

跟 ChatGPT 唠嗑,有风险?

成为历史上增长最快的消费者应用程序 火遍全网的 ChatGPT 究竟是啥?

新华社电 美国人工智能公司 OpenAI 的大语言模型 ChatGPT 在推出约两个月后,1月已达到1亿月活跃用户,成为历史上增长最快的消费者应用程序。相关专家预计,ChatGPT 不仅是新一代聊天机器人的突破,也将为信息产业带来巨大变革,但由此带来的学术造假、技术滥用、舆论安全等风险亦不容忽视。

新一代操作系统平台的雏形

多语言撰写充满想象力的诗歌,编写可运行的程序,快速生成论文摘要,自动制作数据表格,纠正文章中的语法和表达错误,把一周大事写成新闻综述……ChatGPT 不仅能理解很多人类问题和指令,流畅展开多轮对话,也在越来越多领域显示出解决多种通用问题的能力。

ChatGPT 还轻松通过一些对人类难度较高的专业级测试:它新近通过了谷歌编码 L3 级(入门级)工程师测试;分别以 B 和 C+ 的成绩通过了美国宾夕法尼亚大学沃顿商学院 MBA 的期末考试和明尼苏达大学四门课程的研究生考试;通过了美国执业医师资格考试……业界形容它的诞生是人工智能时代的“iPhone 时刻”,意味着人工智能迎来革命性转折点。

“ChatGPT 的成功不应仅仅被看作新一代聊天机器人的突破,而应该重视其对人工智能乃至整个信息产业带来的革命。”北京智源人工智能研究院院长黄铁军接受记者专访时说,人工智能领域的过去十年是深度学习的十年,但产业总体上并没有出现移动互联网和云计算级别的爆发,“ChatGPT 的出现,具有划时代意义,大模型+ChatGPT 已形成新一代操作系统平台的雏形”。

黄铁军说,ChatGPT 在技术路径上采用了“大数据+大算力+强算法=大模型”路线,又在“基础大模型+指令微调”方向探索出新范式,其中基础大模型类似大脑,指令微调是交互训练,两者结合实现逼近人类的语言智能。ChatGPT 应用了“基于人类反馈的强化学习”训练方式,用人类偏好作为奖励信号训练模型,促使模型越来越符合人类的认知理解模式。

“这样的 AI 可帮助人类进行真实创造,尤其是帮助人类提高创造效率,比如提高获取信息的效率或提出新颖想法,再由人解决其真实性问题。创造效率的提高将产生巨大效益和多方面影响,可以改变世界信息化格局。”中国科学技术大学机器人实验室主任陈小平对记者说。



美国人工智能公司 OpenAI 标识和智能聊天机器人 ChatGPT 网站页面 新华社发

引发新一轮人工智能科技竞赛

ChatGPT 的问世正在人工智能领域引发新一轮科技竞赛。北京时间2月8日凌晨,微软推出由 ChatGPT 支持的最新版本必应搜索引擎和 Edge 浏览器,宣布要“重塑搜索”。微软旗下 Office、Azure 云服务等所有产品都将全线整合 ChatGPT。

北京时间2月7日凌晨,谷歌也发布了基于谷歌 LaMDA 大模型的下一代对话 AI 系统 Bard。同一天,百度官宣正在研发的大模型类项目“文心一言”,计划在3月完成内测,随后对公众开放。阿里巴巴、京东等中国企业也表示正在或计划研发类似产品。

人工智能大模型领域的全球竞争已趋白热化。黄铁军认为,ChatGPT 未来有望演变成新一代操作系统平台和生态。这种变革似移动互联网从个人电脑到手机的转化,大部分计算负荷将由大模型为核心的新一代信息基础设施接管。这一新范式将影响到基础设施各层面,引发整个产业格局的巨变,大模型及其软硬件支撑系统的生态之争将成为未来十年信息产业焦点。

值得注意的是,ChatGPT 有时会“一本正经地胡说八道”,存在事实性错误、知识盲区和常识偏差等诸多问题,还面临训练数据来源合规性、数据使用的偏见性、生成虚假信息、版权争议等人工智能通用风险。多家全球知名学术期刊为此更新编辑准则,包括任何大型语言模型工具都不会被接受为研究论文署名作者等。

“学术论文的署名作者须满足至少两个条件,其一是在论文工作中做出‘实质性贡献’,其二是能承担相关的责任。目前这两个条件 ChatGPT (以及其他 AI 系统) 都不满足。”陈小平说。

ChatGPT 也有应用在舆论信息战方面的潜力。加拿大麦吉尔大学研究团队曾使用 ChatGPT 前代模型 GPT-2 阅读加拿大广播公司播发的约 5000 篇有关新冠疫情的文章,然后要求其生成关于这场危机的“反事实新闻”。连 OpenAI 也警告使用 ChatGPT 的用户,它“可能偶尔会生成不正确的信息”,“产生有害指令或有偏见的内容”。

“针对这些问题,需要我们在发展技术的同时,对于 ChatGPT 应用边界加以管控,建立起对人工智能生成内容的管理法规,对利用人工智能生成和传播不实不良内容进行规避。同时加强治理工具的开发,通过技术手段识别人工智能生成内容。这对于内容检测和作品确权都是重要前提。”北京瑞莱智慧科技有限公司副总裁唐家渝说。

■ 相关新闻

ChatGPT 能考上美国医生?

美国执业医师资格考试以难度大著称,而日前美国研究人员发现,聊天机器人 ChatGPT 无需经过专门训练或加强学习就能通过或接近通过这一考试。对此,有人对人工智能在临床医学的应用充满期待,有人则开始反思美国医学教育及相关考试的不足。

ChatGPT 意为“聊天生成预训练转换器”,是美国开放人工智能研究中心研发的大型语言模型,去年11月发布,由于能在数秒内按用户需求写出论文、诗歌或编程代码而引起轰动。美国科学促进会运营的全球科技新闻服务网站“尤里卡警报”说,与现有的多数聊天机器人不同,ChatGPT 不能上网搜索,而是经由内部数据处理生成类似人类语言的文本。

