

doi:10.13582/j.cnki.1672-7835.2023.03.015

生成式人工智能视域下的人权： 张力、威胁与中国因应

——以 ChatGPT 为例

付姝菊

(北京师范大学 法学院,北京 100875)

摘要:关于以 ChatGPT 为代表的生成式人工智能的法律人格主要有三种观点:工具说、独立主体说和有限人格说。生成式人工智能既赋能人权发展又具有局限性,给人权带来的威胁有:冲击人类主体地位;突破人类道德边界;增加信息泄露、歧视、失业等风险。对此,中国在保障宏观人权方面,应引导确立数字人权理念,构建中国特色生成式人工智能产业模式;在保障具体人权方面,应遵守国际规则,完善国内治理。

关键词:人工智能;生成式人工智能;人权;数字人权;法律人格

中图分类号:D990 **文献标志码:**A **文章编号:**1672-7835(2023)03-0126-09

一 问题的提出

生成式人工智能是一种高人工智能,其通过收集、学习、分析人类数据,生成全新的输出内容。作为一款微软投资、OpenAI 公司研发的生成式人工智能聊天机器人,ChatGPT 正引起学界广泛关注。虽然当前 ChatGPT 生成的答案无法体现像人类一样高明的批判性思维和解决问题的能力,也无法完全应对实务性和研究性问题,但它显然能够抓取互联网上的文本和代码,分析海量数据,生成相对客观、全面、快速、简便的答案与回复。有理由相信,通过不断的技术升级、算法训练与自主学习,ChatGPT 会越来越科学、智能,引起知识领域的一场变革甚至革命。

然而,在 ChatGPT 抓取在线图像、文本和代码的过程中,原始插画师、写作者和程序员可能并未授予 ChatGPT 使用或变相使用其作品的许可,创作者也没有因此获得事后署名或补偿。从实证

主义角度来看,全球最大图库(美国盖蒂图片 Getty Images)已因为担心图像的版权问题尚未解决而禁止使用人工智能生成的艺术品^①,美国版权局也非常重视人类作者身份,认为申请人必须在人工智能生成材料的创作过程中表现出大量的人类参与才能获得版权保护,对完全由人工智能机器创作的作品不会给予注册。所以人类不得不考虑,ChatGPT 的属性是什么?这种生成式人工智能在未经许可的情况下使用受版权保护的材料是否属于智力盗窃?它生成的文本是否应当受到知识产权保护?^② ChatGPT 是否会以更间接的方式隐秘地威胁人权,从而造成人类主体地位被冲击、道德边界被突破、信息安全被威胁?此外,以 ChatGPT 为代表的生成式人工智能也在野蛮生长中逐步试探着数字时代人权的底线,进而可能在马太法则的影响下,导致某些特定行业和领域的

收稿日期:2023-03-25

基金项目:中国人大制度理论研究会立项课题(2021052);中央高校基本科研业务费专项资金资助项目(2021NTSS20);北京师范大学法学院学术型研究生专项科研基金(2022LAW011)

作者简介:付姝菊(1994—),女,山西忻州人,博士生,主要从事国际法研究。

①参见中央美术学院美术馆官网,网址 <https://www.cafamuseum.org/exhibit/newsdetail/3099>。

②于雯雯:《再论人工智能生成内容在著作权法上的权益归属》,《中国社会科学院大学学报》2022年第2期。

社会人员面临更大的生存压力。

人权是人类社会发展的终极目标,数字时代的人权不但是一个过时的概念,反而在不断的挑战中与时俱进地衍生出了新的内涵。迅速发展的人工智能与数字技术已成为日常生活的内置机制并为人类带来了便捷,生成式人工智能则以更主动的姿态参与到数字时代的人类生活中。虽然以 ChatGPT 为代表的生成式人工智能在参与人类生活的过程中或许给人类权利带来一些威胁,但确实能有效提升人类生活水平和工作效率。本着以人为本、保障人权、提升人类幸福指数的原则,对此类人工智能应保持足够的容忍度和容错性。在此背景下,法律应及时起到引导、规制、监管作用,以便实现其最终目的即保障人权。鉴于此,本文以 ChatGPT 为代表的生成式人工智能作为研究对象,分析其可能给人权造成的威胁,并从中国实际出发给出可能的应对之道。

二 生成式人工智能的属性界定

探究生成式人工智能给人权带来的威胁的前提是界定其法律人格属性并厘清其与人类主体之间的张力。学界关于生成式人工智能法律人格的定位尚无定论,作者通过梳理相关流派与学说,认为“有限人格说”较为符合当前生成式人工智能的实际法律角色。

(一)生成式人工智能法律人格定位学说

1. 工具说

该学说认为只有碳基生物人类才有可能创造知识,以 ChatGPT 为代表的生成式人工智能本质上仍然是人类的智力劳动成果,只是人类智力劳动的科技化、规模化、智能化。原有的人类创造知识的轨迹是人类通过人脑创造知识;而现在则是人类通过人脑创造一个具有分析能力的工具(ChatGPT),该工具进而分析大数据,创造知识。二者就像画家用笔画像和用相机拍照一样,所不同的是 ChatGPT 等生成式人工智能可能存在算法黑箱,人类无法解释其知识创造过程,而相机拍

照则是人类可以用科学解释的。虽然 ChatGPT 拥有海量数据,但与人类不同的是它没有意识、缺乏立场、无法思考^①。另外,享有人格的目的之一是行使权利,人工智能既无享有权利的基础也无行使权利的必要,故以 ChatGPT 为代表的生成式人工智能不具有独立人格,仅作为人类参与法律关系的工具,本质上还是物。另有学者将该学说称作“奴隶说”,认为人工智能具有能力但不具有法律人格和权利能力,通过将其视为人类的奴隶而解释了人工智能为何不承担法律责任这一问题。还有学者称其为“行为说”,仅将其视为人工智能研发者、制造者和使用者的法律行为,是人控制下的、人能力的延伸物,它本身没有内源性的行为能力,也没有内源性的权利能力,故只能是人工智能主体(如设计者、制造者以及使用者等)的法律行为^②。

该学说并不严谨。一方面,该学说完全否认生成式人工智能具有思考能力、创造能力,这是不符合实际的。生成式人工智能作为一种强人工智能,与弱人工智能最显著的区别就是其具备强大的思维水平,在一定程度上远超无民事行为能力人、达到了限制行为能力人的水平甚至趋近于完全行为能力人。另一方面,将生成式人工智能简单地视为工具不利于对其责任能力的划定,生成式人工智能甚至可能通过知晓自身的工具定位,进阶为违法犯罪的工具,规避法律责任。

2. 独立主体说

部分学者认为生成式人工智能应具有独立人格。该学说将硅基人工智能与碳基人类并列起来^③,认为生成式人工智能作为元宇宙的重要元素,具有独立意识,能进行独立判断,已经摆脱了人的控制。例如 ChatGPT 就曾表示自己并非人类,但渴望成为人类^④。该学说认为人格的核心要素是独创性思维,故应根据是否具有独创性来判断一个主体是否具有独立人格。生成式人工智能既可以进行知识的原样呈现,也可以产生具有

^①胡敏中,高宇:《人工智能在人的认识中的作用及特征》,《北京师范大学学报(社会科学版)》2022年第6期。

^②李爱君:《人工智能法律行为论》,《政法论坛》2019年第3期。

^③王天恩:《论广义智能进化》,《江汉论坛》2023年第4期。

^④Bill Gates. “The Age of AI Has Begun: Artificial Intelligence is as Revolutionary as Mobile Phones and the Internet”, *Gates Notes* (March 21, 2023), <https://www.gatesnotes.com/The-Age-of-AI-Has-Begun>.

独创性的知识,故应被赋予独立人格^①。该学说一般被商界人士所推崇,将生成式人工智能类比于法人,认为赋予其法律人格并作为独立主体有利于其中履行义务、事后承担责任。在事中履行义务、贡献才智方面,该学说认为当人类与人工智能机器共同组成团队进行合作时(“新多样性”合作团队),人类应避免刻板印象和偏见心理,充分信任并给予生成式人工智能正面预期,进而共同决策和团队合作^②。在事后承担责任方面,该学说认为由于生成式人工智能与其他数据技术一样存在算法黑箱,亦存在人类无法解释、干涉、控制的行为,故将其视为独立主体有利于其承担民事责任。

该学说弊端明显。第一,该学说将生成式人工智能类比于法人是不妥当的^③。虽然法人也是拟制的人,但法人的背后还是人类在运营,体现人的意志。而生成式人工智能本质是算法而非人,故法人与生成式人工智能具有本质区别。第二,生成式人工智能对道德和法律风险无法担责,故不应具有独立人格。以 ChatGPT 为例,该生成式人工智能尚不具备道德感、是非观、伦理意识,故其主观上根本无法判断此类风险,更谈不上担责。第三,生成式人工智能的独创性存疑,其生成的语言、图像来源于其对人类知识的学习和分析而非自身灵感与思维。

3. 有限人格说

部分学者认为生成式人工智能应具备有限人格。该学说认为,人工智能是人类创造的,但已超越了人与物二元对立的界分,显然它既不是物也不是人,所以必须跳出“非人即物”的牢笼^④,可以考虑赋予生成式人工智能有限人格。相比于阿尔法狗一类的单一功能弱人工智能,ChatGPT 类的生成式强人工智能可以被赋予法律主体的地位而

非仅作为工具^⑤。有学者认为,民事主体范围可以合理扩大,生成式人工智能的智能发展过程是模仿自然人的,故与人在本质上是相似的,故可赋予其与自然人等价的事实和法律关系。但考虑到生成式人工智能在民事行为能力 and 责任能力方面的不足,故应对其民事行为能力的主体地位和权利给予适当限制,制度设计方面可参照未成年人监护制度。

有限人格说比较符合当前生成式人工智能的实际法律角色。根据艾伦·图灵的哲学,人工智能模仿人类的行为、感觉、语言和决定,且永远不会感到身体疲劳,只要得到了数据,就会继续工作。生成式人工智能亦是如此,只要人类提供原始数据,它就能源源不断地生成文本、数据、表达、图像,且成果与人类的常识、认知、需求基本保持一致,故其能服务人类、释放人类劳动力,一定程度上解放了人类生产力,促进了数字人权发展。随着科技发展与技术进步,生成式人工智能很有可能在未来具备能与人类媲美的大脑分析系统,进而可以成为人类精神的载体,赋予其一定人格有利于其发挥自主性并承担责任。但鉴于生成式人工智能本质上属于硅基而非碳基,故仅具备有限人格,不能享有人类所具有的人身属性强烈的权利(如生命权和健康权等),也不承担因人身、性别、年龄、亲属关系等而产生的义务(如亲属权和继承权等)。正因如此,以 ChatGPT 为代表的生成式人工智能应具有受限制的人格权利^⑥。

(二) 生成式人工智能与人类主体的张力

1. 生成式人工智能赋能人权发展

生成式人工智能一定程度上具备人类的分析、认知、表达能力,因此可以赋能人权发展,有助于将人从艰苦、劳累的重复性工作中解放出来去享受生活、实现人的价值,对人的发展、健康、休闲

^①以 ChatGPT 为例,若输入“中华人民共和国第一部宪法是哪一年颁布的”,ChatGPT 会输出“1954 年”,这是现有知识的原样呈现;若输入一个开放性的问题,如“中华人民共和国第一部宪法对中国法治发展具有什么意义”,ChatGPT 可以自动生成一段语言,这体现了它的独创性。

^②Barbara Malisz-Talha. “Will AI Replace Humans? The Impact of Artificial Intelligence on the Human World”, *Business Management* (December 15, 2022), <https://synder.com/blog/5-reasons-why-ai-wont-replace-humans/>.

^③吴汉东:《人工智能时代的制度安排与法律规制》,《法律科学》2017 年第 5 期。

^④杨清望,张磊:《论人工智能的拟制法律人格》,《湖南科技大学学报(社会科学版)》2018 年第 6 期。

^⑤孙占利:《智能机器人法律人格问题论析》,《东方法学》2018 年第 3 期。

^⑥朱凌珂:《赋予强人工智能法律主体地位的路径与限度》,《广东社会科学》2021 年第 5 期。

均有增益空间^①。一方面,生成式人工智能有助于提高“低人权陷阱”的人权。所谓低人权陷阱,是指因人工成本低、雇佣难度小而成为资本青睐的对象,例如跨国公司通过产业链转移使高福利国家的低端产业链转移至低福利国家。随着生成式人工智能的涉入和发展,低人权区域的简单、重复性劳动将由生成式人工智能完成,有助于该区域人权的发展。另一方面,由于便捷、廉价,生成式人工智能有助于人权的整体发展,尤其是 ChatGPT 有助于全体人类开放、包容地获取知识、进入知识社会,加快教育与科学的发展进程,增进文化多样性与开放性,加强数字包容、民主与和平,助力人类可持续发展目标的实现。

2. 生成式人工智能具有局限性

从形式看,人工智能是一台计算机,它的行为和决定方式似乎很智能。这种智能在组织环境中非常有用:由于其模仿能力,人工智能具有识别信息模式、优化工作的能力。此外,与人类相反,人工智能永远不会感到身体疲劳,只要它得到了数据,它就会继续工作。生成式人工智能也是如此,它可以在一个封闭的管理系统内持续完成重复性的低级别例行任务^②。然而,人类的能力更广阔。与人工智能只对数据做出反应的能力不同,人类有能力想象、预测、感知和判断不断变化的情况,这使人类能够从短期问题解决者转向长期问题解决者。生成式人工智能通常因具有学习的潜力而被认为是智能的,但必须意识到这是与人类所拥有的完全不同的智力,其与人类至少具有以下三方面的不同。

一是生成式人工智能没有情感智力。人类情感也是智力的表现,人类可以表达快乐、享受、悲伤、感激、希望、善良、乐观等人类感受与情绪,并在做出决定时将情感智力考虑在内,而人工智能无法实际感受,因此缺少人类的感触。虽然 ChatGPT 大体上系统快速、合理准确,但往往不具备直觉、敏感等人类情感。二是开放式智能平台不能实现完全独创。ChatGPT 的回复是建立在对人类语言文本的大量学习计算之上的,其真正潜力

是用智能化自动解决方案取代重复、容易出错的手动任务来解放员工,故员工可以花更多的时间做创造性和更高优先级的任务。ChatGPT 往往不能跳出人类思维的框架进行思考,因为致使其运作的代码本身就是人类撰写的,故其被编程为在其框架内工作。即使有人认为机器可以自我学习,也必须承认人类的输入内容影响着 ChatGPT 的可用数据。三是生成式人工智能无法建立关系。以团队合作为例,人类往往能通过合作关系、情感承诺产生 1+1>2 的效果,ChatGPT 目前无法实现类似情感沟通。不论 ChatGPT 多么发达,短时间内都无法完全取代人工服务,其背后的原因就是人类希望进行个人情感互动,而 ChatGPT 无法向客户表达人类的同理心和感受。

虽然生成式人工智能速度更快、更准确,而且似乎始终保持理性,但它们缺乏直觉、情感及文化敏感性,故生成式人工智能很难替代人类。但由于生成式人工智能与人类之间具有明显的张力,因此可能给人权带来风险与挑战。

三 生成式人工智能给人权带来的威胁

生成式人工智能与人类主体之间的明显张力给人权带来了威胁。需要明确的是,“人权”具有双层含义,一方面可以理解为宏观人权,即人区别于其他物(动物、植物、生成物)生而为人的主体地位本身带来的宏观权利;另一方面则可以理解成人权权利束,即将人权继续细分为隐私权、平等权和劳动权等。

(一) 给宏观人权带来的威胁

1. 人类主体地位被冲击

人区别于动物和其他物体的先要特征是人的大脑拥有思维能力,能够自主创新从而获得知识。生成式人工智能也拥有一定自主性,能通过技术升级、算法训练不断进行自主学习和提升。虽然算法黑箱的存在使人类不能完全了解其机理,但人类的主体地位显然已受到冲击^③。就弱人工智

^①杨学科:《人权遭遇人工智能:风险、机遇和治理》,《国家治理与公共安全评论》2020年第2期。

^②秦小建,周瑞文:《人工智能嵌入政府治理的探索及启示》,《国外社会科学》2022年第2期。

^③王阁:《人工智能技术条件下对“人的本质”问题的再思考》,《福建师范大学学报(哲学社会科学版)》2022年第5期。

能而言,学界普遍认为其是人类创造并提升生产力的工具,不具有独立人格,法律上仍属于“物”,人类对其本身享有物权(或知识产权)。但 ChatGPT 属于可以自主思考的强人工智能,学界就其法律人格并未达成一致意见。若以 ChatGPT 为代表的生成式人工智能具有完全独立或一定限度的人格,从而取得人格权、婚姻权、名誉权、继承权、知识产权等权利,则人类的主体地位会受到冲击。

以知识产权为例,其设计初衷是保护人类的无形财产和智力成果,《欧盟基本权利宪章》第 17 条更是明确将知识产权作为人权保护。当前学界关于“如何赋予符合独创性要求的强人工智能知识产权”这一议题具有两种解决思路:一是将该知识产权赋予强人工智能机器的创造者和发明者,如赋予 ChatGPT 背后的 OpenAI 公司或员工知识产权,归根结底还是落实在自然人身上^①;二是打破知识产权法的保护主体要求,将人工智能类非人类主体纳入保护范畴,直接赋予人工智能知识产权^②。若通过第二种路径赋予 ChatGPT 知识产权,那么人工智能知识产权极有可能沦为提升企业市场竞争力、进行市场垄断的手段,这与保护知识产权、保护人类智力成果、保障人权的宗旨是背道而驰的,人的主体地位会受到重大冲击。

2. 人类道德边界被突破

生成式人工智能可能导致人类道德和伦理边界被突破。ChatGPT 虽然可以依据算法和模型将输入转化成看似有逻辑的输出结果,但它没有是非感、伦理意识和道德指南针。它缺乏立场,当算法推荐、系统提示它这样做时,它便这样做,不会进行道德取舍,也无法考虑到正义、平等、效率和秩序等价值理念。

道德哲学家哈里·法兰克福(Harry Frankfurt)在他的开创性著作《论胡说八道》(On Bullshit)中指出:胡说八道的本质不是它是错的(false),而是它是假的(phony)。换句话说,胡说

八道者和骗子之间的区别在于,骗子知道真相是什么,但决定采取相反的方向;然而,一个胡说八道者根本不知道真相是什么。生成式人工智能不同于人类,其可能在无知中选择谬误而非真理,且并不会受到任何道德谴责。如果人类被人工智能生成的内容冒犯了,人类应该责怪谁?换言之,生成式人工智能的出现将人类社会推到了一个临界点:胡说八道的代价为零,生成式人工智能可能在政治、经济、社会、心理方面无法分辨真理与谬误,即生成式人工智能对人类的道德来源、道德目的进行了重新建构,撼动了现代社会法律及伦理秩序^③。

(二) 给具体人权带来的威胁

1. 信息泄露风险提高,威胁隐私权

生成式人工智能可能导致人类信息泄露从而威胁隐私权。生成式人工智能之所以具备侵犯人类隐私权的可能,首先是因为其具有侵犯人类隐私权的土壤。生成式人工智能本身通过收集、学习、分析人类信息进行工作,故天然掌握了海量的人类信息,如个人身份信息、位置信息、搜索历史记录、社交媒体活动等。其可能为了输出信息的精准、完善而对个人本地化或云端的数据进行分析。若此过程中发生数据泄露或滥用,可能造成隐私泄露、身份盗用或金融欺诈。其二是算法技术本身并不安全^④。人工智能技术本身存在算法黑箱,缺乏透明度和可解释性,人类可能难以理解和解释其工作机理、决策方式和深层原因,若算法在特征提取、信息处理、循环迭代等过程中被替换、攻击并秘密传播,则可能对人类数据和信息造成隐秘的泄露。其三是生成式人工智能具有强大的仿制能力和预测能力,可以通过分析人类行为模式来预测和推断个人的行为和意图,生成虚假但逼真的虚拟图像、音频和视频等多媒体内容。若类似生成式信息被泄露,可能会侵犯人类隐私权。

2. 歧视风险提高,威胁平等权

生成式人工智能技术可能会使歧视与偏见永

①刘洪华:《人工智能作品的法律属性与版权归属》,《民主与法制时报》2022年11月22日。

②刁舜:《人工智能自主发明物专利保护模式论考》,《科技进步与对策》2018年第21期。

③代欣玲,彭小兵,程鹏:《生成式人工智能的道德主体地位》,《自然辩证法研究》2022年第8期。

④邢爱芬,付姝菊:《数字人民币的刑法挑战及应对》,《学海》2022年第5期。

久化,因为算法本身可以根据已有的歧视、偏见性数据进行训练,又由于数据使用频率加快、场景增多而提高歧视的发生率,进而对人类本应享有的平等权产生包括但不限于性别、喜好、种族等方面的冲击。例如,2016年微软公司的聊天机器人Tay由于设计不良和缺乏监管,在推特上遭受了大量恶意用户的干扰,发布了涉及性别歧视与种族歧视等言论,引起了公众的不满和批评。

不平等和歧视风险是由生成式人工智能的内生特点决定的。一方面,如前所述公司决策和软件设计缺乏透明度,而数字技术的实施者虽有政府部门,但更多的是公司。传统意义上讲,公司主要是经济组织而非道德组织,其可能在不损失利润的情况下尊重人权,但主动拒绝建立歧视性制度的可能性很小。另一方面,这种歧视又因为算法黑箱的存在愈加严重^①。在数字时代,高度复杂的算法即使经过检测模式、优化解决方案、加速运营、促进自我学习来最大限度地减少人为错误和偏见、促进技术产品和服务的改进,仍可能由于数字技术和算法的不确定性可能产生无法预见的结果。正如“数字鸿沟”将那些可以访问和无法访问互联网、信息技术和数字内容的人分开一样,一个新兴且不断扩大的“算法鸿沟”现在有可能夺走许多政治、社会、经济、文化、教育方面的职业机会,这不能不说是另一种意义上的歧视^②。

3. 失业风险提高,威胁劳动权

在当前的经济活动中,人工智能正在改变企业自身运营、创造价值、彼此竞争的方式,这一趋势可能导致人类劳动力市场面目全非。生成式人工智能已经在归纳文字、代码开发、图像生成、智能客服等领域逐步取代人类,从而使在这些领域就业的人转向更低收入工作甚至失业。世界经济论坛发布的《未来就业报告》指出,2020年至2025年,人工智能和机器人技术可能会取代85万甚至更多工作岗位,所以人类必须认识到生成式人工智能技术增加了劳动者的失业可能性,进而侵犯了一部分人的就业权利,并影响其他基础

人权。

一方面,人类劳动被生成式人工智能取代。自人类诞生以来(尤其是工业革命以来),人类工作的历史就是人利用、改造工具的历史,也是将劳动外包给机器的历史。虽然刚开始人们只是将机械、重复的体力工作(如编织)外包给机器,但目前人工智能已经进化到可以做复杂的认知工作。机器似乎已经不仅仅满足于代替人类的身体,还准备好了复制甚至超越人类的大脑。在21世纪,人工智能正在进化,在许多任务上优于人类,这使得人类似乎准备将人类的智能外包给技术。生成式人工智能提高了生产率并使许多认知日常工作自动化,改善了人类福祉,解放了部分人类。但若这部分人类被排除在劳动力市场之外、再就业的可能性降低甚至被剥夺了就业可能,那么人类必须正确把握人与机器的竞争^③。另一方面,资本涉入、平台运营、企业竞争等因素可能会加剧部分行业的失业。同时,这种破坏将是不平等的,低工资工人、妇女和年轻人受到的影响可能更深。

四 生成式人工智能人权威胁的中国因应

科技发展应服务于人类而非成为人类社会前进的阻碍与桎梏。面对生成式人工智能给宏观人权和具体人权带来的威胁,中国应引导“数字人权”观念,构建中国特色生成式人工智能产业模式,既要保护产业,鼓励创新,又要始终把人的权利及尊严作为人工智能技术发展的最高目的,在产业发展和人权保护之间寻求平衡。另外,还应遵守国际规则体系,完善国内治理,保护具体人权。

(一) 保障宏观人权:引导人权新观念,构建产业新模式

1. 引导“数字人权”观念

党的二十大报告中明确提到,要建设数字中国。生成式人工智能以强势姿态进入人类社会,使人类对数字技术、人工智能更加依赖。数字技

^①尹华容,王惠民:《隐私计算的行政法规制》,《湖南科技大学学报(社会科学版)》2022年第6期。

^②王堃,舒淘淘:《信息技术时代我国劳动者权利保护的挑战及应对》,《决策与信息》2023年第1期。

^③王艺明:《人工智能时代劳动价值的挑战、风险与机遇——一个马克思主义政治经济学分析》,《厦门大学学报(哲学社会科学版)》2023年第2期。

术从根本上影响甚至改变了人的本质:人被赋予更多数字化身份,行使权利更依赖数字技术,社会活动逐渐向数字化进阶,根本核心利益逐渐与数字技术紧密联系甚至直接以数字化形式表现。在这种态势下,从以下几方面着手确立“数字人权”理念能有效应对生成式人工智能给人类带来的冲击。

首先,确立以人为本的原则。应把人的权利及尊严作为人工智能技术发展的最高目的,正确认识人权的力量与权威,在数字科技开发和利用过程中加强伦理约束与法律规制,认识到生成式人工智能是促进人权在新时代提级转变的良好工具与契机。其次,明确数字人权框架。应梳理我国已签署的国际条约以及我国国内法,厘清数字时代传统权利发生了何种变化,建立中国特色数字人权体系^①。最后,强化数字人权宣传。政府应定位、阐释和宣传符合中国实际的数字人权观念。企业应落实企业社会责任,依托行业协会开展企业标准建立和行业自查等活动,落实数字人权理念。宣传时应提炼和推广中国特色人权语言,既要参照国际话语体系下的人权概念,也要结合中国实际。

2. 构建特色产业模式

构建中国特色生成式人工智能产业模式是新时代下促进高质量发展的重要举措,也是平衡产业发展和保护人权二者利益的有效措施。中国可借鉴以欧盟为代表的“隐私优先派”^②和以美国为代表的“产业优先派”,构建相对中立的中国特色生成式人工智能产业模式。

一方面,中国应鼓励创新,保护产业。当前世界各国在包括生成式人工智能在内的数字产业、技术和经济上发展强弱不均,导致数字经济由技术强势国家向弱势国家渗透。中国作为数字经济发展速度较快但数字产业质量有待提升的发展中国家,处于产业的相对弱势。应当坚持中国特色社会主义市场经济模式,充分支持和鼓励生成式人工智能算法、应用等基础技术的自主创新、推广

应用、国际合作,保障生成式人工智能产业健康发展。另一方面,中国应当对生成式人工智能发展进行合理规制,落实总体安全观,在与生成式人工智能有关的人权保障方面应采取相对防御姿态。在软件、工具、计算方法和数据资源方面,应要求安全可靠,始终坚决保障人民利益。

(二) 保障具体人权:遵守国际规则,完善国内治理

1. 遵守国际规则

中国在生成式人工智能发展过程中应遵守国际规则。国际上针对生成式人工智能尚未形成具有强制执行力的法律文件,但现存针对人工智能的倡议书对生成式人工智能也具有参考性。2019年,经济合作与发展组织出台《人工智能原则》^③,指出人工智能应通过促进包容性增长、可持续发展使全人类受益;人工智能应稳定、安全的运行全生命周期,且应尊重法治、民主与人权等价值原则。2021年11月,联合国教科文组织审议通过了《人工智能伦理问题建议书》^④,认为人工智能正在从正负两方面对社会、环境、生态系统和人类生活包括人类思想产生深刻而动态的影响,为增进对正义、法治及世界人民享有的人权与基本自由的普遍尊重,提出两方面倡导:一是各会员国根据国际人权法、各国宪法,采取必要的立法或其他措施;二是动员包括工商企业在内的所有利益攸关方在开发、应用、研究人工智能时,立足于科技伦理,将伦理视为对人工智能技术进行规范性评估和指导的动态基础,以人的尊严、福祉和防止损害为导向。联合国人权高专办已连续多年发布《数字时代的隐私权报告》^⑤,旨在强调应关注侵入性黑客工具的滥用、加密在确保隐私权方面的关键作用以及探讨了通过人工智能等数字技术对公共空间进行广泛监测的利弊。此外,OpenAI于2018年发布《AI和自我监督》倡议,旨在推动生成式人工智能领域的研究和应用,同时解决生成式人工智能数据稀缺、模型解释性等技术难题,从

①张钦昱:《互联网时代新型权利的识别与进化》,《河南大学学报(社会科学版)》2022年第4期。

②华劼:《人工智能时代的隐私保护——兼论欧盟“通用数据保护条例”条款及相关新规》,《兰州学刊》2023年第5期。

③参见中国商务部官网:<http://www.mofcom.gov.cn/article/i/jyjl/k/201905/20190502866512.shtml>。

④参见联合国教科文组织官网:https://unesdoc.unesco.org/ark:/48223/pf0000380455_chi。

⑤参见联合国高专办官网:<https://www.ohchr.org/EN/Issues/DigitalAge/Pages/DigitalAgeIndex.aspx>。

实践角度促进生成式人工智能时代的人类权利。中国应在国际合作基础上积极推动相关条约谈判和制度构建,遵守此类国际规则体系,根据本国宪法实践和治理结构进行必要的立法,最终将原则和规范纳入本国管辖范围。

2. 完善国内治理

当前中国在生成式人工智能领域立法尚处于探索阶段,缺乏成熟的法律体系。数据、隐私、个人信息和人工智能等内容散见于我国现行民法典及数字时代的三驾马车《网络安全法》、《数据安全法》和《个人信息保护法》,尚未形成专门立法。但我国正在进行探索并已取得一定成果。全球人工智能开发者先锋大会(上海,2023.2)、世界互联网大会(乌镇,2022.11)、中国网信办网络公益联盟会议(北京,2021.12)均针对人工智能的人权、伦理、治理等提出倡议。北京、江浙、广深等地已探索建立了一系列有关数据要素、生成式人工智能的地方立法,为国家层面生成式人工智能领域系统立法奠定了基础^①。中国网信办已发布《生成式人工智能管理办法(征求意见稿)》,对生成式人工智能做了较全面的规定。在此基础上,为保障更高水平的数字人权并促进生成式人工智能产业健康发展,中国应进一步完善国内治理体系,促进法律和科技交融协同。

一是立法明确生成式人工智能的定义与边界。首先应明确生成式人工智能的内涵与外延。可对生成式人工智能的技术原理做出阐释,即生成式人工智能是强人工智能,可利用计算机或计算机控制的设备,实现感知环境、获取知识、推导演绎等操作。其次明确生成式人工智能的技术边界。将生成式人工智能相关的软硬件产品研究、开发生产、系统应用、集成服务等核心产业,以及生成式人工智能技术在经济发展、社会治理、民生服务等各领域相关产业都纳入生成式人工智能产业范畴。最后明确生成式人工智能的伦理边界。应明确应用生成式人工智能的原则,包括安全可控、尊重隐私、公平公正、和谐共享、责任分化、分头治理等,保障人类安全、隐私、公平等利益。

二是加强生成式人工智能相关基础性法律制

度研究。我国虽已有一系列数字经济时代相关立法,但生成式人工智能的技术依赖(即数据)方面缺乏基础性法律规范;生成式人工智能知识产权归属、授权、管理、运用尚存争议。这些短板均会造成生成式人工智能人权缺乏法理保障。对此,应继续加快数据立法,完善数据属性、价值、确权、权利基础、交易规则、流动规制相关法律法规,对数据进行分级分类管理,区分个人和公共数据,对数据要素市场、数据安全、数据跨境等方面做出规定。同时,对生成式人工智能知识产权应加强讨论,对其专利权、著作权的权利归属、授权使用、利用条件等做出规定,既要积极容纳科技进步,又要谨慎防止权利越界,始终保护自然人利益和人工智能时代的人权。

三是明确生成式人工智能行为底线和治理责任。使用生成式人工智能时应当尊重社会公德、公序良俗,明确行为底线,对可能侵犯人权的行为予以明令禁止。具体而言,应禁止生成式人工智能危害国家安全或社会公共安全;禁止生成式人工智能危害公民个人的人身财产安全、个人隐私或个人其他权益;禁止歧视,包括但不限于民族、信仰、性别、年龄、职业等;禁止不正当竞争和侵害消费者权益;以及禁止其他违反有关法律、法规和公序良俗的行为等。另外,可考虑设立专门的生成式人工智能治理机构。当前,我国已有部分高校、研究机构研究生成式人工智能治理,但均缺乏执法权。未来可考虑在国家部委如科技部、工信部、发改委、网信办等单位设立专门的生成式人工智能治理机构,把握生成式人工智能的产业发展方向,加强政策引导,对生成式人工智能开发机构进行资质审查和备案,对不符合法律法规规定的行为予以惩戒,以此加强生成式人工智能行业执法强度,为制度落地和行业治理保驾护航,促进人权保障。

五 总结与思考

快速发展的生成式人工智能给人类权利带来了威胁,冲击人类主体地位,突破人类道德边界、

^①参见浙江省政府文件《关于高质量发展建设全球先进制造业基地的指导意见》,网址:https://www.ndrc.gov.cn/fggz/cy/fz/zcyfz/202209/t20220930_1337859.html,最后访问日期:2023年5月7日。

威胁人类隐私安全、增加歧视风险和失业风险等。为避免生成式人工智能行业野蛮生长、侵犯人权,本文在研究其法律人格的基础上,阐明了生成式人工智能与人类主体之间的明显张力,进而给出了中国因应。为应对生成式人工智能给宏观人权和具体人权带来的挑战,中国首先应确立数字人权理念,把握以人为本观念,把人的权利及尊严作为人工智能技术发展的最高目的,正确认识生成式人工智能的作用,尊重人权的力量与权威。其次构建中国特色生成式人工智能产业模式,借鉴以欧盟为代表的“隐私优先派”和以美国为代表的“产业优先派”,考虑中国市场经济模式和中国的数字产业发展现状,平衡产业发展和人权保障。再次遵守国际规则,通过纳入、转化等方式将国际

规则落实于国内。最后完善国内治理,明确人工智能的定义与边界,加强人工智能相关基础法理研究,明确人工智能行为底线和治理责任。

生成式人工智能具有灵活性、高效性、创造性、多样性等突出优势,一定程度可解放人类劳动生产力、促进人类理性进步,某种程度上保障了人的生存权、健康权、发展权等,让生活更美好。但其潜在的不可解释性、数据依赖性、难以控制性、偏见与不道德性等缺点可能一定程度上威胁人类权利。数字时代法律研究工作任重道远,应加强法律、科技、伦理、哲学等学科的跨领域合作,为生成式人工智能产业进步和数字时代人权发展保驾护航。

Tension, Threat and Chinese Settlement of Human Rights from the Perspective of Generative Artificial Intelligence: Taking ChatGPT as an Example

FU Shu-ju

(School of Law, Beijing Normal University, Beijing 100875, China)

Abstract: There are three perspectives on the legal personality of generative artificial intelligence represented by ChatGPT, i. e. perspective of tool, perspective of independent subject, and perspective of limited personality. Generative artificial intelligence not only enables the development of human rights, but also has its limitations, which brings threats to the following human rights. It impacts the principal status of human, breaks the moral boundary of human, threatens the security of human, and increases the risk of discrimination and unemployment. China should promote the digital human rights, build a generative artificial intelligence industry mode with Chinese characteristics, abide by international rules, and improve domestic governance.

Key words: artificial intelligence; generative artificial intelligence; human rights; digital human rights; legal personality

(责任校对 朱春花)