



大家好奇，人工智能也搞了十几年了，AI 对话更是没有激起什么水花。无论是微软的小冰，还是苹果的 siri 还是阿里的天猫精灵，留给大家记忆深处的更多的是一堆“笑柄”。它怎么一下子突然变得这么“聪明”了？

其实说来也简单，ChatGPT 如此强大，离不开背后丰富的“知识储备”。

用于训练 ChatGPT 的数据集至少包含 4990 亿个 token (token 为语言中最小的语义单位，在英文中等于一个单词，在中文中等于一个词语)，覆盖海量的博客、书籍、新闻资讯、网站内容等。

ChatGPT 是如何学习的呢？

首先是 AI 学习来自网络的文本数据，得出随机答案；然后根据人工标注的问答数据微调模型，使 AI 一定程度上理解人类的意图，从文字接龙进入引导接龙阶段；再接着，人工对 AI 生成的结果进行排序，就此数据训练奖惩模型，使 AI 模仿人类评分标准给不同的回答打分，进入模仿喜好阶段；最后，根据打分更新参数，强化 AI 学习，训练 AI 得出高质量回答。说得通俗一点，这类训练就是让 AI 从人类反馈中强化学习，再用数据重新微调从而使得 AI 更像人。

其实整个训练过程和教小朋友数学差不多。就是先上课给例题，然后给作业，再针对作业打分后挑出一些总是出错的题目再进行有针对性的练习。最关键的一点是，对于不同的学科和知识点，我们并不需要从头训练一个新模型，之前的机器学习都是这样的，而这一次，人类不再需要含辛茹苦数十载，进行数据训练，只要拿出之前预训练好的模型，给一些指导，一个新技能就 got 了。

眼下，ChatGPT 正在通过和全球上亿用户的互动，不断增强自身的能力。如果说以谷歌、百度为代表的搜索引擎在“寻找问题的答案”，那么 ChatGPT 就是直接跨越到了“自主完成任务”。

程序员用它生成代码，公务员用它生成报告，设计师用它一键生成设计草图、图标；营销人员用它生成竞品分析，撰写经营报告以及其他业务文件；老师用它生成教学资料、测验和课程内容，学生们用它写起了论文……一天 24 小时在线，几乎零成本的专业人士在你身边做助手，不用它，用谁？比起抓耳挠腮没灵感的人类，AI 机器人更像一台灵感喷涌的永动机。

如果说以前的客服机器人只能僵硬地被某些特定的关键词触发，难以理解用户个性化的问题，那么现在的 ChatGPT 则更符合人们期待的 AI 的样子——它正在朝着人的方向发展，而且

史上最强聊天机器人？

ChatGPT 是什么？

ChatGPT (Chat Generative Pre- Trained Transformer, 聊天生成式预训练转换器) 是人工智能公司 OpenAI 于 2022 年 11 月 30 日推出的人工智能聊天机器人模型，它能够通过学习和理解人类的语言来进行对话，还能根据聊天的上下文进行互动，并协助人类完成一系列任务，因此有望成为提高办公、学习效率的工具。

AIGC (AI Generated Content, 是指利用人工智能技术来生成内容)，AIGC 也被认为是继 UGC、PGC 之后的新型内容生产方式，AI 绘画、AI 写作、ChatGPT 等都属于 AIGC 的分支。对 AIGC 来说，2022 年被认为是其发展速度惊人的一年。

ChatGPT 现在有多火？

问世以来，ChatGPT 就风靡全球互联网，上线 5 天，用户数便突破百万。2023 年 1 月末，ChatGPT 月活用户数突破 1 亿，成为全球用户增速最快的爆款应用之一。

ChatGPT 用户数增长飞快



—○ 全球用户数破百万所花费的时间

注：① 指一百万夜订单预订数量；② 指应用下载量