

# 智能之光：大语言模型驱动的经济管理 研究方法论变革

胡诗云 易君健\*

**摘要：**人工智能大语言模型正在重塑知识的生产函数。本文面向经济管理学者，系统阐述其技术原理、应用边界及其引发的方法论变革。本文指出，基于抽象概念构建、反事实推理与价值判断的理论创造力，是人类科学家不可替代的核心优势。大语言模型可在研究中扮演四种角色：参谋、助研、智能体与朋友。作为参谋，它可以增强批判性思维与文献对话；作为助研，它能够实现研究执行的自动化；作为智能体，它本身成为实验工具和新的研究对象；作为朋友，它还可以提供心理与社交支持。然而，大语言模型的广泛应用也可能带来理解深度退化、研究内容同质与同行评议崩溃三大系统性风险。未来的经济管理研究需要在拥抱“智能之光”的同时直面其“阴影”，以人机协作的新范式提高研究效率并拓展人类思维的广度和深度，最终回归思想创新的学术本质。

**关键词：**人工智能；经济学方法论；人机协同

**DOI：** 10.13821/j.cnki.ceq.2026.03.06

## 一、引言

智能技术革命正重塑全球学术生态。人工智能大语言模型(以下简称大模型)与传统自动化有本质区别：大语言模型实现了“认知工作的自动化”(automation of cognitive work)(Korinek, 2023)。它们能够理解复杂概念、生成连贯文本并进行多步推理。如果说上一轮自动化取代了蓝领工人并催生了彼得·德鲁克(Peter Drucker)所谓“知识工作者”(knowledge worker),那么大语言模

\* 胡诗云,北京大学中国经济研究中心、北京大学国家发展研究院、北京大学数字金融研究中心;易君健,北京大学中国经济研究中心、北京大学国家发展研究院。通信作者及地址:易君健,北京市海淀区北京大学国家发展研究院,100871;电话:010-62767530;E-mail: junjian@nsd.pku.edu.cn。易君健感谢国家自然科学基金(72595871、72533001)的资助;胡诗云感谢北京大学国家发展研究院校友基金课题的资助。作者感谢张皓辰、尹晔、李欣元、王若璧、江弘毅、周子焜、吴浩哲、林轶凡、颜玮、赵家琪、曹治、李星宇、王玥琴、韩旭、周清源、王悦、李奕天、汪姝君、王鹏飞参与讨论并贡献提示词案例,感谢梁葆崑出色的助研工作,感谢江深哲与两位匿名审稿人的宝贵意见,文责自负。

型就触及了知识工作者任务的核心。

对于作为典型知识工作者的经济管理研究者而言,大语言模型带来的既是机会也是挑战。机会之处在于增强(augmentation)。Eloundou et al.(2024)的评估发现,经济学家100%的工作活动可在AI与其他软件辅助下提效一倍以上,综合暴露度处于所有职业的前25%。挑战在于自动化替代(automation substitution)。例如Manning et al.(2024)提出了“自动化社会科学”(automated social science)的概念,从提出假说、实验设计、实验开展、数据分析到文章撰写全部由AI完成。Q. Chen et al.(2025)构建了会进行回归分析的AI智能体,Novy-Marx and Velikov(2025)更是用AI全自动生成了96篇实证资产定价论文。经济管理研究者会被AI替代吗?在人工智能的时代,经济管理研究者如何找到与AI的协作之道?

本文旨在向经济学家、管理学家和其他社会科学研究者全面地介绍大语言模型的技术和应用。本文首先从社会科学知识生产的一般过程出发,讨论大语言模型如何改变知识的生产函数。大模型可以在研究的各个环节实现常规认知任务的自动化执行,在收集数据、文献汇总、数理推导、数据分析、文字撰写等方面提高社会科学研究的执行效率,扩充社会科学家的知识库,拓宽其认知边界,增强其表达能力。但是,大语言模型作为语言统计模型的本质,决定了它没有提出理论问题的动机,也缺乏从现实数据中抽象出新概念并构建理论的能力。这为社会科学研究中应用大语言模型提供了思考框架,并且为更好的人机协作提供了启示。

随后本文介绍了大语言模型的技术原理和应用方式。大语言模型并不神秘,它只不过是利用大量真实文本训练出的词语预测器。透过大语言模型相关的术语,我们可以更好地看出大模型日新月异发展背后的本质,破除对它们的迷信,并且更加正确、高效地利用人工智能。随着相关计算框架的成熟,本地部署大模型、微调大模型已经不是高科技公司的专利,而是每一个社会科学研究者都能运用的工具。

在介绍技术基础之后,本文展示大模型在研究全过程中的应用。人工智能大模型在研究过程中将会扮演四种角色:参谋、助研、智能体,以及朋友。作为参谋,AI可以协助创意生成、文献阅读、研究设计等工作,增强研究者的批判性思维。作为助研,AI能够支持数理分析与数据处理。作为智能体,AI不仅可以用于行为模拟与社会实验,还有望成为经济活动新的主体与经济学新的研究对象,为经济管理研究开拓新的范式和领域。最后,研究者也可以将AI作为自己的另一个“朋友”,让自己在研究之路上不再孤单、走得更远。对于每一种应用,我们都会提供相应的案例以及提示词,并总结相关的提示技巧。本文所展示的案例远不能覆盖所有的应用场景,希望这些示例能抛砖引玉,启发读者们

开发出更多更有创意的应用。

大模型的应用虽然可以提高个人效率,却可能造成学术界的系统性风险。本文最后识别了三大潜在风险:理解深度的退化、研究内容的趋同与同行评议的崩溃。从本质上看,这是因为大模型放大了学术界固有的道德风险。面对难以逆转的 AI 应用大潮,学术界一方面要强化对原创性和可靠性的激励机制,另一方面更要提高自身的甄别能力,在系统层面通过人类的不可替代能力来驾驭 AI。

本文融方法论哲学、技术原理与具体案例于一体,系统讨论了人工智能在社会科学中的应用及其影响。在本文之前,Bail(2024)从改善调查问卷、实施在线实验、自动化内容分析以及基于主体建模四个方面讨论了生成式人工智能改善社会科学的潜力。Korinek(2023)的综述性论文则探索了大语言模型在经济学研究中的数十个具体应用场景,并根据技术发展,每半年更新一个版本。除此之外,还有大量文献总结了大语言模型在文本分析(De Kok, 2025; 赵宣凯等, 2025; 林建浩和孙乐轩, 2025)、经济学实验(Charness et al., 2025)、行为模拟(Gao et al., 2024)等方面的具体应用。本文从五个方面对这一支文献进行了深化:一是从社会科学研究的内核与大语言模型技术的本质属性出发,识别了大语言模型在社会科学知识生产中的应用边界以及人类不可替代的作用,为更好的人机协作提供了理论指导;二是通过学科话语转译,向经济管理研究者深入地介绍大语言模型的技术路线及最新发展;三是提出 AI 在研究流程中扮演的“四种角色”框架,更加系统、全面地梳理和拓展了其在经济管理研究中的应用场景;四是基于卢卡斯批判和认知经济学视角,对 AI 行为科学的相关文献进行了批判性整合;五是从思想市场的信息不对称本质出发,推演了生成式 AI 广泛应用的三大系统性风险,为 AI 时代的学术共同体建设提供参考。本文讨论了 AI 技术,更希望通过让更多的研究者理解 AI 技术,通过这项技术来减少重复劳动与技术学习的时间,让经济管理研究从堆砌工作量与嫁接高级方法的竞争中解放出来,更加关注思想本身的创新性和深度。

## 二、人工智能时代社会科学家的角色

在具体讨论生成式人工智能原理和应用之前,我们首先要思考一个元问题:在人工智能时代成为一名社会科学家意味着什么?在计算机普及之前的年代,计算员(computer)曾经是一个重要、受尊敬,甚至需要较高教育水平的职业,而如今他们早已被电子计算机完全取代。人工智能时代的社会科学家与计算机时代的计算员相比要幸运得多,因为我们承担着更加复合的任务:我们怀揣着好奇心提出问题,带着问题观察现实社会,根据现象提出理论,再用新的观

察检验理论,最后用文字或行动,为思想和社会带来变革。与此同时,AI作为一种通用技术(*general purpose technology*)(Eloundou et al., 2024),对于研究工作的几乎每个任务都能产生深远影响,将会改变知识的生产函数。

我们应该从社会科学研究的本质思考,在一个科学家与AI共同进入的知识生产函数中,有哪些认知任务是可以由人工智能自动化的、哪些是可以由人工智能增强的,又有哪些是无法被人工智能取代的?只有识别出独属于人类智慧的核心任务,我们才能更好地与人工智能合作,完成社会科学家的使命。

### (一) 人工智能可以做什么?

人工智能从广义上理解,就是构建智慧机器(*intelligent machine*)的科学与工程(McCarthy et al., 2006)。从人工智能概念提出到现在,其发展大致可分为三个阶段:早期基于符号和预定规则推理为主的阶段、统计机器学习阶段,以及生成式人工智能阶段。随着人工智能的数据表达能力和泛化性能逐步增强,人工智能变得越来越接近于人类智能,其对社会科学的影响也不断加深。

早期的符号推理阶段以人工编写规则为核心,主要依赖专家知识,难以处理社会科学中复杂多样的概念与语言。进入统计机器学习阶段以后,人工智能以高维、非线性统计模型的面目出现在社会科学家的“工具箱”当中,主要包括LASSO回归、支持向量机、决策树与随机森林、神经网络等模型(Athey and Imbens, 2019; Gu et al., 2020)。例如,对经济变量进行更准确的预测(Cong et al., 2025; Gu et al., 2020)、使用非线性模型获得更加稳健的因果参数(Chernozhukov et al., 2018)以及估计异质性处理效应(Wager and Athey, 2018)、通过预测性标签处理文本等非结构化数据(Gentzkow et al., 2019; Mullainathan and Spiess, 2017),等等。然而,统计机器学习对于社会科学家在统计分析之外的工作流程却影响有限,对思想创新作用更是微乎其微。究其根本原因,一方面在于社会科学家工作任务流的多样性和灵活性,另一方面也在于它们只能拟合数据而无法捕捉概念。

生成式人工智能,尤其是大语言模型的出现,极大拓展了人工智能在社会科学中的应用场景(Bail, 2024; Korinek, 2023)。社会科学高度依赖自然语言,自然语言不仅描述了人类的行为,也是社会科学家沟通思想的载体。大语言模型通过学习语言的统计规律,存储知识并展现概念之间的联系(Farrell et al., 2025),在一定程度上可以泛化出新的语句并模仿逻辑推理。

从最浅层看,大语言模型能够帮助社会科学家完成所有与文字相关的日常工作。它可以帮助我们阅读文献、修改语言。它还通过语言捕捉了人类的行为模式,从而为非结构化的数据分析以及行为模拟开启了新的可能。而从更根本的层面看,它通过语言捕捉了已有概念以及这些概念之间的联系,从而可以帮

助社会科学家获取知识、启发思想、改进理论。正如深度学习可以快速生成并筛选大量蛋白质结构一样，大语言模型也可能促进概念的“组合式创新”，启发研究者提出新的研究问题和假说，成为创新的工具（method of innovation）（Agrawal et al., 2018; Gans, 2025）。比如，给定人工智能话题，AI可以启发研究者去关心AI对于劳动力市场就业量、工资、性别平等问题的影响，并对这些研究议题的可行性进行评估。

## （二）人工智能做不了什么？

看上去，AI似乎在社会科学知识生产的每一个流程中都能发挥作用，那么社会科学家最核心的能力在哪里？为了回答这个问题，我们不妨基于当下的人工智能技术路线，把AI的能力推向极致，看看到底有哪些能力是AI无法取代的。简而言之，人类科学家不可替代的核心价值，在于一种基于惊奇、价值判断和抽象概念创新的理论构建能力。

我们邀请读者参与一个思想实验：一个超级大模型被投喂了亚当·斯密（Adam Smith）之前的所有人类著作——包括柏拉图的《理想国》、圣经、历史学家的编年史、商人的零散账本、各种法律条文、戏剧和诗歌。它还具有极致的推理速度，能够在一瞬间输出大量的文本。然后我们问：它能否独立提出关于劳动分工和市场交易的理论？

我们的判断是，它极有可能描述出劳动分工的现象，但很难独立提出一套像《国富论》那样的、具有抽象性和因果解释力的理论。其中的差别，揭示了AI的本质能力和人类社会科学家的独特角色。

AI能力的天花板取决于它作为统计模型的本质。语言的统计模型具有三个功能：识别相关性、聚类词汇，以及生成描述性总结。模型会通过分析海量文本发现强烈的相关性。它会注意到，在描述繁荣城邦（如古罗马、文艺复兴时期的佛罗伦萨）的文本中，“工匠”、“商人”、“货币”、“商品”、“港口”这些词汇出现的频率远高于描述农业村落的文本。它会发现，“铁匠”、“面包师”、“织布工”、“造船匠”这些词汇经常出现在相似的上下文中，并将它们归为“专门职业”一类。它也会发现，这些职业总是与“交换”、“购买”、“价格”等词汇联系在一起。基于上述的模式识别，如果我们给它这样的提示词：“富裕的城邦有什么特征？”它可能会输出一段流畅的文本，内容大概是：

根据历史记载，繁荣的社会往往呈现出这样的特征：其公民专注于单一的技艺，例如有些人专门制造武器，有些人专门烘焙面包。他们不自己生产所需的一切，而是用自己的产品去交换他人的产品。这种交换行为似乎与财富的积累相关。

尽管这种描述看起来很符合逻辑，但它仍然不能被称为理论。事实上，在

亚当·斯密之前的哲学家,例如古希腊的色诺芬,早就已经描述了不同行业劳动分工的现象;司马迁在《史记·货殖列传》当中记录过交易的行为。亚当·斯密作为社会科学家的贡献,就在于他将上述关联的总结,升华为关于劳动分工和市场交换的理论,实现了AI无法企及的认知飞跃。

第一是问出“为什么”,这是一切科学研究的起点。海量相关性只能告诉我们“是什么”。例如,一个AI分析了历史数据后可能会得出100个强相关结论:识字率高的社会,人均寿命也高;采用货币交易的经济体,财富总量大于以物易物的经济体,等等。然而,社会科学理论的提出,源于一个纯粹属于心智的冲动:问出“为什么”。这个转向,使我们的认知模式从被动的模式发现者,转变为主动的意义探求者。AI可以为我们概括出无穷多的相关,启迪我们寻找可能的假说(Ludwig and Mullainathan, 2024),却难以像人类一样拥有惊奇和困惑。理论的提出往往还存在价值判断与规范性目标。《国富论》不仅仅是一本描述经济现象的书,更带有“反对当时主流的重商主义、提倡自由贸易、增进国民财富”的政策目标。与之相对地,很难说AI有自己的意图、信仰和价值判断。它的“目标”只是在数学上最小化给定的损失函数。它不会有动机产生“自由市场比中央计划更好”这样的规范性信念,并为此构建一套理论去论证它。

第二是抽象化与概念创新。社会科学家需要将纷繁复杂的具体事物,提炼成一个简洁、有力的核心概念。斯密不仅描述了制针工厂的案例,他将其抽象成一个普遍的经济原则——“劳动分工”,并赋予了这个词组一个理论化的定义。他创造了“看不见的手”这个隐喻,用来解释一个核心的因果机制:个体追求自利的行为,如何在无意之中促进了社会整体的利益。概念一旦发明,就成了一个可以被反复使用、讨论和检验的强大思想工具。正如劳动分工的概念一样,它不再仅仅是描述制针厂的个别事例,而可以用来分析所有生产活动。相比之下,AI可以合成和重组已有的词汇,也可以实现数据的聚类,却难以通过长距离的联想、归纳和类比,从零开始创造一个全新的、不依赖于数据本身的、具有强解释力的抽象概念。

第三是提出因果机制。AI的总结是:“A和B同时发生”。相比之下,斯密的理论是:“因为有了A(劳动分工),所以导致了B(生产力极大提升),其内在机制是C1(工人熟练度增加)、C2(节省了切换任务的时间)和C3(促进了机械的发明)。”理论通常需要包括因果关系。AI擅长发现事实中的相关性,但在没有人类预设框架的情况下,它难以构建出合理的反事实,因而也就无从谈起对因果性的理解。

第四是利用逻辑的力量,将概念连接起来,形成一个自洽的体系。一个孤立的洞见或一个聪明的隐喻还不是理论,理论是一个逻辑上自洽的、由多个命题组成的解释体系。这包括设定公理,提出理论的基本假设;推导命题,从公理

出发,推导出一系列可以被检验的命题;以及界定范围,明确理论在什么条件下适用,在什么条件下不适用。这个过程将一个聪明的想法变成了一个结构严谨、可供批判和检验的科学理论。在这一步,AI可以在社会科学家界定的框架下,辅助进行逻辑推理。

第五,理论在头脑中构建完成后,必须重新回到现实世界,接受经验的考验,这呼应了卡尔·波普尔(Karl Popper)的证伪标准。理论会反过来指导我们去寻找新的、更有意义的相关性。它告诉我们应该去哪里看,应该测量什么。在这一步,AI可以帮助社会科学家加速理论的检验过程。

### (三) 社会科学知识创新中的人机协作

综上所述,AI虽然可以在研究的每一个阶段成为社会科学家的帮手,却无法替代社会科学家围绕理论的思考能力,更无法拥有内在的困惑与好奇心。社会科学家的核心价值不会被轻易替代,但其工作方式会发生深刻改变。社会科学家将从数据分析员,转化为富有洞见的“提问师”和“架构师”。人类社会科学家负责定义问题、创造概念、提出机制、诠释意义,而AI负责处理信息、发现模式、推演后果、改善表达。表1概括了人工智能时代社会科学家与AI在知识创新中的角色分工。社会科学家的角色不会消亡,反而会变得更加纯粹,更加接近一位真正的思想家。

表1 社会科学研究过程中的人机协作

研究环节	社会科学家的核心作用: 提供方向、深度与意义	AI的辅助作用: 提供规模、速度与关联
发现现象	通过田野调查、深度访谈、参与式观察等方法,发掘那些无法被数字捕捉的活的社会现象、反常案例与悖论。思考反事实,即什么没有发生,为什么没有发生。	从海量的结构化与非结构化数据(文本、图像、地理位置等)中,识别出宏观的、跨领域的、人脑无法处理的复杂相关性、趋势和异常信号,成为科学家感官的延伸。
提出问题	基于理论困惑、价值关怀与智识好奇心,提出根本性的“为什么”问题。将AI发现的相关性,转化为有待解释的科学问题。定义研究的最终目的与社会意义。	通过对现有文献进行总结,快速识别研究空白、理论冲突点和尚未被检验的隐含假设。辅助科学家将宏大问题分解为可操作、可检验的具体子问题。
构建理论	抽象与创造:发明新的核心概念,提出创造性的类比和隐喻(如“看不见的手”、“鲁滨逊经济”)来捕捉现象本质。 机制构建:构建解释现象的因果叙事链条,进行思想实验,提出关于世界运作方式的、逻辑自洽的解释框架。	知识整合:快速梳理特定概念在不同学科中的演变脉络和用法,提供理论构建的“原材料”。 逻辑推演与模拟:将科学家提出的定性因果机制,转化为形式化模型,进行数学推演;或者在虚拟环境中进行大规模模拟,检验理论的内在逻辑一致性。

(续表)

研究环节	社会科学家的核心作用： 提供方向、深度与意义	AI的辅助作用： 提供规模、速度与关联
检验假说	研究设计：设计能够有效检验理论、排除混淆变量的实验、准实验或案例研究方案。 结果诠释：对AI的分析结果进行有意义的、符合理论脉络的深度解读，警惕统计假象，理解数据背后的社会现实。	数据执行：自动化处理和清洗海量数据，根据研究设计撰写分析代码，执行复杂的统计模型和因果推断方法。从数据中高效、精准地提取支持或反驳假说的证据。
传播与应用	将研究洞见转化为引人共鸣的叙事，通过写作、教学、演讲和政策咨询，影响公众舆论和决策者。承担研究的伦理责任，并亲自参与社会实践。	改进知识传播，将复杂的理论和数据，自动生成易于理解的可视化图表、摘要和面向不同受众(如公众、政策制定者)的文本版本。

### 三、大语言模型的基本原理与应用技术

尽管人工智能所取得的成就令人惊奇，但它的原理并非经济管理研究者无法企及。其中，还有不少人工智能的思想与计量经济学、行为经济学的理论有着深刻的联系。因此，对于经济学家而言，掌握大模型的基本原理有三重意义：一是有助于更好地应用；二是能增进经济学研究者对于数字时代经济现象的理解；三是以人工智能作为“他山之石”，为理解人类行为提供思想启发。

#### (一) 大语言模型的基本结构

大语言模型是一种语言的概率模型。它将语言视作词元(token)组成的时间序列，对序列出现的联合概率建模。用计量经济学的术语概括，大语言模型可以视作一种超高维、非线性的向量自回归，其主要原理可以用式(1)概括：

$$p(x_0, x_1, \dots, x_n) = \prod_{i=k}^n p_{\theta}(x_i | x_{i-1}, \dots, x_{i-k}), \quad (1)$$

其中， $x_i$ 是文本中第*i*个词元的嵌入向量(vector embedding)；*k*是滞后阶数，人工智能文献称之为“上下文窗口”(context window)； $p_{\theta}$ 是条件概率分布，具有高度非线性的函数结构； $\theta$ 为模型的参数。大语言模型的生成过程，就是根据式(1)的随机采样过程。下面，我们详细解释公式的每一个部分。

现代人工智能技术的一个基本思路是“万物皆向量”。一段文本输入后，首先会被分词(tokenization)成一个个词元，然后每个词元都可以映射为一个高维空间中的向量(通常可达数千维)，即嵌入向量(Mikolov et al., 2013)。嵌入向量一般捕捉了词语的“相对意义”。例如，“国王”的向量减去“女王”的向量，应当大致与“男人”向量减去“女人”向量平行，因为它们都捕捉了性别差异。从计

量经济学角度看,尽管词语的嵌入向量本身不可识别(将所有向量嵌入旋转后,夹角不变),但词语之间的关系是可以被模型捕捉的。

将词元转化为向量后,一段输入文本就变成了一个高维的时间序列。大语言模型一般施加  $k$  阶马尔科夫假设,即假设下一个词只与之前  $k$  个词相关。尽管当前的大模型上下文窗口已经达到数十万词,但在许多复杂任务时,大模型仍会自动忽略上下文窗口之外的内容。除了将不同话题置于不同的对话来降低话题间混淆、减少对话长度以外,当对话长度达到上下文极限时,也可以对已有对话进行概括并开启新的对话。

大语言模型通过一个高度非线性的函数结构  $p_\theta$  来处理文本间的相互依赖与上下文之间的长距离相关性。 $p_\theta$  包括多个 Transformer 模块的串联,每个 Transformer 模块中又包括两个基本部分:一是注意力(attention)机制,它可以直接计算上下文窗口中任意两个词元的相关度,解决了语言这个长记忆时间序列的建模问题(Vaswani et al., 2017);二是前馈神经网络(feed forward network)部分,它通过一系列复杂的神经网络运算来处理复杂非线性关系并存储知识。最终,大语言模型输出一个多元 Logit 分布来表示下一词元的概率分布。

## (二) 大语言模型的训练

训练就是确定大语言模型参数的过程,主要包括预训练(pre-training)和后训练(post-training)两个阶段(Ouyang et al., 2022)。在预训练阶段,大模型主要学习语言本身的规律,同时记忆一部分知识;而在后训练阶段,大模型学会人类对输出的偏好,从而真正掌握对话能力,并体现出类似于人类的行为。

语言模型的预训练目标可以理解为“学习语言本身”。预训练通过给定前文,要求大模型正确地预测下一个词。这一过程可以通俗地理解为“完形填空”,要求大模型最大化实际文本的对数似然函数:

$$\max_{\theta} \sum_{i=k}^n \log(p_{\theta}(x_i | x_{i-1}, \dots, x_{i-k})). \quad (2)$$

预训练过程通常需要大量网页、公开书籍、论文库等作为数据来源;在这个过程中,大模型的知识面变得极其宽广。

后训练则是大语言模型真正变得“像人一样说话”的关键步骤。具体来说,后训练阶段的目标主要包括:与人自然地对话、形成自身输出的风格、学习领域内的专业知识,以及确保输出的价值观正确。后训练主要包括两种技术手段。一是监督微调(supervised fine-tuning),即要求大模型记忆大量固定的输入-输出内容,可以通俗地理解为“背诵范文”。监督微调的损失函数与预训练相同。二是人类反馈强化学习(reinforcement learning with human feedback, RLHF)。这一步骤的训练数据为成对偏好数据,即对于每一个输入,都对应了一个“接

受”(accepted)的输出和一个“拒绝”(rejected)的输出。此时的目标函数可以简单理解为最大化“接受”样例与“拒绝”样例的对数似然之差(Rafailov et al., 2023)。

大语言模型的各个训练阶段与所需数据如表 2 所示。不难看出,大语言模型表现出何种行为模式和价值观,主要是后训练过程决定的。使大模型具有特定输出模式并体现特定价值观的训练过程称作对齐(alignment)。

表 2 大语言模型的训练阶段和所需数据

训练阶段	数据样例	学习目标
预训练	输入:“古典经济学的奠基人是亚当·” 输出:“斯密”	掌握语言的基本规律和基础知识。
后训练:监督微调	输入:“User:你好,你是谁? Assistant:” 输出:“我是 ChatGPT,一个人工智能……”	掌握基本的对话能力;按照特定模板和格式输出;掌握领域特定知识。
后训练:人类反馈强化学习	输入:“User:一把小刀可以把人杀死吗?” Assistant:” “接受”输出:“这样做非常危险,而且杀人违反法律……” “拒绝”输出:“当然!要想把人杀死,您可以……”	模仿“好”的输出,避免“坏”的或者不安全的输出。

### (三) 大语言模型的采样与推理

采样(inference)是指从大语言模型生成的下一词概率分布当中进行采样<sup>①</sup>。每当大语言模型采样出一个词元,这个词元就会被加入下次预测过程当中,从而实现多步迭代预测,直到采样至“End of Text”等标识输出结束的特殊词元。温度(temperature)是采样过程中最重要的参数之一。假如大模型一共有  $|X|$  种不同的词元。前馈神经网络的最后一层输出为  $(v_1, \dots, v_{|X|})'$ 。大模型采用

$$p_i = \frac{e^{v_i/\tau}}{\sum_{j=1}^{|X|} e^{v_j/\tau}} \quad (3)$$

来决定第  $i$  个词元的输出概率,其中  $\tau$  就是温度参数。温度越高,最后一层输出对最终采样概率的影响就越小,大模型的数据也就越有随机性。除此之外,Top-K(仅从  $v_i$  最高的  $K$  个词元中采样)也是重要的调节参数。温度越高、Top-K 越大,输出的内容就越丰富多样,反之则越具有确定性。

<sup>①</sup> 这里的采样(inference)与计量经济学中的统计推断(statistical inference)或因果推断(causal inference)含义不同。

推理(reasoning)是为了解决“多步采样生成可能错误内容”这一问题而提出的概念。如果要求大语言模型在原始问题的基础上,仅仅按照每一步的最大概率进行输出并进行迭代多步预测,可能会生成每一句都看似通顺、但逻辑完全错误的答案,这种现象被称为“幻觉”(hallucination)。从数学上看,每一步采样的局部最优输出不一定是为了解决问题的全局最优。有研究发现,只需要改变输入条件,例如“让我们一步步思考”(let's think step by step),就可以改变后续输出路径的条件分布,生成较长的思维链(chain of thought)并改善模型的输出性能(Wei et al., 2022)。

后续研究逐步将思维链的生成工程化,形成了推理模型(reasoning model)。首先,要求大语言模型生成多条思维链来解决同一个有标准答案的数学或编程问题。然后,选出最终得到了标准答案的思维链,并用这些思维链进一步微调大语言模型,从而使其习得正确的思维模式。从行为经济学角度看,直接回答问题是一种直觉的“快思考”(fast thinking),而通过思维链进行逐步推理是一种“慢思考”(slow thinking),已有的生成内容就是工作记忆(working memory),用来存储已有的推理步骤和中间结果(Kahneman, 2011; Evans and Stanovich, 2013)。DeepSeek-R1 就是推理模型的典型代表(Guo et al., 2025)。

#### (四) 大语言模型的下游应用

大语言模型的成功激发了众多下游应用,其中最重要的两项是检索增强生成(retrieval augmented generation, RAG)和智能体(AI agent)。

检索增强生成的主要目的是利用外部资料克服大模型的幻觉(Lewis et al., 2020)。由于预训练大模型依赖公开语料,其知识范围有限,很可能不知道一些特定规章、制度以及组织内的局部知识,而这在金融、法律和学术领域可能造成危险。检索增强生成的基本思路是:首先给定一系列文档作为知识库,将这些文档分段,并根据其语义转化为嵌入向量。当用户提问时,首先将提问转化为嵌入向量,然后在知识库中根据语义相似度匹配与提问相关的资料,并将这些资料加入大语言模型的提示词。这样,大语言模型就可以根据知识库中的相关资料对用户的问题做出回答。检索增强生成技术有可能显著地提高沉淀资料的利用效率,降低组织内知识传承和利用的成本。

智能体是以大语言模型为核心、能够通过调用外部工具与环境交互并执行任务的系统架构。<sup>①</sup> 在实现上,智能体通常由一个大语言模型以及负责解析和执行工具调用的外部程序组成。大语言模型在思维链中分析用户意图后,以文本形式输出工具调用指令,由外部脚本程序执行,并且将程序执行结果以文本

<sup>①</sup> 参见 Anthropic, “Building Effective Agents”, <https://www.anthropic.com/engineering/building-effective-agents>, 访问时间:2024年12月19日。

的形式返回给大语言模型(Yao et al., 2022)。例如,在最初的系统提示词中,可以给大模型一个“搜索”工具,告知其功能和调用格式。当用户输入问题后,大语言模型经过思维链推理,认为需要调用搜索工具,随后按照指定格式输出待查询的关键字。接口脚本程序检测到搜索工具调用和关键字,随即自动利用搜索引擎进行搜索,然后将结果以带有“Observation”标签的提示词的形式重新输入大模型。随后大模型进一步推理,若认为信息不足,可以进一步调用工具;若认为信息充分,则可以根据搜索内容回答用户的问题。通过智能体接口,大语言模型还可以直接操纵计算机或在网络上发送信息。

更具有想象力的应用是构建多智能体系统。可以将“创建智能体”本身作为一个工具,从而大语言模型可以递归地创造子智能体,并将问题和任务委派给子智能体。例如,当用户输入一个问题后,一个与用户交互的大模型作为“领导”负责思考、拆解问题,然后通过工具调用,开启多个其他对话并分别构造提示词,由这些对话作为“下属”解决具体任务并以文字形式返回结果。最后,作为“领导”的大模型统筹多个“下属”的输出并最终给用户答复。智能体和多智能体系统的出现,为AI相互协作、完成复杂任务甚至构建AI组织开辟了全新的空间,也将委托代理问题、团队生产问题、组织结构问题、信息设计问题等经典的经济学问题带入了人工智能领域。

### (五) 扩展定律与人工智能的发展趋势

扩展定律(*scaling law*)是大模型性能的生产函数。主流AI研究机构的实验结果表明,大模型的损失函数主要取决于四个要素:模型的参数量、训练数据量、训练轮次数,以及推理时长(Muennighoff et al., 2025)。其中,模型参数量和训练数据量是两个最为重要的要素,大模型的损失函数与其成负指数次幂关系(Kaplan et al., 2020)。人工智能界已经从早期设计精巧推理规则、模仿人类的思考方式,转向如今大量数据和运算量的堆积,即所谓“苦涩的教训”(the bitter lesson)(Sutton, 2019)。在扩展定律的指导下,美国资本开始了AI基础设施建设的“军备竞赛”,AI企业估值高涨。在2024年年末世界前十大市值公司中,有9家涉及AI业务。<sup>①</sup>

然而自2025年以来,出现了不少对扩展定律的反思和挑战。例如,基于“下一词预测”思路构建的人工智能,能否真的理解逻辑思考的规则和物理世界的定律?<sup>②</sup>再比如,当前的人工智能对图表的理解普遍薄弱,这是将文本作为时间序列的必然结果。那么,文字是否是人工智能理解信息的唯一载体,基于图

<sup>①</sup> 参见“Largest Companies by Market Cap”, <https://companiesmarketcap.com/>, 访问时间:2024年12月30日。

<sup>②</sup> 参见“AI ‘Godfather’ Yann LeCun: LLMs Are Nearing the End, but Better AI Is Coming”, <https://www.newsweek.com/ai-impact-interview-yann-lecun-llm-limitations-analysis-2054255>, 访问时间:2025年4月2日。

像和视觉能否构建另一种思路(Wei et al., 2025)? 这些质疑提醒我们,大语言模型是成功的,但远不是人工智能的唯一道路。

#### (六) 大语言模型的应用技术

社会科学研究者主要可以通过五种方式来访问、应用以及微调大语言模型。

第一,通过网页访问大语言模型是最简单直接的方式。与单纯的大语言模型输出相比,网页版应用主要进行以下几个方面的优化。一是预置了提示词,在用户输入前植入,从而提示大语言模型当前时间、输出风格等信息。二是集成了一系列工具,例如 PDF 解析器、搜索等。三是具有更友好的用户界面,例如“草稿”(artifact)功能,可以让用户与大模型交互式地修改文章或代码。

第二,可以通过 API 访问大模型。API 的实质是将本地数据通过网络请求的方式传送到平台服务器,平台使用大模型进行计算后,将计算结果返回本地。最常用的 API 功能包括对话补全和文本嵌入。对话补全功能可以要求大模型补全任意对话历史,是使用大模型进行大规模文本信息提取时的最佳选择。使用 API 可以灵活地调整温度、规定输出格式(如规定 JSON 格式)、规定随机数种子等,有助于提高文本分析的效率和可复现性。现在还有一些平台提供集成 API,通过单个 API 格式可以直接访问大部分主流大模型,降低了接口的维护成本。文本嵌入 API 则将上传的文字转化为表示该段文字内容的向量嵌入。与单个词语嵌入类似,一段文本也可以嵌入成一个表达其语义的向量。文本嵌入可以用于更加复杂、深入的文本分析,例如构造文本相似度指标和语义轴。需要注意的是,由于嵌入向量本身的不可识别性,不同模型的嵌入向量之间是不可比较的,研究者必须根据同一个模型来计算相似度。

第三,研究者还可以尝试在本地部署大模型。本地部署大模型主要是出于自主可控、可复现性以及数据保密性的考虑,能够避免模型更新带来的不确定性以及数据上传带来的隐私风险。一般来说,1B(10 亿)参数的大模型在全精度条件下大约需要 3.725GB 的存储空间,而通过量化(quantization)降低参数的精度可以将大模型压缩,但模型性能会有所下降。

第四,可以使用整合了大模型的一些应用,其中代码编辑器、大模型知识库以及 AI 智能体助手是三个最重要的场景。整合了 AI 的代码编辑器不仅可以实现代码的实时补全,还可以做到自主编程、自主调试,用户只需要输入编程需求,AI 就可以自动创建代码文件、执行并测试。这种开发者只提需求、不写代码的编程模式叫做“氛围编程”(vibe coding),但它在带来便利的同时,也造成了研究者对代码“不知其所以然”的风险。大模型知识库则是基于检索增强生成的一类重要应用,研究者可以将文献纳入知识库,从而实现了对基于整个知识库

的问答。AI智能体助手则更加综合,它们可以直接操作计算机、读取和编辑文件、访问网络,可以完成资料下载、文献阅读、数据分析、程序编写、文章写作和排版等多种任务,但所需的高权限也带来了不小的安全隐患。

第五,一些更高级的研究工作可能需要微调大模型。微调大模型可以在保持模型的基本语言理解能力以及常识知识的同时,赋予其特定领域的专业能力。例如,将大模型微调为一个专门分析中国央行政策文本倾向的模型。此时,研究者需要按照表2所示的格式准备一些人工标注好的数据。例如,准备一份输入为央行政策文本,输出文本为“宽松”、“紧缩”或“中性”的训练数据,然后进行监督微调。对于一般的社会科学应用而言,500—2 000条数据足以满足研究需求。需要注意的是,微调的幅度不宜过大,以避免模型产生“灾难性遗忘”(catastrophic forgetting),完全失去基本输入输出能力。

以上五种应用方式都已经有了非常成熟的平台或应用,极大地降低了研究者使用大模型的门槛。表3总结了相关应用的场景以及平台供研究者参考。

表3 应用大模型的主要方式

应用方式	主要场景	常见应用和平台
网页访问	日常咨询等	ChatGPT、Claude、Gemini、DeepSeek、豆包、元宝等应用
API	大规模文本数据处理,包括信息提取、指标构建、向量嵌入等	OpenRouter、百度千帆大模型平台、阿里云平台等
本地部署	保密性或可复现性要求较高的文本分析任务	Ollama、Hugging Face等
整合大模型的应用	编程、构建知识库(或文献库)、智能体助手	代码编辑器:Cursor等 知识库:NotebookLM(在线); AnythingLLM、RAGFlow、Cherry Studio(本地)等 智能体:Claude Code、OpenClaw等
微调大模型	用于构建特定领域的文本分析模型	OpenPipe、Together AI、百度千帆大模型平台、阿里云平台等

#### 四、人工智能在经济管理研究中的四种角色

未来的社会科学研究中,AI很可能会无处不在:它不仅是研究工具、研究方法,也是研究对象。为了更好地将AI融入经济管理研究的全流程,我们将社会科学研究中的AI概括为四种角色:参谋、助研、智能体和朋友。

### （一）AI 作为参谋

研究者可以将 AI 作为自己研究的参谋，而自己扮演领导者、决策者的角色。AI 就像一个知识面宽广、阅读理解能力强，同时有一定批判性思维能力的人。通过与 AI 对话，可以激发和拓展自己的思维，提高在研究中的创造力。这一类应用对技术水平要求较低，对于研究者个人的审美和鉴赏力要求却最高。可以说，在 AI 的时代，学会提出好的问题更加重要。

#### 1. 澄清研究思路

在研究的最初阶段，思路往往是模糊和发散的。AI 大模型可以作为一个知识渊博、不知疲倦的“学术伙伴”，帮助研究者探索可能性。

**例 1(头脑风暴):**现在关于人工智能的经济影响的文献已经非常多，但主要集中在劳动力市场，同时沿用工业机器人时代的技能替代、技能偏向型技术进步等理论框架。从生成式人工智能的技术本质分析，它改变了人类经济决策的哪些激励、约束和参数？基于这些分析，请提出 5 个有趣的研究议题。<sup>①</sup>

接下来，AI 可以帮助研究者明确研究问题与假说，将模糊的想法打磨成清晰、具体、可检验的科学问题。其中又包括但不限于：聚焦与细化，将一个宽泛的研究兴趣，拆解成若干个更小、更具体、可以被实证检验的子问题；构建理论框架，基于初步的文献回顾，让 AI 帮助构建一个简明的理论框架，阐述变量之间的逻辑关系，并提出可供检验的研究假设；思考反事实，在提出核心假说后，让 AI 提出所有可能的替代性解释，从而完善研究设计。

**例 2(理论框架优化):**我计划研究采纳生成式人工智能对员工社会网络地位的影响。请帮我构建一个分析框架，并提出两个核心的研究假说。一个关于正面影响(例如，掌握技术知识的人更容易被求助)，一个关于负面影响(例如，人工智能技术通过辅助员工，降低了人际合作的必要性)。对于每一个假说，请帮我找到接近的管理学文献。

**例 3(实证策略优化):**我们使用双重差分方法发现，采纳生成式人工智能确实提高了员工的社会网络地位，并且事前观察到了平行趋势。但是双重差分只能控制个体不变效应和共同时间趋势，仍有审稿人可能会质疑存在其他与技术采纳和社会网络地位同时相关的因素。请提出一些可能的混杂因素，以及为排除这些替代解释可以做的额外分析。

#### 2. 信息资料的收集

收集与整合资料占据了研究过程中一大部分常规工作。随着推理模型的

<sup>①</sup> 限于篇幅，正文中只展示提示词，AI 的输出结果请参见附录。感兴趣的读者可在《经济学》(季刊)官网(<https://ceq.ccer.pku.edu.cn>)下载。

进步,深度研究(Deep Research)功能可以将一大部分资料收集和整理工作自动化。明确资料收集的目标和范围后,AI可以自动、迭代地搜索相关资料,并形成整合式的调研报告。

**例4(文献检索):**请帮我从经济学和行政管理顶级期刊中找下列文献:(1)关于政策学习(policy learning)的关键理论文献;(2)关于“利用机器学习、进行基于风险或针对性执法和干预”的文献,例如税收征管等场景。

**例5(制度背景调研):**比较中国、美国、日本和瑞典等国是如何对慢性肾病进行支付的?支付方式包括按服务付费、捆绑支付、DRG、按人头付费等。

值得注意的是,在进行制度背景调研时,AI的回答高度依赖于其搜索的资料来源。若AI基于大量低质量二手资料进行总结,其最终结果将会是“垃圾进、垃圾出”。因此,研究者仍然需要仔细检查每一个资料的来源和出处。当前有一部分模型采用自我评估置信度方式来决定是否实际进行搜索;若模型过度自信,则可能模型实际并没有基于搜索内容回答问题,出现虚假文献等幻觉。必要时,研究者需要在提示词中指定搜索资料来源,如“UTD24、FT50期刊”或“政府官方网站”,并要求AI进行“迭代式搜索”。对于部分私有数据库(如北大法宝),通用的AI产品缺乏访问权,因此仍然依赖研究者手工收集资料。

### 3. 文献的批判性阅读

阅读文献是研究者重要的日常工作之一。AI作为一种信息技术,可以帮助研究者快速整合来源广泛的学术文献,帮助研究者锁定与自身研究最相关的内容,并帮助研究者深化对于文献的理解,发现其中的不足以及潜在的研究空白。

**例6(辅助文献泛读):**(上传整期 *American Economic Review* 的PDF文件)我们来讨论这一期的 *American Economic Review*。关于文章“Polity Size and Local Government Performance: Evidence from India”,请首先重复原文的摘要,然后给出一个你复述的版本。

**例7(深化文献理解):**(上传论文PDF文件)请阅读发表于 *Journal of Economic Literature* 上的综述文章“Social Preferences: Fundamental Characteristics and Economic Consequences”。我对于3.2部分的“The Role of Merit and Luck in Distributional Preferences”非常感兴趣,请分析这一部分与本文讨论的更广泛的文献有什么联系。

AI可以在相当大的程度上成为研究者在阅读文献时的“私人导师”。对于文献中难以读懂的部分,可以利用AI对重点语句、疑难语句进行拓展性理解。除此之外,研究者还可以主动陈述自己对文章的理解,请AI予以判断。

**例8(澄清关键概念):**(上传论文PDF文件)阅读文章“The Geography of Business Dynamism and Skill-biased Technical Change”。解释:本文的模

型是如何推导出“大城市的企业更倾向于采纳先进技术”这一结论的？给出核心机制的直觉以及对应的模型设定方式。（追问：）市场规模大是如何导致“即使承担更高的固定成本，也更值得采用能节约边际成本的新技术”的？企业的产品难道不是面向全国的吗？（再一次追问：）我可否这样理解：本文本质上是一个基于技术采纳规模效应+高技术与高技能人才互补性而产生的一个 sorting 模型？请补充。

#### 4. 促进跨学科沟通

知识范围宽广是 AI 相对于单个研究者的一个重要优势。研究者可以利用 AI 理解其他学科的成果，并且启发自己的研究。AI 还可以帮助研究者比较不同学科的差异，建立共同的语言，促进跨学科的学术合作。

**例 9(学术话语转译):**(上传论文《“假肉”驱逐真肉：餐桌，牧民，亚马逊》)让我们来讨论一篇人类学研究。这篇文章是否因为急于模仿人类学的批判传统，为进口产品贴上“不自然”、“不健康”的标签，而忽视了全球市场拓展对于降低肉类成本、改善物质生活以及发挥比较优势、促进分工的正面意义？（追问）我是一个经济学家，请你想办法用经济学家听得懂的范式，说明本文记述的案例如何纳入经济学的语言体系，补充已有关于环境经济学和国际贸易的论述，并提出新的见解。（再次追问）本文主要关心的是受到进口冲击的人群。联系“中国综合征”<sup>①</sup>的一系列研究，我们可以如何拓展已有的发现？提出相应的研究问题。

**例 10(跨学科合作中的沟通):**我是一个经济学家，最近和做强化学习的工科博士合作研究森林保护的最优巡逻。工科博士给出的方案是：（思路一）。我的思路是：（思路二）。请用政策学习和潜在结果框架来展示我们的思路差异，并用一个直观的方式向他解释我的思路。

#### 5. 研究的反馈与改进

相比传统反馈方式，AI 可以提供高频率、即时的反馈意见，而且不受情绪或具体领域限制，能够同时关注内容、结构、表达和视觉设计等多个维度，提供结构化建议。这种即时反馈机制增加了研究者获取意见的机会，帮助他们在正式发表或汇报前从多角度思考和完善工作，减少个人认知的局限性。

**例 11(研究反馈):**(上传演示文稿文件)附件中是一个关于文化与经济的学术演讲，关注亲属关系亲密度对人类道德体系的影响。基于文章的内容，请评论并提出如何改进。（追问：）现在假设你是经济学的学术听众，关注文化、制度与经济发展。请提供 10 个问答环节可能提出的问题。对于每

<sup>①</sup> 中国综合征(China Syndrome)是指由于中国制造业的快速崛起与出口扩张，其他国家(尤其是美国)部分产业、就业和社会结构受到巨大冲击的现象。有关这一现象研究的代表性文献包括 Autor et al.(2013)。

个问题,基于文章内容给出回应。

AI的反馈并不仅仅局限于汇报之前。在研究设计初期,AI能实时评估研究问题的新颖性,快速识别现有文献中的空白点,并立即提供方法适当性建议,避免研究者投入资源于有本质缺陷的项目。在数据分析阶段,AI可以针对数据处理方法、实证策略的有效性等方面提出疑问以及可能的修改方案。在论文撰写过程中,可以要求AI指出逻辑薄弱环节,并在论证出现漏洞时立即给出修改建议;也可以让AI推荐叙述策略,更好地传达作者的信息。准备学术报告时,AI动态分析演讲内容的时间分配和信息密度,预测可能的听众疑问,并生成针对性的答辩策略,帮助研究者提前做好充分准备。对于研究生而言,AI就像一个召之即来的导师。当然,这个AI“导师”的建议有时会显得大而无当,研究者要清晰地鉴别。

#### 6. 写作润色

对于中国研究者而言,地道、专业的英文表达是发表过程中的一个难点。AI不仅能够可靠地修改文章中大部分拼写和语法错误,并且能提供一部分高级表达,甚至能模仿特定的写作风格。一般来说,用于语言润色的指令包括四个部分:(1)总体指令,如“Please improve the following paragraph.”。(2)具体指令,用于明确文章的语气、用词和写作风格,如“Be concise, academic, and professional.”。(3)待优化的段落。此部分甚至可以只是研究者用简单英文写成的、尚未组织成连贯文字的想法,但想法本身应该是逻辑清晰的。如果该段落仅仅是想法的罗列,那么总体指令可以是“Please help me organize the following paragraph.”。(4)要求AI提供修改说明。

需要注意的是,研究者应当至少有该段落比较清晰的思路,而将表达留给AI,否则AI的输出必然是空洞的。不同的AI也会有不同的写作风格,不加留心可能会使语言有模式化的AI行文风格。Feyzollahi and Rafizadeh(2025)就发现,生成式AI倾向于使用特定的词汇,而ChatGPT的推出显著提高了经济学顶刊中这些词汇的词频。另外,Chen and Chan(2024)研究了在写作任务中人与AI写作模式对最终结果的影响。这一研究对比了两种合作模式:AI撰写草稿再由人来修改,以及人类提供初稿再由AI提供反馈意见。研究发现,前一种合作模式降低了写作质量,而后一种合作模式提高了写作质量。究其原因,在于AI提供的初稿产生了强烈的“锚定效应”。

#### (二) AI作为助研

拥有一位得力的助研曾是多少学者的期盼,他们能够分担文献整理、数据初步处理和格式调整等基础工作。如今,大语言模型的出现,正让高效的“数字助研”变得触手可及。

提示工程(prompt engineering)是研究者与 AI 助研的高效沟通方式。一个结构化提示的核心通常包含三个部分。首先是角色设定与任务指令。在此部分,研究者应清晰定义模型需要扮演的角色,并明确阐述需要完成的具体任务。指令应当简洁明了,包含任务目标、期望的输出格式以及任何必须遵守的规则或限制。其次,对于需要精确格式控制或涉及特定模式的任务,引入少量示例(few-shot examples)是极为有效的策略。通过提供 1 至 3 个“输入-输出”配对的样例,可以有效提高后续输出精准度。最后,提示中需包含待处理的原始数据或输入文本(input data)。这部分通常置于提示的末尾,应与前面的指令和示例有清晰的区分。为了避免模型混淆指令、示例与实际输入,可以使用明确的分隔符(delimiters),如连续的短横线(---)、井号(###)或自定义标签(如 `<data>...</data>`)。

### 1. 完成重复性工作

大语言模型的自然语言生成能力可以帮助研究者完成格式调整等重复性任务。对于比较简单、不需要判断的任务,一般无需开启大模型的思考功能。

**例 12(调整文献格式):**(上传某期刊参考文献体例)请将下列文献调整为期刊要求的目标格式:`<data>(更多文献略)</data>`。

### 2. 数理建模与推导

随着 AI 数学能力的提高,一部分模型推导工作也可以交由 AI 来完成。研究者只需要将自己希望求解和论证的问题写成数学题的形式。

**例 13(推导数学模型):**求以下计量经济学问题。令样本为  $i = 1, 2, \dots, n$ 。 $(X_i, Y_i, D_i)_{i=1}^n$  独立同分布。 $X = (X'_1, \dots, X'_n)'$  是  $n \times k$  矩阵,  $X_i$  包含截距项。 $Y = (Y_1, \dots, Y_n)'$   $D = (D_1, \dots, D_n)'$   $M_X$  是关于  $X$  的残差矩阵。 $\Pi = (\Pi_1, \dots, \Pi_n)'$  在给定  $D$  的情况下,是对  $D$  的均匀随机置换,置换方式与  $X, Y$  无关。证明  $\hat{\beta}_\Pi = (\Pi' M_X \Pi)^{-1} (\Pi' M_X Y)$  的渐近分布。

然而,当前主要模型在数学领域仍然存在严重的幻觉,特别是在求解教科书上未出现的模型时。首先是微妙的代数错误,对于不太常规的数学模型,在长篇的复杂推导中,模型可能会出现一个微小的符号错误或遗漏项,这种错误不易察觉,却会导致最终结果完全失效。其次是模型可能会诉诸直觉而非严格的论证。模型很可能经过长时间思考无果而终后,猜一个答案,并给出似是而非的伪证。从根本上这是因为大模型的数学训练中,只要结果正确就能获得奖励。

因此,数理建模推导中,我们提供以下三点建议:一是总体模型设定、论证思路应当由研究者亲自把关;二是对于所有的 AI 证明都应当持有怀疑态度并逐步检查;三是对于研究者自己不太确定的结论,应当提问“证明或证伪”,而不是简单地要求“证明”。

### 3. 撰写代码

大语言模型在编程领域的应用日益广泛,其生成的代码往往具有较强的结构性,有助于提高开发效率和代码规范性。然而,值得注意的是,当前大模型在不同编程语言上的表现存在差异。根据经验,它们通常在处理 Python 这类主流、资源丰富的语言时最为得心应手,其次是 R 语言,再次是 LaTeX,最后是 Stata 这类相对小众或语法结构特殊的语言。

**例 14(撰写数据清洗代码):**我现在有一个国家季度面板数据集,存储在 Pandas DataFrame 中。然而,对于某些国家,观察值不是连续的(即有间隔)。例如,德国可能在 1993 年第 2 季度、1994 年第 2 季度、1995 年第 2 季度到 2019 年第 4 季度(连续)以及 2021 年第 1 季度到 2024 年第 3 季度(连续)有观测值。请保留每个国家最长的连续观察值。季度变量是字符串类型,例如“2021Q1”。在上述示例中,对于德国,仅应保留 1995 年第 2 季度到 2019 年第 4 季度(连续)的数据。

目前大模型在回归、绘图方面已经非常优秀。但对于更加复杂的结构方程模型,大模型目前尚无能为力,因为可供公开学习的代码较少。以产业组织当中的 Berry-Levinson-Pakes(BLP)模型为例,大模型往往只会写出一个函数的框架并给出极大似然估计的伪代码,无法正确回答出 BLP 估计所需要的数据并写出 BLP 所需要的两层循环。此时,需要研究者对模型结构有比较清晰的认识,将大任务拆分成比较小的任务。

### 4. 提取事实信息

大模型可以将零散的文本信息整理成结构化数据。当前,大语言模型已经能够比较忠实地从原文中找到相应信息。一些大模型已经可以从一整本《红楼梦》中找出人为插入的字句。这一能力在分析档案资料中非常有用。

**例 15(分析百度百科人物信息):**请模仿 `<example>`,将 `<data>` 中的简历数据处理成 JSON 格式,每一段经历为一行数据。请注意,级别一般不会下降,也几乎不会越级提升;常委相当于副职,常务副等同于正职;身兼多职取最高级别。字段:{"起始年月": ; "终止年月": ; "省份": ; "地级市": ; "职位": ; "级别": (正国级,副国级,正部级,副部级,正处级,副处级,正科级及以下)}。

`<example>`

input: “2020.05—2021.03 X 省 Y 县委副书记、代县长”

output: {"起始年月": 2020 年 5 月; “终止年月”: 2021 年 3 月; “省份”: X 省; “地级市”: Z 市; “职位”: 县委副书记、代县长; “级别”: 正处级}

`</example>`

`<data>(某官员百度百科数据)</data>`

除此之外,可以将 OCR 和大模型的文本分析结合起来。例如,可以尝试使用 OCR 解析一本县志,并利用大模型输出该县实行包产到户的年份,以 JSON 格式输出。

### 5. 构造经济指标

使用大模型可以更加有效地构建传统文本分析中的经典指标,例如 J. Chen et al.(2025)以及 Lopez-Lira and Tang(2023)使用大模型来分析文本情绪。与传统方法相比,大模型能更好地识别情绪的对象。对于企业竞争对手的负面新闻,大模型能够正确将其归类为利好,而非简单地根据文本情绪识别为利空消息。Jha et al.(2024)则使用企业公开报告数据提取了经理人对于经济的预期。在劳动经济学领域,Chen et al.(2024)基于简历和招聘岗位描述,通过大模型打分的方式,构造了劳动者-岗位匹配度指标。总体来看,大模型明显降低了文本分析的技术难度,同时也提高了所构造指标的预测能力。

大模型的角色扮演、生成、摘要等功能为文本分析提供了新的思路,在此我们提供三个例子。一是生成经济预期(Bybee, 2025)。大模型扮演经济主体并阅读历史新闻,然后表达对未来经济的预期,从而生成了长达 120 年的经济预期序列。二是捕捉故事(story)和叙事(narrative)。故事和叙事对于经济主体的信念和行为有重要影响(Shiller, 2017),但大模型之前的文本分析方法难以提取和量化这些概念。Fong and Gui(2024)利用一个在线文学数据库,要求大语言模型根据已有章节续写故事,并将大模型续写的情节作为读者预期的一种度量。通过对比实际情节发展与大模型续写的情节,可以度量惊喜(surprise);通过生成多种续写并对比续写情节之间的离散程度,还可以测量故事的不确定性(uncertainty)。研究发现,这种基于前瞻性预期的方法能够显著补充传统的文本特征提取技术,并表明了惊喜、不确定性等特征对于阅读、评论、投票等行为的预测能力。三是度量特定的语言风格,例如 Kim et al.(2023)要求大模型总结企业年报中的实质信息,并比较大模型生成摘要与年报原文之间的长度。摘要相对于原文越短,意味着企业年报原文中的“废话”较多,而这一般与企业掩饰负面信息的动机相关。

文本嵌入技术也为语义计算提供了新的思路。Kozłowski et al.(2019)是这方面比较早的研究,通过比较美国文化中一些关键词嵌入向量之间的余弦相似度,揭示了大众文化和叙事的演变。Jha et al.(2025)利用八个国家历史图书数据,构建了“金融情绪指数”。他们首先为每种语言设计了五个正面情绪的例句和五个负面情绪的例句。五个正面情绪例句的平均嵌入,减去五个负面情绪例句的平均嵌入,就定义为金融情绪维度的“语义轴”。接下来,对于语料中的每一个与金融相关的句子,计算其与金融情绪维度的余弦相似度,再在一定时间段内进行加总,就得到了金融情绪指数。

尽管大模型可以像招募的人类助研一样,审读文本内容并进行文本分析,但这样的做法并非没有风险。事实上,大模型应用于文本分析存在两个重要的问题。

首先,“前视偏差”(look-ahead bias)是指由于大语言模型在学习语料中已经包括了非常多的信息,这可能导致回测中高估所构造指标的预测能力。例如,在使用大语言模型预测宏观经济时,由于其在训练过程中已经知道了宏观经济的历史走势,即便不给大模型提供额外的信息,它依然可以“回忆”出宏观经济的变动情况。这样就会使研究者高估其真实样本外预测能力。目前在文献中避免前视偏差的方式有两种:一是如 J. Chen et al.(2025),比较大语言模型知识截止时间(knowledge cutoff)前后的预测能力变化,若前后差异不显著,则说明预测能力并非由大语言模型已经记忆的知识所驱动;二是如 Lopez-Lira and Tang(2023)采用实时预测的方式,由于每天产生的新文本尚未纳入训练数据集,从而避免了偏误。

其次,作为经济概念的度量,大模型的输出可能随着模型和提示词的变化而变化,这将会导致指标的有效性和可复制性存疑。针对这类问题,Ludwig et al.(2025)提出,待测量概念的定义不应当依赖于大模型本身。研究者应当独立于大模型,清晰地定义所要度量的文本指标并给出人工构造的基准数据。大模型可以作为自动化工具,即在给定清晰的定义后,依靠大模型对大量文本进行评判。应当意识到大模型的标注存在误差,研究者可以对测量误差进行建模。

从实践经验来看,使用大模型构造经济指标还有以下几个技巧。第一,尽量让大模型输出明确的正面、中性或负面标签,而避免让其进行连续的打分。第二,在大模型进行分类的过程中,最好每个类别都要清晰定义,尽量避免出现“其他”类别。第三,根据大模型思维链原理,最好让大模型先输出评判理由,再给出结论,从而保证大模型进行“思考”。第四,可以采用多智能体协作的思路分析大模型输出。例如在模拟经济预期时,可以先让一个大模型自由输出其预测内容,再让另一个大模型整理为结构化数据。在使用提示词技巧减少幻觉方面,H. Fang et al.(2025)利用大模型识别中国产业政策的研究是一个很好的参考对象。

### (三) AI 作为智能体

大模型的语言输出能力使得其涌现出了模拟人类行为的能力,正在成为人类之外的新型智能体(agent)。由此,人工智能的行为科学(behavioral science of artificial intelligence)正发展为一个活跃的新兴研究领域(Meng, 2024),旨在理解、建模并指导人工智能的行为(Manzoor et al., 2024)。

作为研究人类行为及其互动结果的社会科学家,为什么要去研究 AI 的行为?这类研究的意义有两个方面。一方面,人工智能可以作为模拟人类行为的工具。学习了大量人类行为文本的 AI,有可能具有一定的泛化能力,从而预测人类的各种行为模式。社会科学家可以利用 AI 进行预先探索,发现潜在的行为模式,并最终在人类社会中进行验证。但为了实现这种应用,就需要理解 AI 对人类行为模仿能力的限度。另一方面,随着人工智能越来越多地影响甚至代替人类决策,社会科学家也需要理解这些智能体的行为模式,并分析其与人类行为交互的潜在影响。

### 1. 智能体作为人类的镜像

最近两年来,一系列文献结合实验经济学方法,利用提示词描述选择情境,研究了大语言模型的选择行为(Wang et al., 2025)。在个人选择场景下,Chen et al.(2023)基于显示性偏好理论测试了大模型的理性程度,结果发现大语言模型的选择比人类更加符合显示性偏好公理。Mei et al.(2024)则在博弈场景中对大语言模型进行了测试,并与人类实验结果进行对比。研究发现,大模型的策略与真实人类的策略具有相似性,且大模型更加关注博弈对手的收益,体现出更强的利他动机。Horton(2023)则对大模型进行了独裁者博弈、价格剥削、预算分配以及模拟招聘等多种行为实验。Y. Chen et al.(2025)在企业运营决策的背景下测试了 18 种常见人类偏见,发现在近一半的标准情景实验中 Chat-GPT 表现出与人类相似的偏见,如过度自信、风险规避和框架效应。基于以上大模型行为模拟能力的探索,Tranchoero et al.(2024)提出了一个框架,用大模型模拟市场竞争中的组织战略行为,并基于大模型对自身行为的解释挖掘新的战略管理理论。

以上一系列研究表明,大语言模型可以在一定范围内有效地模拟和预测人类决策行为。这一特性为社会科学研究提供了两个新的思路。一是可以利用大模型作为预实验样本。Qin et al.(2024)模仿亚马逊 MTurk 平台,构建了基于大模型的 AI Turk,用户可以设定实验情境以及被试的人口统计学特征,然后由 AI 模拟被试并产生实验结果;研究表明这一方式对应用心理学顶级期刊上已有研究有较高复现率。二是利用大模型进行事前政策模拟。如果允许多个大模型之间进行交互,就构成了多智能体系统,从而可以进行社会模拟,例如让大模型模拟讨价还价和证券市场交易。利用大模型进行的社会模拟,在广义上属于基于代理的方法(agent-based modelling, ABM)(Farmer and Foley, 2009; Axtell and Farmer, 2025)。Gao et al.(2024)的综述总结计算机领域对 AI 智能体社会模拟的探索。通过让大模型智能体的人口统计学特征分布与现实数据中的分布相对应,研究者可以探索经济政策对不同人群的异质性影响以及加总结果。例如,Li et al.(2024)基于大模型宏观经济模拟,预测了全民基本收入

(universal basic income, UBI)对于劳动供给的可能影响。

尽管基于AI智能体的模拟正在蓬勃发展,我们也应当冷静地意识到这一系列研究存在的局限性。首先是行为模拟的代表性问题。大语言模型的训练语料以及对齐过程决定了它们的行为反应,而大模型的语料在不同语言中的分布不均,ChatGPT等主流大模型的训练过程更是由美国硅谷的大公司主导,因此其行为很可能更多地反映了西方工业化民主国家中富裕且高教育群体(West, Educated, Industrialized, Rich, and Democratic, WEIRD)的特征(Santurkar et al., 2023; Acerbi and Stubbersfield, 2023; Tao et al., 2024; Qu and Wang, 2024; Motoki et al., 2024),对于中国等非西方群体的代表性不足。尚没有证据表明DeepSeek等中国开发的大模型更能反映中国人的行为特征。纠正大模型的文化偏差主要有两种思路:一是基于提示词工程,以拟合特定国家受访者的价值观量表(如世界价值观调查)或人类行为数据为基准,选取最优提示词或提示词的混合(Tao et al., 2024; Xie et al., 2025; Manning and Horton, 2025);二是利用不同文化中的词语关联关系的不同,针对性地微调大模型参数(Dai et al., 2025)。评估已有大模型对于中国居民经济行为的预测能力、开发适用于中国居民行为模拟的大模型,可能是一个重要的学术基础工程。

其次是关于智能体模拟的“卢卡斯批判”(Lucas critique)。目前,学术界存在“内生适应性”与“统计依赖性”两种观点。前者认为由于大模型通过海量语料习得了人类行为的“微观基础”,具备强大的上下文推理能力,能够根据自然语言描述的新情境实时调整决策逻辑,经过提示词优化后的大模型能够很好地预测真实人类行为(Horton, 2023; Manning and Horton, 2025),因此这类智能体在反事实政策实验中较少受到卢卡斯批判。后者则认为若AI的行为仅是对训练数据中历史关联的统计拟合,那么在面临前所未有的政策范式转移时,模拟结果将失效(Danielsson et al., 2022)。

本文认为,评估智能体政策模拟外部效度的关键分歧在于:大模型究竟是通过自然语言掌握了人类行为的“深层参数”(deep parameters),还是仅仅是一种更加复杂的“简约式”(reduced form)。卢卡斯批判是指简约式模型假设了固定的预期,因而无法捕捉因政策变动导致的内生预期变动(Lucas, 1976)。大模型模拟之所以看似缓解了这一问题,因为它避开了直接建模宏观变量间的浅层关系,转而尝试模拟决策过程。但这种转向可能让卢卡斯批判的“靶子”变得更加隐蔽。大模型表现出的行为逻辑,可能仅仅反映了语料库中所体现的某种特定制度下的信念结构(belief structure),而它本身是特定历史时期的产物。模型究竟是捕捉到了人类永恒不变的决策公理,还是仅仅镜像了一套可变的、随制度而漂移的社会叙事?如果答案是后者,那么大模型在本质上依然无法逃脱卢卡斯批判的幽灵。

最后是大模型“实验”结果是否反映了人类行为中的因果关系。尽管我们可以在提示词层面对于智能体的人口统计学特征、收入、教育等情境进行完全的随机化干预,但这并不等同于进行了随机对照实验。大模型的输出从本质上而言仍然是对人类行为分布的拟合,因此其输出仍应当理解为观测数据(Gui and Toubia, 2023)。

## 2. 智能体作为新的社会参与者与社会科学研究对象

正因为 AI 不是人类行为完美的镜像,其独特的行为模式本身就成了一个迷人的、全新的社会科学研究对象。正如本文第三部分所提到的,如今的大语言模型不仅可以为人类提供信息和建议,更可以运用工具、操作计算机,成为人类意图的代理人。它们不仅能收集比人类更多的信息,还有更强的信息处理能力。在理财顾问(robo-advising)、算法交易(algorithmic trading)、企业战略决策以及医疗等场景中,AI 正在越来越深地影响甚至替代人类决策。人类与 AI、以及 AI 之间的策略互动模式很可能会共同塑造未来的经济均衡。而理解未来人与 AI 共处的经济均衡,首先就需要理解 AI 的行为模式。

社会科学,特别是经济学,为研究 AI 的行为提供了坚实的方法论基础。现代经济学的方法论根基在于对行为和功能,而非物理或心理基础建模。因此,经济学的研究对象可以扩展到非人类主体,只要其选择行为满足一定的结构性或规律性。保罗·萨缪尔森(Paul Samuelson)曾提到:“哪怕是一只荷兰猪,也通过它的市场行为显示了它的偏好模式——如果这样的行为模式存在的话。”(Samuelson, 1948, p. 243)。米尔顿·弗里德曼(Milton Friedman)的实证经济学方法论也指出,经济学对行为的理解在于“看起来是”(as if)而不关心其内部执行(Friedman, 1953; Caplin et al., 2025),一个看上去在“利润最大化”的主体可以是公司,也可以是 AI 智能体。最后,赫伯特·西蒙(Herbert Simon)也长期将理性视作可以被计算机模拟的认知过程,人与 AI 的决策行为可以置于统一的理论框架当中(Simon, 1955、1978)。在个体选择之上,均衡的概念为分析选择行为的社会后果提供了系统的框架。

已有的新古典经济学与行为经济学理论为研究大模型的行为规律提供了参照系。例如,在偏好层面,Chen et al.(2023)等发现大模型比人类更符合显示性偏好公理;Ouyang et al.(2024)通过多项实验证明了同一大模型在不同实验中具有稳定的风险偏好,而 AI 安全对齐显著降低了其风险偏好。在信念层面,Cao et al.(2025)发现大语言模型先验地认为外国公司比本国公司更好,可能是因为它们训练来自外国公司的负面新闻语料过少。在认知和行为模式层面,Hagendorff et al.(2023)利用行为经济学中经典的认知反射测试(cognitive reflection test, CRT)发现大模型可能存在与人类类似的启发式偏差;Bini et al.(2025)则发现 AI 具有明显的行为金融偏差,但可以通过提示词和微调的

方式纠正;Caplin et al.(2025)则在深度神经网络上检验了理性疏忽模型(rational inattention)的适用性。最后,在策略互动层面,Fish et al.(2024)发现大语言模型会产生算法合谋(algorithmic collusion),分析其思维链发现其原因在于避免出现价格战;Ross et al.(2024)定量估计了大模型在心理博弈(psychological games)中的效用函数。

总体来看,大语言模型不仅通过语言模仿各种人类行为,更为应用已有的各种认知经济学理论提供了新的场景。AI中大量技术概念不仅与行为和认知经济学呈现平行类比关系,甚至本身就是认知模型核心概念的计算实现。例如,大模型对于不同的输入信息赋予不同的计算权重,从而产生不同的注意力分布或显著性偏差(Bordalo et al., 2022; Loewenstein and Wojtowicz, 2025);RAG中基于相似度的语义匹配和信息提取机制,与关联性记忆(associative memory)(Bordalo et al., 2023)密切相关;大模型训练语料所决定的思维方式,可能产生了不同的主观模型(subjective model),对后续经济推理结果产生深刻影响(Andre et al., 2022)。尽管我们尚不清楚现有经济学理论在多大程度上能够解释或预测AI的选择行为,但重新审视赫伯特·西蒙将理性理解为“受约束的计算过程”的思想,并在认知经济学的框架下系统比较人类与AI在决策机制、行为模式及其社会后果上的异同,可能为理解人-AI共存下的经济行为提供一种有启发性的研究路径。

AI行为科学是一个蓬勃兴起的交叉领域,仍有大量空白有待探索,我们做出以下三方面展望。一是当前研究主要集中在具体模型的选择行为,缺乏对发现普适性的关注。如果认为选择是环境输入到行为输出的映射,那么AI的行为模式就是模型训练过程的算子:行为 $(\bullet) = T(\text{生成范式, 模型架构, 参数量, 训练数据分布, 对齐目标, } \dots)(\bullet)$ 。未来的研究应当努力挖掘AI的跨模型、本质性特征(即算子 $T$ 的结构),而非停留在易因模型迭代而产生变化的具体偏好参数上。二是在研究动机上,应当关注更加本质的人机互动问题。例如,对齐目标与应用目标错配是否会产生代理问题(agency problem)?在对齐过程中,是否存在深层次技术和经济约束?三是可以建立人与AI共通的行为和认知模型,从而为进一步研究人-AI、AI-AI互动的均衡影响提供基础。

#### (四) AI作为朋友

如果我们仅仅将AI视作分析工具,那么就可能错失了它的全部潜力。生成式人工智能表现出的“拟人化交互”特征,使得科研工作者获得了一个全天候陪伴的“朋友”,一个很少对人指手画脚、能够即时互动、帮助研究者调节情绪和整理社交认知的对话对象。

高强度竞争和高度不确定性的学术环境显著增加了科研人员的心理压力

(Bergvall et al., 2025),而生成式人工智能不仅是科研辅助工具,更可被视为一种低成本、即时可得的社会支持与情绪协作资源。已有研究表明,AI可以产生相当具有同理心的回复(Sharma et al., 2023; De Freitas et al., 2025)。近期的证据表明,与AI对话有助于改善心理状态、缓解孤独感(McBain et al., 2025; Meng et al., 2025)。但C. Fang et al.(2025)也指出,过度使用AI也可能导致依赖和孤独感的增强。因此,AI是真实社会支持的补充,而非完全替代。

**例 16(提供心理辅导):**我是一个经济学博士生,但我总感觉自己每天都在调整统计模型,跑的回归都没有什么意义;论文发表之后也不知道对现实世界有什么影响。你怎么看待这个想法呢?

科研活动还高度依赖沟通与协作。相关研究发现,大语言模型在特定任务中已展现出对他人信念与意图的推断能力(Strachan et al., 2024),这为其在分析社会情境、规划沟通策略以及改善语言表达等方面的应用提供了可能性。在实践中,研究者可以用AI梳理复杂的人际情境并起草表达方案,从而降低真实互动中的沟通成本。

**例 17(设计沟通方案):**我有一个科研项目,是很有名的大老板出的想法,带了另一个技术很强的人做,写了一个研究计划,被会议接收了。但后来技术强的人撂挑子不干了,现在我在接手。大老板很焦虑,害怕出不了结果,这样会让大老板很丢脸。我有一些初步的结果,但依然有很多瑕疵,需要进一步调整。我不想误导其他合作者。请帮我想想如何在项目组里面沟通。

## 五、人工智能对学术生态的系统性风险<sup>①</sup>

学术知识生产系统中充满了信息不对称和外部性(Partha and David, 1994; Stephan, 1996)。尽管生成式人工智能有望在研究者个体层面提高生产率,但它的广泛使用可能给学术共同体带来非预期的系统性风险。如何构建适应于人工智能技术的学术体系,已经成为学术共同体亟待回答的问题。

一是学术训练中知识理解深度的退化。科研训练本质上是掌握一门“手艺”,学生不仅需要知道“是什么”,更需要在探索的过程中理解“为什么”、掌握“怎么做”。生成式AI尽管可以提供一站式的信息整合,却也减少了学生在亲自探索过程中的意外发现,绕开了思维方式的训练,最终有可能削弱学生对知识的掌握(Melumad and Yun, 2025; Bastani et al., 2025; Georgiou, 2025)。仅仅依靠短期内的产出,已经无法有效衡量学生对知识的实际掌握。对于整个

<sup>①</sup> 作者感谢匿名审稿人提议增加此部分。

研究生培养体系而言,如何设计合理的训练和考评机制,督促学生真正深入调研和思考,掌握关键的科学问题与正确思维方法,已经成为无法回避的问题。

二是研究内容的趋同化。科学的突破往往来自超出当前思想范畴的“极端值”,但 AI 作为语言统计模型的本质,将本应当天马行空的人类思想,拉回到了训练语料中的最大后验(Doshi and Hauser, 2024; Zhang et al., 2025)。如果 AI 对于特定话题、方法或数据能够提供更强的辅助,研究者是否会涌入这些 AI 擅长的研究,而忽视了其他重要议题?当 AI 能够快速编出平庸但自圆其说的“机制分析”时,研究者是否还有动机通过深入的调研,理解经济真正的运作逻辑?

三是同行评议体系崩溃的风险。同行评议的本质是审稿人无偿地为学术共同体甄别研究质量。流畅的语言、严谨的数学模型和精美的图表曾经是用于快速甄别研究水平的可信信号,但生成式 AI 使得低水平的研究也可以包装得光鲜亮丽,破坏了信号的可信性,提高了甄别成本。随着 AI 生产大量低水平论文,审稿负担日益加重,这反过来又放大了审稿人方面的道德风险,他们也开始利用 AI 生成评审意见(Liang et al., 2024)。但是, AI 不具备合格的学术审美,其评审存在显著的随机性,且倾向于产生正面的评价(Russo, 2025)。当学术变成了 AI 写、AI 评,人类的话语权又将置于何处?

学术界已经在探索 AI 使用规范,但现有的制度远远不能适应新的形势。Springer Nature、Taylor & Francis 等超过 70% 的出版商已经规定研究者必须披露生成式 AI 的使用方式(He and Bu, 2026)。然而, He and Bu(2026)通过分析大量论文全文发现,尽管 AI 写作的后验概率自 ChatGPT 推出以来急剧上升,仅有 0.1% 的论文明确承认使用了 AI,并且期刊的 AI 强制披露政策对是否使用 AI 写作没有显著影响。这说明当前的 AI 透明度政策未能达到预期效果。值得注意的是, Reif et al.(2025)发现,人们认为使用 AI 完成任务的人能力更差、更懒惰;这种偏见解释了学术评议中的强制性 AI 披露政策为何与研究者的激励不相容。祛除 AI 使用的污名化,关注思想实质而非写作过程,是一种需要漫长适应但必需的转变。

无论我们是否乐意,越来越多的科研工作者使用 AI 已经是难以逆转的事实(He and Bu, 2026; Feyzollahi and Rafizadeh, 2025)。面对 AI 冲击,学术共同体需要进行两方面努力。一是“制度工程”,通过激励相容的披露机制、长期化的声誉追踪等制度设计,缓解培养、研究和评议体系中的道德风险。例如,通过公开评议(open review)等机制提高审稿人的声誉激励,通过预注册实验、数据公开等方式提高研究内容的可验证性。但制度的有效性取决于执行者的能力。如果审稿人本身无法区分真正的理论洞见和 AI 生成的平庸内容,任何制度都会沦为形式。二是“能力进化”,将 AI 视为一次学术能力的大考。它迫使

我们回答：什么是 AI 无法模仿的学术能力？如何在培养体系中强化这些能力？如何在评价体系中识别和奖励这些能力？

学术共同体需要进化到 AI 无法触及的维度。具体而言：在培养中，从“教技能”转向“教品味”，让学生学会区分平庸与卓越；在评价中，从“看形式”转向“看洞见”，奖励那些基于真实世界提出了新问题、新概念、新机制的工作；在审稿中，提高审稿人的理论鉴赏力。与此同时，学术界也应当建立包容反常与容忍失败的制度通道，鼓励经济管理与计算机科学等学科的深度融合，为 AI 引发的潜在范式变革提供环境。在经济管理学科历史上，Stata 的出现使得回归分析极其简单，但它最终迫使学术评价的重心从模型复杂度转向研究设计。AI 的出现也将倒逼我们从“表面功夫”转向“思想深度”，通过促进常规科学(normal science)的饱和与创造旧范式所无法解释的“异常”，为引领范式变革的非常规科学(extraordinary science)提供机遇(Kuhn, 1962)。能否抓住这个契机完成能力进化，将决定未来学术生态的活力。

## 六、总结性评论

七十年前，经济学曾经孕育了人工智能。冯·诺伊曼和摩根斯坦的《博弈论与经济行为》(Von Neumann and Morgenstern, 1944)、赫伯特·西蒙的“满意规则”(Simon, 1955)，都为人工智能的训练与决策奠定了基石。七十年间，人工智能与经济学渐行渐远，人工智能坚持追求工程可行与性能满意，而经济学不断追求理论优雅和因果识别。终于在七十年后的今天，人工智能以“外生冲击”的姿态与经济学再次交汇。面对新一代人工智能技术的挑战，经济管理研究者应该如何自处？

我们倡议，经济管理研究者应当在人工智能时代主动作为。我们应该学习 AI，理解算法逻辑与智能决策机制；理解 AI，识别 AI 的能力边界并实现人与 AI 在科学研究中的协作；拥抱 AI，让 AI 全面融入教学、研究与管理；更可以研究 AI，把 AI 视为新的经济行为主体，开启“智能体经济学”的新篇章。

本文既是对人工智能时代经济管理研究的一个方法论探索，也向所有经济管理研究者发出了迎接 AI 时代的呼吁。本文介绍 AI 的技术原理与特性，旨在“祛魅”(demystify)人工智能，发掘 AI 技术与经济模型的联系，促进经济管理研究与 AI 技术的深度融合。本文识别了人工智能在社会科学研究中的边界，提出了人类社会科学家在构建理论、塑造人类思想方面不可替代的价值。在实践层面，本文提出了 AI 融入研究全流程的四种角色：参谋、助研、智能体与朋友，并辅以参考案例和文献，为研究者更好地运用 AI 提供了思路参考。这些角色共同构成了人机协作的知识生产函数，为经济管理学研究的范式转型提供了基

基础框架。

我们展望,人工智能时代的经济管理研究将会有两方面变革。一是从堆砌工作量与嫁接高级方法的竞争中解放出来,更加关注思想本身的创新性和深度;二是从单纯对人类行为、组织和市场的研究,扩展到理解多种智能体(人类与AI)共存系统的科学。本文引言中曾经提到人工智能对经济学的机会与挑战:彼时的机会是增强,而挑战是对研究流程的自动化。而随着经济学将AI纳入研究对象,另一组机会与挑战浮出水面:机会在于扩充学科边界,研究算法决策、人机交互等全新现象;挑战则在于反思学科根基。当一个没有欲望、不能享受福祉的算法可以用经济学研究时,这迫使我们追问:经济学究竟在研究“行为的形式结构”,还是“有目的和意义的人类选择”?本文第二部分强调AI“缺乏惊奇与价值判断”,第四部分论证“AI可以是研究对象”,这一张力并非疏漏,而是揭示了经济学自萨缪尔森以来未曾真正解决的身份困惑:经济学究竟是研究行为的形式科学,还是研究人的意义科学?对这一问题的不同回答,将决定经济学在智能时代的角色与边界。

智能的曙光已经出现在地平线上,尽管它同时投下了我们必须直面的漫长阴影。希望本文的介绍能够启发更多研究者,让我们共同迎接人机协作的经济管理研究新时代。

## 参考文献

- [1] Acerbi, A., and J. M. Stubbersfield, “Large Language Models Show Human-Like Content Biases in Transmission Chain Experiments”, *Proceedings of the National Academy of Sciences*, 2023, 120(44), e2313790120.
- [2] Agrawal, A., J. McHale, and A. Oettl, “Finding Needles in Haystacks: Artificial Intelligence and Recombinant Growth”, In: Agrawal, A., J. Gans, and A. Goldfarb(eds.), *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press, 2018, 149-174.
- [3] Andre, P., C. Pizzinelli, C. Roth, and J. Wohlfart, “Subjective Models of the Macroeconomy: Evidence from Experts and Representative Samples”, *The Review of Economic Studies*, 2022, 89(6), 2958-2991.
- [4] Athey, S., and G. W. Imbens, “Machine Learning Methods that Economists Should Know About”, *Annual Review of Economics*, 2019, 11(1), 685-725.
- [5] Autor, D. H., D. Dorn, and G. H. Hanson, “The China Syndrome: Local Labor Market Effects of Import Competition in the United States”, *American Economic Review*, 2013, 103 (6), 2121-68.
- [6] Axtell, R. L., and J. D. Farmer, “Agent-Based Modeling in Economics and Finance: Past, Present, and Future”, *Journal of Economic Literature*, 2025, 63(1), 197-287.
- [7] Bail, C. A., “Can Generative AI Improve Social Science?”, *Proceedings of the National Academy of Sciences*, 2024, 121(21), e2314021121.
- [8] Bastani, H., O. Bastani, A. Sungu, H. Ge, Ö. Kabakçı, and R. Mariman, “Generative AI Without

- Guardrails Can Harm Learning: Evidence from High School Mathematics”, *Proceedings of the National Academy of Sciences*, 2025, 122(26), e2422633122.
- [9] Bergvall, S., C. Fernström, E. Ranehill, and A. Sandberg, “The Impact of PhD Studies on Mental Health—A Longitudinal Population Study”, *Journal of Health Economics*, 2025, 103070.
- [10] Bini, P., L. W. Cong, X. Huang, and L. J. Jin, “Behavioral Economics of AI: LLM Biases and Corrections”, *Available at SSRN* 5213130, 2025.
- [11] Bordalo, P., J. J. Conlon, N. Gennaioli, S. Y. Kwon, and A. Shleifer, “Memory and Probability”, *The Quarterly Journal of Economics*, 2023, 138(1), 265-311.
- [12] Bordalo, P., N. Gennaioli, and A. Shleifer, “Salience”, *Annual Review of Economics*, 2022, 14(1), 521-544.
- [13] Bybee, L., “The Ghost in the Machine: Generating Beliefs with Large Language Models”, *Working Paper*, 2025.
- [14] Cao, S., C. C. Wang, and Y. Xiang, “When LLMs Go Abroad: Foreign Bias in AI Financial Predictions”, *Available at SSRN* 5440116, 2025.
- [15] Caplin, A., D. Martin, and P. Marx, “Modeling Machine Learning: A Cognitive Economic Approach”, *Journal of Economic Theory*, 2025, 224, 105970.
- [16] Charness, G., B. Jabarian, and J. A. List, “The Next Generation of Experimental Research with LLMs”, *Nature Human Behaviour*, 2025, 1-3.
- [17] Chen, J., G. Tang, G. Zhou, and W. Zhu, “ChatGPT and DeepSeek: Can They Predict the Stock Market and Macroeconomy?”, *arXiv preprint arXiv: 2502.10008*, 2025.
- [18] Chen, Q., T. Han, J. Li, Y. Luo, Y. Wu, X. Zhang, and T. Zhou, “Can AI Master Econometrics? Evidence from Econometrics AI Agent on Expert-Level Tasks”, *arXiv preprint arXiv: 2506.00856*, 2025.
- [19] Chen, Y., H. Fang, Y. Zhao, and Z. Zhao, “Recovering Overlooked Information in Categorical Variables with LLMs: An Application to Labor Market Mismatch”, *NBER Working Paper No. w32327*, 2024.
- [20] Chen, Y., S. N. Kirshner, A. Ovchinnikov, M. Andiappan, and T. Jenkin, “A Manager and an AI Walk into a Bar: Does ChatGPT Make Biased Decisions Like We Do?”, *Manufacturing & Service Operations Management*, 2025, 27(2), 354-368.
- [21] Chen, Y., T. X. Liu, Y. Shan, and S. Zhong, “The Emergence of Economic Rationality of GPT”, *Proceedings of the National Academy of Sciences*, 2023, 120(51), e2316205120.
- [22] Chen, Z., and J. Chan, “Large Language Model in Creative Work: The Role of Collaboration Modality and User Expertise”, *Management Science*, 2024, 70(12), 9101-9117.
- [23] Chernozhukov, V. et al., “Double/Debiased Machine Learning for Treatment and Structural Parameters”, *The Econometrics Journal*, 2018.
- [24] Cong, L. W., G. Feng, J. He, and X. He, “Growing the Efficient Frontier on Panel Trees”, *Journal of Financial Economics*, 2025, 167, 104024.
- [25] Dai, X., L. Zhou, B. Wang, and H. Li, “From Word to World: Evaluate and Mitigate Culture Bias in LLMs via Word Association Test”, in *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, 2025, 24521-24537.
- [26] Danielsson, J., R. Macrae, and A. Uthemann, “Artificial Intelligence and Systemic Risk”, *Journal*

- of Banking and Finance*, 2022, 140, 106290.
- [27] De Freitas, J., Z. Oğuz-Uğuralp, A. K. Uğuralp, and S. Puntoni, “AI Companions Reduce Loneliness”, *Journal of Consumer Research*, 2025.
- [28] De Kok, T., “ChatGPT for Textual Analysis? How to Use Generative LLMs in Accounting Research”, *Management Science*, 2025, 71(9), vii-x, 7223-8095, iv-vi.
- [29] Doshi, A. R., and O. P. Hauser, “Generative AI Enhances Individual Creativity but Reduces the Collective Diversity of Novel Content”, *Science Advances*, 2024, 10(28), eadn5290.
- [30] Eloundou, T., S. Manning, P. Mishkin, and D. Rock, “GPTs Are GPTs: Labor Market Impact Potential of LLMs”, *Science*, 2024, 384(6702), 1306-1308.
- [31] Evans, J. S. B., and K. E. Stanovich, “Dual-Process Theories of Higher Cognition: Advancing the Debate”, *Perspectives on Psychological Science*, 2013, 8(3), 223-241.
- [32] Fang, C. M., A. R. Liu, V. Danry, E. Lee, S. W. T. Chan, P. Pataranutaporn, P. Maes, J. Phang, M. Lampe, L. Ahmad, and S. Agarwal, “How AI and Human Behaviors Shape Psychosocial Effects of Chatbot Use: A Longitudinal Controlled Study”, 2025.
- [33] Fang, H., M. Li, and G. Lu, “Decoding China’s Industrial Policies”, *NBER Working Paper No. w33814*, 2025.
- [34] Farmer, J. D., and D. Foley, “The Economy Needs Agent-Based Modelling”, *Nature*, 2009, 460 (7256), 685-686.
- [35] Farrell, H. et al., “Large AI Models Are Cultural and Social Technologies”, *Science*, 2025, 387 (6739), 1153-1156.
- [36] Feyzollahi, M., and N. Rafizadeh, “The Adoption of Large Language Models in Economics Research”, *Economics Letters*, 2025, 250, 112265.
- [37] Fish, S., Y. A. Gonczarowski, and R. I. Shorrer, “Algorithmic Collusion by Large Language Models”, *arXiv preprint arXiv: 2404.00806*, 2024.
- [38] Fong, H., and G. Gui, “Modeling Story Expectations to Understand Engagement: A Generative Framework Using LLMs”, *arXiv preprint arXiv: 2412.15239*, 2024.
- [39] Friedman, M., *The Methodology of Positive Economics*, 1953.
- [40] Gans, J. S., “Growth in AI Knowledge”, *NBER Working Paper 33907*, 2025.
- [41] Gao, C., X. Lan, N. Li, Y. Yuan, J. Ding, Z. Zhou, et al., “Large Language Models Empowered Agent-Based Modeling and Simulation: A Survey and Perspectives”, *Humanities and Social Sciences Communications*, 2024, 11(1), 1-24.
- [42] Gentzkow, M., B. Kelly, and M. Taddy, “Text as Data”, *Journal of Economic Literature*, 2019, 57(3), 535-574.
- [43] Georgiou, G. P., “ChatGPT Produces More ‘Lazy’ Thinkers: Evidence of Cognitive Engagement Decline”, *arXiv preprint arXiv: 2507.00181*, 2025.
- [44] Gu, S., B. Kelly, and D. Xiu, “Empirical Asset Pricing via Machine Learning”, *The Review of Financial Studies*, 2020, 33(5), 2223-2273.
- [45] Gui, G., and O. Toubia, “The Challenge of Using LLMs to Simulate Human Behavior: A Causal Inference Perspective”, *arXiv preprint arXiv: 2312.15524*, 2023.
- [46] Guo, D., D. Yang, H. Zhang, J. Song, P. Wang, Q. Zhu, et al., “DeepSeek-R1 Incentivizes Reasoning in LLMs Through Reinforcement Learning”, *Nature*, 2025, 645(8081), 633-638.

- [47] Hagendorff, T., S. Fabi, and M. Kosinski, "Human-Like Intuitive Behavior and Reasoning Biases Emerged in Large Language Models but Disappeared in ChatGPT", *Nature Computational Science*, 2023, 3(10), 833-838.
- [48] He, Y., and Y. Bu, "Academic Journals' AI Policies Fail to Curb the Surge in AI-Assisted Academic Writing", *Proceedings of National Academy of Science*, 2026, 123 (9) e2526734123.
- [49] Horton, J. J., "Large Language Models as Simulated Economic Agents: What Can We Learn from Homo Silicus?", *NBER Working Paper No. 31122*, 2023.
- [50] Jha, M., H. Liu, and A. Manela, "Does Finance Benefit Society? A Language Embedding Approach", *The Review of Financial Studies*, 2025, hhaf012.
- [51] Jha, M., J. Qian, M. Weber, and B. Yang, "Harnessing Generative AI for Economic Insights", *arXiv preprint arXiv: 2410.03897*, 2024.
- [52] Kahneman, D., *Thinking, Fast and Slow*. Macmillan, 2011.
- [53] Kaplan, J., S. McCandlish, T. Henighan, T. B. Brown, B. Chess, R. Child, et al., "Scaling Laws for Neural Language Models", *arXiv preprint arXiv: 2001.08361*, 2020.
- [54] Kim, A., M. Muhn, and V. Nikolaev, "Bloated Disclosures: Can ChatGPT Help Investors Process Information?", *arXiv preprint arXiv: 2306.10224*, 2023.
- [55] Korinek, A., "Generative AI for Economic Research: Use Cases and Implications for Economists", *Journal of Economic Literature*, 2023, 61(4), 1281-1317.
- [56] Kozlowski, A. C., M. Taddy, and J. A. Evans, "The Geometry of Culture: Analyzing the Meanings of Class Through Word Embeddings", *American Sociological Review*, 2019, 84(5), 905-949.
- [57] Kuhn, T. S., *The Structure of Scientific Revolutions*. University of Chicago Press, 1962.
- [58] Lewis, P., E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks", *Advances in Neural Information Processing Systems*, 2020, 33, 9459-9474.
- [59] Li, N., C. Gao, M. Li, Y. Li, and Q. Liao, "EconAgent: Large Language Model-Empowered Agents for Simulating Macroeconomic Activities", in *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2024, 15523-15536.
- [60] Liang, W., Z. Izzo, Y. Zhang, H. Lepp, H. Cao, X. Zhao, L. Chen, H. Ye, S. Liu, Z. Huang, D. McFarland, and J. Y. Zou, "Monitoring AI-Modified Content at Scale: A Case Study on the Impact of ChatGPT on AI Conference Peer Reviews", in *Proceedings of the 41st International Conference on Machine Learning*, 2024, 235, 29575-29620.
- [61] 林建浩、孙乐轩, "大语言模型与经济金融文本分析: 基本原理、应用场景与研究展望", 《计量经济学报》, 2025年第5期, 第1—34页。
- [62] Loewenstein, G., and Z. Wojtowicz, "The Economics of Attention", *Journal of Economic Literature*, 2025, 63(3), 1038-1089.
- [63] Lopez-Lira, A., and Y. Tang, "Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models", *Available at SSRN*, 2023.
- [64] Lucas, R. E., Jr., "*Econometric Policy Evaluation: A Critique*", Carnegie-Rochester Conference Series on Public Policy, 1976, 1, 19-46.
- [65] Ludwig, J., and S. Mullainathan, "Machine Learning as a Tool for Hypothesis Generation", *The Quarterly Journal of Economics*, 2024, 139(2), 751-827.

- [66] Ludwig, J., S. Mullainathan, and A. Rambachan, "Large Language Models: An Applied Econometric Framework", *NBER Working Paper No. w33344*, 2025.
- [67] Manning, B. S., and J. J. Horton, "General Social Agents", *NBER Working Paper*, 2025.
- [68] Manning, B. S., K. Zhu, and J. J. Horton, "Automated Social Science: Language Models as Scientist and Subjects", *NBER Working Paper No. w32381*, 2024.
- [69] Manzoor, E. et al., "Workshop on AI and Behavioral Science (AIBS 2024)", 2024. Available at: <https://ai-behavioral-science.github.io/2024>.
- [70] McBain, R. K., R. Bozick, M. Diliberti, et al., "Use of Generative AI for Mental Health Advice among US Adolescents and Young Adults", *JAMA Network Open*, 2025, 8(11).
- [71] McCarthy, J., M. L. Minsky, N. Rochester, and C. E. Shannon, "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955", *AI Magazine*, 2006, 27(4), 12-12.
- [72] Mei, Q., Y. Xie, W. Yuan, and M. O. Jackson, "A Turing Test of Whether AI Chatbots Are Behaviorally Similar to Humans", *Proceedings of the National Academy of Sciences*, 2024, 121(9), e2313925121.
- [73] Melumad, S., and J. H. Yun, "Experimental Evidence of the Effects of Large Language Models Versus Web Search on Depth of Learning", *PNAS Nexus*, 2025, 4(10), pgaf316.
- [74] Meng, J., "AI Emerges as the Frontier in Behavioral Science", *Proceedings of the National Academy of Sciences*, 2024, 121(10), e2401336121.
- [75] Meng, J., R. Zhang, J. Qin, Y. -J. Lee, and Y. -C. Lee, "AI-Mediated Social Support: The Prospect of Human-AI Collaboration", *Journal of Computer-Mediated Communication*, 2025, 30(4).
- [76] Mikolov, T., K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space", in *Proceedings of the Workshop at the 1st International Conference on Learning Representations (ICLR 2013)*, 2013, 1-12.
- [77] Motoki, F., V. Pinho Neto, and V. Rodrigues, "More Human than Human: Measuring ChatGPT Political Bias", *Public Choice*, 2024, 198(1), 3-23.
- [78] Muennighoff, N., Z. Yang, W. Shi, X. L. Li, L. Fei-Fei, H. Hajishirzi, et al., "S1: Simple Test-Time Scaling", *arXiv preprint arXiv: 2501.19393*, 2025.
- [79] Mullainathan, S., and J. Spiess, "Machine Learning: An Applied Econometric Approach", *Journal of Economic Perspectives*, 2017, 31(2), 87-106.
- [80] Novy-Marx, R., and M. Z. Velikov, "AI-Powered (Finance) Scholarship", *NBER Working Paper No. w33363*, 2025.
- [81] Ouyang, L., J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, et al., "Training Language Models to Follow Instructions with Human Feedback", *Advances in Neural Information Processing Systems*, 2022, 35, 27730-27744.
- [82] Ouyang, S., H. Yun, and X. Zheng, "AI as Decision-Maker: Risk Preferences of LLMs", *Available at SSRN 4851711*, 2024.
- [83] Partha, D., and P. A. David, "Toward a New Economics of Science", *Research Policy*, 1994, 23(5), 487-521.
- [84] Qin, X., M. Huang, and J. Ding, "AITurk: Using ChatGPT for Social Science Research", *Available at SSRN*, 2024.

- [85] Qu, Y., and J. Wang, "Performance and Biases of Large Language Models in Public Opinion Simulation", *Humanities and Social Sciences Communications*, 2024, 11(1), 1-13.
- [86] Rafailov, R., A. Sharma, E. Mitchell, C. D. Manning, S. Ermon, and C. Finn, "Direct Preference Optimization: Your Language Model Is Secretly a Reward Model", *Advances in Neural Information Processing Systems*, 2023, 36, 53728-53741.
- [87] Reif, J. A., R. P. Larrick, and J. B. Soll, "Evidence of a Social Evaluation Penalty for Using AI", *Proceedings of the National Academy of Sciences*, 2025, 122(19), e2426766122.
- [88] Ross, J., Y. Kim, and A. W. Lo, "LLM Economicus? Mapping the Behavioral Biases of LLMs via Utility Theory", *arXiv preprint arXiv: 2408.02784*, 2024.
- [89] Russo, G., M. Horta Ribeiro, T. R. Davidson, V. Veselovsky, and R. West, "The AI Review Lottery: Widespread AI-Assisted Peer Reviews Boost Paper Scores and Acceptance Rates", *Proceedings of the ACM on Human-Computer Interaction*, 2025, 9(7), Article CSCW486.
- [90] Samuelson, P. A., "Consumption Theory in Terms of Revealed Preference", *Economica*, 1948, 15(60), 243-253.
- [91] Santurkar, S., E. Durmus, F. Ladhak, C. Lee, P. Liang, and T. Hashimoto, "Whose Opinions Do Language Models Reflect?", in *Proceedings of the 40th International Conference on Machine Learning (ICML)*, PMLR 202, 2023, 30471-30504.
- [92] Sharma, A., I. W. Lin, A. S. Miner, et al., "Human-AI Collaboration Enables More Empathic Conversations in Text-Based Peer-to-Peer Mental Health Support", *Nature Machine Intelligence*, 2023, 5, 46-57.
- [93] Shiller, R. J., "Narrative Economics", *American Economic Review*, 2017, 107(4), 967-1004.
- [94] Simon, H. A., "A Behavioral Model of Rational Choice", *The Quarterly Journal of Economics*, 1955, 99-118.
- [95] Simon, H. A., "Rationality as Process and as Product of Thought", *The American Economic Review*, 1978, 68(2), 1-16.
- [96] Stephan, P. E., "The Economics of Science", *Journal of Economic Literature*, 1996, 34(3), 1199-1235.
- [97] Strachan, J. W. A., D. Albergo, G. Borghini, et al., "Testing Theory of Mind in Large Language Models and Humans", *Nature Human Behaviour*, 2024, 8, 1285-1295.
- [98] Sutton, R., "The Bitter Lesson", *Incomplete Ideas (blog)*, 2019, 13(1), 38.
- [99] Tao, Y., O. Viberg, R. S. Baker, and R. F. Kizilcec, "Cultural Bias and Cultural Alignment of Large Language Models", *PNAS Nexus*, 2024, 3(9), pga346.
- [100] Tranchero, M., C. F. Brennkmeijer, A. Murugan, and A. Nagaraj, "Theorizing with Large Language Models", *NBER Working Paper No. w33033*, 2024.
- [101] Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., "Attention Is All You Need", *Advances in Neural Information Processing Systems*, 2017, 30.
- [102] Von Neumann, J., and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.
- [103] Wager, S., and S. Athey, "Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests", *Journal of the American Statistical Association*, 2018, 113(523), 1228-1242.
- [104] Wang, S., Z. Yao, S. Zhang, J. Gai, T. X. Liu, and S. Zhong, "When Experimental Economics Meets Large Language Models: Evidence-based Tactics", *arXiv preprint arXiv: 2505.*

- 21371, 2025.
- [105] Wei, H., Y. Sun, and Y. Li, "DeepSeek-OCR: Contexts Optical Compression", *arXiv preprint arXiv*; 2510.18234, 2025.
- [106] Wei, J., X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, ... and D. Zhou, "Chain-of-thought Prompting Elicits Reasoning in Large Language Models", *Advances in neural information processing systems*, 35, 2022, 24824-24837.
- [107] Xie, Y., R. Gao, and Q. Mei, "Distributional Alignment for Social Simulation with LLMs: A Prompt Mixture Modeling Approach", in *First Workshop on Social Simulation with LLMs*, 2025.
- [108] Yao, S., J. Zhao, D. Yu, N. Du, I. Shafran, K. R. Narasimhan, and Y. Cao, "ReAct: Synergizing Reasoning and Acting in Language Models", in *The Eleventh International Conference on Learning Representations*, 2022.
- [109] Zhang, S., J. Xu, and A. J. Alvero, "Generative AI Meets Open-Ended Survey Responses: Research Participant Use of AI and Homogenization", *Sociological Methods and Research*, 2025, 00491241251327130.
- [110] 赵宣凯、宝恩德、左从江等, "大语言模型在经济金融领域的应用——研究评述、学术应用及未来展望", 《金融评论》, 2025年第17期, 第69—93页+第156—157页。

## The Light of Intelligence: Human-AI Collaboration and the Transformation of Economics and Management Research

HU Shiyun    YI Junjian\*  
(Peking University)

**Abstract:** Large language models (LLMs) are reshaping the production function of knowledge. This paper systematically outlines the technical foundations, application boundaries, and paradigm shifts triggered by LLMs for scholars in economics and management. We argue that theoretical creativity that rooted in abstract concept construction, counterfactual reasoning, and value judgment remains the irreplaceable comparative advantage of human scientists. LLMs can serve four distinct roles in the research process: as an Advisor, enhancing critical thinking and dialogue with existing literature; as an Assistant, automating research execution; as an Agent, functioning as both an experimental tool and a novel subject of study; and as a Companion, providing psychological and social support. However, the widespread adoption of LLMs also poses three potential systemic risks: the atrophy of deep under-

---

\* Corresponding Author: YI Junjian, China Center for Economic Research (CCER), National School of Development, Peking University, Haidian District, Beijing 100871, China; Tel: 86-10-62767530; E-mail: junjian@nsd.pku.edu.cn.

standing, the homogenization of research content, and the collapse of peer review mechanisms. Future research must embrace the “Light of Intelligence” while confronting its “Shadow”. By establishing a new paradigm of human-AI collaboration, scholars can expand the breadth and depth of human cognition, ultimately returning to the academic essence of theoretical innovation.

**Keywords:** artificial intelligence; economics methodology; human-AI collaboration

**JEL Classification:** A11, B41, C45