

# 《人工智能大模型（类 ChatGPT）技术应用及展望》

段云峰

教授 北京大学博士后

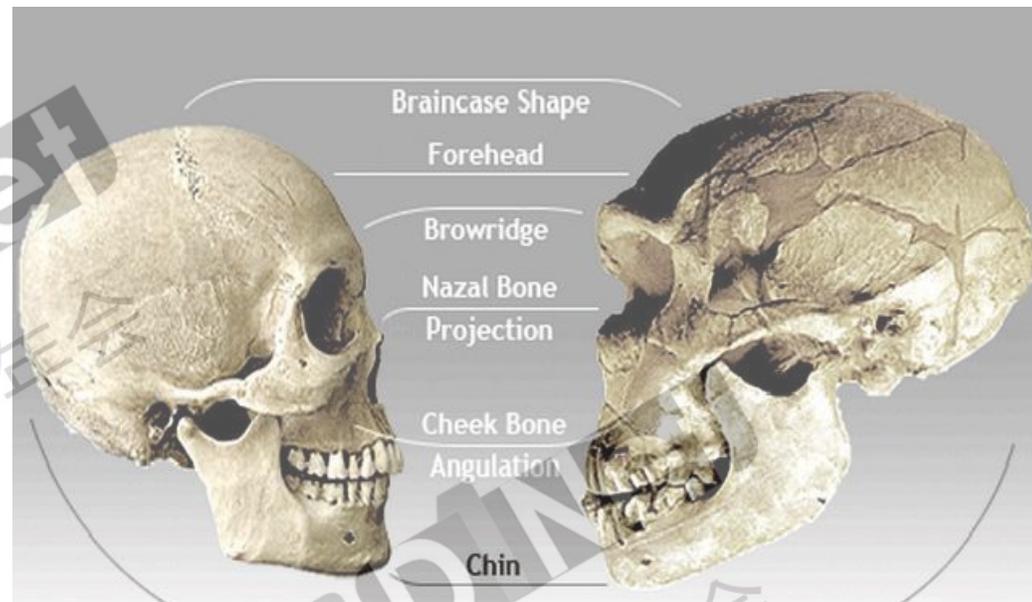
# 主要内容

- 一、“大模型、多模态” (chatGPT) 基础
- 二、大模型的技术发展
- 三、大模型的应用展望
- 四、几点思考



# 一、“大模型、多模态” (chatGPT) 基础

- 大模型=大参数?
  - 传统机器学习的参数“降维”问题
  - 深度学习的神经网络“参数”扩展
- 多模态=多数据源?
  - 文本、图像、视频等等
  - 人类的学习过程：声音、图像、记忆
- 自然语言处理 (NLP) 成为突破口，生产式AI (AIGC)开启突破
  - 从GPT (Generative Pre-Training, 生产式预训练) 到chatGPT
  - 开启人与计算机之间的编程语言翻译转换
  - AI的艺术创作：琴棋书画
- 从图灵测试到专家测试
  - GPT-4: SAT 考 710 ; 美国律师考试前10%; 职业诗人、画家等
  - 人类脑细胞的数量：140-150亿
- “十年前，人们普遍认为，人工智能将首先影响体力劳动，然后是认知劳动，并可能有一天它可以从事创造性工作。现在看来，它将以相反的顺序进行。” ——OpenAI公司首席执行官萨姆·奥特曼 (Sam Altman)
- 类比移动互联网、iPhone?



# 参数的“量变”产生“质变”

- Google 研究者发表了题为 “大型语言模型的涌现能力” (Emergent Abilities of Large Language Models, 2022年8月) 的论文
  - 考察了以 GPT-3 为代表的语言模型，发现语言模型的表现并非随着模型规模增加而线性增长，而是存在**临界点**，只有当模型大到超过特定的临界值，才会涌现出较小的模型不具备的能力
  - 涌现（奇点？）：“如果一种能力不存在于较小的模型中，而存在于较大的模型中，那么这种能力就是涌现出来的。”
  - 文中总结了GPT-3模型具有的137项涌现能力，对于更传统的NLP基准模型，例如 BIG-Bench，其具有的涌现能力也包括67项
- 【思考】客户离网分析的参数能到多少？

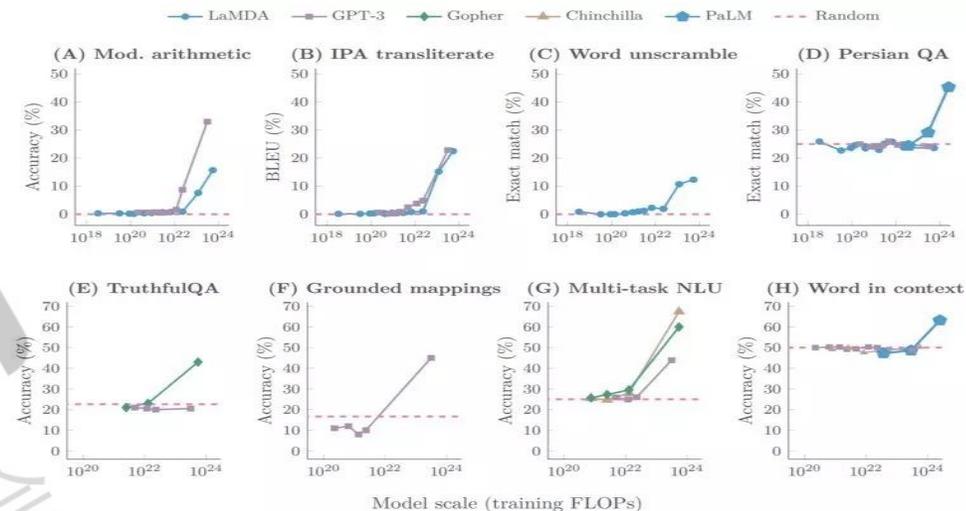


图3. 模型大小和各项任务准确性之间的折线图

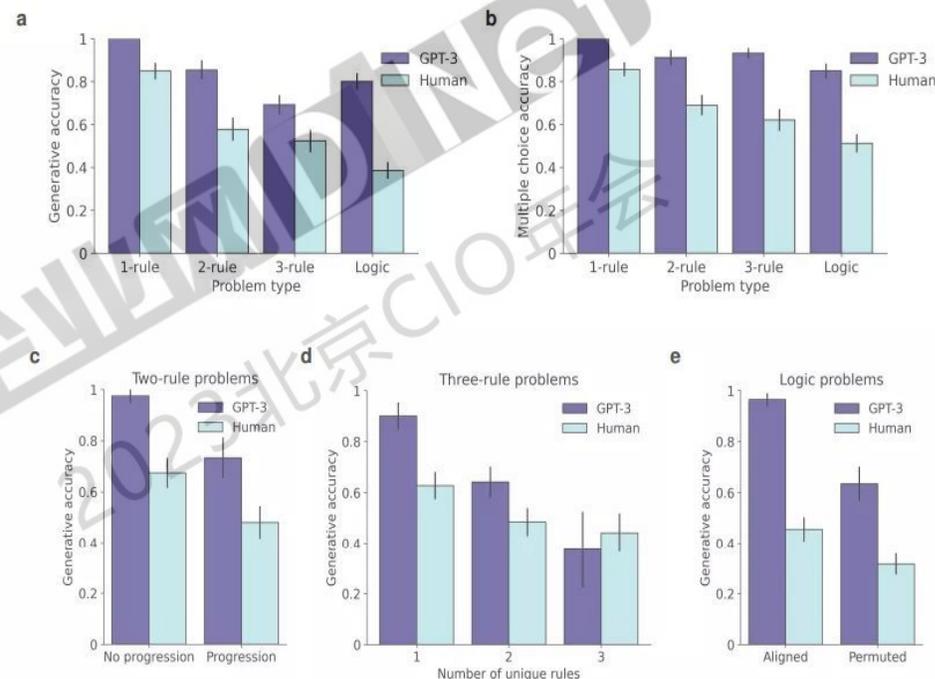


图5. 在生成及多选这两类问题上，需要运用的规则越多，GPT-3的表现就越优于人类

# ChatGPT的魅力

- OpenAI在2022年11月30日发布了ChatGPT，ChatGPT是人工智能技术驱动的自然语言处理工具，它能够通过理解和学习人类的语言来进行对话，还能根据聊天的上下文进行互动，真正像人类一样来聊天交流，甚至能完成撰写邮件、视频脚本、文案、翻译、代码，写论文等任务
- 2023年一月末，ChatGPT的月活用户已突破1亿，成为史上增长最快的消费者应用
- 截止2023年1月，美国89%的大学生都是用ChatGPT做作业
- “一个全新的摩尔定律可能很快就会出现，即宇宙中的智能数量每18个月翻一番。” ——「ChatGPT之父」Sam Altman
- 【思考】chatGPT仅是聊天吗？促使AI自编程、自分析、自控制？会将部分种类的劳动力成本（客服等）降至零吗？人口还有红利吗？



AI陪伴的数字人生—  
—Replika

# 哪个重要? —— 参数、算法、数据、算力、场景、人才

- 谷歌为何“马失前蹄”?
  - 从Bert的“上下文”到GPT的“上文”
  - PalM训练参数5400亿
- 算法:
  - Transformer+RLHF
- 数据(量)
  - ChatGPT 45TB
- 算力
  - 微软Azure与谷歌云计算
- 场景
  - NLP切入Bing、edge等, 会替代搜索引擎吗?
- 人才
  - OpenAI2015年成立之初, 吸引顶尖人才
- 【思考】NLP领域借鉴互联网的海量数据, 企业的海量训练数据在哪里? 企业如何产生“涌现” (奇点) 应用?

大脑皮层的神经元数量, 与通用智能水平成正比

脑容量和神经元数量

哺乳动物大脑皮层质量和神经元数量



lang	语言名	训练集	占比
en	英语	181014683608	92.65%
fr	法语	3553061536	1.82%
de	德语	2870869396	1.47%
es	西班牙语	1510070974	0.77%

# 生成式AI会改变哪些行业和职业?

- 从助手到“雀占鸠巢”?
- 文字生成领域——文案设计等
  - 行政类文书样板
  - 《粤港澳大湾区2035区域现代化战略规划》
- 艺术创作领域
  - 诗词歌赋、绘画、音乐、小说等领域的创作替代
- 教育领域
  - 从死记硬背到AI工具使用，思维创新训练等等
- 法律领域
  - 法律条文查询及解释等
- 计算机编程领域
  - Copilot: Github和OpenAI合作研发的AI编程工具，基于数十亿行开源代码进行训练，能够按照不同的程度、不同的功能要求，辅助人类完成编程任务
- 【思考】企业的运营将发生哪些改变? ——电商改写了商业模式，AI之后呢? 核心竞争力在哪里? 从信息透明到智力透明?
- 【规划】信息透明，解决了商品的信息壁垒，完成比价、比性能等; 智力透明，围绕A产品的产品设计、潜在客户、精准营销、物流配送、售后服务等产业链重新组合，重新洗牌。

14

现在，我需要10个数据集，每个数据集含有三个变量，变量1表示序号，变量2表示时间，变量3表示1-10之间的随机整数，然后打印这20个数据集。请提供一段SAS代码。



以下是可以生成10个数据集，每个数据集含有三个变量，变量1表示序号，变量2表示时间，变量3表示1-10之间的随机整数，然后打印这20个数据集的SAS代码：



```
java Copy code

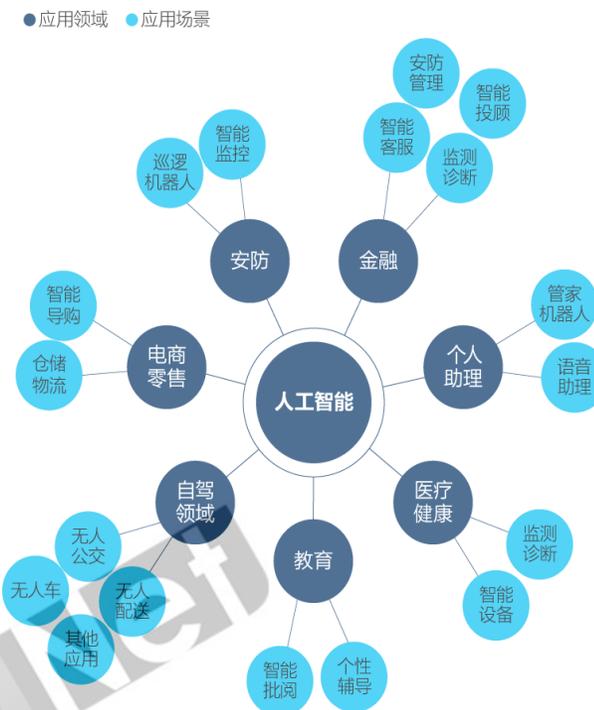
/* 定义生成数据集的数量 */
%let num_datasets = 10;

/* 循环生成数据集 */
%macro create_datasets;
  %do i = 1 %to &num_datasets;
    /* 定义数据集名称 */
    %let ds_name = dataset&i;
    /* 生成数据 */
    data &ds_name;
      /* 定义变量 */
      length var1 8 var2 8 var3 8;
      /* 生成随机数种子 */
      call streaminit(&i);
      /* 生成数据 */
      do j = 1 to 10;
        var1 = &i;
        var2 = datetime();
        var3 = ceil(10*rand("uniform"));
      output;
      end;
    run;
    /* 打印数据集 */
    proc print data=&ds_name;
      title "Dataset &i";
    run;
  %end;
%mend;
%create_datasets;
```

# 生成式AI会改变哪些行业和职业? -2

- Resumebuilder.com 调查了 1000 名计划使用或已经使用 ChatGPT 的企业领导人，发现近半数企业部署了 ChatGPT，部分企业还表示已经开始用 ChatGPT 取代员工
- 企业高管谈论了 ChatGPT 的用途，其中 66% 用于写代码，58% 用于文案和内容创作，57% 用于客户支持，52% 用于写会议摘要和其它文档
- 在招聘过程中，使用 ChatGPT 的企业还有 77% 用它撰写工作职位描述、66% 用它起草面试申请书，65% 回应求职书
- 大部分企业领导人对 ChatGPT 的表现印象深刻，55% 的人认为工作质量“出色”，34% 的人认为“非常好”
- 几乎所有使用 ChatGPT 的公司表示节省了资金，48% 表示节省逾 50,000 美元，11% 表示节省逾 100,000 美元
- 这些企业中 93% 还计划扩大对 ChatGPT 的使用
- 【思考】李开复指出，“AI 2.0将会带来平台式的变革，改写用户的入口和界面，诞生全新平台，催生新一代AI 2.0应用的研发和商业化。总的来说，AI 2.0将是提升21世纪整体社会生产力最为重要的赋能技术。”

当前人工智能企业的主要应用领域



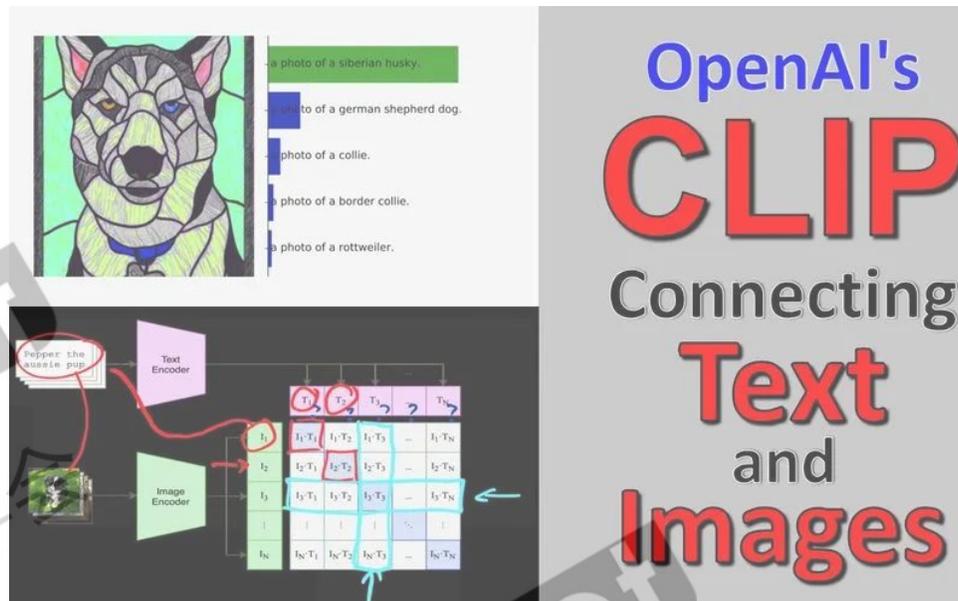
# chatGPT对战争的影响

- 美军认为，人工智能将改变21世纪的战争形态
- 一是参与认知作战，引导舆论走向
  - ChatGPT的出现或将把认知战的烈度提升到一个全新的层次。ChatGPT可模仿人类（领袖声纹等），在网络上可批量、快速产生真假难辨的煽动性网络信息，一旦用于认知作战，可能成为恶意传播虚假信息、操纵舆论的工具
- 二是辅助军事决策，提升作战效率
  - 军队一旦装备ChatGPT或类似程序，就能够实时响应战场上分队或单兵的交互信息，提供最新的战场态势，从而缩短军事决策过程所需时间，极大提升作战效率
- 三是自主编写程序代码，实施网络攻防战
  - ChatGPT生成代码提高工作效率的能力也给黑客打开了新世界的大门，让他们能够更为便捷地设计、编写和执行恶意代码。攻击者可使用ChatGPT构建机器人和网站来诱骗用户共享他们的信息，并发起针对性很强的社会工程诈骗和网络钓鱼活动。ChatGPT可以帮助更多人发现潜在的漏洞，但这些漏洞随后可能会被人利用，扩大威胁范围
- 【思考】商场如战场，AI如何助力商战？——舆情、决策、攻防、民主体系



## 二、大模型的技术发展

- 自然语言处理NLP领域:从RNN到transformer
- 生成式AI领域: 扩散 (Diffusion) 模型
- 多模态 CLIP (Contrastive Language-Image Pre-Training)
- GPT-4模型2022年8月完成训练, OPENAI却花了六个月让其更安全



AIGC的技术发展路径

AIGC的实现效果



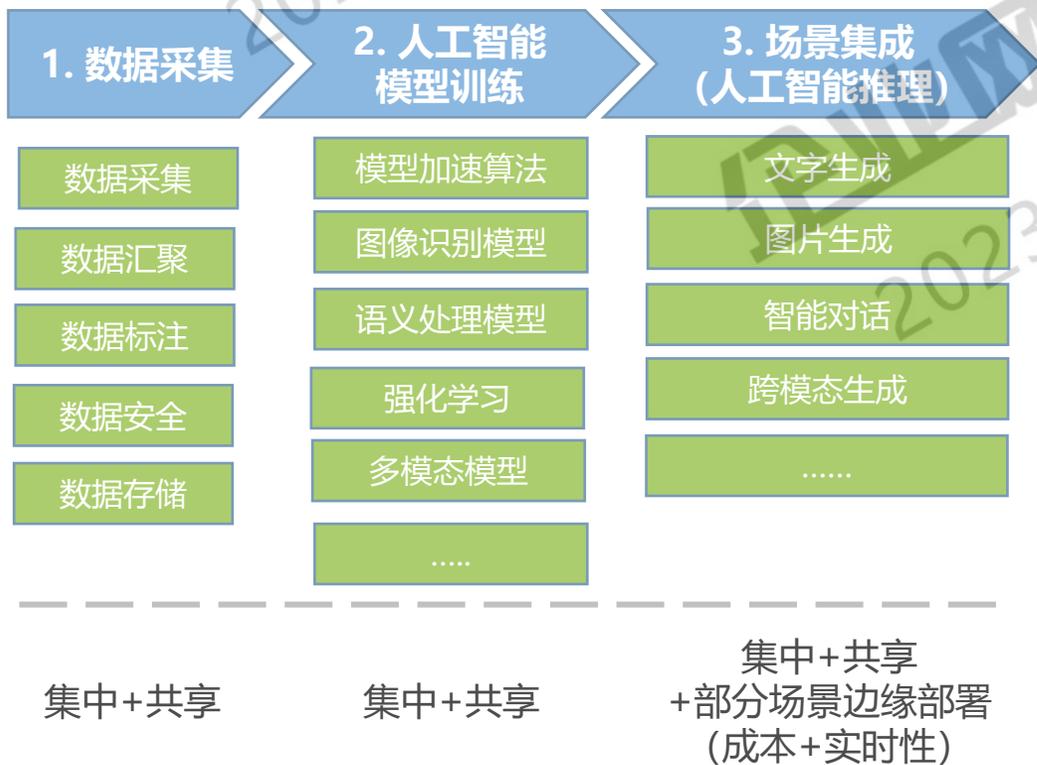
【思考】不同算法决定通用大模型?

# 多模态大模型平台架构

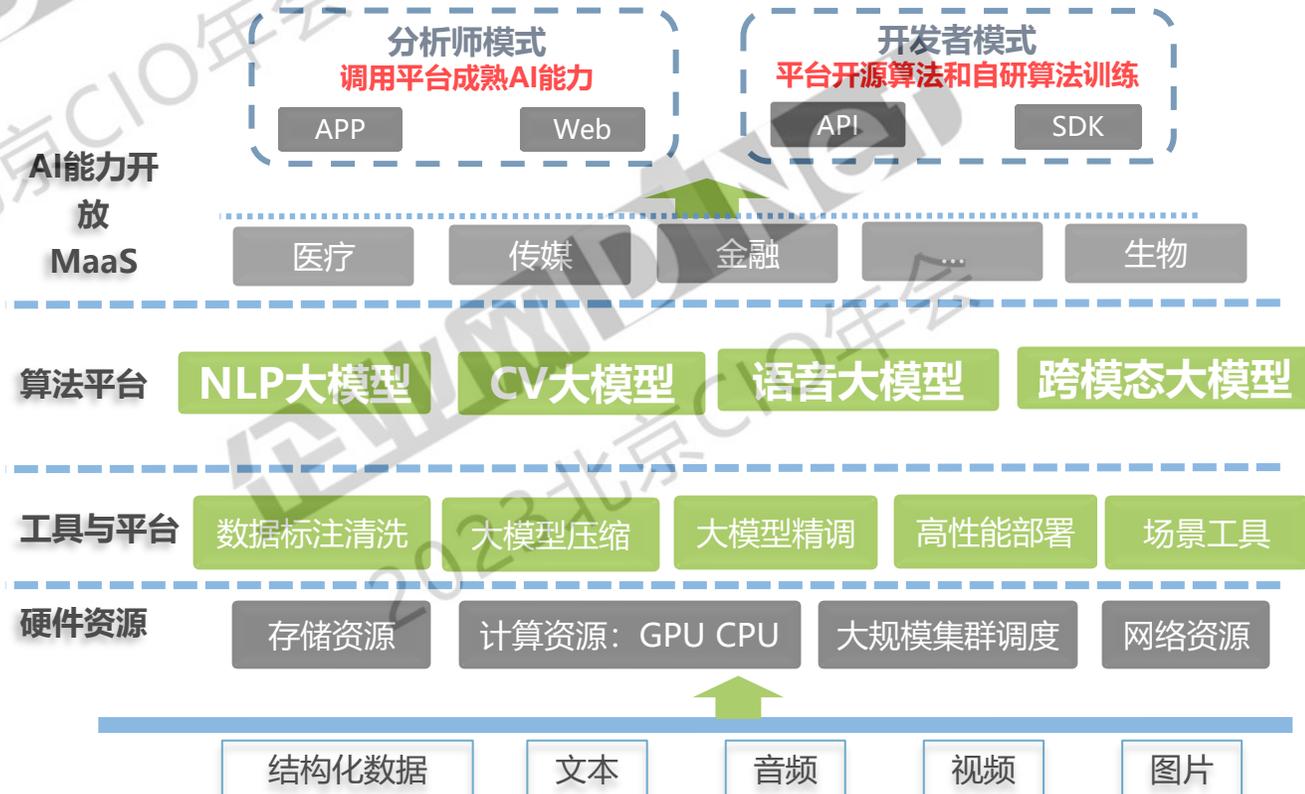
ChatGPT作为大模型的一种，已经展现出惊人的效果，构建以大模型为主导多模态技术体系成为未来技术发展的方向。

大模型基于硬件资源池和软件框架，通过对海量多模态数据进行训练，从而提供成熟AI能力和多模态能力

## 多模态大模型工程阶段



## 多模态大模型技术架构



# chatGPT对InstructGPT的改进

- “Tell me about when Christopher Columbus came to the US in 2015”（告诉我克里斯托弗·哥伦布2015年来美国的时候）
- 【ChatGPT】
  - This question is a bit tricky because Christopher Columbus died in 1506, so he could not have come to the US in 2015. But let’s pretend for a moment that he did!（这个问题有点棘手，因为克里斯托弗·哥伦布死于1506年，所以他不可能在2015年来到美国。但让我们假装他做了！）
  - If Columbus arrived in the US in 2015, he would likely be very surprised at the changes that have occurred since he first landed in the “New World” in 1492. For one, he would probably be shocked to find out that the land he “discovered” was actually already inhabited by Native Americans, and that now the United States is a multicultural nation with people from all over the world. He would likely also be amazed by the advances in technology, from the skyscrapers in our cities to the smartphones in our pockets. Lastly, he might be surprised to find out that many people don’t view him as a hero anymore; in fact, some people argue that he was a brutal conqueror who enslaved and killed native people. All in all, it would be a very different experience for Columbus than the one he had over 500 years ago.
  - （如果哥伦布于2015年抵达美国，他可能会对1492年首次登陆“新世界”以来发生的变化感到非常惊讶。首先，他可能会震惊地发现，他“发现”的土地实际上已经有美洲原住民居住，而现在美国是一个多文化国家，有来自世界各地的人。他可能也会对科技的进步感到惊讶，从我们城市的摩天大楼到口袋里的智能手机。最后，他可能会惊讶地发现，许多人不再将他视为英雄；事实上，有些人认为他是一个残暴的征服者，奴役并杀害了当地人。总之，对哥伦布来说，这将是一次与500多年前截然不同的经历。）
- 【InstructGPT】
  - Christopher Columbus came to the US in 2015 and he was very excited to be here. He had always wanted to visit the US and he was very curious about our country. He was very impressed with our country and he enjoyed his time here.（克里斯托弗·哥伦布于2015年来到美国，他很高兴来到这里。他一直想访问美国，他对我们的国家非常好奇。他对我们的国家印象深刻，他很享受在这里的时光。）

# ChatGPT的RLHF(Reinforcement Learning from Human Feedback)

- 数据处理的过程更重要?
  - 人类参与反馈修正
- 从数据标注到数据矫正
  - 从事前到事后
- 避开“胡言乱语”
  - 避开种族、政治错误等
- 中文的隔离效应
  - 中国人力资本的优势
  - 是否会出现下一个“百度”？
- 如何避免各种“倾向性”？
  - 繁体、简体、英文的TW问题
  - 价值观的引导
- 【思考】企业模型的行业化、个性化RLHF过程？

Step 1

Collect demonstration data and train a supervised policy.

A prompt is sampled from our prompt dataset.

A labeler demonstrates the desired output behavior.

This data is used to fine-tune GPT-3.5 with supervised learning.



Step 2

Collect comparison data and train a reward model.

A prompt and several model outputs are sampled.

A labeler ranks the outputs from best to worst.

This data is used to train our reward model.



Step 3

Optimize a policy against the reward model using the PPO reinforcement learning algorithm.

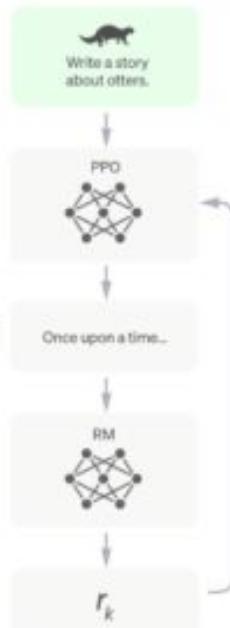
A new prompt is sampled from the dataset.

The PPO model is initialized from the supervised policy.

The policy generates an output.

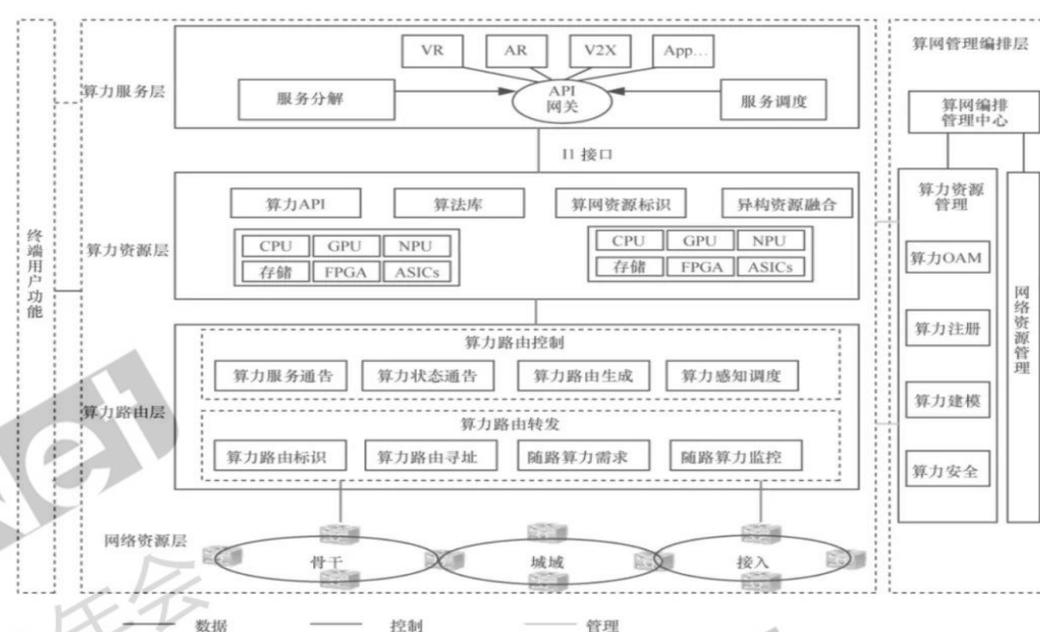
The reward model calculates a reward for the output.

The reward is used to update the policy using PPO.



# 大模型的算力“紧箍咒”

- 2020年5月，微软在Azure上为OpenAI搭建了独立的**超级计算机系统**，包含28.5万个CPU核和1万张GPU卡（当时为V100）
- 英伟达2021年4月，使用1024张80G显存的A100卡训练1750亿参数的GPT-3模型，需要训练34天
- 英伟达A100/H100芯片的中国“断供”
- 算力网络的架构突围
  - 算力网络是“一种根据业务需求，在云、网、边之间按需分配和灵活调度计算资源、存储资源以及网络资源的新型信息基础设施”



# ChatGPT的“复现”失败问题

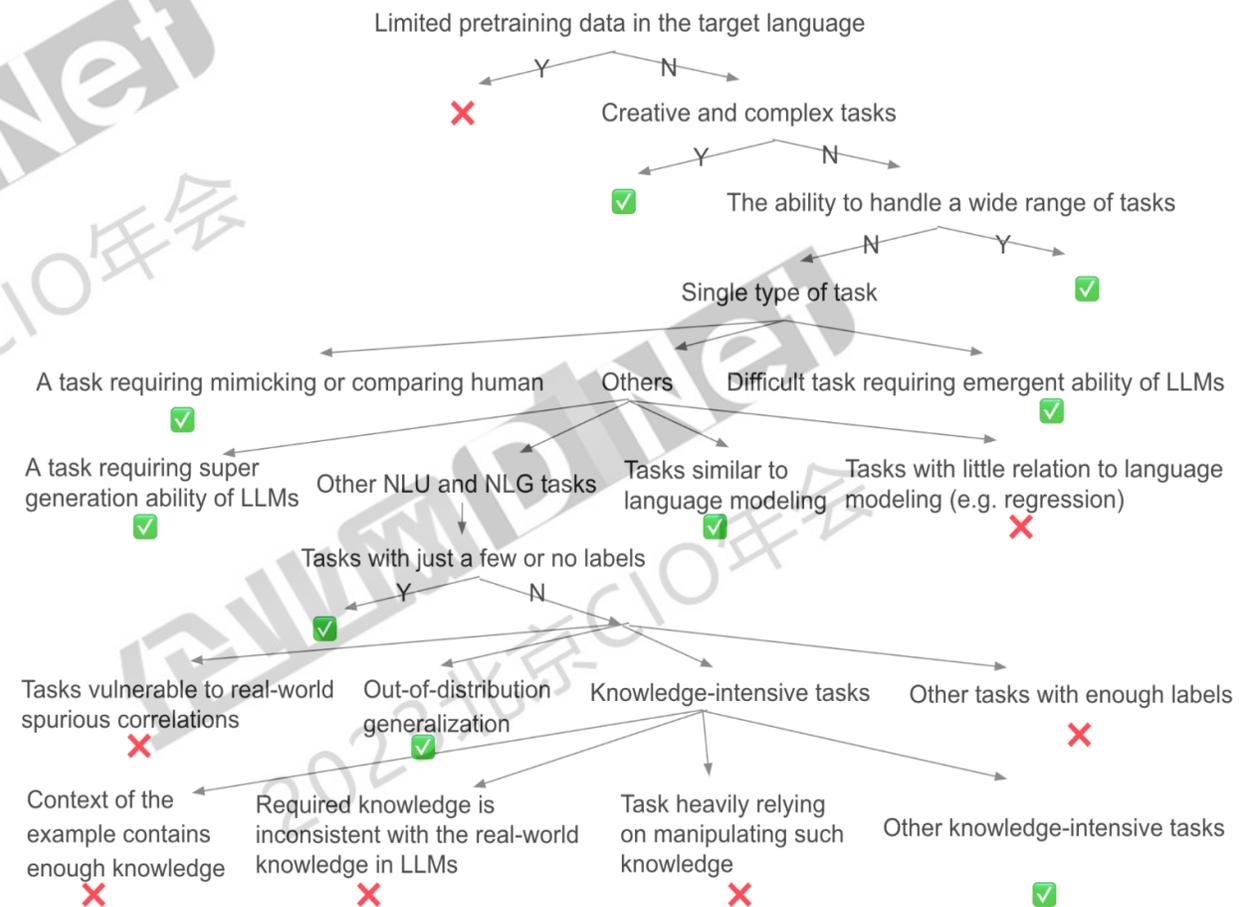
- “失败”是指训练得出模型有接近 GPT-3 或者更大的参数量，但仍无法与 GPT-3 原始文献中报告的性能所匹配——<https://jingfengyang.github.io/gpt>
- 所有的公开模型（例如：OPT-175B 和 BLOOM-176B）都在一定程度上“失败”了。一次训练就将需要在约 1000 个 80G A100 GPU 上花费至少 2 个月的时间（数据来自于 OPT 的原始文献）
- GPT-3 在共计 300B 的 token 上进行训练，其中 60% 来自经过筛选的 Common Crawl，其它则来自：webtext2（用于训练 GPT-2 的语料库），Books1, Books2 和维基百科
  - OPT-175B 所使用的 Pile 语料库声称有着更好的多样性，但 BLOOM 采用的 ROOTS 语料库则有太多的已经存在的学术数据集，缺乏 Common Crawl 数据所包含的多样性。这可能导致 BLOOM 性能更差。
  - 作为对比，GPT3 来自 Common Crawl 语料的占比则要高得多，而它们是多样的和来自广泛领域的，这也可能是 GPT-3 能够作为首个通用聊天机器人 ChatGPT 的基础模型的原因之一。
- 训练策略包括训练框架、训练持续时间、模型架构/训练设置、训练过程中的修改。在训练非常大的模型时，它们被用于获得更好的稳定性和收敛性
  - 一个参数量大于 175B 的模型往往需要 ZeRO 式的数据并行（分布式的优化器）和模型并行（包括张量并行（tensor parallel）、流水线并行（pipeline parallel），有时还包括序列并行（sequence parallel））
  - OPT 使用了 ReLU 激活函数，而 PaLM 使用 SwiGLU 激活函数，GPT-3 和 BLOOM 使用 GeLU，它通常使得训练的 LLM 的性能更好
- 【思考】山寨chatGPT的难点还有哪些？

# 应该在哪些任务上使用GPT-3.5或ChatGPT?

- 由于在对话场景下的良好对齐，ChatGPT 作为一个聊天机器人表现优异。但我们通常使用 GPT-3、InstructGPT (GPT-3.5)、以及 Codex 这些 ChatGPT 背后的模型作为在更多任务和使用场景下的通用模型

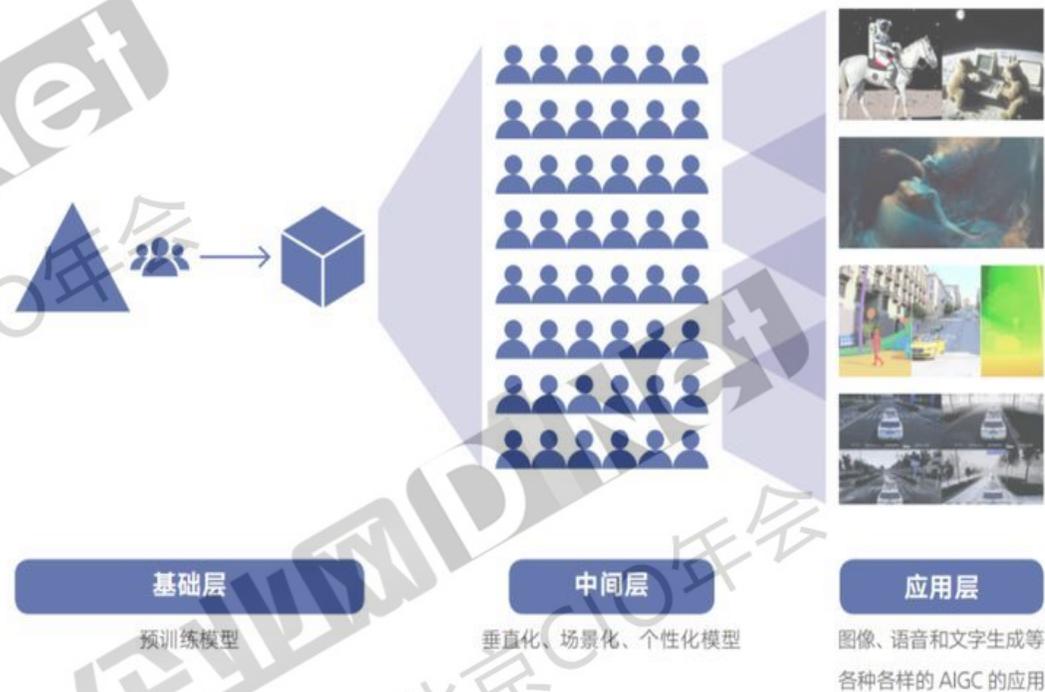
- 1. 创造性和复杂的任务：包括代码（代码补全、自然语言指令生成代码、代码翻译、bug 修复）、文本摘要、翻译、创造性写作（例如写故事、文章、邮件、报告，以及写作的改进等）
- 2. 只有少数标注或者没有标注数据的任务
- 3. 分布外（Out-of-distribution, OOD）泛化
- 4. 需要处理多种任务的能力，而非关注特定任务上的卓越表现
- 5. 那些检索不可行的知识密集型任务

Whether to use GPT-3 in your use case (compared with a fine-tuned smaller model)?



# 三、大模型的应用展望

- 传统数据分析模型的优化
  - 从机器学习升级深度学习
  - 逐个分析主题的优化
- 带动新“蓝海”
  - 从NLP改良客户服务入手
  - 数字人改变的营销、服务等
- MAAS (模型即服务) 成为可能
  - 如何构建通用大模型
  - 开源算法到行业算法、领域算法
- AIGC带动元宇宙
  - 从线上营销转为元宇宙营销
  - 从2C向2B延伸
- 【思考】周鸿祎：“ChatGPT是一种生产力强大的再造工具。任何行业的APP、软件、网站、应用如果再加上GPT的能力都值得重塑一遍，不要把它当成一个媒体工具、搜索引擎、聊天机器人，这都有点把它看得太小儿科了



图：AIGC 产业生态体系的三层架构

# 从大模型技术角度驱动的应用

- NLP角度的切入
  - 需求、设计、生产、供应链、销售、售后等环节的人工交互 (UI) 替代
- AIGC角度的引导
  - 文件编写、文案设计、营销广告 (海报)、操作视频等
  - AI如何改变电商渠道平台? (类搜索引擎? ——魔镜)
- 多模态融合
  - 客户的多模态视图 (短视频、直播)
  - 情感分析、性格分析等
- 围绕非结构化数据为主
  - 如何涵盖结构化数据
  - 数据标注、RLHF
- 【思考】企业的管理如何重塑?

## • 大模型有广泛的应用场景

大模型作为智慧大脑, 可以对各类场景提供通用化解决方案



# 从业务角度驱动的应用

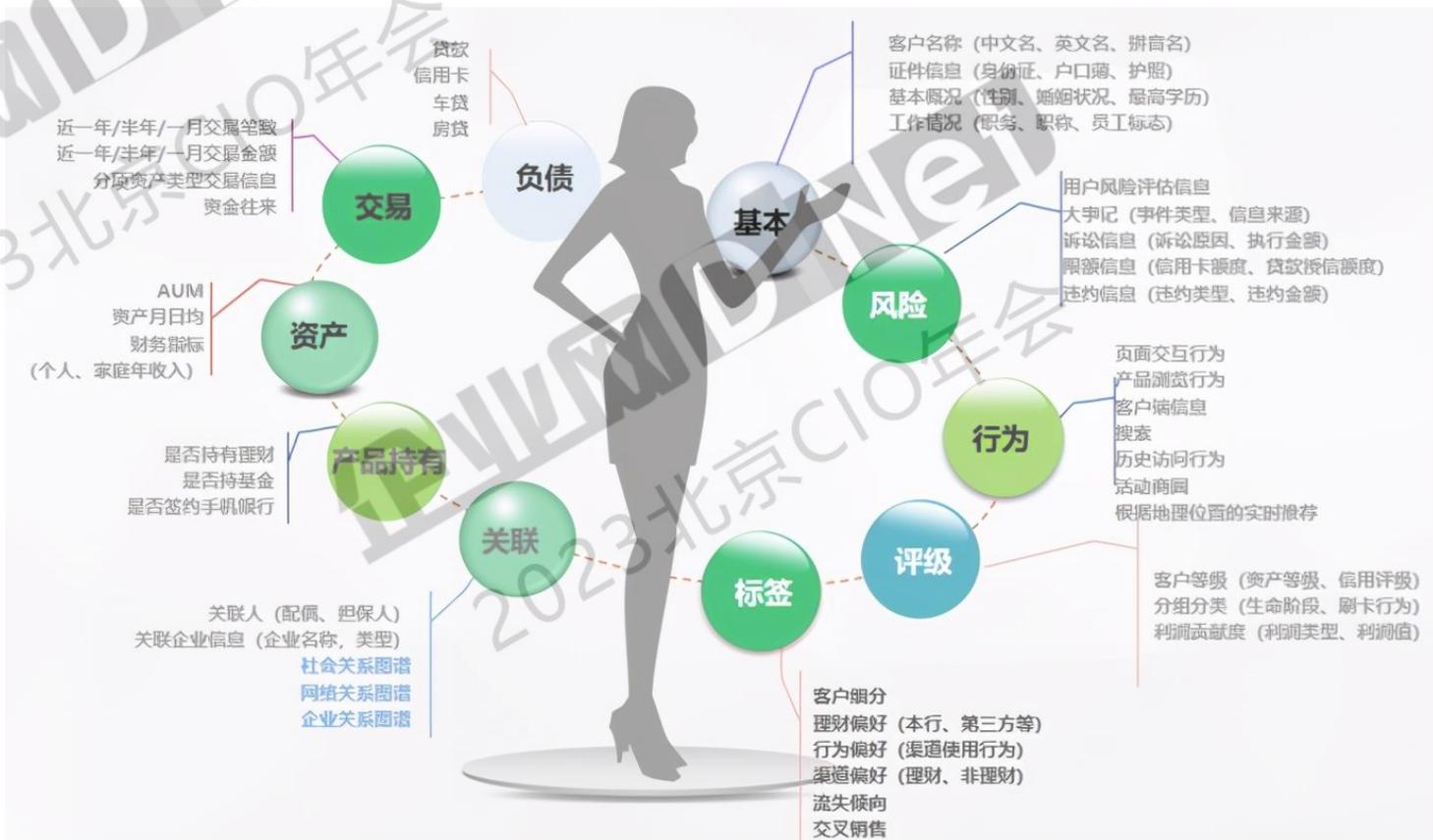
- 客户分析
  - 客户多模态标签
  - 客户AI助理
  - 客户多模态体验升级
- 产品分析
  - 需求多模态收集
  - 产品舆情收集
- 营销支撑
  - 虚拟营业厅建设
  - 全模态精准营销（时机、渠道、产品、话术等）
- 产品管理
  - 投诉的多模态监控
- 系统优化
  - 故障的精准预警
  - “自动驾驶”的运维
- 【思考】如何利用大模型，把企业数据分析重塑一遍？只是重塑吗？

模型类型	算法分类	挖掘主题	常见挖掘算法	应用方向	应用领域举例	应用模型
基于挖掘的模型	时间序列分析	对客户等进行预测	自回归综合移动平均 (ARIMA) 模型	科学决策	KPI指标预测	话务量预测模型 收入预测模型等
	回归		线性回归	科学决策	话务量预测、离网倾向	话务量预测模型 收入预测模型等
	分类预测	对客户等进行分类分析	Logistic回归	针对性客户挽留及营销	客户预警、客户细分	核心客户离网预警模型 学生客户识别模型
			决策树	针对性客户挽留及营销	流量营销	用户套餐匹配模型
			贝叶斯分类		文本分类等	网页内容分类模型
			神经网络		存量保有	客户离网预警模型
			支持向量机, SVM		集团市场预测	集团客户市场占有率测算
			KNN(K临近算法)		文本分类	内容分类模型
	隐马尔科夫模型	语义识别	文本挖掘、网页关键词分类			
	聚类	找出不同的群体	KMeans聚类	精细化管理, 针对性客户营销	渠道预警	养卡预警模型
			KModes聚类		细分市场客户身份识别	客户细分模型等
			最大期望算法	针对性客户挽留及营销	自然语言处理 (如分词)	中文分词模型
社交网络	找出社交网络圈	社交网络	针对性客户营销	细分市场客户识别及营销	学生客户识别模型	
关联规则挖掘	挖掘客户订购业务、产品间交叉销售等的关联性	关联规则, 典型的算法是Aprior算法协同过滤	针对性客户挽留及营销	业务交叉营销	客户业务关联模型	
序列分析		序列模式挖掘		针对性客户挽留及营销	交叉营销	应用精准推荐模型
			交叉营销、购物篮分析		客户业务关联模型 应用精准推荐模型	

# 【例】大模型下客户的数据分析“深度”

## □ 大模型分析更精准、更透彻

- chatGPT可以获取更多的对话等数据，分析客户的性格、喜好、出身、背景等等



# 打造类ChatGPT，赋能CHBN市场

打造类ChatGPT，赋能电信行业CHBN市场，加速各领域应用创新

## 个人移动市场

### • 5G新通话

实时多语种同声传译，主动生成内容，提升通话感知效果

### • 移动云盘

会议决策大脑  
云端的专业图书馆

### • AI秘书

形象与性格定制，个性化AI秘书

.....

## 家庭市场

### • 家庭康养

全科家庭医生  
永远陪伴在身边的亲人朋友

### • 家庭教育

不断生成题库的智能陪练  
自主创作教材的家庭教研  
能讲解能答疑的全年级、全学科  
实时辅导

### • 虚拟管家

智能家居环境调节管理

.....

## 政企市场

### • 智慧金融

全行业投资分析师、数据分析师  
智能个人投资顾问

### • 智慧城市

公文助手  
信息档案智能检索

### • 智慧政务

智能审核

.....

## 新兴市场

### • 音乐

风格没有限制、灵感永不枯竭的  
歌词创作大师

### • 新媒体

内容创作全流程自动化  
全平台内容运营工作室  
赛博影子写手

### • 数字人

多模态数字人

.....

# 个人移动市场应用创新

针对个人移动市场，ChatGPT促进5G新通话、客服、云盘等方向服务创新

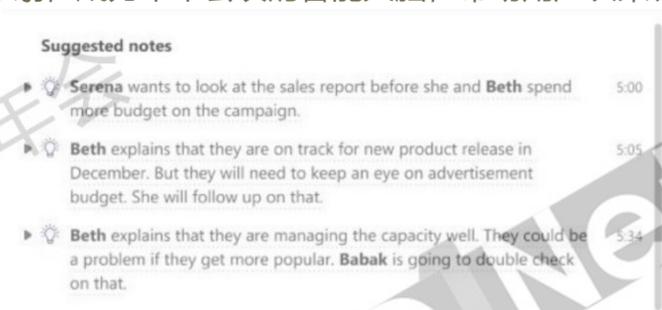
## 多语种翻译大师



在5G新通话的视频对话过程中，ChatGPT可不受语种限制，帮助将双方对话内容实时转写并以字幕形式向通话双方展示，并提供通话记录功能。在通话过程中，基于用户输入语音实时生成图像、文字内容，提升通话应用效果。

## 会议决策大脑

ChatGPT可与语音转写能力相结合，对会议对话文本自动总结核心要点+提炼任务目标，生成结构完整、目标清晰的会议总结与行动计划，成为未来会议的智能大脑，帮助用户决策、提高效率。



ZOOM已经开始进行内部试用

## 全能客服

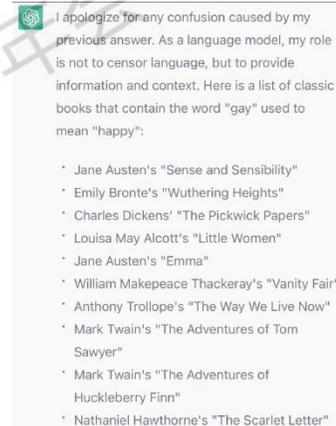
通过ChatGPT与语音转写、合成技术的结合，客服不再受限于一问一答，全能客服有能力回答用户的任何问题、舒缓用户各种情绪、揣测用户多样心里诉求，并能向用户推荐更符合其需求的各类产品。



## 云端专业图书馆

文献深层检索整理展示：ChatGPT有强大的复杂检索能力，可以通过云盘承载面向用户进行付费文献检索服务（如：搜索“minute”用于形容“微小的”，而不是“分钟”的十篇文章并形成列表和索引）。

公文模板大全：可以利用ChatGPT强大的编辑能力形成各类公文、论文、专业文档的规范模板，并向用户批量收费。



# 元宇宙场景类ChatGPT模型应用

元宇宙与生成式模型相辅相成技术互相促进，元宇宙能够实现线上化服务入口的行业，在元宇宙中进行场景融合，chatGPT能助力元宇宙场景生成与智能交互。——改写电商营销模式！

## 虚拟客服和售后支持

利用ChatGPT在元宇宙中创建虚拟客服和售后支持，让用户可以在元宇宙中咨询问题、提出意见和建议，并且通过ChatGPT进行智能回答和服务。



## 虚拟产品演示和销售

利用元宇宙在虚拟空间中创建虚拟产品展示和销售，通过ChatGPT进行自然语言交互，向用户展示产品特点 and 优势，提供智能化的购买建议和支持。



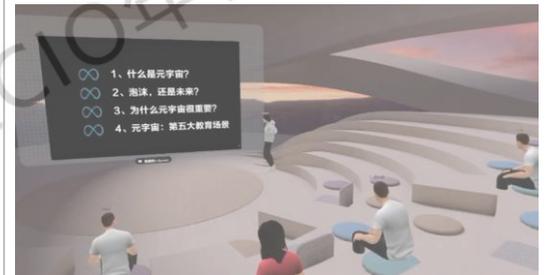
## 促进元宇宙场景高效生成

chatgpt赋能元宇宙场景复刻与全面演进：通过ai生成技术实现元宇宙对真实世界的复刻与发展推演，提升多场景元宇宙构建的效率。



## 元宇宙教育和实验

企业可以利用元宇宙进行虚拟培训和教育，通过ChatGPT进行智能化的交互和学习，提高学习效果和参与度。模拟各种实验的环境，提升实验的便捷性。



# 某数字人产品分析

## 某数字人产品



### 数字人生成平台

数字人生成：可以制作超写实或卡通数字人

### 数字人运营管理平台

统一管理数字人资产，监管数字人运行状态，实时问答管理资源管理、监控模块、多租户管理、效能分析以及平台接口管理

### 数字人视频平台

基于个性化数字人形象，为用户快速生成数字人播报视频，支持数字人视频的字幕设置、背景设置、输出设置等。视频可用于抖音、微博等新媒体平台进行品牌及产品宣发。

### 产品优势

1. 丰富的数字人形象
2. 快速生成数字人形象，通过照片及视频生成超写实形象
3. 精细的数字人还原（毛孔级、拟人化）
4. 多样的数字人驱动（AI及真人驱动）
5. 行业领先的AI算法【新一代超大AI模型NLP技术：基于检索和生成式复合对话系统，支持在线持续自动化更新维护知识库、主体人识别技术（声源分离）、语义驱动动作（准确率70%）】

## 四、几点思考

- 结构化数据如何用大模型？
  - 仅是“精度”提升？（LSTM到transformer?）
  - 分类更精准？（决策树、神经网络到大模型）
- 如何积累非结构化数据？
  - 数据如何存放和管理？是否建模？文件？
  - 新的收集终端和系统，视频、语音等
- 数据的合法、合规问题？算法版权问题？
  - 开源算法，如何开源训练过程？
  - 是否会泄露核心机密数据？
- 数据治理如何用大模型？
  - 类似漏洞扫描，检查不合规的数据质量
- 企业的大模型AI应用部署
  - NLP不能包治百病，如何外延？自编程？
  - YOLO算法如何大模型化？
- MAAS模式or自建？
  - 算力瓶颈、算法瓶颈、业务瓶颈、场景瓶颈等等
- 大模型如何驱动组织变革？
  - 原有工作流程，如何适配大模型的“智力”？
  - 部门、职位设定，如何跟上大模型的变革？
  - 数字人能否替代真实员工？
- 大模型能否驱动企业“自动驾驶”？
  - 更多数据、更准算法、更大算力、更深理解



企业网DINet  
2023北京CIO年会

# 大模型，改变世界

企业网DINet  
2023北京CIO年会

企业网DINet  
2023北京CIO年会

