

为何要为人工智能 设“红绿灯”



大
讲
堂

近期,人工智能热潮风靡全球,它可以聊天、写作、作画,甚至在部分领域可与人类一较高下,甚至表现出可以超越人类的能力。

如何面对新的变化?有人叫好,有人好奇,但也有人担心。上个月,许多业内专家和企业联家联名发表公开信,呼吁暂停巨型人工智能实验,理由是人工智能可能表征着地球生命历史上的一次深刻变化,应该投入相匹配的关切和资源来规划和管理人工智能发展。

就在人们充满好奇地探索人工智能非凡能力之时,大家也不能忽略它可能带来的负作用甚至是毁灭性的危害。

2023年4月11日,国家互联网信息办公室起草《生成式人工智能服务管理办法(征求意见稿)》,并向社会公开征求意见。该办法中首先明确了“生成式人工智能”的定义,是指基于算法、模型、规则生成文本、图片、声音、视频、代码等技术,为人工智能设置“红绿灯”。

主讲人

宁波市慈湖中学
龚迪颖

A 生成式人工智能 技术的大爆炸

ChatGPT 问世,直接拉开了科技大爆炸的帷幕。ChatGPT 所属的 Open AI 公司,推出了 GPT-4;百度的类 ChatGPT 产品文心一言,正式面世;微软发布了集成 GPT-4 大模型的重磅革命性应用 Microsoft 365 Copilot; AI 生成图片工具 Mid-journey 推出了被视作“开启视觉革命”的 V5 模型;谷歌开放 PaLM 大模型 API 后,开放测试自家对话式 AI 服务 Bard; Adobe 推出最新 AI 模型 Firefly……

近日,来自斯坦福大学和谷歌的研究人员创造了 25 个 AI NPC(基于人工智能的非玩家角色),他们有着不同的身份、性格和年龄,会在一个共同社区中生活,每个 AI NPC 都有自己的基础行程,包括起床、洗漱、吃饭、睡觉等,根据不同的身份设定,又会做出不同的行动,共同维持社区的运转。不仅如此,这些相互之间可以产生互动的 AI NPC,甚至还能听取研究者的要求,并采取对应的行动。例如告诉其中一个 AI NPC,希望在社区举办一个“情人节”的活动,他能自发组织这场活动,包括时间地点和到场人物等。

同时,来自上海的一个视频博主利用 AIGC(生成式人工智能)“复活”已经去世的奶奶,并且还和奶奶进行视频通话。视频中,“奶奶”声音、容貌以及语气与真人极其相似。其实,看过《流浪地球 2》的观众可能对这则视频并不陌生,AIGC 奶奶其实就是另一种形式上的数字生命。博主“复活”奶奶,与电影中图恒宇“复活”丫丫的片段很相似。赞成博主行为的人认为利用 AIGC 奶奶来寄托哀思的方式是有温度的,值得肯定的;反对博主行为的人认为 AIGC 奶奶有违伦理问题,科技从业者不应为了自己的私欲而滥用技术。

以 ChatGPT 为代表的 AIGC 之前, AI 技术可以被笼统地看成是一种“关在盒子里的游戏”。相较于之前的 AI 技术, AIGC 之所以引起人们激烈的反应,主要源于它对人类独有的思考、推理领地的入侵。它一方面会采集用户数据进行训练,这将导致大规模的隐私数据泄露,无论是对于个人的隐私还是商业公司的机密与数据、文件隐私,甚至一个国家政府层面的数据隐私都有可能被 AIGC 采集到,从而被大规模扩散。另一方面 AIGC 会输出价值观、文化、思想,对于一个国家来说有着其固有的共同的文化与价值观,这是一个国家的重要组成部分,从这个角度来看, AIGC 有文化渗透的风险。第三则是剽窃问题, AIGC 需要采集大量的用户数据进行训练,这导致很多版权内容、包括文章、新闻与文献、论文资料全部被它所用,它会免费帮助用户去剽窃整个互联网的版权内容。

“科技向善”是科技创新与技术创新的最终目的,推动世界进步的是科技,颠覆道德伦理的是潘多拉魔盒。



B 生成式人工智能的达摩克里斯之剑

着眼长远, AIGC 代表的是未来,它不能被扼杀,也无法被阻挡。科技是他律性的东西,想要很好地利用其为人类服务,必先将其装入符合伦理规范和道德标准的盒子中。

数据安全、隐私安全、道德和法律,则是 AIGC 治理的基础。基于上述原因,国家出台《生成式人工智能服务管理办法(征求意见稿)》,秉承着发展的主基调,划定责任主体,制定行业底线,为中国的生成式人工智能健康发展亮起红绿灯。《办法》首先明确,国家支持人工智能算法、框架等基础技术的自主创新、推广应用、国际合作。因此,制订管理办法正是为了未来更好地发展该技术,而不是限制该技术。当前需解决三大主要问题:

一是应对隐私安全问题。对此,《办法》规定,生成式人工智

能提供者需“承担该产品生成内容生产者的责任;涉及个人信息的,承担个人信息处理者的法定责任,履行个人信息保护义务。”

二是应对技术滥用问题。在与技术滥用问题相关的虚假信息管理和他人权益问题上,《办法》规定“利用生成式人工智能生成的内容应当真实准确,采取措施防止生成虚假信息。尊重他人合法利益,防止伤害他人身心健康,损害肖像权、名誉权和个人隐私,侵犯知识产权。禁止非法获取、披露、利用个人信息和隐私、商业秘密。”

三是应对知识产权和他人权益问题。《办法》也作出了初步规定,包括生成式人工智能的训练“不含有侵犯知识产权的内容”“数据包含个人信息的,应当征得个人信息主体同意或者符合法律、行政法规规定的其他情形”等。

C 教育中的生成式人工智能

在教育领域, AIGC 对现有的教学体系、模式、内容带来了巨大的冲击与挑战。事实上,一些负面效应已经潜滋暗长。就在今年 1 月, 1000 名 18 岁以上的美国学生接受了在线课程平台 Study.com 的一项调查。经统计发现,在接受调查的学生中, 89% 以上的学生使用过 Chat-GPT 完成家庭作业,约 48% 的学生使用过 ChatGPT 完成测验,超过 50% 的学生使用 Chat-GPT 写论文, 22% 的学生曾要求 ChatGPT 提供论文大纲。

AIGC 同时提出了一个重大问题,即当前的教育应当培养什么样的人才? ChatGPT 能够成为一种技术,说明机械化文本写作已经不再是个体能力的体现,教育应更加关注逻辑思维、批判性思维和创造性思维等高阶能力。因此,有效使用智能工具的能力将成为学生核

心素养的重要组成部分,在这一点上已经达成世界共识,即全面推进核心素养教育,从知识体系教育走向素养教育,促进学生高阶思维能力与非智力因素的全面发展,启迪学生智慧,培养具备解决复杂问题能力的创新型人才。

知识为基、能力为重、价值为先。我们要占据两个“价值高地”:其一是人工智能伦理诚信建设。技术应该为人合理、合法地使用。注意伦理规范和学术诚信的问题,明确用人工智能作弊和代考答题是失信行为;其二是高阶思维的培养。小胜在智,大胜在德。习近平总书记指出:“育人的根本在于立德。”要从立德树人的根本任务出发,突破传统高阶思维的培养途径,充分发挥人工智能优势,促进创造力、批判性思维等关键能力的提升。

“君子生非异也,善假于物也”,生成式 AI 应是人类发展的工具。康德认为,人类是理性动物,而理性的真正使命在于产生“善良意志”。一种行为之所以是“善”的行为,是因为这种行为本身即包含“善良意志”,这种“善良意志”构成了人们行为的绝对律令,是一切“善”的根基。生成式人工智能时代,同样也要用“科技向善”关闭科技的潘多拉魔盒。