已有技术搭配大模型 让ChatGPT成聊天机器人『顶流』

◎实习记者 裴宸纬

“写一首诗，赞美诗人李白。”“诗人李白，吟诵仙女长夜。万千情思，尽在深沉诗中……”人类写一首诗，可能需要构思良久才能写出，但有一位“作者”却下笔如有神——只用几秒就能成诗，它就是ChatGPT，一款最近火爆全网的聊天机器人。

如果你认为这款聊天机器人的能力仅限于和人类聊聊天、写写诗，整点风花雪月，那你可真小瞧它了，写代码、编邮件、翻译，它都“样样精通”，甚至还可以写论文。

不久前，美国北密歇根大学哲学教授安东尼在为世界宗教课程作业评分时发现，全班得分最高的论文竟然是学生用ChatGPT写的。

正是由于其强大的功能，ChatGPT上线不到一周，日活用户就突破了百万。

综合运用已有技术，ChatGPT更像人

“它的长时记忆力、上下文关联推理能力和组织语言的逻辑性令人惊艳，和它聊天感觉就像和真人交流一样。”这是一位ChatGPT用户的评价。其实，这款“像人”的聊天机器人的技术原理并不新鲜。

要让聊天机器人更像人，首先要让它理解人类的语言——自然语言。“这就需要机器人对自然语言进行处理。基于大模型的对话智能技术可以帮助其快速‘学会’如何与人类交流。”思必驰科技股份有限公司联合创始人兼首席科学家、上海交通大学教授俞凯告诉科技日报记者。

记者了解到，以往的人工智能模型通用性较差，从一个应用场景切换到另外一个应用场景时，往往会变得不适用，需要重新进行训练。而ChatGPT的大模型所包含的参数在千亿级以上，是一种通用模型，无需再采集额外的数据进行小规模的深入训练，具有泛化性强的优点。

大模型训练对于聊天机器人而言，就如同基础性、通识性的大学本科培养，通过对众多“课程”的学习，机器人有了处理一般事物的能力。

但是光有这种“本科”能力还远远不够。机器人在输出文本时，还必须考虑到不同领域的知识信息以及对话的前后关联，输出更符合语言逻辑与人类价值观的高质量文本。这就需要对聊天机器人进行更为专业的“研究生教育”，使之能完成更为复杂的任务，具有更接近人类的语言表达习惯。

“ChatGPT的核心是大模型技术和对话智能技术，其中大模型技术主要有三块。”俞凯对记者说，“一是基于上下文的学习技术，这种技术可以让人工智能理解人们说话的语境；第二是ChatGPT引入的思维链技术，这种技术可以让人工智能根据语境进行相应的推理；第三是指令学习技术，它能让人工智能理解人类发出的指令，进而完成相关操作。”

俞凯总结道：“从技术迭代发展的角度看，ChatGPT是一个基于深度学习的、统计类的对话模型，它的基础技术都是我们已有的技术。但当这些技术搭配参数量很高的大模型，就产生了像ChatGPT这样的惊艳效果。”

应用前景广阔，或将变革相关产业

ChatGPT的应用场景，远不止陪你说说话、聊聊天那么简单。据悉，微软已经将ChatGPT整合到其办公软件套件中，并尝试将其嵌入到微软旗下的浏览器内，以辅助用户总结网页内容、撰写邮件和在社交平台上发帖。“它还可以纠正文章的错误，甚至按照简单的指令生成相应的文章模板。”俞凯告诉记者。

由于ChatGPT具备推理功能，因此它也能够辅助人类进行推理性工作。“比如可以辅助律师们判断案情。”俞凯说。

未来，聊天机器人或许会越来越多地走进人类的日常生活、工作，其最主要的应用场景，很可能是代替传统的搜索引擎。

“现在的搜索引擎，是用户输入关键字，然后系统给出搜索结果，这种交互是单轮的。”俞凯表示，“而ChatGPT的交互性更强，可以不断地用自然语言同用户进行交互，在这个过程中推理出用户想要的答案，然后呈现给用户。”

此外，在搜索相关信息时，ChatGPT可以快速地给出精确的唯一答案，而不会给用户一连串信息，让用户自己去“淘”想要的信息。这种方式特别符合人类想快速查找知识、获取新信息的要求。

只要能够保证其所提供知识的精准性和相关算法的精确性，聊天机器人就有很大可能在未来代替传统的搜索引擎。

如果聊天机器人替代传统的搜索引擎成为现实，所有与信息搜索相关的产业或许都将发生改变。“虽然目前的ChatGPT还不能为我们的生产生活方式带来根本性的变革，但是它却代表着人工智能的发展已经进入了一个全新的阶段。”深圳市信息服务业区块链协会会长郑定向此前在接受媒体采访时说。

“需要注意的是，ChatGPT现在还只是一个开始。未来，它一定会从文本人工智能向多模态人工智能转换。那个时候，它的能力会变得更强。”俞凯总结道。

人工智能技术的飞速发展，同时也引发了一部分人的担忧。

避免技术伦理风险、遵守人类社会的道德，可以说是聊天机器人必须遵守的底线。因此，人类需要制定相关政策，对聊天机器人之类的人工智能进行约束。

用技术和法规防控伦理风险

为生成式AI这匹“黑马”套上“缰绳”

◎实习记者 李诏宇

谈到AI领域的未来发展，就不得不提到生成式AI。如今，一经推出就迅速在各大社交媒体上走红的、由美国人工智能公司OpenAI发布的聊天机器人ChatGPT便是一种生成式AI。

近日，OpenAI又以“区分人工编写还是AI生成的文本”为目的，趁热打铁地推出了一款AI检测器。

滥用生成式AI可能带来哪些危害？AI检测器对于消除此种危害有何助益？如何利用包括AI检测器在内的多种手段，在推进生成式AI产业化的同时，促进生成式AI的规范使用？针对上述问题，科技日报记者采访了相关专家。

> 一方面，生成式AI固然便利了生产、生活，带动了社会的进步；另一方面，生成式AI可能会被人们无意滥用，甚至被目的不纯之人恶意利用，对社会造成危害。
>
> 闫怀志
> 北京理工大学网络与安全研究所所长

生成式AI是把双刃剑

“通常来说，生成式AI是指使用机器学习等各种AI算法，让人工智能能够利用数据进行学习，进而创建或生成全新的原创内容的一种技术。”北京理工大学网络与安全研究所所长闫怀志表示，“目前，生成式AI能生成文本、图像、音频、视频或代码等多模态的原创内容。”

ChatGPT作为一款人工智能技术驱动的自然语言处理工具，通过学习和理解人类的自然语言来进行对话，并能根据对话的上下文与人类进行相应的互动。与此前同类型的模型不同，ChatGPT不仅在对话的真实感与流畅性上取得了很大的提升，更在互动性上几乎“无出其右”。

以ChatGPT的横空出世为代表，生成式AI在技术上的不断突破，正驱动着全球生成式AI产业的加速发展。优秀的生成式AI有望成为新一代智能助手与信息检索工具，其不仅可以带动科技研发、工业设计等领域的深刻变革，还能在艺术和内容创作等的过程中减少以往调色、勾画轮廓等繁琐的程序性工作，极大地推动各类数字化内容的生产与创造。随着产业集群效应的进一步扩大，生成式AI正迅速成为全球科技领域最具投资价值的热门赛道，极大地影响了全球的科技创新与产业发展格局。

然而，随着生成式AI的爆火，其安全性也遭到了质疑。近日，一份关于ChatGPT的报告显示，在500名IT行业决策者中近一半的人认为，2023年内，将会出现个人恶意使用ChatGPT发动的网络攻击。无独有偶，此前的AI绘画等其他生成式AI同样遭受过相应的质疑。

“生成式AI是一个‘可怕的好东西’。”闫怀志表示，“一方面，生成式AI固然便利了生产、生活，带动了社会的进步；另一方面，生成式AI可能会被人们无意滥用，甚至被目的不纯之人恶意利用，对社会造成危害。”

闫怀志进一步解释道，生成式AI能够产出逼真的文本、图像、音频、视频等内容，可能会被滥用于诈骗、欺诈、剽窃、造谣、伪造身份等不当甚至违法的场景。

“以AI检测AI”避免技术滥用

ChatGPT不仅能像人一样进行聊天对话，它还能根据用户的要求作诗、编代码，甚至写论文。

正因如此，OpenAI近日还推出了AI检测器，其目的正是为了检测一段文本是否为AI生成，从而帮助人们更好地识别文本来源，避免AI文本生成器被滥用，造成一系列不好的影响。

“在很多场景下，对文本、图片等是否由AI生成进行识别是必要的。而这种识别所用到的检测工具，就是AI检测器。”闫怀志说，“AI检测器在技术架构上与生成式AI类似，以生成式AI创作的内容与非生成式AI创作的内容作为数据进行训练，通过大量的训练数据来‘培养’识别能力。”

这种“以AI检测AI”的方式，充分发挥了AI学习的优势。

闫怀志表示，通常来说，目前的AI检测器并不会给出是或否的精确判断，而是根据置信度给出“很有可能”“可能”“不清楚”“不太可能”“非常不可能”等模糊判断，将最终评判权交由人类自身。

然而，目前AI检测器的检测水平并不理想。闫怀志说：“OpenAI研发的AI检测器的检测成功率仅为26%左右，还处于较低水平。”

据悉，OpenAI推出的AI检测器能较轻松地区分单独的人工编写文本和AI生成文本。然而，当人工编写文本与AI生成文本混合在一起时，该检测器就难以进行准确识别了。

目前AI检测器的检测水平，显然无法有效解决滥用生成式AI带来的危害。“如果AI检测器的检测成功率能够得到较大提升，必将有助于解决目前使用生成式AI所带来的部分问题。但很显然，目前离达到这个目标，还有很长的路要走。”闫怀志表示。

多管齐下促进生成式AI健康发展

在享受生成式AI所带来的巨大利好的同时，我们应该采取措施来规范其使用，而这些措施里，既包括AI检测器在内的各类技术手段，也包括制定相关规范性的政策法规。

相关专家曾提出，生成式AI是由人类设计的。因此，人类也需要在生成式AI的程序中嵌入相应的限制手段，使它的行为以保障人类安全为底线。换句话说，就是要用优先级更高的技术规则去限制其他技术。

“ChatGPT规避技术伦理风险的方案大概就是遵循了这个思路。”思必驰科技股份有限公司联合创始人兼首席科学家、上海交通大学教授俞凯解释道，“人类通过对一些特定的问题进行优先判断，并以此构建具有约束力的模型，再让这个模型在ChatGPT的整个模型训练过程中发挥作用，以保证其回复符合人类伦理。此外，应该还有一些‘兜底技术’，让ChatGPT对于一些敏感问题不予回复。”

闫怀志表示：“除了这些技术手段外，我们还需要明确界定生成式AI的应用场景和使用范围，并在必要的情况下向人们告知哪些内容是通过生成式AI生成的。”通过相关的技术规范，界定生成式AI的应用场景和使用范围，有利于保障生成式AI遵循相应的技术伦理要求。

除此之外，相关政策、法律法规的制定，也正在推动着生成式AI产业的健康有序发展。

据悉，国务院于2017年发布了《新一代人工智能发展规划》，明确“初步建立人工智能法律法规、伦理规范和政策体系，形成人工智能安全评估和管控能力”。国家新一代人工智能治理专业委员会也相继于2019年和2021年发布了《新一代人工智能治理原则——发展负责任的人工智能》和《新一代人工智能伦理规范》。上述文件为我国目前生成式AI的发展提供了坚实的法律法规和标准保障。

“我们应该完善生成式AI的法律监管机制，坚持正义的普适价值标准、安全的核心价值目标、创新的根本目的，构建相应的伦理规范体系。”首都经济贸易大学法学院诉讼与司法制度研究中心主任陈磊表示，“具体说来，法律监管可分为事前、事中、事后三个阶段。事前应注意预防，投资者在研发生成式AI之前，应向有关部门提交申请书，将其人工智能的主要内容进行阐述和备案；事中须加强管控，在发现生成式AI侵犯他人著作权或者内容涉及伦理道德不当时，应当及时进行管控；事后要注意监督，对相关侵权行为及时进行处罚，保障他人合法权益。”

“可以想见，未来生成式AI技术与监管的博弈必将展开。”闫怀志表示。

人工智能进入“深度学习+”阶段

◎本报记者 刘 艳

虽然从底层技术看，ChatGPT并不算创新，但其社会影响远远超出了预期。这款由美国人工智能公司OpenAI开发的聊天机器人，2022年11月推出后火遍全球，成为史上增长最快的消费者应用程序。

让机器和真人自由对话，一直是人工智能领域的重要目标之一。ChatGPT的爆火背后，其实是深度学习技术的十年发展。

不久前，在百度Create AI开发者大会上，深度学习技术及应用国家工程研究中心主任、百度首席技术官王海峰表示，当前规模化的AI大生产已然形成，深度学习逐渐在技术、生态、产业等多个维度成熟，人工智能的技术创新和产业发展，进入“深度学习+”阶段。

> 我国的产业体系品类齐全、体量庞大，深度学习驱动的人工智能创新应用，有助于形成产业良性循环，促进底层技术突破，加快现代化产业体系升级。

深度学习让AI应用领域再进一步

要了解“深度学习+”，首先要了解什么是深度学习。

基于神经网络算法的深度学习，它的“深”，是相较于传统机器学习算法而言。

虽然传统机器学习算法在指纹识别、人脸检测等领域的应用基本达到了商业化要求，但要“再进一步”却很难，直到深度学习算法出现。

深度学习属于无监督学习，不需要通过人工方式进行样本标注，就能自动完成学习。需要指出的是，深度学习十分依赖硬件设施，因为它需要的计算量实在太大，且需要花费大量时间以及大量数据来进行训练。

一项技术能够将它的触角延伸至各行各业，靠的是其底层通用性。

“深度学习具有很强的通用性，呈现出标准化、自动化和模块化的工业大生产特征。”王海峰从更具广泛支撑价值的角度指出，规模化的AI大生产已形成。人工智能的技术创新和产业发展，进入“深度学习+”阶段。

深度学习促进各行业加速发展

深度学习让机器同时从海量数据和大规模知识中融合学习，效果更好、效率更高。例如，百度研制的文心产业级知识增强大模型，具备跨模态、跨语言的深度语义理解与生成能力，可应用于搜索、信息流、智能音箱等互联网产品，并通过飞桨深度学习平台赋能制造、能源、金融、通信、媒体等各行各业。

芯片、深度学习框架、模型及应用构成了深度学习良性生态，使得应用需求和反馈传递到深度学习技术的各个环节，各环节持续迭代优化，加速AI技术创新和产业发展。此外，生态中的产学研用各方，也在携手培养人工智能人才。

各行各业应用深度学习技术降本增效，创新产品和业务加快产业智能化进程，努力实现高质量增长。我国的产业体系品类齐全、体量庞大，深度学习驱动的人工智能创新应用，有助于形成产业良性循环，促进底层技术突破，加快现代化产业体系升级。比如，智能交通中“智能调度系统”，就是深度学习+交通融合创新的智能应用。城市交通复杂多变，缺乏全局感知数据，难以全域协同控制。应用深度学习技术，可实现对整个区域交通流量的全局调控，最大限度地减少各方向绿灯的空放，减缓道路拥堵，节省出行时间。