



青年记者
Youth Journalist
ISSN 1002-2759, CN 37-1003/G2

《青年记者》网络首发论文

题目：生成式人工智能中的关键语言问题解析
作者：余梦珑，沈阳
DOI：10.15997/j.cnki.qnjz.20240520.001
网络首发日期：2024-05-21
引用格式：余梦珑，沈阳. 生成式人工智能中的关键语言问题解析[J/OL]. 青年记者.
<https://doi.org/10.15997/j.cnki.qnjz.20240520.001>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

生成式人工智能中的关键语言问题解析

● 余梦琰 沈 阳

摘 要：语言作为人类社会交往的核心工具和文化载体，其在生成式人工智能中的模拟与应用不仅关系到技术的发展与边界，更涉及社会价值传达与伦理风险规范等深层次影响。聚焦于生成式人工智能中的关键语言问题，从人机对话的自然语言演进脉络出发，可以剖析大语言模型所生成的“类人”语言文本背后的底层逻辑与局限性，探索语言生成方式改变与数字语言技术全方位嵌入社会情境后对认知行为、权力表达及文化价值等维度的综合影响，并围绕语言伦理探讨生成式人工智能的道德规约，对未来人机对话语言系统发展路径做出启发性思考。

关键词：生成式人工智能；大语言模型；人机交互；语言媒介；智能融合

生成式人工智能技术的突破标志着新一轮人工智能革命的来临。在此浪潮中，语言是人类与智能设备共同参与和对话的核心媒介，是信息编码与解码的表征以及认知与沟通的桥梁，也是生成式人工智能技术理解、模拟与反映人类思维和文化的关键载体。由 OpenAI 开发的 ChatGPT 等生成式人工智能已近乎实现人类语言的表达与理解能力，除在技术层面推动了人机交互的自然性，还在更广泛的社会文化层面对人类的语言使用与社会功能提出了重新考量的可能性。对生成式人工智能中的关键语言问题展开研究，不仅有助于理解生成式人工智能语言背后的内在逻辑，亦对揭示新型语言生态与社会文化的交互影响具有重要价值。

一、文献回顾与研究问题

生成式人工智能正引发跨学科研究与应用变革，带来学界与业界关于人工智能和社会发展的新思考。目前，国内外相关研究主要集中在：(1) 技术领域的语言模型结构设计与训练算法^[1]；(2) 应用领域的新闻、医疗、教育、法律等场景功能与机遇挑战^{[2][3]}；(3) 社会领域的技术伦理、文化偏见与安全性等问题^{[4][5]}。值得关注的是，语言在这场技术变革中扮演着核心的角色。作为人机交互的桥梁，语言既是输入的媒介，

也是输出的表达形式，在对话沟通、理解需求并提供高质量响应等方面起到至关重要的作用^[6]。从语言学的角度来看，生成式人工智能如何捕捉、理解与生成自然语言，尤其当其处在复杂、模糊和具有歧义的情境中时，正成为研究新焦点^[7]。国内外相关研究较为关注模型多语言性能^[8]、模型处理各种语言任务能力^[9]与适应性研究^[10]。

但整体而言，目前从语言学视角对生成式人工智能的研究还不够充分，特别是在探讨其对人类语言深层结构的复杂性模拟与动态表征，语言价值观内在联系与差异以及语用层面的交互影响等方面，缺乏系统性研究。这些缺陷可能会遮蔽对生成式人工智能实质性语言能力、潜在局限性与风险的全面认识，从而制约其在更为复杂的语言应用场景中的表现与发展。基于此，本研究将围绕生成式人工智能涉及的关键语言问题，从语言的结构性、功能性与社会性维度出发，探讨生成模型在语言技术、语言特点、语言影响与语言伦理方面的内在逻辑及其如何与人类语言的社会属性相互作用，从语言学视角对生成式人工智能的未来发展做出启发性思考。

二、技术与话语：人机交互何以自然

生成式人工智能在人机对话系统的发展可从两

个视角看待：一是作为问答系统的最新进展，从简单的基于规则与模板的响应生成转向更复杂与动态的语言模型；二是作为全新的突破，实现了人机对话的交互变革，体现了从单向编码指令到以人类语言进行对话，再到具有生成与涌现性的动态交互语言演进过程。

（一）代码至语言：交互自然驱动技术变迁

自1940年代计算机进入社会视野，如何让其“理解”执行任务所需要的指令，一直是“语言工程师”面临的重要挑战。作为区分于“人类语言”的“人工语言”，计算机语言（Computer Language）是指用于人类与计算机之间传递信息的语言，总体可分为机器语言、汇编语言、高级语言三大类^[11]。从二进制代码指令到高级程序语言，其交互需严格遵守预设语法规则且操作门槛较高。随着个人电脑的问世与普及，为了使计算机成为非专业人士也能有效操作的工具，语言工程师在开发中强调了“用户友好”概念，使用改编的人类语言即创建可以模仿人类语言知识与表现的程序^[12]，以此实现人机交互的“自然性”。交互自然成为人机对话技术发展的重要驱动力，而自然语言处理（NLP）则致力研究、提出与开发以编写自然语言为主要对象的计算系统^[13]，提供并实现了以人类语言进行人机交互的可能性。

（二）指令到对话：自然语言处理构建桥梁

自然语言处理构建了以人类语言进行人机交互的桥梁。根据算法所使用语言知识的获取方式，自然语言处理任务可以分为基于规则的方法与涉及机器学习的方法。其中，基于规则的系统由语言学家与计算机科学家以某种形式将语言知识进行直接编码，通常是对某些特定开发数据的分析；而基于机器学习的系统通常是通过计算上下文中语言属性出现频次来编译，在机器学习方法中，即使是无监督方法，语言知识也可以通过其他方式发挥作用。语言学家与计算机科学家在自然语言处理中的协作体现在，语言学家基于人类语言研究为专业系统开发提供必要的语言理论与知识，计算机科学家则将这些语言学知识编码进算法和模型中，使机器能够理解与生成人类语言。其理解与生成能力包括文本分析、情感分析、机器翻译、语音识别等，也推动

了苹果Siri等智能语音虚拟助手（Virtual assistants）的发展。但此阶段的智能语音助手与问答系统主要依赖于与语料库的匹配和固定回答，而大语言模型所具备的生成与涌现能力让人机交互体验开始迈入新阶段。

（三）生成与涌现：大型语言模型智能变革

大语言模型（LLM）是自然语言处理技术的一种实现形式，被训练用以理解与生成人类语言，包括语境、语法与语义规则。ChatGPT等大语言模型所具备的处理系列语言任务能力，印证了人机交互中可被广泛应用的语言力量（power of language）^[14]。Transformer（转换器）架构^[15]与底层注意力机制的使用大大提高了语言模型处理自然语言文本中远程依赖关系的能力，即在生成预测时使用自注意力机制（self-attention mechanism）来确定不同部分的相关性^[16]，使模型能够更好地理解输入不同元素之间的关系，从而根据上下文和用户需求，动态生成符合语境的回复，打破了基于预设规则与模板，机器只能根据固定指令给出有限回应的传统人机交互模式。这些技术发展也推动了ChatGPT等大语言模型的“涌现”，即模型可执行开发者未明确编程的任务，通常基于大规模语料库的学习与训练，是模型自我提取与深层理解的语言模式。这种涌现能力使得大语言模型能够在复杂动态与充满不确定性的人机交互场景中，除了可以生成连贯与可理解的类人语言文本并用自然语言保持灵活对话，还能处理更为复杂的语言任务，是人机交互迈向自然的重要环节并推动内容生产领域的智能变革。

三、模拟与再塑：类人文本的拟像解构

人类语言与人工智能生成语言正共塑全新语言样态，解构大语言模型所生成的类人语言文本拟像，需要从形式层面的模拟逻辑、理解层面的响应轨迹以及语境层面的处理机制等多个维度展开。

（一）语言形式：表层拟真与幻觉事实

从形式语义的角度来看，语法与语义之间存在组合关系。二者的合规组合能够使大语言模型生成语法流畅的类人语言。在预训练阶段，大规模语料库中的深度训练，使其学会关于人类语言

的知识,包括语法、事实与推理能力等。微调阶段,通过更具体的数据集对模型进行人工微调,使其能在特定任务或领域内有更好的表现并遵循系列指导原则。基于大语言模型所学习的语言模式、常识信息与各类语境信息,最终生成符合语法、逻辑连贯且符合人类表达习惯的拟真文本。然而,生成式人工智能的显著问题之一是通过提示语生成的看似高质量的回答,可能是“幻觉事实”(hallucinated facts)与误导性结论,这是自然语言处理模型中的常见问题^[17],体现出生成式人工智能语言表面合理性与现实真实性间的差异。有研究者也将其称为“令人信服的华夫饼干”(cogent waffle)^[18],即将数据训练材料中提取的“文本块”按照准确语法拼接为“饼干”,而这些文本块可能并无意义或包含错误信息。人工智能幻觉的产生可以从两方面进行阐释。

一是基于统计的预测,相关研究指出大语言模型并不是为了以人类思维“理解”语言,也不是为了实际产生语法结构而编写,而是生成基于统计计算组合在一起的词汇元素串。通过“词嵌入”(word embeddings)即一个词相对于其他词出现的统计概率,模拟人类语言行为。也有研究者将其称为“统计鹦鹉”(statistical parrots)^[19],这个预测过程是基于统计学而非基于事实校验。因此,模型可能生成在语法与语境方面看似合理但在事实真实性方面存在问题的输出。

二是数据驱动的特性,大语言模型是通过大规模文本数据集进行预训练学习语言模式,并基于输入文本与训练数据中的模式匹配来生成响应。当训练数据中包含错误、偏见或其他形式的误导信息时,模型可能会将这些信息学习并内化,进而在生成过程中产生幻觉事实。由于语言模型并非以人类感知思维对文本内容进行理解,当其以同样高的置信度陈述事实与谎言或幻觉时,将导致人机交互中的信任问题。

(二) 语义理解:“中文房间”与映射响应

ChatGPT等大语言模型展现出了强大的语言理解能力,能够对人类提问进行自然有效的回答,但在人工智能领域关于机器能否拥有类人语言理解

能力或意识一直是争议焦点。“中文房间”(Chinese Room)是哲学家约翰·希尔斯(John Searle)在1980年提出的思想实验^[20],旨在对机器“真正理解”与“模仿理解”进行探讨。该实验中一名只懂英语的实验者在“房间内”借助中文指南回应“房间外”传入的中文问题,在不懂中文的情况下遵循指南规则给出了正确的中文答案,让房间外误认为他(或整个房间)理解中文。该实验旨在质疑计算机或人工智能只是在遵循预设规则进行操作,并不能真正地理解语言。事实上,人类语言理解充满了特定案例、特定受众与特定领域。究其根本,大语言模型与人类理解语言不同的核心差异是缺乏类人功能语言能力所需的概念理解,即在现实世界中强有力地理解与使用语言的能力。

人类理解语言(以及非语言信息)需要拥有语言(或其他信息)所描述的概念,而不仅是语言符号的统计特性。与机器相比,人类对科学与日常生活中的理解形式具有强烈的内在驱动力^[21],对语言的理解无论是在个体还是集体层面,都可视为在构建高度压缩且基于因果关系的世界模型。由于现阶段基于语料训练的生成式人工智能尚未有参与世界的经验或心智模型,因此,本质上在大量文本集合中预测词频的训练所形成的是语言形式而非感知意义。正如“流泪”一词人类映射为一种感觉,而语言模型是映射为另一组词。

(三) 语境感知:软性模型与价值编码

语境在语言学和社会交往领域被视为影响语言使用与理解的环境因素,是信息交流的背景与参考框架。在人机对话中正确识别与处理语境不仅是信息有效传递的关键,更是确保双方能够在动态对话场景中保持高效与深度交流的前提,但人类与机器的语境处理机制存在本质差异。

人类语言中的“语境感知”,涉及复杂的认知机制。通常基于先前经验、文化背景和情境信息,其解析受到个体认知与情感驱动并与内在价值观相对应。语境中的价值传达不仅涉及表层信息,还有系列关于认同、属性与情感的复杂编码。

生成式人工智能的“语境处理”,在某种程度上可以识别与适应对话并提供合乎语境的回应,其背

后的价值编码更侧重于数据中的统计规律。从语境输入层面来看,生成式人工智能的训练数据不仅包含文本“能指”,还蕴含着数据背后深层次的社会文化背景与价值观。正如 Crawford 所言,人工智能系统并不是中性工具,而是具有政治性的社会技术,人类价值观与信仰不可避免地影响了其设计^[22]。从语境输出层面来看,人工智能生成结果会映射出其训练数据中的语境并对已编码价值观进行再现。因此,尽管 ChatGPT 等大语言模型声称并不具有价值观或意识形态,但其训练数据与语言学习过程中所吸收的“语境”因素以及模型设计者设置的具体规则,会让模型输出符合特定价值观或道德标准的内容,其呈现的价值观可能并不总是中立,这就需要在使用时意识到价值映射的底层逻辑并尽可能地消除或减轻负面影响。需要注意的是,生成式人工智能对语境的塑造更贴近于“软性模塑”,即并非严格遵循某一确定性规则,而是在多样性和概率性的框架下进行操作。除了调整其输出以适应不断变化的对话语境,还在人机对话迭代中构建出新语境。

简言之,生成式人工智能输出既反映了训练数据的语境,又在特定语境中涌现新语言观点,并以此对语境进行重新解释与塑造。厘清生成式人工智能语言与人类语言在形式层、理解层、语境层的基本性质与关联,能够为正确衡量生成式人工智能如何与现实世界进行社会互动及其潜在影响奠定基础。

四、交互与融合:生成式人工智能的语言影响

语言作为人类重要的交流工具与文化符号,对个体认知、群体共识与社会文化产生综合影响。随着人工智能语言模型的迭代与嵌入教育、医疗、法律、新闻、决策等现实应用场景,其社会效应将逐渐显现,审视其语言实践如何影响语言意识形态和语言认识论具有必要性。

(一) 语言作为媒介:沟通载体形塑认知行为

语言是人类交流与文化传播的基本载体,在将语言作为特定历史、话语与政治条件下形成的社会建构实体进行讨论时,物质与技术实践发挥着重要

作用。从造纸术、印刷术到现代电报、电话再到互联网,媒介技术对语言的建构、传播与使用方式产生了深远影响。书写与印刷技术推动了语言的标准化,催生了如现代汉字、拉丁字母等书写系统,并推动构建了国家语言;电报与电话改变了远程交流的语境,塑造了新的语言风格和交际礼节;互联网的普及催生了网络语言,诸如热词与表情符号等,新的语言元素不断深化人类对语言动态变化的理解。随着计算机科技、人工智能与自然语言处理技术的发展,机器能够理解与生成人类语言,颠覆了对语言创作的传统理解,为语言建构提供了新的视角与可能性。正如 Gershon 所指出的,技术结构允许和限制通过该媒介进行交流的方式,也塑造着通过该技术结构的内容形态^[23]。生成式人工智能驱动的语言技术将作为物质媒介链接社会各运行体系,并形塑群体认知与行为。与此同时,从技术的社会建构^[24]视角来看,人类产生与传达意义的语言符号本身也是一种物质类型,是在社会关系中的产物和工具。这种形塑与被形塑的认知行为实践过程是数字语言技术与人类社会发展的双向建构。

(二) 语言作为权力:意识形态渗透群体共识

语言既是一种政治工具,也是一种权力工具。作为形塑与传播意识形态的重要途径,国家语言标准的建设和国家归属感的培养离不开语言文字与传播技术^[25]。正如书面语言构成了人类社会的大部分系统内容、规则规范、期望惯例,人工智能语言技术发展带来人类语言生成与感知方式的系列变化时,也将带来权力隐性渗透。大语言模型的训练文本与生成文本不仅反映了语言的表层结构,也隐含了深层的语言意识形态(language ideologies)^[26]。通过语言选择与使用,语言技术背后的话语结构与社会语言等级制度可以潜移默化地影响公众价值观与行为并塑造群体共识。

值得注意的是,机器学习技术需要投入大量计算资源与数据资源,目前先进的生成式人工智能语言系统主要由大型商业公司研发。这就不可避免地暗示了发展语言技术的目标是资本主义性质,即盈利与高效的工作并让用户满意。照此逻辑,商业数字语言技术不是基于教育、审美或语言规范,而是

基于客户（机器可读）的语言实践。从更深层次的角度来看，数字人工智能系统行业与全球社会政治权力等级制度纠缠在一起，Crawford 指出，“人工智能系统的建立是为了以主要有利于它们所服务的国家、机构和公司的方式看待与干预世界”^[27]。从这个意义上说，生成式人工智能系统是来自经济与政治力量层面的权力表达，并影响个人与群体（如少数民族）获得资源、表达意见以及在社会中得到代表的平等机会。

（三）语言作为文化：交往实践共现惯习思维

生产式人工智能对社会文化的影响，既包括模型输出所反映与强化的社会文化价值观，也包括模型广泛使用对社会交往模式与文化实践的影响。在语言人类学（linguistic anthropological）中，语言被视为互动实践而非独立的抽象系统，是构建与维持社会关系、传递文化知识、表达身份与情感的重要工具，也被称为语言化（linguaging）^[28]。在大语言模型的技术创新扩散过程中，通过协助人类完成各项语言任务的方式广泛参与到社会实践进程，并在互动实践中对文化与思维产生影响。在语言与文化的研究中，“萨丕尔—沃尔夫假说”（Sapir-Whorf hypothesis）指出，“人类语言塑造了人类思维模式，不同语言表达决定了认识世界的不同方式”，该假说也被称为“语言相对主义”（linguistic relativity），是关于语言、文化与思维三者关系的重要理论^[29]。这一假说强调了语言不仅是思维的工具，也是塑造思维的力量。在人机交互实践中，双方需要不断接受与适应彼此的“思维”方式，尽管机器“思维”在一定程度上是模型训练数据中反映的人类思维模式，但大语言模型基于概率与统计处理语言的方式以及量化与数据驱动的生成对话方式，也会潜移默化地改变人机交互的语言习惯与语言实践方式。此外，人机交互实践中大语言模型的涌现可能促成新的社会文化现象产生与传播，同时也蕴含着未被意识到的伦理风险。

五、规约与边界：生成式人工智能的语言风险

生成式人工智能语言伦理研究关乎技术与社会

责任的交织，涉及如何保证语言技术系统在复杂的社会与文化背景中做出符合人类道德规约的决策。美国人工智能研究员 Eliezer Yudkowsky 曾指出“人工智能最大的危险也许在于人们过早下结论认为可以理解它”^[30]。对于研究者与实践者而言，应该对技术发展保持警惕，意识到人工智能的缺陷并留意人工智能应用的潜在陷阱。从语言学的视角来看，生成式人工智能语言模型潜在的伦理风险问题包括但不限于以下四个方面。

（一）语言真伪：语言工具武器化应用

随着生成式人工智能与大语言模型的技术突破，生成式人工智能及其下游应用（downstream applications）的机会、破坏性社会影响与道德问题逐渐得到关注。作为一项颠覆性技术，生成式人工智能技术可能被“武器化”应用，并作为传播阴谋论、制造假新闻、支持伪科学等的工具。一方面，当大语言模型生成的内容在不同媒介中进行传播且未被标识时，可能被误认为是人类言语，制作出深度伪造内容，造成信息混淆并损害公众信任与利益。另一方面，由于模型无法确保提供完全正确的内容输出，当其提供错误或劣质信息而用户又无法对这些信息进行辨别时，就会产生明显的信息与价值误导。因此，对大语言模型的使用需要进行严格监管与控制，目前中国、美国、欧盟等都加快了对生成式人工智能的监管与立法进程，确保其在合理合法范围内使用，引导生成式人工智能技术的向善发展，避免其被武器化应用。

（二）语言黑箱：可解释性与监视主义

大语言模型的黑箱性涉及语言的可解释性与透明度。可解释性强调模型对生成语言过程的解释能力，即模型的理解能力如何被人类理解；透明度主要关注模型内部工作机制的可见性。尽管大语言模型可以生成符合语义与语法规则的自然语言，但对于普通用户而言往往难以理解其决策过程与预测模型行为，增加了模型使用风险。此外，和语言黑箱性相关的伦理风险还有数据隐私与控制问题。技术扩散在一定程度上意味着监视资本主义（surveillance capitalism）蔓延^[31]。数据是信息时代的“黄金”，从内容生产、选择到消费都围绕其展开，而数据驱

动经济模式的背后是权力资本介入。在缺乏有效监管的情况下,用户数据可能被滥用于商业盈利、信息交易甚至政治操控,这种对用户数据的过度获取与使用,被称为监视资本主义。大语言模型通过分析 & 学习人类语言行为提供多功能语言服务,在该过程中语言数据被广泛收集与使用,个体往往没有对于生成数据的实际控制权,而是技术平台方对其进行数据掌控,黑箱性使得监视资本主义更难以被察觉。

(三) 语言价值: 偏见歧视与道德困境

Curchod 曾指出,生成式人工智能的广泛使用可能导致人们越来越依赖算法作为重构社会关系的机制^[32],算法对人类互动的影响将逐步深入认知、行为与决策并加剧道德困境^[33],例如偏见与不公平。在语言学研究中“书面语言偏见”(written language bias)早已被证实^[34],这种偏见也伴随着大语言模型技术的发展与应用。麦肯锡(McKinsey)公司强调人工智能面临着克服“潜伏在更广泛的互联网与社会暗流中的偏见挑战”^[35],这对于生成式人工智能尤为重要。大语言模型中偏差在输入端的蔓延,可能导致传播错误信息与助长偏见,模型受到数据可用性、选择和确认偏差(confirmation bias)影响^[36]。事实上,如果没有用于训练大语言模型的数据质量详细信息,就很难衡量这种偏见的规模。

Rozado 对 ChatGPT 进行了 15 种不同政治取向的测试,其中有 14 个回答呈现出“左倾”观点偏好。该研究指出,有道德的人工智能系统应围绕问题向用户提供平衡的观点,并避免在“声称中立”的同时,生成存在明显政治偏见的内容^[37]。ChatGPT 中嵌入政治偏见的原因是其从互联网收集大量文本语料并进行模型训练,而用于训练的语料主要由西方社会中有影响力的机构主导和塑造。这些机构基于自身政治立场与利益诉求,可能在信息生产和传播过程中有意或无意地引入政治偏向。大语言模型基于这些带有偏向性的文本数据上进行训练,可能习得并内化文本中既有的政治倾向。同时大多数算法的架构设计、训练与参数调节都与语言相关,模型和过滤器设计中有意或无意的架构决策也可能造成政治偏见。

(四) 语言等级: 价值排序与数字鸿沟

大语言模型所生成的语言涉及复杂的社会语言学问题。社会语言学等级(sociolinguistic hierarchies)即多语种社区中不同语言的社会地位与权威^[38],包括语言标准、语言价值、语言使用方式等。诸如哪些语言享有较高社会地位,被认为是“标准”“正式”或“高级”语言,而哪些语言可能被视为次级语言或边缘语言。社会语言学等级涉及语言使用和变异如何反映社会结构与权力动态的问题。在数字语言技术普及的时代,哪些语言资源与信息被默认为有价值,将与大数据语料库、算法机制、平台语言意识形态以及由此产生的用户实践相关。目前国际上先进的大语言模型训练首先围绕英语展开,这与美国在人工智能技术的行业领先地位和英语在世界范围内的普及程度相关。这也导致了在 ChatGPT 中并非所有的语言都是平等的^[39]。尽管中文、西班牙语和法语是世界常用语言,但模型响应丰富程度与语言书写的可理解性明显不如英语。Bender 等人的研究指出,“全球超过 10 亿人使用的语言中,90% 以上目前在语言技术方面仍然没有得到支持”^[40]。大语言模型的“反馈循环”效应可能进一步扩大数字语言技术鸿沟与提升英语在全球语言等级体系中的优势地位。研究不足的语言,甚至可能被排除在这场生成式人工智能革命之外。

六、结语

生成式人工智能与人机对话正处于关键的技术转折点,其变革不仅启发研究者从外部视角对人类语言的深层结构与功能展开全新认知,也将为人类与机器间的交互协同以及跨界领域融合提供思考。从狭义的语言对话形式到更广泛意义上社会文化维度的对话,从问答式响应交互到对话式沟通交互,从文本语言交互到多模态感知交互,语言学视角下生成式人工智能的未来发展充满多元可能,伴随其广泛嵌入应用场景带来的社会变革将是未来研究需要关注的重要议题。

【本文为国家社科基金重大项目“基于机器博弈的网络信息传播安全多准则动态管控策略研究”(批准号:19ZDA329)阶段性成果】

参考文献:

- [1] J. Gui, Z. Sun, Y. Wen, D. Tao and J. Ye. A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(4): 3313–3332.
- [2] Ferres JML, Weeks WB, Chu LC, Rowe SP, Fishman EK. Beyond chatting: The opportunities and challenges of ChatGPT in medicine and radiology[J]. Diagn Interv Imaging, 2023, 104(6): 263–264.
- [3] 彭兰. 从 ChatGPT 透视智能传播与人机关系的全景及前景[J]. 新闻大学, 2023(04): 1–16+119.
- [4] Alawida M, Mejri S, Mehmood A, et al. A comprehensive study of ChatGPT: Advancements, limitations, and ethical Considerations in natural language processing and cybersecurity[J]. Information, 2023, 14(8): 462.
- [5] 刘艳红. 生成式人工智能的三大安全风险及法律规制: 以 ChatGPT 为例[J]. 东方法学, 2023(04): 29–43.
- [6] Schneider B. Multilingualism and AI: The regimentation of language in the age of digital capitalism[J]. Signs and Society, 2022, 10(3): 362–387.
- [7] 袁毓林. 人工智能大飞跃背景下的语言学理论思考[J]. 语言战略研究, 2023, 8(04): 7–18.
- [8] Meo R, Nai R, Sulis E. Explainable, Interpretable, Trustworthy, Responsible, Ethical, Fair, Verifiable AI. What's Next? [C]. European Conference on Advances in Databases and Information Systems. Cham: Springer International Publishing, 2022: 25–34.
- [9] 郭全中, 朱燕. 广布局、重应用: 生成式大语言模型的新进展[J]. 新闻爱好者, 2023(08): 21–25.
- [10] 肖辛格, 储小静, 刘云. ChatGPT 对我国语言战略的冲击与应对[J]. 天津师范大学学报(社会科学版), 2023(05): 65–73.
- [11] 罗钧旻, 钟联炯. 计算机语言与计算机语言学[J]. 西安工业学院学报, 1994(04): 319–323.
- [12] 赵京胜, 宋梦雪, 高祥. 自然语言处理发展及应用综述[J]. 信息技术与信息化, 2019(07): 142–145.
- [13] Grishman R. Computational linguistics: an introduction[M]. Cambridge University Press, 1986: 29.
- [14] Chaudhry I S, Sarwary S A M, El Refae G A, et al. Time to Revisit Existing Student's Performance Evaluation Approach in Higher Education Sector in a New Era of ChatGPT—A Case Study[J]. Cogent Education, 2023, 10(1): 2210461.
- [15] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[C]. Proceedings of NAACL-HLT, 2019: 4171–4186.
- [16] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]. Proceedings of the 31st Conference on Neural Information Processing Systems, 2017: 1–15.
- [17] Rohrbach A, Hendricks L A, Burns K, et al. Object Hallucination in Image Captioning[C]. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018: 4035–4045.
- [18] Vincent J. OpenAI's new chatbot can explain code and write sitcom scripts but is still easily tricked[EB/OL]. (2022–12–02). <https://www.theverge.com/23488017/openai-chatbot-chatgpt-ai-examples-web-demo>.
- [19] Bender E M, Gebru T, McMillan-Major A, et al. On the dangers of stochastic parrots: Can language models be too big? [C]. Proceedings of the 2021 ACM conference on fairness, accountability, and transparency. 2021: 610–623.
- [20] 蔡曙山. 哲学家如何理解人工智能: 塞尔的“中文房间争论”及其意义[J]. 自然辩证法研究, 2001(11): 18–22.
- [21] 邵俊宗. 语言发展论纲[M]. 北京: 中国社会科学出版社, 2017: 104.
- [22] Crawford K. The atlas of AI: Power, politics, and the planetary costs of artificial intelligence[M]. Yale University Press, 2021: 201.
- [23] Gershon I. Media ideologies: An introduction[J]. Journal of Linguistic Anthropology, 2010, 20(2): 283–293.
- [24] Bijker W E. Social Construction of Technology[M]. A Companion to the Philosophy of Technology. UK: Oxford Press, 2009: 88–94.
- [25] 辛斌. 语言、权力与意识形态: 批评语言学[J]. 现代外语, 1996(01): 21–26+72.
- [26] Blodgett S L, Barocas S, Daumé III H, et al. Language (Technology) is Power: A Critical Survey of “Bias” in NLP[C]. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020: 5454–5476.
- [27] Crawford K. The atlas of AI: Power, politics, and the planetary costs of artificial intelligence[M]. Yale University Press, 2021: 47.
- [28] Love N. On languaging and languages[J]. Language Sciences, 2017, 61: 113–147.
- [29] 胡壮麟. 语言学教程[M]. 北京: 北京大学出版社, 2002: 69.
- [30] Ali M J, Djalilian A. Readership awareness series paper 4: Chatbots and chatgpt—ethical considerations in scientific publications[C]. Seminars in ophthalmology. Taylor & Francis, 2023: 1–2.
- [31] Zuboff S. The age of surveillance capitalism[M]. Social Theory Re-Wired. Routledge, 2023: 203–213.
- [32] Curchod C, Patriotta G, Cohen L, et al. Working for an algorithm: Power asymmetries and agency in online work settings[J]. Administrative Science Quarterly, 2020, 65(3): 644–676.
- [33] 朱毅诚, 刘玉涵, 蒋雪妍. ChatGPT: 生成式人工智能形塑社会认知的能力与风险[J]. 中国传媒科技, 2023(08): 7–13.
- [34] Linell P. The written language bias in linguistics: Its nature, origins and transformations[M]. Routledge, 2004: 88.
- [35] McKinsey & Company. What is generative AI? [EB/OL]. (2023–01–19). <https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-generative-ai>.
- [36] Rich A S, Gureckis T M. Lessons for artificial intelligence from the study of natural stupidity[J]. Nature Machine Intelligence, 2019, 1(4): 174–180.
- [37] Rozado D. The political biases of chatgpt[J]. Social Sciences, 2023, 12(3): 148.
- [38] Makihara M. Linguistic syncretism and language ideologies: Transforming sociolinguistic hierarchy on Rapa Nui (Easter Island)[J]. American Anthropologist, 2004, 106(3): 529–540.
- [39] Seghier, M. L. ChatGPT: not all languages are equal[J]. Nature, 2023, 615(7951): 216.
- [40] Bender, Emily, Angelina McMillan-Major, Timnit Gebru, and Shmargaret Shmitchell. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? [C]. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. Association for Computing Machinery, New York, 2021: 610–623.

(余梦琰: 清华大学新闻与传播学院博士后;
沈阳: 清华大学新闻与传播学院教授、博士生导师)