

从 GPT 看 AGI 的本质突破:传媒业新挑战与未来

□郭全中 袁柏林

【摘要】大语言模型 ChatGPT 的诞生带来的最大突破在于促使 ANI(专用人工智能)转向 AGI(通用人工智能),GPT-4 的发布则是 AGI 实现多模态的质变。AGI 时代的到来为传媒业带来了新挑战与新机遇,预示着 AI 生成技术能助力传媒业实现技术平权,提升传媒内容生产力以及推动广告、营销乃至媒介经营的智能运作,最终助力元宇宙的“脱虚向实”。基于目前 AGI 的实践应用,我们预判未来 AGI 将利用多模态+大模型+RLHF 技术提升智能性能,并将成为产业智能升级以及元宇宙落地的关键底层技术。然而与此同时,我们也需要警惕 AGI 带来的一系列伦理风险。

【关键词】ChatGPT;GPT-4;AGI;人工智能;技术平权

2022 年 11 月 30 日,OpenAI 团队发布了基于 GPT-3.5 的 ChatGPT 的 AI 聊天工具,它可以自动生成语言文本,应用于文本翻译、撰写文章、生成代码等,由于其操作简单、功能广泛,12 月 5 日,仅仅 5 天时间,其用户数便超过了 100 万。两个月后,用户数再次突破纪录,达到月活用户过亿的惊人成绩。2023 年 3 月 15 日凌晨,OpenAI 又发布 GPT-4 模型,是对 AGI 的进一步改善。毫无疑问,GPT 模型的火爆,让我们不得不去探索其背后的深层机制。

ChatGPT 的本质实际上是通用人工智能即 AGI,这是人工智能的本质性突破。2023 年 2 月 24 日,OpenAI 对外更新了 AGI 的发展路线,其创始人兼 CEO Sam Altman 提出了一系列关于 AGI 的规划和建议,包括如何定义和衡量 AGI、如何确保 AGI 与人类价值观和利益相一致、如何促进 AGI 的公平和透明、如何应对 AGI 对社会和经济的影响等。3 月中旬,延续 ChatGPT 研发出的多模态大模型 GPT-4,取得了文本上量的突破和图像上质的突破。尽管人类距离研发出真正的 AGI 还有很长的路要走,然而什么是 AGI? AGI 与传统 AI 有何不同? AGI 从何而来,又会走向何方?对于传媒业而言,AGI 又将带来哪些挑战与机遇?这些问题都需要我们进一步厘清。因此,本论文将重点梳理 AGI 的新突破是什么,与之前的专用人工智能的本质区别是什么,并详细分析其对传媒业带来的严重挑战和新机遇,以及 AGI 的未来与传媒业的未来。

一、AGI 及其本质性突破

AGI 是长期以来人类在人工智能领域孜孜以求的核心目标,但由于技术阻碍、认知不够、算力不够

等原因,而没有实现根本性突破,ChatGPT 和 GPT-4 让我们看到了 AGI 取得本质性突破的曙光。

(一)AGI 的含义与本质

AGI(通用人工智能)的概念最早可以追溯到 2003 年瑞典学者 Nick Bostrom 的论文 *Ethical Issues in Advanced Artificial Intelligence*,该论文指出,AGI 是一个系统,能够在没有被程序化的情况下实现人类认为可取的各种目标,具备递归自我改进的能力,将会达到一个智能爆炸,其智能水平可能会远远超过人类智能。^①因此,我们可以理解 AGI 利用人工智能系统可以完成人类能做到的所有工作以及人类不能做到的其他工作,能够帮助人类获取思维上需要的任何帮助,突破人类创造性的极限,其最终意义是可以创造出终极智能世界。作为国际顶级 AI 研究机构 OpenAI 在其企业文化中提到,其公司使命是确保在最具经济价值的工作中表现优于人类的高度自主的系统 AGI 能够造福于全人类。

(二)AGI 的发展路径

自人工智能概念诞生以来,AGI 一直是前进目标,但由于技术的阻碍,AI 发展并未直接走上 AGI 道路,而是先选择了更具可行性的 ANI(专用人工智能)道路,在文本、视觉、声音等多领域开展持久深入的智能技术探索,专用性人工智能为实现通用性奠定了技术、资金、人才等基础条件。ChatGPT 的出现将 AGI 拉回舞台中央,ChatGPT 的火爆出圈,打破了人们对 AGI 乌托邦式的质疑,让 AI 技术有了本质性突破,人工智能通用性显著提升,并引发人工智能领域对算法学习范式的思考。综上,AGI 的发展路径可以初步划分为最初设想、技术阻碍和本质突破三个阶段。

(1)最初设想:通用人工智能的提出。人工智能技术可以追溯到“图灵测试”。1950 年“人工智能之

父”阿兰·麦席森·图灵发表了一篇题为《机器能思考吗?》的论文,提出著名的“图灵测试”,人工智能思想的种子自此种下。五年后,美国达特茅斯学院展开了一场讨论计算机相关内容的达特茅斯会议,会议上 John McCarthy, Marvin Minsky, Nathaniel Rochester, 以及 Claude Shannon 等人首次提出“人工智能”概念。自此,人类一直在憧憬并探索和人脑思维一样的人工智能技术。1956 年关于人工智能的第一次尝试是由 Allen Newell 与 Herbert A. Simon 共同研发的通用问题求解器,其目的是为了解决数字公式能解决的任何问题,但由于经费不足、技术不成熟,通用性人工智能并没有得到更多进展。1980 年代,日本投入 8.5 亿美元预算,开展第五代计算机研发项目,创造新一代智能社会,但在 1991 年由于目标实现的艰难加上资金不足,该研究项目被迫停止。在探索通用性人工智能的道路上,由于对通用人工智能的认知不明确、人工智能技术不成熟以及资金成本不充足等原因,AGI 的相关实践一直没有成功。

(2)技术阻碍:转向专用人工智能研发。关于通用性人工智能的研发一直失败但研究者也一直没有放弃,而是转变创新的思路即转向特定的领域问题并寻找专用目的的解决方案。2003 年 Nick Bostrom 给出 AGI 的具体概念后,自 21 世纪初我国相继展开大量有关通用人工智能的研究。但鉴于此前的研究经验,研究者普遍对通用人工智能的探索放慢了脚步,并没有追求一蹴而就,而是从专用人工智能开始。如 2014 年谷歌 DeepMind 公司开发出阿尔法围棋(AlphaGo)的人工智能围棋软件,并在 2016 年打败世界级围棋大师李世石,入围世界围棋等级分排名。此外,还诞生出音乐行业人工智能 DopeLearning 帮助生成说唱歌词,农业人工智能机器人 Robotbees 帮助授粉,传媒行业的虚拟主播“小漾”“小净”,支持情绪监测的 Emotion AI,擅长 Dota 游戏的 Dota Five, AI 作诗小程序, AI 绘画小程序等。

(3)本质突破:人工智能通用性显著提升。在专用人工智能研发的经验积累上,OpenAI 一直在探索通用人工智能,以求人工智能的质变进展。2015 年,OpenAI 成立后,提出“all in AGI”,即为了全人类利益福祉让人工智能朝向 AGI 的方向推进,集中精力展开通用人工智能的相关实践研究。2016 年 4 月,OpenAI 为解决机器学习中的算法问题,发布了第一批人工智能软件即人工智能系统工具包 OpenAI Gym,在机器学习方面取得了重要进展。2016 年 12 月,为进一步打造复杂多样的 AI 环境,Open AI 推出了开源平台 Universe。2017 年,Open AI 发布了单

次模仿学习算法,人工智能只需单次训练便可学习领悟。在技术的不断进步上,OpenAI 开启了大模型的探索之路。2021 年,OpenAI 公司基于 GPT-3 系统,发布 DALL-E 模型,用于多场景下文本自动生成多种视觉效果图像。2022 年 OpenAI 改进了图像分辨率、扩展应用场景,发布 DALL-E2。2022 年底,基于 GPT-3.5 系统,OpenAI 发布大语言模型的 AI 聊天软件 ChatGPT。2023 年 3 月,OpenAI 研发出适应多模态的 GPT-4。

以 ChatGPT 为代表的新一代 NLP(自然语言处理)模型基于大模型与 RLHF(从人类的反馈信号中强化学习)的结合,具备更自然的语言能力,拥有更强的通用性与适应性,但其仅仅局限于 3000 词篇幅的文本格式。GPT-4 发布后,将文本篇幅扩展到了 25000 词,并支持图像的识别。ChatGPT 的诞生颠覆了从专用人工智能到通用人工智能的传统研究范式,使得直接进行通用人工智能研发成为可能,带来了人工智能的本质性突破以及通用人工智能的复兴;GPT-4 的乘胜追击让人类对通用人工智能信心满满。

第一,多模态大模型提升人工智能的可迁移性。GPT 模型是通过一个大一统模型认识世界,再用训练出来的认知对具体领域进行降维打击,而之前我们通常认知里的 AI 都是针对具体任务进行训练,只能完成某项具体任务,而不能在不同任务间通用和迁移。如基于 GPT-3.5 的 ChatGPT 大语言模型将海量的文本语料直接喂给模型进行学习,在其中模型对词性、句法的学习会沉淀在模型的参数当中,其中语料高达 3000 亿单词,参数为 1750 亿,参数是沉淀下来的 AI 对世界的理解和认识,其中一部分是对各类语法、句法的学习,另外一部分则是 AI 对于事实的认知。^②而多模态大模型 GPT-4 则基于更多的数据训练,实现文本和图像多场景的运用。AI 基于多模态大模型能够实现多任务间的通用和迁移,所以虽然目前多模态大模型需要强大的技术能力、庞大的数据库、巨大的资金量,短期内综合成本依然会很高,但是从长期来看,其成本将快速下降,智能水平将显著提升,逐步成为人工智能发展的主导方向。

第二,RLHF 降低人工智能的内生偏见。基于 GPT-3.5 的 ChatGPT 已经运用上 RLHF 的训练方法,能够通过语言训练降低 AI 内生偏见。RLHF 首先通过建立预训练语言模型,生成文本初稿数据,再将数据集运用到奖励模型,用来接收人类反馈信号,以此判断生成的文本是否符合人类思维,进而强化学习优化或微调初始语言模型^③,以此提高内容质量,达到生成的结果最大程度符合人类偏好,降低人工智

能在自行学习时预测形成的内生偏见,给用户带来更好的体验感和质感。

第三,强通用性缩减特定领域研发成本。此前针对各行业推出的专用人工智能,对于企业来说需要花费大量人力物力财力进行投资,而产出只能针对特定领域的特定问题,对于用户来说,只能获得小范围的AI应用,需求不能被满足。而GPT-3.5及GPT-4模型具有强通用性,能够产生研发者一次开发、多元生产的效果,达到对特定领域研发成本的缩减、总体用户市场的扩大,满足多样用户的个性需求。

二、AGI对传媒业的新挑战与新机遇

(一)AGI对传媒业的根本性挑战是技术平权

从万维网诞生之际,标志着互联网时代的到来,至今经历了Web1.0、Web2.0以及正在从Web2.0向Web3.0的演进中。1989年到2004年的Web1.0时代,由于门户网站等的兴起,互联网用户利用初级的搜索网站、博客等能够具备一定的信息获取权和表达展示权。2004年以后的Web2.0时代,由于社交媒体的繁荣发展,互联网用户主要具备丰富的信息获取权和表达权以及一定的经济权,但由于Web2.0时代互联网平台化发展成为主流,互联网平台借助大数据算法技术能够汇聚大量用户信息和数据且掌控了核心经济获取渠道,可以对用户权利进行掌控,用户权利无形中被削弱,由此用户在互联网上的一系列权利并没有得到充分发挥。如用户在获取信息的时候需要付出隐私泄露的成本,在表达展示的时候需要付出管理多重数字身份的成本,而在获取经济利益方面又存在分配不均的困境。然而,在现今时代即Web2.0时代向Web3.0时代的不断迭代过程中,用户权利中的信息获取权、表达展示权及经济权必须得到根本性的扩张即通过新技术实现技术平权,人类才能真正实现进入Web3.0时代。因此,从社会长远发展来看,AGI在传媒业的根本性挑战是通过通用性AI技术最大程度地助力传媒业实现在Web3.0时代的技术平权。

无论是在内容生产环节、传播环节还是接收反馈环节,传媒业都是对信息内容高度依赖的行业,因此在信息获取权上,人工智能技术需要充分给予各行各业用户专业知识,保证满足用户巨量的知识需求。在表达展示权上,普通用户必须能够获得更便捷和更低门槛的内容生产工具,消除数字鸿沟现象,让每位参与者都能得到表达展示的权利。在经济权上,必须保证经济地位的去中心化。AGI代表了人工智能

技术的又一发展前沿,技术的进步降低了用户参与专业内容生产的门槛。而AGI只有通过非盈利的方式,对普通用户开放所有功能,普通用户才能够通过百科全书式的AGI获取想要的资讯,保证信息获取权,也能够获得更便捷的内容生产工具来生产出质量更高的内容产品。此外,AGI需以开源的方式促进技术的共创共享,防止AGI最终依旧成为资本牟利的工具。总的来说,AGI要以非盈利及开源的方式推动传媒业在现阶段的技术平权,实现用户权利的显著性增强,进一步提升广大用户的表达展示权甚至经济权等重要权利,实现Web3.0的去中心化及共建共创和共享机制。

(二)AGI带来传媒内容生产力的提升与创造

随着互联网的发展,无论是专业从业者进行高质量内容生产的PGC(专业生产内容)内容生产范式,还是随着社交新媒体兴起而盛行的普罗大众尤其是自媒体生产多元化内容的UGC(用户生产内容)内容生产范式,均由于当时AI能力和技术水平较低,所以需要大量人力的参与,尤其是UGC兴起时,产生大量内容,需要更多倍的人力参与,而由于人的精力有限、知识水平参差不齐及思维创新能力也有限,更容易产生内容同质化、低俗化等现象,在此情况下,又增加了内容监管和审核的难度。因此,随着互联网的发展,在内容生产范式的不断推演中,虽然解决了人类对信息需求量及信息多样化选择的问题,但并没有解决内容生产力及内容质量根本性提升的难题,机构媒体、自媒体尤其是企业媒体仍然存在严重的内容产能不足的难题,也导致它们难以提供更多和更好的内容。

在AI能力的突破进展下,产生了AIGC(人工智能生成内容)的内容生产范式,并由之前的专用人工智能内容生成转向通用人工智能内容生成。具体来说,AGI一方面提升了传媒内容生产力的效率,以AI为技术支撑的内容生产模式可以全天候不间断地进行内容生产,满足人们随时随地寻求信息的需求,且AGI的无监督式学习可以为人类处理很多平常小事,如ChatGPT可以直接撰写新闻采访大纲、新闻报道导语等,能帮助用户做好基本资料搜集、基本文字处理等基础性工作,减轻基础工作量,为人类提供了创造更高价值内容的机会;另一方面为传媒行业内容生产的创新提供了突破点,AGI能够作为用户的好助手,帮助从业者从不同视角看待问题,拓展自身的知识边界,释放用户创新活力。此外,在GPT-4的发布会上,展示了GPT-4能够在几秒内识别用户预设的网站草图并生成相应的HTML代码,直接形成实用性网站,大大缩减了内容生产力成本。总而言之,在AGI加持

之下的 AIGC 将成为新的内容基础设施,能够极大地提升内容生产效率和生产能力,帮助机构媒体、自媒体、企业媒体等有效解决内容生产能力的短板。

(三)AGI 推动广告、营销乃至媒介经营智能运作

随着 AGI 的快速进步,技术平权将大大降低广告营销业的门槛,使得绝大多数企业都可以创办自身的媒体,也使得更多人可以从事广告营销业的创业,这将从根本上彻底改变之前由 4A 等公司把持的广告营销业或者互联网企业主导的媒介经营,而给广告营销业乃至整个媒介经营带来新变化和新活力,重构广告营销业和媒介经营版图,具体有如下三点。

第一,企业创办为自身服务的媒体将成为主流。长期以来,企业由于自身媒体能力有限,基本上采取把广告等相关业务外包的方式,而随着 AGI 的快速发展,AGI 能够通过专业微调,适应特定领域的工作任务,尤其是 GPT-4 实现图像的突破后,广告主可以借助 AGI 工具,以较低成本组建自身的媒体机构,从事大量的基础性的广告营销业务,当然高端业务依然会采取外部方式。

第二,AGI 能够作为人工智能体,参与到媒介经营智能运作当中,为专业人士提供分析、判断乃至决策参考,促使程序化广告、计算广告进一步智能化运作。广告营销等服务机构和创业者可以把主要精力集中于创意创新等工作上,而把基础性及杂乱的工作交付给 AGI 工具,进而显著提升工作效率和创意能力。

第三,对现有的搜索广告、电商广告将产生革命性影响。一是在搜索广告方面。谷歌是上一代广告商业模式最大的既得利益者,而集成了对话功能的搜索引擎如 Bing 等可能会对谷歌每年接近 1600 亿美元的搜索广告收入带来重大影响。据微软官方称,Bing 只要多获得 1% 的市场份额,就可以赚取 20 亿美元的广告收入。二是在电商广告方面。电商既需要更自然的交互方式,也需要升级真正的智能客服,这都是 AGI 的专长。2023 年 2 月 10 日,京东宣布旗下言犀人工智能应用平台将整合过往产业实践和技术积累,推出产业版的类 ChatGPT 产品——ChatJD。

综上所述,AGI 可以助力广告营销业最大程度地去中心化且促进企业有效率地高质量发展,其商业版图将可能实现重新按贡献分配。

(四)AGI 助力元宇宙“脱虚向实”

就目前而言,元宇宙还是一个基于技术集成的遥远设想,而 AGI 的兴起也导致元宇宙的热度下降。但实际上,AGI 作为元宇宙落地过程中的基础性技术,是元宇宙组建的一个重要部分,其不断地发展演进将会为元宇宙提供更强大的技术支持,同时降低

众多其他技术的研发成本。

第一,AGI 能够为元宇宙的数字孪生提供助力。数字孪生是通过接收来自物理对象的数据而实时演化的数字模型,基于数字孪生可进行分析、预测、诊断、训练等,即可以仿真,并将仿真结果反馈给物理对象,从而帮助物理对象进行优化和决策^[1],因此数字孪生技术是元宇宙中的核心技术之一,为元宇宙的高质量内容优化生产提供技术支撑,促使虚实融合。如 AGI 工具借助 RLHF 等技术可以从人类信号中得到分析和反馈,为数字孪生技术提供更多数据模型。

第二,AGI 能够为元宇宙的分布式运作提供助力。元宇宙是以 Web3.0 为网络空间基础的,Web3.0 的核心技术区块链将构成元宇宙空间分布式运作模式。所谓分布式运作即以区块链为技术基础形成的分布式协议、分布式账本、分布式计算等,其主要特征即通过共识机制实现去中心化。如在助力专用人工智能向通用人工智能进化的去中心化市场平台 SingularityNET 上,买卖双方通过分布式账本,形成智能合约,进行 AI 服务的交换,不需要中间平台,最大限度保证经济权的自主性。AGI 的实现将进一步促成分布式运作的去中心化。

第三,AGI 能够赋能各产业的数字化、智能化升级,降低元宇宙接入成本。专用人工智能只能赋能某一种行业,满足特定群体用户需求,无法做到全面且平民。因此,很多行业由于 AI 人才和技术的缺失无法紧跟时代脚步去进行数字化创新转型,元宇宙接入成本较高。而 AGI 能够汇聚各产业专业知识,为各行业乃至普通大众提供 AI 服务,这将降低各产业数字化、智能化转型的专业门槛及接入元宇宙建设的成本,各行从业者也可以借助 AGI 工具学习数智化知识,提高创新意识,推动建设数字中国。

第四,AGI 能够参与人工智能相关技术迭代,赋能技术研发。人工智能可以分为弱人工智能与强人工智能,AGI 的发展让弱人工智能开始迭代到强人工智能,而强人工智能的进一步发展则需要更先进和创新性的技术,AGI 可以作为科技从业者进行技术迭代工作的辅助工具,借助 AGI 巨量算法和泛化能力低成本完成多方面技术迭代任务,激活大量人工智能技术开发人员和团队创新思维,共同探讨和共创共建。

三、AGI 的未来趋势与风险

(一)技术趋势:多模态+大模型+RLHF 技术提升 AGI 强智能性能

专用人工智能下其智能性较弱,如阿尔法狗实

现了深度学习的进展,精通于围棋规则,甚至超越了几千年来人类对围棋的知识学习,却没有迁移学习的能力,不能举一反三,若换到象棋,阿尔法狗则一窍不通,更不用说教育、传媒、金融等其他非棋类行业了。因此,专用人工智能的智能性是极其有限的,而 AGI 则不然。

AGI 强大的通用功能得益于 AI 技术的发展。中科院自动化研究所研究员、“紫东太初”大模型研究中心常务副主任王金桥指出:其一,ChatGPT 无所不知,得益于大规模预训练语言模型 GPT-3.5 的支持。GPT-3.5 拥有千亿级参数,在训练时使用了海量的人类在互联网上编辑生成的多种类文本数据集,因此可以对世界上已有的知识进行编码,且不同于只能处理单一任务的传统人工智能,ChatGPT 能够对多任务统一编码,通过多任务统一编码,可实现一个模型做很多事,进而达到一专多能。其二,ChatGPT 对答如流,是因为引入了人类反馈的强化学习(RLHF)。OpenAI 的开发团队通过 RLHF 不断对 ChatGPT 模型进行微调,以实现最优化输出效果。由此 ChatGPT 得以像人一样对任意输入的问题做出非常流畅的回答,而且拥有文档概括、文本创作等多种能力。^[2]ChatGPT 实现了大模型+RLHF 的强大功能,而 GPT-4 模型在 GPT-3.5 的基础上,实现多模态的功能发挥,将 AGI 的通用性从文本界限又扩大至图像领域。

AGI 工具已经实践于方方面面,体现出多模态+大模型+RLHF 技术下 AGI 强智能性能的发展。其不仅能够相当于专业家教的知识水平辅导家庭作业,帮助教师备课,还能够将学习的知识进行迁移运用,包括辅助程序员撰写代码,帮助学生撰写邮件等。除此之外,还具备情绪化思维,按照用户的需求,结合上下文逻辑,对用户产生或关怀、或愤怒、或温柔、或激动的话语。

因此,未来 AGI 的技术发展上将可以使用多模态+大模型+RLHF 技术不断提升 AGI 的强智能性能,以实现更高价值的技术飞跃。

(二)应用趋势:成为产业智能升级以及元宇宙落地的关键底层技术

AGI 工具功能强大、应用范围广,因此各大相关企业积极进行产业智能升级,向 AGI 相关人工智能产业方向投资,以求抓住 AGI 的新红利,让企业产品适应人工智能时代,变得更加数智化,有效利用 AI 技术降本增效,扩大用户市场。

目前,西方发达国家多家互联网巨头瞄准 AGI 的产品市场,如 2022 年谷歌多项落地业务大多与 AI

有关,研发改进多类别大模型 GLaM、PaLM、Imagen 等,以求提供更多社会影响力;Meta 也计划组建顶级 AI 团队,直接对标 AGI;OpenAI 发布 AGI 路线图,表示公司要构建更强大的 AI 系统和更可控的模型,最终实现构建超级智能 AGI,形成宏大的超级智能世界;不断投资 OpenAI 的微软将 OpenAI 的语言模型与搜索引擎 Bing 结合,用户可以使用 Bing 得到与使用 ChatGPT 类似的结果,增强了 Bing 的搜索准确性及语言通用性,提高 Bing 在搜索引擎行业的竞争力。此外,马斯克也加入了 AGI 的研发项目中,在收购推特后,致力于目标构建更值得信赖的通用语言模型,即 TruthGPT。

在国内,自 OpenAI 发布 ChatGPT 以来,我国各大互联网企业也在积极探索中国式 ChatGPT。阿里、腾讯、华为、科大讯飞、京东、网易等纷纷探索相关业务。其中,阿里推出达摩院版 ChatGPT,能够实现文本、绘画等多模态任务;腾讯的混元助手项目组在构建大参数语言模型的工作任务上有序进行;华为积极布局鹏城盘古大模型,并发布多模态人工智能产业联合体、智能流体力学产业联合体等促进大模型产业化;智能语音引领者科大讯飞将计划采用类 ChatGPT 技术应用于 AI 学习机中;京东则推出产业版的 ChatGPT 即 ChatJD,实现业务与技术相结合;网易也计划推出与在线教育行业相匹配的 ChatGPT。除了各大互联网企业在积极布局 ChatGPT 相关产品外,高校也在积极拓展自身产品智能升级,如北大研究者发布 ChatExcel 应用,只需输入文字即可生成需要的表格。

全球各大巨头企业纷纷对标 ChatGPT,开展产业智能升级,助力 AGI 的全面快速发展,将促进 AGI 技术更快地成熟,为推动元宇宙空间提供全新内容生成范式,给予元宇宙充足且高质的内容生成。可以看出,AGI 将在今后很长时间内催生着产业智能升级以及成为元宇宙落地的关键底层技术。

(三)伦理风险:更深度信息伪造风险与更不可控的道德风险

AGI 具备功能强大、智能化强的优势,在现今阶段是最具潜力的发展对象,但技术进步所带来的不仅是丰厚的社会价值和经济价值,还会产生新的伦理风险,这值得我们深入思考和解决,因此,对于通用人工智能阶段,我们仍需在确保安全性的前提下积极稳妥地进入。

其一,AGI 无法做到辨别真假,ChatGPT 可能会产生虚假新闻或不合常理的回答,包括最新发布的 GPT-4 也无法做到完全准确。一方面,AGI 工具还不

能够准确分辨事实的真假,在传媒行业,AGI可以帮助直接生成新闻稿,若用户给予的前提条件是凭空捏造的,那么AGI将直接产出虚假新闻,这将引起舆论激化,甚至可能利用AGI生成大量不实新闻,引起舆论信息战。另一方面,由于目前GPT模型遇到没有及时更新的最新信息以及算法无法识别的问题时,AGI工具可能会一本正经地胡说八道,导致用户无法获取最佳答案,甚至会被AI的思维带偏。

其二,AGI生产的内容归属权引发争议。由于AGI具有信息编排和一定的自主创作能力,使得程序算法和独立思考的界限进一步模糊,无法按照之前评定AIGC作品版权的问题来规定,使得著作权争议以及知识产权侵权纠纷发生的频率攀升。尤其是在过亿用户的ChatGPT上,不同用户同时输入相同问题(甚至差异性问题的)得出相同内容的利益归属及相关纠纷解决路径未臻明确。^[3]因此,关于AGI生成的内容版权归属问题变得模糊不清,且现今也尚未有详细的法律界定和充分的案例经验。AI能否成为作者进行署名以及其生成的内容是否可以界定为作品,从而拥有著作权,著作权又该归属于谁等一系列问题,还需在技术实践过程中不断去解决。

其三,对于用户来说,AGI是使用工具,若不正当使用,将会扰乱行业生态。特别是在教育行业,大学生使用ChatGPT直接生成毕业论文的案例已经屡见不鲜,造成学术不端行为,甚至还有期刊因过多ChatGPT撰写的文章投稿而暂停征稿。而AGI在多行业都可以协助生成内容,若从业者只为方便完成任务而直接利用AI进行内容生成,一方面将使知识的真实性难以辨认、知识的权威性难以保证、知识碎片化加剧以及对普遍性知识瓦解;另一方面可能会导致人类思维的停顿与自身能力水平的后退,易消弭人类的自主能动性,使人类沦为技术的奴隶以及身体隐退和创造性的隐退与遮蔽^[4],不利于社会系统长期稳定发展。

其四,对于企业品牌来说,由于技术不成熟有可能给上市公司带来声誉风险,进而利益受损。AGI行业虽然蕴含着巨大的潜力和价值,但是当前依然不够成熟。例如,ChatGPT火爆之后,微软、谷歌、百度等多家互联网巨头企业纷纷涌入赛道,引起股市新一轮动荡。但在2023年2月8日,谷歌在巴黎发布会上介绍的人工智能搜索引擎Bard因错误回答导致谷歌市值一夜间骤降7000多亿元人民币。且随着愈来愈烈的股市炒作,ChatGPT、AIGC相关概念已经出现泡沫化现象,相关概念股股价大涨、初创企业估值虚高,

随着ChatGPT相关概念热度升温,行业投资者、创业者以及相关从业者需要警惕风险。^[5]

对此,目前国内网信办等相关部门也正在围绕AGI可能带来的风险进行调研。因此,各互联网企业需要和相关部门积极沟通,充分了解监管政策和监管红线,在保证产品和服务安全性的前提下再推出相关产品。

四、结语

ChatGPT的发布是ANI转向AGI的开始,GPT-4的升级是AGI的新突破,但现有GPT模型均未通过图灵测试,因此,真正实现完全的AGI还有很长的路要走。在现阶段我们需要把握AGI带来的机遇与挑战,并预测未来AGI在各方面的趋势与风险,为AGI未来的发展制订合理计划。虽然AGI的发展既带来了新的发展机遇,也带来了新的风险,但各行业尤其是传媒业应乐观应对,积极拥抱新技术并善于正确使用和掌控新技术。此外,成功通往AGI的道路并不是仅凭一己之力,而是需要全国乃至全球AI技术前沿企业的联合,共同探讨AGI的下一步,真正实现用AGI为全人类谋福祉的使命。

(本文为北京市社会科学基金规划重点项目“首都互联网平台企业社会责任与协同治理体系研究”的阶段性成果,项目批准号:22XCA002)

注 释:

①总师:《知识科普:什么是AGI?》,微信公众号“总师”,2023年2月16日。

②做产品的马丁:《万字长文:AI产品经理视角的ChatGPT全解析》,微信公众号“马丁的面包屑”,2023年1月30日。

③卖萌薯:《ChatGPT背后的算法——RLHF》,微信公众号“夕小瑶的卖萌屋”,2022年12月18日。

参考文献:

[1]郭全中.元宇宙的缘起、现状与未来[J].新闻爱好者,2022(1):26-31.

[2]赵广立.ChatGPT敲开通用人工智能大门了吗[N].中国科学报,2023-02-22(003).

[3]邓建鹏,朱怿成.ChatGPT模型的法律风险及应对之策[J/OL].https://doi.org/10.14100/j.cnki.65-1039/g4.20230228.001.

[4]周洪宇,李宇阳.ChatGPT对教育生态的冲击及应对策略[J/OL].https://doi.org/10.14100/j.cnki.65-1039/g4.20230224.001.

[5]雷晨.ChatGPT风口A股镜像[N].21世纪经济报道,2023-02-08(009).

(郭全中为中央民族大学新闻与传播学院教授,江苏紫金传媒智库高级研究员;袁柏林为中央民族大学新闻与传播学院硕士生)

编校:董方晓