2024 年 04 月 03 日

计算机行业研究

买入（维持评级）

行业深度研究

证券研究报告

计算机组
分析师：孟灿（执业 S1130522050001）　联系人：赵彤
mengcan@gjzq.com.cn　zhaotong3@gjzq.com.cn

如何实现 AGI：大模型现状及发展路径展望

投资逻辑：

目前大模型能力仍处于 Emerging AGI 水平，就模型成熟度而言，语言大模型>多模态大模型>具身智能大模型。根据 DeepMind 的定义，AGI 应能够广泛学习、执行复杂多步骤的任务。模型的 AGI 水平可分为 Level-0 至 Level-5 共 6 个等级，现阶段大模型在处理任务的广泛性上还有很大提升空间，即使是国际顶尖的大模型也仍处于 Level-1 Emerging AGI 阶段。不同类型大模型成熟度差异较大，目前大语言模型能力相对完善，落地应用场景丰富，底层技术路线较为成熟；多模态大模型已经能够面向 B\C 端推出商业化产品，但细节优化空间较大；具身智能类大模型还在探索阶段，技术路线尚不清晰。

现阶段讨论 AGI 能力提升仍需聚焦于多模态大模型的训练和应用。目前学界和业界重点关注 Scaling Law 的有效性，以及模型算法的可能改进方向。

- Scaling Law 仍有深入空间。根据 OpenAI 研究，随模型参数量、数据集规模、训练使用的计算量增加，模型性能能够稳步提高，即 Scaling Law。从训练样本效率、训练时长、各类资源对模型的贡献维度来看，目前 Scaling Law 仍是提高模型性能的最优方法。OpenAI 测算在模型参数量扩展到 88 万亿及之前，Scaling Law 依旧有效，则中短期仍可延续此路线进行训练。
- 模型骨干网络架构尚未演变至终局，微调及稀疏结构成为提升模型性能的重要方法。目前主流大模型均采用 Transformer 作为底层骨干网络，但针对编码器\解码器选择、多模态融合、自注意力机制等方面的探索仍在持续推进。微调使用更小的数据量、更短的训练时间，让模型能够适应下游任务，以降低边际落地成本。以 MoE 为代表的稀疏结构通过分割输入任务并匹配专家模型，能够提高模型的整体性能。

开源模型性能优化速度快于闭源模型。我们认为，目前第一梯队 AI 大模型纷纷进军万亿参数，且不远的将来大模型将逐步逼近十万亿参数收敛值，对于本轮 AI 浪潮而言，找场景或优于做模型。在场景选择方面，对“幻觉”容忍度高且能够替代人工的场景可实现应用率先落地，如聊天机器人、文本/图像/视频创作等领域；而对“幻觉”容忍度较低的行业需要等待大模型能力提升或使用更多场景数据训练。

投资建议

算法、数据、算力是影响模型性能的关键因素，相关企业能够直接受益于大模型训练的持续推进，推荐国内 AI 算法龙头科大讯飞等，建议关注数据工程供应商以及算力产业链相关公司。对于行业类公司而言，寻找通过 AI 赋能带来效率提升的场景更为重要，建议关注 AI+办公领域的金山办公、万兴科技，AI+安防领域的海康威视，AI+金融领域的同花顺等公司。

风险提示

底层大模型迭代发展不及预期；国际关系风险；应用落地不及预期；行业竞争加剧风险。

内容目录

图表目录

2022 年 11 月 ChatGPT 推出后，自然语言处理领域取得重大突破，正式进入大模型时代，2023 年被称为“大模型元年”；2023 年 3 月，具备多模态能力的 GPT-4 惊艳发布，海内外科技巨头、研究机构等纷纷跟进；至 2024 年 2 月 Sora 面世，大模型在视频生成领域实现代际跃迁，虚拟现实成为可能。在此背景下，学界和业界对于大模型终局，即是否能够实现 AGI（Artificial general Intelligence，通用人工智能）的讨论热度日益提升。

本文主要盘点目前各类主流大模型性能情况，试图讨论大模型性能提升并最终实现 AGI 的可能路径，并分析在实现 AGI 过程中的相关产业链投资机会。

1. 距离 AGI 还有多远：语言大模型较为成熟，处于 Emerging AGI 水平

根据 DeedMind 的创始人兼首席 AGI 科学家 Shane Legg 的定义，AGI 能够执行一般人类可完成的认知任务、甚至超越这个范围。具体而言，AGI 应能够学习广泛任务，能够执行复杂、多步骤的任务。DeepMind 根据 AI 模型性能和学习处理任务的广泛性对 AGI 水平进行分类，从 Level-0 无人工智能，到 Level-5 超越人类共 6 个等级。

图表1：AGI 可以根据性能和广泛性划分为 6 个等级

等级	主要特征
Level-0 无人工智能（Narrow Non-AI）	·只能完成明确定义的任务，比如计算器软件或编译器
Level-1 初现（Emerging AGI）	·性能相当于或略优于一个不熟练的人类。比如一些前沿语言模型在某些任务上已经达到了初现 AGI 的水平
Level-2 熟练（Competent AGI）	·至少能够在大多数任务上达到熟练人类的水平。目前的前沿语言模型在某些任务上已经接近熟练 AGI 的水平
Level-3 专家（Expert AGI）	·在大多数任务上能够达到专家人类的水平
Level-4 大师（Virtuoso AGI）	·在大多数任务上能够达到顶尖人类的水平
Level-5 超越人类（Superhuman AGI）	·在所有任务上都能超过 100% 的人类

来源：《Levels of AGI: Operationalizing Progress on the Path to AGI》，国金证券研究所

现阶段大模型在处理任务的广泛性上还有很大提升空间，虽然 GPT-4、Gemini 1.5、Claude 3 等模型已经能够处理文本、图像、视频等多模态输入，但尚未具备独立决策和执行行动的能力。此外，现阶段更多的模型仍聚焦在某单一领域进行性能提升，比如 Kimi 在处理长文本输入领域表现突出，但尚不能进行图片生成；Sora 能够高质量完成文生视频任务，但不具备问答功能。因此，现阶段评价大模型性能情况、分析模型演进方向，仍需根据模型专长领域进行分类。

图表2：大模型可根据功能进行分类

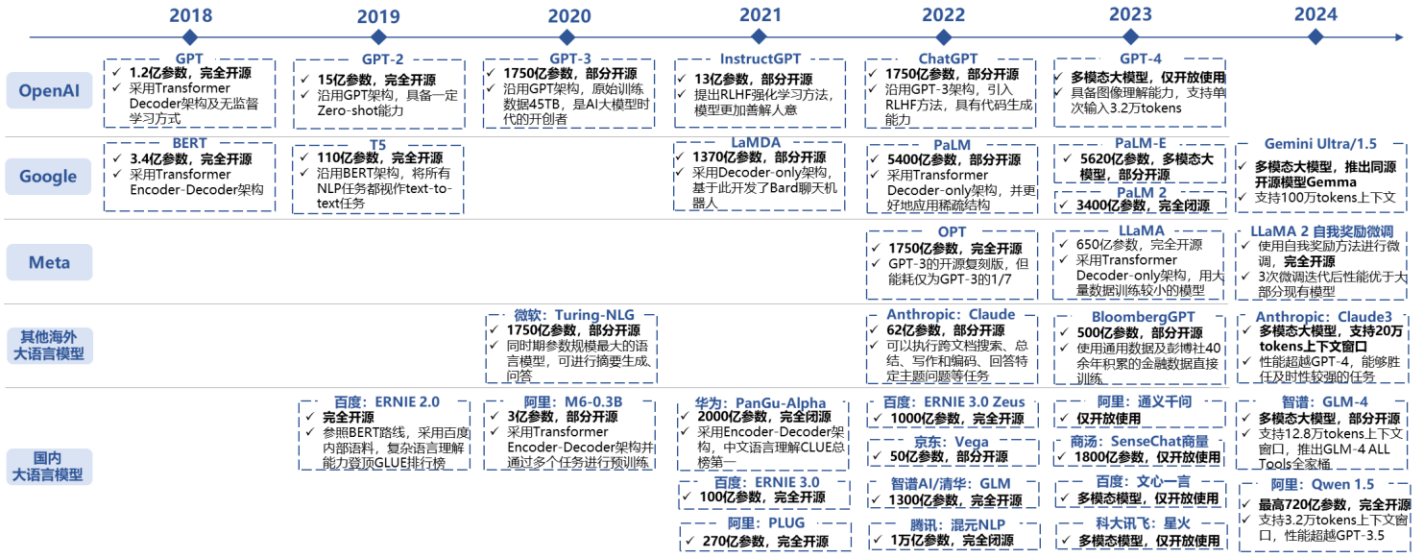
模型分类	主要内容	代表模型
语言大模型	·专注于处理自然语言，能够理解、生成和处理大规模文本数据 ·用于机器翻译、文本生成、对话系统等任务	ChatGPT、Llama
视觉大模型	·专注于计算机视觉任务，如图像分类、目标检测、图像生成等 ·能够从图像中提取有关对象、场景和结构信息	ViT、SAM
多模态大模型	·能够处理多种不同类型的数据，如文本、图像、音频等，并在这些数据之间建立关联 ·多模态大模型能够处理文图融合、图像描述、文生视频等任务	GPT-4、Claude3
策略大模型	·专注于进行决策和规划，能够在面对不确定性和复杂环境时做出智能决策，可用于机器人控制	AlphaGo、RT-1/2/H

来源：金科应用研究院公众号，国金证券研究所

- 在语言大模型以及偏重问答能力的多模态模型领域，自 2020 年 GPT-3 发布后进入爆发期，各主流玩家加速模型迭代，包括 OpenAI 的 GPT 系列、Google 的 Gemini 系列、Meta 的开源 LLaMA 系列等。目前定量测评分数最高的为 Anthropic 旗下的 Claude 3 Opus，在 MMLU（Undergraduate Level Knowledge）、GSM8K（Grade School Math）、MGSM（Multilingual Math）等多个测试项目中准确率超过 85%；模型参数量最高的为 23 年 3 月谷歌发布的 PaLM-E，参数量达到 5,620 亿，是 ChatGPT 的 3.2 倍，模型能够理解自然语言及图像，还可以处理复杂的机器人指令；谷歌于

24 年 2 月发布的 Gemini 1.5 能够处理的上下文长度高达 100 万 tokens（相当于 70 万单词，或 3 万行代码，或 11 小时音频，或 1 小时视频），为目前长文本处理能力的上限。

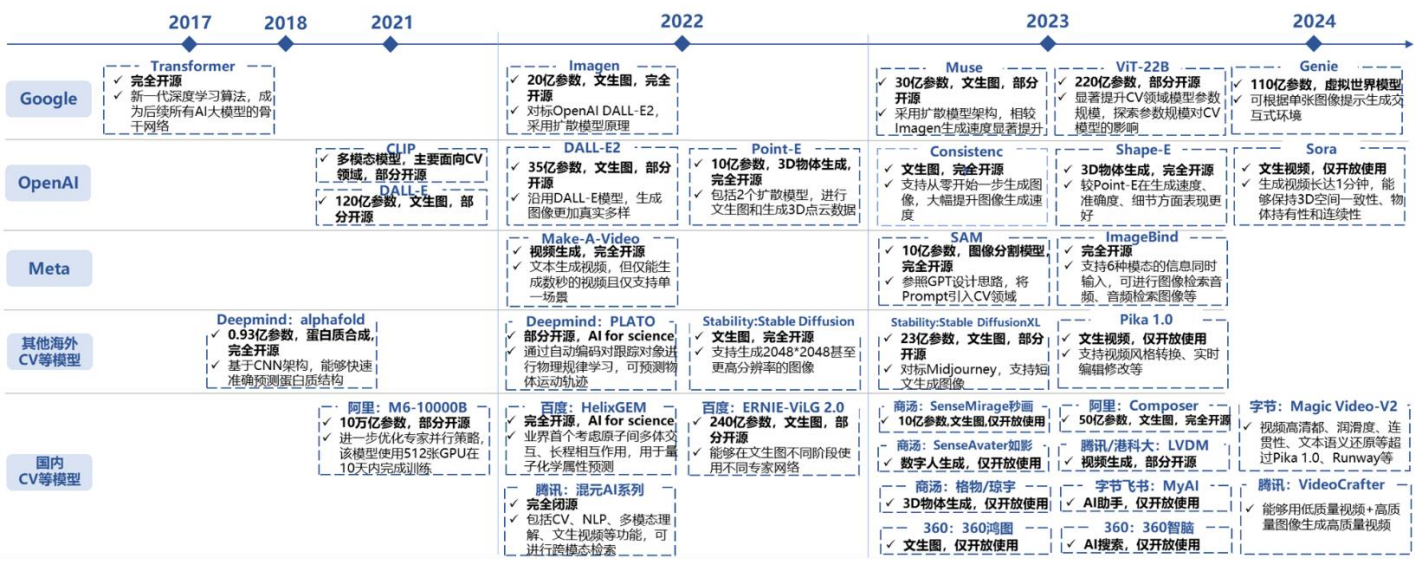
图表3：海内外语言及多模态大模型进展概览



来源：《Large Language Models: A Survey》，《A Survey of Large Language Models》，洞见学堂公众号，机器之心公众号，级市平台公众号，新智元公众号，阿里云开发者社区，京东技术公众号，中国科学基金公众号，数据派 THU 公众号，浙江省软件行业协会公众号，深圳大学可视计算研究中心公众号，量子位公众号，钛媒体 AGI 公众号，彭博 Bloomberg 公众号，腾讯科技公众号，百度 AI 公众号，鹏城实验室公众号，CSDN 公众号，文心大模型公众号，中国人工智能学会公众号，腾讯开发者公众号，阿里云公众号，商汤智能产业研究院公众号，36 氪，科大讯飞公众号，科大讯飞开发者平台，GLM 大模型公众号，阿里通义千问公众号，国金证券研究所

- 文生图、文生视频类模型可追溯至 2014 年的 GAN 框架，2021 年 OpenAI 发布 DALL-E 后图像生成类模型开始爆发，包括谷歌的 Imagen、OpenAI 的 DALL-E 2、Stability 旗下的 Stable Diffusion；至 2023 年文生图功能与大语言模型相结合，并出现文生视频技术，24 年 2 月 OpenAI 发布文生视频模型 Sora，在生成视频长度和质量上均为目前最优水平。

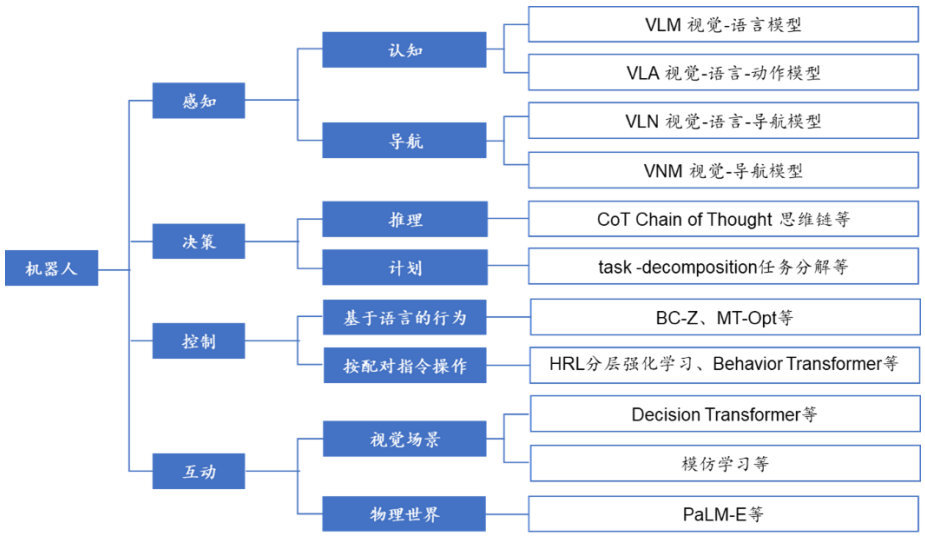
图表4：海内视觉及其他大模型进展概览



来源：《Large Language Models: A Survey》，《Improved protein structure prediction using potentials from deep learning》，《High-Resolution Image Synthesis with Latent Diffusion Models》，机器之心公众号，新智元公众号，信息与电子工程前沿公众号，级市平台公众号，AI 科技评论公众号，AIGC 开放社区公众号，腾讯研究院公众号，中国生物技术网公众号，数据派 THU 公众号，阿里云公众号，智源社区公众号，百度 AI 公众号，中国企业家俱乐部公众号，商汤科技 SenseTime 公众号，商汤智能产业研究院公众号，AIGC 视界公众号，飞书公众号，搜狐科技公众号，AIGC Research 公众号，智东西公众号，国金证券研究所

- 机器人模型包括感知、决策、控制、交互 4 个部分，涉及视觉、图像、声音、导航、动作等多个模态，在实际应用中需要根据特定的环境、动作、障碍、反馈等数据进行决策，因此，机器人对算法的跨模态、泛用性要求更高。

图表5：机器人涉及到的模型种类较多



来源：《Large Language Models for Robotics: A Survey》，国金证券研究所

将语言大模型的底层框架和训练方式应用于机器人的感知、决策、控制成为现阶段重要趋势。2021 年 OpenAI 推出基于 Transformer 架构和对比学习方法的 VLM（视觉-语言模型）CLIP；2022 年起，谷歌先后推出 RT-1/RT-2/RT-X/RT-H 系列模型，同样采用 Transformer 架构，能够将语言描述的任务映射为机器人行动策略；24 年 3 月，初创公司 Figure 与 OpenAI 合作推出机器人 Figure01，由 OpenAI 提供视觉推理和语言理解能力，Figure01 能够描述看到的一切情况、规划未来的行动、语音输出推理结果等。

图表6：将 Transformer 架构应用于机器人决策、控制等成为现阶段重要趋势

模型名称	发布时间	发布机构	功能类别	主要内容
CLIP	2021	OpenAI	感知-VLM	· 网络结构主要包含 Text Encoder 和 Image Encoder 两个模块，分别提取文本和图像特征，然后基于比对学习让模型学习到文本-图像的匹配关系； · CLIP 使用大规模数据(4 亿文本-图像对)进行训练，基于海量数据，CLIP 模型可以学习到更多通用的视觉语义信息，可应用于图像文本匹配、图像文本检索等任务。
LM-Nav	2022	谷歌	计划	· LLM\VLM\VNM 三个模型的结合，LLM 用于提取指令中的地标，VLM 用于将文本地标与图像关联，而 VNM 用于执行导航任务； · 系统以目的地环境的初始观察结果、以及用户给的文本指令作为输入，通过系统中的三个预训练模型得出执行计划。
RT-1	2022	谷歌	决策、控制	· 建立在一个 transformer 架构上，该架构从机器人相机中获取瞬时图像以及以自然语言表达的任务描述作为输入，并直接输出 tokenized 动作； · RT-1 可以以 97% 的成功率执行 700 多个训练指令，并且可以泛化到新的任务、干扰因素和背景。
PaLM-E	2023.3	谷歌	感知-VLM、控制	· 通过 PaLM-540B 语言模型与 ViT-22B 视觉 Transformer 模型相结合，PaLM-E 最终的参数量高达 5620 亿，其训练数据为包含视觉、连续状态估计和文本输入编码的多模式语句； · PaLM-E 不仅可以指导机器人完成各种复杂的任务，还能生成描述图像的语言。
RT-2	2023.7	谷歌	感知、决策、控制	· 使用 Transformer 架构的视觉-语言-动作模型，能够从网络和机器人数据中进行学习，并将这些知识转化为机器人可以控制的通用指令 · 在机器人训练中未见过的场景中，准确性由 RT-1 的 32%提高到 62%
RT-X	2023.10	谷歌	感知、决策、控制	· 由基于 Transformer 的 RT-1-X 模型和视觉语言动作模型 RT-2-X 组成。RT-1-X 模型在特定任务上的平均性能比 RT-1 模型和原始模型提升 50%。RT-2-X 的涌现能力约为 RT-2 的 3 倍，动作指令也可从传统的绝对位置拓展至相对位置
RT-H	2024.3	谷歌	感知、决策、控制	· 能通过将复杂任务分解成简单的语言指令，再将这些指令转化为机器人行动，来提高任务执行的准确性和学习效率； · RT-H 的 MSE 比 RT-2 低大约 20%，这表明行动层级有助于改进大型多任务数据集中的离线行动预测

来源：极市平台公众号，DeepTech 深科技公众号，机器之心公众号，OSC 开源社区公众号，国金证券研究所

按照 DeepMind 的 6 级 AGI 水平分类，目前国际顶尖大模型仍处于 Level-1 Emerging AGI