

语言智能技术对数字经济影响

毕森均 王 伟

重庆工商大学, 重庆 400067

摘要: 基于2010-2024年31个省份面板数据的实证检验发现, 语言智能技术可显著驱动数字经济发展。数据传输规模扩张、市场化程度提升、产业结构高级化是核心传导机制, 该赋能效应在东西部地区与高研发强度地区更为突出。

关键词: 语言智能技术; 数字经济

DOI: 10.64649/yh.shygl.2026020002

0 引言

语言是市场经济运行中核心的信息传递与交易沟通工具, 在降低交易成本、畅通市场循环方面具有不可替代的基础作用。随着数字技术与人工智能深度融合, 传统语言交互模式持续革新, 衍生出语音识别、自然语言处理、智能翻译、知识图谱构建等各类语言智能技术。该技术早已突破单一工具属性, 成为驱动经济社会数字化转型的新型基础设施。然而, 语言智能技术的宏观经济效应尚缺乏实证支撑。当下数字经济已成为引领科技革命、重塑全球经济格局的核心动力, 挖掘其核心驱动因素成为研究焦点。那么, 语言智能技术能否推动数字经济发展? 其内在作用机制是什么?

本文选取2010-2024年中国31个省份样本, 通过构建语言智能技术专利与数字经济指数的数据集, 实证检验二者之间关系。本文的边际贡献在于: 第一, 率先量化了语言智能技术对数字经济的贡献度, 丰富了语言智能技术的宏观效应研究; 第二, 揭示了语言智能技术作用于数字经济的三个关键路径, 拓展了数字经济增长驱动因素研究; 第三, 证实了语言智能技术赋能效应的差异, 为各地区因地制宜制定相关政策提供参考。

1 研究假说

1.1 语言智能技术赋能数字经济的理论逻辑

语言智能技术对数字经济的正向作用, 根植于其“使能技术”本质属性。从技术经济学理论来看, 语言智能技术赋予了机器对海量语言类数据的智能化处理、解析与生成能力, 从根源上破解了数字经济运行过程中的信息不对称难题。同时, 该项技术重塑了传统人机交互模式, 简化了数字服务的操作流程, 显著降低全社会数字服务的获取门槛与使用成本。此外, 语言智能作为典型的通用目的技术, 具备极强的技术溢出效应与产业渗透性, 能够推动数字产业化与产业数字化, 持续为数字经济注入内生动力。由此, 本文提出研究假说:

H1: 语言智能技术对数字经济具有显著促

进作用。

1.2 语言智能技术作用于数字经济的机制分析

第一, 语言智能技术通过破除语言壁垒、激发数据需求, 有效扩大数据传输规模。智能机器翻译、跨语言交互等技术打破了跨境、跨区域信息流通的语言障碍, 拓宽数据传输的覆盖范围, 提升数字网络连接密度与信息流通效率。而数据传输规模的扩大, 既是数字经济活跃度提升的直接体现, 也能借助网络效应为数字经济纵深发展提供支撑。

第二, 语言智能技术通过削减市场信息摩擦、降低交易成本, 推动市场化程度提升。在市场环境优化层面, 语言智能技术可应用于政务智能咨询、商事服务智能化办理等场景, 有效降低企业制度性交易成本与市场准入隐性壁垒。这将推动各类生产要素向高收益、高效率领域顺畅流动, 充分释放制度红利, 强化市场机制在数字资源配置中的决定性作用。

第三, 语言智能技术通过低端替代与高端赋能双重路径, 加速产业结构高级化演进。对于传统劳动密集型产业, 语言智能技术可实现部分替代, 释放劳动力资源向高附加值、高技术含量产业转移。对于知识密集型产业, 它能大幅提升知识生产与专业服务效率与质量, 推动产业附加值稳步提升。这种双向变革, 为数字经济夯实产业基础。

基于上述分析, 本文提出研究假说:

H2: 数据传输规模、市场化程度、产业结构高级化是语言智能技术驱动数字经济的作用机制。

2 研究设计

2.1 样本及数据

本文以全国31个省份为研究对象, 时间跨度为2010—2024年。为缓解极端值影响, 对所有连续变量在1%和99%分位上作缩尾处理。数据来源于中国研究数据服务平台(CNRDS)和国家知识产权局。

2.2 模型设定

为检验 H1, 本文构建以下双向固定效应模型:

$$DE_{it} = \alpha_0 + \alpha_1 LIT_{it} + \sum CV_{it} + \sum Year + \sum Province + \varepsilon_{it} \quad (1)$$

其中, DE 为数字经济, LIT 为语言智能技术, CV 是一系列控制变量, 本文还控制了时间固定效应 Year 和省份固定效应 Province, ε 是随机扰动项, i 、 t 分别为省份和年份。

为检验 H2, 本文借鉴江艇 (2022) 的做法构建如下中介效应模型:

$$M_{it} = \beta_0 + \beta_1 LIT_{it} + \sum CV_{it} + \sum Year + \sum Province + \varepsilon_{it} \quad (2)$$

式中, M 为机制变量。

2.3 变量定义

(一) 被解释变量: 数字经济 (DE)。本文借鉴赵涛等 (2020) 的方法, 从互联网发展和数字金融两方面衡量数字经济, 前者选取互联网普及率、相关从业人员情况、相关产出情况和移动电话普及率 4 项指标, 后者使用“北京大学数字普惠金融指数”, 最终通过主成分分析法合成数字经济指数。

(二) 解释变量: 语言智能技术 (LIT)。结合数据可得性和权威性, 本文通过国家知识产权局关键词检索筛选语言智能相关专利, 以

省份年度申请量加 1 取自然对数来度量。

(三) 机制变量。数据传输规模 (DA) 用移动互联网接入量衡量; 市场化程度 (MA) 用樊纲市场化指数表示; 产业结构高级化 (IS) 用第三产业与第二产业增加值之比衡量。

(四) 控制变量。借鉴既有研究, 选取以下省份层面特征变量: 人口规模 (Peo, Ln 常住人口数)、人力资本水平 (Hc, 高校在校生人数 / 总人口)、城镇化水平 (Ub, 城镇人口 / 总人口)、大中型工业企业数 (Lme, 数量取自然对数)、对外开放水平 (Open, 货物进出口总额 \times 汇率 / GDP)、经济发展水平 (GDPp, 人均 GDP 取自然对数)。

3 实证分析

3.1 基准回归

表 1 为逐步回归结果, 各列中 LIT 的系数均显著为正, 初步表明语言智能技术可有效促进数字经济发展。列 (4) 纳入全变量并控制省份和年份固定效应, 回归系数为 0.0064, 其经济含义为: 其他条件不变时, 语言智能技术每提升 1 个单位, 数字经济将对应提升 0.64%, 假设 H1 得到验证。这说明语言智能技术是数字经济的新型驱动要素。

表1 基准回归结果

	(1)	(2)	(3)	(4)
	DE	DE	DE	DE
LIT	0.0851 ^{***}	0.0469 ^{***}	0.0138 ^{***}	0.0064 ^{**}
	(21.16)	(7.99)	(4.19)	(2.11)
Peo		-0.0251		0.1546 ^{***}
		(-1.46)		(2.62)
Hc		-8.0619 ^{***}		0.4441
		(-9.36)		(0.36)
Ub		-0.1831 [*]		0.2398 [*]
		(-1.71)		(1.78)
Lme		0.0943 ^{***}		0.0092
		(7.71)		(0.68)
Open		0.1089 ^{***}		-0.0375 [*]
		(5.92)		(-1.80)
GDPp		0.0431 [*]		0.1232 ^{***}
		(1.82)		(3.66)
_cons	-0.0384 ^{***}	-0.6987 ^{***}	-0.1599 ^{***}	-2.6793 ^{***}
	(-2.97)	(-2.85)	(-4.38)	(-6.17)
省份/地区固定	不控制	不控制	控制	控制
N	424	424	424	424
adj. R ²	0.514	0.840	0.973	0.976

注: 括号中为 t 值, *, **, *** 分别表示在 10%、5%、1% 的水平上显著。后同。

3.2 稳健性检验

本文从三方面开展稳健性检验: 第一, 倾向得分匹配法 (PSM)。采用核匹配法重新估计, 剔除样本选择性偏差。第二, 调整样本区间。

考虑长周期数据波动干扰, 将样本缩至 2015-2024 年重新回归。第三, 工具变量法。针对双向因果内生性问题, 选取 LIT 滞后一期作为工具变量。上述回归结果均显示, 本文核心结论依然成立。

3.3 机制检验

根据模型(2)进行机制检验,结果见表2。LIT的回归系数分别为0.0853、0.3602、0.0032,并通过显著性检验,说明语言智能技术能够推动区域数据要素扩容、市场化水平提升与产业结构优化升级。结合前文理论分析,语言智能技术能够通过上述三大渠道实现有效传导,推动数字经济发展,假说H2得到验证。

表2 机制检验结果

	(1)	(2)	(3)
	DA	MA	IS
LIT	0.0853**	0.3602***	0.0032*
	(2.28)	(4.79)	(1.79)
控制变量	控制	控制	控制

省份/地区固定	控制	控制	控制
N	311	424	424
adj. R ²	0.843	0.876	0.970

3.4 异质性分析

考虑到区域资源禀赋差异与研发强度差距,会影响语言智能技术的赋能效果,本文进一步将样本划分为东、中、西部三组,并根据研发强度(R&D经费占GDP比重)中位数分为高低两组,回归结果见表3。易知,东部、西部语言智能技术的赋能效应均显著为正,中部省份效应不明显;研发强度较高地区技术驱动效应显著,其他地区则不显著。这表明,语言智能技术对数字经济的赋能效果依赖区域禀赋与研发支撑。

表3 异质性检验结果

	(1) 东部	(2) 中部	(3) 西部	(4) 研发强度较低	(5) 研发强度较高
	DE	DE	DE	DE	DE
LIT	0.0143*	0.0051	0.0057*	0.0018	0.0076*
	(1.86)	(1.45)	(1.67)	(0.54)	(1.77)
控制变量	控制	控制	控制	控制	控制
省份/地区固定	控制	控制	控制	控制	控制
N	190	87	147	212	212
adj. R ²	0.975	0.972	0.959	0.977	0.987

4 结论与启示

本文基于2010-2024年省级面板数据,实证检验显示,语言智能技术可显著驱动我国数字经济发展,核心传导机制为数据传输规模、市场化程度和产业结构高级化,且该赋能效应在东西部地区及高研发强度地区更为突出。

本文的政策启示在于:第一,强化语言智能技术战略定位。将其纳入新型基建规划,加

快多模态语言大模型研发,释放其新质生产力引擎作用,夯实数字经济发展技术根基。第二,疏通技术赋能传导路径。打破数据流动壁垒、提升市场化交易效率、加速产业智能化升级,构建技术突破与机制协同的良性循环。第三,实施区域差异化策略。东部聚焦技术前沿引领,中西部补齐基建短板,同时优化研发资源配置,防范数字鸿沟扩大。

参考文献:

- [1] 张卫国. 论语言的经济效应: 理论、经验与启示 [J]. 上海交通大学学报(哲学社会科学版), 2022, 30(06): 120-129.
- [2] 李宇明, 梁京涛. 语言数据的生产要素功能与产权制度构建 [J]. 语言教学与研究, 2024, (02): 1-11.
- [3] 屈哨兵, 王海兰. 数字经济发展中的四大基本语言服务能力建设 [J]. 广州大学学报(社会科学版), 2023, 22(05): 112-121.
- [4] 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, (05): 100-120.
- [5] 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): 65-76.

作者简介: 毕森均(2000—), 男, 汉, 山西高平, 硕士研究生, 研究方向: 经济学。

王伟(1984—), 男, 汉, 湖北黄冈, 教授、博士生导师, 研究方向: 经济学。

项目信息: 重庆市语言文字科研项目“语言智能技术对数字经济贡献度研究”(yyk23237)。

重庆工商大学研究生科研创新项目“AI创新应用政策赋能地方企业碳绩效提升的机制及路径研究”(yjscxx2025-269-269)。