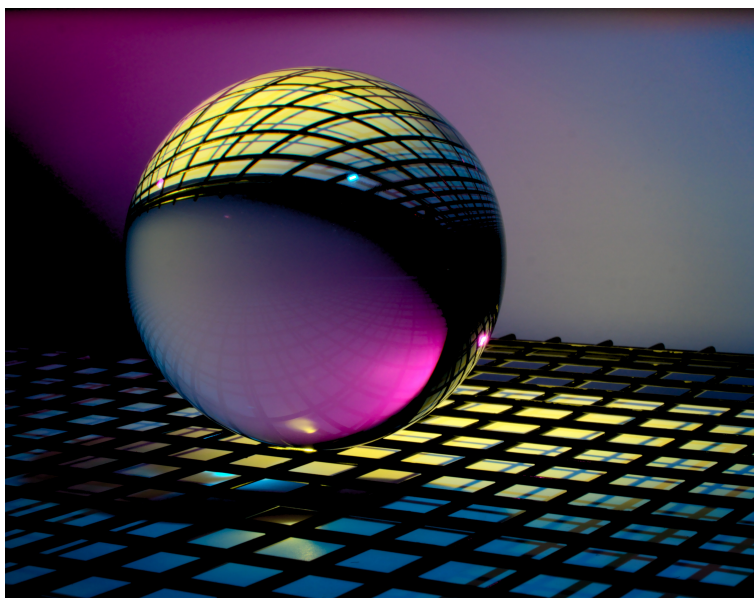


Artificial Generate Intelligence(AGI) 行业分析

张桐林



题目

Step 1 简要综合行研

1. 整合信息得到一个相对简答的 AGI 行研报告。
2. 自主学习并选择一个逻辑去表达你的 AGI 行业研究成果。
 - (a) 包含中国及北美各自的行业格局的分析。
 - (b) 选择该逻辑组织信息的原因。
 - (c) 得到至少 3 个行研得到的洞见。

Step 2 模拟决策

1. 假设有 1000 万资金，基于 Step 1 形成一个投资策略。
 - (a) 二级市场投资占比 $\leq 20\%$, 标注理由。
 - (b) 一级市场股权投资或者孵化策略 $\leq 100\%$ 。
 - (c) 解释一级市场股权投资的细分赛道和下注方式。
 - (d) 如上策略不需要考虑实际操作难度和成本。

2. 条件不变，资金变为 1 亿元人民币。

Step 3 信息趣味

1. 如果可以和 Sam Altman 等 AGI 领域的人交流，最想问的 5 个问题？
2. 推荐最受益的信息源，包括最看好的 VC、最欣赏的从业者，每个不超过 5 个。并阐释理由。

Step 4 信息推送

模拟一天的推送信息，格式和形式不限。

1 简要 AGI 综合行研

* 注：因为大模型是迈向 AGI 目前为止可行性最高的实现路径，因此以下分析和研究均围绕大模型或者基于大模型的多模态模型展开。

1.1 大模型发展标志性事件梳理

1. 2015 年，OpenAI 成立。
2. 2017 年，Google 推出 Transformer 模型，参数量首次达到上亿规模。
3. 2018 年，Google 推出 Bert 模型，参数量达到 3 亿规模；OpenAI 推出 GPT。
4. 2019 年，OpenAI 推出 GPT-2 模型，参数量达到 15 亿规模。
5. 2020 年，OpenAI 推出 GPT-3 模型，参数量达到 1750 亿；微软和英伟达发布 MT-NLG 模型，参数量达到 5300 亿。
6. 2021 年，Google 发布 Switch Transformer 模型，参数量达到 1.6 万亿。
7. 2022 年，OpenAI 推出基于 GPT-3.5 的 ChatGPT，成为现象级产品。
8. 2023 年，OpenAI 推出 GPT-4，参数规模和细节尚未公开。

大 (语言) 模型作为人类在人工智能领域最先进的产出，同时也是最接近实现 AGI 的产品，其发展极其迅速。并且有进一步加速的趋势。

1.2 大模型时代下的核心竞争力

1. 算法
2. 数据
3. 算力

算法或者模型本身，本质是其相关的研发实力，（同时也是在 AGI 领域科技树的某个技术分支的押注）是一个公司的最强的壁垒和发展护城河。越先进的模型，越可能打造出优质的产品，才能吸引到广大消费者群体和后续的商业化，因此算法是做 AGI 公司的核心竞争力。

数据是训练的基础，训练数据质量的好坏、范围直接影响模型训练的结果，从这点来看，模型决定了效果的上限，但是数据决定了效果的实际完成度。当数据和模型都配套好才可以发挥模型的最好的效果。

算力是发展大模型的前提。在当前的模型参数下，没有庞大算力的公司意味着可能直接错失进入 AGI 行业的入场券，这也意味着当前 AGI 行业只能是算力充足、资金雄厚、研发实力强劲的巨头可以参与。同时因为 NVIDIA 的 A100 销售禁令，可能会造成国内算力缺口。

1.3 中国和北美大模型竞争格局

首先需要理清一个逻辑是：不是领先的大模型占据高端市场，落后一些的大模型占据低端市场，而是领先的大模型几乎会占领一切市场。和传统的高端——中端——低端市场有与之相对应的高端——中端——低端消费群体不同，民用大模型的使用成本已经降低到一个较低的水平（按照 GPT-4 的订阅费用来衡量），因此高端中端低端用户都可以消费起行业最领先的模型和产品，同时因为行业头部的模型（GPT-4）的使用体验要远大于中部的模型（百度的文

心一言、阿里的通义千问等国产大模型),会导致各个市场的用户会集中使用最好的模型(除非有不可抗因素,如 IP 禁用、大陆地区封号等等)。

因此这对于中国发展大模型是有严重劣势的,因为国产大模型可能和北美有 1-2 代的版本代差,而这个追赶期目前仍不确定。“李彦宏称自己前段时间接受采访时说跟 *ChatGPT* 的差距大约是两个月,有点断章取义,因为自己后面紧接着说:‘这不是重点,重点是这两个月的差距我们要用多长时间才能赶上,也许很快,也许永远也赶不上。’”

我们从“大模型时代下的核心竞争力”梳理的三个角度来进行分析。

模型:

- 北美有顶流——OpenAI 的 GPT-4 及 GPT-3.5,号称 ChatGPT 的最强竞品——anthropic 的 Claude,传统 AI 巨头 Meta 的 LLaMA,Amazon 的 Amazon Titan 等等。
- 国内有百度的文心一言、阿里的通义千问、华为的盘古、科大讯飞的星火认知大模型、商汤的 SenseChat、360 智脑、腾讯的混元等。

由于大部分均仍在内测阶段,无法对其进行完整的测试。但是根据一份来源于真格基金的测评来看,国内最强的模型是百度的通义千问,综合表现较好的商汤的 SenseChat。但是仍然不及 ChatGPT-3.5。从直接的测试结果来看,国内最好的模型效果是 ChatGPT-3.5 的 75% 左右。

数据:从中文语料库和英文语料库来看,都应该是足够训练大模型的,但是国内的数据标注产业可能相较于北美较为落后。

算力:算力问题可能是国内最明显的短板,最主要的原因就在于 Nvidia 的 A100 禁令,国内有超大规模算力(1 万块 A100 规模)的公司屈指可数。

1.4 洞见

1. 整体来看,大模型产业北美从各个方面都处于领先地位,国产大模型在追赶,但是处于一个十分危险的境地(好在有 Great firewall)。这次的 AGI 竞赛可能会直接关系到一些互联网科技公司的生死存亡。
2. 国内有可能的发展是数据标注行业,虽然相对北美较为弱势,但是行业门槛较低,如果发挥一定人力成本优势,或许存在发展空间。
3. 比较看好的是百度和华为的模型,因为具有先发优势,同时两家都在 AI 领域也有积累。其次是字节的模型,字节在 AI 领域的积累也比较深厚,但是由于发展时间较晚,可能还存在一定的落地期。
4. 在这种情况下,大厂可能需要两条腿走路——既要发展自己的大模型,又要积极支持和收购在技术路线上有正确探索的中小公司。
5. 在 GPT 部署使用场景下,云服务和向量数据库可能会更大规模使用,需要关注有相关业务的公司。

2 模拟决策

2.1 1000 万人民币筹码的投资策略

1000 万的资金按照 2:8 的二级市场投资和一级市场投资划分。个人的建议如下：

1. 200 万中按照 4:2:2:1:1 的比例分别投资 Nvidia、Microsoft、Google、Amazon、百度股票。
2. 800 万中投资 3-4 家做**数据标注**的 Startup 的天使轮。* 如果有可能，投资 1-2 个**非常规半导体算力**的 Startup 天使轮。或者一些做**向量数据库**的 Startup 天使轮。前提是技术路线可落地、前景好、靠谱等。

2.2 1 亿元人民币筹码的投资策略

相较于 1000 万筹码，1 亿元可以到达给几个小型公司 A——B 轮融资的程度。但是个人认为，和 1000 万人民币级别的投资策略不会有太大区别。因为可投资的范围（一级市场）到不了做算力、模型的公司上。可能到 20 亿人民币规模会存在一些明显的改善，因为当前做的一些比较优秀的公司估值普遍在 20 亿美元以上。2023 年 3 月 9 日 *Anthropic* 筹资 11 亿美元，投资者包括谷歌、*Skype* 联合创始人 *Jaan Tallinn* 和 *FTX* 前首席执行官 *Sam Bankman-Fried* 等。因此对于 1 亿人民币规模的筹码而言，个人的建议如下：

1. 20% 资金按照 4:2:2:1:1 的比例分别投资 Nvidia、Microsoft、Google、Amazon、百度股票。
2. 80% 资金投资 3-4 家做**数据标注**的 Startup 的 A 轮。* 如果有可能，投资 1-2 个**非常规半导体算力**的 Startup 的 A 轮。或者 1-2 个做**向量数据库**的 Startup 的 A 轮。前提是技术已经有一定落地，发展成本可控、前景较为明确，靠谱等。

3 信息趣味

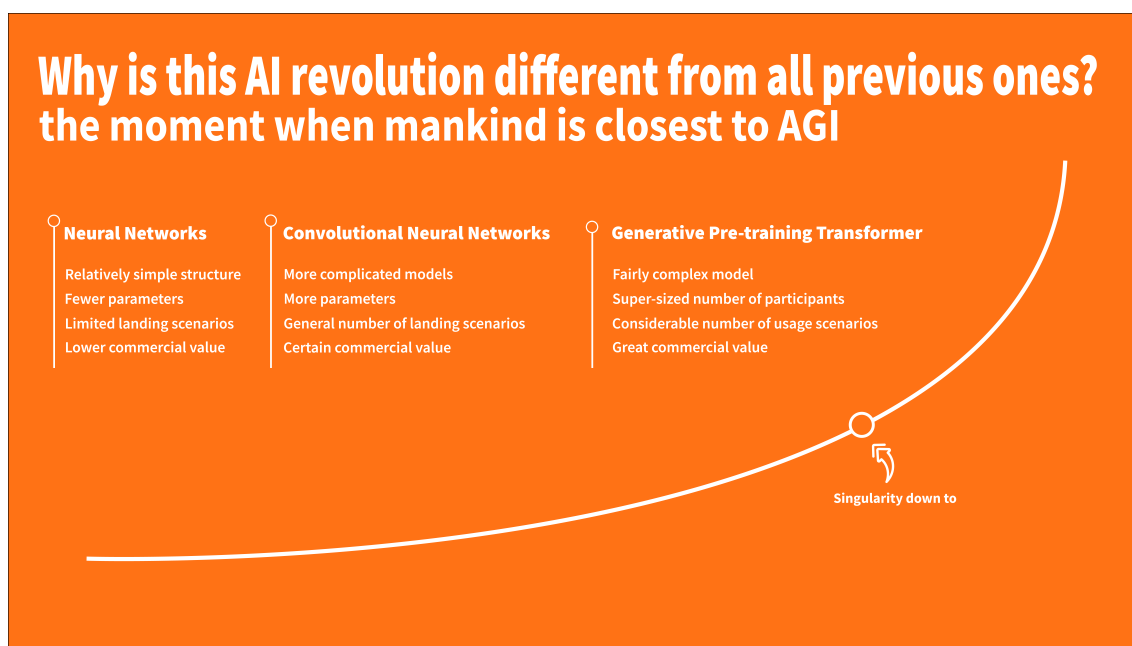
3.1 向 AGI 领域相关人员提问

1. openAI 的 GPT 如何解决用户隐私数据安全? (向 Sam Altman)
2. 您真的认为 AGI 会失控吗? 如果是的话, 人类为什么控制不了它呢? (向 Elon Musk)
3. 您曾否定了自回归模型, 那您是如何看待 GPT-4 的成果和括“涌现”现象的发生呢? (向 Yann LeCun)
4. 您觉得您和 OpenAI 相比, 百度在技术领域最大的优势是什么? (向李彦宏)
5. 国内互联网公司其实较早都布局了大模型, 但是字节今年才成立相关研究团队, 明显慢于同行, 您觉得其中的原因是什么呢? (向 Liang Rubo)

3.2 推荐的信息源

- 1. tier 0: Github Daily Trending (Don't talk, show me your code.)
- 2. tier 1: ArXiv.org (一手资源、及时性更强)
- 3. tier 2: ICML、ICLR、NeurIPS (三大顶会); CVPR、ICCV、ECCV (计算机视觉领域); ACL、EMNLP、NAACL、COLING (NLP 领域) 等 (一手资源、及时性较弱)
- 4. tier 3: MIT Technology Review、Google AI Blog、OpenAI Blog (二手资源、但是技术针对性较强); AI 领域著名从业者的推特的推文等。
- 5. tier 4: forbes、VentureBeat (二手或三手资源、技术针对性较弱)
- 最看好的 VC 机构:
 1. Sequoia Capital: 投资了最知名的一些企业, 是最经典的、最著名的、最成功的 VC 之一。
 2. a16z: 投资过一批经典的科技企业、在 Web3.0 领域布局很深。
 3. Y Combinator: 硅谷最知名的创业孵化器。
 4. Accel: 投资过一批经典的科技企业。
- 最欣赏的从业者:
 1. Elon Musk: 马斯克是 OpenAI 的联合创始人 Tesla 和 Neuralink 的 CEO。Tesla 的 FSD 和 Neuralink 的 OS 都是实际的、潜在的或者已经大规模落地的、技术上领先的 AI 应用。
 2. Sam Altman: OpenAI 创始人, 没有股权, 因为热爱从事 AI。
 3. Yann LeCun: 著名计算机科学家、图灵奖获得者, 否认自回归模型, 提出“世界模型”架构。

4 信息推送



The Revolution in Large Language Models: A Pivotal Moment in Our Quest Towards Artificial General Intelligence

4.1 Introduction

The quest to design and develop Artificial General Intelligence (AGI) has seen numerous significant breakthroughs, with each phase of technological advancement bringing us a step closer to this pinnacle of AI. The current revolution in large language models like OpenAI's GPT-3, however, stands out as a unique turning point in this journey. Unlike previous advancements, this revolution exhibits superior capabilities, including the understanding of context, generation of creative content, and the ability to learn from fewer examples, thus bringing us the closest yet to AGI.

4.2 Understanding Context

One of the key differences that make this revolution different from the ones before is the ability of large language models to understand context. Previous AI models were significantly limited in their understanding of human language due to their inability to comprehend context. They could process keywords and syntax but struggled with idioms, metaphors, and expressions that required contextual interpretation.

Modern large language models, however, are trained on vast volumes of text data from the internet, enabling them to infer the meaning from context and generate responses that are contextually accurate. This contextual understanding goes beyond simple keyword recognition, bringing us closer to AGI, which would possess the ability to understand and interact with its environment in a way that is indistinguishable from human beings.

4.3 Generative Capabilities and Creativity

Another distinguishing feature of the current revolution in large language models is their generative capabilities. Earlier AI models were predominantly task-specific, designed to perform specific tasks within strict parameters. They lacked the capability to generate creative solutions or produce unique content.

In contrast, large language models can generate coherent and contextually appropriate content, whether it be completing a story, writing an essay, or even composing poetry. This ability to generate creative output is a significant step towards AGI, which would be expected to exhibit creativity and innovation, traits currently unique to human intelligence.

4.4 Learning from Fewer Examples

The current revolution in large language models also represents a significant leap in the efficiency of learning. Previous models required vast amounts of training data and explicit instructions to learn a new task. This was time-consuming and often led to models that were excellent at one specific task but useless outside of it.

In contrast, large language models can learn from fewer examples, often requiring only a few instances of a task to understand it. This ability to learn efficiently and generalize from few examples brings us significantly closer to AGI, which would be expected to learn and adapt to new tasks and environments with minimal instruction, much like a human being.

4.5 Conclusion

The current revolution in large language models signifies a significant leap in our pursuit of AGI. By understanding context, generating creative content, and learning from fewer examples, these models have taken us closer than ever to achieving AGI. However, it is essential to bear in mind that while we are closer than ever, we are not there yet. There are still significant challenges to overcome, such as imbuing AI with common sense understanding and the ability to form meaningful relationships between different knowledge domains. Nevertheless, the progress made with large language models is an encouraging sign of what the future may hold.