

基于语料库的中国媒体人工智能话语建构研究

丁彦心, 裴佳敏*

浙江工商大学外国语学院, 浙江 杭州

收稿日期: 2026年5月6日; 录用日期: 2026年7月17日; 发布日期: 2026年7月28日

摘要

面对人工智能技术对全球经济结构、社会运行模式及大众日常生活的深远影响, 以及国内AI产业的新发展态势, 主流媒体的报道对公众理解AI技术发挥着关键作用。本文以中国主流媒体AI新闻报道为研究语料, 运用语料库分析方法, 分析媒体对人工智能的话语建构。主题词分析显示, 媒体主要关注AI技术创新与发展、AI产业应用与赋能、以及AI经济影响与战略地位三大核心议题。搭配词分析揭示出AI被描述为需要积极发展、广泛应用和规范治理的领域。同时, 亦涉及深度伪造、AI幻觉、算法黑箱、数据隐私等风险议题, 反映报道在突出AI发展的同时也涉及对潜在风险的关注。总体而言, 中国主流媒体对AI的话语建构呈现发展与风险并存的多维特征。

关键词

人工智能, 语料库, 主题词, 搭配词

A Corpus-Based Study on the Discursive Construction of Artificial Intelligence in Chinese Media

Yanxin Ding, Jiamin Pei*

School of Foreign Languages, Zhejiang Gongshang University, Hangzhou Zhejiang

Received: May 6, 2026; accepted: July 17, 2026; published: July 28, 2026

Abstract

Given the far-reaching implications of artificial intelligence technology for global economic structures,

*通讯作者。

文章引用: 丁彦心, 裴佳敏. 基于语料库的中国媒体人工智能话语建构研究[J]. 现代语言学, 2026, 14(7): 955-961.
DOI: 10.12677/ml.2026.147726

modes of social operation, and public daily life, coupled with new developmental dynamics in China's domestic AI industry, mainstream media reporting serves a pivotal function in facilitating public comprehension of AI technology. This paper examines AI news coverage in Chinese mainstream media as research corpora, utilizing corpus linguistic methods to analyze the media's discursive construction of artificial intelligence. Thematic analysis indicates that media attention centers on three principal domains: AI technological innovation and advancement, AI industrial applications and enablement, and AI's economic ramifications and strategic significance. Collocation analysis reveals that AI is discursively framed as a field necessitating active development, extensive application, and normative governance. Meanwhile, the concordance lines also surface risk-related issues such as deep-fakes, AI hallucinations, algorithmic black boxes, and data privacy, indicating that the coverage attends to potential risks alongside its emphasis on AI development. Overall, the discursive construction of AI in Chinese mainstream media exhibits a multi-dimensional feature in which developmental and risk-related discourses coexist.

Keywords

Artificial Intelligence, Corpus, Keywords, Collocations

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

人工智能技术正在全方位地改变着我们的世界。它不仅影响着全球经济结构调整、社会运行模式的革新,也日益融入大众的日常工作与生活之中,从智能制造、智慧医疗到个性化推荐、智能家居,其应用场景不断拓展。近段时间,DeepSeek 等国内 AI 企业的崛起,让中国人工智能产业正呈现出新的发展态势。在这一过程中,媒体的报道对公众理解这项技术起着关键作用,媒体对 AI 的报道不仅反映了技术发展的最新动态,也在一定程度上影响着人们对 AI 的看法,中国媒体对人工智能的报道,不仅是国内 AI 发展态势的一面镜子,也在无形中塑造着社会大众对这项技术的理解和态度,因此,对中国媒体如何论述人工智能进行研究,不仅有助于把握 AI 技术在国内的发展态势,也可以获得关于 AI 技术社会认知的宝贵信息,为相关政策制定和技术发展提供参考。

2. 相关研究回顾

人工智能是近年来学术界日益关注的主题。研究主题方面,国外研究主要关注人工智能对经济、社会的影响,以及相关的伦理问题,比如算法偏见、数据隐私泄露、潜在的失业焦虑[1]。Sanguinetti & Palomo (2024)的研究分析了欧美媒体中 AI 所引发的焦虑情绪,发现 AI 被媒体塑造为一个具有自主性、不透明乃至“异己”的存在,这可能加剧公众的恐慌,而国内研究更加关注 AI 的技术发展、产业应用和政策解读[2]。张梦和陈昌凤(2021)在其综述中系统梳理了国内智能媒体研究的脉络,指出其相关研究主要集中在人工智能技术如何改变新闻工作的各个环节,包括新闻的采集、制作、发布和核实,以及人工智能技术普及后,新闻行业需要面对的新伦理问题及其引发的思考等几个方面[3]。李明德和邝岩(2021)研究了大数据和 AI 在网络舆情治理中的应用,分析了它们的积极作用和潜在风险,强调既要利用技术优势,也要加强规范管理[4]。

在研究视角与方法上,政策文本分析被用于解读国家 AI 战略与媒体话语的关联;内容分析法常用于

考查 AI 报道的议题分布、情感倾向, 穆军芳和陈紫涵(2023)以趋近化理论为基础, 运用语料库分析工具分析社交媒体中人工智能报道发现媒体话语中应用了大量的空间趋近化策略以展现 AI 对经济发展和进步的推动作用, 同时, 也有少量报道提示 AI 可能带来的风险, 促使公众保持警惕[5]。话语分析, 特别是批评话语分析的理论框架, 也被引入以揭示媒体表征背后的深层含义。阿力亚·艾尼与张正坤(2024)的研究运用语料库辅助的批评话语分析三维框架, 考查美国主流媒体在北京冬奥会中对中国国家形象的建构[6]。然而, 尽管相关研究已取得显著进展, 但仍存在一些不足之处。首先, 基于语料库的人工智能媒体话语研究尚不多见。多数研究或侧重于理论思辨, 或样本量有限, 未能充分利用语料库的分析工具。其次, 将自建新闻语料库与大型参照语料库进行对比, 从而提取出相关的主题词, 并在此基础上进行深入的主题分析的研究相对缺乏。本研究力图通过严谨的语料库方法, 为中国主流报纸媒体的 AI 话语建构研究提供思路。

3. 理论框架与研究方法

3.1. 理论框架

本研究以批评话语分析为理论基础。批评话语分析认为, 语言不仅是交际工具, 更是社会权力关系的体现和建构手段。媒体话语通过词汇选择、句法结构、隐喻运用等语言策略, 建构特定的社会认知和意识形态框架[7]。

在分析层面, 本研究借鉴 Baker 提出的语料库辅助话语研究范式, 将语料库语言学的量化方法与批评话语分析的质性阐释相结合。这一范式强调通过关键词、搭配和索引行等语料库工具揭示话语模式, 再运用批评话语分析的框架进行深度阐释[8]。

3.2. 研究方法

3.2.1. 语料建设

语料来源于慧科新闻数据库。该数据库全面收录了国内众多主流媒体及地方报纸, 本研究从中选取了“国内最具影响力综合性报章”作为语料抽样范围。选择这些报章作为语料来源, 是因为它们通常具有较大的发行量、广泛的社会影响力以及较高的权威性, 其对人工智能的报道能够在一定程度上反映中国主流媒体对 AI 议题的呈现方式与话语特征。本研究设定的新闻报道检索时间段为 2025 年 2 月 1 日至 2025 年 3 月 31 日。2025 年 1 月, 国内大模型企业 DeepSeek 推出了推理模型 DeepSeek-R1, 引发主流媒体广泛报道与社会热议, 选择这一时间段, 旨在捕捉这一时期内媒体对人工智能议题的集中关注和话语特征。在慧科数据库中检索“人工智能”“AI”“A.I.”以及“Artificial Intelligence”等关键词, 确保所获新闻报道的核心议题与人工智能高度相关, 人工剔除无关结果, 最后建成中国主流媒体人工智能新闻语料库, 总语料库共 1092 篇文章, 总形符数 817,044。

为了更科学地提取 AI 新闻语料库中的主题词, 本研究引入了参照语料库进行对比。参照语料库选用的是 Torch2014 数据库(<https://corpus.bfsu.edu.cn/info/1072/1015.htm>)。Torch 通常被学界认为是代表当代汉语普通话通用书面语的大型平衡语料库, 其词汇和语用特征能够反映现代汉语的一般使用情况。通过与 Torch 语料库进行对比, 可以有效地突显出 AI 新闻报道领域中那些使用频率异常高、具有显著性的词汇, 即主题词。

3.2.2. 分析步骤

本研究主要采用语料库语言学的研究方法, 结合定量统计与定性分析。语料库方法能够基于大规模真实文本数据, 客观、系统地揭示语言使用的模式和特征, 避免了单纯依赖直觉或小样本分析可能带来

的主观性和片面性[9]。研究方法分为两个步骤：第一，利用 AntConc4.3.1 获取新闻语料库主题词，从宏观方面观察主流媒体关于人工智能报道中所关注的核心词汇和主要议题。第二，筛选出与“人工智能”搭配的前二十个高频搭配词从微观层面分析这些搭配词项与“人工智能”共同出现的具体语境，探究其构成的话语模式以及所反映出的媒体叙事和价值取向。

4. 数据分析及讨论

4.1. 主题词分析

利用 AntConc 软件,将自建的 AI 新闻语料库与 Torch 参照语料库进行对比,并设置其 p 值小于 0.01,提取了 LL 值较高的一系列主题词。通过将排名前 50 位的主题词进行分类后可以看出,媒体关注的议题可大致分为三类: AI 技术创新与发展、AI 产业应用与赋能、AI 经济影响与战略地位(见表 1)。

Table 1. The top 50 keywords grouped by topic
表 1. 前 50 个主题词按主题分类

主题	主题词
AI 技术创新与发展	AI, 人工智能, DeepSeek, 模型, 技术, 智能, 创新, 科技, 生成, 深度, 数据, 平台, 算力, 研发, 开源, 精准, 融合, 新, 算法, 工具, 智能化, 建议, 表示
AI 产业应用与赋能	产业, 应用, 机器人, 场景, 领域, 赋能, 政务, 行业, 服务, 视频, 提升, 用户, 发布, 接入, 高质量, 数字
AI 经济影响与战略地位	企业, 推动, 深圳, 广州, 广东, 发展, 全球, 未来, 提供, 打造, 人才

在 AI 技术创新与发展主题中,“技术”“模型”“算法”“算力”“数据”等词汇的出现,反映了媒体对于人工智能核心技术要素、前沿进展以及支撑技术发展的基础设施的关注。这些主题词展现出媒体对 AI 快速发展、持续创新的特点的关注,“模型”一词的提及,指向了当前 AI 技术发展的热点方向,即大数据模型,而“算力”为人工智能系统提供必要的计算资源支持,“数据”则为机器学习模型训练提供基础素材,这两类基础设施的持续优化与完善,是推动 AI 技术不断发展的重要支撑条件。DeepSeek 一词出现,说明作为国内 AI 企业的代表,中国本土的 AI 技术和创新成果正受到越来越多的关注,体现出媒体对行业动态的及时跟进,也在一定程度上反映了国内在人工智能核心技术研发取得的实质性突破,既体现出我国科技创新的实力,也为全球人工智能发展作出了重要贡献。

值得注意的是,“技术”“算法”等词也涉及到关于 AI 的风险,如深度伪造技术滥用(见例 1)以及算法黑箱、数据偏见等问题(见例 2),表明媒体在报道 AI 技术创新与发展的同时,也关注其潜在的技术风险与治理需求。

例 1: AI “换脸拟声”乱象的出现,除了违法成本低、监管法规滞后、利益驱动等之外,技术的便利性也是主要原因。深度伪造(Deepfake)技术自 2017 年问世以来,正以指数级速度渗透日常生活。(北京晚报, 2025/03/15)

例 2: 我国人工智能迅猛发展,在赋能千行百业、促进经济社会高质量发展的同时,技术狂飙突进带来的“算法黑箱”“数据偏见”“深度伪造”等问题也日益凸显。(人民日报, 2025/03/27)

在 AI 产业应用与赋能主题中,“产业”“应用”“赋能”“行业”“政务”“服务”“高质量”等词反映出人工智能正在快速进入各个领域,改变传统的工作方式,给各行各业带来效率提升和新的发展机会,这表明 AI 已经不仅仅是一项新技术,而是推动产业升级、优化政务服务、改善民生的重要工具。“高质量”强调了 AI 在应用中所追求的效果标准和发展方向,这与中国经济社会“高质量发展”的国家战略相契合,而“数字”的出现,表明在数字化转型和数字经济的宏大背景之下, AI 被视为构建数字社

会、驱动数字经济发展的关键力量,“赋能”一词则更是生动体现了人工智能技术在各领域的实际应用价值,形象地表达了 AI 技术正在给各行各业带来实实在在的改变,工厂用 AI 优化生产线,政府用 AI 提高办事效率,医院用 AI 辅助诊断,这些应用都在证明,人工智能技术正在从理论概念转化为切实可行的生产力工具,有效推动传统行业的效率提升和创新发展(见例 3、例 4)。可以预见,人工智能将持续发挥其技术优势,成为促进产业升级和经济发展的新动力。

例 3:接下来要支持人工智能企业与制造业龙头企业形成赋能联合体,加快工业人工智能应用模式创新探索……(人民日报,2025/03/08)

例 4:从“拿地即动工”的审批革命,到“AI 员工”赋能的智慧政务,中山以刀刃向内的改革魄力与数字创新的敏锐嗅觉……(南方日报,2025/03/30)

与此同时,本主题下的“行业”“数字”等词亦呈现出 AI 在产业应用过程中的风险,如 AI 诈骗治理缺乏行业合力(见例 5)以及 AI 换脸拟声违法侵权(见例 6),说明媒体在报道 AI 产业应用的同时,也注意到其在不同行业落地过程中暴露出的治理难题。

例 5:AI 诈骗治理缺乏行业合力。各个行业数据泄露的情况仍然存在,即便一些用于保护用户数据的基础性制度开始实行……(新京报,2025/03/14)

例 6:近期,小米集团创始人雷军建议加强“AI 换脸拟声”违法侵权重灾区治理,再次引发公众对 AI 数字人使用的讨论。(新华每日电讯,2025/03/14)

在 AI 经济影响与战略地位主题中,媒体经常使用“发展”“全球”“未来”这些词汇,揭示了 AI 已经超出普通的技术创新范畴,转而上升到成为影响国家竞争力和经济转型的重要因素,“深圳”“广州”“广东”这些具体地域名称的出现代表着 AI 产业发展中的关键区域集群和创新高地,这些地名在报道中常与 AI 产业布局、政策推动、战略定位等语义相关联(见例 7、例 8),在媒体话语中被赋予了经济引擎和战略支点的象征意义,反映媒体对 AI 区域发展模式 and 经验的关注,“人才”是 AI 经济影响和战略地位中最核心的人力资源要素,在报道中“人才”常与“培养”“缺口”等搭配共现(见例 9),体现媒体对智力资本和人才战略的深层关注。分析发现,中国媒体倾向于将人工智能置于国家发展大局和国际科技竞争的宏观背景下进行报道,更加强调它在推动经济转型、增强国家竞争力以及提升国际科技地位等方面的重要作用,这种话语建构方式与中国将人工智能提升到国家战略高度的政策导向高度相关。

例 7:在人工智能相关企业的数量上,以及在基础层、技术层、应用层的企业分布上,均位居全国之首,成为人工智能“密度”最高的城市……深圳向人工智能和机器人开放城市全域全时全场景……(南方都市报,2025/02/28)

例 8:广州正加快构建“12218”现代化产业体系,把“人工智能”定位为战略先导产业,深入开展“人工智能+”行动。(广州日报,2025/03/27)

例 9:韩民春:“人才是创新的第一资源。目前,我国精通人工智能知识的专业人才缺口较大……”(人民日报,2025/02/20)

以上可以看出,随着 AI 技术的快速发展和广泛应用,社会各界对这项技术的认识也在不断深化,从最初单纯关注其技术突破和应用前景,逐渐转向更加全面、辩证的思考。媒体话语在突出 AI 产业与战略机遇的同时,也呈现出对 AI 风险议题的关心,反映出在推进 AI 技术发展过程中对治理框架建设的关注。

4.2. “人工智能”搭配分析

对节点词“人工智能”进行搭配分析,可以进一步揭示其在新闻报道中的语境和语义组合特征。为此,本研究提取以“人工智能”为节点词,跨距分别为 L5-R0 和 L0-R5 的前 20 的高频词表,并按词性和

主题分别分类(见表 2 和表 3)。

Table 2. The top 20 high-frequency collocates of “artificial intelligence” classified by part of speech
表 2. “人工智能”的前 20 个高频搭配词按词性分类

词性	主题词
名词	技术(762), 产业(754), 领域(506), 创新(487), AI(459), 企业(411), 教育(323), 机器人(304), DeepSeek (255), 场景(249), 建设(226), 深度(211), 治理(210), 行动(204)
动词	发展(1015), 应用(711), 推动(408), 赋能(335)
形容词	全球(272), 生成式(232)

将搭配词按词性分类(见表 2)观察发现“人工智能”在新闻报道中呈现与名词、动词、形容词的词汇组合, 这些组合从不同角度展现了人们对人工智能的理解。名词性搭配如“人工智能技术”直接点明了其作为一种前沿科技的本质属性。“人工智能产业”“人工智能领域”“人工智能企业”“人工智能经济”等名词搭配, 表明媒体倾向于将人工智能视为一个新兴的、具有巨大潜力的产业领域和经济增长点。动词性搭配词如“发展人工智能”“应用人工智能”“推动人工智能”“赋能人工智能”, 揭示了媒体话语中与人工智能相关的核心行为。AI 被描述成一个需要积极发展、广泛应用、大力推动的领域, 并且是驱动其他领域变革的重要力量。形容词性搭配如“全球人工智能”“生成式人工智能”, 这类搭配反映了媒体对 AI 发展阶段、技术水平和未来发展态势的描述。

Table 3. The top 20 high-frequency collocates of “artificial intelligence” classified by topics
表 3. “人工智能”的前 20 个高频搭配词按主题分类

主题	主题词
技术核心	技术(762), 深度(211), 生成式(232), AI(459)
发展与规范	发展(1015), 创新(487), 推动(408), 建设(226), 行动(204), 治理(210)
产业生态	产业(754), 企业(411), DeepSeek (255), 全球(272)
应用与赋能	应用(711), 领域(506), 赋能(335), 教育(323), 机器人(304), 场景(249)

将这些搭配词经主题分类(见表 3)发现主要包括以下几个方面: 首先, 是 AI 的技术核心, 聚焦 AI 的核心算法与基础理论, 包括机器学习、深度学习等前沿技术方向, 构成 AI 发展的理论基础; 其次, 是 AI 的发展与规范, 描述了 AI 作为一项事业不断向前演进的过程, 以及为引导和管理这一过程所采取的行动和规范, 涉及技术伦理、数据安全、行业标准等规范体系的建立(见例 10、例 11); 再次, 是 AI 的产业生态, 包括 AI 产业链的构成与布局, 相关的企业及其在全球的竞争与布局; 最后, 是 AI 的应用与赋能, 如应用、领域、赋能、教育、机器人、场景, 展示了 AI 在各行各业的实际运用及其带来的能力提升和影响(见例 12、例 13)。

例 10: 中国积极参与并推动全球人工智能治理规则的制定。2024 年, 第七十八届联合国大会协商一致通过中国主提的“加强人工智能能力建设国际合作”决议…… (人民日报, 2025/03/03)

例 11: 构建人工智能科技伦理治理体系。坚持“以人为本、智能向善”的人工智能发展原则。针对数据安全、算法歧视、隐私保护、知识产权等问题, 加强对人工智能发展的安全…… (人民日报, 2025/03/21)

例 12: 北京海淀人工智能集群能级跃升启航仪式举行, 将建设具有全球影响力的人工智能创新策源地和产业高地。(新京报, 2025/02/17)

例 13: 广州发挥重点场景示范效应, 启动“数据要素×”行动, 广州低空数据开发利用相关工作, 已入选国家重点联系示范场景, 绿色低碳数据产品入选国家“数据要素×”典型案例。(广州日报, 2025/02/17)

总体而言, 中国主流媒体较多报道 AI 在算法优化和计算能力等方面的突破进展, 同时也突显了 AI 领域的动态性、创新性以及将其转化为实际生产力和社会服务动力的紧迫性, 反映了推动 AI 技术落地应用和治理的迫切需求。然而, 媒体话语并非单纯的技术乐观叙事, 而是在肯定 AI 产业价值的同时, 也保留了对技术风险的提示与对规范治理的关注。这种多角度的建构方式, 共同塑造了人工智能在媒体话语中既是技术焦点、又是发展热点、更是战略重点和时代引擎的复合形象。

5. 结语

本研究通过对 AI 新闻语料库进行主题词和搭配分析, 呈现并阐述了中国主流媒体在 2025 年 2 月至 3 月期间关于人工智能的话语建构特征。研究发现, 主题词分析展示出媒体主要关注的三个核心议题: AI 技术创新与发展、AI 产业应用与赋能、AI 经济影响与战略地位。这与当前中国将人工智能上升至国家战略高度的背景相呼应, 显示出媒体对 AI 作为前沿技术, 新兴产业和战略引擎的关注, 进一步的搭配词分析, 揭示了人工智能在新闻报道中常与界定其技术本质、产业范畴和应用场景, 突出其推动和赋能作用, 关注其发展与规范的词语组合, 结合发现, 中国媒体的报道在总体上以积极导向为主, 同时兼及对 AI 技术风险的关注。媒体话语呈现出发展与风险、应用与治理并存的多维特征。通过本研究, 有助于更系统地理解中国主流媒体对 AI 议题的话语建构方式, 并为后续研究提供参考。需要指出的是, 本研究存在若干局限。其一, 在样本时间范围上, 语料采集仅覆盖 2025 年 2 月至 3 月这一较短时段, 主要捕捉 DeepSeek 引发媒体集中报道的特定时期, 研究结论难以直接推及更长时段的话语演变趋势。其二, 在样本代表性上, 语料仅来自国内主流综合性报章, 未纳入地方媒体、新媒体平台及社交媒体话语, 可能存在媒体类型上的偏向。其三, 在话语层面上, 本研究侧重对词汇搭配及其语境的描述, 对不同报章之间可能存在的差异性与复杂性关注不足。后续研究可在延长时间跨度、拓展媒体类型、引入跨媒体对比的基础上, 进一步结合文本层面的深入阐释, 考察 AI 媒体话语中发展话语与风险话语的互动关系。

参考文献

- [1] Brause, S.R., Zeng, J., Schäfer, M.S. and Katzenbach, C. (2023) Media Representations of Artificial Intelligence: Surveying the Field. In: Lindgren, S., Ed., *Handbook of Critical Studies of Artificial Intelligence*, Edward Elgar Publishing, 277-288. <https://doi.org/10.4337/9781803928562.00030>
- [2] Sanguinetti, P. and Palomo, B. (2024) An Alien in the Newsroom: AI Anxiety in European and American Newspapers. *Social Sciences*, 13, Article No. 608. <https://doi.org/10.3390/socsci13110608>
- [3] 张梦, 陈昌凤. 智媒研究综述: 人工智能在新闻业中的应用及其伦理反思[J]. 全球传媒学刊, 2023, 8(1): 63-92.
- [4] 李明德, 邝岩. 大数据与人工智能背景下的网络舆情治理: 作用、风险和路径[J]. 北京工业大学学报(社会科学版), 2021, 21(6): 1-10.
- [5] 穆军芳, 陈紫涵. 趋近化视角下社交媒体关于人工智能报道的话语建构研究[J]. 南京邮电大学学报(社会科学版), 2023, 25(6): 55-64.
- [6] 阿力亚·艾尼, 张正坤. 美国主流媒体冬奥会涉华报道中的中国形象探析——基于语料库辅助的批评话语分析[J]. 上海理工大学学报(社会科学版), 2024, 46(3): 218-225, 232.
- [7] Wodak, R. and Meyer, M. (2016) *Methods of Critical Discourse Studies*. 3rd Edition, Sage.
- [8] Partington, A., Duguid, A. and Taylor, C. (2013) *Patterns and Meanings in Discourse: Theory and Practice in Corpus-assisted Discourse Studies (CADS)*. John Benjamins Publishing Company. <https://doi.org/10.1075/scl.55>
- [9] 许家金. 后经典时代语料库研究方法及其理论启示[J]. 外语教学与研究, 2023, 55(3): 442-455.