

公共数据开放能否促进人工智能创新水平的提升?

——来自政府开放数据平台上线的准自然实验

李星乐

重庆理工大学重庆知识产权学院, 重庆

收稿日期: 2026年4月3日; 录用日期: 2026年5月20日; 发布日期: 2026年5月29日

摘要

公共数据开放是数字经济发展的关键基础性制度安排。本文将我国城市公共数据开放平台上线视为一项准自然实验, 运用多期双重差分模型, 基于2009~2023年全国292个城市的面板数据, 系统考察了公共数据开放对人工智能创新水平的影响及其内在作用机制。研究显示, 公共数据开放显著促进了城市人工智能创新, 其作用机制并非单纯的数据可得性提升, 而是通过降低信息摩擦与构建制度信任, 推动创新要素优化重组。异质性分析表明, 政策效果在知识产权保护完善、经济基础较好的东部地区更为突出。本文从要素配置与协同创新双重视角, 揭示了公共数据开放推动人工智能创新的内在路径, 为构建数据驱动的高质量创新生态系统提供了理论依据与政策参考。

关键词

公共数据开放, 数据要素, 人工智能创新

Can Public Data Openness Enhance Artificial Intelligence Innovation Levels?

—A Quasi-Natural Experiment from the Launch of Government Open Data Platforms

Xingle Li

Chongqing Institute of Intellectual Property, Chongqing University of Technology, Chongqing

Received: April 3, 2026; accepted: May 20, 2026; published: May 29, 2026

Abstract

The opening of public data constitutes a vital institutional provision for digital economic development. Treating the launch of China's urban public data platforms as a quasi-natural experiment, this study employs a multi-period difference-in-differences model. Utilising panel data from 292 cities nationwide spanning 2009~2023, it systematically examines the impact of public data openness on artificial intelligence innovation levels and its underlying mechanisms. Findings reveal that public data openness significantly stimulates urban AI innovation, with its mechanism extending beyond mere data accessibility. Instead, it facilitates the optimisation and reorganisation of innovation factors by reducing information friction and fostering institutional trust. Heterogeneity analysis indicates that policy effects are more pronounced in eastern regions with robust intellectual property protection and stronger economic foundations. From dual perspectives of factor allocation and collaborative innovation, this paper reveals the intrinsic pathways through which public data openness empowers AI innovation, providing theoretical foundations and policy references for constructing data-driven high-quality innovation ecosystems.

Keywords

Public Data Openness, Data Factors, Artificial Intelligence Innovation

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

在数字经济持续深化的背景下，数据已然成为关键生产要素，其在经济社会各领域发展中的重要作用持续增强。近年来，我国积极布局公共数据开放工作，一系列重要政策相继出台，2015年，国务院印发《促进大数据发展行动纲要》¹，明确将“加快政府数据开放共享”作为重要任务，以推动资源整合与治理能力提升。2022年，《中共中央 国务院关于构建数据基础制度更好发挥数据要素作用的意见》²（即“数据二十条”）发布，进一步明确了公共数据授权使用等重要机制。这些政策为地方政府数据开放提供了制度依据与行动指引。自2012年起，各地方政府积极推进数据资源的开放和共享，推出了公共数据开放平台，这些平台不仅集成了海量的政务数据，也吸纳了来源多样的社会数据，为数据的深度开发与融合利用奠定了重要基础，截至2024年7月，我国已有243个省级和城市的地方政府上线了数据开放平台，全国已建成省级平台24个（不含直辖市和港澳台），城市级平台219个（含直辖市、副省级及地级行政区）^[1]，初步形成了多层次、广覆盖的公共数据开放体系。从内容看，公共数据开放内容广泛渗透至社会生活的多个维度，包括生活服务、教育文化、城建住房、工业农业、财税金融等关键领域，长期以来处于沉睡状态的公共数据价值正被逐步唤醒，其作为生产要素的潜能开始得到有效释放与充分涌流。现有研究表明，公共数据开放能够通过资源补给^[2]与动力激发^[3]双重机制，显著促进市场主体的创新行为。一方面，公共数据为企业发展提供了关键的数据要素和研发基础；另一方面，也激发了其探索新技术、新应用的内在动力。目前学界已经意识到了公共数据开放在推动社会发展中的重要作用，但较少深入微观产业层面进行探讨。作为高度依赖数据驱动的关键领域，人工智能创新过程与公共数据的供给质量密切

¹参见：国务院关于促进印发大数据发展行动纲要的通知，<https://www.mmac.org.cn/UploadFile/UserWeb/laws/223.pdf>.

²https://www.gov.cn/zhengce/2022-12/19/content_5732695.htm

相关, 现有研究较多关注人工智能创新所产生的后端影响, 而亦有学者指出政府数据治理能提供覆盖广泛、权威性高的数据“燃料”, 从而激发人工智能领域创新活力[4]。有研究表明, 数据要素促进人工智能产业技术创新与规模扩张[5]。与此同时, 为确保此类高质量数据的可持续供给, 构建激励性与规范性的数据治理与法律规制框架[6], 已成为学界讨论的前沿议题。

选择该领域, 不仅为检验公共数据开放这一新兴制度供给对前沿技术的直接效应提供了理想场景, 更有助于揭开当前人工智能创新热潮背后的关键驱动因素, 深化了对人工智能发展动因与路径的理解。公共数据的开放能否实质促进人工智能领域的创新水平? 如果是, 那么其背后的驱动机制又是什么? 基于这些考量, 本文在公共数据开放这一政策性框架下, 重点考察了城市公共数据开放平台对人工智能创新的影响, 并且尝试从信息壁垒、协同创新等角度给与揭示。研究结论既可以为分级、分类推进公共数据平台建设提供参考, 也为理解数据如何促进创新增添来自人工智能领域的视角。

2. 理论分析与研究假设

(一) 公共数据开放对人工智能创新的影响

数据资源是驱动人工智能发展的基础性战略要素, 其规模与质量直接决定了生成式大模型等前沿技术的训练效果与迭代上限。在我国, 庞大的人口与市场规模铸就了数据资源在体量与多样性上的显著优势, 然而, 数据获取渠道不畅与流通壁垒却使大量高价值数据处于“沉睡”状态, 未能转化为现实生产力。在此背景下, 推动公共数据开放成为破解数据困境、激活创新源流的重要举措, 其将原本分散于政府各部门的静态信息, 转化为可被社会广泛获取和使用的动态资源, 从而为人工智能创新提供基础性的数据支撑。从资源补给机制来看, 公共数据开放可从供给端解决人工智能创新“用什么”的问题。从性质上看, 公共数据主要涵盖政府业务数据、社情民情数据以及基础环境数据等类别。若企业独立从事此类数据的采集与整理, 将面临高成本与高技术门槛的约束[7], 一方面, 基础环境数据等生产要素需依托物联网感知层设备及相应的云计算资源支持, 初始投入规模较大, 普通企业难以独立承担。另一方面, 公共数据具有衍生性和场景依附性, 通常伴随公共管理、服务及运营活动产生, 主要由国家权威机关、事业单位及其他公共组织负责生成与管理, 来源具有制度赋予的天然垄断性。因此, 政府作为关键的数据供给方, 通过系统性地开放与共享数据资源, 能够为企业提供稳定的数据来源。从动力机制来看, 该机制主要回应需求端“为何创新”的问题。从信息经济学视角出发, 经济运行中的效率损耗主要归因于信息摩擦。现实中, 信息并非完全流通, 往往呈现不对称与不完全分布状态, 其获取、甄别与利用均需付出显著成本, 数据作为非竞争性、非损耗性的生产要素, 其边际成本趋近于零, 边际收益则呈现递增特性, 在数字经济的长期发展中容易形成高度集中的市场结构, 甚至引发“赢家通吃”的局面。而随着公共数据应用频率的提升与应用场景的扩展, 企业参与数据创新的意愿增强, 竞争主体增多。这一变化一方面丰富了数据利用的方案与范式, 另一方面也降低了新进入者在不确定情境下的战略判断压力[8]。同时, 数据要素的深入应用能够优化组织内部的信息传递与决策机制, 缓解了因有限理性和机会主义行为带来的效率损失, 进而提升了创新资源的配置效率。

基于此, 本文提出如下假设:

H1: 公共数据开放推动城市人工智能创新。

(二) 创新创业活力的中介作用

高水平的公共数据开放向市场传递了明确的支持信号与可信的承诺, 它不仅是政府提升透明度的举措, 更是一种构建创新友好型制度环境的关键行动。这种由制度预期激励所激发的创新创业活力, 直接扩大了人工智能创新的主体规模与实验多样性, 从而成为孕育突破性成果的土壤。具体而言, 首先, 数字经济发展降低了人工智能企业创新底基的难度和成本, 公共数据开放通过提供免费、结构化的数据, 直接降低了

企业和相关研发机构在数据获取、清洗、验证环节的成本与不确定性，这种资源补给效应为人工智能模型的训练、学习和迭代提供了不可或缺的“燃料”，从而直接促进了更多人工智能技术方案的产生与专业化。其次，新进入者常利用新的技术或商业模式从边缘市场切入并最终颠覆现有格局。城市的公共数据开放，实质上是为本地人工智能初创企业和技术团队进行了一次关键的数据资本注入，使其能够绕过初始的数据壁垒，将有限资源集中于核心算法研发与商业探索上。对于初创企业和中小型科技公司，公共数据开放使他们能够获得与行业巨头相近的数据资源，从而开发出更具创新性、专注利基市场的人工智能应用。这种“鲶鱼效应”激发了两类主体的竞争动力：一是新进入者为了生存和扩张而积极进行专利布局；二是现有巨头为维持优势而加大研发投入以应对挑战。双方竞争的加剧整体上拉高了行业的创新节奏和创新产出。最后，质性知识的碰撞与重组是创新的重要源泉。公共数据开放为大学、科研院所、企业、个人开发者等提供了共同的数据实验场，催生了基于数据应用的联合研发、技术合作与人才交流。这种互动形成了密集的协同创新网络。该网络促进了隐性知识的传播和显性技术的融合，催生了更多复杂、前沿、复合型的人工智能解决方案，最终体现为联合申请专利数量的增加和专利质量的提升。

H2：创新创业活力在公共数据开放与城市人工智能创新之间起中介作用。

(三) 要素优化配置的中介作用

信息经济学理论揭示，信息不对称是阻碍要素高效配置与流动的关键因素[9]。在此理论视角下，公共数据开放可视为一项重要的制度安排，它通过提供规范、可信、低成本的公共信息资源，有效缓解了市场主体之间的信息不对称。第一，公共数据开放通过揭示高价值的本地化应用场景，引导创新资源进行定向汇聚与精准配置。已有研究指出，公共数据有助于企业识别市场机会、优化决策[10]。公共数据天然根植于本地城市治理与服务的具体情境，其开放实质上是將一座城市的“创新需求清单”进行数据化、透明化呈现。这有助于本地人工智能人才、风险资本与企业研发力量，从追逐宽泛的技术概念，转向聚焦于解决本地的真实痛点上来。这种基于真实场景的需求牵引，提高了创新资源的配置效率，也为人工智能技术的验证、迭代与商业化提供了明确的市场入口，从而加速创新成果的转化。第二，公共数据的规范性与权威性，能有效增强企业间的认知信任与制度信任，为产学研等多方协作创造可预期的制度环境[11]。依据协同创新理论，创新的涌现依赖于不同系统间资源的深度整合与协同互动。公共数据开放作为关键的制度供给，为数字技术与实体经济的深度融合提供高质量的数据要素。具体而言，当制造业、服务业等实体部门获取并利用政府开放的公共数据时，能够有效优化生产流程、开发智能产品与服务，从而为人工智能技术的产业化落地创造丰富的应用场景与明确的市场需求。这一数据驱动形成的“数实技术融合”生态，正是协同创新所强调的跨系统资源整合过程，它通过拓展应用边界与提升商业化潜力，显著增强了人工智能创新的规模效应与实际效能。

H3：要素优化配置在公共数据开放与城市人工智能创新之间起中介作用。

3. 研究设计

(一) 模型设定

为有效识别政策实施的真实效应，本文运用双重差分模型，对样本在公共数据平台上线前后的时间变化与个体异质性加以控制，以分离政策实施带来的影响。基于上述思路，设定的回归模型如式(1)所示：

$$Innovation_{it} = \beta_0 + \beta_1 DATA_{it} + \beta_2 CONTROLS_{it} + \delta_t + \delta_i + \varepsilon_{it} \quad (1)$$

解释变量 $DATA_{it}$ 指识别公共数据平台上线政策的虚拟变量，上线当年及之后取 1，否则取 0。为更准确地估计政策效应，模型中加入了控制变量及固定效应，以涵盖其他影响因素。考虑到同一城市内样本随机扰动项可能潜在相关，本文将标准误聚类到城市层面。系数 β_1 为衡量了公共数据开放的政策效果，预期其显著为正，表明该政策对区域人工智能创新具有促进作用。

(二) 数据来源

本文使用的数据来源主要为以下三个方面：其一，人工智能创新。本文采用人工智能发明专利授权数来衡量人工智能创新，参考姚加权等(2024) [12]的研究，构建人工智能词典库，利用该词典，检索专利的标题和摘要文本中是否涉及相关关键词，以此测度各地区在观测年份内的人工智能专利产出，并以发明专利数加 1 的自然对数衡量人工智能创新水平；其二，公共数据开放。本文对城市公共数据开放平台网址进行人工测验与采集，并借鉴《2024 年中国地方公共数据开放利用报告(城市)》³，对初步筛选得到的平台数据进行交叉验证，最终确定截至 2023 年底，全国共建成 189 个地市级公共数据开放平台；其三，城市层面的控制变量主要取自 2009~2023 年《中国城市统计年鉴》，对其中部分缺失数据，本文结合各省统计局发布的相应数据予以补充。倘若仍然存在缺失值则采用线性插值法补充。此外，对核心变量观测缺失的城市样本进行剔除，最终获得我国 292 个地级市面板观测值作为本文研究样本。

(三) 变量定义

1) 被解释变量

人工智能创新。专利作为创新活动的成果载体，常被视为观测创新行为、衡量创新水平的关键代理指标[13]。在各类专利中，发明专利的授权须通过严格的新颖性、创造性与实用性审查，就表征效果而言，能更为稳健地反映一个地区在人工智能领域所实现的实质性技术进步与知识积累。因此，本文聚焦于发明专利，以更准确地捕捉人工智能领域的核心创新产出。

2) 核心解释变量

公共数据开放。各政府相继推动公共数据开放平台上线，为研究数据开放的效能构建了一个理想的准自然实验环境。为更全面考察公共数据开放的实际效果，本研究在进一步分析中引入了公共数据开放质量维度，着重探讨公共数据开放质量对人工智能创新的影响机制。

3) 机制变量

(1) 数字企业创业集聚。指在一定时空范围内，从事人工智能及相关数字技术研发、应用与服务的新创企业，在特定城市或区域内集中出现、协同共生并形成规模效应的现象。该变量反映了公共数据开放对微观创业活动的催化效果。本文将国家统计局《国民经济行业分类》⁴中的“信息传输、软件和信息技术服务业”企业界定为数字企业。该行业直接对应信息技术开发、软件服务和数据处理等核心数字经济活动，是衡量地区数字经济发展水平最常用且最具代表性的行业类别。在此基础上，本研究利用各城市工商企业注册信息，整理得到每年新设立的数字企业数量，以此测度城市层面数字经济主体的集聚程度。

(2) 要素资源优化配置。指在公共数据开放的推动下，区域内的传统要素与新兴数据要素得以在产业、部门间更高效地流动、匹配与组合，从而实现整体创新效率提升的状态。该变量进一步解构为两个子维度。第一，传统要素流动增强。本文采用城市人均邮政业务收入作为传统要素流通水平的代理变量，该指标综合反映区域内物流、资金流与信息流等方面的流通与整合能力。第二，数实技术融合深化。为衡量数实技术融合程度，本文将企业实体技术创新中对数字技术的吸收、应用与再造能力作为数实技术融合的测度标准。

4) 控制变量

本文结合现有研究和理论分析，选取了如下控制变量：(1) 经济发展规模，采用 GDP 水平进行衡量，并以对数形式回归；(2) 产业结构，使用第三产业与第二产业比重度量；(3) 城市金融发展水平，以年末

³参见：2024 年中国地方公共数据开放利用报告(城市)，

[http://ifopendata.cn/static-report/2024%E4%B8%AD%E5%9B%BD%E5%9C%B0%E6%96%B9%E5%85%AC%E5%85%B1%E6%95%B0%E6%8D%AE%E5%BC%80%E6%94%BE%E5%88%A9%E7%94%A8%E6%8A%A5%E5%91%8A_%E5%9F%8E%E5%B8%82.pdf](http://ifopendata.cn/static/report/2024%E4%B8%AD%E5%9B%BD%E5%9C%B0%E6%96%B9%E5%85%AC%E5%85%B1%E6%95%B0%E6%8D%AE%E5%BC%80%E6%94%BE%E5%88%A9%E7%94%A8%E6%8A%A5%E5%91%8A_%E5%9F%8E%E5%B8%82.pdf)

⁴参见：国民经济行业分类，<https://www.stats.gov.cn/sj/tjbz/gmjjhyfl/202501/P020250116506795831658.pdf>

金融机构贷款余额与当地 GDP 的比值衡量；(4) 财政压力水平。采用各地方财政一般预算内支出占地方财政一般预算内收入比重量化；(5) 科研支持力度，直接体现政府对研发活动的财政投入强度，采用城市科学事业费支出占财政支出比重度量；(6) 教育发展水平，能够用来衡量城市人力资本储备与基础研究能力，为创新提供人才与知识源泉，以城市内高等学校数量来衡量。

(四) 描述性统计

在对变量进行了定义之后，表 1 汇报了各变量的描述性统计信息，从表 1 的统计结果来看，核心被解释变量城市人工智能专利创新的均值为 1.330，中位数为 0.693，部分城市并没有人工智能专利产出，这表明当前我国城市间人工智能创新水平存在较大差异。核心解释变量公共数据开放的均值为 0.224，说明在样本期内，有 22.4%的城市处于公共数据开放平台已上线状态。

Table 1. Descriptive statistics

表 1. 描述性统计

变量	样本量	均值	标准差	最小值	中位数	最大值
Innovation	3,663	1.330	1.672	0.000	0.693	9.420
Data	3,663	0.224	0.417	0.000	0.000	1.000
LnGDP	3,663	16.759	0.925	14.312	16.678	19.973
InduStr	3,663	1.122	0.622	0.114	0.987	6.387
Fina	3,663	1.133	0.702	0.118	0.948	12.817
Pre	3,663	2.943	2.067	0.649	2.323	26.541
Sci	3,663	0.019	0.020	0.001	0.013	0.276
Uni	3,663	9.927	15.779	0.000	4.000	93.000

4. 实证分析

(一) 基准估计

根据式(1)的基准回归结果见表 2。表 2 第(1)列汇报了仅包含核心解释变量的基准回归结果，此时，公共数据开放(Data)的系数在 1%的水平上显著为正，初步支持了公共数据开放对城市人工智能创新的促进作用。列(2)进一步纳入了城市层面的控制变量以及城市与年份的双向固定效应，结果显示，核心变量 Data 的系数估计值仍保持显著且符号为正，公共数据开放推进人工智能创新提升 29.9%。

Table 2. Benchmark results: public data openness and urban artificial intelligence innovation

表 2. 基准结论：公共数据开放与城市人工智能创新

	(1)	(2)
	Innovation	Innovation
Data	0.395*** (0.071)	0.299*** (0.069)
LnGDP		0.532*** (0.162)
InduStr		-0.084 (0.097)
Fina		0.222** (0.091)

续表

Pre		-0.022 (0.029)
Sci		5.514*** (1.551)
Uni		0.056** (0.023)
_cons	1.241*** (0.016)	-8.408*** (2.707)
样本量	3663	3663
城市固定	YES	YES
年份固定	YES	YES
调整 R ²	0.858	0.869

注：*、**、***分别表示在 10%、5%、1%的水平上显著。

(二) 平行趋势检验

本文参考事件分析法[14]进行平行趋势检验，设定如下估计模型：

$$Innovation_{it} = \beta_0 + \beta_1 DATA_{it}^{-7} + \dots + \beta_{15} DATA_{it}^7 + \beta_2 CONTROLS_{it} + \delta_i + \delta_t + \delta_i t + \varepsilon_{it} \quad (2)$$

$DATA_{it}^n$ 为相应年份对应的 $Treat_i \times Post_t$ 代理变量。为降低随机误差的干扰，更清晰地捕捉政策效应的总体趋势和持续性，本文参照黄先海与虞柳明[15]的做法，对政策时点进行归并处理：将政策实施前第七期及更早时期归并为“前七期”(Pre_-7)，将政策实施后第七期及之后的时期归并为“后七期”(Post_+7)，并以-7期为基期进行平行趋势检验。平行趋势检验结果见图 1，政策效应在实施后显著为正且持续增强，证明了基准结果的稳健性。值得注意的是，政策效应的动态检验结果显示，在政策实施后的第六期，促进效应出现回落。这可能与政策推行的阶段性特征有关：政策初期，公共数据开放平台建设数量有限，政策推力相对较弱，其所带来的创新激励作用较为有限；而至 2018 年后，平台建设数量实现倍增，政策覆盖范围迅速扩大，政策信号与数据要素的扩散强度显著提升，系统性地提升了人工智能创新的整体水平。

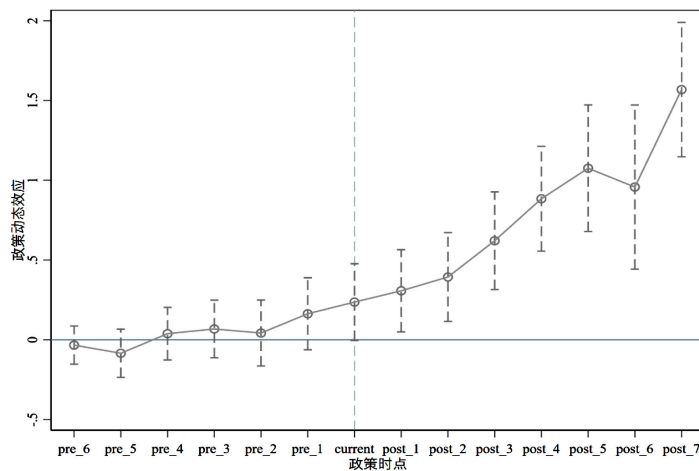


Figure 1. Parallel trend test
图 1. 平行趋势检验

(三) 倾向得分匹配模型

本文进一步采用多时点倾向得分匹配双重差分法(PSM-DID)缓解可能由样本选择偏差导致的内生性问题。以前述控制变量作为匹配协变量,使用最近邻匹配(1:4)为处理组城市构造特征相似的对照组,以在保留样本信息与控制匹配偏误之间取得平衡。基于匹配后的样本重新进行估计,结果见表 3,可以看出,核心解释变量的系数依然显著为正,表明公共数据开放对人工智能创新的促进效应并非源于选择性偏差,基准结论具备良好的稳健性。

Table 3. Regression results of propensity score matching test

表 3. 倾向得分匹配检验回归结果

	(1)
	Innovation
Data	0.171** (0.080)
LnGDP	0.322* (0.179)
InduStr	-0.056 (0.101)
Fina	0.198** (0.087)
Pre	-0.038 (0.033)
Sci	5.213*** (1.430)
Uni	0.051* (0.028)
_cons	-4.846 (2.997)
样本量	2792
城市固定	YES
年份固定	YES
调整R ²	0.851

注: *、**、***分别表示在 10%、5%、1%的水平上显著。

(四) 安慰剂检验

为检验基本结论是否受到某些未被观测到的、同时影响政策实施和城市创新的随机因素的干扰。本文采取随机化分组的方式进行安慰剂检验,重复进行 500 次随机抽样与回归估计,最终得到 500 个伪政策冲击变量的估计系数。图 2 显示,安慰剂检验中的核心解释变量估计系数与表 3 中基准回归结果中列(2)的系数具有显著差异,说明未观测变量对城市人工智能创新几乎没有影响。根据安慰剂检验结果,可以认为基准结论并不受到遗漏变量等因素的偏误影响,进而证实城市政府数据开放对城市人工智能创新水平具有显著的提升效应。

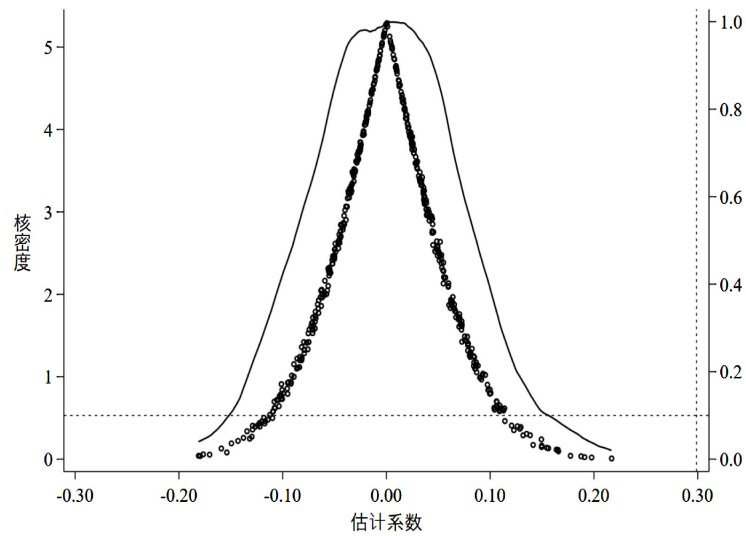


Figure 2. Placebo test of artificial intelligence innovation
图 2. 城市人工智能创新的安慰剂检验

(五) 异质性处理效应诊断及检验

考虑到传统双向固定效应模型在处理多时点政策冲击时可能因异质性处理效应导致估计偏误[16] (Goodman-Bacon, 2021), 本文进一步对基准结果的稳健性进行诊断。本文通过排除异质性处理效应的稳健估计量方法进行检验, 具体通过 Borusyak 等(2024) [17]提供的一种基于插补的反事实方法解决 TWFE 的估计偏误问题。结果如图 3 所示, 在政府公共数据开放平台实施前, 处理组与对照组之间未呈现显著差异; 平台上线后, 政策对人工智能创新的促进效应依然显著。由检验结果可知, 异质性处理效应并未对基准结论构成实质性干扰, 表明本文的研究结论具有较好的稳健性。

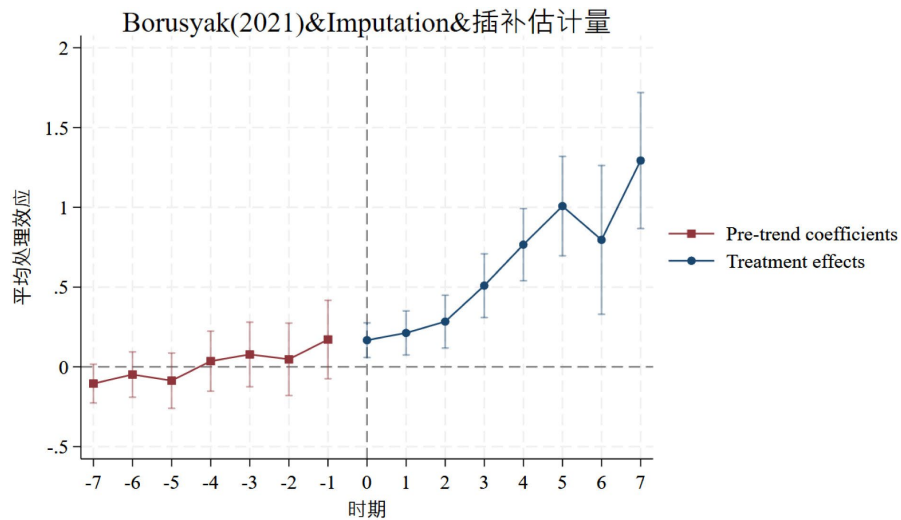


Figure 3. Imputation estimator plot
图 3. 插补估计量图

(六) 替换被解释变量

替换城市人工智能创新衡量指标。依据国家知识产权局于 2023 年 9 月公开的《关键数字技术专利分

类体系(2023)》⁵中“人工智能技术专利分类体系表”模块，匹配出主分类号符合文件标准的相关人工智能专利，由此统计各省市原人工智能发明专利数据，回归结果如表4所示，列(2)中解释变量的估计系数显著为正，表明本文的研究结论具有较好的稳健性。

Table 4. Regression results with replaced dependent variable
表 4. 替换被解释变量回归结果

	(1)	(2)
	Innovation	Innovation
Data	0.225*** (0.052)	0.165*** (0.050)
LnGDP		0.531*** (0.132)
InduStr		-0.036 (0.078)
Fina		0.074** (0.030)
Pre		0.022 (0.030)
Sci		2.169** (0.911)
Uni		0.018*** (0.006)
_cons	2.791*** (0.012)	-6.430*** (2.262)
样本量	3659	3659
城市固定	YES	YES
年份固定	YES	YES
调整R ²	0.939	0.941

注：*、**、***分别表示在 10%、5%、1%的水平上显著。

(七) 排除同期政策干扰

数字经济政策冲击是当前热点议题，实行公共数据开放政策的试点城市往往在同一时期受到类似导向的政策叠加影响，为准确识别公共数据开放的效应，本文系统梳理了近年来的相关数字经济政策，并在实证中逐一控制这些政策的潜在混杂影响。

1) 智慧城市试点政策。智慧城市试点政策是我国推动城市数字化转型与治理现代化的一项国家级重大战略，智慧城市试点可能对人工智能创新产生影响，表5第(1)列结果显示，在控制该政策因素后，公共数据开放的政策变量估计系数依然显著为正，而智慧城市试点的系数未通过显著性检验，表明公共数

⁵参见：关键数字技术专利分类体系(2023)，https://www.cnipa.gov.cn/art/2023/9/25/art_75_187769.html

据开放对人工智能创新的促进作用并未受到智慧城市政策的干扰。

2) 宽带中国试点政策。鉴于网络基础设施建设的重要作用, 2014年至2016年间, 工信部与国家发改委分三批共设立了120个“宽带中国”示范城市, 旨在以高速网络推动经济高质量发展, 其政策目标与公共数据开放具有一定相似性。表5第(2)列结果显示, 在纳入“宽带中国”政策变量后, 公共数据开放变量的系数仍然显著为正。

3) 国家大数据综合试验区试点。该试验区的设立旨在系统推进大数据产业发展与应用。表5第(3)列结果表明, 在控制该试点政策的影响后, 公共数据开放变量的系数依然呈现显著正向效应。

在上述基础上, 本文进一步将三类政策虚拟变量与公共数据开放的虚拟变量一并纳入模型进行回归, 表5第(4)列展示了对应的估计结果。在控制多项相关政策的潜在影响之后, 公共数据开放的系数依然在统计上显著为正。这表明, 公共数据开放对城市人工智能创新水平具有独立于其他数字经济发展政策的正向影响, 从而进一步验证了基准回归结论的稳健性。

Table 5. Regression results excluding interference from relevant policies

表 5. 排除相关政策干扰回归结果

变量	(1) Innovation	(2) Innovation	(3) Innovation	(4) Innovation
Data	0.300*** (0.069)	0.288*** (0.068)	0.302*** (0.069)	0.293*** (0.069)
智慧城市	0.064 (0.102)			0.056 (0.100)
宽带中国		0.337*** (0.085)		0.341*** (0.085)
国家大数据综合试验区			0.094 (0.102)	0.116 (0.097)
_cons	-8.297*** (2.724)	-8.498*** (2.636)	-8.555*** (2.714)	-8.583*** (2.659)
控制变量	Yes	YES	YES	YES
样本量	3663	3663	3663	3663
城市固定	YES	YES	YES	YES
年份固定	YES	YES	YES	YES
调整 R ²	0.869	0.871	0.869	0.871

注: *, **, ***分别表示在 10%、5%、1%的水平上显著。

(八) 异质性分析

1) 区域异质性

由于受到自然地理环境和社会经济基础等多方面影响因素的制约, 经济发展在空间维度上往往是非均质、不平衡的, 因此公共数据开放对我国非均质发展的地理空间模块也可能具有差异性影响。东部地区算力基础设施完善, 数据要素市场较为成熟, 为公共数据的高效存储、调用与分析奠定了技术基础, 地方政府可依托成熟市场机制, 选择优质供应商承担平台开发运维, 在稳定性、更新频率与接口可用性

上形成质量优势。此外东部地区产业结构层次高，数字技术应用场景丰富，企业自身在数据吸收与转化方面具备更强的动态能力，能够更快将公共数据资源转化为实质性的人工智能创新产出；而非东部地区在资源禀赋、发展阶段与技术承接能力上则面临不同的挑战，数据开放的效应可能受到数字化基础、人才储备等条件的约束。基于此，本文按照东部和非东部地区划分样本，非东部地区包括中部和西部地区，并采用分组回归的方式考察公共数据开放对城市人工智能创新水平的影响。表 6 第(1)列对应东部地区，公共数据开放回归系数为 0.222，且显著为正，表 6 第(2)列对应非东部地区，公共数据开放回归系数为 0.149，且显著为正，这说明公共数据开放对东部地区与非东部地区人工智能创新均产生明显的促进作用，使用 Chow 检验来检验分组回归后的组件系数差异，p 值为 0.037，证实公共数据开放在促进人工智能创新水平方面存在明显的区域异质性，其对东部地区的边际贡献显著更大。

Table 6. Regional heterogeneity analysis

表 6. 区域异质性分析

变量	(1)	(2)
	Innovation	Innovation
	东部	非东部
Data	0.222** (0.099)	0.149* (0.080)
_cons	-18.198*** (3.471)	-7.760** (3.656)
控制变量	YES	YES
观测值	1395	2268
城市固定	YES	YES
年份固定	YES	YES
调整R ²	0.911	0.838
系数差异p值	0.037	

注：*、**、***分别表示在 10%、5%、1%的水平上显著。

2) 知识产权保护

专利所有者在其权利受到侵犯时能得到的司法保护强度是影响专利创新的重要方面，一个城市知识产权案件的胜诉率较高，不仅反映了当地司法系统在处理知识产权纠纷中的专业性与公正性，更能够传递出对知识产权高度尊重与有效保护的制度信号。这种制度环境营造了良好的创新成果保护氛围，降低了企业在创新过程中面临的技术溢出风险。在此背景下，企业更愿意将创新成果以专利等形式固定下来，形成“以公开换保护”的良性循环：即企业通过申请专利获得法律保护，而强有力的司法保障又进一步激励企业加大合作创新投入。因此本文采用各城市知识产权案件中的胜诉案件数量占该城市总案件数量的比例，作为该城市知识产权保护强度(IPP)的代理指标。参考易巍的(2023)的研究[18]，利用裁判文书网公布的知识产权类案件的判决书，获取案件的原被告双方所在地以及判决结果，被侵权方胜诉的定义为被告赔偿或停止侵权。回归结果如表 7 所示，公共数据开放对两组城市的人工智能创新能力均呈现正向显著影响，使用 Chow 检验来检验分组回归后的组间系数差异，P 值为 0.001，显著拒绝了原假设，说明分组回归的系数在两组之间存在显著差异。

Table 7. Heterogeneity analysis of intellectual property protection
表 7. 知识产权保护异质性分析

变量	(1)	(2)
	Innovation 司法保护强	Innovation 司法保护弱
Data	0.366** (0.088)	0.217*** (0.079)
_cons	-5.904* (3.115)	-6.373** (3.145)
控制变量	YES	YES
观测值	2048	1594
城市固定	YES	YES
年份固定	YES	YES
调整R ²	0.866	0.896
系数差异p值	0.001	

注：*、**、***分别表示在 10%、5%、1%的水平上显著。

5. 进一步分析

(一) 公共数据开放如何影响城市人工智能创新

1) 数字企业集聚创新的驱动机制

本文参照孙伟增等(2023) [19]的做法, 通过以下方式测度数字企业集聚水平: 依据行业分类, 将类别属于信息传输、软件和信息技术服务业的企业视为数字企业。继而基于工商注册数据, 统计各城市历年新增数字企业数量, 并以该数据表征城市数字经济的集聚发展程度。根据江艇(2022) [20]所倡导的因果机制分析框架, 本文采取两步法进行实证检验。表 8 列(1)报告了公共数据开放对数字企业集聚程度的实证结果。结果显示, 政策实施后当地新注册的数字企业数量显著增加, 且数字企业的创业集聚能够促进人工智能创新水平的提升。从理论角度出发, 数字企业的空间集聚有助于促进知识溢出与经验共享, 从而激励企业更加注重开展高质量的数字创新活动, 此种空间集聚天然形成了一个高密度的区域信息与创新网络, 地理邻近性极大地便利了隐性知识的交换与传播[21]。这种知识往往难以通过远距离通信有效传递, 却可通过企业员工间的非正式交流、面对面的技术研讨与人才流动, 加速 AI 技术创新扩散, 产生“1 + 1 > 2”的协同效应。

2) 要素配置优化的资源供给效应

区域要素的畅通流动是缩小区域差异以及实现区域协调发展的前提, 本文参考张小宇等(2025) [22]的研究, 以城市人均邮政业务收入代表生产资源要素的流通水平, 该指标能够较好地反映区域内在物流、信息流等方面的整合能力。其次, 为识别公共数据开放是否通过促进数字技术与实体经济的深度融合来促进创新, 本文参考黄先海等(2025) [23]的研究, 构建“数实技术融合”的中介变量, 以实体技术专利每引用一项数字技术专利记为一次数实技术融合事件, 在城市 - 年份维度加总此类融合事件数, 并进行对数化处理, 最终形成测度指标。该指标的本质是知识流动与技术融合的过程, 即传统实体技术领域对数字技术知识的吸收与整合程度, 关注数字技术在非数字领域的渗透与引用, 其与人工智能创新构成了清晰的逻辑链条: 数实技术融合是人工智能创新的前置技术基础。表 8 中列(2)与列(3)分别汇报了公共数据

开放对传统生产要素流动与数实技术融合的影响结果，以及两者对于人工智能创新的作用效果，两组系数估计值均显著为正，表明公共数据开放有助于优化区域要素配置，其作用路径主要体现为两方面：一是提高了传统生产要素的流动与使用效率，二是推动了数字技术与实体经济的系统性融合。有研究发现[24]，传统生产要素与新型生产要素的创新性配置有助于激发区域创新创业活力，基于此可以推断，公共数据开放能够通过优化生产要素配置，促进传统要素与数字要素的协同整合，进而增强城市人工智能创新水平。

Table 8. Mechanism analysis
表 8. 中介检验

变量	(1)		(2)		(3)	
	创业活力	Innovation	人均邮政业务收入	Innovation	数实技术融合	Innovation
Data	0.111** (0.056)		1.321*** (0.414)		2.048** (0.852)	
创业活力		0.241*** (0.035)				
人均邮政业务收入				0.044*** (0.011)		
数实融合						0.019*** (0.005)
_cons	-2.360 (2.422)	-8.910*** (2.712)	-34.755** (14.810)	-7.807*** (2.649)	-162.666*** (59.072)	-6.268** (2.650)
控制变量	YES	YES	YES	YES	YES	YES
观测值	3663	3663	3663	3663	3663	3663
城市固定	YES	YES	YES	YES	YES	YES
年份固定	YES	YES	YES	YES	YES	YES
调整R ²	0.879	0.871	0.568	0.875	0.645	0.876

注：*、**、***分别表示在 10%、5%、1%的水平上显著。

(二) 公共数据开放与城市人工智能创新：数据开放质量的影响

政务在管理公共事务和履行政府职能的过程中积累了大量公共数据，公共数据开放并非单一数据发布行为，而是一个涵盖数据供给、生态培育以及价值实现的系统工程。人工智能创新活动在研发初期依赖高质量、结构化、覆盖多领域的原始数据以进行模型训练与算法优化，在成果转化阶段，需要验证数据应用的可行性与价值，并依赖良好的数据生态获取支持、拓展与合作。因此，仅考虑数据平台是否上线单一指标，无法解释公共数据开放如何作用于人工智能创新链条。为进一步探讨公共数据开放的质对人工智能创新的影响，本文参考“中国开放树林指数”，选取利用层(Digital_use)、数据层(Digital_data)、综合指数(quality)三个维度的评估分数作为衡量平台公共数据开放深度与广度的特征指标。鉴于“中国开放树林指数”自 2017 年起发布，本文选取 2017~2023 年间的指数数据进行分析。其中，利用层指数因自 2018 年才开始公布，故在评估该指标时剔除了 2017 年数据，具体结果见表 9。分析表明，公共数据开放的质量越高、利用越充分、整体开放水平越高，其对人工智能创新的促进作用也越显著。为充分发挥公

共数据开放的价值，需实现数据质量、开放综合水平与利用效果三者的协同支撑。

Table 9. Quality of public data openness
表 9. 公共数据开放质量

	(1)	(2)	(3)
	Innovation	Innovation	Innovation
Quality	0.045*** (0.016)		
Data		0.040* (0.021)	
Use			0.147*** (0.032)
_cons	-4.776 (3.460)	-4.908 (3.510)	-6.308 (3.895)
观测值	2027	2027	1744
城市固定	YES	YES	YES
年份固定	YES	YES	YES
调整R ²	0.919	0.919	0.927

注：*、**、***分别表示在 10%、5%、1%的水平上显著。

6. 结论与建议

本文将各城市公共数据开放平台的正式上线视为一项准自然实验，基于公共数据开放影响人工智能领域创新水平的理论分析，以 292 个地级及以上城市 2009~2023 年的面板数据为研究样本，构建多期双重差分模型，探究公共数据开放对城市人工智能领域创新水平的影响。研究结论如下：(1) 在基准回归分析方面，公共数据开放可以显著促进人工智能领域创新，这一结果在经过稳健性检验之后依然成立。(2) 在中介效应方面，公共数据开放通过“数字企业集聚”与“要素配置优化”两条渠道间接促进人工智能创新。(3) 异质性分析显示，公共数据开放对东部、知识产权保护水平高城市的人工智能领域创新的受益更明显。(4) 公共数据开放质量方面，公共数据开放的数据质量、利用程度、总体开放水平对于人工智能创新产生促进作用。本文的结论对支持地方政府公共数据开放，实现智能创新具有重要的启示，相关政策建议如下：

第一，多措并举持续提升公共数据开放质量与可及性。应持续推进政府及公共机构数据的系统性开放与动态更新，尤其提升数据集的完整性、可用性与标准化水平。可制定分级分类开放目录，优先开放与人工智能研发密切相关的科技、产业、环境等高质量数据集。同时建立覆盖数据采集、治理、开放全过程的质量规范与评估机制，推动跨部门数据互联互通，破解逻辑孤岛。在安全可控前提下，探索公共数据授权运营与可信流通模式，实现“数据可用不可见”，促进高价值数据在合规环境中赋能人工智能创新。

第二，构建数据驱动型区域创新生态。当前，人工智能产业的技术深度与产业发展均处于初级阶段，商业化过程仍将十分漫长，充分借助社会资本，加强与其他企业、高校、研究所等机构及个人的合作，鼓励形成以人工智能为核心的产业集群。具体而言，各地可围绕人工智能产业链的数据层基础需求，规

划并开放具有区域特色与产业针对性的高质量公共数据集。通过提供稳定、可靠、结构化的关键数据资源，吸引算法研发、模型训练、行业应用等上下游企业及专业人才在本地集聚，促进数字企业在空间上的高效集聚。此外，可鼓励政府、企业、高校及研究机构共建数据创新平台。协同开展数据清洗、标注与融合，形成适用于人工智能训练的高质量语料库。通过多元协作，推动公共数据转化为支持创新的关键要素。

第三，因地制宜制定差异化数据开放与产业发展战略。东部及发达地区可扩大高价值、高精度公共数据开放范围，支持其成为人工智能技术策源地；中西部等地区应结合本地产业与资源特点，开放特色公共数据集，推动数据开放与本土产业发展相衔接。此外，可探索建立跨区域“数据开放协同区”，推动数据标准互通、平台互联与场景共建，促进数据要素在更广范围内有序流动。各类区域均需配套加强数据安全、隐私保护与知识产权司法保障，稳定创新主体预期。

基金项目

2024年重庆市研究生科研创新项目“数据知识产权登记实施制度研究”(CYS240725)。

参考文献

- [1] 复旦大学数字与移动治理实验室. 中国地方公共数据开放利用报告——城市(2024年度) [R/OL]. 2024-09-26. http://ifopendata.cn/static/report/2024%E4%B8%AD%E5%9B%BD%E5%9C%B0%E6%96%B9%E5%85%AC%E5%85%B1%E6%95%B0%E6%8D%AE%E5%BC%80%E6%94%BE%E5%88%A9%E7%94%A8%E6%8A%A5%E5%91%8A_%E5%9F%8E%E5%B8%82.pdf, 2026-01-06.
- [2] 沈坤荣, 林剑威. 链“岛”成“陆”: 公共数据开放的技术创新效应研究[J]. 管理世界, 2025, 41(2): 83-104.
- [3] 胡雪萍, 王谊巧. 公共数据开放能否提升城市创新水平——基于有效市场和有为政府的机制分析与对策, 2026, 43(5): 1-12.
- [4] 冀淑辛. 政府数据治理与人工智能企业创业活力[J]. 技术经济与管理研究, 2025(11): 122-127.
- [5] 徐维祥, 石伊凡, 周建平. 数据要素集聚与人工智能产业高质量发展[J]. 财经论丛, 2026, 42(2): 29-40.
- [6] 谢潇, 罗世杰. 人工智能大模型训练数据供给的激励性法律规制[J]. 广东财经大学学报, 2025, 40(6): 110-119.
- [7] Jetzek, T., Avital, M. and Bjorn-Andersen, N. (2014) Data-Driven Innovation through Open Government Data. *Journal of theoretical and applied electronic commerce research*, 9, 15-16. <https://doi.org/10.4067/s0718-18762014000200008>
- [8] 赵斌, 刘军, 王洋洋. 从开放到创新: 政府公共数据开放的企业渐进式创新效应研究[J]. 财会月刊, 2026, 47(3): 105-112. <https://link.cnki.net/urlid/42.1290.F.20260130.1327.028>, 2026-01-31.
- [9] Diamond, P.A. (1981) Mobility Costs, Frictional Unemployment, and Efficiency. *Journal of Political Economy*, 89, 798-812. <https://doi.org/10.1086/261003>
- [10] 乔子轶, 滕曼茹, 戴明禹. 公共数据资源共享能否提升供应链韧性: 基于公共数据开放平台的准自然实验[J]. 世界经济研究, 2025(12): 88-102+133-134.
- [11] 杨志安, 侯耀威, 任珊珊. 数据要素视角下公共数据开放的产学研合作创新效应[J]. 华东经济管理, 2026, 40(1): 32-41.
- [12] 姚加权, 张锬澎, 郭李鹏, 冯绪. 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, 40(2): 101-116+133+117-122.
- [13] 王宸威, 钞小静. 数据要素应用如何影响企业分布式创新?——来自上市公司微观视角下的经验证据[J]. 现代财经(天津财经大学学报), 2024, 44(11): 3-21.
- [14] Beck, T., Levine, R. and Levkov, A. (2010) Big Bad Banks? The Winners and Losers from Bank Deregulation in the United States. *The Journal of Finance*, 65, 1637-1667. <https://doi.org/10.1111/j.1540-6261.2010.01589.x>
- [15] 黄先海, 虞柳明. 政府数据开放能提升城市创新水平吗——来自中国地级市的证据[J]. 浙江学刊, 2023(4): 88-98.
- [16] Goodman-Bacon, A. (2021) Difference-in-Differences with Variation in Treatment Timing. *Journal of Econometrics*, 225, 254-277. <https://doi.org/10.1016/j.jeconom.2021.03.014>
- [17] Borusyak, K., Jaravel, X. and Spiess, J. (2024) Revisiting Event-Study Designs: Robust and Efficient Estimation. *Review*

of Economic Studies, **91**, 3253-3285. <https://doi.org/10.1093/restud/rdae007>

- [18] 易巍, 龙小宁. 行政边界与专利知识传播[J]. 数量经济技术经济研究, 2023, 40(10): 159-180.
- [19] 孙伟增, 毛宁, 兰峰, 等. 政策赋能、数字生态与企业数字化转型——基于国家大数据综合试验区的准自然实验[J]. 中国工业经济, 2023(9): 117-135.
- [20] 江艇. 因果推断经验研究中的中介效应与调节效应[J]. 中国工业经济, 2022(5): 100-120.
- [21] 郑梓若, 李欣欣. 数据要素共享、创业活跃度与城市创新发展[J]. 统计与信息论坛, 2025, 40(10): 102-113.
- [22] 张小宇, 张旻. 数据要素共享与新质生产力——基于公共数据开放的经验证据[J]. 上海经济研究, 2025(8): 29-42.
- [23] 黄先海, 高亚兴. 数字知识存量、数实技术融合与企业实体技术创新边界[J]. 经济学动态, 2025(3): 36-53.
- [24] 姚登宝, 阮倩倩, 施志超. 数实融合对安徽新质生产力的影响: 理论机制与实证检验[J]. 华东经济管理, 2026, 40(4): 9-20.