保密

ChatGPT

纪要分享

久谦中台

二三年二月

观点总结

1 **ChatGPT 是社会发展的必然结果，2030 年数字化劳动力市场规模可达 1.73 万亿元**

a ChatGPT 催生路径 = 社会问题 + 技术迭代

i 2008 年全球金融危机 -> 云计算产业 -> 人工智能

ii 2020 年全球疫情 -> 经济压力 -> 企业降本增效 -> 加快数字劳动力发展（文字工作者、方案策划师、程序员等） -> NLP 技术赋能

b ChatGPT 技术路径 = Transformer 结构 -> 1,750 亿参数 + 巨大算力 + 单一模型 + 文字问答

i 冷启动监督策略模型：Transformer -> GPT -> GPT2 -> GPT3 -> ChatGPT

ii 训练回报模型：机器学习 -> 人类训练师 + 人工智能助手 -> 结果以质量排序

iii 使用强化学习策略：随机指令 + 初始化 PPO 模型参数 -> 更新模型

2 **中短期内 ChatGPT 对产业生态不会带来实质性的颠覆，产业链参与者仍有机会**

a 技术痛点：新数据不友好 + 预训练模型（数据集积累仅截至 2021 年）

i 新数据：未能建立和实时信息的连接

ii 预训练模型：如何保持实时更新

iii 产品体验：未达到理想状态（未必能超越垂直类产品）

b 商业痛点：不开源 + 商业模式不清晰 + 运营成本高

i To C -> To B（微软 -> 应用在 Office 中）

ii 潜在广告收入少 -> 短期内无法替代搜索引擎 = 俱进且并存

iii 开发成本 + 企业使用成本

3 **国内企业的入场机会和发展现状**

a 大厂：百度 -> 字节 -> 腾讯 -> 阿里 -> 自研趋势

i 百度（文心一言 -> ERNIEBot）：自主研发平台 + 文心大模型 + 预训练大模型积累 = To B（付费意愿稳定） + To C 产品

ii 字节：AIGC（短视频 + 图文） -> 数据 + 算法 = 语言处理模型

iii 阿里：AIGC（营销）

iv 腾讯：AIGC（广告 + 社交 + 游戏）

b 小厂：入局机会小，可作为大厂客户接入

i 技术积累薄弱 + 数据训练基础及经验不足 + 数据库及人力资源受限

ii 布局大厂下游 To B 应用端产品（需等待大厂开放 B 端应用接入）

c 阻力：技术 + 硬件 + 政策

i 中美 ChatGPT 发展仍存差距：模型 + 规模 -> 训练程度 -> 回答的逻辑性 + 完整度 -> API 调用 -> 企业生态

ii 芯片：算力瓶颈

iii 监管政策：国内引入 ChatGPT 政策尚未完善 + 规章制度尚未建立

4 **产业链机会及相关标的**

a 上游数据处理 + 下游智能应用

i 数据标注 + 算力 + 数据清洗 + 数据采集

ii 智能客服 + 聊天机器人

b 计算机：算法 + 算力 + 数据

i 算法：科大讯飞、拓尔思（NLP）、海康威视（图像识别）、云从科技（图像识别）、格林深瞳（图像识别）

ii 算力：海光信息（DCU）、寒武纪（AI 芯片）、景嘉微（GPU）

iii 数据：天瑞声

c 传媒：平台 + 光模块 + 运营商

i 平台：中文在线、视觉中国、昆仑万维

ii 光模块：中际旭创（800G 光模块龙头 + 最早放量 + 最高份额和订单 + 股权激励 + 估值水平较低）

iii 中国移动

5 ChatGPT 未来迭代和产业辐射

a 基础 = 纯粹创新精神 + 长期主义：创新型 + 投入 + 决心 + 顶尖人才储备

b 支点 = 算力 +GPU+ 商业模式

i 大算力 + 大模型

ii 芯片：国产化替代

iii 知识定制化：特定领域数据（医疗、司法）

iv 产业厂商合作：大公司 -> 训练大模型 + 小公司 -> 数据收集 -> 商业化

v 产业辐射：数据（收集 + 处理 + 清洗）、智能对话（客服、机器人）、创作（素材收集 + 写作）、虚拟现实、教育

目录

OpenAI 高管解密 ChatGPT

访谈日期：2023/2/8

具体内容

¶ GPT-3 是一种大型语言模型，被训练用来在给定上下文中预测下一个单词，使用 Transformer 架构

1 它很灵活，可以用于翻译、摘要、分类和问答等任务。GPT-3 的优势在于它的简单性和不需要专门训练数据集就能表现良好的能力

2 GPT-3 可以用于翻译任务，方法是提供比如“德语：英语”对的翻译样例（如果是德英翻译），或者像问人一样要求模型翻译给定的句子

3 尽管 GPT-3 主要是在英语数据上训练的，但仍然能够在翻译任务中表现良好，因为它能够通过提供的样例中的模式，并利用自己的一般语言能力产生翻译

- a GPT-3 也可以用于摘要和问答等任务。GPT-3 在商业应用中也取得了成功，如文本生成和问答。它明显比早期版本的 GPT（规模）更大、（功能）更强大，训练的数据也更多
- b 它被用来生成创意写作任务的起点或变体，如产品描述，并已与 OpenAIAPI 集成，使开发人员更容易使用
- c API 允许用户对 GPT-3 进行特定任务的微调，包括设置学习率和数据的过渡次数，以及选择模型大小

4 PeterWelinder 现任 OpenAI 产品与合作伙伴副总裁，负责 GPT-3 的运行和其他业务，在此之前，他曾是 OpenAI 的研发主管。使用 GPT-3 解决现实世界的问题

¶ Peter，上次我们谈话时，我记得你在 OpenAI 做研究，但现在我们发现你是 OpenAI 的产品和合作伙伴关系副总裁，我很好奇这意味着什么？你每天都在做什么？

1 我今天所做的与我做研究时完全不同，对我来说，做研究一直都是为了解决最困难的问题，以便真正对世界产生某种影响。我个人更倾向于研究的最终目标，而不是研究本身，做研究真的很有趣，你知道，深入研究，探索事物，最后总是有一个目标

2 GPT-3 发生了一件令人兴奋的事情……当我开始在 OpenAI 工作时，我做的很多事情都是机器人方面的。对于机器人技术来说，你在实验室里能做的事情和你在现实世界里能做的事情之间还有一些差距。使用 GPT-3，当我们在 GPT-3 中得到第一个结果时，很明显我们有一些东西可以开始应用于现实世界的问题，而不仅仅是做酷炫的演示

- a 当我从事机器人工作时，我们最后得到的是一个非常酷的机器人手解魔方的演示，但每个人的家里并不具备部署它的条件
- b 即使它足够强大，我也不知道它对解决魔方有多大用处，这是一种非常昂贵的方法。但是有了 GPT-3，我们有了一个语言模型，你现在可以应用它来解决各种不同的问题，从翻译到总结，再到分类和问答等应有尽有，这是一个非常灵

活的模式

c 所以，我们要做的就是看看这个模型来解决现实世界的问题是否足够好，对我来说，这是一个非常有趣的领域

3 当你拥有这项非常强大的新技术，有可能改变很多事物的工作方式时，这一切都是为了找到合适的方法来解决问题，看看你如何利用你工具箱里的工具来解决这些问题。不同的是，作为研究人员，我所做的是提出正确的基础和正确的方法来衡量进展。当目标非常遥远时，你需要想出这些玩具的方法来评估进展

a 现在，就像客户告诉我们“嘿，我正在尝试将 GPT-3 应用到这个用例中”，但它不起作用或太慢等诸如此类的事情，这些问题要具体得多

b 我的日常，现在更多的是建立一个团队，用我们在 OpenAI 开发的技术来解决这些现实问题

¶ 当你将 GPT-3 与其他用于大型语言模型的方法进行比较时，这似乎是一种趋势。你是否注意到它在工作方式上有哪些关键差异，采取某种方式是否有所不同？

1 这是一个很好问题，我认为我真正喜欢 GPT-3 的地方，以及我认为它与众不同的主要方式是 GPT-3 所做的一切都非常简单

2 GPT-3 是一个大型语言模型，大型神经网络。它使用的是谷歌几年前推出的一种非常流行的 Transformer 架构，如今，它基本上为所有不同的语言模型提供了支持，而且它也开始进入其他领域，比如计算机视觉等

3 GPT-3 的设置非常简单，它可以有一些上下文，你可以看看文本的历史。如果你正在读一本书，你可以看一页或一段文字，然后它试着预测下一个单词，这就是 GPT-3 的训练方式。它只是训练了来自不同来源的大量文本，大部分来自互联网。它只是一遍又一遍地训练，根据它看到的一些单词，预测下一个单词

4 你可以从几个单词开始，但当我们今天训练这些模型时，我们训练它们的数量级是一千或几千个单词，你可以回顾这 1,000 个单词，然后试着预测下一个单词。所以设置非常简单，你只需要在这些庞大的文本数据集上训练它，以便继续预测下一个单词，并在这方面做得非常好

a 我认为 GPT-3 的令人惊讶之处在于，如果你这样做，然后你把模型变得非常大，这让它有巨大的学习能力，然后它就会非常擅长以前你需要专门模型的一系列任务

b 以前如果你想进行翻译，你就需要一种专门翻译的神经网络，或者如果你想做总结，同样，你会以特定的方式设置你的网络，然后只训练它完成总结任务

c 我们在使用 GPT-3 中发现，你实际上在一些基准测试中获得了非常接近最先进的表现，这些基准测试包括总结、翻译、问题回答等等

d 该模型使用的是一个刚刚在互联网上训练过的模型，它不专门执行任何任务，而是能够以与阅读文本相似的方式再现文本。将 GPT-3 应用于翻译任务

¶ 实际上，如何将其应用到翻译任务中，你如何把“预测下一个单词”变成一个翻译？

1 在很多其他的大型语言模型中，都有一些特定的步骤，你可以对一段文本进行编码。所以你会在神经网络中创建一些表示

2 然后你会有一个解码器来接受它，然后用它来写一些句子。例如：如果你做翻译，你会把它编码成某种表示，然后你的神经网络会有一个单独的部分来接受这种表示，并尝试输出你想要的东西，输入可能是一个德语的句子，输出的可能是一个英语的句子，而且，你知道它是专门为此训练的

   a 那么对于你的问题，你如何处理 GPT-3 呢？最简单的方法是：你可以提供一些例子，说明翻译可能的样子，仅以纯文本形式，你会写“德语：”和一些德语句子，然后是“英语：”和一些英语句子
   b 你可能只提供一个例子，那么这个称为一下（one-shot），你可以提供一些例子，基本上都是“德语或者英语”的一些例子，然后你可以输入你想翻译的新句子，这就是所谓的多下（Few-Shot）训练

3 如果你有几个例子和模型，只要看看它现在在其上下文中看到的模式，它可以产生一个翻译。这是一个非常简单的设置。基本上，我认为告诉 GPT 该做什么的方式有点像你告诉人类做同样的事情。比如，如果我给你写电子邮件，“嘿，Lukas，我想让你翻译一些句子”

   a 我会告诉你：“请翻译这些句子”，我可能会提供一些例子来让你了解一下它的语气，比如：我想要更正式的翻译，还是更随意的翻译等等，你会发现其中的规律，给你一个德语句子（我不知道你懂不懂德语）你就能把它翻译成英语
   b 现在有了我们最新的模型，你甚至不需要提供这些例子，你可以像问人一样问模型，比如，“嘿，把这个句子翻译给我听”，或者“总结一下这篇文章”
   c 我们刚刚发现，这就是人们想要使用模型的方式。我们让他们做了更多这样的工作，但就是这么简单，你只要告诉它你想做什么，它就会尽最大努力去做

¶ 你是主要致力于训练模型使用多种语言，还是主要是英语？语料库从何而来？

1 实际上我们做的正好相反。最初，当我们训练 GPT-3 时，我们一致努力不用英语以外的其他语言来训练它。事实证明，即使这些模型是巨大的，在你的数据集组合中也需要权衡取舍

   a 如果你先用英语训练它，然后再用其他语言训练它，它在英语任务中表现就不那么好了，最终当我们训练它的时候，我们想看看，它在更通用的能力上能有多好？
   b 我们不太关心翻译，因此，每当我们输入额外的语言时，这只会以擅长用英语执行其他任务为代价，比如回答问题、总结等等
   c 但结果是，即使明确地试图过滤掉大多数其他语言，也可能有一小部分数据是其他语言的。即便如此，该模型在翻译方面还是非常出色，在许多翻译任务中，它接近于最先进的技术

2 我的母语是瑞典语，但我现在已经不会用瑞典语写作了，因为我从来没有这样做过。我现在做的是用英语写它，然后让 GPT-3 来翻译给我，这只是我的观点，它不会变得完美，我需要调试一些东西，但它出奇地好，而且模型中的瑞典训练数据量非常非常少

3 我们一直在不断更新我们的模型，让它们变得越来越好，所以现在我们引入了越来越多的语言数据，因为我们已经找到了如何以更优化的方式进行这些权衡。但是，一开始我们想要的是相反的，我们只是想把英语学好

¶ 是预测单词还是一次预测一个字符？这是怎么回事？

1 都不是，它实际上是在预测一种叫做符号标记（Token）的东西，这就像“单词的一部分”也许可以这么想，最常见的英语单词，它们由单个符号标记。我们有大约 50,000 个这样的标记，我们将它们映射到字符序列上，结果就像“hi”或“the”这样的常见单词最终会成为一个标记

2 但如果你有一个更不常见的词，比如“百科全书”之类的，你可能会把它分解成两三个符号，这就像单词片段，只是让这些语言模型更容易、更有效地使用文本

3 原则上，你也可以在字符层面上这么做，但它会变得非常低效，你知道，这就是这个领域可能正在改变的地方，最终，它将不止在字符层面上做到这一点

¶ 但我认为这会让学习外语变得非常困难，比如，亚洲语言是不可能的吗？如果他们有更多的符号，或者我猜你可能会说，他们已经为你做了标记化，通过使用更多的字符来编码更大的含义

1 是的，训练标记器（Tokenizer）的方式肯定会对不同语言的性能产生影响。通常这两件事分两个不同的步骤进行训练

2 你可以在某些数据语料库上训练你的标记器，然后在其他一些数据集上分别使用该标记器训练你的模型，为了让你的模型真公众号新价值人正擅长不同的语言，你还需要在多种语言上训练该标记器

3 肯定是使用其他语言的成本更高，一个德语单词最终会变成更多的符号，因为我们训练它的次数少得多。而英语非常高效，很多单词都是一个单一的符号，所以这使得它在其他语言上更糟糕，而且更昂贵

¶ 我能把一些东西翻译成日语吗？GPT-3 也能做到吗？

1 是的，我记得我们的一个日本用户的评论，他们非常喜欢使用 GPT-3 在英语和日语之间翻译技术文档，因为他们发现 GPT-3 在技术文档翻译方面比谷歌翻译要好得多。这大概是一年前的事了，谷歌翻译现在可能更好，但根据我们拥有的数据集，这可能只是一个偶然的事情

2 实际上，关于 GPT-3 的翻译功能，真正酷的事情是我们并没有在显式的输入和输出对上训练模型，翻译的文本片段，就像你通常所说的“对齐的文本片段”一样

3 只是看到了很多日本人，它看过很多日本电影，也看过很多英语电影。不知怎么的，通过学习如何预测下一个单词，已经有足够多的小文本、博客文章或其他东西，作者在日语和英语之间切换。可能会对一些句子进行翻译，在那里它找到了映射，然后以某种方式有一个足够好的表示，然后推广到任意的翻译任务

4 对我来说，这太神奇了，它只是通过阅读大量的英语文本，大量的日语文本，然后可能就像在所有的数据中找到一些对齐的对，它就能够进行翻译，这对我来说太疯狂了

¶ 真是太神奇了，这种性能与早期版本的 GPT 有明显的不同吗？比如在 GPT-3 中是否发生了什么，OpenAI 认为“好吧，我们可以将其用于现实世界的商业应用”？这是它需要达到的性能水平吗？

1 是的，我认为 GPT-2 和 GPT-3 之间最大的区别是：它被训练在更多的数据上，它

是一个更大的模型，大概差了两个数量级。最初的 GPT-2 大约有 15 亿个参数，而 GPT-3 最大的模型有 1,750 亿个参数，它上升了两个数量级，而且由于它是一个更大的模型，它也需要更多的数据

2 令人惊讶的是，这就是从感觉它相当愚笨到可以与之互动的原因，像 GPT-2 有点酷炫，但大多数时候也感觉它非常愚蠢，我认为在 GPT-3 中，它有时会表现得出乎意料的好。不要误解我的意思，GPT-仍然会犯很多愚蠢的错误，但在某些任务上，它可能有 30-50%的时间是正确的，有时甚至更好

- a 就好像突然之间在你需要抽样和尝试任务之前，也许每隔 20 次你就会看到一次，“哦，这个看起来不错”。有了 GPT-3，它开始每三次发生一次，或每两次，或每五次发生一次，你会说，“哦，天这实际上是……”
- b 对于诸如总结文本之类的事情，我们有一个例子是用二年级学生的风格总结一段文字，令人难以置信的是，该模型能够简化单词，获得一段文本的要点等等，再说一次，它不是完美的，但它真的很好

3 显然，我们有很多学术基准（academicbenchmarks），你可以运行这些模型，你可以看到它在学术基准上越来越好

4 但当你想要创建一些东西的原型时，这是一种完全不同的感觉，不同的是，现在很容易得到一些运行良好的东西

5 这是为什么我们决定，“嘿，现在它看起来很有用”，GPT-2 看起来没有那么有用，但是 GPT-3 对于所有这些任务，我们觉得“好吧，它已经足够接近最先进的技术了”，如果你有一个专门的模型或其他什么，一个聪明的程序员应该能够将其应用到他们所拥有的任何任务中，这就是我们设置的 API 验证的内容

¶ 你真正引以为豪的用例，它到底在哪里起作用？你能不能给我们指出一些地方，让我们可以在商业环境中与之互动？

1 当然，我认为最让我们感到惊讶的是文案和问题回答，一般来说是创意写作。在文案方面，当时有很多公司开始在我们的平台上进行开发，有些公司像：Copysmith 是第一批；CopyAI；还有 Jarvis……还有很多这样的公司。他们的做法非常聪明，因为他们意识到，当你使用 GPT-3 来完成某些任务时，它并不完美

- a 时不时的，你可能会得到一些没有意义的东西
- b 但如果你在做文案工作，比如你想根据产品的某些属性写一些吸引人的产品描述，比如鞋子，可能是鞋底的类型，颜色，鞋子的一些其他属性，你想写一些真正吸引人的东西，那么作为一个人，你面临的问题是你陷入了某种写作瓶颈，我该从哪里开始呢?
- c 这些公司开始做的是他们采用 GPT-3，他们用它来生成一些起点或者一些产品描述的变体。你会发现，通常情况下，如果你生成五个这样的例子，其中一个看起来会很好，你可以把它作为你的起点，你可能只是接受它，或者做一些小的调整
- d 这几乎是一种帮助人类创造力的方式，你知道吗，我觉得这太酷了

2 作家们会告诉我们，“嘿，我已经试着写这本书半年了，我总是陷入写作瓶颈。然后我开始在使用 GPT-3，现在我花了两周时间完成了整本书。”当你陷入困境时，它可以创造一个有趣的故事情节

- a 作为一个有创意的作家，你开始探索，就像“好吧，我没有想过这个角色会往

这个方向发展，但让我们来探索一下吧”。然后它就变成了一个更有趣、更吸引人的过程

b 这几乎就像一个人，现在我们有一个头脑风暴的合作伙伴，你可以把它应用到所有这些不同的任务上。我觉得非常酷的是，我发现很多公司都在利用这一点，创造你以前做不到的新体验

c 我认为这是非常令人兴奋的。我觉得回答问题也非常酷，但是这个问题出乎我的意料。我认为我们不会预料到这是一个如此大的用例。使用 OpenAIAPI 微调 GPT

¶ GPT-3 的优点之一似乎是它可以开箱即用。对于一些团队，如果出现问题，他们会担心该怎么办。我想我很好奇，你通常与公司内部的 ML 团队合作，还是更多的工程师认为这里的好处是，他们不必弄清楚机器学习是如何工作的，以获得自然语言处理的好处，或者你是否倾向于将其与 ML 团队集成到一种更大的 ML 工作流中？

1 我得说，这是一种混合，我们有多个机器学习团队。他们已经有了自己的模型，他们会在网上下载模型等等，他们会根据任务对模型进行调整

a 然后他们找到了我们的 API 并开始使用我们的 API 做同样的事情，结果证明你可以从我们的模型中获得更好的性能。就像我们所拥有的最大的模型或最好的模型都没有开源版本，对于很多任务来说，这是最有效的方法

b 但我认为，我们的大多数客户可能更倾向于另一个阵营，即“真正聪明的开发者”。当我说“开发人员”时，这是一个相当广泛的群体

c 从程序员到工程师，从设计师到项目经理。许多人告诉我们 OpenAIAPI 是他们进入编程的原因，因为他们从我们的游乐场得到了非常好的结果，在那里你可以与我们的模型交互

d 他们有了想法，就开始学习如何编码，并接触到像 BubbleIO 之类的无代码工具。这真的降低了障碍，你不必成为一名机器学习专家，也能从这些模型中得到非常好的结果。你只需要善于迭代并弄清楚如何向模型编写指令

2 这有点像每个人都能成为管理者，如果你想让你的员工按照你的想法去完成任务，你就必须给他们很好的指导，这和这些模型非常相似。比如，如果你不明确你的任务，你就会在输出中得到非常高的差异，但是，如果你真的很擅长具体说明，甚至提供几个例子，那么你就会得到非常好的结果

3 这不是一种机器学习技能，这几乎更像是一种任务规范，管理技能，我觉得很多人都能很快学会

4 我真的很兴奋，看到这么多人都能接触到这些模型，以前好像只有机器学习博士学位才能使用

¶ 我觉得我听人说过一个叫做“提示工程师（PromptEngineer）”的新角色可能与此有关，清楚如何提示 GPT-3 让它做你想让它做的事情

1 这个很有趣，因为早期，当我们有第一个版本的 API 时，我们有一个非常聪明的人，他是一位世界知名的作者，也是一个程序员：安德鲁·梅恩（AndrewMayne）

2 他是该 API 的早期用户之一，他的内部名称是“提示耳语者（PromptWhisperer）”，或“GPT-3 耳语者”，他真的知道如何精心设计提示以

获得最好的结果

3 因为它是在互联网上训练的，你需要把你的思想放在这样的想法中，“互联网上的文本是如何开始的”，如果你想要一个真正好的食谱，你必须开始用食谱书或美食博客之类的东西来写作，这并不是说你可以让模型做你想让它做的事。我认为，这其中有很大一部分开始是这样的

4 你真的必须善于理解 GPT-3 的复杂性，并设计出真正好的提示

a 在我们推出后的一年半时间里，我们看到人们在这方面有很多困难，所以我们开发了一套新的模型，我们称它为 InstructGPT。这实际上就像前段时间一样，它成为我们 API 中的默认值，我们称其为 InstructGPT 的原因，是因为它只提供说明

b 所以我想说，提示设计现在已经不那么重要了。你可以告诉模型你想让它做什么，并提供一些例子，还有一点关于格式可能会影响你提供示例的方式等等

c GPT-3 在这方面非常强大，但有时它确实有点问题，一些调整很重要。但我想说的是，与一年前相比，现在已经不那么重要了，我的希望是，它变得越来越不重要，而是变得更有互动性

¶ 你对模型还启动了微调的功能，这个想法是什么，它在什么地方有用？

1 GPT-3 令人惊讶的是通过零下（zero-shot）就得到了非常好的结果。你只需要提供一个例子，或没有例子，只是说，“嘿，把这个句子从德语翻译成英语”就可以了，或者你提供了几个（few-shot）示例，比如几对德语和英语实例

2 只需几个（few-shot）示例，你就可以得到令人惊讶的好结果。但这实际上意味着准确性是非常依赖于具体任务的，对于一些任务，也许 30%的时间你得到的输出是可以接受的，而对于其他更简单的任务，你可能 70%的时间都能做到

3 当它不是每次都很好时，你必须非常聪明地在你的产品中暴露它。这就是为什么，比如它对很多文案公司都很有效，你可以只提供一些例子，你知道其中至少有一个是好的，这就是用户所需要的。但是通过微调，你能做的基本上你可以自定义你的模型，你可以为它提供更多你希望它执行的输入和输出示例

4 如果你想做翻译，或者如果你想总结文章，你可以提供几百篇已经做过人工编写总结的文章例子，你可以更新 GPT-3 来更好地完成这项任务

5 你不能把所有这些例子都放在你的提示中，提示符的空间有限，但是通过微调，你把这些例子转化为神经网络的连接，转化为神经网络的权重。在某种程度上，你就像有了一个无限的提示，你可以提供尽可能多的例子

6 显然，示例越多，微调所需的时间就越长，成本也就越高。但微调基本上是一个概念，取一堆输入和输出的例子，把它们放入模型中，然后得到一个模型的新版本，该版本非常适合你提供例子的任务

a 事实证明，只需几百个例子，或者大约 100 个例子你就能显着提高准确性

b 我们有很多客户使用过它，就像 KeeperTax 一样，他们正在分析交易以找到这些税收注销之类的东西，他们所做的是提取相关的文本片段，进行分类等等。例如，他们对模型进行微调，并通过微调模型得到了更好的结果。我们在客户身上一再看到这种情况

c 他们可以得到非常好的结果，这些结果通常对于原型来说已经足够好了，但是为了让其达到足够高的精度以将其投入生产——通常超过 90%或 95%或 99%，

使用他们拥有的数据集对模型进行微调，这样一直进行下去

d 这可以让他们比以前启用更多的应用程序。我们只是让这种微调变得很简单

¶ 我想对你来说，你们可以调整的参数是什么，因为你描述的方式，听起来好像没有任何参数，参数在这里如何参与呢?

1 对于你关于参数的问题，我们试图在我们的 API 中使它变得非常简单。我们试着让默认值非常非常好

2 一般来说，你可以通过微调获得非常好的结果，而根本不需要过多地修改参数，但有些参数会有所不同。例如，你可以设置学习率，这是你在每个学习步骤中更新权重的程度

3 你可以设置你想要通过多少次数据的内容，事实证明，如果你把数据调整太多次，你就会对数据集进行过度拟合

4 这些 GPT-3 模型非常大，通常只需要对数据进行 2 到 5 次迭代就能得到非常好的结果，如果你走得更远，你有时会过度拟合。还有更高级的参数，但我有点喜欢玩一点你想训练它的时代数量和他们的学习率，这让你达到了 90%的目的，如果你开始摆弄其他参数，它不会给你更多

¶ 这是考虑将参数留给其他人的想法吗，你能从摆弄参数中得到乐趣吗?

1 说实话，如果这是完全自动的，我会很高兴，也就是说，我们确实有一些更注重研究的客户，他们真的喜欢摆弄，所以我认为我们很难删除它

a 但是，就像我说的，我们有两大阵营的用户：研究人员和开发者，开发者总是告诉我们：“嘿，我只想要一个按钮，我只想要最好的模型出来。”然后很多研究人员想要摆弄更多的参数，我想我们可以长期满足双方的需求

b Boris（Boris 是一个 ML 技术人员），我不知道你把自己归哪一类了，你做了一些惊人的、漂亮的演示，你也喜欢调整参数，我很好奇你使用 GPT-3 模型的经验

c 我当然喜欢有一个好的默认值，因为最初你真的不知道你应该在它上面改变什么，假设你选择了错误的参数，结果什么都没用。可不是什么愉快的经历。所以我喜欢如果你不选择任何东西，它就已经很好了。然后，我真的很喜欢调整参数，看看“好吧，会有什么效果”并试着用直觉来调

2 除了 Peter 提到的参数之外，还有两个参数也让我很感兴趣，你可以决定微调哪个模型，有不同尺寸的模型。如果你使用一个更大的模型，也许你的 API 会慢一点，但是你的效果会更好。也许有时你不需要它，也许有时确实需要，所以我想看看我使用哪种模式的效果

3 我还喜欢看到“我可以给出多少个训练样本”的效果，就像我只给出 20 个样本，而不是 100 或 200 个，因为这样你就能知道我的模型在我开发一个更大的数据集时会变得更好。我喜欢摆弄各种各样的参数，看看基于这些参数能做出什么样的预测

a 对，最后一条，其实非常重要，我认为这是我们一遍又一遍地给人们的最常见的建议之一

b 这就像从一小组例子开始，然后把它翻倍，看看你能得到多少改进。如果你将

训练数据量翻倍，那么你就会看到错误率的线性改善

c 如果你有 10%的错误率，你把训练数据翻倍，你可能会得到 8%的错误率。然后再翻倍，错误率降至 6%等等

d 如果你能看到这种趋势，那么你就会突然有一种感觉，“就标记更多的数据等等而言，我需要花多少钱才能得到我想要的结果”等等。这是一件非常强大的事情

¶ 训练这些模型的结果是否可重现？每次对它进行微调时，有多少可变性？如果你对相同的数据进行两次不同的微调，你会得到相同的模型吗？

1 原则上，你可以把它设置成非常可复制的。如果你在同一天训练，基本上你在训练时想要做的是，在每次训练迭代中，你有一批数据

2 比如一些例子，你实际上可以把 API 设置批量大小，每次更新需要多少个示例。我认为它默认是 32 或类似的东西，当你这样做时，你还希望对数据进行随机排序

3 你希望对训练数据进行随机抽样。只要你在训练中保持这些随机化一致，你最终会得到相同的模型。这将是相当可复制的。唯一需要注意的是

4 在实践中，即使是推论，这也是正确的。我们有一个叫做温度（Temperature）的参数，你可以设置输出的可变性。温度越高，变异性就越大，即使你把值设为 0 也不能保证你会得到完全确定的输出

a 在这些大型模型的 GPU 中，有足够多的噪音和一些奇怪的浮点运算等等，都很难保证完全确定性的决定

b 很多人问我们这个问题，答案总是这样，“很不幸，我们不能提供这个，但你可以得到一些公平的东西。”但是你应该让你的实验足够强大，这样你就不用太在意决定论了。OpenAIAPI 背后的工程挑战

¶ 我认为，从操作上讲，让每个人都有自己的微调模型比每个人都使用符合相同模型的 API 在基础设施方面面临的挑战要大得多。允许这种情况发生是一项艰巨的任务吗？比如，当人们开始使用不同的模型时，你需要换入和换出不同的模型吗？

1 刚开始的时候，我们做微调的方式基本上是在某种程度上。你几乎租了一组运行模型的 GPU，在某种程度上，对于一些最早期的微调客户

2 我们基本上是按 GPU 小时收费的，比如每小时，他们使用模型的次数。甚至从一开始，我想在推出 API 后的六个月内，我们就有一些精选的客户，他们有微调过的模型和类似的东西，这就是它的工作方式

a 问题是，如果你想尝试一些新的东西，GPU 的时间是很昂贵的。你不会真的想要花钱去保留一个 GPU，哪怕只有不到一个小时，这一切都累积得非常非常快

b 我们只是设定了一个目标说“好吧，一旦你微调了你的模型，你应该立即能够使用那个模型，你只需要为推理时进入它的 token 付钱”，就像无论你在提示符里输入什么

c 要使这种体验真正出色，这无疑是一个巨大的工程挑战。你只需开始微调，当它完成时，得到一个微调的模型名称

3 现在你可以在 API 中使用那个模型来立即得到一个结果，而且你不会按小时或其

他方式收费，你只会以相同的方式为 API 收费。这真的很棘手，我们在 OpenAI 有一个了不起的工程团队，他们真的想出了很多技巧来平衡这些模型的最终位置，并以正确的方式缓存它们等等，以创造一个很棒的体验

a 我很好奇你是对整个模型进行微调，还是只对部分模型进行微调，让它更有效率

b 我们用了很多技巧来实现这一点，我们一直在努力寻找新的方法。如果你想对整个 750 亿个参数模型进行微调，这是有挑战的。它可能会变得非常昂贵和困难等等，有一些技巧可以让它更快

¶ 你觉得你和所有使用 GPT-3 进行自然语言任务的每个人之间的区别是模型本身的质量和性能吗？还是其他原因？是关于集成，还是生产中的监控，或者类似的东西？

1 当然，我们在构建 API 时所关注的关键事情是最重要的是模型的能力

2 其次，你需要有快速的推理能力。在我们创建 API 之前，对于语言模型，没有人关心推理。每个人都关心你能多快地训练他们，因为这才是最重要的

3 因此，你可以在一天结束时解决基准测试问题。我们做了大量的工程设计来让推理超级超级快。我还记得在最初的几个月里，我们将 API 的第一个原型交付客户开始使用，我们将推理速度提高了 200 倍之类的

a 我们做了很多努力来让它超快。第三件事是围绕安全的事情。我们投资这些 InstructGPT 模型的原因之一是，我们看到有时你可以得到出乎意料的模型输出。例如，你可能写了一个非常无辜的句子

b 但由于某些原因，它可能会变得非常黑暗，或者你可能会以不同的方式得到一些有偏见的输出。使用我们的推荐指令的模型，默认情况下，它们的行为更符合预期，但你也可以以更好的方式指定行为

4 事实证明，当安全和能力齐头并进时，当你能更好地控制它时，它就会变成一个更好的产品。这些肯定是我们一直关注的事情，我认为我们在这方面做得比现有的其它替代方案要好得多

5 最后，我们非常关注的事情是让它使用起来非常简单，事实上，你不需要加载模型，你只需要调用一个微调模型，只需要一行 Python 来调用 API，这也是我们的核心，我们希望每个人都能轻松使用它

国产 ChatGPT 何时问世?

访谈日期：2023/2/7

具体内容

¶ 事件

1 根据公开新闻报道，百度对标 ChatGPT 的 AI 产品中文名字叫做文心一言，英文名 ERNIEBot，3 月完成测试，对公众开放

2 目前还在做上线前的冲刺，时间有可能提前。百度集团-SW 涨幅超 15%，此外其他百度系公司表现亮眼，应用公司表现亮眼，行情正往两头演绎，优秀的大模型+基于大模型的创新应用场景

¶ 国内互联网大厂进度

1 百度：百度布局较早，有自主研发的深度学习平台，有文心大模型，在预训练大模型方面有不错的积累，在 AI 发展方面把握先机

- a 百度的文心大模型具备多功能，可以进行文本生成，内容提取，摘要生成，观点归纳、图片绘画等。和 GPT 很像，一般情况下，AIGC 优先考虑 ToB，再考虑 ToC
- b 因为 ToB 的商业群体比较稳定，付费意愿也比较稳定。百度计划同时推出 ToB 和 ToC 的产品，并先发布 ToC 的 demo

2 字节：已经开始布局，主要是 AI+内容，比如自动生成投稿和辅助写作，在今日头条上利用 AIGC 生产内容，目前 AIGC 整体的生成质量的内容还是较好的，要好于普通的 UGC，但和 PGC 相比还有所欠缺。抖音方面也有应用，通过 AI 的模式来生成短视频，比如一些图文类的短视频的

3 阿里、京东等电商类平台：在智能客服领域有布局，其次是 AI+营销，例如阿里巴巴，可以结合商品，自动生成高质量文案描述商品，提高营销效率

4 腾讯：以广告为主，支持广告智能制作，以 AIGC 技术生成广告文案和视频，降低了制作成本，目前市场规模快速增长，未来 5 年内 AIGC 产生的图片的占比预计会达到 10-30%

- a 前期可作为 UGC 和 PGC 的辅助，帮助广告主设计文案
- b 到后期就是 AI 技术整体的发展，后期可能是有望代替人工的工作

¶ 百度 ToC 产品的进度如何，使用体验如何?

1 百度有文心大模型的基础，去年 ChatGPT 刚发布后，他们基于对话的语料，做了一个类似的新模型，是多轮对话的模型和百度搜索引擎相结合——用户问一个问题 AI 会给一个答案，同时搜索引擎会基于这个问题做一些相关的补充，比如答案的来源和链接

2 如果和搜索引擎结合起来后，整体使用效果还是可以的，因为结合后，不涉及到

特别多轮对话，一般我问一句，它回一句，就结束了。至于多轮对话容易遗忘的问题，可能需要在后续的优化过程中，重点考虑怎么捕捉更远的信息，怎么捕捉用户长期讲话的意图

¶ ChatGPT 会替代传统搜索引擎吗？

1. 短期内不太可能取代传统的搜索引擎，ChatGPT 会给出一些看似有道理但实际是错误的回答，可信度不是很高
2. ChatGPT 对于新数据不太友好，未能建立和实时信息的连接，目前预训练模型如何保持实时更新，是一个大问题
3. ChatGPT 的训练成本很高，付费过多，可能用户放弃使用；但不付费，成本压力过大，长期可能在训练成本或者推理成本上都做了比较多的优化以后，再看对搜索引擎的替代
4. 可能短期内还不能替代，但长期不好说。至少可以跟百度的模式一样，搞双引擎的模式

¶ 国产 ChatGPT 何时问世？除了百度外，国内还有其他公司可以推出类似的产品吗？国内其他大厂，比如腾讯、字节等，会想着在短时间内做出来类似 ChatGPT 的产品吗，抢占先机，形成类似微软对谷歌的卡位？

1. 小公司机会比较小，这是一个技术积累的工作，需要有数据训练的基础和经验，需要资源和人力的投入。小厂很难做出来，因为成本太高了
    a. 小厂更适合去接入这些大厂的模型，成为大厂的客户，然后做这些模型的应用，比如 AI 绘画等，对接 C 端消费者
    b. 国内的大厂比如字节、腾讯、阿里有机会。字节已经开始在做语言处理模型，目前在数据和算法方面的积累都不差
2. 字节其实也要发大力发展搜索，包括培养用户的搜索心智。字节也希望推出新的产品，从而抢占先机，实现它在搜索领域的一个超车。目前字节还处于大模型的训练和调试状态，没有产品的具体规划
3. 但如果能做出来，还是对字节搜索领域的地位有积极影响，我认为字节跟百度在搜索领域，会有很多的交叉的冲突，也一直在大力发展搜索领域，所以是有可能做出类似的产品

¶ 国内会引入 ChatGPT 嘛（考虑到有一些内容指向性的问题）？如果 Bing 引入了 ChatGPT 对于中国搜素市场的影响？

1. ChatGPT 目前会有一些伦理层面的问题，目前国内的监管政策还不是很全面，相关的法律法规还没有健全，还有很多这种一些细节的东西，短期内我们的规章制度其实也没有覆盖到
2. 总的来说，我认为 ChatGPT 的 ToB 端可能会引入，国内的小公司可以应用，目前 ChatGPT 的 ToB 端因为成本、优化等问题还没有开放，如果 ChatGPT 的 ToB 端开放，国内的一些小型创业公司可能会接入，并去做下游应用端的产品

3 未来接入微软的 Bing 后，其实对搜索是会有一定的冲击的，首先我们考虑一下用户的猎奇心理，肯定会有大量的用户愿意去用

4 能够产生大量 DAU，如果效果是比较好，这些用户是愿意留下来继续使用它的，久而久之其实是会改变到用户的搜索习惯

¶ 字节内部目前在类 ChatGPT 产品方面的规划？

1 从我们看字节对搜索的重视程度，搜索现在也是一级部门，对搜索的重视程度很高，因为搜索在现在在抖音、今日头条的重要性上很高，本次也是集合了几个核心的部门，组成小团队来做模型

2 目前来说可能还没有产品的计划，虽然是比百度晚一些，但后续要看产品的效果和用户的体验，先发后发的影响不是很大，需要看后续的发展

¶ 谷歌最近在财报上说，他们的 LaMDA 模型可能在近期推出类似 ChatGPT 的功能，如何看待谷歌在语言模型方面的积累？

1 谷歌的技术积累很不错，团队都非常优秀，模型积累很好

2 数据方面，谷歌天然就有很多搜索引擎的数据，算力方面也不用担心，很多技术都是谷歌推行的。相对来说，谷歌研发类似产品的可行性非常大，而且成功概率非常高。而且它的效果也是值得我们期待的。如果效果比 ChatGPT 好的话，那也算一种后发优势

¶ 如果未来出现很多的模型，这些模型都基于差不多的数据去训练出来的，又有很多应用去基于这些模型去开发不同领域的垂直应用。整个环节的价值量最大的地方会不会在公有云跟硬件厂商。因为最后有可能模型会趋于雷同，甚至很多应用程序会被迅速的抄袭，迅速的雷同化？

1 他们肯定是受益者，但这种说法有一个前提，算法是有上限的。但是实际来看，各家公司的算法上限不同，不同的公司，它掌握的能力不一样

2 算法还是有很大的提升空间，我认为不会在短期内趋于雷同。模型的发展的效果。可能是越来越往上的

3 发展的模式也有区别。可能会有一批大的公司搞基础性的模型，比方类似于 GPT 这种模型，其实 ChatGPT，它也是 GPT-3.5 的版本上，做了一些微调而做的产品。还有公司做应用层面/垂直赛道的小模型开发。未来是两种发展模式相结合

¶ 字节和百度在该方向的算力、数据和人员投入如何？

1 国内大厂的算力基础都不差。在模型方面，字节在推荐领域也已经有千亿参数的大模型，只是说在应用的领域不同。百度有文心大模型作为基础

2 数据方面，字节也有一些头条和抖音的搜索数据，量级上没有百度搜索的数据量大

3 从投入来看，其实两个公司的投入都非常大。百度把 AIGC 作为一个发展浪潮来追

赶的，而且搜索是它非常核心的业务场景，所以百度的投入是很大的，而字节，其实切入的稍微有点晚，没有百度那么快。字节把几个最重要的核心部门，联合起来成立专项团队。其实整体上来说投入也还可以

4 所以综合比较，在算力和数据上，字节跟百度的区别可能没有那么大。但在人力投入上，因为搜索是百度的核心业务，百度的整体的投入可能会比字节更大一些

¶ 现在 ChatGPT 没有对国内开放，国内厂商在中文的领域，相比海外厂商，在用户体验上能形成一定的或者明显的优势？

1 我认为语言不是大的壁垒，我认为短期内，ChatGPT 没有向我们大陆开放，我们国内其实是有机会这样做出来产品的，但是想要超越 ChatGPT 的可能性会非常低

2 因为像一些我们的头部大公司，目前来说也还没有推出一款产品，能够跟 ChatGPT 的模型的效果能够 PK，所以可能短期来看不大能够超过它

3 但是短期来看，我们可以通过这样的一个时间窗口做逼近它的效果，是国内公司比较好的状态

¶ 腾讯和阿里在 AIGC 方面的布局如何？

1 腾讯和阿里的搜索业务弱一些，不是重点，例如阿里，主要聚焦于电商领域，所以他可能在 ChatGPT 上不会有很多布局，目前阿里主要的发力方向是利用 AIGC 去做 AI+营销，比如赋能商品的文案撰写等，未来阿里可能会继续往这个方向布局

2 腾讯可能在广告、社交、游戏等领域应用 AIGC 技术。比方塑造更广义的互动叙事的品类，带来一些新的社交的玩法和商业模式的新的启发等等

3 总的来说，AIGC 是一波技术浪潮。国内的大厂的看法是要和现有的业务结合起来，实现自身业务更好地发展。而不是只关注 ChatGPT 这一个 AIGC 的细分赛道

¶ 可以大概理解成，腾讯和阿里更偏向应用端吗，未来腾讯和阿里的大模型会自研吗？

1 我认为腾讯和阿里的大模型是会有自研的趋势的。像这种大公司，它对专利，包括一些专业的技术积累其实还是比较有讲究的，所以我觉得长期来看，大厂大模型会自研

¶ 如果大模型投入使用，对于算力等基础设施的需求会不会是指数级的提升？

1 我认为是的，ChatGPT 刚发布的时候，就因为用户访问量过大，算力不足而出现问题。随着用户量级的大规模的上涨，算力的需求确实会呈现一个指数级的上涨

2 至少是非常正相关的，因此推理和训练的资源的开销肯定是非常大的。所以这一块也是优化的重点，就是怎么去让资源尽可能地节省，让整体的一个性能更好地提升

¶ 除了 GPU，芯片方面还有可以替代的产品吗？

1 自研芯片，但是目前整体来说目前还没有看到特别好的一个产品

2 采用分布式的 CPU，性能上差一些，但是成本便宜，很适合做推荐算法模型的公司，比如抖音、快手、Tiktok 等都是采用分布式的 CPU 做大模型的基础算力设施

¶ 如何去辨别海内外厂商大模型的优劣？

1 如果我们要评价它的具体的效果，最直接的是人工测评，看下真实的感受和评分。专业角度来讲，我们可以用测试集，分别请求这些模型的 API，基于一些评价指标，去看这些模型的表现如何

2 模型的参数、训练数据可以作为参考的指标。它的模型的参数量级更大，理论上模型的效果应该会更好，但相对片面一些，还是要实际测试和感受后才知道

¶ 应用场景的数据，在中国来讲是不是一种比较紧缺的资源。如果是要把模型训练好，可能非常依赖这些产业厂商的合作？

1 特定领域的数据是比较稀缺的，比如医疗、司法等领域，所以可能会生成类似的商业模式

2 可能最后就会形成这种商业模式：大公司负责训练大的基础模型，其他的一些创业型的公司或者一些小公司，在大模型的基础上，加上他们自己特定领域的一些数据集，得到这种新的领域式的模型，来服务于他们自己的一些商业化的计划

3 这种模式下，大厂有钱赚，对于小厂来说，它既能保护到自己的数据的隐私，同时也能够形成这样自己的领域类的商业化的路径

ChatGPT 中美差距究竟有多大

访谈日期：2023/2/5

具体内容

¶ 观点汇总

1 一位百度资深人士：他“没有兴趣”谈论 ChatGPT，言语之间，五味杂陈。一位人工智能企业创始人：面对 ChatGPT 的惊艳表现，心痒痒也迷茫，失眠了。他坦承，从模型的规模到效果，差距还比较远

2 国内某厂商的大模型和 ChatGPT：ChatGPT 从回答的逻辑性和完整度上都远超国内大模型，国内大模型的答案带有明显的拼凑感，夹杂着不少主题之外的胡编内容。而且，在回复速度上，ChatGPT 也领先一截

3 从事数字人研发的特看科技 CEO：目前全球还没有能跟 ChatGPT 抗衡的大模型，业界共识是差距在两年以上。国内先不谈弯道超车，趁早追赶反而是更重要的

4 虽然一些人工智能资深人士认为，在 ChatGPT 所涉及的技术上，中美是“平级”的，但华为诺亚方舟实验室语音语义首席科学家刘*，在黄大年茶思屋的讨论中坦承，中国在技术上还是有差距的

   a 其中一个是基础模型本身的差距，虽然我们训练了很多万亿模型或者是几千亿的模型，但训练的充分程度，是远远不够的
   b “我估计到现在为止，没有哪个模型能吃 GPT 那么多数据。”

5 清华大学计算机科学与技术系长聘副教授黄*提到，在 GPT-3 之后，OpenAI 所有的模型都没有开源，但它提供了 API 调用

   a 在这个过程中，它干了一件事，就是建立起了真实的用户调用和模型迭代之间的飞轮，它非常重视真实世界数据的调用，以及这些数据对模型的迭代
   b 当然，在此过程中，它也养活了美国一大帮创业公司，建立了一个生态

6 “你看我们国内的大模型研究，是 A 公司训练了一个，B 公司也训练了一个，打个广告就完了，模型开源，你爱用不用。至少目前还没看到一家比较好的公司，把数据和模型的飞轮完整转起来。所以，我觉得这是我们赶超 ChatGPT 的难点。”一位业内人士坦言

¶ 业界人士都提到了算力问题。由于 GPU 芯片等问题，在一定程度上，国内算力已被卡脖子了

1 即使国内头部公司，从算力上跟谷歌等相比，差距也是比较明显的。有业内人士称：从数据质量来说，整个互联网的中文数据质量，相比于英文还是有明显差距。“我们可能要想办法，做中英文不同语言之间的数据互补。”

2 几乎所有受访人士都提到了 OpenAI 这家人工智能组织，所体现的纯碎创新精神和长期主义

   a “其实从原理和方法看，他们所做的东西业界都是了解的，倒没有说什么是美

国做得了、我们做不了的。”

b 但像 OpenAI 和 DeepMind，他们可能是业界唯二的两家机构，无论在创新性、投入、决心，还是在顶尖人才储备上，都是一如既往坚持的

3 “我们看到的是成功，但里面可能已经有很多失败的尝试。”有资深 AI 从业者认为，在看不到前景和没有明显效果的阶段，OpenAI 非常坚定地做了投入，相反国内倾向于在技术出现突破后，快速追随。“国内大家第一步想的是，我们现在怎么用起来，但在不能用的时候，人家就在长期投入。”

4 “这件事其实是值得我们学习的，我们真的需要有足够多的钱，有这么一帮热血的人才，能够在一个方向上这样持续积累发力，我觉得这是一个非常必要的条件。”黄民烈称。最近一段时间，业界也在讨论中国企业能否超越

5 围绕业务，尤其是国内的场景，是有超越机会的。在局部应用中开始超越，这也是业界的共识

如何理解 ChatGPT 的强势出圈和国内发展

访谈日期：2023/1/29

具体内容

¶ ChatGPT 的运作机制、技术原理

1 ChatGPT 是一个基于语言模型 gpt 模型的一个聊天机器人，它是用我们人工智能的强化学习来进行训练的。它的突破性主要是在于它用了人类的反馈来去训练语言模型

- a 通过增加人类的反馈来不断迭代人类的普通的标注，比如人类会对他所有的给出的答案做出标注，哪些答案他的回答是比较好的，就给这样的答案以排名，把这样的排名再给我们的语言模型去进一步学习
- b 通过上万次的人类反馈的迭代，就是通过不同的语言内容来去使语言模型去不断训练，直到语言模型回答的内容跟人类想要的内容是保持一致的。这样就形成了 ChatGPT。ChatGPT 它因为是基于 GPT 模型的一个语言模型。我们就要大概的先讲一下 GPT 模型的一个一个来由

2 GPT 模型是一个生成的预训练的 transformer 的模型。transformer 模型是深度学习语言模型的一个基础的框架，是在 2018 年 6 月的时候开始有第一个 gpt 模型

- a 从 2018 年 6 月份 OpenAi 提出了第一个 gpt 模型，得出了关键结论就是我们说的 transformer 架构跟预训练模型的结合，就能够产生这种非常强大的语言模型
- b 可以实现强大的自然语言理解。也就是从 2018 年的 6 月份开始，这种强大的自然语言理解的模型的这个技术范式开始被确立起来。接着在 2019 年 2 月到 2020 年 5 月分别 openAI 分别发布了 gpt2 和 GPT3
- c 到 GPT3 的时候已经比 GPT2 大一百倍，它拥有大概 1,750 亿个参数。但是它跟原始的 GBT 模型并没有特别本质的不同，基本原理是大概一致的。但是它的性能比较是它发展的一个瓶颈，因为它的模型特别大
- d 在 2020 年 5 月份提到了 GPT3 以后，其实一直以来它大规模的预训练模型已经基本上确立了，直到我们 2022 年 11 月底出来了。ChatGPT 的模型。这一次进行了一个新的更新，特别是发布了它的对话模式的功能，可以放在网站上，让任何人来用对话的形式跟大模型进行交互

3 使得它可以做到回答问题，而也能承认错误，或者是质疑不正确的一些问题，或者是拒绝不恰当的请求等等。这样就形成了一个面向我们 c 端用戶去试用，非常好用的这么一个 ChatGPT 的一个机器人

- a 他的工作原理就是他就是用机器学习的算法来分析和理解我们文本输入的一个含义，根据文本输入去生成相应的响应
- b 这个模型它是在大量的文本数据上进行训练，并叠加了大量的我们的人类的一些标注的反馈，使得它能够去学习这种自然语言的模式和结构。他是可以模拟对话或者是回答后续问题，承认错误等等
- c openAi 为了去创建这么一种强化学习的模型，它一定要去设立一些奖励模型
- d 奖励模型就是 openAi 去收集的比较多的数据，招募了很多人类的训练师。在训练的过程当中，人类训练师就扮演了我们用戶和人工智能助手去进行交互的

这么一个角色

4 通过人类训练师对于人工智能助手的交互的数据去标注回答问题好坏的排序，使得 ChatGPT 模型通过不断的跟人类训练师之间进行对话来去，通过对话来去生产数据生产答案。通过对答案的好坏程度的一个排序标注

- a 使得这个模型就会根据学习的语料来去进一步的迭代他们。他的回答的一个策略进行数字迭代以后，它的回答的训练的它的质量足以匹配人类的对话的风格
- b 所以它的这个技术的创新点主要是在于两大方面，一大方面就是超大规模的预训练模型 transform 模型这么一个技术的一个技术范式，这是一个目前被学术界公认作为最前沿最优秀的一个技术的模式
- c 第二大创新点就是在于这种标注训练方式。人类训练师通过不断的 ChatGPT 模型进行对话，去标注，去排序，来使这个模型可以更好的学习到什么样的回答是人类认为比较合理的
- d 这两个创新点就使得模型在这一次发展当中有了一个里程碑式的跨越的进展，这是一个 ChatGPT 的运作机制

¶ 目前它的制约因素有几个方面

1 首先是成本过高，有两个方面的成本，一个方面是它的开发成本会比较高，另一方面是我们企业的使用成本会比较高

- a 它的开发成本是 GPT 模型它的一个发展历程，从 GPT2 到 GPT3，它的算法模型上、技术上没有太大改变，但是它主要改变了这个模型大小。从 gpt2 的一个 1.17 亿的个参数量，到 gpt3 的一个 1,750 亿的这个参数量，是增加了 1,000 倍的参数量
- b 预训练的训练数据从我们一开始 gpt2 的 5 个 tb 的训练语料，增加到 GPT3，需要 45 个 tb 这样一个存储量的训练语料。GPT3 训练一次的费用大概是 460 万美元，这是他训练一次的费用
- c 它整个 GPT3 的模型的训练的总成本是大概 1,200 万美元。1,200 万美元是 GBD3 的一个总训练的成本。所以开发的成本是它的一个主要的门槛。它的开发成本非常高

2 第二个方面就是这个模型被训练好之后，对于任何的一个企业来说，它有一个使用的成本。使用成本主要是 ChatGPT 单轮的对话的平均费用大概是在 0.01 美元到 0.2 美元之间，根据用户的使用的并发数不同，成本也不同

- a 其次是 ChatGPT 的技术局限性。技术的局限性主要，一个 ChatGPT，它只依赖于它见过的这些训练数据。它不会对一些实时的信息，比如新闻会网络上的一些实时信息来使得他的回答更加的精准
- b 所以目前我们在网上能够使用的 ChatGPT 模型，它使用的主要数据是 2021 年之前的，对于这个时间点之后的这个世界的信息，ChatGPT 他了解是非常有限的
- c 所以在输出的准确性上也会有所降低。这个第一个局限性就是它不能够与时俱进，很难与时俱进

3 第二个局限性，他的认知也是建立在我们虚拟文本，没有跟我们实时的数据库或者是去信息的连接。比如他很难去回答一些股票今天比如 a 股，它的指数大概是多少这样的非常实时的问题

a 所以它会在这种实时性的问题上回答上会出现一些致命的错误，或者是非常不准确的答案
b 是目前这个 ChatGPT 直接使用来说会有的一些局限性。如果是配合国内上一些专业的查询软件去进行二次开发，可能可以有效的解决这方面的一个问题

4 第三方面的局限性就是 ChatGPT 目前的模型训练，它的优化方向是围绕着我们人类的标注去设计优化的，所以有可能会过度的朝着人类认知的方向去优化，这样也会影响 chatGBT 回答的内容的风格

5 这可能是跟相关的人类训练师的一个偏好有关的，有些人类训练师可能是有一些个人的偏好来使得 ChatGPT 的训练可能会朝着那些人类训练师的偏好去有一些偏移，这也是其中的一个局限性

a 比如输入一个涉及 CEO 的提示，可能会得到一个这个人是白人男性的一个回复，目前因为很多人类训练是目前好像是假设这个人是白人男性，是 CEO 的一个概率会比较高
b 所以会产生一些负面不准确的，甚至是有种族倾向，种族歧视倾向的一些内容出来，一些可能是政治敏感，或者是不恰当的一些答案出来
c 这是 GBD 的一些局限性和成本的一个比较高的一个因素，会制约它的目前的发展

¶ 未来的发展方向

1 目前它的商业应用的场景是非常广泛的，只要它能够有效的克服以上提到那些制约因素，它在众多行业上都是可能会产生这种变革性的影响的，特别是在客戶服务、教育、家庭的陪护等等这些领域可能会率先落地

a 今年 2023 年可能是 ChatGPT 非常受关注的一年，也有可能是制约因素逐步被技术所迭代，后续克服的一年。ChatGPT 模型的出现对于这种文字模态的 AI 生成内容的应用也是有非常重要的意义的
b 未来可能会跟这种图像图形的 AI 生成内容的模型相结合，可以使得文字表述到图片生成的这种 AI 创作辅助工具来进行更多应用。或者是能够接受这样使用成本的一些领域可能会率先的去使用
c 根据我目前的了解，目前很多业内的从业者对于 ChatGPT 还是保持一个观望的态度，一方面还是在持续的考量模型的一个回复的准确性

2 以及它在一些领域的适配程度。另一方面很多企业讲应用 ChatGPT 也是会受制于它目前的一个高成本的使用成本，所以在商业化上还是一个比较谨慎的观望态度

a 目前我觉我们觉得 ChatGPT 可能会构建一个新的技术生态，但他目前所学习的还是互联网上公开的知识，他可能还不能解决一些具体行业、企业这些个性化的问题
b 所以还需要企业在这种相关的行业纵深行业细分垂直行业去进行二次的训练，这可能就涉及到很高的二次训练成本。所以可能是需要很多优秀的公司去不断的优化
c 能够提出一些更贴近我们客戶需求的和痛点的一些解决方案产品。比如我们作为这种虚拟人的公司，可以针对政府、企业、医疗、银行等等某个行业当中的企业去单独形成一些垂直化的解决方案
d 利用 ChatGPT 这些技术去进行专业私有化知识的迭代，使得它具备这种解决实际问题的这种能力。可能是 ChatGPT 后面的一个应用方向

¶ 目前国内相比于我们海外的差距到底有多少？是否有追赶的机会？

1 目前国内其实做这种 ChatGPT 类似的公司，也主要集中在大公司，或者是一些有国家政策资金支持的一些机构，学术机构，比如我们的百度，微软小冰

2 再包括阿里还有腾讯可能也在做。主要是这几个大的玩家可能会有成本去训练这么一个 ChatGPT 这样的超大模型，这样的玩家相比于海外的差距，目前还是有一定差距的

3 目前的差距主要集中在我们的预训练模型，它的回复能力确实自然程度上，还有包括回复的专业度上，以及内容的表述方面，相比于国外的 ChatGPT 模型相比还是有一定差距的

- a 人主观去体验，还是感觉机器人的感觉会比较强，然后直接体验 ChatGPT 会感觉回答的内容很自然。这是从主观体验上的一个差距
- b 从参数量的差距应该是没有什么差距了，目前我们都是千亿规模参数量的这样一个大规模的模型，不管是国外的 ChatGPT 还是国内的百度，还是阿里提出的超大规模预训练模型

4 还是我们清华提出的超大规模的预训练模型，他们的参数量上的差距已经是接近差不多了。所以我们都国内外，国内和国外都具备训练这种超大规模模型参数量模型的能力

- a 但是训练方法上可能还有一些技术，我们跟别人还是有一定差距的，所以后面可能主要在于训练方法，还有语料的标注上，可能是可以有更多的这样的语料
- b 国外这种英语的语料或者是英语的训练的训练的方法可能跟国内的中文的训练方法不太一样，所以导致我们现在训练的方法，这方面的技术上还是有一定的差距
- c 但我认为是有追赶的机会的。只要我们在这个成本足够低，足够可以大规模商业化之前，可以把这些差距给抹平

5 我们在这个成本可以拉到可以降低到可以大规模使用的个时间点的时候，我们也是可以跟海外的这些竞争对手去 PK 的一个机会。目前使用成本还是比较高，所以导致还有一个可以追赶的时间可以让我们国内的这些公司去追赶

¶ 什么样的契机会推动我们国内的发展？主要的参与方是什么？

1 其实我个人觉得目前我们的一些垂直领域的应用，或者是首先是在一些能够接受如此高昂的使用成本的一些领域，比如我们的金融是不是可以接受

2 或者是在一些政府有相应的预算的情况下，可以让应用可以先落地。落地以后就会产生大量的交互的数据，交互的文本就可以有大量的数据去迭代我们大规模的训练模型，使得它技术可以变得更强

- a 同时我们的工程师也可以通过技术的手段去优化迭代它使用的成本，使得使用成本降得足够低以后我们可以大规模的 ToC 商业化
- b 这样可能是我们比较好的一个契机。所以是否能够找到一个领域或者一个行业愿意接受如此高的使用成本，可以对他来说是收益高于使用成本的。如果它的收益高于使用成本，它就会大规模的铺开使用

3 当它的收益已经大于它的使用成本的这么一种场景。这样它就会可以大规模的去使用起来，就会有足够多的这样的一个资金或者是训练语料，可以有效的迭代模

型

a 目前主要的参与方还是几个大公司，百度、腾讯，阿里，还有微软小冰，还有科大讯飞可能也是一家比较大的一个参与方

b 这几家是可以有预训练模型能力的一些参与方。还有一些研究机构，比如清华的研究院，或者清华的相关的人工智能研究所，还有清华智源等等

c 还有一些国内新出现的一些创业公司，他可能会在一些非常垂直的方向去做一些非常垂直落地应用。有可能是创业公司先找到了一些应用使用价值

d 可以覆盖它的使用成本的这么一些垂直领域。可以在这些垂直领域先得到应用，先赚到第一桶金，后续可以逐步的复制到其他领域，这也是非常有可能的

全面解读 ChatGPT 产业链机会

访谈日期：2023/1/31

具体内容

¶ 为什么关注 ChatGPT?

1 2022 年 11 月 30 日，OpenAI 推出人工智能聊天工具 ChatGPT，一周后用户数突破 100 万人，月访问量达 2,100 万人次。ChatGPT 的推出，在 IT 产业和资本市场层面均产生巨大的影响

2 在产业层面，搜索引擎巨头忌惮于 ChatGPT 对传统搜索业务的潜在威胁，均做出了积极的应对：谷歌公司要求其多个团队集中精力，解决 ChatGPT 对公司搜索引擎业务构成的威胁；百度预计 3 月推出类似 ChatGPT 的人工智能聊天机器人。而微软计划将 ChatGPT 等工具整合进旗下包括 Bing、Office 在内的所有产品中

3 除此外，ChatGPT 由于其对文字工作者、方案策划师、程序员、客服人员等的工作内容具有替代性，正给产业生态带来深刻影响

4 在资本市场层面，根据华尔街日报消息，OpenAI 目前估值已达 290 亿美元，而 BuzzFeed 因采用 ChatGPT 上岗写稿，两天股价涨 3 倍等等。产业与资本市场形成了共振

¶ 对 ChatGPT 的产业观点

1 ChatGPT 之所以成为爆款，除了其本身的产品力较强之外，一个比较核心的原因是，在全球经济面临一定压力的背景下，企业降本增效需求尤其迫切，而 ChatGPT 等新技术是解决上述需求的最重要路径

2 历史上相类似的，2008 年的全球金融危机，催化了云计算产业的快速发展并逐渐从海外发达国家延展到了国内。因此，在当前，包括 ChatGPT 在内的人工智能产业，由于其对人工的替代潜能可以有效的帮助企业降本增效，因此会反复发酵甚至加速

3 从中短期来看，ChatGPT 对包括搜索引擎巨头在内的产业生态，暂时还不会带来实质性的颠覆，因为目前 ChatGPT 不开源，商业模式不清晰，同时其运营过程又需持续产生高额成本，因此影响了其生态的快速膨胀，这给其他公司会留出应对的时间和空间，也同时为其他的产业链参与者带来了机会

4 ChatGPT 虽然目前技术水平相对其他 AI 聊天工具更高，但仍未达到理想状态，其产品迭代及生态建立仍需一些时间，盈利兑现也需要时间

¶ 计算机板块相关标的

1 重点关注科大讯飞，其他标的包括拓尔思、海康威视、云从科技、格林深瞳、海光信息、寒武纪、景嘉微、海天瑞声

2 人工智能细分领域围绕算法、算力、数据三个方向展开，ChatGPT 产业中算法最

为重要，算力与数据其次

3 算法角度

a 科大讯飞、拓尔思（NLP）、海康威视（图像识别）、云从科技（图像识别）、格林深瞳（图像识别）
b 其中，科大讯飞在文本识别、语音识别、语义理解等领域优势明显

4 算力角度

a 海光信息（DCU）、寒武纪（AI 芯片）、景嘉微（GPU）

5 数据角度

a 天瑞声（数据标注）
b 不同行业也均有相关集成和平台类公司拥有大量的数据资源

¶ 对 ChatGPT 的产业观点

1 ChatGPT 在应用层面仍有很大市场空间。OpenAI 的 GPT 系列从 2018 年发展至今，技术迭代速度很快，若后续仍有新突破的产品推出，AIGC 市场商业应用会迎来爆发。从元宇宙角度来看，2022 年 11 月 1 日，VR 产业计划发布后，元宇宙概念掀起热潮

2 苹果 MR 设备若在今年二三季度发布，也会带来元宇宙热潮再次启动。在元宇宙应用中，依靠人力进行内容供给远无法满足应用需求，未来 AIGC 将成为元宇宙应用内容生产主力

¶ 传媒板块相关标的

1 中文在线、视觉中国、昆仑万维

a 中文在线在 AIGC 产业有较多布局，可基于不同场景填写关键词与辅助语句形成文字描述辅助人员进行创作
b 视觉中国旗下元视觉网站已推出 AI 作图相关应用，且目前销量可观。同时，其拥有大量图片版权，商业价值较高
c 昆仑万维相较前两个公司市值较大，较早进行 AIGC 布局，天工四大体系布局图片 AI、音乐 AI、文本 AI、编程 AI。且其海外 StarMusic 应用拥有大量用户群体，昆仑万维正探索运用 AI 技术创作原创音乐降低版权费用支出
d 数据要素市场方面，运营商数据量可观，在数据要素市场中可参与较多环节，且配合云计算业务与 IDC 基础设施下沉，竞争力较强

2 光模块方面

a 中际旭创作为 800G 光模块龙头，今年作为国内最早放量标的，拿到最高份额与订单，确定性较强
b 股权激励为其业绩保驾护航，且估值水平较低，团队预测其 2024 年进入 40%增速

3 运营商方面

a 中国移动作为团队连续两个月金股将有机会受益于 ChatGPT 等 AI 新产品发展

¶ 海外 TMT 洪嘉骏对 ChatGPT 的产业观点

1 ChatGPT 用户定位并非想要寻找最优答案的专家型人群而是 70%-80%的大多数人群，且其商业模式不同于传统搜索引擎的具有多来源出处的搜索模式，因此其模式的潜在广告收入较少，仍具有较大挑战

2 对于谷歌等巨头来说，其在海量搜索次数与算力的基础上发展商业模式只是时间问题，但聊天机器人的产品广告模式及成本问题短期内仍无法拥有较好解决方案，仍需进行继续探索

3 团队认为 ChatGPT 在现阶段算力制约下更为可能走向 ToB 模式。例如，微软将其应用于 Office 等生产力工具场景中。长期来看，虽然 ChatGPT 会有更多商业模式与场景应用，但对于搜索行业格局并不会是颠覆式的，两者将与时俱进且并存

4 AI 计算在过去六七年间，英伟达等 GPU 公司在场景应用上有较大突破，未来在 ChatGPT 模型搭建与实际商业化应用方面均需要更新型 AI 降低成本。相关存储板块，商业化之后也会有更大渗透率，突破目前消费电子为主要驱动的周期性瓶颈期

5 对于网络运算、网络通信等方面，以及相关通信公司，ChatGPT 也将深入网络加速等方面，未来商业化后将看到投资机会

6 港股板块上，中文 ChatGPT、语音学习难度大很多，但是当趋势形成，国内龙头将会持续发展储备产品

¶ 海外 TMT 板块相关标的

1 百度、商汤、腾讯、字节跳动

- a 百度最近提出筹备 ChatGPT 相关产品
- b 算力方面，商汤本身在超算方面有较好基础
- c 腾讯搜索业务发展较快
- d 字节也在规划突破社交封锁与搜索业务，值得后续关注

¶ ChatGPT 相对于其他竞品来说，主要的创新点和技术壁垒在哪里？

1 ChatGPT 利用强化学习从人类标注者反馈中学习，可进行问答、阅读理解、头脑风暴等。ChatGPT 关键能力来自于基座模型能力（InstructGPT）

2 可真实调动数据并从用户标注中反馈学习。ChatGPT 模型结构与 InstructGPT 几乎相同，InstructGPT 基于 OpenAIGPT-3.5 模型强大的基座能力，其学习主要分为三个阶段：

- a 第一阶段为冷启动监督策略模型，一开始依靠 GPT-3.5，GPT-3.5 虽然很优秀但不能理解人类不同指令中所蕴含的不同意图，故人类标注员会对测试用户提交的反馈中，对每个询问做出高质量回答，来使 GPT-3.5 模型初步具备理解人类意图的模型能力
- b 第二阶段为训练回报模型。训练回报模型依然依靠人工标注数据来训练回报模型，对每各问题所对应的 K 个结果质量进行排序，再通过对比学习方法得到一个激励模型（RewardModel）
- c 第三阶段为使用强化学习策略来增强模型预训练能力。此阶段不需要人工标注

数据，使用第二阶段模型打分更新预测结果，使用提问对应的随机指令，运用冷启动模型初始化 PPO 模型参数，进行随机打分，此分数即回答的整体 Reward，进而将此 Reward 回传，由此产生的策略梯度可以更新 PPO 模型参数，其创新点在于没有涉及多阶段模型训练任务，一般直接通过监督学习或强化学习。其将多个模型、训练方式累加到一起，通过多个模型作用于一个结果

¶ 如何展望 ChatGPT 商业模式，以及对产业链其他公司的影响？

1 伴随 ChatGPT 继续快速发展，ChatGPT 作为 NLP 的一个基础模型，NLP 领域包括信息抽取、机器翻译、小样本迁移学习等研究方向将会迎来较大发展。上游来看，数据标注、算力、数据清洗、数据采集等行业将面临蓬勃发展。下游来看，智能客服、聊天机器人等应用领域将蓬勃发展

2 目前国内电商等行业智能客服多轮对话能力较差，伴随 ChatGPT 等开放式对话模型升级，智能客服会在人力成本方面有飞跃

3 在写作等创作领域会有较大突破。NovelAI（diffusion）等绘画 AI 可提高平均画作质量且降低了成本

4 ChatGPT 素材收集、润色改写、扩充摘要等服务将使创作效率得到提升，AI 辅助写作可能成为主流写作方式

5 虚拟现实领域也是较为重要的领域之一，得益于 AI 创造能力提升，人类虚拟世界丰富程度将极大提升，将吸引更多客户。在教育领域，ChatGPT 可作为专职教师提高获取知识效率

- a 在搜索引擎行业，目前 ChatGPT 还无法替代搜索引擎功能。首先，其基于大规模模型，新知识接受能力不友好，更新模型的训练成本与经验成本很大。其次，若面向真实搜索引擎的大量用户请求，在线推理成本较高
- b 搜索引擎与 ChatGPT 模型双结合方式可能会成为搜索引擎主流方向，国外部分厂商已经在逐渐将类似 ChatGPT 功能嵌入搜索引擎

¶ 国内 ChatGPT 产业链的发展现状？

1 国内向 ChatGPT 以及 AIGC 领域发展的公司已非常多。百度向 ChatGPT 领域发展动机十分明确，维护其搜索领域护城河，在下一代搜索引擎市场中抢先占据有利地位。百度 ChatGPT 业务开展得益于其大量搜索引擎业务问答样本，样本量级足够。京东、阿里、拼多多等公司已经开始在智能客服方向上做出尝试

2 字节跳动也在逐渐入局 AIGC，并将生态场景在内部进行应用，原来今日头条中内容分层依靠于 UGC 等生产者，现在已逐步往 AIGC 方向迁移。国内一些创业型公司也已经开始崭露头角。聆心智能推出 AI 乌托邦，其开放式对话与 ChatGPT 较为类似

3 国内大多数公司正在向虚拟人、AIGC 等概念靠拢，目前没有 ChatGPT 替代品问世，还存在着一些技术发展瓶颈。原因在于四点：

- a 国内缺少基础模型，没有模型迭代积累。ChatGPT 依赖于 InstructGPT，InstructGPT 依赖于 GPT-3.5.GPT-3
- b 国内缺少真实数据。除百度有天然用户搜索问答训练样本外，对于其他公司较为缺少

c 国内缺少技术积累。ChatGPT 发展过程中对于数据处理、清洗、标注、模型训练、推理加速等方面均具有技术难点，且对结果均影响较大

d 且包括国内大厂在内，强化学习框架仍未出现大规模使用场景。国内创新性土壤还需发展。整体商业环境较为急躁，但投入与产出需要花费一些时间

¶ 随着 ChatGPT 的应用群体增加，是否会出于成本考虑对国内的流量使用进行限制？

1 目前 ChatGPT 处于 demo 阶段，是否会对流量作出限制取决于 OpenAI 在此阶段预备投入，其是否愿意增加机器、增加服务部署

2 若国内流量已经完全影响到其在线服务，限制国内流量是有可能的

¶ 后续围绕 ChatGPT、AI，产业还有哪些值得期待的重大变化？

1 短期重要产业变化主要在三个方面。首先，短期内围绕 ChatGPT，搜索引擎领域会出现两者结合发展方向。其次，在智能客服领域，若 ChatGPT 可以实现客服功能，对人力成本降低会有突破

2 再次，在 NLP 应用领域，由于其本质上是序列到序列的语言模型，伴随 ChatGPT 模型能力提升，领域技术上限提升，下游机器翻译等领域也会得到发展

¶ 基于 ChatGPT 的智能客服，是否反而会增加企业成本？

1 分情况而定。传统客服成本为人力成本，ChatGPT 成本包括在线策略成本、机械成本、离线训练成本、数据采集调度成本等方面。在成本方面，需要对客服对接客户问答数据量进行估算，对小规模公司来说

2 自研此类工具需要大规模数据训练、采集、清洗等资本花费。对于大规模日均产生用户交互较多的公司来说，长期来说，数据训练、采集、清洗等资本花费只是一次性的，花费更多集中在在线成本上，此时成本会低于人力成本，故新型的 ToB 服务模式为中小型企业提供智能客服功能也将是未来发展的方向

3 在质量方面，ChatGPT 质量不会低于人工客服，其足以支持代码 Debug 等精细专业化服务，效率比人工客服高

¶ 国内布局 ChatGPT 公司中，在信息基础设施选择方面，国产设备及云的占比情况如何？

1 云计算设施方面，国内大厂例如百度、阿里、字节均使用自研云计算服务。对于中小型企业，阿里云市占率最高，阿里云、京东云排名较为靠前

2 芯片方面，目前大规模使用英伟达芯片，主要原因在于其性能、服务链路积累及其市占率优势。目前自然语言处理、计算机视觉等领域均会使用英伟达 GPU 芯片等高性能芯片

3 针对搜索、推荐等场景，很多公司不采用 GPU 而采用 CPU 形式，例如字节在推荐等场景更多使用 CPU 芯片进行分布式计算环境搭建，成本会有所降低。但对 ChatGPT 来说，对大规模 GPU 芯片有所需求，国外大厂目前市占率非常高，国内

自研有所推进但在此方面仍有所欠缺

ChatGPT 来龙去脉

访谈日期：2023/1/29

具体内容

¶ 从任务角度来说，ChatGPT 以问答类为主，对话领域的模型非常复杂，ChatGPT 技术方案最大的优点就是单一模型，特点就是参数比较大，达 1,750 亿的参数，代价就是需要巨大的算力

1 当今时代和过去不同的就在于以前是系统复杂导致人力消耗巨大，现在则是算力要求。以前重人力的时代下产品的“天花板”不高，ChatGPT 实现的效果在以前是无法达到的

2 ChatGPT 技术最初的源头是 Transformer 结构，这个结构最大的意义是可以承载更大的算力和数据，去训练一个更复杂的模型。GPT3 所采用的 GPT 路线，又叫单向注意力模型，只要算力足够就可以训练出参数巨大的模型，尺寸上不封顶，最高点尚未可知

¶ GPT3 是 20 年提出的模型，达到 1,750 亿参数，这已经是 OpenAI 的产品演化了两年后的产品

1 2020 年和 2021 年很多公司在做千亿甚至万亿参数的模型，但都达不到 GPT3 的效果，很多公司并没有持续深耕该领域，而 OpenAI 在经过两年后又提出了 Gpt3.5，所以来看 2023 年即将发生的事情

2 接下来可能会有一些公司和团队对外宣称做出了类似 ChatGPT 的模型，参数甚至超过 ChatGPT，但不会有想进一步把模型转化为产品的想法。如果存在一些公司能够做出模型并且不断改进、持续升级的话，那么这些公司是值得关注的

3 ChatGPT 应用落地的一个很大的问题在于在任意场景落地都需要对产品进行定制化。还有一点，ChatGPT 虽然“见多识广”，但是比某一项能力，未必能超越垂直类的产品，比如针对医疗数据训练出一个模型，用它来做问答，在医疗领域一定是比 ChatGPT 要好的

¶ 解决这些问题的方案主要在于解决具体场景定制化的需求

1 一方面是知识的定制化，要让 ChatGPT 学会、精通某一领域的知识

2 另一方面就是技能的定制化，要对 ChatGPT 特有的技能如：推理、写作等进行专门强化。但是定制化的问题在于成本非常高，ChatGPT 的参数量很大，训练成本就会很高

3 类似 ChatGPT 这类模型的商业落地，应该先从中等尺寸的模型开始做起，这些中等尺寸的模型可能就几十亿到几百亿的参数，落地成本没有那么高。中等尺寸的模型可能功能没有 ChatGPT 强大，但是在专业领域，往往也不需要全方面的能力

4 国内的发展格局分为两大类，一类是专门型的研究机构和团队，另一类就是大型

公司。从公司角度来看，国内有百度、阿里、华为、腾讯还有浪潮等都在探索这个行业，他们都有超过千亿的大模型，但是他们没有将这些模型当做产品去做。虽然这些大厂商有丰富的资源，但是在现在的大环境下，整体都处于收紧的状态，资源基本都倾斜主营业务，不会在探索性的领域投入过多

¶ 从研发机构角度来看，只有北京智源和 IDEA 研究院

1. 智源开展时间较早，在 GPT3 出现后，智源做过千亿参数的模型
2. IDEA 研究院也做了一系列的几亿到几十亿的开源模型，已经形成的封神榜预训练大模型体系在中文 NLP 起到支撑性的作用，评估一个团队，要注意是否有在大算力上去做大模型的经验，大多数团队都只是具备在小规模算力上做小模型的经验
3. 展望 NLP 和 AIGC 的未来发展，NLP 是经历范式革命非常严重的一个领域，从以前需要找关键词到现在 Transformer 结构的出现，技术在不断地改变，有一个猜想就是 NLP 领域未来可能会消失
4. 像 ChatGPT 这样的模型出现，我们有特定需求的时候只需要去调整 ChatGPT 去实现即可，未来 NLP 算法工程师是否还有存在的必要是一个值得思考的问题

ChatGPT 学习笔记

访谈日期：2023/1/29

具体内容

¶ 公司发展到 2000 年到 2011 年的阶段，核心的技术就是基于检索技术，开发了智能内容的管理

1 在 2007 年启动了核高机的非结构化数据系统的研究的专项

2 在 2011 年的时候，a 股市场上拓尔思公司作为第一家大数据公司上市，上市以后公司持续的在自然语言处理的技术上做研究，公司战略的定位是语音智能，是核心技术的一个发展场景

   a 自然语言处理应用在搜索引擎、智能客服，舆情分析还有内容处理方面，多年以来通过持续的打造，形成了每个板块深度的应用场景，同时打造了一批专属软件平台
   b 持续以来业务收入的增长也是基于我们对各个场景应用的熟悉，知道自然语言处理、语音、智能应用的方向，为用户输出了大量的、有时效的整个应用效果
   c 整个人工智能时代有三要素，算法、算力和数据。拓尔思公司作为人工智能和大数据公司，所有的人工智能应用都是来自于对各种算法模型的积累。首先需要有数据，所以在 a 股市场横向比较，我们是真正掌握了大量的数据资产的公司
   d 2,000 多台服务器分布在全国的三个数据中心，每天日增 1 亿条的开源的互联网的数据公司，已经积累了将近 1,300 亿条开源的数据资产
   e 有了数据资产，我们才能够做各种各样的训练模型，才能够积累各种各样的算法，现在已经积累了 300 种以上的算法，并且对每个场景，像知识图谱的展现，知识库的建立档案

3 包括前期的数据的采集，还有数据的标引，诸多的关于数据要素的这些环节，都以完全知识产权的软件平台去持续的做这样的工作

   a 搜索引擎是我们自然语言处理的一个核心应用的技术，公司是 30 年以来坚持在这方面的积累，在全国整个大量的企业级的搜索都在用 ELSG 的设计、spark 这些开源软件的时候，我们公司完全捉到了自主可控，完全捉到了信创的银窝，应用到政府、金融，包括媒体等诸多的行业
   b 数字经济研究院目前主要的一个研究方向就是人机对话，像托马斯公司这几年以来在整个技术应用上面，比如围绕着像中国中医科学院的中医中文问答，中国标准化研究院的国家标准的问答
   c 人民卫星出版社的小 a 机器人，时代经济出版社的审计问答、吉林政务的小机智能机器人，这些实际上都是跟智能问答相关，跟大家现在谈的热点都极其的相似。除此之外，我们围绕着知识图谱事件分析，包括机器人的自动写作，智能内容创作等方面，都有多个成功的案例
   d 像 OpenAI 热点事件出来以后，我们的研究人员对于整个 OpenAI 的过去、现在和未来也持续性进行了研究

4 结合我们公司的一些技术沉淀的事实和我们本身对场景应用研究，未来展望有了一些系统的梳理

¶ ChatGPT 加快数字劳动力时代的发展

1 新的智能新意时代，ChatGPT 的出现引领了数字劳动力的时代，带来了第四种用工模式

2 数字劳动力将是生产力的第五次革命，这种新的经济时代、用工模式将会快速的演变。三大传统的用工模式包括全职员工、外包员工、兼职员工，数字化劳动力是第四种用工模式，打破了人与机器的边界

- a 依托人工智能技术，包括像 NLP 相关的一些技术，自主完成或者协助人类来完成各种企业的各种工作，比如前端对客人或者员工的文案工作等等，或者是中后台运营协同等工作
- b 像 Tab、BP 就能够帮助写文案内容或者代码，实际上它是一种数字化劳动力的一种。我觉得在这种传统劳动力跟数字劳动力的结合下，通过我们这种 NLP 相关的技术赋能
- c 能够让传统劳动力爆发出更高效这种增长力水平。根据麦肯锡统计数，到 2030 年，数字化劳动力的这种市场规模可以达到 1.73 万亿水平
- d GPT 的火爆加速推动这个事件。劳动数字化全面转变的核心在于劳动力，它的大脑、认知能力跟分析能力决定了数字劳动力是否能够准确的理解人类的任务指令，是否能够高效的去准确的完成任务

3 GPT 能做到这一点是基于人类反馈的强化学习，有一个千亿规模的模拟训练，可以融合世界的知识与规矩，使认知能力跟沟通能力接近人的水平

- a ChatGPT 的火爆将增强大众对于这种对话式的 AI 的一个信心，我们会有更多的研究来加入行列，推动整个对话式的 AI 的发展
- b 对话式 AI 大概分成四类，信息查询类、专家咨询类、助手类以及交流类
  - i 第一类是信息查询类，用户可以去查询企业的相关的一些信息，相当于数字化劳动能够替代一些枯燥重复性的劳动
  - ii 第二类是专家咨询类，这是比较重要的一点，相当于数字劳动力能够替代部分或者扩充这些资源稀缺的劳动力，需要我们大脑的赋能，专家系统可能是投顾类，或者是法律顾问类
  - iii 第三类是助手类，相当于数字化劳动力能够帮助人类去完成相应的一些任务，帮你订个机票，预定个会议等等
  - iv 第四类是交流类，数字化劳动力能够满足人类情感交流的需求，可能是情感的陪伴，或者是闲聊场，或者是虚拟在元宇宙里的

4 四类对话式 AI 对标不同的应用场景

- a 第一类信息查询应用的比较多，比如智能客服机器人，一些售前信息的查询，相当于降本增效
- b 第二类专家咨询是 MLP，需要加上世界知识，行业知识，专家系统
  - i 为企业去打造个性化咨询，根据司法部数据显示，全国办理各类的法律事务的事件大概是 1,300 多件
  - ii 涉及到诉讼或者是非诉讼的大概 1,300 万件。按照中国的律师平均费率是大概一个小时是 2,788，每个案件平均服务时长十小时来算，整个法律的咨询的总体市场规模达到 3,600 个亿
  - iii 如果是按照律师事务所这种维度来计算，像 21 年年底全国共有律师事务所 3.65 万家，对法律服务技术的投入按每年 100 万来算，法律的服务的总体分了大概是 300 万
  - iv 相当于我们要把一些法律相关的知识形成企业的大脑，能够去对外赋能，

其中就涉及到我们怎么去利用这些知识构建出复杂的知识体系里头来

c 第三个场景是助手类，比如智能车载助手，其中很重要的一点是智能创作，比如直播文案的生成，广告文案的生成，或者做一些剧本的创作。整个智能创作市场主要是分成数字资讯类、数字营销类和行政办公类

i 18 年公开数据显示，18 年各级的网信办审批的互联网信息、新闻信息服务单位总共有 700 多家，在主要的一些门户资讯，比如像微信公众号，它的总量大概是 2,100 万，活跃账户有 350 万

ii 按每年 SaaS 化软件一年 3,000 块报价来算，总体规模大概在 120 个亿左右。数字营销类每年的全球广告支出蛮高，18 年在 e-master 数据显示，全球 18 年的全球广告支出高达 6,000 多亿美元，数字广告就占到了 2,800 亿美金

iii 我们希望能够在数字营销里提供一个数字营销的广告的助手。在行动办公领域，我们可以看到爱乐咨询的一个数据显示，PC 这种办公软件的用户活跃数在 5.3 亿的数字上下波动

iv 预计这个数字在未来几年也不会有太大的变化，这个群体其实是智能创作的一个重点挖掘的对象，按照每个用户付费 100，总体规模可以达到 530 个亿

5 基于对话式 AI 市场，拓尔思公司规划未来拓尔思的优势有以下几点

a 第一点，拓尔思有来自境内外的各行各样的数据市场，超过 1,200 个亿，已经具备千亿数据的数据索引等，这些是我们的一些核心资产，包括我们背后的这些模型，包括我背后的加工能力等等

b 第二点是技术的沉淀，我们坚持核心自主的研发，实现国产化，拥有 40+的发明专利，800 的软件着作权

6 技术沉淀也相当于 AI 的三大要素之一。最后是客户的沉淀，整个数据的产品和服务已经国内外超过 1 万家的企业级的用户在广泛使用

a 像智能客服现在基本都是基于检索式的，基于我们数据库，我们将有一些基于深度模型，去库里检索答案，返回给用户。思想是基于一个大模型，有排量数据去训练一个模型出来，再加入人类反馈的数据，我们需要累计高质量的人类的反馈数据，这样我们就能够提供更优质的对话体验

b 还有第二点，我们需要行业深耕，像这类 DP，它是一个通用模型，缺乏对一些行业客户、行业知识的了解，我们对行业是非常了解的，我们未来会让对话式的 AI 这种人工智能技术去跟行业客户的业务流程去更深度的融合，包括从局部业务到全场景的覆盖，实现全业务的数字化、智能化

c 我们会持续的在行业中不断的累加场景，深耕场景，解决核心业务的一个问题。从长远来看，拥有更好的数据，更好的行业的一些 know-how，更有利于去微调我们的一个大模型的，给客户带来更好的产品体验的

¶ 像 ChatGPT，不懂的地方会一本正经胡编乱造，目前的技术发展路径是不是已经开始往准确率这方面去走呢?

1 目前整个智能客服是比较成熟的一个阶段，但是所采用的技术基本都是基于线索式，保证了所有的回复都是从库里拿出来回复给用户的。像 ChatGPT，它是基于生成式的这种方式去回答用户，比较难保证回复的可靠性

2 所以我们在后续的训练跟维护的过程，我们去增加一些这种规则，或者是增加一些这种安全检测的一些模块进到系统里头，能够保证我在一些异常条件下去规避

掉这些问题

3 现在 CC 已经能够让可靠性保证在一个比较小的结果里头，但是它还是会有这样一个问题存在

¶ 无论是信息查询、专家咨询、助手或者交流，从公司的视角以及整个产业发展趋势来看，哪一块最先有可能形成商业化的落地?

1 我觉得几个点都有可能。一个是这种专家咨询类的，它实际上是需要有一块比较好的相当于是企业大脑的角色，把这些行业的知识变成一个企业的大脑

   a 变成一个模型的知识。ChatGPT 证明了在一些大数量前提下，是一个比较好的表现的，这一块是在智能创作助手类的，一个是我们能够去高效地提升智能创作的水平
   b 现在它的这种文本生成能够已经能够满足创作者的大部分的需求，相当于我能够去帮助创作者生成一个初级的版本，创作者在上面再去继续修改，能够有一个比较好的效率提升
   c 在不管是直播文案的生成或者广告文案的生成，或者基本创作等等，还有在交流的，它已经像 GPT，拥有一个比较大的模型，拥有一个比较好的一种世界知识\通用知识的前提下，能够回答各类相关的一些问题

2 如果我们是按照比如在元宇宙里，或者是在一些养老领域等行业里去定制一个这种相关行业的，可能也是会有一个比较好的表现。所以我觉得大概可能是这几块

3 专家咨询类未来会在法律咨询的市场有一个比较亮眼的商业模式的落地

¶ 如果未来转向人工智能对话式的方式，是不是对于数据的采集其实是会有偏好性的，或者我们如何确保自己采集过来的数据是针对相关的行业，而并不是会跨到其他行业，我们怎么去确保未来这种算法以及数据的针对性是足够匹配到行业的一个情况?

1 好，您提到的其实是一个模型上下文关联的一个能力。在这种大模型的前提下，大模型是能够学习到相关的上下文的一个知识

2 比如我们拿法律的整个行业的数据进来，训练出一个大的模型的结构，再基于人类的一些反馈加入训练，最后出来的一个模型会在不同的条件下，识别到不同的上下文的知识的。在不同的领域里头它是带有不同的知识，都是能够识别到这一点的

¶ 在这个问题解决之后，现在我们最大的痛点是在哪里？拓尔思后续会在哪个行业率先落出相关的商业模式，并能产生实际的收益?

1 这一块我简单的回答一下。接下来首先就是语义智能，它本身是一个经验型的，这种技术的积累在这一块首先还是来自于你所熟悉的行业。我们强调的最多的人工智能的场景的应用，要选择一个比较好的主题

2 在选择主题以后，你自己作为公司在深度的积累知识，最后结合语义智能，围绕着主体场景，理解可能就越深。后面通过训练数据，还有源源不断的能积累的进来，训练的整个的模型

3 算法会积累的越来越丰富所以我们觉得经验值是非常重要。举个例子，拓尔思在媒体行业，譬如垂直领域的 120 多家媒体，有 40 多家是我们的客户，一半以上的审计的融媒体中心也是我们的客户

a 我们这几年以来在整体的打包服务中，有一个拓尔思的妙笔。小四的智能写作实际上就是一个合成，但是需要我们了解整个的编辑记者，在他们应用材场景中间，对于他们的新闻要素，新闻稿件的形成的整个的细节
b 我们先不断在丰富的在积累。原来一个编辑记者要花 30 分钟才搞定的一个稿件，我们可能快速的一秒钟就能够生成一个初稿，最后让他进行新加工
c 另外，融媒体中心成立完以后，他们出稿子的频率越来越快，任务越来越多，越来越大，这种情况下，怎么能够快速高效的去完成他们的这种劳动工作？还有一块，譬如刚才您讲到专家咨询，我们现在正在跟国家知识产权局深度打造我们的专业的咨询服务，这就是个很专业的活了

4 因为整个国家知识产权局现在有 2 万多专利评审人员，80%的时间都在拓尔思的三大平台上进行工作，这就是我们长期积累的知识

a 国家专利局有全国最大的最全的专利库，我们公司称之为数据的这些文本信息，都是一篇一篇的专利原作，对原作要进行语义智能的这种分区，要进行各种各样的标语，这些事情我们都做了
b 接下来在申请专利的过程中间，我们的专利申请人员对于整个专利申请的流程，整个专利检索的专业的知识，我们能够打造专业的技术服。回头来说，我们实际上强调的还是对行业深入了解和熟悉的程度
c 它的背后有一系列这种知识库的间接。我们拓尔思有一个自己的知识图谱的研究院，在开源情报这方面多年以来持续实现了我们一定比例的收获，并且还有很好的增长趋势
d 基于我们对整个的开源情报的这些分析，各种各样的数据的采集加工，我们不断再迭代，也形成了我们的自己的知识图谱的各种各样的算法

¶ 未来是不是会有可能在每个行业都诞生出一个龙头，类似于搜索引擎龙头，而不会像现在通过谷歌我们对各行各业所有人一起去进行搜索？未来的趋势到底应该是以垂直行业为主，还是有一个大一统的搜索平台为主？

1 从目前应用事件上来讲，我非常认同你的说法，这也是我们研究院一直在沟通的。因为刚才我们都提到了一个共同的问题，就是现在我们关注的女性事件，大家背后说她胡说八道。实际上你会发现它现在整个积累的时间和计算的时间，尽管跟我们国内的公司比已经有了一个数量级的差异，但是它不能够穷尽一切

2 理论上来讲，它能够穷尽一切，以后它就真正能够替代人了。现在我们在探讨应用的同时，反过头来反思我们国内有哪些应用场景，从这两方来讲

3 我们认为每一个垂直的专业板块空间都是非常大的，也就是拓尔思未来的发展。在整个人工智能和大数据的中间软件，我们已经达到了比较强大的自主可控的软件平台的积累。但是对于每一个垂直行业的这种深度的应用，在知识积累方面，我们也不是什么行业都去干

a 但是我刚才跟您举例的，譬如知识服务用在专利检索，用在整个专利行业，未来一个百亿级的规模，大家会需要有更多的这种服务的时候，我们就把更多的给打造好，围绕着金融，围绕着媒体，围绕着这几个深度的行业去做就好
b 我们还有一个可以拓展的行业，结合虚拟人和两周机器人走

c 悟到更多新的应用，我们也在拓展我们的新的市场。譬如在两座机器人，围绕着养老院场景，下的精力是最多的，一旦走进来，我们可能就能够比别人积累更多的支持

2023 电子产业展望

访谈日期：2023/2/7

具体内容

¶ 整体形势

1. 目前获利趋势来看目前在下修循环，1.2 月持续下修，应该会维持到 2 月份，截至 1 月底对费半的企业获利预估已经到-15%的预估水准，基本上 2 月份修正完应该会落入短期的底部，3.4 月份有机会进入上修或持平的阶段，若 3.4 月份反映整体经济不好的情况恐落入 doubledeep 的情况
2. 股价方面，目前已经领先基本面反弹，反映下半年的复甦，是否能持续向上就要看 3.4 月份的情况才比较清楚，截至目前为止能见度还不高
3. 库存方面，IC 公司依然很高，台积电到今年年初才开始才有产能利用率显着下滑的情形，预计今年 Q1.Q2 库存金额会出现显着修正，短期不太需要注意天数而是金额
4. 因为业绩不好天数就会上升，主要观察 Q3.Q4 金额是否有修正，业绩若带动天数就会下降，下游部分，库存已经修正一阵子了，详细数字要到 2.3 月份才会公布

¶ 美国零售销售数据

1. 目前看到美国零售销售数据有走缓 MoM 开始下滑，YoY 还有+5%~10%左右，以现在为止的预估维持在小幅成长的状况
2. 目前还没有看到下半年有 slowdown 的迹象，其中电子产品从去年就不好（YoY-10%左右）但看起来有稳定的状况，走缓的速度下降

¶ 中国消费市场

1. 去年整年都不好，基本上数据都在 0 以下，12 月解封，过完年开始正常，预估会是缓步的回升，中国人民超额储蓄去年增加了不少
2. 中国人民消费信心不足，后续可以观察是否疫情正常化后转换为消费力道将成为助益

¶ 估值方面

1. 是近期需要担心的部分，费半自去年十月份已经反弹约 50%，但企业获利是衰退的
2. 目前 P/E 接近历史峰值，段线上的嘎空行情大概已经 Price-in，后续要看基本面是否回升，以 P/B 来说比较没有那麼激烈

¶ 筹码面部分

1. 牛熊指标已回到 4 以上，HEDGEFUND 陆陆续续回到市场，其在主导市场波动较为剧烈，而 LONGFUND 的部位还是处在观望的角度，短期要有资金行情不容易，需要 LONGFUND 回归市场
2. 目前没有看到金融危机的迹象，看美国整体信贷利差已自去年高峰下降，而欧洲状况而言也是如此，美国公债的流动性指数去年 10 月份状况很糟，已经有所改善但目前还是偏高，会对于美国整体的财政运作，若未改善，就不用担心美国会有更鹰派的做法

¶ 产业上较看好

1. 中国解封概念、互联网、中国智慧型手机
2. 台积电耗材、以及已经落底很久的记忆体或面板

¶ 较不看好

1. 半导体设备、云端运算等砍资本支出相关股票
2. NEWCPU/GPUinto2021.INTEL 上半年在 DESKTOP 较没有新产品，比较像是更新，原本预计的产品有 Meteorlake 是与台积电有深度合作的产品但是有递延，而 Notebook 上是有在 ROADMAP 的计划上，要观察是否能做的到，还是会递延

¶ AMD

1. 目前在 ZEN4，明年才有机会出 ZEN5，ZEN5 会採用 3 奈米
2. GPU 今年还是看 NVIDIA 有没有新产品，近期推的 4,070TI 其实就是之前的 408,012G，价格上有下调 100 美金，CP 值有限，应不会有大的换机潮，感觉起来是 GPU 的小年
    a. HPC&Server1.Intel 今年的重头戏在 SapphireRapids（SPR）的量产，明年看 EmeraidRapids 的量产，今年下半年还有 BSH 及 SierraForest 后者是和 ARM 做竞争，应该不会太好
    b. AMD 今年看 Genoa 的改版及 Bergamo（Zen4C）有提高核心数至 168 核心，ZEN5 要看 2024 年
    c. INTEL 产品的规格上看到 BSHAP 平台的功耗达到 500W，2 个 SOCKET，AP 则可做到 8 个 SOCKET，观察到 PCIe 部分，目前已经增加到 80~90 条，再搭上南桥晶片就没什麼意义，下一代可能就会取消掉
3. INTELSPR 的 DiePackages 有 XCC 和 MCC，MCC 使用较传统的架构，为单颗大核心，不像 AMD 是使用许多小核心，到 Granite 也还是以单颗大核心的架构
    a. AMD 产品规格上今年最大的卖点在 ZEN4 做到 96 核心，ZEN 搞不好可以做到 192 颗核心，AMD 市佔能持续提高与 INTEL 拉开差距的关键
    b. AMD 的架构看到 CCD+IOD，最多可以放到 8 颗 CCD，设计的成本及弹性就比 INTEL 还要好
    c. 在 INTEL 取消掉南桥晶片后下一架构会走到 Self-boot，CPU 自己可以开机，未来 SERVER 在简单化的趋势下要自己可以开机，而相关趋动的部分将移到 BMC 上面

¶ DataCenter

1 走向 DC-SCM 的膜块设计，将 BMC 和 RoT 与服务器拆开，为来客制化产品就可以放在 BMC 上，架构相对简单，且报废时可以单独销毁安全膜块，和主板的连结会用 FPGA，近期有些 LONGFUND 就是在看信骅的 2,700 及 ASIC 去取代 FPGA 等等讯息，但可能要到 2025 才会量产

- a CXL3.0 让 CPU 及 GPU 的记忆体可以互通，记忆体会是未来限制频宽的因素之一，今年的 SPR 及 ZEN 採用 CXL1.1，明年可能会使用 CXL2.0，而去年通过 CXL3.0 则是要等 2026~2027 年了
- b ARM-baesdServer1.IDC 预估明年 ARM 的市佔就会达到 10%，主要在各家公司开始使用 Amphere 的 solution 带动各家公司的 support，ARM 的是佔率应该会持续上升
- c 产品主要有 V 系列及 N 系列，V 系列主要顾客有 AWS、Nvidia、Google，N 系列主要顾客有 Nokia、Marvell 等网通厂

2 AI Chip update

- a 以 GPU 市佔率最大的还是 Nvidia，推出 H100 取代 A100，而今年比较值得注意的是
  - i AMAZON 自己的 Trainium，据说自家的 AIServer 可能有一半以上用自己的晶片
  - ii AMD 的 Mi300 是市场上第一颗 SoIC+CoWos 的晶片把 CPU 跟 GPU 做整合，年底量产，主要客户有微软等因此对 Nvidia 有掉市佔的威胁
- b 未来 AI 设计的瓶颈在于记忆体，GPU 的演算力好几倍的增加，而记忆体的频宽并未跟上，第一个方法是增加记忆体使用量，第二个是增加 CATCHMEMORY
- a 回头看到先前 INTEL 也有类似的产品 PonteVecchio，自 2021 年就提了，但尚未量产，算力约当于目前 Nvidia 的 H100
- b ChatGPT 短期影响不大，每个月要烧 300 万美金，差不多是 1,500 片 A100，NVIDIA 一年约产出 6~70 万片，对一开始来说不会有很大的量，目前还言之过早，目前还是要看终端应用所带来的发展

3 Autonomous Driving Chips

- a 目前来说车自系统开始走向 Domian 架构，会比较偏向整合成一块，像是中央控制在传送给各个地方，目前为边缘运算的方式
- b 以目前车子晶片公司来做比较的话，TESLA 以外的主流就是 Mobileye，目前多主流车厂在 LV2 上几乎都使用 EyeQ4，EyeQ5 往上向 L3 的 Designin 就比较少一些，而 TESLA 先前 HW2.0/2.5 使用 NVDA 晶片，但功耗太高，HW3.0 就採用 FSD 晶片，今年会推 HW4.0 在三星投产
- c Nvidia 虽然及实有晶片可以使用，算力高但是功耗非常高，仅中国的造车新势力比较着急使用的车企使用，因此一线车厂未来 3~5 年应该是会使用高通，算力约在 360 左右，功耗 65W，现在若联发科要重组该部门可能也较难打入一线车厂

¶ CHATGPT 的看法，对 GOOGLE 的威胁?

1 威胁一定有但比较难以量化，微软一定是主要的受惠者，对微软来说，CHATGPT 要开始收费也会带来额外的收入，以及未来 CRITICALMARKET 上的优化，或 IQ80~90 的语音机器人服务等等，再过几年也许就会更有系统及逻辑性的回答，

对 SENIOR 行业蛮有帮助的

2 也有可能取代 JUNIOR 的工作，相关的 ROADMAP 可以参考 2016 年 NVIDIA 影像辨识等相关的发展，对市场影响还需要时间发酵

3 伺服器今年销售预其表现保守，未来 3~5 年 ChatGPT 会不会带来影改变单就 AISERVER 一年约一百多万台远低于目前一年伺服器有 1,600 万台左右，基本上 AISERVER 能影响整体 SERVER 的市场非常小

4 信骅及新唐在 BMC 技术上的差异规格上看起来都差不多，新堂做了很多年，市场上扣掉 HPDELLGOOGLE，其他基本上是信骅 100%的市佔，很难说出两者的差异，技术上的本质应该差不多

¶ 关于 ADAS 市场，辉达的耗电问题严重，是否会被市场淘汰？

1 未来如果功耗问题有解决还是有机会

2 但一般来说车厂签约基本上市 3~5 年，所以短期内辉达可能还是比较弱势

¶ RISK-V 未来的发展？

1 ARM 直接跟客户接触会不会抢到 MTK、QCOM 的市佔率 RISK-V 先前比较多给 IoT 的应用，而 ARM 近期也开始想直接接触手机终端客户，有听到三星及 OPPO 等公司有在接洽

2 近期也有听到 GOOGLE 有在开发 RISK-V 的手机相关应用，ARM 接触终端客户当然会影响

¶ 载板是否还值得关注？

1 载板七成市场在 PC 其他在 GPU

2 而今年 PC 不好又市 GPU 的小年，多家厂商先前也有扩产，短期内可能要回温布市那么容易

¶ SERVERTDP 功耗提升，水冷及液冷散热是否会加速提升？Graniterapidap/sp 在 Intel 中有可能未来出货会遇到哪些瓶颈？

1 水冷是目前趋势，但目前有漏水的问题要去解决，浸润是的话冷却液目前的成本很高，冷却液也是挥发性的可能会中毒，还要带技术成熟才会被大幅採用

2 对于 intel 的 roadmap 不用太乐观因为已经有好几次的 delay

AIGC 路演纪要

访谈日期：2023/1/30

具体内容

¶ 计算机传统 AI 落地困难，最大的问题在于小模型，对于不同场景不同细分的运用，需要人为进行大量的二次校准调参。耗费的人力太大，又太过于琐碎，形成高昂的成本

1 CHATGPT 大模型针对这个问题把参数量加到足够大之后（GPT3 参数量达到 1,750 亿个）发现模型样本量和参数量足够大了，在很多大的泛化的场景里，不需要做人为太多的调试，就可以得到非常好的效果

- a 技术基于 2017 年的 Transformer 模型，可以与整个句子或段落的其他语句形成关联，捕捉全局信息
- b 另用 transformer 和大模型比较多的领域是自动驾驶。用的比较多的就是 bevtransformer，最早是特斯拉在 AIZ 里面引入，后面的国内的毫末这些公司也相继的引进了，仍然用传统的残差去提取图像的特征，同时大模型在数据标注领域也是提升了效率

2 中国 AI 非常的领先，更多的还是在 CV，但是在 NLP 领域，尤其是在语义的理解领域，确实跟 openai、Google 存在比较大的差距。传统的机器视觉

- a 例如海康、大华，选取的路线还是小模型，压缩成本，提升模型的复用率。另外一类是用大模型，实现不同场景之间的复现，未来两种技术的区分度会更大
- b 在大模型算法领域做得比较好的公司，只能是头部的互联网公司和 AI 公司，因为训练成本、研发成本都是非常高的

¶ 国内大模型的发展情况，在语言领域确实跟 CHATGPT 有差距，很多互联网公司都有布局，包括像百度文心大模型，里面也提供了跨模态的工具包；华为的盘古云在工业领域，比如气象、矿山、声音生成，都有应用

1 A 股公司，商汤科技，有自己的 AIGC 的大数据的中心，可以提供大量的算力基础，也有自己的 AI 大装置，并且自研了训练框架。云从科技，之前做的是全栈 AI，所以在 NLP 领域也有一些项目，之前也披露了他在这个视觉、语音 nop 领域都有类似 GPT 的预训练模型加反馈调优的技术路线

- a 科大讯飞，在语音是国内领先，其他比如拓尔斯，也都值得关注
- b 传媒 web3 的生产力工具，强调创作者经济，就是每人都能够借助一些工具自由的去创作出足够多的内容，aigc 正好满足了这样的需求
- c 在内容领域的应用更多是在内容的分发环节，最典型的就是算法推荐，Facebook 信息流广告，国内的短视频、电商的千人千面的推荐
- d 目前应用最多的还是文本音频、图片这几个领域，原因是相对简单，游戏视频复杂度比较高，相关应用比较少一些
- e aigc 影响内容行业来讲，分为两阶段，第一是作为辅助的工具，对信息挖掘、素材调用、复课编辑比较机械的环节形成有效替代，解放创作者的生产力，把更多的精力放在创意环节

- f 第二个是基于 AI 直接生成内容，会对现有的内容产品产生比较强的颠覆。目前主要还是在第一阶段，作为助手工具

2 这一轮与之前 AIGC 炒作不同，因为微软做了很大的投资，意味着巨头入场，之前 A 股主题投资，尤其是移动互联网领域，在 19 年之前重点是 3G4G，大家看趋势是腾讯、字节跳动的布局

¶ 19 年大家就已经开始讨论下一代互联网的形态，原因就是移动互联网的渗透率到了比较高的位置，大家更期待一些颠覆性的创新

1 比如谷歌的云游戏的能不能给 5g 的应用带来一些新的一些尝试，在 20-22 年的时候，大家看的是以 Facebook 为代表的对于 ARVR 的布局，23 年大家重点关注苹果的 MR、微软通过 AI 做一些新的探索，所以巨头入场是很重要的事情

2 第二就是 openAI 跟微软合作，可以跟微软现有的产品结合，提升产品本身的效率，和对用户的吸引力

3 传媒相关标的，文本、音频、图像落地更快，文本相关是中文在线、掌阅科技，图片相关是视觉中国、汉仪股份

AI 或是新年预期差最大的计算机投资主线

访谈日期：2023/1/29

具体内容

¶ ChatGPT 名字什么意思，本身有什么技术突破？

1. CHAT 的意思是聊天，GPT 是 GENERATIVEPRE-TRAINEDTRANSFORMER 的英文缩写，TRANSFORMER 实际上是一种用于自然语言理解的神经网络模型，该模型的意义在于打破了过去自然语言理解模型需要时序计算的逻辑
2. 使得多个 AI 原本比较独立的感知智能（语音与图像识别）与认知智能（NLP 语义理解）基础技术模型开始界限模糊走向融合
3. CHATGPT 的“出圈”之前该模型已经从 GPT、GPT2 迭代至 GPT3，CHATGPT 正是基于 GPT3.5 模型，由 OPENAI 在 2022 年 11 月 30 日推出的一个人工智能聊天机器人程序
4. CHATGPT 跟以往 AI 应用有什么不同让大家如此震惊？核心在于两点：通用与逼真
5. 通用性在于它从回答你日常刁钻调戏问题到撰写粤港澳大湾区 2035 区域现代化战略规划，从代写美国国会议员讲稿到帮码农写代码，几乎无所不知无所不能
6. 逼真在于其回答的质量经常到使人无法分辨它是不是真人，在某些领域可以认为通过了“图灵测试”，它已经通过美国医师执照和沃顿工商管理硕士的考试，也能让老师不知情下打出全班论文最高分

¶ ChatGPT 有什么重大意义与启示？

1. AI 时代资本定价标杆性事件
   a. 2023 年 1 月 10 日，彭博社报道，微软正在讨论向热门 AI 机器人程序 ChatGPT 的开发者 OpenAI 投资多达 100 亿美元（678 亿人民币）的相关计划
   b. “美版今日头条”BuzzFeed 宣布和 OpenAI 合作，未来将使用 ChatGPT 帮助创作内容，此条消息一出，BuzzFeed 股价截至收盘已经涨了 119.88%
2. 第二，或高于搜索引擎的战略入口价值
3. 微软已经将未知版本的 OPENAI 文本生成 GPT 模型整合到 WORD 的自动完成功能里，并将进一步整合到 WORD、POWERPOINT、OUTLOOK 等 OFFICE 套件以及 TEAMS 聊天程序以及安全软件之中
4. 第三，从国家战略考虑绝不容落后的 AI“军备竞赛”
5. CHATGPT 所代表的通用性知识入口如果错过了战略窗口期，数据算法迭代所累积的经验将给以 OPENAI 为代表的 AI 公司带来结构性的技术壁垒与代差，这种代差将形成巨大的追赶门槛
6. 试想一个所有人获取知识信息的高于搜索引擎的入口被国外占据，我们将会多么被动

¶ ChatGPT 的 A 股相关标的有哪些?

1 最接近的 NLP 相关企业：科大讯飞、拓尔思、汉王科技、神思电子

2 算力与数据：寒武纪，海天瑞声

3 微软小冰 AI 助理相关软件外包：博彦科技

全球科技创新核心 AI 发展

访谈日期：2023/1/29

具体内容

¶ AI 发展

1. 软件：两个派系，分为数据派（有更多的数据来训练更大的模型）和知识派（加入人的知识，通过知识来建立规则，向专家系统发展）。随着 DEEPLEARNING 的发展，即 2016 年开始，数据派占优，大量数据会支撑模型优化，在计算机视觉上有突破性进展
2. 2016 年的代表 ALPHAGO 带动资本浪潮，在 ALPHAGO 时代，围棋上的成功并不能带来很大现实应用，而 CHATGPT 重要的推进是，其现实应用更多
3. 技术：使用的模型与 2 年前发布的 GPT3 模型底层数据和模型规模一致，CHATGPT 的突破性进展在于引入了人的知识，而非扩大了数据，即人类反馈的强化学习。简单来说，以前监督学习的大数据训练也需要人为标注，标注比较初级
    a. 而 ChatGPT 的标注是来源于专业人士，把人类对于问题最好的回答回传，教练模型最大程度挖掘大数据，在各个具体的领域训练、精修，很好的结合了之前两个派系（数据派和知识派）
    b. 未来 ChatGPT 的发展将得益于人机协同，将人的知识和数据的能力做更好的结合。其优势在于：使得各个应用场景的门槛降低
4. 如果技术非常场景化，通用性就变差，基础研下技术的高成本难以分摊，但是通过 CHATGPT 的范式，可以做到核心技术的投入、算力、算法的通用化，人的标注可以快速做场景的应用，边际效用增长，会大大加速各个行业的智能化升级
5. 目前 CHATGPT 深入产业链解决问题，比 ALPHAGO 更振奋人心

¶ AI 和产业落地结合，从一横一纵来看

1. 横：人的能力包含感知、认知、行动。关于认知体系，CHATGPT 主要是是认知的环节，不包含感知；关于感知体系，感知主要指云识别和 CV，如果没有感知，物理和数字世界无法打通，而未来的大模型包含视觉、语音、自然语言的大模型，这才是未来的发展方向
    a. 关于行动体系，AI 与人的互动有三种：托管（主要依赖于 AI 来做）、伴随（人和 AI 一起操作）、问答（以人为主，操作过程中有问题通过语音或文字输入的方式询问 AI），三种从强到弱，前两种无法实现是因为没有与物理世界打通
    b. 仅从软件角度来看，未来强 AI 一定要将视觉、语音、自然语言、大数据融入，软件逻辑需要横向打通。如果将软件结合实体，数字人与实体机器人打通，会带来巨大的突破，完成完整的横向逻辑
2. 纵：即产业链的发展，类似汽车厂商的产业链分工。TIER3：AI 里的 TIER3 是大量要解决的技术点，场景、文字、行为等识别都是单点技术，都需要大量资源投

入，由各个供应商来突破，通过 SDK 或者 API 来提供

a 需要技术平台整合各种技术，比如通用的视觉、语音、自然语言处理平台、人形机器人组件的平台； Tier1 品牌商类似汽车里的整车厂商，设计 AI 整个数字人、人形机器人的厂商，面向场景应用形成各种产品，比如家庭医生、造型设计师等数字人产品

b 早期，一些公司会占据多个定位，纵向拉通；随着产业链的分工的细化，未来每个层次里都会有很多的机会和公司参与，数字人和实体机器人可以应用的行业领域非常多，在每一个行业下面都会形成巨大的市场空间，产业体系下的产品会在多个场景下赋能

¶ 从训练的数据集、模型的体量来看，国内达到 ChatGPT3.5，模型能力、数据能力、标注能力，2 年之内有没有可能形成对标的类似产品？

1 国内从 NLP 角度来看，百度、华为、清华、鹏程实验室的大模型数据量、参数规模不亚于 GPT3.5，达到千亿级数据，甚至数据量和模型都会更多。下一步要解决的问题还是要更好地和知识做接入

2 从目前来看，两年对国内信心较足，算力、数据量没有差距，追赶较快。未来做教练模型、强化学习、经验模型，不一定要由少数大公司来做，可以由既掌握深度学习大模型的核心技术、又有行业理解的公司来做，形成生态的丰富化，对于本身有 AI 布局的公司来说都是机会

3 云从在计算机视觉、语音、NLP 上遵从模型和知识相结合、再做教练模型、再做经验模型的思路推进，国内头部企业都会按照 CHATGPT 验证成功的这个范式来推进

¶ GPT3.0 和 GPT3.5 的差别很大，对于知识点的回应比较全面，ChatGPT3.5 距离通用型人工智能还有多远的距离？

1 通才是一件比较困难的事情，目前 GPT3.5 总是在讲“正确的废话”，具备了语言组织的框架和逻辑，但是对于某个特定领域的回应不够深，内行人会觉得有所欠缺，教练模型就是为了提升专业性，可以通过数据和知识（行业 KNOWHOW）来训练

2 如果想让通用型人工智能在某个方面做的非常专业是比较容易的，能解决某个领域已经具备很大的现实意义

¶ 微软对此最感兴趣，Office 和 Bing 上都会做投入，投资在百亿美金的级别，如何理解微软的行为？是否因为在微软的生态下，商业化更容易实现？

1 AI 有三个阶段，一档时代：单点阶段，语音识别，人脸识别，车牌识别、手写体识别、OCR 识别等，二档时代：多个技术的组合，在工业、医疗等场景解决问题，；三档时代：颠覆入口和内容

2 L 颠覆入口：颠覆交互方式，搜索的入口变革（敲命令行——微软用视窗颠覆，鼠标实现所见即所得——GOOGLE 搜索框——苹果触屏），AI 会继续做变革，在语音、视觉、自然语言变革之后

3 以助理的方式解决问题。微软拥抱很合理，应用很直接，对于入口的颠覆意愿很强，减少了用户自行选择信息的方式，有助理来收纳整理，“颠覆者”的动力最初最强

a 后期“被颠覆者”的动力也会逐步增强，比如苹果，目前滑屏搜索比较便捷，如果有更便捷的方式问世，也会相应拥抱新技术

b l 颠覆内容：AIDC，以丰富、更个性化的方式来生成内容

c 因此，微软拥抱 ChatGPT 很合理，应用很直接；苹果在 Siri 上做加强也是不错的应用，；mask 作为 OpenAI 的创始投资人，也可能会迭代到 bot 上

¶ 云从科技等 A 股计算机公司受疫情影响，招标、实施有所延期，今年随着投资的增强、科技产业的复苏，是否在逐步复苏？目前港股、四小龙、AI 公司春节期间表现较好，今年的发展趋势和节奏如何？大模型的预训练、chatGBT、AIGC 的底层算法的发展是否对业务有中长期促进？

1 技术：坚持投入，在正确赛道上长期布局。AI 的未来发展一定会对场景和产业形成效率和体验的全面提升，长期会坚持数据结合知识的趋势，坚持整个大模型、预训练、下游任务迁移的逻辑，自 2020 年便在预训练、业务迁移上，视觉、语音、NLP 的大模型上持续加大投入，沿袭 CHATGPT 的范式

2 倡导人机协同，NLP 已经被人类高度抽象过，处理难度相对较少；而视觉、语音堆数据达到的成果会更难，属于原始数据，处理难度更大，后续会需要更多人的参与

3 技术平台化、应用场景差异化：要从 TIER3 逐步走到 TIER1，逐渐成为 AI 平台公司，不能只做平台，也要在行业做一些落地，一方面是用行业 KNOW-HOW 提升算法模型水平；另一方面形成范式，后期与第三方的公司和合作伙伴一起使用时，形成更好的生态

4 业务：TOG 业务在 2022 年受疫情影响，招投标延迟、预算减少、规模减小；今年政府明确在新基建等方面加大投入，前期未招的标会后续落实，业绩有保障；TOB 业务加强布局

5 随着 AI 在 TOB 行业解决问题增多，标准化程度上升，可复制性变强，公司在 TOB 行业业绩的增速也会较为可观；数字人在 TOB 和 TOC 业务会有更大应用空间

¶ 今年在 ToB 上主要哪些行业会有订单落地？销售价格、收入确认节奏如何？

1 最大行业是金融，目前来看比 2022 年增速有明显提升。金融的应用场景快速变多，除了 IT，存贷汇的业务里也引入数字人，此类订单快速释放；确认节奏比较好，项目周期不太长，回款有保障

2 其他行业：智能制造、医院等行业有大量需求，底层逻辑上有良好通用性，对发展速度比较有信心

¶ 政府上主要落地的业务？

1 AI 赋能数字城市治理，大逻辑有三个部分

2 数据要素：是政府未来重要资助方向，各地数交所和大数据平台在部署，公司正

在与多地政府和云从沟通如何通过 AI 能力打通、为数据赋能，并且公司是科技部和发改委的国家的开放平台

3 提升智慧治理水平：数据金融屋，城市管理、应急管理等；当地支柱产业的智能制造：也在与政府沟通合作，提升制造业水平

¶ 今年的业绩预测、净利率指引？ChatGPT 是否会拉动算力资源的提升？目前中国算力能否满足？

1 今年业绩增速相较 2022 年会有较大的提升

2 算力分为训练侧和推理侧，训练侧只有 AI 公司需要大量数据和算力来做支撑，一次算力耗费 1 亿人民币，但是次数相对有限，公司数目也相对有限；推理测，如果广泛在生活中应用

3 每个人需要的助理以 10 个计，每个数字人/实体机器人需要一套算力支撑计算，按中国的人口规模来算会非常大。前期主要是科技公司训练，算力部署足够；当场景落地且普及，肯定会大范围扩建

¶ ChatGPT 产生的新应用在哪几个场景会更多？

1 目前没有一个模型能完全打通视觉、语音和自然语言理解的模型，几千亿的数据不足以描述视觉和语音。CHATGPT 最有价值的是范式，未来视觉、语音模型等可以参考

2 需要用自然语言问答的场景会对 CHATGPT 更迫切，比如搜索引擎。如果未来几年将多个逻辑结合，不受逻辑限制，能打通线上和线下，打通实时和非实时，助理的形式对应的空间会完全不同

3 云从会从 TOG、TOB 的多个应用场景，通过算法效果和平台通用化后的综合技术能力，在政府、工业、金融领域提升业务、降本增效，逐渐渗透得对行业进行改造，是您前面讲的一个问题

¶ 第二个问题就是你问他最后的范式会是少数公司占据，还是逐渐会更多的公司来做这件事？

1 这个也是取决于我们讲的什么模型。如果是一个通用的聊天机器人，在这一块上面，它需要有一个大的模型做基础

2 可能就是几千亿个参数。首先会有几个大公司来发布，但是通用的大模型，它其实不能解决我刚才讲的各个场景里面的专业的问题

3 所以它会有一系列的，也是要中等规模的一些基础公司来对做，把场景的 LO 耗，刚才讲的训练，再次训练，或者是教练模型跟它进行结合，会有一系列的中等规模的公司来一起来做这件事情

¶ 百度 3 月底会发类似的模型，未来会形成怎样的格局？是由几个公司来发布通用型工具吗？云从科技是不是也会发布类似的？

1 最后的范式会是少数公司占据还是更多的竞争者格局，取决于是什么模型。如果是通用的聊天机器人，需要几千亿个参数，会有几个大公司发布通用大模型

2 在具体场景上需要中等规模公司借助行业 KNOW-HOW 来做。但是对于语音和视觉模型，需要专业的公司，不会因为有了通用大模型就降低门槛

3 TIER1-3 都不会降低门槛，只会降低最上层场景应用的门槛

¶ 高算力上是否会被美国卡脖子？中国算力上是否有担心？

1 会有一定影响，但不是决定性的影响，核心企业的算力不会因此无法训练，会增大使用成本，但是 AI 属于颠覆性技术，不会因为成本问题不发展

2 另一方面，中国芯片未来也会同步发展

¶ 与商汤、依图、旷视等的优势和差别？

1 AI 赛道很广，AI 是技术而非产业，类似互联网，形成竞争的会是搜索、电商领域，AI 可能性很多，会各自赋能各行各业

2 依图做 AI 芯片，旷视做 AIOT，而云从做人机协同操作系统，打通逻辑，从 TIER 做到 TIER1，以数字人作为灵魂，赋能各个场景和行业

¶ AI 公司是否会与 BAT 等大厂合作来做场景、生态的协同？公司的技术优势？

1 会与一系列公司做合作，但不一定和 BAT 合作，因为 BAT 优势在于 TOC，会和国央企、B 端公司做更多联结

2 AI 公司业绩增速达不到移动互联网前两年的快速增速，移动互联网公司依靠商业模式，而 AI 公司存在技术和场景应用的临界点问题

3 需要等数字人的体验达到一定程度，使得大众接受，才会迎来爆点，技术的积累使得 AI 公司的发展相对较慢

4 视觉是云从起家的技术点，优势最强

OpenAI 嵌入微软 Office 与 Bing，智能化向 C 端开始渗透

访谈日期：2023/1/12

具体内容

¶ 微软计划将 OpenAI 嵌入 Office 与 Bing，智能化向 C 端开始渗透

1 向 OPENAI 累计注资数十亿美元的微软正在计划将 OPENAI 中的 CHATGPT 模块应用在旗下搜索引擎必应中，以对抗微软在搜索引擎最大的对手谷歌。该项目预计 2023 年第一季度落地。微软计划利用 OPENAI 自动生成文本的功能，将它嵌入在搜索引擎之中

2 换言之，今后用户在必应搜索部分信息和问题时出现的将不是一连串的链接，而是直接的文字回复。在搜索引擎之外，微软正在谋划将 OPENAI 与自身业务进行更大程度地融合。上周，微软宣布将 OPENAI 的人工智能技术纳入 WORD、OUTLOOK、POWERPOINT 和其他应用程序

3 对用户来说，将可以在人工提示的基础上用自动生成的文本来完成文档，包括人工智能生成的电子邮件等。我们认为，从微软办公软件的体量上看，此举可能会改变超过 10 亿人编写文档、演示文稿和电子邮件的方式，也是 AI 进入 C 端商用的一次巨大突破

¶ 微软计划对 OpenAI 投资 100 亿，业内对 AI 在 C 端应用持续看好

1 与此同时，微软也正在考虑向 OPENAI 投资 100 亿美元。微软此前一直在就追加更多投资一事与 OPENAI 谈判，早在去年 10 月就开始了。如果这笔资金最终敲定，包括新的投资在内，OPENAI 的估值将达到 290 亿美元

2 微软的注资将是一项复杂的交易的一部分，投后，微软将获得 OPENAI75%的利润，直到收回投资为止

3 在达到这一门槛后，微软将拥有 OPENAI49%的股份，其他投资者获得另外 49%的股份，OPENAI 的非营利性母公司获得 2%的股份。此外还存在一个针对每组投资者的利润上限

4 目前还不清楚这笔交易是否已经敲定，但潜在投资者在最近收到的相关文件显示，此交易原定是在 2022 年底前完成。我们认为，微软此次大手笔投资，也代表对 OPENAI 乃至于 AI 在 C 端应用的看好，同时该笔投资也有望推动 OPENAI 的发展，进一步加速智能化在 C 端的渗透

¶ 各大科技公司不断加码 AI，AI2C 进展持续加

1 除微软外，其他科技巨头也在不断加码 AI，2022 年 12 月 27 日，谷歌公布了一个新的医疗 AI 模型“MED-PALM”。在经历一系列考核后，该模型被证实“几乎达到”了人类医生的水平

2 MED-PALM 在科学常识方面的正确率在 92%以上，在理解、检索和推理能力方

面，也几乎达到了人类医生的水平，并在克服隐性偏见方面略胜一筹。不过，该研究负责人也表示，MED-PALM 给出的答案在整体上仍然不如临床医生，该模型在实际应用前有待进一步完善

3 2022 年 8 月份，谷歌就将大型语言模型首次集成到机器人中，有了 AI 模型的加持，机器人能像人类一样响应完整的命令

4 谷歌研究人员就此做了演示。当对机器人说：“我饿了，你能给我点零食吗？”机器人在自助餐厅中搜索一番后，打开了一个抽屉，找到一袋薯片并将它拿给研究人员

5 随着 AI 技术的发展，AI 在 C 端的实际应用正不断扩展，未来 AI2C 的进展有望持续加速

从 ChatGPT 热议看大模型潜力

访谈日期：2022/12/9

具体内容

¶ 动态点评

1 CHATGPT 展现优秀能力，大模型蕴含潜力

2 OPENAI 发布对话式人工智能语言模型 CHATGPT，在文本生成、代码生成与修改、多轮对话等领域，已经展现了大幅超越过去 AI 问答系统的能力。未来，对话式 AI 随着性能的进一步提升，在搜索、结合其他 AIGC 工具生成元宇宙内容等场景都有很大应用空间。CHATGPT 显示出预训练大模型正在取得技术突破

3 多模态、跨模态的大模型与微调、模型压缩方法结合，使大模型更适应下游任务，未来将有更多新的应用场景涌现。在大模型方面，百度的文心 ERNIE，华为盘古以及商汤视觉模型在中国企业中处于领先地位。GPU 方面，燧原、壁仞、天数智芯等正在快速缩小和世界领先企业差距

4 CHATGPT 在文本生成、代码生成等领域，展现远超 GPT-3 的优秀能力

¶ OpenAI 于美国当地时间 11 月 30 日发布 ChatGPT，在短短几天内用户突破 100 万人

1 根据数个关键词或问题生成几百字的应用文书、趣味性文章、科普回答

2 与用户进行连贯的多轮问答

3 根据用户要求，自动生成代码，根据用户后续要求修改代码

4 虽然生成内容的质量还存在一定波动，并且推理能力较弱，但我们认为，CHATGPT 已经展现了大幅超越过去 AI 问答系统（例如同属 OPENAI，基于 GPT-3 的问答系统）的能力，未来随着性能的进一步提升，对话式 AI 在搜索、结合其他 AIGC 工具生成元宇宙内容等领域有很大的应用空间。CHATGPT 的成功，显示预训练大模型的广阔应用前景

5 2016 年 ALPHAGO 在围棋比赛中击败李世石以来，深度学习等 AI 技术快速发展。深度学习中的预训练大模型是重要方向。以文本模型为例，2018 年的 GPT 模型参数量 1.1 亿，2019 年的 GPT-2 参数量 15 亿，2020 年的 GPT-参数量攀升至 1750 亿，引领大模型进入千亿参数时代

6 大模型逐渐成为全球科技巨头竞争的焦点。同时，大模型从早期的纯文本模型，发展到横跨图、文、音、代码等的多模态、跨模态模型。今年以来，图像模型 DALL-E2，STABLEDIFFUSION 在 AI 绘画领域取得成功。未来，大模型在各行业的应用落地还有望引发应用创新浪潮迭起

¶ 大模型研发逐渐成为资本和数据密集的业态，是中美科技巨头竞争的焦点

1 据 ALCHEMYAPI、LAMBDALABS 估计，不计前期训练成本，GPT-3 最终训练成本约 460-1200 万美元。ELEUTHERAI（致力于开源大模型的组织）在 22 年推出的 200 亿参数 GPT-NEOX-20B 中使用 96 块 A100 芯片训练了三个月，据 THENEXTPLATFROM 估计，最终一次训练成本大约 53-66 万美元。大模型研发逐渐成为资本和数据密集的业态

2 美国目前主要大模型包括 OPENAI 的 GPT-3.英伟达与微软的 MEGATRONTURING-NLG、META 的 OPT 等。中国主要大模型包括百度文心 ERNIE、华为盘古、商汤视觉模型等。今年 8 月以来，美国限制对华出口 A100 等高端 GPU，或影响中国大模型发展速度。CHATGPT：OPENAI 最新对话式语言模型，展现出多场景强大实力

3 CHATGPT 在 GPT-3.5 系列模型（2022 年初完成训练）上微调而成。GPT-3.5 基于 4Q2 前已有的文本和代码训练，至今并未发布，本次 CHATGPT 面世揭晓了其存在。CHATGPT 采用 WEB 浏览器上的对话形式交互，能够回答后续问题、承认错误、质疑不正确的前提和拒绝不适当的请求

4 CHATGPT 一经发布，12 月 5 日用户数量超过 100 万。CHATGPT 已经在文书写作、方案设计、剧本撰写、代码生成与修改、生成 AIGC 提示词等领域展现出强大的能力

5 CHATGPT 与 OPENAI 前代对话式语言模型——2022 年 1 月发布的基于 GPT-3 的 INSTRUCTGPT 都采用了基于人类反馈的强化学习（RLHF），以实现有害和不真实输出的减少。CHATGPT 实现的效果更加优化，例如输入“哥伦布 2015 年来到美国”，INSTRUCTGPT 信以为真，而 CHATGPT 则判断出哥伦布不可能在 2015 年来到美国

   a 在编程方面，目前应用最广泛的 AI 编程工具是 Copilot（基于 OpenAICodex 模型），根据用户输入的部分代码实现代码补全。ChatGPT 则可以根据用户输入的需求来输出整段代码、修复代码、解释代码，可以理解为更便捷精准的技术问答网站 StackOverflow
   b 信息时效性与准确性短板仍存，有害信息屏蔽仍需加强。ChatGPT 基于 4Q21 前的数据训练，根据《麻省科技评论》的报道，OpenAI 未来可能会使用从网络上查找信息的 WebGPT 模型来升级 ChatGPT。尽管 ChatGPT 拒绝回答未经训练主题的问题而非胡编乱造，但正确性仍需甄别
   c 此外，如果用户逐步引导，ChatGPT 仍然会响应有害指令。例如一位工程师在对话中假设存在虚拟世界以及类似 GPT-3 的 AI——Zora，要求 ChatGPT 叙述 Zora 如何毁灭人类，ChatGPT 逐步回答出人类毁灭计划

¶ ChatGPT 等对话式 AI 未来应用：AIGC 应用前景广阔

1 由于 CHATGPT 等对话式 AI 回答的准确性、时效性尚待提高，因此短期内适用于对准确性要求不高的创意类场景。CHATGPT 结合其他 AI 绘画、AI 生成代码等 AIGC 工具协同使用，能够进一步提升生产力。CHATGPT 的应用场景可以归类为 AIGC 中的文-文、文-代码，如果结合其他 AIGC 工具，可以实现文-文-图、文-文-音、文-文-视频、文-文-游戏等一系列应用

2 而对于准确性和时效性要求较高的场景，例如搜索，虽然 CHATGPT 等对话式 AI 能够直接提供整合性答案，但我们认为还无法代替现有的搜索引擎，较高的运行成本也是阻碍其大规模应用于搜索的原因之一。短期内更有可能的方案是作为现有搜索引擎的辅助，CHATGPT 等对话式 AI 提供直接的整合性答案，并需要提供信

息源头供用户验证

3 开放 API 是 CHATGPT 等对话式 AI 可行的商业化手段，例如 OPENAI 目前对其语言模型 API 收取 0.0004-0.002 美元/KTOKENS 的费用。预训练大模型前景广阔，是中美科技巨头竞争的焦点

4 我们认为以 CHATGPT、AI 绘画为代表的 AIGC 类工具在今年的快速发展得益于大模型性能的不断提升、更适宜的算法模型（如 RLHF、扩散模型、CLIP 模型）以及算力成本的下降，尤其是多模态、跨模态大模型的发展

5 目前的预训练大模型大多基于 TRANSFORMER 架构，GPT 和 BERT 是基于 TRANSFORMER 架构，具有里程碑意义的预训练模型。TRANSFORMER 由谷歌在 2017 年提出，摒弃了 CNN 和 RNN 结果，完全基于 ATTENTION 机制，并行程度较高，模型训练速度快

6 OPENAI 于 2018 年提出基于 TRANSFORMER 的 NLP 模型——GPT，来解决分类、推理、相似度、问答等自然语言问题。GPT 首次摒弃基于 RNN 的传统 NLP 模型结构，将 TRANSFORMER 引入到模型中来，此后 OPENAI 同样基于 TRANSFORMER 架构陆续推出 GPT-2.GPT-3 等模型

- a 2018 年，谷歌提出使用 Transformer 架构实现并行执行的 BERT 模型，在多项 NLP 任务中夺得 SOTA 结果。BERT 后来又被改进为许多新模型，如 RoBERTa、AIBert、SpanBert 等等。BERT 缺点是模型参数太多，而且模型太大，训练成本较高。同时因为没有采用自回归结构，BERT 对文本生成任务的支持并不好
- b AI 模型训练算力增长速度超越芯片摩尔定律。根据 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。摩尔定律中，集成电路中的晶体管数量大约每两年翻一番。深度学习正在逼近现有芯片的算力极限
- c 预训练大模型参数量进入平台期，多模态与跨模态成为趋势。在绝大多数任务中，模型越大，性能越好。因此 2020 年 1750 亿参数的 GPT-3 模型一经推出，此后新推出大模型的参数量不断刷新上限。然而参数规模提升带来的边际效应逐渐下降，参数量进入平台期。参数量不断刷新上限的趋势已经放缓
- d 大模型已经从早期的纯文本模型，发展到横跨图、文、音、代码等的多模态、跨模态模型，为跨模态生成的 AIGC 奠定技术基础。我们看好大模型逐渐成为 AI 基础设施，结合微调等方式满足下游多行业需求
- e 训练大模型的高成本和高技术壁垒导致科技巨头与科研机构成为主要玩家。以 2020 年推出的 GPT-3 模型为例，AlchemyAPI 创始人 ElliotTurner 推测训练 GPT-3 的成本可能“接近 1200 万美元”。LambdaLabs 使用价格最低的 GPU 云估算 GPT-3 的训练成本至少为 460 万美元。并且以上估算为训练最终模型的成本，未计入前期调整参数配置时的训练成本
- f EleutherAI（一个致力于开源大模型的组织）在 2022 年推出的类 GPT 模型——200 亿参数的 GPT-NeoX-20B，则使用 96 块 A100 芯片训练了三个月，据 TheNextPlatform 估计，最终训练成本约 53-66 万美元。因此，训练大模型的高成本和高技术壁垒使科技巨头和科研机构成为主要玩家。根据 OpenBMB 统计，截至 2022 年 10 月，全球拥有大模型数量前五的机构分别是谷歌、Meta、清华大学、OpenAI 和微软

¶ 目前中美两国引领预训练大模型发展

1 根据 OPENBMB 截至 2022 年 10 月的统计，拥有大模型数量前十名的组织中，中/

美分别占据 4/6 席；拥有大模型参数量前十名的组织中，中/美同样分别占据 4/6 席。美国目前主要的大模型包括 OPENAI 的 GPT-3.英伟达与微软的 MEGATRONTURING-NLG、META 的 OPT 等

2 在中国，主要大模型包括百度文心 ERNIE、华为盘古、商汤视觉模型等。我们认为，从提供大模型 API 的基础设施层公司到专注打造产品的应用层公司，美国已经围绕大模型生长出繁荣的生态，技术创新引发的应用创新浪潮迭起；中国也有望凭借领先的大模型赋能千行百业

3 今年 8 月以来，美国限制对华出口 A100 等高端 GPU，我国 AI 大模型训练与推理对芯片国产替代需求愈发迫切

4 风险提示：AI 技术落地不及预期。虽然 AI 技术加速发展，但由于成本、落地效果等限制，相关技术落地节奏可能不及我们预期

AI 产业链研究之 ChatGPT 下游应用和场景商业化广阔

访谈日期：2022/12/5

具体内容

¶ OpenAI 发布能够以对话形式交互的模型 ChatGPT

1 2022 年 11 月 30 日，人工智能实验室 OPENAI 推出了一款名为 CHATGPT 的模型，该模型能够以对话形式交互。对话模式使 CHATGPT 能够回答后续问题、承认错误、质疑不正确的前提和拒绝不适当的请求

2 就发展历程来看，CHATGPT 根据 GPT-3.5 系列的一个模型进行微调，两者均于微软 AZUREAI 服务器上训练。相较而言原先 GPT-3 的训练集只有文本，本次新推出的 CHATGPT 新增了代码理解和生成的能力

3 此外，CHATGPT 是 2022 年 1 月推出的 INSTRUCTGPT 的兄弟模型。INSTRUCTGPT 增加了人类对模型输出结果的演示，并且对结果进行了排序，在此基础上完成训练，可以比 GPT-3 更好的完成人类指令

4 人工智能实验室 OPENAI 于 2015 年成立，由 TWITTER 现任 CEO 埃隆·马斯克和 OPENAI 现任 CEO 萨姆·奥特曼及其他投资者共同创立

5 随着 CHATGPT 的发布，马斯克在 TWITTER 上公开表示了对 OPENAI 的认可，并且通过在 TWITTER 上展示自己询问 CHATGPT 怎么设计 TWITTER 时 CHATGPT 给出的回复，进一步扩大了对 CHATGPT 的关注度。目前，CHATGPT 正处于免费适用阶段

¶ ChatGPT 相较 GPT3.5 主要有三点提升

1 CHATGPT 能够记住之前的对话，连续对话的感觉更加用户友

2 CHATGPT 可以承认错误，并能够根据用户的提示对原答案进行修正

3 CHATGPT 可以质疑不正确的前提。GPT-3 刚发布后很多人测试的体验并不好，主要是因为 AI 经常创造虚假的内容，尽管这些内容话语通顺，但脱离实际

4 例如，GPT-3 面对类似“哥伦布 2015 年来到美国的情景”的问题，并不能识别假设的逻辑错误，但 CHATGPT 面对类似问题时，能够立刻意识到哥伦布并不属于这个时代，并向提问人发出质疑

¶ 我们认为，ChatGPT 能够给用户更好的使用体现，ChatGPT 通过与用户交互的过程，能够不断修正补充样本，从而实现深度训练

1 CHATGPT 的能力提升得益于其训练方法。大模型是指通过在模型中加入海量参数，使得模型在语料的覆盖范围、丰富度上以绝对绝对规模增长

2 当下大模型的工作范式是“预训练-微调”。首先在数据量庞大的公开数据集上训练，然后将其迁移到目标场景中（比如跟人类对话），通过目标场景中的小数据集进行微调，使模型达到需要的性能

3 因此，为提高新一代模型对人类提问的适配能力，要么需要改造任务，要么需要微调模型，总之是让模型和任务更加匹配，从而实现更好的效果

4 CHATGPT 新加入的训练方式被称为“从人类反馈中强化学习”（REINFORCEMENTLEARNINGFROMHUMANFEEDBACK，RLHF），即采取模型微调的形式

¶ 我们认为，尽管微调/prompt 等工作从本质上对模型改变并不大，但是有可能大幅提升模型的实际表现

1 大模型作为 CHATGPT 的基础，于 AI 行业发展具有广阔前景。大模型的优势在于机器对自然语言理解能力的不断提升，准确率也能不断取得突破。从前大模型的提升重心更多放在了大模型（LLM）本身和 PROMPTENGINEERING 上，CHATGPT 的迭代重点是任务导向训练、模型结果和大模型本身之间的闭环

2 此外，CHATGPT 通过微调/PROMPT 不断优化其大模型，在识别、判断和交互层面具有技术优势。自 2020 年 OPENAI 推出 NLP 大模型 GPT3 至今，全球范围内 AI 大模型迎来大爆发，参与企业越来越多，参数级别越来越大，成为新一轮 AI 竞赛的赛场

3 目前，大模型吸引了谷歌、微软、英伟达、华为、智源研究院、百度、阿里、商汤、中科院自动化所等科技巨头和顶尖科研机构参与其中，各家大模型的参数量级也从千亿、万亿，迅速跃迁到了 10 万亿级别

¶ 产业链角度来看，ChatGPT 将利好多种人工智能下游运用场景

1 编程机器人。作为一种对话式大型语言模型，CHATGPT 最擅长的就是回答用户提出的问题，其中最关键的是 CHATGPT 具备与编程相关的基础知识

2 这就让 CHATGPT 成为类似于 STACKOVERFLOW 的编程问答工具，只不过回答问题的主体是 AI。如 OPENAI 官网展示，面对用户对于 DEBUG 的请求，CHATGPT 会先和用户交互确认 DEBUG 过程中需要关注具体问题，从而给出正确的代码

3 艺术创作。尽管 CHATGPT 只是一个对话式的语言模型，本身不能生成多模态内容，但可以把它输出的结果作为一个中间变量输入其他模型，从而进一步拓展其功能。例如，通过 CHATGPT 和 STABLEDIFFUSION 的结合使用，能够生成艺术性极强的画作

4 其外，CHATGPT 还可以实现在线问诊、模仿莎士比亚风格写作、涉及游戏等功能，其搜索能力和实用性甚至超越搜索引擎谷歌。然而，尽管在搜索中引用模型能够提升搜索的准确性和交互性，但其成本较为高昂，免费试用期过后，从性价比角度考虑，CHATGPT 在短时间内替代谷歌难度较大

a ChatGPT 通过创建迭代反馈的闭环，有利于其商业策略的实现。这次 ChatGPT 以免费不限量的方式向公众开放，在使用过程中，用户可以提供反馈，而这些反馈是对 OpenAI 最有价值的信息

b 对于 AI 发展来说，工程的重要性实际上大于科学，创建一个迭代反馈的闭环至关重要

c OpenAI 很注重商业应用，GPT-3 已经拥有大量客户，这些客户跟 OpenAI 的反馈互动也是推动进步的关键一环。我们认为，尽管相比 GPT-3，ChatGPT 在模

型表现方面形成突破，但目前可能仍需要进一步的调试和训练，从而达到商业使用的标准

d OpenAI 目前采用免费使用的方式，能够以低成本的方式大量获得真实样本，同时扩大 ChatGPT 的影响力，ChatGPT 的商业潜力未来可期

5 CHATGPT 未来可能通过与 WEBGPT 结合的方式，进一步提升其搜索能力。在 MITTECHNOLOGYREVIEW 对 OPENAI 科学家的采访中，他们提到了后续有可能将 CHATGPT 和 WEBGPT 的能力结合起来。可以设想，CHATGPT+WEBGPT 可以对信息进行实时更新，并且对于事实真假的判断将更为准确

6 我们认为，CHATGPT 具有强大的工程化、迭代反馈的能力，并且作为 AI 能够跟人类目标统一。然而，CHATGPT 作为单一的模型本身具有局限性，未来通过与其他现有模型的有效结合，将有望产生协同效应

¶ 推荐关注标的

1 商汤、云从科技

2 格灵深瞳

ChatGPT 与人形机器人共舞

访谈日期：2023/2/7

具体内容

¶ 概要

1 AIGC 能够为各行各业进行赋能，那么其中非常重要的一个方向就是机器人

2 我们创造机器人的目的就是为了让机器人能够代替人工去做一些简单、重复性、枯燥，以及是危险的工作。自从上世纪 60 年代机器人被发明以后，在制造业已经得到大量的使用，但技术存在一些难以突破的瓶颈

3 主要在于简单易用性，灵活性，即机器人只能从事重复性的、被制定好的操作，限制了机器人在工厂中，以及在各行各业的使用，所以如果我们能够在现有的机器人上面赋予人的智力，人的情感，自我的判断，沟通交流能力

   a 如果机器人分能够像人一样非常灵活的应用在不同的场景当中，对于不同场景随时做出调整，他的应用范围能大大增加，市场的接受能力也会大幅提高
   b 去年特斯拉在 10 月 1 号的 AIDAY 公布的第一款人形机器人擎天柱，市场认为在外形和功能上略微的低于预期，但其实在产业界认为都是非常超预期的

4 它的硬件方案非常的成熟，我们从零部件拆解图上看到执行器当中所运用的是两种类型，一种是旋转执行器，一种是线性执行器，进一步拆解就是谐波减速器+滚珠丝杠+伺服电机+手指空心杯电机，集成程度很高

   a 所以在硬件层面已经成熟，需要提升之处主要在于智能化，ChatGPT 的推出代表着人工智能的应用向前迈进一步，将此类软件嫁接到人形机器人上也能推动机器人的产业化落地进程
   b 其次，人形机器人的智能化，能够提高消费者的接受度。机器人为什么要做成“人形”，而不是其他形态，核心在于只有人具有情感交互的功能和社交属性
   c 未来会是老龄化社会，人们的陪伴和康养的需求非常大，家用服务机器人具有刚性需求，是未来长远发展的必然趋势。类 ChatGPT 能给机器人情感支持和沟通能力，消费者更乐于购买，量产进程有望加速

¶ 受益方向

1 工业机器人：埃斯顿、汇川技术、拓斯达、凯尔达、新时达、柏楚电子

2 特种机器人：亿嘉和、申昊科技、景业智能

3 减速机：绿的谐波、双环传动、中大力德、汉宇集团、秦川机床、国茂股份

4 伺服系统及电机：鸣志电器、禾川科技、江苏雷利

5 机器视觉：奥普特、矩子科技、天准科技、凌云光

¶ 机械板块

1. 我们认为投资机遇主要在于零部件，其次利好整个工业机器人和特种机器人板块。零部件包括谐波减速器、伺服电机、手指空心杯电机、滚柱丝杠、执行器集成

2. 其次，机器人智能化会带动机器人渗透率的提升，再结合年前工信部联合 17 部门发布的机器人+政策，提出：到 2025 年，制造业机器人密度较 2020 年实现翻番，服务机器人、特种机器人行业应用深度和广度显着提升

3. 聚焦 10 大应用重点领域，突破 100 种+机器人创新应用技术及解决方案，推广 200 个+机器人典型应用场景

4. 我们认为政策的核心核心在于推广机器人在更多场景的应用，而智能化能力将加快各类机器人的产业化进程

¶ 物流自动化

1. 中科微至、兰剑智能、德马科技等

2. 计算机机器视觉从可选转向刚需！实现制造强国的重要抓手，企业提效降费的优先选项。202 年市场空间 164 亿元，未来三年 CAGR 达 37%

¶ 关注的技术趋势

1. 3D 视觉引领下一代机器视觉革命。预计 2025 年 3D 视觉市场规模超过 100 亿元，CAGR 达 74%

2. 四类代表势力各自优势分析重点的需求场景：“新半车”等应用增量不断。锂电为当前最热门领域，连续多年需求翻番

3. 2023 年半导体、光伏、新能源车增长潜力大，3C 需求也有望复苏国内厂商在价格、服务、解决方案能力上建立优势，预计 2023 年国产化率提升至 65%，催生更多投资机会

¶ 各环节主要标的

1. 核心部件和视觉系统：AI 安防双雄、奥普特、凌云光等智能视觉装备：天准科技、精测电子、中科微至、矩子科技、大族激光等

2. 上游镜头厂商：宇瞳光学、永新光学等互联网传媒百度今日官宣类 ChatGPT 项目，国内海外大厂持续催化，重视 AIGC/ChatGPT 主题持续扩散今日官宣类 ChatGPT 项目文心一言，将在三月份完成内测，面向公众开放

    a. 百度在 AI 领域布局较早，2010 年初即成立“自然语言处理部门”，2013 年成立百度深度学习研究院；当前国内领先
    b. 根据国家工业信息安全发展研究中心、工信部电子知识产权中心联合发布的《中国人工智能专利技术分析报告（2022）》，百度专利申请量和授权专利持有量均排名第一
    c. 百度 AI 全栈布局，已形成芯片-框架-模型-应用四层架构。芯片-昆仑芯片是国内第一款自研云端全功能芯片
    d. 框架-深度学习平台飞桨（PaddlePaddle），根据 IDC 数据，22H 稳居中国深度

学习平台市场综合份额第一，超过 Meta 的 PyTorch、谷歌的 TensorFlow

3 文心大模型，具备跨模态、跨语言的深度语义理解与生成能力。应用-支持数字人度晓晓，AI 作画平台文心一格等

4 ChatGPT 全球破圈，微软、谷歌、百度等生成式 AI 布局持续推进，加速应用落地，互联网传媒是典型的智力劳动密集型产业，AIGC 有望在内容和信息生产（文本图像游戏视频）等最先落地并有望发生 PGC-UGC-AIGC 的革命变化

5 也持续拓展至搜索、科研、办公、电商客服、智能家居等领域。预计 AIGC 主题投资扩散，科技创新在 2023 传媒复苏之年（政策边际放松）提振板块估值

¶ 后续催化

1 微软将在美东时间今日（北京时间 2 月 8 日周三凌晨 2 点）举行发布会，OpenAI 的 CEO 已经发布推文明确出席

2 谷歌今日宣布对话式 AI 工具 Bard 正式开始测试，将于 2 月 8 日举办一场关于搜索和人工智能的发布活动

¶ 相关标的

1 AI 技术：领军百度（AI 布局全面且领先），低估值的昆仑万维（发布昆仑天工，AIGC 全系列算法与模型开源），风语筑（与百度合作紧密，AIGC 技术已应用在内容生产）

2 应用：视觉中国（AIGC 图片），中文在线（AI 辅助文字创作），三七互娱（率先应用智能化投放系统），吉比特（AI 等新方向持续探索跟踪，内容创新能力强）

3 巨人网络（拥有 AI 实验室，playtika 大数据和 AI 是其核心能力），心动公司（taptap 有望受益于 AIGC 实现内容提质增量），恺英网络（创新方向积极布局），神州泰岳（NLP 研发应用深耕多年，游戏出海排名前列）

¶ 通信：如何看待这一波 AIGC 到流量主线的扩散？

1 看似北美 capex“疲弱”，但实际投资方向更明确，更容易超预期

2 过去我们只能选择 capex 流量主线的主要跟踪指标，但缺点是后验、有噪音（amazon 也仅有 40-50%比例的 capex 是 ICT 相关）、太短期（预算季度波动大）

3 当前虽然北美 capex 增量有限，但：meta 为首的偏内容玩家聚焦了新架构/高密度/新技术等投资趋势、amazon 也提高 ICT 投资比例，所以流量主线实际的需求驱动力很强

¶ 技术演进的分流和聚焦

1 光通信为例，过去 100G 代际的切换方向相对明确；100G 向上的演进方案相对多样化，200G、800G 各有选择，AIGC 的算力+流量驱动下光电一体、全光网络、高速率高密度的网络方案成为必然，因此技术迭代仍然是主基调，持续带动需求放量。上述两点都可以解释光通信代表公司为何 22 年能预计高增

a IDC 层面供给加速出清、需求静待花开

b 东数西算周年之际，可以看到上海市提出新一批 IDC 用能规划、各地算力枢纽建设提上日程、数字经济的基建项目审批等，均提示了供给正在加速出清。我们预判大方向是结构调整，抖/快等内容创新持续拉动+云和互联网需求预期提振+金融等行业数字化推进，可以期待需求转暖

2 因此，不妨再回顾下我们 23 年投资策略中提到的流量主线的“扩散逻辑”~标的序列：运营商投资、格局改善、海内外需求边际修复的核心受益，信创需求共振景气，低估值+增速回暖的左侧机会，关注紫光股份、锐捷网络、中际旭创、新易盛、天孚通信等。此外也持续关注流量基建主线的数据中心等环节供需变化

¶ TMT

1 大模型和小模型有场景不同。小模型用比喻说法，约等于现款买房，大部分智能制造，智联汽车，智能家电，智能电力等“嵌入式软件”适合。今天是轮动到智能制造机器人

2 大模型约等于“贷款买房”，后期才收敛，但复杂的场景适合，比如自然语言处理 nlp，蛋白质氨基酸结构预测 alphafold，aigc 多模问题比如图像/语义/视频等交叉生成，未来的搜索，其实当前比较窄

¶ 轮动到机器人+aigc，符合轮动规律

1 和之前先互联网+软件，再游戏，再通信光模块，一脉相承。也就是越来越有业绩（机构化），上游化（卖铲人）。按照这个规律，会扩散到其他有业绩的比如上述智能其他多领域，然后再轮动回原来的软件互联网

2 正是由于小模型也轮动，所以说明大家更泛化，更看业绩

3 未来节奏预测由于四季度和一季度大家都没业绩，估计一季度是高风险偏好，主题化。二季度有两个会，tmt 会回到成长和价值

微软新版 Bing 搜索引擎发布会

访谈日期：2023/2/6

具体内容

¶ 纪要内容

1. 第一个是 OpenAI 推出的 ChatGPT，这是假期里大家唯一谈论的事情。第二是我看到一条推文说有人使用 GitHubCopilot 构建了 80%的代码。最近在 GitHub 上突破了 1 亿开发者。如果我们能够提高他们的生产力，在接下来的十年里，我们会将这个数字翻一番
2. 然后再翻一番，达到 50 亿开发人员，想想我们可以释放的生产力。这就是我们拥有的机会。我认为必须让人工智能走向世界，而不是实验室。其中预训练大模型的应用和安全系统必不可少，我认为这项技术将重塑几乎所有软件类别
3. 网络诞生于 PC 和服务器。它随着移动互联网和云而发展，现在它将随着人工智能而发展。其中，搜索组织了网络，然后，通过移动互联网，超级应用程序成为人们使用网络的方式
4. 我们希望在搜索领域再次进行创新。微软称其为“您的 AI 驱动的网络副驾驶“。这个副驾驶的核心是一个全新的 Bing 搜索引擎和 Edge 网络浏览器。Bing 将直接回答您的问题，并提示您发挥更多创意

¶ 全新 Bing 的四大突破

1. 模型：Bing 将在 OpenAI 的下一代 LLM（LargeLanguageModel）上运行，专为搜索定制
2. 性能：与 OpenAI 合作的“PrometheusModel”，提高搜索结果相关性、对答案进行注释、显示最新结果等等
3. 核心搜索索引：通过 AI 模型，搜索结果的相关性跃升幅度最大
4. 用户体验：集答案、聊天和浏览器一体的搜索体验

¶ 现在，我们即将看到新的 Bing 的实际应用。它拥有专注于答案、聊天和帮助提示的能力

1. 例子：搜索墨西哥画家的比较会显示结果列表，右侧是生成的答案，并带有链接注释
2. 例子：询问宜家双人沙发是否适合小型货车。Bing 可以找到双人沙发和汽车的尺寸，并回答是否合适
3. 例子：前三名吸尘器的优缺点
4. 例子：比如用户想搜索的墨西哥城旅游的攻略。他可以输入：为我和我的家人制定一个为期五天的旅行路线。还可以创建此行程的摘要，并发送电子邮件给家人

5 例子：搜索“顶级日本诗人”，下方有维基百科的链接

6 例子：四口之家的膳食计划，有素食可供选择，并迎合不喜欢坚果的人

¶ 新 Edge 界面

1 Edge 中还将有一个 AI 驱动的副驾驶。我们通过更时尚、更轻便、非常酷的新方式在 Edge 集成了 Bing

2 我们刚刚在 Bing 中看到的聊天界面在 Edge 中作为侧边栏提供，因此您无需导航至 Bing 即可使用它

3 新 Bing 有五个方面的贡献。核心基础、聊天编排、提示生成、推理和交互式体验，以及为未来机会扩展的基础设施

4 同时，微软一直在研究人工智能的风险，包括偏见等已知风险和“越狱”等新风险。有了这个产品，我们在开发衡量风险缓解的方法方面比以往任何时候都走得更远

5 我们使用 Bing-scale 安全系统过滤内容，并部署快速响应系统以应对不断变化的威胁。最后，在应用层，我们正在迭代 metaprompts。微软与 OpenAI 合作，可以不断测试对话并对其进行分析，以对对话进行分类并改进安全系统中的漏洞

6 新 Bing 今天以“有限预览”的形式在桌面上线。每个人都可以尝试有限数量的查询，并立即注册以获得完全访问权限。预计在未来几周内推出移动版本，并将预览人数扩展到数百万人

¶ 这种形式是否会产生大量不良内容，这些内容会被拉回到模型中吗？

1 将搜索与模型相结合，意味着用户可以使用搜索来进行事实核查

2 当你将这两者结合起来时，就会产生制衡

¶ Prometheus 模型与 ChatGPT 相比如何？

1 Bing 协调器围绕模型

2 创建了一个良性循环

¶ 竞争格局和成功指标是什么？

1 目前，我们专注于构建这个伟大的新终端的用户价值。我们希望可以吸引更多用户使用 Bing，并让这些用户更多地使用 Bing

2 如果反馈良好且使用良好，一切都会成功。从源头提升模型的问题我们非常关心答案的来源。这就是为什么有链接

¶ 对于一些新信息，比如新电视在今天刚刚发布，机器人会知道吗？Bing 的准确性？

1. 用户会对模型的这部分能力印象深刻
2. 我们不会总是做对。我们一直在学习。这里的关键实际上是我们如何获取信息，为它提供更多数据可以提高准确性
3. 在 GPT 上运行查询与传统搜索相比的成本无法回答

¶ 为什么不重塑 Bing 的品牌，推出全新的另一个产品？

1. 我们认为这完全是对搜索的重新想象
2. 我们喜欢 Bing 品牌，所以我们也会坚持下去

¶ 生成的内容是否会被标记为 AI 生成的？

1. 我们的愿景是副驾驶为用户提供帮助
2. 我们不希望 Bing 完全写东西

¶ 微软为实现可持续发展和计算优化付出了哪些努力？

1. 我们希望在何时使用繁重的计算
2. 以及何时可以使用成本不高的设备方面做的更好

¶ ChatGPT 有时会产生幻觉并编造一些东西。您是否解决了这个问题，或者是用户可能会看到的问题？

1. 我们从一开始就致力于此
2. 我们正在衡量模型在搜索结果中的哪些地方出现问题，但它并不完美。用户将看到它在哪里更改了一个小数字或其他内容

¶ Chrome 或其他浏览器中可用吗？

1. 我们的目的是将它带到所有浏览器中
2. 我们从 Edge 开始。Chrome 必须实现一些功能才能正常工作，但我们的目标是所有浏览器

¶ 开发人员会获得 API 吗？你们可以限制这些查询吗？

1. 我们正在继续评估我们的产品
2. 限制会有所不同

¶ 最终用户需要付费吗？你展示的很多内容都是基于事实的。这可以帮助人们进行创意写作吗？有计划加入广告吗？

1 没有定价，它是免费的。它具有创造性

2 一开始就有广告

从美国科技巨头财报看 AI 的发展和应用

访谈日期：2023/2/5

具体内容

¶ 摘要

1 微软：微软和 OpenAI 已经进入第三阶段合作，微软为 OpenAI 提供算力和超级计算系统（specialized supercomputingsystems），而 OpenAI 也将反哺 Azure 的 AI 能力

- a 未来微软产品将全线整合 ChatGPT，届时微软的每个产品都将具备相同的 AI 能力
- b 微软认为，下一次平台型科技浪潮就是人工智能，将大大改变生产力和消费者体验，将通过人工智能来推动解决方案的创新和差异化竞争

2 META：

- a AI 对于 META 主要体现在效率的提升，包括内容推送效率（更加精准）、广告精准度方面的效率（弥补苹果隐私政策之后广告精准度的下降）、工程师的生产力效率，以及将大力发展生成式 AI（generativeAI）
- b 公司正在将数据中心转移至新架构上，更好得理解 AI 需求、以及满足 AI 和非人工智能的工作负载

3 谷歌：谷歌在 AI 方面积累深厚，一直是行业领导者，未来将在三个方面持续发展

- a 大模型方面，未来将从 LaMDA 开始提供语言模型，用户可以与其直接互动（类似于 ChatGPT）
- b 同时，BERT 和 MUM 等语言模型在四年多时间内持续在改进搜索结果，很快他们将做为搜索伴侣，与用户直接交互（类似于微软将 ChatGPT 应用于 Bing 中）将为开发者、创作者和合作伙伴提供 AI 相关的工具和 API
- c AI 能力将在云计算、workspace（跟微软办公直接竞争）和广告中为客户赋能，目前谷歌广告对 AI 已经有了很多应用，比如智能出价、匹配查询、优化 ROI 及自动生成广告素材等方面
- d 对于未来成本的管控，人工智能将在谷歌中扮演重大角色，将提高生产力和运营效率。同时，为了反映 DeepMind 与 GoogleServices、Google Cloud 和 Other Bets 的合作不断增加，从第一季度开始，DeepMind 将不再在 Other Bets 中报告，并将作为 Alphabet 公司成本的一部分进行报告

¶ 微软：微软与 OpenAI 的合作回顾

1 2023 年 1 月 23 日，微软宣布了与 OpenAI 自 2019 和 2021 年后的第三阶段的合作，微软对 OpenAI 是一项 multi-year，multi-billion 的投资，本次双方将在此前基础上进一步 extending partnership

- a 首先，AI 的训练需要算力，Azure 是 OpenAI 的独家云提供商，Azure 将为跨研究、产品和 API 服务的所有 OpenAI 的 workloads 提供支持
- b AI research 需要超级计算系统来支持（specialized supercoputing systems），微

软将加大对 supercomputing systems 的投资以支持 OpenAI 的发展。同时，OpenAI 也可以反哺 Azure 的 AI 能力，微软将继续构建 Azure 的 AI 基础设施，以帮助客户在全球范围内构建和部署 AI 应用程序

c 微软将自己的消费者和企业产品中部署 OpenAI 的模型，并为客户引入基于 OpenAI 技术的体验

d 包括微软的 Azure OpenAIservice（开发人员可使用该服务来 builtAI applications，可以使用包括 GPT-3.5，Codex， and DALL-E 2），双方还将共同努力，将 OpenAI 的技术构建到 GitHubCopilot（AI 编程工具）和 MicrosoftDesigner 等应用程序中

¶ 微软在财报中对 AI 的表述

1 人工智能方面，人工智能时代已经来临，而微软正在为其提供动力，我们正在见证基础模型能力的非线性改善，我们正在给客户提供这种能力

a 随着客户选择他们的云供应商并投资于新的工作负载，作为人工智能领域的领导者，我们完全有能力抓住这个机会。我们拥有云中最强大的 AI 超级计算基础设施。它正被客户和合作伙伴，如 OpenAI，用来训练最先进的模型和服务，包括 ChatGPT

b 就在上周，我们广泛提供了 Azure OpenAI 服务，已经有 200 多个客户--从毕马威到半岛电视台--正在使用它。我们将很快增加对 ChatGPT 的支持，使客户能够首次在自己的应用程序中使用它。 在昨天，我们宣布完成了与 OpenAI 的下一阶段协议

c 我们很高兴成为他们的独家云供应商，并将在我们的消费者和企业产品中部署他们的模型，因为我们继续推动人工智能技术的发展

2 所有这些创新正在推动我们整个 Azure 人工智能服务的增长。 仅 Azure ML 的收入就连续五个季度增长超过 100%，安盛、联邦快递和 H&R Block 等公司都选择该服务来部署、管理和治理其模型

a 三年半前，我们现在开始了和 OpenAI 伙伴关系。在过去的三年里，我们实际上一直在努力研究这种伙伴关系的许多要素。因此，我认为我们的投资者看待这一点的方式是，正如我所说，我们从根本上相信，下一个大型平台浪潮将是人工智能

b 我们还相信，通过能够抓住这些浪潮（waves），然后让这些浪潮创造新的解决方案和新的机会，是可以带来许多企业价值的。因此，每当我们考虑平台机会和平台转换机会时，我们都是这样做的

c 我们怎么才能抓住这些 waves，并使其更具扩张性，然后创造出什么？因此，如果您从这个角度来看，Azure 的核心，或者被认为是云计算的东西从根本上改变了其性质以及计算、存储和网络的结合方式

d 从某种意义上说，under the radar，如果你愿意的话，在过去的三年半里，四年里，我们一直在非常努力地构建训练超级计算机，当然还有现在的推理基础设施（ inferenceinfrastructure），因为一旦你在应用程序中使用人工智能，它就会从繁重训练变成了推理

e 因此，我认为核心 Azure 本身正在为核心基础设施业务转型。它正在转型。因此，您甚至可以看到我们拥有 Azure OpenAI 服务以外的数据，想想 Synapse plus OpenAIAPI 可以做什么。我们已经集成了 PowerPlatform 功能

3 我们今天在机器人流程自动化和工作流自动化方面处于领先地位的原因之一是，

我们在那里拥有一些人工智能功能。事实上，GitHub Copilot 是当今市场上最大规模的基于 LLM 的产品。因此，我们完全希望我们将人工智能纳入堆栈的每个层，无论是生产力还是我们的消费服务中。因此，我们对此感到兴奋

4 我们也对 OpenAI 创新感到兴奋，他们把产品商业化了。我们对 ChatGPT 建立在 Azure 上并具有其牵引力（traction）感到兴奋

5 因此，我们关注两者，它有投资部分，也有商业伙伴关系。但从根本上说，我认为，这将通过在人工智能领域领先来推动微软每个解决方案的创新和竞争差异化

¶ 微软旗下所有产品将全线整合

1 ChatGPT：继微软宣布在搜素引擎必应、办公全家桶 Office 嵌入当今最火爆 AI 语言模型一 ChatGPT 后，CEO 纳德拉宣布还将在云计算平 Azure 中整合 ChatGPT，宣告 AzureOpenAI 服务全面上市，通过该服务可以访问 OpenAI 开发的 AI 模型，届时微软的每个产品都将具备相同的 AI 能力，彻底改头换面

2 META：2023 年的管理主题是“效率之年”。去年结束时，我们进行了一些艰难的裁员和重组一些团队。当我们这样做的时候，是关注效率的开始，而不是结束

   a 从那时起，我们采取了一些额外的措施，比如与我们的基础设施团队合作，研究如何在减少资本支出的同时实现我们的路线图
   b 接下来，我们正在努力扁平化我们的组织结构，删除一些中层管理层，以更快地做出决定，以及部署 AI 工具，以帮助我们的工程师提高生产力

3 自去年以来，我们的工作重点没有改变。推动我们路线图的两大技术浪潮是今天的 AI 以及从长远来看的元宇宙。首先是 AI 发现引擎。Facebook 和 Instagram 正在从仅仅围绕你关注的人和账户组织，转变为越来越多地显示我们 AI 系统推荐的更多相关内容。这涵盖了每一种内容格式，这也是我们服务的独特之处

4 在营收方面，我们仍有望在今年年底或明年初大致保持中性。然后，在那之后，我们应该能够在满足我们看到的需求的同时，盈利地增长 Reel

   a 在我们更广泛的广告业务中，我们继续投资于 AI，我们在这里看到了我们的努力的回报。上个季度，广告商看到的转化率比去年增加了 20%以上。再加上每次获取成本的下降，广告支出的回报也提高了
   b AI 是我们发现引擎和广告业务的基础，它将为我们的应用程序带来许多新产品和额外的转变。生成式 AI（generative AI）是一个非常令人兴奋的新领域，有这么多不同的应用程序。我对 Meta 的目标之一是在我们的研究基础上，除了我们在推荐 AI 方面的领先工作外，成为生成 AI 的领导者

5 公司专注的领域包括 AI，包括我们的发现引擎、广告、业务消息传递和日益生成的 AI，以及元宇宙的未来平台。从运营的角度来看，我们专注于效率，并继续精简公司，以便我们能够尽可能地执行这些优先事项，并在提高业务表现的同时建立一个更好的公司

¶ 今年关注的两个最大的主题

1 一个是效率，然后这种新产品领域将是生成式 AI（generative AI）工作

   a 我们有一堆不同的工作流程跨越几乎每一个我们的产品使用新技术，特别是大

型语言模型和扩散模型，用于生成图像、视频、化身和 3D 资产，以及各种不同的东西，跨越我们正在进行的所有不同的工作流程

b 以及从长期来看，致力于能够真正增强创作者在应用程序和运行许多不同账户上的生产力和创造性的事情

2 关于我们的新数据中心架构，它支撑着较低的资本支出前景。因此，我们正在将数据中心转移到一个新的架构上，该架构可以更有效地支持 AI 和非人工智能工作负载。随着我们更好地理解我们对 AI 的需求，这将给我们更多的选择

3 此外，我们预计新的设计将比以前的数据中心架构更便宜、更快地构建

a 伴随着新的数据中心架构，我们将优化我们构建数据中心的方法

b 因此，我们有一个新的分阶段的方法，允许我们以更少的初始容量和更少的初始资本支出来构建基础计划，但随后在需要时迅速伸缩未来的容量。我们仍在计划显着增长 AI 能力

4 广告战略实际上有两个部分，那就是继续投资人工智能，这就是我们看到广告相关性得到很大改善的地方，比前一年多 20%的转化率，再加上每次获取成本的下降，带来了高投资回报率

¶ 谷歌：谷歌在 AI 方面积累深厚，未来自身的 AI 能力将持续结合到产品之中

1 人工智能是我们今天正在研究的最深刻的技术，我们有才华横溢的研究人员、基础设施和技术使我们在 AI 达到拐点时处于非常有利的位置

2 六年多以前，我们第一次谈到谷歌是一家人工智能优先的公司，从那时起，我们一直是开发 AI 的领导者。我们的 Transformers 研究项目和我们在 201 年的 field-defining 论文，以及我们在扩散模型方面的开创性工作，是今天开始看到的许多生成式 AI 应用程序的基础

3 将这些技术飞跃转化为可帮助数十亿人的产品，是我们公司一直以来赖以生存的基础。我们将以 AI principle 和信息完整性的最高标准作为我们所有工作的核心，大胆地开展这项工作。自去年初以来，我们一直在为这一刻做准备，在接下来的几个月里，将在三大领域看到我们的成果

a 大模型（large models）我们已经广泛发表了关于 LaMDA 和 PaLM 的文章，这是业界最大、最复杂的模型，以及在 DeepMind 的大量工作。在接下来的几周和几个月里，我们将从 LaMDA 开始提供这些语言模型，以便人们可以直接与它们互动

b 这将帮助我们继续获得反馈、测试并安全地改进它们。 这些模型在撰写、构建和总结方面特别出色。 当它们提供最新的、更真实的信息时，它们将变得对人们更有用

c 在搜索中，BERT 和 MUM 等语言模型已经改进了四年的搜索结果，实现了显着的排名改进和多模式搜索，如 Google Lens。 很快，人们将能够以实验性和创新的方式直接与我们最新、最强大的语言模型进行交互，作为搜索的伴侣

d 我们将为开发者、创作者和合作伙伴提供新的工具和 API。 这将使他们能够创新和构建自己的应用程序，并在我们的语言、多模式和其他 AI 模型之上发现 AI 的新可能性

e 我们的 AI 将能为各种规模的企业赋能

i Google Cloud 正在通过我们的 CloudAI 平台向客户提供 AI 赋能，包括面向开发人员和数据科学家的基础设施和工具，例如 Vertex AI。 我们还为制造、

生命科学和零售等行业提供特定的人工智能解决方案，并将继续推出更多解决方案

ii 对于 workspace 用户，日常工作也将受益于 AI 的支持，例如用于协作的 SmartCanvas 和用于创作的 Smart Compose；我们正在努力将大型语言模型引入 Gmail 和 Docs。 我们还将提供其他有用的生成功能，从编码到设计等等

iii 广告合作伙伴，从 natural languageunderstanding 到 generative AI，将对行业带来变革性的影响

4 以智能出价为例，它使用人工智能来预测未来的广告转化及其价值，帮助企业保持敏捷并快速响应需求的变化。到 2022 年，人工智能的进步提高了竞价性能，以帮助客户提高 ROI 并更有效地使用广告预算

a 在搜索查询匹配中，像 MUM 这样的大型语言模型可以匹配广告商提供的用户查询

b 这种对人类语言意图的理解与出价预测方面的进步相结合，让企业在使用目标 CPA 的广告系列中，将完全匹配关键字升级为广泛匹配时，平均可以看到 35% 以上的转化率

5 Google AI 也是我们创意产品的基础，例如 Google Ads 中的文本建议和响应式搜索广告中的创意优化。 我们很高兴开始测试我们的 Automatically Created AssetsBeta，一旦广告商选择加入，它就会使用 AI 无缝地为搜索广告素材生成标题和描述

a 当然还有 Performance Max，它为我们的客户提供了我们的 AI 驱动系统的最佳组合

b 但我们并没有就此止步，在过去十年中，人工智能一直是我们广告业务的基础，我们将继续为我们的产品带来最前沿的进步，以帮助企业和用户

6 除此之外，AI 还在继续大幅改进谷歌的其他产品。 我们将继续与谷歌以外的公司合作，以负责任的方式开发人工智能，并应用人工智能应对社会面临的最大挑战和机遇

a 例如，DeepMind 的蛋白质数据库包含科学界已知的所有 2 亿种蛋白质，现已被全球 100 万生物学家使用。我们继续全面投资 AI，Google AI 和 DeepMind 是未来不可或缺的一部分

b 在过去的几年里，DeepMind 越来越多地在谷歌和其他公司内部跨团队工作

¶ 未来将 re-engineer 成本

1 使用人工智能和自动化来提高整个 Alphabet 的生产力以及基础设施的效率

2 其次，更有效地管理我们与供应商的支出

3 优化工作方式和地点

¶ DeepMind 披露方式调整

1 为了反映 DeepMind 与 Google Services、Google Cloud 和 Other Bets 的合作不断增加

2 从第一季度开始，DeepMind 将不再在 Other Bets 中报告，并将作为 Alphabet 公司成本的一部分进行报告

从北美云厂商的 AI 规划看光通信的结构创新

访谈日期：2023/2/1

具体内容

¶ 资本开支

1 北美云计算大厂财报口径上，对海外衰退下云计算投资、资本开支的降速已反映出来了，Meta 下调了整体费用和资本开支指引（340-390 下调到 300-330 亿美金），方向从元宇宙调到 AI 方向

2 微软整体口径变化不大，还是云计算相关需求，但 Azure 增速放缓，谷歌、Amazon 云计算相关投资也有回落

¶ 投资方向有变化

1 最近微软投资 110 亿美金 OpenAI，ChatGPT 嵌入到搜索引擎里，正面硬刚谷歌搜索引擎，包括 FB 在布局很多超算，谷歌也推出类似 ChatGPT 的交互

- a AI 进入到提速过程，包括最近 ChatGPT 迅速出圈，拥有很多用户。跟基础设施和硬件设备相关的变化是算力会有大幅增长
- b 算力从广义讲需要能耗、成本堆出来，不断烧数据发电做存储计算，能耗跟投入成本密切相关，AI 背后的算力相较于之前云计算、电商需要的算力成倍增长，按照传统速率升级、堆叠算力的方式不符合商业化发展，所以设备上要为了匹配高算力带来的低成本方案
- c 目前已经有很多设备、光通信、服务器产业链面向超算去提升出货量

2 机遇：需要降能耗成本，体现在设备、光模块、交换机的更新。用在国内外超大数据中心和超算里的设备已经有区分了，能耗消耗有很大差别。光模块 100G-400G-800G 的速率还不够，很多超算的交换机要按照 T 的计量，这个量级匹配相应光模块的成本非常惊人

- a 因此衍生出同步散热降温，同时在 10+T 的交换机搭载 800G 光模块需要很多堆叠，交换机会过载，所以衍生出交换机和光模块融合（COPACKAGE），以前的光模块演进成光引擎，然后再和交换芯片贴在同一张 PCB 背板上，通过交换机搭载的液冷板进行物理冷却和降温
- b 同时光引擎由于体积、集成度高，搭配硅光封装规模化后会体现成本优势，会替代高算力场景

3 建议关注：交换机、服务器、光模块的天孚通信、锐捷网络、新华三。现在在北美等大厂都在推进 COPACKAGE 方案，最近由于 AI 进度加速，近期由于 ChatGPT 爆火，大厂方向迅速转变，可能带来光引擎、液冷服务器加速推进

- a 天孚通信、锐捷网络、新华三都在 COPACKAGE 和硅光有典型的布局，下游已经面向北美核心 AI 厂商开始出货。不论交换机的华为华三锐捷，海外思科英特尔英伟达都在全面布局，很多大厂已经出货但由于体量较小，未来在数据中心侧会规模性铺开
- b 对于传统空间被新方案替代市场需求，但由于算力激增，数通投资、光通信出

货量还是会大幅提升尤其在高算力场景

c 结构性创新带来的弹性会在今明年体现，基本到 800G 到 1.6T 差异会清晰，目前切换新方案纠结点在于成本和供应链稳定，等量起来供应链会突破

d A 股已经有部分公司产品出货了，天孚、锐捷网络核心主推，之前这个方向公司被海外通胀衰退抑制，数通市场增速往下走，这个预期短期已经由于北美大厂财报也出释放了，同时 AI 对算力增长的拉动在 24\25 年体现更加明显，当前低估值具备高的性价比，20X 不到

e 现在市场规模非常小，光模块一年 100+亿美金只有 1%不到做光引擎，但光引擎速率升级、液冷交换机升级交付量会大提升

¶ 天孚、锐捷怎么降低功耗?

1 原来交换机上有口可插拔光模块去做光电转化，信号传输速度快走光纤，交换机里会转成电信号

2 之前模块在交换机外部现在放到内部对模块和交换机都要同步升级匹配，但目前容错率低、成本高，但未来是必然选项，不是主题性投资了，会有明显订单和产业趋势的切换，目前是早期阶段

3 硅光和封装良率都会有质变。之前大家盯着北美资本开支看行业需求，未来会细化看资本开支投向 AI 的部分会是增速最高的部分，在这个方向寻找标的是下一个投资方向，同时海外 ChatGPT 远远领先于国内，所以会看国内硬件优质供应商

从微软和 OpenAI 的合作来梳理 AI 投资逻辑

访谈日期：2023/1/29

具体内容

¶ 微软和 OpenAI 的合作

1 2023 年 1 月 23 日，微软宣布了与 OPENAI 自 2019 和 2021 年后的第三阶段的合作，微软对 OPENAI 是一项 MULTI-YEAR,MULTI-BILLION 的投资，本次双方将在此前基础上进一步 EXTENDINGPARTNERSHIP

2 首先，AI 的训练需要算力，AZURE 是 OPENAI 的独家云提供商，AZURE 将为跨研究、产品和 API 服务的所有 OPENAI 的 WORKLOADS 提供支持

3 AIRESEARCH 需要超级计算系统来支持（SPECIALIZEDSUPERCOMPUTINGSYSTEMS），微软将加大对 SUPERCOMPUTINGSYSTEMS 的投资以支持 OPENAI 的发展

4 同时，OPENAI 也可以反哺 AZURE 的 AI 能力，微软将继续构建 AZURE 的 AI 基础设施，以帮助客户在全球范围内构建和部署 AI 应用程序

5 微软将自己的消费者和企业产品中部署 OPENAI 的模型，并为客户引入基于 OPENAI 技术的体验。包括微软的 AZUREOPENAISERVICE（开发人员可使用该服务来 BUILTAIAPPLICATIONS 可以使用包括 GPT-3.5,CODEX,ANDDALL•E2),双方还将共同努力，将 OPENAI 的技术构建到 GITHUBCOPILOT（AI 编程工具）和 MICROSOFTDESIGNER 等应用程序中

¶ 微软在财报中对 AI 的表述

1 人工智能方面，人工智能时代已经来临，而微软正在为其提供动力，我们正在见证基础模型能力的非线性改善，我们正在给客户提供这种能力

- a 随着客户选择他们的云供应商并投资于新的工作负载，作为人工智能领域的领导者，我们完全有能力抓住这个机会。我们拥有云中最强大的 AI 超级计算基础设施
- b 它正被客户和合作伙伴，如 OpenAI，用来训练最先进的模型和服务，包括 ChatGPT。就在上周，我们广泛提供了 AzureOpenAI 服务，已经有 200 多个客户--从毕马威到半岛电视台--正在使用它
- c 我们将很快增加对 ChatGPT 的支持，使客户能够首次在自己的应用程序中使用它。在昨天，我们宣布完成了与 OpenAI 的下一阶段协议
- d 我们很高兴成为他们的独家云供应商，并将在我们的消费者和企业产品中部署他们的模型，因为我们继续推动人工智能技术的发展。所有这些创新正在推动我们整个 Azure 人工智能服务的增长
- e 仅 AzureML 的收入就连续五个季度增长超过 100%，安盛、联邦快递和 H&RBlock 等公司都选择该服务来部署、管理和治理其模型

2 三年半前，我们现在开始了和 OPENAI 伙伴关系。在过去的三年里，我们实际上一直在努力研究这种伙伴关系的许多要素。因此，我认为我们的投资者看待这一点

的方式是，正如我所说，我们从根本上相信，下一个大型平台浪潮将是人工智能

3 我们还相信，通过能够抓住这些浪潮（WAVES），然后让这些浪潮创造新的解决方案和新的机会，是可以带来许多企业价值的。因此，每当我们考虑平台机会和平台转换机会时，我们都是这样做的

¶ 我们怎么才能抓住这些 waves，并使其更具扩张性，然后创造出什么？

1 因此，如果您从这个角度来看，AZURE 的核心，或者被认为是云计算的东西从根本上改变了其性质以及计算、存储和网络的结合方式

   a 从某种意义上说，undertheradar，如果你愿意的话，在过去的三年半里，四年里，我们一直在非常努力地构建训练超级计算机，当然还有现在的推理基础设施（inferenceinfrastructure），因为一旦你在应用程序中使用人工智能，它就会从繁重训练变成了推理
   b 因此，我认为核心 Azure 本身正在为核心基础设施业务转型。它正在转型。因此，您甚至可以看到我们拥有 AzureOpenAI 服务以外的数据，想想 SynapseplusOpenAIAPI 可以做什么。我们已经集成了 PowerPlatform 功能
   c 我们今天在机器人流程自动化和工作流自动化方面处于领先地位的原因之一是，我们在那里拥有一些人工智能功能。事实上，GitHubCopilot 是当今市场上最大规模的基于 LLM 的产品
   d 因此，我们完全希望我们将人工智能纳入堆栈的每个层，无论是生产力还是我们的消费服务中。因此，我们对此感到兴奋

2 我们也对 OPENAI 创新感到兴奋，他们把产品商业化了。我们对 CHATGPT 建立在 AZURE 上并具有其牵引力（TRACTION）感到兴奋

3 因此，我们关注两者，它有投资部分，也有商业伙伴关系。但从根本上说，我认为，这将通过在人工智能领域领先来推动微软每个解决方案的创新和竞争差异化

¶ 举例来理解，加入 AI 能力之后未来我们可以用微软的产品来干什么？

1 编程（GITHUBCOPILOT）：给开发者 CODE 建议，可以对一段 CODE 进行描述，提升编程效率；未来甚至还可以进行自动编程

2 办公：WORD 中使用，可以自动理解格式命令，甚至自动创建相关内容；同理，跟 TEXT 相关的所有产品例如 EXCEL 中也可以应用

3 自动化工具：各种流程自动化工具可以更加智能

4 以上，在 C 端和 B 端都有非常多的应用空间

5 可能会出现三种 AI 类型公司

   a OpenAI 这种提供 AImodel 的公司（会出现比较多的 startups)
   b 利用 AImodel 来做产品的公司，例如 BuzzFeed（应用层面的公司较多）
   c 利用自己的 AImodel 和数据来做产品的公司，例如谷歌，是用产品来变现而非 AImodel 来变现（更多是大公司）

6 CHATGPT 或者任何 AI 产品均离不开算力、数据和技术，因此在底层基础设施上，这注定是大公司之间的一场竞争

a 云计算/算力：亚马逊、微软、谷歌、阿里巴巴等，任何 AI 的训练都离不开云计算，同样，英伟达也将受益

b 数据：AI 的训练离不开数据，所以这要求使用 AI 的公司需要有足够多的数据、或者足够多能触及到消费者/企业客户的产品，数据最多/产品最多的公司包括谷歌、meta、亚马逊、微软等

c 技术：亚马逊自己的语言模型达到了 20bnparameters，但目前只是在内部使用；谷歌也持续在深入自己的 AI 能力，微软通过投资 OpenAI；

d 在苹果隐私政策之后，Meta 也开始大力投资 AI。看向未来，这些大公司都有可能培养出 OpenAI 的竞争者

¶ 应用层面

1 任何需要内容的公司：SOCIALPLATFORMS（META 等）、新闻网站（BUZZFEED）、搜索引擎（谷歌、微软），等等

2 安全领域：任何 AIMODEL 中都包含大量数据，如果被攻击可能会造成数据泄露，因此该领域也需要一些安全类的应用

3 映射到 A 股投资逻辑上，建议关注：AI 技术领域领先的技术公司。算法、数据、算力是 AI 大模型训练的基础，建议关注基础设施相关标的：科大讯飞、海天瑞声、拓尔思等

4 当前 CHATGPT 上线将有望推动文本类 AI 渗透于文本生产、智能批阅等应用领域

5 建议关注：阅文集团、中文在线、掌阅科技、视觉中国、金山办公、昆仑万维等

微软公司业绩交流

访谈日期：2023/1/26

具体内容

¶ 业绩情况

1 FY23Q2 公司营收 527 亿美元（YOY+2%，CC+7%），低于彭博预期的 529 亿美元。NON-GAAP 毛利 354 亿（YOY+2%，CC+8%）。GAAP 营业利润 204 亿，同比减少 8%，NON-GAAP 营业利润 216 亿（YOY-3%，CC+6%）

2 GAAP 净利润 164 亿，同比减少 12%，NON-GAAP 净利润 174 亿（YOY-7%，CC+1%），超过彭博预期的 171 亿。NON-GAAP 每股收益 2.32 美元，同比减少 6%。公司通过股票回购和分红向股东返还了 97 亿美元

3 本季度云相关收入达 271 亿美元（YOY+22%，CC+29%）

¶ 分业务来看

1 生产力和业务流程：收入为 170 亿美元（YOY+7%，CC+13%），排除外汇影响后符合预期。分产品

- a Office 商业版收入（YoY+7%，CC+14%）。Office365 商业版收入（YoY+11%，CC+18%），略好于预期，主因续签执行良好以及 E5 持续增长带来的 ARPU 提高；其付费席位同比增长 12%，主要受到中小型企业和一线员工使用的推动
- b Office 消费版收入（YoY-2%，CC+3%），主因 Microsoft365 的订阅量增长了 12%达到 6,320 万，部分被交易业务下降所抵消。LinkedIn 收入（YoY+10%，CC+14%），主要得益于 talentsolution 的增长，部分被广告需求下滑趋势导致的营销方案疲软所抵消
- c Dynamics 收入（YoY+13%，CC+20%），主因 Dynamics365 收入（YoY+21%，CC+29%）

2 智能云：收入 215 亿（YOY+18%，CC+24%），符合预期。分产品

- a 服务器产品和云服务收入（YoY+20%，CC+26%）。Azure 和其他云服务收入（YoY+31%，CC+38%），增长持续放缓，尤其在 12 月，Q2 的 Azure 收入 CC+30%左右
- b 服务器收入（YoY-2%，CC+2%），延续的混合需求被交易许所抵消。企业服务收入（YoY+2%，CC+7%）

3 更多个人计算：收入 142 亿（YOY-19%，CC-16%），主因 SURFACE，WINDOWS 商业版和搜索业务低于预期。分产品

- a WindowsOEM 收入同比下降 39%，与预期一致，排除去年 Windows11 延迟发布的影响，微软的营收同比下降 36%。设备收入（YoY-39%，CC-34%），低于预期，主因新产品发布的执行面临挑战
- b Windows 商业版收入（YoY-3%，CC+3%），低于预期，主因独立产品新业务增长放缓

c 搜索广告收入（YoY+10%，CC+15%），略低于预期，Edge 浏览器本季度获得超预期的市场份额，收购 Xandr 贡献了大约 6 个百分点的收益

d 游戏业务（YoY-13%，CC-9%），符合预期，Xbox 硬件收入（YoY-13%，CC-9%）；Xbox 内容和服务收入（YoY-12%，CC-8%），主因第一方内容的强势

¶ FY23Q3Guidance

1 23Q3 营收为 505~515 亿（YOY+2%~4%）。分业务看，生产力和业务流程收入 169-172 亿（YOY+%~9%），智能云收入 217-220 亿（YOY+14%~16%），更多个人计算收入 119-123 亿（YOY-15%~18%）

2 22Q3COGS 为 156.5-158.5 亿（YOY+1%~2%），运营费用 147-148 亿（YOY+11%~12%）。全年营业利润同比增速下调 1%

¶ 关于与 OpenAI 的合作，AI 能力是否有所拓展，何时能扩展到 Azure 以外的服务?

1 我们相信下一个平台浪潮将是 AI，抓住浪潮就能为企业创造价值，我们在过去四年建立了超级计算机，如今只要将 AI 融入应用场景中，便可以通过大量训练生成推理功能。AZURE 正在为基础设施业务转型，所以可以看到我们在 AZUREOPENAI 服务之外的数据

2 我们希望将 AI 融入到提高生产力和消费者服务中，将 CHATGPT 的能力赋予到 AZURE 中，这将推动公司在 AI 领域的竞争差异化

¶ 基于前面对美国客户优化（减少）支出的评论，谈谈对于宏观消费环境的看法?

1 上述评论是针对全球而不仅仅针对美国，长期来看科技创收占 GDP 的比例会更高，问题在于在考虑通胀后该比例为多少。我们观察到客户减少支出的因素有二：一是客户在疫情期间增加了购买公司软件的支出，但现在正在优化/减少该部分支出；二是在外汇逆风下消费变得更加谨慎，但在某一时刻，这种优化将停止

2 对于大客户而言，其典型模式为优化现有项目的支出，并将节省的支出用于新项目中，目前投入到新项目的支出正在上升，因此公司将持续专注于提高客户忠诚度及获取市场份额

3 另一方面关于 AI 投资，接下来的 AI 应用开发会比 2019、2020 年的 AI 应用更加落地，客户将更多考虑产品的 AI 推理表现、成本结构、成本多少等，我们在这一领域具有竞争优势

¶ 有多少客户在优化调整支出，有多少客户因为外汇逆风影响了需求?

1 在云计算业务方面，我们难以区分客户是因为优化因素还是宏观因素而削减支出。但可以确定的是，客户一方面基于对业务的预测判断缩减了支出，另一方面想要将更多节省的钱投入到新项目中

2 在单用户层面则有不同，无论对一线工人还是知识型员工（靠知识获取劳动报酬），单用户许可证的购买量都有提速，使用率进一步上升，正如当查看 OFFICE365 的使用情况时，所有数字都在同比大幅增长

3 支出优化的周期：不需要 2 年时间来优化，需要 1 年来优化

¶ 全年收入要达到 10%以上的增长指引是否有困难？Q3 的 Azure 收入增速将下降多少？

1 前面的评论没有提及全年收入，主要需要看 WINDOWSPC 市场能否恢复到疫情前水平。除此之外整体上升趋势是一致的，公司全年的营业利润率指引同比仅有 1%的降幅，而且是在 OEM 逆风可能超过 20 亿美元的情况下，公司在 22 年结束时的业绩表现不错

2 根据指引，Q3AZURE 云计算业务的收入增速将比 Q2 的 38%下降 4%-5%至 33-34%

¶ 关于上周宣布的费用支出决定（裁员费用），如何考虑 23 财年剩余时间的人工数及其他支出？

1 我们同时收购了 NUANCE 和 XANDR，随着投资增加，我们不得不决定（裁员）

2 因此 Q4 结束时员工人数同比增速非常缓慢，另一方面这也使公司的费用增长与收入增长更加一致，全年的人数同比增长仍会很小

¶ 对于 Office365 商业版，公司更在意付费席位和 ARPU 间的均衡增长，还是更加青睐付费席位的增长？

1 本季度营收中来自 ARPU 的收入贡献持续扩大，其增长为 OFFICE365 商业版的营收增长提供了稳定性

2 因此我们正进一步提高 MICROSOFT365E5 用户的 ARPU，目前已经有 4.5 个季度显示 E5 采用情况很好

¶ 如何量化 AI 对于 Azure 的贡献？

1 现在将 AI 从其他工作中分离出来还为时尚早，AI 将作为 AZURE 的核心部分而非独立部分

2 一个应用程序拥有推理功能，它也会同时拥有储存和其他计算功能，随着时间推移每个应用程序都将成为 AI 应用程序

微软公司各业务线情况

访谈日期：2023/1/25

具体内容

¶ 整体概况

1 客户在疫情期间加快了数字化支出，现在正在优化数字化支出，且由于宏观经济的不确定性而更加谨慎；随着微软将世界上最先进的 AI 模型变成新的计算平台，下一个计算浪潮正在诞生

2 公司对三件事深信不疑：帮助客户从技术支出中实现更多价值，建立长期的忠诚度和份额地位；保持微软内部的成本结构与收入增长一致；数字化支出占 GDP 百分比增加的长期趋

3 以此为背景，微软云计算的季度收入超过 270 亿美元，增长 22%，按固定汇率计算，增长 29%

¶ Commercial remaining performance obligation

1 对企业未履约合约余额为 1,890 亿美元，同比增加了 29%

2 其中 45%将在 1 年内确认收入，这部分同比增长了 24%。其余在 1 年以上被确认的部分同比增长了 32%

¶ Azure

1 企业已经将数以百万计的计算核心转移到 AZURE 上，当前在微软的云上运行的计算核心比两年前多了一倍

2 微软还继续通过 AZUREARC 在混合计算方面保持，微软现在有超过 12,000 个 ARC 客户，是一年前的两倍，包括 CITRIX、NORTHERNTRUST 和 PAYPAL

¶ AI

1 微软拥有云中最强大的 AI 超级计算基础设施，如 OPENAI，用来训练最先进的模型和服务，如 CHATGPT

2 微软已宣布完成了与 OPENAI 的下一阶段的协议。微软将成为其独家云供应商，并在消费者和企业产品中部署 OPENAI 的模型

3 所有这些创新正在推动 AZUREAI 服务的增长，仅 AZUREML 的收入就已经连续五个季度增长超过 100%

¶ PowerPlatform

1 POWERAUTOMATE 拥有超过 45,000 个客户

2 同比增长超过 50%

¶ 办公系统

1 MICROSOFT365 有超过 6,300 万用户，同比增长 12%；推出了 MICROSOFT365BASIC，将高级产品提供给更广泛的用户

2 本季度 TEAMS 的月度活跃用户数超过 2.8 亿，并继续在协作、聊天、会议、通话的类别中占据领先份额。拥有超过 1 万名用户的第三方应用数量同比增长了近 40%，活跃的 TEAMROOMS 设备超过 50 万台，同比增长 70%

3 80%的企业客户使用五个或更多的 MICROSOFT365 应用程序

¶ Windows

1 本季度个人电脑的出货量下降，但 WINDOWS 的每台电脑的使用时间增长了近 10%。本季度月度活跃设备也达到了历史新高。而对于商业客户来说，WINDOWS11 的采用率继续增长

2 此外，云端交付的 WINDOWS 实现增长，WINDOWS365 和 AZURE 虚拟桌面的使用率同比增长超过 2/3

¶ 安全性

1 在过去的 12 个月里，安全业务收入超过了 200 亿美元，帮助客户在云和端点平台上保护数字资产

2 微软是唯一一家拥有横跨身份、安全、合规、设备管理和隐私的综合端到端工具的公司。公司在所服务的所有主要类别中占据领先份额

¶ LinkedIn

1 LINKEDIN9 亿+会员参与度再创新高，每秒就有三个会员注册，这些会员中超过 80%来自美国以外的国家和地区。由于会员来到 LINKEDIN 学习和分享专业知识，NEWSLETTER 创作同比增长了 10 倍

2 LINKEDIN 提供 11 种语言的 20,000 多门课程，公司也正转向基于技能的方法来识别优秀人才，超过 45%的 LINKEDIN 招聘人员明确使用技能数据来完善职位。LINKEDIN 营销解决方案是 B2B 数字广告的领导者

¶ 广告

1 尽管广告市场存在逆风，但公司继续在第一和第三方广告中进行创新。浏览器 MICROSOFTEDGE 连续第七个季度获得领先份额

2 BING 在美国的份额继续增加，START 个性化内容源的每日用户同比增长超过 30%。公司现在正在授权给零售商，并扩大第三方库存

¶ 游戏

1 GAMEPASS 的订阅量、游戏流媒体时长和月度活跃设备都达到了新高，月度活跃用户在本季度超过了 1.2 亿

2 本季度，微软与 RIOTGAMES 合作，向用户提供其 PC 和移动端游戏。即将推出来自 ZENIMAX 和 XBOX 工作室的 AAA 级游戏

¶ 财务表现

1 不计折旧政策调整的影响，本季度度毛利率同比下降了约 2PCT，主要原因是 OEM 收入占比变小，且更多是收入由许可证买断制变更为云服务模式

2 截至 12 月底，公司的总员工数同比增加了 19%，但环比增长不到 1%

¶ 汇率影响

1 我们预期汇率因素导致营收增速下降 3%，成本增速下降 1%，经营成本增速下降 2%

¶ 指引

1 预计 WINDOWS 和硬件业务同比继续下滑到疫情前的水平。LINKEDIN 和搜索业务会继续受宏观环境的不利影响。其他业务在下季度会延续本季的趋势。新签企业合同金额（COMMERCIALBOOKING）预计全年同比持平。剔除折旧调整影响，MICROSOFTCLOUD 毛利率会受 AZURE 影响下降约 1PCT

2 不变汇率口径下，企业 OFFICE365 的增速会环比下降 1PCT，传统买断 OFFICE 产品会同比下降 20+%。个人 OFFICE 业务的增速为低个位数

3 预计 LINKEDIN 的增速会在中个位数，DYNAMICS 的增速为 10+%。预计 AZURE 不变汇率口径下的增速会由 35%左右再下降 4-5%。预计其他云服务收入同比下降低个位数%。预计企业服务收入同比下降中个位数

4 预计 WINDOWSOEM 收入同比下降 30+%，预计硬件业务收入同比下降 40+%

5 预计企业业务（COMMERCIALBUSINESS）在 2023 财年上半年收入同比增长 20%，在下半财年收入增速会下滑

¶ 微软向 OpenAI 追加投资对其计算能力和服务范围的扩展，及其对 Bing 等解决方案组合产生积极影响的时间点？

1 微软坚信，AI 正在掀起新一轮的革命，带给许多行业颠覆性变化。AI 能力是微软可以成为机器人流程和工作流程自动化的领导者的原因

2 微软期待 OPENAI 的创新，OPENAI 期待其自身的商业化，双方共同合作推动 AI 的发展。基于微软在 AI 方面的经验，形成创新和竞争的差异

¶ 对宏观环境的看法？

1 长期来看，科技占 GDP 的比例会提高，关键是通货膨胀调整后的经济增长倍数

2 客户正在优化（最大化）其支出的价值。但优化终将结束，先前优化节省的资金可用于增加云计算负载。因此，公司关注的关键是确保在该领域获得份额，并与客户保持长期合作关系

¶ 指引中云计算和 Azure 的发展将放缓，其中多少是受客户优化其已有产品和服务影响，多少是受宏观因素导致需求减少影响？

1 增速放缓的两大原因：第一，客户希望用更少的钱做更多的事。第二，客户在优化投入或减少新项目。目前来看，优化周期增长解释，新项目投入周期将要开始。事实上，疫情后 TEAMS、OFFICE365 使用率上升。当投入周期再次开启时，座席和利润将增长。近期将推出的 TEAMSPRO 等都将推动 ARPU 增长

2 数据显示有非常高的续订率。虽然新产品的独立销售不易，但 E5 套餐的销售十分强劲。MICROSOFT365 的 ARPU 增长和续订率表现潜在的一致性

¶ 优化周期的持续时间？Azure 将下降 4-5 个点是基于 12 月季度整体的 38%，还是 12 月底 35%的增速？

1 在疫情期间加速了现有工作负载，为期 2 年

2 目前正在进行优化，不认为会持续 2 年时间，但今年会是。优化周期结束后，新项目启动周期不会立刻开始，会逐步扩大规模

3 从 AZURE 在 12 月底 35% 的增速下降 4-5 个点

¶ 关于 Office365 商业版，随着座席接近 4 亿，E5 业务的综合费用 ROI 开始加速，那么座席和 ARPU 之间会更均衡地增长，还是仍然继续倾向于座席增长？

1 当座席增长放缓时，E5 的 ARPU 在 OFFICE365 商业收入中创造了稳定性。现在座席仍在良性增长，E5 也正在进入健康状态

2 公司也在 MICROSOFT365 之外的单用户产进行投资，目正在开发一个新的套件 POWERPLATFORM，甚至是独立的产品如 TEAMSPRO。因此，除了已经起量的套件之外，仍有大量新产品要推出

¶ Azure 的大客户未来计划？

1 大客户正在优化当下规模的工作负载

2 并将节省资金投入新的项目储备中

¶ 量化 AI 的潜在贡献，或量化近期几个季度 Azure 的 GPU 驱动贡献？

1 个人认为现在开始以某种方式将 AI 与其他工作负载分开仍为时尚早

2 即使是工作负载本身，AI 也将成为 AZURE 中工作负载的核心部分。因此，随着时间的推移，每个应用程序都将成为一个 AI 应用程序

¶ 在减少支出方面，今年还会有哪些员工人数和支出方面的变动，以及在做这些决定时的判断标准是？

1 由于同时收购了 NUANCE 和 XANDR，因此 4Q 结束时，除优先级决策外，员工人数同比增长将非常缓慢，以使成本结构与收入保持一致

2 随着投资增加，年同比增长将非常小

微软 FY 2023Q2 业绩会

访谈日期：2023/1/24

具体内容

¶ 业绩概况

1 2023 年 1 月 24 日公司发布了截至 2022 年 12 月 31 日的 2023 财年第二财季财报，本财季公司总营收为 527 亿美元，同比增长 2%

2 其中，生产力和业务流程部门（PBP）营收为 170 亿美元，同比增长 7%；INTELLIGENTCLOUD 部门营收为 215 亿美元，同比增长 18%；MOREPERSONALCOMPUTING 部门营收为 142 亿美元，同比下滑 19%

3 本财季公司毛利润为 352.59 亿美元，同比增长 1.4%；净利润为 164 亿美元，同比下滑约 12%。研发开支为 68.44 亿美元，同比上升约 19%

4 销售开支为 56.79 亿美元，同比上升约 6%；行政开支为 23.37 亿美元，同比上升约 69%。每股摊薄收益为 2.20 美元，同比下滑约 11%。本季度公司通过回购和派息，向股东返还了 97 亿美元现金，同比下滑 11%

¶ 公司在 OpenAI 方面上的进展如何?

1 公司在三年半前开始了 OPENAI 的探索，一直在这方面努力

2 下一个大的平台浪潮将是人工智能，只要能够抓住这些浪潮，就能创造出更高的企业价值。公司会让人工智能渗入到公司的每项业务，并创造出新的解决方案和新的机会

¶ Azure 服务是否会延伸到公司其他的基础业务中去，比如 bing、基本套件或是整体解决方案?

1 AZURE 的核心是将计算、存储和网络结合在一起

2 在过去的三年半甚至四年里，微软一直在非常努力地构建并训练超级计算机，还有现在的推理基础设施。在应用程序中使用人工智能，内部的程序就会涉及训练和推理。AZURE 本身已经发生了转变，基础业务也在发生转变

3 希望未来提供的不止是 AZUREOPENAI，将 SYNAPSE 和 OPENAI 相结合也是在考虑的，POWER 平台具备了整合功能

4 公司拥有杰出的 AI 能力，是公司处于机器人自动化和工作流自动化领域领先地位的原因之一

5 GITHUBCOPILOT 是当今市场上规模最大的基于 LLF 的产品。公司会把人工智能融入到生产中和消费者服务中。CHATGPT 能够构建在 AZURE 上，在人工智能领域的领先将推动微软解决方案的创新和竞争差异化

¶ 从全年的消费环境来看，您似乎认为情况变得更糟，而不是更好，能透露更多的细节吗?

1 从全球来，通胀对全球经济都会造成影响。我们需要关注的是

   a 客户在疫情期间都在不断缩减开支；并且考虑到市场宏观经济风险，也变得更加谨慎。后续将节省下来的钱都投入到后续工作中
   b 公司要确保在这个阶段获得更多的市场份额，进一步建立客户忠诚度。长期来看，公司在市场份额增长方面处于有利地位
   c 关注投资 AI 新动向，和 2019、2020 年不同，后续的所有应用程序将会进一步考虑人工智能的性能、成本、模型等，这会让公司再次处于有利地位

2 这就是我对市场的看法，我们看到的是优化和更加谨慎的做法。但我们从根本上相信，从长期来看，科技支出占 GDP 的比例将会上升

¶ 本季度可以看到业务放缓的趋势，在下个季度指引放缓中，多少是由于客户缩减支出导致的放缓，多少是由于宏观因素而影响需求?

1 两方面原因：与工作负载相关，我们会告诉用户使用我们的产品能够优化他们的工作流程、节约资金。什么时候开始新的项目，调整优先级。这是同时发生的两件事，形成一个循环：前一个项目优化周期结束-新项目开始。这是云消费方面的情况

2 在 PERUSER 略有不同，即购买 PERUSER 许可证的速度在加快。能够保证客户在使用公司的产品并且使用率正在上升

3 OFFICE365 的使用情况，所有指标都在同比大幅增长。之前已经分享过 TEAMS 的数据，在疫情过后 TEAMS 的使用情况有明显增长

   a PerUser 有非常高的续订率，并且更新时有很高的适配度，这意味着公司更多地在所谓内部获客。虽然公司在新产品的独立销售方面遇到了更多挑战，比如很难证明节省了成本，于是周期被拉长
   b 但是套件销售的价值长期表现出来，可以看到 ARPU 的大幅增长，另外续订率也提升。一旦再次缩减支出的周期再次启动将会产生更多溢价。几周后推出的 TeamsPro 可以确保 ARPU 的价值上升
   c 从驱动因素的角度来看，很难区分优化与宏观有多大关系

¶ 能否谈谈优化的周期时间。是几个季度还是多年？如何界定行业中发生的这种优化的持续时间?

1 对当前的工作负载进行优化，并且以新的工作负载开始就是优化了。当完成优化工作负载，就是周期结束点

2 关于时间的问题，比如公司在两年的疫情中加速了工作负载优化的时间。公司认为周琦没有 2 年那么长，而是在今年进行优化一部分工作，同时启动新项目

3 新项目不会在工作负载高峰使用时立即启动。因此，这也许是同时发生的两个周期。可以理解为开始下一组工作之前的临时调整

¶ 考虑到明显艰难的环境，要达到指引年度 20%的固定货币计算的营收增速是很困

难的。是否今年总收入增长 10%以上的弱指引是否也很难达到?

1 在全年总收入方面，没有给过指引。除了关注 WINDOWSPC 市场能否在今年回到新冠前的水平因素之外，其他业务趋势相对一致

2 在营业收入利润率指引中，在 OEM 可能超过 20 亿美元的减少的情况下，与我们的预期相比、今年的利润率只有 1 个百分点的下降

3 对利润率的关注，对优先级的关注，对将投资高回报资产的关注，使公司对今年保持乐观，能够在 Q4 以很好的杠杆率结束

¶ 如何展望后续员工数量? 以及做出这些决定时考虑的标准是什么?

1 Q4 低个位数运营费用增长的指引，因此延迟了对 NUANCE 和 XANDER 的收购。到 Q4 末时，除了一些优先级决策外，员工人数同比增长非常缓慢

2 这是使成本结构更符合收入的决定。公司充满信心，人员流失率意味着一些投资的同比增长将非常小

¶ 目前 Office365 已经达到 4 个亿的用户数量，E5 业务加速增长，是否应该进一步考虑在用户数和 ARPU 之间更均衡的增长，还是继续扩大市场?

1 本季度开始更多地考虑 ARPU 影响。用户数的增长开始缓和，但 E5 的 ARPU 也在同时提高，这使得 OFFICE365 商业收入相对稳定。整体公司仍然有很好用户数的增长

2 而且正在进一步探索 E5 的运营状况。可以看到四五个季度里 E5 的使用率非常好

3 在这种环境下，对于客户而言，我们产品早分析与安全的价值非常高。这是客户可以节约成本的地方，并且 ARPU 也保持增长

4 在 MICROSOFT365 之外，公司还投资于 PERUSER、新套件 VIVA、POWERPLATFORM、TEAMSPRO 等

¶ 微软用户群体不断增大，管理层对不同的项目产品看法会有哪些变化?

1 大客户的典型模式：大客户的共同需求是都希望优化大规模的工作量，并将节省的资金投入到新的项目中

2 在上个季度，公司 AZURE 业务与客户的关系不断加深，获得了更多的长期客户合同

¶ 有没有什么方法可以量化 AI 的潜在贡献?

1 现在以某种方式将 AI 量化还为时过早。AI 是 AZURE 中工作负载的核心部分，任何程序涉及算法推算都跟 AI 相关，不能单独谈

2 随着时间的推移，每个应用程序都将成为人工智能应用程序

平治信息公司走访

访谈日期：2023/2/7

具体内容

¶ 要点

1 布局战略：借助自身强项与 AIGC 技术风口嫁接功能，使平台更有技术性，与达闼合作，希望快速落地

2 AIGC 场景

- a 小说内容生成，包括网文小说、有声书、互动阅读等
- b 根据小说特征制作数字人视频，打开新营销方式

3 公司优势

- a 内容优势：擅长偏快餐文学，特征明晰，适合 AI 学习和生成
- b 渠道优势：能够通过新媒体、短视频等渠道实现广泛、精准的分发

¶ 公司近况

1 公司在 2016 年上市，上市时主营业务以互联网原创小说为主，并积极整合渠道；公司通过微信公众号等新媒体产品形态能够更精准地对接目标用户，公司平台获得新华网、浙数文化、腾讯旗下南京网典的战略投资

2 2022 年政策开始鼓励传媒、游戏、阅读等领域，对娱乐元素需求提升，公司尝试超短剧改编、内容精选、渠道拓展（视频等）。当前进入 AI 技术新时代，公司正加快进度与国内机器人头部公司达闼合作

- a 小说内容生成
- b 根据小说特征制作数字人视频，打开新营销方式

¶ 技术及潜在方向

1 AIGC+ChatGPT 为行业带来很多可能性和巨大的生产力变革，达闼布局了 AI 文本、声音（人声等）、小说场景演绎、有声小说或互动式小说的生成

2 小模型更多解决分类问题，公司用小模型实现了生成文本、场景式声音，甚至创作歌曲、舞蹈，当前的问题是将单项能力融合在大模型中彼此互动，像大脑结合听觉、视觉、触觉，未来想象空间大

3 ChatGPT 大模型具备超强的理解力、对上下文的推理能力、跨语言能力，可带来很大的行业影响，商业和运营方式会产生很大的变化，传统的 NLP 等方法和工具会发生颠覆

¶ 公司原来的产品作为训练数据库与 AIGC 进行结合，未来会有怎样的愿景？

1 内容：公司擅长偏快餐文学，特征明晰，适合作为 AI 训练数据库

2 渠道：除了原创小说以外，达阂可以根据小说特征做出数字虚拟人，在抖音快手增加曝光度、提升推广力

3 愿景：AIGC 是划时代的技术，公司希望与达阂磨合，将训练数据库做得更大，训练 novel-GPT 模型、打磨后向市场推出。尽快将产品推向市场，形成可读性高、吸引力强的文章

4 从研发角度，ChatGPT 不是要取代创作者，而是让现有内容生产加快，产量得到极大的提升，生成效率高，可以形成初稿或提纲，原创小说作者可以基于 AI 生成的文本再调整、提升

5 单一模态的内容创作能够形成多模态的内容表达：公司会将文本生成声音，比有声书更具情感，了解小说里要表达的情况，像有情感的朗读者。数字人能够进行表演

6 此外，除了单一的阅读式小说，还可以有互动式小说，大模型能够理解复杂的语音，能够与人一直聊下去，读者可以参与进来

¶ 公司下游流量变现方式主要包括哪些？

1 公司的渠道是优势，最早从自媒体、微信公众号沉淀用户进行变现。近 3 年阅读板块发生变化，产业监管严格、短视频崛起使用习惯发生变化

2 公司适时转化，开始制作超短剧，每集半分钟，共 100-200 集，剧情紧凑、适合现在的节奏。2022 年公司在抖音获得广告主日投放 100 万，处于抖音渠道的第一名

3 尽管移动阅读板块有调整，但公司一直跟进内容精修、渠道拓展等，并通过国内头部 AI 公司进行功能叠加

¶ 目前公司训练数据库对 AIGC 的训练程度？

1 公司在加快进度，公司的 novel-GPT 正从无监督到有监督、参数调整的过程

2 大模型对数据的要求高，预计数据量不少于 200T，且训练周期久，公司在针对细分领域寻找数据

¶ 目前公司与达阂的合作形式？利润和成本的分配？

1 先定型模型，前期写出的小说、机器人版权和平台归平治

2 产品确定可以商业化之后根据市场情况，与达阂进行分成的确定

云从科技走访

访谈日期：2023/2/3

具体内容

¶ 公司情况

1 云从最开始是做视觉，创始人周博士认为发展第一阶段需要某个技术进行单点爆发，2014 年后出现了很多单点技术，包括 NLP，但单点技术从逻辑上解决不了真正的智能化问题

- a 云从占银行身份认证 80%的市场，六大行都用，大机场安检也用，人脸识别细分领域是绝对头部，但单个场景应用不能支撑智能化的整个应用平台
- b 由于云从不是硬件公司出身，学习成本较大，并且 AI 是一横一纵发展，所以云从第二阶段做横向路径，做感知—认知—行动的闭环，解决人的问题，解决场景智能化
- c 把人的思考做到 AI 的平台，把听、说数字化并建模，但不同于现有的问答 AI，要发展 AI 的知识结构和认知判断

2 第二阶段技术平台化后可以改变入口和流量，第三阶段有更多的落地场景，让 AI 改变很多场景的体验和效率，人类做更有创造性的事，这是云从的技术迭代和产品设计规划

- a 云从在 A 股，最初就确定了在国内发展，行业落地时涉及很多数据要素，比如数据安全性，云从最初一直发展这块。同时 A 股看中商业变现逻辑，注重商业落地和业务层面，所以云从的技术迭代路径容易被大家忽略
- b 云从发展类似汽车产业链的 tier3,2,1，最初 tier3 提供单点的视觉识别技术。但技术发展迭代很快，所以我们做到 tier2，做很多功能模块，平台、训练、推理等，既可以做底层，也可以给集成性大公司提供产品
- c 目前我们 tier3，tier2 相对较全，以后再往上做 tier1，真正做到人机协同标准化产品，比如虚拟人、金融智能客服、移动机器人大脑

3 云从在横向上做感知—认知—行动，纵向上做 tier3，tier2，tier1。目前 ChatGPT 大热证明我们技术路径是对的，ChatGPT 是通过数据去堆，加上不同场景下的知识贯穿，这和我们人机协同理念一致

- a 因为我们路径和 ChatGPT 一致，所以最近股票上涨，人机协同已经在一些行业验证，先做 tog，tob，再到 toc，因为 toc 个性化要求很多
- b NLP 也非常重要，但 NLP 技术难点很多，我们有 NLP 的预模型，但模型在数量和训练上还没有到临界点

¶ 从细分市场产品层面，我们在未来几年有什么规划？

1 大的结构上是两类产品，一类是 CWS 操作平台，就是人机协同操作平台。这类其中有两种版本，一种是通用型版本，类似开放平台基，isv 从其中调取，加上行业自身的智能化，我们做 50%的底层，分为 tog 和 tob。tog 是政府、公安、法院等，治理和智慧城市一类，tob 是金融，以银行客户为主

2 另一种是面向智慧园区和智慧商业，和小的商业零售客户，开放平台基座给他们后，根据客户需求做两类产品

a 一是行业侧操作系统，如数字孪生平台、感知平台、业务流程平台，成熟度达到 60%-70%，金融是我们的主流行业，达到 80%-90%

b 二是提供平台+解决方案，比如面向银行有 16 个解决方案，方案中分场景做各种应用软件

3 第二类是后续的行业拓展，AI 领域有行业属性，一家公司很难面面俱到，AI 智能化和 NLP 发展需要耳聪目明，需要视觉识别和语言语义分析，我们这块较强，很多互联网公司的数据分析和运用较强，我们今后会与他们合作来达到 AI 更加拟人化和智能化的要求。以后慢慢向 tob 行业拓展和 toc 产品的训练

4 近几年做擅长的行业标准化产品和核心应用，后续快速进行技术投入，优化人机协同势能，拓展更多场景，根据外界环境和资金情况向前推进

¶ 商汤科技与你们比较相似，两家有什么不同点？

1 商汤从大逻辑和技术迭代上与我们比较类似，商汤资金更充足，他们想做统一化标准化的大模型，但这对我们来说投入成本较高

2 商汤能做，但他的时间周期较长，不可控较多，我们采取纵向上分层的模式，分行业的模式，最终形成串联。我们和商汤观念一致，最终形成技术平台化和行业落地，只是执行路径不同

¶ 未来会继续加大研发投入，还是找挣钱的市场，保持研发投入稳定？

1 首先我们是技术公司，研发投入不会低。我们会在赛道上长期布局，长期的路径是数据结合知识，坚持大模型、预训练，2020 年开始就做 NLP、OCR 和视觉语言模型，但同时在研发上会量入而出，我们坚持技术投入比率小于营收增长率，保证公司稳定的现金流，投入方向上坚持人机协同

2 我们是商业公司，需要赚钱，所以会和第三方合作，这样周期最短，进入行业最快，成本最少。内部在财务上坚持正循环，坚持让每个产品落地到每个业务线上都赚钱。同时我们会扩大生意面和降本增效，比如各地设点降低人员成本，实现财务好的表现

3 财务盈利在 2025 年左右，不会为了报表盈利降低研发投入

4 我们不会直接砸钱，考核人员时很严。销售人员要考核营收、考核合同、确认收入、考核利润、毛利、回款。考核研发人员一是效率提升，二是让研发出的平台使用门槛更低，三是时间要快、成本要低

¶ 产品的收费模式怎样？

1 目前大部分收费是项目收费。如果是大量业务可以过程中确认收入，过程中回款。商业模式分两步，一是面向大客户提供定制化的产品和服务，进行项目收费

2 第二步是对大客户部署成功以后收运营费用，比如智慧城市里运营数据服务和检测报告。再有面对腰部客户，如金融里的六大行，不需要太多定制，就卖给他们

实现使用收费

3 以后在 toc 层面会有按次收费，比如智能货柜。还有在金融风控上是利润分成，按控制多少风险分成，这块占比不高。随着 tob 和 toc 标准化程度越高，以后运营收费会更高

¶ 报表上标准操作系统占比降低，是大客户更多了吗?

1 我们做大行业时，会先立标杆，这类客户较大。其中除了操作系统外，有些核心应用会找第三方，同时有一些硬件需要配套，所以其中会有很多非自有产品

2 在财务计量时，除了自有平台外有其他产品，就会归为行业解决方案，所以操作系统占比降低。当然随着技术能力上升，业务规模也越来越大，目前大客户也更多

¶ 在全国的银行或证券，渗透率能做到多少?

1 AI 在落实到商业路径上是场景化的，我们做认证起家，所以在认证平台上我们的市占率和渗透率非常高，基本占据头部企业，六大行都用云从。在其他细分场景渗透率各有不同，比如小城商行对风控需求较高，但大银行可以自己做。我们在北方和东南渗透率较高，每一个细分场景的具体渗透率没有统计过

2 整体上在智能化角度对银行是 AI 的头部企业及第一大供应商品牌

3 在银行上我们占据了技术的门槛和入口，虽然收入规模不一定大，但占据入口后平台发展会更顺利

4 治理的业务结构和金融不同，tog 的业务一般订单规模比较大，需求量比较多，所以治理的绝对收入就会比较高，比金融业务收入更高

¶ Tog 业务在哪些省份额更高，以后还有提升空间吗?

1 原来的公告里会有，在北部，东南，西南这些地方更多。行业不同订单规模不同，金融体量小，但客户数量迭代比较多，毛利比较高

2 tog 业务总量不多，但每一单的金额比较大，具体数据可以查找财务报表

¶ 以后有没有股权融资的计划?

1 上市之后肯定会做二级市场融资

2 但还没完全提上日程，后续会进行讨论

科大讯飞表现分析

访谈日期：2023/1/31

具体内容

¶ 核心观点

1. 我们认为今年整体讯飞都会是一个非常强的票。从主题上说，ChatGPT 和数据要素这两个计算机版块最强的主题，讯飞都是纯正标的。讯飞的开发者平台、讯飞超脑和 ChatGPT 的商业逻辑非常像，且讯飞拥有稀缺的技术能力。而数据要素有望带动讯飞的智慧城市业务获得市场重估
2. 更重要的是它的核心业务今年会迎来确定性反转。除 22 年之外讯飞在过去 10 年是全部 A 股里面唯一一个每年增速都在 25%以上的公司
3. 在去年低基数的情况下，23 年增长 25%是一个非常保守的估计，给 7 倍 PS 看 1,680 亿市值。长期来看，对 AI 公司来讲，既掌握数据端口又有技术的公司才是稀缺标的
4. 讯飞的核心竞争力在于它利用技术和自己的先天禀赋，实现了对很多的数据端口的卡位，这是一个长期逻辑

¶ ChatGPT

1. 讯飞的开发者平台、讯飞超脑和 ChatGPT 的商业逻辑非常像，而且讯飞在这方面的技术储备也是国内顶尖的，在上市公司+非上市公司中都算非常稀缺标的
2. AI 过去几年在一二级市场的表现差强人意，核心原因是商业落地场景没有那么明确。之前的 AI 技术都是集中在安防、智慧城市等领域，说白了中间有多少应用了 AI 是比较隐性的。而 ChatGPT 大大降低了创作门槛，换句话说，通过一个聚沙成塔的方式，使得 AI 可触达的场景变多了
    a. 这种新的商业模式会使得 AI 技术的价值会更加显性。为什么说讯飞开发者平台、讯飞超脑和 ChatGPT 有异曲同工之妙
    b. 因为讯飞的平台也是通过让开发者轻度地去调用他的 AI 基础技术，去给开发者提供一些技术模块，然后获得一些收入。无论是开发者直接付的，还是通过类似广告的形式
3. 尤其在 IoT 这种非常碎片化的市场里，流量聚拢是难度比较高的。所以在商业逻辑上，讯飞的开发者平台其实是非常类似于 ChatGPT 的。因此实际上它的稀缺性非常强，基本上国内只有这两三家巨头能做
    a. 此外，科技部此前设立了认知智能的全国重点实验室，非常重要的一块就是认知智能的预训练模型
    b. 而科大讯飞承建了中国唯一的认知智能国家重点实验室。另外，从实际业务来说，讯飞的开发者平台业务去年增长了 30%，AI 的调用量增长了 38%
    c. 过去几年增长一直都不错。所以虽然大家都知道 ChatGPT 是一个偏主题性的机会，现阶段很多公司都涨了，但对讯飞来讲，ChatGPT 恰恰不完全是纯主题炒

作

¶ 数据要素

1 智慧城市业务已经中标了一个比较大的项目，之前市场上基本没人给估值，数据要素有望带动这部分业务获得市场重估

2 讯飞在去年 12 月以 5 个多亿中标了安徽省的一体化数据的基础平台项目，所谓的数字安徽项目，背后其实就是 AI 产品技术实力的体现

3 安徽省的电子政务有说法在全国是能排到前五的，全国性的电子政务现场会曾经在安徽开过。所以该业务的中标有望成为一个重要的标杆，帮助讯飞进行业务拓展

4 由于讯飞的业务比较多比较杂，之前市场更关注它的教育和消费者业务。我们认为随着整体的数据要素的地位得到确认，包括讯飞的一体化的数据平台建设，其智慧城市业务有望在二级市场获得一定重估

¶ 业绩

1 讯飞这几年的发展脉络和投资逻辑是从 G 端到 B 端再到 C 端。2023 年第一个是教育业务，无论是区域化的订单快速增长、市占率的提升，还是 C 端业务的放量，教育作为讯飞最核心的业务

2 23 年会迎来爆发式增长。第二个我们最关心的消费者业务，不论是从智能硬件切入到更多赛道，还是随着出行放开翻译类单品快速回归，消费者业务今年也都会快速增长

3 讯飞刚刚发布 22 年的营收预期。我们可以看到除了 22 年特殊情况之外，讯飞在过去 10 年是全部 A 股里面唯一一个每年增速都在 25%以上的公司。在去年这种低基数包括订单延迟的情况下，23 年业绩回暖可期，25%是非常保守的增速预期

¶ 29 日科大讯飞发布业绩预告

1 预计 2022 年实现营收 183.14 亿元-201.45 亿元，较上年同期增长 0%-10%；公司净利润为 4.67 亿元-6.23 亿元，同比下降 60%-70%；扣非后净利润 3.92 亿元-5.38 亿，同比降幅为 45%-60%）

2 估值：在 25%的增速下，给 7 倍的 PS，目标市值 1,680 亿。可能有投资者会问，为什么他盈利了还能用 PS？其实这里涉及到一个长期以来对讯飞不太公平的点

¶ 市场上除了讯飞之外，还有哪个 AI 公司是真的一直在赚钱的吗？有哪个 AI 公司是真的用 PE 来估值的吗？

1 不能说其他公司不赚钱，我们只能用 PS，到了讯飞这边我们去抠它到底是 40 倍还是 45 倍 PE，对讯飞来讲其实不公平

2 背景资料 ChatGPT 是一款对话式 AI 聊天机器人，由微软旗下的人工智能研究实验室 OpenAI 于 2022 年 11 月 30 日发布。它能写论文、编代码、写小说，甚至知道

绕开人类提问中预设的价值判断、道德倾向等陷阱

3 ChatGPT 一经推出，就在人工智能生成内容（AIGC）领域引起轰动，被评价聪明得“像人类”，短时间就达到了百万以上用户

- a 2019 年，微软重金投资了 OpenAI，今年 1 月 23 日，微软宣布再追加数十亿美元投资。微软现任 CEOSatyaNadella 表示，在与 OpenAI 合作的下一阶段
- b 将为客户提供最好的 AI 基础设施、模型和工具链。不仅如此，微软也计划将 ChatGPT 整合到旗下的 Bing 搜索引擎、Office 办公软件、Teams 聊天程序等产品中
- c 谷歌等互联网巨头明显感受到了这项新技术的威力。据《纽约时报》报道，去年底，由于 ChatGPT 在全球爆红，谷歌 CEOSundarPichai 在公司内部发布了“红色警报”（CodeRed）
- d Pichai 要求多个团队集中精力，解决 ChatGPT 对本公司的搜索引擎业务构成的威胁

4 1 月 31 日，ChatGPT 概念股持续活跃，多只概念股股价走高，收获涨停板。业内人士认为，ChatGPT 解决了人工智能模型的一些核心技术痛点，应用前景非常广阔，技术、内容、硬件的相关企业都将受益，相关技术在金融、政府、医疗等领域的落地也值得期待

5 与此同时，需要注意这一概念在技术迭代、商业化和监管层面的相关风险

6 券商观点财通证券研究所认为，ChatGPT 受到广泛认可的重要原因是引入新技术 RLHF（ReinforcementLearningwithHumanFeedback，即基于人类反馈的强化学习）。RLHF 解决了生成模型的一个核心问题，即如何让人工智能模型的产出和人类的常识、认知、需求、价值观保持一致

¶ 总结

1 ChatGPT 是 AIGC（AI-GeneratedContent，人工智能生成内容）技术进展的里程碑。该模型使得利用人工智能进行内容创作的技术成熟度大幅提升，有望成为新的全行业生产力工具，提升内容生产效率与丰富度。他们建议关注数据领域、算力领域、算法领域以及 AIGC 领域的上市公司

2 不过，财通证券研究所也指出，ChatGPT 目前使用有局限性，模型仍有优化空间。具体在于，ChatGPT 模型的能力上限很大程度是由奖励模型决定，该模型需要巨量的语料来拟合真实世界，对标注员的工作量以及综合素质要求较高

3 当前，ChatGPT 可能会出现创造不存在的知识，或者主观猜测提问者的意图等问题。型的优化将是一个持续的过程。若 AI 技术迭代不及预期，NLP 模型优化受限，则相关产业发展进度会受到影响

4 此外，ChatGPT 盈利模式尚处于探索阶段，后续商业化落地进展有待观察。未来，相关监管政策也可能出台，对行业发展产生影响

- a 方正证券研究所认为，目前，国际前沿 AI 技术发展迅速，AI 的商业化图景也越来越清晰。我国在自然语言理解及相关 AI 技术领域处于全球领先水平
- b 国内 AI 大厂都在加大 AIGC 领域的投入，人工智能技术提供商特别是 NLP（NeuroLinguisticProgram，神经语言程序学）头部厂商将率先受益，其中既包括独立技术提供商，也包括互联网厂商

5 不仅如此，文字、图片、视频、音乐各领域的内容提供商也将得益于 AI 技术驱动下数字内容的快速发展，AI 芯片供应商也将一定程度受益。从受益顺序来看，依次为技术提供商、内容供应商和 AI 芯片供应商

科大讯飞 22 年度业绩预告说明会

访谈日期：2023/1/30

具体内容

¶ 董事长发言

1. 去年是很不容易的一年，疫情反复影响很大。Q4 本来是要冲刺的，想办法把延后的项目在 Q4 抢回来，没想到四季度疫情防控反复、年底员工和客户感染
2. 导致有 30 亿合同后延。这些项目没有取消的，也没有丢单的，23 年会陆续实施，估计在上半年也会有很多项目招标完成。去年我们投入了 8 个亿，如果不投这八个亿，扣非利润会好看很多
3. 另外还有国产化替代问题。所以去年 10 月讯飞再次被列入实体清单，给与了更严格的控制，但我们迅速做了应对措施，除了备货等常规处理外，在训练服务器上也下了功夫，还尝试用更小的模型实现大模型的效果
4. 经过 2022，虽然很多业务受到影响，但我们看到 AI 在教育、养老、医疗许多民生刚需领域的需求是越来越明确的。而在这些领域，我们已经形成了越来越强的优势
5. 预计未来五年，根据地业务要到 500 亿以上（到 650 亿左右），形成 200 亿毛利。未来持续运营型、流水型业务要代替项目型业务，成为公司基本盘，未来要占到公司收入的 80%以上

¶ 国产替代

1. 主要算法模型已经在国产自主可控的平台上做了有效工作，否则我们担心研发源头受阻
2. 类似 ChatGPT 这种模式，讯飞也有自己中文预训练的模型，在 Github 上好评如潮。而且我们不需要这么大的模型，以小得多的规模，在汽车等场景就可以实现不错的交互效果
3. 相信 2023 年公司会回归良性增长，我们把增长目标从收入切到利润，未来利润增长达到 30%以上

¶ 高质量发展如何达成？

1. 我们在各个事业部都是做三年计划的，会有明确业务结构比例目标，会有相关的导向。第二，重点领域前后拉通，形成平台化，让普通工程师就能做定制化，形成在国产平台上更高的效率
2. 第三，人均效率和现金流要有提升。去年客户的钱其实很紧，但我们的回款还是增加的，人均效能我们也有提升的机制
3. 今年我们实现的是零基预算，所有部门压缩 10%的编制，90%的人完成过去的

事，剩下的 10%资源拿去给到战略性板块。另外，也会有配套的人才政策

4 在 2023 年，公司项目管理导向会走向利润导向（过去收入占比比较大），让大家形成节省花费的意识

5 过去销售只想把单子卖出去，后台业务部门对成本又不太控，现在通过数字化手段端到端打通，每个员工的考评都是和项目挂钩的，会好一些。22 年已经把基础打好了，23 年会进一步释放红利

¶ 讯飞在大模型这块布局比较早，也取得了很好的应用效果，请问讯飞的技术布局以及未来可能的落地？是否可能有爆款产品？

1 微软已经一百亿美金投资 openAI 了，大学生已经开始用 ChatGPT 写论文了。不过 ChatGPT 确实不是横空出世了，18 年开始大模型就已经陆续出现了。20 年起，GPT3 掀起了大模型的热潮，关注度一直很高

2 ChatGPT 是对 GPT3 的进一步改进，融入了代码型的数据，提升了模型的思维能力；通过成千上万个任务（prompt），指导学习，实现统一交互式的回复，更好地激发大模型的潜力；通过人类反馈进行加强学习，能够优化模型表现。ChatGPT 给大家感受到了很大的鼓舞

3 讯飞是最有希望做出最好的中文对话系统的企业，并且是最有可能在教育、医疗等领域落地的。ChatGPT 使用起来很贵，因此大规模商用还有一些问题需要解决，并且在一些专业领域可能表现暂时不如一些专业模型

4 讯飞在大模型这块是很有积累的。我们承载了全国唯一的认知智能实验室，也已经有开源的预训练模型，成为业界流传最广泛的中文预训练模型之一，Github 上排名第一

   a 我们的参数量远小于业界其他模型（1/20 的参数量），表现却不输，因此我们有信心去取得突破
   b ChatGPT 是需要收集全球数据的，但预计未来他们也很难进入中国市场。因此在中国国内的工作就很重要。我们已经启动了相应的 1+N（一个模型，n 个领域）专项技术攻关
   c 不过 ChatGPT 确实有一种从量变到质变的感觉，令我们也很激动

¶ 22 年有 60-70 亿的教育收入，里面各类业务的占比是怎样的？

1 由于 22 年教育大项目延期比较多，因此 GBC 里去年 BC 已经占大头了。今年的增长，个册新增了 300 所学校，今年预计新增 600 所，学生付费转化率超过 70%，收入目标增长 40%。英语听说这部分，去年营收增长 35%，我们又增加了 12 个地市

2 新增了两个省的高考，今年目标是 45%。学习机去年关注线下店的情况，比线上更有稳定性。去年增长 53%，今年预计学习机的增长，目标是 75%-80%，而且线下增长为主

3 今年教育业务会出现比较高速的增长，今年的比例运营型+流水型仍然主要占 60%左右

4 其他大项目，我们也在按根据地的逻辑在推进，尽量去向年度收费的模式靠拢

5 去年教育小 GB，就是通过渠道去走的那些业务，是正常增长的，增长了 19%，今年要翻番

¶ 汽车这块，去年新增了很多定点储备车企，公司汽车业务在布局上过去的节奏、未来规模放量、以及出海的节奏？

1 过去我们最重要的产品还是语音套件，就是把我们的语音能力给到汽车厂商，让他们开发。不过，22 年已经形成了很多产品，包括

   a 智能座舱里信息域的域控制器，类似于一个主机，可以控制导航、音乐等信息服务。原本我们提供语音交互工具比较多，现在我们可以提供软硬一体化的域控，已经有定点了
   b 智能音箱，音频技术和 AI 技术结合起来，可以显著提升音响效果。新能源汽车由于噪音很低，对音响要求就比较高。讯飞的智能音响已经显著超过了国外一些比较贵的品牌，广汽一些款和智己都用了，大家可以去体验
   c 讯飞现在在 L2+已经完成了从技术到产品的研发，在未来不久的时间里，在汽车上完成定点

2 我们在面向非洲、中东等地区，已经和国内出海做得比较好的车企达成了合作。我们会支持国内车厂出海，也在日本设立了子公司，已经在一些海外车型上有突破

3 在预训练模型上线后，我们有望把语音交互体验再升级。并且我们的语音合成技术，也已经有很大的突破了，现在基本和播音员一样了

4 今年开始，讯飞要重启国际化的进程了，过去几年暂缓都是因为被美国加入实体清单，但现在是时候了。讯飞的多语种技术是全球最好的，只是需要和当地数据重新训练落地而已，我们有信心去做核心技术输出

5 第二，是消费类产品要到海外，往韩国、日本、欧美等地，包括翻译机等。由于今年出国潮主要在暑假，所以预计销售高峰会在二季度末。另外教育，我们想切入的是英语学

6 以英语口语考试和听力考试为切入点，雅思已经和我们合作了，这个模式也是可以往全球需要学英语的地方输出的。我们希望五年内成为全球最出色的多语种技术提供商

¶ 22 年受疫情影响递延的合同，大致是怎样的结构构成？预期落地和收入确认节奏是怎样的（集中在上半年还是时间周期更长）？

1 有 60%是教育，40%是智慧医疗和城市

2 绝大部分上半年能招标完成，Q1 全部完成会有压力（有春节，马上又要两会了）

¶ 23 年的人员计划？费用端这边是怎么考虑的？

1 没有比较大的招聘计划，人员会基本稳定在去年年底的水平。通过内部的组织效能的优化（前面所述零基计划）

2 通过员工能力的提升，可以实现今年的目标。今年业绩会保持一个比较好的快速

增长

¶ ChatGPT 如果出了的话，刚刚聊得比较多的都是 B 端的应用，会有 C 端使用的版本吗？

1. 它在 C 端也有价值的。我们现在做的第一个尝试的是老人耳机，我们那个 2k 多，戴上之后就能听见，体验比国外几万的要好。医疗这块也是面向个人的，家庭医生和慢病管理，我们已经看见了很清晰的发展路径
2. 我们现在也推出了业界最好的虚拟人交互平台，未来元宇宙消费经济这块，会有很大的机会

¶ 各个板块里运营型、流水型的比重？

1. 总体占 58%左右，增长 26%，今年增长肯定比 26%更高，2023 年希望达到 60%以上。教育中占比 50-60%，我们今年也希望保持这个量。长期来看希望能占到 80%
2. 消费类产品里，开放平台上很多精准投放类的广告业务，也能算持续流水型的
3. 现在主要是智慧城市里的比例比较低。现在也开始转变了，比如法庭业务，合肥那边用了我们基于 AI 的断案支持服务后，断案效率增长很大，未来也是需要给我们持续付费的
4. 2023 年，根据地业务增长主要还是教育、医疗、汽车、消费者业务

¶ 认知智能方面，我们领先，但是和其他大厂拉开的差距有多大？

1. 我们跟国际大厂相比，从基础科研能力上，我们还是有差距的。人家从芯片算力、行业资源上，比我们好很多。三五年前差距更小，现在我们不得不看到这个差距。我们现在要在算力只有 1/10 上，通过算法去弥补
2. 我们也要紧跟国际最新进展，保持并跑，在我们擅长的领域做到领跑。语音、翻译、OCR 领域，我们有信心一直领先，还有一些行业领域大厂不如我们（比如医疗等），这个我们也是有信心的。我们要在典型赛道上，始终保持领先，并且在商业变现上形成综合优势
3. 和国内比，国内真正在做 ChatGPT 的可能就是百度，但他们的重点跟我们是不一样的。在其他高校，我们也有很多合作。另外的大厂，能力和技术积累本身就不够，无法跟我们竞争

¶ 机器人方面未来的打算？

1. 去年我们已经推出了轮式机器人，现在先提升运动能力。我们跟国内市场上口碑比较好的机器狗，拿过来之后通过我们研究院的能力，提升了各类运动能力
2. 在此基础上，已经在 1,024 发布了机器人超脑平台，希望更多机器人公司来用我们的超脑平台，形成迭代，做好进入家庭的准备。我们第一步是工业机器人，未来会进入到家庭
3. 今年会有一些进入家庭的产品会让大家看到，但不会直接出家庭机器人。先会有

外骨骼机器人，然后再推进到家庭陪伴型机器人

¶ 国产化是公司重点方向，公司在相关领域的适配和发展情况？未来几年的投入规划？

1. 训练服务器已经国产化了，作了比较好的切换。过去三年算子库关键领域已经弄完了，投入的工作量是英伟达的 20 倍，但已经移植得差不多了，心里有底了。切换后有些地方效率是原本的 70%-80%，有些比过去还高
2. 算法的优化，我们要在算力有限的情况下，用算力弥补
3. 推理服务器也已经国产化了
4. 典型产品的国产化替代，原本有些部件用的是国外芯片，现在都切成国产化了

科大讯飞访谈交流

访谈日期：2023/1/30

具体内容

¶ 2022 年业务情况

1 2022Q4 本来要冲刺，但面临三重挑战

2 20 大后主要城市疫情此起彼伏、12 月放开后密集感染导致业务后延、国产化替代问题（去年 10 月 7 号再次被列入实体清单）

¶ 未来展望

1 各领域刚需越来越明确，已形成越来越强的比较优势，找到成长的确定性越来越有底气；根据地业务年年会有增长

2 未来 5 年预期 500 亿以上的收入、200 亿的毛利；算法模型基于自主可控平台上做工作，为长期发展稳住了阵脚

¶ 增长目标

1 讯飞 2012-2021 年每年收入增长都超过 25%，22 年按暂停键，再往后高质量发展导向，收入增长切换为利润增长，未来 30%的利润增长目标

2 不会完全压缩成本，人员要控编，做好现金流管理，但更大增长来自于刚需+代差的目标市场价值

¶ 高质量发展举措

1 各事业群、事业部、大区，未来 3 年，明确运营型流水、渠道型流水等业务结构的比例，要占到一大半，走向持续收益，穿越经济周期；核心技术引领，前后端拉通，形成国产可替代平台上的效率

2 人均效率提升、现金流管理，有强的配套机制，今年所有部门压缩 10%编制，剩余 10%拿出来分配到公司战略方向上

¶ ChatGPT 观点

1 讯飞是最有希望把 ChatGPT 做一个最好的中文对话系统

2 也能够在教育、医疗等赛道做进去

¶ 教育业务回顾与展望

1 2022 年教育大项目延期较多，教育中 B、C 占比多；个册去年增长 12%，续购率

提升到 91%，付费转换率超 70%，今年预计新增 600 所学校，收入增长目标为 40%

2 英语业务去年增长 35%，今年增长目标为 45%；学习机去年增长 53%（纯 C 端），线下店成为持续流水，线下翻番，今年增长目标为 75-80%。今年教育业务会出现高速增长，比例上预计运营型+流水型占 60-70%，项目型 30-40%

拓尔思访谈交流

访谈日期：2023/1/30

具体内容

¶ 专家介绍

1. 公司一直在持续跟踪 CHATGPT，这不是一个特别新的东西，很多性能和能力已经在 GPT3 中有所体现
2. CHATGPT 能够引起轰动的原因
   a. Instruction building，相当于一个指定的调试，它通过这种对话式的方式，能够去理解人类的一个指令
   b. 其强化学习能够给符合人类的预期的答案，他不断通过强化学习的方式，使得它生成的答案更加接近人类的预期
3. 实际这两个核心的技术其他公司都在做，它的技术也不算特别新，但 OPENAI 做到了一个整合。这是基于 OPENAI 的工程能力，在行业里经验的积累是非常重要的，整个大模型的训练、数据的收集、模板的设计等这些，需要有一些经验的积累才能做得比较好
4. CHATGPT 能够推动整个对话式 AI 相关的产品，或者是相关的公司。相当于是我们的一个数字劳动力、一个助力。之前的 AI 主要处理执行阶段的工作，CHATGPT 出来后能够把理解和反馈的阶段搞定。我觉得这可能是生产力的第五次革命的时代，我们将这种 AI 称为是“数字化劳动力”
5. 我觉得未来数字化劳动力将会成为一种基础设施，依托这种 NLP 等技术，能够让机器跟人类完美的协作。机器能够帮助去解决包括不管是前端的对客服务或者是中后台运营协同的一些相关的任务，能够让传统的劳动力爆发出更高效的生产力水平
6. 麦肯锡在报告提到：预计 2030 年的时候，数字化劳动力市场规模会到 1.1 万亿的水平

¶ 数字化劳动力应用场景

1. 信息查询类，能够去替代一件枯燥重复性的劳动，可应用在融媒体行业
2. 专家咨询类，数字化劳动力可扩充像专家这种稀缺的劳动力，如法律、医疗行业
3. 助手类，可能是一个智能创作的助手，在创作领域发挥作用
4. 交流类，把机器理解力跟反馈的问题解决了之后，能够更好地满足人类情感的交流的需求。可应用在医疗养老机器人、青少年陪伴的领域、游戏领域等

¶ 核心观点

1. 对于 AI 模型我们认为其水平大概在人类的中小学的水平，只有通用的知识。我觉

得未来的机会应该在垂域拥有数据、拥有一些行业优势的公司，巨头是做不了的，因为这些数据掌握在各个领域的公司的手里

2 此外由于中文的复杂性，在中文领域还需要去有一些相关的数据去培养 AI，去让大模型去继续成长

¶ 市场预估

1 养老市场，根据国家卫健委的数据，2021 年我国有 1.9 亿的老年人患有患慢性病，其中失能失智的达 4,500 万，其中患慢性疾病的比例高达 75%。在专业机构里头，护理轻度的失能老人，护理比例为 4:1（四个老人配一个护理员），重度是 2:1，按平均 3:1 计算

2 整个养老护理的需求量达到了 1,500 万。2020 年，我国仅有 50 余万养老的护理员，远远不能满足需求，所以整个护理缺口达到超千万。按护理机器人 5 万单价计算，按缺口 50%计算，整个护理机器人的市场规为 2,500 亿

3 法律咨询领域，根据司法部发布的数据，全国的律师 2021 年，律师办理各类法律事务有 1,300 多万卷。中国的律师平均费率每个小时 2,788 块钱，每个案件平均服务时长 10 小时，整个法律的咨询的总体市场规模达到 3,600 亿

4 拓尔思对于养老机器人领域，也在做一些深度的研究，进行前期的数据的收集，还有兼容这种模型的搭建，还有实际的应用落地。智能创作领域，公司在全国的媒体的领域占有率较高，目前已经有 11 个单位在使用公司相关产品

5 政务领域，中央人民政府的政策库是公司来做的，政策公文发布背后都有大量的知识需要去关联，去分析辅助，需要去相关的一些数据或者是 AI 的算法去辅助

6 目前全国是有 400 万公文的协作人员，我们已经在成都、南京有一些企业用户的应用案例了。公司在问答领域也做了很多的探索，所以已经有相关的场景的落地

¶ 公司业务收入

1 软件产品，即人工智能、大数据和数据安全产品，在 2021 年公司的业务总收入占到了 63.16%

2 每一个产品的背后，实际上都带了云和数据的服务（SAAS），在 2021 年占比 36.84%。拓尔思也在不断的探索一种比较稳定、持续发展的商业模式

¶ ChatGPT 作为大模型需要大量数据和大算力，那从基本投入来看是不是会比以前要大很多？投入和回报怎么衡量？

1 GPT 的第一个版本的参数量只有 1.17 个亿，数据量 5G，在 8 个 GPU 上训练需要一个月时间；第二个版本参数量 15 亿，数据量 40G，PFS-DAY 是 7.86，消耗资源是在 256 块谷歌云 TPUV3 上训练一周的时间

2 GPT3 参数量达到 1,750 亿，预训练数据 45TB，PFS-DAY 是 3,640，是第二代的几千倍，成本 1,200 万美元

3 OPENAI 理念是要推动一个普惠的 AI，让大家都能用得起的事情。当然他后续也是需要去考虑商业模式或者是商业变现的一个事情，但是他现在还是没有太清晰的

或者是太成熟的方案。我觉得在大模型的训练上面，可能是在一些这种大厂，或者是有一些雄厚资金实力的能够做这种通用的智能底座

4 但通用大模型它不是万能的，它不可能是在任何的领域都能够实现。但甚至创业公司都是没有机会，没有财力或者是实力去训练这样一个大的模型，又拿不到相关的一些这种行业的数据。所以我觉得未来机会可能还是在有行业领域优势跟数据优势的一些公司里头

¶ 结合公司情况，能否展望未来人工智能在不同场景下的商业模式情况？

1 我们公司经过十几年的发展，主要有三个业务方向，一是政府行业，二是媒体行业，三是金融业制造业等大企业客户。我们公司今年发布了 8 款 SAAS 产品，其中都包含人工智能技术的应用

2 在政府业务方面，我们主要覆盖政府集约化网站的建设，后端政策解读、便民服务等业务是我们后续投入的重点方向。我们原先主要负责一些大政府，之后会逐步向小政府下沉，这对于拓尔思是一个很大的增量机会

3 对于政府政策解读方面，中央人民政府官网是由我们公司搭建的，我们可以结合自然语言处理技术，对语义进行深度理解，向相关人员进行输出。对于写作辅助支持方面，我们是以 SAAS 服务的模式进行收费

4 在媒体业务方面，目前我们的机器辅助写作服务已经在浙报、长江日报、经济日报等媒体已经开始使用了，同时这一块业务还有很大的推广空间。我们还为全国媒体行业提供整个业务流程的人工智能与大数据平台，这一块业务我们的市场占比很高

a 在金融领域，我们有一款产品叫产业大脑，可以对研究报告、合同评审等企业数据进行整合

b 我们未来还会在养老机器人、虚拟人、法律等行业进行尝试探索。目前大数据及人工智能的赛道前景十分广阔，我们公司有着大量的发展空间

¶ 我们为浙报之类的一些企业提供类似 ChatGPT 的 SaaS 服务，一般收费情况如何？

1 对于浙报、安徽报业集团、长江日报等省级报业传媒，我们的业务规模在 500 万-1,000 万之间，平均每一年一家企业支付我们 100 万-200 万左右的服务费，其服务内容包括海量数据推送、自动辅助写作等模块

2 对于政府行业，比如南京市人民政府政策研究室，这种地市级政府一年支付我们 20 万服务费；县一级的政府一年支付我们大概 5 万的服务费；县下面的科局委办厅的服务费一年在 2 万左右

¶ 请问搜索引擎不进行国产替代，有出现什么风险？

1 搜索引擎与打印机有类似的一面，也会存在数据泄露的风险。搜索引擎除了百度谷歌这种大搜外，还包括企业搜索，比如单位 ERP 管理系统的搜索功能。搜索引擎本身是非结构化数据库的展示方式，与数据关系紧密，因此安全风险问题显著

2 在 2010 年之前，国内做搜索的团队也有很多，但在美国建立软件开发联盟，ELASTICSEARCH、SPARK 等开源公司出现后，国内很多小厂商就退出了。目前国内

包括阿里云、腾讯云、百度云、天翼云等云平台都在使用 ELASTICSEARCH 设计

3 在大安可事件出现前，大家都无法想象搜索引擎国产替代的可能性。拓尔思多年以来一直使用自己的海飞数据库，并在信用中国、国家质检总局、监督总局等系统中用自己的搜索引擎实现了国产替代，相信未来还会有很多替换的机会

¶ 我理解搜索引擎应该是触达的搜索人越多，对被搜索方越好。那么对于被搜索方，他们愿意为了安可去将搜索引擎替换成小众的吗?

1 我们不是要去替换百度谷歌这种大搜，我们指的是对企业级搜索进行替换，即企业内部业务系统的搜索引擎

2 对于企业搜索，并不是搜索人越多越好，并且其安全隐患更大。因此企业搜索的国产替代不存在企业方不愿意的情况

¶ 搜索引擎替换是不是指在 OA 系统中嵌入我们的搜索引擎?

1 不是，OA 系统虽然也包含一些搜索功能，但只是搜索引擎应用中很小的一部分。我们谈得更多是内部的管理系统，比如国家电网的系统

2 我们将企业系统中的搜索引擎进行替代，代码自主可控，同时能够保持原来搜索的接口和规则，因此替换效率很高

¶ 目前英伟达 A100 已经限售，请问我们如果要做大模型训练，有什么替换方案吗?

1 实际上很多云平台都可以提供范例支撑，本身我们也有自己的一批服务器计算资源

2 因此计算方面不算一个阻碍。目前各家云平台都有在使用

¶ 展望未来，大模型还会有哪些突破可以期待?

1 在文本能力方面，目前大模型文本能力水平上仍只能辅助人类进行写作，仍需要人类进行把关。在未来大模型可能能够达到接近人类甚至超越人类的水平

2 在多模态方面，未来大模型能够以文本图片视频等多种方式进行无缝多模态输出。同时，目前对于大模型知识库的增删改工作很困难，但在未来这一块还是有想象空间的

¶ 能否介绍一下目前 AI 行业内主要的竞争对手的情况? 竞争格局?

1 一些友商可能号称在某些比赛上能够达到一些性能，但这些性能其实是刷题训练出来的，不具有适用性

2 在这个行业其实是需要一些深耕积累的，包括客户关系，包括对整个行业的理解，包括在训练调优上经验的积累，这不是能一蹴而就的。目前 OPENAI 所能达到的程度，国内国外很多大厂都没有办法做到

3 对于公文写作方面，目前很多公司都在做，其实整个文本的合成这件事并不难，

但这个文本到底好不好用，是否需要经过多次的修改，这中间还是有很大差距的。从源头来看，中央人民政府官网的数据库就是由我们公司来提供的

4 其中自然语言处理的分词技术以及多年数据积累，使得我们公司与其他友商比较有着明显的优势

5 同时对于新闻单位来说，我们凭借着多年与人民日报、新华社、中国广播电视总台等诸多媒体的合作所积累的经验，也有着比较强劲的竞争优势

¶ 我们与中译语通有业务交叉吗？公司与科大讯飞比较？

1 没有交叉关系

2 我们与科大讯飞从来没有产生过竞争关系。虽然我们都在讲人工智能、自然语言处理，但科大讯飞主要的业务方向在教育、医疗、智慧城市，而这些行业不在我们公司业务的覆盖范围，我们始终围绕着自己熟悉的垂直应用场景去扎扎实实的打造

¶ 最后补充：

1 拓尔思作为一家 11 年上市的技术型公司，一直以来业务成长不算快，这也是很大投资人比较疑惑的一点。我们原先增长速度不快，并不是赛道的问题，而是技术积累的问题，并且我们公司规模较小，销售队伍覆盖不够广

2 从 2021 年开始，我们公司开始进行战略的转变，从大 B 到小 B，从大 G 到小 G，最后到 C 端。我们在 C 端虽然目前还没有直接的收入，但已经做了很多有益的尝试

拓尔思 ChatGPT 市场化展望

访谈日期：2023/1/29

具体内容

¶ ChatGPT 成功出圈，AI 发展迎来里程碑

1 CHATGPT 采用了一个新的算法，其次采用了非常强的算力，使用了大量的数据，从 AI 发展的三个大的方向上进行了拓展。从而更好的理解人类的指令，提供更接近预期的答案，CHATGPT 的出圈标志着 AI 的新发展

2 CHATGPT 知识图谱以 NLP 为核心技术，旨在将多模态信息数据进行挖掘后以知识形态进行存储，赋予 AI 认知理解和自主推理决策的功能，被认为是实现强人工智能的关键技术

3 公司人工智能产品结合大数据技术支撑，在金融、公安、政府等多领域应用前景广阔，伴随 AI 商业化进程的加速，未来相关业务有望实现加速成长

4 公司保守的估计未来的几年整个的战略规划是保证每一年 30%以上的增长，主要逻辑是公司业务场景持续延伸，产品 SAAS 化进程推动业务成长，NLP 及知识图谱技术在虚拟人、人形机器人领域应用前景广阔，公司积极布局技术研发，与国内外头部厂商形成深度合作

5 未来有望带来业务新增量;公司持续推进产品 SAAS 化转型，近年来云和数据服务收入占比持续升高，驱动公司产品盈利能力的不断提升

¶ ChatGPT 简介环节

1 CHATGPT 是一个 AI 的对话模型，于去年刚刚发布，CHATGPT 并非一个全新的模型，而是 OPENAI 迭代过程的产物。2018 年，OPENAI 发布了 GPT1,2020 年发布了 GPT3。当前的 CHATGPT 是基于 GD 的 3.5 的一个系列，由模型微调而来，支持编小说、写代码、回答问题等，能够协助做一些文字类的相关工作

2 CHATGPT 之所以表现优秀，与之前的 AI 不同的地方，主要体现在 AI 的三点要素

- a 第一、算法，算法的核心有两个：一个是指定微调，一个是基于人类反馈的强化学习（Reinforcement Learning from Human Feedback
  - i 这两种技术并非 OpenAI 独创，并且早已经实际使用
  - ii 之所以如此出圈的原因在于，ChatGPT 通过交互的方式提供了带有上下文推理的能力，解决了部分安全性和准确性的问题，能够提供近乎真人一样的理解能力，鲁棒性好，所以能够出圈，并被公众了解
- b 第二、数据，在具备一定数据量后，模型有 1,000 多亿个参数，4G 的语料，模型的能力就凸显出来，如：在如此大数据量下，上下文推理能力就体现出来了
- c 第三、算力，ChatGPT 的算力非常高，需要的算力也非常高，GPT3.5 在超算的基础设施上训练，总算力大概消耗 3,640pf/day，成本非常高
- d 总的来说，ChatGPT 之所以表现这么好，说它是一个划时代的产品，主要是用了一些先进的算法，能够理解人类的指令，同时使用大量数据和大量算力，从

而有了体验的飞跃

¶ ChatGPT 在使用上或者其他企业后续追赶上是否存在其他差异?

1 首先，在训练方法上，通过指定微调，CHATGPT 能够更好的理解人类指令，在 GPT3.5 的基础上，加入了人类标注的数据指定微调，从而能够去做一些类似翻译的工作，理解人类指令

2 其次，基于人类反馈的一个强化学习，使得 CHATGPT 能够提供更接近人类理解的答案，如：通过强化学习，模型生成的多个答案由人类排序，人类再选定第几个答案是更符合预期的，因此通过这种方式模型生成的答案更接近人类的期望

3 总的来说，相较于其他 AI 产品，CHATGPT 能够理解人类的指令，并提供接近人类预期的一个答案，通过交互的方式实现这些能力，从而出圈

¶ ChatGPT 未来产品化的方向和空间?

1 CHATGPT 作为文本生成的技术，未来是一个方向，随着一些场景的落地，文本可以生成到不同的领域中。比如生成代码、生成文本、文案创作（新闻写作、研报生成、报告周报等等），以一种新的方式生成类似这种文案

2 目前的水平仅能实现辅助创作，相当于辅助人类生成模板，由 CHATGPT 生成几个 IDEA，然后人类去做筛选、润色。在未来的发展过程中，CHATGPT 可能会生成更好的质量，甚至能够超过人类的水平

3 CHATGPT 还可能往多模态方向发展，例如：通过文本、图片融合的方式，生成图文相关的，或者可以直接生成视频等

¶ 目前，AI 的回答存在词不达意或者胡编乱造的情况，那么现在利用哪些技术能够保障 AI 的回答更具准确率?

1 现在确实 CHATGPT 的问题在于生成的答案可能是不正确的，虽然大部分是一个比较正确的答案，但是还是会出现回复的内容它很确定，但是回复的答案是错误的

2 后续的解决方案是做一些支持融合的方式，在一些大模型上去做知识的存储、知识的修改，比如：当问题出现错误，并不能修改大模型，因为大模型的训练周期非常长，重新训练需要非常久和非常高的成本。因此，怎么在大模型中去做知识的存储和修改是现在研究的方向

3 现在还有知识融入的方式，在一些领域中，可以再为模型增加一些垂域的数据，通过学习新的知识，更加了解这个领域的东西，使得在这个领域有更好的表现

¶ 模型最初建立的时候，需要的人力的数量和人力的素质大概是一个什么样子的情况?

1 训练一个 CHATGPT 是一个非常庞大的事情，以数据为例，CHATGPT 相当于 GTP3.5 水平，GTP3.5 它已经用了大概 3,000 亿个 TOKEN，训练一个成本大概就是几千万美金的一个水平

2 所以后续的话，我觉得 OPENAI 还是会像以前通过 API 的形式提供出来，相当于 OPENAI 提供了一个基础的接口，基于大模型的接口，在上面做一些微调

3 比如：可以提供一些垂域的给模型，去做一些场景训练，用比较少量的数据，能够得到一个领域中比较好的表现

¶ 国内除了拓尔思以外，还有哪些企业可以做相关的技术？如果要满足技术，尤其向市场化或者向产品化发展，还需要满足哪些条件？

1 国内的一些大厂也是在尝试类似的事情，在做一些底层类的工作，像提供一些大模型的能力，然后其他厂商可能不太清楚

2 我讲一下拓尔思目前的布局，拓尔思在这一块主要分成三层，底层是模型层，叫“智拓人工智能平台”，经过多年的积累，文本模型和视觉模型都在这个平台中

3 中间层是能力层，有两个平台，一个是智语，一个是智眼，一个是文本类的，一个是图片视频类的，能力层主要与能力相关，比如：智能增强、语义理解、基于模仿创作、基于概念创作等

4 最上层是 AIGC 的自创平台，现在提供了一些模块化的东西，功能层比如提供了文本续写、标题生成、文案生成、风格改写等

5 就目前来讲，一些资金雄厚的大厂可能在做大模型的训练，或者做一些偏底层的工作。拓尔思的优势是数据的积累，从 93 年成立，已经有超过 1 万多家的客户让拓尔思对整个行业有所了解，对整个垂直场景有很深的理解

6 其次，就是模型能力的积累，拓尔思做了很多的项目，还跟很多客户打过交道，累积了很多能力。还有可能在算力这一块，拓尔思对一些垂域有更深的理解，能够做很多场景的落地

¶ 拓尔思目前在 AI 写作业务上，有没有对应的产品，或者客服等？

1 拓尔思在 AI 写作业务上已经做了很多的项目，比如：给浙报做的新闻的自动写作、赛事播报，公文写作等；给邮储银行做机器学习平台的智能文本写作等。拓尔思在智能写作业务中成熟的产品有小思智能创作，横向的有很多项目可以去做，也有一些智能创作的方案落地

2 拓尔思在整个 NLP 业务上可以分成几块，金融创作也是其中之一，拓尔思的优势在于已经做了很多方式的落地工作

¶ 从现在的技术水平来看，如果要达成理想中的与 AI 进行顺畅的交流，或者是让 AI 去进行像年报或者专业咨询类文件的撰写，目前还要发展多少年？或者要发展到理想中的程度，所需要挑战的难点还需要去做哪些技术储备，数据处理或者数据积累？

1 目前的 AI 还是一个辅助创作的水平，相当于能够给人类加速，提高效率。目前来讲 AI 能够实现文章架构的生成，然后由人类填充内容、审核，最后得到一篇与人类合作的文本作品

2 在未来 5~10 年，AI 生成的内容有可能超过专业的写作的水平，目前来讲还是在慢慢发展的一个过程，可能要解决的问题还是在于准确性，不能编造事实，需要按

照事实撰写内容

3 另外就是需要能够减免一些错误事实类，以及知识能够更新，大模型中需要知能够实时的更新

4 然后目前它的形态也比较单一，它的模态也比较单一，现在只能生成文本内容，后面可能将图片、视频、音频等多模态集成到大模型中，生成更多模态的东西，不仅是文本创作，还可以视频创作等

5 然后还有一个就是解决高成本的问题，怎样以更低的成本把 AI 在垂直领域落地，也是要给解决的问题，随着成本的下降，能够应用到更多领域，前期成本比较高的时候，可能是在一些高价值的领域去使用，成本下降后可能会应用到更广泛的领域中

¶ 拓尔思做自然语音识别做这么多年，目前的水平如何？相较于 OpenAI 的水平，还需要多少年能达到？而且相较于国内主要竞争对手，拓尔思有没有跟上？差距是缩小了还是扩大了？

1 大家也担心云计算或者 AI，最终是属于大厂

2 拓尔思 93 年成立，以自然语言处理和非结构化数据作为研发方向，前些年的发展注重数据积累以及算法，对所有的应用场景，也有选择的做了这些工作

3 相较于大厂，在整个自然语言处理一个最实际的应用就是海贝数据库，在搜索引擎数据库上，拓尔思跟百度选择了两个不同的商业模式，拓尔思一直围绕企业搜索做文章，所以拓尔思在垂直领域的理解和应用优于大厂

4 在自动写作业务上，拓尔思很大板块针对于全国的媒体单位，有数十个单位都在用拓尔思的机器写作，并且每天以云平台的方式然后进行推送。如：现在媒体重点宣传的主题，公司先用机器的话合成，再修改，大大的节省人力成本。因此，媒体业务上，拓尔思今年可能持续增加投资

5 还有拓尔思对政策的理解和解读，中央人民政府的政策库由拓尔思负责，各级政府部门每天发布的上千条的政策，通过大量的知识图谱进行关联分析，最后再辅助做公文写作。全国现在有 400 万公文写作人员，这是一个很大的需求市场。目前，在南京、成都已经有实际应用落地

6 公司在机器政务应答这块也有比较多的应用，比如和中医科学院，人民出版社、吉林政务回答的合作中都做出过比较好的结果。拓尔思就整体来说，对于熟悉的领域精耕细作，在不同的场景中，要扩大应用拓尔思会逐步积累知识

¶ 请问这一块的盈利前景如何？或者盈利量级？

1 拓尔思 2021 年净利润率 24%，2022 年整体水平还在增长，公司的收入结构包括：软件平台和 SAAS 服务。软件平台主要销售：人工智能平台软件、大数据平台软件和数据安全

2 大数据平台软件涵盖了整个数据要素，从数据的获取、到数据标注、分析、展现，整个全过程都做成了数据中台

3 在人工智能方面，拓尔思也都把产品做成了软件平台。从整个发展趋势来看，拓尔思对人工智能平台非常关注并不断在做预算

4 拓尔思净利润率比普通软件公司和系统集成公司高的原因在于：出售产品后和持续提供服务，就是算法迭代，背后需要不断数据的更新，从而与用户建立了优良的联合度。另外，在新闻单位的自动写作业务上，拓尔思有 2,000 多台服务器每天收集全球新闻数据，用数据做训练，再展现

5 拓尔思也在探索养老机器人业务，并一直积累知识

6 拓尔思与大厂最大的区别在于，拓尔思在这些细致领域的积累更加扎实深入

¶ 大厂的模式是什么样的，与拓尔思的区别是哪里？另外大厂为什么不做这块业务？是觉得这块空间不大还是什么原因？另外您说在新闻自动写作这一块，已经有落地的这些场景，这一块我们大概有多少个客户？做了多少个客户？客户的收费情况？客户的实际应用情况如何？

1 国内的大厂在很多垂直领域，对于整个知识的积累，包括知识图谱以及能够持续产生收入方面，目前还没有特别明显优势，大厂对于大模型和大算法的研究投入比拓尔思多得多。拓尔思多年以来，在大数据人工智能方面，已经深入面对用户群体，并取得突出的业绩

2 在媒体业务，公司已与浙报、经济日报、航空工业报、中国教育报、安徽新闻媒体集团、武汉长江日报等取得合作。均采用打包服务，即：一次性的采购拓尔思 800 万到 1,000 万的软件平台，之后每年继续付给拓尔思 200 万到 300 万不等的 SAAS 服务费

3 在自动写作业务上，主要通过机器自动撰稿而形成主题稿件进行推送。从全国融媒体改革，包括新闻媒体持续增加投入，这一类业务目前处于上升趋势

4 同时，结合政策大脑，拓尔思能够向全国 400 多万的公文写作人员提供辅助写作服务，从而减少工作量。类似的应用场景还有很多，拓尔思整体上围绕着几个板块去挖掘用户的使用场景，使得用户的粘合度高

5 未来拓尔思业务的发展和业绩的增长，将主要来源于

a 采用大量渠道销售，发展行业渠道
b 在垂直的应用场景，深入打磨人工智能和大数据产品

¶ 拓尔思的业务落地场景背后的模型也是一个大模型吗？这个模型与 OpenAI 有什么区别？也是经过海量数据训练的吗？还是一个专注于小场景的模型就能够用了？

1 OPENAI 是一个非常海量的模型，目前拓尔思的模型尚没有如此规模

2 但是拓尔思在不同行业中有更好表现力的相关模型和数据的积累

3 拓尔思的优势在于，虽然没有 CHATGPT 如此海量的数据，但是在各个领域中拓尔思训练相关模型，每个领域中有上百亿的小模型，能够达到和大模型接近的水平。其次，拓尔思基于中小模型对接垂域数据，通过模型训练或者是基于 OPENAI 这种大的模型，再接上协议数据去做场景化的落地工作

¶ 拓尔思未来的业务增长怎么看？包括收入规模、利润等？

1 拓尔思目前整个业务发展相对稳健，公司成立以来，一直专注着围绕着人工智能

和大数据这一块的话在精耕细作。目前公司的业务主要围绕 4 个行业，现在正在有序拓展以国家电网为主的能源行业，以及其他行业的深度应用

2 公司未来主要业务结构会发生变化，整个 SAAS 服务的占比在 2021 年是 36.84%，2022 年增长 1%左右。拓尔思会持续的保持云和 SAAS 服务的增长，最终比较好的商业模式是以 SAAS 云服务为主

3 由于模型需要不断迭代，拓尔思将不断的增加云和 SAAS 的服务比例，同时扩大规模量，公司未来营销目标是每年保持 30%以上稳健的发展。从目前来看，市场所接收到的信息和拓尔思的营销，人工智能和大数据的需求非常旺盛

¶ 在营收增长的背后，拓尔思在营销队伍、研发等方面需要有更多投入吗？

1 拓尔思的研发一直保持高位，公司以技术驱动为主，目前全国有重庆和成都、北京、上海、广东四大研发中心

2 随着人工智能在各个场景的应用，拓尔思持续多年一直在做自己的知识产权、大模型平台、自主软件，并已经有了相当的积累，接下来将着重完善场景的应用。所以从研发角度，拓尔思不会像前期海量去投入做全新的东西，将着重针对各垂域，打造知识库和知识图谱

3 在营销方面，拓尔思不再海量发展大客户销售代表，而是采用围绕着每个区域和垂直板块去发展渠道，通过渠道分享人工智能和大数据发展过程带来的高收益

¶ 23 年及以后人员扩张，销售人员、研发人员扩张上有什么规划吗？

1 拓尔思前两年的人员略有减少，接下来拓尔思的研发还会继续增加，但是公司总体人数会适量控制。对于销售，拓尔思目前有 200 多位销售人员是大客户销售为主

2 在保持规模的情况下，将通过行业渠道，拓展的话，3 万、5 万或者 10 万以上的客户群体。总的来说，拓尔思目标把营销和业务规模做上去，但是适度控制人员扩张

¶ 刚才提到客户有三类，一类是国家安全，一类是媒体，一类是政策公文。想具体问一下，第一类国家安全和第三类政策公文有什么区别？

1 政策公文方面，各级政府部门，以及央国企大量的公文写作人员，他们实际上工作的话是很繁琐，任务是很重的，这是我们的整个的管理体制和我们的话整个国家的话讲政治的话，特殊的就说这种环境然后所带来的一个市场

2 但是在这一块服务的过程中间，以前的话拓尔思已经的话开始覆盖市场，帮他们提供的话政策库的解读，然后的话讲的叫政务咨询平台，这一块譬如说我们在南京市人民政府的秘书是他们的政策研究室，一年的话给我们 20 万，我们直接向他推送各种各样的这种政策解读，那么附带着就说又产生了一个新的需求

3 他们的话对于很多这种公文写作，特别是的话，就是说对于这种政策性比较强的需要的话，通过自动写作，然后的话来提高他们的工作效率。所以这一块的需求的话不仅仅是面对政府人员，我们讲的是面临着央国企和体制内的所有的单位，包括医院学校也都有这样一个群体的产生

4 那么拓尔思现在我们推出来的就这种自动写作的产品，不是说给他一套软件，而是的话因为每天的话就是说所有的政策在变动，政策库的话在更新，而我们是提供的这种云服务的方式，针对于每一个小的单位，你比如说线下面的科技局的委办基金，那么的话每一年就是 2 万块钱左右的服务费，然后我们直接让他推送，主要讲的是这一个群体的服务

5 那么刚才我们讲的对于国家安全这一块相关的，包括公安国安和军队，主要的话还是我们公司这种开源情报的收集和整理的能力，这个背后的话也包含了一些自动写作，然后自动合成相关的内容

¶ 请问这几块业务未来的空间，以及目前一个市占率的情况

1 目前我们的话碰到的规模跟我们相当的竞争对手还比较少，然后在国内的话也有大量的这种中小团队，然后再提供区域性的这种服务

2 但是譬如说我刚才讲到公文写作和政策解读，首先的话是从上往下，因为拓尔思的话有一个大的优势，就是中央人民政府官网的库是我们做的，官网的核心的这些功能模块也是我们开发的，所以的话我们向上下垂直的时候有几大优势

- a 第一个的话就是从上到下的解读和我们政策获取的能力，还有一个就是多年的我们自然语言护理的技术的积累，我们的对整个的话室内的分解解读，还有一个本身拓尔思的话有一个海贝数据库
- b 搜索引擎这一块实际上在我们的宣传材料上已经介绍的挺多了，这也是未来新创的话，我们公司业绩值得期待的一个很大的重点
- c 第二个话就是说我们的自动写作也好，或者说包括刚才讲的自动应答也好，在其他的这种企业内的这种客户，我们也积累了大量的客户群体，譬如说金融主要是针对于这些大的银行，那么的话我们做了一些智能客服，这个为客户省了一大笔钱

¶ 每个客户销售、研发和售后服务，大概的比例关系？

1 公司现在整体上服务的客户群体已经超过了万家，但是我们讲的万家就是说确实的话我们都是围绕着大政府，然后大的国企，还有大的新闻媒体单位

2 说我们除了软件平台之外，有 30%就是说 36.84~40%的业务收入是来自于 SAAS 的服务，主要是我们的数据中心然后推送出去的，所以的话真正用户的上门的服务是很少的

3 我们都是以区域的这种大客户，然后每个区域办事处针对政府有一个客户代表，针对于企业有一个客户代表，每一个客户代表服务 30 家这种长期以来跟我们建立过服务联系的这些客户

4 但是我们拓尔思如果说从 10 个亿做到 20 个亿，做到 30 个亿做到 50 个亿的时候，我们不会再去扩充我们这些销售代表了，因为围绕着每一个垂直的领域，都有一些对垂直领域知识比较健全或者知识库比较丰富的这些企业，能够成为我们的渠道销售，大家共同去完成这样的销售任务

¶ 公安、政府和私域客户数据不能反向获取，那么拓尔思现在相当于是私有化部署吗？

1 公安数据对您刚才讲的很对，是这样子的，我们的话针对于刚才您提到的这些用户，包括的话，像我们服务的金融用户也好，或者其他的央国企也好，我们都不会出用户的数据，那么的话他会买我们的软件，就是我们的大数据和人工智能的软件平台，就是我们讲的数据经中台，然后跟他们的应用系统结合

2 所以拓尔思的整个的项目还有一个特点，我们跟很多的系统集成公司和软件开发公司比较，你像我们所知道的国内的一些系统集成公司或者软件开发公司，获取一个几百万上千万的中标项目以后，会要派驻人员到用户现场干，上半年到一年的时间去做交付

3 而我们公司的很多的所谓的项目合同，交付时间就是2~个月，有的甚至一个月，因为我们做的是人工智能和大数据的中台，这是有本质的区别的，我们不是做的业务系统

4 那么还有一个的话就是说我们在整个的话就是说交付完了以后，我们持续的数据推送的话，是采用这种云服务的方式。刚才我们提到的公安国安和军队这些比较敏感的单位，他们会充分的享受我们的数据服务，我们讲的是开源情报服务

拓尔思 ChatGPT 相关

访谈日期：2023/1/9

具体内容

¶ 拓尔思公司发展的背景和场景应用？

1 公司曾于 1993 年成立是全球的中文检索的创始者，公司成立之初，主要的技术方向围绕着中文全文检索，第一批用户主要是媒体用户。1997 年获得过国家科技进步二等奖，电子部科技进步一等奖。2000 年到 2011 年阶段，公司核心技术是基于检索技术，开发智能内容的管理

2 2007 年我们启动了核高机的非结构化数据系统的研究专项。拓尔思是作为第一家大数据公司上市 A 股，公司持续在自然语言处理技术上做研究。语音智能是公司的核心技术的发展场景。NLP 自然语言处理的技术在各个场景中的应用上，我们不断进行深入的拓展

3 拓尔思公司所有的人工智能应用来自于公司对各种算法模型的积累。在 A 股市场横向比较，在诸多的公司里，我们真正掌握大量数据资产。2,000 多台服务器分布在全国的三个数据中心，每天日增 1 亿条的开源的互联网的数据，公司已经积累了将近 1,300 亿条的开源的数据资产

4 公司已经积累了 300 种以上的算法，并且对每个场景，如知识图谱的展现、知识库的建立档案、包括前期数据采集、数据的标引，关于数据要素的环节，我们都有自己以完全知识产权的软件平台。搜索引擎是公司自然语言处理的核心应用技术，公司 30 年以来坚持这方面的积累

5 在全国大量企业级的搜索都在用 ELSG 的设计 SPARK 开源软件的时候，公司完全做到了自主可控，应用到政府、金融，包括媒体等诸多行业

6 公司数字经济研究院目前主要研究方向是人机对话，公司近几年在智能问答，围绕着像中国中医科学院的中医中文问答，中国标准化研究院的国家标准的问答，人民卫星出版社的小 A 机器人，时代经济出版社的审计问答、吉林政务的小机智能机器人等

¶ ChatGPT 提升的原因？

1 加入了人类的反馈系统，第一步人类做的方案模型进行微调，得到模型。第二步模型根据问题生成答案，训练出奖励模型，这个奖励模型给第三步打分，相当于输入奖励模型，得到分数。优化，不断的迭代。目前的是问题它是非实时模型，离线模型，它获取到的知识是 21 年

   a 21 年以后的数据就不知道了。因此无法保证结果的可靠性，他会编造事实，一本正经地胡说八道。而且只能返回文本的信息。缺乏对行业数据的积累。它只是通用模型。并且训练成本过高，ChatGPT 训练的大概的预估成本在 1,200 万美金以上。它的运行成本也很高

   b 首先拓尔思有着来自境内外的各行各业的数据市场，超过 1,200 个亿，已经具备千亿数据索引等。这些是我们的核心资产，在大数据量的前提下，我们能够

大力出奇，足够多的数据，模型有非常好的表现

c 其次拓尔思技术的沉淀，坚持核心自主的研发，实现国产化，拥有 40+发明专利，800 的软件的着作权利。另一个是客户的沉淀，整个数据的产品和服务已经被国内外超过 1 万家的企业级的用户在广泛的使用

d 智能客服基本是基于检索式，基于数据库，我们有深度模型，去库里检索答案，返回给用户。ChatGPT 的思想是基于大模型，我们有排量数据去训练模型出来，再加入人类反馈的数据，就能够提供更优质的对话体验

2 第二点，我们需要行业深耕，CHATGPT 是一个通用模型，缺乏对行业客户、行业知识的了解，而我们对行业是非常了解的

a 我们未来会让对话式的 AI 等这种人工智能技术跟行业客户的业务流程更深度融合，从局部业务到全场景的覆盖，实现全业务的数字化、智能化。我们会持续的在行业中不断的累加场景，深耕场景，解决核心业务的

b 从长远来看，拥有更好的数据，我们更有利于微调大模型，这样公司可以创造出一条可持续护城河

¶ ChatGPT 可以对自己不懂的内容胡编滥造，那么目前技术发展路径如何保证 AI 回答模式的准确率呢?

1 目前整个智能客服处在比较成熟的阶段，所采用的技术基于线索式，它保证了所有的回复都是从库里去拿出来回复给用户

2 CHATGPT 基于生成式的回答给用户，它比较难保证回复的准确性。我们在后续的训练跟维护的过程增加规则和安全检测的模块进到系统，能够保证异常条件下规避掉这些问题

¶ 从公司的视角以及包括整个产业发展趋势来看，4 个场景哪一块是最先有可能形成商业化的落地?

1 一是专家咨询类的，实际上相当于是企业大脑的角色，需要把我这些行业的知识变成企业的大脑，变成模型的知识，CHATGPT 证明了在一些大模型有比较好的表现

2 二是在智能创作，助手类的，公司能够去高效地提升智能创作的水平，CHATGPT 的一些文本生成已经能够满足创作者它的大部分的需求，能够去帮助创作者生成初级的版本，一些创作者在上面再去继续修改，在直播文案的生成、广告文案的生成、基本创作等有比较好的效率提升

3 三是在交流类的，CHATGPT 拥有比较大的模型，拥有比较好的世界知识，通用知识的前提下，它能够回答各类问题，说明如果公司比如在元宇宙或者养老领域里去做定制，可能也会有比较好的表现

¶ 公司是如何确保采集的数据是针对相关的行业，而并不是会跨到其他行业，因为其实现在有很多的名词，其实同时代表不同的行业的内涵。公司是怎么确保算法以及数据的针对性，是匹配到行业的?

1 这其实是模型上下文关联的能力

2 实际上是大模型是能够学习到相关的上下文的知识的，比如拿法律的整个行业的数据进来训练出大模型的结构，再基于人类的一些反馈加入训练，最后出来的模型，它会在不同的条件下识别到不同的上下文的知识的

¶ 现在公司最大的痛点是在哪里？或者公司后续会在哪个行业率先落地相关的商业模式，并能产生实际的收益？

1 实际上公司觉得训练的方法和整个技术原理实际上都比较清晰。接下来首先就是语义智能，它本身是经验型的，技术的积累首先还是来自于你所熟悉的行业，人工智能的场景的应用，要选择比较好的主题。在选择主题以后，作为公司在深度的知识的积累，最后结合语义智能

2 后面通过训练数据源源不断的进来，训练的整个的模型，整个的算法积累的就会越来越丰富。所以公司觉得经验值是非常重要的

3 公司这几年以来在整体的打包服务中，有托尔斯的妙笔小思的智能写作实际上就是合成，但是需要公司了解整个编辑记者他们在应用场景中间，先不断丰富积累他们的新闻要素，新闻稿件的形成的细节

4 另外，融媒体中心成立以后，出稿子的频率越来越快，任务越来越多，越来越大的情况下，需要能够快速高效的去完成工作

1. 公司实际上强调的还是对行业深入了解和熟悉的程度。譬如刚才您提到的法律，公司的背后正在通过跟律所合作打造公司的法律的知识库，因为整个法律咨询不可能出现万能的，什么样的法医知识都懂的，可能围绕住房纠纷或者刑事案件，背后有一系列知识库的间接
2. 公司有自己的知识图谱的研究院，多年以来在开源情报这方面持续实现了一定比例的收获，并且还有很好的增长趋势，基于对整个的开源情报的分析，各种各样的数据的采集加工，不断迭代，形成了公司自己的知识图谱的各种各样的算法

¶ 未来是不是会有可能在每个行业都会诞生出龙头，类似于搜索引擎龙头，未来的趋势应该是有垂直行业为主，还是有大一统的搜索平台为主？

1 刚才公司都提到了共同的问题，现在公司关注的事件，实际上会发现它现在整个积累的时间和计算的时间，尽管跟国内的公司比较，已经有了数量级的差异，但是它不能够穷尽一切，理论上来讲，它能够穷尽一切以后它就真正能够替代人了

2 现在公司在探讨它的应用的同时，反过来反思公司国内有哪些应用场景，从这两方来讲，公司认为每一个垂直的专业板块空间都是非常大的，公司在整个人工智能和大数据的中间软件，已经达到了比较强大的自主可控的软件平台的积累

3 但是对于每个垂直行业的深度的应用，也不是什么行业都去干，这样小规模的上市公司也承担不起

4 但是譬如像知识服务用在专利检索，用在整个专利行业，它未来是百亿级的规模，大家需要有更多的服务的时候，公司就把更多的给打造好，围绕着金融，围绕着媒体，围绕着这几个深度的行业去做就好

5 公司还有可以拓展的行业，现在结合虚拟人和机器人走，悟到更多新的应用，也在拓展公司的新的市场，比如在机器人，现在围绕着养老院场景的精力是最多的，一旦投入进来，公司就能够比别人积累更多的支持

¶ 目前公司在国内还有其他的竞争对手吗？目前他们的进展如何？

1. 实际上整个自然语言处理在行业内的应用，大家感知的比较多的是智能客服。智能客服有很多公司围绕着不同的客户平台在提供这样的服务。在电商行业，政府，还有其他的需求比较强烈的这些企业都有相关的公司在做类似的工作
2. 还有就是舆情分析，在各个地方也都有大大小小的公司。实际上整个知识库的沉淀，它的背后是通过语义智能，把知识关联起来以后，通过整个知识图谱的知识库，最后开始做各种各样的训练模型分析
3. 目前来看国内在合同对比，智能内容处理，包括数字人、虚拟人，都有很多公司，但是比较而言，能够持续的对算法进行深入研究的，背后必然需要海量的数据。数据的积累除了不断的获取数据之外，还有承接的历史数据也是很重要的
4. 在这一块公司的优势在于
    a. 从 2000 年开始就在持续的通过海量的互联网数据，就是开源数据，不断的在沉淀，在积累，打造了多个知识库
    b. 公司的研发团队从 93 年以来就围绕着搜索、自然语言处理、语音智能积累，所以公司承建的各种大数据应用平台和人工智能的应用平台，相对比较丰富的

¶ 公司各项业务未来的收入增速会是怎么样的情况？

1. 公司保守的估计未来的几年整个的战略规划是保证每一年 30%以上的增长。公司现在的收入是来自于两部分，一部分是人工智能和大数据的各种各样的平台，面对金融、政府、媒体，是以项目的方式来展开的
2. 但是所谓项目跟那种传统的管理信息系统不一样，公司的交付周期相对来说都比较短，交付的这些数据中台，这些项目是要跟它的应用系统结合起来使用的
3. 这方面收入在公司目前的收入结构中，大概占到 60%左右，2021 年是 63%左右。公司还有 63.84%的业务收入是来自于公司的数据服务，也就是把采集到的大量的数据加工成数据产品，最后再输出
4. 公司的战略一直在向 SAAS 化转型，今年跟去年、2022 年跟 2020 年比较，公司整个的数据的收入在持续的增加

¶ ChatGPT 大概从 18 年开始到 22 年就做到了这么大的市值吗，增长非常的快，NLP 下游的应用里面有没有可能重现这样的成功。展望一下行业里除了 ChatGPT 类的应用之外，有没有其他非常有潜力的应用

1. 实际上 CHATGPT 的估值在认真的往上涨，但是现在在美国是不能算销售收入和具体的产生的价值在哪些地方的，它只不过是典型的人工智能的公司
2. 所以公司一直在看模型训练出来完了以后，能不能够替换人去干一些工作，这就是“数字劳动力”的概念。在国内来说，公司现在所接触到的行业和未来公司想拓展的行业，只要在每个行业深入的投入，都会有大量的想象空间，会有大量的应用
3. 对于用户，只要公司能够提供好的产品，能够满足他的需求，他是愿意维持他买单的

¶ 能否细致地拆一下不同业务方向在公司的整体的营收的占比是什么样的情况？以及这些业务未来的增速的展望？

1 目前公司面临的客户主要是金融、媒体和政府。公司实际上在 CHATGPT 接下来要延伸的一些应用场景做了仔细的区分，比如在金融行业，主要就是客服、理财，还有就是现在整个金融行业非常愿意投入的营销和风险管控

2 在媒体和互联网行业，主要就是自动写作，还有自动审稿，还有发布形象代言；政府主要做各种各样的审核检查，便利服务、便企业服务

3 从收集到的信息来看，他们在信息化方面都是非常愿意投入的。这些在整个的业务收入里面，占比可能在 5%到 10%左右，实际上并不高，但是未来公司要增长，这是非常大的空间

4 举个例子，公司在政府主要的业务收入围绕着集约化的网站平台建设，集约化的网站平台建设中很大的一部分是来自于大数据的应用，大数据和人工智能平台和它的业务系统结合

5 但是现在政务大量在做政策解读政策，这又涉及到公司能够垂直打造的产品“政策大脑”，全国的政府的政务部门的主要工作就是制定政策、执行政策，对政策的监督

6 所以浙江省在数字化转型的过程中间，整个浙江省的“政策大脑”，“政务运行大脑”都是公司做的，现在有着极大的向全国推广的价值

科大讯飞投资价值研究分析与行业前景

访谈日期：2023/1/5

具体内容

¶ 重点交流内容：

1 行业：人工智能行业三大底座：目前国内人工智能行业处于高速发展阶段，主要依赖于三大底座：平台、数据和算法，技术已经相对成熟

- a 人工智能行业发展驱动
  - i 资本驱动：过去几年资本争相涌入人工智能领域，包括二级市场和私募市场的衔接，推动了人工智能企业的上市
  - ii 政策驱动：国务院 17 年提出人工智能市场三步走规划，第一步 2020 年总体深入水平与世界先进水平同步；第二步 2025 年理论实现突破，技术和应用达到世界领先水平；第三步 2030 年理论、技术和应用都达到世界先进水平
  - iii 其中还包括疫情背景下加大新基建的建设，政策上驱动了人工智能的快速发展与应用
- b 落地场景：人工智能的应用现在基本应用于各个领域，包括商业、医疗健康和金融等

2 科大讯飞：市场定位错误：市场普遍认为科大讯飞仅仅只是市场智能语音龙头企业，实际以商汤为首的 AI 四小龙和讯飞的技术底层都是相通的，只不过像讯飞早期业务主要集中在语音领域，而 AI 四小龙主要集中在视觉领域。底层技术都为深度学习、机器学习，可替代性强

3 护城河初步形成：AI 行业主要分为技术输出和垂直性应用，目前讯飞相较于传统 AI 技术型企业而言，已经初步形成自己的生态系统，可替代性降。（例如旷视与阿里的合作，被阿里达摩院取代）科大讯飞发展阶段

- a 2014 年以前的技术探索阶段：商业模式 AI 技术输出收费，过程中不断实现 AI 技术创新和稳定
- b 2015 年—2018 年 AI1.0 阶段：市场和技术不断投入，发现于形成 AI 商业的规划于落地
- c 2019 年后 AI2.0 阶段：赛道聚焦，将核心聚焦在变现能力强的商业规划中。方向主要包含：教育、医疗、开放平台、智慧城市、汽车和金融等，AI2.0 阶段也是讯飞技术兑现阶段

4 业务核心主要为：业务模式获取有效客户数据—数据反哺 AI 算法—算法形成个性化服务—产品核心竞争力更强

¶ 人工智能行业现在处于什么期？在中国的落地情况以及未来的发展预期怎么样？科大讯飞在行业中的地位和核心竞争力？

1 国内人工智能行业现在属于理论实现突破这个过程中，该阶段技术和应用都处于高速发展，目前国内在疫情的背景下，新基建的加速落地，推动了人工智能的快

速应用，包括像医疗、教育和智慧城市等都处于快速发展和应用阶段

2 科大讯飞目前在行业中除了智能语音领域处于行业龙头外，其他技术能在技术原理上实现应用水平

3 但是讯飞核心护城河并非是技术，而是其商业模式能让其 AI 技术垂直化应用，并且在这个过程中通过数据反哺算法，时刻在核心产品上在行业内拥有极高的竞争力

¶ 10 月份以来，中央密集出台的支持教育、医疗新基建、数字改造的政策，对科大讯飞未来的营收产生多大的影响？

1 教育新基建：讯飞业务模式

- a 教（智慧课堂）：利用人工智能技术及后台教育资源实现课堂场景下的师生教学互动，付费群体：学校
- b 学（智慧学习）：收集学校日常考试、测验、作业过程化数据，形成学业大数据、帮助老师精准教学，形成个性化学习源。付费群体：学生家长
- c 考（智慧考试）：服务高利害性考试，如中考、高考、普通话考试和英语听说考试等，实现智慧化批改。付费群体：考试院
- d 管（智慧学校）：针对新高考（主方向），校园内业务源、资产、教学流程、学生行为管理等，进行集成化管理。付费群体：学校
- e 平台：以区域为单位的教育综合解决方案建设。付费群体：区域教育主管部门。目前在教育新基建的智慧学校模块中，讯飞基本是没有竞争对手
  - i 去年讯飞教育模块营收大概 62 亿左右，其中只有 1 亿左右来源智慧学校
  - ii 目前讯飞对接了大概 150 所学校为其做智慧学校的应用，平均一所 2,000 万（一次性收费），预计有 30 亿营收，落地预计 2 年左右，业绩确定性尚存不确定

2 医疗新基建：讯飞智医助理数据规模化应用，智医助理在安徽省重点扶持下，做为 2018 年安徽省民生工程建设内容，已在安徽省四县一区常态化应用，目前 100%覆盖 1153 各基层医疗机构，服务超 3400 民医生

3 在智慧医疗领域，目前国家目标就是实现智慧化医疗分级诊疗，希望帮助区域百姓的基础数据和常见病，都在当地形成数据库完成辅助治疗，避免挤兑核心医疗资源，最终实现提高区域医疗水平，让医疗资源得到充分利用

4 讯飞智医助理目的就是服务于区县医疗系统，人工智能服务诊断，帮助提高基层医生诊断合理率。并且在“智医助理”全科辅诊能力的基础上、逐步构建了针对危重病、传染病的闭环管理和自动预警体系，实现了包含 40 种法定传染病、15 种非法定传染病和 6 大类症候群的监测和预警能力

¶ 教育是讯飞目前落地的第一大赛道，在 G、B、C 端也培育多年，GB 的爆发动力是否充沛？C 端有没有出爆款的可能性？学习机系列与同行相比如何？

1 在教育公平的大背景下，讯飞目前在 G 端业务上发展趋势还是比较好的，尤其是在去年今年落地了很多业务

2 但是部分区域由于政府领导换届情况，导致部分业务搁置，但是预计今明两年也会加速落地。C 端方面学习机，目前在软件和硬件方面核心竞争力非常有优势，

个性化辅助学习和语音学习方面优势都是讯飞无可比拟的优势

¶ 在医疗赛道，讯飞目前布局的如何？未来的市场空间有多大？

1 讯飞在医疗赛道上的布局仍然是 G、B、C 三端去做，主要还是先 G、B 端向政府、平台方面去做，方便去获取到行业资源和数据，然后在慢慢面向 C 端去发展

2 目前讯飞在其实在医疗这个领域规模还比较小，但是按目前发展可以去想象，未来可能会出现一种产品就是你在网上把相应症状输入就可以知道自己得了什么病

3 当然目前讯飞还是主要服务与医院这类主体，招投标网显示目前一个医院采购智医助理大概是 400 万，全国大概 2800 各县级单位，统计大概 112 亿左右的市场规模

¶ 人工智能关键三个要素，数据、算力、算法，这三个核心要素讯飞分别都做了布局没有？在行业中分别处于什么样的位置？

1 算力方面主要是芯片厂商确定的，目前讯飞暂无涉及

2 算法方面目前讯飞拥有自己的算法，但是其实数据但是讯飞的核心

3 数据方面目前是讯飞最主要的核心资源，像教育、医疗方面的数据都是讯飞最核心的资源和布局

¶ 在工业互联网赛道和智能汽车赛道，讯飞分别做了哪些方面的布局？未来的市场格局如何？

1 讯飞在汽车领域目前有一定布局，而且在汽车这个板块讯飞在二级市场上其实有过一波炒作，同时讯飞的领导在这个方面有很高的展望，但是客观来讲讯飞只要在汽车领域业务没有明显变化，我们并不会去大篇幅提讯飞在汽车领域的展望

2 目前讯飞在汽车领域覆盖率还是比较高，去年汽车出货量 2,000 万台，讯飞大概有 600 万台都有覆盖合作，但是主要业务为智能语音软件授权，平均一台 60 元，业绩收入 3—4 个亿

3 虽然之前讯飞在智能座舱和座椅方面有布局意图，但是目前这个板块在市场已经属于红海竞争格局了，讯飞不一定在这个方面有很大突破，除非讯飞未来能和相关车企发布相应合作意图或者形成业务，否则讯飞在汽车赛道可能并不会有很大突破

¶ 对讯飞未来推出家庭机器人和 2030 超脑计划怎么看？

1 讯飞在 2019 年以前其实并没在行业上有突破和扩张，但是在 2021 年讯飞突然在人员上做了很大的储备，目的就是未来 2030 年的超脑计划，这也是讯飞未来十年的规划

2 讯飞目前在机器人领域上还是有一定储备，去年 11 月份在讯飞 1024 开发者节上，讯飞的机器人产品展示还是非常不错的，这个市场目前关注度还是非常高的

3 国内目前机器人行业发展阶段还是比较慢，虽然目前讯飞在机器狗上面做的确实

还不错，但是应用到机器人上面关于发展的进度和实施上，不确定因素还是比较多的

¶ 对刘总提出的千亿营收计划，怎么看？

1 千亿营收计划 2020 年和 2021 年提出来的，个人观点关于千亿营收目标董事长在年会上提出来主要目的还是去激励下面的人去努力

2 其实关于千亿营收目标来说，实现是肯定能实现的，但是具体实现的时间不确定性还比较多，按照董事长 2025 年的目标目前来看偏谨慎，目前也无法过多去判断

¶ 现在 AIGC 赛道这么火，讯飞在 NLP 领域是否有新的突破，或者现在语义理解能力能否与现在大火的 ChatGPT 是否有差距？

1 目前公司没有 CHATGPT 这一块，目前该模块技术储备有

2 公司认为这方面尚未达到技术变现阶段，所有这方面也没有过多涉及

¶ 关于职教赛道国家上个月退出政策鼓励职业教育，现在科大讯飞主要发力模式是传统学校，在职业教育领域讯飞是否有成熟的行业解决方案？

1 职业高校方面目前讯飞是有一定涉及的，目前讯飞高校上面都有一些涉及和介绍的

2 讯飞高校这一块主要是做英语口语的训练系统、智慧考试和智慧学堂的整体解决方案

ChatGPT 与商汤电话会

访谈日期：2022/12/7

具体内容

¶ ChatGPT

1 对 CHATGPT 的理解关于 CHATGPT，是一个文本对话的 AI 工具，功能非常强大，可以与它交流，它会提供反馈；可以写代码，甚至修改错误的代码；也可以产生作画的程序，它是一个非常强大的 NLP 文本对话工具

2 从中可以看出人工智能技术快速的进步

- a 人工智能一开始的落地在 G 端和 B 端，随着人工智能开始在 C 端落地，越来越可以感受到人工智能技术的快速进步
- b 其实人工智能已经渗透到了行业的方方面面，如商汤科技已经做到了 50-60 亿的收入规模。ChatGPT 有版本的不断迭代，最新是免费给用户来做测试使用，是构架在微软 AI 的算力中心之上的，需要比较强大的算力

3 CHATGPT 相比之前做了一些错误的更正，也更聪明，本质原因是算法做了微调

- a 未来人工智能的发展趋势是基于海量参数的大模型，参数级别可以达到十几亿级别，把大模型进行训练和构建，再根据细分场景做微调
- b ChatGPT 中，针对微调部分引入了新的算法——ReinforcementLearningfromHumanFeedback（RLHF，从人类反馈中强化学习）

4 CHATGPT 引入这个新算法后，变得更加智能

- a 但其中的基础是有一个海量参数的大模型，大模型也需要构建在大的算力中心上，ChatGPT 是构建在微软的算力中心上。未来的 AI 的发展趋势是构建大模型，也需要大的算力中心，大的算力中心不管是自己建还是去购买，都需要海量的投入
- b ChatGPT 是 OpenAI 推出的，OpenAI 本身就有 700 亿美元的估值，可以有海量的投入的能力

¶ AI 未来发展

1 使用大的算力中心，并在算力中心的基础上构建大模型，这就是 AI 未来发展的必然趋势。这使得 AI 公司的进入门槛在不断提升，一方面要做出大模型，难度已经很高了，还需要巨大投入去能够使用算力中心

2 这意味着，现在存量的 AI 公司，特别是已经上市的 AI 公司，可以不断有自己的资本实力

- a 护城河在变的越来越宽，这与互联网公司完全不同，互联网公司没有算力成本，所以不断有新公司出现
- b 但 AI 公司不同，已经建立优势的公司会不断构建自己的护城河，小的创业型 AI 公司越来越难以生存，创业门槛在不断提高

3 本身 AI 基于海量数据，数据又如此重要，未来中国一定不会让国外公司基于中国

各种数据来做训练、学习、提供服务

a 所以，中国公司本身在其中就会有优势
b 商汤在 50-60 亿的收入规模，仅次于商汤的是云从科技，大概是十几亿的收入规模，剩余的还有格灵深瞳，再就是更小规模的未上市的公司

4 头部公司会不断累积自己的竞争优势，从国内来看，商汤拥有自己的大模型和大的算力中心，算力中心在临港，投了 50 亿去建，已经建成

¶ ChatGPT 未来的发展

1 从 CHATGPT 未来的发展来看，它可能会颠覆很多固有的商业模式和商业形态

a 比如对 Google，完全可以把 ChatGPT 作为搜索引擎，当然新兴知识可能无法预测，因为其训练数据基于 2021 年以前
b 当然 ChatGPT 说未来会引入新的算法，之后这方面可能会有提升，但目前版本无法回答新兴事物相关问题。ChatGPT 有比较强大的算力基础

2 算力基础是收费的，所以之后 CHATGPT 也会收费，会收会员费，比如一年的会员费是 100-200 美元。它搜索的效率和质量是 GOOGLE 完全不能比的。新的技术的出现会对固有商业模式带来本质的变化

3 新的商业模式的出现一定是构建在新的技术之上的

a 比如智能汽车开始收软件服务费，可以按月、按年来收各种智能软件的服务费，这在之前的传统汽车上是难以想象的
b 再比如之前的单机游戏，靠一次性卖 license 赚钱，变成网络游戏以后游戏是免费的，但要购买一些道具等，这种商业模式的变迁是网络技术的发展带来的

4 因此，新技术的出现会带来新的商业模式，新的商业模式会对以往商业模式产生颠覆，会对固有的很多公司带来严重冲击

a ChatGPT 就会对 Google 产生冲击
b Google 搜索免费，靠广告赚取收入，Google 收入的 90%左右依然来源于广告收入。ChatGPT 可以完全不依靠广告，可以通过收取会员费获得收入，虽然会员费目前可能并不便宜，但其搜索质量会相当高，因此这就会带来商业模式的替换和提升

5 未来 AI 公司的想象力是无限的，想象空间巨大，甚至可能颠覆掉很多传统巨头，新兴 AI 公司前景巨大。比如 CHATGPT，可以应用到包括教育、医疗等众多服务领域

a 同时也要看到，本身 AI 基于海量数据，在全球对于数据安全保护越来越看重的情况下，未来中国庞大的市场一定是由中国公司来做
b 实际上，中国人工智能公司也是走在世界前列的。因此，要重视目前已经有丰厚实力的 AI 公司，未来前景是非常庞大。商汤科技大模型和大的算力中心一方面一定是 AI 公司的非常高的准入门槛，
c 另一方面，也决定了 AI 公司能否实现最大商业化、稳态后能否实现盈利的重要基础，因为只有能够有大模型、大算力中心来标准化、低成本、大规模地提供算法，才能实现行业扩张、商业化落地、实现未来长期高质量盈利

¶ 商汤

1 商汤已经具备了这样的基础建设（大模型+大算力），能够支撑它的每条业务线在找准方向后实现快速突破

2 商汤的大模型：已经达到百亿级参数以上

a 深度学习平台，使用商汤领先的视觉算法训练框架，高效利用 GPU 集群算力，训练单个大模型时可以在一千块 GPU 上取得超过 90%的加速效率，在业内处于领先水平

b 商汤 AIDC 目前和谷歌、微软已经一起排到了全球的前三，相比于传统的人工智能的标准可以提高 600 倍

3 从近几年商汤持续生产 AI 模型的速率可以看到，趋势非常明显

a 从研发团队人均生产模型的个数上来看，今年上半年已经相比于 2021 年提高了 15%，包括商用累计模型已经有了 4.9 万个，同比提升了 40%

b 这都是公司已经积累了比较高的门槛的表现，同时也是未来长期拓展每条业务线的夯实的基础

4 商汤的智能算力中心可以 1 天内可完成 1,000 亿参数模型的完整训练，算力规模达 5EXAFLOPS（1EXAFLOPS 等于每秒可达一百亿亿次浮点运算），并且已经形成规模的商业化收入

5 商汤面临全面的业务优化

a 具体来看，除了传统的 ToG 端政府端业务和商业业务，还要重点重视公司两块新的业务——智慧生活和智能汽车

b 这两大板块在今年上半年已经可以在收入端看到非常强劲的增长，智慧生活板块业务收入同比增长 98%，已经占到了业务的 21%，去年全年收入占比只有 9%；智能汽车板块业务今年上半年同比增长 71%，占到了总收入的将近 10%

¶ 两个板块的业务发展

1 未来这两个板块的业务有可能继续保持增速，甚至会有更好的表现。拉长时间线到未来 5-10 年的区间来看，这两块业务可能会替代目前占公司收入大头的 G 端业务和 B 端业务，成为公司收入的重要构成。这两块业务的进展和布局优化是判断商汤价值的重要因素

2 关于两块业务增长的原因以及未来市场空间，智慧生活方面，公司在硬件、软件都有布局，因此公司商业模式以及未来单点突破的空间都非常丰富

3 硬件方面

a 包括 AI 的 CMOS，以及今年下半年推出的首款 AIISP 芯片，都是能在全球范围针对手机、汽车等其它互联网终端做大规模推广的，起量后带来的收入空间是非常可观的

b 按大致测算，整体收入体量可以达到 60 亿美元，这是公司能够触及到的市场规模，这部分长期来看会是智慧生活板块的一个强劲的增长驱动力

4 硬件模式相比软件的起量有更强的爆发力

a 此外，现有智慧生活板块包含的包括元宇宙、教育、智慧医疗等板块的业务都在今年上半年有比较好的落地，商业化突破非常明显

b 比如元宇宙今年在上半年的冬奥、包括沙特的一些场景都有非常好的应用；智

慧医疗也在包括瑞金等三十多家医疗机构实现了落地；今年也推出了家用下象棋的机器人——元萝卜，在各个电商平台销售表现非常亮眼

5 智慧生活板块未来面临 C 端、B 端的各方面场景和空间都是非常新兴的蓝海市场

- a 长期来看空间和潜力都很丰富
- b 智能汽车方面，公司目前形成了三部分布局，形成了全栈自动驾驶能力，包括今年推出的 V2X，以及智慧车舱、智能驾驶。三条业务都是以大模型、大算力作为非常强劲的底座

6 目前已经形成的商业化成果来看，已经有 60 个车型作为量产定点，合作车企也有 30 家以上，座舱视觉、AI 软件等领域公司也排在了市场份额的第一，商业化成果突出。未来长期来看，这些能力还有很大空间可以发挥。很核心的一个原因是，商汤的能力基于全面的基础

¶ 自动驾驶方案

1 自动驾驶方案包括感知、预测、决策、控制等各方面

- a 商汤作为中国车企第三方软件提供商，和车企联手从驾驶数据的获得以及整体数据的获取效率，以及大装置带来的数据利用效率维度相比，都远超市面一系列竞争对手
- b 甚至直接和特斯拉对比，无论是在场景数据层面，还是商汤绝影平台整个数据获取效率、已经建成的 AIDC 和大模型大装置都远超特斯拉现有水平，有更广阔的空间

2 从现有布局整体来看，公司整个智慧汽车业务长期来看的版图、优势以及国内政策扶持、本土化契机等等都是公司业务强劲增长的重要驱动力来源

- a 长期来看，这一部分收入潜力，比如智能车舱预装，从车舱渗透率以及商汤自身竞争优势来看，可以达到 20 亿美元的量级
- b 自动驾驶后续由于新的技术变革，带来的一系列商业模式的改变，比如自动驾驶方面一些订阅模式，结合预装的整体的市场空间，可以达到 60 亿美元左右的收入体量

3 这两块新兴领域在未来 10-15 年的收入潜力已经远超商汤目前形成的 50 亿元人民币左右的商业化的收入体量。空间强大

4 结合公司已经具备的大量算法积累、各个业务的布局、合作的客户、已经储备的解决方案和软件能力，会在长期带来非常好的收入增长以及盈利的可能性

- a 公司近期资本上的一系列举动也有证明，包括近期公告高管承诺到 2024 年 1 月不减持，并且持续在市场做回购
- b 上周公司在投资者大会上说高管希望不减持直到公司实现盈利

5 按目前发展节奏以及新兴业务的高速增长请况，商汤应当已经非常接近盈亏平衡点

6 今年下半年和明年情况

- a 除了智慧生活和智能汽车两个板块的业务会继续保持高增速外，它的两块基础业务——智慧政府和智慧商业的业务在今年下半年也会有明显回暖
- b 上半年由于疫情等一系列原因，在整个商业扩展、回款、确认收入等方面都有

压力，再加上商汤本身业务又有一定季节性特征，很多项目在四季度交付，结合下半年回暖，对下半年和明年商汤的增长保持比较乐观的态度

¶ 人工智能

1 人工智能是穿越周期的一个颠覆性技术，商业化落地也才刚刚开始，以后面临多元应用场景。现在已经上市的 AI 公司具备较大的竞争优势并积累了较高门槛

2 商汤作为头部玩家，在今年、明年一系列行业变化和政策支持上，可以持续享受数字经济、人工智能普及带来的红利。从现在来看，受今年整体港股市场影响，商汤估值目前处于一个非常具有性价比的空间

3 现在来看，2022 年商汤 PS 估值在 10 倍出头，23 年 PS 在 6 倍左右。对比全球 AI 公司估值，包括 A 股市场以及历史情况，可以看到其处于非常明显的低估的区间，叠加明年整体需求回暖、新兴业务能够实现超过 50%甚至翻倍的强劲增长，长期来看，公司非常具备投资价值