doi:10. 16104/j. issn. 1673-1883. 2023. 02. 013

GPT模型下法律人工智能的风险与对策研究

张　恒

(安徽大学法学院,安徽 合肥 230031)

摘　要: ChatGPT是搜索模式的一次颠覆性变革,依托"自然语言处理(NLP)"与"搜索引擎"两大底层技术,GPT模型为法律咨询、类案检索、法律文书撰写的智能化发展创造了广阔空间。同时GPT模型也面临着伦理性风险、准确性风险、法律责任承担风险等方面的质疑。将新一代人工智能技术应用于法律领域,需要通过完善规则体系建设、强化法律专业监管、加强GPT模型的法律专业训练等方法来规范其使用,推动我国新一代法律人工智能的发展。

关键词: ChatGPT;法律人工智能;司法辅助;法律监管

中图分类号:D916.4;TP18　**文献标志码:**A　**文章编号:**1673-1883(2023)02-0087-05

Opportunities and Risks of Legal AI under GPT Technology

ZHANG Heng

(Law School, Anhui University, Hefei, Anhui 230031, China)

Abstract: The launch of Chat GPT can be described as an upheaving change in the search mode, and it has also triggered the imagination of the future form of legal artificial intelligence. Relying on the two underlying technologies of "natural language processing (NLP)"and"search engine", the GPT model has created a broad space for the intelligent development of legal consultation, similar case retrieval and legal document writing.At the same time, GPT models also face ethical risks, accuracy risks, and legal liability risks.To apply the new generation of artificial intelligence technology to the legal field, it is necessary to promote its role by improving the construction of the rule system, strengthening legal profsional supervision, and optimizing the construction of GPT models, so as to promote the development of a new generation of legal artificial intelligence in China.

Keywords: ChatGPT; Legal artificial intelligence; judicial assistance; legal regulation

2022年末,由Open AI公司发布的ChatGPT引爆全球社交媒体,被视为人工智能技术的又一重要突破,标志着以大模型为核心的智能计算范式的确立,具有巨大通用潜力,至2023年1月末,月活用户已突破1亿,成为史上用户数量增长最快的应用程序[1]。此后各大互联网头部企业纷纷布局这一领域,谷歌于2023年2月宣布推出ChatGPT的竞品Bard,马克•扎克伯格宣称要让自己的公司成为内容生成式人工智能的领导者[2],复旦大学发布了国内第一个对话式大型语言模型moss,在任务完成度和知识储备量上,还有很大提升空间[3]。

ChatGPT展现的巨大技术潜力使法学界已经不需要回答是否应该引入此项技术,而是要回答如何引入。法律人工智能集合了人工智能的便捷性与法律自身的专业性,增强了法学界对依靠法律人工智能提高生产力的信心,但由于对技术的稳定性、专业性要求更高,对待技术升级更为谨慎。GPT技术带来法律人工智能发展新机遇的同时,也面临更大的监管风险。对此需要从理论与实践的双重角度来思考如何最大限度地利用其机遇,最小化其风险。

一、GPT模型的运行逻辑与法律应用前景

(一)GPT模型的运行逻辑

ChatGPT是基于GPT-3.5开发的新一代对话式人工智能系统,随后更新到GPT-4版本。不同于传

收稿日期:2023-03-21

基金项目:国家社科基金重大项目:核心价值观融入法治建设中西比较研究(17VHJ010);安徽大学学科建设基金项目:多源诉讼证据融合机理与算法研究(S030162001/009)。

作者简介:张恒(1999—),男,山东临沂人,硕士研究生,研究方向:刑事诉讼法学。

统的列举式检索方式,ChatGPT通过模拟人类自然语言为用户提供检索结果,可谓是搜索模式的一次颠覆性变革。传统的搜索方式中,用户需要在检索框内输入内容,系统根据关键字匹配等技术,以关联度升序的方式提供海量条文信息,用户需要在条文信息中进行二次筛选。而在ChatGPT中,用户在对话窗口输入指令后,系统会基于数据库检索结果,结合自然语言处理技术,以对话形式为用户提供单一、指向明确的检索结果。ChatGPT的这种输出方式得益于自然语言处理技术(NLP)。自然语言处理技术是人工智能领域中的一个重要分支,旨在使计算机能够理解和处理人类语言。通过建构海量语料数据库,结合语言学习微调训练模型,通过系统自主预训练学习的方法使其掌握语言理解与文本分析的能力,能够理解用户指令并以自然语言的形式输出结果[4]。GPT模型的运行逻辑主要依托“自然语言处理”与“搜索引擎”两大底层技术。Open AI公司研发人员曾对NLP技术下GPT模型的运行逻辑进行了详细介绍,GPT模型的语言学习过程类似于人类学习语言的过程,系统会基于语料数据库进行自主语言学习与训练,掌握语言的语法、语义、上下文联系及俚语等,再借助Transformer模型构建高度类似人类表达逻辑和方式的语言算法模型[2]。用户输入指令后,GPT为知晓指令涵义会分步对指令进行词法分析、语义分析、语言情感分析等,使用NLP技术将文本从自然语言翻译为机器语言来理解。在知晓用户指令涵义后,GPT模型会依托搜索引擎技术在数据库中检索相应内容,然后基于对用户指令的分析结果,将搜索结果由数字序列形式逆向翻译为自然语言形式,并为用户输出。

但目前GPT模型在运行过程中还存在一些制约因素。首先,诸多科技公司布局GPT领域更注重其经济效益而非技术的安全性。科技公司如同民法中的先占制度一般抢先推出自己的GPT软件,在追求速度的过程中忽视了安全性。例如谷歌在展示巴德(Bard)时,该聊天机器人就回答错误一个有关韦伯太空望远镜的问题,使其股票暴跌[5]。其次,现有GPT模型仍是基于固定的数据库,还未实现联网式检索。GPT模型历经GPT、GPT-2、GPT-3等多个版本更迭,目前GPT-4是训练数据最大、智能化水平最高的模型。但由于数据时间范围的限制,以及不能联网搜索,无法获取即时数据,ChatGPT对2021年以后发生的事情了解有限。当后续版本的GPT模型可以联网时,将会展现出更强的创造力。有研究者认为GPT模型以后可能具有改造人类的思想和创造能力,这一系统可能具备人类的思维能力,也可能在某一方面或其他方面替代人类[6]。最后,目前GPT模型仍处于“即兴创作阶段”,有生成错误答案的风险。在使用过程中,不断有用户反映ChatGPT会编撰一些不存在的事实,类似于人类的谎言,这也引发了对其回答真实性的质疑。

(二)GPT模型的法律应用前景

1.GPT模型与法律咨询

法律人工智能为用户提供法律咨询应用十分广泛,“法狗狗”“法咚咚”“法小淘”等系统不断涌现,主要是为用户提供智能的法律咨询服务。这些法律咨询型人工智能仍是基于文本分析方法,通过理解用户的整体语义,根据不同用户的不同要求做出预测,结合基本案情、证据收集情况、相关案件的证据采信率、用户自身客观条件等实际情况,为用户提供符合自己利益的最有效的建议[7]。ChatGPT具有提供法律咨询的先天优势,可以较为准确地分析用户指令,理解案件事实,咨询结果具有较好的针对性。以ChatGPT为代表的深度学习道路应是未来法律咨询型人工智能的发展方向。近年来文本分析方法也逐渐走上了深度学习道路,其中作为文本结构化表示方式的词向量的提出和发展,以及深度神经网络文本分类模型的发展,对于文本分类任务的研究和发展具有重要意义。但应用GPT模型进行法律咨询存在一定风险[8]。GPT模型提供的法律咨询结果一般较为具体,指向性明确,但由于不同国家、地区的法律规则存有差异,GPT模型作出的回答并不一定符合用户所处地的法律规定,存在误导用户的风险。

2.GPT模型与类案检索

“类案同判”在司法实践中具有重要价值,是实现司法公正、保障法律面前人人平等的重要举措。但目前类案检索技术仍处于初始阶段,主要是采用了“穷举式”而非“精准式”的推送方式,搜索结果的精准度不高。同时“穷举式”推送方式也增加了法官筛选负担,有研究显示,相较于其他司法智能辅助平台,法官对类案搜索平台的使用意向并不强烈[9]。ChatGPT提供了升级类案检索技术的机遇,有效提高类案检索的效率和准确性。首先,GPT模型的检索结果较为准确,指向性明确,用户不需要在检索结果中进一步筛选,检索结果与用户指令的贴合度更高,可以极大提升类案检索的用户体验感。当下制约类案检索率提高的重要因素之一就是检索过程过于烦琐,且检索结果的贴合度不高,需要二次筛选。其次,GPT模型降低了类案检索门

槛,扩大了类案检索适用主体范围。现有的类案检索系统多是以“关键词检索+标签检索”模式构建,检索标签多属于法律术语,专业化水平较高,对非法律专业人员不友好。而GPT模型则可以通过NLP技术自主理解用户指令,降低使用门槛,扩展应用类案检索的主体范围。但GPT模型具备自主创造力,存在自行编撰案例的可能,应用GPT模型进行类案检索也存在一定风险。

3.GPT模型与法律文书撰写

人工智能已经被广泛应用于法律文书的撰写。例如上海法院办公办案平台的法律文书智能辅助模块,支持“判决书”“裁定书”“调解书”“决定书”等法律文书的在线制作,并提供相应模板。法律文书撰写的准确性和合法性对于保障案件正确审理、保护当事人的诉讼权利具有重要意义,因此,法律文书的撰写必须严格遵守法律规定和格式。目前法律人工智能在法律文书撰写领域的主要功能仍是提供模板、在线协作等,尚未达到针对某一具体案件自动生成文书的水平。GPT模型使系统自主生成法律文书成为可能,目前GPT模型已经可以完成写文章、写诗等创造性写作任务。用户只要对写作内容提出明确指令,GPT模型就可以依据指令进行命题作文。但其中也暴露了一些风险,由于GPT模型是基于大量文本数据训练出的,因此其回答并不完全准确,存在偏差风险。有用户指出ChatGPT模型在一些情况下会“跑题”,若该情况发生在法律文书的撰写过程中,造成的危害是显而易见的。同时由于法律领域的复杂性和不确定性,GPT模型的回答也会存在漏洞或者遗漏。此外GPT模型生成错误法律文书的责任承担也是一个重要问题。但不可否认的是,GPT模型的应用会极大减轻司法实务人员撰写法律文书的压力,有效节约司法资源。所以需要从规则与技术层面共同推进GPT模型的使用,实现便捷性与安全性的平衡。

二、GPT模型的法律应用风险

(一)伦理性风险

法律人工智能是依托代码与算法的运行结果,自身并无伦理观念,但隐藏在代码和算法中的情感偏向会使法律人工智能具备一定的道德敏感性。西方有学者开始探索AMAs(人工道德智能体)的可能性,通过嵌入道德代码来降低法律人工智能的伦理风险[10]。当GPT模型展现出类人类的自主创造能力与自主深度学习能力时,便引发了其是否具备主体资格的伦理问题讨论[11]。首先,GPT模型涉及数据隐私问题。用于训练GPT模型的数据可能包含个人隐私信息,如不能妥善处理则会对个人隐私权造成侵害。GPT模型在进行信息传播时也涉及伦理性问题,GPT模型的信息输出归根到底是依托数据库的算法运算结果,必然会使用数据库中的相关信息,其中就涉及隐私泄露等风险。其次,GPT模型存在公正性风险。由于数据偏差风险的存在,GPT模型可能会对特定群体产生偏见,如种族、性别、年龄等,而“法律面前人人平等”原则是法律不可突破的底线。如果GPT模型存在公正性问题,例如在审判过程中对被告产生偏见,将会极大地挑战法律的权威性。同时,定向偏差性数据训练会影响GPT模型的结果生成,造成不正当竞争等问题。最后,GPT模型的主体资格问题。解决法律人工智能的责任承担问题的首先要解决法律人工智能的主体资格问题。理性主义认为,具有主体资格须具备一般的辨识是非、合理预见和控制行为的能力[12],法律人工智能可以拟制为法律主体。但感性主义认为,主体应该具备通过感官获得快乐或痛苦的能力[13],GPT模型具备自主创造力,但不具备感性能力,故不具备法律主体资格。

(二)准确性风险

法律语言是人们基于语言学基础,在长期法律实践过程中形成的一种语言变体,由于使用主体的专业性、应用目的的严谨性,准确性是法律语言的首要风格特点。语言学中的准确性是指通过语句能够与表达者的主观意思高度贴合,法律语言的功能要求其必须准确表达事实,避免歧义与模糊性[14]。GPT模型作为基于NLP技术开发的语言模型,如何达到法律语言所要求的准确性是一个重要问题。首先,GPT模型通常依托大量文本数据进行训练,这些语料库不仅包括法律语言数据库,还包括其他语言数据库,数据中可能存在偏差,导致模型的结果不够准确。如果仅使用法律语言数据库进行模型自主学习,则会因训练数据不足而导致模型不够“智能”。复旦大学认为其moss模型与ChatGPT模型最大的差距就在于训练语料不足,由此导致moss相对不够“智能”。同时对于法律中的不同观点,GPT模型如何进行选择,如何避免技术人员通过偏差数据诱导GPT模型倾向等,均需审慎解决。其次,GPT模型的语言灵活性不足以满足法律语言的要求。数据偏差、相关语料占总体语料比例偏低等因素导致GPT模型的语言灵活性不足。法律文件通常具有高度灵活性,可以通过解释、引用和法律原则来解释和推断,这种复杂性很难被语言模型

所理解。而且,法律文件和案例通常建立在特定的历史背景、政治、经济和社会因素的影响下,语言模型如果缺乏对这些关键背景的了解,会导致结果不准确。

(三)法律责任承担风险

法律人工智能可以自动分析和解释大量法律文本,帮助律师和法官更快速、准确地作出决策。但是法律人工智能技术的引入也导致数字时代司法呈现出"去责任化"趋势。司法实务人员出于"趋利避害"本能,倾向于将责任转移给法律人工智能或软件开发者承担,导致出现极端形式的规则迷恋、程序主义、僵化思维和责任推卸等不良后果,引发了对法律责任承担风险的关注[15]。使用法律人工智能生成的文书,需要明确责任主体。法律人工智能自身是否具备责任主体能力涉及伦理性问题。在"技治主义"观念下,可以依托技术可靠性赋予法律人工智能一定的责任能力,而反对者则基于感性主义予以反驳。无论法律人工智能是否具备责任能力,其都为"人"提供了转移责任的可能,对此必须解决参与法律人工智能开发、生产、使用人员的责任分配问题。作为系统运行的实质参与者,使用者在将原本应自主完成的任务交给法律人工智能时,就应负担由此产生的审查义务与责任后果。但如果由于法律人工智能的自身缺陷而出现错误,是否可以将责任转移到开发、生产法律人工智能的企业或个人身上,还是仍由对生成结果负有审查义务的使用者负责,目前仍无定论。若是使用者能够证明已经尽了合理的努力来验证生成文书的准确性和可靠性,在这种情况下是否可以进行责任转移,这些均需在规则层面予以明确。同时在进行责任分配的过程中,必须认识到转移责任会导致责任主体缺位的风险。多个主体相互推诿、指责,或将责任推卸给系统承担,架空原有的追责制度。

三、探索GPT模型的法律应用路径

(一)明确GPT模型的法律应用规则

建章立制是发挥法律人工智能效能的必要基础。实践中法律人工智能已被广泛应用,但有关法律人工智能应用的相关规定仍然匮乏。中华人民共和国最高人民法院2022年发布《关于规范和加强人工智能司法应用的意见》,明确提出法律人工智能应用的五大基本原则:安全合法原则、公平公正原则、辅助审判原则、透明可信原则、公序良俗原则,主线思想是必须在确保法律人工智能安全合规的基础上,才能对发挥法律人工智能的便捷性进行实践。合规的前提是有规可依,我国法律人工智能应用规则相对缺失,必须加紧制定相关法律,明确法律人工智能的应用原则、应用路径、责任承担等。特别是GPT模型在使用过程中会不断自主学习,其中涉及用户个人信息安全,需在结合《个人信息保护法》的基础上作出明确规定。基于强化对数据产业的合规监管目的,我国近年来陆续出台了《个人信息保护法》《数据安全法》等多部法律法规。企业在开发软件时,需要在合规基础上,结合具体技术使用情况对用户协议进行调整。但由于目前相关法律多为总领性规则,可操作性不强,特别是企业在开发GPT模型这种具备自主创造力的应用时,明确的法律依据不足。这会导致因法规保障不到位而制约技术发展。对此,一方面要优化规则架构,提升法规制定的规范性、科学性与系统性,为新兴技术的开发预留合规空间。另一方面要加紧具体细则的制定,通过对新兴技术进行二级分类,分别出台相应规定。例如,根据法律人工智能的智能化水平制定不同等级的安全规则。此外,发挥GPT模型的便捷性不能突破安全性的限制。目前GPT模型已经通过系统设置来规避一些不合法、不合规的内容,未来应对这些内容进行类型化归纳,并从法规层面予以明确。这有利于通过强制性规定促使其他GPT模型建构相应的系统设置,也可为平台发展提供坚实的法律保障。

(二)加强GPT模型的法律专业训练

应用法律本质上是一个思维决策过程,而每个人的决策都会受到自身多重因素的影响,故个体会对他人的决策表现出不信任。但在社会现实中,人们对司法裁判者的决策表现出极大的信赖,这根源于人们相信司法裁判者通过前期大量学习,可以较为客观地进行决策。司法裁判者决策的相对客观性是前期大量训练的结果,因此将具备一定自主创造力的人工智能应用于法律领域时,必须要优化GPT模型的训练,保证其客观、公正以获取公众信赖。在使用GPT模型前,需要针对算法和数据构建事前合规机制,确保司法数据和算法的透明性[16]。保证使用的信息不涉及数据隐私与保护问题,不会泄露敏感信息。首先,要进行数据准备。打包一份高质量的法律案件文本数据集,选择高质量的法律文件和案例作为训练数据以减少数据偏差,并对案件进行标签化分类。然后,利用已有的NLP框架如TensorFlow、PyTorch等对GPT模型进行训练。模型训练结束后需要使用独立的数据集对模型进行准确性测试、安全性测试和验证,确保模型可靠。若

测试结果符合法律规定和预期使用效果,可以通过提供一个Web界面或API接口进行实用化部署。使用过程中还需要对GPT模型进行日常维护与性能监测,定期更新数据集并重新训练。

(三)强化GPT模型的法律专业监管

在应用法律人工智能的过程中,最广泛的质疑声便是其是否可靠、专业。回答这个问题不仅需要技术迭代升级,还需要强化对法律人工智能的专业监管。法律人工智能的自主创造力与可控性、安全性呈反比例关系,但提升法律人工智能的智能化水平本质上就是提升其自主创造力。当GPT模型代表的深度学习、自主创造型人工智能被应用于法律领域时,对其进行法律专业监管将变得更为困难,为此必须构建体系化的监管机制。首先,强化模型构建过程的监管。GPT模型需要大量数据进行训练和优化,这些数据可能包含敏感信息,因此应当对GPT模型使用的训练数据进行合规监管,明确在数据收集、存储和使用方面的要求,并对违规行为进行处罚。数据偏差与系统架构问题会导致模型生成的结果不够客观,需要在模型建构过程中与法律专业人员合作,确保模型生成的结果符合法律要求。其次,强化GPT模型使用过程中的监管。GPT模型的使用可能导致一些不良后果,如造成虚假信息传播、引导不当决策、不公平竞争等。应当明确GPT模型使用过程中的追责标准与程序,对其生成的法律文件和决策进行审查,保障符合道德和法律的要求,禁止将GPT模型用于违法、歧视、虚假等行为。最后,需要加强对GPT模型生成结果的监管。在GPT模型生成结果后,使用者应负有实质审查义务,及时发现问题并采取相应的措施,避免损害扩大。

四、结语

GPT模型的每次迭代都会使其生产能力呈指数级增长,在司法资源相对紧张的背景下,法律人工智能技术的发展为解决该问题提供了新思路。当下GPT模型的构建与训练主要面向通用型,未来可以通过使用大量司法数据来培育法律专业GPT模型,但需注重解决数据隐私、系统安全、合规应用、法律监管等一系列问题。未来新兴技术一定会不断地丰富到法律领域以提高生产力,需要从规则框架中为其预留应用空间,推动法律领域智能化水平的提高。

参考文献:

[1] 任晓宁.进击的ChatGPT[N].经济观察报,2023-02-06(017).
[2] 邓建鹏,朱怿成.ChatGPT模型的法律风险及应对之策[J/OL].https://www.qkl456.com/391922.html.
[3] 侯树文,王春.复旦MOSS距离ChatGPT还有多远?[N].科技日报,2022-02-23(02).
[4] 张夏恒.ChatGPT的逻辑解构、影响研判及政策建议[J/OL].https://doi.org/10.14100/j.cnki.65-1039/g4.20230228.001.
[5] 刘霞.出师不利! 谷歌Bard答题犯下事实性错误[N].科技日报,2023-02-10(04).
[6] 朱光辉,王喜文.ChatGPT的运行模式、关键技术及未来图景[J].新疆师范大学学报(哲学社会科学版),2023(4):113-122.
[7] 李丹.人工智能技术能否应用于律师行业? ——基于情感、效率和执业监督维度的分析视角[J].东南大学学报(哲学社会科学版),2019,21(S1):58-62.
[8] 杜思佳.基于深度神经网络的法律咨询用户意图理解研究与实现[D].哈尔滨:哈尔滨工业大学,2019.
[9] 左卫民.如何通过人工智能实现类案类判[J].中国法律评论,2018(2):26-32.
[10] 马长山.人工智能的社会风险及其法律规制[J].法律科学(西北政法大学学报),2018(6):47-55.
[11] 令小雄,王鼎民,袁健.ChatGPT爆火后关于科技伦理及学术伦理的冷思考[J].新疆师范大学学报(哲学社会科学版),2023(4):123-136.
[12] 龙文懋.人工智能法律主体地位的法哲学思考[J].法律科学(西北政法大学学报),2018(5):24-31.
[13] 杨志航.人工智能法律主体资格之否定[J].财经法学,2022(4):83-98.
[14] 杨淑芳.确保法律语言准确性应注意的问题[J].政法论丛,2004(2):96-97.
[15] 高童非.数字时代司法责任伦理之守正[J].法制与社会发展,2022(1):151-172.
[16] 刘金松.数字时代刑事正当程序的重构:一种技术性程序正义理论[J].华中科技大学学报(社会科学版),2023(2):18-29.