DOI:10.16339/j.cnki.hdxbskb.2025.03.008

生成式AI如何变革经济学文本分析范式*

肖柏高[1],周艺珮[2],谢锐[1]
(1. 湖南大学 经济与贸易学院,湖南 长沙 410079;2. 长沙理工大学 经济与管理学院,湖南 长沙 410076)

[摘 要] 生成式AI强大的语言能力和使用便捷性极大地影响了文本数据的使用范围、方式、角度和深度,有效地解决了已有文本分析方法中存在的准确率低、难度高、成本大和文本数据利用水平不足的问题,进而带动研究广度和深度的提升。在生成式AI时代,文本分析还可能存在以下三个潜在趋势:生成式AI成为处理各类文本任务的标准工具、生成式AI提供智能辅助以及生成式AI"端对端"成为研究新过程。提示语是生成式AI高质量完成文本分析的关键,可以通过提供示例、应用思维链和反复测试等途径来优化提示语。使用生成式AI还要注意避免"幻觉"和保护数据私密性。

[关键词] 生成式AI;文本数据;ChatGPT
[中图分类号] F49 [文献标识码] A [文章编号] 1008-1763(2025)03-0063-09

How will Generative AI Revolutionize Text Analyses Paradigm in Economics

XIAO Baigao[1], ZHOU Yipei[2], XIE Rui[1]
(1. School of Economics & Trade, Hunan University, Changsha 410079, China;
2. School of Economics & Management, Changsha University of Science and Technology, Changsha 410076, China)

Abstract: The powerful language capabilities and ease of use of generative AI have significantly influenced the scope, methods, perspectives, and depth of textual data utilization. It effectively addresses issues in existing text analysis methods, such as low accuracy, high complexity, substantial costs, and insufficient utilization of textual data, thereby enhancing the breadth and depth of research. In the era of generative AI, text analyses may exhibit three potential trends: generative AI becoming a standard tool for handling diverse text-based tasks, generative AI providing intelligent assistance, the "end-to-end" of generative AI becoming a new research process. Prompts are critical for generative AI to achieve high-quality text analyses, optimization can be achieved through providing examples, applying chain-of-thought, and iterative testing. Additionally, when using generative AI, special attention must be paid to avoiding "hallucinations" and protecting data privacy.

Key words: generative AI; text data; ChatGPT

* [收稿日期] 2024-09-30
[基金项目] 教育部人文社会科学研究青年基金项目:智能算力封锁与AI行业发展:机制分析、效应评估和转型路径研究(24YJC790195);教育部哲学社会科学研究重大课题攻关项目:疫情常态化下我国制造业全球供应链重构和数字化转型研究(21JZD026)
[作者简介] 肖柏高(1996—),男,湖南娄底人,湖南大学经济与贸易学院助理教授,博士,研究方向:技术创新。

一 引 言

文本是人类记录和交流信息的核心载体之一，很多经济、管理等领域的关键信息也记录在文本中，例如党中央和国务院的各类文件能体现中国经济走向，央行公告中的细微措辞差异代表了货币政策的可能转变，政府工作报告包含了当年的经济工作计划等信息，财经帖子反映了广大中小投资者对股价的看法，上市企业年报披露了详细的企业经营和发展动态……因此经济学研究将会不可避免地、越来越广泛地使用文本数据。统计数据也印证了这一趋势，自2016年以来，国内部分权威经济管理学期刊中使用文本数据的文章迅速增长(图1)。随着文本数据的受重视程度增加，文本来源范围广泛，文本开发程度加深和文本分析工具的多样化，可以预见文本数据相关文献的增长趋势还将继续。

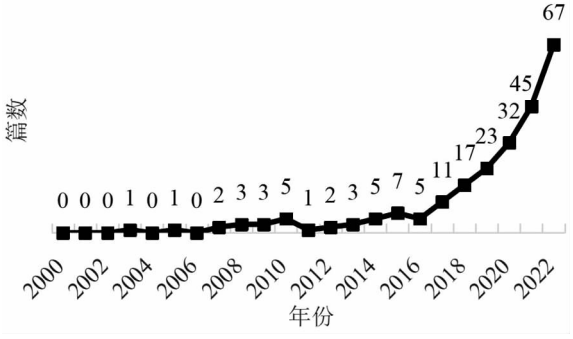

图1 部分中文经济管理学期刊文本数据相关文献篇数

期刊统计范围为《经济研究》《世界经济》《中国工业经济》《经济学(季刊)》《管理世界》，统计标准为文章的标题、摘要或关键词中包含“文本”或者是正文中包含“文本数据”或“文本分析”。

然而，文本数据具有结构复杂、信息含量高、来源广、语调语义多样等特征[1-2]，目前的分析方法难以兼顾准确性和高效性，对文本的信息挖掘和利用程度也不够深。例如，人工阅读法存在成本高昂、可重复性差的不足，词频法非常依赖分词质量，而且无视了文本叙事结构、语法语义和其他文本信息，统计和机器学习则存在数据标注工作量大、使用门槛较高、模型选择标准不一等问题。此外，当前文本研究中也缺乏“数据驱动”，即研究主题和思路等很少来自对文本数据的启发和追问，而是先由作者基于经济学知识提出再经文本数据验证，反映出研究者对原始文本中丰富有效信息的利用不足。这一现象可能是因为研究者对原始文本的不够重视，但更可能是因为缺乏高效有力的分析工具，使研究者在面对海量的文本数据而显得“有心无力”。

生成式AI强大的语言能力和使用便捷性解决了准确性和高效性难以兼顾的问题，特别是其“生成”功能可以完成已有方法所不能完成的任务，因此生成式AI一面世，就被广泛运用。生成式AI是利用大算力对大数据进行训练后得到的可以依据提示语(prompt)生成新内容(当前主要是文本)的超大模型，它在人类语言的理解和运用方面取得了革命性突破，在包含经济学研究等在内的各行各业得到广泛使用。如Garg和Fetzer利用生成式AI提取并分析了4万余篇经济学论文中的因果推断信息[3]，Choi和Kim发现GPT-4可以识别企业年报中税务审计表述的细微用词变化及含义[4]。Bai等证实了生成式AI可以发现企业高管发言中的新信息，进而帮助投资者挖掘隐藏信息[5]。Kim等测试了生成式AI基于财报预测企业盈余的准确性，发现生成式AI表现出比分析师更好的水准[6]。Lopez-Lira和Tang的实验表明生成式AI可以通过阅读新闻标题来预测股价[7]。Hansen和Kazinnik指出生成式AI能阐释美联储报告中充满技术性术语的内容[8]。此外，还有研究探讨了生成式AI在非文本领域的应用潜力，如测试了其面临实验测试或历史事件时是否表现出跟人类相似的反应[9-10]。

当前，生成式AI拉开了新一轮AI科技革命的序幕，经济学中的文本研究范式也迎来了重大变革。虽然不少研究已经把生成式AI用于文本分析，但还有以下三个关键共性问题亟待解决。第一，用生成式AI完成各类语言任务已成为AI界的共识和趋势，那么生成式AI是否也会取代分词等成为经济学中的标准文本分析工具？第二，由于文本量巨大、缺乏合适工具等因素，当前经济学研究对文本数据的分析和利用远比不上历史等人文社会学科，处于较为粗浅的阶段，可能导致研究主题不够丰富、研究结论不够严谨等结果，那么生成式AI能否解决这一问题？第三，不断进化的生成式AI是否将主导经济学文本研究，即变革经济学文本研究范式？当前生成式AI的应用仍然是局限于文本处理环节，生成式AI在研究主题构建、实证因果推断、机制分析等跟原始文本息息相关的环节尚未得到使用，但是随着生成式AI能力的不断提高，它有无可能渗透各个环节甚至是主导整体研究？这些问题对经济学文本研究质量和主题多样性的提升有重要意义。

目前已经有文献探讨了生成式AI对经济学等学科的影响，它们虽然与本文有一定关联，但由于研究侧重点不同，尚未全面深入回答上述问题。一些研究将生成式AI视为一个能处理人类经济社会的

系统分析工具，深入分析了生成式 AI 破解经济学中的“维数灾难”问题、弥合宏微观之间的鸿沟、影响理性人假设等[11-12]。也有学者讨论了生成式 AI 在经济研究中的六个用途：帮助构思、协助写作、背景研究、分析数据、撰写代码和数学推导[13]。李春涛等撰写了一个可以在 Stata 中应用生成式 AI 的命令，进一步方便经济学研究者使用生成式 AI，并介绍了其使用方法以及相较于传统方法的优越性[14]。政光景和吕鹏从哲学社会科学的一般视角研究了生成式 AI 的影响，主要包含生成式 AI 作为新突破将提供新的研究工具、数据、研究主题和方向等[15]。Birhane 等特别指出了生成式 AI 应用于科学研究时潜在的负面影响，例如生成式 AI 不具备科学推理能力、存在“幻觉”等，所以研究者要更加负责任地使用生成式 AI[16]。

生成式 AI 作为通用型的人工智能技术，其影响是广泛的，但目前生成式 AI 主要是处理语言的大模型，其对于文本数据分析的影响是最直接的，尤其是在文本数据的使用越来越普及，而文本分析工具显著落后于 AI 技术前沿、文本数据的利用和研究不够深入的多重背景下，我们有必要专门探讨生成式 AI 对经济学文本研究的广泛影响。首先，本文分析和总结了已有文本分析方法的局限性，并说明如何用生成式 AI 来解决局限性问题。接着，本文探索了生成式 AI 时代下经济学文本分析的三个重要问题，即生成式 AI 是否会成为文本分析的标准工具，文本分析质量和深度不足的问题能否通过生成式 AI 得到有效解决，以及生成式 AI 如何更加深入地应用于数据分析之外的其他研究环节，乃至主导研究。然后，本文还讨论了研究者与生成式 AI 的沟通桥梁“提示语”(prompt)。最后，本文对生成式 AI 用于研究时的一些潜在问题进行了讨论，并尝试给出了建议。

二　现有文本分析方法及文本数据利用的局限性

(一)现有文本分析方法的局限性

1. 人工阅读的局限

人工阅读和分析文本是最基础的处理文本数据的方法，即便这一方法后来随着技术发展和数据增长逐渐不再使用，但是人工阅读和总结文本特征仍然是用计算机处理文本的重要先行步骤。人工阅读的最大优点无疑是准确率相对较高，因为文本本来就是人类交流的工具，但是缺点也很明显，尤其是当文本数据量较大时会出现以下问题：第一，人工阅读的时间成本极其高昂。第二，由于主观因素的存在，不同的研究者对于同一个文本任务可能有不同的看法和结果，即便是同一个研究者处理的标准也可能前后不一致，这可能会导致结果无法复现，影响研究结论的稳健性和可靠性。解决上述问题一般是采用增加人数、多人交叉验证的方法，但这并不能从根本上解决时间和金钱成本高的难题。因此，后续的研究基本上都是采取计算机来处理大量的文本数据。

2. 字词分析的不足

字或者词往往是语言中具有独立语义的基本组成单位，句子、段落、文章以及它们传达的信息都是由一个个字词累积而成，通过对字词进行分析可以在一定程度上推断其所属的句子、段落和文章的语义或情感，这就是基于字词的文本分析方法的基本逻辑。这类方法的优点在于相对简单和便捷，在文本数据可得性不高、深度学习尚未流行、计算机算力有限的背景下，基于字词的方法一度成为经济、金融和管理学等研究中文本分析的主流。但是基于字词的方法也存在分词误差、语义丢失、忽略文本叙事结构等问题，导致结果出现难以预测的偏差，下面对这三个问题进行说明。

中文不像英文有空格作为天然的词语分割标志，因此基于字词进行文本分析时首先需要对中文句子进行分词。目前文献中主流的分词算法都存在一定的误差，例如经管学界目前最常用的分词工具“结巴”(jieba)[17-19]，其分词的总体准确率为 81% 左右，这些误差对后续的研究会产生一定影响。语义丢失是基于字词方法的另一问题，字词组合起来后，各自词义会受其他字词的影响，也会形成更为复杂的整体语义。句子一旦被切分，则词义影响关系和整体语义就会丢失。例如多义词的具体词义取决于上下文，刘青和肖柏高指出“人工”在“节约人工”中是表示“人力”(名词)，在“人工关节”中是表明“人类制造的”(形容词)[20]。无视整体语义导致偏差的情况更为严重，在计算产品相似度时，“制作芯片的沙子”和“沙子制作的芯片”中的字词完全相同，已有研究通过计算两段文本中相同字词的比例来衡量产品相似度[21]，则误将这两个产品视为完全相同。难以反映文本叙事结构的复杂性是另一不足。不同作者阐述事情有不同的逻辑、价值取向、风格乃至偏见等，作者也会采取叙述、议论、修辞等多种表现手法[22-23]。叙事结构蕴含了作者的态度、事件的发展脉络、政策导向、实际成效等，而基于字词的方法经常难以捕捉到隐藏在文字背后的真实含义，通常需要结合文中独特的表达方式才能实现对文本的正确识别。

3.机器学习的缺点

随着机器学习的兴起,自然语言处理领域也开始应用机器学习来分析文本。机器学习相对于字词分析方法的优点在于,通过机器学习和一定样本量的人工标注数据可以挖掘一些文本信息,其对文本的理解相对会更好,在文本分析任务中的准确率也会更高。其缺点也非常明显,主要有难度大、成本高和应用单一等,这阻碍了经管学界基于机器学习来分析文本。自然语言处理历来被视为AI中最为困难的研究领域之一,非专业学者要踏入这一领域更是困难重重。即便目前针对典型的文本任务已经有了一些较为成熟的模型和方法,但是如果要将其应用于经管课题中,要求研究者掌握一定的AI知识和语言知识。机器学习还需要人工标注数据,要使得训练后的模型能较好地理解语言和完成任务,人工标注的数据通常是万级起步,多的甚至高达百万级,正所谓"有多少人工,就有多少智能"。巨大的工作量意味着高昂的人工成本支出,而且在工作量如此大的情况下,如何保证人工标注数据的一致性也是一个难题。除人工成本外,硬件成本也难以负担,复杂的模型和海量的训练样本还要求高端的GPU等硬件设备,例如参数为亿级的BERT语言模型的训练成本为2.5万美元以上[24]。机器学习模型还面临应用单一的问题,譬如用于文本分类的模型无法用于分词、阅读理解等其他任务,甚至可能出现用于新闻主题分类的模型无法用于新闻情感分类的情况。这意味着即便每一项研究课题面临的文本任务差别很小,但可能需要各自构造和训练相应的语言模型,导致资源的极大浪费。

(二)现有文本数据利用的局限性

尽管越来越多的文本数据得到利用,文本分析工具也不断改进,但是经济学对于文本资料的利用一直停留在构造实证指标层面,使用程度远远不如历史等学科。当前公文、书籍、年报、新闻、评论、专利、网页等文本已经在经济学研究中得到了广泛使用,分析工具也从传统的人工到字词再进化为统计和机器学习,生成式AI的出现将进一步丰富分析工具,文本数据和分析工具对研究的限制已经大大减轻。然而,目前对文本数据的利用思路和程度仍然处于抽取特定信息构建实证指标的阶段,对于文本中其他跟研究主题、机制等息息相关的信息关注不足。例如,Baker等从大量新闻中构建了一个衡量经济政策不确定性的指标,由此引发了大量的跟经济政策不确定性相关的实证论文的发表。然而这些研究均没有去探讨经济政策不确定性在新闻中出现的背景和原因,它跟哪些主题或现象相关,新闻工作者撰写经济政策不确定性主题的潜在目的是什么。这些信息跟经济政策不确定性的研究息息相关,可以成为研究灵感来源、提高结论可信度、丰富研究主题等。因此,在利用文本数据构建出指标后,我们也可以尝试像历史等学科一样去了解甚至掌握研究所用文本材料,为后续的经济学研究和分析打下坚实基础。

当然,文本数据利用的局限性也跟缺乏高效优质的分析工具有关。历史研究一般苦于史料缺乏,因此能对不多的史料进行精细地研读和分析,在此基础上开展研究。然而经济学研究面临着文本数据过多的问题,例如上市企业的一份年报高达上百页,一年有几千份年报,研究者无法一一阅读这些文本材料。此外,这些材料的信息密度往往也低于史料,所以人工阅读的"性价比"不高,更适合用计算机自动化处理。然而对于材料的背景、叙事、逻辑等复杂信息,计算词频、统计和机器学习等工具难以处理。虽然自然语言处理领域已经有了各种各样的模型尝试去做文本摘要、文本分类、实体识别等工作,但是实际效果还难以令人满意。所以,在生成式AI出现之前难以深入地利用文本大数据。

三 生成式AI处理文本数据的优势

鉴于已有的文本分析方法存在的各种问题,AI学界一直在寻找更为高效、简单、通用的工具,经过不断迭代,利用超大算力对超大数据进行训练后得到的能生成新内容的超大模型生成式AI逐渐替代传统机器学习成为主流方向。原因有以下三点:

首先,生成式AI极大地拓宽了文本数据的处理范围,将丰富研究主题。目前经济学研究领域的文本处理主要是计算词频以及文本分类,例如文本情绪等指标的构建既有通过对文本分词后计算情绪的词频来实现[25],也有利用统计与机器学习进行分类来完成[26]。对于一些更为复杂的文本处理任务,已有方法难以取得预期效果,导致了研究受限。功能强大的通用型生成式AI出现之后,探讨如何更加充分利用文本信息来完成复杂任务、探索新主题的文献陆续出现,例如使用生成式AI抽取和分析文献中的因果推断信息[3]、识别用词细微差异的含义[4]、挖掘年报中蕴含的投资机会[5]、阐释美联储报告中的技术性词汇[8]以及模拟人类的偏好和反应[9-10]。可以预见,生成式AI带来的文本研究热潮方兴未艾,其应用场景有待进一步开发,但其必将带来文本研究的大繁荣。

其次,生成式AI极大地降低了文本分析的门

槛，让更多研究者可以轻松进入。为了完成特定任务，机器学习等传统做法需要通过撰写代码操控模型来完成语言任务，其中甚至可能涉及 AI 等领域的专业知识，这对人文社科研究者而言具有一定挑战。而生成式 AI 能够较好地依据人类语言指令来完成特定任务，极大地降低了使用难度。比如刘青和肖柏高基于 BERT 语言模型识别劳动节约型专利的过程比较复杂，需要人工标注样本、训练和预测等多个环节[20]。而本文经过测试，只需要将劳动节约型专利的定义和专利摘要输入生成式 AI 中，生成式 AI 就可以给出更加准确的分类结果①。

最后，生成式 AI 掌握了大量的语言和客观知识，并且具备推理、深思等智能，大幅提升了文本处理的准确率。人类语言和客观知识的高度复杂性注定需要大数据并且由大模型来进行学习。通过引入自监督学习（即利用没有打标签的原始文本来作为机器进行学习的样本），几乎所有的人类文本都可以用作训练集，极大地扩充了语言模型能学习到的知识范围。同时新的神经网络的出现使得语言模型的学习效果能随模型参数和学习样本的增加而持续提升，2018 年的 BERT 模型参数为亿级，2022 年的 ChatGPT 则已经是千亿级别。参数的增加不只是线性地提升了语言模型性能，更是出现了“涌现”（emergence）现象，即参数的量变带来模型能力的质变，语言模型出现了一些未预料到的新能力[27]。目前 AI 学界主要关注的“涌现”现象包括“语境学习”（in-context learning）和“思维链”（chain-of-thought）、推理和深思等。这些能力大大拓展了生成式 AI 能应对的任务范围，很多经济学等领域以往无法处理的文本任务或将迎来希望。

四　生成式 AI 时代下文本研究新范式探究

（一）生成式 AI 成为文本分析标准工具

在 AI 领域，生成式 AI 取代其他文本、音视频等数据处理工具、方法和模型已经成为趋势。随着 AI 自然处理领域研究的繁荣，各种子课题、子任务涌现，例如阅读理解、文本摘要、文本分类、问答对话等，每个子任务衍生出了各自的解决思路和模型，此时的趋势是各个任务之间所用思路和工具的差异越来越大。而 ChatGPT 等生成式 AI 的出现扭转了这一趋势，由于生成式 AI 在上述任务上都展示出更加优异的表现，阅读理解等子任务和相关模型的研究开始消亡，自然语言处理领域的研究重心转向如何让生成式 AI 更通用、更强大、更安全，以应对更加复杂的各类文本任务[28]。此外，由于 ChatGPT 等文本生成式 AI 展示出了对文本这一非结构化数据强大的处理能力，音视频等非结构化数据的处理也借鉴和利用 ChatGPT 的底层机制来构建生成式 AI，甚至目前已经出现了能同时处理文本、音视频的多模态生成式 AI。可以预见，生成式 AI 将成为一个最为基础和通用的 AI 工具，成为各行各业的“水电”。

经济学领域或将同样跟随这一潮流。在生成式 AI 出现之前，由于研究者在文本处理能力方面的差异，且不同文本处理方法各有优劣，经济学研究中的文本分析方法和工具繁多。例如技术分类研究中，既有使用词频[29]，又有机器学习中的狄利克雷分布[30]、贝叶斯模型[31]、word2vec 模型[32]以及 AI 语言模型 BERT[20]。同时，这些工具也日趋复杂和昂贵，例如使用 BERT 不仅需要深入了解 BERT 的诸多参数含义和撰写代码，还要价值不菲的显卡进行驱动。此现象一方面反映了文本相关经济学研究的繁荣和发展，另一方面也意味着文本相关研究的进入壁垒增高，以及研究者的很大一部分精力被文本处理工具的选择和应用消耗，反过来遏制文本相关研究的深化。前文中已经展示了生成式 AI 比上述方法明显更准确、更简单，因此，生成式 AI 的出现将极大地“解放”研究人员，可以让研究者把更多的精力放到经济学含义的深挖，而不是文本分析工具的选择和使用上。

（二）生成式 AI 成为文本研究智能助手

较历史等学科的学者而言，经济学学者对文本数据的使用还处于相对粗浅的阶段，这可能会带来研究结论偏误和研究主题狭隘等后果。同样作为极度依赖文本史料的实证科研，历史研究建立在对史书等文本史料的严谨搜寻、考证、理解和辨析之上，体现出历史学者将研究建立在对史料充分掌握基础之上的严谨态度。然而，经济学研究使用文本材料的严谨和充分程度尚有不足，主要体现为不重视阅读和了解所使用的文本材料，或者是简单机械地使用文本数据。例如当前衡量企业技术转型的通常做法是计算企业年报中技术词汇的频数，没有去阅读企业文本的具体内容，即便是技术词汇所在句子、段落也未得到阅读，这会带来不少的问题。例如，技术是导致企业新旧交替的重要因素，被新技术替代而面临淘汰压力的企业在年报中谈及自身受到了技术负面冲击，这些技术词汇会被错误地识别为在进行

① 受制于篇幅，具体过程并未列出。

技术转型,最后我们利用回归分析可能会得到技术转型导致企业被淘汰的相反结论。

此外,对文本材料的深挖和追问还有助于深入和拓展研究主题。在以文本为核心数据来源的研究中,研究命题很少有来自所用文本的启发和直接验证,更多的是研究者自身提出想法,然后基于文本构建实证指标,最后用回归的方式来验证命题。机制等也一般源于成熟的经济学理论,缺乏对文本内经济现象的剖析。同样以技术转型为例,在借助企业年报研究技术转型的各种影响时,已有文献很少从年报里面去找寻企业进行技术转型的原因、目标、方式、机制、影响等特征。实践是理论产生和发展的核心来源,我们以企业为样本研究技术转型,以企业年报为数据来源,应当关注样本现实中的技术转型实践,特别是这些实践就记录于研究所用的企业年报中,便捷可得。这种现象一方面意味着学术研究在关注实践方面的不足,另一方面也可能导致了宝贵信息和潜在研究课题的埋没,不利于研究主题的深化和拓展。

当然,导致上述问题的关键并非研究者"不愿",而是"不能"。经济学研究要处理的文本量往往远超历史等学科,所以像历史研究那样精细、严谨地使用文本在客观上几乎无法实现,在技术的限制下不得已才采取了简单的处理手段。文本数据爆炸成为经济学等各领域学者面临的共同难题,如何在浩如烟海的文本资料中找出跟研究主题相关的内容,并加以分析、总结和阐释,成为破解难题的关键。词典、传统机器学习等方法虽然可以处理大规模文本,但是它们对文本的处理存在以下问题:一是处理简单、机械,智能程度太低,远远谈不上像研究者一样去阅读、理解、总结文本,例如分词后匹配关键词的做法把语言进行了"肢解",导致文本失去了原有含义。二是处理过程完全依照研究者预设特定规则进行,在本质上而言它们是研究者智能的延伸、想法的实现,没有创造出新的智能活动,即这些方法无法通过处理原始文本资料自主生成新发现、新想法等在内的新内容。因此从本质上而言,所有的跟文本相关的高阶智能活动还是得由人类大脑完成。

生成式AI将成为经济学文本研究的有力助手,它有能力把所有原始文本数据仔细"阅读、分析、总结",为研究者提供准确参考乃至新证据、新发现。生成式AI在处理文本方面已经逐渐逼近人类的水平,一方面,如前文指出,通过给予生成式AI劳动节约型专利的文字定义,它就可以根据定义和专利摘要准确识别。在此过程中我们并没有设定特定明确的规则,生成式AI是根据对于定义和摘要内容的理解灵活完成识别工作,说明了生成式AI能如同人类一样完成文本相关任务。另一方面,一系列的研究表明生成式AI可以通过阅读文本去尝试理解用词细微差异背后的含义、挖掘高管发言中的潜在投资机会、预测企业盈余等[4-6],说明生成式AI能够像人类一样"思考",具备较高程度智能。可以预见,之后的经济学文本研究中,生成式AI将成为一个研究助手,自动进行文本数据阅读、理解、分析和总结工作,从而能更加精细、深入地理解和运用文本数据。

(三)生成式AI"端到端"成为文本研究新过程

"端到端"(end to end)是指将原始数据输入模型后模型直接输出最终结果,不需要清晰地定义或设定中间分析步骤。同样以基于企业年报的技术转型实证研究为例,一些实证研究所需的解释变量(技术转型)、结果变量和控制变量(如财务数据等)等都存在于年报中,传统做法需把这些信息从年报中提取出来,构造数值变量,然后借用因果推断回归模型对这些数据进行分析,并得出结论。而"端到端"的模式则是将原始年报数据输入生成式AI,生成式AI对所有原始材料进行处理后直接得到技术转型和结果变量之间的关系。也就是说,"端到端"将数据提取、模型构建、因果分析的步骤全部内化、融合到生成式AI模型中,生成式AI直接基于原始数据得出最终结论。

这一思路理论上可行,现实中其他领域也正在应用这一思路。从理论上而言,数据提取、变量构建和因果分析均可以被视为函数,例如因果推断函数的输入为各种变量,输出为因果结论。从更广的视角来看,基于文本等原始数据推断出结论也可以被视为一个更加复杂的函数。而根据通用逼近定理(universal approximation theorem),只要神经网络的层数和神经元数量足够多,神经网络可以在任何精度上逼近任意函数[33-34]。所以理论上而言生成式AI可以无限逼近传统研究范式。在现实中,同样需要处理复杂、非结构化视频数据的自动驾驶的趋势就是"端到端",即自动驾驶模型将车辆上传感器获取的道路相关信息直接转化为加减速、转向角度等驾驶决策。而传统自动驾驶模型会先把自动驾驶分解为感知、预测、规划、控制等模块和步骤,然后对各个模块进行设计和优化,最终将它们整合到一起以完成自动驾驶任务。"端到端"自动驾驶模型开始取代传统驾驶模型,并广泛在汽车行业推广开来充分说明了"端到端"大模型的趋势性。

生成式AI"端到端"的分析路径具有信息损失少、效率高、发现能力强等优点。信息损失是指,在

传统研究过程中每一步骤本身的信息处理和步骤之间的信息传递存在损失。以基于词频统计的技术转型为例，“企业正在加速推进技术转型”和“企业正在逐渐推动技术转型”中的技术转型程度信息就丢失了，这是提取信息步骤内部的信息损失。而为将文本数据用于回归分析，我们不得不对文本进行某种处理，不管是词频还是向量化，均会带来信息的损失，这就是步骤之间的信息传递损失。而生成式 AI 通过融合上述步骤，实现了从数据端到结果端的直接转化，在一定程度上避免了信息损失问题。而且，相较于研究者通过数据整理、数据分析等步骤得出结论，生成式 AI 的处理更加便捷和高效，因而也能加快研究速度。此外，在文本大数据时代，囿于研究能力的不足，可能存在大量的蕴藏于文本中的研究议题和机制尚未得到揭示，一方面“端到端”的研究过程不需要研究者介入或者只需耗费较小成本，另一方面它通过对超大数据的学习和分析，可能具备发现研究者所未关注到的新现象、新机制的能力，从而能为学术界贡献一定成果。

当然，实现这一目标目前还面临许多难题。首先，虽然从理论上而言传统的研究过程函数可以通过神经网络无限逼近，但是如何找出神经网络是一个极大的挑战。当前的生成式 AI 是在通用型文本上训练得到的，训练出的生成式 AI 处理的是通用型任务，对于专精深的研究类任务自然要求大量高质量学术研究成果作为训练集，学术文献的样本量能否满足这样的神经网络训练需求尚无法确定。其次，“端到端”意味着内部机制的不可知。在清晰的研究步骤和机制被深度神经网络取代后，神经网络的“黑箱”特征意味着结论的可解释性面临着极大的考验。只知其然不知其所以然对于科学研究来说不可接受，所以如果要使得“端到端”的研究方式得到大范围应用，必须解决这一问题。当然，这并非意味着在没有解决这一问题前就不能使用，缺乏过程但结果正确的结论也是可以大大推动研究进程的，例如人类先是实现飞行然后研究飞机的机理，所以“端到端”对学术研究中的稳健性检验等方面而言大有助益。最后，“端到端”相当于生成式 AI 完全自主做研究，其研究深度、创新性等能否达到人类水平还有待验证。

五　使用生成式 AI 的建议

（一）注重提示语的撰写

提示语（prompt）是研究者与生成式 AI 交互沟通的桥梁，在很大程度上决定了研究者使用生成式 AI 做研究的质量，因此本节对提示语进行简要介绍和说明。提示语是引导和控制生成式 AI 的语句，一般而言它包含任务指令、上下文、输入数据和输出形式四个部分。任务指令是指需要生成式 AI 完成的任务，例如翻译、分类、阅读、总结。上下文是一些跟任务相关的信息，例如背景、说明、定义、示例等。输入数据是指使用者输入生成式 AI 的文本材料，如需要分词或分类的文本。输出形式是指生成式 AI 应该以什么形式输出任务结果。

在给定生成式 AI 能力的情况下，我们所得到的答案质量取决于提示语质量。学会提问一直就是科学研究中的重要技能，利用生成式 AI 来帮助研究时，提示语就好比向其发问，以期得到一个高质量的答复。好的提问，意味着研究者对想要解决的问题有了清晰而全面的认识，进而引导生成式 AI 解决问题。因此，提示语的质量从根本上反映了研究者对问题的理解和思考程度。提示语中和经济学议题相关的部分一般需要研究者自己撰写，除此之外也要掌握一些共性的技巧。根据 AI 界的研究，为了得到高质量答案，尤其是在任务较为复杂的情况下，一个好的提示语往往具有以下特点：提供示例、应用思维链、反复测试与精准表达等。

1. 提供示例

生成式 AI 的突出能力之一就是像人类一样可以通过少量甚至零示例就可以学会处理新任务。生成式 AI 如果要成为文本分析的标准工具，意味着其要具有处理各种复杂问题的能力。在此之前，用于文本情感分类的语言模型面对经济政策不确定性分类任务，往往需要人工标注新数据并重新训练，人工成本和训练成本高昂。而且在一些情况下，可供机器学习的数据集是非常有限的，例如让生成式 AI 去发现年报文本中的欺诈表述，可能不会有太多的欺诈案例可供生成式 AI 学习。而生成式 AI 涌现出来的“语境学习”能力就是可以通过几个示例的学习（few-shot learning）甚至零示例的学习（zero-shot learning）来完成新任务[35]。

是否需要提供示例取决于生成式 AI 的能力和任务的新颖度与难度，一般而言，提供示例可以优化语言模型的表现。生成式 AI 在训练过程中会积累很多先验知识，可直接用于后续任务。研究发现如果语言模型的参数越多，则完成任务时的表现越不依赖示例。因为参数越多意味着模型能掌握更多的知识，那么任务所需的相关知识可能在训练阶段就已经被语言模型掌握，无需示例来注入新知识。任务的新颖度也是一个重要考虑因素，如果任务很常见，例如“情感分类”任务所涉及的跟情感相关的各

种词汇和知识在文献中大量存在,那么语言模型处理此任务就无需示例。但是,如果让生成式AI处理新任务,任务涉及的知识和定义可能没有在生成式AI训练数据中出现过,则提供示例能让模型表现更好。

2.应用思维链

生成式AI在处理复杂、多步骤的任务时"涌现"出一步步解决问题的能力就是思维链。面对复杂的专业问题时,生成式AI思维链在完整度和正确性等方面相对研究者而言会有一定不足,容易出现方向错误、步骤缺失等问题。因此在提示语中,研究者需要"循循善诱",即利用自身的专业知识和见解,明晰问题的解决步骤,让生成式AI更高效地完成任务。提示语中使用思维链的另一个好处是可以根据生成式AI的每一步结果来进一步修改提示语。如果不让生成式AI分步骤完成任务,而是直接给出一个最终的答案,并且这个答案还是错误的,此种情形下,由于生成式AI的内部对于我们而言是一个黑箱,我们将无法得知哪里出了问题,也无法根据生成式AI的答案改进提示语。

3.反复测试与精准表述

一个好的提示语还需要反复调整测试、表述具体精准。尽管生成式AI能力很强,但仍然会出现错误。而且,由于生成式AI内部运作原理在很大程度上仍然是一个黑箱[8],当生成式AI在测试样本上出错时,我们可能难以直接找出症结点,通过不断修改提示语,并且在测试样本上反复测试,能大幅度优化生成式AI的表现。在调整测试过程中,思维链式的提示语可以大大加快这一过程,这也是在提示语中利用思维链的优势之一。提示语的表述要尽量具体精准,不要有模棱两可的词汇。例如,当一段文本内同时有"国内生产总值""GDP""生产总值"等表述,在尝试利用生成式AI找出这些指标时,应该使用的提示语是"找出国内生产总值指标(或同义指标)"的提示语,而不是"找出国内生产总值、GDP等指标"。因为"等指标"也可能被理解为财政收入等其他经济指标。除此之外,当处理大量数据,需要生成式AI按特定格式返回结果时,也要描述清楚格式要求。

(二)避免生成式AI的"幻觉"

"幻觉"(hallucination)是指生成式AI生成一些看似正确、实则虚假的信息。例如,要生成式AI给出一些某领域的参考文献,虽然它给出的参考文献的格式完全正确,但是文献并不存在。生成式AI"幻觉"出现的原因尚未有明确答案,可能是由于对训练集中缺乏的信息的错误推断、学习的训练集中存在错误信息、无监督学习机制下模型只学习而不思考材料的对错等。当前,生成式AI的开发者均在努力避免"幻觉"问题。对于生成式AI的使用者而言,避免"幻觉"的途径有以下几个:第一,在提示语中尽可能地给出任务细节、约束条件和思维链,让生成式AI沿着人为设定的正确路径运行。第二,给生成式AI提供相关正确信息,例如在询问2020年湖南省的经济增长目标时,如果给生成式AI提供政府工作报告全文,则可以避免"幻觉"的发生。

(三)保护数据私密性

在使用生成式AI时,文本数据会被发送至生成式AI拥有者,这可能会导致数据泄露的问题。目前先进的生成式AI均为企业所有,研究者在使用这些模型时,输入数据会被发送至企业,可能会被用来进一步训练生成式AI,甚至还可能会被生成式AI输出给其他使用者,导致研究主题或数据的泄露。因此研究者在使用生成式AI时要注意数据的私密性,特别是不能使用生成式AI处理保密数据,最好只使用生成式AI处理公开的文本数据。当然,处理公开的文本数据时要特别注意这些数据可能已经被生成式AI在训练阶段学习过了,用生成式AI处理这部分数据可以考虑所谓的样本内和样本外的问题,例如Chen等就测试了生成式AI能否基于训练样本外的新闻来预测股价,以作为稳健性检验[36]。

六 总 结

生成式AI的出现是迈向通用型人工智能之路的标志性事件,对各行各业都产生了深远影响。就文本数据分析而言,生成式AI具备的大量语言知识和世界知识,以及"涌现"出上下文学习、思维链、深思和推理等能力,显著提升了文本分析的准确度和效率,将研究者从烦琐的文本数据整理过程中解放出来,让他们可以更加专注于所研究问题本身。生成式AI将逐步取代已有的其他文本分析方法,成为文本数据分析的标准工具。

本文指出已有的文本分析方法存在效率低、语义丢失、难度大、成本高、应用单一等问题,并说明了生成式AI如何解决上述问题。接着本文探索了生成式AI时代下文本分析的潜在趋势,以及如何撰写和优化提示语以让生成式AI更好地完成任务。最后,本文指出了使用生成式AI时还需注意的关键要点。

本文的贡献和意义需要在如下背景下理解:作为人类社会信息主要载体的文本数据在经济学研究

中越来越受到关注，相关文献迅速增长。与此同时，文本分析的难度、准确度和文本数据利用程度不高等问题一直没有得到很好的解决，阻碍当前经济学研究中对文本数据的进一步使用。本文一一剖析了上述问题，提出用生成式 AI 来解决文本分析难题的思路及操作步骤，并对提示语等要点进行了说明。这些工作对加快经济学研究中文本数据分析工具的更新、经济学现象的剖析、研究质量的提升、经济学文本分析范式的变革有一定推动作用。

[参　考　文　献]

[1] Gentzkow M，Kelly B，Taddy M. Text as data[J]. Journal of Economic Literature，2019(3)：535－574.

[2] 唐国豪，姜富伟，张定胜. 金融市场文本情绪研究进展[J]. 经济学动态，2016(11)：137－147.

[3] Garg P，Fetzer T. Causal claims in economics[R]. I4R Discussion Paper Series，2024.

[4] Choi G Y，Kim A G. Economic footprints of tax audits：a generative AI-driven approach[R]. Chicago Booth Research Paper，2023.

[5] Bai J J，Boyson N M，Cao Y，et al. Executives vs. Chatbots：unmasking insights through human-AI differences in earnings conference Q&A[R]. Northeastern U. D' Amore-McKim School of Business Research Paper，2023.

[6] Kim A，Muhn M，Nikolaev V. Financial statement analysis with large language models[R]. arXiv preprint arXiv：2407.17866，2024.

[7] Lopez-Lira A，Tang Y H. Can ChatGpt forecast stock price movements? Return predictability and large language models[R]. arXiv preprint arXiv：2304.07619，2023.

[8] Hansen A L，Kazinnik S. Can Chatgpt decipher fedspeak[R]. Available at SSRN，2023.

[9] Bybee J L. The ghost in the machine：generating beliefs with large language models[R]. arXiv preprint arXiv：2305.02823，2023.

[10] Goli A，Singh A. Frontiers：can large language models capture human preferences? [J]. Marketing Science，2024(4)：709－722.

[11] 洪永淼，汪寿阳. 人工智能新近发展及其对经济学研究范式的影响[J]. 中国科学院院刊，2023(3)：353－357.

[12] 洪永淼，汪寿阳. ChatGPT 与大模型将对经济学研究范式产生什么影响[J]. 计量经济学报，2024(1)：1－25.

[13] Korinek A. Language models and cognitive automation for economic research[R]. National Bureau of Economic Research，2023.

[14] 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[J]. 数量经济技术经济研究，2024(5)：197－216.

[15] 政光景，吕鹏. 生成式人工智能与哲学社会科学新范式的涌现[J]. 江海学刊，2023(4)：132－142.

[16] Birhane A，Kasirzadeh A，Leslie D，et al. Science in the age of large language models[J]. Nature Reviews Physics，2023(5)：277－280.

[17] 沈艳，陈赟，黄卓. 文本大数据分析在经济学和金融学中的应用：一个文献综述[J]. 经济学(季刊)，2019(4)：1153－1186＋224.

[18] 吴璇，田高良，李玥婷，等. 经营信息披露与股票收益联动：基于财务报告文本附注的分析[J]. 南开管理评论，2019(3)：173－186.

[19] 张成思，孙宇辰，阮睿. 宏观经济感知、货币政策与微观企业投融资行为[J]. 经济研究，2021(10)：39－55.

[20] 刘青，肖柏高. 劳动力成本与劳动节约型技术创新：来自 AI 语言模型和专利文本的证据[J]. 经济研究，2023(2)：74－90.

[21] Hoberg G，Phillips G. Text-based network industries and endogenous product differentiation[J]. Journal of Political Economy，2016(5)：1423－1465.

[22] 胡百精，安若辰. 认知落差、媒体叙事与冬奥会的全球想象[J]. 上海体育学院学报，2020(1)：31－38.

[23] 何天平. 数字新闻叙事的结构、话语和文化研究[J]. 新闻大学，2023(1)：28－38.

[24] Sharir O，Peleg B，Shoham Y. The cost of training NLP models：a concise overview[R]. arXiv preprint arXiv：2004.08900，2020.

[25] 姜富伟，孟令超，唐国豪. 媒体文本情绪与股票回报预测[J]. 经济学(季刊)，2021(4)：1323－1344.

[26] 张宗新，吴钊颖. 媒体情绪传染与分析师乐观偏差：基于机器学习文本分析方法的经验证据[J]. 管理世界，2021(1)：170－185.

[27] Wei J，Tay Y，Bommasani R，et al. Emergent abilities of large language models[R]. arXiv preprint arXiv：2206.07682，2022.

[28] Zhao W X，Zhou K，Li J，et al. A survey of large language models[R]. arXiv preprint arXiv：2303.18223，2023.

[29] 吴非，胡慧芷，林慧妍，等. 企业数字化转型与资本市场表现：来自股票流动性的经验证据[J]. 管理世界，2021(7)：130－144＋10.

[30] Bellstam G，Bhagat S，Cookson J A. A text-based analysis of corporate innovation[J]. Management Science，2021(7)：4004－4031.

[31] Mann K，Püttmann L. Benign effects of automation：new evidence from patent texts[J]. Review of Economics and Statistics，2023(3)：562－579.

[32] 姚加权，张锟澎，郭李鹏，等. 人工智能如何提升企业生产效率：基于劳动力技能结构调整的视角[J]. 管理世界，2024(2)：101－116＋133.

[33] Cybenko G. Approximation by superpositions of a sigmoidal function[J]. Mathematics of Control，Signals and Systems，1989(4)：303－314.

[34] Hornik K，Stinchcombe M，White H. Multilayer feedforward networks are universal approximators[J]. Neural Networks，1989(5)：359－366.

[35] Brown T，Mann B，Ryder N，et al. Language models are few-shot learners[J]. Advances in Neural Information Processing Systems，2020(3)：1877－1901.

[36] Chen J，Tang G，Zhou G，et al. ChatGPT，stock market predictability and links to the macroeconomy[R]. Available at SSRN 4660148，2023.