人工智能生成语言的治理现状和问题[①]

郭书谏　沈　骑[②]
（同济大学语言规划与全球治理研究中心）

摘　要：本文聚焦于人工智能大模型生成语言的治理现状与挑战，探究了生成式大模型带来的AI生成语言治理的新议题。本文回顾传统语言规划到现代语言治理的理论演进，提出生成语言治理需适应治理目标、路径及体系的深刻变化，尤其是在价值观对齐等关键问题上亟待应对。研究总结了国内外生成式AI治理的主要模式：以美国为代表的“弱监管”，依托公司自治和领域监管的结合；欧盟采取“强监管”模式，强调风险分级与透明性要求。然而上述模式在价值观塑造、公众知情权保障以及语料公平利用等方面仍存显著不足。通过分析AI生成语言在教育、舆论和媒体等领域的潜在影响，本文认为未来应提高公众数字素养，建立分级化、情境化的本地化治理方案以化解价值观风险。基于此本文提出了AI生成语言的研究进程：未来，应深化对生成语言价值观、用户认知与行为影响及其治理模式研究，构建符合社会公共利益与文化核心价值的AI价值观治理框架。从学科发展而言，语言学对AI话语的研究应坚持学科本位，融合语言学的传统范式与新技术背景下的理论创新，以守正创新的视角应对生成语言治理的复杂性与多样性挑战。

Abstract: This paper focuses on the governance status and challenges of artificial intelligence (AI) large language models (LLMs), exploring new issues in the governance of AI-generated language brought about by systems such as ChatGPT. It reviews the theoretical evolution from traditional language planning to modern language governance and emphasizes the need for governance frameworks to adapt to profound changes in objectives, approaches, and systems, particularly addressing key issues like value alignment. The study summarizes the primary governance models of generative AI both domestically and internationally, highlighting the “light regulation” approach represented by the United States, which relies on a combination of corporate self-regulation and sectoral oversight, and the “strong regulation” model of the European Union, which emphasizes risk grading and transparency requirements. However, these models exhibit notable deficiencies in areas such as value orientation, safeguarding

① 本文系教育部人文社科基金项目“人工智能生成语言的治理策略和路径研究”（23YJC740016）阶段性的成果。
② 通讯作者。

public knowledge rights, and equitable use of linguistic resources. By analyzing the potential impacts of AI-generated language in education, public opinion, and media, the paper advocates for enhancing public digital literacy and establishing localized, tiered, and context-sensitive governance frameworks to mitigate risks related to value systems. Based on this analysis, the paper outlines the future research trajectory for AI-generated language, calling for deeper investigations into value alignment, user cognition and behavioral impacts, and governance model design. It proposes constructing governance frameworks for AI values that align with societal public interests and core cultural values. From the perspective of disciplinary development, the paper argues that linguistic research on AI discourse should maintain a disciplinary focus, integrating traditional linguistic paradigms with theoretical innovations informed by new technological contexts. By adopting a perspective of balanced innovation, this approach aims to address the complexity and diversity challenges in the governance of AI-generated language.

关键词：ChatGPT；语言治理；大模型；生成式人工智能

Key Words: ChatGPT; language governance; large language model; generative artificial intelligence

一、语言治理研究的新课题

2022 年开放人工智能(OpenAI)公司发布了基于注意力机制的自注意力(transformer)模型 ChatGPT,这一人工智能的产生迅速引起了广泛关注。2025 年中国深度求索公司国产自主研发的大模型深度求索(DeepSeek)在数学问题、复杂推理等任务上达到了先进水平,宣布开源,这标志着中国在人工智能大模型竞争中取得了领先地位。大语言模型(large language model,简称 LLM)能够理解人作为对话者的提示词(prompt),并根据上下文提供接近人类自然语言的回答。随着 ChatGPT、深度求索等语言大模型性能的不断改善,其语言处理能力不仅能够满足日常会话,在知识问答、文章摘要、图形描述等各个细分领域甚至能够超越针对单一任务的自然语言处理算法。大语言模型自问世以来,人工智能生成内容(AI generated content,简称 AIGC)开拓了广阔的应用空间,大模型生成语言(又可以称为“人工智能生成语言”)成为理论研究的前沿议题。

从语言规划到语言治理,经历了语言本体、语言地位、语言教育(习得)、语言声望和话语规划和治理等理论阶段(沈骑,2019)。理论发展史契合了 20 世纪 50 年代以来殖民体系瓦解之后,各民族国家通过语言文字恢复国家民族意识、提高国民语言文字能力和建构身份共同体的现实诉求。实践中,国家通过语言文字立法、文字改革、语言文字规范化、语言保护、语言能力和语言服务等措施,推动语言文字事业不断向前发展(周庆生,2019)。

数字社会对语言治理提出了新的机遇和挑战(王春辉,2021)。面对生成式大模型不断发展,社会影响日益加深,语言治理理论产生了重要的理论鸿沟。从语言规划到语言治理研究,自始至终是以人类自然语言作为对象,其学科范畴一直是社会语言学的重要领域。由于语言使用者是人,具备社会属性,语言治理的目标通常和社会治理的宏观目标相契合,语言治理的手段主要以立法、教育、宣传等社会治理方式开展,正如语言治理学者巴尔多夫所说,“规划语

言就是规划社会”(沈骑,2021)。但随着大模型生成语言的广泛产生,语言治理的对象从人类自然语言向着机器生成语言拓展,作为“硅基生物”的人工智能算法并不具备社会性,传统的语言治理理论面临失效的难题。

当前关于大模型生成语言的研究尚属于起步阶段,它和人类自然语言治理的差异性主要体现在三个方面,考虑从以下三个方面进行理论化探索。首先从治理目标而言,大模型本质上是基于自注意力架构的预训练算法,高度复杂的基于注意力机制的神经网络算法是“黑箱”,不具备可解释性。这就带来一个关键问题:如何确保其生成语言的价值观,能够符合人类道德和国家社会总体发展目标?其二,在治理路径上:国外当前主要采取公司自治的方式;国内相关部门已出台若干治理意见,如《生成式人工智能服务管理暂行办法》等,但总体仍处于探索阶段,尚未形成成熟体系。其三,从治理体系来看:当前大模型的应用具有高度场景化特点。基于这一特性,如何针对不同的场景和使用者,构建不同领域和用户的分级治理体系,成为亟待解决的问题。

关于生成式人工智能的治理,有学者认为应该“将大模型工具化要求我们以工具观对其进行治理”,治理主要放在技术使用和场景上(饶高琦等,2023)。相关研究综述了国内外在大模型治理中的各项法案,将治理模式归纳为以美国为代表的“弱监管模式”和以欧盟为代表的“强监管模式”(王晓丽、严驰,2025)。当前研究的局限在于,大模型生成语言的版权问题、数据跨境安全等传统安全问题已经受到了一定重视和研究,而语言负载的价值观等非传统安全问题,当前研究相对匮乏。人工智能生成语言(AI generated language)的价值观导向和意识形态风险,以及人机互动中用户的认知受人工智能影响的程度,未来或将成为理论发展的探索方向。

二、大模型生成语言的国外治理现状

在大模型研发和应用领域,当前世界范围内主要是中美两大参与者。美国以开放人工智能、谷歌(Google)、微软(Microsoft)等公司为代表。中国以深度求索、华为、月之暗面等公司为代表。目前美国人工智能生成语言的治理具有“公司自治+领域监管”的总体特征。公司自治即由大模型的研发者和商业受益者负责生成内容的治理。“公司具有治理 AI 的重大责任”(Askell 等,2019),在 ChatGPT 的研发公司文件中也具有相关表述(OpenAI, 2023)。领域监管,是不同行业和应用场景下,相关行业和机构对大模型应用的监管,使大模型生成语言在该领域能够符合伦理道德地推广和应用。

在公司自治阶段,企业在研发和部署阶段采取技术措施规定语言产出的价值观。作为一种生成式大模型,大模型的研发经历了预训练、微调、奖励模型和强化学习四个阶段。在第三和第四阶段,以 OpenAI 公司为代表的技术企业会将大模型的生成结果与人类价值观参照,基于“帮助性(helpfulness)、真实性(honesty)以及无害性(harmless)”的 3H 原则,通过人工标注和反馈的形式,不断让大模型学习人类的价值观偏好,产出和人类价值观对齐的语言内容(张奇等,2023)。

在大模型的研发和版本迭代阶段,生成语言和人类价值观对齐是一项长期持续的工作。尽管早期版本的大语言模型,曾经生成过违背伦理的语料内容,存在一些漏洞能够绕开伦理审查机制。但随着学习机制的强化,大模型能够不断改善生成的语言,逐步修补相关的漏洞和问题。开放人工智能公司官方宣称它们实行了大模型的全生命周期(lifecycle)的管理,具体包

括预训练数据过滤、微调模型、风险分析、筛选有害模型输出、审查监督使用、监控滥用迹象、研究模型影响等方面(OpenAI, 2023)。

在大模型的应用阶段,领域监管是语言治理的第二重保障。公司自治的目标在于让大模型生成语言和人类基本伦理和价值观相一致,然而在不同的领域和行业,如何应用大模型生成语言是一个高度场景化的问题。在教育领域,不少学术机构及组织对大模型的使用,制定了相关的治理措施。哈佛大学(2023)、加利福尼亚大学洛杉矶分校(2023)、康奈尔大学(2023)等大学已出台了AI在教育教学中的使用指南。在AI教育使用指南中强调道德与伦理考量,确保公平、公正和透明,特别是在数据隐私和安全方面遵循严格规定。AI被视为辅助工具,增强个性化教学和评估,而非替代传统教育模式,且其决策过程需具备透明度和可解释性。此外,高校特别注重防止AI系统中的偏见,确保其应用公正无歧视,并促进教育机会的普及性。教师和学生需接受相关培训,确保能够有效使用AI工具,同时持续评估和改进AI系统的效果与影响。上述教育界的实践表明,AI在教学中难以完全禁止,人机协同是未来的发展趋势,应该确立合理的治理方案和教育政策,合理规范生成式AI介入教学的程度和使用场景。

目前AI治理的立法方案在国外已相继展开。2022年美国政府公示了AI立法蓝图,确立了五项基本原则。其中涉及生成内容和语言治理的,主要体现在"算法歧视保护"(algorithmic discrimination protections)原则(White House, 2023),即人工智能不应根据用户的种族、肤色、民族、性别、宗教、年龄和身份被区别对待。但实践中,由于AI算法的黑箱特征,很难确认其语言生成不以用户信息为基础,进行区别对待。2024年,欧盟出台《人工智能法案》,将人工智能技术划分为四个风险等级,不同的分类决定了企业分级的合规性要求。同时要求通用型人工智能满足特定透明性要求,保证用户和公众的知情权和投诉权利。从上述趋势判断,人工智能生成语言的治理,未来可能会向着分级治理、场景应用和确保公开的方向发展。

三、大模型生成语言的问题

国外当前采取的"公司自治+领域监管"的大模型生成语言治理策略,虽然具有一定的合理性,在技术和应用两个层面实现了人工智能生成语言总体符合社会发展需要,但当前在理论和实践方面仍存在不可忽视的问题和挑战。

第一,价值观对齐这一大模型生成中的关键步骤,由少数专家研究者定义。以ChatGPT为例,在奖励模型和强化学习阶段,参与标记、反馈和伦理审查的专家预计仅有数百人。在生成过程中,生成结果的价值观是由参数调整决定的,具体是生成哪一种语料,取决于调参时的人为因素,其中的一个关键步骤是价值观对齐。该过程通常由少数专家参与,他们主导了大模型生成内容的方式。然而,尽管训练过程中所使用的语料是全球共享的语言数据,数以亿计的用户却无法参与生成决策,只能被动接受AI生成的结果。可以认为ChatGPT生成语言在研发阶段,并非与人类价值观对齐,而是与几百个专家的价值观对齐。2020年,GPT相关科研和产业研发人员在一次学术会议中针对如何治理生成语言的偏见(bias)进行了讨论。由哪些人来对输出语料进行价值观对齐,成为该问题的关键(Tamkin等,2021)。从实际应用中发现,西方中心主义的价值观在大模型生成语言中屡见不鲜。ChatGPT在生成涉及中国问题的语言中,始终罔顾中国立场和利益。可见,大模型并非是没有国界的价值观中立的产物,其生成语言具有高度的"意义建构性"和"价值观导向性"。尽管OpenAI公司声称立场中立,但在14项政治倾向性的研究中,ChatGPT的回答具有鲜明的左倾政治观点(Rozado, 2023)。

第二，大模型的数据在接入各种社交媒体、搜索引擎和知识问答等工具之后，语言生成即传播。普通用户难以分辨自己面对的语言内容，到底是由人产生，还是人工智能的产出。人工智能生成语言事实上已经成为虚拟世界的舆论参与者，存在舆论战、数据污染等风险（陈积银、贾超然，2023）。用户无法参与甚至难以获知大模型价值观对齐背后的决策者，在大模型的使用和传播中，无意识地接受了生成话语的价值观和偏见。未来对于大模型在舆论和媒体中的使用，需要有更加严格的监管和透明度要求。明确对生成内容的监管责任，完善 AI 内容生成规则，在社交媒体和舆论阵地中对大模型生成内容进行适度标识。

第三，训练大型语言模型的语料库是人类共有的知识、文本和信息的集合，它们是全球人类共同劳动的产出和智力成果。科技公司在研发这些模型时，往往以“免费”方式，甚至未经授权地使用了这些宝贵的语料资源。2023 年，《纽约时报》对开放人工智能公司（OpenAI）和微软提起诉讼，指控这两家公司未经授权使用其数百万篇文章来训练人工智能聊天机器人，如 ChatGPT。这一诉讼揭示了在训练大型模型过程中，未经授权使用相关语料（United States District Court，2024）。

由于大型模型的预训练依赖于人类共有的知识语料，这些模型应当被视为人类知识和语言文字的公共设施。它们的产出和收益应当类似于水电煤气等公共设施，由社会共享其成果。这意味着，大型语言模型的开发者和使用者应当考虑到这些模型对公共资源的依赖，并确保其使用和收益能够公平地回馈语料的原始创作者。由于成本收益不匹配，长期缺乏回馈机制，机器生成将大量替代人类的原始创作。

第四，大型语言模型，依托于海量人类知识语料库，通过自注意力架构进行预训练，形成了一种先进的计算机算法。这些模型生成的语言是基于给定提示词后的概率计算结果。尽管在注意力机制和神经网络等方面，大型语言模型与人类大脑存在一定的相似性，但由于其生成内容受限于概率计算，其创新能力终究有所局限。这些生成语言往往语义重复度高，雷同性强，并可能包含计算机幻觉所产生的虚构内容。

广泛使用大型语言模型生成的语言可能会限制人类创造性地使用语言的能力。特别是在教育领域，独立思考和批判性思维的培养对学生至关重要。如果学生依赖于大型语言模型生成的语言来解答本应由自己独立思考和批判性思考的问题，这可能会对这些关键思维能力的培养产生负面影响。

总之，AI 时代的语言治理，特别是大模型生成语言的治理，本质上已从传统意义上的语言本体和语言地位的治理，向着话语和数据的治理转变（郭书谏、李晓阳，2024）。需保障公众对大模型价值观对齐的知情权和参与权，确保本国大模型生成的语言符合社会公共利益和社会主义核心价值观。同时应提高公众的数字素养和语言安全意识，让使用者能够更好地理解和评估大模型生成语言的可靠性和安全性。

四、语言规划视域下大模型生成语言研究的研究议题

基于上述分析，大型语言模型生成的语言具有一系列独特的属性：它们承载着特定的价值观，依赖于公共所有的语料库，基于概率计算，具有可迭代性，并且能够处理多模态数据。人工智能生成语言的治理不再仅仅是传统语言规划研究中所涉及的“使用哪种语言”或“如何规范使用语言”的问题。随着语料的不断丰富、技术的持续进步和模型的迭代更新，未来大型语言模型有望精通所有语种，并能以比人类更规范的方式生成任何语言。目前，大型语言模型在

某些情况下表现出的性别歧视、语言暴力等问题,有望在后续的迭代训练中得到逐步纠正。

因此,对人工智能生成语言的研究不应仅仅停留在列举个案的"捉虫"式研究上。对于某位研究者生成的特定语料,很可能在后续的研究中无法复现。人工智能生成语言的研究应当以问题为导向。具体而言,以下三个细分领域可能成为语言规划视域下大模型生成语言关键的研究议题。

第一,大模型生成语言的价值观研究。特别在当前大国博弈的背景下,ChatGPT 生成语言承载和传播的是哪一种价值观?尽管生成的语料可以多种多样,但由于大模型在研发过程中经过了价值观对齐的流程,它所体现的价值观一定具有较为稳固的特征。

第二,人与大模型交互过程中的认知和行为研究。研究聚焦于分析大型模型输出内容对人的认知、理念和理解的影响,以及这些影响如何进一步转化为行为变化。例如,研究学生在与大型模型互动时,其输出内容如何塑造学生对学科概念和学习材料的认识。这些研究有助于我们理解大型模型在教育和信息传播中的作用,以及它们如何塑造学习者的知识结构和思维方式。行为研究进一步明确了大型模型生成语言应用的边界,探讨了在何种程度上,这些模型生成的内容能够改变人的认知和行为。这涉及对大型模型输出内容的深度分析,以及对用户反应的细致观察。通过这些研究,我们可以确定大型模型在社会互动中的角色,评估它们在塑造公众意见、影响决策过程和促进知识传播方面的潜力和限制。

第三,大模型生成语言的治理模式研究。当前的"公司自治+领域监管"的模式,只能算是新技术产生之后摸着石头过河的探索性方案。大模型在教育领域的应用缺乏足够的参考数据和理论支持,而 AI 生成语言的价值观风险在这一领域尤为突出。由于学生在缺乏适当限制和引导的情况下广泛使用 AI 工具,这一风险得到了进一步的强化。随着 AI 技术在教育中的普及,学生对其生成内容的依赖可能导致他们在无意识中接受特定价值观的灌输,从而影响其独立思考和价值判断能力。因此,确保 AI 在教育场景中的价值观对齐和风险管控变得尤为重要。未来或可针对不同学科和学龄的特点,制定具有可参考性的指导意见。适时可以通过实名制管理,避免学生因过早接触生成式大模型造成学术能力发展的缺陷。

五、探讨大模型生成语言的治理建议

大模型领域的竞争门槛极高,不仅需要海量的语料和算力资源,还需要高水平的科研人才。目前,全球范围内,主要的大模型竞争者仅限于中美两国。在此背景下,开放人工智能公司所提出的治理方案,前文已有讨论,其决策过程由少数技术和伦理专家"黑箱"式地控制,数十亿用户则被动地接受并传播这些决策。这种治理模式不仅忽视了科技伦理中的用户知情权,同时数以亿计的用户无法了解 ChatGPT 的价值观和伦理规则是如何确立与调整的。更为严重的是,ChatGPT 在回答问题时输出了带有偏见的语言,违背了科技向善的基本立场。因此,理论和实践都需要中国及相关企业提出更为合理且具有伦理保障的治理方案。

相较于国外人工智能大模型的价值观对齐和生成内容由数家大型科技企业垄断,中国人工智能深度求索让 AI 成为开源的公共产品,有效破解了传统闭源模式下的技术和伦理黑箱困境。此模式创新性地实践了"部署者即责任主体"原则,将伦理审查、内容监管等治理职能分布式让渡到各社会组织和个体,形成"基础模型研发者、领域部署商和用户"的三级治理体系。在创新生态构建方面,任何组织和个体可通过微调(fine-tuning)、提示工程(prompt engineering)及插件扩展等路径,微调符合自己价值目标的 AI 生成语言。中国的人工

智能治理，将更有利于实现价值多元、开放共享的人工智能愿景，让不同行业、不同领域、不同需求的用户都能在这个生态中发挥作用，共同推动人工智能技术朝着更加健康、可持续的方向发展国产人工智能治理。然而，在取得显著成效的同时，中国人工智能治理仍存在若干亟待完善的关键方面。

首先，当前国产大模型亟需提高并拓展其多语种能力，以更好满足国际受众的多语种需求。然而，现阶段主要国产大模型主要局限于中英文服务能力，在面对法语、西班牙语、阿拉伯语等非中英语种的提示词时，普遍缺乏准确的理解和反馈能力。由于国产大模型在多语种支持上的短板，大量非中英母语用户只能选择 ChatGPT 或其他国家的生成式人工智能，强化了西方中心主义和价值观的传播。

第二，未来大模型生成语言的价值观应该体现更广泛的代表性和公开性。相关高科技企业在大模型研发过程中，需重视并投入更多资源研究 AI 价值观与人类价值观的对齐问题，提高透明度，接受公众监督，减少决策和筛选的“黑箱”操作。AI 应该是充分体现人类多元价值的人工智能，不能仅代表部分科技精英群体的价值观。

第三，重视和中国相关关键词和提示语的生成语料质量。作为未来重要的信息基础设施，大模型不只是信息获取的渠道，也是讲好中国故事的重要窗口。关涉本国本民族的语料生成，应体现正确的价值立场和利益诉求。增进民族情感，促进民族团结，推动中华民族共同体意识的形成（高伟，2023）。

第四，大模型生成语言治理的领域监管应该符合中国国情，制定符合中长期发展的制度设计和发展规划，逐步建立覆盖教育、医疗、文化创意、传媒等领域的分类治理方案。以教育为例，大模型生成语言适合应用于哪一类的教学场景、哪一层级的课程，以及应以何种方式引入其生成的语料，需要结合科研产出和实践数据进行综合评估，实现有效利用。在这一点上未来国内外 AI 生成语言的治理可能存在显著差异。

大模型生成语言的治理方案应坚持中国语言治理的核心价值体系（王春辉，2020），实现善治良治，而非走西方在技术治理方面的老路——以新自由主义为标准，或以技术先进和效率至上为唯一准则。回顾西方新技术治理模式的发展史，其文化价值观层面的治理逻辑和生态环境领域存在相似性，从工业革命以来的“先污染后治理”的老路，不应在大模型治理的过程中重现。

六、结语

自 ChatGPT 和 DeepSeek 等生成式大模型出现以来，AI 生成语言的治理成为语言治理研究领域重要的新兴课题。过去的语言治理理论脱胎于 20 世纪 50 年代以来的民族国家语言规划，伴随着语言管理理论、批判范式和话语治理走向 21 世纪。今天语言治理第一次面向机器生成语言，而非人类自然语言的治理，其理论和实践之间形成了显著的鸿沟。大模型生成语言治理的研究问题、理论视角和研究范式，是这一领域的框架性问题。作为该新兴领域的探索性研究，本文综述了当前大模型生成语言治理的基本模式：“公司自治+领域监管”，批判性地提出了这一模式存在的三大问题。特别是生成式人工智能成为舆论场的重要参与者之后，AI 生成语言给受众群体带来的价值观风险，以及人机交互、人机共创过程中的认知影响，均值得关注。

以 OpenAI、Google DeepMind 为代表的西方科技巨头通过闭源模式构建技术壁垒，形成了高度集中的 AI 治理体系。这种集中化模式虽然确保了技术标准的统一性，但也带来了价值单

一性和创新垄断等结构性风险。中国人工智能 DeepSeek 开创性地探索出一条差异化发展路径。通过将 AI 技术开源化、公共产品化,DeepSeek 不仅打破了技术垄断的藩篱,更重要的是构建了更加多元开放的新型治理范式。

在人工智能技术迅速发展的背景下,AI 治理成为各个学科的重要研究议题,AI 的版权、数据安全、经济效益和社会意义等相关研究已经相当丰硕。然而从语言学学科进行的研究还不多见。“学科是学术研究发展到一定水平的产物。解决学科的问题,是为了让学术力量更强大,更好地研究社会问题。”(李宇明,2020)在人工智能时代与语言学学科问题转型的双重背景下,AI 的价值观、话语机制、认知影响及其治理路径,正是新技术发展背景下催生的新理论和实践问题。AI 时代的语言学研究应坚持守正创新,将这些研究范畴和概念纳入学科话语体系,避免学科话语权被边缘化。

参考文献

[1] Askell, A., Brundage, M., and Hadfield, G. “The Role of Cooperation in Responsible AI Development.” <http://arxiv.org/abs/1907.04534n> (accessed 2024-05-12).

[2] Cornell University. “Artificial Intelligence (AI).” <https://it.cornell.edu/ai> (accessed 2024-05-12).

[3] Harvard University. “Guidelines for Using ChatGPT and Other Generative AI Tools at Harvard.” <https://provost.harvard.edu/guidelines-using-chatgpt-and-other-generative-ai-tools-harvard> (accessed 2024-05-12).

[4] OpenAI. “Lessons Learned on Language Model Safety and Misuse.” <https://openai.com/research/language-model-safety-and-misuse> (accessed 2023-10-07).

[5] ——. “Our Approach to AI Safety.” <https://openai.com/blog/our-approach-to-ai-safety> (accessed 2023-10-07).

[6] Rozado, D. “The Political Biases of ChatGPT.” *Social Sciences*, (3)2023: 148.

[7] Tamkin, A. et al. “Understanding the Capabilities, Limitations, and Societal Impact of Large Language Models.” <http://arxiv.org/abs/2102.02503> (accessed 2024-05-12).

[8] United States District Court Southern District of New York. <https://www.courthousenews.com/wp-content/uploads/2023/12/new-york-times-microsoft-open-ai-complaint.pdf> (accessed 2024-03-26).

[9] University of California, Los Angeles. “Guidance for the Use of Generative AI.” <https://teaching.ucla.edu/resources/ai_guidance/> (accessed 2023-10-18).

[10] White House Office of Science and Technology Policy. “Blueprint for an AI Bill of Rights.” <https://www.whitehouse.gov/ostp/ai-bill-of-rights/> (accessed 2023-10-15).

[11] 陈积银、贾超然.ChatGPT 技术对国家安全领域的挑战与应对策略研究——基于传播学的分析视角.《青海社会科学》, 2023(3): 127-131.

[12] 高伟.中华民族共同体意识与国家通用语言文字互塑的历史演进.《青海社会科学》, 2023(2): 140-145.

[13] 郭书谏、李晓阳.基于大数据的语言治理研究: 内涵、方法与应用.《云南师范大学学报(哲学社会科学版)》, 2024(1): 46-53.

[14] 李宇明.语言学研究：问题的“问题化”.《东北师大学报(哲学社会科学版)》, 2020(5)：21-29+192.
[15] 饶高琦、胡星雨、易子琳.语言资源视角下的大规模语言模型治理.《语言战略研究》, 2023(4)：19-29.
[16] 沈骑.中国话语规划：人类命运共同体建设中语言规划的新任务.《语言文字应用》, 2019(4)：35-43.
[17] ——.全球语言治理研究的范式变迁与基本任务.《语言文字应用》, 2021(3)：30-40.
[18] 王春辉.论语言与国家治理.《云南师范大学学报(哲学社会科学版)》, 2020(3)：29-37.
[19] ——.学科建构视角下的语言治理研究.《陕西师范大学学报(哲学社会科学版)》, 2021(6)：155-163.
[20] 王晓丽、严驰.生成式 AI 大模型的风险问题与规制进路：以 GPT-4 为例.《北京航空航天大学学报(社会科学版)》, 2025(2)：17-27.
[21] 张奇等.大模型理论与实践. <https://intro-llm.github.io/> (accessed 2024-05-12).
[22] 周庆生.中国语言政策研究七十年.《新疆师范大学学报(哲学社会科学版)》, 2019(6)：60-71+2.