生成式人工智能驱动的管理研究方法：发展现状、应用模式与未来方向*

□赵　晨　林　晨

领域编辑推荐语：

生成式人工智能作为一种前沿的研究工具，已经广泛应用于管理学研究中。论文采用文献计量和内容分析方法，总结出生成式人工智能应用于管理学研究中的显著优势，为研究人员更好地利用生成式人工智能提供了指导。

——程聪

摘　要：生成式人工智能为管理学研究范式创新提供了机遇，也对其方法体系提出挑战。为推动生成式人工智能在管理研究中的规范应用，采用文献计量与内容分析法，对国内外文献进行全景梳理，力图提炼可资借鉴的方法体系。研究发现：生成式人工智能在管理研究中的应用呈初期积累后爆发增长的态势，集中于文本挖掘与信息提取等应用，并在宏中微观各级管理议题中展现广泛潜力；该类方法具有高效性、精准性、客观性、可解释性和可扩展性5大优势，但在可靠性和可重复性上仍存隐忧；通过整合文献中的应用方法、信效度保障及检验技术，提出12C－PIPE矩阵模型，通过覆盖生成式人工智能应用的4阶段12项关键流程，为研究者提供方法论指导；并从方法检验、适用边界、生态建设和伦理规范等方面，为后续研究提出展望。

关键词：人工智能；管理研究方法；研究工具；大语言模型

一、引言

自OpenAI于2022年推出ChatGPT 3.5以来，生成式人工智能凭借其突破性的语义理解能力，为管理学研究方法革新带来新机遇。尽管该技术已在情感分析

*　本文受国家社会科学基金重大项目“人才链支撑创新链产业链深度融合的机制与对策研究”（23&ZD052）的资助。作者感谢审稿专家和领域主编提出的宝贵意见，感谢《管理学研究》编辑部在本文投稿与修改过程中的耐心与专业精神。作者文责自负。

（Hur et al.，2024；Abdurahman et al.，2024）、非结构化数据识别（Schwitter，2025）、理论建构（Rathje et al.，2024；赵晨等，2025）及因果关系分析（Guo & Yang，2024）等方面展现出巨大潜力，许多学者仍就其在管理研究中仓促应用表达了忧虑（de Kok，2025；Wang et al.，2025；Harding et al.，2024）。其核心症结在于，现有应用探索多呈碎片化、零散化特征，尚未形成系统性的方法论共识（Abdurahman et al.，2024）。因此，如何对当前零散的方法论探索进行系统性整合，以探寻生成式人工智能与管理研究科学融合的合理路径，已成为亟待回应的关键议题。

当前，学界围绕生成式人工智能应用的理论分歧，主要源于其在带来效率革命的同时，也潜藏着方法论准备不足的深层隐忧（Abdurahman et al.，2024）。一方面，持积极观点的学者认为，相较于传统文本分析方法，生成式人工智能在分析成本、结果信效度及研究可重复性上实现了显著突破（李春涛等，2024），其接近人类水平的复杂语义解析能力（Jha et al.，2024）与经济学理性决策模拟能力（Chen et al.，2023），为管理学研究范式创新提供了重要突破口。另一方面，持审慎态度的学者则强调，该技术固有的局限性使其在注重严谨性的学科范式中面临合法性质疑。例如，其重语义逻辑、轻事实验证的底层算法所衍生的“幻觉”现象（de Kok，2025；Cheng et al.，2025），以及输出结果易受模型版本、提示工程等因素影响的内在不稳定性（Fišar et al.，2023），均对研究的精确性与可复现性构成了严峻挑战。尽管该技术已在质性分析（赵晨等，2025）、心理学测量（Ye et al.，2025）等领域催生了一系列新兴研究，但学界尚未就其局限性与学科范式冲突建立起系统化的应对规范，也未形成生成式人工智能在管理研究中的标准信守，这种方法论层面的不足已成为制约该技术进一步科学化应用的关键瓶颈。

面对生成式人工智能带来的机遇与挑战，本研究主张以审慎积极的态度，既拥抱其为管理学科带来的创新潜力，也直面其方法论不完善的现实（van Dis et al.，2023）。当前学界普遍呼吁在把握技术机遇的同时保持批判性反思（de Kok，2025；Wang et al.，2025；Harding et al.，2024；Abdurahman et al.，2024），并由此引发对操作规范、研究严谨性、结果不确定性与可验证性等问题的追问。为回应学界对方法创新的迫切期待，本研究以中国知网与 Web of Science 核心合集为数据源，结合文献计量分析与质性内容分析，对国内外应用生成式人工智能开展管理研究的文献进行全景扫描与深度解析，系统考查其研究格局、技术实现及方法论创新与局限，揭示该技术与管理学科深度整合的关键瓶颈，进而凝练标准化工作流程，提出前瞻性发展路径，以推动其在管理学科研究中的规范应用。

二、文献来源与梳理

（一）从研究对象到研究工具：生成式人工智能的定位转变

近年来，生成式人工智能在管理学研究中的定位明显转变，学界视角已从将其作为研究对象的浅层探讨，深化为作为研究工具的方法论创新应用。

具体而言，前者将生成式人工智能作为独立议题，探究其技术伦理与影响机制。伦理层面的研究系统考查了该技术应用于管理研究的合理边界（Harding et al.，2024）、潜在道德偏见（Pellert et al.，2024）及技术局限导致的系统性偏差（Abdurahman et al.，2024），强调在管理研究中需保证人类参与者和客观数据的核心地位以及对生成式工具的审慎使用。机制层面的研究则分析了其对员工培训（Gezdur & Bhattacharjya，2025）、财务决策（Saxena & Rishi，2025）、审计（Eulerich et al.，2024）、人力资源（Aguinis et al.，2024）等管理流程的再造，对金融市场（Beckmann & Hark，2024）和劳动力市场（Hui et al.，2024）的宏观影响，以及对使用动机（Lee & Park，2023）、消费态度（Karakaş & Jaeger，2025）、决策行为（Yamamoto，2024）、任务表现（Xu et al.，2025）等微观机制的作用。但这些研究范式仍基于传统分析框架，本质上以生成式人工智能为对象而非工具。

随研究深入，学界将其深度整合到方法论体系中，形成可用性论证、工具性开发与操作性应用三方面探讨。可用性论证验证了其在经济数据分析（Korinek，2023）、心理学理论建构（Tong et al.，2024）及质性研究（Smirnov，2025；高虎源等，2025）中的适用性，强调其具备从非结构化文本中提取核心概念、识别复杂指标并进行因果推断的语义解析能力（Schwitter，2025）。工具性开发既构建生成式人工智能在管理研究中的标准化应用框架，如提示工程（Naeem et al.，2025）和语义解析工作流（Li et al.，2025），也面向金融（Huang et al.，2023）、货币政策（Leek & Bischl，2025）等领域开发垂直模型，为解决特定管理问题提供专业化工具支持。操作性应用中，学者不仅利用生成式人工智能的文本分析，构建组织或群体的特征画像（张志豪等，2025；赵晨等，2025），还将其应用于供应链风险识别（Fan et al.，2025）、投资行为分析（Wu et al.，2024）等动态过程。这种从研究对象到研究工具的定位转变，反映了生成式人工智能不仅拓展了管理研究的理论视野和问题场景，同时正在重塑推动管理科学发展的方法论体系。

基于此，为系统梳理生成式人工智能在管理学科研究中的应用，本研究确立三方面文献筛选标准：一是检验方法可用性、弥合技术与研究方法鸿沟的研究；二是开发研究工具与标准化流程、回应文本解析精度、可解释性与伦理合规诉求的文献；三是依托生成式人工智能实现操作化创新、为应用范式提供操作模板的代表性成果。

（二）数据来源与处理

为确保样本文献的权威性与代表性，本研究将数据来源聚焦于 CSSCI 与 Web of Science（WoS）核心合集两大数据库。该搜索范围与国内外管理学多项文献计量研究一致，以确保数据来源的充分性与合理性（Naskar & Merigó Lindahl，2025；Zhuang & Sun，2025；李鸿磊 & 王凤彬，2025）。文献检索工作于 2025 年 6 月 5 日完成，搜索时间区间为 1985 年至检索时间。

在具体检索策略上，英文文献基于 WoS 核心合集，将研究领域限定在管理学、应用心理学、运筹学及经济学等学科类别，采用布尔检索式 TS = ("Generative artificial intelligence" OR

"Generative AI" OR "Large Language Model" OR "GLLM" OR "LLM")进行主题检索。中文文献则依托中国知网（CNKI），限定在 CSSCI 来源期刊且文献分类为经济与管理科学，使用检索式 TKA = ('生成式人工智能' + '大语言模型' + '大模型')。文献筛选流程如图 1 所示。

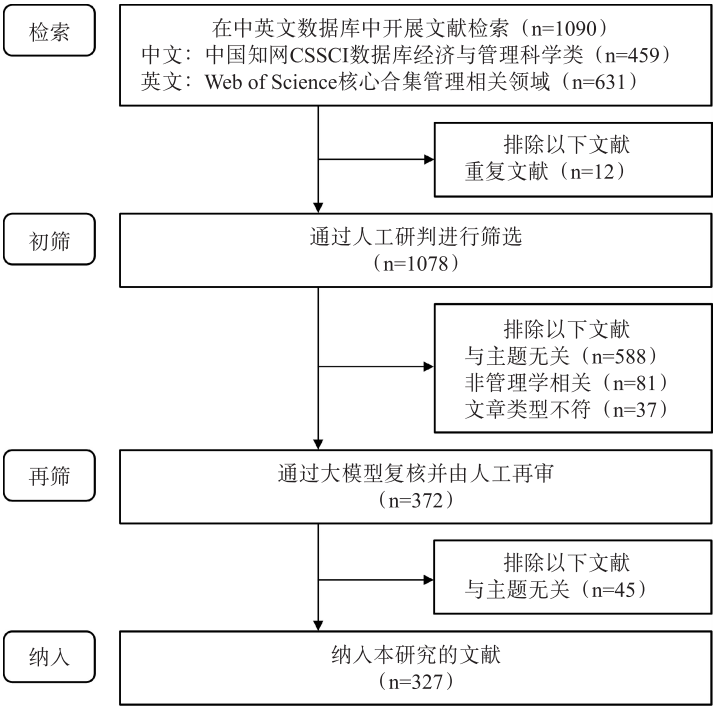

图 1　PRISMA 文献筛选流程

针对检出的题录，先剔除重复项，再经人工比对排除因数据不一致导致的隐性重复，构建初始文献库。由于所搜集文献中存在较多以生成式 AI 为主题，但仅以观点论述、文献梳理或用传统方法探讨其前因机制与影响效应，未涉及与管理研究方法结合的文献，故予以筛除：由两名研究者独立审阅标题与摘要，剔除未涉及 AI 与研究方法结合、非管理学或文章类型不符者，保留明确探讨生成式 AI 在管理研究中的适用性、工具开发或方法论融合的文献，分歧经讨论达成一致，共得中文文献 130 篇、英文文献 242 篇。

为验证人工筛选的可靠性，引入 DeepSeek - V3 大语言模型（版本 DeepSeek - V3 - 250324 - 1）进行复核。通过 Python 调用百度千帆平台 API，设 Temperature 与 Top P 均为 1、其余参数默认，要求模型判断摘要是否将生成式 AI 作为研究工具或方法，即探讨其在研究中的适用性、开发面向研究的 AI 工具或将其融入研究方法，而非将 AI 作为自变量或因变量的实质性研究，要求模型输出 1 或 0 并附简短解释。依据 Wu 等（2025）的自洽性验证方法，对每篇摘要清除历史记录后连续提问 5 次，以多数投票确定结果，以抑制幻觉、提升稳定性。初步复核的英文文献人机一致率为 0.905、中文为 0.831，团队对人机不一致的文献进行复核。经上述流程，最终纳入英文文献 219 篇、中文文献 108 篇。

三、文献计量分析

（一）文献特征

中英文文献随时间的发表趋势如图 2 所示，其中左图呈现了各时间节点的累计发文量，右图反映了不同时间段中英文文献的发表数量。

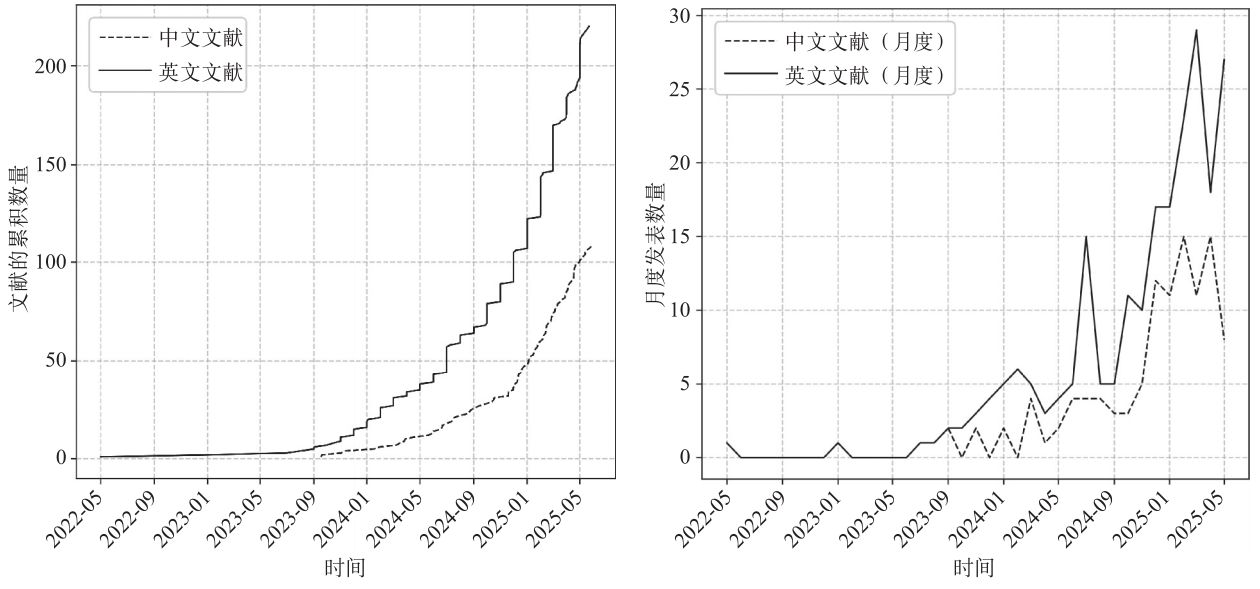

图 2　中英文文献的发表趋势

通过对发文趋势的分析发现，中英文文献的发表量分布均呈现阶段性特征。就中文文献而言，2023 年 9 月至 2024 年 10 月为初始积累期，月度发文量维持在 5 篇以下的较低水平，但已显现稳步上升态势；2024 年 10 月后进入快速发展期，发文量跃升至 10～15 篇区间。英文文献的演进轨迹与之类似：2023 年 7 月至 2024 年 9 月为缓慢增长期，月度发文量多数处于 10 篇以下；2024 年 9 月后则进入爆发式增长阶段，发文量迅速突破 10 篇，并在 2025 年达到 15～30 篇的较高水平。两段曲线加速增长的时间节点（2024 年 Q4 左右）可能与生成式人工智能的技术迭代及学术出版规律有关。尽管 ChatGPT－3.5 于 2022 年年底发布，但真正具备严谨逻辑推理能力、能辅助学术研究的关键事件是 2023 年 3 月 GPT－4 的发布。考虑到从工具应用于实证研究到论文撰写、同行评审及最终发表普遍存在的时间滞后性，发文高峰在技术成熟一年半后集中涌现符合学术规律。此外，2023 年年底多家权威期刊发布的生成式 AI 特刊征稿也进一步推动了该领域成果在 2024 年年底的集中产出。总体而言，中英文文献的发文量与增长速度表明，生成式人工智能与管理学的学科交叉融合已度过初步探索期，展现出发展潜力。

（二）关键词分布

运用 CiteSpace 做关键词共现与聚类分析：时间切片 1 年，节点为关键词，采用 Top N 法取各切片前 50 位高频词，中心度采用介数中心度（Chen，2006）。

接着，采用战略坐标图分析法（Law et al.，1988），系统考察生成式人工智能在管理学科的研究应用格局，如图 3 所示。在分析过程中，为避免通用词掩盖具体研究议题，剔除“大语言模型”“生成式人工智能”“large language models”等作为检索主题的关键词，重点关注具有信息增量的具体应用词汇。构建以关键词频次和中心度的均值为基准的虚线坐标系，其中，位于第一象限的关键词具备高频次、高中心度特点，代表研究热点，表明中英文文献共同聚焦于知识图谱、提示工程及文本挖掘等核心领域，揭示了当前管理学界主要聚焦于生成式人工智能在文本挖掘与信息提取方面的创新应用；此外，英文文献还形成“Deep learning”“In – context learning”等机器学习与自然语言处理交叉领域的关键词热点，反映了国际学术界在管理工具开发及多种人工智能方法融合方面的前沿探索。

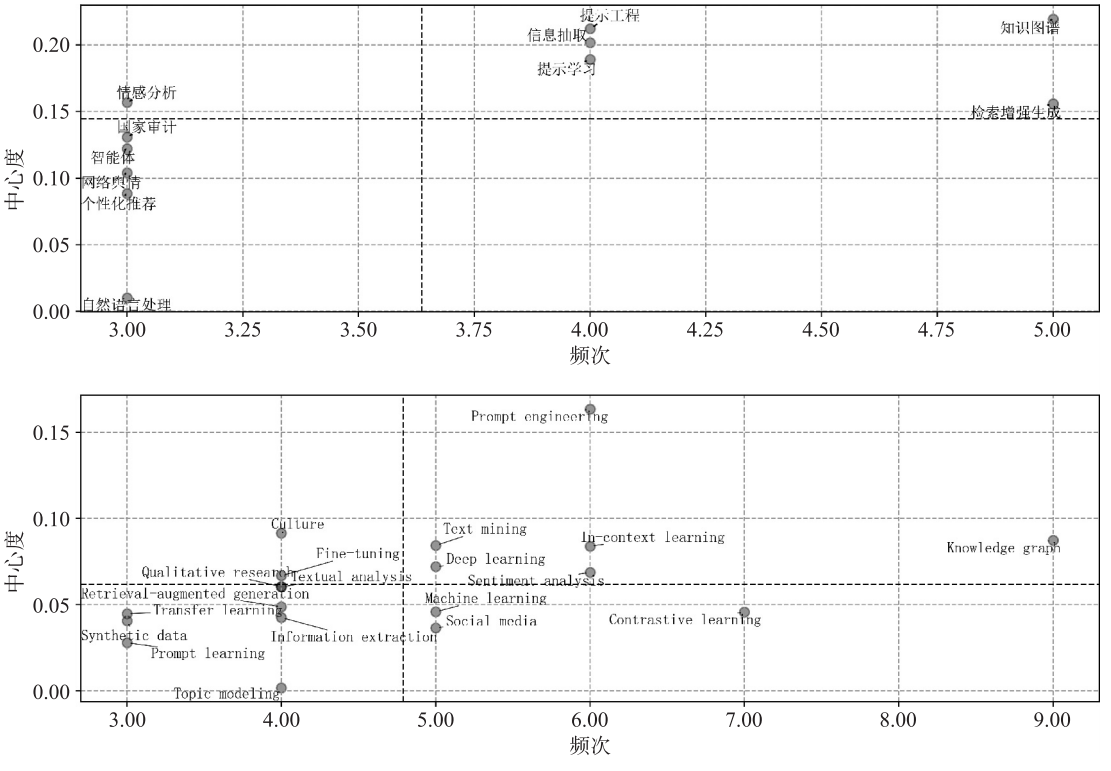

图 3　生成式人工智能与管理学科交叉研究的关键词分布

进一步关键词聚类分析显示，中文文献模块度指数为 0.939、英文为 0.677，聚类轮廓性指数分别为 0.945 和 0.979，结果具备良好可靠性与结构效度（陈悦等，2015）。结合 Ferreira 和 Otley（2009）的管理控制模型，本研究识别出生成式人工智能与管理学交叉研究的 4 个核心维度（见表 1）：①战略规划机制维度，探讨其在宏观管理中的应用，如政策分析、趋势预测、战略资源配置，体现为地理空间数据分析、中美贸易战评估等；②战术决策机制维度，考查中观决策应用，涵盖市场预测、消费者行为分析、资产配置优化与舆情监测；③业务处理机制维度，聚焦微观运营中的安全、质量、医疗管理等流程自动化重构与效率提升；④管理

信息系统维度，运用其探索面向管理问题的技术体系，包括系统架构设计、算法训练优化与知识图谱构建等。

表 1　　生成式人工智能与管理学科交叉研究的关键词聚类

分类	聚类名称	节点数量	代表性节点
管理信息系统	系统开发	115	novelty score prediction；pre－trained language model；recommendation system；outfit recommendation；contrastive reinforcement
	算法训练	36	information extraction；multi－lingual training；cross－lingual transfer；disentangled training；transfer learning
	知识提取	32	incremental knowledge graph；full knowledge graph；static knowledge graph；security knowledge graph；knowledge representation learning
		15	知识抽取；知识图谱；提示构建；提示工程；信息抽取
战略规划机制	宏观战略	17	multi－source geospatial data；land－use classification；Sino－US trade war；exchange rate forecasting
战术决策机制	市场分析	52	destination marketing；social media influencers；text mining；supply chain finance；internet－sourced data；financial analytics
	营销分析	33	new energy vehicle；user satisfaction；customer experience；social media；topic modeling
	金融分析	13	asset selection；financial advice；stock picking；accounting information systems；data extraction
	舆情分析	12	网络舆情；舆情分析；情感分析；公共突发事件；态势恶化
业务处理机制	安全管理	34	incident reporting；information retrieval；safety analysis；hazard analysis；visual perception
	质量管理	32	corporate social responsibility；business ethics；energy efficiency；quality control；defect manufacturing
	医疗管理	28	automatic ICD coding；synthetic data ；international classification of diseases；event knowledge graph；imbalanced label distribution

四、文献内容分析

（一）管理研究中生成式人工智能的机遇与隐忧

现有文献强调，社会科学研究领域的方法论创新需首先思考“为什么采用新方法”的基本问题（Abdurahman et al.，2024）。通过对文献的系统性梳理与理论分析发现，生成式人工智能驱动的管理研究方法论创新呈现出五个方面优势：

一是分析效率上的突破。与传统人工编码或专家研判相比，生成式人工智能可连续不间断运行、自动化处理海量文本语料（李春涛等，2024），既降低了研究者主观判断偏差对结果的影响（Beheshti et al.，2024），又突破性地减少了人工与时间成本。

二是文本语义解析上的优势。生成式人工智能可依托上下文语境实现语义理解（Lyman et al.，2025），精准捕捉词汇在不同语境中的语义变化（Davidson & Karell，2025），进而识别更复杂的语义模式（李春涛等，2024），既能挖掘文本间隐含的深层语义关联和潜在因素特征（Beheshti et al.，2024），又能识别因表述片段化而常被人类编码员忽视的细微语义差异（Davidson & Karell，2025）。

三是研究分析中的客观性。生成式人工智能在专业化场景中展现出与人类专家相媲美的经济理性（Chen et al.，2023），可规避研究者的主观认知偏见。得益于预训练阶段对海量多领域语料的学习（Dillion et al.，2023），其知识储备在多数领域接近人类专家水准（李春涛等，2024）；并且，前沿模型还能模拟人类思考与决策过程（Davidson & Karell，2025），生成高度拟合人类主观判断的结果（李春涛等，2024），同时因非生物特性而避免疲劳、情绪等因素引发的质量波动。

四是可解释性上的优势（de Kok，2025；Kern et al.，2026）。相较于传统预训练模型主要依赖词汇共现的统计规律（Sun et al.，2021），生成式人工智能可建构具逻辑的解释性论述；深度推理模型尤能模拟人类认知的逻辑链条，系统揭示结论生成的内在理据，这种透明化决策既符合科学研究的可验证性原则，也为研究者评估结果效度提供依据。

五是可扩展性优势。研究表明其具备零样本泛化能力，无须额外训练即可应用于新任务（Davidson & Karell，2025）；主流大语言模型本质是融合多领域先验知识的通用智能体，可作为支撑各类研究的基础架构；通过微调与检索增强生成等方法可快速获取特定领域知识（Davidson & Karell，2025；Ren et al.，2025；Zou et al.，2025）；多模态模型更融合图像、音频等多维数据的处理能力（Liu et al.，2025；Zhang et al.，2025），拓展了方法适用边界。

然而，在展现优势的同时，该技术也引发学界对潜在隐忧的关切（Davidson & Karell，2025）。现有文献揭示了其存在伦理争议、数据偏差、隐私泄露及产权风险等方面的问题（de Kok，2025；李春涛等，2024），也有学者指出其可能诱发研究者能力退化与路径依赖等结构性风险（Roberts et al.，2024）。仅从研究方法适切性看，该技术在结果的可靠性与可重复性方面亦面临严峻挑战：就可靠性而言，生成式人工智能基于词元预测的机制本质上是概率驱动的文本续写，而非具备事实认知的推理演绎（de Kok，2025；陈小平，2024），易产生为保持语义连贯而生成虚构内容的幻觉现象（Ren et al.，2025），其输出还可能存在内容偏见（Lyman et al.，2025）、过度自信（Akata et al.，2025）及迎合用户观点（Roberts et al.，2024）等偏差。就可重复性而言，模型容易产生片面且不稳定的回答。相同提示在不同时间或版本间会诱发差异化输出（Chen et al.，2024），模型精度（Feng et al.，2024）、参数量级（Zhou et al.，2024）与温度参数（de Kok，2025）等因素均可能引发结果变异。

总之，生成式人工智能作为新兴研究工具正深刻重塑管理学科的研究方法，其在分析效率、语义识别及洞见生成等方面展现独特优势，但也面临信效度挑战。

（二）管理研究中生成式人工智能的应用

针对生成式人工智能在管理研究中的应用路径，现有研究主要呈现在模型微调、信息提取、定性与定量分析三方面。

（1）模型微调是通用生成式人工智能向领域专家系统转型的核心技术路径，其方法论价值已获国内外学界验证（余池等，2025；Leek & Bischl，2025）。微调能显著提升模型在垂直领域的语义理解与知识推理能力（Leek & Bischl，

2025），其成效已在专利文本挖掘（余池等，2025）、政策文本解析（Xu et al.，2024）、金融决策支持（Huang et al.，2023；Xia et al.，2025）、舆情检测（霍朝光等，2024）等管理学关键场景中得到检验。已有文献以监督式微调为主，核心机理是采用 LoRA 等高效参数优化算法（Hu et al.，2021），通过构建 Alpaca 格式问答对数据集，在冻结基础模型参数的前提下仅优化适配层参数，即可实现领域知识迁移（黄永文等，2025；余池等，2025）。其关键挑战在于高质量标注数据的获取，现有研究主要采用三类方案：一是领域专家标注（Leek & Bischl，2025；Xu et al.，2024），人工构建经过严格验证的训练和测试数据集；二是对现有标准数据集的迁移应用（Xia et al.，2025），但受垂直领域数据稀缺约束，往往需要对源数据集进行适应性调整；三是由生成式人工智能生成用于训练的问答对数据，如让大模型从政策文本（Ren et al.，2025；Zou et al.，2025）或真实案例（Huang et al.，2025）中生成问答数据，反过来训练模型。

（2）尽管生成式人工智能具备创造新内容的能力，但因其生成内容不稳定，主要作为文本信息提取的辅助工具应用于管理研究（Aguinis et al.，2024），可归纳为三方面。一是特征识别，即判定文本中特定因素的存在与否及程度。典型实践是替代传统基于词典法的情感分析（Hur et al.，2024；Abdurahman et al.，2024）；此外，相关研究还将生成式人工智能扩展至其他判别任务，如从生平文本识别心理特质的“有/无”（赵晨等，2025；Rathje et al.，2024）、从制造商描述识别服务类型（Niu et al.，2024）、从资产描述识别多样性收益程度（Ko & Lee，2024）等。二是内容提取，即提炼研究所需关键信息，常用于扎根编码等质性分析（高虎源等，2025），如从社会文本提取事件、人物、时间（Schwitter，2025），从公众反馈识别政治立场及其理由（Fu et al.，2024），或辅助识别个人传记中的成长因素（赵晨等，2025）。多数应用停留在关键信息提炼层面，部分研究则探索复杂语义逻辑提炼，如提取经济学文本中的因果关系（Guo & Yang，2024）或从多模态数据识别故障来源（Zhang et al.，2025）。三是整理归纳，通常与前两者结合，基于语义关联进行聚类或构建更高层次元分类体系，如在识别循环经济的前因机制基础上，对这些因素进行聚类整合（Beheshti et al.，2024），或对半结构化访谈作归纳性主题分析（Paoli，2024）。

此外，部分研究基于生成式人工智能由海量人类语料训练而来，且能在一定程度上反映人类行为判断平均水平的特性（Pellert et al.，2024），尝试将其作为研究对象用于心理学实验与问卷（Dillion et al.，2023；吴胜涛等，2025）。但此做法引发争议（Harding et al.，2024）：有学者指出其现阶段仅能作辅助分析工具，输出既不等同于人类真实认知，也不具备揭示客观规律的科学效度，贸然作为研究对象可能威胁结论可靠性。

（3）在使用生成式人工智能提取信息的基础上，已有研究往往会将所获得的数据应用于后续的定性与定量分析。定性方面，已有研究多结合人工专家校验保障结果质量（高虎源等，2025），在此基础上，从提炼信息中凝练理论框

架（颜赫隆等，2025；赵晨等，2025）。定量方面，一是在提取环节即生成量化数据，如大模型打分或从离散结果生成虚拟变量与有序赋值（Fu et al.，2024；Niu et al.，2024；Xia et al.，2025；Hur et al.，2024）；二是依据特定信息在语料中的分布密度形成量化指标，如考查体现特定因素的段落占比（赵晨等，2025）、特定时段相关文本出现频率（Xu et al.，2024），或借助文本聚合算法（吕成双等，2025）、指标计算方法（陆瑶等，2025）将非结构化数据转化为可分析数据。此外，部分研究还探讨了其在工业故障诊断（Zhang et al.，2025）、任务调度（Yu et al.，2025）、工程质量评估（Pu et al.，2024）等管理流程优化中的应用，为运筹管理研究提供技术支撑。

（三）管理研究中生成式人工智能的方法保障与验证

现有研究不仅系统构建了生成式人工智能在管理学研究中的应用方法，也提供了相应的稳健性保障与验证机制。本研究将其归纳为架构规划、设计规范与评估检验三方面。

首先，在架构规划方面，现有文献主要包含模型选择与微调策略两个维度。①模型选择需系统考量生成式人工智能与特定管理研究问题的多维适配性：一是功能匹配，当前生成式人工智能已出现文本生成、图像视觉、音频识别等功能类型，研究者需依据问题本质界定所需功能，并重点评估模型在中英文语境下的语义理解与生成能力；二是技术经济性，模型性能与部署成本存在权衡（Dettmers et al.，2022），参数量、量化精度的提升虽增强对复杂问题的表征能力，却带来高昂算力成本；三是版本选择，宜采用经学术验证的主流模型以降低技术风险，如 ChatGPT（de Kok，2025）、DeepSeek（刘彦辉等，2025）、Llama（霍朝光等，2024）、Qwen（赵志枭等，2024）、Ernie（赵晨等，2025）。②微调策略宜遵循辩证原则。在知识更新方面需兼顾高质量数据的成本与知识扩展的收益，尽管专业领域微调可构建垂直知识体系（Leek & Bischl，2025），但数据偏差会引发风险，如类别缺失导致认知偏差（Mai et al.，2024）、有害数据污染模型认知（Betley et al.，2025）；同时需警惕性能退化，例如顺序微调可能导致灾难性遗忘（Chen & Garner，2024；Pfeiffer et al.，2021）；还需平衡算力与能力，例如小参数模型微调成本低、更适配 LoRA 等高效算法，但会牺牲大模型的潜在性能。

其次，在设计规范方面，现有文献主要涵盖任务计划、部署与执行三方面。①任务计划需关注模型固有的认知局限，包括自利性倾向（Akata et al.，2025）及处理复杂文本时的注意力损失。为此，建议任务拆解，将整体目标分解为独立子任务；采取合理的文本切片策略（Paoli，2024），控制单次输入量在模型处理容量内（Siano，2025；Wu et al.，2025）；并避免长对话模式，以防历史信息干扰与记忆遗忘。②任务部署环节，可采取多种策略以优化精度：采用检索增强生成整合外部知识库（Ren et al.，2025；Zou et al.，2025）、设置零温度参数（de Kok，2025）、谨慎采用对齐模型（Davidson & Karell，2025），以降低幻觉与偏差风险；亦可嵌入因果知识图谱增强推理可靠性（Tong et al.，2024），或构建递进式框架，初筛用词典法或轻量模型、深度分析切换高性能模型（Niu et al.，

2024），实现资源与精度的动态平衡。③任务执行需合理设计提示工程，采用链式思考提示，以“先解释后评估”的推理路径提升可解释性（de Kok，2025；Kern et al.，2026）；输入上可将重点信息直接置入对话流或智能体设定（赵晨等，2025）、设计结构化提示（de Kok，2025）、提供人类研判示例（Kern et al.，2026）、设置预设选项（de Kok，2025）；输出上需规范呈现形式以便量化（Niu et al.，2024），并以重复提问（Ko & Lee，2024）和投票法（Wu et al.，2025）确保其稳健性。

最后，在评估检验方面，现有研究主要有关联效度与对比信度两类途径。①关联效度通常将提取指标作为预测变量纳入经典结果变量框架，通过影响系数或增量解释力验证效度，如量化财务文本违规风险并证实其对信用违约预测的增量解释力（Xia et al.，2025），或检验文本特征对汇率预测精度的改善（Han et al.，2024）。②对比信度采用多方法三角验证，系统比较模型输出与人类专家标注、其他 AI 模型研判（Paoli，2024；Niszczota & Abbas，2023）及词典法（Xu et al.，2024）、LDA 主题建模（Fu et al.，2024）等主流文本分析方法的一致性。

五、生成式人工智能用于管理研究的流程总结

本部分通过系统性考查生成式人工智能在管理学科研究中的应用现状，旨在对现有管理学及相关领域文献中的零散操作进行归纳、提炼与系统化，力图为管理学界提供一套规范且可参考的流程框架，即 12C－PIPE（管道）矩阵模型。如图 4 所示，该模型由两个维度构成：其一为 PIPE 四阶段研究流程（Preparation 准备阶段、Integration 集成阶段、Processing 处理阶段、Evaluation 评估阶段），系统刻画了生成式人工智能赋能管理研究的完整周期；其二为每个阶段配套实施的“3C”操作策略，由此形成 4×3 的 12C 细分操作矩阵。

首先，准备阶段（Prepare）聚焦于根据研究问题和任务类型确定基础模型及部署策略，包含能力选择（Capability）、参数配置（Configuration）、部署策略（Cloud/Locality）的“3C”策略：①模型能力选择需围绕下游任务，重点考量推理、长文本、多语言及多模态能力（Davidson & Karell，2025；Kong et al.，2024）。如赵晨等（2025）针对中文人物传记文本，选用中文语义理解较强的 Ernie 4.0；Pu 等（2024）针对图文融合的工程管理任务，选用多模态模型。②参数配置需基于算力预算与数据规模，权衡参数量（如 8B、14B、30B）、微调与内存效率及压缩技术，同时需在 4bit、8bit 整数与 FP16 浮点数之间做出选择，遵循“适度够用”原则（de Kok，2025）。如 Kwon 等（2024）用 Llama 3.1 8B 完成文本嵌入等基础任务，在情感判定环节切换至 70B，实现性能与算力的动态平衡。③在部署环节，研究者需依据数据隐私级别和项目预算，决定采用 API 云端部署还是本地部署。云端便利但需关注隐私安全（Kwon et al.，2024），本地门槛较高但在大规模长期项目中更具成本优势。例如，de Kok（2025）针对财报电话会议的识别任务，系统对比了 API 云端调用与本地部署开源模型两种路径，并指出数据可共享且推理规模经济时优先云端，数据敏感

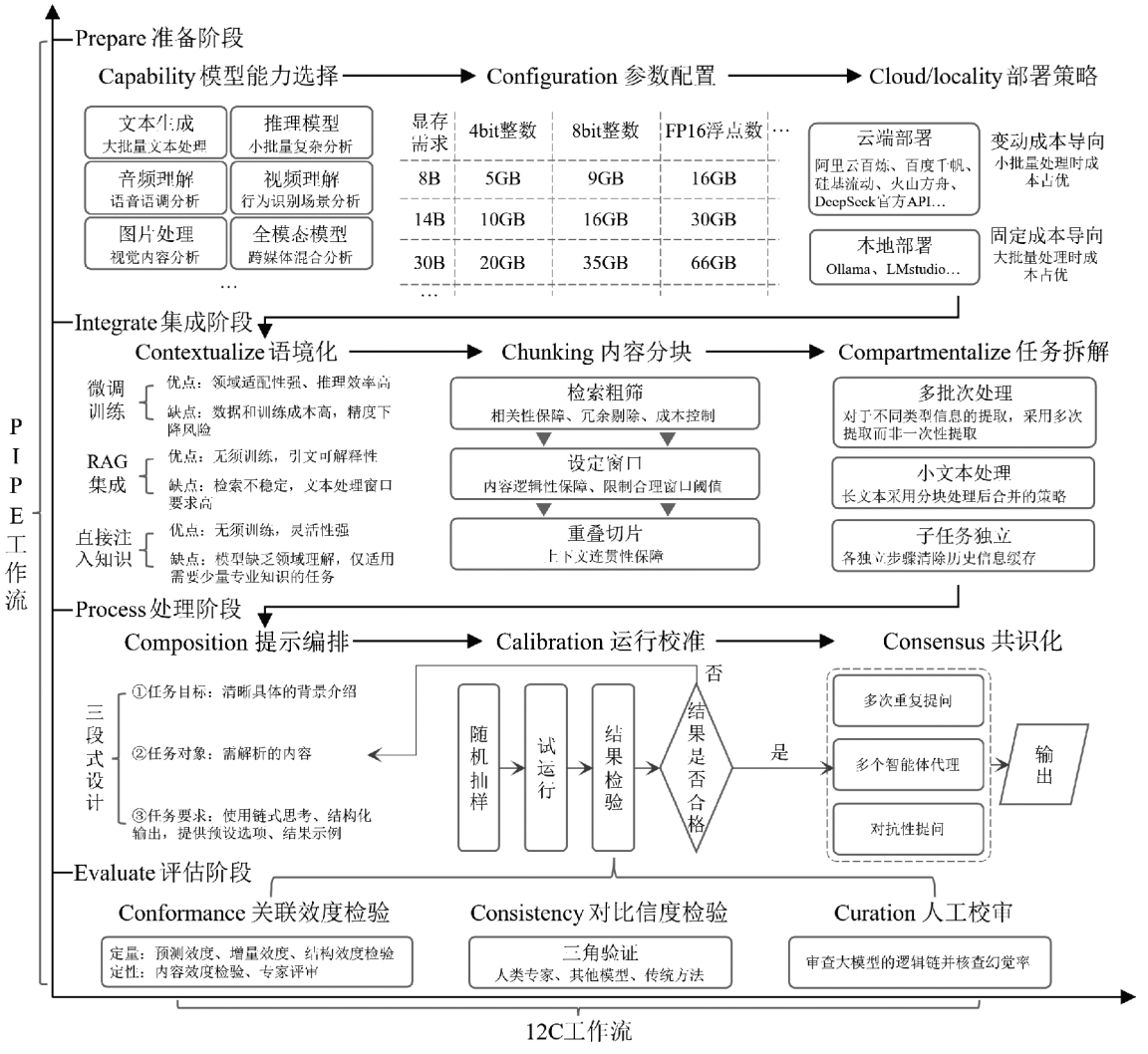

图 4　生成式人工智能驱动管理研究的 12C – PIPE 矩阵

或样本量过大时转向本地。

其次，集成阶段（Integrate）聚焦于将模型与研究情境结合，采取语境化（Contextualize）、内容分块（Chunking）、任务拆解（Compartmentalize）的“3C”策略：①语境化。以领域知识与专业词典为输入，经微调、RAG 增强或提示词优化提升模型的专业能力。如 Leek 和 Bischl（2025）用专家标注语料微调模型，以识别货币政策压力。微调需关注查全率、查准率及 F1 指数等指标（Leek & Bischl，2025；余池等，2025），建议采用 LoRA 等参数高效策略；RAG 架构需检查命中率、忠实度（Ren et al.，2025）及准确率（Zou et al.，2025）；轻量任务可直接在提示词中注入知识。②内容分块。针对年报、访谈稿等长文本，需转化为标准切片并确保语义完整与逻辑连贯，建议基于段落、章节等结构切片并保留适度重叠以防上下文丢失（Paoli，2024）。③任务拆解。通过多批次、小文本量处理，避免模型的注意力分散，关键在于将复杂任务拆为独立子任务（Kong et al.，2024），或者采用多个相互关联的链式提示词处理（de Kok，2025）。如 Paoli（2024）将主题分

析拆为“初始编码—去重—归纳—复核—命名”五个顺序子任务，各阶段独立提示词并以上阶段输出为输入，利用模型无会话记忆的特性形成对话隔离，确保各环节在纯净语境下完成。

再次，处理阶段（Process）旨在执行具体任务，采取提示编排（Composition）、运行校准（Calibration）、共识化（Consensus）的“3C”策略：①提示词编排应遵循“目标—输入—要求”三段式：先精准界定分析任务的目标范畴，通过严谨的操作化定义确保分析任务边界清晰；再规范任务输入对象的格式标准，保障待分析内容的标准化和可比性；最后要求模型运用链式思考并以 JSON 等结构化格式输出（de Kok，2025），并通过预设选项与示例引导高质量回答。②正式运行前需基于小规模样本试运行，通过查全率、查准率及 F1 指数等指标判断参数与提示词设计是否合理并优化（de Kok，2025）。③共识化输出建议多次运行或多智能体协作以保障稳健性，可重复提问 5 次以上（Wu et al.，2025），或用多个主流模型共答并经投票机制整合。如 Hao 和 Xie（2025）构建多模型框架，利用 DeepSeek－V3、GPT－4o、Gemini、Claude、Llama 在相同任务上独立作答，通过比较不同模型输出的差异来分析结果的稳健性。

最后，评估阶段（Evaluate）可采用关联效度检验（Conformance）、对比信度检验（Consistency）、人工校审（Curation）的“3C”策略：①关联效度检验，是指将生成结果与外部变量或专家意见对比。定量研究可借助层次回归、结构方程等检验预测效度、增量效度及结构效度（Xia et al.，2025），如 Xia 等（2025）利用生成式人工智能从借贷文本提取违约概率指标，证实其对违约预测有显著增量解释力；定性分析则可借助专家评审与内容效度检验（Beheshti et al.，2024）。②对比信度检验，是指衡量模型内部一致性及人机一致性，应将结果与人类专家（de Kok，2025）、传统 NLP 方法（Fu et al.，2024）或其他大模型（Niszczota & Abbas，2023）对比，并计算 Cohen’s Kappa（Paoli，2024）。③人工校审针对模型逻辑链及抽样乃至全量数据进行复核，评估幻觉率并修正数据，若错误较多或查全、查准率过低则需重新调整前序步骤（Kong et al.，2024）。

上述各阶段的具体策略、输入输出要素及关键考量因素详见表 2。

表 2　　12C－PIPE 矩阵的关键性操作及考量因素

阶段	12C 策略	输入	产出	关键检查点	关键考量因素
P－准备阶段（Preparation）	Capability	研究问题、任务类型	选定的基础模型	根据特定需求选择合适的模型	下游任务类型；数据类型以及模型推理、长文本处理、多语言支持能力（Davidson & Karell，2025；Kong et al.，2024）
	Configuration	算力预算、数据规模	模型参数设定方案	参数量、成本与任务匹配度	微调效率、内存效率和模型压缩；量化精度（de Kok，2025）
	Cloud/Locality	数据隐私级别、项目预算	部署环境（API/本地）	数据合规性与成本效益比	API 可用性、数据隐私及安全（Kwon et al.，2024）、成本与经济性、技术门槛与便利性（de Kok，2025）

续表

阶段	12C 策略	输入	产出	关键检查点	关键考量因素
I – 集成阶段（Integration）	Contextualize	领域知识、专业词典	由微调、RAG 或提示词增强的模型	领域知识的注入效率	微调：查全率、查准率、F1 指数（Leek & Bischl, 2025；余池等, 2025）； RAG 增强：检索结果的命中率、忠实度（Ren et al., 2025）、准确率（Zou et al., 2025）
	Chunking	长文本数据（年报、访谈稿）	标准化的文本切片	语义完整性；切片间的逻辑连贯性	模型处理长度、为模型回复预留空间、文本块间适度重叠（Paoli, 2024）；基于文本结构切片（Jimeno – Yepes et al., 2024）
	Compartmentalize	复杂分析任务	子任务序列/独立会话	任务间逻辑链条的清晰度	将复杂任务拆解为多个相互关联的链式提示词（de Kok, 2025）或多个独立子任务（Kong et al., 2024）
P – 处理阶段（Process）	Composition	变量定义、任务目标、对象、要求等	结构化提示词	提示词结构是否包含“目标— 输入—要求”	思维链提示、精简 Token 数量、反复测试与调优、提供结果示例、采用 JSON、XML 等结构化输出、提供必要的上下文信息和清晰的任务要求（de Kok, 2025）
	Calibration	小规模样本数据	优化后的参数与提示词	试运行效果并优化参数和提示词	考察小样本测试的查全率、查准率、F1 指数等，以判断当前参数和提示词设计是否合理（de Kok, 2025）
	Consensus	多次运行结果/多智能体输出	最终的整合分析结果	结果的收敛程度；投票机制的有效性	进行 5 次及以上重复提问（Wu et al., 2025）；或采用多个主流模型共同回答以分析结果的异质性和不稳定性（Hao & Xie, 2025）
E – 评估阶段（Evaluation）	Conformance	生成结果、外部变量、专家意见	效度检验报告	预测效度是否显著；内容效度是否达标	定量：预测效度检验达到显著水平（Xia et al., 2025）； 定性：组织具备管理研究经验的专家小组开展内容效度检验（Beheshti et al., 2024）
	Consistency	生成结果、人类专家编码、其他模型结果、传统方法结果等	信度系数报告	模型内部一致性；人机一致性	与人类专家评判（de Kok, 2025）、传统自然语言方法（Fu et al., 2024）、其他大模型评判（Niszczota & Abbas, 2023）的结果进行对比，并计算 Cohen's Kappa 指标（Paoli, 2024）
	Curation	模型逻辑链、结果的抽样或全量数据	经过修正的数据集	人工介入的比例；错误的拦截率	手动纠正错误（de Kok, 2025），若结果存在较多错误或查全率、查准率、F1 指数等较低，需重新生成数据（Kong et al., 2024）

六、总结、挑战与未来方向

（一）研究总结

本研究综合文献计量与内容分析法，全景式考察生成式人工智能驱动的管理研究方法。研究表明，生成式人工智能已从单一研究对象跃升为核心研究工具，在宏观战略、中观决策及微观运营议题中前景广阔。其凭借高效性、精确性、客观性、可解释性和可扩展性五大优

势，提升了研究者对非结构化数据的处理能力，但在结果可靠性与可重复性方面仍存挑战。据此，本研究整合模型微调、提示工程及信效度检验等分散经验，提出 12C – PIPE 矩阵模型，与现有文献形成三方面对话。其一，针对方法论探索碎片化、缺乏系统共识的问题，回应学界对建立规范化应用框架的诉求；其二，针对现有研究多聚焦单一环节、缺乏全周期框架的局限，PIPE 四阶段框架覆盖准备、集成、处理、评估四阶段，并匹配能力选择、语境化、共识化检验等 12 项策略，将分散经验整合为闭环流程；其三，区别于多数研究将技术客体化、作为变量考察其效应的范式，本研究将生成式人工智能定位为重塑研究范式的方法论主体，力图将技术实践抽象为框架，为研究者应对新模型、新任务提供流程指引。

（二）生成式人工智能对管理研究方法论的潜在挑战

现阶段，生成式人工智能在管理研究中主要扮演辅助工具角色，多数研究仍依赖传统方法。然而，随着该技术向研究流程核心环节持续渗透，其对管理研究方法论逻辑的潜在挑战已在测量、变量构建、理论生成、因果识别及知识生产方式五个维度上显现端倪。

第一，测量逻辑重构，却面临理论脱节挑战。传统管理研究遵循假设 – 演绎逻辑，研究者先基于理论推导构念，再通过操作化定义将其转化为可观测指标，最后借助预定义的量表、词典或编码方案实施测量。生成式人工智能正在动摇该逻辑：模型本身成为测量工具，其语义解析直接生成测量结果，不完全依赖研究者预设的理论框架与操作化规则。例如，基于大语言模型的情感分析无须预设情感词典（Hur et al.，2024；Abdurahman et al.，2024），心理特质识别也无须预编码方案（赵晨等，2025；Rathje et al.，2024）。这种从演绎测量向生成测量的转变虽拓展了可测量构念的范围，但也带来双重挑战：一方面弱化了测量与理论间的显性关联，另一方面引入了系统性且不透明的新型测量误差，同一文本在不同模型版本间可能产生差异化结果，构成难以校正的混杂因素。

第二，变量来源方式丰富，却面临效度挑战。传统变量操作化受限于可获取的结构化数据，许多理论构念因缺乏合适的测量手段而难以进入实证模型。生成式人工智能使研究者能够从非结构化文本中直接构建变量（Xia et al.，2025；Niu et al.，2024；赵晨等，2025），极大丰富了变量来源。然而，核心挑战在于构念效度：模型所提取的变量究竟在多大程度上精确对应了目标理论构念，而非同时捕获了文本语境中的混杂信息（Davidson & Karell，2025），目前仍缺乏系统性的检验手段。

第三，理论生成效率提升，却面临理论原创性不足。传统理论建构依赖研究者基于经验素材的归纳推理与深层洞察。生成式人工智能虽可辅助扎根编码、主题归纳乃至假设生成（Tong et al.，2024），但其底层仍是基于概率的词元预测机制（de Kok，2025；陈小平，2024），依据的并非对现象机制的真实理解，而是语料中理论表述的统计分布。管理学理论建构不仅要求发现变量间的关联，更需理解其因果机制与边界条件。当大语言模型从文献中提取变量关系时（Guo & Yang，2024），这种关系本质上是语义复现而非独立的机制识别，可能导致理

论的创新上限。

第四，因果实证路径扩展，却面临技术应用的新型偏差。一是模型输出天然携带其内部的概率性，这种误差进入回归模型后，可能导致估计系数有偏或不一致，且这种非经典测量误差难以通过传统方法校正。二是模型处理文本时可能同时捕获与因变量相关的混杂信息，引入新的内生性来源。三是提示词设计可能导致研究者自由操纵，不同的措辞与输出约束可能系统性地改变变量取值分布（Chen et al.，2024），引发类似于传统研究中分析策略灵活选择所导致的不可重复性问题（Fišar et al.，2023）。

第五，知识生产方式转型，却面临认识论挑战。当生成式人工智能深度参与测量、变量构建和理论生成等核心环节时，管理研究的知识生产方式正从纯粹的人类认知活动向人机协同模式过渡。这种转型引发的根本性挑战在于：由生成式人工智能辅助产出的研究发现，其知识论地位如何界定？有学者担忧过度依赖该技术可能导致批判性思维退化与路径依赖（Roberts et al.，2024），也有学者质疑其输出是否等同于人类认知（Harding et al.，2024；Dillion et al.，2023）。这些争议表明，生成式人工智能对管理研究的挑战已触及“何为有效的学术知识”的深层问题（van Dis et al.，2023）。

（三）面向方法论挑战的未来方向

上述潜在挑战表明，推进生成式人工智能在管理研究中的规范应用，关键在于发展与之配套的方法论保障机制。结合本研究归纳的 12C – PIPE 矩阵，未来研究可从以下五个方向加以应对。

一是建立生成式测量与变量构建的效度检验规范。针对测量误差不透明、构念效度难以保障的问题，未来研究可将 12C – PIPE 矩阵评估阶段的检验环节制度化。对于由模型生成的变量，可参照增量效度的检验逻辑，考察其在纳入成熟变量后是否仍具备独立解释力（Xia et al.，2025；Han et al.，2024），以判断其是否真正对应目标构念。在信度层面，可将人机一致性指标设定为变量可用的合格门槛，例如要求 Cohen’s Kappa 或 F1 指数达到既定水平（Paoli，2024；Niu et al.，2024）。此外，鉴于模型版本差异会引入系统性偏差，研究者可在同一任务上比较不同模型及版本的输出一致性，并将结果稳定性作为变量进入实证模型的前置条件。

二是厘清生成式人工智能在因果识别中的适用边界。生成式人工智能的长处在于模式识别而非因果推断，未来研究需区分二者的适用场景，例如在描述性测量与特征提取环节可充分发挥其语义解析能力，而在涉及因果识别的环节则应保持审慎。当模型生成的指标作为解释变量进入计量模型时，研究者可通过多次独立生成取均值、报告输出分布等方式，抑制随机波动对估计一致性的干扰（Wu et al.，2025；Ko & Lee，2024）。在变量构建阶段，应明确界定提取目标，避免将与因变量相关的无关语义一并纳入，以减少新的内生性来源。针对提示词措辞可能改变估计结果的风险，研究者可在分析前对提示词与判定规则进行预注册，并报告若干等价提示词下的系数分布，以检验结论对提示设计的敏感性（Chen et al.，2024）。

三是确立人机分工的理论建构机制，并以垂直生态提供支撑。鉴于模型的理论生成本质

上是既有语料表述的统计复现，未来研究可遵循“模型承担信息处理、研究者主导机制阐释”的分工原则：由模型完成信息提取与模式发现（Schwitter，2025；Beheshti et al.，2024），由研究者负责因果机制判断与边界条件界定，并通过人工校审对模型的推理逻辑链逐步复核（de Kok，2025；Kong et al.，2024）。为提升模型在专业机制识别上的可靠性，可面向管理学子领域开发垂直模型，整合核心文献与经典案例，构建高质量训练数据（Leek & Bischl，2025；余池等，2025），并尝试将因果知识图谱嵌入分析流程，为推理提供结构化依据（Tong et al.，2024）。

四是构建支撑人机协同知识生产的透明披露机制。针对知识生产方式转型引发的认识论争议，未来研究可通过提升过程透明度来界定人机协同产出的可信度。研究者可完整披露所用模型的版本号、关键参数及提示词编排（Korinek，2023；Abdurahman et al.，2024），使分析过程可追溯、可复现。对于涉及敏感数据或人类受试者的研究，可设置专门的偏见审查环节，防止模型固有偏差扭曲研究结论（Harding et al.，2024）。此外，由于常用语料可能已进入模型预训练数据，研究者在效度检验中可补充使用模型训练截止日之后的新语料作为外部样本，从而剥离记忆效应对结论的影响。

五是建立面向数据安全与产权合规的治理规范。生成式人工智能的应用还伴随数据隐私泄露与产权归属不清等风险（李春涛等，2024），未来研究可将这类伦理考量前置到 12C－PIPE 矩阵的准备阶段，使部署方案与数据敏感程度相匹配。对于涉及企业商业机密或个人隐私的数据，可优先选择本地化部署，或在调用云端接口前完成脱敏处理，以控制数据外泄风险（Kwon et al.，2024）。在语料来源上，研究者应核实训练数据与分析文本的使用权限，避免因版权或授权不清引发合规争议。在此基础上，可在研究流程中固定设置合规审查节点，明确各阶段的数据处理边界与责任归属（de Kok，2025），为基于 12C－PIPE 矩阵的研究提供合规保障。

接受编辑：程聪
收稿时间：2025 年 6 月 23 日
接收时间：2026 年 7 月 7 日

作者简介

赵晨，现任北京邮电大学经济管理学院教授、博士生导师，2012 年在清华大学经济管理学院获得管理学博士学位。他的研究方向为人才管理与人力资源管理，研究成果发表于《管理世界》《南开管理评论》《中国软科学》以及 *Journal of Applied Psychology*、*R&D Management* 等国内外重要学术期刊。

林晨（通讯作者，E－mail：lauterate@ foxmail. com），现任中国社会科学院数量经济与技术经济研究所助理研究员，2026 年在北京邮电大学经济管理学院获得工学博士学位。他的研究方向为人才管理、算法管理及其交叉领域中的相关问题，研究成果发表于《管理世界》《中国软科学》《南开管理评论》等学术期刊。

参考文献

［1］陈小平：《大模型的逻辑增强与人工智能驱动

的行业创新》,《技术经济》, 2024 年第 12 期。

[2] 陈悦、陈超美、刘则渊、胡志刚、王贤文:《CiteSpace 知识图谱的方法论功能》,《科学学研究》, 2015 年第 2 期。

[3] 高虎源、唐啸、付帅泽:《基于生成式大语言模型的社会科学定性分析——研究方法与应用示例》,《社会发展研究》, 2025 年第 1 期。

[4] 黄永文、马玮璐、鲜国建、李娇、罗婷婷、孙坦:《大语言模型驱动的科学数据自动分类研究》,《情报理论与实践》, 2025 年第 6 期。

[5] 霍朝光、尹卓、杨媛、杨万诚、茹润钰、霍帆帆:《基于大模型的政策反讽评论自动识别方法研究》,《情报学报》, 2024 年第 12 期。

[6] 李春涛、闫续文、张学人:《GPT 在文本分析中的应用:一个基于 Stata 的集成命令用法介绍》,《数量经济技术经济研究》, 2024 年第 5 期。

[7] 李鸿磊、王凤彬:《企业裂变创业研究——基于理论成熟度阶梯的述评》,《管理世界》, 2025 年第 5 期。

[8] 刘彦辉、张海涛、周红磊、庞宇飞:《融合大语言模型的情报智库政策内容问答服务研究——以粮食安全政策为例》,《图书与情报》, 2025 年第 1 期。

[9] 陆瑶、施函青、周欣怡:《中国企业数字技术风险暴露对企业价值的影响——来自大语言模型的文本分析证据》,《经济研究》, 2025 年第 2 期。

[10] 吕成双、王彤、孙浩然:《融合情绪指标的股价波动率预测研究 - 基于微调大语言模型与 GAT - TCN 网络》,《运筹与管理》2025 年第 10 期。

[11] 吴胜涛、高承海、胡琬莹、王宁、彭凯平:《集体主义促进亲社会正义感:共同责任的作用》,《心理学报》, 2025 年第 4 期。

[12] 颜赫隆、刘江峰、王子怡、裴雷:《数字经济时代的数据要素流通政策:构成要素、理论框架、实践路径》,《信息资源管理学报》, 2025 年第 3 期。

[13] 余池、陈亮、许海云、牟琳、夏春姊、贤信:《基于大语言模型的专利命名实体识别方法研究》,《数据分析与知识发现》, 2025 年第 6 期。

[14] 张志豪、刘思航、马天骥、安嘉骏、何喜军:《大模型驱动下基于企业画像的专利交易个性化推荐研究》,《情报理论与实践》, 2025 年第 5 期。

[15] 赵晨、林晨、王宏飞、杜鹏、李建新:《企业科技领军人才的多重构型及成才路径:基于大语言模型(LLMs)的质性分析》,《中国软科学》, 2025 年第 2 期。

[16] 赵志泉、胡蝶、刘畅、沈思、王东波:《人文社科领域中文通用大模型性能评测》,《图书情报工作》, 2024 年第 13 期。

[17] Abdurahman, S., Atari, M., Karimi - Malekabadi, F., Xue, M. J., Trager, J., Park, P. S., Golazizian, P., Omrani, A., & Dehghani, M. 2024. Perils and opportunities in using large language models in psychological research. *PNAS Nexus*, 3 (7): pgae245.

[18] Aguinis, H., Beltran, J. R., & Cope, A. 2024. How to use generative AI as a human resource management assistant. *Organizational Dynamics*, 53 (1): 101029.

[19] Akata, E., Schulz, L., Coda - Forno, J., Oh, S. J., Bethge, M., & Schulz, E. 2025. Playing repeated games with large language models. *Nature Human Behaviour*, 9 (7): 1380 - 1390.

[20] Beckmann, L., & Hark, P. F. 2024. ChatGPT and the banking business: Insights from the US stock market on potential implications for banks. *Finance Research Letters*, 63: 105237.

[21] Beheshti, M., Mahdiraji, A. H., & Rocha - Lona, L. 2024. Transitioning drivers from linear to circular economic models: Evidence of entrepreneurship in emerging nations. *Management Decision*, 62 (9): 2714 - 2736.

[22] Betley, J., Tan, D. C. H., Warncke, N., Sztyber - Betley, A., Bao, X., Soto, M. I. N., Labenz, N., & Evans, O. 2025, Emergent misalignment: Narrow

finetuning can produce broadly misaligned LLMs：*Proceedings of the ICML* 2025 *Conference*, Vancouver.

[23] Chen, C., 2006, CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, 57 (3): 359 – 377.

[24] Chen, H., & Garner, P. N. 2024. Bayesian parameter – efficient fine – tuning for overcoming catastrophic forgetting. *IEEE/ACM Transactions On Audio*, *Speech*, *and Language Processing*, 32: 4253 – 4262.

[25] Chen, L., Zaharia, M., & Zou, J. 2024. How is ChatGPT's behavior changing over time? *Harvard Data Science Review*, 6 (2): 1 – 47.

[26] Chen, Y., Liu, T. X., Shan, Y., & Zhong, S. 2023. The emergence of economic rationality of GPT. *PNAS*, 120 (51): e2316205120.

[27] Cheng, F., Li, H., Liu, F., van Rooij, R., Zhang, K., & Lin, Z. 2025, Empowering LLMs with logical reasoning: A comprehensive survey: *Proceedings of the Thirty – Fourth International Joint Conference on Artificial Intelligence*, Montreal.

[28] Davidson, T., & Karell, D. 2025. Integrating generative artificial intelligence into social science research: Measurement, prompting, and simulation. *Sociological Methods & Research*, 54 (3): 775 – 793.

[29] de Kok, T. 2025. ChatGPT for textual analysis? How to use generative LLMs in accounting research. *Management Science*, 71 (9): 7888 – 7906.

[30] Dettmers, T., Lewis, M., Belkada, Y., & Zettlemoyer, L. 2022, LLM. int8: 8 – bit matrix multiplication for transformers at scale: *Proceedings of the Advances in Neural Information Processing Systems* 35 (*NeurIPS* 2022), Louisiana.

[31] Dillion, D., Tandon, N., Gu, Y., & Gray, K. 2023. Can AI language models replace human participants? *Trends in Cognitive Sciences*, 27 (7): 597 – 600.

[32] Eulerich, M., Sanatizadeh, A., Vakilzadeh, H., & Wood, D. A. 2024. Is it all hype? ChatGPT's performance and disruptive potential in the accounting and auditing industries. *Review of Accounting Studies*, 29 (3): 2318 – 2349.

[33] Fan, S., Wu, Y., & Yang, R. 2025. Measuring firm – level supply chain risk using a generative large language model. *Finance Research Letters*, 77: 107111.

[34] Feng, G., Yang, K., Gu, Y., Ai, X., Luo, S., Sun, J., He, D., Li, Z., & Wang, L. 2024, How numerical precision affects arithmetical reasoning capabilities of LLMs: *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, California.

[35] Ferreira, A., & Otley, D. 2009, The design and use of performance management systems: An extended framework for analysis. *Management Accounting Research*, 20 (4): 263 – 282.

[36] Fišar, M., Greiner, B., Huber, C., Katok, E., & Ozkes, A. I. 2023. Reproducibility in management science. *Management Science*, 70 (3): 1343 – 1356.

[37] Fu, X., Sanchez, T., Li, C., & Junqueira, J. 2024. Deciphering public voices in the digital era benchmarking ChatGPT for analyzing citizen feedback in Hamilton, New Zealand. *Journal of the American Planning Association*, 90 (4): 728 – 741.

[38] Gezdur, A., & Bhattacharjya, J. 2025. Innovators and transformers: Enhancing supply chain employee training with an innovative application of a large language model. *International Journal of Physical Distribution & Logistics Management*, 55 (4): 394 – 408.

[39] Guo, Y., & Yang, Y. 2024, EconNLI: Evaluating large language models on economics reasoning: *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Bangkok.

[40] Hao, Y. & Xie, D., 2025, A multi – LLM – agent – based framework for economic and public policy analysis, *arXiv*: 2502. 16879.

[41] Han, D., Guo, W., Chen, H., Wang, B., & Guo, Z. 2024. Lest: Large language models and spatio – temporal data analysis for enhanced Sino – US exchange rate forecasting. *International Review of Economics & Finance*, 96 (A): 103508.

[42] Harding, J., D Alessandro, W., Laskowski, N. G., & Long, R. 2024. AI language models cannot replace human research participants. *AI & Society*, 39 (5): 2603 – 2605.

[43] Hu, E. J., Shen, Y., Wallis, P., Allen – Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. 2021, LoRA: Low – rank adaptation of large language models. *arXiv*: 2106. 09685.

[44] Huang, A. H., Wang, H., & Yang, Y. 2023. FinBERT: A large language model for extracting information from financial text. *Contemporary Accounting Research*, 40 (2): 806 – 841.

[45] Huang, C., Tang, Z., Hu, S., Jiang, R., Zheng, X., Ge, D., Wang, B., & Wang, Z. 2025. ORLM: A customizable framework in training large models for automated optimization modeling. *Operations Research*, 73 (6): 2986 – 3009.

[46] Hui, X., Reshef, O., & Zhou, L. 2024. The short – term effects of generative artificial intelligence on employment: evidence from an online labor market. *Organization Science*, 35 (6): 1977 – 1989.

[47] Hur, J. K., Heffner, J., Feng, G. W., Joormann, J., & Rutledge, R. B. 2024. Language sentiment predicts changes in depressive symptoms. *Proceedings of the National Academy of Sciences*, 121 (39): e2321321121.

[48] Jha, M., Qian, J., Weber, M., & Yang, B. 2024. ChatGPT and corporate policies. *NBER Working Paper*.

[49] Jimeno – Yepes, A., You, Y., Milczek, J., Laverde, S., & Li, R. 2024. Financial report chunking for effective retrieval augmented generation. *arXiv*: 2402. 05131

[50] Karakaş, N., & Jaeger, B. 2025. Changes in attitudes toward meat consumption after chatting with a large language model. *Social Influence*, 20 (1): 2475802.

[51] Kern, F. B., Wu, C., & Chao, Z. C. 2026. Assessing novelty, feasibility and value of creative ideas with an unsupervised approach using GPT – 4. *British Journal of Psychology*, 117: 741 – 760.

[52] Ko, H., & Lee, J. 2024. Can ChatGPT improve investment decisions? From a portfolio management perspective. *Finance Research Letters*, 64: 105433.

[53] Kong, Y., Nie, Y., Dong, X., Mulvey, J. M., Poor, H. V., Wen, Q., & Zohren, S. 2024. Large language models for financial and investment management: applications and benchmarks. *Journal of Portfolio Management*, 51 (2): 162 – 210.

[54] Korinek, A. 2023. Generative AI for economic research: Use cases and implications for economists. *Journal of Economic Literature*, 61 (4): 1281 – 1317.

[55] Kwon, B., Park, T., Perez – Cruz, F. & Rungcharoenkitkul, P., 2024, Large language models: a primer for economists. *BIS Quarterly Review*: 10 December 2024.

[56] Law, J., Bauin, S., Courtial, J. P., & Whittaker, J., 1988, Policy and the mapping of scientific change: A co – word analysis of research into environmental acidification. *Scientometrics*, 14 (3): 251 – 264.

[57] Lee, S., & Park, G. 2023. Exploring the impact of ChatGPT literacy on user satisfaction: The mediating role of user motivations. *Cyberpsychology, Behavior and Social Networking*, 26 (12): 913 – 918.

[58] Leek, L., & Bischl, S. 2025. How central

bank independence shapes monetary policy communication: A large language model application. *European Journal of Political Economy*, 87: 102668.

[59] Li, Y., Liu, Y., & Yu, M. 2025. Consumer segmentation with large language models. *Journal of Retailing and Consumer Services*, 82: 104078.

[60] Liu, Y., Bu, N., Li, Z., Zhang, Y., & Zhao, Z. 2025. AT-FinGPT: Financial risk prediction via an audio-text large language model. *Finance Research Letters*, 77: 106967.

[61] Lyman, A., Hepner, B., Argyle, L., Busby, E., Gubler, J., & Wingate, D. 2025. Balancing large language model alignment and algorithmic fidelity in social science research. *Sociological Methods & Research*, 54 (3): 1110-1155.

[62] Mai, Z., Chowdhury, A., Zhang, P., Tu, C., Chen, H., Pahuja, V., Berger-Wolf, T., Gao, S., Stewart, C., Su, Y., & Chao, W. 2024. Fine-tuning is fine, if calibrated. *arXiv*: 2409.16223.

[63] Naeem, M., Smith, T., & Thomas, L. 2025. Thematic analysis and artificial intelligence: A step-by-step process for using ChatGPT in thematic analysis. *International Journal of Qualitative Methods*, 24 (7): 1-18.

[64] Naskar, S. T. & Merigó Lindahl, J. M., 2025, Forty years of the theory of planned behavior: A bibliometric analysis (1985-2024). *Management Review Quarterly*: Online ahead of print.

[65] Niszczota, P., & Abbas, S. 2023. GPT has become financially literate: insights from financial literacy tests of GPT and a preliminary test of how people use it as a source of advice. *Finance Research Letters*, 58 (A): 104333.

[66] Niu, Y., Wu, J., Jiang, S., & Jiang, Z. 2024. The bullwhip effect in servitized manufacturers. *Management Science*, 71 (1): 1-20.

[67] Paoli, S. D. 2024. Performing an inductive thematic analysis of semi-structured interviews with a large language model: an exploration and provocation on the limits of the approach. *Social Science Computer Review*, 42 (4): 997-1019.

[68] Pellert, M., Lechner, C., Wagner, C., Rammstedt, B., & Strohmaier, M. 2024. AI psychometrics: assessing the psychological profiles of large language models through psychometric inventories. *Perspectives On Psychological Science*, 19 (5): 808-826.

[69] Pfeiffer, J., Kamath, A., Rücklé, A., Cho, K., & Gurevych, I. 2021, Adapterfusion: Non-destructive task composition for transfer learning: *Proceedings of the* 16*th Conference of the European Chapter of the Association for Computational Linguistics*, Association for Computational Linguistics, Online.

[70] Pu, H., Yang, X., Li, J., & Guo, R. 2024. Autorepo: a general framework for multimodal LLM-based automated construction reporting. *Expert Systems with Applications*, 255: 124601.

[71] Rathje, S., Mirea, D., Sucholutsky, I., Marjieh, R., Robertson, C., & Van Bavel, J. J. 2024. GPT is an effective tool for multilingual psychological text analysis. *PNAS*, 121 (34): e2308950121.

[72] Ren, R., Ma, J., & Zheng, Z. 2025. Large language model for interpreting research policy using adaptive two-stage retrieval augmented fine-tuning method. *Expert Systems with Applications*, 278 (10): 127330.

[73] Roberts, J., Baker, M., & Andrew, J. 2024. Artificial intelligence and qualitative research: The promise and perils of large language model (LLM) "assistance". *Critical Perspectives On Accounting*, 99: 102722.

[74] Saxena, A., & Rishi, B. 2025. Designing an artificial intelligence-enabled large language model for financial decisions. *Management Decision*, 63 (12): 4135-

4153.

[75] Schwitter, N. 2025. Using large language models for preprocessing and information extraction from unstructured text: A proof – of – concept application in the social sciences. *Methodological Innovations*, 18 (33): 61 – 65.

[76] Siano, F. 2025. The news in earnings announcement disclosures: Capturing word context using LLM methods. *Management Science*, 71 (11): 9831 – 9855.

[77] Smirnov, E. 2025. Enhancing qualitative research in psychology with large language models: A methodological exploration and examples of simulations. *Qualitative Research in Psychology*, 22 (2), 482 – 512.

[78] Sun, Y., Wang, S., Feng, S., Ding, S., & Pang, C. 2021, Ernie 3.0: Large – scale knowledge enhanced pre – training for language understanding and generation. *arXiv*: 2107.02137.

[79] Tong, S., Mao, K., Huang, Z., Zhao, Y., & Peng, K. 2024. Automating psychological hypothesis generation with AI: When large language models meet causal graph. *Humanities and Social Sciences Communications*, 11 (1): 896.

[80] van Dis, E. A. M., Bollen, J., van Rooij, R., Zuidema, W., & Bockting, C. L. 2023. ChatGPT: Five priorities for research. *Nature*, 614 (7947): 224 – 226.

[81] Wang, A., Morgenstern, J., & Dickerson, J. P. 2025. Large language models that replace human participants can harmfully misportray and flatten identity groups. *Nature Machine Intelligence*, 7 (3): 400 – 411.

[82] Wu, S., Ma, X., Luo, D., Li, L., Shi, X., Chang, X., Lin, X., Luo, R., Pei, C., Du, C., Zhao, Z., & Gong, J. 2025. Automated literature research and review – generation method based on large language models. *National Science Review*, 12 (6): nwaf169.

[83] Wu, Y., Wang, J., & Yang, P. 2024. How supply chain digitalization investment affects firm's financial and non – financial performance: Evidence from listed companies in China. *International Review of Financial Analysis*, 96 (A): 103639.

[84] Xia, Y., Shi, Z., Du, X., & Zheng, Q. 2025. Extracting narrative data via large language models for loan default prediction: When talk isn't cheap. *Applied Economics Letters*, 32 (4): 481 – 486.

[85] Xu, Z., Li, J., Hua, X., & Ren, P. 2024. Is the tone of the government – controlled media valuable for capital market? Evidence from China's new energy industry. *Energy Policy*, 184: 113917.

[86] Xu, Z., Song, T., & Lee, Y. 2025. Confronting verbalized uncertainty: Understanding how LLM's verbalized uncertainty influences users in AI – assisted decision – making. *International Journal of Human – Computer Studies*, 197: 103455.

[87] Yamamoto, Y. 2024. Suggestive answers strategy in human – chatbot interaction: A route to engaged critical decision making. *Frontiers in Psychology*, 15: 1382234.

[88] Ye, H., Jin, J., Xie, Y., Zhang, X., & Song, G. 2025. Large language model psychometrics: A systematic review of evaluation, validation, and enhancement. *arXiv*: 2505.08245.

[89] Yu, W., Lv, J., Zhuang, W., Pan, X., Wen, S., Bao, J., & Li, X. 2025. Rescheduling human – robot collaboration tasks under dynamic disassembly scenarios: An mLLM – kg collaboratively enabled approach. *Journal of Manufacturing Systems*, 80: 20 – 37.

[90] Zhang, Q., Xu, C., Li, J., Sun, Y., Bao, J., & Zhang, D. 2025. LLM – TSFD: An industrial time series human – in – the – loop fault diagnosis method based on a large language model. *Expert Systems with Applications*, 264 (10): 125861.

[91] Zhou, L., Schellaert, W., Martínez – Plumed,

F., Moros – Daval, Y., Ferri, C., & Hernández – Orallo, J. 2024. Larger and more instructable language models become less reliable. *Nature*, 634 (8032): 61 – 68.

[92] Zhuang, J. & Sun, H., 2025, Systematic bibliometric analysis of entrepreneurial intention and behavior research. *Administrative Sciences*, 15 (8): 290.

[93] Zou, Y., Shi, M., Chen, Z., Deng, Z., Lei, Z., Zeng, Z., Yang, S., Tong, H., Xiao, L., & Zhou, W. 2025. ESGreveal: An LLM – based approach for extracting structured data from ESG reports. *Journal of Cleaner Production*, 489 (15): 144572.