

DOI: 10.16538/j.cnki.fem.20260420.401

数据知识产权保护与企业数据技术创新

——来自生成式大语言模型的经验证据

杨靖¹, 陈志斌¹, 封思贤^{2,3}

(1. 东南大学 经济管理学院, 江苏 南京 211189; 2. 南京师范大学 商学院, 江苏 南京 210023;
3. 南京师范大学 金陵女子学院, 江苏 南京 210023)

摘要: 企业数据技术创新是实现数据要素价值转化的关键路径,然而,数据知识产权保护能否以及如何有效激励企业开展数据技术创新,尚缺乏系统的实证检验。本文以全国首批数据知识产权工作试点为准自然实验,利用2015—2024年A股上市公司数据,运用大语言模型识别企业的数据技术专利,构建双重差分模型考察数据知识产权保护对企业数据技术创新的影响。研究发现,数据知识产权工作试点显著促进了企业数据技术创新。机制检验表明,数据知识产权保护主要通过激励专用要素投入和促进知识溢出发挥作用。异质性分析显示,该促进效应在国有企业、数据原生企业以及数字基础设施较完善地区更为突出,在不同科创属性企业间未呈现显著差异。经济后果分析表明,数据知识产权保护通过促进企业数据技术创新,进而提升了企业全要素生产率。研究为理解数据知识产权保护的微观创新效应提供了新证据,对建设全国一体化数据市场、培育新质生产力与推动数字经济高质量发展具有重要政策启示。

关键词: 数据知识产权保护;数据技术创新;生成式大语言模型;知识溢出

中图分类号: F270 **文献标识码:** A **文章编号:** 1001-4950(2026)08-0095-18

一、引言

伴随数字经济的蓬勃发展,数据作为新型生产要素的战略价值日趋重要(李健等,2023)。国家数据局发布的《数字中国发展报告(2024年)》显示,2024年我国数据产量达41.06泽字节(ZB),同比增长25%,累计数据存储量达2.09ZB,数据资源规模优势持续扩大。现有研究普遍支持数据在促进经济增长和价值创造方面的积极作用(Cong等,2021;杨俊等,2022),但在实践中,数据资源的潜在价值能否真正转化为现实生产力,核心在于企业是否具备以大数据分析为核心的数据技术能力,以有效提取信息并支撑数据要素参与企业生产经营活动(陈楠等,

收稿日期: 2025-12-26

基金项目: 国家社会科学基金重大项目(24&ZD092)

作者简介: 杨靖(1997—),男,东南大学经济管理学院博士研究生;

陈志斌(1965—),男,东南大学经济管理学院教授,博士生导师(通信作者, seu0216@163.com);

封思贤(1973—),男,南京师范大学商学院/金陵女子学院教授,博士生导师。

2024)。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》提出要“引导企业和科研机构推动数据要素相关技术和产业应用创新”。2024年,国家发展改革委等部门印发的《关于促进数据产业高质量发展的指导意见》(下文简称《指导意见》)明确将“加快数据技术创新”列为重点任务,强调着力推动数据采集、存储、处理、分析等关键环节的技术突破,鼓励企业加大在数据技术领域的研发投入,以提升数据资源利用效率,激活数据要素潜能。

然而,我国企业在推进数据技术创新过程中面临显著的制度性约束。数据要素具有非竞争性、易复制性等特征,其产权属性与传统财产权体系存在显著差异(周文和韩文龙,2023)。现有产权制度对数据权属缺乏清晰、统一的界定标准,导致企业在创新过程中难以形成稳定的权利预期(申卫星,2020)。例如,企业在采集多源异构数据以支撑算法研发时,常因数据来源的权属边界模糊而面临非故意侵权;而在创新成果落地后,其研发的结构化数据集、定制化数据处理算法等成果,又因相关权益保护机制尚不完善,可能被竞争对手低成本复制,造成创新投入与收益严重失衡(郭丰等,2025)。产权保护的缺失直接削弱了企业开展数据技术创新的内生动力。因此,如何有效激励企业开展数据技术创新,充分释放数据要素价值,已成为推动数字经济高质量发展中亟待回应的重大议题。

为摆脱上述困境,探索建立适配数据要素特征的数据知识产权保护制度,成为赋予企业稳定预期、激发创新活力的关键制度安排。相较于传统知识产权制度,数据知识产权制度更加注重为依法收集、经算法加工、具备实用价值与智力成果属性的数据集提供产权保护。2021年,中共中央、国务院印发的《知识产权强国建设纲要(2021—2035年)》首次提出“研究构建数据知识产权保护规则”。次年,国家知识产权局在北京市、上海市等地区开展首批数据知识产权工作试点,旨在探索建立数据登记、审查、公示与运用机制,为数据相关权益的登记、识别、保护和流通利用提供制度支持。党的二十届四中全会通过的《中共中央关于制定国民经济和社会发展第十五个五年规划的建议》提出“健全数据要素基础制度,建设开放共享安全的全国一体化数据市场,深化数据资源开发利用”。在此背景下,一个亟待回答的问题是:数据知识产权工作试点能否有效激发企业数据技术创新活力?科学评估这一政策效应及其作用机理,能够为完善数据基础制度、释放数据要素潜能提供有益启示。

主要有三部分文献与本文密切相关。第一部分是数据基础制度建设经济后果的相关研究。数据基础制度涉及数据产权界定、流通交易、收益分配、安全治理等多个方面。既有研究多从数据交易平台建设视角出发,探讨数据要素流通对企业全要素生产率(戴魁早等,2023)、资源配置效率(徐晔等,2025)、新质生产力发展(谢地等,2025)等方面的影响。该类研究主要聚焦于流通交易这一市场化后端环节,对数据知识产权保护这一基础性前置制度安排的微观创新效应关注不足。第二部分是关于企业数据技术创新的研究。现有文献多将数据技术创新涵盖于广义的数字创新或数字化转型框架下,从政策环境、市场条件、企业家特征等视角探讨了其驱动因素(胡增玺和马述忠,2023;Corvello等,2023;陶锋等,2025),并主要采用年报关键词检索或IPC专利分类加以测度(Wang等,2023)。上述研究在概念界定上较为宽泛,难以区分围绕数据采集、存储、处理与分析等环节的技术创新与一般数字化应用行为,且传统关键词检索与IPC分类法难以精准识别企业的数据技术攻关活动。第三部分是关于知识产权保护与创新的研究。现有文献广泛验证了传统知识产权保护对企业创新的激励效应(郭丰等,2025)。然而,数据知识产权作为适配数据要素特征的新型产权制度安排,其保护对象和规则设计与传统知识产权存在明显差异,相关实证研究仍较为有限。

基于上述背景,本文以2022年全国首批数据知识产权工作试点为准自然实验,利用2015—2024年A股上市公司数据,运用生成式大语言模型识别企业数据技术专利,并采用双重差分模

型考察数据知识产权保护对企业数据技术创新的影响。相较于既有研究,本文的边际贡献在于:第一,拓展了数据基础制度建设经济后果的研究边界。区别于关注数据流通交易经济效应的文献,本文从作为基础性前置制度安排的数据产权制度入手,利用数据知识产权工作试点的准自然实验,考察数据知识产权保护对企业数据技术创新的影响,为理解数据基础建设的经济效应提供了新的经验证据。第二,丰富了企业数据技术创新驱动因素的相关研究,并在测算方法上作了有益补充。当前专门探讨企业数据技术创新影响因素的文献较为有限,代表性研究包括张杨等(2026)考察了智能制造对数据技术创新的影响。本文将数据技术创新从广义的数字创新框架中剥离出来,创新性地引入生成式大语言模型对数据技术专利进行识别,克服了传统IPC分类与关键词检索方法的效度局限,为企业数据技术创新的量化研究提供了方法论补充。在此基础上,本文揭示了数据知识产权保护对企业数据技术创新的作用机理,为从制度环境维度理解企业数据技术创新的驱动因素提供了新视角。第三,深化了对数据知识产权保护政策效应的理解。现有文献主要关注传统知识产权保护对企业创新的影响,关于数据知识产权这类新型产权安排的研究多围绕理论构建与规则设计展开(王伟玲,2024),对其微观实施效果特别是对企业技术创新行为的实证检验尚显不足。本文考察了数据知识产权工作试点对企业数据技术创新的因果效应、作用机制及异质性影响,有助于全面评估该政策的实施效果,不仅拓展了知识产权保护研究的理论边界,也为优化数据知识产权制度设计、更好适应数字经济发展需求提供了实证依据。

二、制度背景与理论分析

(一)制度背景

2022年,国家知识产权局印发《关于确定数据知识产权工作试点地方的通知》,确定北京市等8个地区作为首批试点,围绕数据知识产权登记制度开展系统性探索,推动形成可操作、可复制的实践路径。随着试点工作的深入推进,2023年12月末,国家知识产权局进一步将试点范围扩展至天津市、河北省、山西省等9个地区,以深化数据知识产权登记管理体系建设,加强数据知识产权保护和运用,为全国范围内数据知识产权制度的正式建立积累经验、奠定基础。

从数据知识产权工作试点内容来看,形成了以下制度功能。一是产权界定功能。试点地区通过制定数据知识产权登记管理办法,经地方登记机构依其规则审查后予以登记并颁发证书,明确数据权属关系,为数据要素市场化流通奠定了产权基础^①。并且,该试点政策突破传统物权式的独占强保护模式,将保护客体限定为“依法依规获取、经过一定规则或算法处理、具备实用价值与智力成果属性的数据集合”,而一般不直接对底层原始数据本身赋予类似传统知识产权的绝对排他权,通过“登记赋权+行为规制”的制度安排构建了一种以有限排他性为特征的数据知识产权保护模式。在这一模式下,登记证书保护数据处理者的实质性加工成果,规制他人以不正当方式获取和使用数据,但允许其他主体依法依规获取同源数据并进行加工分析,从而在保护数据处理者合法权益的同时,为他人正当开发数据留下空间。二是信息公示功能。试点地区统一规范登记信息的公示要素,要求登记主体在数据知识产权登记平台上对数据集合的名称、规模、数据来源、处理规则及应用场景等关键信息进行标准化披露。这使部分原本以商业秘密形式存在的数据技术信息得以在权利边界清晰的前提下呈现,潜在的数据使用者能够了解数据集合的来源合法性、加工规范性与应用价值,从而降低了知识搜寻成本,为数据要素的

^①关于数据知识产权登记证书的效力,在全国首个涉数据知识产权登记证书效力认定案——Y公司与S公司不正当竞争纠纷上诉案中,法院首次在司法裁判中确认数据知识产权登记证书对数据持有以及数据收集行为合法的证明效力,认为S公司就涉案数据集取得的《数据知识产权登记证》“可作为证明S公司享有涉案数据集相关财产性利益的初步证据”以及“可作为涉案数据集收集行为合法的初步证据”,为数据知识产权登记实践提供了有力的司法支撑。

跨主体流通和再利用创造了条件。三是价值转化功能。试点地区积极探索数据知识产权的价值转化路径,通过拓展市场化应用场景、明确相应规则与操作流程,推动数据知识产权融入交易授权、质押融资、作价入股等运作环节。这有助于引导金融机构、数据企业等市场主体深度参与价值实现过程,为数据资源的合规开发、流通利用和价值实现创造条件,进一步激发数据要素市场活力。

从实践成效来看,数据知识产权工作试点在制度建设、登记实施与价值转化等方面取得阶段性进展。截至2025年6月,17个试点地区均已出台专门登记管理办法,规范数据知识产权的登记流程与审查标准。在具体实施层面,各试点地区累计受理登记申请近5.8万件,颁发登记证书超过2.9万份,累计实现融资增信超100亿元,交易许可金额超5.8亿元。2025年4月,17个试点全部通过总结验收,其中北京、上海等7地获评优秀,标志着试点工作取得阶段性成效。这一实践为培育统一开放、竞争有序的数据要素市场奠定了坚实基础,也为研究数据知识产权保护如何影响企业创新行为提供了宝贵的准自然实验场景。

(二)理论分析与研究假设

数据知识产权工作试点通过构建有限排他性的数据知识产权保护模式,可从内部资源积累与外部知识获取两个层面影响企业数据技术创新:一方面,数据知识产权工作试点通过登记赋权强化对数据加工成果的保护,激励企业加大数据资源积累、数据技术人才等专用要素投入,为数据技术创新夯实内部资源基础;另一方面,数据知识产权工作试点依托信息公示功能改善外部知识流动环境,使企业能以更低成本获取前沿技术与场景知识,为数据技术创新提供更为丰富的外部知识来源。基于此,本文将从专用要素投入与知识溢出两个维度,阐释数据知识产权工作试点促进企业数据技术创新的内在机理。

1. 专用要素投入渠道

数据技术创新是企业在全生命周期各环节中,通过引入新技术或改进现有技术,实现数据要素价值跃升的创造性活动。不同于传统生产要素,数据要素具有虚拟性与非均质性特征(Jones和Tonetti, 2020),其原始形态往往碎片化、非标准化。企业在数据采集、治理、分析等环节进行技术突破时,必须投入专用性要素,如购建定制化的软硬件数据资产以支撑海量数据的采集、存储与处理,引进算法工程师等专业人才以完成数据的清洗、建模与技术开发(Goldfarb和Tucker, 2019; Brynjolfsson等, 2021)。这些专用要素投入构成企业数据技术创新的物质基础与人力资本保障,其规模与质量直接影响数据技术创新的能力与潜力。然而,这类投入具有典型的资产专用性特征(彭正银等, 2024),高度依赖特定业务场景与技术体系,一旦投入便形成较高的沉没成本。

根据不完全契约理论,契约的天然不完全性使得事前产权界定成为影响经济主体投资激励的关键制度安排(Kaplan和Strömberg, 2003; 余明桂等, 2013)。在数据技术创新领域,由于数据来源复杂、权属界定困难,企业在采集与整合多元数据时可能面临侵权风险与合规不确定性(彭辉, 2022)。若数据权属存在争议,即便企业已投入大量专用性资产,创新活动仍可能中断,进而抑制其持续追加专用要素的意愿。同时,数据要素具有非竞争性与低边际复制成本特征(蔡跃洲和马文君, 2021),在缺乏明确产权保护的情况下,企业耗费大量资源所积累的数据技术创新成果容易被低成本复制,难以通过市场独占实现收益内部化,从而进一步弱化其进行专用要素投入的收益预期。

数据知识产权工作试点通过确立确权与登记规则,为企业增加数据专用要素投入提供了关键激励。首先,登记审查要求申请人提供数据来源合法性证明,从程序上规范了数据获取行为,降低了企业在数据采集环节面临的侵权风险与合规不确定性。当企业能够确认所使用数据

的合法来源时,其对前期投入可能因权属争议而沉没的担忧将显著缓解,从而更愿意对专用要素进行长期性投入。其次,登记赋权强化了对数据加工成果的保护,提升数据加工成果权益边界的可识别性和保护预期(申卫星,2020)。这种有限排他性的保护方式在保障企业创新收益预期的同时,也为他人合法开发同源数据保留了空间,从而增强企业持续投入专用要素的动力。由此,数据知识产权工作试点通过降低投入端风险、强化产出端收益,畅通了数据专用要素投入渠道,为企业数据技术创新提供了坚实的要素基础。基于上述分析,本文提出如下研究假设:

假设1:数据知识产权工作试点通过激励专用要素投入促进企业数据技术创新。

2.知识溢出渠道

数据技术创新不仅依赖于企业内部的专用要素投入,还高度受益于外部知识的流动与扩散。根据知识溢出理论,知识在组织间的传播能够降低后续研发的重复成本,促进技术交叉融合与新知识的生成(Jaffe等,1993;Bloom等,2013)。相较于传统生产要素,数据要素具有可复用性与正外部性特征,其价值不仅体现在持有者的内部使用中,更能在跨主体流动中实现持续增值(Jones和Tonetti,2020)。然而,在数据知识产权保护制度缺位时期,企业为防止数据算法模型等被无偿使用,往往采取数据封闭和低共享策略,限制数据技术相关成果的披露与共享,从而阻碍了数据要素的跨组织流动,抑制了知识溢出对数据技术创新的促进作用。

数据知识产权工作试点有效改善了数据技术创新领域的知识流动环境,为企业吸收外部知识创造了有利条件。一方面,试点政策通过引入标准化的信息披露机制,显著降低了企业搜寻与识别外部知识的成本。试点地区将保护对象限定于加工后的数据集合,且不禁止其他主体合法获取同源数据。在此基础上,要求登记主体公开数据集合的规模、处理规则、应用场景等关键信息,形成标准化的信息披露机制。这一安排使数据技术相关信息在权利边界清晰的前提下得以更广泛传播,企业能够通过公示平台快速了解行业技术动态、识别潜在合作伙伴的技术优势,从而更高效地定位可借鉴或可合作的知识资源(黄宏斌等,2023)。同时,信息的标准化披露还有利于促进不同领域数据标准的逐步趋同,使企业得以从其他行业的数据技术创新中汲取灵感,实现跨领域知识融合(张杨等,2026)。另一方面,试点政策通过强化数据技术成果的合法性与可交易性,有效拓宽了技术扩散与知识获取的路径。试点地区通过产权登记确立了数据要素的合法持有与经营权,推动数据知识产权融入交易授权、质押融资、作价入股等环节,将原本依赖商业秘密保护的数据集、核心算法等转化为可在合规框架下披露、授权与交易的资产。这不仅可以提升企业对外授权数据技术成果的意愿,也使其能够更便捷地获取其他主体已登记的技术成果信息,进而为试点地区企业吸收外部知识、借鉴已有技术方案提供更多路径。

此外,数据知识产权工作试点还通过强化创新网络效应,进一步放大了知识溢出的促进作用。试点政策带来的信息环境优化,有助于降低企业间的信息不对称与搜寻成本,促进企业集聚与网络连接,进而形成稳定的创新合作网络(宋周莺和刘卫东,2012)。在创新合作网络中,企业通过持续的技术交流实现创新资源的优势互补,采取技术引用、联合研发等方式完成外部知识的吸收与再创造(Bernal等,2022)。随着更多企业将数据技术成果纳入登记体系,信息公示平台可能发挥信息汇集与检索作用,为知识交流提供基础,能够持续吸引数据类企业加入,而新企业的进入又会进一步促进知识流动,从而形成知识溢出与主体集聚相互强化的良性循环(胡峰等,2020),最终推动区域数据技术创新水平的整体提升。由此,数据知识产权工作试点通过降低搜寻成本、促进技术扩散、强化网络效应,畅通了数据技术领域的知识溢出渠道,为企业数据技术创新提供了更丰富的外部知识来源。基于上述分析,本文提出如下研究假设:

假设2:数据知识产权工作试点通过推动知识溢出促进企业数据技术创新。

三、研究设计

(一)企业数据技术创新的测度

1.企业数据技术创新的定义

本文对企业数据技术创新进行概念界定,界定标准主要依据国家发展改革委等部门发布的《指导意见》及其附件《数据技术和产业重点发展方向》(简称《重点方向》)。《指导意见》以深化数据要素市场化配置改革、构建高水平数字经济体系为战略导向,明确将数据采集汇聚、流通交易、安全治理等纳入数据产业核心范畴,为理解数据技术创新的政策内涵提供了基本遵循。《重点方向》进一步从技术实施维度,系统阐释了数据采集、存储、治理、分析、交易、应用、安全七大领域的发展路径,为界定数据技术创新的边界提供了操作指引。基于上述政策框架,本文将企业数据技术创新定义为:企业通过引入新技术、新方法或改进现有技术,在数据采集、存储、治理、分析、交易、应用及安全等全生命周期环节中,提升数据处理效率、优化数据质量、增强数据安全与合规保障能力,最终实现数据要素价值跃升的创造性活动。

数据技术创新是数字经济领域技术体系的重要构成,但既有研究多侧重广义数字创新(胡增玺和马述忠,2023;陶锋等,2025),较少将数据技术作为独立范畴加以考察。本文聚焦于数据要素全生命周期的技术突破,尝试厘清数据技术与人工智能、大数据等通用数字技术之间的边界,以更准确地刻画数据要素特有的技术创新特征。在具体操作层面,本文采用企业专利申请数据识别微观层面的数据技术创新活动,主要基于以下考量:一方面,专利是企业保护创新成果、获取市场竞争优势的重要制度工具,专利申请行为能够有效表征企业的创新意愿与实际创新能力;另一方面,基于专利数据测度特定领域技术创新的方法已在学术界形成广泛共识(刘青和肖柏高,2023),具有较好的可比性与可操作性。

综上所述,本文构造的企业数据技术创新指标具有以下合理性:一是政策一致性。本研究严格遵循国家数据要素市场化配置改革导向,以《指导意见》等政策文件为基本依据,围绕数据全生命周期的关键技术突破构建专利测度指标,确保研究设计与数据要素市场化建设的目标要求一致。二是文献契合性。本文将研究视角从传统的广义数字创新细化至数据要素这一新兴领域,弥补现有研究对数据技术细分领域关注不足的局限,推动数字创新领域研究的深化。三是实践可行性。基于数据全生命周期的技术领域划分,既能覆盖从数据采集、存储等基础技术到数据交易、安全等应用技术的完整创新链条,又能与企业专利文本中“技术目标—技术手段”的表述逻辑匹配,从而为基于专利数据识别企业数据技术创新提供可操作的判断标准。

2.基于生成式大语言模型的专利识别

本研究的难点在于如何准确识别数据技术专利。现有研究主要依据国际专利分类(international patent classification, IPC)体系或结合关键词检索的方法进行特定领域的专利分类(陈楠等,2024),但这些方法在识别数据技术专利时存在明显不足:一方面,IPC分类体系基于技术主体(如物理、电学等)设立分类标准,与本文按数据生命周期功能划分专利的逻辑并不完全一致,这种逻辑错配使得IPC分类体系在数据技术专利识别中存在类别缺失与错误归类问题,可能降低特定数据技术专利的识别准确性^①。另一方面,基于专利文本的关键词检索法存在语义理解的局限性,该方法对文本语义的理解停留在表层,容易因表述多样性和语义复杂

^①数据生命周期中的诸多关键创新环节在IPC体系内缺乏精准对应的分类类目,如数据资产化相关的估值与定价技术、数据可信流通场景下的新型隐私计算技术等前沿专利,往往只能被归入“G06F”(电数字数据处理)等宽泛类目,导致其技术独特性被掩盖,且那些具有创新实质但表述或实现方式特殊的数据技术型专利可能因此被遗漏。同时,宽泛的IPC分类号下又混杂着与数据技术创新无关的专利,例如,“G06F”大类中既包含具备实质创新的数据处理算法与架构专利,也涵盖常规软件应用及计算机硬件维护等无关专利。

性出现遗漏或误判的情况^①。

为解决上述方法在分类准确性上的不足,本研究引入生成式大语言模型(large language model, LLM)作为数据技术专利的识别工具。近年来,生成式大语言模型凭借其卓越的语义理解能力,在经济学与管理学领域的学术研究中逐渐得到应用(Chen和Chan, 2024; Doshi等, 2025)。该类模型能够基于提示语(Prompt)工程直接生成结构化判断,克服传统分类方法的局限,并显著降低对大规模人工标注数据的依赖,在提升识别效率的同时有效控制研究成本,为专利识别这一关键任务提供了兼具准确性、可扩展性的创新解决方案(赵宣凯等, 2025)。在具体操作中,本研究采用Qwen大语言模型,通过系统化的Prompt工程对专利摘要文本进行深度语义理解,从而识别数据技术专利,实施步骤如下:

第一步,构建包含基础判断标准、典型示例与输出格式约束的多层级Prompt指令体系。在基础判断标准层面,本文明确将“数据技术专利”定义为以提升数据要素价值为目标、涉及数据采集、存储、治理、分析、交易或安全等环节的实质性技术创新;引导示例提供正、负两类典型样例,如“提出基于联邦学习的跨机构数据协同分析方法,解决数据隐私保护与共享矛盾”被判定为数据技术专利,而“改进计算机硬件散热结构,仅在数据中心场景应用”则判定为非数据技术专利;输出约束要求模型输出二元分类结果,并附200字以内的判定依据,以便后续复核与溯源。

第二步,开展模型优化与迭代验证。从国家知识产权局数据库抽取2 000条2015—2023年间的专利摘要文本构建验证集,通过人工标注与模型输出的比对,实施多轮Prompt优化。针对初期将“常规数据库备份流程优化”误判为数据技术专利的问题,在Prompt中补充“仅涉及常规流程优化、未实现技术原理突破的除外”的排除规则;同时扩展“隐私计算”“数据确权”等关键术语的语义覆盖范围,降低漏判率。

第三步,实施批量识别与数据集构建。将优化后的Prompt与Qwen模型集成,对上市公司专利摘要进行自动化处理,逐篇输出分类结果与判定依据,并同步提取申请年份、申请人等元数据,最终构建企业—年度层面的数据技术专利面板数据,为后续实证研究提供数据基础。

3.数据技术专利识别的有效性检验

为评估基于大语言模型的数据技术专利识别结果的有效性,本文从时间趋势、地区差异和行为一致性三个维度展开检验。

第一,从时间趋势看,企业数据技术专利的申请行为呈现出显著的“政策驱动型”演化特征。本文以各年度申请数据技术专利的上市公司占比衡量创新广度,以企业平均专利申请数量衡量创新强度。如图1所示,2015年至2021年,申请企业占比由5.5%逐步攀升至14.9%,平均专利申请量在0.35件至0.77件之间波动,反映出数据技术从萌芽走向稳步积累的阶段特征。随着2022年数据知识产权工作试点的启动,两项指标均实现跨越式增长:申请企业占比从2021年的14.9%跃升至2024年的29.9%,增幅超过

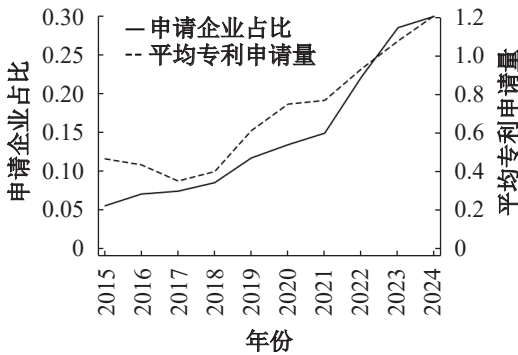


图1 数据技术专利发展趋势

^①具体表现为:第一,表述多样性导致漏判。同一数据技术概念往往存在多种表述方式,如“隐私计算”技术可能表述为“多方安全计算”“联邦学习”或“差分隐私”等,传统关键词库难以覆盖所有语义变体。第二,语义复杂性导致误判。某些专利文本中虽包含数据技术相关词汇,但实际并不涉及实质性创新。例如,“数据库备份方法改进”专利虽包含“数据”关键词,但可能仅涉及常规流程优化;而“数据血缘关系追踪”专利可能因未包含预设关键词而被遗漏。

一倍;平均专利申请量同步从0.77件增长至1.21件,增长幅度达57.1%^①。这一变化与数据知识产权制度建设的政策脉络高度吻合,有力验证了本文数据技术专利识别结果的有效性,表明其能够捕捉制度激励下的微观创新行为。

第二,从地区差异看,企业数据技术创新活跃度呈东高西低的梯度格局。本文以各省级地区上市公司中申请数据技术专利的企业比例衡量地区数据技术创新活跃度。如图2所示,北京、广东、上海、江苏、福建、浙江等东部省级地区的申请比例显著较高,而青海、甘肃等西部省份的申请比例相对较低。这一分布特征与我国数字经济发展的梯次格局高度吻合。据中国信息通信研究院《中国数字经济发展研究报告(2024年)》,2023年北京、上海、广东、福建、浙江等地数字经济占地区生产总值的比重均超过50%,浙江、上海、北京、山东、江苏、广东等地的数字经济增速也高于全国平均水平。上述结果表明,本文识别的数据技术专利指标具有良好的外部效度,能够较好反映我国区域创新发展的实际态势。

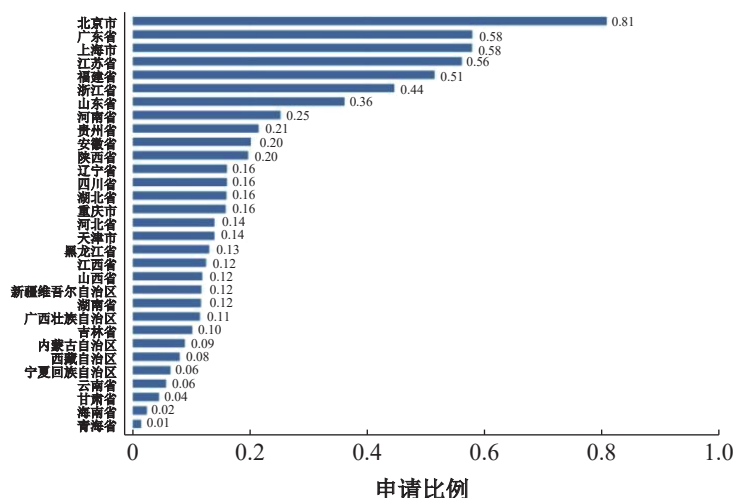


图2 各省级地区上市公司中申请数据技术专利的企业比例

第三,从行为一致性看,企业的数据技术专利产出与年报中的数字化叙事形成正向关联。数据技术专利从创新产出维度反映企业实质性数字化能力,数字化叙事则从信息披露维度体现企业的数字化认知,二者相互印证。理论上,企业在数据技术领域的持续创新会促使其在年报中更积极地强调数字化议题,以传递竞争优势信号。因此,有效的数据技术专利测度应对数字化叙事具有解释力。本文参考吴非等(2021)的研究,以上市公司年报数字化关键词词频的自然对数作为数字化叙事的代理变量,并将其作为被解释变量,以当年申请数据技术专利数量加1的自然对数作为解释变量进行回归^②。结果表明,数据技术专利对数字化叙事具有显著正向影响,验证了专利识别结果与企业数字化导向的一致性,为识别有效性提供了有力支撑。

(二)模型设定与变量定义

本研究以全国首批数据知识产权工作试点为准自然实验,通过构建双重差分模型评估数据知识产权工作试点政策效果,以检验数据知识产权保护对企业数据技术创新的影响,模型设定如下:

$$DataInvo_{j,i,t} = \alpha_0 + \alpha_1 Treat_i \times Post_t + \alpha_2 X_{j,i,t} + \mu_i + \lambda_t + \psi_{j,t} + \varepsilon_{j,i,t} \quad (1)$$

①全样本均值因大量零值企业存在而偏低。以2024年为例,若仅统计有专利申请的企业,平均申请量约为6件,表明创新活跃企业的实际产出强度更高。

②限于篇幅,检验结果留存备案。

其中, $DataInvo_{j,i,t}$ 是被解释变量, 指 j 行业中 i 企业在 t 年的数据技术创新水平, 本文采用企业当年申请的数据技术专利数量加1的自然对数值衡量。 $Treat_t \times Post_t$ 是解释变量, 表示企业是否受到数据知识产权工作试点影响。 $Treat_t$ 是处理组虚拟变量, 若企业所在城市为数据知识产权工作试点区域, 则取值为1, 否则为0^①。 $Post_t$ 为政策时点虚拟变量, 2022年及之后年份取值为1, 之前年份取值为0。 $X_{j,i,t}$ 表示一系列控制变量, 借鉴已有研究(刘青和肖柏高, 2023; 陈楠等, 2024), 本文控制了企业特征与宏观经济两类变量, 变量定义详见表1。 μ_i 、 λ_t 分别表示企业固定效应和年份固定效应, 由于不同行业的数据密集度、创新需求及政策敏感度存在显著差异, 这些行业特有的时间趋势可能独立于数据知识产权工作试点政策而影响企业数据技术创新, 从而破坏处理组和对照组的平行趋势假设。 因此, 本文还控制了行业与年份交互固定效应 $\psi_{j,t}$ ^②。 $\varepsilon_{j,i,t}$ 为随机扰动项。 系数 α_1 反映了数据知识产权工作试点对企业数据技术创新的净影响, 若 α_1 显著为正, 则说明数据知识产权工作试点能够促进企业数据技术创新。

表 1 主要变量定义

变量符号	变量名称	变量定义
$DataInvo$	企业数据技术创新	当年数据技术专利申请数量加1的自然对数值
$Treat$	处理组虚拟变量	所在城市为数据知识产权工作试点地区时为1, 否则为0
$Post$	政策时点虚拟变量	2022年及之后为1, 否则为0
$Size$	企业规模	总资产的自然对数值
Lev	资产负债率	总负债/总资产
Roe	净资产收益率	净利润/净资产
$Indep$	独立董事比例	独立董事数量/董事会规模
Top	前十大股东持股比例	前十大股东持股数量/总股本
$Instshare$	机构投资者持股比例	机构投资者持股数量/总股本
Age	企业年龄	当前年份减去企业成立年份加1的自然对数值
Soe	产权性质	国有企业取1, 非国有企业取0
$Lnpgdp$	经济水平	所在城市人均GDP的自然对数值
$Fiscal$	财政支出	所在城市财政一般预算内支出/地区生产总值
$Industry$	产业结构	所在城市第三产业增加值/地区生产总值
Fin	金融资源	所在城市年末金融机构贷款余额/地区生产总值

(三)样本选择、数据来源与描述性统计

考虑到2015年国务院正式印发《促进大数据发展行动纲要》, 首次从国家战略层面系统部署大数据发展工作, 标志着我国数据要素市场化建设的起步, 因此, 本文选取2015—2024年A股上市公司为初始样本, 并剔除金融类、ST/*ST及关键变量缺失样本, 最终得到33 880个观测值。 本文使用的企业专利申请数据来源于国家知识产权局专利数据库, 其他企业层面数据来源于国泰安(CSMAR)数据库。 城市特征的相关数据主要来自历年《中国城市统计年鉴》、国家统计局等权威部门。 本文对连续型变量实施上下1%缩尾处理。 表2报告了主要变量的描述性统计结果。

四、实证结果

(一)基准回归

表3汇报了模型(1)的逐步回归结果, $Treat \times Post$ 的估计系数均显著为正, 说明数据知识

①由于第二批数据知识产权工作试点于样本期末(2024年)才开始实施, 缺乏政策实施后的有效观测数据, 无法满足双重差分法对政策冲击前后均须具备足够观测期的基本前提, 因此本文将首批数据知识产权工作试点地区(2022年启动)企业作为处理组。

②例如, 信息技术、互联网等行业通常具有较高的数据密集度和创新需求, 其对数据知识产权保护政策的敏感度可能天然强于农林牧渔等传统行业。

产权工作试点能够显著促进企业数据技术创新。在经济意义上,根据第(4)列的结果, $Treat \times Post$ 估计系数为0.0759,表明在其他条件不变的情况下,相较于未开展数据知识产权工作试点的地区,试点地区企业的数据技术创新水平提升0.0759,相当于均值的35.89%(=0.0759/0.2115 $\times$ 100%),提升效果明显。因此,无论在统计意义上还是经济意义上,数据知识产权工作试点确实促进了企业数据技术创新。

表2 描述性统计

变量	观测值	均值	标准差	最小值	中位数	最大值
<i>DataInvo</i>	33 880	0.2115	0.5902	0.0000	0.0000	3.4490
<i>Treat</i>	33 880	0.6763	0.4679	0.0000	1.0000	1.0000
<i>Post</i>	33 880	0.3631	0.4809	0.0000	0.0000	1.0000
<i>Size</i>	33 880	22.3169	1.3037	20.0067	22.1164	26.3746
<i>Lev</i>	33 880	0.4121	0.1997	0.0595	0.4026	0.8895
<i>Roe</i>	33 880	0.0468	0.1440	-0.7866	0.0655	0.3241
<i>Indep</i>	33 880	0.3784	0.0533	0.3333	0.3636	0.5714
<i>Top</i>	33 880	0.5849	0.1514	0.2407	0.5914	0.9049
<i>Instshare</i>	33 880	0.4212	0.2483	0.0037	0.4307	0.9092
<i>Age</i>	33 880	3.0365	0.2972	2.1972	3.0445	3.6636
<i>Soe</i>	33 880	0.3120	0.4633	0.0000	0.0000	1.0000
<i>Lnpgdp</i>	33 880	11.6348	0.4552	10.3274	11.7307	12.2560
<i>Fiscal</i>	33 880	0.1551	0.0531	0.0777	0.1407	0.3692
<i>Industry</i>	33 880	0.5842	0.1238	0.3433	0.5731	0.8584
<i>Fin</i>	33 880	1.3526	0.7238	0.2840	1.0185	3.2480

表3 基准回归结果

变量	(1)	(2)	(3)	(4)
<i>Treat</i> $\times$ <i>Post</i>	0.0816*** (7.3190)	0.0794*** (7.1566)	0.0829*** (7.5293)	0.0759*** (6.7448)
企业层面控制变量	否	是	是	是
城市层面控制变量	否	否	是	是
企业固定效应	是	是	是	是
年份固定效应	是	是	是	是
行业—年份固定效应	否	否	否	是
观测值	33 880	33 880	33 880	33 880
调整后的 R^2	0.7063	0.7074	0.7079	0.7177

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

(二)平行趋势检验与敏感性分析

本文采用事件研究法检验平行趋势假设,模型设定如下:

$$DataInvo_{j,i,t} = \beta_0 + \beta_1 EVENT_{i,t}^n + \beta_2 X_{j,i,t} + \mu_i + \lambda_t + \psi_{j,t} + \varepsilon_{j,i,t} \quad (2)$$

其中, $EVENT_{i,t}^n$ 是样本观测年份相对于数据知识产权工作试点的时间距离变量,其余变量设定与基准回归一致。本文以数据知识产权工作试点前1年为基期,考察试点前7年至后2年,处理组与对照组企业数据技术创新水平相对于基期的差异。由图3可知,数据知识产权工作试点前的系数估计值均不显著,而试点后的回归系数呈显著上升趋势,满足平行趋势检验假设。此外,本文还借鉴Rambachan和Roth(2023)的研究进行敏感性分析。检验结果显示^①,在设定的相对偏离限制下,数据知识产权工作试点当年、后一年及后两年,数据知识产权保护对企业数据技术创新的促进作用依然稳健成立。

^①限于篇幅,平行趋势敏感性检验结果留存备索。

(三)稳健性检验^①

本文从以下几个方面进行了稳健性检验:(1)匹配—双重差分估计;(2)排除同期干扰政策;(3)替换被解释变量;(4)政策实施强度检验;(5)安慰剂检验;(6)排除性检验;(7)剔除直辖市、滞后解释变量及剔除2024年样本等其他稳健性检验。上述一系列检验结果表明,数据知识产权工作试点显著促进企业数据技术创新的结论依然稳健。

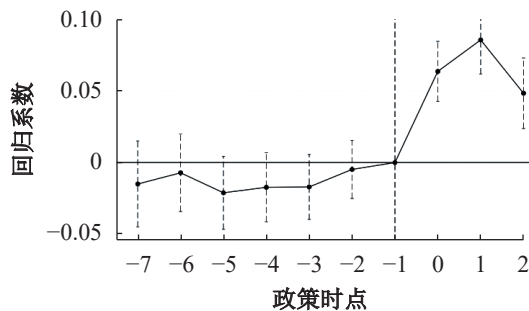


图3 平行趋势检验

五、进一步分析

(一)机制检验

根据理论分析,数据知识产权工作试点通过专用要素投入和知识溢出渠道促进企业数据技术创新。为验证上述作用机制是否成立,本文借鉴江艇(2022)的研究思路,设定如下模型检验核心解释变量对机制变量的影响:

$$M_{j,i,t} = \gamma_0 + \gamma_1 Treat_i \times Post_t + \gamma_2 X_{j,i,t} + \mu_i + \lambda_t + \psi_{j,t} + \varepsilon_{j,i,t} \quad (3)$$

其中, $M_{j,i,t}$ 是待检验的机制变量,其余变量与基准回归模型保持一致。

1.专用要素投入渠道

理论分析指出,数据知识产权工作试点能够激励企业增加数据技术研发所需的专用要素投入。为验证这一机制,本文从数据人才招聘与数据资产积累两个维度选取代理变量进行检验。

首先,数据技术研发高度依赖具备数据清洗、算法开发、建模等技能的专业人才。当前网络招聘是企业获取人才的重要渠道,本文借鉴余典范等(2025)的研究思路,基于上市公司互联网招聘文本信息构建企业数据人才投入指标($Data_Employee$)。具体而言,通过爬取各大招聘网站发布的上市公司招聘公告,采用Qwen大语言模型对招聘岗位的职位描述进行语义识别,判断其是否属于数据劳动者,包括算法工程师、数据分析师、大数据架构师等。将企业当年经模型识别为数据职业相关的招聘公告数量加1后取自然对数,以衡量企业在数据技术研发领域的人力资本投入强度。根据表4列(1)的回归结果,数据知识产权保护显著提升了企业对数据技术人才的需求。

表4 机制检验结果

变量	(1) <i>Data_Employee</i>	(2) <i>Data_Asset</i>	(3) <i>Data_Know</i>	(4) <i>Data_Entry</i>
<i>Treat</i> × <i>Post</i>	0.1687*** (53.7939)	0.0825*** (4.2436)	0.4692*** (10.0594)	0.1202** (2.5094)
控制变量	是	是	是	是
企业/城市固定效应	是	是	是	是
年份固定效应	是	是	是	是
行业—年份固定效应	是	是	是	否
观测值	33 880	33 880	33 880	2967
调整后的 R^2	0.7431	0.7565	0.4043	0.9189

注:列(4)城市层面控制变量包括经济水平、财政支出、产业结构和金融资源。

^①限于篇幅,稳健性检验结果留存备索。

其次,数据资产是企业开展数据技术创新的物质基础,涉及数据采集、存储、处理与分析等环节。本文借鉴陈志斌和杨靖(2025)的研究,基于企业财务报表附注中披露的无形资产明细项,同样采用Qwen大语言模型对无形资产项目名称及描述进行语义识别,判断其是否属于数据相关资产,包括数据库、数据平台、数据管理系统等。将企业当年经模型识别为数据资产的金额加总后取自然对数,构建数据资产规模指标(*Data_Asset*),以衡量企业在数据相关领域的专用性资产投入水平。根据表4列(2)的回归结果,数据知识产权保护显著扩大了企业的数据资产规模。

上述结果表明,数据知识产权工作试点通过激励企业增加数据人才储备与数据资产积累,为数据技术创新提供了坚实的人力资本与物质基础,假设1得以验证。

2.知识溢出渠道

理论分析表明,数据知识产权工作试点能够通过改善知识流动环境,促进知识在企业间溢出与扩散,从而为企业数据技术创新提供更丰富的外部知识来源。为验证这一机制,本文从企业层面的知识吸收与区域层面的创业涌现两个维度选取代理变量进行检验。

首先,借鉴Acemoglu等(2016)、张杨等(2026)的研究思路,本文基于企业数据技术专利的引用关系构建知识溢出吸收指标(*Data_Know*)。具体而言,定义引用权重 $\omega_{i,m,t} = CI_{i,m,t} / \sum_{m=1}^M CI_{i,m,t}$,其中 $CI_{i,m,t}$ 为企业*i*在第*t*年对企业*m*的数据技术专利引用次数,分母则代表企业*i*当年的引用总数(覆盖全样本公司),该权重刻画了企业*m*对企业*i*数据技术创新的相对重要性。在此基础上,测度企业*i*吸收的外部知识溢出水平: $DataKnow_{i,t} = \sum_{m \neq i} \sum_{\tau=1}^5 \omega_{i,m,t-\tau} \ln H_{m,t-\tau}$ 。式中, $H_{m,t-\tau}$ 为企业*m*滞后 τ 年的数据技术专利数量。该指标刻画了企业*i*在数据技术创新过程中所吸收的知识溢出规模。该指标值越大,表明企业从外部吸收的数据技术知识规模越大,知识溢出效应越强。表4列(3)的回归结果显示,数据知识产权保护显著促进了企业对外部数据技术知识的学习与吸收。

其次,知识溢出不仅体现为技术成果的直接吸收利用,还表现为对区域创新创业生态的整体赋能。根据知识溢出创业理论,知识禀赋丰富的区域能够产生广泛的创业机会,新进入者可借助在位企业溢出的知识实现商业化,从而提高创业成功率(Acs等,2009)。而新企业进入又会进一步促进知识流动,为在位企业提供更多可供学习与合作的创新主体,从而强化知识溢出对技术创新的持续推动作用(胡峰等,2020)。为此,本文在城市层面考察数据知识产权工作试点对新注册数据类企业的影响。具体而言,基于全国工商企业注册数据,采用Qwen大语言模型对企业经营范围进行语义识别,判断其是否属于数据类企业(如从事数据采集、标注、交易服务等业务的企业)。在此基础上,将企业所在城市当年新注册数据类企业数量加1后取自然对数,构建区域数据企业进入指标(*Data_Entry*),以间接检验数据知识产权工作试点的知识溢出效应。表4列(4)的回归结果显示,数据知识产权保护所激发的知识溢出吸引了更多的数据类企业进入市场。

上述结果表明,数据知识产权工作试点通过畅通知识溢出渠道,在促进企业吸收外部技术成果的同时有效激发了区域创业活力,为企业数据技术创新提供了持续的外部知识来源与创新生态支撑,假设2得以验证。

(二)异质性分析

1.企业产权性质

产权性质的不同可能影响数据知识产权保护对企业数据技术创新的提升效果。与非国有企业相比,国有企业在执行政府政策时更为坚决,对数据知识产权工作试点这类政策的引导具有更强行动力(姚加权等,2024),能够快速响应试点要求推进数据确权、技术研发等工作。同

时,数据技术研发既需要硬件投入,也依赖高端技术人才支撑,会显著增加企业资金压力与人才需求,国有企业拥有更丰富的资源储备与更高的风险承担水平,其数据技术研发的投入强度与转化效率相对更强。因此,数据知识产权工作试点对国有企业数据技术创新的激励效果可能更为显著。为检验这一假设,本文按照企业产权性质进行分组回归。表5列(1)至列(2)的回归结果显示,数据知识产权保护对国有企业与非国有企业的技术创新均存在显著促进效应,且组间系数差异检验结果表明,该促进效应在国有企业中更强。

表 5 企业产权性质与上市板块属性的异质性分析

变量	企业产权性质		上市板块属性	
	(1) 国有企业	(2) 非国有企业	(3) 中小科创企业	(4) 主板企业
<i>Treat</i> × <i>Post</i>	0.1084*** (4.4800)	0.0666*** (5.3806)	0.0868*** (3.8478)	0.0729*** (5.5097)
组间系数差异 <i>p</i> 值	0.0400		0.3040	
控制变量	是	是	是	是
企业固定效应	是	是	是	是
年份固定效应	是	是	是	是
行业—年份固定效应	是	是	是	是
观测值	10571	23309	9545	24335
调整后的 <i>R</i> ²	0.7547	0.7000	0.6974	0.7214

注:组间系数差异*p*值为经过500次自抽样的费舍尔组合检验方法得到。下表同。

2.上市板块属性

企业所属的上市板块映射其规模特征与科创属性。相较于业务相对成熟的主板企业,科创板、创业板及北交所汇聚了大量高成长性中小科技企业,其创新导向更为突出,对数据知识产权保护的政策信号也可能更为敏锐。为此,本文将科创板、创业板与北交所企业合并为中小科创板块,将沪深主板企业归为另一组,进行分组回归。表5列(3)至列(4)的回归结果显示,数据知识产权工作试点对两类企业的数据技术创新均有显著促进效应,且组间系数差异未通过显著性检验。这一结果表明,数据知识产权保护对企业数据技术创新的激励具有普惠性特征,制度红利广泛覆盖不同类型企业。无论是中小科创企业还是主板企业,均在数据知识产权制度完善过程中受益。这说明,在数据要素时代,传统主板企业同样具有强烈的数据技术创新需求,其创新响应强度并不弱于中小科创企业,数据知识产权保护的激励效应未因企业规模特征与科创属性而呈现显著分化。

3.数据价值链分工

既有研究指出,企业在数据价值创造中的角色可划分为数据原生、数据次生与数据应用三类,各自聚焦于不同的价值链环节,其价值创造方式与所需数据资源特质存在显著差异(魏江和应瑛,2025)。这一分工特征可能引致数据知识产权保护对数据技术创新的促进效应呈现异质性。为验证这一推断,本文参考《数字经济及其核心产业统计分类(2021)》,结合证监会行业代码对样本进行类型划分:将信息传输、软件和信息技术服务业(I63、I64、I65)界定为数据原生企业,将计算机、通信和其他电子设备制造业(C39)及部分专业技术服务业(M73、M74)界定为数据次生企业,其余行业归为数据应用企业。表6列(1)至列(3)的回归结果显示,数据知识产权工作试点对三类企业的数据技术创新均具有促进效应。组间系数差异检验进一步表明,试点对数据原生企业的促进效应显著强于数据次生与数据应用企业,而后两类企业之间不存在显著差异。这可能是因为,数据原生企业作为数据要素的源头供给者,其业务活动覆盖数据价值链

的采集存储、加工分析、场景应用及流通交易等全生命周期环节,数据技术创新活动贯穿始终,对数据处理的持续投入依赖度高,因而对数据知识产权保护更为敏感。相比之下,数据次生企业侧重加工分析与流通交易等环节,数据应用企业则聚焦于场景化应用,两者的数据技术创新多围绕特定环节展开,创新活动的广度与复杂度相对有限。随着数据知识产权工作试点推进,数据来源、处理规则及应用场景等信息逐步公示,后两类企业得以凭借更低的信息搜寻成本获取外部知识资源,从而在一定程度上弥补其在数据获取与处理环节的劣势,支撑其数据技术创新活动。

表 6 数据价值链分工与地区数字基础设施的异质性分析

变量	数据价值链分工			地区数字基础设施	
	(1) 数据原生企业	(2) 数据次生企业	(3) 数据应用企业	(4) 完善	(5) 薄弱
<i>Treat</i> × <i>Post</i>	0.2557*** (2.6974)	0.1111*** (2.8121)	0.0752*** (6.8645)	0.0721*** (4.3848)	0.0216 (1.1488)
组间系数差异 <i>p</i> 值	0.0420	0.0340	0.1940	0.0240	
控制变量	是	是	是	是	是
企业固定效应	是	是	是	是	是
年份固定效应	是	是	是	是	是
行业—年份固定效应	是	是	是	是	是
观测值	2298	3904	27678	23845	10035
调整后的 <i>R</i> ²	0.6823	0.7281	0.6946	0.7219	0.6502

注:数据价值链分工异质性分析的组间系数差异*p*值为两两组合检验所得,数据原生企业与数据次生企业组间系数差异*p*值为0.0420,数据原生企业与数据应用企业组间系数差异*p*值为0.0340,数据次生企业与数据应用企业组间系数差异*p*值为0.1940。

4.地区数字基础设施

数据具有非竞争性与高流动性,其价值实现高度依赖于跨主体传输、匹配与再利用,而数字基础设施正是支撑这一过程的关键底座。完善的网络通信、大数据平台及算力资源等,有助于降低数据传输与处理成本、提升匹配效率,为数据要素流通与利用创造必要条件,同时也为企业开展专用要素投入提供技术支撑。在数字基础设施较为完善的地区,数据知识产权工作试点的信息公示功能更容易发挥效用:公示所披露的数据来源、处理规则及应用场景等信息,能够被其他市场主体及时获取并加以利用,进而更有效地促进数据资源与技术信息的匹配与再利用。相反,在数字基础设施相对薄弱的地区,数据传输、处理与应用各环节均受技术条件制约,这不仅限制了企业进行专用要素投入的能力,也削弱了其对外部知识的获取与吸收,从而抑制了数据知识产权保护对企业数据技术创新的促进效果。基于此,本文借鉴赵涛等(2020)的方法,选取光缆密度、互联网普及率与移动电话普及率等指标,运用熵值法构建数字基础设施建设指数。该指数取值越大,表示城市数字基础设施发展水平越高。本文依据该指数中位数将样本划分为数字基础设施完善与薄弱两组进行分组检验。表6列(4)至列(5)的回归结果显示,在数字基础设施较完善的地区,数据知识产权保护对企业数据技术创新的促进效应更强。

(三)经济后果分析

前述分析验证了数据知识产权工作试点能够促进企业数据技术创新。以往研究表明,大数据及其相关技术的应用在优化资源配置、降低信息摩擦等方面发挥着积极作用,可以有效提升企业全要素生产率(史丹和孙光林,2024)。而数据技术创新是推动数据资源开发利用、挖掘数据价值创造潜力、发挥数据生产要素作用的关键支撑。那么,数据知识产权保护所激发的数据技术创新,能否转化为企业全要素生产率的提升,提升企业生产效率和价值创造能力?为此,本

文使用OP法测度企业未来一期全要素生产率(TFP_{t+1}),采用逐步回归模型考察数据技术创新与企业全要素生产率之间的关系。如表7第(1)列所示,数据知识产权工作试点对企业全要素生产率具有促进作用;第(2)列为基准回归估计结果;第(3)列为同时纳入 $Treat \times Post$ 和 $DataInvo$ 的回归结果, $DataInvo$ 估计系数在1%水平上显著为正,同时 $Treat \times Post$ 估计系数依然显著为正,但数值较第(1)列有所减少。以上结果表明,数据知识产权保护能够通过促进数据技术创新,有效提升企业全要素生产率。

表 7 经济后果检验结果

变量	(1) TFP_{t+1}	(2) $DataInvo$	(3) TFP_{t+1}
$Treat \times Post$	0.0367*** (2.6716)	0.0759*** (6.7448)	0.0347** (2.5270)
$DataInvo$			0.0250*** (3.7567)
控制变量	是	是	是
企业固定效应	是	是	是
年份固定效应	是	是	是
行业—年份固定效应	是	是	是
观测值	33 880	33 880	33 880
调整后的 R^2	0.8854	0.7177	0.8887

六、结论与启示

在数字经济时代,数据技术创新是企业释放数据要素价值、构筑核心竞争力的关键引擎,更是推动数字经济高质量发展、赋能中国式现代化的重要支撑。本文以2022年全国首批数据知识产权工作试点为准自然实验,利用2015—2024年A股上市公司数据,创新性地引入生成式大语言模型识别企业数据技术专利,构建双重差分模型考察数据知识产权保护对企业数据技术创新的影响。研究发现,数据知识产权工作试点显著促进了企业数据技术创新。这一结果在多项稳健性检验中保持稳定。机制检验表明,数据知识产权保护主要通过激励专用要素投入以及促进知识溢出发挥作用。异质性分析显示,上述促进作用在国有企业、数据原生企业以及数字基础设施较完善地区更为突出,而在不同科创属性企业间未呈现显著差异。此外,经济后果分析表明,数据知识产权保护通过促进企业数据技术创新,进一步提升了企业全要素生产率。

基于上述结论,本文提出如下政策启示:

第一,深化数据知识产权制度建设,夯实企业数据技术创新基础。本文研究发现,数据知识产权工作试点显著激励了企业数据技术创新,凸显了数据知识产权保护在激发企业创新动力中的基础性作用。因此,应持续推进数据知识产权制度在全国范围的落地与完善,加快构建统一规范、权责清晰的数据产权登记体系。具体而言,可进一步细化数据知识产权登记审查标准,明确数据来源合法性、处理合规性等核心要件,明确登记证书在权利主张、交易审查和争议解决中的证明功能及适用边界。同时,应加强数据知识产权保护与运用的协同,推动登记信息在融资、交易、合作等场景中的广泛应用,增强企业对数据资产的价值预期,引导其加大数据技术研发投入,为企业数据技术创新活动提供稳定、可预期的制度环境。

第二,健全专用要素投入与知识溢出机制,增强政策赋能效果。本文研究表明,数据知识产权保护通过激励企业加大数据人才、数据资产等专用要素投入,并促进知识溢出,有效推动了企业数据技术创新。为此,应着力构建支撑专用要素积累与知识流动的制度生态。在专用要素

投入层面,鼓励企业扩大数据资产购置与数据技术人才引进、培养和实际投入规模,综合运用税收优惠、补贴激励等政策工具降低专用要素投入成本,夯实数据技术创新的物质基础与人力资本保障。在知识溢出层面,一方面推动建设跨区域、跨行业的数据知识产权信息服务平台,促进数据技术成果的开放共享与有效传播,降低企业搜寻与合作成本;另一方面,持续完善数据要素市场规则,提升数据权属界定与交易规则的清晰度和可预期性,为更多市场主体进入数据领域营造良好环境,不断增强区域创新主体密度与协作活力。

第三,注重差异化推进路径,提升政策适配性。本文异质性分析揭示,数据知识产权保护的创新激励效应在国有企业、数据原生企业以及数字基础设施较完善地区更为突出,且在不同科创属性企业间未观察到显著的效应差异。因此,在政策推进过程中,应充分考虑企业类型与区域发展基础的差异。对于国有企业与数据原生企业,应发挥其“头雁效应”与创新引领作用,通过赋予更大创新自主权、支持组建创新联合体等方式,鼓励其在数据确权、流通与标准制定中先行先试,开展前沿数据技术研发。对于数字基础设施薄弱地区,应同步加强算力网络、数据中心等基础设施建设,为政策落地提供硬件支撑。同时,鉴于政策效应在不同科创属性企业间具有普遍性,应推动数据知识产权登记与服务体系向各类企业全面覆盖,鼓励企业围绕数据全生命周期开展核心技术攻关,推动数据技术创新与产业升级深度融合,为经济高质量发展注入持续动能。

主要参考文献

- [1]蔡跃洲, 马文君. 数据要素对高质量发展影响与数据流动制约[J]. 数量经济技术经济研究, 2021, 38(3): 64-83.
- [2]陈楠, 蔡跃洲, 马文君. 企业数据技术能力对市值和利润的影响——基于大数据专利的机制检验和互补性投入分析[J]. 财贸经济, 2024, 45(1): 106-123.
- [3]陈志斌, 杨靖. 数据资产化的企业价值效应研究——来自数据资产入表新规颁布的经验证据[J]. 会计研究, 2025, (5): 18-30.
- [4]戴魁早, 王思曼, 黄姿. 数据交易平台建设如何影响企业全要素生产率[J]. 经济学动态, 2023, (12): 58-75.
- [5]郭丰, 任毅, 段云桂, 等. 知识产权保护对企业数字技术创新的影响效应研究——来自中国上市公司的证据[J]. 管理评论, 2025, 37(6): 77-89.
- [6]胡峰, 裘讯, 黄登峰, 等. 协同创新知识溢出风险管理框架: 表征与认知[J]. 科学学研究, 2020, 38(6): 1048-1056.
- [7]胡增玺, 马述忠. 市场一体化对企业数字创新的影响——兼论数字创新衡量方法[J]. 经济研究, 2023, 58(6): 155-172.
- [8]黄宏斌, 梁慧丽, 许晨辉. 数字化转型驱动了企业协同创新吗?[J]. 现代财经(天津财经大学学报), 2023, 43(11): 96-113.
- [9]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.
- [10]李健, 董小凡, 张金林, 等. 数据资产对企业创新投入的影响研究[J]. 外国经济与管理, 2023, 45(12): 18-33.
- [11]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, 58(2): 74-90.
- [12]彭正银, 楚伯微, 罗贯擎, 等. 数据专用性对平台包络的影响效应研究[J]. 经济管理, 2024, 46(5): 72-91.
- [13]申卫星. 论数据用益权[J]. 中国社会科学, 2020, (11): 110-131, 207.
- [14]史丹, 孙光林. 数据要素与新质生产力: 基于企业全要素生产率视角[J]. 经济理论与经济管理, 2024, 44(4): 12-30.
- [15]宋周葛, 刘卫东. 信息时代的企业区位研究[J]. 地理学报, 2012, 67(4): 479-489.
- [16]陶锋, 翟少轩, 王峤. 数字经济政策与传统企业跨界数字创新[J]. 中国工业经济, 2025, (2): 118-136.
- [17]王伟玲. 中国数据产权制度构建研究[J]. 经济纵横, 2024, (1): 78-85.
- [18]魏江, 应瑛. 企业数据基础观的构建[J]. 中国社会科学, 2025, (12): 53-68, 200.
- [19]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, 37(7): 130-144, 10.
- [20]谢地, 王荣基, 贺城. 数据要素市场化配置赋能企业新质生产力发展[J]. 经济学动态, 2025, (5): 19-37.

- [21]徐晔, 王志超, 陶长琪. 数据要素市场化建设对企业资源配置效率的影响研究[J]. 科研管理, 2025, 46(9): 25-34.
- [22]杨俊, 李小明, 黄守军. 大数据、技术进步与经济增长——大数据作为生产要素的一个内生增长理论[J]. 经济研究, 2022, 57(4): 103-119.
- [23]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, 40(2): 101-116, 133.
- [24]余典范, 张宇, 宋晴. 数字人力投入与企业价格加成——来自招聘大数据的证据[J]. 中国工业经济, 2025, (7): 62-80.
- [25]余明桂, 李文贵, 潘红波. 民营化、产权保护与企业风险承担[J]. 经济研究, 2013, 48(9): 112-124.
- [26]张杨, 张钟文, 喻彦杰, 等. 智能制造对企业数据技术创新的影响研究[J]. 管理学报, 2026, 23(1): 97-106.
- [27]赵涛, 张智, 梁上坤. 数字经济、创业活跃度与高质量发展——来自中国城市的经验证据[J]. 管理世界, 2020, 36(10): 65-75.
- [28]赵宣凯, 宝恩德尔, 左从江, 等. 大语言模型在经济金融领域的应用——研究评述、学术应用及未来展望[J]. 金融评论, 2025, 17(1): 69-93, 156-157.
- [29]周文, 韩文龙. 数字财富的创造、分配与共同富裕[J]. 中国社会科学, 2023, (10): 4-23.
- [30]Acemoglu D, Akcigit U, Kerr W R. Innovation network[J]. Proceedings of the National Academy of Sciences of the United States of America, 2016, 113(41): 11483-11488.
- [31]Acs Z J, Braunerhjelm P, Audretsch D B, et al. The knowledge spillover theory of entrepreneurship[J]. Small Business Economics, 2009, 32(1): 15-30.
- [32]Bernal P, Carree M, Lokshin B. Knowledge spillovers, R&D partnerships and innovation performance[J]. Technovation, 2022, 115: 102456.
- [33]Bloom N, Schankerman M, Van Reenen J. Identifying technology spillovers and product market rivalry[J]. Econometrica, 2013, 81(4): 1347-1393.
- [34]Brynjolfsson E, Rock D, Syverson C. The productivity J-curve: How intangibles complement general purpose technologies[J]. American Economic Journal: Macroeconomics, 2021, 13(1): 333-372.
- [35]Chen Z N, Chan J. Large language model in creative work: The role of collaboration modality and user expertise[J]. Management Science, 2024, 70(12): 9101-9117.
- [36]Cong L W, Xie D X, Zhang L T. Knowledge accumulation, privacy, and growth in a data economy[J]. Management Science, 2021, 67(10): 6480-6492.
- [37]Corvello V, Belas J, Giglio C, et al. The impact of business owners' individual characteristics on patenting in the context of digital innovation[J]. Journal of Business Research, 2023, 155: 113397.
- [38]Doshi A R, Bell J J, Mirzayev E, et al. Generative artificial intelligence and evaluating strategic decisions[J]. Strategic Management Journal, 2025, 46(3): 583-610.
- [39]Goldfarb A, Tucker C. Digital economics[J]. Journal of Economic Literature, 2019, 57(1): 3-43.
- [40]Jaffe A B, Trajtenberg M, Henderson R. Geographic localization of knowledge spillovers as evidenced by patent citations[J]. The Quarterly Journal of Economics, 1993, 108(3): 577-598.
- [41]Jones C I, Tonetti C. Nonrivalry and the economics of data[J]. American Economic Review, 2020, 110(9): 2819-2858.
- [42]Kaplan S N, Strömberg P. Financial contracting theory meets the real world: An empirical analysis of venture capital contracts[J]. The Review of Economic Studies, 2003, 70(2): 281-315.
- [43]Rambachan A, Roth J. A more credible approach to parallel trends[J]. The Review of Economic Studies, 2023, 90(5): 2555-2591.
- [44]Wang X C, Li Y, Tian L W, et al. Government digital initiatives and firm digital innovation: Evidence from China[J]. Technovation, 2023, 119: 102545.

Data Intellectual Property Protection and Enterprise Data Technology Innovation: Empirical Evidence from a Generative Large Language Model

Yang Jing¹, Chen Zhibin¹, Feng Sixian^{2,3}

(1. School of Economics and Management, Southeast University, Nanjing 211189, China;

2. Business School, Nanjing Normal University, Nanjing 210023, China;

3. Ginling College, Nanjing Normal University, Nanjing 210023, China)

Abstract: Enterprise data technology innovation is a crucial process for realizing the value conversion of data elements. However, there is still a lack of systematic empirical research on how to effectively incentivize it through data intellectual property protection. Based on data from China's A-share listed companies from 2015 to 2024, this paper innovatively employs a generative large language model to identify enterprise data technology patents. Using the first batch of national pilot projects for data intellectual property as a quasi-natural experiment, it constructs a DID model to systematically assess the impact of data intellectual property protection on enterprise data technology innovation. The results show that the data intellectual property pilot significantly promotes enterprise data technology innovation. Mechanism testing indicates that this effect is primarily realized through incentivizing specialized factor inputs and facilitating knowledge spillovers. Heterogeneity analysis reveals that the promoting effect is more pronounced in state-owned enterprises, data-native enterprises, and regions with better-developed digital infrastructure, while no significant difference is observed among enterprises with different technological innovation attributes. Further research demonstrates that by fostering enterprise data technology innovation, data intellectual property protection effectively enhances enterprise TFP. This paper provides new evidence for understanding the micro-level innovation effect of data intellectual property protection and offers important policy implications for building a unified national data market, cultivating new quality productive forces, and promoting high-quality development of the digital economy.

Key words: data intellectual property protection; data technology innovation; generative large language model; knowledge spillovers

(责任编辑: 王雅丽)