

公共数据开放与企业人工智能创新

王艳 赵超 吴鸿亮 王邦昊

摘要：本文以地方政府上线公共数据开放平台为准自然实验，运用多时点双重差分模型，考察公共数据开放对企业人工智能创新的影响及作用机制。研究发现，政府公共数据开放显著提升了企业人工智能创新水平。机制分析表明，降低经营管理成本、拓展大数据应用是公共数据开放促进企业人工智能创新的重要渠道。异质性分析表明，公共数据开放对小规模企业、城市发展水平较高企业中人工智能创新的正向影响更显著。本研究证明了政府公共数据开放对企业人工智能创新的积极作用，为激活数据要素潜能、赋能人工智能创新发展提供了政策启示。

关键词：公共数据开放；人工智能创新；经营管理成本；大数据应用

中图分类号：F49；F273.1 **文献标识码：**A **文章编号：**2095-8838 (2026) 02-0040-14

一、引言

在大数据时代，数据已成为国家的基础性战略资源和关键生产要素。作为重要的数据生产和采集主体，政府部门所掌握的海量数据资源不仅为政策的科学制定、政府的数字化转型和治理能力现代化提供支撑，也为公众参与监督、挖掘数据潜在价值创造了有利条件。2012年6月上海市政府数据服务网上线，随后北京、无锡等城市相继跟进。2025年11月复旦大学数字与移动治理实验室发布的《中国地方公共数据开放利用报告》显示，截至2025年7月，我国已上线257个省级和城市政府数据开放平台，实现了省级平台的广泛覆盖，并持续向地市、区县层级深化普及。

现有研究主要从宏观与微观两个层面探索公共数据开放的经济后果。在宏观层面，学者们发现政府公共数据开放能够显著促进经济增长（张晨等，2023）、推动区域协调发展（方锦程等，2023）、提升金融市场运行效率（Du等，2024）。在微观层面，已有研究表明公共数据开放有助于提高企业全要素生产率（彭远怀，2023）、劳动收入份额（卫志民等，2024）与经营绩效（程波辉和罗培锴，2024）。进一步地，公共数据开放也深刻影响着前沿技术领域的发展。一方面，企业能够利用高质量公共数据训练更准确的算法模型，提升产品性能与市场竞争力（杨来峰，2025）；另一方面，跨领域数据的共享与融合催生了更多智能化应用场景与服务模式（于同同等，2024）。此外，公共数据的低成本

收稿日期：2025-08-19

基金项目：国家社科基金重大项目（21ZDA039）；广东外语外贸大学和南方电网能源发展研究院共建数据资产安全与财会监督联合实验室课题；广东省哲学社会科学规划2026年度一般项目“‘两个毫不动摇’促进各种所有制经济优势互补、共同发展研究”

作者简介：王艳，广东外语外贸大学粤港澳大湾区会计与经济发展研究中心/国家安全研究院教授，博士生导师；
赵超，广东外语外贸大学会计学院硕士研究生；
吴鸿亮，通讯作者，南方电网能源发展研究院有限责任公司投资与财务研究所正高级经济师，博士；
王邦昊，西南财经大学金融学院本科生。

本获取也有助于降低企业研发与交易成本,激发其持续开展相关创新活动(郑志强等,2025)。然而,虽然现有文献已考察公共数据开放对微观企业整体创新的影响(陈艳利和蒋琪,2024),但未能深入揭示其作用于人工智能创新的内在逻辑与微观机制。相对整体创新而言,人工智能创新需要企业具备数据处理、算法研发、算力支撑等专用性能力,其本质是数据驱动的技术突破,因此对公共数据开放更为依赖。仅仅考察公共数据开放对企业整体创新的平均处理效应,难以准确评估其对人工智能发展战略的价值,无法提供具有针对性的政策建议。

作为助力数字化转型的核心战略举措之一,政府公共数据开放在挖掘数据资源价值、增强治理效能方面具有不可替代的重要作用,为人工智能技术的创新提供了有力支撑。近年来,我国密集出台多项政策,系统布局人工智能发展。2017年国务院印发《新一代人工智能发展规划》,明确提出构建开放协同的人工智能科技创新体系;2025年8月发布的《国务院关于深入实施“人工智能+”行动的意见》中指出,到2030年,新一代智能终端、智能体等应用普及率超90%,智能经济成为我国经济发展的重要增长极。在此背景下,公共数据开放平台所释放的海量、多维度、高价值的政府数据,恰好为人工智能模型的训练与迭代提供了不可或缺的数据燃料。相较于传统技术创新,人工智能创新的核心驱动力在于数据,这使得公共数据开放对其影响可能更为直接和显著。研究公共数据开放对人工智能创新的驱动作用,不仅关乎特定技术的进步,更关乎如何重塑产业模式、提升全要素生产率乃至重构区域竞争力格局。尽管相关政策勾勒出数据赋能人工智能的宏伟蓝图,但公共数据开放能否以及在何种条件下有效转化为人工智能创新动能,尚缺乏严谨的微观证据支持。因此,本文聚焦人工智能创新,利用政府推动公共数据开放平台上线的准自然实验,构建多时点双重差分模型,旨在精准识别公共数据开放对前沿技术突破的驱动作用,丰富“数据—人工智能”链条的实证研究,为理解数据要素如何赋能新一轮科技革命与产业变革提供关键的微观解释。

相比于已有文献,本文的边际贡献在于:

第一,在研究主题上,本文将政府公共数据开放的影响扩展至人工智能领域,丰富了政府公共数据开放经济后果的相关研究。公共数据的高质量供给对数据要素体系进行了重塑,其经济影响具有重要研究价值。现有文献多聚焦于公共数据开放对微观企业全要素生产率和整体创新的影响(彭远怀,2023;陈艳利和蒋琪,2024),本文则进一步沿人工智能领域进行延伸,探讨地方政府公共数据开放是否以及如何促进企业人工智能技术创新,对公共数据开放经济效应的文献形成了有益补充。

第二,在研究视角上,本文聚焦企业人工智能创新这一特定前沿领域,深化了对数据要素驱动企业前沿技术创新的理论认知。现有研究多关注传统技术创新或整体数字化转型,而本文聚焦人工智能这一具有高复杂度、强数据依赖性和广泛赋能潜力的新兴技术,揭示了数据要素在企业人工智能创新中的独特作用,拓展了企业创新理论在人工智能情境下的应用边界,为企业有效利用公共数据推动人工智能创新提供了经验证据。

第三,在影响机制上,与现有文献多从优化营商环境、打破资源垄断等方面对数据公开影响进行研究不同,本文基于公共数据自身的信息内涵和要素属性,分析并论证了公共数据开放通过降低企业经营管理成本、拓展大数据应用的渠道发挥作用。这不仅可以丰富和深化对公共数据价值生成机制的认识,还可以为政府部门在数字化背景下更好地推动经济发展提供理论依据和政策建议。

二、制度背景、文献综述和研究假设

(一)制度背景

国家战略层面高度重视数据要素的市场化配置与开放共享。2021年发布的《国民经济和社会发展第十四个五年规划和2035年远景目标纲要》与2022年发布的《要素市场化配置综合改革试点总体方案》均明确指出,要加强公共数据开放共享。2025年8月1日起施行的《政务数据共享条例》推动了政务数据共享的标准化、规范化,为包括人工智能在内的各类创新应用提供数据支撑。《数字中国建设2025年行动方案》将公共数据的统筹管理、运营和应用提上日程,进一步为公共数据资源的开发利用奠定了制度基础。在此背景下,全国各级地方政府加速推进公共数据开放平台建设,已实现省、市两级政府数据开放平台的广泛覆盖。

近年来,生成式人工智能技术突破式发展,也促使人们深刻思考如何构建有利于人工智能产业取得创新应用的政策环境。全球多个国家和地区纷纷出台相关政策,推动公共数据的开放和利用,以加速人工智能等前沿技术的创新与发展,例如欧盟的《人工智能法案》、美国的《人工智能倡议》等。中国政府同样高度重视公共数据开放和人工智能发展,近年来相继发布《新一代人工智能发展规划》《新一代人工智能治理原则——发展负责任的人工智能》等政策文件,为人工智能产业发展提供了有力支持。2025年8月发布的《国务院关于深入实施“人工智能+”行动的意见》更是系统地为人工智能技术赋能千行百业描绘了路线图。公共数据开放与人工智能创新已被置于同一战略框架下协同推进。在人工智能产业的发展过程中,企业亟需数据供给和应用场景等方面的支持,而公共数据的开放可以为企业提供丰富的数据资源和应用场景。

(二)文献综述

大数据时代,公共数据开放已成为学界研究的热点议题,受到国内外学者的广泛关注。学者们主要从公共数据开放的发展现状、政策法规框架、运行保障机制及评估体系构建等维度展开探讨。部分学者采用系统文献分析法,对公共行政与信息领域的相关研究进行整合分析,深入探究了政府公共数据开放的用户体验、数据质量等关键影响因素,并基于窗口理论对平台未来发展趋势作出前瞻性预测(Matheus和Janssen, 2019)。在探讨公共数据开放的经济效应时,现有文献主要从宏观与微观两个层面展开。在宏观层面,研究表明公共数据开放通过优化要素配置效率、降低制度性交易成本等途径,促进区域协调发展(方锦程等, 2023)以及城市群协同发展(杜振华, 2020; 沈坤荣和林剑威, 2025)。同时,数据开放显著激发了城乡居民的创业活力,并对改善收入分配结构产生积极影响(谢尚和赵鹏军, 2024)。在微观层面,相关研究发现公共数据开放通过打破信息孤岛、增强市场透明度等机制,提升企业的全要素生产率(彭远怀, 2023; Wang等, 2023; 李晓等, 2025)。程波辉和罗培锴(2024)的研究发现,政府公共数据开放通过优化区域营商环境提升企业经营能力(Zhou等, 2023)。李荣华等(2025)研究发现,公共数据开放能够通过强化创新基础支撑、提升创新意愿与增强创新动力促进企业数字创新产出。

在人工智能创新效应方面,现有研究从产业角度分析人工智能对产业结构调整、产业链协同和产业技术安全防范等方面的影响(Cheng等, 2019; 沈超和赵正宇, 2025)。谭玉松(2024)的研究表明,人工智能主要通过绿色创新引致新兴环保型产业扩张实现产业生态化转型。吕越等(2024)研究发现,智能化技术有助于企业拓展和优化产业链,保障产业链和供应链的安全稳定。Haefner等(2021)提出,人工智能正在重塑企业创新管理的组织架构,并提供了一个分析框架,阐释了人工智能在取代人类工作方面的潜力。Babina等(2024)的研究则发现,人工智能投资通过推动产品创新,不仅提升了企业的销售额和雇佣水平,还增强了市场估值,从而促进企业市值的增长和超级明星企业的崛起。

总体而言,国内外关于公共数据开放的研究已形成较为丰富的理论成果,然而仍存在明显不足:一是在研究视角上,现有文献多关注企业整体创新和运营效率,未能深入特定前沿技术领域,尤其是作为新一轮科技革命核心驱动力的人工智能技术。二是在机制阐释上,现有研究多从优化营商环境、打破资源垄断等相对间接和宏观的渠道进行解释,未能充分结合人工智能创新的自身规律与数据需求,剖析公共数据开放如何通过降低经营管理成本、拓展大数据应用场景等更为直接的路径赋能人工智能创新。这为本文的研究提供了契机。

(三)理论分析与研究假设

公共数据开放能够系统性降低企业经营管理成本,从而为人工智能创新构建关键的成本优势。首先,根据交易成本理论,企业在构建自有数据资产或从市场获取替代性数据时,面临着高昂的数据搜寻、验证、谈判与整合费用,这些成本构成企业进行人工智能创新的重要壁垒(权小锋和李静蕾, 2025)。公共数据的开放为企业提供了低成本、多维度的数据要素,可减少企业在外数据获取上的大量现金支出,将原本必须资本化或费用化的数据资源开支转化为可自由支配的现金流(彭远怀, 2023)。其次,在信息不对称框架下,人工智能创新与应用往往伴随着合规成本与代理成本,如因训练数据偏见导致的模型歧视会引发潜在的法律诉讼及声誉损失,并且人工智能存在技术黑箱,使得决策不够透明,难以监督和追责。通过引入权威公共数据,一方面能为算法模型提供无偏的基准数据集,有效

辅助模型校验与去偏,降低企业合规风险与相关预期损失;另一方面能使委托人相对便捷地获取可以验证决策的基础数据,而且能够为独立审计机构、行业协会、公众提供监督依据,加强多元共治的约束机制。最后,可被人工智能模型直接调用的公共数据能够赋能企业优化生产、供应链及营销流程(郑志强等,2025),实现运营自动化与智能化,降低传统业务流程中的人力、物力和时间等经营成本。上述经营管理成本的节约,能够显著提升人工智能创新项目的预期净现值,进而激励企业在人工智能领域的持续性研发投入与探索性创新。

公共数据开放能够促进不同领域之间的数据共享和融合应用,推动人工智能技术创新和应用。企业利用政府开放的公共数据并整合自身掌握的数据,基于人工智能技术挖掘数据的潜在价值,拓展大数据应用场景、优化现有资本配置效率(Nikiforova和Lnenicka,2021),可以实现业务、产品、模式等创新,不断推动企业发展。首先,在数据广度上,政府开放的公共数据往往覆盖民生、经济、环境等众多领域,且多为企业内部难以获取的数据维度。将这些外部公共数据与企业内部数据进行融合与碰撞,能够催生单一数据源无法实现的新型分析视角和人工智能应用场景,这种数据化学反应直接拓展了人工智能创新的可能性边界。其次,在应用深度上,公共数据开放本身即向社会展示了数据驱动的巨大潜力,发挥了应用示范与市场教育的作用,通过呈现真实的数据集和应用案例,降低了企业对数据应用价值的认知模糊性,启发并激励企业探索更复杂、更深层次的人工智能解决方案。同时,基于公共数据的成功开发,人工智能应用也能形成正向反馈,验证技术路径,增强企业加大投入、深化应用的信心。

基于此,提出本文的核心研究假设。

研究假设:公共数据开放会显著促进人工智能创新。

三、研究设计

(一)样本选择与数据来源

本文选取2009~2022年沪深A股上市公司为初始研究样本,并对初始样本进行如下处理:(1)剔除金融行业的公司样本;(2)剔除ST类公司样本;(3)剔除主要变量数据缺失的样本。经上述处理,最终得到32146个公司年度观测值。此外,对所有连续变量均进行上下1%的缩尾处理。政府公共数据开放平台上线时间及平台特征数据来源于《中国地方公共数据开放利用报告》,并通过手工方式查找核对。公司财务数据和治理数据来源于CSMAR数据库,城市层面数据来源于《中国城市统计年鉴》。

(二)模型设计与变量定义

1. 模型设定

本文以各地级市分批上线政府公共数据开放平台为准自然实验,将所在城市上线该平台的企业设定为处理组,其余企业设定为控制组,采用多时点双重差分模型来研究政府公共数据开放对人工智能创新的影响。具体回归模型设定如下:

$$AIpatent_{i,t} = \alpha_0 + \alpha_1 did_{i,t} + \gamma Controls_{i,t} + \eta_i + \tau_t + \varepsilon_{i,t} \quad (1)$$

其中,被解释变量 $AIpatent_{i,t}$ 为企业 i 在 t 年的人工智能创新水平。该指标越大,表明人工智能技术创新水平越高。解释变量 $did_{i,t}$ 为公共数据开放平台上线的虚拟变量,当企业 i 所在城市在 t 年上线公共数据开放平台,上线当年及以后年份取1,否则取0。 $Controls_{i,t}$ 为控制变量。 η_i 和 τ_t 分别为公司个体和年份固定效应,并在企业层面进行标准误的聚类。系数 α_1 是基准回归重点关注的系数,反映政府公共数据开放平台对人工智能创新的净影响。若系数 α_1 显著为正,则表明政府公共数据开放提高了人工智能创新水平。

2. 变量定义

(1)被解释变量:人工智能创新($AIpatent$)。根据Acemoglu和Restrepo(2019)的定义,人工智能创新的核心在于优化算法、软件或机器来推动智能化。本文借鉴尹志锋等(2023)、任英华等(2023)的做法,在国家知识产权局专利检索网站以机器学习、深度学习、人工智能、人机交互、机器人、无人驾驶等26个关键词进行检索,获取专利申请数据,作为人工智能创新的代理变量。选择专利申请数作为人工智能创新代理变量的主要原因在于:一是专利作

为技术创新产出，能够衡量创新技术水平，该指标在人工智能领域的实证研究中已被广泛应用(Mann 和 Püttmann, 2023)；二是相较于已有文献中使用工业机器人、信息技术投资指标作为代理变量，专利作为一种创新结果更能反映创新水平(尹志锋等, 2023)。本文选取 2009~2022 年上市公司人工智能专利申请数加 1 取自然对数，得到相应的人工智能创新变量作为实证分析的核心被解释变量。

(2)解释变量：地方政府公共数据平台上线与否(did_{it})。若企业所在城市上线公共数据开放平台，则当年及以后年份取值为 1，否则为 0。若企业所在城市未上线公共数据开放平台，则取值为 0。

(3)控制变量。为了控制其他影响因素，本文参考已有研究(方锦程等, 2023 ；李晓等, 2025)，在模型(1)中引入 5 个企业层面的控制变量，具体包括：公司规模(Size)、资产负债率(Lev)、资产收益率(ROA)、现金流状况(CashFlow)、固定资产占比(Fixed)。参考张慧等(2023)、陈婷等(2024)的研究，选取以下 6 个城市层面的控制变量，同时借鉴周波等(2024)的做法进行对数化处理，具体包括：经济发展水平(PCG)、对外开放水平(TIE)、产业结构(TPT)、政府支出(PEL)、社会消费水平(PTR)、金融发展水平(BDL)。具体变量定义见表 1。

表 1 变量定义

变量类型	变量名称	变量代码	变量含义及说明
被解释变量	人工智能创新	AIpatent	人工智能技术专利申请数，加 1 取自然对数
解释变量	公共数据开放	did	企业所在城市上线数据开放平台当年及以后取值为 1，否则为 0
控制变量	公司规模	Size	年末公司总资产取自然对数
	资产负债率	Lev	总负债占总资产的比例
	资产收益率	ROA	净利润占总资产的比例
	现金流状况	CashFlow	经营活动产生的现金净流量除以总资产
	固定资产占比	Fixed	固定资产净额占总资产的比例
	经济发展水平	PCG	人均地区生产总值的自然对数
	对外开放水平	TIE	进出口总额占地区生产总值比重的自然对数
	产业结构	TPT	第三产业总产值占第二产业总产值比重的自然对数
	政府支出	PEL	地方政府财政支出占地区生产总值比重的自然对数
	社会消费水平	PTR	社会消费品零售额占地区生产总值比重的自然对数
	金融发展水平	BDL	金融机构存贷款额占地区生产总值比重的自然对数

四、实证结果

(一)描述性统计分析

表 2 列示了主要变量的描述性统计结果。人工智能专利申请数量(AIpatent)均值为 0.203，中位数为 0，标准差为 0.576，表明样本企业的人工智能专利申请数量呈右偏分布。地方政府公共数据开放(did)的均值为 0.401，表明有 40.1% 的样本受到地方政府公共数据开放平台建设的影响。

(二)基准回归结果

表 3 报告了公共数据开放影响人工智能创新的实证检验结果。列(1)为仅加入解释变量、公司个体和年份固定效应的回归结果。结果显示，公共数据开放(did)的系数为 0.028，且在 5% 统计水平上显著。列(2)在列(1)基础上加入了企业层面控制变量，did 的系数为 0.026 且在 5% 统计水平上显著。列(3)在列(1)基础上加入了城市层面控制变量，did 的系数为 0.032 且在 1% 统计水平上显著。列(4)加入全部控制变量，did 的系数为 0.030 且在 5% 统计水平上显著。以上结果表明，公共数据开放显著提高了企业人工智能创新水平。

(三)稳健性检验

1. 堆叠平行趋势检验

$$AIpatent_{i,t} = \beta_0 + \beta_1 EventN4_{i,t} + \beta_2 EventN3_{i,t} + \beta_3 EventN2_{i,t} + \beta_4 EventN1_{i,t} + \beta_5 Event0_{i,t} + \beta_6 EventP1_{i,t} + \beta_7 EventP2_{i,t} + \beta_8 EventP3_{i,t} + \beta_9 EventP4_{i,t} + \gamma Controls_{i,t} + \eta_i + \tau_t + \varepsilon_{i,t}$$

(2)

表 2 描述性统计分析

Variable	Obs	Mean	SD	Min	P25	Median	P75	Max
Alpatent	32 146	0.203	0.576	0.000	0.000	0.000	0.000	2.944
did	32 146	0.401	0.490	0.000	0.000	0.000	1.000	1.000
Size	32 146	22.117	1.274	19.415	21.189	21.940	22.857	26.452
Lev	32 146	0.420	0.207	0.028	0.253	0.412	0.576	0.908
ROA	32 146	0.037	0.066	-0.556	0.014	0.038	0.069	0.222
CashFlow	32 146	0.046	0.071	-0.222	0.008	0.046	0.087	0.282
Fixed	32 146	0.207	0.158	0.001	0.085	0.174	0.292	0.772
PCG	32 146	11.394	0.537	9.918	11.065	11.478	11.825	12.156
TIE	32 146	-1.120	1.146	-4.307	-1.824	-0.988	-0.105	0.831
TPT	32 146	-0.662	0.222	-1.287	-0.799	-0.657	-0.494	-0.217
PEL	32 146	-1.962	0.309	-2.565	-2.180	-2.001	-1.742	-1.023
PTR	32 146	8.258	0.215	7.692	8.123	8.260	8.403	8.801
BDL	32 146	1.250	0.424	0.223	0.967	1.280	1.603	2.016

表 3 基准回归

	(1) Alpatent	(2) Alpatent	(3) Alpatent	(4) Alpatent
did	0.028** (2.269)	0.026** (2.129)	0.032** (2.626)	0.030** (2.508)
Size		0.024*** (2.956)		0.025*** (3.004)
Lev		0.071** (2.448)		0.072** (2.460)
ROA		0.136** (2.302)		0.140** (2.372)
CashFlow		-0.031 (-0.696)		-0.032 (-0.715)
Fixed		0.067* (1.916)		0.070** (2.010)
PCG			-0.042 (-1.595)	-0.045* (-1.723)
TIE			0.007 (0.578)	0.008 (0.694)
TPT			0.014 (0.214)	0.020 (0.311)
PEL			-0.017 (-0.521)	-0.016 (-0.506)
PTR			0.052 (1.479)	0.054 (1.541)
BDL			0.005 (0.148)	0.001 (0.035)
Constant	0.192*** (39.183)	-0.395** (-2.191)	0.213 (0.464)	-0.351 (-0.728)
Observations	32 146	32 146	32 146	32 146
Adjusted R ²	0.499	0.500	0.499	0.500
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes

注：***、**和*分别表示在1%、5%和10%的水平上显著；括号内为t值。

上述模型中，本文进一步采用堆叠事件研究法对模型（1）进行平行趋势检验。Event 为样本所处年份相对于政府公共数据开放年份的时间距离变量。鉴于各地区公共数据开放平台的上线年份存在时间异质性，且事件研究法要求在事件发生前后具备相对均衡且可比的观测窗口，本文以公共数据开放年份为中心，构建前后对称的事件窗口，从而更准确地识别公共数据开放的动态效应。具体而言，本文样本区间为 2009~2022 年，能够完整覆盖 2013~2018 年六个不同公共数据开放年份向前和向后各延展 4 年所对应的事件窗口。基于此，本文分别以 2013、2014、2015、2016、2017 和 2018 年作为事件发生年份构建事件研究子样本，其中处理组为所在城市于当年上线公共数据开放平台的企业，对照组为样本期内所在城市未上线公共数据开放平台的企业。随后，对上述子样本进行堆叠估计。模型（2）中，EventN4、EventN3、EventN2、EventN1 分别为公共数据开放前 4、3、2、1 年观测值，Event0 代表公共数据平台开放当年，EventP1、EventP2、EventP3、EventP4 分别表示数据开放平台构建后第 1、2、3、4 年，其余变量的定义与模型（1）相同。为了规避时间变量间的多重共线性，本文以事件窗口最早一期作为基准期，即在回归中省略 EventN4，检验结果如表 4 所示。列（2）加入控制变量后，系数 β_2 、 β_3 、 β_4 均不显著，系数 β_0 、 β_7 、 β_8 、 β_9 均显著为正。以上结果表明在地方政府上线公共数据开放平台前，处理组和控制组企业在人工智能创新方面不存在显著差异，而公共数据开放平台的上线能够显著提高人工智能创新水平，且这一推动作用能够持续多期。

2. 安慰剂检验

本文采用随机设定公共数据开放平台上线城市的方法进行安慰剂检验，500 次随机重复实验结果如图 1 所示。估计系数的分布高度集中于 0 附近，大量落在 95% 经验区间，表明在随机分配处理状态的情况下，模型通常不会产生系统性的正向处理效应，并显著异于基准回归系数。以上结果验证了基准回归结果的稳健性。

3. 高维固定效应

本文进一步采用行业与年份的交互固定效应，以消除不同行业在不同时期所面临的行业特定冲击对人工智能专利申请的影响。如表 5 第（1）列所示，did 的系数在引入行业年份交互固定效应后依然显著。

4. 双重聚类标准误

鉴于样本时间跨度较长，基准回归不仅可能受到公司个体问题的影响，也可能受到时间相关因素的影响，本文对标准误进行了公司和年份层面的双重聚类。检验结果如表 5 第（2）列所示，did 系数显著为正，从而排除标准误聚类层级差异对基准回归结论的影响。

5. 替换被解释变量

为进一步验证研究结论的稳健性，本文借助国家知

表 4 堆叠事件研究法检验		
	(1) Alpatent	(2) Alpatent
EventN3	0.007 (0.539)	0.011 (0.769)
EventN2	0.021 (1.159)	0.026 (1.890)
EventN1	0.035* (1.661)	0.036 (1.546)
Event0	0.005 (0.219)	0.010 (0.375)
EventP1	0.071** (2.054)	0.075** (2.044)
EventP2	0.101*** (2.825)	0.108*** (2.929)
EventP3	0.075* (1.913)	0.085** (2.174)
EventP4	0.072** (2.085)	0.079** (2.322)
Constant	0.168*** (59.905)	0.552 (0.657)
Observations	55 998	55 998
Adjusted R ²	0.483	0.485
Controls	No	Yes
Year FE	Yes	Yes
Firm FE	Yes	Yes

注：***、**和* 分别表示在 1%、5% 和 10% 的水平上显著；括号内为 t 值。

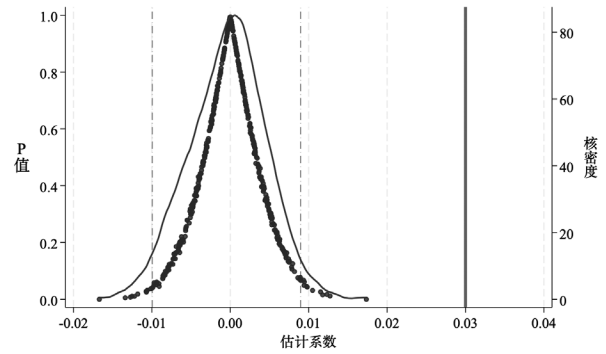


图 1 安慰剂检验

表 5 采用高维固定效应等的稳健性检验

	(1) 高维固定效应 Alpatent	(2) 双重聚类标准误 Alpatent	(3) 替换被解释变量 Alpatent1	(4) 倾向得分匹配 Alpatent
did	0.023** (1.974)	0.030** (2.376)	0.020** (2.424)	0.100*** (2.814)
Size	0.031*** (3.851)	0.025** (2.288)	0.035*** (5.248)	-0.005 (-0.261)
Lev	0.012 (0.423)	0.072 (1.328)	-0.034 (-1.357)	0.124* (1.650)
ROA	0.124** (2.095)	0.140 (1.717)	-0.041 (-1.058)	0.264* (1.863)
CashFlow	-0.064 (-1.442)	-0.032 (-0.719)	-0.085*** (-2.915)	0.091 (0.640)
Fixed	0.048 (1.353)	0.070 (1.731)	0.026 (0.752)	0.003 (0.029)
PCG	-0.032 (-1.274)	-0.045* (-1.890)	-0.014 (-0.654)	-0.047 (-0.426)
TIE	0.009 (0.772)	0.008 (0.782)	0.000 (0.031)	0.042 (1.611)
TPT	-0.007 (-0.110)	0.020 (0.224)	-0.013 (-0.304)	-0.261 (-1.072)
PEL	-0.028 (-0.867)	-0.016 (-0.624)	-0.021 (-0.856)	0.081 (0.897)
PTR	0.044 (1.290)	0.054 (1.088)	-0.002 (-0.093)	-0.030 (-0.211)
BDL	-0.008 (-0.222)	0.001 (0.031)	-0.009 (-0.333)	-0.027 (-0.298)
Constant	-0.549 (-1.181)	-0.351 (-0.607)	-0.528 (-1.423)	1.016 (0.599)
Observations	32 141	32 146	32 146	7 933
Adjusted R ²	0.529	0.500	0.496	0.462
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Industry × Year FE	Yes	No	No	No

注：***、**和*分别表示在1%、5%和10%的水平上显著；括号内为t值。

识产权局发布的《关键数字技术专利分类体系(2023)》，依据人工智能技术相应的国际专利分类号(IPC)识别企业人工智能专利申请数，并以其自然对数(Alpatent1)替换基准回归中的被解释变量，数据来自CNRDS数据库。检验结果见表5第(3)列，did系数显著为正，表明本文结论具有稳健性。

6. 倾向得分匹配

为避免研究结论受到自选择问题的干扰，本文采用倾向得分匹配法，将模型(1)中的控制变量作为匹配变量进行1:1的最近邻匹配，匹配后的样本估计结果见表5第(4)列。结果表明，利用倾向得分匹配缓解选择性偏误问题后的估计结果与基准回归结果并无实质性差异，进一步证明了结论的稳健性。

7. 排除其他政策冲击的干扰

为排除同期实施的其他相关政策对估计结果造成的潜在干扰，本文进一步在基准回归模型中控制了三项可能与公共数据开放政策存在叠加或替代效应的国家级试点政策(信息消费示范城市、智能制造试点、国家级大数据综合试验区)，并分别进行回归分析。表6结果显示，在纳入相关政策变量后，did的系数至少在5%的水平上显著为正，

表 6 排除其他政策冲击的稳健性检验

	(1) Alpatent	(2) Alpatent	(3) Alpatent
did	0.040** (2.526)	0.030** (2.507)	0.034*** (2.606)
infoconsum_treat	0.041** (2.532)		
smartmanu_treat		0.542*** (6.631)	
bigdata_treat			-0.019 (-1.006)
Size	0.021** (2.242)	0.024*** (2.964)	0.025*** (3.000)
Lev	0.076** (2.282)	0.076*** (2.616)	0.073** (2.480)
ROA	0.175** (2.347)	0.133** (2.284)	0.139** (2.354)
CashFlow	-0.065 (-1.257)	-0.037 (-0.817)	-0.032 (-0.709)
Fixed	0.101*** (2.592)	0.064* (1.872)	0.069** (1.981)
PCG	-0.059* (-1.813)	-0.044* (-1.712)	-0.047* (-1.795)
TIE	0.011 (0.864)	0.004 (0.348)	0.007 (0.586)
TPT	0.065 (0.815)	0.019 (0.297)	0.008 (0.121)
PEL	-0.001 (-0.017)	-0.023 (-0.709)	-0.016 (-0.494)
PTR	0.050 (1.196)	0.044 (1.278)	0.055 (1.577)
BDL	-0.021 (-0.530)	0.000 (0.012)	0.009 (0.269)
Constant	-0.022 (-0.040)	-0.287 (-0.610)	-0.354 (-0.734)
Observations	24 579	32 146	32 146
Adjusted R ²	0.488	0.506	0.500
Year FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes

注：***、**和* 分别表示在 1%、5% 和 10% 的水平上显著；括号内为 t 值。

表明本文结论稳健。

(四) 影响机制

为考察公共数据开放对企业人工智能创新的影响机制，进一步借鉴毛其淋(2020)的做法，对基准回归模型(1)进行拓展，构建如下模型：

$$Alpatent_{i,t}=\alpha_0+\alpha_1did_{i,t}\times M_{i,t}+\alpha_2did_{i,t}+\alpha_3M_{i,t}+\gamma Controls_{i,t}+\eta_t+\tau_i+\varepsilon_{i,t}$$
 (3)

其中，M_{i,t}表示机制变量，包括降低经营管理成本和拓展大数据应用。

1. 降低经营管理成本

参考席卫群和凌兴(2025)的做法，本文采用管理与销售费用之和占营业收入之比作为企业经营管理成本的代

理变量 (cost), cost 数值越大, 代表企业的经营管理成本越高。表 7 第 (1) 列展示了经营管理成本对企业人工智能创新的回归结果, 可以发现经营管理成本显著抑制了人工智能创新。表 7 第 (2) 列汇报了回归模型 (3) 的估计结果, 可以发现交互项 $did \times cost$ 的系数在 1% 水平上显著为负, 该结果支持了本文的理论预期, 即公共数据开放通过降低经营管理成本, 进而促进企业人工智能创新。

2. 拓展大数据应用

参考张叶青等 (2021) 的做法, 使用 Python 对上市公司年报进行文本分析, 用大数据、数据挖掘、数据可视化等关键词在年报中出现的次数加 1 后取自然对数表示企业大数据应用程度 (bigdata)。表 8 第 (1) 列展示了大数据应用对企业人工智能创新的回归结果, 可以发现大数据应用显著促进了人工智能创新。表 8 第 (2) 列汇报了回归模型 (3) 的估计结果, 可以发现交互项 $did \times bigdata$ 的系数为正, 并且在 5% 水平上显著, 该结果支持了本文的理论预期, 即

表 7 降低经营管理成本的机制检验

	(1) Alpatent	(2) Alpatent
did × cost		-0.162*** (-3.028)
did		0.056*** (3.562)
cost	-0.263*** (-6.213)	-0.202*** (-4.153)
Size	0.019** (2.282)	0.018** (2.194)
Lev	0.062** (2.144)	0.062** (2.139)
ROA	0.051 (0.824)	0.038 (0.623)
CashFlow	-0.044 (-0.974)	-0.049 (-1.086)
Fixed	0.068** (1.967)	0.071** (2.038)
PCG	-0.037 (-1.410)	-0.046* (-1.758)
TIE	0.008 (0.751)	0.009 (0.812)
TPT	0.024 (0.372)	0.024 (0.363)
PEL	-0.013 (-0.395)	-0.012 (-0.367)
PTR	0.049 (1.395)	0.052 (1.484)
BDL	-0.002 (-0.054)	-0.006 (-0.162)
Constant	-0.193 (-0.399)	-0.113 (-0.236)
Observations	32 142	32 142
Adjusted R ²	0.501	0.501
Year FE	Yes	Yes
Firm FE	Yes	Yes

注: **、* 和 * 分别表示在 1%、5% 和 10% 的水平上显著; 括号内为 t 值。

表 8 拓展大数据应用的机制检验

	(1) Alpatent	(2) Alpatent
did × bigdata		0.029** (2.492)
did		0.009 (0.734)
bigdata	0.048*** (6.012)	0.031*** (3.119)
Size	0.015* (1.862)	0.015* (1.876)
Lev	0.075** (2.554)	0.069** (2.373)
ROA	0.151** (2.528)	0.157*** (2.629)
CashFlow	-0.029 (-0.637)	-0.027 (-0.606)
Fixed	0.076** (2.174)	0.077** (2.210)
PCG	-0.032 (-1.257)	-0.041 (-1.598)
TIE	0.009 (0.801)	0.010 (0.936)
TPT	0.039 (0.591)	0.049 (0.747)
PEL	-0.008 (-0.240)	-0.006 (-0.199)
PTR	0.052 (1.498)	0.054 (1.563)
BDL	-0.003 (-0.078)	-0.004 (-0.113)
Constant	-0.252 (-0.522)	-0.164 (-0.342)
Observations	32 095	32 095
Adjusted R ²	0.502	0.503
Year FE	Yes	Yes
Firm FE	Yes	Yes

注: **、* 和 * 分别表示在 1%、5% 和 10% 的水平上显著; 括号内为 t 值。

公共数据开放通过提高大数据应用水平，进而促进企业人工智能创新。

(五) 异质性检验

1. 企业资产规模

本文按各年度企业资产规模的中位数将样本划分为大规模组和小规模组，并基于模型(1)进行分组回归。分组检验结果如表9列(1)和列(2)所示，did系数在大规模企业组为0.012，不具有统计显著性，而在小规模企业组中的系数为0.046，在1%的水平上显著，并且两组系数通过了组间差异检验。该结果可能的原因在于，小企业组织结构更为灵活，具有更强的技术敏捷性和创新动机，能够更快速地将公共数据转化为创新应用。

2. 城市发展水平

不同城市的发展水平也具有显著差异，发展水平较高的城市不仅在经济、文化、科技等方面优势明显，往往还拥有政策优势以及较强的资源配置能力，这可能会导致政府公共数据开放对企业技术创新的影响表现出异质性。对

表 9 企业特征异质性分析

	(1) 大规模 Alpatent	(2) 小规模 Alpatent
did	0.012 (0.661)	0.046*** (2.979)
Size	0.023 (1.277)	0.081*** (6.249)
Lev	0.152** (2.478)	0.031 (0.953)
ROA	0.219** (2.070)	0.083 (1.116)
CashFlow	-0.016 (-0.223)	-0.017 (-0.288)
Fixed	0.005 (0.073)	0.087 (2.023)
PCG	-0.054 (-1.257)	-0.003 (-0.104)
TIE	0.028 (1.533)	-0.001 (-0.056)
TPT	0.169 (1.617)	-0.151* (-1.917)
PEL	0.037 (0.747)	-0.062* (-1.672)
PTR	0.103* (1.889)	-0.033 (-0.743)
BDL	-0.068 (-1.223)	0.037 (0.965)
Constant	-0.311 (-0.382)	-1.538** (-2.487)
Observations	15 929	15 904
Adjusted R ²	0.553	0.459
Year FE	Yes	Yes
Firm FE	Yes	Yes
组间差异	0.049**	

注：***、**和*分别表示在1%、5%和10%的水平上显著；括号内为t值。

表 10 城市特征异质性分析

	(1) 高发展水平 Alpatent	(2) 低发展水平 Alpatent
did	0.042*** (2.768)	0.012 (0.540)
Size	0.037*** (3.197)	0.008 (0.698)
Lev	0.090** (2.157)	0.049 (1.251)
ROA	0.078 (0.924)	0.244*** (3.177)
CashFlow	0.036 (0.629)	-0.133* (-1.813)
Fixed	0.078 (1.555)	0.055 (1.152)
PCG	-0.002 (-0.049)	-0.006 (-0.158)
TIE	0.015 (0.536)	0.021* (1.727)
TPT	0.014 (0.110)	0.078 (0.951)
PEL	0.042 (0.735)	-0.076 (-1.503)
PTR	0.055 (1.019)	0.059 (1.196)
BDL	0.054 (0.734)	0.036 (0.789)
Constant	-1.072 (-1.312)	-0.559 (-0.883)
Observations	19 020	13 106
Adjusted R ²	0.511	0.475
Year FE	Yes	Yes
Firm FE	Yes	Yes
组间差异	0.071*	

注：***、**和*分别表示在1%、5%和10%的水平上显著；括号内为t值。

此,本文参考李婉红等(2022)、马永军和崔闯盛(2024)的研究,根据新一线城市研究所2020年发布的《城市商业魅力排行榜》,将一线、新一线、二线城市归为高发展水平城市组,其他城市归为低发展水平城市组,分别进行检验,回归结果见表10的(1)(2)列。在高发展水平城市样本中,did的回归系数显著为正,但在低发展水平城市样本中,did的回归系数不显著,并且两组系数通过了组间差异检验。这表明政府公共数据开放对企业人工智能创新的促进效应在发展水平较高的城市更强。

五、研究结论及启示

本文基于人工智能创新视角,以地方政府公共数据开放平台上线为准自然实验,实证检验地方政府公共数据开放对企业人工智能创新的影响及作用机理。研究发现:政府公共数据开放显著提升了企业人工智能创新水平。机制分析表明,政府公共数据开放通过降低经营管理成本、拓展大数据应用作用于人工智能创新。异质性分析表明,政府公共数据开放对小规模企业、高发展水平城市企业的促进作用更为明显。

基于上述结论,本文提出以下建议:

第一,加强公共数据开放平台的建设与运营。政府应加大对公共数据开放平台的投入,不断完善其功能与架构。如提升平台的数据存储、安全防护及检索能力等,确保数据能稳定、高效地开放给企业等主体。同时,注重运营管理,安排专业人员及时更新数据、处理数据使用过程中的问题,保障平台良好运行,为人工智能创新筑牢数据基础。

第二,增强公共数据开放对企业和城市的精准赋能。在企业层面,应重点关注并支持数据资源相对匮乏的小规模企业,通过出台专项扶持政策,帮助其突破数据瓶颈,充分发挥公共数据对技术创新的促进作用。对于已具备较完善数据系统的大规模企业,则应引导其进一步拓展数据融合与应用场景,提升数据要素的协同价值。而在城市层面,应着力推动区域间公共数据开放与人工智能创新的协调发展,避免进一步扩大城市发展差距。

第三,深化政企合作共同推动人工智能创新体系建设。政府应积极与企业携手,搭建沟通交流平台,使企业能及时反馈使用公共数据时的需求与困难。双方共同探索创新合作模式,如联合开展人工智能创新项目、共建研发中心等,借助公共数据开放的契机,打造完善且高效的人工智能创新体系,全方位助力人工智能创新发展。

主要参考文献:

- [1] 陈婷,段尧清,吴瑾.数据要素市场化能否提升城市创新能力——一个准自然实验[J].科技进步与对策,2024,41(3):74-83.
- [2] 陈艳利,蒋琪.数据生产要素视角下开放公共数据与企业创新——基于建立公共数据开放平台的准自然实验[J].经济管理,2024,46(1):25-46.
- [3] 程波辉,罗培镡.政府数据开放能提升民营企业绩效吗——基于多时点双重差分法的验证[J].吉林大学社会科学学报,2024,64(4):189-203.
- [4] 杜振华.政府数据开放与创新驱动经济增长的关系[J].首都师范大学学报(社会科学版),2020,(2):63-71.
- [5] 方锦程,刘颖,高昊宇,等.公共数据开放能否促进区域协调发展?——来自政府数据平台上线的准自然实验[J].管理世界,2023,39(9):124-142.
- [6] 李荣华,王娇娇,张磊.公共数据开放的新质生产力效应——基于企业数字创新的视角[J].财经研究,2025,51(10):50-63.
- [7] 李婉红,刘芳,刘天森.国家高新区提升了城市创新效率吗?——基于空间集聚调节效应的实证检验[J].管理评论,2022,34(5):93-108.
- [8] 李晓,杜宜萱,于晓旭.数据要素与企业产能利用率——基于公共数据开放平台的准自然实验[J].财务研究,2025,(3):54-71.
- [9] 吕越,张昊天,高恺琳.人工智能时代的中国产业链“延链补链”——基于制造业企业智能设备进口的微观证据[J].中国工业经济,2024,(1):56-74.

- [10] 马永军, 崔闯盛. 政府数据开放能否促进城市技术创新? [J]. 西部论坛, 2024, 34 (5) : 18-33.
- [11] 毛其淋. 贸易政策不确定性是否影响了中国企业进口? [J]. 经济研究, 2020, 55 (2) : 148-164.
- [12] 彭远怀. 政府数据开放的价值创造作用: 企业全要素生产率视角 [J]. 数量经济技术经济研究, 2023, 40 (9) : 50-70.
- [13] 权小锋, 李静蕾. 人工智能与企业协同创新——基于国家新一代人工智能创新发展试验区的准自然实验 [J]. 经济研究, 2025, 60 (8) : 72-93.
- [14] 任英华, 刘宇钊, 李海彤. 人工智能技术创新与企业全要素生产率 [J]. 经济管理, 2023, 45 (9) : 50-67.
- [15] 沈超, 赵正宇. 人工智能技术安全风险探析 [J]. 国家安全研究, 2025, (1) : 29-45.
- [16] 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.
- [17] 谭玉松. 人工智能技术与产业生态化转型 [J]. 求是学刊, 2024, 51 (3) : 79-90.
- [18] 卫志民, 杨修博, 王浚懿. 政府数据开放能否提升企业劳动收入份额 [J]. 山西财经大学学报, 2024, 46 (7) : 1-15.
- [19] 席卫群, 凌兴. 税收优惠与高新制造企业的数字化转型——来自 A 股上市公司的经验证据 [J]. 财政科学, 2025, (8) : 101-118.
- [20] 谢尚, 赵鹏军. 公共数据开放与城乡居民收入增长: 来自政府数据平台上线的证据 [J]. 北京交通大学学报 (社会科学版), 2024, 23 (3) : 61-75.
- [21] 杨来峰. 公共数据开放能改善企业的产品市场表现吗——基于公共数据开放平台上线的准自然实验 [J]. 当代财经, 2025, (2) : 141-152.
- [22] 尹志锋, 曹爱家, 郭家宝, 等. 基于专利数据的人工智能就业效应研究——来自中关村企业的微观证据 [J]. 中国工业经济, 2023, (5) : 137-154.
- [23] 于同同, 张涛, 马海群, 等. 我国公共数据开放领域知识融合测度研究 [J]. 情报理论与实践, 2024, 47 (7) : 160-169.
- [24] 张晨, 万相昱, 姜智超, 等. 开放政府数据的经济增长效应研究 [J]. 中国软科学, 2023, (2) : 1-11.
- [25] 张慧, 易金彪, 徐建新. 数字化变革如何影响城市创新——基于国家大数据综合试验区建设的经验证据 [J]. 科学学研究, 2023, 41 (8) : 1484-1494.
- [26] 张叶青, 陆瑶, 李乐芸. 大数据应用对中国企业市场价值的影响——来自中国上市公司年报文本分析的证据 [J]. 经济研究, 2021, 56 (12) : 42-59.
- [27] 郑志强, 何佳例, 陈尧. 公共数据开放与企业数字技术创新 [J]. 财经研究, 2025, 51 (2) : 19-33.
- [28] 周波, 孔欣悦, 刘钟墨. 数字经济对财政支出效率的影响与机制检验 [J]. 统计与决策, 2024, 40 (4) : 146-151.
- [29] Acemoglu, D., Restrepo, P. Automation and New Tasks: How Technology Displaces and Reinstates Labor [J]. Journal of Economic Perspectives, 2019, 33 (2) : 3-30.
- [30] Babina, T., Pedyk, A., He, A., et al. Artificial Intelligence, Firm Growth, and Product Innovation [J]. Journal of Financial Economics, 2024, 151: 103745.
- [31] Cheng, H., Jia, R., Li, D., et al. The Rise of Robots in China [J]. Journal of Economic Perspectives, 2019, 33 (2) : 71-88.
- [32] Du, J., Gao, H., Wen, H., et al. Public Data Access and Stock Price Synchronicity: Evidence from China [J]. Economic Modelling, 2024, 130: 106591.
- [33] Haefner, N., Wincent, J., Parida, V., et al. Artificial Intelligence and Innovation Management: A Review, Framework, and Research Agenda [J]. Technological Forecasting and Social Change, 2021, 162: 120392.
- [33] Mann, K., Püttmann, L. Benign Effects of Automation: New Evidence from Patent Texts [J]. Review of Economics and Statistics, 2023, 105 (3) : 562-579.
- [35] Matheus, R., Janssen, M. A Systematic Literature Study to Unravel Transparency Enabled by Open Government Data: The Window Theory [J]. Public Performance & Management Review, 2019, 43 (3) : 503-534.
- [36] Nikiforova, A., Lnenicka, M. A Multi-Perspective Knowledge-Driven Approach for Analysis of the Demand Side of the Open

Government Data Portal[J]. Government Information Quarterly, 2021, 38 (4) : 101-122.

[37]Wang, X., Li, Y., Tian, L., et al. Government Digital Initiatives and Firm Digital Innovation : Evidence from China[J]. Technovation, 2023, 119 (3) : 102545.

[38]Zhou, M., Wang, Y., Huang, X., et al. Can Open Government Data Policy Improve Firm Performance? Evidence from Listed Firms in China[J]. Managerial and Decision Economics, 2023, 44 (5) : 2593-2603.

Public Data Openness and Corporate Artificial Intelligence Innovation

WANG Yan, ZHAO Chao, WU Hongliang, WANG Banghao

Abstract: Taking the establishment of local government public data open platforms as a quasi-natural experiment, this paper employs a multi-period difference-in-differences (DID) model to examine the impact of public data openness on artificial intelligence innovation and its underlying mechanisms. The results show that government public data openness significantly enhances firms' levels of artificial intelligence innovation. Mechanism analysis reveals that reducing operating and management costs and improving big data application capabilities are important channels through which public data openness promotes corporate AI innovation. Heterogeneity analysis indicates that the positive effect of public data openness on AI innovation is more pronounced for small-sized firms and firms located in highly developed cities. This study confirms the positive role of government public data openness in promoting corporate artificial intelligence innovation and provides policy implications for leveraging data as a production factor to empower AI innovation.

Key words: public data openness; artificial intelligence innovation; operating and management costs; big data application

(责任编辑 张雨吟)