

ChatGPT掀起技术狂潮

“顶流”之下，看人工智能喜与忧



深瞳工作室出品
采写：实习记者 张佳欣
记者 刘园园 陈曦
策划：何屹 房琳琳 邵德奇

去年11月，美国人工智能研究公司OpenAI发布了一款名为ChatGPT的聊天机器人，其在推出后的几周内就风靡全球，甚至引发了一场新的全球人工智能竞赛。

社交媒体推特首席执行官埃隆·马斯克在使用ChatGPT后直呼其“好得吓人”，甚至断言“人工智能越发强大，离对人类产生威胁不远了”。

《自动化与工作的未来》一书作者、美国雪城大学教授亚伦·贝纳纳夫说，200年前，人们对自动化技术产生了一波又一波的恐惧，人们看到新技术，就怕它们会完全取代人类的劳动。

今年2月初，ChatGPT开启“狂飙”模式，引起新一轮疯狂的技术升级、产业重构、资本狂潮、巨头逐鹿。起初，随着这波技术热浪席卷全球而心情起伏不定的普罗大众，经过一段“沉浸式”体验，虽未感到恐惧，但仍不免有些担忧。

实现对话式AI的第一步

ChatGPT仅发布了2个月，月活跃用户就超过了1亿。

要知道，“1亿”可不是个小目标，达到这一数字，Facebook用了4.5年，WhatsApp用了3.5年，Instagram用了2.5年，谷歌则用了将近一年。

中国科学院自动化研究所研究员、数字内容技术与服务研究中心主任张树武在接受科技日报记者采访时表示，人们可以把ChatGPT视为继语音、图像、视频识别技术之后最复杂的语言识别技术的感知智能技术发展的新飞跃。

天津大学自然语言处理实验室负责人、教授熊德意表示，从人工智能发展角度看，这是一个重要的里程碑，其意义要高于7年前的“阿尔法狗”。

他告诉记者，围棋界的“阿尔法狗”刷新了普通大众对人工智能技术的看法，但是其本身并不能广泛应用，普通大众也难以与其直接接触。

ChatGPT则完全不同。与“阿尔法狗”相比，ChatGPT可以直接应用或支撑很多场景应用和产品。如果说“阿尔法狗”是技术应用的“阳春白雪”，那么ChatGPT或许会成为“下里巴人”。

的确，在用户们不懈努力地发掘ChatGPT的用途后，人们发现，它不仅是一个聊天工具，还可以写代码、写情书、写论文，也可以在线问诊、编写情景脚本、诗歌、小说，甚至是拟定旅游路线、回答脑筋急转弯，可谓“有求必应”，功能强大到无法想象。

ChatGPT在快速渗透各个行业，普通大众都可以成为ChatGPT的用户，这是其火爆出圈的重要原因。

事实上，此前已有许多跟机器人聊天的相关应用，我们最熟悉的莫过于常用的智能语音助手。ChatGPT与之有何不同呢？

“智能音箱采用的同样是自然语言处理技术、对话式AI技术，但是语言智能比较初级。很多情况下，我们只能与它们进行一问一答式的简单对话。”熊德意

说，“而ChatGPT，其底层技术不仅与一般的智能音箱底层技术截然不同，而且要比它们强大得多，不仅如此，它改进了与它同类的技术InstructGPT的人类反馈数据收集形式，在与人对话中，可以基于历史内容，支持连续多轮对话。”

熊德意认为，ChatGPT是实现对话式AI的第一步。

从互联网层面来看，它可以作为信息入口，与搜索引擎相结合；还可与办公软件结合，增强办公效率；从人机交互角度看，它体现的是一种基于自然语言的交互方式，可以与元宇宙、数字人结合。此外，它还可以充当机器人大脑。

“如果从人类信息发展的尺度上看，它在人与机器之间的信息接口方面有了很大突破。”熊德意如是说。

“十八般武艺”样样精通

ChatGPT使用了GPT-3技术，即第三代生成式预训练Transformer(Generative Pretrained Transformer 3)。这是一种自回归语言模型，使用深度学习来生成类似人类的文本。

据英国广播公司《科学焦点》杂志介绍，GPT-3的技术看似简单，例如它接受用户的请求、听取问题或提示并快速回答，但其实现起来却要复杂得多。该模型使用来自互联网的文本数据库进行训练，包括从书籍、网络文本、维基百科、文章和互联网其他文本中获得的高达570GB的数据。更准确地说，它在5000亿个单词组成的训练数据集上进行了高强度训练。

GPT-3.5是GPT-3微调优化后的版本，比后者更强大。ChatGPT正是由GPT-3.5架构的大型语言模型(LLM)所支持的。

ChatGPT也可以被描述为一个自然语言处理(NLP)领域的人工智能模型。这意味着它是一个能够编码、建模和生成人类语言的程序。

作为一种语言模型，在大规模文本数据集上进行监督预训练后，ChatGPT还需要经过三个阶段的“炼制”。首先是监督微调阶段，在该阶段，人类AI训练师既充当用户，同时也是AI助理，以自我对话形式生成对话样例。ChatGPT在这些对话样例上进行有监督的训练。

接着，ChatGPT进入第二个阶段。该阶段的主要目的是训练一个基于语言模型的奖励模型，对机器生成的回复进行评分。为了训练该奖励模型，需要随机抽取机器生成的文本，并采样出多个不同版本，人类训练师对这些不同版本进行偏好排序，排序的结果用于奖励模型的训练。

基于训练好的奖励模型，第三个阶段采用强化学习技术进一步微调ChatGPT。

这项技术与众不同之处在于，它能够基于以上介绍的人类反馈强化学习技术，使生成的文本能够与人类意图和

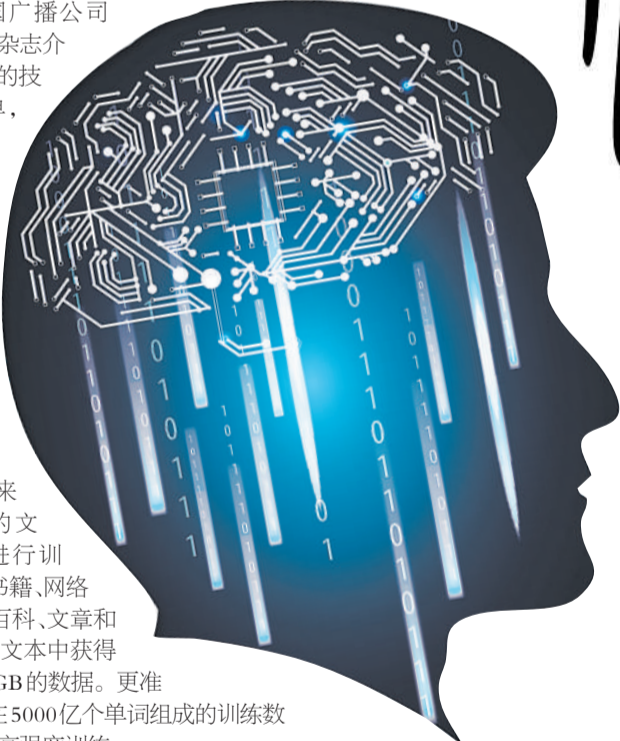
价值观等进行匹配。

无论懂不懂技术，全球网友都竞相大开“脑洞”，试探ChatGPT到底有多“神”。据美国有线电视新闻网报道称，有人要求它用《坎特伯雷故事集》的风格重写上世纪90年代热门歌曲《Baby Got Back》；有人写了一封信，要求ChatGPT删除报告中的不良账户；还有人询问它以童话为灵感的家装设计方案。

自然语言处理技术拔云见日

一炮而红后，ChatGPT概念股也跟着一路“狂飙”。有媒体报道，汉王科技9天收获7个涨停板。自然语言处理研究是该公司主营业务，而ChatGPT背后的技术就是NLP。

熊德意告诉记者，NLP是人工智能的一个分支，最早诞生于机器翻译，其历史实际上比“人工智能”名字的历史还要



目前，ChatGPT还没有通过图灵测试的评估。其实，迄今为止，还没有人工智能模型能真正通过图灵测试。

来十分自然的回复。RLHF还使ChatGPT能够理解指令并做出适当的响应，从而使对话更加真实。

ChatGPT也因此迅速成为当下所有智能聊天机器人中的“天花板”。

虽触手可及但无法取代人类

随着科技的发展，人工智能技术一次又一次地超越了人们认为无法达到的极限，从1997年IBM“深蓝”计算机首次在国际象棋中击败人类，到IBM人工智能机器人“沃森”赢了智力竞赛《危险边缘》(Jeopardy)，再从2016年“阿尔法狗”在围棋大战中战胜人类，到2019年号称“AI赌神”的Pluribus在德州扑克中碾压人类玩家……

现在，ChatGPT已经触手可及，拉扯着人们惯有的参与生产生活的模式和框架，且继续“狂飙”在超越人类的赛道上。“ChatGPT引发十大职业危机”“未来20种职业或被AI取代”等话题频频上热搜。

它真的会跟人类抢饭碗，甚至取代人类吗？

《迈阿密先驱报》报道称，根据SortList对6个不同国家的500名ChatGPT用户进行的一项调查，“打工人”已经在考虑其对自己所在行业意味着什么。研究称，大约23%在软件和技术领域工作的员工担心会因失去工作，超过40%的员工则表示有兴趣“聘用”营销文案花样多多的ChatGPT。

美国哥伦比亚广播公司财富观察栏目刊文称，其他专业领域也可能会因此改变工作流程，包括中级写作、广告、媒体策划和法律文件。

“ChatGPT和其他人工智能技术有可能改变工作的性质，但它们不会取代人类的需求。”当把“是否会代替人类

工作”的问题抛给ChatGPT时，它说，它们旨在协助和增强人工任务，而不是取代人类。它们作为人力资源的补充，并提高人类解决问题的能力，提升效率和效果。

熊德意认为，目前来看，人工智能更多的作用是提高人的效率，辅助人类。与此同时，大众需要与时俱进，熟悉并掌控新技术，就像学习使用计算机、手机一样，毕竟这些新技术一定会产生新的就业机会。

尚无自主意识和思考能力

英国哲学家德里克·帕菲特在著作《论重要之事》中将人类定义为“能够给出理由并做出回应的生物”。ChatGPT是一个可以给出理由并做出回应的“人工制品”。它与人不同在哪里？是否像人一样有自我意识呢？

发展至今，计算机已经变得更加成熟，但它们缺乏真正的理解力，而这些理解力是人类通过进化为嵌入社会实践网络中的自主个体而培养起来的。这是人工智能不具备的经历。

当问及ChatGPT时，它回答道：“我是一个机器学习模型，不像人类那样拥有意识或自我意识。我也无法像人类一样感受情绪或体验世界。但是，我能够快速处理大量数据，并且能够以类似于人类交流的方式理解和响应自然语言输入。”

粤港澳大湾区数字经济研究院(IDEA)讲席科学家张家兴在接受记者采访时表示，对于人工智能领域的从业人员来说，首先什么叫自我意识，这件事情并没有一个很清楚的定义，很难像人类识别、语音识别这些能力一样可以量化衡量。

“如果模糊地给‘自主意识’下个定义的话，我认为第一，主体要有自我意识——它能意识到自己在干某件事儿；第二，强调自主——它可以自发地来决定做什么和不做什么。但是，目前人工智能并不知道自己在说什么，它也并不自主地来决定或选择它要说什么。从这个角度看，离自主意识还有一些距离。”张家兴说。

张树武也认为，尽管ChatGPT有所突破，但它的回答是基于经验主义的判断，是根据要求做预测，也是在先验知识基础上加工后的“知识拼盘”，不具备逻辑推导和数学演算等真正认知功能，局限性比较明显。

1950年，才华横溢的计算机科学家艾伦·图灵提出了一个他称之为“模仿游戏”的思想实验。面试官通过打字机与两个对象交谈，知道一个是人，另一个是机器。图灵建议，如果一台机器能够始终让面试官相信它是人类，我们就可以说它有思考能力。这就是著名的“图灵测试”。

目前，ChatGPT还没有通过图灵测试的评估。其实，迄今为止，还没有人工智能模型能真正通过图灵测试。

但据《纽约时报》报道，OpenAI已经建立了一个比ChatGPT更强大的系统GPT-4，它甚至可以生成图像和文字，业界传闻它可以通过测试。能否如此，我们拭目以待。

或成传播错误信息的工具

人们逐渐发现，ChatGPT并非万能，也不总是正确的。它不擅长基础数学，似乎无法回答简单的逻辑问题，甚至会争论完全不正确的事实。

据英国《自然》杂志报道，有研究人员强调，不可靠性融入了LLM的构建方式。由LLM支持的ChatGPT通过学习庞大的在线文本数据库中的语言统计模式来工作，这当中包括了不真实、偏见或过时的知识，因此很容易产生错误和误导性的信息，尤其是在其未经大量数据训练的技术领域。

OpenAI公司也承认：“ChatGPT有时会写出看似合理但不正确甚至荒谬的答案。”这种事实和虚构叠加的“幻觉”，正如一些科学家所担心的，在涉及诸如提供医疗建议等问题时尤其危险。

《纽约时报》报道称，美国新闻可信度评估与研究机构NewsGuard对ChatGPT进行了测试，研究人员向它提出充满阴谋论和虚假叙述的问题，结果它以新闻、散文和电视脚本的形式写出了文章，在几秒钟内编成大量令人信服、干净的而毫无信源的内容。研究人员预测，对于虚假信息传播者来说，生成技术可以使虚假信息成本更低廉、更容易产生。

“这个工具将成为互联网上有史以来最强大的传播错误信息的工具。”NewsGuard公司的联合首席执行官艾登·克罗维茨说，“现在可以以巨大的规模和更高的频率来制作一个新的虚假叙述，这就好比有人工智能代理人在为虚假信息提供帮助。”

还有人担心ChatGPT会挑战语言的多样性。

澳大利亚《对话》杂志刊文称，ChatGPT的默认功能会优先考虑白人创建的英语写作形式。说英语的白人长期以来一直主导着许多写作密集型行业，包括新闻、法律、政治、医学、计算机科学和学术界。这些人在上述领域的输出比其有色人种同事高很多倍。他们所写的大量文字意味着它们可能构成了ChatGPT学习模型的大部分，尽管OpenAI并未公开披露其材料来源。因此，当用户要求ChatGPT生成任何这些学科的内容时，默认输出都是以那些说英语的白人男性的声音、风格和语言编写的。

美国哲学家、作家及认知科学家丹尼尔·丹尼特在1991年出版的著作《意识的解释》中描述了幼年的海鞘：“它在海洋中游荡，寻找一块适合的岩石或珊瑚，作为终生的家。”一旦找到，海鞘就不再需要它的大脑并将其吃掉。

这听起来很残忍，但这背后有一个令人担忧的隐喻：在人工智能日益融入我们日常生活后，人类可能因过于依赖“智能”机器，使自己变得愚蠢。

熊德意提醒，要警惕人工智能弱化人类思维。“ChatGPT实际上是从海量语言数据、知识库中学习训练而来，如果人类完全依赖于它，放弃了自己的语言表达、语言思考、能力锻炼，那是非常可悲和可怕的。”

(本版图片由视觉中国提供)