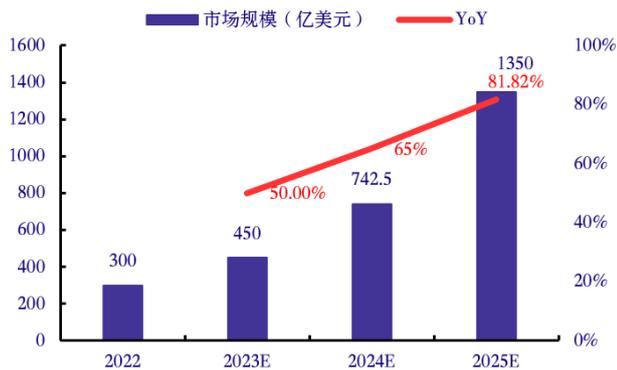


资料来源: Statista, 中国银河证券研究院

人工智能时代 AI 服务器优势凸显。随着 AI 技术升级应用, CPU 的串行处理架构不能满足 AI 时代的算力需求, 企业需要为人工智能、机器学习和深度学习建设全新的 IT 基础架构, 逐渐由 CPU 密集型转向搭载 GPU、FPGA、ASIC 芯片的加速计算密集型机构, 且越来越多地使用搭载 GPU、FPGA、ASIC 等加速卡的服务器。

全球 AI 服务器市场预计 2025 年达到 1350 亿美元, 未来三年 CAGR 有望超过 50%。研究机构 Aletheia 报告指出, 预估 AI 服务器市场规模将在 2024 年翻倍、2025 年达到 1350 亿美元, 是 2022 年规模的 4.5 倍。此外, 根据 TrendForce 预测, 2026 年全球 AI 服务器出货量将进一步提升, 2022 至 2026 年 CAGR 达 10.8%。

图36: 全球 AI 服务器市场规模及预测



资料来源: Aletheia, 中国银河证券研究院

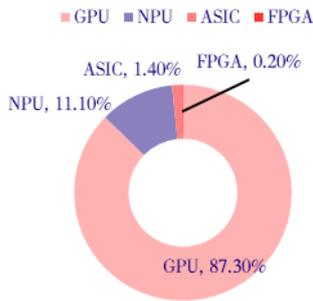
图37: 全球 AI 服务器厂商出货量及预测



资料来源: TrendForce, 中国银河证券研究院

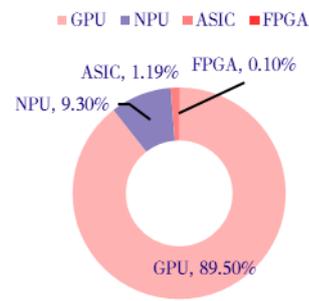
从搭载芯片种类上来看, 目前全球以 GPU 服务器为主流。据 IDC 统计, 2022 年全球 GPU 服务器出货量占比 87.3%, GPU 服务器销售额占比 89.5%。

图38：全球 AI 服务器出货量结构



资料来源：IDC，中国银河证券研究院

图39：全球 AI 服务器销售额结构



资料来源：IDC，中国银河证券研究院

中国 AI 服务器市场存量替换需求叠加增量需求，有望迎来量价齐升。根据中商产业研究院数据及我们预测，中国 AI 服务器市场预计 2026 年市场规模超千亿元，未来三年复合增长率 21.65%，预计 2026 年出货量 64.5 万台，未来三年复合增长率 15.26%。中国 AI 服务器受益于人工智能等相关新兴领域的应用以及“东数西算”政策下，云计算、超算中心的蓬勃发展，数据计算、存储需求呈几何级增长，算力需求持续释放，AI 服务器作为算力基础设备保持较快增速。1) **存量来看**：服务器平均寿命 3-5 年更换一次每年根据算力需求使用需求变化产生比较明显的更新需求。2) **增量来看**：伴随人工智能浪潮以及数字中国建设，未来对智能算力需求将持续爆发增长，且智能算力增长速度远超算力总体增速，中国 AI 服务器市场将迎来爆发增长，占比将逐步提升。2021-2026 年我国 AI 服务器市场规模由 350 亿元增长至 1089.4 亿元，2021 至 2026 年 CAGR 为 20.83%。

图40：中国 AI 服务器市场规模及预测



资料来源：《2023 年中国 AI 服务器市场前景及投资研究报告》（中商产业研究院），中国银河证券研究院

图41：中国 AI 服务器出货量及预测

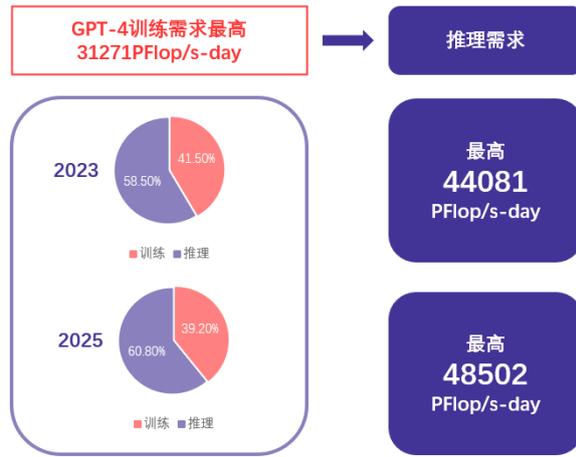


资料来源：《2023 年中国 AI 服务器市场前景及投资研究报告》（中商产业研究院），中商产业研究院，中国银河证券研究院

2. Sora 等多模态加速应用端落地，推理服务器需求激增

多模态对推理算力需求指数级增长，推理服务器占比将持续提升。根据 IDC 预测，2023 年 AI 服务器训练需求占比达 41.5%，随着大模型的应用，该比例在 2025 年将降低至 39.2%；将 GPT-4 的推算结果作为训练需求，进一步推算 2023、2025 年推理需求最高达 44081、48502PFlop/s-day。

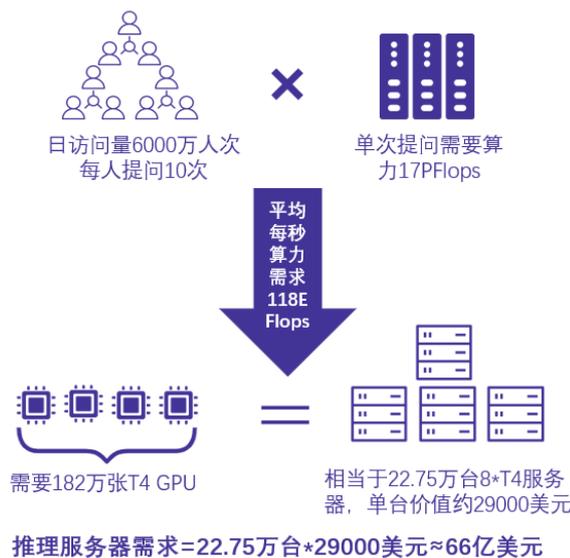
图42：2023年、2025年推理需求预测



资料来源：IDC，ARK 投资，浪潮信息，中国银河证券研究院

单个 AI 应用如 ChatGPT 可以带动推理算力 66 亿美元需求。假设平均针对 20 字的提问生成 200 字的响应，对应 267tokens，根据 OneFlow 的数据和《Scaling Laws for Neural Language Models》，在推理过程中每个 token 的计算成本约为 $2 \times N$ Flops，其中 N 为模型参数数量，则在 ChatGPT4 一万亿参数中每个 token 需算力 2 万亿 Flops。假定 GPT-4 训练期间 FLOPS 利用率为 32%，则每人每次提问需要算力： $2 \text{ 万亿} \times 267 \text{ tokens} / 32\% = 17 \text{ PFlops}$ 。据官网 9 月数据，ChatGPT 目前拥有超过 1 亿用户，每月产生 18 亿次访问量，假定每日访问量为 6000 万人次，每人提问 10 次，且假设一天平均分布，则每秒算力需求为 118EFlops，目前 AI 推理使用的主流 GPU 是 T4，提供混合精度算力 65TFlops，则需要 182 万个 T4GPU 可满足单日访问量，对应 22.75 万台 8*T4 服务器，一台 8*T4 服务器的价格约为 29000 美元，则目前来看推理服务器的需求在 66 亿美元。

图43：推理服务器需求推算



资料来源：《Scaling Laws for Neural Language Models》(Jared Kaplan 等)、NVIDIA 官网、OpenAI，中国银河证券研究院

Sora 训练一次所需算力或可达到 2.6×10^{24} Flops，相当于 GPT-3 175B 的 8.2 倍。多模态大模型在训练端算力需求通常在计算大语言模型算力需求通常与参数量及 token 数量成正比，而 Sora 大模型中可以将 Patch 类比作大语言模型中 token，基于大语言模型计算

算力需求方法框架及以下三大假设，对 Sora 算力需求进行分析测算。

假设一：Sora 训练数据集为 60 亿张图片，分辨率为 1980×1024 ；3500 万个视频，每个视频平均时长为 30 秒，分辨率为 1980×1024 ，帧率为 60FPS。

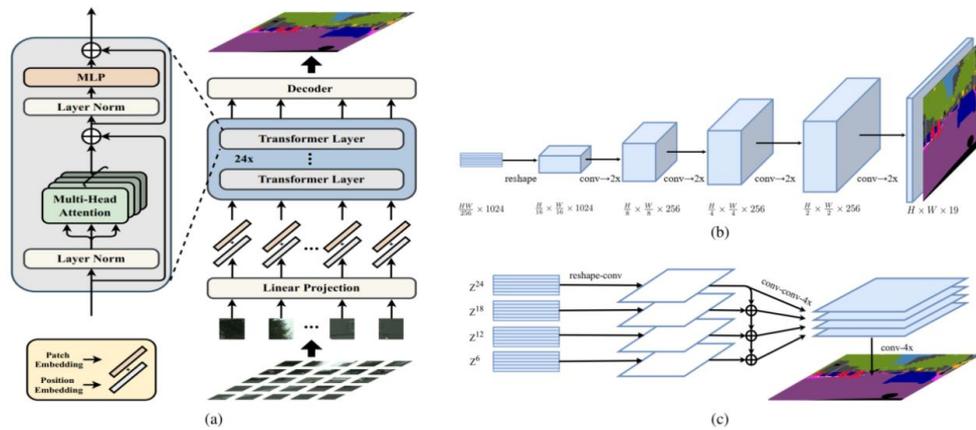
根据阿里联合浙江大学、华中科技大学提出的文生视频模型 I2VGen-XL，研究人员收集了大约 3500 万单镜头文本-视频对和 60 亿文本-图像对来优化模型。我们暂且保守假设 Sora 训练数据集与 I2VGen-XL 相同，同时二维向量空间图片表示为 $H \times W \times C$ （其中 H 为长度，W 为宽度，C 为 RGB 颜色通道数，假设 $C=3$ ）。

我们估算 Sora 训练数据集中视频类数据 Patch 规模= $3500 \times 10^4 \times 60 \times 30 \times 3=1.89 \times 10^{11}$ ；图片类数据 Patch 规模= $60 \times 10^8 \times 1024 \times 1980 \times 3=3.65 \times 10^{16}$ ；**训练数据集总 Patch=图片类数据 Patch+视频类数据 Patch= 3.65×10^{16} 。**

假设二：Sora 中 Patch Size 为 16×16 ，将 Patch 转化为 token。

根据谷歌论文《AN IMAGE IS WORTH 16×16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE》，Transformer 的输入是一个序列，对于一张图像来说如果把每个像素点当作一个 token，那就会需要相当庞大的计算量，该文则将图像划分为 16×16 大小的一个个 Patch，然后将每个 Patch 当作一个 token 组成一串序列作为 Transformer 的输入，减少了计算成本。我们假设 Patch Size 为 16×16 ，通过将 Patch ($N \times P \times P \times C$) 转换为 token，N 大小为 $H \times W / (P \times P)$ ，每个 token 的大小为 $P \times P \times C$ ， $P=16$ ，通过计算得到 $\text{token} = 3.65 \times 10^{16} / (16 \times 16) = 1.43 \times 10^{14}$ 。

图44：Transformer 语义分割方法



资料来源：《Rethinking Semantic Segmentation from a Seq-to-Seq Perspective with Transformer》（Sixiao Zheng 等），中国银河证券研究院

假设三：Sora 模型参数为 30B，训练一次所需总算力=模型参数量 \times token 数量 $\times 3 \times 2$ 。

根据 OpenAI 论文，T5 模型由于采用编码器-解码器模型，在向前和向后传播的过程中只有一半 token 处于激活状态，而 BERT 与 GPT 基于 Transformer 的自然语言监督模型，每个 token 都处于活跃状态，而每个 token 都在向前传播过程中涉及一次加法和一次乘法，论文添加一个 $3 \times$ 的乘数来计算向后传递的计算量，故推出 GPT 模型所需算力：训练所需总算力=模型参数量 \times token 数量 $\times 3 \times 2 \times$ 训练轮数。通过上述公式我们计算得到 Sora 训练一次所需算力= $30 \times 10^8 \times 1.43 \times 10^{14} \times 3 \times 2=2.6 \times 10^{24}$ Flops。

图45：不同模型训练一轮所需算力需求

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》(Tom B. Brown 等)，中国银河证券研究院

根据上述测算，基于 Sora 参数量大概在 30 亿（待确认）水平，同时采用 I2VGen-XL 训练数据集水平进行估算，我们保守估计，Sora 训练一次所需算力或可达到 2.6×10^{24} Flops，相当于 GPT-3 175B 的 8.2 倍（测算采用参数和训练数据集规模会与实际有一定出入）。

假设四：Sora 模型训练不考虑利用及其他成本，大约需在 1 万张 A100 上训练 154 天。

单张 A100 算力为 19.5TFlops，暂时不考虑模型训练利用率及其他训练成本，如果在 10000 张英伟达 A100 进行训练，所需时间= $2.6 \times 10^{24} / (19.5 \times 10^{12} \times 10000) / (24 \times 60 \times 60) \approx 154$ 天。

3. “人工智能+”鼓励数字基础设施“适度超前”发展，国产服务器迎来曙光

2024 年 3 月 5 日，国务院总理李强在政府工作报告中提出，要大力推进现代化产业体系建设，加快发展新质生产力。同时，深化大数据、人工智能等研发应用，开展“人工智能+”行动，打造具有国际竞争力的数字产业集群。另一方面，报告中也提出，要深入推进数字经济创新发展，制定支持数字经济高质量发展政策，积极推进数字产业化、产业数字化，促进数字技术和实体经济深度融合。适度超前建设数字基础设施，加快形成全国一体化算力体系。

政府工作报告中首次提出“人工智能+”行动，数字基础设施建设未来将成为经济核心抓手。在政府工作报告中被首次提出的“人工智能+”行动，正是推进数字产业化、产业数字化的重要举措。另一方面，在实践中，政府和企业都需要加大对人工智能领域的研发投入，提高对我国人工智能产业发展中技术创新、跨界人才等资源的储备力度。

图46：算力对数字经济和 GDP 的影响



资料来源：IDC《2022-2023 全球算力指数评估报告》，中国银河证券研究院

“适度超前”建设算力基础设施，国产服务器大势所趋。报告中提出，适度超前建设数字基础设施，加快形成全国一体化算力体系。我们认为，“适度超前”一方面将驱动数字产业化加速发展，算力基础设施国产化提上日程，芯片、服务器国产化率将进一步提升，建议重点关注上游算力基础设施如国产芯片、国产 AI 服务器、光模块、液冷等细分赛道机会；另一方面，将持续推动产业数字化升级转型，算力基础设施将大范围赋能千行百业，降本增效，为新质生产力发展提供新动能。

出口禁令影响海外供应，AI 服务器国产化进程提速。2023 年 10 月 17 日，美国商务部工业和安全局（BIS）发布了针对芯片的出口禁令新规，更加严格的限制了中国购买重要的高端芯片。一方面，从 ChatGPT 面世以来，国内各企业和研究院在短短半年多的时间内先后推出了超过 130 款大模型，其中领跑玩家已经开始着手于将大模型应用于特定场景，打造爆款应用。另一方面，为了构筑算力底座，各地政府纷纷上马智算中心建设，铺设大数据时代的信息高速，推动产业创新升级，降低企业调用以大模型为代表的科技成果的成本。根据华为昇腾计算业务总裁张迪焯在 2023 世界人工智能大会上的揭示，大模型所需的算力相对于 2020 年预计将增长 500 倍，这个算力缺口正在不断扩大。

A800、H800 被禁后，英伟达继续推出新款芯片，单卡性能 H20 弱于昇腾 910b。2023 年 11 月 9 日，相关报道称英伟达已开发出针对中国市场的最新改良版系列芯片——HGXH20、L20PCle 和 L2PCle。最新三款芯片是由 H100 改良而来，就单卡性能而言 H20 弱于昇腾 910b。

表11：英伟达 H20、L20、L2 性能参数对比

性能参数	HGX H20	L20 PCIe	L2 PCIe
GPU 架构	NVIDIA Hopper	NVIDIA Ada Lovelace	NVIDIA Ada Lovelace
GPU 内存	96 GB HBM3	48 GB GDDR6 W/ ECC	24 GB GDDR6 W/ ECC
GPU 内存带宽	4.0TB/s	864 GB/s	300 GB/s
INT8 FP8 Tensor Core	296 296 TFLOPS	239 239 TFLOPS	193 193 TFLOPS
BF16 FP16 Tensor Core	148 148 TFLOPS	119.5 119.5 TFLOPS	96.5 96.5 TFLOPS
TF32 Tensor Core	74 TFLOPS	59.8 TFLOPS	48.3 TFLOPS
FP32	44 TFLOPS	59.8 TFLOPS	24.1 TFLOPS
FP64	1 TFLOPS	N/A	N/A
RT Core	N/A	Yes	Yes
MIG	Up to 7 MIG	N/A	N/A
L2 Cache	60 MB	96 MB	36 MB
Media Engine	7 NVDEC 7NVJPEG	3 NVENC(+AV1) 3 NVDEC 4 NVJPEG	2 NVENC(AV1) 4 NVDEC 4 NVJPEG
Power	400 W	275W	TBD
Form Factor	8-way HGX	2-slot FHFL	1-slot LP
Interconnect	Pee Gen5 x16: 128 GB/s NVLink: 900GB/s	PCIe Gen4 x16: 64 GB/s	PCIe Gen4 x16: 64 GB/s
上市时间	生产排期: Nov 2023 量产: Dec 2023	生产排期: Nov 2023 量产: Dec 2023	生产排期: Dec 2023 量产: Jan 2024

数据来源：NVIDIA 官网，中国银河证券研究院

华为昇腾芯片为 AI 体系提供强大算力，昇腾 910b 单卡性能接近英伟达 A100。华为昇腾芯片是华为发布的两款人工智能处理器，包含昇腾 310 用于推理和 910 用于训练，均采用自家的达芬奇架构。昇腾 910 是一款高性能 AI 芯片，采用了 7nm 工艺制程，集成了数千个达芬奇核心，能够提供高达 256TOPS 的算力，在业界其算力处于领先水平。昇腾 310 是一款入门级 AI 芯片，采用了 12nm 工艺制程，集成了数百个达芬奇核心，能够提供高达 8TOPS 的算力，适合用于边缘计算和物联网等应用场景。**2023 年科大讯飞与华为昇腾启动专项攻关，合力打造我国通用人工智能新底座，让国产大模型架构在自主创新的软硬件基础之上，当前华为昇腾 910B 能力已经基本做到可对标英伟达 A100。**

表12：华为昇腾 910b 与英伟达 A100 参数性能对比

性能指标	昇腾 910b	英伟达 A100
架构	自研华为达芬奇架构	NVIDIA Ampere 架构
峰值算力 FP16	376TFlops	312TFlops
峰值算力 FP32	94TFlops	19.5TFlops
GPU 显存	64GBHBM2E	80GBHBM2E
CPU-NPUPCIe 规格	PCIe5.0 × 16512GB	PCIe4.0 × 16256GB
NPU-NPU 带宽	392GB/s (HCCL)	400GB/s (Nvlink)
芯片 RDMA 出口带宽	200GB/sRoCE 芯片直出网口	1000Gb-2000Gb/sIB 需要通过 IB 网卡扩展出口
TDP	400W	300W
制程工艺	7nm	7nm

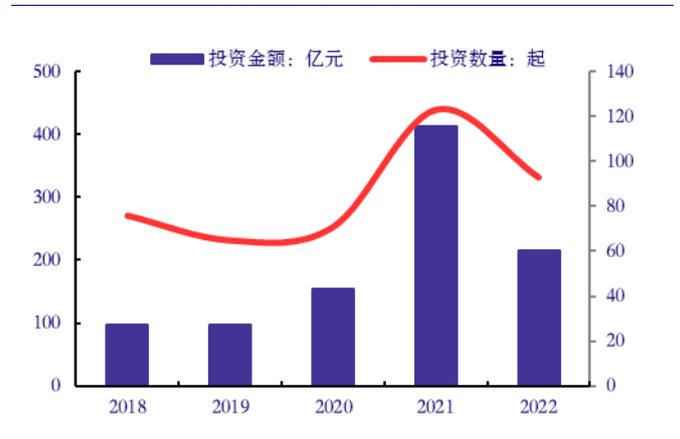
资料来源：华为海思官网、NVIDIA 官网，中国银河证券研究院

华为昇腾生态打开市场空间，国产算力产业链有望持续受益。我们认为，国内第一批大模型厂商使用的基本都是英伟达 A100、A800 的芯片，因为英伟达构建了完善的 CUDA 生态，贸然换生态，意味着学习成本、试错成本、调试成本都会增加。目前华为基于“鲲鹏+昇腾”双引擎正式全面启航计算战略，打造算力底座，未来国产替代趋势下，华为昇腾市场份额将不断提升，产业链细分赛道上市公司有望持续受益。**15%到全国产业化是大概率事件，国产化空间巨大。根据 IDC 数据，2022 年中国 AI 芯片出货量约 109 万张，其中英伟达市占率约为 85%，华为在内的国产 AI 芯片市占率约为 15%，国产化仍有很大空间。**

图47：中国 AI 芯片市场规模



图48：2018-2022 年中国 AI 芯片投资情况

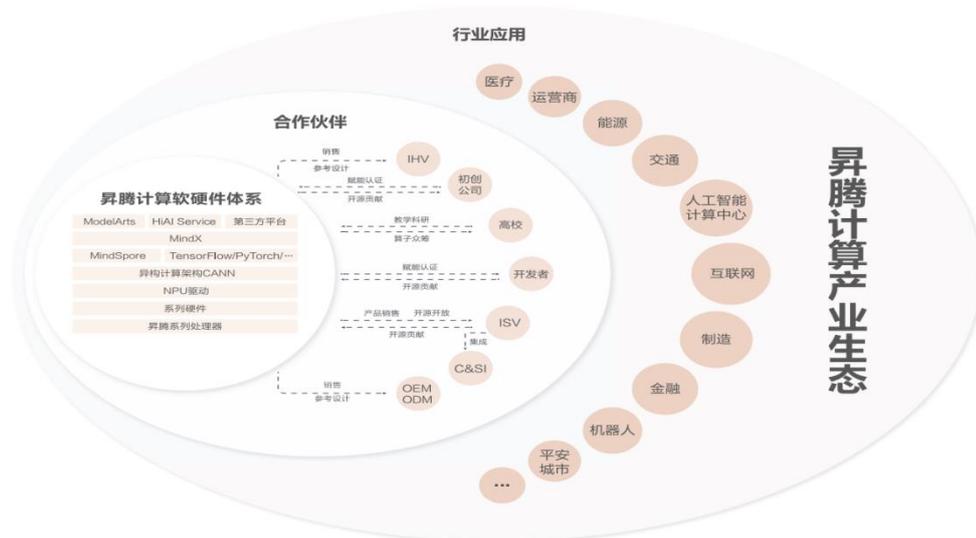


资料来源：中商产业研究院，中国银河证券研究院

资料来源：中商产业研究院，中国银河证券研究院

昇腾计算产业链是基于昇腾系列处理器和基础软件构建的全栈 AI 计算基础设施、行业应用及服务，包括昇腾系列处理器、系列硬件、CANN、AI 计算框架、应用使能、开发工具链、管理运维工具、行业应用及服务全产业链。

图49：华为昇腾计算产业生态



资料来源：《昇腾计算产业发展白皮书》（华为、中国信通院联合发布），中国银河证券研究院

百度等率先开始采购昇腾 910B，释放多重积极信号。据相关报道，百度为 200 台服务器向华为订购了 1600 颗昇腾 910BAI 芯片，作为英伟达 A100 的替代品，订单总价值约 4.5 亿人民币，预计今年年底前完成交付，截至 10 月已交付约 60%，同时 360 集团创始人周鸿祎表示，360 也采购了华为 1000 片左右的 AI 芯片。**我们认为，此次采购意义重大，虽然此订单规模相较过去从英伟达采购的数千颗芯片较小，但是此次采购证明国产昇腾 910B 可以满足大模型训练需求，伴随百度、360 等互联网大厂竞相采购，有望掀起互联网行业乃至千行百业采购浪潮。**

华为昇腾服务器，国产算力之王。搭载华为昇腾系列 AI 芯片和业界主流异构计算部件的算力集群，具有超强计算性能，可以广泛用于中心侧 AI 推理、深度学习模型开发和场景训练，根据训练需求不同可以分为 AI 训练服务器和 AI 推理服务器。

AI 训练服务器：1) Atlas800 训练服务器（型号：9000）是基于华为鲲鹏 920+昇腾 910 处理器的 AI 训练服务器，具有超强算力密度、超高能效与高速网络等特点。2) Atlas800 训练服务器（型号：9010）是基于 Intel 处理器+华为昇腾 910 芯片的 AI 训练服务器，具有超强算力密度、高速网络带宽等特点。这两款服务器广泛应用于深度学习模型开发和训练，适用于智慧城市、智慧医疗、天文探索、石油勘探等需要大算力的行业领域。

AI 推理服务器：1) Atlas800 推理服务器（型号：3000）是基于昇腾 310 芯片的推理服务器，最大可支持 8 个 Atlas300I 推理卡，提供强大的实时推理能力，广泛应用于中心侧 AI 推理场景。2) Atlas800 推理服务器（型号：3010）是基于 Intel 处理器的推理服务器，最多可支持 7 个 Atlas300I 推理加速卡，支持 560 路高清视频实时分析，广泛应用于中心侧 AI 推理场景。

表13：华为服务器对比

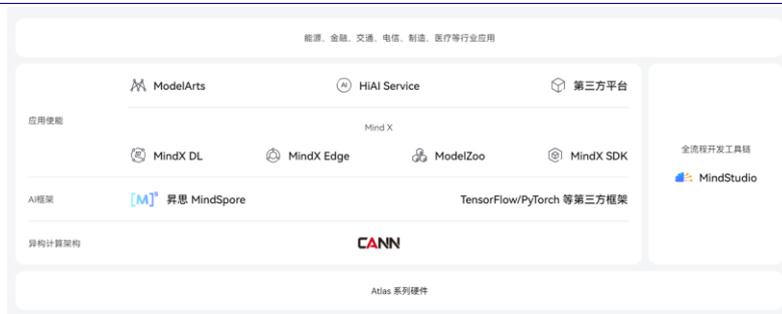
产品名称	算力	AI 加速卡	CPU	CPU 内存	PCIe	最大功耗
------	----	--------	-----	--------	------	------

Atlas800 训练服务器 型号: 9000	2.24PFLOPSFP16 1.76PFLOPSFP16	8*昇腾 910	4*鲲鹏 920	最多 32 个 DDR4 内存插槽, 支持 RDIMM 内存速率最高 3200MT/s, 单根内存条 容量支持 16GB/32GB/64GB	最多支持 2 个 PCIe4.0 扩展插槽	5.6kW
Atlas800 训练服务器型 号: 9010	2.24PFLOPSFP16 1.76PFLOPSFP16	8*昇腾 910	2*IntelV5CascadedLake 处理器	最多 24 个 DDR4 内存插槽, 支持 RDIMM	最多支持 6 个 PCIe3.0x8 扩展插槽	5.6kW
Atlas800 推理服务器型 号: 3000	最大 704TOPSINT8	最大支持 8 个 Atlas300I 推理卡	2*鲲鹏 920	32 个 DDR4 内存插槽, 最高 3200MT/s	最多支持 9 个 PCIe4.0PCIe 接口, 其中 1 个为 RAID 扣卡专用的 PCIe 扩展槽位, 另外 8 个为标准的 PCIe 扩展槽位	67W
Atlas800 推理服务器型 号: 3010	最大 616TOPSINT8	最大支持 7 个 Atlas300I 推理卡	1/2 个 IntelXeonSPSkylake 或 CascadeLake 处理器	24 个 DDR4 内存插槽, 最高 3200MT/s	10 个 PCIeGen3.0 接口(含 1 个 RAID 控制卡+1 个灵活 LOM)	205W

数据来源: 华为海思, 中国银河证券研究院

软件端: 昇腾全栈 AI 软硬件平台, 面向“端、边、云”的全场景 AI 基础设施。AI 处理器和基础软件构建 Atlas 人工智能计算解决方案, 包括 Atlas 系列模块、板卡、小站、服务器、集群等丰富的产品形态, 打造面向“端、边、云”的全场景 AI 基础设施方案, 覆盖深度学习领域推理和训练全流程, 充分释放硬件性能。

图50: 昇腾全栈 AI 软硬件平台



资料来源: 华为昇腾官网, 中国银河证券研究院

昇思 MindSpore 崛起, 打造国产 AI 开发新生态。昇腾 910 配套的 AI 开源计算框架 MindSpore 更方便 AI 科学家和工程师使用, 该框架可满足终端、边缘计算、云全场景需求, 能够更好地保护数据隐私。昇腾在硬件上为合作伙伴提供 Atlas 昇腾计算模组和计算卡, 发展 20+硬件合作伙伴(整机、工控机等), 15 家一体机伙伴;在软件上开源昇思 MindSporeAI 框架, 社区开源模型 400+, 下载量超 390 万+; 生态发展上, 昇腾采用行业+教育的方式, 行业打通 1000+合作伙伴, 认证解决方案 2000+, 超过 110 所高校开设昇腾 AI 课程, 有着 120 万+开发者。

图51: 华为昇思 MindSpore 框架



资料来源: 华为昇思官网, 中国银河证券研究院

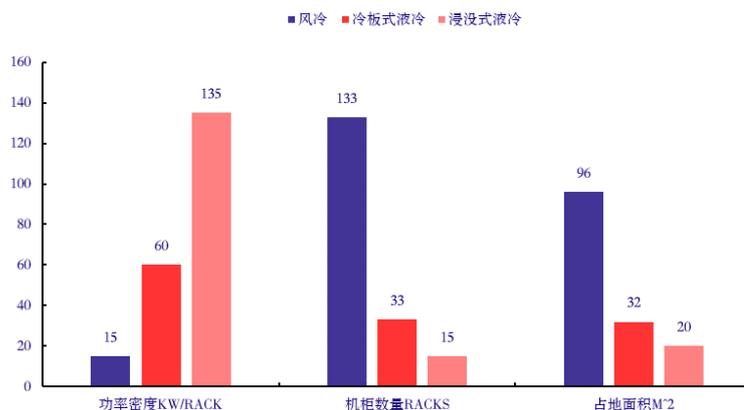
4、高性能计算对液冷需求爆发式增长, 液冷服务器打开千亿市场空间

数据中心 PUE 要求愈发严苛。随着云计算、大数据、人工智能、元宇宙等信息技术的快速发展和传统产业数字化的转型, 数据呈现几何级增长, 算力和硬件部分能耗也在持续增加, 而在“双碳”政策的持续推进下, 国家、地方政府、企业层面均在积极推动绿色低碳转型和可持续发展, 通讯领域对数据中心节能降耗要求越来越严格。

液冷未来有望逐渐替代风冷, 成为 AI 服务器、数据中心标配。A1 训练及推理应用、超算等高算力业务需求持续推升, 由此带来的芯片性能需求、服务器功率需求不断提高, 场景侧, 以英伟达 DGXA100640GB 服务器为例, 系统最大功率为 6.5KW, 传统风冷无法做到及时散热, 相比之下, 液体比热容为空气的 1000-3500 倍, 导热性能是空气的 15-25 倍, 利用自然冷却显著降低耗电量, 使得液冷成为风冷的不二选择。

我们认为, 人工智能浪潮下, 对算力需求进一步提升, 液冷预计将成为最优冷却方案, 未来中国液冷服务器市场有望进一步打开竞争格局, 产业相关上市公司将受益。目前, 中国液冷服务器普及率不足 5%, 普及率并不高。受制于: 1) 数据中心国家 PUE 标准收紧; 2) 受制于面积等因素, 机柜密度逐渐提升; 3) 温度过高, 芯片故障率升高等客观因素, 未来液冷服务器将成为调和快速的算力需求与有限数据中心承载力的共识方案。

图52: 风冷与液冷散热能力对比



资料来源: 中兴通讯《液冷技术白皮书》, 中国银河证券研究院

表14: 全国主要数据中心 PUE 要求

城市	年平均气温℃	数据中心 PUE 要求
北京	12.3	年能源消费量小于 1 万吨标准煤的项目 PUE 值不应高于 1.3; 年能源消费量大于等于 1 万吨标准煤, 且小于 2 万吨标准煤的项目, PUE 值不应高于 1.25; 年能源消费量大于等于 2 万吨标准煤且小于 3 万吨标准煤的项目, PUE 值不应高于 1.2; 年能源消费量大于等于 3 万

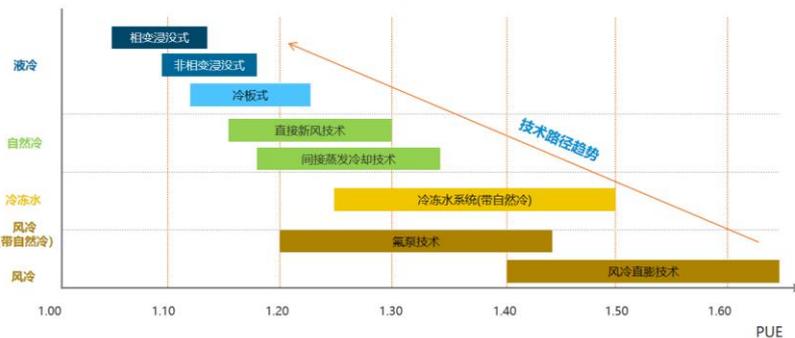
请务必阅读正文最后的中国银河证券股份有限公司免责声明。

		吨标准煤的项目，PUE 值不应高于 1.15； 14 < PUE ≤ 18，每度电加价 ¥0.2； PUE > 18，每度电加价 ¥0.5
上海	16.6	到 2024 年，新建大型及以上数据中心 PUE 降低到 13 以下，起步区内降低到 1.25 以下。推动数据中心升级改造，改造后的 PUE 不超过 1.4。
广东	22.6	新增或扩建数据中心 PUE 不高于 13，优先支持 PUE 低于 1.25 的数据中心项目，起步区内 PUE 要求低于 1.25
浙江	16.5	到 2025 年，大型及以上数据中心电能利用效率不超过 13，集群内数据中心电能利用效率不得超过 1.25
江苏	15.5	到 2023 年底，全省数据中心机架规模年均增速保持在 20% 左右，平均利用率提升到 65%，全省新型数据中心比例不低于 30%，高性能算力占比达 10%，新建大型及以上数据中心电能利用效率（PUE）降低到 1.3 以下，起步区内电能利用效率不得超过 1.25
山东	14.7	自 2020 年起，新建数据中心 PUE 值原则上不高于 1.3，到 2022 年年底，存量改造数据中心 PUE 值不高于 1.4。到 2025 年，实现大型数据中心运行电能利用效率降到 1.3 以下。优先支持 PUE 值低于 1.25，上架率高于 65% 的数据中心新建、扩建项目
青岛	12.7	新建 13，至 2022 年存量改造 14
重庆	18.4	到 2025 年，电能利用效率（PUE）不高于 13。集群起步区内 PUE 不高于 1.25。
四川	15.3	到 2025 年，电能利用效率（PUE）不高于 13。集群起步区内 PUE 不高于 1.25。各市（州）要充分发挥已建在建数据中心作用，除天府数据中心集群外，区域内平均上架率未达到 60%、平均 PUE 值未达到 1.3 及以下的，原则上不得新建数据中心。
内蒙古	4.3	到 2025 年，全区大型数据中心平均 PUE 值降至 13 以下，寒冷及极寒地区力争降到 1.25 以下，起步区做到 1.2 以下
宁夏	9.5	到 2025 年，建成国家（中卫）数据中心集群，集群内数据中心的平均 PUE ≤ 1.15，WUE ≤ 0.8，分级分类升级改造国家（中卫）数据中心集群外的城市数据中心，通过改造或关停，到 2025 年，力争实现 PUE 降至 1.2 及以下。
贵州	15.5	引导大型和超大型数据中心设计 PUE 值不高于 1.3；改造既有大型、超大型数据中心，使其数据中心 PUE 值不高于 1.4。实施数据中心减量替代，根据 PUE 值严控数据中心的能源消费新增量，PUE 低于 1.3 的数据中心可享受新增能源消费量支持。

资料来源：中兴通讯《液冷技术白皮书》，中国银河证券研究院

液冷服务器是大势所趋，数据中心 PUE 可降至 1.25 以下。算力的持续增加，意味着硬件部分的能耗也在持续提升；在保证算力运转的前提下，只有通过降低数据中心辅助能源的消耗，才能达成节能目标下的 PUE 要求。

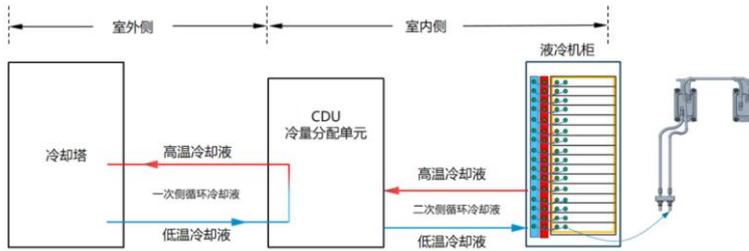
图53：制冷技术 PUE 对比



资料来源：中兴通讯《液冷技术白皮书》，中国银河证券研究院

冷板式液冷服务器与浸没式相变服务器为两大主流液冷服务器。冷板式液冷服务器技术利用工作流体作为中间热量传输的媒介，将热量由热区传递到远处再进行冷却。在该技术中，工作液体与被冷却对象分离，工作液体不与电子器件直接接触，而是通过液冷板等高效热传导部件将被冷却对象的热量传递到冷媒中。

图54：冷板式液冷服务器散热原理



资料来源：中兴通讯《液冷技术白皮书》，中国银河证券研究院

该技术将冷却剂直接导向热源，同时由于液体比空气的比热大，散热速度远远大于空气，因此制冷效率远高于风冷散热，每单位体积所传输的热量即散热效率高达 1000 倍。该技术可有效解决高密度服务器的散热问题，降低冷却系统能耗而且降低噪声。

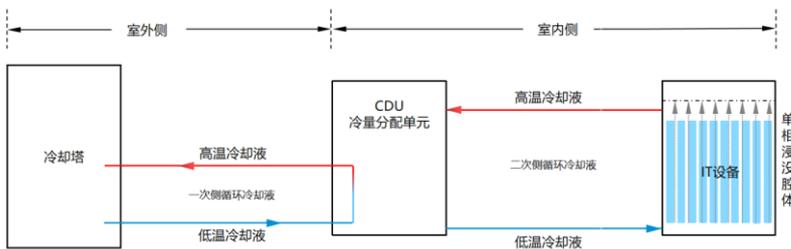
图55：曙光冷板式液冷服务器



资料来源：曙光数创，中国银河证券研究院

浸没式液冷服务器又可以分为单相浸没式液冷服务器和两相浸没式液冷服务器。浸没式相变换热液冷系统采用进口环保专用冷媒，具有不导电、无闪点、无腐蚀性、无毒性的特性，利用环保冷媒良好的热物理特性，通过控制系统物理参数，利用冷媒工质的气化潜热转移服务器内部热量，极大提高了系统的换热效率，同时保留了高端热源的能量品位。此冷媒较传统冷媒，在系统压力较低的情况下即可实现 50°C ~ 60°C 的蒸发温度，无须利用压缩机进行机械制冷，从而使室外机组的全年自然冷却工作方式成为可能。

图56：单相浸没式液冷服务器散热原理



资料来源：中兴通讯《液冷技术白皮书》，中国银河证券研究院

作为中国液冷服务器第一的曙光数创，目前浸没式液冷服务器技术领先。1) 整机功耗：全浸没方案，无风扇设计，风扇功耗降低为 0。2) 终极的噪音指标：区别于传统风冷机房，全浸没机房噪音控制在 35dB 以下。3) 终极的功率密度：高密度配置，轻松实现整机柜功率 200kW。4) 终极的 PUE 指标：直接利用高品位完成热量转移，可实现 PUE 低至 1.01-1.02。

图57：曙光数创全浸没式液冷服务器



资料来源：曙光数创，中国银河证券研究院

2023H1 中国液冷服务器市场同比增长近 3 倍。根据 IDC 发布的《中国半年度液冷服务器市场（2023 上半年）跟踪》报告数据显示，2023 上半年中国液冷服务器市场规模达到 6.6 亿美元，同比增长 283.3%，预计 2023 年全年将达到 15.1 亿美元。IDC 预计，2022-2027 年，中国液冷服务器市场年复合增长率将达到 54.7%，2027 年市场规模将达到 89 亿美元。

图58：2022 年-2027 年中国液冷服务器市场规模预测



资料来源：IDC、中商产业研究院，中国银河证券研究院

（二）算法：多模态大模型 C 端、B 端加速渗透，未来从云端走向终端

1. 边云算法协同发展，推动大模型终端侧落地

据 Mordor Intelligence 预计，云人工智能市场规模将从 2023 年的 510.4 亿美元增长到 2028 年的 2074 亿美元，预测期内（2023-2028 年）复合年增长率为 32.37%。随着企业数量的不断增加以及企业之间的竞争，企业正在积极尝试将人工智能技术与其应用、分析、业务和服务相集成。此外，将公司致力于降低运营成本以提高利润率，推动云端人工智能快速发展，进一步推动预测期内的市场规模增长。

图59：云端 AI 预计市场规模



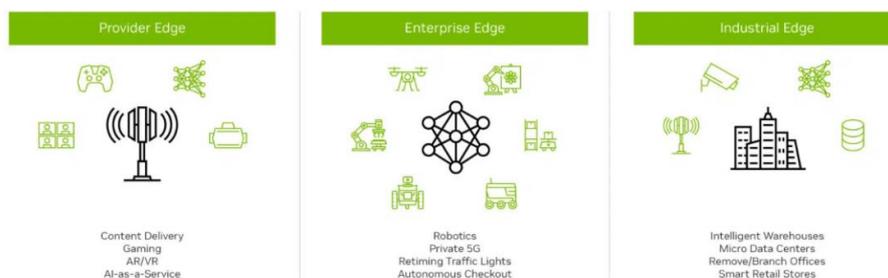
资料来源：Mordor Intelligence，中国银河证券研究院

大模型时代，从云到端（边缘侧）算法不断优化升级，其中边缘智能优化方向包括五大方面，包括边云协同、模型分割、模型压缩、减少冗余数据传输以及轻量级加速体系结构。其中，边云协同、模型分割、模型压缩能够减少边缘智能对于计算、存储和设备的需求；减少冗余数据传输以改善传输效率，降低网络资源浪费；轻量级加速体系结构将在硬件和应用方面支持边缘计算效率提升。边云协同云计算和边缘计算相结合，将数据和计算资源分布在云端和边缘设备，实现数据的高效处理和传输。云边协同优势明显，1）充分利用云端强大的计算资源；2）在需要实时响应或者断网情况下可借助本地设备完成任务；3）保证数据安全性。

云端处理人工智能的成本高昂，致使发展高效“云端处理+边缘设备”混合人工智能处理以落地终端的重要性凸显。混合架构具有不同的卸载选项，以便根据模型和查询复杂性等因素在云和设备之间分配处理。具有超过 10 亿个参数的 AI 模型已经在手机上运行，其性能和准确度水平与云类似，并且具有 100 亿个或更多参数的模型预计在不久的将来在设备上运行。

终端侧大模型主要用于推理，将拓展边缘设备应用。混合 AI 将提振 B 端 AI 产品及服务实力，面向行业开发多场景应用、优化 AI 交互体验感、增强工作自动化程度，从而提升工作效率；混合 AI 将优化生成式 AI 的功能性及可用性，推动企业研发以全面化终端 AI 功能，拓展 C 端软件及硬件产品的多样性。混合人工智能方法几乎适用于所有生成式人工智能应用程序和设备领域，包括智能手机、笔记本电脑、汽车、机器人和智能物联网终端（AIoT）等。

图60：云+边缘落地方向



资料来源：NVIDIA，中国银河证券研究院

2. 多模态大模型角逐日益激烈，自研模型本地部署是大方向

3月15日，工信部旗下赛迪工业和信息化研究院（集团）四川有限公司发布的《2024中国人工智能多模态大模型企业综合竞争力20强研究报告》综合评估大模型实力，衡量标准涵盖经济抗衡力、行业影响力、技术竞争力、商业角逐力、营收规模、数据优势、模型规模、算力能级等十余个指标。排名前五的依次是，腾讯混元大模型、科大讯飞星火大模型、阿里巴巴通义千问大模型、华为盘古大模型和智谱华章智普清言大模型。

OpenAI的“文生视频”大模型Sora的发布，标志着多模态大模型在处理和生成与多模态（包括文本、图像、音频和视频）相关数据方面的能力。多模态大模型可理解和生成不同形式的复杂数据，其重要性骤增，并将应用于包括数字人、游戏、广告商拍、社交媒体、智能营销、教培、健康医疗、新闻媒体及艺术创作等多个行业。据赛迪研究院分析，从多模态大模型的技术场景来看，20强企业中有15%的企业具备文生视频能力；从应用场景来看，相较语言大模型赋能行业，多模态大模型更加偏重落地场景。

表15：AI多模态大模型企业20强

排名	企业名称	大模型名称
1	腾讯	混元
2	科大讯飞	星火
3	阿里巴巴	通义千问
4	华为	盘古
5	智谱华章	智普清言
6	百度	文心一言
7	网易	丹青
8	360	360智脑
9	昆仑万维	天工
10	万兴科技	天幕
11	商汤科技	日日新
12	虹软科技	ArcMuse
13	中国电信	星辰
14	抖音	Boximator
15	因赛集团	InsightGPT
16	第四范式	式说
17	京东	言犀
18	开普云	开悟
19	小冰公司	小冰
20	硅基智能	炎帝

资料来源：赛迪四川，亿欧智库，中国银河证券研究院

随着企业自研大模型的发展，轻量级和本地部署已成为趋势。轻量级加速体系结构在大模型部署终端侧应用方面至关重要。应用方面，英伟达在Computex2023大会上发布模组JetsonAGXORIN，算力高达每秒275万亿次TOPS，性能是上一代产品的8倍，同时发布的工业级模组可以帮助客户在农业、建筑、能源、航空航天、卫星等领域部署边缘AI的需求。芯片方面，Fortinet、ARM分别推出适用于边缘计算的芯片FortiSP5和Hercules、Apollo，FortiSP5安全芯片防火墙性能大幅提升17倍，功耗相比业内高性能通用CPU降低88%，为边缘计算提供安全、高效、低成本的数据处理服务；Hercules和Apollo性能最高提升480倍，同时具有低功耗、小体积、高安全性等优点，使终端设备能够更好地集成AI能力，实现更广泛的边缘计算应用。

3. 多模态大模型赋能终端，持续渗透C端及B端

3.1 AI+手机：大模型接入手机，终端功能多样化

3月8日，谷歌发布MediaPipe LLM Inference API，新版本使大语言模型（LLM）跨平台在设备上运行。考虑到LLM的内存和计算需求，新版本内存和计算需求比传统的设

备端模型大 100 多倍。整个设备端堆栈的优化使这成为可能，包括新的运算、量化、缓存和权重共享。该版本简化 Web 开发人员的设备端 LLM 集成，支持 Web、Android 和 iOS，最初支持四个公开可用的 LLM：Gemma、Phi 2、Falcon 和 Stable LM。研究和开发人员可灵活在设备上对公开可用的 LLM 模型进行原型设计和测试。

Android 端, MediaPipe LLM Inference API 仅供实验和研究使用。可通过 Android AICore 在设备上使用 Gemini API 或 Gemini Nano。AICore 是 Android 14 中引入的全新系统级功能，旨在为高端设备提供 Gemini 驱动的解决方案，包括与最新的 ML 加速器、LoRA 适配器和安全过滤器集成。

初始版本支持以下四种模型架构，与其架构兼容的任何模型都可与 MediaPipe LLM Inference API 一起使用。

图61：MediaPipe LLM Inference API 适配四种模型架构

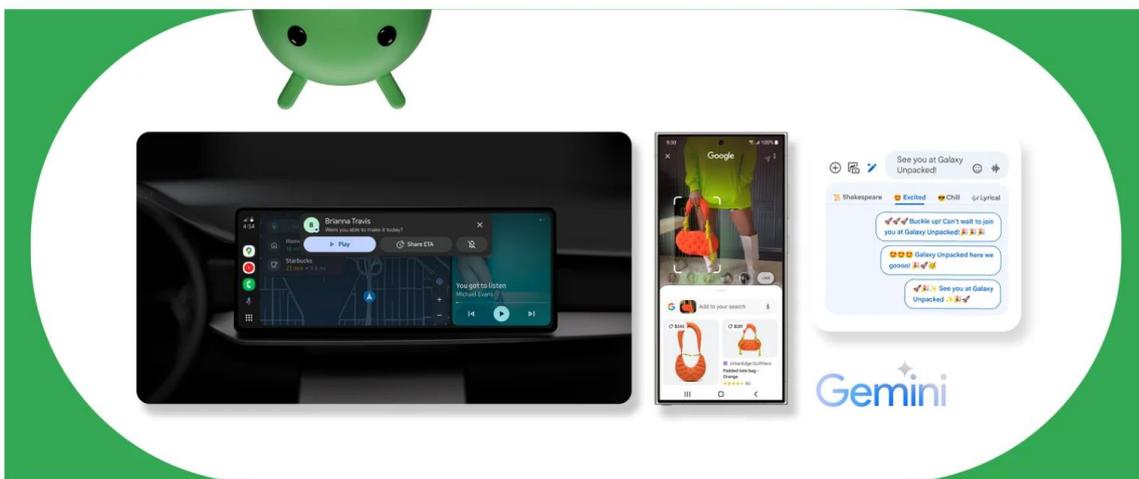
Model	Parameter Size
Falcon 1B	1.3 Billion
Gemma 2B	2.5 Billion
Phi 2	2.7 Billion
Stable LM 3B	2.8 Billion

资料来源：NVIDIA，中国银河证券研究院

先前，谷歌宣布其与三星的合作，Galaxy S24 用户将能通过三星构建的应用程序和服务访问 AI 模型 Gemini。借助 Gemini Pro 可与三星应用程序进行交互。最新 Galaxy S24 系列，三星的 Notes、Voice Recorder 和 Keyboard 应用程序可使用 Gemini Pro 优化摘要功能。例如，可录制讲座并快速获得关键部分摘要。借助 Google “文生图” 扩散技术 Imagen 2，可在 Galaxy S24 Gallery 应用程序中使用 Generative Edit 访问照片编辑功能。

Galaxy S24 系列还将内置 Gemini Nano，可在 Google Messages 中启用新功能，并确保数据存储在三星设备上。三星将成为首批测试 Gemini Ultra 的公司之一，并将在 2024 年内向开发人员和企业客户提供。

图62：三星手机接入谷歌 AI 大模型 Gemini



资料来源：Google，中国银河证券研究院

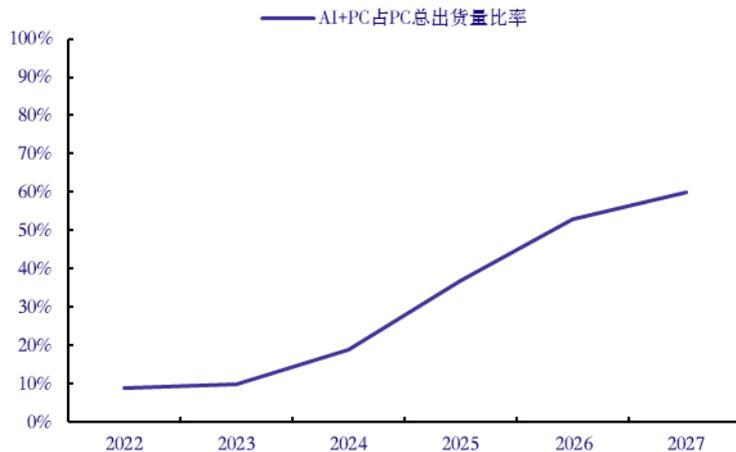
3.2 AI+PC: MaaS To C 终端智能化需求待开发, AI 手机成 C 端落地第一场景, 引发换机需求

C 端需求仍处起步阶段, 终端侧产品研发有望拓展并覆盖更多 AI 智能设备产品品类、持续优化平台性能, 并优先在 AI 手机设备落地, 进而引发换机需求。终端部署 AI 将提高终端设备易用性, 并加强数据隐私及安全, 降低开发者开发门槛。AI 终端部署集成 AI 功能, 可完成在不同智能设备上的应用, 进而提升用户生活品质及工作效率, 提振终端设备智能化需求。

2023 年 7 月, 高通发布公告称将和 Meta 合作, 2024 年推出“手机版 Llama2”, 高通已将 AI 边缘计算定为未来发展方向。大模型需要同时在云端和边缘终端上运行, 高通和 Meta 此举预计将推动 AIGC 加速推广到 C 端市场; 边缘端算力需求将成为算力侧未来新的增长点。高通和 Meta 正在努力优化 Meta 的 Llama2 大型语言模型的执行, 而不只依赖于云服务。在智能手机、PC、VR/AR 耳机和汽车等设备上运行 Llama2 等生成人工智能模型, 使开发人员能够节省云成本, 并为用户提供更私密、可靠和个性化的体验。因此, 高通计划提供基于 Llama2 的人工智能设备上实现, 以创建新的人工智能应用程序。这将允许客户、合作伙伴和开发人员构建用例, 如智能虚拟助理、生产力应用程序、内容创建工具、娱乐等。这些新的设备端人工智能体验将由骁龙 8Gen3 芯片驱动, 可以在没有网络连接的区域工作, 包括飞行模式。

据 Canalys 预计, 2023 年 AIPC 占 PC 总出货量比率大约为 10%。2024 年起 AIPC 将加速普及, 预计至 2027 年 AIPC 占 PC 总出货量将达到 60%; AI 产品研发引发换机率上升的背景下, 终端 AI 渗透率有望逐年攀升。

图63: 2024 年起全球将加速普及 AI+PC



资料来源: Canalys, 中国银河证券研究院

3.3 AI+机器人: 大模型上身机器人, 预示机器人 AGI 时代的到来

Figure 于 3 月 13 日发布 Figure 机器人的进展, 公司公布了一段 Figure 机器人与人类互动的视频。Figure 团队联合 OpenAI, Figure01 展现出与人类对话的卓越能力。视频中 Figure01 展示的能力使用端到端神经网络, 不涉及任何远程操作; 可做到描述其视觉体验、规划未来行动、反思记忆并口头解释推理过程。我们认为, Figure01 的发布将影响未来人

机交互模式，大模型上身机器人，预示着机器人 AGI 时代的到来。通用人形机器人的商业化进程预计加速，公司计划高速批量生产以降低单个人形机器人的单位成本，致力于实现可持续的规模经济，并实现量价齐升。未来 AI 机器人有望广泛应用于家庭、医疗、教育、服务等多个领域，并带动相关产业链，如传感器、执行器、控制系统等核心零部件的研发及生产。

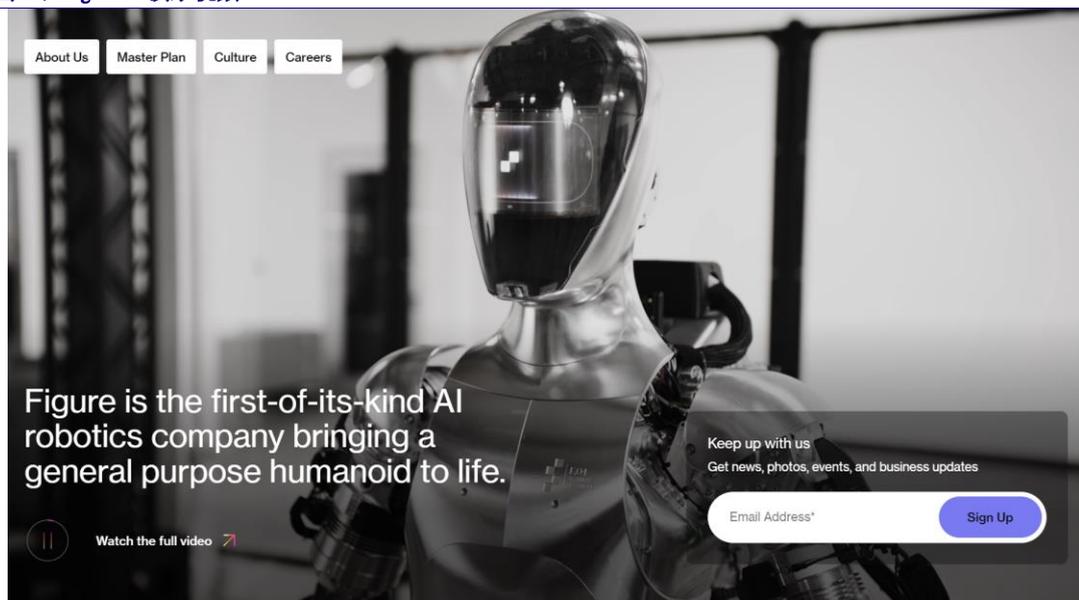
系统硬件：Figure 目标是开发出具有类人物理能力的硬件，通过运动范围、有效载荷、扭矩、运输成本和速度以衡量，并通过较短发展周期持续改进。

单位成本：Figure 旨在通过高速批量生产以降低单个人形机器人的单位成本，致力于实现可持续的规模经济。在高批量生产率下，公司认为单位成本将降至可承受水平。

安全性：Figure 开发的类人机器人能够安全地与工作场所的人类互动，以符合行业标准和企业要求。

批量生产：Figure 预计，为提供高质量的产品并以高产量交付，从原型设计转向批量生产的过程中可能面临困难。因此，公司严密制定制造设计、系统安全性、可靠性、质量及其他生产计划。

图64：Figure01 演示视频



资料来源：Figure，中国银河证券研究院

（三）数据：跨境流动监管趋严，大国经济支持数据要素价值释放

1. 全球数据跨境流动监管趋严，“大国经济体”将凸显数据优势

当前，数据跨境流动已成为全球资金、信息、技术、人才、货物等资源要素交换、共享的基础。3月22日，国家互联网信息办公室公布《促进和规范数据跨境流动规定》（下称“《跨境流动新规》”），规定明确了数据出境安全评估申报标准，规定了免于申报数据出境安全评估、订立个人信息出境标准合同、通过个人信息保护认证的数据出境活动条件。

《跨境流动新规》对数据出境安全评估、个人信息出境标准合同、个人信息保护认证等数据出境制度进行了优化调整，其实施将有利于促进数据依法有序自由流动，扩大高水平对外开放，并激发数据要素价值的释放。

《跨境流动新规》明确了应当申报数据出境安全评估的两类数据出境活动条件，一是**关键信息基础设施运营者向境外提供个人信息或重要数据**；二是**关键信息基础设施运营者以外的数据处理者向境外提供重要数据，或是自当年1月1日起累计向境外提供100万人以上个人信息（不含敏感个人信息）或1万人以上敏感个人信息**。同时，明确了应当订立个人信息出境标准合同或者通过个人信息保护认证的数据出境活动条件，即**关键信息基础设施运营者以外的数据处理者自当年1月1日起累计向境外提供10万人以上、100万人以下的个人信息（不含敏感个人信息），或1万人以下的敏感个人信息**。

《跨境流动新规》规定了**免于申报数据出境安全评估、订立个人信息出境标准合同、通过个人信息保护认证的数据出境活动条件**：一是国际贸易、跨境运输、学术合作、跨国生产制造和市场营销等活动中收集和产生的数据向境外提供，不包含个人信息或者重要数据的；二是在境外收集和产生的个人信息传输至境内处理后向境外提供，处理过程中没有引入境内个人信息或者重要数据的；三是为订立、履行个人作为一方当事人的合同，确需向境外提供个人信息的；四是按照依法制定的劳动规章制度和依法签订的集体合同实施跨境人力资源管理，确需向境外提供员工个人信息的；五是紧急情况下为保护自然人的生命健康和财产安全，确需向境外提供个人信息的；六是**关键信息基础设施运营者以外的数据处理者自当年1月1日起累计向境外提供10万人以下个人信息（不含敏感个人信息）的**。

《跨境流动新规》体现了**宽严并济的态度**，对于敏感个人信息保护、关键信息基础设施运营者等问题的监管尺度进行了重申和明确，做到力求坚持促进数据发展与安全并重的原则。

此外，监管部门还同步发布了《数据出境安全评估申报指南（第二版）》和《个人信息出境标准合同备案指南（第二版）》，进一步完善与确定了合规义务履行的流程和细节。可以预见的是，**中国数据跨境合规监管将进入新阶段，未来更加明晰、具体细化的监管体系有望有序推出，相关执法与监管活动将基于该体系稳步开展和推进**。

近年，我国跨境数据传输的法律框架及实际执行取得重大进展。网信办跨境数据传输安全审查和标准合同条款方面进展卓越，旨在解决原先审查规则缺乏明确性、跨境数据传输审批程序延误严重等问题。2023年12月7日，国务院印发《全面对接国际高标准经贸规则推进中国（上海）自由贸易试验区高水平制度型开放总体方案》，以下简称《总体规划》，将上海定位为“建设国家制度型开放示范点”，为上海自贸试验区进一步放宽跨境转移业务指明方向新框架下的限制。

为落实《总体规划》试点，上海市人民政府于2024年2月3日印发了《全面对接国际高标准经贸规则推进中国（上海）自由贸易试验区高水平制度型开放总体方案》的实施方案，以下简称《实施方案》。2024年2月6日，上海市人民政府新闻办公室召开新闻发布会，介绍《实施方案》重点内容，提出一系列规范和促进数据跨境传输的措施，如制定重要数据目录，探索建立合法、安全、便捷的跨境数据传输机制；在自贸区（包括临港新片区）建立跨境数据服务中心。具体而言，《实施方案》规定，**允许金融机构根据适用的安全政策和措施将业务数据转移到境外**。

根据《实施方案》中概述的原则，临港特别片区将为跨境数据传输开发“通用数据清单”和“重要数据目录”。临港新片区政府近日发布了《临港新片区数据跨境流动分类分级管理办法（试行）》，将跨境数据分为三个等级：1）“核心数据”、2）“重要数据”和3）“一般数据”。禁止将“核心数据”转移到中国境

外；涉及“重要数据”的传输必须经过临港新片区跨境数据服务中心的初步验证和申请程序，之后必须向当地国家互联网信息办公室提交安全评估；如果满足相关的数据保护管理要求，则允许“一般数据”自由传输。临港新区政府成立了由行业利益相关者组成的工作组，负责制定“一般数据”清单和“重要数据”目录，以解决各个行业（例如智能网联汽车、金融理财、高端航运、国际贸易、生物医药、文化出海）的特定数据传输场景。

表16：核心数据、重要数据定义

概念	法条	规定
核心数据	《数据安全法》	第二十一条第二款：关系国家安全、国民经济命脉、重要民生、重大公共利益等数据属于国家核心数据，实行更加严格的管理制度。
	《网络安全法》	第二十一条：国家实行网络安全等级保护制度。网络运营者应当按照网络安全等级保护制度的要求，履行下列安全保护义务，保障网络免受干扰、破坏或者未经授权的访问，防止网络数据泄露或者被窃取、篡改；采取数据分类、重要数据备份和加密等措施。
重要数据	《网络安全法》	第三十七条：关键信息基础设施的运营者在中华人民共和国境内运营中收集和产生的个人信息和重要数据应当在境内存储。因业务需要，确需向境外提供，应当按照国家网信部门会同国务院有关部门制定的办法进行安全评估；法律、行政法规另有规定的，依照其规定。
	《数据安全法》	第二十一条第一款：国家建立数据分类分级保护制度，根据数据在经济社会发展中的重要程度，以及一旦遭到篡改、破坏、泄露或者非法获取、非法利用，对国家安全、公共利益或者个人、组织合法权益造成的危害程度，对数据实行分类分级保护。国家数据安全工作协调机制统筹协调有关部门制定重要数据目录，加强对重要数据的保护。
		第二十一条第三款：各地区、各部门应当按照数据分类分级保护制度，确定本地区、本部门以及相关行业、领域的重要数据具体目录，对列入目录的数据进行重点保护。

资料来源：上海源泰律师事务所，中国银河证券研究院

大国经济+全球化循环赋能数据要素：数字经济作为典型的大国经济，具规模性特征、结构性特征及内源性特征，大国经济的核心优势是内部可循环，具备较稳定的国内产业链、供应链和价值，推动经济自主协调发展。中国依托内循环为基础，加以政策促进数据跨境传输，推动全球化双循环进程，将促进数据要素自由流动，推动人工智能、数字经济发展，进而成为国内经济发展的主要推动力。

数字主权的重要性随着全球数字化进程的加速日益凸显，数据已成为国际竞争和合作的新焦点。数字主权不仅关系到国家的经济发展、社会稳定和国家安全，也与公民的隐私权、信息自由等基本权利密切相关。数字化转型战略目前存在以下卡点：

1) 技术发展不平衡：不同国家在数字技术领域的发展水平存在差异，可能导致某些国家在全球数字经济中处不利地位。为缩小此差距，国家需投资于教育和研发，培养本土人才，同时鼓励创新和创业精神。

2) 国际法律框架缺失：目前尚缺乏统一的国际法律框架以规范相关行为。国际社会需要共同努力，建立公平合理的国际规则，以促进跨境数据流动和技术交流，同时保护各国的数字主权。

3) 数据治理标准不一：不同国家和地区对数据的处理和保护存在不同法律及规定。为促进国际间合作，需制定共通的数据治理标准，同时尊重各国的法律和文化差异。

4) 网络安全威胁：网络攻击和数据泄露事件频发，对国家安全和个人隐私构成严重威胁。国家需加强网络安全防护能力，提高公众的网络安全意识，并与其他国家合作打击跨国网络犯罪。

5) 公民权利保护需求：数字经济发展的背景下，国家需重视公民的基本权利，确保公民的隐私权、言论自由等权利得到充分保护。

6) 国际合作与竞争加剧：数字主权发展过程中，国家需在保护自身利益的同时，寻求与其他国家的合作机会，共同应对包括国际数据治理标准、数据跨境流动、数字贸易

等问题。

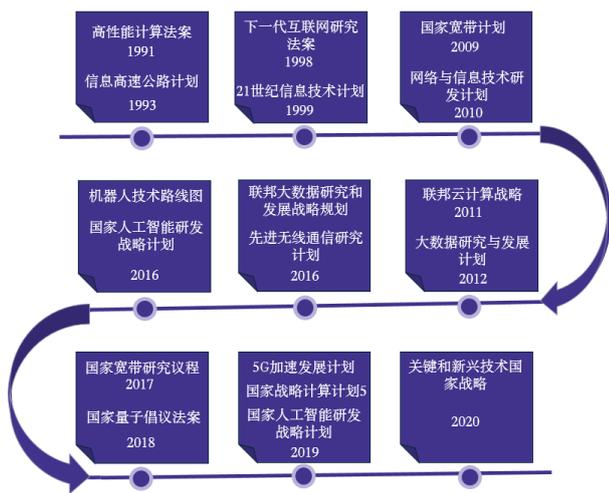
2. 美国、欧盟、中国数据战略比较

全球主要经济体数字化转型进程来看，美国、英国等国家正在从“自由探索”向“政府干预”转变；德国、法国等国家通过政府引领，打造高效产业数字化生态系统，发展数字经济；日韩等国家以解决问题为导向，重视基础研究和技术研发；中国积极推动数据要素市场制度，在确保安全底线的前提下更好地促进公共数据、社会数据流通交易和创新应用，建立全国数据采集、存储、处理、分析等方面的强制性标准，统筹和协调相关部门有效履职，实现国家层面数据的汇聚和共享，为各个领域提供数据资源支持。

2.1 美国：行政命令侧重安全与隐私保护，进一步限制数据跨境传输。

美国数字战略发展主要围绕技术领域、国防领域、政务领域、与产业领域四个方面所展开，在聚焦国际竞争为重点的时代，美国仍将技术领先尊为首要原则。美国和中国在监管数字经济方面遵循截然不同的模式。美国政府过去一直采用市场驱动的模式，目前部分领域趋向于向“政府干预”转变。

图65：美国数字战略布局历史沿革



资料来源：《美国数字战略的演进与发展》（胡微微等、中国电子科学研究院学报），中国银河证券研究院

图66：美国现阶段数字战略布局



资料来源：《美国数字战略的演进与发展》（胡微微等、中国电子科学研究院学报），中国银河证券研究院

2023年，人工智能正式步入美国的政治对话，除辩论还采取了系列行动，致使拜登政府于10月底发布了关于人工智能的行政命令（行政命令是法规，不是立法，在到期或被撤销之前一直有效），侧重要求提高人工智能产业的透明度、并制定新标准的指令。该行政命令主要从以下方面对人工智能发展做出规范限制：

表17：美国人工智能行政命令的主要规范限制

1 人工智能安全新标准	该行政命令要求最强大的人工智能系统的开发人员与美国政府分享其安全测试结果和其他关键信息。目的是这些措施将确保人工智能系统在公开之前是安全、可靠和值得信赖的。美国国家标准技术研究院将制定严格的测试标准以确保安全，并由政府部门实施。政府部门还将制定防止人工智能欺诈的指南，并制定开发人工智能工具以协助网络安全的计划。
2 保护美国公民的隐私，尤其是数据隐私	拜登政府利用行政命令呼吁国会通过数据保护立法，以保护所有美国人，特别是儿童的隐私权利。它还指示联邦政府将优先支持加速数据隐私的发展。

3 促进公平和公民权利	该行政命令发布指示，通过制定指导、培训和最佳实践，防止人工智能被用来加剧司法、医疗健康和住房领域的歧视、偏见和其他滥用行为。
4 保障消费者、患者和学生权益	通过部署适当的人工智能工具和建立安全计划来防止不安全行为，将促进人工智能在医疗健康和教育领域的负责任使用。
5 支持工人权利	将制定原则和最佳实践，以减轻人工智能对工人的危害并最大限度地提高其效益。
6 促进创新和竞争	该行政命令旨在引领创新和竞争，重点关注研究，特别是医疗健康和气候变化等重要领域，并为小企业提供援助和资源，使他们能够将人工智能商业化。
7 提升美国在海外的领导力	该行政命令指出“人工智能的挑战和机遇是全球性的”，指导与国际合作伙伴的合作和接触。
8 确保政府负责任且有效地使用人工智能	为了确保美国政府负责任地使用人工智能，将发布指南，制定明确的标准来保护权利和安全。行政命令发布后，管理和预算办公室（OMB）发布了一份关于推进人工智能机构使用的治理、创新和风险管理的政策草案。该指南将在联邦机构中建立人工智能治理结构，推进负责任的人工智能创新，提高透明度，保护联邦工作人员，并管理政府使用人工智能带来的风险。

资料来源：WIND，中国银河证券研究院

表18：除行政法令外的美国人工智能政策

政策	内容
《人工智能权利法案蓝图》	人工智能系统开发和实施的五项基本原则为：建立安全有效的系统、防止算法歧视、数据隐私保护、通知和解释、保留人工评估和选择
《人工智能风险管理框架》	可信人工智能的特征；通过形成风险承受度、风险等级排序等因素的风险框架
《人工智能问责政策（征求意见稿）》	征询人工智能问责目的、问责主体、可借鉴立法等内容，重点关注算法歧视、数据隐私保护、透明度等领域
《透明自动化治理法案》	要求联邦机构在使用特定人工智能或其他自动化系统进行互动或受其影响时，通知个人与人工智能系统互动
《国家人工智能委员会法案》	建立国家人工智能委员会，引导相关立法

资料来源：赛迪咨询，中国银河证券研究院

2024年2月28日，拜登政府依据《国际紧急经济权力法》（IEEPA）发布了一项保护美国公民敏感个人数据免遭“受关注国家”利用的行政命令。美国司法部同时发布了执行该行政命令的拟议规则制定的预通知（ANPRM）Fact Sheet，概述了实施该命令的规则。此项举措表明美国政府决意限制向特定“受关注国家”主要包括中国（不包括中国台湾省）、俄罗斯、伊朗、朝鲜、古巴和委内瑞拉等国家传输美国公民个人数据。被限制数据主要包括美国公民个人与敏感信息，主要有基因组数据、生物特征数据、个人健康数据、地理位置数据、财务数据与某些类型的个人可识别信息。拜登政府认为数据使用者可能使用这些数据跟踪美国公民（包括军人），窥探个人生活，并将数据传给数据经纪人，出售给外国情报机构、军队或外国政府控制的公司。“受关注国家”可能通过技术分析和操纵大量敏感个人数据，对美国安全形成威胁，人工智能技术将加剧此风险。我们认为，美国此举是以往贸易限制的一种延续，是在人工智能三要素的数据领域的一种影射，我国人工智能数据侧发展势在必行。

2.2 欧洲议会3月13日通过《人工智能法案》，数据主权原则体现“内松外严”，

2024年3月13日，欧盟议会以523票赞成、46票反对和49票弃权审议通过全球首个AI监管法案《人工智能法案》（EU AI Act），离落地越来越近。但乐观估计欧盟议会、欧盟理事会和欧盟委员会最快今年内会形成《人工智能法案》终稿并通过。考虑到实操中普遍设置约为2年的“过渡期”，该法案通过后的全面施行预计需

要等到 2026 年前后。《人工智能法案》的通过标志着全球人工智能领域监管迈入全新时代。该法旨在保护基本权利、民主、法治和环境可持续性免受高风险人工智能的侵害，同时促进创新并确立欧洲在该领域的领导者地位。该法规根据人工智能的潜在风险和影响程度规定了人工智能的义务。

图67：《欧盟人工智能法案》发展时间线



资料来源：中国银河证券研究院

欧盟近年来受内部发展不足影响，数字经济发展的战略核心目标是夺回“数字主权”。欧盟数字经济战略的三份文件提到“数字主权”概念，并从三个方面对其数字经济战略发展规划进行展开，即基础设施和工具、标准和法律规则、以及价值观和社会模式。

21 世纪初，欧盟计划在十年内将欧盟建设为全世界最具竞争力和活力的数字经济体，但截至 2019 年，其在全球数据所占份额极低，与其在世界经济中所占比重较不匹配，原因主要系：1) 欧洲在研究和创新领域的投资占比较低；2) 欧洲缺乏一个集研发和创新一体的研发中心，以统筹整合各经济体的资源。因此欧盟至 2030 年对数字经济的中期目标为：欧盟在全球数字经济中的占比应与其在全球经济中占比相匹配，成为中美外的第三极。战略层面上看，欧盟在科技领域的战略规划中更关心的是其独立技术发展能力和对技术的运用与控制能力，即控制技术的权力。

欧盟数据战略相关法案陆续推出及落地。目前欧盟在数字化战略领域有四部重要的立法，分别是欧盟的《数据法案》(Data Act)、《人工智能法案》(AI Act)、《数字市场法案》(Digital Markets Act) 和《数据服务法案》(Data Service Act)。 欧盟理事会于 2023 年 11 月 27 日通过了《数据法案》(Data Act)。欧盟于 2020 年提出《数字服务法》和《数字市场法》草案，经历了一波三折的立法过程后，2022 年 7 月，欧洲议会以压倒性多数通过了这两部法律。欧盟《数字服务法案》及《数字市场法案》分别自 2024 年 2 月 17 日、2024 年 3 月 7 日起已正式落地生效。《数字服务法》从内容管理、广告推送、商品交易等维度为平台方设置了一系列义务，侧重从内容及形式等方面规范数字企业提供的服务；《数字市场法》旨在规范数字市场，尤其是数字企业之间的竞争，主要规制根据法案中的客观标准被认定为“守门人”的大型在线企业，通过加强守门人义务对平台进行规制与监管，避免跨国科技巨头凭借垄断优势在欧洲市场过度扩张，两项法案对数字服务尤其是线上平台进行了一系列规制，将共同组成欧盟对线上平台的监管框架。

欧盟的数据主权原则可以概括为四个字：“内松外严”。欧盟数据主权战略可将其分为“单一市场”战略和“共同空间”战略两个部分。“单一市场”战略是欧盟的对外战略，体现了欧盟维护数据主权的本质追求；“共同空间”战略是欧盟的对内战

略，是欧盟建设内部数据市场、实现内部数据流通的联结机制。

欧盟“对外”采取的是“数据跨境限制+保护性域外管辖”的模式，以 GDPR 为典型代表，体现“外严”数据主权原则，包括较为严格的限制数据跨境以及保护性域外管辖；而欧盟“对内”战略，则在“内松”数据主权原则下，尽可能实现欧盟内部数据市场的协调与统一，促进个人和企业的数据/信息共享，创造坚实和公平的数据驱动型经济，并指导欧盟到 2030 年的数字化转型。

欧盟议会通过的《人工智能法案》重点关注五个主要优先事项：人工智能的使用应该安全、透明、可追溯、非歧视和环境友好。该立法还要求人工智能系统应由人而非自动化监督，对人工智能建立技术中立的统一定义，适用于已开发的系统以及未来构建的系统，要求人工智能系统提供商遵守该法规。此外，如果该系统在欧盟境内使用，位于欧盟以外的提供商和用户也有义务遵守这些准则。尽管如此，将人工智能用于军事目的组织和欧盟以外国家的公共机构不受此规定的约束。其主要亮点包括根据风险级别对人工智能系统进行分类，从最小风险到不可接受的风险，并风险类别强制实施监管，强调人工智能的隐私和安全。

其中，明确禁止的人工智能实践主要有三种：

第一类：指运用超出个人意识的潜意识技术和有目的的操纵或欺骗技术，上述技术明显损害个人或群体做出知情决定的能力，扭曲其行为并导致做出意料之外的决定，从而对其产生重大伤害的系统。

第二类：指利用特定个人或群体的弱点，扭曲其行为，或可能以对其造成重大伤害为目的或效果的系统。

第三类：指利用个人及其群体的社会行为或已知及潜在的个性特征，在一定时期内对其进行评估或分类的人工智能系统。

图68：《欧盟人工智能法案》的监管分级



资料来源：EU，中国银河证券研究院

该法案对高风险人工智能系统从设计、实施和上市后进入阶段提出了一系列要求。包括：风险管理制度（第 9 条）；数据和数据治理（第 10 条）；技术文件（第 11 条和附件 IV）；记录保存（第 12 条）；透明度和向用户提供信息（第 13 条）；人工监督（第 14 条）；准确性、稳健性和网络安全（第 15 条）；质量管理体系（第 17 条）；基本权利影响评估。虽然有限风险系统不会面临相同的合规审查，包括合格评定和产品安全审查，但它们也将这些类别下进行评估。

欧盟人工智能法案的处罚：

对违反人工智能法案的罚款按照违规公司上一财年全球年营业额的百分比或预定金额确定，以较高者为准。违反被禁止的人工智能应用程序的罚款为 3500 万欧元或 7%，违反人工智能法案义务的罚款为 1500 万欧元或 3%，提供不正确信息的罚款为 750 万欧元或 1.5%。

2.3 中国《促进和规范数据跨境流动规定》近日正式施行，力求坚持数据发展与安全并重原则

2024 年 3 月 22 日，国家网信办公布《跨境流动新规》正式施行。《跨境流动新规》于 2023 年 11 月 28 日就已通过，在新规征求意见期间，全球数据跨境流动的相关立法频繁变化，特别是美国频繁出台法案及行政令，而从中国实际发布的《跨境流动新规》来看，新的数据跨境流动监管框架力求坚持数据发展与安全并重原则，进一步明确监管机制与要求的同时，释放较强促进正常商贸活动的数据流动信号，符合国家促进数据要素发展战略指引，也标志着中国数据跨境合规监管将进入新的发展阶段，相关执法与监管活动将基于更明晰、具体的监管规定稳步地开展和推进，有利于数据产业生态健康发展。

《跨境流动新规》主要对下列内容进行了规定：

- I. 明确重要数据出境安全评估申报标准；
- II. 明确不涉及重要数据和个人信息的数据跨境传输免于履行数据出境安全评估、个人信息出境标准合同订立及通过个人信息保护认证机制；
- III. 设立自由贸易试验区负面清单制度；
- IV. 调整三大数据出境合规专项机制的触发阈值，可使企业合规成本降低；
- V. 延长数据出境安全评估结果有效期，增加数据处理者可以申请延长评估结果有效期的规定。

中国在发展人工智能方面拥有的数据优势主要体现在以下几个方面：

- I. **海量数据资源：**中国拥有世界上最多的互联网用户，中国的网民数量超过 7 亿，这些数据来源于消费者行为、社交媒体互动、在线交易、移动应用使用等多个方面，为人工智能算法的训练和优化提供了丰富的素材。
- II. **多样化应用场景：**中国在多个领域如电商、金融科技、智慧城市、医疗健康、教育等场景都有着广泛的人工智能应用需求。这些应用场景的多样性和复杂性为人工智能技术的发展和 innovation 提供了广阔的试验场。
- III. **政策支持数据要素价值释放：**
 - i. **政策制定与实施：**中国政府高度重视数据要素的高质量发展，出台了一系列政策文件，如《关于构建数据基础制度更好发挥数据要素作用的意见》（简称“数据二十条”），明确了数据基础制度体系的基本架构，为数据要素发展奠定了制度基础。
 - ii. **数据要素市场建设：**国家数据局的组建优化了数据管理体制，增强了数据要素发展的统筹力度。各地区、各行业、各企业纷纷加快数据要素领域的布局，围绕公共数据的开发利用、场内外数据交易、数商生态培育，推动数据要素价值释放。
 - iii. **数据产权界定：**探索建立数据资源持有权、数据加工使用权和数

据产品经营权“三权分置”的数据产权制度框架，解决市场主体遇到的实际问题，促进数据要素的流通和交易。

- iv. **数据资产登记与评估：**建立国家数据资产登记存证平台，完善数据资产评估体系，确保数据资产的安全性、合规性、质量和价值。
- v. **数据资产定价与人表：**推动数据资产纳入会计核算体系，构建全新的数据资产估值和定价逻辑，促进数据要素交易流通。
- vi. **技术创新与应用：**鼓励数据要素相关技术研发创新，如向量数据库、图分析技术、时空大数据平台等，支持新兴业务场景下的数据要素价值释放。
- vii. **数据流通与交易：**建立数据流通和交易制度，包括数据交易所的设立，如广州数据交易所、深圳数据交易所等，促进数据资源的最优配置。

IV. 产业发展与生态建设：通过数据要素产业服务平台等措施，提供数据动态和解决方案服务，支持数据要素产业主管机构、供给方、运营方、服务方、需求方等主体快速把握趋势，充分释放数据要素价值。

通过这些措施，中国正在加速从“数据大国”向“数据强国”转变，不断推动数据要素市场化配置改革，加速释放数据要素的价值，促进数字经济与实体经济的深度融合。

中国目前的人工智能监管以鼓励人工智能技术发展与创新为主，监管并重。随着 AGI 的发展，国内立法将更为全面。2023 年 6 月，国务院宣布“人工智能法”列入立法议程。2023 年 7 月 13 日，中国国家网信办会同六部委共同发布《生成式人工智能服务管理暂行办法》（以下简称《暂行办法》），并宣布该办法将于 2023 年 8 月 15 日正式施行。适用范围为“面向中华人民共和国境内公众提供服务”的主体，无论该提供者是境内还是境外主体，也无论该产品是否在境内研发或者使用，都需要符合征求意见稿的相关要求。**我们认为，相对于之前的征求意见稿，《暂行办法》在鼓励创新与技术发展上有所侧重，并将强监管严打击的态度转变为兼顾发展与合规。**《暂行办法》第三条新增《科学技术进步法》作为上位法依据。同时，《暂行办法》新增第四条、第五条，明确了在合理监管的前提下，鼓励生成式人工智能技术的平台建设、自主创新、国际交流、以及各领域适用的多线发展。在处罚方面，第二十一条删除了罚款及终止利用生成式人工智能提供服务等较为严厉的措施，并在现行有效的中国法律及行政法规没有明文规定的情况下更倾向于采取先引导改正的态度，而拒不改正或严重违规时，由主管部门责令暂停提供相关服务。

《暂行办法》侧重于数据安全、内容合规和知识产权保护三方面。针对提供生成式人工智能服务的行为，设置了四项新制度。对 AIGC 技术提供者的法律责任做出了规定，其中包括算法评估和备案，数据标注，标识义务，实名认证，防沉迷设置及科学引导义务，投诉接收处理机制，服务稳定性要求，以及模型优化义务。

表19：《生成式人工智能服务管理暂行办法》的四项制度

《暂行办法》第十六条规定了，国家有关主管部门应当以行业和领域为单位，制定相应的分类分级监管规则或者指引。从横向来看，该条明确了提供生成式人工智能服务的行为未来将实行分类分部门管理，而不是简单地将所有事项纳入网信部门的数据保护责任范畴。从纵向来看，该条明确了生成式人工智能技术未来将进行分级监管。而分类分级的具体准则，仍待各行业主管部门的进一步细则。值得参考的是，欧盟《人工智能法案》将人工智能系统的风险划分成不可接受的风险、高风险、有限风险和轻微风险四种类型，并针对不同类型施加了禁用、高度监管、自我监管等不同程度的措施。

分级分类监管

提供者和使用者之间的服务协议	<p>《暂行办法》第九条规定了“提供者应当与使用者签订服务协议，明确双方权利义务”。这一条款相比于《征求意见稿》将责任大部分落在提供者一方的做法，为提供者 and 使用者共享风险开启了一扇新大门。然而，就服务协议的具体内容和实施方式，《暂行办法》并未给出详细规定。考虑到提供者 and 使用者之间存在的信息不对称，如果允许双方自由定义服务协议的内容，可能会导致使用者在无力反抗的情况下接受大量免除提供者法律责任的不公平条款。因此，参照中国现行的个人信息出境标准合同的监管模式，普华永道预测，未来可能会有针对服务协议文本内容的相关细则或指导文件陆续出台。</p>
境外生成式人工智能服务的监管	<p>《暂行办法》第二十条规定了“对来源于中华人民共和国境外向境内提供生成式人工智能服务的违法违规行为，国家网信部门有权通知相关机构采取技术措施处置”。这意味着，如果未来在跨境服务过程中发现违法违规的行为，国家网信部门有权以技术手段先进行屏蔽、拦截等处置，为中国进一步管理来源于境外生成式人工智能的服务提供法律依据。</p>
外商投资生成式人工智能服务	<p>《暂行办法》第二十三条规定了“外商投资生成式人工智能服务，应当符合外商投资相关法律、行政法规的规定”。虽然现行的外商投资法律并未对生成式人工智能服务进行规定，但从该条可以窥视到生成式人工智能服务或将是未来外商投资领域的监管重点。</p>

资料来源：中国国家网信办等六部委，中国银河证券研究院

此外，我国在各个行业部门与地方对人工智能的监管将日益完善。

表20：人工智能相关部门规章与地方政策梳理

发文机关	政策	内容
全国信息安全标准化技术委员会秘书处	《网络安全标准实践指南——生成式人工智能服务内容标识方法》	为贯彻落实《生成式人工智能服务管理暂行办法》关于对生成内容进行标识的要求，指导生成式人工智能服务提供者等有关单位做好内容标识工作，此文件围绕文本、图片、音频、视频四类生成内容给出了内容标识方法，可用于指导生成式人工智能服务提供者提高安全管理水平。
工业和信息化部、国家互联网信息办公室等七部门	《生成式人工智能服务管理暂行办法》	《暂行办法》中明确了一系列法律文件作为上位法，包括《中华人民共和国网络安全法》《中华人民共和国数据安全法》（简称“《数据安全法》”）、《中华人民共和国个人信息保护法》《中华人民共和国科学技术进步法》等。同时，《暂行办法》关联了《互联网信息服务深度合成管理规定》（简称“《深度合成规定》”）和《互联网信息服务算法推荐管理规定》（简称“《算法规定》”）等规范性文件，在生成式人工智能服务的内容标识、安全评估手续等方面进行支持和补充，使生成式人工智能服务得到更加全面和细致的管理。
科技部等六部门	《关于加快场景创新以人工智能高水平应用促进经济高质量发展的指导意见》	场景创新是以新技术的创造性应用为导向，以供需联动为路径，实现新技术迭代升级和产业快速增长的过程。推动人工智能场景创新对于促进人工智能更高水平应用，更好支撑高质量发展具有重要意义。我国人工智能技术快速发展、数据和算力资源日益丰富、应用场景不断拓展，为开展人工智能场景创新奠定了坚实基础。
科技部	《新一代人工智能伦理规范》	人工智能治理专业委员会发布了《新一代人工智能伦理规范》（以下简称《伦理规范》），旨在将伦理道德融入人工智能全生命周期，为从事人工智能相关活动的自然人、法人和其他相关机构等提供伦理指引。
全国信息安全标准化技术委员会秘书处	《网络安全标准实践指南—人工智能伦理安全风险防范指引》	为进一步确保人工智能安全可控，统筹协调人工智能发展与安全，促进人工智能对国家经济、社会、生态等方面的持续推动作用，相关组织或个人在开展人工智能研究开发、设计制造、部署应用等相关活动时，应充分识别、防范、管控人工智能伦理安全风险。此实践指南依据法律法规要求及社会价值观，针对人工智能伦理安全风险，给出了安全风险防范措施，为相关组织或个人在各领域开展人工智能研究开发、设计制造、部署应用等活动时提供指引。
国家标准化管理委员会等五部门	《国家新一代人工智能标准体系建设指南》	《指南》编制遵循指导性、可用性、阶段性原则，旨在对人工智能标准化工作进行顶层设计，构建了基础共性、支撑技术与产品、基础软硬件平台、关键通用技术、关键领域技术、产品与服务、行业应用、安全/伦理八部分组成的国家新一代人工智能标准体系框架。

教育部等三部门	《关于“双一流”建设高校促进学科融合加快人工智能领域研究生培养的若干意见》	《意见》提出，要以国家发展人工智能的重大战略需求为中心，以“需求导向、应用驱动”“项目牵引、多元支持”“跨界融合、精准培养”为基本原则，瞄准“理论、算法、平台、芯片和应用”等急、断、缺的短板领域，构建基础理论人才与“人工智能+X”复合型人才并重的培养体系。
林草局	《国家林业和草原局关于促进林业和草原人工智能发展的指导意见》	《意见》提出建设生态保护人工智能应用体系。实施创新驱动发展战略，充分运用大数据、物联网、卫星遥感、图像识别、无人机、机器人等新一代信息技术，在森林生态系统保护领域、草原生态系统保护领域、湿地生态系统保护领域、荒漠生态系统保护领域、生物多样性保护领域，创新监管模式，开展智能监测，搞好预警，提供科学决策依据，激发生态保护新动能，实现生态保护智能化，形成生态保护新模式。
最高人民法院	《关于规范和加强人工智能司法应用的意见》	坚持以习近平新时代中国特色社会主义思想为指导，深入贯彻习近平法治思想，坚持司法为民、公正司法工作主线，加快推进人工智能技术与审判执行、诉讼服务、司法管理和治理社会等工作的深度融合，规范司法人工智能技术应用，提升人工智能司法应用实效，促进审判体系和审判能力现代化，为全面建设社会主义现代化国家、全面推进中华民族伟大复兴提供有力司法服务。
上海市经济和信息化委员会	《上海人工智能示范应用清单（2023）》	清单采用“3+4+3+9”的总体架构：3个类别，大型综合应用、引领性应用和成熟应用。其中，大型综合应用展现多种先进技术产品融合的应用场景，引领性应用展现前沿技术探索、有潜在发展空间的应用场景，成熟应用展现已落地、可规模化推广的应用场景。

资料来源：WIND，中国银河证券研究院

3. 高质量数据是全球人工智能竞争的“胜负手”，数据要素市场空间广阔

数据资源的价值可量化使得 2024 年成为数据要素三次价值释放的元年。

数字要素市场空间测算：根据清华大学社科院刘雄涛教授等人根据增值法对数据资本的测算所得，2020 年，我国数据资本存量约为 17.4 万亿。本报告基于该测算做如下假设，并测算十四五期间数据资本预测值：1) 假设 2024-2025 年 GDP 增速为 5.0%、4.9%；2) 假设 2021-2025 年数据资本存量占 GDP 比重的增速均为 5%。

基于以上假设的测算结果为：预计 2025 年，我国数据资本空间约为 32.0 万亿元，增速为 10.1%。

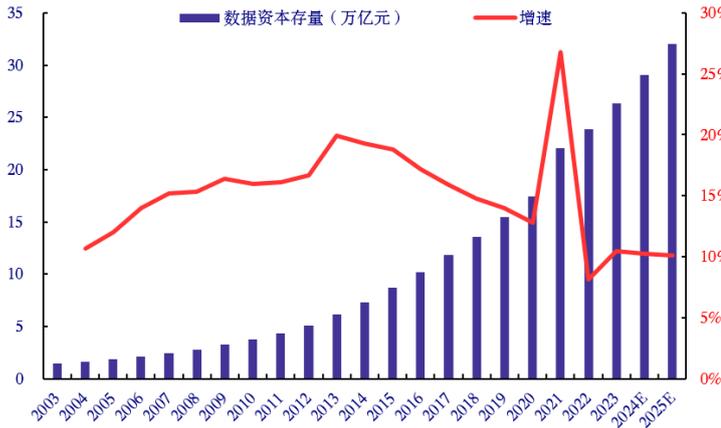
表21：数据要素市场规模及增速预测

年份	数据资本存量（亿元）	GDP（亿元，不变价）	数据资本存量/GDP	GDP 增速（调整后）	数据资本存量占 GDP 比重增速
2012	51110	486983	10%	7.9%	8%
2013	61311	524803	12%	7.8%	11%
2014	73130	563774	13%	7.4%	11%
2015	86891	603471	14%	7.0%	11%
2016	101818	736036	14%	6.8%	-4%
2017	118014	787170	15%	6.9%	8%
2018	135443	840303	16%	6.7%	8%
2019	154367	890305	17%	6.0%	8%
2020	174138	910236	19%	2.2%	10%

2021	220803	1099198	20%	8.4%	5%
2022	238773	1131632	21%	3.0%	5%
2023	263774	1191037	22%	5.2%	5%
2024E	290811	1250589	23%	5.0%	5%
2025E	320314	1311868	24%	4.9%	5%

资料来源：WIND，《数据资本估算及对中国经济增长的贡献-基于数据价值链的视角》（刘涛雄等，《中国社会科学》2023年第10期），中国银河证券研究院

图69：数据要素市场规模及增速预测



资料来源：WIND，《数据资本估算及对中国经济增长的贡献-基于数据价值链的视角》（刘涛雄等，《中国社会科学》2023年第10期），中国银河证券研究院

注：数据资本空间的测算与假设主要参考《中国社会科学》2023年第10期刘涛雄等的方法，主要有以下步骤：一，基于其文章中表2中2020年之前测算所得资本存量数据，假设数据资本存量占比增速为5%，得出相应数据资本存量占比，再根据2021年与2022年已公布的GDP数据与增速，得出2021与2022年资本存量数据；二，假设2023-2025年GDP增速为5.2%、5.0%、4.9%，数据资本存量占比增速为5%、5%、5%，测算出2023-2025年数据资本空间。

高质量数据是全球人工智能竞争的“胜负手”：数据质量及其包含的有用信息量是决定机器学习算法学习能力的关键因素。因此在将数据集提供给机器学习算法之前，确保对数据集进行检查和预处理至关重要。高质量的数据可提高大型语言模型（LLM）的SOTA（例 phi-1），同时可大幅减少数据集大小和训练计算，并可显著降低 LLM 的训练成本。

图70：尽管 phi-1 的训练规模较小，但其表现明显优于其他模型

Date	Model	Model size	Dataset size (Tokens)	HumanEval (Pass@1)	MBPP (Pass@1)
2021 Jul	Codex-300M [CTJ*21]	300M	100B	13.2%	-
2021 Jul	Codex-12B [CTJ*21]	12B	100B	28.8%	-
2022 Mar	CodeGen-Mono-350M [NPH*23]	350M	577B	12.8%	-
2022 Mar	CodeGen-Mono-16.1B [NPH*23]	16.1B	577B	29.3%	35.3%
2022 Apr	PaLM-Coder [CND*22]	540B	780B	35.9%	47.0%
2022 Sep	CodeGeeX [ZXZ*23]	13B	850B	22.9%	24.4%
2022 Nov	GPT-3.5 [Ope23]	175B	N.A.	47%	-
2022 Dec	SantaCoder [ALK*23]	1.1B	236B	14.0%	35.0%
2023 Mar	GPT-4 [Ope23]	N.A.	N.A.	67%	-
2023 Apr	Replit [Rep23]	2.7B	525B	21.9%	-
2023 Apr	Replit-Finetuned [Rep23]	2.7B	525B	30.5%	-
2023 May	CodeGen2-1B [NHX*23]	1B	N.A.	10.3%	-
2023 May	CodeGen2-7B [NHX*23]	7B	N.A.	19.1%	-
2023 May	StarCoder [LAZ*23]	15.5B	1T	33.6%	52.7%
2023 May	StarCoder-Prompted [LAZ*23]	15.5B	1T	40.8%	49.5%
2023 May	PaLM 2-S [ADF*23]	N.A.	N.A.	37.6%	50.0%
2023 May	CodeT5+ [WLG*23]	2B	52B	24.2%	-
2023 May	CodeT5+ [WLG*23]	16B	52B	30.9%	-
2023 May	InstructCodeT5+ [WLG*23]	16B	52B	35.0%	-
2023 Jun	WizardCoder [LXZ*23]	16B	1T	57.3%	51.8%
2023 Jun	phi-1	1.3B	7B	50.6%	55.5%

资料来源：Medium，中国银河证券研究院

据 Epochai 预测，2030 年至 2060 年，大模型训练将耗尽视觉数据的存量；2026 年之前将耗尽高质量语言数据；2030 年至 2050 年，将耗尽低质量语言数据的存量；该预测基于当前机器学习（ML）数据使用和生产的趋势将延续，且数据效率无重大创新的假设前提。发展数据要素市场势在必行。

四、AIGC 驱动千行百业加速裂变，行业应用加速渗透

（一）AIGC+金融：国内金融大模型抢滩布局，场景应用加速渗透

海外：彭博的大语言模型 BloombergGPT，是 LLM 在垂直金融场景的落地加速的象征。BloombergGPT 聚焦于金融场景，将其多年积累的金融领域报告术语等对 LLM 进行训练，相较普适性类 ChatGPT 产品，其对金融术语的理解更为高效与准确。

对比通用大模型，小模型在特定领域表现更优。BloombergGPT 采用 500 亿参数与 7000 亿数据集规模对大模型进行训练，远小于 GPT-3 的 1750 亿参数量与 45TB 数据量。其数据集主要包含两个部分，1) 数据量占比 49% 的通用型文本数据集，主要包括 ThePile、C4、Wikipedia 等通用常识类数据集，2) 数据量占比 51% 的金融预料数据集，主要包括新闻、研究报告、公司财报、网络爬取的金融文件以及提取到的社交媒体消息等内容。根据彭博发布的论文，BloombergGPT 在通用领域能力与 GPT-3 几乎持平，但在金融垂直领域文本撰写和问答能力更为突出。BloombergGPT 的成功说明 LLM 在垂直领域应用的可用性和训练成本可控的可行性。

表22：Bloomberg 在金融权益类测试中较大模型表现更优

	BloombergGPT	GPT-NeoX	OPT66B	Bloom176B
EquityNews	79.63	14.17	20.98	19.96
EquitySocialMedia	72.4	66.48	71.36	68.04
EquityTranscript	65.06	25.08	37.58	34.82
ESNews	46.12	26.99	31.44	28.07
CountryNews	49.14	13.45	17.41	16.06
AllTasks(avg)	62.47	29.23	35.76	33.39
AllTasks(WR)	1.00	0.00	0.67	0.33

资料来源：Bloomberg，中国银河证券研究院

国内：2023 年 9 月 18 日，首个金融行业大模型标准发布。信通院牵头，联合腾讯云、奇富科技、科大讯飞等四十多家企业共同编制的《面向行业的大规模预训练模型技术和应用评估方法第 1 部分：金融大模型》正式发布。该标准主要涵盖了金融大模型的关键能力要求，包括场景适配度、能力支持度和应用成熟度三大方面，覆盖了投资研究、投资顾问、风险管理、市场营销、客户服务等多个金融应用场景，并详细规定了金融大模型在数据合规性、可追溯性、私有化部署、风险控制等方面的要求。为大模型在金融领域的应用提供了一定的指导，在金融行业的合规可信方面提供保障，有望加速行业应用

落地。

国内各厂商纷纷抢滩布局金融大模型，应用加速落地。

同花顺：深耕 AI 技术多年，为用户赋能具备先发优势。公司 AI 技术的研发和应用早在 2009 年就率先市场开始，成立“i 问财”部门，为用户提供股票信息相关搜索服务。2017 年，同花顺 AI 开放平台上线，可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；公司自主研发的同花顺智能语音平台，在中文金融场景语音识别准确率达到 98% 以上，中英文通用场景识别准确率超过 95%，多种方言识别准确率超过 90%；同时，平台还具备高度拟人的语音合成能力，以及语音转换、歌声合成、情感识别和声纹识别等智能语音技术能力。目前基于自主智能语音技术的产品已应用于多家证券公司、基金公司及电信运营商。同花顺虚拟数字人对话平台，运用多模态数字人对话技术，实现与真人用户“面对面”的交互体验，该产品目前已落地多个大型客户；同花顺机器翻译平台，目前支持中（简体）、中（繁体）、英、西班牙、法、德、日、韩、阿拉伯、越南、印尼、马来等 12 种文字，132 种语向的翻译，支持 PDF、Word、PPT 等 11 种主流文件格式的翻译，以及图片翻译、语音翻译等跨模态翻译。目前公司自研金融大模型“问财 HithinkGPT”，已进入内测与备案阶段。

图71：同花顺 AI 开放平台核心技术



资料来源：同花顺，中国银河证券研究院

恒生电子：2023 年 6 月，恒生电子发布金融行业大模型 LightGPT 以及金融智能助手光子、基于大模型技术升级的投研平台 WarrenQ。9 月，LightGPT 及四款大模型产品已面向 20 家金融机构开启内测。目前公司已发布基于 LightGPT 打造的四款光子产品：面向投顾创作的光子·文曲、面向合规管理的光子·方圆、面向投顾咨询的光子·善策和面向运营管理的光子·慧营。

华为：2023 年 9 月 20 日，华为全联接大会“智领睿变，共建数智金融未来”金融峰会成功举办。会上，华为发布金融大模型解决方案，包括场景层、模型层、底座层三层。以高质量的数据为基础，通过三层方案加速创新生产力、跃升智力、澎湃算力三力协同，逐步实现场景的价值落地、体系构建、能力固化。

百度：度小满于 2023 年 9 月 22 日宣布将“轩辕 70B”金融大模型开源，所有用户均可自由下载和试用，该升级版本加入了公司业务场景中的海量金融数据，在预训练阶段，能够处理更长的金融报告、研究和分析。该大模型在 C-Eval、CMMLU 两大权威大语言模型榜单上排名第一。

科大讯飞：2023年10月24日，科大讯飞于1024开发者节智慧金融论坛上发布星火金融行业大模型V1.0及场景应用最新成果。此次星火金融大模型V1.0具备金融知识问答、金融文本理解、金融内容生成、金融会话分析、金融数据分析、金融数学计算、金融逻辑推理七大核心能力。该大模型已与知识运营、客户服务、人机协同、APP交互、营销展业、软件研发等行业细分场景深度融合，并已孵化出星火智能客服、大模型坐席助手、大模型知识库、金融APP助手、金融展业助手、智能编程助手等产品应用。

宇信科技：2023年10月17日，公司推出大模型应用产品和解决方案，包括4个应用级产品和一个开发平台——开发助手CodePal、金融数据安全分级分类助手DataSherpa、AI+信贷助手、AI+营销助手以及大模型应用开发平台。本次发布的了5个产品及解决方案：1)开发助手CodePal基于开源编程大模型，为金融软件开发人员提供代码补全、自动化bug检测等功能，能够有效加速软件开发流程。2)金融数据全分级分类助手DataSherpa主要面向数据资产管理，能够快速理解监管指南要求，输出标准化的定级结果。目前该产品分类准确率超过95%，效率较人工提升10倍。3)AI+尽调助手是面向信贷客户经理的客户尽调辅助工具，能够辅助了解对公信贷客户，提供客户资料收集、智能核验、风险分析、尽调报告生成等功能。目前该产品的财报数据准确率达95%以上，报告生成可接受率达85%以上，能帮助客户经理提升60%以上的效能，降低处理成本。4)AI+营销助手是面向零售客户的营销精细化运营工具，具有外呼、呼入、人机协同、虚拟坐席等功能，适用于客户拉新、产品推荐、促活回访等营销场景。相较纯人工服务，该产品可节约80%的时间，同时转化率也提高了10+倍，产能提高3倍以上。5)大模型应用开发平台预置了知识库问答、报告生成、会议纪要等应用场景，支持以Agent为中心的应用开发新范式，能够帮助企业用户以较佳的体验快速实现大模型应用场景落地，提高开发效率，降低开发成本。除了目前已经发布的5个产品之外，公司未来还规划发布信贷助手、监管合规助手、经营分析助手等产品。其中信贷助手预计覆盖资产保全、信贷申请手里、审查审批、贷后管理及催收等。

表23：国内金融大模型进展

厂商	AI应用方向及进展
恒生电子	2023年6月,恒生电子发布金融行业大模型LightGPT以及金融智能助手光子·基于大模型技术升级的投研平台WarrenQ;9月,LightGPT及四款大模型产品已面向20家金融机构开启内测。目前公司已发布基于LightGPT打造的四款光子产品:面向投顾创作的光子·文曲、面向合规管理的光子·方圆、面向投顾咨询的光子·善策和面向运营管理的光子·慧营。
同花顺	深耕AI技术多年,为用户赋能具备先发优势。同花顺AI开放平台可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;公司自主研发的同花顺智能语音平台,在中文金融场景语音识别准确率达到98%以上,中英文通用场景识别准确率超过95%,多种方言识别准确率超过90%;同时,平台还具备高度拟人的语音合成能力,以及语音转换、歌声合成、情感识别和声纹识别等智能语音技术能力。目前公司自研金融大模型“问财HithinkGPT”,已进入内测与备案阶段。
华为	2023年9月20日,华为全联接大会“智领睿变,共建数智金融未来”金融峰会成功举办。会上,华为发布金融大模型解决方案,包括场景层、模型层、底座层三层。以高质量的数据为基础,通过三层方案加速创新生产力、跃升智力、澎湃算力三力协同,逐步实现场景的价值落地、体系构建、能力固化。
百度	度小满于2023年9月22日宣布将“轩辕70B”金融大模型开源,所有用户均可自由下载和试用,该升级版本加入了公司业务场景中的海量金融数据,在预训练阶段,能够处理更长的金融报告、研究和分析。该大模型在C-Eval、CMMLU两大权威大语言模型榜单上排名第一。
科大讯飞	2023年10月24日,科大讯飞于1024开发者节智慧金融论坛上发布星火金融行业大模型V1.0及场景应用最新成果。此次星火金融大模型V1.0具备金融知识问答、金融文本理解、金融内容生成、金融会话分析、金融数据分析、金融数学计算、金融逻辑推理七大核心能力。该大模型已与知识运营、客户服务、人机协同、APP交互、营销展业、软件研发等行业细分场景深度融合,并已孵化出星火智能客服、大模型坐席助手、大模型知识库、金融APP助手、金融展业助手、智能编程助手等产品应用。

宇信科技

2023年10月17日，公司推出大模型应用产品和解决方案，包括4个应用级产品和一个开发平台——开发助手 CodePal、金融数据安全分级分类助手 DataSherpa、AI+信贷助手、AI+营销助手以及大模型应用开发平台。

资料来源：恒生电子，同花顺，宇信科技，华为，百度，科大讯飞，中国银河证券研究院

表24：国内金融大模型主要应用场景

	智能客服/问答/投顾	运营辅助/管理	数据分析	智能营销	内容创作	合规与风险管理
恒生电子	✓	✓			✓	✓
同花顺	✓	✓			✓	
科大讯飞	✓		✓	✓	✓	
宇信科技	✓	✓		✓	✓	
华为	✓	✓	✓		✓	✓
百度	✓	✓	✓		✓	✓

资料来源：恒生电子，同花顺，宇信科技，华为，百度，科大讯飞，中国银河证券研究院

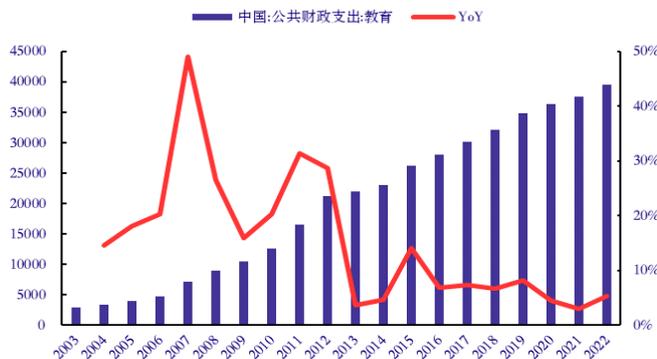
（二）AIGC+教育：大模型重构教育生态

ChatGPT 带来的大模型时代，将深入影响学校生态、教育教学改革、教学评价、教师发展等各个层面。随着大模型发展和落地，可以实现“因材施教”、“个性化学习”等教育最为核心的需求，从而给教育行业带来“质变”。

人工智能技术在教育领域的应用场景广阔，可大致分为四个层面：1) 知识内容检索：AI 可以帮助学生快速检索相关知识内容，满足对知识点查询的需求。这一应用场景知识密度要求较高，但对学生思维引导性较弱；2) 语言学习：AI 可以评估和纠正学生的口语、写作等语言表达能力。这一应用在通用智能达到一定阈值后即可实现，知识密度要求不高，但对学生思维起到部分引导作用。3) 内容生成：AI 可以根据教材和课程设置自动生成相关学习内容，如生成习题、案例分析等。这一应用知识密度需求不高，但对思维引导性要求增加。4) 灵感索引：灵感索引是“因材施教”的关键，也是传统教育中老师较难被替代的环节，例如，在学生若干结题步骤中精准定位出错点，判断思维误区，进行正确引导，目前国内外 AI+教育产品均在探索中，尚无成熟案例。

教育投入长期以来是公共财政第一大支出。2018年—2022年，国家财政性教育经费累计投入 21.4 万亿元，年均增长 7%，占国内生产总值的比例连续 11 年保持在 4% 以上。

图72：2003年至2022年国家教育公共财政支出（亿元）



资料来源：WIND，中国银河证券研究院

在线教育市场规模稳定增长，未来大有可为。根据艾瑞咨询数据，2025 年我国在线教育市场规模将达到 4905 亿元，年复合增长率 17.5%。这主要受益于以下几个方面的因素：1) 国家对于教育事业的重视和支持，不断增加教育经费投入，推动教育信息化和现代化的发展，为在线教育提供了政策红利和基础条件。2) 互联网技术的不断进步和创新，提升了在线教育的技术水平和服务质量，为在线教育提供了技术驱动和效率提升。3) 用户对于在线教育的接受度和付费意愿的不断提升，以及对于个性化、多样化、国际化的在线教育需求的不断增长，为在线教育提供了市场需求和消费动力。4) 在线教育机构的不断创新和优化，以及与其他行业的跨界合作和融合，为在线教育提供了产品多元化和竞争优势。

图73：Sora 生成视频截图



资料来源：OpenAI，中国银河证券研究院

Sora 实现跨代级提升，教育或将成为优质应用场景之一。Sora 可以将 Prompt 转换成长达 60 秒视频，相对于此前文生视频大模型 Runway、Pika、Stable Video 等提升几个级别。同时在视频分辨率以及质量方面，Sora 可以生成 1080P 清晰度视频，并且能够相对完整实现对世界及物体运动规律理解及模拟，在镜头切换方面保持稳定性。此外，Sora 还支持图片格式输入、视频扩展、视频拼接等，是文生视频领域突破性技术变革。

Sora 将彻底颠覆在线教育，拓宽应用场景。在线教育在于情境化学习、跨学科学习、主题性学习、项目式学习和学科实践，都与教学中创设情境有关。Sora 作为一种视频生成工具，初步能模拟物理世界及运动规律，具有创设情境的天然优势，可以成为情境教学的得力助手。Sora 具有能力涌现的功能，加上强大的算力，几乎可以生成人们能够想象和无法想象的任何逼真视频。包括文学体验、历史场景、气候变化和天体运行，这些在教学中能够用到的情境都可以让 Sora 生成，而且无限丰富、场景具体。

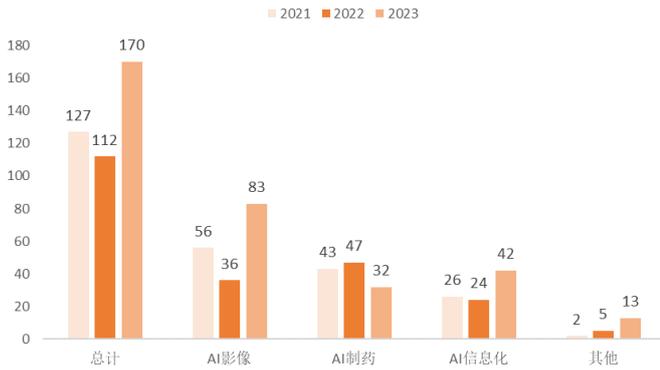
（三）AIGC+医疗：AI 药物研发和 AI 医学影像深具潜力

全球 AI 医疗市场保持快速增长，AI 药物研发和 AI 医学影像占比最高。根据 Global Market Insights 报告，2022 年全球 AI 医疗市场规模超过 50 亿美元，并预计将以超过 29% 的年均复合增速增长至 2032 年的 700 亿美元。细分子行业里，AI 药物研发和 AI 医学影像占比最高，合计占比超过 50%。预计 2032 年 AI 药物研发市场规模将超过 205 亿美元。

中国市场起步较晚，处于高速增长期，年增速超 30%，AI 药物研发和 AI 医学影像

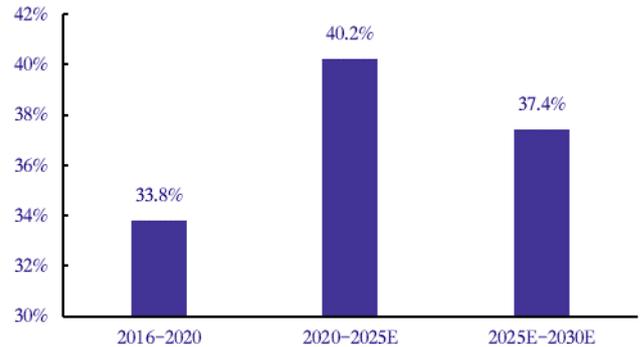
最具潜力。根据 IQVIA 数据，2021 年中国 AI 医疗市场规模不足 40 亿元，主要以智慧病案、信息系统和 CDSS（临床决策系统）等应用为主，而国外较成熟的 AI 药物研发和 AI 医学影像占比较低。据国家卫健委的数据，我国 AI 医疗市场规模预计到 2025 年将达到 500 亿元，显示出强劲的增长势头。未来 10 年，预计中国 AI 医疗市场将以超 30% 的年均复合增速增长，AI 医学影像（超声、CT、X 光、病理等）和 AI 药物研发将成为增长最快的细分市场。从 2021-2023 年中国 AI 医疗不同赛道融资情况来看，2023 年融资数有大幅增长，AI 影像、AI 制药、AI 信息化依次位列前三。

图74：2021-2023 年中国 AI 医疗不同赛道融资情况（单位：个）



资料来源：动脉网，中国银河证券研究院

图75：智能医疗行业阶段年复合增长率

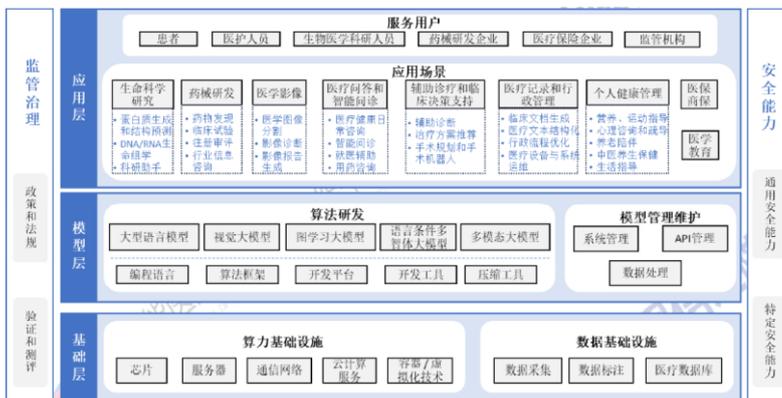


资料来源：沙利文《医疗智能行业白皮书》，中国银河证券研究院

大模型成为 AI 医疗行业的主要创新驱动力，多模态架构、自监督学习技术以及上下文学习能力使得大模型有望实现医疗领域的通用人工智能。未来的医疗大模型有望灵活解释多种数据模态组合，比如来自医学影像、电子病历、检验科、实验室结果等等的图表和文本数据，并产生更强更丰富的表达输出，可输出文本、图表或者语言，甚至有医学推理能力。

基于大模型驱动的 AI 医疗产业链可以分为基础层、模型层与应用层：1) **基础层**，除数据服务外，芯片与通信等基础核心领域与其他行业共通，而医疗数据由于其特殊性，数据开放度较低，获取存在一定壁垒，其数据的确权与流通问题尚待解决；2) **模型层**，AI 算法、框架需要长期的研发投入，目前各大科技企业与互联网巨头均在加速布局，大模型在医疗领域的应用越发广泛，解决了一些关键问题；3) **应用层**，应用层可触达全医疗服务场景，如院内临床决策系统、手术机器人、智慧病案系统、医疗影像、药企新药研发与基因检测。

图76：基于大模型驱动的 AI 医疗产业链



资料来源：人工智能医疗器械创新合作平台，中国信通院，中国银河证券研究院

模型层，参与者主要有大型科技公司和医药类企业。大型科技公司基于自己的通用大模型，开发垂直化服务平台进行赋能，主要代表公司有英伟达、英特尔、腾讯、百度、阿里等等，比如英伟达推出了生物医药基础模型云服务平台 BioNeMo，提供生成式化学、蛋白质语言和结构预测相关开源模型，可以进行蛋白质、DNA 和生物化学数据处理，从而加速药物研发。医药类企业具有丰富的生物医学行业数据、用户资源和细分领域专业能力，以调用接口或基于开源模型自研的方式切入，微调改进自身产品或开发智能化产品。如 FermaAI、Nuance、Wondercise 集成 Chat-GPT 等模型，提供更加快捷智能的生物技术及制药咨询、自动生成病例、医学文档处理、个人健康管理功能。

表25：医疗大模型主要参与者与应用场景

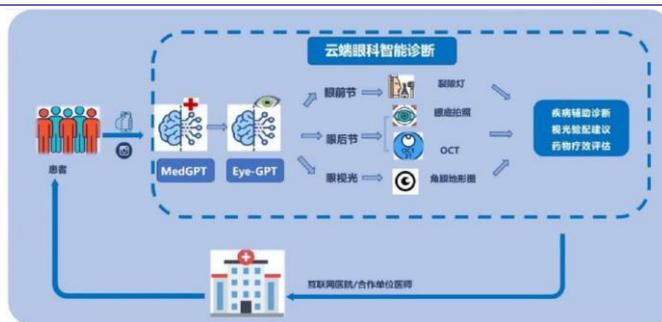
公司	医疗大模型	应用场景
英伟达	生物医药基础模型云服务平台 BioNeMo	提供生成式化学、蛋白质语言和结构预测相关开源模型，可以进行蛋白质、DNA 和生物化学数据处理，从而加速药物研发。
英特尔	科学生成式模型 Aurora genAI	根据生物学、化学、材料学、物理学、医学的科学文本、代码和结构化科学数据进行训练，可以加速科学研究和新药研发。
腾讯	腾讯医疗大模型	赋能文案生成、智能问答、病历结构化和检索、影像报告和辅助诊断。
百度	灵医大模型	已全面向公立医院、互联网医院、连锁药房等机构开放体验，并支持其他生态合作方进行 API 调用、插件集成。
阿里	通义医疗行业大模型	助力医疗、医药业务场景模型构建和全生命周期二次训练、推理、评测和模型加速应用服务。
商汤	商量大医	利用超 200 亿 tokens 的高质量医学知识数据训练而成，其聚焦智慧大健康、智慧患者服务、智慧临床以及数智建设四大领域，覆盖智能自诊、体检咨询、健康问答、导诊、预问诊、用药咨询、诊后随访管理、智慧病历、诊室听译机器人、智慧医助、智慧随访、影像报告结构化及病历结构化共 13 个细分医疗健康场景，推动医疗健康全产业链数智化转型。

资料来源：人工智能医疗器械创新合作平台，中国信通院，中国银河证券研究院

应用层，大模型的应用使得多个领域的新场景涌现，按成熟度由高到低为生命科学研究领域、药械研发领域、医学影像和图像领域、医疗问答和智能问诊领域、辅助诊疗和临床决策领域、个人健康管理领域、医学教学领域。

生命科学研究领域：AI 大模型在该领域发展较为成熟，起步早，模型数量多，迭代发展快。例如，AI 大模型可以完成蛋白质语言理解和生成的任务，协助蛋白质结构预测和从头设计合成；可以赋能 DNA/RNA 等生命组学计算，为病因推断、疾病预测、精准医疗提供新思路；可以作为科研助手辅助生物医学研究开发工作，优化科研流程。由温州眼视光国际创新中心(中国眼谷)开发的眼科专用大型语言模型 EyeGPT，目前主要应用于科研场景以及临床医疗辅助等医疗相关内容，已在研究性论文、病例报告生成、医学学术文章润色、医疗文档写作等方面有较为成熟的落地，未来将进一步向眼健康早期检测和智能诊断场景拓展。

图77：EyeGPT 眼健康智能诊断一体化



资料来源：人工智能医疗器械创新合作平台，中国信通院，中国银河证券研究院

药械研发领域的应用：AI 大模型可服务于药品和器械从研发到上市的各个环节，包括药物发现、临床前研究、临床试验、注册申请、上市后再评价等，实现提速降本增效。例如，AI 大模型可助力药物研发早期阶段，可以协助进行分子性质预测和靶点发现；可以助力药械注册和审评自动化，提升药械企业和监管机构的办公效率；可与生物医药行业数据相结合，驱动行业信息咨询服务新发展。通义医疗行业大模型由阿里云研发，是以通义生成式语言大模型为基底，在通义 Qwen 整体训练数据超过 3 万亿 tokens 基础上，融入海量医学知识文献与医疗数据所训练出来的行业大模型。通义行业大模型通过 API 与交互式问答形式提供服务,并提供用于模型二次训练与评测的完整操控平台，已经开始在行业应用层面探索落地。

图78：阿里巴巴通义大模型结构分层



资料来源：阿里巴巴，中国银河证券研究院

医学影像和图像领域：在该领域，AI 大模型可以辅助医学影像诊断分析，并可自动生成影像诊断报告。大模型基于医学影像的图文对数据集进行训练，可以将视觉理解与文本知识相结合，以对话方式解释医学图像，回答相关问题。近期，深睿医疗颅内动脉瘤 CT 造影图像辅助检测软件通过创新通道获批 NMPA 三类证,成为国内首个创新人工智能颅内动脉瘤 CT 检测系统。本次获证的颅内动脉瘤 AI 产品也是深睿医疗 Deepwise MetAI 智慧影像&大数据通用平台中睿影辅助诊断系统的重要组成部分。结合睿影系列的 ASPECT 评分、头颈 CTA、脑灌注(CTP)、脑出血等多个系统,针对缺血性脑卒中和出血性脑卒两大临床应用场景。

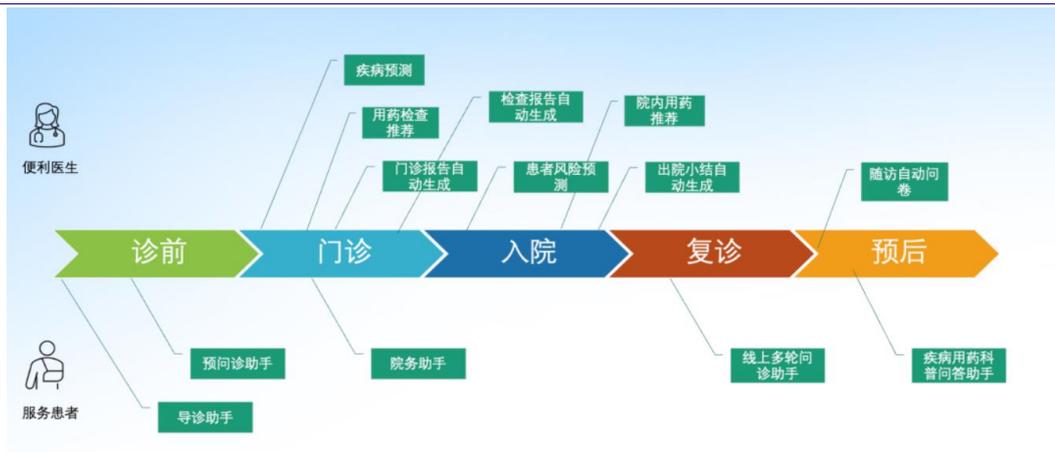
图79：深睿医疗颅内动脉瘤 CT 造影图像辅助检测软件



资料来源：新浪新闻，中国银河证券研究院

医疗问答和智能问诊领域：AI 大模型可以对话方式回答用户的医疗健康问题并支持连续自由对话和多场景的功能，并且提升了问诊类产品的准确性和智能化程度。腾讯医疗大模型包括文案生成、智能问答、病历结构化和检索、影像报告和辅助诊断等场景，为医疗决策全流程提供更精准的辅助，助力患者就医体验以及临床医生、药剂师服务效率和质量的双向提升。此外，基于腾讯医疗大模型的家庭医生助手能够提升基层服务能力，畅通医患沟通渠道，为医生提供更加智能的签约、咨询、随访、宣教能力，也为居民提供精细化的健康管理。

图80：腾讯医疗大模型的应用场景



资料来源：腾讯健康，中国银河证券研究院

辅助诊疗和临床决策领域：AI 大模型可以预测疾病风险，并生成诊断和治疗建议，为临床决策提供支持。百度灵医大模型使用了千亿 tokens 的训练语料数据，包括海量临床脱敏数据、医学知识图谱、300 万+例多模态影像数据，6 亿+条健康科普内容，70 万+ 临床试验研究信息，实现智能医生助手从辅助诊断、病历生成、文献速览等方面为医生提供服务。

图81：百度灵医大模型



资料来源：人工智能医疗器械创新合作平台，中国信通院，中国银河证券研究院

个人健康管理领域：个人健康管理领域大模型分布广泛，模型种类众多。提供的主要功能有：推动个人健康管理迈向主动化、个性化、智能化；提供营养、运动辅导服务，

多方面支持个人智能健康管理；提供更拟人化的情感疏导与支持，同时带来心理健康评估、疗愈的新工具；生成中医药处方或多维度中医养生方案；智慧康养，满足老年人陪伴、看护、社交娱乐等多样化需求。由 OPPO 打造的运动健康助手依托健康知识图谱训练的大型语言模型 andesGPT，再结合健康大数据感知与分析技术，进一步增强通用大模型的数据分析能力，打造运动健康领域的专属解决方案。

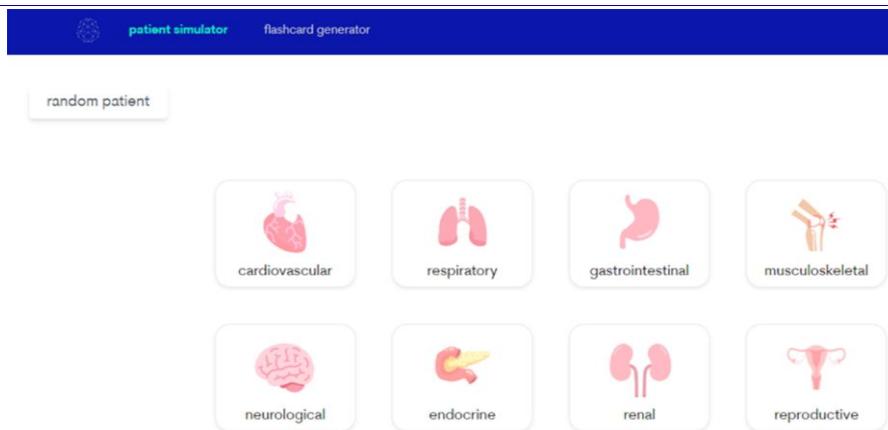
图82：OPPO Salus Copilot 产品架构



资料来源：OPPO，中国银河证券研究院

医学教学领域：模拟不同类型病人，通过提供及时反馈和个性化指导，提高学生知识和技能。Hippocratic AI 产品将医学理论知识应用于现实世界情境，利用大模型的能力虚拟患者，完成和人类医生的对话，其模拟的患者不仅具有不同疾病、性格、情绪和疾病史，还能为医学生的临床诊断技能提供反馈评价。Hippocratic AI 主要提供了 Patient simulator（类 ChatGPT 问答）和 flashcard generator 两种功能。Patient simulator 提供了心血管、呼吸、肠胃、肌肉骨骼、神经学、内分泌、肾脏、生殖 8 种模拟病人问答，并且在整个过程中病人还会表现出愤怒、急躁、焦虑等拟人化情绪，以帮助医生适应不同类型的病人。Patient simulator 还提供了考试、笔记、病例录入等学习功能。

图83：Hippocratic AI 的 Patient simulator 功能



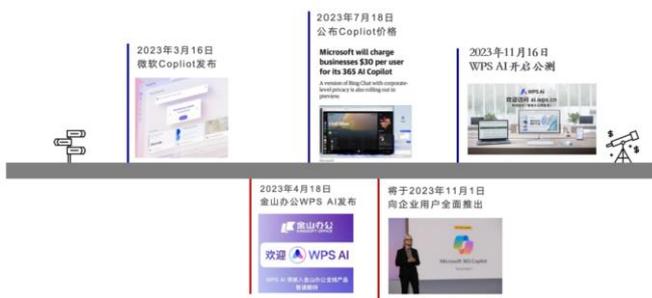
资料来源：Hippocratic AI，中国银河证券研究院

（四）AIGC+办公：ARPU 值与付费率双升

海外，微软 365 Copilot 集成 GPT-4 的功能，以聊天机器人的模式出现在产品的右侧。用户通过向其发号指令，便可自动生成文字、表格、演示文稿等内容。大模型的赋能下，简化了用户的学习路径，通过自然语言就可与机器进行交互并发号指令，将操作流程大大简化，使得工作流程与效率得到了较大的提升。2023 年 7 月 18 日，微软公布其 Copilot 价格为 30 美元/人/月，此举将大幅提升其产品单价。在近期微软发布会上，其公布 Copilot 产品将于今年 11 月 1 日正式面向全部企业用户推出，并将嵌套进全新的 Windows11 系统。我们认为，伴随着微软 Copilot 的全面落地应用，AI 对业绩、用户数据、产品优化的贡献能力将不断显现，24 年将开始逐步反应在财报上。

国内，2023 年 4 月 18 日，金山办公宣布推出 WPSAI，将应用于新一代在线内容协作轻文档，能力包括：从 0 到 1 生成内容、多轮对话以修改内容、以及处理（编辑/改写/扩充）已有文档等功能，并将陆续嵌入全线产品。7 月 6 日，WPS 面向大众招募智能办公体验官。我们认为，AIGC 对办公产品的赋能将提升用户的使用意愿与付费意愿。AI 功能主要在云端进行推理任务，用户只能通过订阅付费方式进行获取，将进一步提升公司付费率。目前公司付费率依然较低，相较于微软 18%左右的付费率仍有较大空间。此外，随着 AI 功能的上线，未来产品单价将进行提升，ARPU 值将得到提升。此外，对比竞争对手微软来说，金山的价格优势有望进一步放大。

图84：AI 在办公领域的应用时间轴



资料来源：微软，WPSAI，中国银河证券研究院

图85：WPS AI 主要功能



资料来源：WPSAI，中国银河证券研究院

AI 对收入空间贡献测算：

微软 365 Copilot：根据业绩会纪要披露，2020 年 Q1 公司 Office 商业版用户数为 2.58 亿，2021Q1、2022Q1、2023Q1 分别同比增长 15%、16%、11%得出 2023Q1 Office 商业版用户数约为 3.82 亿。Copilot 定价为 30 美元/人/月。假设：1) 2024 较 2023Q1 用户增长率为 11%；2) 年费为 30*12=360 美元/人/年；3) Copilot 渗透率中枢为 15%。基于以上假设，当用户增长率是 11%，Copilot 的用户渗透率为 15%时，明年 AI 对公司业绩贡献为 76 亿美元。

表26：微软 365 商业版收入增量测算核心假设

用户数 (亿人)	3.82
用户数增速假设	11%
2024 年用户数 (亿人)	4.24
价格增量 (美元/人/年)	120
渗透率	15%
收入增量 (亿美元)	76.32

资料来源：微软，中国银河证券研究院

表27：微软 365 商业版收入增量测算敏感性分析（亿美元）

		用户增速						
		8.00%	9.00%	10.00%	11%	12.00%	13.00%	14.00%
渗透率	5.00%	25.44	25.44	25.44	25.44	25.44	25.44	25.44
	10.00%	50.88	50.88	50.88	50.88	50.88	50.88	50.88
	15.00%	76.32	76.32	76.32	76.32	76.32	76.32	76.32
	20%	101.76	101.76	101.76	101.76	101.76	101.76	101.76
	25.00%	127.21	127.21	127.21	127.21	127.21	127.21	127.21

资料来源：微软，中国银河证券研究院

金山 WPSAI:

长期空间测算：鉴于公司暂未做出定价策略，我们对公司的预测基于以下假设：1）公司 WPS AI 功能额外定价 5 元/人/月（结合 OpenAI 与百度文心一言价格给出 API 价格参考）；年费为 60 元/人/年；2）公司仅披露了个人付费用户数，并未披露企业付费用户数量，因此，我们采用付费率做假设。假设付费率在 AI 带动下提升至约 24%（目前微软的付费率水平）；3）活跃用户数为 2.59 亿人（PC 用户与移动用户有重叠，此处取最新 2023Q3PC 端活跃用户数）；4）AI 渗透率中枢为 15%。那么，当付费率为 24%，AI 渗透率为 15%时，每年收入增量为 5.59 亿元。

表28：WPS AI 长期空间测算核心假设

用户数（亿）	2.59
价格增量	60
付费率	24%
付费用户数（亿）	0.62
渗透率	15%
每年收入增量（亿元）	5.59

资料来源：WPS AI，中国银河证券研究院

表29：WPS AI 长期空间测算敏感性分析（亿元）

		付费率						
		15%	18%	21%	24%	27%	30%	33%
渗透率	5.00%	1.17	1.40	1.63	1.86	2.10	2.33	2.56
	10.00%	2.33	2.80	3.26	3.73	4.20	4.66	5.13
	15.00%	3.50	4.20	4.90	5.59	6.29	6.99	7.69
	20%	4.66	5.59	6.53	7.46	8.39	9.32	10.26
	25.00%	5.83	6.99	8.16	9.32	10.49	11.66	12.82

资料来源：WPS AI，中国银河证券研究院

2024 年假设性测算：鉴于公司暂未做出定价策略，我们对公司的预测基于以下假设：

1) 假设公司 WPS AI 功能额外定价 5 元/人/月（结合 OpenAI 与百度文心一言价格给出 API

价格参考)；年费为 60 元/人/年；2) 公司仅披露了个人付费用户数，并未披露企业付费用户数量，因此，我们采用付费率做假设。假设付费率在 AI 带动下提升至约 12%；3) 活跃用户数为 2.59 亿人 (PC 用户与移动用户有重叠，此处取最新 2023Q3PC 端活跃用户数)；4) AI 渗透率中枢为 10%。那么，当付费率为 12%，AI 渗透率为 10%时，2024 年收入增量中枢为 1.86 亿元。

表30: WPS AI 2024 年假设性测算核心假设

用户数 (亿)	2.59
价格增量	60
付费率	12%
付费用户数 (亿)	0.21
渗透率	10%
每年收入增量 (亿元)	1.86

资料来源: WPS AI, 中国银河证券研究院

表31: WPS AI 2024 年假设性测算敏感性分析 (亿元)

		付费率						
		9%	10%	11%	12%	13%	14%	15%
渗透率	2.00%	0.28	0.31	0.34	0.37	0.40	0.44	0.47
	6.00%	0.84	0.93	1.03	1.12	1.21	1.31	1.40
	10.00%	1.40	1.55	1.71	1.86	2.02	2.18	2.33
	14%	1.96	2.18	2.39	2.61	2.83	3.05	3.26
	18.00%	2.52	2.80	3.08	3.36	3.64	3.92	4.20

资料来源: WPS AI, 中国银河证券研究院

(五) AIGC+工业: 打造新质生产力, 推动智能制造发展

AI 大模型在工业领域的应用仍处于探索阶段, 主要有以下原因:

第一, 工业领域对安全生产的要求极高。目前生成式 AI 技术通过预训练学习后推理出“合理”的答案, 这个答案可能存在逻辑漏洞或者歪曲事实, 对工业生产来说, 这种小概率的错误可能引发宕机等生产事故。第二, 目前的模型训练数据主要来源与文本、图像与视频, 工业生产中的数据是非标准化的, 并且各领域的生产流程、工业、产线、原材料等差异较大。

根据中国质量认证中心及中关村智用人工智能研究院联合发布的《产业大模型应用白皮书》, 产业大模型主要有五项指标: 1) 可控性: 大模型需要实现预测、分类与检测的准确性, 在应用过程中, 需要保证其决策的透明度和可解释性, 输出需要与社会的道德与法律标准保持一致与可控。2) 能力增强: 产业大模型需要对工作流程拥有逻辑理解, 包括对工厂数据、文本、图像的理解能力与输出能力, 突出决策辅助的准确性。3) 安全性: 需要保障工厂数据的安全与不可泄漏, 具备抵挡外部黑客攻击的能力。4) 部署的可实施性: 大模型与原有的软硬件系统需要兼容, 并且能够低成本地进行维护。5) 算力的性价比: 大模型所需的算力成本和投入需要与其产出相匹配。

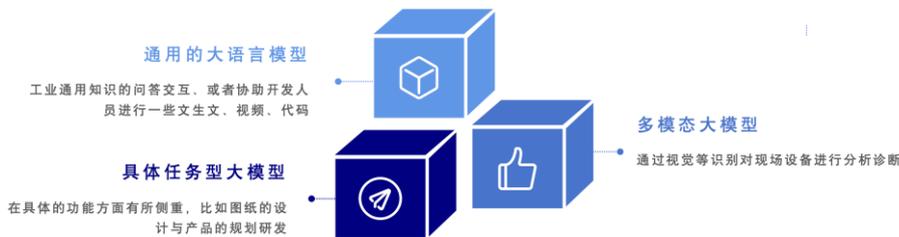
图86: AI 产业大模型的五项指标



资料来源:《产业大模型应用白皮书》(中国质量认证中心、中关村智用人工智能研究院), 中国银河证券研究院

生成式 AI 在工业领域主要通过大模型与行业小模型赋能。主要有几类赋能方式:可直接基于通用的大模型底座对行业进行赋能,例如通用型模型文心一言直接运用在工业企业运营流程中;此外,可将通用模型进行特定行业的调优,在原本的工业软件中台或者工业互联网平台以 API 接口的方式接入;或者在细分的特定生产场景进行针对性的模型开发。模型的表现形式主要有:通用的大语言模型的应用领域主要有工业通用知识的问答交互、或者协助开发人员进行一些文生文、视频、代码,提升工作效率;多模态大模型可以通过视觉等识别对现场设备进行分析诊断;具体任务型大模型在具体的功能方面有所侧重,比如图纸的设计与产品的规划研发。

图87: 大模型赋能工业的核心方式与产品形态



资料来源:中国信通院, 工业互联网产业联盟, 中国银河证券研究院

各工业软件公司相继在大模型领域进行了探索,应用场景主要体现在:第一,在产品开发阶段,生成式 AI 可以在设计的图像与文本上进行辅助,能够压缩设计和迭代的过程。第二, AI 大模型的代码能力可以协助软件工程师提升编程的效率,或者直接替代部分较为低端的代码环节。第三, AI 大模型可以辅助工厂的决策人员进行数据的追踪与分析。第四,大模型有望改变操作人员与现场机器的交互方式,通过自然语言的方式与机器进行“沟通”,降低操作的学习成本与门槛。

中控技术: 中控技术于近期宣布将推出首个面向流程工业运行优化与设计的 AI 大模型,运用海量的生产运行、工艺、设备及质量数据,自主研发生成式 AI 算法架构(AIGC),基于工业多源数据进行融合训练,建立流程工业高泛化、高可靠的大模型,为客户提供 AI+安全、AI+质量、AI+效益、AI+低碳的智能化解决方案,或有望在流程工业的效率上实现革命性的突破。

鼎捷软件：鼎捷行业大模型平台为不同行业提供专业、个性化的 AI 服务和场景化、定制化解决方案，主要功能有创意写作、知识问答、智能客服、报告生成、推理预测和总结。此外，鼎捷开发了雅典娜 ChatFile 机器人，是一款基于鼎捷知识中台和 GPT 技术构建的企业级知识机器人，主要针对 PDF，WORD，PPT，TXT，EXCEL 等非结构化文档做知识交互，实现快速、低成本构建企业知识问答系统。企业用户可通过 ChatFile 实现自然语言与知识间的交互，准确合规安全的获取及使用企业知识，提升员工工作效率及学习能力。

赛意信息：赛意信息与华为深度合作，携手华为盘古打造制造业大模型/AIGC 中台。赛意 AIGC 中台(善谋 GPT)则是深度融合赛意信息在财税、人力、营销、供应链、研发与生产制造等领域知识的最佳实践。善谋 GPT 为企业提供了多种应用场景，包括数字员工（智能引导）、智能单据（智能辅助）、企业知识库（智能问答）、AI 自动报价（智能报价）等。通过与企业现有的系统和数据集成，善谋 GPT 能够为企业提供更加个性化和智能化的服务，提高企业生产与经营管理效率。

索辰科技：索辰在 AI 的布局主要体现在它原有的仿真能力与机器人相结合。在工程设计、建模与仿真的领域中，生成式人工智能技术能够用于改进数据输入、创造场景、优化流程以及合成数据的生成。公司立足于精确的物理原理，在已有的 CAE 软件和新增的人型机器人等业务方向中利用生成式模型的最新进展，将生成式人工智能的使用作为开发和实施数字孪生技术的关键组成部分，并取得了一定的阶段性成果。目前，通过大语言模型进行仿真数据输入和仿真场景创建，结合生成式流程，能够实现从文本到仿真结果及最后的视频渲染。

汉得信息：发布汉得 AIGC 中台与燕千云 SaaS 服务，燕千云 SaaS 服务平台基于汉得 HZERO 原生态 AIGC 能力，对企业内部的 IT 服务、员工服务、客户服务进行场景重铸，并以低成本、短周期、低门槛部署解决方案，为企业提供更高效率、更好体验的“强 IT”服务体系。

表32：工业软件厂商在 AI 的布局

工业软件厂商	相关 AI 布局
中控技术	中控技术于近期宣布将推出首个面向流程工业运行优化与设计的 AI 大模型，运用海量的生产运行、工艺、设备及质量数据，自主研发生成式 AI 算法架构（AIGC），基于工业多源数据进行融合训练，建立流程工业高泛化、高可靠的大模型，为客户提供 AI+安全、AI+质量、AI+效益、AI+低碳的智能化解决方案，或有望在流程工业的效率上实现革命性的突破。
索辰科技	索辰在 AI 的布局主要体现在它原有的仿真能力与机器人相结合。在工程设计、建模与仿真的领域中，生成式人工智能技术能够用于改进数据输入、创造场景、优化流程以及合成数据的生成。公司立足于精确的物理原理，在已有的 CAE 软件和新增的人型机器人等业务方向中利用生成式模型的最新进展，将生成式人工智能的使用作为开发和实施数字孪生技术的关键组成部分，并取得了一定的阶段性成果。目前，通过大语言模型进行仿真数据输入和仿真场景创建，结合生成式流程，能够实现从文本到仿真结果及最后的视频渲染。
赛意信息	赛意信息与华为深度合作，携手华为盘古打造制造业大模型/AIGC 中台。赛意 AIGC 中台(善谋 GPT)则是深度融合赛意信息在财税、人力、营销、供应链、研发与生产制造等领域知识的最佳实践。善谋 GPT 为企业提供了多种应用场景，包括数字员工（智能引导）、智能单据（智能辅助）、企业知识库（智能问答）、AI 自动报价（智能报价）等。通过与企业现有的系统和数据集成，善谋 GPT 能够为企业提供更加个性化和智能化的服务，提高企业生产与经营管理效率。
鼎捷软件	鼎捷行业大模型平台为不同行业提供专业、个性化的 AI 服务和场景化、定制化解决方案，主要功能有创意写作、知识问答、智能客服、报告生成、推理预测和总结。此外，鼎捷开发了雅典娜 ChatFile 机器人，是一款基于鼎捷知识中台和 GPT 技术构建的企业级知识机器人，主要针对 PDF，WORD，PPT，TXT，EXCEL 等非结构化文档做知识交互，实现快速、低成本构建企业知识问答系统。企业用户可通过 ChatFile 实现自然语言与知识间的交互，准确合规安全的获取及使用企业知识，提升员工工作效率及学习能力。
汉得信息	发布汉得 AIGC 中台与燕千云 SaaS 服务，燕千云 SaaS 服务平台基于汉得 HZERO 原生态 AIGC 能力，对企业内部的 IT 服务、员工服务、客户服务进行场景重铸，并以低成本、短周期、低门槛部署解决方案，为企业提供更高效率、更好体验的“强 IT”服务体系。

资料来源：中控技术，索辰科技，赛意信息，鼎捷软件，汉得信息，中国银河证券研究院

五、投资建议

我们认为，我国数字经济发展进入快车道，数字中国按照“2522”整体架构进行布局，大力推进现代化产业体系建设，加快发展新质生产力，深化大数据、人工智能等研发应用，开展“人工智能+”行动。当下，大模型及多模态加速迭代升级，应用端加速落地，传导上游基础设施爆发增长，数字经济产业链迎来投资机会，建议关注：1) 算力基础设施厂商：中科曙光、曙光数创、软通动力、拓维信息、神州数码、烽火通信；2) 大模型及多模态厂商：科大讯飞、海康威视、拓尔思、大华股份、万兴科技、中科创达；3) 数据资源持有商及运营商：易华录、超图软件、上海钢联、深桑达 A、中科星图；4) AI+金融：恒生电子、同花顺、财富趋势；5) AI+教育：科大讯飞；6) AI+医疗：嘉和美康、卫宁健康；7) AI+办公：金山办公、彩讯股份；8) AI+工业：中控技术。

表33：推荐公司盈利预测表

证券代码	证券简称	营业收入（亿元）/增速（%）				净利润（亿元）/增速（%）				PE			
		2022A	2023E	2024E	2025E	2022A	2023E	2024E	2025E	2023E	2024E	2025E	
688111.SH	金山办公	38.85/18.44	47.03/21.05	59.92/27.42	78.46/31.93	11.18/7.33	13.69/22.47	17.73/29.58	31.93/33.86	101.8	78.57	65.18	
002230.SZ	科大讯飞	188.20/2.77	231.53/23.0	300.75/29.9	320.25/25.17	5.61/-63.94	13.12/133.8	19.55/49	25.17/47.16	80.18	50.81	65.72	
300033.SZ	同花顺	35.59/1.4	40.25/13.09	48.16/19.66	48.3/16.13	16.91/-11.51	18.26/7.95	22.39/22.63	16.13/17.66	44.63	36.4	38.50	
600570.SH	恒生电子	65.02/18.3	79.41/22.12	95.98/20.87	109.13/19.08	10.91/-25.45	18.38/68.41	22.89/24.5	19.08/23.03	30.36	24.38	19.09	
300036.SZ	超图软件	15.96/-14.9	19.74/23.68	24.66/24.97	31.19/25.4	0.015568312	2.80/182.77	3.75/33.85	25.4/32.75	35.36	26.4	18.68	
300226.SZ	上海钢联	765.67/16.41	850.38/11.0	949.83/11.7	1110.97	2.03/14.13	2.63/29.42	3.38/28.57	13.11/30.1	33.84	26.33	19.76	
300634.SZ	彩讯股份	11.95/34.71	15.74/31.65	20.50/30.28	25.48/27.34	2.25/50.91	3.09/37.12	3.97/28.66	27.34/26.72	31.07	24.18	18.82	
688246.SH	嘉和美康	7.17/9.97	8.60/20.00	10.53/22.34	14.34/27.76	0.68/37.13	0.85/25.55	1.18/38.31	27.76/43.21	56.47	40.68	20.87	
603019.SH	中科曙光	130.08/15.44	152.67/17.3	180.31/18.10	202.37/17.79	15.44/31.27	20.02/29.64	25.12/25.48	17.79/24.24	28.07	22.37	25.39	
872808.BJ	曙光数创	5.18/27.01	6.72/29.77	8.57/27.54	10.7/25.77	1.17/24.72	1.65/41.23	2.14/29.76	25.77/27.4	53.33	41.12	42.31	
688568.SH	中科星图	15.77/51.62	23.26/47.55	33.21/42.77	49.34/43.01	2.43/10.19	3.42/41.05	4.81/40.37	43.01/40.7	50.88	36.17	29.17	
300229.SZ	拓尔思	9.07/-11.81	11.92/31.34	15.40/29.25	15.67/23.75	1.28/-48.04	2.63/106.29	3.36/27.71	23.75/34.4	50.95	39.88	38.04	
688318.SH	财富趋势	3.21/-1.49	4.08/26.81	5.23/28.37	6.72/26.02	1.54/-45.48	3.07/99.69	3.95/28.91	26.02/24.49	59.28	46.08	32.22	
300496.SZ	中科创达	54.45/31.96	63.15/15.98	80.46/27.40	82.49/23.83	7.69/18.77	9.18/19.39	12.05/31.33	23.83/30.59	40.09	30.54	25.67	
002261.SZ	拓维信息	30.31/114.35	40.22/107.7	53.59/56.08	53.59/33.26	35.52/145.73	32.68/70.14	33.26/44.94	33.26/56.08	145.73	70.14	44.94	
301236.SZ	软通动力	201.1/-11.1	233.31/30.2	270.78/25.88	270.78/16.06	5.27/56.61	16.02/43.45	16.06/34.52	16.06/25.88	56.61	43.45	34.52	
000034.SZ	神州数码	1227.63/23.3	1307.23/18	1397.51/19.2	1397.51/6.91	5.94/17.38	6.48/14.65	6.91/12.29	6.91/19.24	17.38	14.65	12.29	
600498.SH	烽火通信	344.02/26.67	396.9/44.16	461.12/39	461.12/16.18	11.27/44.62	15.37/30.95	16.18/22.27	16.18/39	44.62	30.95	22.27	
300624.SZ	万兴科技	15.18/130.76	19.22/52.6	23.92/40.89	23.92/24.49	28.65/162.83	26.58/106.7	24.49/75.73	24.49/40.89	162.83	106.70	75.73	
002415.SZ	海康威视	831.66/2.14	898.60/8.05	1007.76/12.1	1139.45	128.37	163.19/27.1	184.61	13.88/17.04	19.33	17.08	15.63	
688777.SH	中控技术	66.24/46.56	116.42/22.7	150.97/28.75	150.97/29.67	8.07/37.03	35.08/26.42	29.67/22.71	29.67/28.75	32.41	26.42	20.52	
000032.SZ	深桑达 A	593.9/16.33	695.3/17.07	811.25/16.68	811.25/16.68	16.33/13.13	17.07/19.92	16.68/27.92	16.68/60.5	78.53	60.50	96.91	
002236.SZ	大华股份	328.03/7.32	378.51/15.3	441.43/16.62	441.43/16.62	7.32/53.18	15.39/47.51	16.62/56.75	16.62/19.71	-10.54	19.71	12.65	

300253.SZ 卫宁健康 34.69/12.15 42.59/22.79 52.76/23.87 52.76/23.87 12.15/3.64 22.79/5.34 23.87/7.31 23.87/32.39 41.58 32.39 45.45

资料来源: WIND, 中国银河证券研究院

六、风险提示

技术研发进度不及预期风险; 供应链风险; 政策推进不及预期风险; 下游需求不及预期风险; 行业竞争加剧风险。

图表目录

图 1: 人工智能是核心技术催化剂, 促进颠覆性技术融合	4
图 2: 四次工业革命里通用性技术的经济影响估算	5
图 3: AI 对经济增长的贡献突出	5
图 4: AGI 有望加速到来	5
图 5: DiT 模块架构	6
图 6: 将视频压缩到一个较低维空间, 再将其分解为 patches	6
图 7: 随着训练计算的增加, 样本质量显著提高	6
图 8: Sora 缺陷其一: 生成的运动违反物理原理	7
图 9: Sora 缺陷其二: 物理建模不精确、目标“变形”不自然	7
图 10: 受益于人工智能和计算机视觉, 机器人将能够在非结构化环境中经济高效地运行	10
图 11: 全球劳动力每天工作时长	10
图 12: 全球上网时长	10
图 13: Med-PaLM 2 在 MedQA 上的表现比 Med-PaLM 提高了 19% 以上	12
图 14: 结构性货币政策工具梳理 (截至 2023 年末)	13
图 15: 首轮积极财政政策期间我国基础设施处于起步阶段	15
图 16: 1998 年至今价格指数及 GDP 增速 (%)	15
图 17: “四万亿”计划的主要投资方向	15
图 18: 我国基建投资资金来源与完成额间的资金缺口 (亿元)	15
图 19: 1994 年至今广义财政净融资额变化 (亿元)	16
图 20: 城投债与政府债的融资利差	16
图 21: 2021-2024 年主要财政指标及政策表述	16
图 22: 中国数字经济规模及增速预测	19
图 23: 中国数字经济增速 (名义) 对比 GDP 增速 (名义)	19
图 24: 中美欧印 GDP、投资、人口占全球比重	19
图 25: 中美欧印互联网渗透率	19
图 26: 全球人工智能人才数量占比	20
图 27: 中美人选 AI2000 历年人数对比	20
图 28: 全球人工智能人才数量占比	21
图 29: 截至 2023 年 7 月全球大模型累计数量区域分布情况	21
图 30: 国内外代表性大模型中文测评得分趋势(23 年 7 月-24 年 2 月)	22
图 31: 全球大模型中文测评结果 (24 年 2 月)	22
图 32: 通用服务器拆解图 (以华为 TS200-2280 为例)	23
图 33: AI 服务器拆解图 (以华为 Atlas800TA2 为例)	23
图 34: 全球服务器市场规模及预测	24
图 35: 中国服务器市场规模及预测	24
图 36: 全球 AI 服务器市场规模及预测	24
图 37: 全球 AI 服务器厂商出货量及预测	24
图 38: 全球 AI 服务器出货量结构	25
图 39: 全球 AI 服务器销售额结构	25
图 40: 中国 AI 服务器市场规模及预测	25
图 41: 中国 AI 服务器出货量及预测	25
图 42: 2023 年、2025 年推理需求预测	26
图 43: 推理服务器需求推算	26
图 44: Transformer 语义分割方法	27
图 45: 不同模型训练一轮所需算力需求	28

图 46: 算力对数字经济和 GDP 的影响	28
图 47: 中国 AI 芯片市场规模.....	30
图 48: 2018-2022 年中国 AI 芯片投资情况.....	30
图 49: 华为昇腾计算产业生态	31
图 50: 昇腾全栈 AI 软硬件平台	32
图 51: 华为昇思 MindSpore 框架	33
图 52: 风冷与液冷散热能力对比	33
图 53: 制冷技术 PUE 对比	34
图 54: 冷板式液冷服务器散热原理.....	35
图 55: 曙光冷板式液冷服务器	35
图 56: 单相浸没式液冷服务器散热原理.....	35
图 57: 曙光数创全浸没式液冷服务器.....	36
图 58: 2022 年-2027 年中国液冷服务器市场规模预测.....	36
图 59: 云端 AI 预计市场规模.....	37
图 60: 云+边缘落地方向.....	37
图 61: MediaPipe LLM Inference API 适配四种模型架构.....	39
图 62: 三星手机接入谷歌 AI 大模型 Gemini	39
图 63: 2024 年起全球将加速普及 AI+PC	40
图 64: Figure01 演示视频.....	41
图 65: 美国数字战略布局历史沿革	44
图 66: 美国现阶段数字战略布局	44
图 67: 《欧盟人工智能法案》发展时间线.....	46
图 68: 《欧盟人工智能法案》的监管分级.....	47
图 69: 数据要素市场规模及增速预测.....	52
图 70: 尽管 phi-1 的训练规模较小, 但其表现明显优于其他模型.....	52
图 71: 同花顺 AI 开放平台核心技术.....	54
图 72: 2003 年至 2022 年国家教育公共财政支出 (亿元)	56
图 73: Sora 生成视频截图.....	57
图 74: 2021-2023 年中国 AI 医疗不同赛道融资情况 (单位: 个)	58
图 75: 智能医疗行业阶段年复合增长率.....	58
图 76: 基于大模型驱动的 AI 医疗产业链.....	58
图 77: EyeGPT 眼健康智能诊断一体化.....	59
图 78: 阿里巴巴通义大模型结构分层	60
图 79: 深睿医疗颅内动脉瘤 CT 造影图像辅助检测软件	60
图 80: 腾讯医疗大模型的应用场景	61
图 81: 百度灵医大模型.....	61
图 82: OPPO Salus Copilot 产品架构	62
图 83: Hippocratic AI 的 Patient simulator 功能	62
图 84: AI 在办公领域的应用时间轴	63
图 85: WPS AI 主要功能.....	63
图 86: AI 产业大模型的五项指标	66
图 87: 大模型赋能工业的核心方式与产品形态	66

表格目录

表 1: OpenAI Sora 对比同类文生视频大模型	7
表 2: 主流大模型上下文输入长度对比	9
表 3: AGI 带来千行百业颠覆性革命	9
表 4: 脑机接口系统关键指标	11
表 5: 2019 年至今历次专项债用途扩容	17
表 6: 关于国有资本经营预算的定义变化	18
表 7: 中美人工智能产业发展状况对比, 截至 2023 年 7 月	21
表 8: 国内外代表性大模型参数及预训练数据对比	21
表 9: 部分国内代表性大模型 SuperCLUE 中文测评基准得分(23 年 7 月-24 年 2 月)	22
表 10: 服务器架构	23
表 11: 英伟达 H20、L20、L2 性能参数对比	29
表 12: 华为昇腾 910b 与英伟达 A100 参数性能对比	30
表 13: 华为服务器对比	31
表 14: 全国主要数据中心 PUE 要求	33
表 15: AI 多模态大模型企业 20 强	38
表 16: 核心数据、重要数据定义	43
表 17: 美国人工智能行政命令的主要规范限制	44
表 18: 除行政法令外的美国人工智能政策	45
表 19: 《生成式人工智能服务管理暂行办法》的四项制度	49
表 20: 人工智能相关部门规章与地方政策梳理	50
表 21: 数据要素市场规模及增速预测	51
表 22: Bloomberg 在金融权益类测试中较大模型表现更优	53
表 23: 国内金融大模型进展	55
表 24: 国内金融大模型主要应用场景	56
表 25: 医疗大模型主要参与者与应用场景	59
表 26: 微软 365 商业版收入增量测算核心假设	63
表 27: 微软 365 商业版收入增量测算敏感性分析(亿美元)	64
表 28: WPS AI 长期空间测算核心假设	64
表 29: WPS AI 长期空间测算敏感性分析(亿元)	64
表 30: WPS AI 2024 年假设性测算核心假设	65
表 31: WPS AI 2024 年假设性测算敏感性分析(亿元)	65
表 32: 工业软件厂商在 AI 的布局	67
表 33: 推荐公司盈利预测表	68

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。

首席经济学家章俊；

宏观组分析师：许冬石、詹璐；首席计算机分析师吴砚靖；

计算机分析师邹文倩；计算机分析师李璐昕。

免责声明

本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。

评级标准	评级	说明
评级标准为报告发布日后的 6 到 12 个月行业指数（或公司股价）相对市场表现，其中：A 股市场以沪深 300 指数为基准，新三板市场以三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）为基准，北交所市场以北证 50 指数为基准，香港市场以摩根士丹利中国指数为基准。	行业评级	推荐：相对基准指数涨幅 10% 以上
		中性：相对基准指数涨幅在 -5% ~ 10% 之间
		回避：相对基准指数跌幅 5% 以上
公司评级		推荐：相对基准指数涨幅 20% 以上
		谨慎推荐：相对基准指数涨幅在 5% ~ 20% 之间
		中性：相对基准指数涨幅在 -5% ~ 5% 之间
	回避：相对基准指数跌幅 5% 以上	

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：程曦 0755-83471683chengxi_yj@chinastock.com.cn

苏一耘 0755-83479312suyiyun_yj@chinastock.com.cn

上海地区：陆韵如 021-60387901luyunru_yj@chinastock.com.cn

李洋洋 021-20252671liyongyang_yj@chinastock.com.cn

北京地区：田薇 010-80927721tianwei@chinastock.com.cn

唐曼玲 010-80927722tangmanling_bj@chinastock.com.cn