GPT发展：

自然语言处理是研究如何利用计算机技术对语言文本进行处理、加工和转换的一门学科［１］。由于该学科在理论上面临巨大的挑战，而其技术应用前景极其广泛，因此被誉为人工智能皇冠上的明珠。

随着自然语言处理技术的进步，语言建模（ＬａｎｇｕａｇｅＭｏｄｅｌ，ＬＭ）技术已经经历了从最初的统计语言模型（ＳｔａｔｉｓｔｉｃａｌＬａｎｇｕａｇｅＭｏｄｅｌ，ＳＬＭ）发展至神经网络语言模型（ＮｅｕｒａｌＬａｎｇｕａｇｅＭｏｄｅｌ，ＮＬＭ），再到预训练语言模型（Ｐｒｅ－ｔｒａｉｎｅｄＬａｎｇｕａｇｅＭｏｄｅｌ，ＰＬＭ）的演进过程［２］。

通过扩展预训练语言模型得到的大模型将语言建模技术推向了一个新的发展高度，其发展速度之快、模型能力之强和通用程度之高，都远远超越任何一个历史时期的任何一种方法，令人瞠目。

大语言模型（ＬａｒｇｅＬａｎｇｕａｇｅＭｏｄｅｌ，ＬＬＭ）通常指参数量为百亿级甚至更大规模的神经网络组成的语言模型，它采用自监督学习方式利用大量未标注数据训练而成。

尽管在扩展预训练语言模型时主要是增大模型参数量（模型架构和训练方法基本不变），但这些大规模的预训练语言模型表现出与较小规模的预训练语言模型不同的行为。

该模型是在生成型预训练模型（ＧＰＴ）［９］系列模型的基础之上，通过指令微调并从调试人员的反馈中强化学习训练建立起来的。同其他大模型相比，由于其采用了指令微调和ＲＬＨＦ等技术，ＣｈａｔＧＰＴ具有更加强大的理解人类用户意图和偏好的能力，既可以根据指令生成高质量的回复，也可以针对不恰当的输入拒绝回答，甚至更正对话中的错误。

ＣｈａｔＧＰＴ作为通用模型在几乎所有的自然语言处理任务上都展示了较好的性能和优势。强大的通用处理能力，准确的用户意图理解能力和“随机应变”的交互能力。ＣｈａｔＧＰＴ 几乎能够准确理解和把握人类用户的意图，且能够根据人类的指令和上下文进行自然流畅的人机交互，可随时根据用户的问题和反馈修改模型自身的输出，其看似缜密的推理过程和滴水不漏的应答能力都是已有模型所未能做到的。尽管有时候它也会胡说八道，但其表现仍然一本正经。

自然语言处理 (natural language processing, NLP) 的目标是实现人机之的有效通信, 使得计算机既能够理解自然语言的意义, 也能以自然语言文本来表达意图与思想. 自然语言处理是人工智能领域的重要研究方向, 融合了语言学、计算机科学、机器学习、数学、认知心理学等多个学科领域的知识. 总体而言, 自然语言处理包含自然语言理解和自然语言生成两个主要方面, 研究内容覆盖的粒度包括字、词、短语、句子、段落和篇章等多种层次. 由于语言的复杂性, 高精度、高鲁棒、可解释的通用自然语言处理系统目前还没有成熟解决方案, 仍需进行长期研究.

自然语言处理经历了3 个历史阶段. 基于规则的方法、基于机器学习的方法和基于深度学习的方法. 基于规则的方法其核心思想是通过使用语言学知识, 如词汇和形式文法等, 来制定规则, 从而完成自然语言处理任务.基于机器学习的自然语言处理方法主要采用有监督分类算法, 将自然语言处理任务转化为某种分类任务, 在此基础上, 根据任务特性构建特征表示, 并构建大规模的有标注语料进行模型训练. 深度学习方法则通过构建深度模型, 将特征学习和预测模型融合在一起. 通过优化算法, 模型可以自动地学习出良好的特征表示, 并基于此进行结果预测.

第一个阶段是基础大模型训练, 目的是完成长距离语言模型的预训练, 并使得模型具备代码生成的能力. 第二个阶段是指令微调, 通过给定指令对模型进行微调, 以使其能够完成各种任务. 最后一个阶段是类人对齐, 通过加入更多的人工提示词, 并使用监督学习和基于强化学习的方法, 使得模型的输出更加贴合人类需求. 其中, 基础的大模型预训练给模型带来了语言生成能力, 并且具备了上下文学习 (in-contextlearning) 能力和世界知识, 包括事实性知识 (factual knowledge) 和常识 (commonsense). 指令微调使得模型能够更好地遵循人的指令, 增强了模型的零样本能力. 类人对齐则使得模型的输出符合人类的期望, 不仅可以忠实地输出对人类有用的结果, 也避免了输出有害内容并拒绝模型知识范围之外的问题.

1. 无监督预训练技术

在第一代ＧＰＴ－１模型［２］被提出之前，大多数深度学习方法倾向于采用有监督的训练和学习方式，需要大量人工标注的高质量数据，然而巨大的数据标注成本极大地限制了这类方法在各项任务的性能上限和在任务之间的通用性。ＧＰＴ－１及同时期的其他深度学习模型所解决的两个主要问题为：

（１）如何利用大规模开放获取且无需标注的数据进行模型的训练，也就是通常所说的无监督预训练。

（２）如何将从大规模无标注数据上学习到的模型参数，经过简单的再训练方式适应到下游任务，即基于预训练和精调的迁移学习，而精调的过程通常采用有监督的方式，因此也称为有监督精调（ＳｕｐｅｒｖｉｓｅｄＦｉｎｅ－ｔｕｎｉｎｇ，ＳＦＴ）。

2.无监督多任务学习技术

与ＧＰＴ－１模型采用的预训练—精调范式来完成多个下游任务不同，ＧＰＴ－２模型［３］更加侧重于模型本身的通用性，即在不改变模型结构的基础上，通过引入额外的输入信息来使模型适应到多个不同的下游任务，即多任务学习。简而言之，ＧＰＴ－２并未针对ＧＰＴ－１的模型结构进行调整和创新，而是在ＧＰＴ－１的基础上引入要完成的任务信息（任务名称或标识任务的提示语）作为模型生成的输入条件，并继续增大预训练的数据规模以及模型本身的参数量。在使用过程中，ＧＰＴ－２模型在下游任务中不进行任何训练或精调，不再根据下游任务的数据进行参数上的优化，而是根据给定的输入和任务信息自行理解并完成任务。

3.少样本学习技术

ＧＰＴ－３［４］使用了与 ＧＰＴ－２类似的模型结构，但进一步增加了模型的规模。为了探索模型规模对于性能的影响，ＯｐｅｎＡＩ一共训练了８个不同参数量的模型，并将其中具有最大参数量，即１７５０亿参数（１７５Ｂ）的模型称为ＧＰＴ－３。ＧＰＴ－３最显著的特点就是规模大。具体体现在两方面，一方面是模型具有的参数量大，其结构上具有９６个Ｔｒａｎｓｆｏｒｍｅｒ解码层，每一层有９６个１２８维的注意力头（ＡｔｔｅｎｔｉｏｎＨｅａｄ），词嵌入的维度也达到了１２２８８；另一方面，与大规模参数对应的是训练过程中使用到的数据量大，达到了４５ＴＢ。在这样的模型参数量与数据量的情况下，ＧＰＴ－３在多个任务上均展现出了非常优异的性能，延续 ＧＰＴ－２将无监督模型应用到有监督任务的思想中，ＧＰＴ－３在少样本（Ｆｅｗ－ｓｈｏｔ）、单样本（Ｏｎｅ－ｓｈｏｔ）、甚至零样本（Ｚｅｒｏ－ｓｈｏｔ）设置下，完成相应任务的性能上却得到了显著的提升，因而首次展现了大规模预训练语言模型的能力。

4. 人类偏好对齐技术

尽管ＧＰＴ－３展现出了强大的少样本学习的能力，但是其完成任务的方式仍然需要给出若干个任务示例数据。例如，如果我们希望 ＧＰＴ－３完成一个加法运算，我们需要给出几个加法运算的完整示例，这显然与人类的使用习惯不符。因而，与人类习惯对齐的ＣｈａｔＧＰＴ的出现，更加符合人机交互的自然形式，即人机对话的形式。遗憾的是，ＯｐｅｎＡＩ并没有公布 ＣｈａｔＧＰＴ的技术实现细节。实际上，ＯｐｅｎＡＩ从 ＧＰＴ－３开始就已经 闭 源，因 此 大 家 都 戏 称 ＯｐｅｎＡＩ 可 以 改 名ＣｌｏｓｅＡＩ。但是，结合 ＯｐｅｎＡＩ官方网站上发布的信息，能够了解到比较接近ＣｈａｔＧＰＴ的技术架构是其在２０２２年初提出的ＩｎｓｔｒｕｃｔＧＰＴ模型［５］，其包括三项主要技术：

有监督精调，即利用自动或人工构建的〈Ｉｎｓｔｒｕｃｔｉｏｎ，Ｑｕｅｒｙ，Ｒｅｓｐｏｎｓｅ〉形式的数据，对基础的 ＧＰＴ 模型（通常也称为基座模型）进行精调，让ＧＰＴ能够遵循用户的指令，完成相应的文本生成

任务。

（２）奖励模型的训练（ＲｅｗａｒｄＭｏｄｅｌｉｎｇ，ＲＭ），由于ＧＰＴ解码生成文本回复是一个在词表上通过概率分布进行采样的过程，因此多次生成能够产生互不相同的回复。在ＳＦＴ训练结束之后，由人类按照自身的偏好对多个回复进行排序的过程，称为人类反馈（ＨｕｍａｎＦｅｅｄｂａｃｋ），奖励模型利用人类反馈的结果进行训练，实现对不同的回复进行

打分。

（３）基于近端策略梯度算法（ＰｒｏｘｉｍａｌＰｏｌｉｃｙＯｐｔｉｍｉｚａｔｉｏｎ，ＰＰＯ）的强化学习，即通过强化学习的方式，利用奖励模型对不同回复进行打分，将ＳＦＴ训练之后的模型向着人类偏好的方向进行进一步的调整，生成礼貌、无害、无偏见以及长度适宜的回复。

大模型中的新技术：

提示学习通过将下游任务的输入，转换成与预训练过程中使用的训练数据相似的形式，以此拉近下游任务的数据分布与预训练数据分布之间的距离，进而可以利用大规模预训练语言模型的强大语言建模能力，不经过精调就可以在各种下游任务上获得很好的结果。目标是通过构造提示语拟合预训练数据的分布，进而实现少样本或零样本的模型学习。

指令（Ｉｎｓｔｒｕｃｔｉｏｎ）是一种让模型完成特定任务的提示语，因此，指令精调（ＩｎｓｔｒｕｃｔｉｏｎＴｕｎｉｎｇ）属于提示学习的一种特殊形式。但与提示学习不同的是，指令精调的目标不单是拟合预训练数据的分布，而是希望通过构造指令的方式，学习人类与模型的交互模式，使模型更好的理解人类意图，与人类的使用习惯、偏好等进行对齐。在指令精调中，模型所要完成的任务进一步被泛化，因此模型需要根据不同的任务要求，做出与任务相符的正确回复。当指令所涉及的任务种类达到一定量级后，大模型可以在没有见过的零样本任务上达到较好的处理能力。因此，指令精调可以帮助语言模型训练更深层次的语言理解能力，以及对齐人类的使用习惯和偏好，从而泛化到处理各种不同任务的零样本学习能力。

受到上述人类求解问题过程的启发，谷歌研究人员ＪａｓｏｎＷｅｉ（现 ＯｐｅｎＡＩ员工）等提出了通过连续多步逻辑相关联的提示或指令进行精调，能够使模型在一些特殊的情况和问题中具备多步推理的能力，并将这种学习过程、能力或数据构造形式称为思维链

未来的方向：

在２０世纪９０年代对统计机器翻译［１］和语音识别［２］的研究中，研究者们在对实际应用问题进行建模时发现正确估计由ｎ个词ｗ１，ｗ２，…，ｗｎ 组成的序列的联合概率 Ｐ（ｗ１，ｗ２，…，ｗｎ）非常重要，并将估计词序列概率的模型称为语言模型。

该范式的核心思想是先利用无监督训练方法在大规模无标注文本语料上得到通用预训练模型，再结合具体下游任务的标注数据来微调模型以提升其在相应任务上的性能。该范式的优势在于能够充分利用无标注文本数据提升模型的通用语言表示能力，从而改善通用预训练模型的可迁移性，使其在各类语言理解和生成任务（尤其是标注数据较少的低资源任务）上均达到较好的性能。