

ChatGPT 第二批纪要合集

ChatGPT 的发展前景	2
ChatGPT 对国内产业的影响	9
ChatGPT 发展路径和竞争力	17
ChatGPT 技术壁垒和国产替代	19
投资人谈 ChatGPT	24
ChatGPT 算力知多少	29
ChatGPT 算力需求测算	32
ChatGPT 开启 AI 新纪元.....	37
ChatGPT 掀起 AI 产业新周期	40
汤姆猫访谈纪要	42
微软加速 ChatGPT 落地	44
小冰访谈纪要	47
中文在线访谈纪要	49
数码视讯访谈纪要	50
汉王科技访谈纪要	53
科大讯飞值得关注的 AI 应用方向	56
同花顺 AI 应用方向	65

核心观点

- 1 AI 行业还是在成长期，目前在 A 股的投资更多还是主题投资阶段。今年科技一定会有持续行情
- 2 深层次 AI 行业里面，大概是四类角色
- 3 AIGC 投资可以关注四个层面
- 4 现在 AI 行业有点像五年前甚至十年前的新能源行业
- 5 未来买数字化，很多时候你可能也要买传统行业公司，它会是一种双向奔赴
- 6 科技股未来投资方向，最看好还是数字经济方向，分两块

ChatGPT 是当下大热的一个话题，国外的相关公司可能领先了小半年，他们的发展情况或者他们各自有什么样的特点、优势，可以大概给我们分享一下吗？

- 1 这个爆款的应用来自于 OPENAI，它最早创始的时候，包括像马斯克等很多美国科技巨头参与在其中，而且它甚至不以盈利为目的，是一个类似于科研类的组织
 - a 后来微软看到它的发展非常好，实际上最近也有再投入。所以它跟微软自身的 AI 在这方面的布局结合的非常紧密。微软后面会推“全家桶”，就是微软将搜索跟这个相结合，使它变成超级搜索或者超级引擎的落地的场景
 - b 再包括其实微软最擅长的还是 Office，就是在办公场景下，比如 Word、Excel，其实我们过去用了很多年的工作习惯有可能会再有一个大的功能上的跃升，就是跟 AI 相结合。可能我们写作的效率，甚至刚刚谈到的一些题例式的文章，它都能完成基本撰写
 - c 像 Office 过去也有大量的比如文本校验等功能，但其实还是非常基础的。这次结合了 AI，它真的就相当于一个审核或者高级校验的角色，比如对你提供一些数据，逻辑的梳理、验证等等
- 2 这是微软在这一块儿有很明确的跟存量的搜索引擎、办公软件相整合，其实它已经从 AI 的模型阶段延伸到应用阶段
 - a 但是 Google 也在推 Bard 机器人，后面可能也会正式发布。当然 Google 过去是全球引擎行业的巨头，一直在人工智能方面有非常多的投入，目前它的危机感也比较强
 - b 另外对应国内这一块，像百度、阿里、字节都有类似的部署
- 3 但我自己总结为，在整个深层次 AI 行业里面，我觉得大概是四类角色
 - a 第一是提供底层的基础算力或者云服务的偏硬件层的基础设施，或者云基础设施的公司。这里面涉及通信或者电子相关的公司会更多一些
 - b 第二大层是提供大模型或者算法的公司。ChatGPT 或者 OpenAI 是一个非常典型的做大模型的公司，国外像 OpenAI、Google 会做，当然对应国内像百度、科大讯飞、阿里，这些科技巨头在大模型方面应该会有一些布局，陆陆续续大家应该也会看到，这是第二层
 - c 第三层，就是应该会有一些中间层，稍微偏专业或者垂直侧的小模型的公司。
 - i 其实国内现在也有一些公司已经开始做，它可能基于第二层的大模型来开发，但是它用底层开源的大模型在上面再做一个垂直细分领域的优势场景，比如医生或医疗行业，再比如汽车、法律、律师行业等，就是它可能会做一些垂直模型
 - ii 对于第二层和第三层的关系，大模型更像 K-12（中小学普世义务教育），是一个基础层的培养，非常重要。如果这 12 年或者这个部分没有做好，上面很难长出垂直方向很好的模型

- iii 到第三层，有点像大学和研究生阶段，可能要选专业，选一个目标培养方向。这个阶段不太会大而全，比较难有专业优势
- iv 国内可能会有很多公司做这一类，但是它也要基于一个大模型，在大模型的这一端，最终不太会有非常多的公司做，可能会是一个相对更大的，因为它对算力算法的要求还是蛮高的
- d 最后一层量会更大的就是应用层的公司。有点像毕业了，你要工作了，你有了很好的知识，真的从一个 K-12 教育走向大学教育，甚至研究生教育，最后我要到实践中创造生产力或者创造价值，那可能走向各行各业的 AI 的应用。
 - i 这类公司在国内现在也开始有一些比如游戏行业、搜索引擎，甚至将来的人形机器人等等，会有很多场景
 - ii AI 有点像五年前甚至十年前的新能源

OpenAI 能否成为一个现象级的公司，类似特斯拉，以及由此带来的 AI 能够成为类似新能源行业，它的这个行情将会非常大？

- 1 坦白讲如果现在一定要去判断这两家公司的空间有多大，现在确实还有点早。因为 OPENAI 这家公司最早设立的时候可能不是以盈利商业机构为目的，它更多的还是在创新上面做探索
 - a 从这点来看，可能特斯拉创立的开始也是一样，它也要在这种颠覆式创新上有它的独到之处，或者它设计的初衷就是改变人类现有的对科技的线性的理解。可能会有第一性原理，就是可能会有一个全新的颠覆式创新，这点是相通的。
 - b 但是如果从行业的阶段上来比，现在 AI 行业跟新能源的成熟度肯定是完全不同的，有点像五年前甚至十年前的新能源行业
 - c 当然中间新能源行业也经历过从最早的技术萌芽到产业的成长，像 Gartner 曲线中讲到的任何一种技术至少都会经历五个阶段，从技术的萌芽到快速成长，然后到泡沫期，就是资本或者整个创业者会非常疯狂的涌入这个领域，很多技术都会经历这样的阶段，然后到泡沫的破裂
- 2 泡沫破裂之后就进入去伪存真，就是真正做这件事情的人可能会留下来。但有时候最终到这个阶段可能有一些就扛不住、会被淘汰，那才会进入最后一个周期，就是真正的成熟期。在成熟期，才会有更多的收入、利润、业绩的体现，基本上都会经历这样五个阶段
 - a 如果套用这样的模型，毫无疑问特斯拉、新能源行业在今天其实已经走到了相对比较成熟的阶段，所以它已经在各种财务指标上或者分析师模型里通过销量或者成本模型，能够计算或者预测它的利润，然后给予它一个估值去定价
 - b 但如果回到这个模型，AI 现在处在什么阶段，或者 AI 不同的决策式、生成式可能处在不同的阶段，但整体来讲，AI 行业还是在成长期。当然，有些部分甚至可能在泡沫期，就是大家很热的涌到这个领域
 - c 按照正常的模型，它后面还有三大阶段要走，所以从发展阶段上来讲，跟新能源行业还是不能相提并论，对应的估值模型肯定也是完全不一样的
- 3 映射到投资上来看，我个人觉得目前在 A 股的投资更多还是主题投资阶段，但主题投资相对于价值投资并没有更优
 - a 其实在科技行业，很多新技术一开始都是主题投资的阶段，因为这个阶段会有对未来比较大的想象空间，同时行业的竞争格局没有那么固化，这时可选的标的或者很多公司都会有一种朦胧美，就是我们认为它可能都有做成的可能
 - b 像电动车行业现在也处在这个阶段，就是你现在还没有分化。但比如消费电子其实已经有龙头出来了，那我们就要更多关注它的收入、利润。但其实很多产业现在确实还处在一个板块式的行情中，AI 今年应该还是处在这样的阶段。

- c 所以对应到投资机会上或者从 OpenAI 自身来讲，影响估值的更多的可能不是它的收入、利润

当前我们应该看什么指标？因为对于成长性公司的股指，可能您在 TMT 领域还是非常驾轻就熟的，但是对于一些比如偏重于成熟型公司投资的投资人来说，可能还不是太了解？

- 1 其实在主题阶段，其实最近这个行情之所以能一直持续，跟海外巨头持续在这个部分有很多动作有很大关系
 - a 比如微软会再投 100 亿美金进去，但实际上除了微软之外，Google 马上也会动起来，它也要发它的模型，然后也没有商业化计划，同时别的产业链上下游的巨头其实也都在动。国内也一样，像百度三月份也会推出“文心一言”
 - b 这可能对行情都会有一个持续的催化，就是从主题上来讲它需要持续有新的东西，但是短期可能在 PE 和 PS 上不一定有很明确的指向，但是它也有可能会在比如用户量上有进一步的跃升
 - c 包括 OpenAI 自身在这个阶段还没有推专门的付费版本，我个人觉得可能未来会有两种收费模式。一种是对中小企业来讲，更多的是用云端调用它的接口；另一种是对大型企业，它可以本地化部署，这样我也可以直接付费或直接收费。所以整体来讲，从爆款的应用到用户的暴增到收入阶段，还是能看到的
- 2 最怕的是对于一个初创型的行业，大家一定要看利润，反过来对于一个很成熟的行业，大家又觉得可以看商业模式或者数据，实际上还是要用适宜它的方法去给它估值，其实还要有多样化的审美，这也是注册制以后可能 A 股要面临的，在科技投资方面非常重要

AIGC 可想象的空间很大，但如果从投资角度来看我们可以提前布局，将来可以关注哪些投资方向？

- 1 可能有这几个层面，倒着来看还是可以从刚刚提到的四个层面来理解。它是一个“倒金字塔”的结构，就是越往应用层越分散，应用层面应该是百花齐放的，它可以容纳很多小公司，这个行业集中度不会太高，这是第一大类
- 2 第二层就是在“小模型”这一层，或者比喻为上大学和研究生阶段训练。这个部分应该会有很多大公司做，但很多有数据、有场景优势的公司也可以做
 - a 比如我在图片行业或者音乐行业有大量的版权、原始数据，那它可能也会有，因为这个板块相关公司的核心壁垒第一个肯定要有基础的算法优势，但更重要的是它要有数据，就是要有很多原创数据
 - b 那就去挖掘哪些是有场景、有数据的公司，这部分也有投资机会
- 3 第三层就是“大模型”，前面也有提到，这里就不再展开
- 4 第四层，就是提供底层算力、云基础和相关支撑的公司
- 5 2022 年年初，国务院发了一个有关数字经济十四五规划。“十四五”规划中对数字经济定义时提到“数据要素是核心生产力”，这就有点像“未来的石油”，我可以通过石油加工出很多如尼龙、化工行业的东西，未来数据也一样
 - a 所以从这点来讲，提供底层算力的公司可能不会是涉及面很多的公司，因为它是一个门槛相对比较高的行业，现在像做云计算、做通信、做技术支持服务还是是一些大公司或国资类的公司比较多，但是它的确定性会很高
 - b 底层有关的这种“云”，不管是 IDC 服务、网络服务、提供 ICT 设备，甚至是 CPU 和 GPU（芯片）都会受益。它可能没有前面提到的那些应用、做模型的公司那么直接，但反过来讲，它的确定性会很高。工业数字化投资确定性很高，未来传统行业公司也要买

在数字经济发展中，新型的基础设施建设、硬件设备、软件开发、应用场景可能都有爆发点，哪些领域有望率先突破？

- 1 未来十年，数字经济最根本的是把数据作为生产的一个要素，把 AI 作为生产工具，这跟过去十年不同，过去十年是网络和信息化
 - a 如果这样，我认为首先在教育端或者未来最重要的抓手应该是谁能把数据用好，谁能用 AI 的工具创造新的生产模式或新的应用场景，这是数字经济的最根本
 - b 当然，数字化的前提是云化。云化就是如果它都没有上云，数据都是离线的，或者生产数据、管理数据可能没有全部在线或在云上，这很难称之为数字化企业，所以从这点来讲，云计算肯定还是确定性很大的。所以这两年云背后包括存储、服务器，实际上是对算力的支撑
- 2 另外在数字化投资里，各地有各种东西出来可能和过去十年还会有很大不一样。过去十年，互联网时代更多的是 TOC 的场景比较成功，C 是“CONSUMER”，就是更多改变我们的生活，比如社交、电商还是对生活质量的提升，让我们的生活更加便利
- 3 但未来十年在数字化里面，本质上还是 TOB 的场景，B 就是生产，实际上会对它有更大的跃升，比如数字化在智能制造，在制造业里面对降本增效的效果。现在对企业来讲，数字化可能不是一个选择题，而是一个必修题，就是如果你不做，可能未来你的商业模式就要被干掉
- 4 在过去一段时间，汽车行业就是非常典型的，电动车来了以后，汽车供应链都重组了。汽车供应链以前都讲 TIER1 供应商，但它不是一个垂直化的、定制的生产模式，这就是一个很传统的制造业生产模式的改变
 - a 汽车可能不是最后一个在能源行业的数字化行业，我们当时研究数字化场景时发现一个很有趣的现象，就是越集中度高、越传统的行业反而越容易数字化，而供给侧改革过的行业会更容易数字化，为什么？
 - b 因为它的行业结构相对巨头林立，产业分工很清晰。比如煤炭、钢铁、有色拥抱数字化的态度非常积极，行业竞争格局已经很清晰
 - c 想再提效率或者有增长，那就要有新的工具和技术手段。反而对于一些比较分散或者集中度没那么高的行业，比如医药行业好像就没有那么快，因为它的中间环节还很多
 - d 这是汽车能源，当然中国还有一块非常大的机会就是工业，就是去年二十大报告提到的新型工业化。新型工业化的背后，工业数字化就是很大一块，在投资上的确定性都很高
 - e 所以我去年提出一个观点，以前投数字化或者投科技就是投 TMT 行业本身，大家去买那些做 IT、做 CT、做芯片的公司。但未来买数字化，很多时候你可能也要买传统行业公司，它会是一种双向奔赴

那会不会像我们之前理解的互联网+，就是在传统行业的颠覆上，互联网这个应用各个行业都开始了，也颠覆了传统行业，会类似吗？

- 1 会类似，但是我不用一个词叫“颠覆”
 - a 原来在 ToC 的领域确实是颠覆，就是零售行业做大会对传统的线下零售有很大影响。但是这次 ToB 或者跟很多传统行业的融合，TMT，就是单纯做技术的公司，还是挺难颠覆的
 - b 因为在一生产场景，相比于一个消费场景，首先它是非标准化的，就是每个场景长得都不一样，所以你很难出现一个巨头，在互联网时代赢家通吃
 - c 但是 ToB 不会，比如在制造业和医疗行业可能完全不同，它的 know-how（场景）完全不同，这个时候如果传统行业公司能够用好 AI，用好新的数字化工

具，它可能会更有优势

d 数字化可能会成为一个基本工具，所以这点很多传统行业现在还处在估值比较低的位置，比如能源行业、工业中也有很多非常优秀的巨头已经开始全面拥抱数字化

2 过去一年疫情中大家对数字化还是保持非常开放的心态，过去它觉得你要颠覆我，甚至是一种排斥，现在因为整个经济增速放缓过程中，大家觉得这可能会是一个非常重要的工具让自己重生，所以企业主愿意在这上面进行投入，它会对整个行业的推动起到很重要的作用

那会不会在集中度比较高的如国企领域，或者在偏垄断、竞争不激烈的行业率先发生？

1 如果按行业分，TOC 先不说，在 TOB 的场景因为有很大，更大的如分服务业的数字化，农业、工业等，稍微小一点工业里面又分很多种，但整个研究下来，首先中国的产业数字化水平比全球低

2 其实就是在 TOC 互联网里很发达，在消费部分发达，但是在生产的部分相比于美国、德国、日本，他们的渗透可能都在 60% 以上，我们可能只有百分之三十几，这个本来就是要低一些，所以空间会比较大

3 第二个单就在产业数字化的场景里面，就是 TOB 的行业里面，我刚刚举了三个比较看好的方向——汽车、工业、能源，它的行业集中度相对较高，对产业链分工比较清晰，巨头或者龙头对行业的技术和模式创新引领就达成共识，这是比较好的场景。当然，比如金融行业也在发生类似的事情，也会延伸包括一些现代服务业等

在 AI 发达的时代，我们如何分享科技股的盛宴。近期在券商的策略会当中，大家集中看好科技股，科技股也已经有了不小的涨幅，未来您怎么看？

1 上次你们办论坛的时候，我其实还是坚定的看好，我当时有三点理由

a 第一个是本身技术变迁到了一个新周期的开始。TMT 的每一轮行情最重要的因素可能不是估值，因为它足够便宜，当然这肯定是其中一个很重要的因素，但不是唯一的因素

i 最底层的因素应该是到底有没有质的技术创新，就是现象级的应用，或者有没有爆款的东西出来。其实 2015 年那一波科技行情，移动互联网的爆款应用出来很多，包括后来像抖音、快手这样爆款的现象级的应用

ii 现在又有现象级的应用出来，而且它到上亿用户的速度可能是历史上最快的，只用两个月时间，这是一个很重要的标志，就是它需要有全新的、让人感觉很兴奋的东西出来，而这种东西可能是一个颠覆式的创新，当然过去几年科技有很多场景有点低预期，那就没有达到想象

iii 但是这次 AI 在这个方向其实是有超预期的表现，这是一个很重要的驱动科技股的因素。当然，如果基于这个模型，后续还有没有爆款应用出来，是有可能的

b 第二个影响科技股行情最根本的因素是流动性，我去年提到美联储加息的趋势停止，其实这也很重要。因为总体来讲，确实整个估值体系里面，科技股历来是估值比较贵的，但是它贵背后有一个“G”（Growth），就是有增长在里面

i 所以我更习惯于去看 PEG，而不是简单的看动态 PE 更有价值。就是相比于 15 倍估值，但是只有 15% 增长的东西，我更愿意选择一个 50 倍估值，但是有百分之百增长的东西。这是典型的科技研究员的审美，一定是要有增长、有变化、有未来潜在的非线性的拐点

ii 像当年亚马逊大概有十年只有收入的增长，但 AWS 没有利润的增长，但它一旦开始盈利以后，你会发现收入可能每年增长 50%，但利润增长百分之

百，然后毛利率或者利润率有一个边际成本迅速下降，软件、互联网行业其实都有这样的特点

iii 所以这是第二个审美的重点，就是它今年有一个流动性的优势，同时最新的流动性的溢价会更直接的体现在科技股上，当然在去年那样反向的时候就会更痛苦

c 最后一点，从今年整个 A 股来看，确实在科技股的投资板块上，估值还是有一些吸引力。因为从 2018 年的四季度应该是一个低点，但经过后来 2019、2020 年有一轮反弹，但总体来讲跟 2018 年以前还是有差距的

d 就是实际上高的时候看 2018 年之前，像今天很热的计算机行业在火的时候，实际上也有 50 倍以上的估值中枢的阶段，但是其实我们就觉得 50 倍可能很贵

e 相比于 2013、2014 年非常疯狂的阶段，还处于中位数，估值还有一定空间，没有拥挤到已经泡沫化了，那可能确实后面会比较难

2 所以基于以上三点，今年科技一定是会有持续行情

从这三点来看，或者结合一些历史数据，您觉得科技股的牛市一般会持续多久？

1 其实还是取决于前面几个因素，要不断有新东西出来。比如移动互联网，就是 3G 那一波，大概从 2012 年游戏作为第一个爆款应用出来到 2015 年大概有三、四年的时间

2 再看 2018 年的半导体，其实就是以信创、自主可控为代表，也有差不多三年时间。当然可能结构会不一样，我以前经常讲“硬三年，软三年，商业模式再三年”的周期性，过去几年其实更多的还是在硬件上

长期逻辑不看新鲜概念或主题，怎么把握未来 5 到 10 年科技股的投资方向呢？

1 简单总结，我最看好的还是数字经济方向。数字化其实是一个周期比较长、确定性很高的东西

2 具象一点，可能有几个板块，数字经济就是两大块

a 第一大块是数字产业化。就是 TMT 行业本身会有一些基础设施部分（新基建），这个部分确定性比较高的像云计算，现在它又是一个需求跃升或者拉动的过程刚刚开始，对应到不管是做云自身的公司，还是做云基础设施的，像配套设备或者是里面的芯片、ICT 设备的公司，它会受益

b 第二个在基础设施部分我比较看好的还是大安全的主线

i 现在在科技的自主可控，以及信创，包括网络安全，它是一个不仅仅因为全球竞争的问题，其实也有应用从 ToC 走向 ToB，它自然会对安全的要求很高

ii 所以在相关的比如网络安全、底层设备的投入上，它对安全的要求就会更高，从这点上信创板块应该还有确定的机会

c 最后就是在产业数字化方面，在应用端下一个爆款的应用会在哪里，这也是大家所关注的。但整体来讲，可能 ToB 的机会会大一些

d 这三类未来都会产生投资机会，但不一定会是那种千亿市值的特别大的公司，ToB 的应用可能会容纳一些专精特新，小而美的公司也会很多，因为它往应用端去走。但是在数字产业化，就是提供云基础设施这一方面会有一些大公司。现在已经有一些大公司了，无非是这些大公司能不能从千亿市值涨成更大市值，成长空间到底有多大

5G 的建设目前是不是到了中后期，它的投资逻辑是否发生变化，未来怎么看？

1 关于 5G 我之前总结了两句话，第一句话是从新基建走向新应用。这个观点很明确，就是基建的投资肯定是告一段落

2 下一步我们要思考怎么用好这么好的基础设施，就是这么好的高速公路修通了，

我怎么保留大量的车来走

a 硬件和软件没有孰优孰劣之分，它是一个螺旋式创新，所以今年 5G 的应用就很重要。到底有没有一些好的应用，比如自动驾驶、工业互联网、VR、AR，这是 5G 潜在的一些比较大的场景

b 所以从新基建走向新应用，在 5G 的投资上今年应用的机会应该更多。

3 第二个，在投资上要注意从终端走向云端，这个可能听起来比较绕，但很好理解

a 过去产业链的价值是承载在终端上的，比如我们买一个手机、电脑很贵，而且我们会关注它的内存、CPU 指标。但现在我们买手机有可能关注它的屏幕大不大、摄像头多不多，原因是什么呢

b 算力或者计算的功能是在云端的，承载在云上面的，就是英伟达做的这个工作。因为网速足够快，我的内容没有存在本地，实际上我没有必要存本地。以前大家丢了手机很慌张，现在丢了马上想找回账号密码，其实硬件本身变成一个显示器，所以很好理解

c 其实你所有产业链的价值量或者算力是在云端的，所以在 5G 投资上也要注意。以前我们投 5G 可能大家会关注手机产业链，但现在你可能要关注应用，云端有没有投资机会。再比如汽车，去年对于智能车或者电动车，我们关注硬件本身的投资机会，但接下来的重点应该在自动驾驶和智能座舱

ChatGPT 出来之后，对我们国内相关产业的影响？国内相关的厂商，阿里、百度未来一段时间落地情况？

- 1 从国家层面，GPT 这个产品推出以后，网信办已经发出了一些政策相关的东西
 - a 从国家层面的角度来说，短期内我们很难看到 ChatGPT 这些产品直接跟国内的应用，或者做比较深度的结合，因为这块不管是基于信息安全，还是国内的一些产业保护的角度来说，国内可能都得需要有这样一个窗口期，得需要有逐步缓冲的时间
 - b 所以，国家后续会出台相关的政策，给国内的玩家们提供追赶的时机
- 2 从我们之前对百度文心类似产品的使用体验来看，内容的质量上百度文心和 CHATGPT 差距不是特别大，只是在内容的多样性上有差距
 - a 咱们国内研究相关的大模型的，目前主要靠工程人员和研发人员，大概几百号人或者上千号人这么研发，在这个过程中没有引入像 OpenAI 这种用户反馈机制，没有大量的用户和技术模型的迭代过程中参与进来
 - b 所以在内容的多样性上会有所欠缺
- 3 可能跟目前实际没有放开政策有关系，就是比如像 CHATGPT 响应能力，一个 Q 过去，A 回来大概是 1-3 秒，响应能力比较快。目前百度文心这边我们能够体验到的，大概短的在 20 秒左右，长的甚至在 80 秒以上
- 4 当然这个不是技术瓶颈问题，需要在模型研发完成以后，我需要在服务器做部署，部署完之后支持数以百万计，甚至数以千万计高并发访问的需求，这块属于常态的部分。未来 3-6 个月左右的时间，像百度文心、阿里推出类似于 CHATGPT 的产品应该可以达到目前 CHATGPT 60%-70% 左右的水平

ChatGPT 出来之后，一开始在美国程序圈特别火，比如说从应用层面，或者算力层面有什么对行业的影响？百度或者其他有没有正在加大投入的？

- 1 先说一下整个参与的玩家，只说国内。像百度已经宣布了，阿里达摩院也宣布了会投入 3 个亿的基金做这块。当然还有我们想不到的，比如说美团前合伙人也宣布下海做这件事情。当然可能在其他的领域，包括您提到的字节，也对之前一些 GPT2.0、3.0，可能以开源的模型作为基础，也做了一些相应的优化，也跟这些客户推相关的解决方案
- 2 像字节也做了一些优化。主要是训练过程，在云服务架构上，通用的算力部署，大概会有 1.5-2.5 倍的效率提升。可能跟我们接触部门有关系，它自己没有单纯的训练大模型，更多的想把服务包给愿意做大模型的公司
- 3 这个模型训练的成本主要有三个方面，单纯从基础算力投入这块大概是 384 块到 3,000 块左右的 A100 的投入，基本上可以满足要求。这块的费用大概 400-1,000 万美金，最多到 3,000 万美金的规模，所以对于很多巨头、独角兽甚至一些初创团队，其实都有能力做这个投入。而且它的算法的基础都是基于预训练这样一种机制。所以参与的玩家会比我们想象的要多
- 4 关于技术跟行业结合的影响，短期内主要是有两个大的方向
 - a 一个方向就是不管是谷歌也好，还是微软也好，肯定会跟搜索引擎做结合。这块主要是用来提升搜索引擎的信息获取的效率
 - b 第二个大的方向，对于微软来说更加有优势，就是把类似于 ChatGPT 的这种技术跟办公软件做结合。因为跟办公软件做结合有一个好处，它可以提高日常生活工作的效率。比如说处理表格，或者写一写小作文，类似这种可以提高效率。而且可以很好的在工具软件的定位上，把一些资讯的敏感性过滤掉
 - c 如果只是单纯的跟搜索引擎结合的话，这里面可能会有一些政策、政治、意识

形态，类似这方面的影响。而且从内容的提供信息给终端用户使用的角度来说，AIGC 这种方式也存在一个被大家忽略掉的风险，以百度、谷歌来举例子，比如说传统的搜索引擎的提供商都是通过快照这种方式把信息抓取过来，实际上用户看到的是一个一个的链接

d 用户点进去这些链接查看自己想要的信息，出了问题承担责任的话，其实是可以从这个链接找到一个最终的责任方。即使像之前百度跟莆田系的关系一样，尽管可能口碑不好，但是我可以找到莆田系的医院承担最终的责任

5 但是目前以类似 CHATGPT 的技术，相当于对这些信息做了二次加工，这个责任的主体其实已经在这个过程中发生了变化。所以这时候其实如果大家没有意识到这一点的话，可能承担责任的话，谁提供这个信息，谁提供这个产品，就变成了最终的责任方。所以在具体的内容呈现上，可能还需要做一些优化

投入成本方面，您指的是比如说需要买算力芯片、服务器芯片的成本，还是算法工程师上面的投入成本？

1 要做大模型的话，算法成本一共是包括三部分

a 第一，刚才已经提到的这部分，基础的算力投入，这块大概是三四百万美金到一千万美金左右的规模，无非是我投入的少，训练的时间长一些，比如说 170 天，200 天，大概这样的

i 如果投入的多一些，像英伟达也在合作做大模型，五千亿的参数，我投入了 3,072 块 A100 做训练，这块效率可以缩短到 20-30 天

ii 这个成本在整个大模型训练过程中最少的。真正的大头在数据采集、模型的技术迭代和优化过程中。这一块需要堆集大量的研发人员、工程师、科学家做这方面的工作

iii 比如说数据采集，我们要抓取整个新浪、网易、搜狐、知乎这些网站的数据，因为这些网站都有反爬虫的策略，所以需要堆集大量的人员去写破解反抓取策略的策略，这块会耗费大量的人力和工程师的时间

b 第二，引入用户反馈机制，这时候不仅仅是研发人员了，有大量的标注人员，有大量的用户参与，这块耗费的一个是时间长，另外一个成本高。对于 OpenAI 来说，每年在这块的投入在 1 亿到 1.5 亿左右美金的规模，这是比较大的地方。因为数据采集和处理的时间会远大于单纯训练的时间

c 第三，模型正式上线了，运维和部署。像 ChatGPT 一样，上线以后很快获得日活一亿左右的庞大的用户群体，这个过程中我要处理数以十万计，甚至数以百万计的请求。过程中需要堆集大量的服务器，需要做分布式部署，需要做负载均衡，甚至搭载网络安全策略，防止黑客攻击

2 这个过程中还会产生大量的比如说电力的消耗，这块每年的成本至少在五千万美金以上。所以主要由这么三大部分构成的

您说到的这几块，是不是也要分训练相关的成本，以及运营过程中的三块的成本。这块会有什么差异吗？

1 比如说前两部分，我们都可以归纳为是训练的成本

2 第三部分可以归纳为服务正式部署上线之后运营方面的成本

像国内的浪潮、海光，他们跟英伟达这块的高算力相关的芯片，或者是服务器能不能支撑上。如果说英伟达他们的芯片不能卖给中国的话，中国有没有类似的解决方案？

1 从两个方面说，单纯讲芯片和算力的研发能力这块，我们肯定跟英伟达这些公司是有差距的。不管您提到的浪潮、华为、寒武纪，哪怕研发出来最顶级的芯片，还是有比较大的差距，这是一块短期内应该还是比较难跟上的。即使我们在某些技术上有优势，但是最大的问题不是在单纯的技术上，而是在整个生态体系，这块英伟达非常完善

- 2 我们可以分享一些数据，比如说在全球的超算中心，英伟达大概可以有 90% 的份额，全球的云服务中心，那边可以有 80% 的份额，这个其实在短期内很难改变的状态
- 3 再从国内的这些不同类别的厂商，从采购高端显卡的实际情况来看，这种限制肯定会有影响，但是影响可能没有我们想象的这么大，可以说政策是政策，生意是生意。比如说目前我们国内的四大巨头，像字节、阿里、百度、腾讯这四家公司，在 CHATGPT 火之前，去年的这些数据每年大概采购 A100 这种级别的显卡的数量，大致是在 2.5 万片左右
- 4 像浪潮这些公司，英伟达那边提供了相当于稍微有点阉割的 A100，或者 H100 同级别的显卡和芯片。基本上没有特别大的影响，至少目前来看

像百度刚才您说到的一些，您觉得训练出来可能也差不多。从算法模型来说，因为 ChatGPT3.5 它是闭源了。像百度、字节他们未来研发出自己的 GPT 产品，从过去的 3.0 开源版本找这些数据，去直接拿它的模型在上面改，还是说需要重新去搭建一个大的模型，这个具体在算法上大概是怎么样去做？

- 1 这块目前比如说从字节那边了解到的情况，从阿里那边了解到的情况，基本上都是基于 GPT3.0 以前，就是开源那些东西拿过来去改，这个相当于是一个基础。当然 GPT3.0 的基础也是以预训练模拟训练作为最底层的算法逻辑的。所以从这个角度讲，你可以理解为大家在通用模型上是一套东西
- 2 尽管不会从 0 开始的这么一种状态，但是训练的时候 OPENAI 尽管细节我们没有办法复现，但是至少方向上是明确的。比如说过程中引入用户反馈机制，用户反馈机制说起来高大上，但是实际上就是堆人，可以简单的这么理解，针对同一个 A，我组织成百上千的人去组织或者提问不同的 Q，针对同一个 Q，组织成百上千的人去回复不同的 A
- 3 这个其实就是为了满足内容生成，就是预训练的时候我有大量的多样化丰富的数据，才能够满足 AIGC 这块基础训练的需要
- 4 其实目前百度文心这块欠缺的可能就是在这一点上，在内容的多样性上，丰富度上有差异。因为之前我们不要说国内的公司了，包括国外的公司，包括谷歌这边其实也是一样的，它的整个研发过程中，其实普通的用户很少参与，都是堆积大量的工程师、科学家、研究人员做这个工作，在正式公测之前普通用户很少参与
- 5 但是 OPENAI 经过前几个版本的迭代，在这个过程中发现了普通用户参与的好处，所以才推出了用户反馈机制，大家顺着这个思路做就好了。细节上可能没有完全复现，但是效果上不会有太大的差距

在短期没有能够创造收益的情况下，您觉得这些互联网巨头可以持续做这个投入吗？年化下来费用大概总共是多少？

- 1 其实对于 OPENAI 推出 CHATGPT 来说，在整个的领域，我们现在已经可以看到很多的迹象，不管是前美团的合伙人王慧文总，他自筹资金，搭建团队研发模型，还是说从目前一些招聘渠道，可能对 NLP、AIGC 相关人才的需求，我们已经可以看到这些迹象，大家都会增加这方面的投入
- 2 之所以会增加这方面的投入，其实是因为 OPENAI 推出 CHATGPT 有点像一条鲶鱼，大家如果不跟上这波浪潮，有可能在未来会被甩掉，这是极有可能的。所以目前大家基于被动防御的策略，我们可以看到不管是谷歌也好，百度也好，阿里也好，这些公司都在宣布往里面投入资金做模型
- 3 所以相对比较短的时间内，因为还有后续的比如说 GPT4 这些东西不断的推出，所以在未来 2-3 年的时间里，在这种类似 CHATGPT 大模型的资本的输出，肯定是一个可以预见的一个地方

如果放在中长期来看，最终的格局大概率国内就是一两个，两三个大模型最后能跑出

来。现在可能有很多家去投，或者很多家已经有布局，最终要去做这个事情。您怎么看待这个事，终极来看的话？

- 1 如果终极来看的话，个人认为不管国内还是国外，一定会存在多家类似 CHATGPT 这种大模型。比如说国外谷歌和微软的关系，包括把苹果也拉进来，甚至特斯拉，我们举个简单的例子。这些巨头都有可能研发自己的模型，而且他们所处的生态位或者产业略有差异
- 2 比如说像苹果，它有大量的智能终端，把这个能力，或者自己研究大数据的模型跟 SIRI 做结合。它通过智能设备，比如说打开谷歌的应用，或者打开浏览器去做搜索，会更加的直接，有更高的效率
- 3 再比如说像谷歌和微软，虽然短期内比如说微软借着这个风头，看起来具有领先的优势，它在短期内也的确可以影响到互联网、金融、投资、科技、人工智能相对比较精英的人士
 - a 比如说从日常习惯使用谷歌，转移到使用 Bing 上，但是它无法改变更大的格局
 - b 这个格局比如说像谷歌全球的市占率大概在 90% 还要多一些，Bing 大概只有 3%-4% 之间，因为这些应用嵌入到很多硬件设备当中的，比如说电脑，甚至更多的属于中低端的智能手机设备，所以很难短期内有非常大的市场渗透率
- 4 所以真正决定胜负关键的不是一时的得失，而是参与的玩家所处的生态位，这个生态位决定他可以快速的把这种能力推给他的用户，达到非常高的市占率

如果我们分技术和应用来看，先说技术这个层面，微软做的这个东西，OpenAI 做的这个东西和谷歌现在做的东西，从技术层面会有很大的差异吗？还是更多是应用场景的差异，大家所拥有的数据资源的差异导致最终的应用效果的差异？

- 1 其实单纯从大模型上看，本质上没有特别大的差异，因为现在三大系列的模型，不管 GPT，还是 BARD，还是 XONES（音），都是基于预训练和 TRANSFORM 这两种基础，都是谷歌大概在 2017 年开源的 TRANSFORM 那一套基础之上做的
 - a 只不过大家以前有一个创新者困境，而且这些年人工智能除了很久之前的 AlphaGo 能够让大家感觉到很兴奋以外，在整个过程中一直没有让人感觉特别兴奋的一个点
 - b 所以创新者困境不仅困住了国内和国外巨头的手脚，首先资本上大家不敢过分的往这里投入，谁也不知道我投入 5 年、10 年、8 年到底能不能做出来
- 2 再有一个在团队内部，一个团队负责人工智能的项目，稍微时间长一点，比如说两年，甚至有的短到半年，如果见不到效益，或者没有好的一些东西，这个部门可能就被裁掉了，这个团队也被裁掉了，这跟创新者困境有很大的关系
 - a 现在 OpenAI 把这个东西推出来以后，因为这么多研究大模型的厂商，之前使用的算法、基础、原理、机制这些基本都是一致的，只不过他们之前可能没有采用像 OpenAI 采用的用户反馈机制
 - b 当然有一些细节上的调整，这个肯定没有完全的一致。但是这个用户反馈机制在这里面会非常重要，可能从量变引起质变的效应
- 3 现在大家是创新者困境被破除了，而且 OPENAI 也给众多的厂商指明了一个方向，你引入这个机制，让更多的用户去参与。相信这些巨头应该在相对比较短的时间内，会陆续推出自己类似 CHATGPT 这种产品出来

ChatGPT 其实是基于 GPT 的应用，从后面 GPT 这一套，大模型的这套东西来说，应该各家是原理上基本上是类似的，能力上也不会有太大的差异，更多的是涉及的领域，手里的数据资源等等这些方面。往后看这些东西有没有可能被开源化，类似于像数据湖等等的技术开源化，或者像安卓那样？

- 1 这种开源的模型也有，比如说像法国的 BIGSCIENCE 推出的 BLUM，可能有很多从

谷歌、微软，甚至还有 OPENAI 这些公司出来的，他们推出一个开源的框架，也是一个类似于 CHATGPT 的，只是效果怎么样，现在大家没有真正的去做深入的体验。但是这种开源框架是有的

- 2 考虑到一些实际的情况，特别是跟行业应用做结合，比如说在金融科技领域，在数字政务领域，或者在教育、医疗等等这些实际应用领域的话，单纯的比如说一个模型公司，它能够获取的数据都是基于公开网络上获取的一些数据
- 3 那么像我们刚才说的这些领域的的数据，基本上是不对外公开开放的，还是有人需要在这些算法模型基础之上，拿过来做二次的处理，去往垂直领域做相关模型的进一步研发

如果从技术层面，比如微软和谷歌两个技术层面是类似的，相对他们的不一样更多的体现在他们拥有数据不一样，用户反馈的机制不一样，从而导致使用效果不一样，是可以这么理解吗？

- 1 其实跟他原有的产品结合上。比如说微软靠 BING 还是很难挑战谷歌，但是它有 OFFICE 这套软件系统，有 WINDOWS 操作系统，把这些能力跟自己原有的优势做结合。比如说像谷歌，我可以跟安卓系统，可以跟搜索引擎做结合，这些方向上会更强势
- 2 其实可能在这不同的生态位上，每个公司可能会有自己独特的优势。比如说像苹果完全可以跟 SIRI 结合，直接跟终端硬件绑定了。特斯拉也一样，特斯拉跟 OPENAI 有这种投资人的关系，但是 OPENAI 肯定也是没有办法直接获取特斯拉在智能驾驶上的这些数据的，要么以把 SDK 或者 API 授权给特斯拉使用，要不特斯拉自己干了。所以基本上是这样的

您刚才说在一些细分领域，比如政务、金融这些领域，这些可能是某种程度上属于 GPT 这套，大模型最容易去落地的，比较容易落地的几个场景。如果从落地的难易度的角度，您觉得哪些领域会比较快？

- 1 在落地领域，最快可能还是在工具软件上。因为我们刚才提到了，类似 CHATGPT 这个技术有一个很容易被大家忽略的风险点，就是 AIGC 的过程
 - a 当然我们在前几天，比如说新 Bing 的体验上已经可以看到，比如说它去分析 GAP 的行情是上涨，但是实际上是下跌的，其实问题恰恰出在这
 - b 这里面有信息的二次加工问题，尽管你提供了信息源，但是我们看它有足够的理性，有足够的分辨能力可以甄别这些信息。但是对于绝大多数普通用户来说不具备这些分析能力，也没有这样的理性，也不愿意分辨。这个时候出现问题直接承担责任的的就是信息的提供方。所以具体的交互形态上，我们可以得出这么一条结论
 - i 第一，不是所有的内容适合 AIGC，就像我们谈到的金融科技、数字政务、教育、医疗，很多就是 1,2 就是 2，AIGC 需要划定一个范围
 - ii 第二，细节的算法需要做一些优化。现在有的人在不管是调侃也好，或者别有用心也好，可以对 ChatGPT 回答的内容做诱导，这个也存在问题
- 2 当然，在具体的内容上多少存在偏意识形态或者这些方面的问题，这个可能会更加的危险。所以大家可能在去用这个技术的时候，刚才说的由于内容的真实性和可靠性这块引发的一些小的问题，看起来很小的问题，需要先解决掉。然后才是跟商业做深度应用的结合
- 3 再把这种技术跟行业深度应用的时候，其实有一个最后一公里的问题需要解决，主要是在多模态交互上，我们知道在 CHATGPT，我们跟它交流的时候使用的输入文本的方式
 - a 但是真正在高速上比如说开车，或者在银行网点办理业务，这些场景基本上用户跟这些智能机或者 AI 去交流的时候，基本上采用语音交互的方式为主，不

可能说我还要去提供一个键盘输入文本，几乎不存在这样一种情况

- b 所以我们可能目前像阿里、百度，也包括微软这些公司提供的智能语音的这些产品，90%的识别准确率没办法满足线下对答如流，答非所问这种实际应用的要求，这些基本上可以理解为最后一公里的要求，得先把这些问题解决了，才能更多的跟行业应用做更深度的结合

除了刚才说的这种大的，像微软、谷歌本来就有大的软件，或者大量的流量入口，比如说像谷歌搜索，或者 office，在一些垂直领域，类似于刚才说的金融、政务，小公司比大模型不用比了，完全不是一个量级。最终这些垂直领域的公司，是不是大部分会演变成使用巨头的模型来做行业应用的应用型的公司？

- 1 对，相当于巨头们更多做的基建的产品，他们提供基础的能力。行业里面还是有初创团队，或者中小公司，他们做处理技术应用的这些问题。而且技术应用这块对于大厂来说，他们觉得这些属于苦活，累活，也不愿意去干这些东西
- 2 但是这里面也存在一个巨大的机会，我们刚才已经提到了，多模态交互是一个非常复杂的，比如说我们在银行网点，开车的时候，它的环境，极度嘈杂，车内存存在多个人
 - a 这时候我们说话的话，AI 需要自主的分辨出来跟谁说话，或者这些人说话是不是要听，要不要做出回应等等，它是非常复杂的东西
 - b 这里面就像当初上世纪 70-80 年代的时候，微软和苹果抄袭施乐的界面，有了现在以键盘、鼠标、触控为代表的操作界面一样，谁在服务的过程中把交互的部分冲击出来，ChatGPT 作为核心，实际上就是下一代的操作系统
- 3 我们知道操作系统像谷歌、微软、苹果，三家公司的市值大概 7 万亿美元左右，这个想象力空间是足够的
 - a 再有偏一些智能设备的助手方向的，比如说像苹果的 Siri，小米的小爱同学等等，类似这些产品也会有这样的机会，比如把这种能力跟自己的语音助手做结合，未来一旦做好了，也不排除分拆独立的可能性，成长为更好的，更有价值的产品。因为现在 ChatGPT 回答我们任何人的时候，你可以理解为它就是一个 ChatGPT
 - b 但是未来它完全可以更加的以个性化，比如说你聊的时候，它可能像成龙，我聊的时候可能像肖战，带有比较鲜明的人设特征的，可以满足用户的个性化方面的一些需求

从相对中长期来看，一方面有大模型的大厂，以及掌握大的流量入口的大厂，不管搜索引擎还是 office，具有丰富的用户群体直接受益了。另外一块比如说掌握语音技术的，掌握跟人交流技术的，相当于把后面大模型更好的跟人机交流搭起桥梁的。其他方面您觉得还有一些什么样的方向，或者说细分领域可能会受益于这个浪潮？

- 1 这块其实主要看好五个大的方向
- 2 第一，做基础算法的。就是 OPENAI、百度文心，谷歌的 BARD 等等，有一堆算法的。但是他们的商业模式相对比较简单，一种面向企业和开发者提供的标准的 SDK 和 API
 - a 再有我在这个基础上把它稍微功能化处理一下，变成有点像功能软件，比如说处理表格用来写小作文类似这种功能，以订阅的方式向终端用户提供
 - b 但是这两种商业模式的附加值不够高，这是基础，而且研发成本又投入大，这是第一类方向可能会存在的一个困境
- 3 第二，提供基础算力和应用服务部署这方面的，主要是有这么三类公司，一类是像英伟达、寒武纪这种做基础芯片、显卡的公司
 - a 第二类像华为云、阿里云、百度智能云这种提供云服务的公司。第三类因为我们已经看到了 GPT4 要推出的时候会增强这种多向模态的能力，现在 ChatGPT

更多的处理文本，文本对算力的要求更少

b 未来我要增强音视频、度片类似这方面的处理，会要求更高的算力，还会有一堆专门处理音视频处理，或者图片处理的，提供这些算力支持的公司，比如说有一些新型的，像平行云类似于这些，但是都提供基础的算力支撑，包括服务部署方面的公司

4 第三，刚才已经提到了，主要是技术应用。比如说跟搜索引擎的结合，跟操作系统的结合，跟个人助手的结合，跟工具软件的结合，或者是我就是干脆直接的垂直行业的应用，比如说进入科技领域，比如说像海康威视，它有渠道和客户资源，我把原有的能力跟硬件产品做结合，它也能够在这里面获得足够的成长和机会。相当于渠道和客户在这里面也会非常关键的

5 第四，刚才讲了软的东西，讲了算法，这些东西我要拿过来给终端的用户做交互的话，我都会依赖一个载体。比如说智能手机，可能是一台智能汽车，也可能在未来甚至眼角膜植入芯片，可能满足用户交互的东西。所以相当于在物联网终端这块，跟 CHATGPT 这种技术做结合，也是有很大的机会的

6 第五，在 AIGC 这样一个方向上。因为有了软件，有了硬件，用户跟软硬件结合在一起的智慧体系交互的时候，需要大量的内容。单纯靠人工生产的效率太低了，所以使用 AIGC 的技术，生产比如说不管是文本、图片、音视频内容的提供方，在这个过程中也会有足够的机会

您的公司应该也是做类似于终端这块的是吧？

1 我们在过去的几年时间里面，主要是给银行和政府提供服务，把刚才说的这些技术，比如说智能语音、NLP，甚至还有数字人的这些软硬件结合在一起，解决一些无接触产品和服务的东西，软硬件结合在一起的

a 比如说以金融行业的银行业举例，有一个巨大的需求，我们知道传统的银行业是人力资产密集型的，因为疫情的影响也好，或者现在技术进步的影响也好

b 现在中老年人可能不方便去，年轻人不爱去，这种柜台业务极度萎缩，所以银行面临着巨大的降本增效的压力。

2 要解决这个问题无非就是两条路径可以走

a 第一，对现有的银行网点进行升级改造，提升智能化、数字化的水平。所以这里面可能会部署一套软件，外加上配套的，比如说一个网点部署 5-10 台左右的硬件。原来一个网点要雇 20 个人，现在只要雇两个人就可以了，这是一个方向

b 再有这个网点实在运营成本太高，需要把它裁撤掉，这个柜台业务尽管少还是要有，需要应对这种需求的变化。那怎么办？可能有点类似无人超市的产品，一台售价 70-80 万人民币，原来的网点可以覆盖 20-30 个小区，这种设备前置部署到小区，从总的成本投入上要员大于网点运营的，但是可以增加便捷性

c 以前我要驾车或者乘坐公共交通工具去网点柜台办理业务，现在下楼可能就能办理了。这种技术在 ChatGPT 的加持下有可能取得大的突破。我们知道全国银行网点有 22 万个，这个规模是足够的

您说的软硬一体的表现形式是什么？是类似于我们平时接触的 ATM 吗？是一个机器还是什么？它可能和我们平时接触的 ATM 机器有什么不一样？能处理非现场业务还是什么？具体的形式是什么样的？

1 比如说跟 ATM 这种机具结合在一起，也会有屏幕，就是当你在办理业务的时候，遇到问题的时候，会有一个数字人代表的 AI，通过语音对话的方式直接跟用户交流，这样就不需要在大堂部署客户经理了，相当于这么一种做法

2 这块从 2019 年末，2020 年初就开始做这些事情了比如说在中行网点，在浙江省银行，在很多地方开始部署了

在政务领域呢？能类似的举一个例子吗？因为大家对应用场景没有一个感性的认知。

- 1 在政务领域其实跟我们给银行提供服务是类似的，比如说党建，一个村党支部，或者我智慧社区，我一个社区的办公室，可能也是类似于银行网点部署，它也是软硬件结合在一起去做，只不过从收费模式上略有差距。但是相比较提供基础的模型，它是一个高附加值的，因为软件就是按软件授权，硬件就是按照肯定会做足够的毛利空间，目前大概是 40%左右
- 2 软件因为不同的行业，其实它只是在内容差异，比如说我把银行的数据拿过来，我训练大概有 6-12 个小时，训练完了就可以部署了，同样一套软件，把这个东西拿过来，把政务的数据拿过来也是训练类似这么长的时间就可以处理了
 - a 我是说单纯从交互层面上，在不同的行业差异只是一些不同的 RPA 流程，比如说在银行是办卡，跟柜台业务有关的东西，数字政务更多的是跟政策，跟民生服务相关的，智能流程自动化模块的接入会有差异化，会有一些不同的地方
 - b 但是未来这个方向可以通过类似开放平台的方式去做处理

您现在用的是一套自己的模型和算法，数据是属于银行的，拿到他们的场景里用他们的数据训练，实现一个最终能够对话，或者能够解答问题的效果？如果百度推出这个，或者国内基于中文体系推出这个的情况下，从您自己发展的规划来说，您准备自研一套呢，还是接入到大厂里面去？

- 1 对。单从我们自己来说，我们这个过程中一直和商汤、百度、阿里、腾讯这种级别的公司存在比较直接的竞争关系，所以我们在这个过程中肯定用自己的
- 2 但是我们和他相比我们的优势在什么地方
 - a 我们的优势可能不是在于基础模型的研发，我们在解决最后一公里的问题。比如说一个智能机具，前面你和我，我们两个人都站在智能机具面前，它处于唤醒状态的时候，咱们两个人同时说话，这个时候它需要判断出来需要跟谁去交互，或者咱们俩说的话它是否需要做出回应，它必须得具备这种能力
 - b 这种工作对于大厂来说，它没有下沉到这么深的地方去专门干这些事情，它也认为这些成本对于他们来说，或者收益对于他们来说，完全不成正本的，他们看不上这块的东西

背后的大模型，有些开源的技术可以去采用是吧？站在巨人的肩膀上，不用从头研发的？

- 1 是这样的，但是你获得的数去有独特的优势
- 2 比如说单纯的语音识别上来说，比如说讯飞做语音识别最好的，但是你把 AFR 这个功能放到网点上发现出错的几率太高了所以还是需要专门针对这块做大量的纠错算法的处理，推理的一些处理，还需要做大量的训练
- 3 这就是小的创业团队也好，或者专门做这块服务的企业也好，他可能在这块具有优势

一方面利用已有的开源的，或者框架性的东西，不用从最底层开始做，在某些方面可以深耕是吧？如果软硬一体的话整个毛利率做到 40%以上。如果是软件的话，相当于把软件加到现有的机具里面去，可能是更高毛利的收益的？其他的交付还需要什么？除了硬件采购的成本之外，还有别的成本吗？比如说现场安装配制这些？

- 1 是这样的，硬件大概是 40%-45%
- 2 软件这块可以做到 70%以上，大概这样的，那些成本都包含在内

ChatGPT 算力、研发、进展情况

- 1 初始是谷歌做的 TRANSFORMER，CHATGPT 一代就有 40 几亿个参数。参数可以理解为深度学习的时候讲到的神经网络，一层有 1,000 个节点，节点互联，是 1 兆个参数。CHATGPT 是 1,750 亿个参数。我们实验中做过的最大的模型是 10 亿个参数，所以大概是我们的 175 倍
- 2 TRANSFORMER 核心就是用大量的语料进行训练，做出理解力很强的 AI。语言理解力是我们长期攻克的难题，中文是最难的，首先要做断句分词、根据上下文研究语法模型。但从 TRANSFORMER 开始，不需要做断句分词，所有都从语料库来，实现理解。统一了所有语言的学习方法
- 3 OPENAI 从 1 代到 3 代，模型不断放大，从 43 亿到 1,750 亿。网络复杂度是几何级数的增加，我们估算是 2-3 次方之间
- 4 不能光看参数大小，参数很多，但是节点互联很简单也没有意义。算力投入一方面是规模，另一方面是迭代次数和时间
- 5 CHATGPT 此次是用了 5,000 块 A100，每块节点是 8 块 A100 全互联。美国对我们禁运的首先就是 A100。用 NVLINK 进行互传。如果不用 NVLINK 的话只能用 CPU，传输时间会变长。国内不会很快追上 OPENAI 的算力投入
- 6 截止到 2021 年底：因为新知识可能会产生矛盾，出现之后需要人为处理。知识还需要进行重新校验的，所以不能接入在线数据

中国的 ChatGPT 有哪些瓶颈

- 1 从处理器的角度，我们有很多可以支持计算的处理器，例如海光、飞腾、寒武纪等。最大的问题是做不了高制程，只能做到 28NM，28NM 会导致芯片密度低，能耗更高。导致 NVIDIA 一块卡我们要用 2-3 块
- 2 第二个问题是互联，有一些机器的互联可以，例如飞腾等，但是有产能问题。缺乏 NVLINK 的互联设备，我们在做大语言模型的时候问题就很大
- 3 海光有不到 3,000 块的 A100
- 4 CHATGPT 能够做到 99.9 几的理解能力，我们如果降低要求到 97-98 之间，阿里京东百度等自身的算力就可以在 1-1.5 年内做到这个水平。短时间追上 OPENAI 是很困难的。微软的投入不光是钱，更多是云计算的基础设施，例如 5,000 块的 A100，按照市价，算力投入在 3,000 万。但如果降低要求，现在相关领域做应用的话还是有机会的
- 5 瓶颈最主要的还是在数据交换上，如果互联不好的话，5,000 块卡和 100 块卡的效果是一样的

商业应用价值

- 1 国外很快就会淘汰数据程序员，可能会淘汰一半，交一部分的工作给 CHATGPT
- 2 国内程序员供应量多，如果用 CHATGPT 取代反而成本会更多。但另一方面，国内有些行业，例如 3D 建模的人才，我们目前是缺乏的，3D 建模本身门槛不高，但没有办法跟机器比速度。CG 图片、特效成本可能会迅速降低
- 3 另一个就是销售客服。现在很多地方都是机器客服为主，国内做外贸的很多公司都需要英语销售客服，但他们跟外商的交流很不畅通，如果换成 CHATGPT，可以看到工厂的实时库存、帮助下订单。CHATGPT 计费标准来说，这样的替代也是很便宜的

中美竞争

- 1 如果看大模型的话，中美的差距有可能会拉大。AI 模型、芯片产业链都需要投入
- 2 如果降低 NLP 的理解能力，不影响后续工作的话，也是一个方向

数据分工

- 1 国内主要问题在于复制性的内容太多，国外相对而言能够找到一些权威的语料。这是中文和英语的最大差别。权威就是可以告诉机器，从 NATURESCIENCE 上面找到的文章，权威值就是 99%，我们在万方知网上面能够找到的权威数据有限，这些论文发了中文也都会发英文，我们是比较吃亏的。中文语料训练上没有问题，但是在知识质量上我们会低一些
- 2 微软数据标注团队质量很高，我们的劣势：有劣势但是不大。人工标注的话不会具体到每一篇文章，只会是标注刊物、高引的文章等。例如知乎的高赞文章是不是高质量的需要人为来进行鉴定

确权的问题

- 1 CHATGPT 的创作思路跟我们所说的洗稿很像，据说 OPENAI 会给以后生成的文章里面打技术性水印
- 2 少数作者会启动防止被引用的开关，大部分作者不会启动
- 3 NEWBING 会生成来源
- 4 175G 的参数需要 1T 的内存才可以运行，开源是不太可能的

审核

- 1 现在是基于关键字和简单语义识别
- 2 但如果变成 80% 的 CHATGPT 水平，发出违规文字的可能性就没有了

公司竞争力

- 1 做类似 CHATGPT 需要几个元素
 - a 语料：自有平台或者爬取网页。微信有公众号朋友圈，百度是爬取，京东是产品和用户咨询评价，相对比较少。360 和腾讯有搜索引擎，应该还在爬取。自有平台大一点就是知乎、腾讯、网络小说、门户网站等
 - b 算力：一个是看自有平台的 CPU 和 GPU，现在比较强的是阿里腾讯，都有 GPU 的云。地方的数据超算中心都有一定的算力资源
 - c 钱：OpenAI 拿到了很多投资
 - d 人
- 2 京东淘宝可能需要的是智能客服。百度可能需要的是自动驾驶和 AI 的联动。腾讯可能就是游戏的开发

搜索成本

- 1 谷歌的首次搜索的成本高，但是重复搜索的单次成本不高。谷歌主要成本是不断爬取数据的成本
- 2 如果放到搜索引擎上，直接的问题就是广告放在哪里
- 3 搜索引擎市占率低的公司可以通过 CHATGPT 来抢市场，但是市占率高的公司如果运用的话会把最赚钱的广告业务挤掉

AI 方向

- 1 国内可能会在 3D 建模、网络小说方面发展
- 2 下游的游戏影视也会有变化

核心观点

- 1 CHATGPT 和以前 GPT 大模型的区别：CHATGPT 是基于 GPT3.5 的基础模型框架，核心变化在于通过真实的调用数据以及人类反馈的强化学习进行训练
- 2 国内企业与 CHATGPT 的差距：百度、华为等头部厂商都有大模型，百度文心模型参数量已经达到 2,600 亿，从技术能力上国内相比专家判断比 CHATGPT 约晚 1-2 年
- 3 OPENAI 第一梯队，GOOGLE 第二梯队，百度第三梯队。从数据、算力、模型的维度上，主要是差在模型环节，包括清洗、标注、模型结构设计、训练推理的技术积累

训练成本昂贵

- 1 GPT3.0 离线训练成本高达 1,200 万美元，大模型重新训练成本达 400 万美元，都是租用微软的计算资源，成本与训练数据量成比例，增加 GPU 也可提升训练速度，中小厂商或将无法承担高昂的大模型自研成本
- 2 模型训练所依赖的高性能芯片被美国卡脖子的解决方案：采用国内自研 GPU；用分布式 CPU 替代，部分大厂应用采用分布式 CPU 的方式进行训练

未来商业模式

- 1 OPENAI 或将对外输出模型能力，国内百度等企业也会跟进
- 2 国内应用软件企业未来可能采用 OPENAI 或百度等厂商的 AI 模型作为基础能力。

专家发言

- 1 首先介绍一下 CHATGPT，CHATGPT 去年出来后，在全球 AI 界、创投界掀起热烈的讨论潮。很多人想了解 CHATGPT 具体是做什么的，我们可以理解为通用的问答系统，发布方是 OPENAI 公司，OPENAI 是业内比较知名的人工智能研究机构
- 2 CHATGPT 模型从 18 年开始迭代为 GPT1，到 19 年有 GPT2，到 20 年有 GPT3 等等，通过一系列的模型迭代，在今年推出了 CHATGPT。这个模型是基于 GPT3 的模型框架
- 3 GPT3 的参数数量级非常大，有 1,700 多亿，所以这个模型出来后效果比较好，大家体验下来发现功能强大，在业界受欢迎。

ChatGPT 模型关键的能力来自于三个方面

- 1 前身是 INSTRUCTGPT，通过真实的调用数据以及反馈学习，CHATGPT 在模型结构、学习方式和 INSTRUCTGPT 基本上是完全一样的，INSTRUCTGPT 基于 GPT3.5 的强大能力，整个模型通过 OPENAI 一个系列的迭代，有很多技术积累
- 2 CHATGPT 主要的 3 个学习阶段包括，从 OPENAI 调用数据集中采取 AI 训练编写答案，第一阶段通过监督学习方法训练 GPT3 版本，然后用比较型的数据训练一个 REWARD 模型，用到一些对比学习和 REWARD MODEL，第三阶段是强化学习里面的 PPO 算法和奖励模型的语言生成策略
- 3 跟之前的很多模型比，单纯用监督学习或无监督学习以及强化学习的端到端形式，现在分成多阶段来做，整体模型的训练开销非常大，这个模型有 1,700 亿参数，训练一次要 400 多万美元的成本，对资源的使用比较大

分享一下国内头部玩家，比如百度、阿里、腾讯、字节、讯飞、商汤等目前 AI 大模型的参数量，以及与 ChatGPT 的差距在哪里？我们用多久能追上？

- 1 国内这样的大模型非常多，百度、华为都有大模型。百度文心模型参数量已经达到 2,600 亿，不逊于 GPT3.0。百度今年宣称 3.4 月份推出类似于 CHATGPT 的模型
- 2 可能首期的应用场景会跟搜索结合，会形成 CHATGPT、搜索的双引擎结合模式，会发布 DEMO 出来，整体看，模型的参数量级比 GPT3.0 大了 50%

我们国内的发展水平是不是和 ChatGPT 达到同等程度？

- 1 这个地方可以看一下 OPENAI 发展的技术史，动态看 18 年后一到两年时间国内的技术追赶比较快，提出比较好的改进模型，但真正的差距拉开是 20 年中，也就是 GPT3.0 出来后
- 2 当时可能很多人觉得 GPT3.0 不仅是技术，体现了 AI 模型的发展新理念，这个时候我们跟 OPENAI 的差距拉得比较远，因为 OPENAI 对这一块有技术的独到见解，也领先了国外的 GOOGLE，大概 GOOGLE 比 OPENAI 差了半年到一年的时间，国内差了两年的时间
- 3 所以梯队上，OPENAI 排第一梯队，GOOGLE 排第二梯队，百度排第三梯队，差距不是那么大。动态看事情的发展，百度还是有机会做得更好
- 4 从百度发布的模型看，内测版还没有开放出来，我们体验下来发现效果不错，模型参数量级看上去和 GPT3.0 是不差的，所以目前看上去我们相对比较有期待

国内主要厂商

- 1 比如字节跳动也在做大模型，一方面是应用场景，接下来可能有一些商业化。这个模型不是一两个月就能做出来，需要时间积累，随着后面 AIGC、CHATGPT 的深度学习发展，里面会形成越来越多的商业化内容，对于头部公司来说，做前沿的技术积累，有利于形成先发优势
- 2 所以像字节跳动、阿里都在往这方面做。阿里、腾讯不会复制 CHATGPT 模式，而是做一些与 AIGC 相关的产业化
- 3 阿里现在往智能客服系统、AI+营销做，使得商品介绍、广告介绍不需要人工生成文案，而是通过 AI 生成内容。头条也用 AI 生成内容，因为头条是以内容生产、分发为主要业务的公司
- 4 总结看我们跟国外的区别是，模型大小上我们很多公司可以媲美，从技术角度或者 PAPER 角度，可能我们跟他们的差距可能有一到两年的距离

OpenAI GPT3.0 2020 年推出，现在推出了 ChatGPT 之后市场大火。ChatGPT 相对于 GPT3.0 有什么实质性的改变，造成市场认可度高？热度会不会突然下去？

- 1 现在 CHATGPT 做了会话机器人，不是 IT 圈子里面的人也可能去试用，使得人群受众变广，其实基本的模型、训练方式，两者之间没有多大区别，而是在 GPT3.0 基础上做了特定领域的训练数据，拿过来做训练得到特定领域的模型。所以总结下来不是之前的 GPT3.0 做得不好
 - a 现在比较热，热度会不会过两个月下去？坦率说，从我的认知来讲，短暂时间内热度不会下去。从业界很多企业对这个事情的反馈来看，比如 Google、百度相继在发力
 - b Google 已经列为红色预警，包括将来对搜索引擎的颠覆，有很高的优先级；百度也是因为搜索引擎，对这块非常上心，3 月份会推出新的模型，到时候看市场的效果
 - c 所以从大公司的反应来看，这个事情可能不是噱头，而是新的技术革命的发展。甚至有夸张的说法，业内一些专家认为这是 PC 互联网之后的又一个里程碑式的发展。包括我们自己体验下来，ChatGPT 的功能确实是很智能
- 2 而且随着 OPENAI 的研发，接下来朝着 AGI 方向发展的更大规模的模型，我们有理由期待这个事情不是一波流的，而是从现在开始持续形成大的影响，并且扩散出去

从数据、算法到算力，国内有没有可能做出 ChatGPT 这种有商业化价值的模型？听到一种说法，国内号称模型参数高，但效果特别差？我们中国独特的市场有没有办法培育出比较好的产品出来？

- 1 从 GPT 模型的 PIPELINE 来看，有数据、算力、模型。数据上，国内是不缺的，百度、今日头条有大量用户的真实场景的数据。算力角度，一些大厂比如百度的文心模型达到 2,000 多亿参数量，虽然训练成本非常高，但大公司承担得起
- 2 我们真正缺乏的是技术积累，包括数据怎么做清洗、标注以及模型的结构设计，怎么做训练、推理，这个地方很多都是需要经验和积累。包括里面的很多算法是 OPENAI 和 DEEPMIND 提出来的，作为原作者的理解能力是我们远不及的
- 3 20 年 GPT3.0 出现的时候，我们跟国外的差距就拉大了，当时很少人觉察到 GPT3.0 不仅是一项技术，那个时候我们开始逐步落后，坦率讲国内这块技术方面落后于国外，而数据、算力不会存在瓶颈
- 4 总结下问题，我们跟 OPENAI 的距离确实存在，重点的技术瓶颈可能在这块的模型上面，这也是国内的弱点
- 5 所以我们需要投入更多的人力、资源逐渐弥补上，百度投入也非常多，按照这种速度下去，也许不用花两年的时间，我们能追上步伐

大模型为什么会产生？百度文心模型为什么不如 ChatGPT？

- 1 涉及到模型参数的量级，以及数据的量级。比方说算力提升 10 倍，那么把模型参数做更大，或用更多数据来训练
- 2 如果我们单变量来看这个事情，数据变大，模型效果变好，或者单纯把模型参数变多，模型效果也更好。如果算力提升 10 倍，这两个怎么平衡？光把模型参数变大，不一定会效果好
 - a 数据集的质量也要有保证。单纯的模型大小不足以刻画模型效果，还跟数据集的大小、质量有关。第二点，模型之间存在差异
 - b 百度、OpenAI 模型的本身结构不一样，参数量大不代表模型效果好，因为不一定学到通用能力。百度没有公布模型的细节，我们不太了解是怎么做的，而单纯通过参数量级比较，不大能够看出效果的差别
 - c 百度对这块细节公布不多，数据集大小、模型的框架结构没有特别公布，所以不大了解整体模型的效果，以及跟 OpenAI 技术的差距有多大，我们只是单纯从参数量大小判断，不太好判断
- 3 我们很好奇一件事情，比如一个模型学了很多东西，东西学到哪里去了？这也是一个问题。大家可能会去想，那么通用的东西存到哪里去了
- 4 这一块目前做研究的同行非常多。目前一个主流观点是模型主体结构用 TRANSFORMER 结构，更多做一些知识的存储，包括可能涉及到 KEYVALUE 的形式，涉及到知识检索的能力，但里面的细节就需要比较多的时间来解释

OpenAI 大模型今后是否提供给个人或机构使用？国产办公软件底层的人工智能，会不会用微软 OpenAI 微软产品中各种场景都会接入

- 1 其中 2B 的业务比较有影响，比如订阅 OFFICE 软件的企业的员工可以享受智能化服务。国内软件目前没有这个功能，为了面对微软的竞争，未来都会往这个方向发展
- 2 国内其他公司如百度也会通过输入 AI 能力来进行 2B 的商业化布局，所以对国内的企业来说也可能会有很多选择，可以选择接入 OPENAI 或者 GOOGLE 海外公司，也可以接入百度等国内公司提供的 AI 模型
- 3 市场竞争看模型的整体效果和价格。此外也涉及法律法规的问题，目前 AI 底层能力的使用这方面还没有特别的限制，还是法律盲区，未来是否能够顺利引入微软 OPENAI 的算法模型，还要有待未来立法的规范
- 4 但是 AI 模型未来接入应用场景是大势所趋，具体接入产品看情况

AI 是否会大幅降低人工，未来 AI 为底层生态之后，之后在传统软件的参与者会减

少？国内大模型做的更好之后，小模型的企业存在的必要性还有吗？

- 1 有些行业的工作人员确实有可能被替代，内容生产的，AIGC 的占比会大幅提升，挤占投稿的量级，AI 也帮助提高了生产力，帮助投稿的质量和效率提升
- 2 NLP 领域的会存在这样趋势，小模型的厂商会慢慢被竞争掉，中小厂商没有办法去创造大模型，接入 2B 的服务，没有办法自研，这种模式在将来会存在

训练中采用的硬件设施通过买或者租的形式，每次训练迭代成本怎么计算？买和租各要多少钱？国内那种方式做大模型比较多？

- 1 GPT3.0 涉及 1,700 亿参数，内存达 300+GB，训练过程耗费 1,200+万美金，此外离线训练后，访问请求来的线上服务也需要开销
- 2 基础设施都是微软的计算系统进行，没有自建。训练过程中会出现 BUG，但是都没有敢重新训练这个模型，因为成本太高了，所以没有做及时的修补，披露出的数据是重新训练一次要 400+万美金

据说 ChatGPT 是 40 个人做强化学习，为什么能在人员投入这么少的情况下，在 GPT3 上做出 ChatGPT？核心能力是什么？

- 1 CHATGPT 跟 GPT3 的主要区别，额外 FINETUNE 增加了训练的数据集，增加了通过人工标注的样本，所以 40 个人是不准确，实际上还有大量标注的团队，40 个人可能只是用来判断回答是否符合标准
- 2 从算法的角度来说，GPT3.0 和 CHATGPT 在模型上没有太大区别，只是加了专属领域的数据集进行训练

替换人工的问题，传闻数字媒体已经被替代了，实际情况如何？GPT3.0 专注于专业领域，是否比 ChatGPT 更快一些？

- 1 AIGC 各家都开始鼓吹并开始投入，到 2023 为止，网络内容 20%通过 AIGC 生成的，网络编辑是很容易替代的，比如摘要、新闻等
- 2 头条这边也在逐步用 AI 生成投稿，这块趋势是通过 AI 生成内容会逐步增加，机构预测 2025 年 AIGC 内容占比达 30%+，量级很大，增长也很快，能够帮助大幅提高生产力

和其他 AI 模型相比，对 GPU 的需求区别？

- 1 GPT3.0 对 GPU 算力要求很大，训练过程采用 400+个 GPU。未来对 GPU 需求越来越大，价格变得贵+产业链供应能否跟上需求是重要的限制条件
- 2 这也导致创业公司会更难做相关的事情，可能更多集中在大厂做这些事情

ChatGPT 训练数据截止到 2021，但是对话内容也可以涉及最新的信息，怎么结合在一起？

- 1 模型具有泛化能力，模型可以做到这一块，通过历史数据训练，不可避免有时间限制，2023 年的问题可能在模型里面找不见，但是其他字段能够在模型找到联系，模型可以实现预测
- 2 同时也跟数据分布有关，历史数据的函数关系，已经能够刻画绝大部分关系了，用新的数据训练，可能也会有类似的分布

训练的频次怎么界定，是否要每隔一段时间后更新数据集，对大模型重新训练？

- 1 比较快的是用 FINETUNE 的形式，但是引入新的数据集可能会导致灾难性的遗忘问题，这就是 LM 模型的难点问题，也是主流的研究方向
- 2 我们也可以累计到一定程度进行重新训练，但是这样成本高，未来 LM 模型的训练方式也是会持续优化，如何不断加入新的训练数据，同时又不导致原有模型性能下降

ChatGPT 训练依赖高性能的芯片，中美竞争情况下，被卡脖子的情况？

- 1 大模型依赖 GPU 的芯片，如果对 GPU 芯片封锁，会有比较大影响，解决方式是有

国内自研的芯片，二是我们大量用 GPU 进行训练，供应能力跟不上需求

- 2 也可以用分布式的 CPU 进行替代，会慢一些，字节这边是分布式的 CPU 用的比较多，包括主流的搜索推荐等需求

模型训练+租金是 1,200 万美元，离线训练的阶段需要多长时间？

- 1 依赖于 GPU 的投入量，从而加快训练速度，之前提到中美在这个领域的差距有一年半到两年是指达到整体的效果所需要的时间
- 2 实际人员整体投入应该很大，OPENAI 光论文中研究人员都有 30 多人，还有大量的工程师等，百度这边投入也有 200-300 人。新增数据训练的费用，要看增加的训练数据的量级，费用和数量是成比例的

首选百度还是 bing？预测哪个效果更好？

- 1 不太好判断
- 2 需要产品发布后比较

未来两年的角度，中美 AI 领域的差距是否会因为 GPU 芯片禁运等原因差距拉大？

- 1 GPT4.0 今年会发布
- 2 性能会有进一步提升，差距拉大是有可能的

ChatGPT 类的模型是否对知乎这种内容平台起到更大的作用？

- 1 知乎本身就是问答系统，问题的回答有可能通过 CHATGPT 来实现
- 2 甚至比人工效果更好，包括程序 DEBUG 的角度也可以朝着自动化靠拢

只有大的平台能做大模型，知乎能自己研发吗？

- 1 涉及到商业模式的问题，知乎这类的企业可能不需要自己研发
- 2 未来可能百度等头部厂商会对外开放这种能力，其他企业直接采购使用就可以

技术突破有望构建新的商业模式

- 1 对于投资行业来说，AI 并不是一个全新的领域，大家持续关注也有近十年的时间了
 - a 从 2012 年左右深度视觉模型带来 CV（计算机视觉）的快速发展开始，到 2016 年前 AlphaGo 给行业带来的震动，到这次 OpenAI 的大模型 ChatGPT 引起广泛的触动，底层逻辑是大模型的性能带来的质变提升，从而衍生带来新的垂直场景解决方案的构建，以及新商业模式塑造的可能性
 - b 一直以来，AI 领域较多的解决方案是，通过端到端垂类应用，用小模型优化算法和模型精度，以降低算力去解决垂类场景问题，但如今这一商业思路受到了较大冲击。因为在过去的路径下，行业内的公司大多重复研发功能相近的产品或模型，让 AI 只能在一些有限的场景内，较多是 to B 领域去解决问题，壁垒不太深
 - c 现在，大家开始对大模型在多场景泛化的能力满怀期待。一方面表现在，随着大模型性能的提升，AI 在更多消费级 to C 场景加速渗透，消费级场景潜力巨大。另一方面，AI 也在过去近十年间完成了从感知、理解到推理性能的全面提升，由此带来定义新一代交互、生产方式、生产关系的可能
- 2 对于投资机构来说，大家都对新一代大模型的性能打开和赋能 TO C 场景感到振奋，因为技术的突破使得纵向构建新的商业模式成为可能
 - a 过去几年内，由于 to C 消费级场景没有大的技术性突破，整个行业陷入了突破性的革新停滞，因此行业内只能不断内卷和重复开发。而类 ChatGPT 大模型为基础的公司的出现，让行业重拾新的信心和产品/场景定义的激情
 - b 同时，我觉得它们也给行业带来一定的反思和激励。在过去一段时间内，资本市场对以 OpenAI 为代表的、底层基础学科科研导向公司的态度非常审慎，因为它们大部分从非营利机构起步，专注科学底层突破，在看不到快速商业化前景时持续做投入。但相信以后一段时间内，这类公司会得到更多关注和资本配置
- 3 但这并不意味着相关创业公司的前景一定乐观，因为从目前来看，受限于算力、历史投入、科技人才密度、数据量级、资金实力等方面，大型科技公司的优势明显优于创业公司。大模型通过被持续调用、喂新数据、训练，不断强化性能，并在应用场景上泛化，有较大可能挤压提供小模型闭环方案的一些创业公司的生存空间
- 4 所以，我们目前会更关注大型科技公司在底层模型上的动态，以及业务流能借用大模型的性能、并且在垂类应用上构建出一定壁垒的创业公司
- 5 但以下几种类型的创业公司也会受到我们的关注
 - a 一是市场上已有 AI 公司对大模型的融合
 - b 二是已有消费级公司，借助大模型构建的与现有业务集成起来的新功能
 - c 三是以 AI 为基础研究导向的新科研团队。一旦它们构建了相应的壁垒，未来非常值得期待

生成式 AI 的创业者要技术为先

- 1 CHATGPT 的出现，给全社会带来很大震动。它来源于核心算法模型上的突破（例如：TRANSFORMER），可从海量文本（其中包含部分程序代码）中以无监督学习方式训练出巨大参数量的大型语言模型（GPT 系列），具备强大的自然语言理解和生成能力，并“涌现”出一些之前业界从未想象到的其他能力，例如基础的世界知识、上下文学习、基础推理等

- 2 与此同时，通过程序代码学习、人类反馈的强化学习和有监督学习等进一步的微调训练，可使其在人机交互和内容输出方面的体验更加友好。语言是人类重要的沟通手段，而且还是各类知识和推理的主要载体，GPT 系列大型语言模型的通用性更好，可无示例或少示例直接完成部分下游任务，应用范围也更广，这都让人们看到了更大的发展潜力和空间
- 3 此前，我们一直想做的是让 AI 代替人类完成“简单”的重复性劳动，例如安防、驾驶等，从而解放人类去做更“难”的知识性和创意类工作，因为我们倾向于认为机器不具备这种类型的能力；但如今的生成式 AI，率先挑战的是知识工作者，它掌握了海量知识，可以轻松写出各类文章与程序代码，在某些方面的创意能力甚至超过人类
- 4 OPENAI 的 CEO 山姆·阿尔特曼认为，我们过去对于劳动难易的划分准则可能有问题，现在看来 AI 最先取代的可能是知识工作者和创意工作者的工作
 - a 最近，类似赛道的项目很火，我们也在积极地看。曾有人问我，大家会不会“抢项目”？我的回答是，暂时还好。因为在上一轮的 AI 发展浪潮里，大家都投了很多，但实际产出与效果没有预期的那么明显。所以，我观察下来，现在投资行业持理性观望态度的居多，但我认为这确实是一个很大的机会
 - b 上一轮 AI 热潮，大家投资的主要是以深度学习为代表的判别式 AI。峰瑞在 2021 年下半年开始关注生成式 AI 赛道。我们看到 AI 语言大模型在不断更迭升级，从 2022 年开始很长一段时间内，据说平均每 4 天就会出现一个新的大型语言模型，发展速度很惊人
 - c 不止是 OpenAI，包括美国的很多大厂也推出了自己的大模型，虽然有些模型没有开放出来给外部使用。AI 发展至今，对“参赛者”的筛选是残酷的，入局早的已经有了先发优势，因为其对行业的集中度要求比较高
 - d 从已经推出的大模型本身来看，它主要是靠规模取胜，这背后需要高密度的顶尖科研人才、巨量计算资源、海量训练数据，以及良好的用户或合作者生态。只有同时具备这几个重要条件，才能更好地训练模型、拓展应用边界、给未来的商业化打下基础。在这一点上，规模较大的公司（大厂）相对更有优势
- 5 国内在 AI 领域的发展相较于美国慢了一步，这体现在人才储备、数据积累以及资源投入上。不过随着中国逐渐出现自己的大模型、有更多的人才积累，以及整个研究和行业发展氛围变得更好，以我们的技术迭代效率，相信应该具备追赶的能力
- 6 我们认为，当前的生成式 AI 创业机会分为模型层和应用层两大类，应用层又可分为 TO B 和 TO C 两个子类。但除大厂之外，模型层的创业是比较难的，因为前期投入实在太太大，可能短期内比如两三年之内创业公司跑出来的概率较低。创业公司可以考虑在垂类模型或者应用层有所突破

但是在应用层创业，在做商业模型设计的时候要警惕模型层的边界，最好保持一定的“安全距离”

- 1 以新锐独角兽公司 JASPER.AI 为例，它基于模型提供商 OPENAI 的 GPT3 模型的 API，构建自己的商业模式，但是没想到 OPENAI 后来推出了 CHATGPT 这个新产品，免费提供了部分 JASPER.AI 的功能，某种程度上影响了 JASPER.AI 的发展
 - a 2022 年，我们投资了一家生成式 AI 公司，它的主要技术亮点是视频和直播的数字内容生成，商业模式是围绕直播和短视频电商的内容制作、数据归因、生意参谋等服务。这个商业模式有足够的行业纵深，且具备明确的商业化能力。这类模式相对好很多
 - b 从去年下半年起，我们看了很多生成式 AI 的项目，我觉得这个领域的创业要点是：技术为先，场景为重

- 2 创业者首先要对技术有深刻的认知，要对技术有不断学习深挖应用的能力，而非拿个模型直接简单使用；其次是，当前这个领域处在科研快速迭代和广泛传播的状态，技术的溢出效应很明显，这意味着同一项技术对大小平台的赋能作用相差不多，因此，创业者必须深入到行业场景中去，构建自己独特的壁垒
- 3 比如用户群体、行业知识、产品服务、技术算法等等，从而避免陷入同质化竞争的红海中去。CHATGPT 是 AI 技术范式变化，看好国内创业公司夏令明势资本合伙人 CHATGPT 的大火，并不是突然发生的
- 4 其背后的技术趋势和变革已经发展很多年了。从 OPENAI 的 GPT-1/2/3，再到现在的 3.5，每一代都在进步，终于实现了通用人工智能 AGI 的突破，产生了 AI 技术范式的变化
- 5 明势资本从 2021 年底、2022 年初就在考察、投资和布局这一赛道。在明势资本看来，目前市场上有三类 AI 公司有投资价值：一是专注于大模型的公司，以 OPENAI 为代表；二是既做大模型，又做直接应用垂直一体化的公司，比如 MIDJOURNEY；三是调用大模型 API 的公司，与第一类公司合作，重点开发具体场景的 AI 应用公司，如 JASPER 等
- 6 大模型方面，尽管它的技术壁垒和门槛相对较高，但中国的创业公司从迭代效率和反馈质量上而言，比欧美很多公司更具优势。所以即使国内的大模型现在跟海外有差距，但我们认为这些技术差距是可以追赶上的

在此基础上，我们认为国内创业公司也有很大发展机会

- 1 尽管之前大厂也都在做大模型，但过去他们的投入和重视度并不够，因为他们没有以工程化的方式去做大模型，所以目前大厂的大模型，在实用性上相对不足。而国内的大学研究院也不是基于工程化产品化的视角去做，也存在一定缺陷
- 2 另外，CHATGPT 这类应用，未来可以广泛应用到生活的各个方面，在数字经济的各个方面都有结合的机会，也会相应产生非常多的丰富应用，这肯定会带动一批创业公司去做探索和尝试
- 3 这一背景下，国内已有少数具备很强技术研发和工程化能力的创业公司投入大模型研发，甚至比国内大厂有先发优势
- 4 而通过明势已经投资的公司，我们也能深刻体验到新一代端到端数据驱动模型的价值。最开始的模型相比于上一代 AI 模型并没有显着优势，但在数据驱动下，以两个月一迭代的频率推进，就会发现模型的能力呈现指数型增长，很快就能达到显着超越传统模型的效果
- 5 数字经济时代，无论是工作还是生活场景，都存在着海量的文字、代码、图片、声音、视频的生成和理解需求，人类对于效率、成本和体验的追求是无止境的，所以大家都在期待 AI 能够在各个领域发挥作用，但目前它还无法完成高度抽象、高质量或高可信度的任务
- 6 因此，我认为如今 CHATGPT 只是新一代 AI 技术的起步，即技术与商业刚开始产生交集，这个交集会越来越大，未来技术的天花板以及它带来的商业可能性还是非常广阔的

AI 大模型的未来：赋能万业的“水电煤”

- 1 如今，AIGC 把人工智能技术带向了一个新的高度，是人工智能发展过程中一个非常重要的拐点。AI 企业也从主要集中在深度学习的判别式 AI，发展为产出新内容的生成式 AI，这是技术的一个大的飞跃
 - a 从技术层面分类，我们把整个生成式 AI 分成了三层。底层是生成式 AI 的基础设施层，中间为最重要的模型层，以及最上层的依赖于底座模型的应用层
 - b 在我看来，今天生成式 AI 所依托的超大规模语言模型是一个更先进的算法架构，集中了人类的全部数据，汇集大量算力，进行集约化训练，训练出来的 AI

能力供应所有用户使用，这也是十几年前 AI 界专家所设想的那样，AI 未来一定会成为类似“水电煤”的人类公共资源

- c 只有这样，才能做到 AI 赋能万业。所有行业应用中，目前 AI 技术的渗透率只有百分之二三十，有些行业虽然已经实现了这一渗透率，但只是应用在一些比较浅层的商业节点，离全面赋能行业还很远。因此，我们也考察并划分了更具投资和创业价值的领域
- d 智算平台更适合有规模效应，如已有的公有云平台，或数据中心运营方介入
- 2 而工具链方面，随着大模型持续发展迭代，用更好的工具去帮助大模型的训练和管理去降本增效，一定是有机会的，且会有一些顶级 AI 工程人才不断进入，我觉得这肯定是一个很标准的 VC 投资机会
- a 在闭源模型及 API 服务领域，未来肯定也会有越来越多科技大厂进入，创业公司是否还有进入的机会，我觉得值得观望，因为这对于人才密度的要求和资本的投入都是巨大的
- b 技术开发应用的基础上，我觉得未来的主要趋势，是利用第三方模型构建应用的方式。我们预计会有成千上万个创业公司涌入这个市场。对于创业者而言，不用从头建一个 AI 模型，只需要直接利用这些底座模型的能力，加上他们对于场景和行业的深刻理解，就可以做出一家应用型公司
- 3 此外，某些特定的领域、行业、场景，可能会有一些创业者选择从头做，自建模型及应用，端到端，我觉得也是有机会的
- a 中国在大模型的研发上，需要给自己更多的时间，才能够慢慢地达到全球领先水平。因为围绕着 OpenAI 或者西方科技大厂的大模型已经构建出了一个初步的生态，国内科技大厂或者创业公司的大模型在技术和工程上也有很多地方需要不断追赶
- b 如果只是做出大模型，没有人去用，缺少完整生态能力，也是没有任何意义的
- 4 因此，我非常看好生成式 AI 和大模型，它的出现标志着一个巨大的 AI 开发范式的转换，将会真正的被利用到更多应用场景。但现在也处于科技的炒作周期，叠加了很多情绪和期待，一位美国 AI 研发人员说，IGNORE THE HYPE CYCLE，BUILD BORING BUSINESS，就是忽略这种短暂的炒作周期，而是去建立无聊生意
- 5 在我看来，AI 能力的真正体现，并不仅仅局限于作画、对话聊天领域。AI 很明显是一个能够改变人类发展的技术，它应该融入所有行业、所有应用。想要真的实现这一步，我觉得目前我们才刚刚开始，它远比我们今天看到的一些示范性的 TO C 应用更有价值
- 6 在未来的三五年加速发展后，AI 可能不只是停留在媒体或者人们的幻想中。十年后，它有可能真正去实现大规模落地，赋能万业

大公司“疯狂”，一级市场相对理性

- 1 短短几个月内，CHATGPT 的用户迅速破亿，且还在飞速增长。即使对投资人而言，这也是一个非常现象级的情况。种种迹象都表明，它的产品或技术肯定有独到的地方，或者值得期待的价值趋势
- 2 目前的大背景是，国外的大公司如微软、谷歌等都动作频频，甚至可以称之为“疯狂”，这也会使整体市场发展得非常迅速。不过，从国内目前的状况来看，无论是大公司还是创业公司，包括投资人，更多还处于跟随的状态
- 3 确实从这两年能感受到，AIGC 的进展飞速，尤其在海外，无论是模型端，还是应用层面，甚至包括现在发布的一些产品，在 C 端用户体验层面，都达到了一定的高度，并可以进入消费级市场
- 4 技术上的突破，可能已经到了一个相对临界点的状态。因为人工智能的想象力其实是挺高的，它确实是新的一个生长性方式，包括人机交互的方式，已经从原来

的机器语言（编程、代码）等，到现在可以理解自然语言，其实这是一个非常大的突破。从这方面来说，很多软件和硬件，包括一些互联网产品都有可能因其升级，或被替代

a 一方面，对已有的部分传统产品，人工智能可以加持一些应用。具体体现在产品上，如谷歌或微软的搜索，如果加上类 ChatGPT 技术或者应用，传统的搜索可以进行升级，提升整体使用的体验，对信息的处理和生成也可以更加优化

b 另一方面，人工智能可以产出新的产品或玩法，也就是所谓的替代。比如自动生成图片等功能，此前图片的产出需要人力手工绘图，但现在可以利用 AI 自动生成图片，从而减少人力和时间成本。此外，在 AI 的加持下，可能还有一些新的产品或新的应用面世

5 除了互联网产品这些软件外，一些硬件产品，例如家电等工具，如果配备相应的人工智能技术应用，也可能会有全新的服务和体验。由此衍生出来的未来行业方向或模式，对于相关公司来说，都存在发展机会。且随着人工智能技术的不断发展，它确实可以影响到很多行业，包括不同的一些模式。对于它的未来，我觉得确实是值得期待的

6 但体感上，我觉得现在外部关于行业讨论的声音有点太多了，其实一级市场（投资行业）对于这一赛道的看法还是相对理性的。因为对于国内来说，大部分创业公司或者投资，都还处于早期或跟随状态

在春节假期复工前，我们就非常明确 ChatGPT 会是今年计算机预期差最大的主线观点，也是全市场首篇明确此观点的报告。到了现在大概过了近一个月的时间，大家已经看到整个 ChatGPT 行情的扩展，我们的预测已经得到了验证

- 1 一周前我们不断强调 CHATGPT 投资转向算力与场景，上周二时我们也是全市场唯一明确提示浪潮信息，到今天公司已经是全市场的关注的焦点。大家最关注的是 CHATGPT 的算力和到底如何量化浪潮信息的受益程度
 - a ChatGPT 的总算力消耗约为 3,640PF-days。至少需要 7-8 个算力 500P 的数据中心（投资预计超过 200 亿）才能支撑 ChatGPT 的运行。从数据中心的整个成本构成来看，一半左右的成本需要专门投入到服务器领域，体量约在一两百亿之间
 - b 另外需要关注 ChatGPT 是持续升级的，算力消耗呈几何倍增加。GPT、GPT-2 和 GPT-3（当前开放的版本为 GPT-3.5）的参数量从 1.17 亿增加到 1,750 亿，预训练数据量从 5GB 增加到 45TB，其中 GPT-3 训练单次的成本就高达 460 万美元
- 2 所以从一方面说明了整个 CHATGPT 的算力投入，是一个几乎只有巨头或者大国的科技巨头才能做的事
- 3 第二个可以看到 CHATGPT 会成为大国科技巨头的必争之地，因为没有哪一个重要的国家和科技巨头会把 CHATGPT 这么一个战略级的入口拱手让人。最近我们也做了相对的产业研究，关于国内的各大互联网巨头，如 BAT 字节、三六零等
- 4 原来最大的问题是过去两年国内互联网的资本开支比较收缩，导致浪潮信息的估值在启动的时候也只有一个底部
 - a 现在不同，现在所有的科技巨头都看到了机会，美国的 ChatGPT 也不是传统的科技巨头最先做出来的，是 openAI 这个专门为了通用人工智能成立的实验室做的，所以核心在于原来商业化的公司没有想到往这个方向去劲烈的投入和迭代，没有想到会出现这么一个令人震惊的产品
 - b 所以认为它不是技术上的突破，而更愿意称它为是一个通用工程领域的重要拐点。因为是通用工程的产品所以资本市场的感知和长远落地的速度非常快，另一个方面就是科技巨头看到了会出现这个成果，他们反应的速度也很快。所以在我们产业研究中，一线的国内巨头公司今年大概率会加强在 AI 算力领域的投资

从标的层面，大家应该都会同意英伟达处于所有 AI 芯片整个源头，是最受益的一个领域，可惜英伟达不是在国内的上市的公司，所以我们选择投资标的有两个思路

- 1 第一个，在全球的巨头的资本扩张的情况下，我们有哪一个数据中心底层产业链是在全球占据比较高的份额的，这里面像通信的光模块最近大家已经挖掘了，这个是相当于是整个英伟达 AI 算力的一条产业链
- 2 第二个就是国内的在使用英伟达 GPU 来做 AI 服务器的，在 GPU 和 AI 芯片的下游，而且这个行业份额非常的集中
 - a 大家都知道浪潮信息，它整个的服务器在国内是当之无愧的第一，全球第二，它整个服务器的份额占据全中国大概接近一半的份额，但它 AI 的服务器占据了整个 BAT 为代表的互联网市场的绝大部分的份额，它的份额的程度是超过所有的剩余的参与者的总和的
 - b 而且在 BAT 可能份额会更高，甚至会达到八九十。所以基本上，一旦互联网巨头开始不计代价往 AI 的领域、算力领域投资，浪潮信息将是首要的受益者
- 3 过去两年，因为大家对互联网资本开支的担心，使它成为了一个低估值，最低只有 15 倍的一个公司。新的资本开启的周期，它的估值向上弹性，包括资本链业绩

最先出现的速度，可能都会是 CHATGPT 产业里面非常快的。这也是为什么我们从上周市场还没有声音的时候，周二开始非常旗帜鲜明的开始强推浪潮信息

4 另外一个大家比较关注的几个方向

- a 第一个是大家希望量化一下对它的弹性，ChatGPT 一次版本升级从数据训练量就上了 3 个数量级。从 GB 到 TB 的一个规模，后面可能还会迭代到 GPT4，这是一个指数级增大的算力需求
 - i 以目前的版本去算，一个互联网巨头或科技巨头想要从 0 开始支撑区别的应用，国内有能力往领域去做的估计不少于 6 家
 - ii BAT 字节，京东、360，包括讯飞等等这个里面。浪潮在各类服务器的份额作为龙头地位，将充分受益此类催化。对于估值到今天还只有 20 倍左右的浪潮信息来说，整个估值体系包括业绩弹性都是相当大的
- b 第二个市场比较关注的一点是关于上游的 AI 芯片的美国对我们的制裁或者限制。这里我们注意三点
 - i 第一点，AIGPU 的芯片，它是一个通用的芯片，包括 A100 以上，现在被限制住，但实际上越通用的芯片是越难被限制的
 - ii 第二个就是 AI 芯片的上游。现在绝不是 GPU 一家独大，它是一个多分类的。以浪潮的 AI 服务器为例，它根据客户的需求，包括了 GPU、FPGA，包括寒武帝这 Asic 专用的 AI 的芯片，可选择的范围会非常的多。实际上就规避了单一的芯片，尤其单一芯片里面高算率的供应链的一些影响
 - iii 第三个，现在最大的算力的消耗 AI 的计算跟以往不同，它包括训练和推理。它最大的算力消耗是 AI 训练芯片。AI 训练芯片它本质上不是一个耗损品，不像手机芯片一样卖多少，每年消耗这么多。它本质上是一个后台的工具，它只有在训练的峰值的时候才会消耗到这么大的算力。所以就像你的办公电脑一样，你其实提前存很多并不会消耗，只等你峰值的时候来用它
- c 所以依据这三个方面，其实大家对上游芯片的 AI 芯片的影响或者瓶颈的制裁不用太过去担心，而且这个时候往往体现了浪潮的渠道的地位，它已经是英特尔，包括英伟达在国内最大的分销商之一了，所以浪潮的渠道地位反而更加的显著

5 另外一个，关于 AI 芯片，AI 服务器，还有一点就是 AI 服务器，如果大家关注浪潮的业绩拆分，包括浪潮的 AI 服务器的毛利率基本上是传统的互联网服务器的毛利率的两倍

- a 为什么会这么高？是因为 AI 服务器它本身并不像传统服务器那样，已经是一个非常标准化的 CPU，都是一样的操芯。AI 估计刚刚说了
 - i 第一个，游芯片不同
 - ii 第二个，它包括了推算、训练和推理不同的需求，底层实际上是没有统一的操作系统的。它的训练的框架有的客户可以自己，像一些互联网巨头，像浪潮，它自己会提供底层的 AI 训练和对你的整个底层的类似于超系统的整个深度学习的框架，来提供附加提问
- b 而且浪潮在技术上还做了参数量已经超过了 GPT3，在 21 年就已经推出了叫做源这种大模型系统，所以它本质上也是一个大模型，包含算力大模型的给各个行业。虽然一些科技巨头会自己做，但是它是给各个行业赋能的一个潜在的供应商，本身算力又受到很明确的提升的估值也属于低位的，同时又是潜在的 AI 的大模型的各个行业赋能者
- c 当前 20 倍的估值，虽然最近一周推荐之后涨了很多，但是我自己判断才刚刚开始，还有更大的机会。这是我们今天关于浪潮的判断

d 而按照以往的惯例，互联网的资本开支通常在财报的开始，也就是三四月份会定全年的大概规划，而在近期互联网的动向中，都是要加大在 AI 上面投入，所以应该在二季度就会看到一些业绩的拐点，或者一些标志性的大订单开始出现，所以像浪潮这种基本上是最大的收益面最先会在 ChatGPT 里面，算力层面确实是业绩最容易出现拐点的公司

另外我们有几个应用层面

- 1 第一个就是金山办公，其实就是最直接的，像 OFFICE，它的估值的天花板，但市场还没意识到彻底的会打开。原来靠信创大家会觉得固执，天花比较有限。但是现在我从我们团队对线上办公的了解，一定会开放去引入 CHATGPT 的能力所以它的弹性也是相当大
- 2 第二个，梦网科技，我们最近推荐的 5G 消息是整个 CHATGPT 最佳的落地的窗口。5G 消息在两年前刚推出的时候，大家看运营商的白皮书 CHATBOT，它的核心功能并不是去蹭刚出来的 CHATGPT，一旦引流之后，原来给每个企业发送的短信，实际就变成一个发送的智能的、帮业务客户综合一体的数据。所以它从每个短信 1 毛钱左右，可以完全升级成 SAAS 服务，按广告点击业务分成各种商业模式
- 3 从巨头的角度，跟这种 5G 消息的龙头公司去合作，会一下子获得上百万的客户，同时每一个客户它的垂直场景对话的数据同时拥有。所以对技术来说，它也是第一优先的选择。所以像梦网科技这种去年整个的历史的最低点，今年开始整个业绩反转，基本面急剧向上，再加上 CHATGPT 的历史性窗口，我们认为这公司后面可能会弹力相当的大
- 4 第三个方向，我们觉得在一些金融的，代替初级分析师的，像同花顺等，可能会有非常大的潜意做到整个的，这是我们目前在场景里建议可以重点关注的方向
- 5 最后再强调一下我们的观点，CHATGPT 一定是今年计算机最大的主线，我们是在春节回来前唯一的一个明确的说了今年主线的，在整个市场还没有像现在大家看到的这么全面扩散，引领整个市场的时候，我们依然认为大家可以重点去关注这个方向，我们也会提供更多更好的标的

风险提示：技术落地不及预期，竞争格局加剧

大模型训练热潮下，算力基础设施有望迎来产业机遇

- 1 CHATGPT 发布之后，引发了全球范围的关注和讨论，国内各大厂商相继宣布 GPT 模型开发计划。我们认为，以 GPT 模型为代表的 AI 大模型训练，需要消耗大量算力资源，主要需求场景来自：预训练+日常运营+FINETUNE
- 2 以预训练为例，据我们测算，进行一次 CHATGPT 的模型预训练需要消耗约 27.5PFLOP/S-DAY 算力
- 3 基于此，我们认为，随着国产大模型开发陆续进入预训练阶段，算力需求持续释放或将带动算力基础设施产业迎来增长新周期
- 4 产业链相关公司包括
 - a 算力芯片厂商：景嘉微、寒武纪、海光信息、龙芯中科、中国长城等
 - b 服务器厂商：浪潮信息、中科曙光等
 - c IDC 服务商：宝信软件等

ChatGPT：大模型训练带来高算力需求

- 1 训练 CHATGPT 需要使用大量算力资源。据微软官网，微软 AZURE 为 OPENAI 开发的超级计算机是一个单一系统，具有超过 28.5 万个 CPU 核心、1 万个 GPU 和 400GB/S 的 GPU 服务器网络传输带宽。据英伟达，使用单个 TESLA 架构的 V100GPU 对 1,746 亿参数的 GPT-3 模型进行一次训练，需要用 288 年时间
- 2 此外，算力资源的大量消耗，必然伴随着算力成本的上升，据 LAMBDA，使用训练一次 1,746 亿参数的 GPT-3 模型所需花费的算力成本超过 460 万美元。我们认为，未来拥有更丰富算力资源的模型开发者，或将能够训练出更优秀的 AI 模型，算力霸权时代或将开启

需求场景：预训练+日常运营+Finetune

- 1 具体来看，AI 大模型对于算力资源的需求主要体现在以下三类场景
 - a 模型预训练：ChatGPT 采用预训练语言模型，核心思想是在利用标注数据之前，先利用无标注的数据训练模型。据我们测算，训练一次 ChatGPT 模型（13 亿参数）需要的算力约 27.5PFlop/s-day
 - b 日常运营：用户交互带来的数据处理需求同样也是一笔不小的算力开支，我们测算得 ChatGPT 单月运营需要算力约 4,874.4PFlop/s-day，对应成本约 616 万美元
 - c Finetune：ChatGPT 模型需要不断进行 Finetune 模型调优，对模型进行大规模或小规模的迭代训练，预计每月模型调优带来的算力需求约 82.5~137.5PFlop/s-day
- 2 算力芯片+服务器+数据中心，核心环节有望率先受益我们认为，随着国内厂商相继布局 CHATGPT 类似模型，算力需求或将持续释放，供给端核心环节或将率先受益
 - a 算力芯片：GPU 采用了数量众多的计算单元和超长的流水线，架构更适合进行大吞吐量的 AI 并行计算
 - b 服务器：ChatGPT 模型训练涉及大量向量及张量运算，AI 服务器具备运算效率优势，大模型训练有望带动 AI 服务器采购需求放量
 - c 数据中心：IDC 算力服务是承接 AI 计算需求的直接形式，随着百度、京东等互联网厂商相继布局 ChatGPT 类似产品，核心城市 IDC 算力缺口或将加大。关注 AI 景气周期下，算力基础设施产业机遇我们认为，国产厂商未来或将训练出自己的 GPT 模型，带动算力设施产业迎来景气周期
- 3 相关公司包括

- a 算力芯片厂商：景嘉微、寒武纪、海光信息、龙芯中科、中国长城等
- b 服务器厂商：浪潮信息、中科曙光等
- c IDC 服务商：宝信软件等

4 风险提示：宏观经济波动；下游需求不及预期

ChatGPT 有望带动算力需求需求端

- 1 据 OPENAI 测算，自 2012 年以来，全球头部 AI 模型训练算力需求 3-4 个月翻一番，每年头部训练模型所需算力增长幅度高达 10 倍。而摩尔定律认为，芯片计算性能大约每 18-24 个月翻一番
 - a 因此，AI 训练模型算力需求增长与芯片计算性能增长之间的不匹配，或将带来对算力基础设施供给需求的快速增长
 - b 我们认为，考虑到算力对于 AI 模型训练效果的关键性作用，拥有更丰富算力资源的模型开发者，或将能够训练出更优秀的 AI 模型，算力霸权时代或将开启
- 2 具体来看，AI 大模型对于算力资源的需求主要体现在以下三类场景
 - a 模型预训练带来的算力需求模型预训练过程是消耗算力的最主要场景。

ChatGPT 采用预训练语言模型，核心思想是在利用标注数据之前，先利用无标注的数据，即纯文本数据训练模型，从而使模型能够学到一些潜在的跟标注无关的知识，最终在具体的任务上，预训练模型就可以利用大量的无标注数据知识

 - i 在 Transformer 的模型架构下，语言预训练过程可以根据上下文一次处理所有输入，实现大规模并行计算。通过堆叠多个解码模块，模型的层数规模也会随着提升，可承载的参数量同步增长。与之相对应的，模型训练所需要消耗的算力也就越大
 - ii 我们预计，训练一次 ChatGPT 模型需要的算力约 27.5PFlop/s-day。据 OpenAI 团队发表于 2020 年的论文《Language Models are Few-Shot Learners》，训练一次 13 亿参数的 GPT-3XL 模型需要的全部算力约为 27.5PFlop/s-day，训练一次 1,746 亿参数的 GPT-3 模型需要的算力约为 3,640PFlop/s-day
 - iii 考虑到 ChatGPT 训练所用的模型是基于 13 亿参数的 GPT-3.5 模型微调而来，参数量与 GPT-3XL 模型接近，因此我们预计训练所需算力约 27.5PFlop/s-day，即以 1 万亿次每秒的速度进行计算，需要耗时 27.5 天
 - b 此外，预训练过程还存在几个可能的算力需求点
 - i 模型开发过程很难一次取得成功，整个开发阶段可能需要进行多次预训练过程
 - ii 随着国内外厂商相继入局研发类似模型，参与者数量增加同样带来训练算力需求
 - iii 从基础大模型向特定场景迁移的过程，如基于 ChatGPT 构建医疗 AI 大模型，需要使用特定领域数据进行模型二次训练
 - c 日常运营带来的算力需求 ChatGPT 单月运营需要算力约 4,874.4PFlop/s-day，对应成本约 616 万美元。在完成模型预训练之后，ChatGPT 对于底层算力的需求并未结束，日常运营过程中，用户交互带来的数据处理需求同样也是一笔不小的算力开支
 - i 据 SimilarWeb 数据，2023 年 1 月 ChatGPT 官网总访问量为 6.16 亿次。据 Fortune 杂志，每次用户与 ChatGPT 互动，产生的算力云服务成本约 0.01 美元。基于此，我们测算得 2023 年 1 月 OpenAI 为 ChatGPT 支付的运营算

力成本约 616 万美元

- ii 据上文，我们已知训练一次 1,746 亿参数的 GPT-3 模型需要 3,640PFlop/s-day 的算力及 460 万美元的成本，假设单位算力成本固定，测算得 ChatGPT 单月运营所需算力约 4,874.4PFlop/s-day

Finetune 带来的算力需求

- 1 模型调优带来迭代算力需求。从模型迭代的角度来看，CHATGPT 模型并不是静态的，而是需要不断进行 FINETUNE 模型调优，以确保模型处于最佳应用状态
 - a 这一过程中，一方面是需要开发者对模型参数进行调整，确保输出内容不是有害和失真的
 - b 另一方面，需要基于用户反馈和 PPO 策略，对模型进行大规模或小规模的迭代训练
- 2 因此，模型调优同样会为 OPENAI 带来算力成本，具体算力需求和成本金额取决于模型的迭代速度

供给端：核心环节有望率先受益

- 1 算力芯片：AI 算力基石，需求有望大规模扩张 GPU 架构更适合进行大规模 AI 并行计算，需求有望大规模扩张。从 CHATGPT 模型计算方式来看，主要特征是采用了并行计算
 - a 对比上一代深度学习模型 RNN 来看，Transformer 架构下，AI 模型可以为输入序列中的任何字符提供上下文，因此可以一次处理所有输入，而不是一次只处理一个词，从而使得更大规模的参数计算成为可能
 - b 而从 GPU 的计算方式来看，由于 GPU 采用了数量众多的计算单元和超长的流水线，因此其架构设计较 CPU 而言，更适合进行大吞吐量的 AI 并行计算。基于此，我们认为，随着大模型训练需求逐步增长，下游厂商对于 GPU 先进算力及芯片数量的需求均有望提升
- 2 单一英伟达 V100 芯片进行一次 CHATGPT 模型训练，大约需要 220 天
 - a 我们以 AI 训练的常用的 GPU 产品—NVIDIA V100 为例。V100 在设计之初，就定位于服务数据中心超大规模服务器。据英伟达官网，V100 拥有 640 个 Tensor 内核，对比基于单路英特尔金牌 6,240 的 CPU 服务器可以实现 24 倍的性能提升
 - b 考虑到不同版本的 V100 芯片在深度学习场景下计算性能存在差异，因此我们折中选择 NVLink 版本 V100（深度学习算力 125TFlops）来计算大模型训练需求
 - c 据前文，我们已知训练一次 ChatGPT 模型（13 亿参数）需要的算力约 27.5PFlop/s-day，计算得若由单个 V100GPU 进行计算，需 220 天；若将计算需求平均分摊至 1 万片 GPU，一次训练所用时长则缩短至约 32 分钟
 - d 全球/中国 GPU 市场规模有望保持快速增长。据 VMR 数据，2021 年全球 GPU 行业市场规模为 334.7 亿美元，预计 2030 年将达到 4,773.7 亿美元，预计 22-30 年 CAGR 将达 34.4%。2020 年中国 GPU 市场规模 47.39 亿美元，预计 2027 年市场规模将达 345.57 亿美元，预计 21-27 年 CAGR 为 32.8%
- 3 服务器：AI 服务器有望持续放量 CHATGPT 主要进行矩阵向量计算，AI 服务器处理效率更高。从 CHATGPT 模型结构来看，基于 TRANSFORMER 架构，CHATGPT 模型采用注意力机制进行文本单词权重赋值，并向前馈神经网络输出数值结果，这一过程需要进行大量向量及张量运算
 - a 而 AI 服务器中往往集成多个 AIGPU，AIGPU 通常支持多重矩阵运算，例如卷积、池化和激活函数，以加速深度学习算法的运算。因此在人工智能场景下，AI 服务器往往较 GPU 服务器计算效率更高，具备一定应用优势

风险提示：宏观经济波动。若宏观经济波动，产业变革及新技术的落地节奏或将受到影响，宏观经济波动还可能对 IT 投资产生负面影响，从而导致整体行业增长不及预期。下游需求不及预期。若下游数字化需求不及预期，相关的数字化投入增长或慢于预期，致使行业增长不及预期

核心观点

- 1 CHATGPT, 人工智能的旷世巨作。CHATGPT 是一种聊天机器人软件, OPENAI 于 2022 年 11 月推出的聊天机器人, 具备人类语言交互外复杂的语言工作, 包括自动文本生成、自动问答、自动摘要等多重功能, 应用场景广阔, 相较于上个版本更像人类一样聊天交流
- 2 OPENAI 除了 CHATGPT 还包括 DALL·E2、WHISPER 等项目分别是自动绘图、自然语言翻译等软件。OPENAI 的商业模式即 API 接口收费, 可根据不同项目需求进行收费, 我们认为其商业模式属于底层模型开放性标准化 SAAS 服务模式。我国仍处于初期阶段, 以辅助生成内容服务为主, 我们认为未来有望形成相关 SAAS 模式
 - a ChatGPT 促使 AIGC 快速商业化发展。GPT 系列是 AIGC 的一种商业化方向, 目前 AIGC 已经实现商业化的方向有 AI 写作、AI 作图、AI 底层建模, 未来 AI 生成视频和动画领域有望快速商业化发展
 - b AIGC 也被认为是继 UGC、PGC/UGC 之后的新型内容生产方式, 有望解决 PGC/UGC 创作质量参差不齐或是降低其有害性内容传播等问题, 有望在实现创意激发, 提升内容多样性的同时降本增效, 并大规模使用。目前我国已经有商业化例如百度 AIGC 数字人主播度晓晓、百家号 TTV 等
- 3 AIGC 蓬勃发展, 相关厂商有望受益。我们认为 AIGC 的出世会产生革命性的影响, 同时有望赋能千行百业。我们认为受益厂商分为三类, 分别是
 - a AI 处理器厂商, 具备自研 AI 处理器的厂商可以为 AIGC 的神经网络提供算力支撑, AI 处理器芯片可以支持深度神经网络的学习和加速计算, 相比于 GPU 和 CPU 拥有成倍的性能提升和极低的耗电水平
 - b AI 商业算法商业落地的厂商, 原因是 AI 算法的龙头厂商在自然语言处理、机器视觉、数据标注方面都具有先发优势和技术领先性
 - c AIGC 相关技术储备的应用厂商, 相关厂商有望在降本增效的同时实现创意激发、提升内容多样性并打开海量市场
- 4 投资建议:关注三条投资主线
 - a 具备自研 AI 处理器的厂商, 受益标的为寒武纪、商汤、海光信息
 - b 具备 AI 算法商业落地的厂商, 重点推荐科大讯飞, 其他受益标的为:汉王科技、拓尔思、海天瑞声、虹软科技、云从科技、格灵深瞳
 - c AIGC 相关技术储备的应用厂商, 受益标的为万兴科技、中文在线、阅文集团、昆仑万维、视觉中国
- 5 风险提示:核心技术水平升级不及预期的风险、AI 伦理风险、政策推进不及预期的风险、中美贸易摩擦升级的风险

AI 纪元, ChatGPT

- 1 CHATGPT 是一种聊天机器人软件:全名为 CHATGENERATIVEPRE-TRAINEDTRANSFORMER, 是 OPENAI 于 2022 年 11 月推出的聊天机器人, 该软件使用方便快捷, 只需向 CHATGPT 提出需求, 即可实现文章创作、代码创作、回答问题等功能
 - a ChatGPT 功能强大:ChatGPT 目前仍以文字方式互动, 而除了通过人类语言交互外, 还可以用于相对复杂的语言工作, 包括自动文本生成、自动问答、自动摘要等在内的多种任务
 - b ChatGPT 关注度较高:截至于 2022 年 12 月 4 日, 上线仅 5 天, OpenAI 估计 ChatGPT 已经拥有超过 100 万用户。此外, 根据《纽约时代》报道, GPT 的下一个版本 GPT-4 有望于 2023 年某个时候推出

- c ChatGPT 打开海量应用场景:ChatGPT 应用场景广泛,拥有潜在空前蓝海,其功能覆盖各个板块,我们将其分成生成应用和布局、搜索和数据分析、程序生成和分析、文本生成、内容创作、一般推理和其他七部分
- 2 我们认为,CHATGPT 经历了人工智能发展的三大趋势
 - a 机器学习:2012 年左右,随着基础算力的提升,全球已经开启人工智能热潮即大数据时代
 - b 神经网络:2015 年左右开始繁荣爆发,神经网络是实现 AI 深度学习的一种重要算法
 - c Transformer 算法:2017 年左右,是图型神经网络的一种,Transformer 开始广泛应用于机器翻译之中,逐步取代循环神经网络(RNN),不再使用递归,而是采用了注意力机制进行优化
- 3 从算法模式的版本上,CHATGPT 已经经历了四个版本,未来有望生成新版本即 GPT-4
 - a GPT-3 是 GPT 系列第三代语言预测模型,是 OpenAI 创建的 GPT-2 的继任者:GPT-3 的应用包括帮助 Viable 公司快速理解客户反馈、虚拟人陪伴、论文搜索等
 - b 我们认为,InstuctGPT 可以理解成是 GPT-3 的微调版本:InstuctGPT 模型比 GPT-3 更擅长遵循指令,且回答更真实,且有害情绪输出大幅下降例如暴力情节
 - c 我们认为,ChatGPT 可以理解成是 GPT-3.5 的微调版本:相较于 InstructGPT,ChatGPT 效果更加真实,模型的无害性实现些许提升,编码能力更强
- 4 GPT-4 有望打开海量应用:据 OPENAI 创始人 ALTMAN 消息,参数预计更大,计算模型优化有望实现更优化,且 GPT-4 将是纯文本模型。我们认为其"拟人化"能力更强
 - a ChatGPT 与 InstructGPT 的相同点:使用来自人类反馈的强化学习(RLHF)来训练该模型
 - b ChatGPT 与 InstructGPT 的不同点:可以理解成在“人脑思维”的基础上加入了“人类反馈系统”,是一种奖励模型
 - c ChatGPT 仍有提升空间,仍有一些限制需要突破:可能写出看似合理但不正确或荒谬的答案,对输入措辞的调整或多次尝试相同的提示很敏感,该模型通常过于冗长并过度使用某些短语
- 5 全球 AI 的“领军企业”OPENAI: OPENAI 是美国一个人工智能研究实验室,是促进和发展友好的人工智能,使人类整体受益
 - a OpenAI 不仅仅是 ChatGPT:OpenAI 的业务不仅仅局限于 ChatGPT 领域,还包括 Dall·E2、Whisper 等项目
 - b OpenAI 的商业模式即 API 接口收费:客户可以通过 OpenAI 的强大 AI 模型构建应用程序,公司按照不同项目的 AI 模式和不同需求进行收费,对于 AI 图像系统按不同的图片分辨率定价
 - c 对于 AI 语言文字系统按字符单价收费,对于调整模型和嵌入模型按照文字单价收费,并根据不同的调用模型区别定价
 - d 我们认为 OpenAI 的商业模式属于底层模型开放性标准化 SAAS 服务模式:公用户通过其 API 接口接入,并支付平台费用获取相关图像、语言、代码调整服务,OpenAI 可获得付费订阅式的 SAAS 费用,我们认为该商业模式用户粘性极强
 - e 目前我国 AIGC 商业仍处于初期阶段,部分文本、图像等生成的公司仍处于初期阶段,以辅助生成内容服务为主

AIGC，星星之火可以燎原

- 1 AIGC 的含义:AIGC(AI GENERATED CONTENT)即人工智能自动生成内容可以在创意、表现力、迭代、传播、个性化等方面，充分发挥技术优势
- 2 AIGC 已经实现的商业化方向:AI 写作:例如 AI 生成文字、写邮件、广告营销等; AI 作图:AI 自动生成图片; AI 底层建模:AIGC 生成底层技术开发
- 3 AIGC 未来商业化发展方向:AI 生成视频和动画，目前该领域仍存在连贯性和逻辑性的问题亟待解决
- 4 AIGC 也被认为是继 UGC、PGC/UGC 之后的新型内容生产方式:我们认为 AIGC 在创作成本上具有颠覆性，而且具备降本增效的多重优势，并且有望解决目前 PGC/UGC 创作质量参差不齐或是降低其有害性内容传播等问题，同时创意激发，提升内容多样性
 - a AIGC 有望赋能“千行百业”，我国已有应用案例:百度 AIGC 技术的数字人主播度晓晓正式“上岗”，百家号 TTV 技术验证了 AIGC 的发展潜力
 - b 来自人民网、中国青年网等多家媒体通过百家号 TTV 技术进行内容生产
- 5 我们认为 AIGC 的受益厂商分为三类。AI 处理器厂商; AI 商业算法商业落地的厂商; AIGC 相关技术储备的应用厂商
 - a 具备自研 AI 处理器的厂商可以为 AIGC 的神经网络提供算力支撑:算力作为数据加速处理的动力源泉，其重要性不言而喻。AI 处理器芯片可以支持深度神经网络的学习和加速计算，相比于 GPU 和 CPU 拥有成倍的性能提升，和极低的耗电水平
 - b 具备 AI 商业算法落地的厂商具备相关算法的领先性:AI 算法的龙头厂商在自然语言处理、机器视觉、数据标注方面都具有先发优势。算法上，数据标注属于 AIGC 算法的生成关键步骤，而在自然语言处理、机器视觉等方面，AIGC 已经对此方向应用产生深远影响
 - c AIGC 相关技术储备的应用厂商有望打开海量市场:相关娱乐、传媒、新闻、游戏、搜索引擎等厂商具备海量文本创作、图片生成、视频生成等需求，随着 AIGC 的逐渐成熟，相关 AI 算法不断成熟完善，并结合相关应用，我们认为相关厂商在降本增效的同时，实现创意激发，提升内容多样性，从而打开海量空间

投资建议

- 1 我们认为 AIGC 的出世会产生革命性的影响，同时有望赋能千行百业。我们梳理了三条路径图，积极的推荐以下三条投资主线
- 2 具备自研 AI 处理器的厂商，受益标的为寒武纪、商汤、海光信息
- 3 具备 AI 算法商业落地的厂商，重点推荐科大讯飞，其他收益标的为:汉王科技、拓尔思、海天瑞声、虹软科技、云从科技、格灵深瞳
- 4 AIGC 相关技术储备的应用厂商，受益标的为万兴科技、中文在线、阅文集团、昆仑万维、视觉中国

风险提示：核心技术水平升级不及预期的风险、AI 伦理风险、政策推进不及预期的风险、中美贸易摩擦升级的风险

ChatGPT 问世，AI 赛道关注度有望持续提升

- 1 2022 年 11 月 30 日，OPENAI 公司正式对外发布聊天机器人软件 CHATGPT，并迅速在全球范围获得大量用户青睐。目前 CHATGPT 主要通过提供 API、订阅服务等获得收入，随着盈利模式拓展，未来商业落地有望加速
- 2 我们认为，CHATGPT 受到广泛关注的原因在于能够实现接近人类水平的自然语言处理能力，对话逻辑也已经与人类类似。而 CHATGPT 成功的背后，来自独特算法模型的加持，验证了单一大模型的实战意义，同时也是 AI 大模型路线一次里程碑式的胜利
- 3 在大模型基础上，GPT 得以从 1.0 逐步演化至 2.0、3.0、3.5 等版本。而随着用户数量扩张，模型有望加速迭代，产品力持续升级或将带动 AI 赛道关注度有望持续提升

背景：发源于 OpenAI，成名于生成式语言

- 1 OPENAI 成立于 2015 年，专注于强化学习等 AI 算法、模型的研发与应用。2019 年公司获得微软注资 10 亿美元，就 AZURE 业务开发人工智能技术。2023 年，微软再次追加百亿投资，深入布局生成式 AI 技术
- 2 随着公司 AI 技术不断沉淀，产品模型持续演进，GPT 模型陆续演化出 GPT-2、GPT-3、GPT-3.5 等版本。基于微调后的 GPT-3.5 模型，OPENAI 推出 CHATGPT 软件。目前来看，OPENAI 围绕 CHATGPT 的盈利模式大致分为两种
 - a 提供 API 数据接口，执行自然语言处理任务，按字符数收费
 - b ChatGPT Plus 提供服务器忙碌时段优先响应等特权，并收取 20 美元/人/月的订阅服务费

原理：AI 大模型里程碑式的胜利

- 1 我们看到，CHATGPT 成功的背后，离不开独特训练模型的加持，核心技术采用了监督学习+奖励模型共同进行语言模型训练。拆解来看，训练过程大致分为三个阶段
 - a 训练监督策略模型
 - b 训练奖励模型
 - c 采用近端策略优化进行强化学习
- 2 此外，我们发现，CHATGPT 的训练过程存在几个特点
 - a 采用的是单一大模型
 - b 采用的是小样本学习方法
 - c 采用人类反馈来微调监督学习
 - d 基于此，我们认为，ChatGPT 的成功，验证了单一大模型实战意义，有望推动人工智能训练向大模型方向演进

模型：单一大模型加持下 GPT 快速迭代

- 1 我们认为，GPT 模型强大的语言处理能力，来自底层模型的持续演进。据 OPENAI 官网，从 GPT-1 到 GPT-2、GPT-3，GPT 模型参数量不断提升，从 1.17 亿大幅提升至 15 亿、1750 亿。而单一大模型的形态是 GPT 模型得以快速演进的基础，通过大规模语言预训练的方式，模型迅速积累泛化通用的语言能力
- 2 此外，GPT3.5 开始还引入监督学习过程，人类反馈机制的加入在提升模型准确性、真实性的同时，又降低了参数计算量，使得更大规模的 AI 学习成为可能。我们认为，随着 CHATGPT 用户不断扩张，用户反馈或将带来类似监督学习的效果，GPT 有望在大模型加持下加速进化

投资建议：模型层，应用层，算力层

- 1 我们认为，在中国或将训练出国产的 GPT 模型，而 GPT 模型算法的开发和训练，具有一定的研发门槛。目前国内在模型层面具备产业基础的相关企业包括：百度、商汤、云从科技等
- 2 在应用层持续推出新应用的相关企业包括：科大讯飞、金山办公、同花顺、汉王科技、拓尔思等；在底层基础设施和工具有布局的相关企业包括：寒武纪、景嘉微、海天瑞声等

风险提示：市场需求变动风险，市场竞争加剧风险

概况

- 1 2011 年起，汤姆猫是最早跟玩家、用户有交互功能的虚拟 IP 之一，运营了 11 年的老产品《会说话的狗狗本》去年重新回到了加拿大等地的排行榜第一，超过了 TIKTOK
- 2 海外团队不断地丰富 IP 角色，同时持续强化。我们除了汤姆猫、狗狗本，还有安吉拉等等 6 个角色，围绕这 6 个角色，打造的 20 多款游戏，累计下载已经超过了 200 亿次。截至去年的 MAU 超过 4 亿
- 3 经过了 10 多年在休闲游戏里面的深耕，我们也形成了 2 个具有优势的品类，一个是虚拟宠物养成；一个是跑酷类。根据 APP ANNIE，虚拟宠物养成，公司的市占率达到 86%；跑酷，我们市占率是 15%，也是全球前三
- 4 除了游戏，汤姆猫围绕系列 IP 角色，创造了非常多的动漫、影视作品。动漫累计播放量超过 1000 亿次。汤姆猫也规划了大电影内容，完成了剧本二次创作，会和华纳、索尼等好莱坞公司洽谈。公司陆续布局了线下业态。线下主题乐园、线下衍生品
- 5 公司紧跟 AI 技术，让汤姆猫与用户有更真实的互动。2021 年我们在抖音开启了虚拟直播，去年在海外上线了智能语音机器人，实现日常生活下的语音交流和互动。游戏同时也在拓展渠道，2021 年以来先后拓展了华为电视机，《汤姆猫跑酷》也是华为首款体感类游戏
- 6 去年公司和科大讯飞，也展开了战略合作，目的是实现人工智能交互、语音识别等。我们也注意到 CHATGPT 的技术，我们国内的团队也在和海外的团队合作，已经接入了 CHATGPT 模型，打造更新的产品

我们是帮助 ChatGPT 落地非常好的场景和应用，我们现在和科大讯飞、ChatGPT 的合作，是一个什么具体的合作方式？我们调用 ChatGPT 的成本是否有大概能量化的情况？未来通过什么商业模式去变现？未来产品推向市场的节奏？

- 1 我们一直在寻找新的领域和我们的产品能合作落地的东西。我们一直关注前沿的动态，也在关注怎么让我们的 IP 从会说话，变成会聊天的转变。我们原来和谷歌有很好的关系，像 OPENAI 推出来后，我们的海外子公司在对接。我们国内的同事已经做出来一款比较前期的应用，接的就是 CHATGPT
- 2 成本来说，我们现在没有太多计较这块，还是对产品的探索，看看能不能落地、跟我们的产品相结合和融合。目前尝试下来是可行的。我们去年 2 月份也推出了智能机器人，在汤姆猫跑酷里面，可以互动交流，给与鼓励或者指导等。我们尝试下来以后，在某一点上，在语音和 AI 方面并不是我们擅长的
- 3 所以我们先和科大讯飞合作，去年 4 季度，我们尝试 CHATGPT 文字端的转换的应用。做了一些尝试后，我们在今年 1 月份之后，我们开始尝试设计和创造新的语音聊天的产品。我们是开放的，目前还没法确定到时候上线的确切的模式
- 4 我们目前国内团队和海外团队同时在研究怎么做商业化，今天国内的团队还在跟我探讨。我们尝试免费互动一定次数，用完后有体力值，需要等待体力值恢复，或者购买体力等。我们也在研究怎么设置付费点，付费的金额肯定要能覆盖 CHATGPT 的成本，初步的框架就是刚刚说的

关于主业的情况，2022 年受疫情和宏观经济，广告面临了压力；公司对于今年主业的展望？

- 1 2022 年大体的情况，国内所有的企业都会受到一些影响，我们海外从去年数据来看，美国和欧洲的市场给我们带来的增长还是比较明显的，美国有 18% 的增长。俄罗斯、乌克兰本来也是我们体量比较大的市场，但是去年的事情让我们受影响也比较大。比如俄罗斯，业务数据其实一直在增长，但是收入数据受到了很大的

影响，因为很多企业也缩减了投放广告。我们今年也会加大地区的推广和布局，我们通过两种方式，一个是自己去推，另一个是跟当地企业合作，去扶持他们

- 2 国内来看，版号的发放有所加快，我们也拿到了版号。今年产品节奏来看，汤姆猫本身也有自己的新产品，全球的发行节奏会一直保持。国内还是会有一些不确定性，是否能慢慢恢复到以往的状态，进口版号能否放开，会影响我们业务的节奏
- 3 非汤姆猫 IP，我们现在也在拿全球比较好的 IP，比如索尼克。同时我们自行研发非汤姆猫的 IP
- 4 另外我们线下的布局，线下乐园店，可能覆盖周边 3-5KM 的地区，是家长比较能接受，我们看到我们的数据比较好的，客流量基本处于饱和的状态。我们也探索到了新的模式，在江苏落地的乐园，是和政府的合作的。随着疫情完全放开，各地政策的导向都会出来。长沙等地方都在找我们公司落地线下乐园的项目
- 5 随着市场的回暖，未来 3 年建设 200 家，大部分是加盟的形式，一部分是自营的模式来推广。随着我们衍生品不断地拓展，我们通过品类筛选，去开线下衍生品的专卖店。植入衍生品的售卖
- 6 整体来看，我们觉得今年的话，没有给业务团队特别明确的业务指标。O7（汤姆猫）这边其实非常稳定，随着新业务的拓展，必定有成本的产生，我们做的是长周期的规划

我们现在和 ChatGPT 有合作吗，合作的版本是 3.5 还是 4.0？国内除了我们还有哪家也有合作？

- 1 目前我们接的是最新的版本，应该是 3.5 的版本
- 2 国内我们没有太去了解同行还有哪些接入的，应该比较少，至少上市公司层面，合作的不多

如果添加聊天对话功能，是否会涉及到游戏版号的更新？

- 1 如果我们想通了，或者有很好的商业模式，我们会直接在现有版本里面做更新。同时也会去打造新的产品
- 2 这里面要做区分，如果只是纯粹的应用，语音交互的应用，从目前来看，我们的判断可能是不需要申请版号，当然目前还要求证。如果只是纯粹的语音互动，应该不会界定为游戏产品

产品推出的节奏，国内和海外的节奏会是怎样？

- 1 目前我们国内和海外团队在同时推进这块业务。目前没有给出明确的要求什么时候推出，其实现在推出对我们问题不是很大，主要还是对商业化这块要做更多的探讨
- 2 O7 这个团队是相对比较严谨的团队，汤姆猫这块的产品开发节奏，其实去年年底已经规划到 24 年年底的开发节奏。国内来说相对会更快的推出，海外 O7 那边会考虑全球各地区的长期运营等问题，时间上可能会相对更久

会说话的汤姆猫的 MAU 和 DAU？

- 1 20 多个产品，我们去年平均 MAU 是到 4 亿多
- 2 我们的头部产品，我的安吉拉、我的汤姆猫 2 都是在 5KW 以上，汤姆猫跑酷在国内非常火，如果在春节能到 7-8KW

摘要

- 1 微软发布四款基于 CHATGPT 的融合应用，GPT 系列商用有望提速。公司宣布高级 TEAMS 产品“智能回顾”功能、VIVA-SALES 电邮生成功能、BING 搜索扩展式聊天框、EDGE 浏览器聊天和编写功能，在宣布扩大与 OPENAI 合作后不足一月的时间内实现初步产品落地
- 2 我们认为，这将有效提升用户在会议、电邮、信息搜索和信息提取等领域的使用体验和效率，公司将引领 CHATGPT 产品应用，大规模商业化落地有望提速
- 3 BING 搜索&EDGE 浏览器：基于 CHATGPT 提供智能化高效搜索，革新搜索引擎体验
 - a 新 Bing 主页增加扩展式的聊天框，可以以类似 ChatGPT 的方式回答具有大量上下文的问题
 - b Edge 浏览器新增“聊天”和“编写”功能，“聊天”功能能够总结其正在浏览的网页或文档，并以对话形式回应查询
 - c “编写”功能可以根据少数开头的提示，帮助用户生成各种场景化个性文本
 - d Bing 搜索和 Edge 浏览器的升级，有望提高用户处理信息的交互性和准确性，个性化呈现附有来源的最新资讯
- 4 TEAMS“智能回顾”：CHATGPT 自动生成会议纪要以个性化时间线，提升线上会议体验
 - a 基于 GPT 系列模型，Teams 高级产品“智能回顾”功能将自动生成会议记录和要点，以时间线记录每位与会者发言，帮助未能到会的用户快速查看和分享会议内容
 - b 公司表示这项服务后续将以订阅制收费，6 月份每月收费\$7,7 月份增加至 \$10，AI 技术进一步商业化落地
- 5 VIVASALES：帮助销售人员自动生成适用各种场景的电子邮件内容
 - a 基于生成式 AI 驱动经验的预览功能，VivaSales 可以根据用户的选择生成场景化电邮内容，帮助销售人员更有效地与潜在客户和消费者沟通
 - b VivaSales 还可以提醒销售人员何时该跟进潜在客户或客户，并按照预设格式的生成个性化的电子邮件回复，有望大幅简化销售人员电邮处理过程，提高办公准确度和效率，优化时间分配
- 6 投资建议
 - a 建议关注
 - iii 对标微软办公：金山办公
 - iv 相关技术公司：百度、科大讯飞、云从科技、拓尔思、海天瑞声、阿里、腾讯
 - v 算力公司：海光信息、寒武纪、景嘉微、浪潮信息、中科曙光
 - vi 国内巨头合作或场景型公司：光云科技、同花顺、神思电子、宇信科技、高伟达、当虹科技、寺库、知乎、汉王科技
 - vii 机器人：三花智控、鸣志电器、绿的谐波、晶品特装、微创机器人等
 - b 风险提示：AI 技术迭代不及预期风险；经济下行超预期风险；行业竞争加剧风险

AI 版 Bing&Edge 浏览器：基于 ChatGPT 的拓展聊天式搜索引擎

- 1 公司宣布推出 AI 驱动的新 BING 主页，增加扩展式的聊天框，可以针对用户的提问给出交互式答案

- a 新 Bing 以类似 ChatGPT 的方式回答具有大量上下文的问题
 - b 例如，新 Bing 可以为用户设计旅行路线，也可以进一步回答用户更多的问题，包括策划的旅行要花多少钱，或者是否可以在行程中添加或更改某些内容
- 2 技术方面，BING 搜索采用比 CHATGPT 更先进的 GPT-3.5 模型，可以获取网络最新资讯。OPENAI 的 CEO 阿尔特曼表示，公司将部分 OPENAI 的 GPT-3.5 语言技术融入了 BING，从而开发了 BING 的聊天式交互技能
- a 公司 Edge 浏览器基于人工智能的必应搜索新增“聊天”和“编写”功能。“聊天”和“编写”被嵌入到 Edge 浏览器的侧边栏，“聊天”功能让用户能总结其正在浏览的网页或文档，并以对话形式回应查询，提供附有来源引用的信息
 - b “编写”功能则充当写作助手，根据少数开头的提示，帮助用户生成从电邮到社交媒体帖子在内的文本
 - c 展望未来，我们预计 AI 技术在搜索引擎中的应用将快速推广落地，科技巨头纷纷跟进有望带来产业加速以及持续消息催化
 - d 微软旗下 Bing 搜索和 Edge 浏览器的升级，将为用户带来高交互式的信息搜索和提取功能，并利用 GPT 模型呈现个性化的最新资讯，此举将有效提升公司在搜索引擎领域的产品竞争力
 - e 据每日经济新闻报道，谷歌 2 月 6 日透露计划推出一款聊天机器人 Bard，与 OpenAI 颇受欢迎的 ChatGPT 竞争，寻求在将强大的新语言人工智能引入互联网搜索业务的竞赛中收复失地

Teams“智能回顾”：ChatGPT 自动生成会议记录和个性化时间线

- 1 公司推出基于 CHATGPT 技术的高级 TEAMS 产品，CHATGPT 可以提供会议记录自动生成、推荐任务、帮助参会者创建会议模板等帮助
- a 基于 OpenAI 的 GPT 模型，“智能回顾”功能能够自动生成会议记录和要点，将每位发言者的贡献突出显示在整齐的主题和章节的时间轴下，同时生成个性化时间线标记以使用户快速查看讨论内容
 - b “智慧回顾”将有效提升会议举办者、参与者和未到会者的体验和工作效率，帮助不同时区的参与者快速查看会议内容、处理和分享会议信息
- 2 TEAMS“智能回顾”功能后续将以订阅制收费，标志着 CHATGPT 正式开启商业化，公加速 AI 落地进度。微软科技官方微信公众号透露，这项高级服务将在 6 月份每月收费 7 美元，然后在 7 月份增加至 10 美元，标志着 CHATGPT 相关的应用正式进入商业化阶段
- a 基于 GPT 技术自动生成场景化电子邮件内容 2023/02/08
 - b 此外，VivaSales 还可以提醒销售人员及时跟进潜在客户或客户并自动生成预设格式的电子邮件回复，包括个性化文本、下一步最佳行动，以及产品描述、建议和截止日期等细节

VivaSales：基于 GPT 技术自动生成场景化电子邮件内容

- 1 公司推出 VIVASALES 基于生成式 AI 驱动经验的预览功能，可以根据用户需求生成场景化电邮内容，帮助销售人员更有效地与潜在客户和消费者沟通
- 2 VIVA-SALES 可以为各种工作场景生成推荐的电子邮件内容，比如回复询问或创建提案、提供与收件人具体相关的数据，包括定价、促销和截止日期

风险提示：AI 技术迭代不及预期风险：若 AI 技术迭代不及预期，NLP 技术理解人类意图水平未能取得突破，则对产业链相关公司会造成一定不利影响

- 1 经济下行超预期风险：若宏观经济景气度下行，固定资产投资额放缓，影响企业再投资意愿，从而影响消费者消费意愿和产业链生产意愿，对整个行业将会造成不利影响，NLP 技术应用落地将会受限

- 2 行业竞争加剧风险：若相关企业加快技术迭代和应用布局，整体行业竞争程度加剧，将会对目前行业内企业的增长产生威胁

详细内容

- 1 技术变化：大语言模型，大量参数情况下把人类几十年在互联网上面所存在的这些数据全部都压缩到一个模型里。GPT 变化在于，它以很好的逻辑把它所掌握的数据提供给你。解决了一个核心的问题，海量的信息和数据如何压缩并且再次提取。但目前还不代表通用 AI 的到来
- 2 CHATGPT 最激动人心的地方：证明了大模型这种新的理念是可行的，新的范式突破。整个 AI 行业，找打了天花板向上提高的方法，未来几年时间，整个人工智能行业就会迎来一次新的创新高潮。缺点不改进：结果准确率不高，因为没更新数据。在研发 WEBGPT，把它不一些及时的数据结合
- 3 由于参数量规模过大，所以它的成本也陡然上升。每一次交互成本接近人民币一毛钱，是商用的一个阻碍。GPT4 在尝试更小的参数规模，戒者同样参数量实现更好的效果，算法不算力：算力必须要不断地膨胀
- 4 展望：大模型的方向，毫无疑问的将会成为未来至少三五年时间的一个主流技术方向，现在已经基本上明确

AIGC，生产力工具的方向

- 1 虽然 1 毛钱的成本，可能非常划算的。小冰上个月月底的时候，刚刚跟 NETFLIX 一起推出全球第一个商真正可以达到商用化质量的动画短片，端到端全部都是由人工智能生成的
- 2 GOOGLE 其实没有那么没有那么恐慌。因为并不能去取代 GOOGLE 的搜索业务。无论是从成本角度考量，还是从用户体验角度准确性的考量。现在的 AI 发展还处于一个萌芽阶段，最领先技术往往不是最能落地的，往前数一代往往是能落地的
- 3 垂直。利用精标的垂直领域的数据做好。不是就一家独大戒者几家独大，是一个百花齐放的过程
- 4 希望老黄多配合大家把它的算力尽可能地不断地去扩展。未来 10 年内，会努力希望算力增长去配合技术的进步。希望算力总值快速增长的情况下去追求技术突破，而不会过早地去考虑优化算法模型戒者技术范式来使它摆脱戒者降低对于算力的要求。老黄多配合大家把它的算力尽可能地不断地去扩展
- 5 大力出奇迹就好，不要想那么多精打细算
- 6 数据标注
 - a 第一个环节，以前我们每一个训练数据，我们都是要做精确的标注。而大模型的特点大力出奇迹，我把一大堆数据全都塞到大模型里面去。在这个过程里面，既不会去进行人工干预，也不可能做人工干预
 - b 第二个环节，数据的精标往往是跟模型所需要完成的任务相关
 - c 在数据隔离这件事情还没有的时候，我们国家的这种数据标注人员实际上不光为中国提供服务，为全世界范围的提供服务。增量部分对算力不标注的需求

有没有一种可能是存量的差不多训练完了，后面新增的数据不太需要有特别多的算力不数据标注？

- 1 在 5 年之内应该不会出现这个问题。由于大模型的新的能力天花板的打破，肯定会产生大量的各种各样的创新，这种创新将像雨后春笋一样出现。因为这个原因，所以算力的需求我觉得是会持续地增长的
- 2 数据标注也是一样的，A 公司的数据标注跟 B 公司的数据标注肯定，戒者哪怕是不同模型的数据标注很可能会采用不同的要求，所以甚至于我们说这里面可能会有重复浪费的情况
- 3 针对模型在算力的使用上面进行优化。往往能够得到 10-20 个百分点可能的改

善。通用性：最近实际上开始说同样的一个方法，可以同时去解决语言、语音和视觉的问题

- 4 大模型就是其中的一个代表性的一个主要的范式。大模型的训练很多都是基于多模态，就是把这种文本的，语音的，静态图像的、视频的训练数据全都放在一起的。大模型是一个很通用的东西，但你用它去最终形成什么样的东西，中国的企业很可能更了解中国用

a 例如，百度更懂的是中国用户的搜索习惯和他们的真实需求

b 认为中美从技术角度来讲是没有什么太大的差距的

商业模式

- 1 目前现状：今天为止，整个全世界范围内，包括美国，还没有看到一个成功的成熟的 AI 的商业模式。目前人工智能创造内容的企业，它往往会创造很高的价值，但是它自己只能从这个价值里面取得很低的收入

- 2 包括整个人工智能。客单价是往往难以提高的。所有的 API 调用形式的，这种类型的人工智能商业模式，戒多戒少都存在着一个天然的弊端。因为没有从 CONTENT 的不同的价值回报中获得不同的收益

- 3 基座模型的商业模式：人工智能的底层的，我们说基座模型就是 FOUNDATIONMODEL 公司

a 如果要想很好地去攫取它的商业价值，并且让它的商业模式比较稳固，它必须要向前走一步，走到至少是端到端的环节

b 如果只是停留在大家可以来调用我，我来做基础服务层，水电煤这一层，我认为的它的商业模式是很难稳固的，并且它的价格提升也是会有很多困难的

- 4 要不要做通用：基座大模型一定是通用的

a 大家有一个基本的认识，我要去做一个足够通用的大模型。我需要投入很多钱，我会得到一个即使是我现在都看不明他到底有多大价值，多大用处的，可能还有很多新奇的用处的。得到这样一个黑盒子，对肯定是一致的

b 技术壁垒/扩散：通用的模型上面，各家都会有一些独特之处，例如一些新的训练方法

c 但是这些独特之处如果是有效的，一般也会很快的被同行者所学习到。所以技术壁垒在这个上面不是特别高

- 5 数据的价值：行业所最珍贵的数据，其实是我们叫 FEEDBACKLOOP 是什么？

a 是在实际使用过程中，由于用户交互或者由于客户交互，由于使用者交互所反馈给平台的。这种反馈信息是随着你的产品而获得的。搜索引擎的抓取技术是很成熟

b 但抓取的几乎所有的数据世界的的数据，公开有机会可以获得的，任何一家公司在这上面的领先优势没有那么大

c 假设字节跳动的平台上面的它的用户量大，它的反馈机制，数据的回路建立得好，它的数据最大部分可能就是超过百度，甚至超过很多的，这都是可能

公司对 AIGC 的看法

- 1 2018 年的 CHATGPT1.0 版本，2022 年的 3.5 版本，年底爆发关注度。技术路线上，1.0-2.0 版本相比谷歌的大模型没有优势，2020 年的 3.0 版本发生了突变，2022 年 STABLEDIFFUSION 等软件带来热潮，叠加 2022 年 12 月 CHATGPT 的推出，致使 23 年 2 月当下市场对 AIGC 的非常关注
- 2 CHATGPT 为什么在短期内能有 1 亿的用户数，在于人机交互上第一次做到了自然语言理解。目前 AIGC 以国外科技巨头引领，几个巨头的博弈会带动全球 AI 的发展，包括国内的一些科技公司和创业公司，因此 2023 年国内 AI 的基础设施建设会迅速提升
- 3 2023 年预计会发布 CHATGPT4.0 版本，据说其参数模型要比 3.5 版本大十几倍，推理能力会更强，且预计是多模态的，不仅包括文字，还包括声音、图片等内容形态
- 4 AIGC 将作为中文在线非常重要的内容生产方式。每年市场产出的文字内容最多（创作门槛低），几百万的文字作品，万集的漫画等，每次内容形式的复杂度提升，都意味着创作门槛的巨大提升。新的技术发展所带来的多媒体形式的统一，会使得生产成本急剧下降，可能会改变内容产品数量级的金字塔式关系，AIGC 带来内容产业的想象空间
- 5 AIGC 领域，公司一直有布局，已有相关产品上线，提升创作质量和效率
 - a AI 图、AI 文字转语音，将文学作品录成有声作品，做到非常自然的人声音色，大幅降低生产过程中的录制成本
 - b AI 小说创作，辅助作者，已在 17k 小说网上线，针对作者开放，通过简单的描述然后 AI 进行续写创作，让作者有更多的时间和精力放在整个作品组织架构上，提升作品质量，AI 会根据作品的类别进行语言调整
- 6 公司海外业务收入占比约一半，其中互动式视觉阅读平台 CHAPTERS 运营情况较好，AIGC 的运用可以降低美术成本，是一个非常好的应用场景，这种对话式的剧本创作，每一次对话选择都有不同的故事线，有一些重复性劳动可以用 AI 来进行替代，正在进行尝试和研究。公司在 AI 领域的合作方是澜舟科技，拥有孟子训练模型，调用了其 API 接口

公司介绍

- 1 数码视讯的定位是全球领先的视频技术服务平台公司，公司于 2000 年成立，2 年在创业板上市。一直致力于视频技术、安全加密技术、AI 技术以及 5G 等基础技术的研发和积累。公司非常注重创新和技术研发，每年研发投入占比超过 20%
- 2 研发人员在公司员工的占比也达 50%。在国内各超一线、一线城市以及美国硅谷都建有研发基地。公司与电视台合作进行后期的辅助制作工作，公司的 AI 模型能将原始素材（公司在研究视频领域的 AI 技术有天然优势
- 3 自 2000 年开始公司就服务于电视台广电网络，储备有大量视频数据，客户需求的驱动也促使公司很早就开始了 AI 领域的布局
- 4 2016 年公司在 VR 和 AI 方面基本上是同步立项并推进相应研发工作，从最早的视频识别、后期的视频 AI 内容优化再到现在视频内容的 AI 生产，有一个进化和演变的过程。AIGC 实际上是对公司 AI 内容生产的一种全新提法，也是面向客户的重要服务模式
- 5 现在很多平台的文字内容也都是通过 AI 进行辅助提升的。公司视频识别及 AIGC 模型能够作为短视频平台的创作工具帮助短视频的创作者开展素材生产和剪辑工作、提高生产效率，也可能在未来集成为一个 2C 的工具类产品。腾讯在短视频方面有其重要或略布局（视频号），而我们在视频技术方面有很强的技术积累，与公司签订了保密合作协议
- 6 根据编导要求剪辑成短视频。比如体育赛事的进球精彩瞬间、红黄牌、越位等片段以及一些宣传物料。目前视频领域的 AIGC 模型还需要通过进一步的训练来提高所生产内容的品质及效率

AI 后期制作或者主动创作的成本大概是怎样的情况？提高效率的同时减少了成本

- 1 21 年军网（8：45）的元宇宙春晚原本需要数月时间制作
- 2 使用了公司的系统后只用了两个星期的时间就完成了

AIGC 在降本增效方面的贡献率能否量化地用比例指标进行衡量？

- 1 首先减少了时间投入和相应的人员工时费用
- 2 以中国军网元宇宙 AIGC 视频平台及内容项目为例，内容制作效率的提高是在 50% 以上

公司目前有在视频文本方面与 ChatGPT 或 OpenAI 等 AIGC 模型进行接口的合作？

- 1 公司业务偏重视频领域，上述这些模型主要应用在图文领域，接入价值不大
- 2 因此目前没有考虑这一合作方向

AI 视频的生成方面与抖音、快手等短视频平台的合作状态如何？

- 1 公司 AIGC 视频应用服务的核心面向对象就是短视频平台
- 2 公司最早和腾讯签署了战略合作协议，现阶段计划全力推进该合作方向，验证使用场景及创作者对 AIGC 产品的接受程度，在此基础上再去推进与抖音等其他短视频平台的合作沟通

客户对 AIGC 的关注热度是否有提升？

- 1 整体热度都在提升，包括电视台客户、局领导等，关注度在持续提升
- 2 考虑 AIGC 方向考虑布局

公司现在和腾讯合作的在手订单量级？

- 1 目前尚不能称为订单
- 2 AI 技术及其相应产品，存在发展过程，公司最早的 AI 内容识别、AI 内容优化，有案例

AI 内容生产，尚未有在手订单，目标是进行模型训练，落地到更多应用场景。公司在 AI 目前与电视台已实现长期合作，公司在视频平台与影视公司合作方面是否有更多的拓展？

- 1 与视频平台的合作。公司一直在大力推进，与腾讯签署保密战略合作协议
- 2 对于影视公司，数码视讯存在 4 大业务板块，影视制作、投资以及发行是其中重要的业务方向，影视领域目前已有很多合作伙伴，与影视制作的后期业务进行对接，提高生产效率，缩短制作周期，目前处于积极推进和沟通状态

应用领域的拓展方面还有哪些潜在的空间？

- 1 在 AI 领域的拓展，目前主要是 AIGC
- 2 作为一家科技型的公司，公司立足于广电行业，对于内容布局一直在探索思考，AIGC 会作为重要的技术去落地，以适配更多应用场景

公司在 AI 方面布局较久，感受技术迭代的变化情况如何？

- 1 AI 迭代存在较长演变，每次迭代都伴随功能提升
- 2 16 年早期的时候，公司注重内容识别，对于整个视频，不仅要识别画面，还要识别语音以及字幕，甚至是画面中的物品，并判断是否存在危险性或敏感性等等。持续迭代过程中，升级到对视频的内容优化，如超分 AI，将质量较低内容用 AI 生成质量较高的内容
- 3 最后演变成目前 AIGC 这类更具智能的生产，目前主要作为短视频素材生成工具，未来期望实现更多高品质视频生产。持续迭代，伴随着更多功能和更多应用场景的实现

如何看待公司的研发投入水平与产出成果，公司主要的竞争优势是什么？

- 1 公司的研发投入位居创业板前列，每年研发投入占比超过 20%，研发人员在全公司占比约 50%
- 2 公司非常注重研发，在北京、深圳、武汉、西安、成都、长沙等一线及新一线的城市，以及美国的硅谷，都设有研发基地，大量的研发投入确保公司具备长期核心竞争力

公司 2023 年的业绩展望如何，及 AIGC 业务能否有所贡献？

- 1 公司 2022 年给出 1.2 亿元的净利润业绩指引
- 2 对于 2023 年，公司预期达到 1.8 亿元净利润，2023 年的业绩指引增幅源于新的研发产品及新拓展的客户
- 3 AIGC 是重点发展方向，目前在营收结构上没有单独统计，我们希望实现更多场景，积累更多客户，并不急于一时营收，在未来某个节点所创造的营收将会是水到渠成。公司的其他业务，如超高清业务、安全业务、应急系统业务等，足以支撑公司达成业绩预期

中外的发展路径是否相同，ChatGPT 从文字交互出发，国内存在一些在语音成熟的公司，在图片、视频领域是否会相同？

- 1 发展路径应当比较一致
- 2 都会从最易实现的方向开始，视频方向会做持续布局展望

行业在 AIGC 的文字、图片、语音、视频等，哪个方面有望率先突破？国内的市场份额谁将是最大的？

- 1 文字是最快的，目前推出的 ChatGPT 已证明文字是当前更具智能化的一个方向。视频是最困难的，其次是图片，图片可以看作视频的一个画面
- 2 视频未来会成为最重要的部分，类比人们信息获取从文字到图片，再拓展到视频，无论是长短视频，还是与人沟通采用视频连线，都说明视频的信息承载是最多的

3 AI 从文字起步最容易的，随着发展，最后一定是在视频领域

近期 ChatGPT 比较火，公司与之有何关联？

- 1 CHATGPT 是一个人机交互的语言模型，是目前 NLP 领域最前沿的研究成果之一，由于其高质量的对话内容响应能力
- 2 事实上相当于解锁了多样化的文本内容生成能力，目前成为 AIGC 里程碑式的模型或产品。CHATGPT 是一个通用的大模型，而生成式模型做为一个黑匣子，仍然具有结果不可控的特点
- 3 相对而言，公司基于自身在 NLP 技术领域的全面性以及长期在行业端的深耕，对不同行业客户的数据特点、业务需求的理解更为深刻
- 4 在项目磨炼中，已经形成自身独有的算法模型，更能为行业客户提供满足需求、输出结果更为专业精准的专业化模型，这在前面提到的项目中已经落地并得到实践验证

汉王在文本大数据业务上，主要的竞争优势？

- 1 公司曾获国家科技进步一等奖的手写识别技术、国家科技进步二等奖的 OCR 识别、业内领先的 NLP 自然语言处理等一系列领先技术支撑公司成为文本大数据领域的龙头企业，公司的满文古籍、藏文古籍、民国文献、医学文献等人文数字化业务长期处于行业龙头地位
- 2 公司在文本大数据业务上依托 OCR 识别、自由手写文稿识别、复杂报表识别、自然语言处理 NLP、RPA 机器人流程自动化等技术以“技术+产品+服务”的方式为客户提供服务
- 3 形成“技术平台+行业赛道”的业务模式，在各行业赛道中寻找数字化、智能化的落地场景，并在全国重点战略城市成立分子公司搭建大营销网络，目前在司法、档案、医疗、金融、数字人文、教育等众多领域广泛应用

公司文本大数据业务技术积累，主要应用在哪些领域，未来的趋势是什么？

- 1 公司是人工智能领域的创新型企业，通过三十年的不断演进，形成业内领先的文本图像识别、自然语言处理（NLP）、生物特征识别、智能视频分析、智能人机交互等人工智能产业链关键技术
 - a 公司的文本大数据业务就是基于 OCR 技术、NLP 技术，从文本数据的采集、结构化处理到数字化、智能化利用的大数据服务，经过多年发展已经形成较为完整的产品体系
 - b 并在司法、人文、金融、档案、政务、医疗等行业积累一大批具有粘性的优质客户群，对客户需求有深入理解，在各领域树立了具有示范作用的标杆项目
- 2 从模式上看，公司目前主要是通过嵌入式业务服务+多种 AI 机器人的形态，帮助客户完成业务资料数字化及业务流程的智能再造，提升公共服务的效率与服务水平
- 3 在数字中国、数字经济建设的大背景下，相关行业数字化、智能化建设的政策导向明显，公司希望以自身技术为驱动
- 4 帮助客户打通从文本信息—电子信息—文本数据—结构数据—RPA—数据智能决策（行业机器人）的模式，实现机器代替或辅助人进行相关信息数据的智能处理或决策
- 5 未来随着算力、NLP 技术、大数据技术、RPA 等软硬件技术的发展，AI 在图像及文本数据上的智能处理将得到更广泛的应用，自动生成问答、虚拟员工、虚拟机器人等 AI 智能服务也将得到更多的市场认可

目前公司的业务在国内的开展区域似乎集中在北京，天津，在其他区域有进展吗？

- 1 在国内 TOB 方面，公司通过标杆项目、设立区域中心、资源合作等方式进行业务

扩张及渠道开拓，在全国主要城市设立了十余个分子公司、区域中心辐射国内重点区域市场

- 2 并通过第三方代理体系增强覆盖偏远地区，目前在北京、上海、广州、深圳、杭州、武汉、许昌等重点战略城市均布局标杆项目或区域中心
- 3 在国内 TOC 方面，目前公司主要通过京东、天猫、拼多多及其他平台进行线上产品销售

公司之后如何进一步降本增效，提高净利率？

- 1 公司管理层会进一步加强费控管理，对一些长期亏损且前景不明的业务进行及时调整，重点优化和发展市场潜力足、前景明朗的产品和业务线条
- 2 受疫情等因素影响，2022 年公司的原材料成本处于相对高位，随着库存原材料的消化、产品销售的完成及公司对供应链的调整，在成本上的控制及优化应会有所体现
- 3 为稳固市场地位，公司数字绘画部分产品在 2022 年的售价有所下降，加上原材料成本上升、物流成本上升等因素影响，毛利率有所下降
- 4 但是搭载数字笔芯片的新产品以及高端专业数字绘画产品的陆续上市，使公司数字绘画产品体系更加完善，市场地位更稳定
- 5 随着芯片、屏等原材料以及物流成本的回落，专业产品销售额和数量的提升，应会带动数字绘画整体业务的毛利率回升

如何看待公司的研发投入水平与成果？

- 1 作为人工智能领域的创新型公司，公司一直保持较高水平的研发投入，技术领先是公司的生命线。2021 年，研发投入占比为 177%，研发投入占比较大
- 2 长期的研发投入为公司储备和积累了大量的先进技术，公司管理层会进一步努力经营，加快技术落地转化，通过扩大业务规模，将技术优势转化为市场规模，摊薄研发费用，提升公司的销售利润率，体现公司的技术价值
- 3 未来公司会继续保持对重点项目及前沿技术的投入，确保公司在人工智能领域的技术领先性及新产品的推出，为业务开展奠定技术储备及产品基础

2022 年业绩预告解读？

- 1 整体来讲，2022 年的经营压力跟疫情的波动反复有较大联系。主要体现在以下两个方面：一是对公司业务开展和市场开拓的直接影响，海外数字绘画业务目前占公司业务比重较大，2020 年疫情爆发刺激线下消费向线上消费加速迁移
- 2 同时在线教育的教学互动需求被激发，带动公司数字绘画业务大幅增长，之后受海外防疫政策的变化、高通胀、以及俄乌战争影响，海外数字绘画业务受到影响；同时，国内受疫情反复等影响，公司 TOB 部分项目无法正常实施、新项目开拓受到影响
- 3 二是疫情对产业链影响的传导导致供应链波动，公司的原材料成本处于相对高位、交付延迟等，以及为应对市场波动，稳固市场地位，部分产品降价销售，导致毛利率受到影响
 - a 虽然 2022 年公司面临很大压力，但是公司的战略性、必要性投入并没有停滞，继续保持较高水平的研发投入，确保公司在人工智能领域的技术领先性及新产品的推出
 - b 为之后的业务开展奠定技术储备及产品基础；就 2023 年来讲，目前疫情因素已经基本可控，公司对各项业务制定了较为明确的经营计划
 - c 管理层及业务团队也会继续努力、提升经营效率，克服困难，推进各项业务的顺利开展

目前公司的业务在国内的开展区域似乎集中在北京，天津，在其他区域有进展吗？

- 1 在国内 TOB 方面，公司通过标杆项目、设立区域中心、资源合作等方式进行业务扩张及渠道开拓，在全国主要城市设立了十余个分子公司、区域中心辐射国内重点区域市场
- 2 并通过第三方代理体系增强覆盖偏远地区，目前在北京、上海、广州、深圳、杭州、武汉、许昌等重点战略城市均布局标杆项目或区域中心
- 3 在国内 TOC 方面，目前公司主要通过京东、天猫、拼多多及其他平台进行线上产品销售

概要

- 1 讯飞有望迎来“戴维斯双击”回顾历史，讯飞过去的股价上涨周期大多由行业催化剂和公司基本面改善共振驱动
- 2 随着 23 年疫情好转、22 年项目型延期订单确认及 23 年恢复增长、运营型业务持续快速增长，公司 2023 年营收有望恢复高增长；同时叠加人员控制和人均效益提升，公司 23 年净利润或将实现高增长
- 3 我们认为，在 CHATGPT 的火爆催化和公司基本面 2023 年大概率反转的双重驱动下，公司有望迎来“戴维斯双击”
- 4 鉴于疫情影响合同项目交付节奏，下调公司 22-23 年 EPS 至 0.24、0.77 元，预计公司 24E EPS 为 1.0 元。可比公司平均 23E 76.1X PE (WIND)，考虑到公司在 AI 领域具备深厚技术积累，具备一定产业领先地位，给予公司 23E 80X PE，对应目标价 61.60 元，给予“买入”评级
 - a 2022 年公司业绩和股价双触底，ChatGPT 火爆助推讯飞估值提升 2011 年~2021 年讯飞连续十一年保持营收增速在 25% 以上，受疫情影响，2022 年营收实现同比增速 0%~10%，跌至十余年来最低；扣非净利润同比下滑 45%-60%
 - b 股价方面，2021 年 6 月中旬后，“双减”政策驱动的行市结束，讯飞股价开始回调，2022 年 10 月 11 日公司股价触底，期间跌幅近 50%，同期中信计算机指数跌幅 29%
 - c 近期 ChatGPT 火爆成为讯飞股价上涨的重要催化剂，23 年 1 月 16 日-2 月 1 日讯飞股价上涨 44% 左右，同期中信一级计算机指数上涨 14% 左右

2023 年

- 1 公司业绩大概率迎来反转展望 2023 年，因疫情影响的项目型业务有望在 2023 年确认收入，同时公司运营型业务 22 年营收同比增长 23%，23 年仍有望保持快速增长态势
- 2 据董事长 2023 年年会演讲内容，预计 23 年讯飞英语听说考试、学习机、个性化手册业务营收同比增长 45%、77%、38%，智医助理/开放平台/智能硬件合同毛利同比增长 80%+/30%/35%
- 3 费用端，2023 年公司将严控人员规模增长，更加注重人均效益提升，公司人才进入红利释放阶段。综上，我们预计，公司 2023 年大概率迎来收入及净利润的快速增长
- 4 2023 年起公司迈入利润年增速 30% 以上的高质量发展阶段根据董事长 2023 年年会演讲内容，2023 年起，公司战略由规模扩张阶段转向高质量发展阶段
- 5 体现在四个方面：利润实现年 30% 以上的增长；现金流匹配利润增长；人均效益三年内翻番；运营型业务五年内年收入达到 500 亿元，收入占比达到 80% 以上，业务毛利润达到 200 亿元。我们看好运营型业务助推讯飞实现长期可持续性发展

回顾历史

- 1 公司股价上涨周期多由行业&公司基本面共振驱动 2011 年以来，公司股价历经三次上涨周期：2015 年 1~6 月、2017 年 6~12 月和 2020 年 10 月~2021 年月
- 2 回顾历史，除 2015 年是整体创业板牛市外，之后讯飞的每轮股价上行背后都有公司基本面改善和 AI 行业催化剂的驱动
- 3 AI 行业层面的催化剂成为拉动讯飞估值提升的重要驱动力，公司基本面的边际改善提升了公司 EPS，两者共振驱动讯飞迎来“戴维斯双击”

华泰计算机

- 1 讯飞股价涨幅 155.0%，同期中信计算机指数下跌 3.76% 行业层面，2016 年 3 月，

GOOGLE 旗下的 DEEPMIND 智能系统 ALPHAGO 对战世界围棋冠军、职业九段选手李世石，并以总比分 4：1 的明显优势战胜李世石

- 2 2017 年 5 月，ALPHAGO 迎战世界排名第一选手柯洁再次以 3：0 的绝对优势获胜。ALPHAGO 的接连获胜引发了资本市场对 AI 的热烈讨论
- 3 公司层面，2017 年科大讯飞“平台+赛道”战略初具成效，2017 年公司营收同比增速首次突破 60%，达到 64.0%，较 2016 年上涨了 31.2PCT；扣非净利润同比增长 40.7%
- 4 2020 年 10 月~2021 年 6 月：讯飞股价涨幅 99.1%，同期中信计算机指数微涨 0.17%公司层面，在数十年的深耕和沉淀下，公司智慧教育业务红利开始释放，区域因材施教解决方案大规模推广，陆续在安徽省蚌埠市等地区形成标杆产品。2020 年，公司教育领域营业收入同比增长 67.6%，相较 2019 年上涨了 51.5PCT
- 5 行业层面，2021 年 3 月，教育部颁布“睡眠令”，同年 7 月出台“双减”政策。“双减”政策的出台使得资本助推的校外培训行业受挫，科大讯飞专注于 K12 校内教育业务，“双减”政策的出台对讯飞智慧教育业务形成了重大利好，也推动讯飞股价有了显著的上漲

2022 年

- 1 公司业绩和股价双触底业绩底：2022 年营收增速跌至十余年来最低 2011 年~2021 年科大讯飞连续十一年保持营收增速在 25%以上
- 2 2011 年以来，科大讯飞稳扎稳打，丰富产品矩阵，强调 AI 的规模化应用，积极扩大市场份额，追求营收的高速增长。所以过去十一年，公司营收增速都维持了 25%以上的快速增长，其中 2013 年和 2017 年分别高达 59.9%和 64.0%，201 年增速最低（27.3%）
- 3 受疫情影响，2022 年公司营收增速下滑至 0%~10%，超 30 亿元项目延至 2023 年。根据公司业绩预告，2022 年受新冠疫情反复等不利因素影响
- 4 预计 2022 年营收实现同比增速 0%~10%，跌至十余年来最低；归母净利润同比下滑 60%-70%；扣非净利润同比下滑 45%-60%
- 5 主要原因是全国各地项目招投标延迟，尤其是在 4Q 有超过 20 个项目、超 30 亿元合同延期

股价底

- 1 2021 年 6 月-2022 年 10 月讯飞股价下行近 50%2021 年 6 月中旬后，“双减”政策驱动行情结束，科大讯飞股价开始回调，2022 年 10 月 11 日，公司股价触底，跌至 31.10 元，期间跌幅近 50%，同期中信计算机指数跌幅 29%。之后公司股价跟随计算机指数震荡向上
- 2 CHATGPT 的火爆成为此轮讯飞上涨的行业催化剂。2022 年 12 月 2 日，OPENAI 历经三年研究正式推出 CHATGPT。作为全新的人工智能（AI）聊天机器人，CHATGPT 被认为正在“掀起新一轮 AI 革命”。2 月 1 日，OPENAI 官方发文称将推出 CHATGPT 的试点订阅计划 CHATGPTPLUS，定价每月 20 美元
- 3 CHATGPT 和大模型的火爆引发市场对人工智能的高度关注，国内资本再次热议 AIGC 的投资价值。202 年春节后 CHATGPT 板块迎来“开门红”，开盘第一周 CHATGPT 指数上涨 15.9%

2023 年

- 1 公司业绩大概率迎来反转 2023 年起讯飞将开启高质量发展新阶段 2023 年起，公司战略由规模扩张阶段转向高质量发展阶段。在 2023 年 1 月 18 日科大讯飞 2023 年会上，公司董事长刘庆峰提出，“依托根据地 and 系统性创新，2023 年科大讯飞将开启高质量发展的新阶段
- 2 “同时，公司明确定义了高质量发展阶段的四个关键维度：利润、现金流、人均效

益和可持续性。具体来看

- a 利润：2023 年，公司将开启利润年增长 30% 以上的新阶段。考核机制上从规模优先转变为效益优先，将利润指标变为考核指标的第一权重
 - b 现金流：现金流要匹配利润增长，同步良性发展
 - c 人均效益：通过人才招引和内部培养系统化提升，丰富人才梯度类型，加快高潜员工的发展，实现人均效益在三年内能够翻番，扩大员工的成长空间
 - d 可持续性：运营型根据地业务和持续流水型业务目标在 5 年内实现收入占比超过 80%，项目型业务收入占比控制在 20% 以内
- 3 讯飞未来五年目标：运营型根据地和持续流水型业务直接可触达的年收入规模达 500 亿元。根据公司发展目标，五年内公司运营型根据地和持续流水型业务年收入达到 500 亿元，业务毛利润达到 200 亿元，夯实公司长期可持续发展的基本盘
- 4 其中，教育、消费者领域可预见收入各 200 亿元，医疗领域收入 100 亿元，智能汽车、运营商领域收入各 50 亿元，智慧城市和智慧金融业务收入合计 50 亿元

2023 收入端

- 1 我们预计项目型业务恢复增长，运营型业务继续保持快速增长项目型业务：2022 年项目型业务受疫情影响延期，2023 年有望恢复增长
- a 2022 年受新冠疫情反复等因素影响，对公司项目型业务造成较大扰动。根据公司业绩预告，22Q4 超 20 个项目、超 30 亿元合同延期至 2023 年
 - b 据公司 2023 年 1 月 30 日发布的《2023 年 1 月 30 日投资者关系活动记录表》，2022 年公司延期项目中超过 60% 的项目属于教育领域，剩余 40% 与智慧医疗和智慧城市相关，公司预计大部分项目可在 23H1 招标完成
 - c 2022 年公司运营型业务收入尽显韧性，2023 年有望实现更快增长在疫情背景下，2022 年公司运营型业务收入仍同比增长 23%
- 2 根据公司 2023 年年会董事长演讲内容，运营型业务中教育领域占比最大，在 22Q4 大部分学校提前放假的背景下，22Q4 运营型业务收入依然实现了同比增长 13%，对冲了大项目延迟所带来的压力，成功应对了宏观环境的不确定性

智慧教育

- 1 运营型业务进一步夯实，新领域布局完成迈入规模扩张期讯飞深耕智慧教育领域十余年，搭建了比较完备的产品体系
- 2 经过十多年的深耕，公司目前积累了深厚的行业 KNOW-HOW 和人才储备，建立了紧密的客户关系，成功搭建完备的产品体系。从场景角度来看，讯飞智慧教育提供了覆盖教、学、考、评、管的教育全场景解决方案
- 3 从商业模式来看，讯飞智慧教育形成了面向国家到省、市、县、区、学校、课堂以及个人的全方位智慧教育产品体系
- 4 从产品角度来看，随着智慧体育、智慧心育等新产品的推出，公司已经形成了从学业能力提升、到五育并举、再到底层平台的产品格局
- 5 2022 年，公司智慧教育运营型业务收入保持了稳健增长。同时，公司积极参与新兴业务的建设，在课后服务平台、智慧体育、智慧心育等方向进行了新产品的研发和业务拓展投入。具体来看：1) 英语听说考试：英语听说是典型的 GBC 联动根据地业务，能够在环境波动下体现出良好韧性
- 6 根据公司 2023 年年会董事长演讲内容，2022 年公司英语听说业务营收同比增长 35%，新增 12 个地市的中考和 2 个省市的高考，新增市场学生 100% 覆盖。公司预计，2023 年英语听说考试业务营收有望同比增长 45%

学习机

- 1 学习机业务分为线上、线下两大模式。2022 年整体学习机业务营收同比增长

53%。线下门店开始形成持续流水，营收同比翻番

- 2 同时，学习机的因材施教独特价值进一步得到明确，受到了市场的广泛肯定。公司预计，2023 年学习机营收同比增长会比 2022 年更快，增长目标为 77%

个性化学习手册

- 1 在高中个性化手册领域，2022 年讯飞运营的收费学校新增 300 所，累计达 1,400 校，收入同比增长 12%，学生续购率由 2021 年的 89%提升至 2022 年的 91%，付费转化率超过 70%，成熟运营校的 ARPU 值同比提升了 16%
- 2 公司预计，2023 年新增 600 所学校，目标营收同比增长 38%。同时，公司业务还延伸至中职教育市场，挖掘职教高考用户的需求，目前已在部分省市形成了应用示范
- 3 课后服务平台、智慧心育等新兴业务布局完成，2023 年有望进入规模推广期
- 4 根据讯飞 2023 年年会董事长演讲内容，2022 年科大讯飞积极布局课后服务平台、智慧体育和智慧心育三大新领域，目前均已初步完成前期搭建，有望在 2023 年进入规模推广期

课后服务平台

- 1 “双减”政策聚焦学校主阵地，学校是未来重要的场所。针对“三点半”难题，科大讯飞推出课后服务平台，实现课程接入、管理及必要的资源共享。据公司《2022 年 10 月 29 日投资者关系活动记录表》、《2023 年 1 月 30 日投资者关系活动记录表》
- 2 从收费模式来看，目前有两种付费模式：第一，以县区为单位付费，项目建设后可形成持续稳定的运营服务收益；第二，学校付费建设管理平台，更多收益来自于学生按照课时付费
- 3 从业务进展来看，根据讯飞 2023 年年会董事长演讲内容，2022 年公司新增占位 130 个区县平台，累计覆盖区县 280 个，形成 260 门课程，超过 5,000 所学校开课，市场份额超过 50%

智慧体育

- 1 讯飞智慧体育解决方案依托 AI 教学助手、AI 测评系统、智能穿戴设备三种产品形态，覆盖日常教学、体育比赛、课后训练、国家体测、体育考试、体育教育管理六大应用场景
- 2 2022 年，讯飞智慧体育解决方案支持超 20 项体育考试和体能训练项目结果诊断与运动姿态指导，累计开展 110 多万次评测指导，帮助学生运动成绩平均提升 13.7%

智慧心育

- 1 科大讯飞智慧心育是一款服务于中小学心理健康教育的整体解决方案，引入北京师范大学心理学部、华东师范大学上海市心理健康与危机干预重点实验室、中科院心理所专家的研究成果
- 2 讯飞智慧心育解决方案能够为学生、教师、家长、校长不同人群提供减压星球、心育资源、家庭教育指导和心育大数据服务。根据讯飞 2023 年年会董事长演讲内容，2022 年，公司智慧心育业务已覆盖 20 多个省级行政区，已有学生用户 70 余万人，累计完成 135 万次心理测评
- 3 综上，2023 年讯飞教育业务营收有望实现快速增长。据公司《2023 年 1 月 30 日投资者关系活动记录表》：英语考试、学习机、个册 2023 年营收增长目标分别为 45%/77%/38%

新兴业务

- 1 课后服务平台在 2022 年属于布局期，2023 年起进入课程收费阶段；智慧教育、智慧心育在 2022 年属于探索期，根据讯飞 2023 年年会董事长演讲内容，目前也

已积累了上亿元商机

2 因此，我们预计公司智慧教育业务 2023 年营收有望实现快速增长

智慧医疗

- 1 2023 年智医助理业务将保持 80%以上的高速增长 2022 年智医助理的刚需不断显现。根据讯飞 2023 年会董事长演讲内容，2022 年公司智医助理合同毛利增长 100%，2023 年仍将保持 80%以上的高速增长
- 2 在全科医生助理方面，2022 年科大讯飞已累计覆盖全国 380 个区县，累计提供 AI 辅诊建议次数 5.3 亿次，累计提供有价值的修正判断超过 63 万份，累计识别不合理处方数 4,000 万个
 - a 目前，AI 辅诊推荐 Top1 准确率已经提升至 97%，覆盖了超过 2,000+种疾病。同时，AI 用药审核的正确率也提升至 95%，覆盖药品数量已经超过 5 万种
 - b 在家庭医生助理方面，公司产品广泛应用于慢病随访、通知宣教、体检预约、疫苗接种、满意度随访、疫情防控等多场景，累计服务超过 10 亿人次
 - c 发展智慧医疗多元业务，丰富产品矩阵。除了智医助理外，科大讯飞还孵化了智慧医院、智慧医保、安徽影像云平台等新产品，形成了非常重要的试点应用
- 3 根据讯飞医疗官微发布的 2022 年度报告显示，截至 2022 年底，公司累计向超过 300 家三甲医院提供智慧医院解决方案，为医生提供了 1,100 万次智能规划诊疗和协同办公服务
 - a “人工智能全病案医保审核系统”目前已在吕梁市全面上线应用，常态化开展医保数据合规审核；安徽影像云平台联通了 1,800 多家医疗机构，提供远程影像服务 450 万次，存储了 6,100 万例影像检查数据
 - b “一体两翼”核心技术框架推动智慧医疗不断迭代演进。为了进一步突破智慧医疗领域，科大讯飞提出“以医学知识自学习为基础底座
 - c 以诊疗推理、健康交互为两大技术”应用的医疗人工智能“体两翼”核心技术框架体系。在诊疗推理技术领域，AI 用药审核能力显著提升，图神经网络的时序推理技术方案取得突破
 - d 在健康教育技术领域，基于患者画像和医疗事件图谱的多轮交互推理技术显著提升，多模态 AI 问诊预研取得了突破。在医学知识自学习领域，公司构建了面向医生、患者、AI 的一体化医药知识体系

智慧汽车

- 1 产品布局不断丰富产品布局：从语音到驾舱到音响，公司智能汽车根据地产品不断丰富。根据公司 2023 年 1 月 30 日投资者关系活动记录
- 2 在 2019 年前，公司在汽车领域的主要产品是语音交互套件，提供语音能力给汽车厂商，汽车厂商负责开发相关应用系统

智能座舱

- 1 即智能座舱信息域智能控制器，具体来说，可以将域控制器理解为电脑主机，能够控制汽车上的导航、音乐等各种服务。目前，公司能够提供从软硬件一体化的控制器到相关的应用系统
- 2 根据讯飞 2023 年年会董事长演讲内容，2022 年，讯飞智能汽车新增定点储备车企 110 多个，2023 年公司预计新增车型超过 130 个。2022 年新增定点储备收入达到 38 亿元，同比增长 118%。2023 年，存量定点储备项目可提供 4 亿元营收，新增项目则将提供 3.4 亿元营收

智能音效

- 1 由于新能源车机械结构简单，汽车产生的噪音较传统燃油车更小，安静环境下车内人员对音响效果的敏感度和音质要求均更高。而科大讯飞将音频技术和人工智

能技术相结合，显著提升了车载音响效果。讯飞智能音响一经推出，受到了市场的广泛欢迎

- 2 根据 2023 年 1 月 30 日投资者关系会议纪要，目前一汽、广汽、上汽等汽车厂商的多款车型已经完成定点；音效产品随着智己 L7、传祺 M 等合作车型落地量产，预计未来会有更多的新能源车配置（来源：科大讯飞智能汽车官微）

其他运营业务

- 1 运营商、智慧城市、智慧金融等业务亦实现突破 1) 运营商业务：由于国家管理体系对重大数据安全的要求，运营商在国计民生领域的主导地位重新增强，重回黄金时代
- 2 科大讯飞既有独特技术，又是三大运营商的战略合作伙伴，同时中国移动是公司的第一大股东，有较大的机会开拓市场
- 3 根据公司 2023 年 1 月 30 日投资者关系活动记录，2022 年，公司累计终端用户数 9,100 万，日语音活跃用户数首次超千万
- 4 一老一少等运营型业务收入达到 6,100 万元，同比增长 91%。2023 年公司目标是运营商业务中的运营型业务毛利占比从 2022 年的 32% 提升至 38%，运营型收入同比增长 57%

智慧城市

- 1 2021 年 12 月，科大讯飞中标 5.5 亿元的安徽省一体化平台项目，支撑数字安徽发展。目前，仅有安徽省将全省数据放在同一个一体化平台
- 2 实现了省、市、县、区、乡五级联动，大幅降低单个部门信息化服务成本，提升数据共享效率。2022 年，该项目取得了较大进展，公司已将其报备国务院，有望成为政府数字化转型的标杆项目
- 3 皖事通是在项目期间打造的移动端平台。在公布的 2022 年度政务 APP 排名中，皖事通排第四，仅次于个人所得税、国家反诈中心和国务院 APP，服务了安徽超 7,000 万名老百姓

智慧金融

- 1 科大讯飞的智慧金融业务主要面向银行、保险、证券等金融行业客户，为他们提出数字化管理与数字化经营的整体解决方案，提供包括 AI 中台、数据中台、知识中台、金融营销平台、智能客服等 AI+金融应用
 - a 2021 年起公司与浦发银行联合打造的“客户互动服务实时感知与智能分析平台”荣获“亚洲银行家最佳会话式人工智能项目”称号，在金融行业客户服务领域中率先落地基于人工智能技术的实时风控
 - b 综上，2022 年公司运营型根据地和持续流水型业务呈现良好态势，2023 年公司预期各子领域仍能保持快速增长，拉动科大讯飞整体业务向好
 - c 2023 年费用端：注重人均效益，人才红利开始释放 2023 年公司更加注重人均效益，人才进入红利释放阶段
- 2 根据讯飞 2023 年会董事长演讲内容，202 年科大讯飞期间费用增长 9 亿元，其中 70% 左右是人员费用，这些人员大多用在运营型根据地开拓和新产品研发上，这些投入在 2022 年并没有能够形成较大的当期收益
 - a 根据公司 2023 年 1 月 30 日投资者关系活动记录，科大讯飞预计采取两方面措施为了控制人员增长、提升人均效益：控制人员数量
 - b 据 Wind 数据，2021 年公司总员工数 14,307 人，同比增加 3,301 人。我们认为，随着新员工经过了 1~2 年的训练，有望在 2023 年逐步贡献增量价值
 - c 此外，据公司《2023 年 1 月 30 日投资者关系活动记录表》，2023 年公司除了应届毕业生和高端人才外，公司没有其他大型招聘计划，目标做到人员总数同

比基本不增长

- 3 提高人均效益。为了能够提高人均效益，科大讯飞提出了“零基预算”战略。“零基预算”战略是指所有部门压缩 10% 的编制，90% 的员工完成以前 100% 的工作量，公司将剩余 10% 的人员重新分配到新的战略方向上。通过内部组织的效能优化、现有员工能力的提升以及前后端拉齐，公司有望实现人均效益的增长
- 4 综上，我们预计，讯飞 2023 年收入端有望实现快速增长，费用端有望得到控制，从而带动净利润同比实现高速增长

盈利预测和估值收入及盈利能力预测智慧教育业务

- 1 主要面向 G/B/C 三类客户提供产品及服务：G 端业务主要以市县区等区域建设为主体，涵盖面向区域内所有学校及用户的因材施教整体解决方案等
- 2 B 端业务主要以学校建设为主体，包括校内大数据精准教学、英语听说考试等；C 端业务主要以家长用户群自主购买为主，包括人工智能学习机、个性化学习手册等产品
- 3 智慧教育业务 2021 年收入同比增速为 48.85%，增速较快，主因公司区域级因材施教解决方案规模化扩张加速。考虑到：公司因材施教解决方案持续复制推广，G 端业务收入有望持续高增长
- 4 据公司官网，2022 年 B 端业务中的考试业务营收同增 35%、新增落地 12 地市中考和 2 省市高考，B 端客户持续开拓；据公司官网，2022 年 C 端业务中的学习机产品营收同增 53%，个性化学习手册业务营收同增 12%、续费率同增 2PCT 至 91%，C 端产品持续放量
- 5 据公司 2022 年业绩预告，受疫情反复影响，2022 年公司部分合同项目延期，收入确认滞后。考虑到上述因素，结合公司对 23 年收入预测，我们预计公司 2022-2024 年智慧教育业务营收增速分别为 5.00%、40.00%、35.00%

智慧城市

- 1 公司智慧城市业务以“城市超脑”为核心，构建新型智慧城市的数据中台、智慧中台和业务中台。2021 年智慧城市业务营收增速为 29.62%
- 2 据公司 2021 年报，科大讯飞“城市超脑”业务已应用于安徽铜陵、宣城、芜湖、合肥、亳州等地，并在天津、山东聊城、吉林长春、河南漯河、陕西铜川、内蒙古乌海等国内多个城市落地
- 3 此外，据中国政府采购网，2022 年 12 月 26 日公司全资子公司讯飞智元以 5.5 亿中标安徽省“全省一体化数据基础平台”项目
- 4 我们认为，公司智慧城市业务在全国多地持续推广，持续拿下大额订单，有望形成良好示范效应，或将带来智慧城市业务营收较快增长
- 5 此外，考虑到疫情对公司 2022 年合同项目交付的影响，预计 2022 年公司智慧城市业务营收增速下行，随着疫情好转，2023 年营收增速有望上行
- 6 因此，我们预计公司 2022-2024 年智慧城市业务营收同比增速分别为 4.00%、22.00%、20.00%

开放平台及消费者业务

- 1 讯飞人工智能开放平台是基于人工智能技术与大数据运营能力建设的人工智能技术与生态服务平台，具体形态包括讯飞 AI 营销平台、讯飞智能工业平台等能力平台等。在消费者领域，公司围绕 AI+办公场景，主要提供智能办公本、录音笔、讯飞听见 APP 等产品
- 2 2021 年公司开放平台及消费者业务营收同比增速为 52.19%，增速较快，主因：开放平台下游场景持续拓宽，AI 营销业务 2021 年营收同比增长 55.55%
- 3 智能硬件产品线持续拓宽，2021 年发布二代智能办公本 X2,2021 年消费类硬件销

量同比增长 140%

智慧金融

- 1 公司智慧金融业务主要为银行、保险、证券等金融行业客户提供信息化解决方案，2021 年营收 1.97 亿元，同增 33.82%
- 2 我们认为，随着公司智慧金融业务持续提升在区域性银行等金融客户中的渗透率，有望实现营收较快增长。此外，考虑到疫情对公司 2022 年合同项目交付的影响，预计 202 年该业务营收增速下行，随着疫情好转，2023 年营收增速有望上行
- 3 我们预计 2022-2024 年公司智慧金融业务营收增速分别为 15.00%、25.00%、25.00%

其他业务

- 1 公司其他业务主要包括正在培育中的一些新应用场景业务，如 2019 年公司其他业务包括了智慧医疗业务，随着业务成熟，2020 年智慧医疗业务单独披露
- 2 因此，考虑到业务体量较小、公司持续加大新产品/技术投入，我们预计 2022-2024 年该业务营收同比增速均为 5%

毛利率

- 1 考虑到公司未披露分业务毛利率，因此我们只对公司整体业务毛利率进行假设。公司 2019-2021 年毛利率分别为 46.02%、45.12%、41.13%，2021 年毛利率有所下滑，预计主因
 - a 智慧教育业务中，学习机等硬件产品收入占比提升
 - b 智慧城市业务中，公司总包集成等信息工程业务收入占比提升
 - c 根据地业务人员投入加大，21 年人员规模增长 3,301 人，部分业务营业成本上升
- 2 展望未来，我们认为公司过去战略投入红利或将逐步显现，高质量发展目标下毛利率有望保持小幅上升趋势，预计 2022-2024 年公司毛利率分别为 41.20%、41.50%、41.80%

费用率预测销售费用率

- 1 公司销售费用主要包括职工薪酬、广告宣传费、外包服务费等。2019-2021 年销售费用率分别为 17.66%、16.00%、14.70%，销售费用率呈下降趋势，主因公司业务的规模效应初步显现，营业收入增速超过销售费用增速
- 2 我们认为，公司作为国内人工智能龙头企业，深耕行业数字化十多年，具备一定的技术壁垒和品牌优势，规模效应下销售费用率有望保持稳中有降
- 3 此外，考虑到 2022 年受疫情影响，公司部分合同项目延期，收入确认延后，但销售成本前置，故 2022 年公司销售费用率或将有所提升
- 4 基于此，我们预计 2022-2024 年公司销售费用率为 15.00%、15.00%、14.70%

管理费用率

- 1 公司管理费用主要由职工薪酬、折旧费、租赁费等费用组成。2019-2021 年管理费用率分别为 7.01%、6.58%、6.02%，总体呈持续下行趋势
- 2 我们认为，随着“零基预算”战略的推动，公司管理费用率将进一步得到控制，预计 2022-2023 年公司管理费用率分别为 6.00%、6.00%、5.50%

研发费用率

- 1 2019-2021 年研发费用率分别为 16.27%、16.98%、15.45%，研发费用率相对稳定。公司的研发费用主要由职工薪酬、无形资产摊销、折旧费等组成
- 2 2020 年公司研发费用率同比上行，主因是公司在运营型根据地业务上大规模投入，持续加大新产品研发力度
- 3 据公司 2022 年业绩预告，2022 年公司加大了在教育、医疗等业务领域的研发力

度，新增投入约 8 亿元。基于此，我们预计 202 年公司研发费用率或将上行

- 4 但考虑到随着公司营收恢复较快增长，我们认为，2023 年开始公司研发费用率或将转为下行。因此，我们预计 2022-2024 年公司研发费用率分别为 18.50%、16.00%、15.70%

估值

- 1 采用相对 PE 估值法，给予公司目标价 61.60 元。综上，预计公司 2022-2024 年 EPS 分别为 0.24、0.77、1.06 元。选取国内 AI 行业上市公司虹软科技、恒生电子、四维图新、格灵深瞳作为可比公司
- 2 采用 PE 估值法，可比公司平均 2023E 76.1XPE，考虑到讯飞在 AI 领域具备深厚技术积累，具备一定产业领先地位，给予公司 23E 80XPE，对应目标价 61.60 元，给予“买入”评级

风险提示

- 1 智慧教育业务拓展低于预期。智慧教育业务已是公司第一大主营业务，若公司区域级因材施教项目向其他区域复制推广不及预期，以及学习机销量低于预期，将对公司整体营收产生负面影响
- 2 智能硬件销量低于预期。智能硬件是公司 C 端市场重要业务，若智能硬件销量低于预期，将对公司营收产生负面影响

概要

- 1 AI 领域积累十余年，AI 商业化落地或加速公司在 AI 领域具备十余年的积累，形成了涵盖自然语言处理、语音识别、文字处理、图像处理在内的全方位 AI 技术布局，并以此为依托打造了涵盖面向企业及个人用户，金融、医疗、法律等多行业的产品及解决方案
- 2 我们认为公司凭借 AI 领域的技术、业务、场景积累，有望实现 AI 技术商业化的加速落地
- 3 多维度技术+产品布局，AI 应用落地实践经验丰富公司拥有丰富的 AI 技术积累，包括自然语言处理、图像处理、文字处理技术方面形成了系列的 AI 能力
 - a 公司积极运用文字、语音、视频等多维度的 AI 能力，不断改善用户体验。
 - b 从产品角度看，通过同花顺 AI 开放平台向个体客户提供智能语音、智能客服、智能外呼机器人等服务，并形成了文字识别、AI 理财师、企洞察等产品
 - c 形成了多种技术类型，多种行业领域，多种目标用户的完整产品线。公司通过全方位的 AI 技术及产品布局，积累了丰富的 AI 应用落地经验
 - d 具备良好用户基础+多个内容生成应用场景，商业化落地基础良好 AI 迭代需要大量的数据训练，用户则是数据的重要来源之一
 - e 公司在多年的发展中通过金融信息服务业务积累了良好的用户基础，截至 22H1 末同花顺金融服务网注册用户约 6 亿人，每周活跃用户数约 1,927 万人
- 4 我们认为用户既是推动 AI 落地的重要依托，也是提升数据质量，推动 AI 持续迭代的关键力量
 - a 此外，丰富的应用场景是新技术商业化落地的重要条件，而公司在内容生成相关领域，拥有辅助撰写分析报告、智能问答、智能客服、企业舆情分析报告生成等特色功能，具备丰富的内容生成落地场景
 - b 我们认为公司拥有大量用户及内容生成场景，具备良好的 AI 技术商业化落地基础
 - c 金融行业知识+多行业 AI 落地实践经验，或推动 AI 商业化落地加速公司深耕金融领域，垂直行业应用或进一步突破
- 5 我们认为专业知识积累或将成为 CHATGPT 为代表的 AI 技术突破垂直行业应用的关键因素，公司深耕金融信息服务行业近 30 年，在金融行业积累大量 KNOW-HOW，有望进一步提升 AI 技术在垂直行业应用深度
- 6 此外，公司积极探索 AI 技术的多行业应用，除金融外覆盖高校、企业、科研院所、政府部门、医院、通信运营商等领域。我们认为公司有望凭借良好的场景基础，推动 AI 技术的多场景落地

同花顺

- 1 AI 领域技术及落地经验积累深厚深耕十余年，技术积累深厚十余年积累，AI 领域积累深厚。公司自 2009 年推出 i 问财以来，AI 领域已有十余年积累。大致可分为三大时间段
 - a 萌芽期：2009 年，定位于金融垂直搜索引擎并提供专业股票信息搜索服务的 i 问财成立
 - b 基于此，同花顺在 2012 年上线了问答功能，i 问财也从金融垂直领域拓展出股票、基金、债券、保险、百科、法律法规等领域
 - c 技术突破：2015 年公司自主研发语音识别，为语音应用推广打下良好基础；应用加速落地：公司不断丰富 AI 产品线，推出包括全领域智能语音助理、智能投顾、智能外呼、智能质检、智能客服、智能转写等产品

- 2 拥有丰富的 AI 技术积累。近年来公司在研发方面投入逐年递增。研发人员从 2015 年的 836 人提升至 2021 年的 291 人，研发费用 2013-2021 的 CAGR 高达 33.87%
- a 在自然语言处理方面，同花顺布局语音产品十余年，深耕并优化语音技术，具有 ASR 自动语音识别技术、语音合成技术、NLP 自然语言处理技术
 - b 其研究团队的研发范围覆盖中英文语音识别、语音合成、声纹识别、情感计算等基础研究方向，具有先进的降噪算法模型和语音信号处理技术
 - c 在图像处理技术方面，同花顺拥有智能人脸识别和活体检测技术、身份验证+OCR 技术、智能质检技术
 - d 在文字处理方面，同花顺掌握知识图谱、机器阅读理解和智能审核技术、NLU 及大数据分析技术、文档解析技术、关键信息提取技术、知识关联和知识校验技术
- 3 多维度技术+产品布局，AI 应用落地实践经验丰富多维度 AI 技术优化产品，改善用户体验
- a 公司积极推进 AI 的应用，通过文字、语音、视频等多维度的 AI 能力运用，不断改善用户体验。同花顺 AI 开放平台主要面向个体客户提供智能语音、智能客服、智能外呼机器人等服务
 - b 文字识别产品基于金融领域领先的深度学习技术和海量金融数据，可提供身份证照识别、表格票据识别等功能
 - c AI 理财师依托大数据、人工智能等金融科技技术，融合专业投研能力的覆盖财富管理全流程的理财师智能营销工具。企洞察 app 企业版是融合了智能搜索、标签画像、产业链分析和风险监控等应用
 - i 证券预警系统则能够实现舆情监控、风险追踪、企业云等功能。公司的 AI 技术布局涵盖文字，语音，视频等多个维度，并且通过功能迭代，不断打磨 AI 技术能力
 - ii 已实现 AI 产品落地，形成丰富的产品布局。同花顺的人工智能产品覆盖语音、图像、自然语言等多种技术类型；以及覆盖金融、医疗、法律等多行业的解决方案
 - iii 分用户来看，同花顺目前已经开发了针对上市公司、中小企业、政务机关、医疗机构、金融机构、基金经理、理财咨询师、个人等多元主体的全方位 AI 产品布局
 - iv 拥有包括上市公司路演平台、智能电话机、AI 内镜助手、企洞察 app、文字识别产品、证券预警系统、AI 智能图像处理等全方位产品
 - d 内容生成彰显商业化潜力，良好基础助技术加速落地数据质量重要性提升，内容生成彰显商业化潜力 RLHF 打开新方向，数据质量重要性提升
 - i ChatGPT 发布于 2022 年 11 月 30 日，相比以往的 AI，展现了更自然更有逻辑的对话能力，带动 AI 行业关注度进一步提升
 - ii 从技术原理看，ChatGPT 模型相较于此前的 GPT3 模型，关键差异在于引入人类反馈强化学习技术（RLHF），能够在与人类标注者互动时通过人类的反馈进行强化学习
 - e 人类标注者可对初始模型生成的结果进行修改、比较和排序，帮助模型进行训练。从参数量看，GPT1/GPT2/GPT3/ChatGPT 的参数量分别为 1.17 亿/15 亿/1,750 亿/13 亿。我们认为 RLHF 打开模型迭代新方向，数据质量的重要性进一步提升
 - i ChatGPT 用户突增彰显商业化潜力，内容生成或成 ChatGPT 重要应用。据各公司官网，ChatGPT 从 0 到 100 万用户数仅花费 5 天，速度远超其他平

台

- ii 从活跃用户角度，据 Similarweb，2023 年 1 月期间，ChatGPT 平均每天有 1,300 万独立访客，超出 22 年 12 月一倍

4 用户数量的快速扩大反映了 CHATGPT 具备较强的商业化潜力。从具体应用领域来看，CHATGPT 是专为对话任务设计的模型

- a 作为 AI 内容生成器，能够根据既有的文字、图像、音频创作新的内容，从而大幅提高创作效率，为消费者提供个性化的内容，甚至打造丰富的内容生态，内容生成为 ChatGPT 当前重要的应用场景
- b 同花顺具备良好用户+场景+业务基础，AI 商业化落地有望加速具备良好用户基础+多个内容生成应用场景，商业化落地基础良好同花顺具备良好用户基础
- c AI 迭代的需要大量的数据训练，ChatGPT 通过 RLHF 方式训练，则需要更高质量的数据，由此可见在 AI 发展的过程中，质量高、数量大的数据能够为 AI 迭代打下良好的基础，用户则是数据的重要来源之一
 - i 公司在多年的发展中通过金融信息服务业务积累了良好的用户基础。据公司公告，截至 22 年 6 月 30 日，同花顺金融服务网注册用户约 6 亿人；
 - ii 每日使用同花顺网上行情免费客户端的人数平均约为 1,452 万人，每周活跃用户数约为 1,927 万人
 - iii 公司拥有良好的用户基础，大量用户一方面有助于形成庞大数据积累，另一方面在与用户交互的过程中，能够持续收到来自用户的反馈，有助于推动产品的迭代
 - iv 我们认为公司在网站端、移动端积累的大量用户，既是推动 AI 落地的重要依托，也是提升数据质量，推动 AI 持续迭代的关键力量
- d 业务涉及内容生成等多个环节，同花顺具备丰富落地场景。丰富的应用场景是新技术商业化落地的重要条件
- e 以 ChatGPT 为例，ChatGPT 具备较强的对话能力，因此在搜索引擎，邮件撰写等内容生成相关领域首先得到应用，商业化也有望率先于内容生成领域落地。同花顺同样在内容生成领域具备丰富的应用场景
 - i 如公司通过标签标记文章句式主干，结合 AI 技术自动生成金融类文本内容，辅助机构撰写分析报告；公司凭借自然语言技术生成语音资讯，向互联网用户提供智能问答、智能客服等服务
 - ii 并且通过机器学习不断总结经验，自动更新维护；此外公司还能够一键生成企业关系图谱、企业财务数据可视化图表、企业舆情分析报告、客户画像、晨报推送文章等多元化的内容，具备丰富的内容生成落地场景
 - iii 金融行业知识+多行业 AI 落地实践经验，或推动 AI 商业化落地加速深耕金融领域，垂直行业应用或进一步突破
 - iv 当前 ChatGPT 为代表的 AI 技术在通用领域展现了较强的能力，但距专业性较强的领域展开应用仍有一定的差距，我们认为专业知识积累或将成为突破垂直行业应用的关键因素

5 公司深耕金融信息服务行业近 30 年，拥有丰富的金融数据资源和形式多样、品种丰富的产品及服务，现有业务平台覆盖了证券市场中不同类型的客户群体，机构客户方面覆盖面广，其中证券公司覆盖率超 90%

- a 个人用户方面同样有庞大的客户基础。我们认为公司在金融行业积累的 Know-how，有望进一步提升 AI 技术在垂直行业应用深度
- b 公司触及多个行业场景，有望凭借良好的场景基础，推动 AI 技术的多场景落地

风险提示

- 1 市场交易量不及预期。公司的业务与市场景气度存在一定相关性，若市场交易量不及预期，可能导致下游需求不及预期，影响收入及业绩增长节奏
- 2 技术落地不及预期。若 AI 技术落地不及预期，可能导致公司的相关业务增长不及预期