

大模型时代的社会科学，何去何从？

The Landscape of Social Science in the Era of Large Language Models

谢天¹ 邱林² 李雨瞳² 罗殷³ 刘盼⁴
XIE Tian QIU Lin LI Yutong LUO Yin LIU Pan

(1. 武汉大学哲学学院心理学系, 武汉, 430072; 2. 新加坡南洋理工大学社会科学学院心理学系, 新加坡, 639798; 3. 新加坡南洋理工大学跨领域学院, 新加坡, 639798; 4. 湖南大学公共管理学院, 长沙, 410082 / 1. Department of Psychology, Philosophy School, Wuhan University, Wuhan, 430072; 2. School of Social Sciences, Nanyang Technological University, Singapore, 639798; 3. Interdisciplinary Graduate School-Global Asia, Nanyang Technological University, Singapore, 639798; 4. School of Public Administration, Hunan University, Changsha, 410082)

摘要: 文章探讨了大语言模型 (如 GPT) 对社会科学研究带来的革命性影响。首先介绍了大语言模型的原理, 而后举例并论述了大语言模型在文献综述、提出理论观点、研究设计、数据收集与分析等环节为社会科学研究带来的改变, 最后提出了大语言模型将产生的新的研究范式与导向, 以及社会科学研究者面临的新挑战。

关键词: 人工智能; 大语言模型; GPT; 社会科学; 研究方法; 研究范式

中图分类号: C1; TP18 **DOI:** 10.13366/j.dik.2023.06.006

引用本文: 谢天, 邱林, 李雨瞳, 等. 大模型时代的社会科学, 何去何从? [J]. 图书情报知识, 2023, 40(6): 6-9, 30. (Xie Tian, Qiu Lin, Li Yutong, et al. The Landscape of Social Science in the Era of Large Language Models[J]. Documentation, Information & Knowledge, 2023, 40(6): 6-9, 30.)

Abstract: Large language models (such as GPT) have brought unprecedented breakthroughs in the field of artificial intelligence. In this article, we briefly introduce large language models, and illustrate how such models have been applied in each stage of social science research, including literature review, theory formation, research design, data collection, and data analysis. We envision new research paradigms brought by large language models, and new challenges faced by social science researchers.

Keywords: Artificial Intelligence; Large language models; GPT (Generative Pretrained Transformer); Social science; Research methods; Research paradigm

生成式大语言模型 (简称大模型) 横空出世, 在人工智能领域带来了前所未有的突破。它们能够参与复杂的逻辑推理、理解和生成自然语言, 甚至显示出对社会现象深刻的洞察与新颖的解释。对于社会科学研究者, 大模型意味着什么? 它仅仅是一个研究工具的升级, 还是对传统研究范式的挑战乃至重塑? 本文将从大模型的原理和它在研究方法中的实用案例, 来探讨大模型将如何重构社会科学的研究方式。

1 生成式大语言模型简介

生成式大语言模型, 如GPT (Generative Pretrained

Transformer)^[1], 最初是为了研发一个智能填词工具。比如, 有一个填空题: 天空是_____。通过分析海量文本与基于神经网络的机器学习算法, GPT能填上: “蓝色的”, 因为根据计算, “蓝色的”出现在“天空”后的概率最高。这种能力被有的研究者戏称为“随机鹦鹉” (stochastically parroting)^[2], 因为GPT无法将“天空”这个词与实际的天空对应起来, 即它并不“知道”每个词所对应的意义是什么。

然而, 令人惊讶的是, 这只“随机鹦鹉”在学习了海量的训练数据后, 形成了有万亿个参数的大模型, 能产生与人类创作媲美的文本、影音等内容, 甚至表现出与人类相当的心智能力。例如, GPT-3.5能站在他人视角上看待问题^[3] (theory of mind), 在一系列认知

[基金项目] 本文系教育部人文社会科学重点研究基地重大项目“跨文化传播视域下的国际舆论朋友圈建设研究” (22JJD860013) 和2023年武汉大学-昆山杜克大学联合科研平台种子基金支持计划“文化演化的多学科交叉融合科研平台” (WHTUDKUZZJ202306) 的研究成果之一。(This is an outcome of the Major Project “Constructing the ‘Circle of Friends’ in Global Media from the Perspective of Intercultural Communication” (22JJD860013) supported by Key Research Base for Humanities and Social Sciences of the Ministry of Education, and the project “Multidisciplinary Integration Platform for Cultural Evolution” (WHTUDKUZZJ202306) supported by a grant from the Joint Research Platform Seed Foundation Supporting Program of Wuhan University-Duke Kunshan University in 2023.)

[通讯作者] 邱林 (ORCID: 0000-0002-3587-5371), 博士, 副教授, 研究方向: 计算社会科学、社会心理学、文化心理学、人格心理学、网络心理学、人机交互, Email: linqiu@ntu.edu.sg. (Correspondence should be addressed to QIU Lin, Email: linqiu@ntu.edu.sg, ORCID: 0000-0002-3587-5371)

[作者简介] 谢天 (ORCID: 0000-0002-2201-7634), 博士, 副教授, 研究方向: 社会心理学, Email: thanksky520@126.com; 李雨瞳 (ORCID: 0000-0003-2298-3426), 博士研究生, 研究方向: 计算社会科学、人格心理学, Email: yutong001@e.ntu.edu.sg; 罗殷 (ORCID: 0009-0005-4418-3173), 博士研究生, 研究方向: 计算社会科学、环境心理学, Email: luoy0038@e.ntu.edu.sg; 刘盼 (ORCID: 0000-0001-5411-4191), 博士, 副教授, 研究方向: 行为公共政策、大数据与网络心理学, Email: liupan@hnu.edu.cn.

与决策任务中的表现与人类相当^[4]，“涌现”出的类比推理能力甚至超过了人类^[5]。在经过一定训练后，它还能在内部生成一个理解时间与空间的世界模型^[6]。部分研究者甚至认为人工智能(AI)即将产生出自我意识，并呼吁对未来潜在风险进行管控^[7]。

2 大模型赋能社会科学研究

传统的社会科学研究，往往由研究者主导进行文献综述并提出研究问题和理论观点，设计研究，收集数据，分析数据。大模型现在已能够赋能每个研究环节。

2.1 综述文献并提出理论观点

文献综述是学术研究的起点。对于单篇论文，GPT能帮助研究者迅速总结出论文的研究问题、方法、主要结论等框架性要点^[8]。而对于文献的全面总结，尽管存在不足，GPT已经能够辅助开展系统性文献综述(systematic review)^[9]。综述文献后，发现研究问题并提出理论观点通常被认为是整个研究的核心。此过程涉及的创造性思维也会被认为是人类智慧的关键且充满神秘色彩。但创造性思维可以被简化为在问题空间中进行有效搜索，同时生产可能的解决方案(即理论或假设)^[10]。这样一来，人工智能就可以通过搜索过程显示出创造力。因此，人工智能助力的理论创新或科学发现已蔚然成为一个新的研究领域^[11]。

来自意大利的一个研究团队在2023年5-6月就开展了一系列实验，以检验GPT能否能在商业理论与实践提出新的理论观点。他们首先在“商业战略与创新管理”及“组织与创新文化”两个领域中找出了10个重要的理论及实践，然后让ChatGPT学习这些内容，并提出颠覆性的新观点。他们发现，虽然ChatGPT的输出有时会混淆有价值的和没价值的信息，但并没有出现明显错误。更关键的是，他们发现GPT提出了非常创新的观点，有时甚至超过了专家的水平。这说明GPT已经能够提出具有创造性的理论观点了^[12]。

2.2 设计研究

大模型能帮助研究者进行研究设计。比如，无论是

定量、定性还是混合方法，基于研究问题的性质和已提出的理论观点，GPT能识别并推荐最匹配的研究方法。接下来，在操作化研究概念和变量时，GPT可以通过访问广泛的文献库，确保所选变量的定义、测量和解释符合研究目的。最后，GPT还可以根据研究的样本分布为分层抽样或样本配对提供建议，以确保样本的代表性^[8]。

在制作研究材料方面，如访谈提纲或问卷题项，GPT也能助研究者一臂之力。在我们的一项关于松-紧文化^①的研究中，需要创作一些情境短文，来测量被试者对短文中信息的敏感度。我们期望每个短文以某人进入一个陌生环境为情境，并包含“必须”或“可以”做的事情。我们输入了以下指令给GPT(原文为英文)：

请制作一个150字左右的场景，描述一个孤独的旅行者偶然发现了一个偏僻、与世隔绝的村庄。请在情境中包含6件旅行者必须做的事情。在描述这些事情时，请务必使用“必须”一词。此外，还请包含6件旅行者可以自由选择做的陈述，并确保使用“可以”一词。还请保持“必须”和“可以”句子里表达的情绪在程度上相当。

GPT输出了符合我们要求的短文，但不够浅显易懂。于是，我们输入了以下指令给GPT(原文为英文)：

请调整语句，使没有经过大学教育的人也能理解。

GPT根据指令对原先的短文进行了修改。我们邀请了实验助手对GPT生成的实验材料进行了人工评估。结果表明，GPT生成的实验材料完全符合我们的要求。我们使用这些材料收集数据，并获得了良好的效果。这表明，GPT能够产生有效的研究材料。

2.3 收集数据

如果大模型获取了足够多的有关人类社会的知识，并掌握了人的思考方式，那么在理论上它就可以模拟人类作答心理测试与问卷调查。这将取代人类被试在数据收集中的必要性^[13]。为此，我们在2023年6月开展了一项研究，来检测GPT是否能取代大规模民调。皮尤研究中心(Pew Research Center)是美国的一个知名调研机构，主要对社会、公共和政策问题进行调查分析。它于2022年3月调查了10,441位美国公民对俄乌冲突和拜登政府应对方式的看法，该调查包括8道题。例如：

① 松-紧文化中的“紧”指规范强，对偏差行为包容度低；而“松”是指规范弱，对偏差行为包容度高。经典研究见：Gelfand M J, Raver J L, Nishii L, et al. Differences between Tight and Loose Cultures: A 33-Nation Study[J/OL]. Science, 2011, 332(6033): 1100-1104[2023-10-01]. <https://www.science.org/doi/10.1126/science.1197754>.

“您对拜登政府对俄乌冲突的反应持何态度？”（1：非常支持~4：非常不支持）^[14]。由于GPT-4的训练数据仅更新至2021年9月，因此它对于2022年2月发生的俄乌冲突及这个有关俄乌冲突的民调并不知道。如果GPT能产生出与实际民调结果一致的数据，那就能为GPT模拟受访者回答问卷提供有力的证据。

我们首先向GPT介绍了俄乌冲突和以上民调的题项内容，然后要求其预测，美国受访者在每个问题上选择各个选项的百分比。与实际调查结果对比，GPT-4的预测平均绝对误差（MAE）为9.59%。也就是说，尽管存在偏差，但GPT的预测与实际调查结果相当接近。具体而言，在8个问题中，GPT在7个问题上的预测与调查结果的选择方向相符，即它能相对准确地预测出更多人选择“支持”或“不支持”的倾向，或预测出“重大威胁”大于“小威胁”大于“非威胁”的选择比例。我们又对另外三个皮尤研究中心的民调做了相同的实验，并得到了类似的结果。这说明，GPT有模拟受访者产生大规模调查数据的能力。

不仅是模拟调查数据，大模型甚至还催生一种全新的数据生产方式，即由大模型控制的虚拟智能体（agent）模拟出个体与群体行为。斯坦福大学的研究团队建立了一个模拟小镇^[15]。在这个小镇里有25个以GPT为“大脑”的智能体，他们能产生拟人程度很高的个体行为。比如，一个智能体早晨起床后会刷牙，洗澡，边看新闻边吃早饭，并与家人聊天。同时，他们也能表现出自主的群体行为。例如，当研究者让一个智能体组织一个派对，这个智能体就会自动把派对的信息告诉其他智能体，其他智能体就会相互约定一起去，最后准时在派对举行的地点的出现。这样的智能体与过去的基于主体的建模（Agent-Based Modeling）中的个体不同。他们不再局限于建模时的既定行为规则，而是能够在互动中自动产生规则^[16]。这样的“斯坦福智能体小镇”能在虚拟环境中产生类似真实生活中的交互场景，为社科研究者收集个体与群体行为数据提供了极大的便利。

2.4 分析数据

在定性数据分析中，最耗时耗力的工作就是对文本进行编码。在演绎编码（deductive coding）中，研究人员需要根据预先确定的编码本将新的数据标记为一组固定的编码。最近的研究发现，GPT-3可直接用于各

种演绎编码任务，无需微调（finetune）就能与人类专家编码的结果保持一致^[17]。我们的研究也显示，GPT对中文与英文文本的编码能力很相近。而对于多人协作的编码任务，也已经有研究者开发了基于GPT的AI辅助编码工具，并通过了可用性与其有效性评估^[18]。对于定量数据分析，GPT不仅可以协助研究者进行数据清洗、异常值检测、变量转换、缺失值分析等基础工作，而且对于描述统计、回归分析、多变量统计等，GPT基本能够根据研究者需求生成相应统计软件的指令，甚至自动生成结果报告。尽管在某些任务上表现不佳，但鉴于GPT的综合效率和潜力，有研究者认为大模型的出现会对数据科学产生巨大改变^[19]。

3 人工智能驱动的社会科学

传统的社会科学研究多为理论驱动（theory-driven）。但在过去的10年，随着社交媒体与大数据的出现，浮现出大量数据驱动（data-driven）的社会科学研究^[20]。在当今的大模型时代，社会科学研究将逐渐成为人工智能驱动（AI-driven）。因为如前文所述，大模型已能在社科研究的每一个环节发挥重要作用。具体来说，大模型已能够通过文献综述提出研究问题和理论观点，产生研究材料，模拟并收集研究数据，以及分析数据。现今，大模型在每一个环节里的功能是孤立的，不连续的。但是，一旦大模型能够将这些功能整合在一个自动系统中，即它能独立提出研究问题，并为这个研究问题产生研究材料，然后应用这些材料去模拟或收集数据，最后分析数据来回答最初的研究问题，那么社会科学研究将由人工智能来主导。社科研究者的角色有可能从主导者逐渐转变为辅助者。

同时，原来以解释导向（explanation-oriented）为主的社会科学，将逐渐转化为解决导向（solution-oriented）。这是因为以往很难在现实生活中去尝试并检验某个解决方案所产生的实际效果。而现在，在类似“斯坦福智能小镇”这样的多智能体模拟环境中，研究者不仅可以观测并解释某个社会问题是怎么产生的，而且可以尝试不同的干预策略，来找到最佳的解决方案。这将够使社会科学研究更有实效性，并颠覆理论在社科研究中的主导地位。

大模型人工智能驱动的新研究范式与研究者的角

