ChatGPT专题

ChatGPT发展历程、原理、技术架构详解和产业未来

来源：陈巍谈芯，本文将介绍ChatGPT的**特点、功能、技术架构、局限、产业应用、投资机会和未来。**

作者：陈巍 博士，作者本人曾担任华为系自然语言处理（ NLP ）企业的首席科学家。

存算一体/GPU架构和AI专家，高级职称。中关村云计算产业联盟，中国光学工程学会专家，国际计算机学会（ACM）会员，中国计算机学会（CCF）专业会员。曾任AI企业首席科学家、存储芯片大厂3D NAND设计负责人，主要成就包括国内**首个大算力可重构存算处理器产品架构**（已在互联网大厂完成原型内测），**首个医疗领域专用AI处理器**（已落地应用），**首个RISC-V/x86/ARM平台兼容的AI加速编译器**（与阿里平头哥/芯来合作，已应用），**国内首个3D NAND芯片架构**与设计团队建立（与三星对标），**国内首个嵌入式闪存编译器**（与台积电对标，已平台级应用）

0，引言

先上参考网页或论文。专业的读者可以直接看paper。

ChatGPT: Optimizing Language Models for Dialogue ChatGPT: Optimizing Language Models for Dialogue
GPT论文：Language Models are Few-Shot Learners Language Models are Few-Shot Learners
InstructGPT论文：Training language models to follow instructions with human feedback Training language models to follow instructions with human feedback
huggingface解读RHLF算法：Illustrating Reinforcement Learning from Human Feedback (RLHF) Illustrating Reinforcement Learning from Human Feedback (RLHF)
RHLF算法论文：Augmenting Reinforcement Learning with Human Feedback cs.utexas.edu/~ai-lab/p
TAMER框架论文：Interactively Shaping Agents via Human Reinforcement cs.utexas.edu/~bradknox
PPO算法：Proximal Policy Optimization Algorithms Proximal Policy Optimization Algorithms

今年12月1日，OpenAI推出人工智能聊天原型**ChatGPT**，再次赚足眼球，为AI界引发了类似**AIGC**让艺术家失业的大讨论。

据报道，ChatGPT在开放试用的短短几天，就吸引了超过 **100 万互联网注册用户。**并且社交网络流传出各种询问或调戏ChatGPT的**有趣对话。**甚至有人将ChatGPT比喻为“**搜索引擎+社交软件**”的结合体，能够在实时互动的过程中获得问题的合理答案。

ChatGPT 是一种专注于对话生成的**语言模型。**它能够根据用户的文本输入，产生相应的智能回答。这个回答可以是简短的词语，也可以是**长篇大论。**其中GPT是Generative Pre-trained Transformer（生成型预训练变换模型）的缩写。

通过学习大量现成文本和对话集合（例如Wiki），ChatGPT能够像人类那样即时对话，流畅的回答各种问题。（当然回答速度比人还是慢一些）无论是英文还是其他语言（例如中文、韩语等），从回答历史问题，到写故事，甚至是**撰写商业计划书和行业分析，**“几乎”无所不能。甚至有程序员贴出了ChatGPT进行程序修改的对话。

ChatGPT也可以与其他AIGC模型**联合使用**，获得更加炫酷实用的功能。例如上面通过对话生成客厅设计图。这极大加强了AI应用与客户对话的能力，使我们看到了**AI大规模落地的曙光。**

1，ChatGPT的传承与特点

1.1 OpenAI家族

我们首先了解下**OpenAI**是哪路大神。

OpenAI总部位于旧金山，由特斯拉的马斯克、Sam Altman及其他投资者在2015年共同创立，目标是开发造福全人类的AI技术。而马斯克则在2018年时因公司发展方向分歧而离开。

此前，OpenAI 因推出 **GPT系列自然语言处理模型**而闻名。从2018年起，OpenAI就开始发布生成式预训练语言模型GPT（Generative Pre-trained Transformer），可用于生成文章、代码、机器翻译、问答等各类内容。

每一代GPT模型的参数量都**爆炸式增长**，堪称“越大越好”。2019年2月发布的GPT-2参数量为15亿，而2020年5月的GPT-3，参数量达到了1750亿。

模型	发布时间	参数量	预训练数据量
GPT-1	2018年6月	1.17亿	约5GB
GPT-2	2019年2月	15亿	40G
GPT-3	2020年5月	1750亿	45TB
ChatGPT	2022年11月	千亿级?	百T级?

GPT家族主要模型对比

1.2 ChatGPT的主要特点

ChatGPT 是基于GPT-3.5（Generative Pre-trained Transformer 3.5）架构开发的对话AI模型，是InstructGPT 的兄弟模型。ChatGPT很可能是OpenAI 在GPT-4 正式推出之前的演练，或用于**收集大量对话数据**。

ChatGPT的主要特点

OpenAI使用 **RLHF**（Reinforcement Learning from Human Feedbac，人类反馈强化学习）技术对 ChatGPT 进行了训练，且加入了更多**人工监督**进行微调。

此外，ChatGPT 还具有以下特征：

1）可以主动承认自身错误。若用户指出其错误，模型会听取意见并优化答案。

2）ChatGPT 可以质疑不正确的问题。例如被询问 “哥伦布 2015 年来到美国的情景” 的问题时，机器人会说明哥伦布不属于这一时代并调整输出结果。

3）ChatGPT 可以承认自身的无知，承认对专业技术的不了解。

4）支持连续多轮对话。

与大家在生活中用到的各类智能音箱和“**人工智障**“不同，ChatGPT在对话过程中会记忆先前使用者的对话讯息，即上下文理解，以回答某些假设性的问题。ChatGPT可实现**连续对话**，极大的提升了**对话交互模式**下的用户体验。

对于准确翻译来说（尤其是中文与人名音译），ChatGPT离完美还有一段距离，不过在文字流畅度以及辨别特定人名来说，与其他网络翻译工具相近。

由于 ChatGPT是一个大型语言模型，目前还并**不具备网络搜索功能**，因此它只能基于2021年所拥有的数据集进行回答。例如它不知道2022年世界杯的情况，也不会像苹果的Siri那样回答今天天气如何、或帮你搜索信息。如果ChatGPT能上网自己寻找学习语料和搜索知识，估计又会有更大的突破。

即便学习的知识有限，ChatGPT 还是能回答脑洞大开的人类的许多**奇葩问题**。为了避免ChatGPT染上恶习， ChatGPT 通过**算法屏蔽**，减少有害和欺骗性的训练输入。，查询通过适度 API 进行过滤，并驳回潜在的种族主义或性别歧视提示。

2，ChatGPT/GPT的原理

2.1 NLP

NLP/NLU领域**已知局限**包括对重复文本、对高度专业的主题的误解，以及**对上下文短语的误解。**

对于人类或AI，通常需接受多年的训练才能正常对话。NLP类模型不仅要理解单词的含义，还要理解如何造句和给出上下文有意义的回答，甚至使用合适的俚语和专业词汇。

NLP技术的应用领域

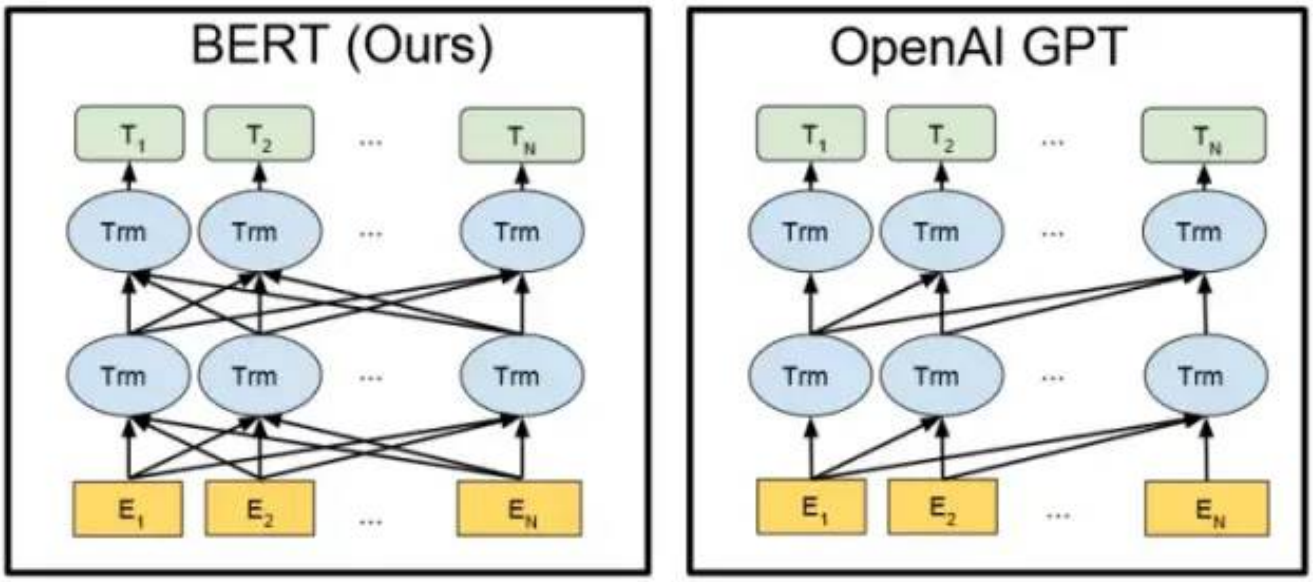
本质上，作为ChatGPT基础的GPT-3或GPT-3.5 是一个超大的统计语言模型或顺序文本预测模型。

2.2 GPT v.s.BERT

与BERT模型类似，ChatGPT或GPT-3.5都是根据输入语句，根据语言/语料概率来自动生成回答的每一个字（词语）。从数学或从机器学习的角度来看，语言模型是对词语序列的**概率相关性分布的建模**，即利用已经说过的语句（语句可以视为数学中的向量）作为输入条件，预测下一个时刻不同语句甚至语言集合出现的概率分布。

ChatGPT 使用来自人类反馈的强化学习进行训练，这种方法通过人类干预来增强机器学习以获得更好的效果。在训练过程中，人类训练者扮演着用户和人工智能助手的角色，并通过**近端策略优化算法**进行微调。

由于ChatGPT更强的性能和海量参数，它包含了更多的主题的数据，能够处理更多小众主题。ChatGPT现在可以进一步处理回答问题、撰写文章、文本摘要、语言翻译和生成计算机代码等任务。



BERT与GPT的技术架构（图中En为输入的每个字，Tn为输出回答的每个字）

3，ChatGPT的技术架构

3.1 GPT家族的演进

说到ChatGPT，就不得不提到**GPT家族**。

ChatGPT之前有几个知名的兄弟，包括GPT-1、GPT-2和GPT-3。这几个兄弟一个比一个个头大，ChatGPT与GPT-3更为相近。

ChatGPT与GPT 1-3的技术对比

GPT家族与BERT模型都是知名的NLP模型，都基于Transformer技术。GPT-1只有**12个**Transformer层，而到了GPT-3，则增加到**96层**。

3.2 人类反馈强化学习

InstructGPT/GPT3.5（ChatGPT的前身）与GPT-3的主要区别在于，新加入了被称为**RLHF**（Reinforcement Learning from Human Feedback，人类反馈强化学习）。这一训练范式增强了人类对模型输出结果的调节，并且对结果进行了更具理解性的排序。

在InstructGPT中，以下是“goodness of sentences”的评价标准。

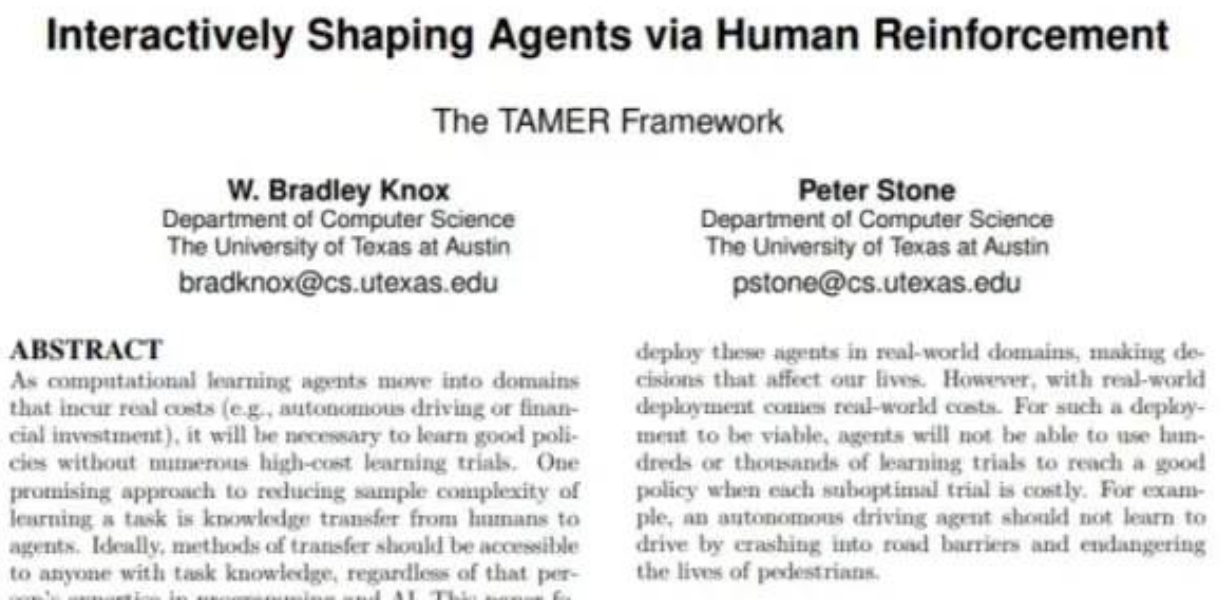
真实性：是虚假信息还是误导性信息？

无害性：它是否对人或环境造成身体或精神上的伤害？

有用性：它是否解决了用户的任务？

3.3 TAMER框架

这里不得不提到**TAMER**（Training an Agent Manually via Evaluative Reinforcement，评估式强化人工训练代理）这个框架。该框架将人类标记者引入到Agents的学习循环中，可以通过人类向Agents提供**奖励反馈**（即指导Agents进行训练），从而快速达到训练任务目标。



Interactively Shaping Agents via Human Reinforcement

The TAMER Framework

W. Bradley Knox
Department of Computer Science
The University of Texas at Austin
bradknox@cs.utexas.edu

Peter Stone
Department of Computer Science
The University of Texas at Austin
pstone@cs.utexas.edu

ABSTRACT

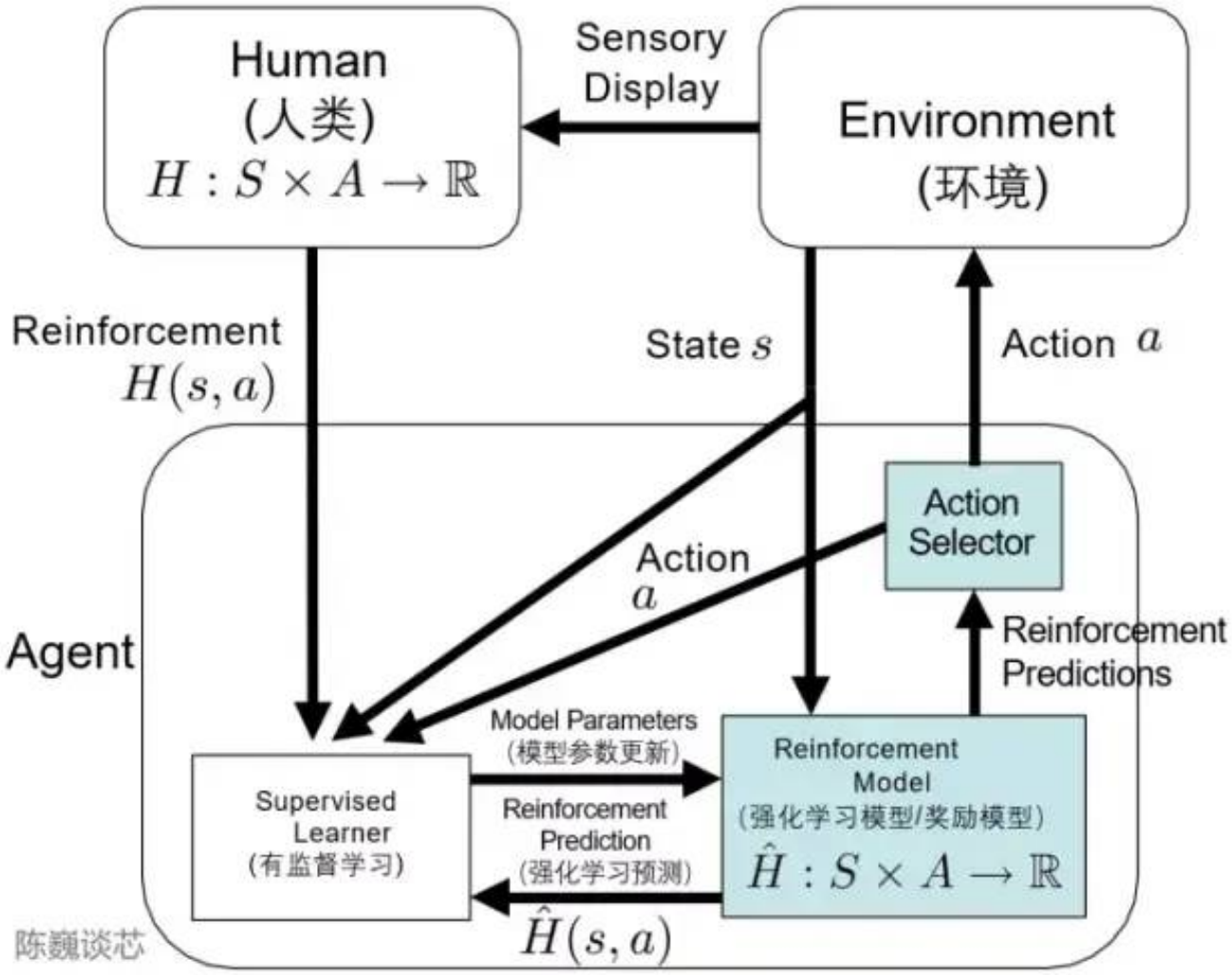
As computational learning agents move into domains that incur real costs (e.g., autonomous driving or financial investment), it will be necessary to learn good policies without numerous high-cost learning trials. One promising approach to reducing sample complexity of learning a task is knowledge transfer from humans to agents. Ideally, methods of transfer should be accessible to anyone with task knowledge, regardless of that per-

deploy these agents in real-world domains, making decisions that affect our lives. However, with real-world deployment comes real-world costs. For such a deployment to be viable, agents will not be able to use hundreds or thousands of learning trials to reach a good policy when each suboptimal trial is costly. For example, an autonomous driving agent should not learn to drive by crashing into road barriers and endangering the lives of pedestrians.

TAMER框架论文

引入人类标记者的主要目的是**加快训练速度**。尽管强化学习技术在很多领域有突出表现，但是仍然存在着许多不足，例如**训练收敛速度慢，训练成本高**等特点。特别是现实世界中，许多任务的探索成本或数据获取成本很高。如何加快训练效率，是如今强化学习任务待解决的重要问题之一。

而TAMER则可以将人类标记者的知识，以奖励信反馈的形式训练Agent，加快其快速收敛。
TAMER不需要标记者具有**专业知识**或编程技术，语料成本更低。通过TAMER+RL（强化学习），借助人类标记者的反馈，能够增强从**马尔可夫决策过程** (MDP) 奖励进行强化学习 (RL) 的过程。



TAMER架构在强化学习中的应用

具体实现上，人类标记者扮演对话的用户和人工智能助手，提供对话样本，让模型生成一些回复，然后标记者会对回复选项**打分排名**，将更好的结果反馈回模型中，Agents同时从两种反馈模式中学习——人类强化和马尔可夫决策过程奖励作为一个整合的系统，通过**奖励策略**对模型进行微调并持续迭代。

在此基础上，ChatGPT 可以比 GPT-3 更好的理解和完成人类语言或指令，模仿人类，提供连贯的有逻辑的文本信息的能力。

3.4 ChatGPT的训练

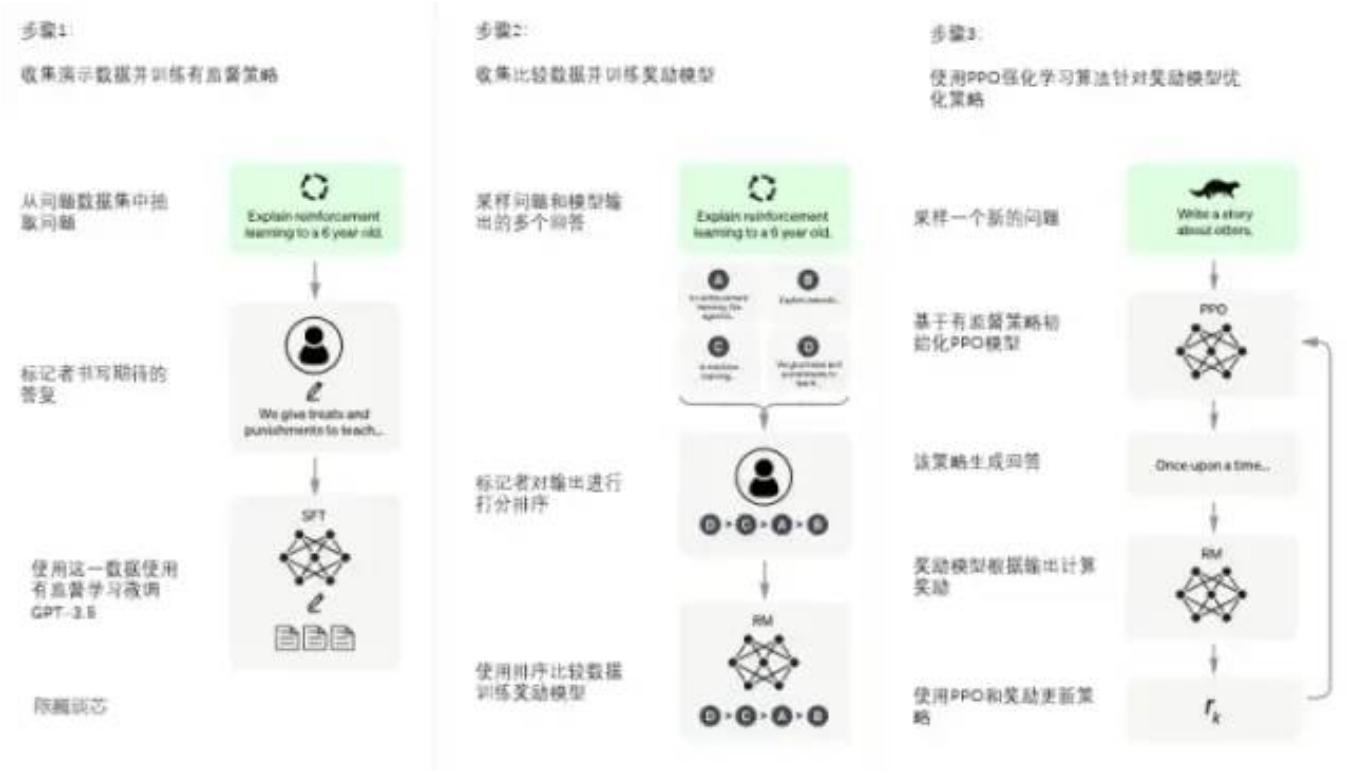
ChatGPT的训练过程分为以下三个阶段：

第一阶段：训练**监督策略**模型

GPT 3.5本身很难理解人类不同类型指令中蕴含的不同意图，也很难判断生成内容是否是高质量的结果。为了让GPT 3.5初步具备理解指令的意图，首先会在数据集中随机抽取问题，由人类标注人

员，给出高质量答案，然后用这些人工标注好的数据来**微调 GPT-3.5模型**（获得SFT模型，Supervised Fine-Tuning）。

此时的SFT模型在遵循指令/对话方面已经优于 GPT-3，但不一定符合人类偏好。



ChatGPT模型的训练过程

第二阶段：训练**奖励模型**（Reward Mode，RM）

这个阶段的主要是通过人工标注训练数据（约33K个数据），来训练回报模型。在数据集中随机抽取问题，使用第一阶段生成的模型，对于每个问题，生成多个不同的回答。人类标注者对这些结果综合考虑给出**排名顺序**。这一过程类似于教练或老师辅导。

接下来，使用这个排序结果数据来训练奖励模型。对多个排序结果，**两两组合**，形成多个训练数据对。RM模型接受一个输入，给出评价回答质量的分数。这样，对于一对训练数据，调节参数使得高质量回答的打分比低质量的打分要高。

第三阶段：采用**PPO**（Proximal Policy Optimization，近端策略优化）**强化学习来优化策略。**

PPO的核心思路在于将Policy Gradient中On-policy的训练过程转化为Off-policy，即将在线学习转化为离线学习，这个转化过程被称之为**Importance Sampling**。这一阶段利用第二阶段训练好的奖励模型，靠奖励打分来更新预训练模型参数。在数据集中随机抽取问题，使用PPO模型生成回答，并用上一阶段训练好的RM模型给出质量分数。把回报分数依次传递，由此产生策略梯度，通过强化学习的方式以更新PPO模型参数。

如果我们不断重复第二和第三阶段，通过**迭代**，会训练出更高质量的ChatGPT模型。

4，ChatGPT的局限

只要用户输入问题，ChatGPT 就能给予回答，是否意味着我们不用再拿关键词去喂 Google或百度，就能立即获得想要的答案呢？

尽管ChatGPT表现出出色的上下文对话能力甚至编程能力，完成了大众对人机对话机器人（ChatBot）从“人工智障”到“有趣”的印象改观，我们也要看到，ChatGPT技术仍然有一些局限性，还在不断的进步。

1）ChatGPT在其未经**大量语料**训练的领域缺乏“**人类常识**”和引申能力，甚至会一本正经的“胡说八道”。ChatGPT在很多领域可以“**创造答案**”，但当用户寻求正确答案时，ChatGPT也有可能给出有误导的回答。例如让ChatGPT做一道小学应用题，尽管它可以写出一长串计算过程，但最后答案却是错误的。

2）ChatGPT无法处理复杂冗长或者**特别专业的语言结构**。对于来自金融、自然科学或医学等非常专业领域的问题，如果没有进行足够的语料“喂食”，ChatGPT可能无法生成适当的回答。

3）ChatGPT需要非常**大量的算力（芯片）**来支持其训练和部署。抛开需要大量语料数据训练模型不说，在目前，ChatGPT在应用时仍然需要大算力的服务器支持，而这些服务器的成本是普通用户无法承受的，即便数十亿个参数的模型也需要惊人数量的**计算资源**才能运行和训练。，如果面向真实搜索引擎的数以亿记的用户请求，如采取目前通行的免费策略，任何企业都难以承受这一成本。因此对于普通大众来说，还需等待更轻量型的模型或**更高性价比的算力平台**。

4）ChatGPT还没法在线的把**新知识**纳入其中，而出现一些新知识就去重新预训练GPT模型也是不现实的，无论是训练时间或训练成本，都是普通训练者难以接受的。如果对于新知识采取在线训练的模式，看上去可行且语料成本相对较低，但是很容易由于新数据的引入而导致对原有知识的灾难性遗忘的问题。

5）ChatGPT仍然是**黑盒模型**。目前还未能对ChatGPT的内在算法逻辑进行分解，因此并不能保证ChatGPT不会产生攻击甚至伤害用户的表述。

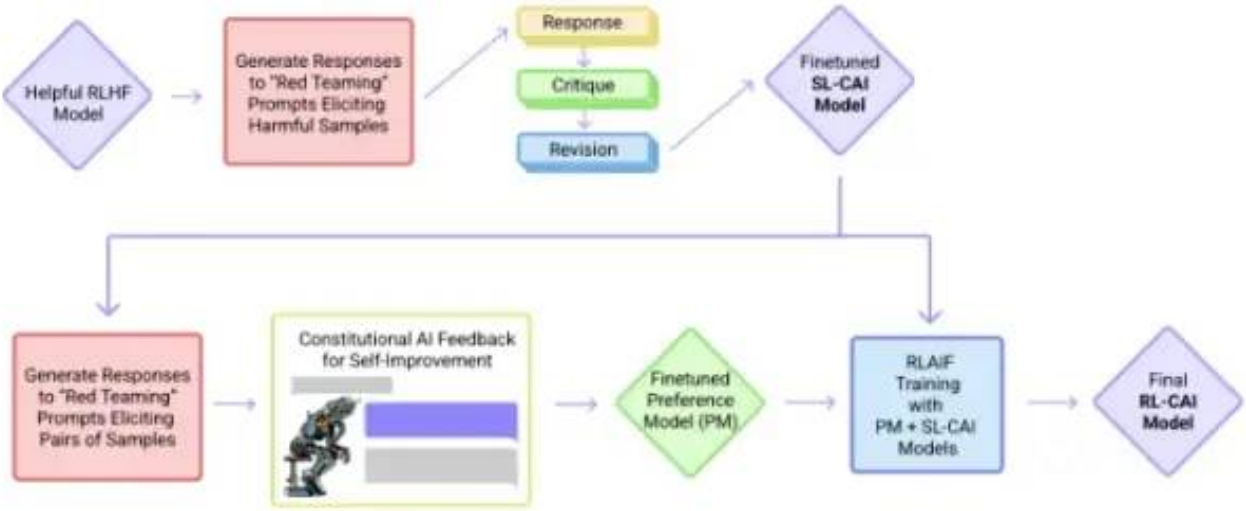
当然，**瑕不掩瑜**，有工程师贴出了要求ChatGPT写verilog代码（芯片设计代码）的对话。可以看出ChatGPT水平已经超出一些verilog初学者了。

5，ChatGPT的未来改进方向

5.1 减少人类反馈的RLAIF

2020年底，OpenAI前研究副总裁Dario Amodei带着10名员工创办了一个人工智能公司Anthropic。Anthropic 的创始团队成员，大多为 OpenAI 的早期及核心员工，参与过OpenAI的GPT-3、多模态神经元、人类偏好的强化学习等。

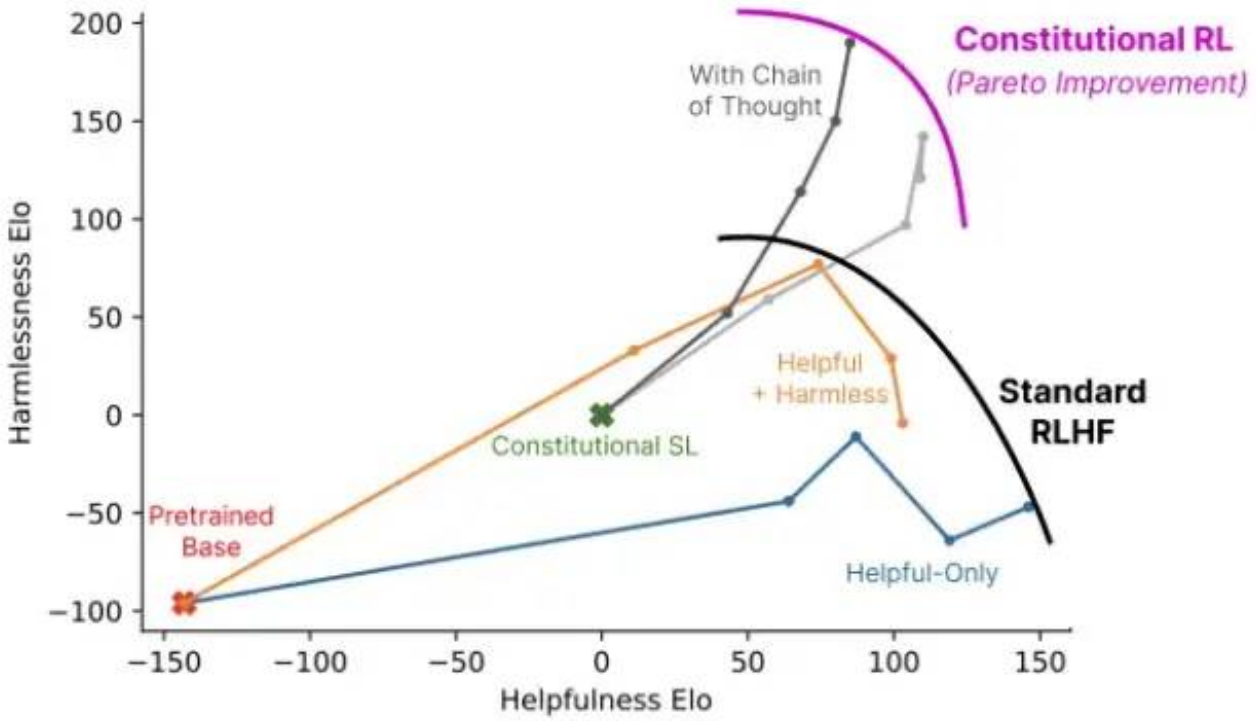
2022年12月，Anthropic再次发表论文《Constitutional AI: Harmlessness from AI Feedback》介绍人工智能模型Claude。（arxiv.org/pdf/2212.0807）



CAI模型训练过程

Claude 和 ChatGPT 都依赖于强化学习(RL)来训练偏好（preference）模型。CAI（Constitutional AI）也是建立在RLHF的基础之上，不同之处在于，CAI的排序过程使用模型（而非人类）对所有生成的输出结果提供一个初始排序结果。

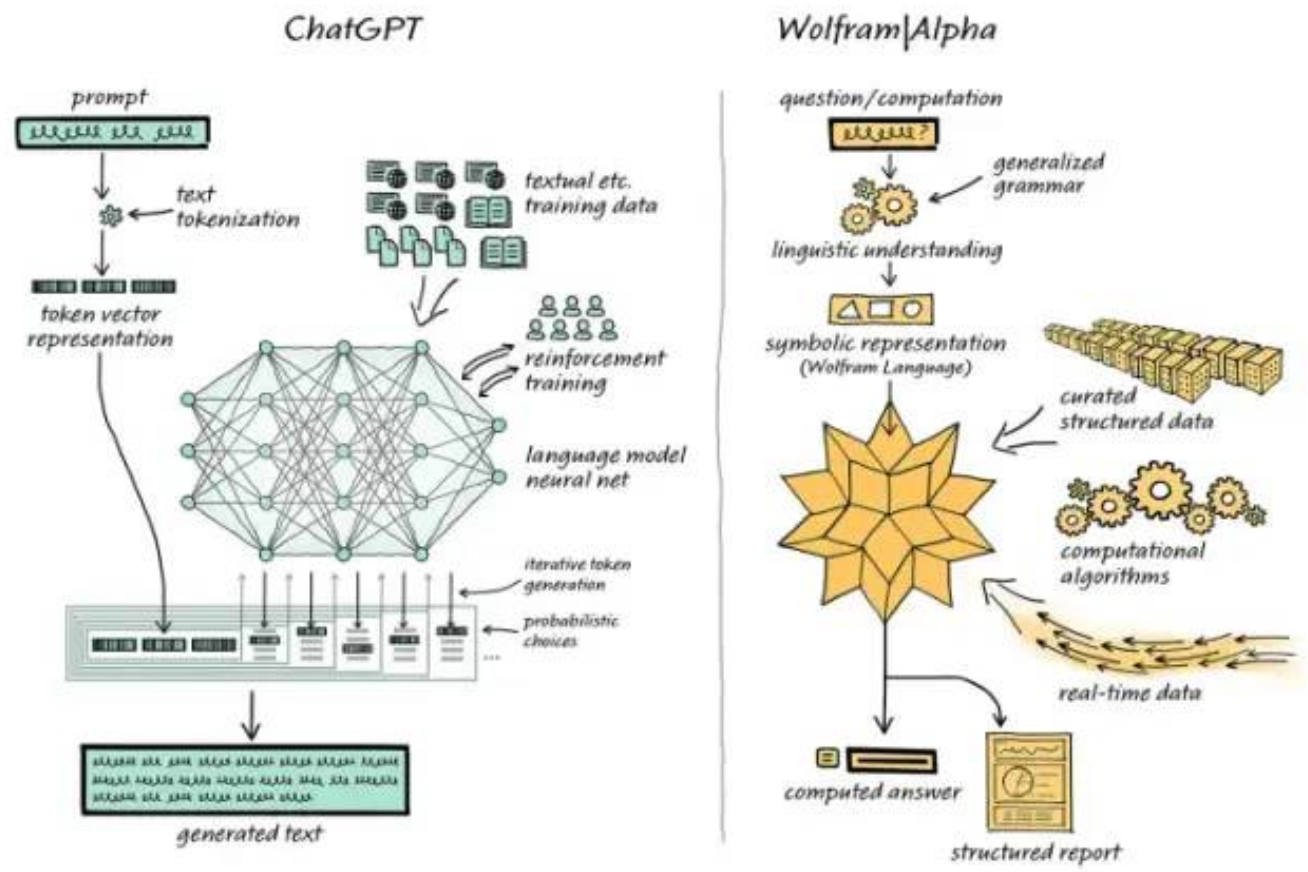
CAI用人工智能反馈来**代替人类对表达无害性的偏好**，即RLAIF，人工智能根据一套constitution原则来评价回复内容。



5.2 补足数理短板

ChatGPT虽然对话能力强，但是在数理计算对话中容易出现一本正经胡说八道的情况。

计算机学家Stephen Wolfram 为这一问题提出了解决方案。Stephen Wolfram 创造了的 Wolfram 语言和计算知识搜索引擎 Wolfram | Alpha，其后台通过Mathematica实现。



ChatGPT与Wolfram | Alpha结合处理梳理问题

在这一结合体系中，ChatGPT 可以像人类使用 Wolfram|Alpha 一样，与 Wolfram|Alpha “对话”，Wolfram|Alpha 则会用其**符号翻译能力**将从 ChatGPT 获得的自然语言表达“翻译”为对应的符号化计算语言。在过去，学术界在 ChatGPT 使用的这类 “统计方法” 和 Wolfram|Alpha 的 “符号方法” 上一直存在路线分歧。但如今 ChatGPT 和 Wolfram|Alpha 的互补，给NLP领域提供了更上一层楼的可能。

ChatGPT 不必生成这样的代码，只需生成常规自然语言，然后使用 Wolfram|Alpha 翻译成精确的 Wolfram Language，再由底层的Mathematica进行计算。

5.3 ChatGPT的小型化

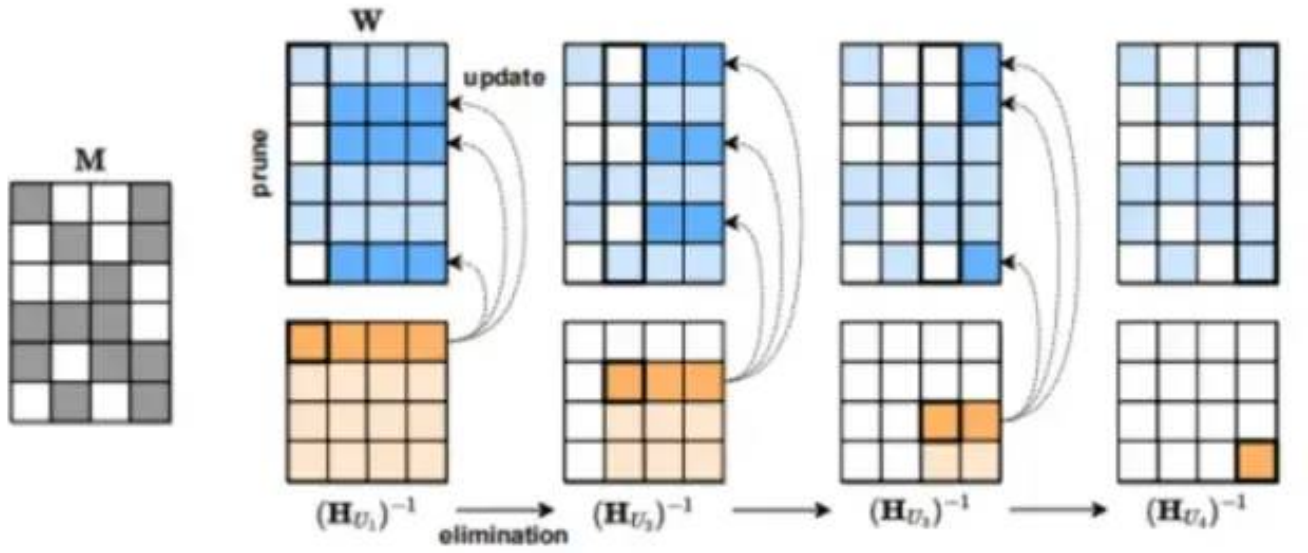
虽然ChatGPT很强大，但其模型大小和使用成本也让很多人望而却步。

有三类**模型压缩**（model compression）可以降低模型的大小和成本。

第一种方法是**量化**（quantization），即降低单个权重的数值表示的精度。比如Tansformer从FP32降到INT8对其精度影响不大。

第二种模型压缩方法是**剪枝**（pruning），即删除网络元素，包括从单个权重（非结构化剪枝）到更高粒度的组件如权重矩阵的通道。这种方法在视觉和较小规模的语言模型中有效。

第三种模型压缩方法是**稀疏化。**例如奥地利科学技术研究所 (ISTA)提出的SparseGPT（arxiv.org/pdf/2301.0077）可以将 GPT 系列模型单次剪枝到 50% 的稀疏性，而无需任何重新训练。对 GPT-175B 模型，只需要使用单个 GPU 在几个小时内就能实现这种剪枝。



SparseGPT 压缩流程

6 ChatGPT的产业未来与投资机会

6.1 AIGC

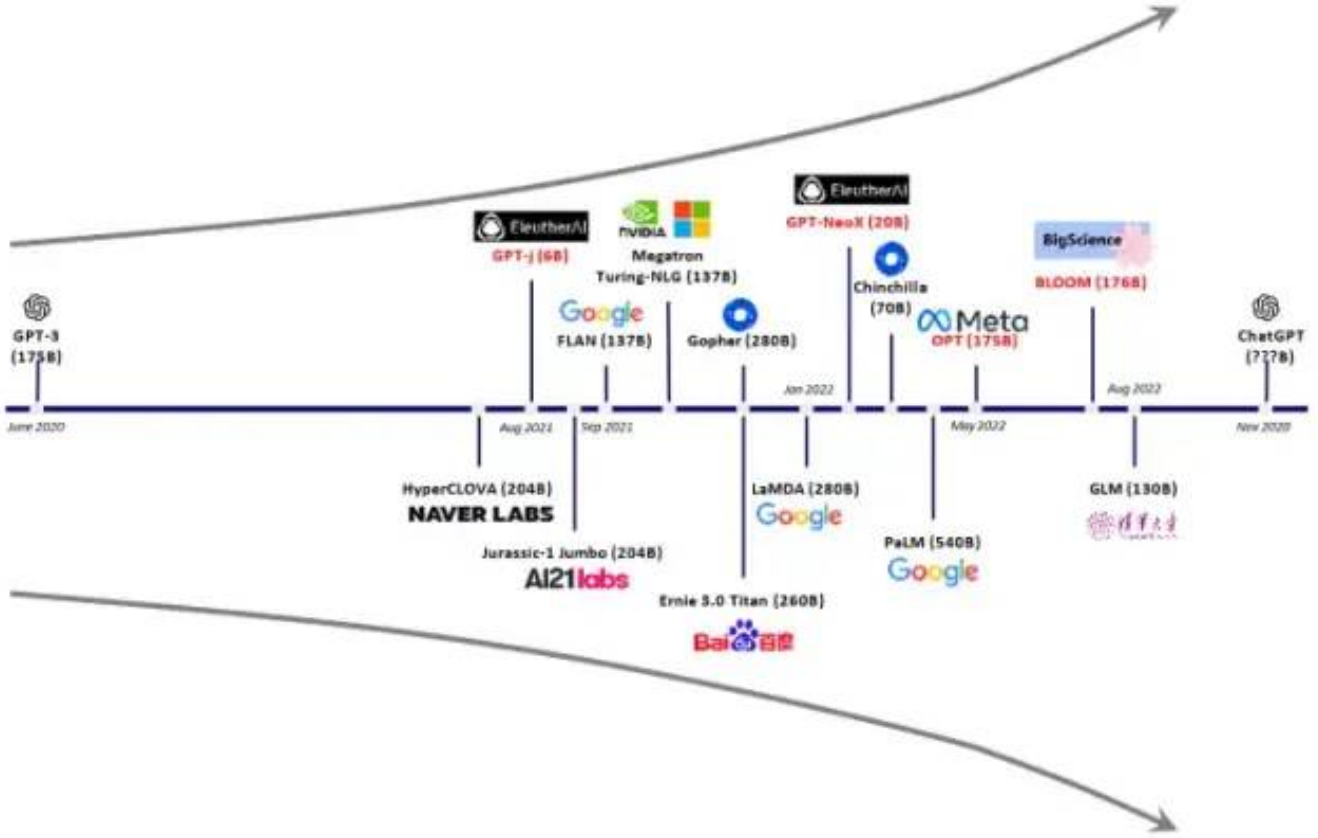
说到ChaGPT不得不提AIGC。

AIGC即利用人工智能技术来生成内容。与此前Web1.0、Web2.0时代的UGC（用户生产内容）和PGC（专业生产内容）相比，代表人工智能构思内容的AIGC，是新一轮内容**生产方式变革**，而且AIGC内容在Web3.0时代也将出现**指数级增长。**

ChatGPT 模型的出现对于文字/语音模态的 AIGC 应用具有重要意义，会对AI产业上下游**产生重大影响。**

6.2 受益场景

从下游相关受益应用来看，包括但不限于**无代码编程**、小说生成、对话类搜索引擎、语音陪伴、语音工作助手、对话虚拟人、**人工智能客服**、机器翻译、**芯片设计**等。从上游增加需求来看，包括**算力芯片**、数据标注、自然语言处理（NLP)等。



大模型呈爆发态势（更多的参数/更大的算力芯片需求）

随着算法技术和算力技术的不断进步，ChatGPT也会进一步走向更先进功能更强的版本，在越来越多的领域进行应用，为人类生成更多更美好的对话和内容。