

CSMAR



人工智能时代下的 实证研究新视点

——CSMAR数据库的科研应用

主讲人：张中岩

深圳希施玛数据科技有限公司

<https://data.csmar.com/>

内容大纲

CSMAR数据库简介 01

- 数据库概况

- 基础操作指引

→ *检索案例：常用AI发展水平测度指标

CSMAR最新数据资源 03

- 最新数据库概览

- 新库资讯获取

02 CSMAR数据库应用

- 企业经济领域

- 宏观经济领域

- 金融领域

小模型范式

与

大模型应用

04 CSMAR科研资讯

- 讲座资源

- 科研频道

01 CSMAR数据库简介

- 数据库概况
- 基础操作指引

→ *检索案例：常用AI发展水平测度指标

CSMAR数据库简介：数据库概况

全 称：China Stock Market & Accounting Research Database

中国经济金融研究数据库

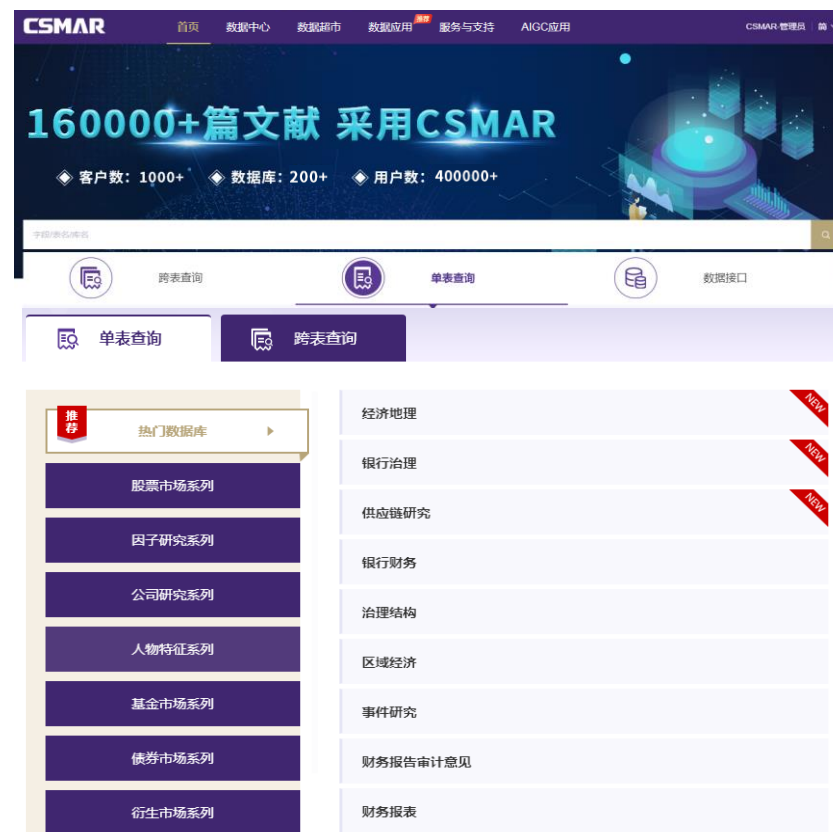
定 位：研究型精准数据库

标 准：CSMAR数据库参照CRSP、COMPUSTAT等权威数据库的标准。

服务对象：以研究和量化投资分析为目的的学术高校和金融机构。

内 容：将数据库分为股票、公司、基金、债券、衍生、经济、行业、海外、资讯系列。涵盖中国证券、期货、外汇、宏观、行业等经济金融主要领域的高精准研究型数据库，是投资和实证研究的基础工具。

官 网：<http://data.csmar.com/>



CSMAR数据库简介：数据库概况

18+1
系列

200+
数据库

70000+
字段

④ 公司研究系列



④ 股票市场系列



④ 债券市场系列



④ 人物特征系列



④ 银行研究系列



④ 基金市场系列



④ 经济研究系列



④ 专题研究系列



④ 商品市场研究系列



④ 海外研究系列



④ 行业研究系列



④ 智能研究系列



④ 绿色经济系列



④ 市场资讯系列



④ 货币市场系列



④ 因子研究系列



④ 历史数据



④ 衍生市场系列



④ 合作数据



CSMAR数据库简介：数据库概况

数据系列	内容简介	热门数据库	特色数据库
公司研究系列	全面的上市公司研究数据，基础数据丰富，最新数据及时	财务报表、治理结构、财务指标分析、股东、上市公司研发创新、财务报表附注、经营困境、分析师预测、上市公司基本信息、股权性质、首次公开发行（A股）、会计信息质量、环境研究	业绩说明会文本、董事长致辞、市场分析报告文本、市场竞争、破产重整、新三板供应链、一致预测、会计差错更正、供应链研究、融资租赁、董监高责任险、政府审计、经营困境、对赌协议、环境研究、家族企业、会计信息质量、产业资本
经济研究系列	汇集了具有重要现实意义与学术价值的研究主题，提供相关数据，包括：新质生产力、数字经济、碳中和、城乡建设、人口老龄化、普惠金融、经济内循环、经济地理等，以及常用的宏观、区域经济指标	宏观经济、区域经济	
专题研究系列	聚焦重要研究主题，提供特色数据指标，包括：人工智能产业链、专精特新企业、金融科技、一带一路、事件研究等		事件研究、投资者情绪、金融科技、资本市场监管、垄断与反垄断
海外市场研究系列	港股、美股上市公司的基础与特色数据		美国年报语调、美国年报可读性、美国年报风险度量、美国报告基本信息、美股财务报表、美股财务指标、美股股本结构、美股红利分配、美股机构持股、美股基金持股、美股董事高管信息
股票市场系列	股票市场交易的基础数据与衍生指标	股票市场交易、股票市场衍生指标、市场指数	龙头股
因子研究系列	提供金融研究的常用衍生指标数据		已实现指标、行为金融、动量因子、DGTW股票特征基准、股票流动性
智能研究系列	运用深度学习、自然语言处理等AI技术从股价预测、舆情分析、财务诊断和风险感知四个方面构建数据库		智能研报基本信息、智能股价预测、智能舆情分析、智能财务诊断、智能风险感知
市场资讯系列	提供市场资讯，包括社交媒体、市场分析报告、证券市场公告以及证券市场监管公告		
绿色经济系列	资源环境与经济主题相关数据，包括：绿色金融、全球暖化、资源		绿色金融
人物特征系列	上市公司高管与基金经理人物特征数据	上市公司人物特征	
行业研究系列	提供十余个行业的数据指标		
合作数据	与权威科研团队共同研发的多个主题数据库		企业数字化转型、企业碳信息披露指数、区域热辐射

CSMAR数据库简介：数据库概况

CSMAR数据库的研发，始终确保：

数据来源权威、采集流程专业、质检流程规范

为广大用户提供高质量数据，助力学术研究。

Available On WRDS		
中国上市公司年、中、季报公布日期数据库	中国上市公司治理结构研究数据库	中国上市公司股东研究数据库
中国股票市场交易数据库	中国海外上市公司研究数据库	中国上市公司增发配股研究数据库
中国基金研究数据库	中国上市公司并购重组研究数据库	中国上市公司红利分配研究数据库
中国上市公司机构投资者研究数据库	中国上市公司财务报表数据库	中国上市公司分析师预测研究数据库
中国上市公司首次公开发行研究数据库(A股)		

权威

01

- **数据源**包括上交所、深交所、上期所、港交易所、中金所等，国家外汇管理局，中证信息,中国年鉴信息网等

基础

02

- CSMAR数据被**摩根斯坦利**选用，作为编制MSCI-A股指数的**基础**。

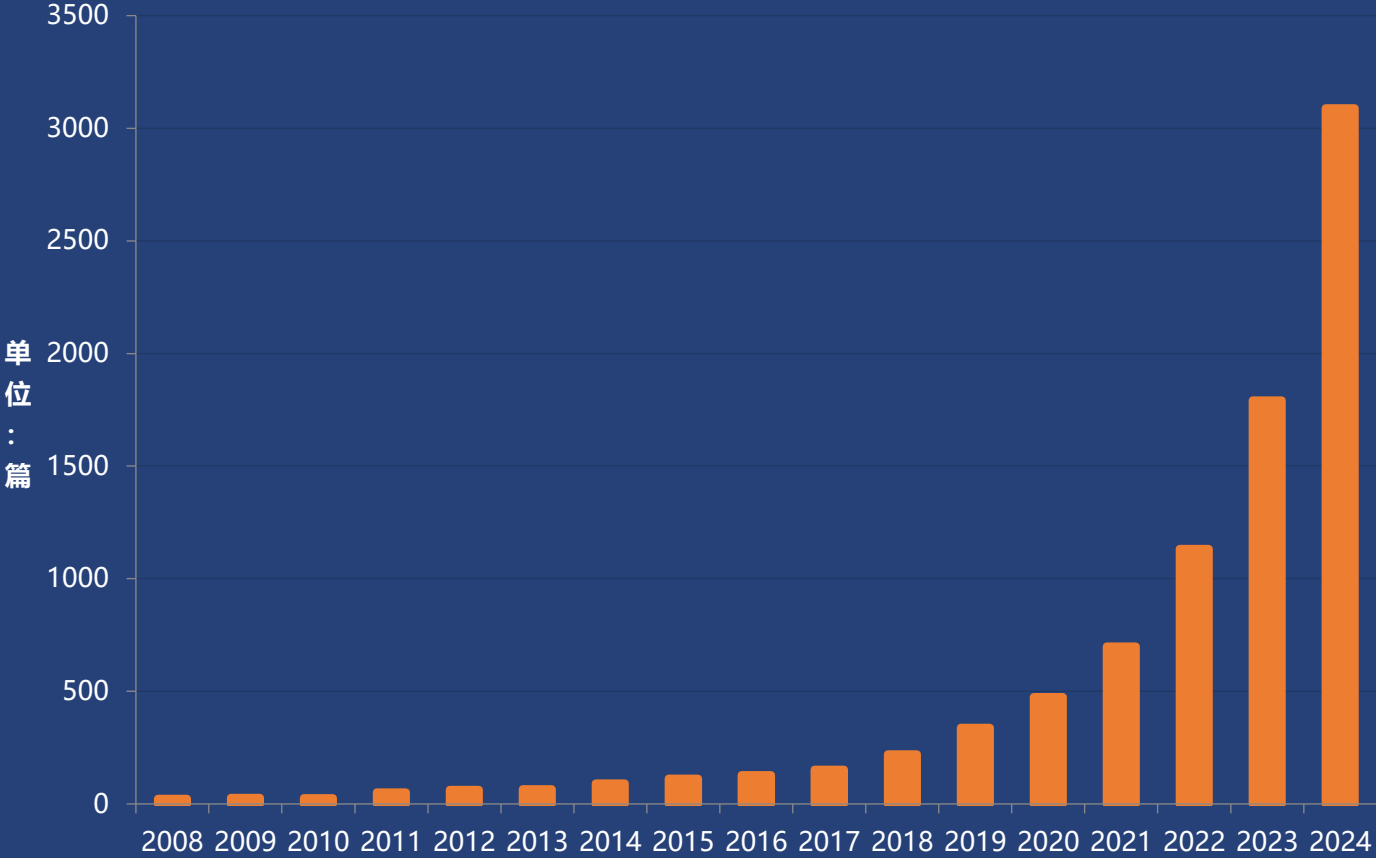
引用

03

- 截至2025年6月30日，国内外使用**CSMAR**数据发表的论文超过18万篇。

CSMAR数据库简介：数据库概况

国际期刊采用CSMAR情况



数据来源于Wiley Online Library 和 ScienceDirect，截止时间2024.12.31



截止时间2024.12.31，国际顶刊累计发表量

CSMAR数据库简介：数据库概况

数据查询

关键字搜索、单表查询、跨表查询、合作数据、数据可视化.....

数据服务

数据接口、数据定制、视频集锦、公开数据.....

数据应用

国内TOP期刊论文、图谱中心、上市公司数据分析、公司财务分析、数据图表分析、数说TOP10、公司多维分析、沪深京股票专题分析、商品期货专题分析、智能财经报告、合作应用.....



CSMAR Solution
主要功能

<http://data.csmar.com/>

CSMAR数据库简介：基础操作指引

➤ 检索案例：常用AI发展水平测度指标

AI关键词词频

上市企业年报中AI技术关键词数量

AI投资水平

人工智能资本投入

AI专利

根据《2020人工智能中国专利技术分析报告》中对人工智能技术的定义，通过专利关键词搜索来识别上市公司的人工智能专利。

地区AI发展水平

某城市AI企业数量

机器人应用水平 / 渗透度

$$robot_{jt} = (\sum_{m=1}^M E_{jmt} \times robot_stock_{mt}) / Emp_{jt}$$

E_{jmt} 表示 t 年城市 j 在 m 行业的从业人数在当年全国该行业总从业中所占的份额， $robot_stock_{mt}$ 为 m 行业 t 年的机器人安装量

综合评价指标体系

从数智产业发展、数字技术创新、数智企业发展等不同维度出发，搭建囊括多个指标的人工智能评价指标体系，使用熵权法/熵值法测算

关联CSMAR子库

人工智能

从人工智能企业（包含企业基本信息、所有者权益情况、管理人员信息、附属公司情况及财务指标）、上市公司AI水平、企业智能算法备案信息、工业机器人安装量多方面统计了人工智能相关数据

人工智能产业链

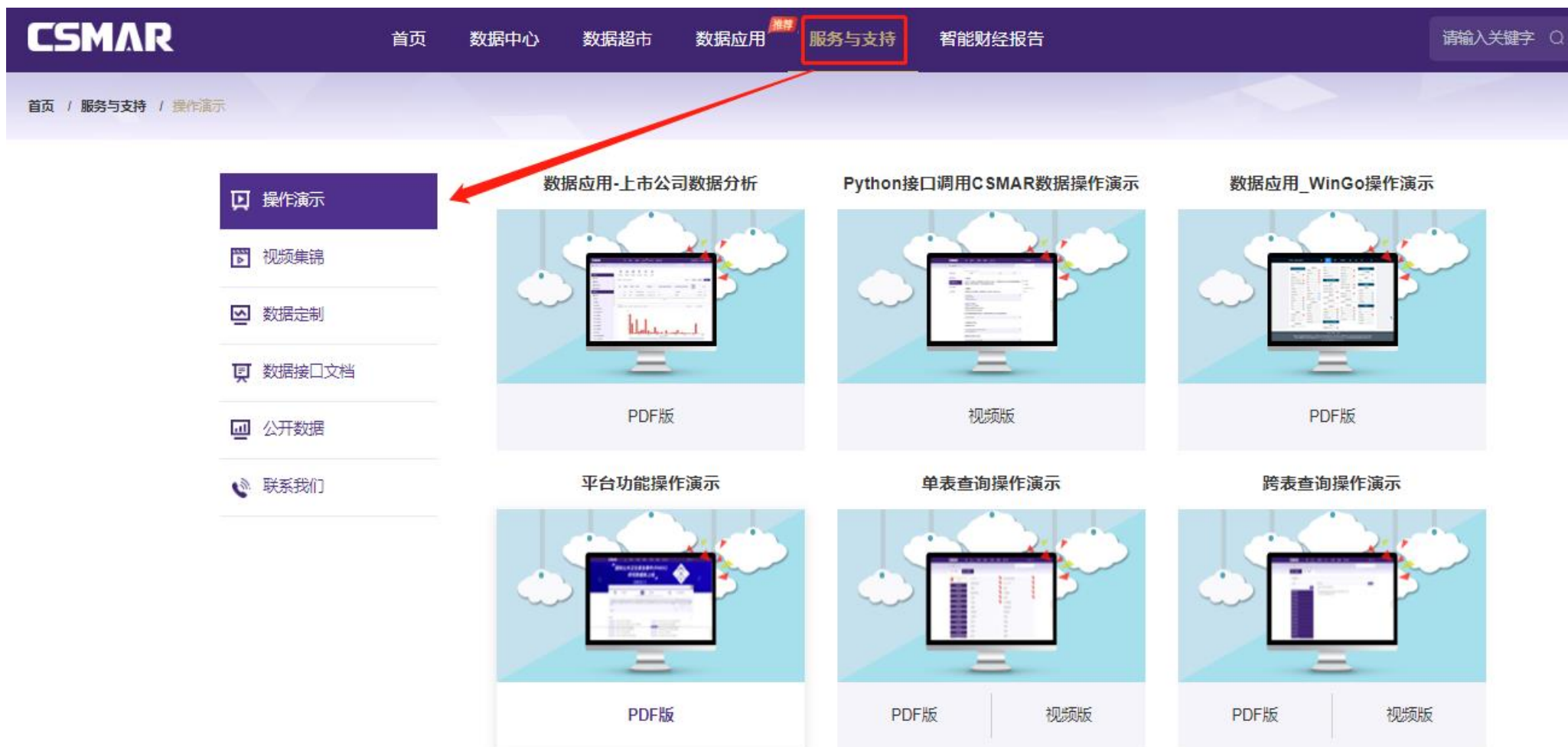
从产业结构、公司指标、政策资讯等方面分析人工智能产业链

人工智能专利

提供上市公司及其子公司人工智能专利授权以及被引用等信息

CSMAR数据库简介：基础操作指引

访问CSMAR数据库主页 (<http://data.csmar.com/>)，在“服务与支持-操作演示”模块可获取PDF及视频版的操作指引。



02 CSMAR数据库应用

- 企业经济领域
- 宏观经济领域
- 金融领域

小模型范式
与
大模型应用



研究领域	热点问题	前沿主题	研究范式
企业经济	公司治理、融资约束、内部控制、盈余管理、投资效率、会计信息质量、企业创新、企业社会责任、审计质量、股权质押、家族企业、混合所有制改革、企业金融化、分析师.....	中国科学院科技战略咨询研究院 – 2024 近期的前沿主题体现出了以人工智能为代表的数智技术与经济学等社会学的交叉融合，关注气候环境和人类健康等研究趋势。 - 政策变化评估及特定事件影响研究 - COVID-19 对不同人群的心理健康影响 - 气候变化与投资决策 - 人类活动对气候变化、生物多样性等影响 - 社会结构因素与健康差异关系研究 - 脑电图（EEG）数据分析在神经科学、心理学、认知科学中的应用 - 深度学习算法与金融预测、资产定价 - 超加工食品消费问题 - 零工经济的算法、人力资源及平台管理 - 人工智能伦理准则	小模型 关键变量的获取、处理与应用
宏观经济	经济增长、全球价值链、环境规制、“一带一路”、创新、产业政策.....	- 新质生产力研究 - 生态文明建设与高质量发展 - 数据要素的创新驱动作用研究 - 人工智能的产业赋能 - 金融服务实体经济高质量发展 - 全球视野下的产业布局与发展 - 中国式现代化进程中的社会保障制度顶层设计 - 建构中国自主的经济学知识体系 - 新一轮财税体制改革 - 新就业形态劳动者权益保障研究 - 防范化解地方债务风险	大模型 BERT GPT ERNIE
金融	股票流动性、股价崩盘风险、金融稳定、数字金融、绿色金融、货币政策.....	《经济理论与经济管理》 经济学研究热点课题组 - 2024	

CSMAR数据库应用：企业经济领域

小模型范式

- [1] 人工智能如何赋能企业新质生产力[J].科技进步与对策, 2025.
- [2] 人工智能应用的绿色创新效应研究[J].经济学动态, 2025.
- [3] 一枝独秀还是百花齐放：人工智能技术创新与企业供应链集中度[J].经济评论, 2025.
- [4] 工业机器人应用与企业“高持现”治理[J].经济管理, 2025.
- [5] 绿色产业政策下投资者绿色关注对企业绿色创新影响研究[J].学术研究, 2025.
- [6] 人工智能、企业数字责任对企业社会绩效的影响——基于中国上市企业ESG表现的证据[J].西安财经大学学报, 2024.
- [7] Impact of regional artificial intelligence development on corporate environmental information[J]. Finance Research Letters, 2025.
- [8] 如何实现数字竞争优势：“专精特新”企业数字投资与专业投资的协同逻辑 [J].中国工业经济, 2025.
- [9] 数字化转型的持续性创新效应研究[J].数量经济技术经济研究, 2025.
- [10] 企业管理层气候关注与ESG表现——来自中国A股上市公司的经验证据[J].南方金融, 2025.
- [11] 政府环境关注度与企业ESG表现——高管团队环保意识与企业创新的机制作用[J].科技进步与对策, 2025.

大模型应用

- [1] 基于人工智能技术的管理层语调识别及其有效性——来自对企业长期投资预效果的证据[J].财经理论与实践, 2025.
- [2] 企业数字化转型的测度难题：基于大语言模型的新方法与新发现[J].经济研究, 2024.
- [3] 中国企业数字技术风险暴露对企业价值的影响——来自大语言模型的文本分析证据[J].经济研究, 2025.



企业经济

CSMAR数据库应用：企业经济领域

相关研究1：人工智能如何赋能企业新质生产力[J].科技进步与对策, 2025.

关键解释变量	被解释变量
年报中人工智能关键词数量加1并取对数	从劳动力和生产工具两个方面构建企业新质生产力评价指标体系

人工智能 新质生产力

$$NP_{it} = \alpha_0 + \alpha_1 AI_{it} + \alpha_2 Controls_{it} + year + ind + pro + \epsilon_{it}$$

变量	NP			
	(1)	(2)	(3)	(4)
AI	1.964 ***(0.373)	1.211 ***(0.296)	1.757 ***(0.373)	1.180 ***(0.280)
变量	技术赋能 (1)Dig	效率赋能 (2)Effi	信息赋能 (3)Asy	
AI	0.465 ***(0.077)	0.121 ***(0.036)	-0.053 ***(0.013)	

其他CSMAR关联子库：治理结构、财务报表、财务指标分析、上市公司基本信息、.....



提升智能化水平是企业加快形成新质生产力的关键举措。文章选取2013—2022年A股上市公司为研究样本，探讨了人工智能对企业新质生产力的赋能机制。结果发现：人工智能通过增强数字创新能力、提高供应链效率和降低信息不对称程度赋能企业新质生产力。

人工
智能



企业
新质生产力

- ◆ **技术创新赋能**：为企业数字技术创新提供新的工具和手段,有助于企业基于海量数据和应用场景提升数字技术创新能力
- ◆ **效率赋能**：提升供应链效率、资源配置效率和生产效率
- ◆ **信息赋能**：有助于企业提升内外部信息收集、整合和处理能力,降低因信息不对称带来的风险和成本

CSMAR数据库应用：企业经济领域

相关研究2：人工智能应用的绿色创新效应研究[J].经济学动态, 2025.

关键解释变量

是否属于引入人工智能建设项目的企业

人工智能 / 财务报表附注

被解释变量

绿色发明专利申请数和绿色实用新型专利申请数

绿色专利

具体构建方法：首先，参考相关文件确定人工智能关键词。其次，基于人工智能关键词对企业“在建工程明细”进行筛选，确定引入人工智能建设项目的企业及引入时间。

其他CSMAR关联子库：治理结构、财务报表、财务指标分析、上市公司基本信息、……



文章基于2007—2022年中国A股制造业上市企业数据考察了人工智能应用对绿色创新的影响与机制。研究发现，人工智能应用显著推动了企业绿色创新“提质增量”。机制分析显示，扩宽企业知识宽度与提升适应能力是企业人工智能应用推动绿色创新的重要机制。



$$GRP_{it} = \alpha_0 + \beta_1 Treat_i \times Post_t + \beta_2 Controls_{it} + \mu_i + \delta_t + \varepsilon_{it}$$

变量	(1)	(2)	(3)	(4)
	lnGRS	lnGRS	lnGRQ	lnGRQ
$Treat \times Post$	0.109*** (0.027)	0.097*** (0.027)	0.085*** (0.022)	0.076*** (0.022)

CSMAR数据库应用：企业经济领域

相关研究3：一枝独秀还是百花齐放：人工智能技术创新与企业供应链集中度[J].经济评论, 2025.

关键解释变量

企业当年申请的人工智能
专利数量

人工智能专利

被解释变量

供应商集中度和客户集中
度

供应链研究

人工智能的测度方式主要基于技术应用和技术创新两种视角。已有研究中，技术应用多采取制造业企业机器人渗透度衡量，技术创新则多采用人工智能专利数据。但工业机器人的应用只能作为人工智能技术应用中的一部分，主要应用在制造业部门，可能无法完全涵盖人工智能技术应用领域。采用专利作为创新视角的测度能更好地捕捉企业人工智能技术的真实应用水平。

其他CSMAR关联子库：财务报表、财务指标分析、机构投资者、.....



文章基于2010-2020年中国A股上市公司数据，系统考察了企业人工智能技术创新对供应链集中度的可能影响与作用机理。研究发现，人工智能技术创新显著降低了企业供应链集中度，并且该效应主要通过改善供需效能、增强营运能力和优化企业声誉三种作用机理得以实现。

基准回归结果

变量	SC		
	(1)	(2)	(3)
AI	-0.007 *** (-4.020)	-0.006 *** (-3.422)	-0.006 *** (-3.310)

机制检验结果

变量	Coor	Eff	Cost	ROA	Cover	Rep	Pos
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
AI	-0.018 ** (-2.312)	-0.070 *** (-4.586)	-0.007 *** (-2.924)	0.002 *** (3.763)	0.044 *** (4.688)	0.072 ** (2.291)	0.008 *** (3.962)

供需效能改善机制

Coor: 供需偏离程度
Eff: 供应链效率损失情况, 以库存
周转天数度量

营运能力增强机制

Cost: 成本费用
ROA: 资产收益率
Cover: 销售收入
降价空间
市场覆盖

企业声誉优化机制

Rep: 企业声誉指标
Pos: 正面新闻占比

CSMAR数据库应用：企业经济领域

相关研究4: 工业机器人应用与企业“高持现”治理[J].经济管理, 2025.

关键解释变量

企业层面工业机器人渗透率

被解释变量

企业现金持有：现金及现金等价物占总资产的比例；其二为货币资金与交易性金融资产之和与总资产扣除货币资金及交易性金融资产后余额的比值

人工智能

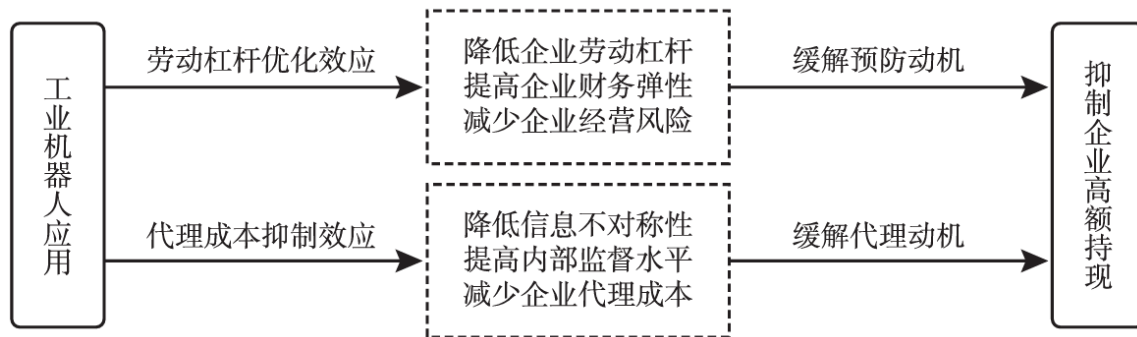
财务报表

$$Cash_{i,j,t} = \alpha_0 + \alpha_1 Robot_PR_{i,j,t} + \alpha CV_{i,j,t-1} + \theta_j + \delta_t + \varepsilon_{ijt}$$

变量	Cash1	Cash2	Cash1 _{t+1}	Cash2 _{t+1}	Cash1 _{t+2}	Cash2 _{t+2}
	(1)	(2)	(3)	(4)	(5)	(6)
Robot_PR	-0.2816*** (-3.18)	-0.5048** (-2.41)	-0.2900*** (-3.33)	-0.5527*** (-2.72)	-0.2515*** (-2.93)	-0.5635*** (-2.82)
变量	Nor_Cash1	Nor_Cash2	Ex_Cash1	Ex_Cash2	Cash1	Cash2
	(1)	(2)	(3)	(4)	(5)	(6)
Robot_PR	-0.3600** (-2.28)	-0.3948** (-2.24)	-2.1275*** (-3.76)	-1.9098*** (-3.15)	0.0808 (1.30)	0.2453 (1.51)
Group					0.1424*** (79.89)	0.2890*** (63.39)
Robot_PR × Group					-0.3899*** (-5.18)	-1.0190*** (-5.36)



文章以2011—2019年沪深A股制造业上市企业作为样本，考察工业机器人应用对企业现金持有的影响。研究发现，工业机器人应用能够显著降低企业现金持有水平，有效抑制企业过度持现的倾向。机制分析表明，工业机器人应用通过降低劳动杠杆，减少企业出于预防性动机而持有的超额现金；同时，通过降低代理成本，抑制企业因代理问题而产生的现金囤积。



其他CSMAR关联子库：上市公司研发创新、财务指标分析、治理结构、……

CSMAR数据库应用：企业经济领域

相关研究5：绿色产业政策下投资者绿色关注对企业绿色创新影响研究[J].学术研究, 2025.

绿色创新数量： 绿色发明专利与绿色实用新型专利申请/授权数	➔	人工智能专利数量： 人工智能发明专利与人工智能实用新型专利申请/授权数	人工智能专利
绿色创新质量： 绿色专利被引用次数/绿色专利知识宽度		人工智能专利质量： 人工智能专利被引次数/人工智能专利知识宽度	
投资者绿色关注： 投资者关系互动平台中绿色相关的投资者提问数量		投资者人工智能/数字关注： 投资者关系互动平台中人工智能/数字相关的投资者提问数量	业绩说明会文本
绿色产业政策： “十二五”、“十三五”规划		人工智能/数字产业政策： 中国制造2025	人工智能产业链/数字经济



文章基于中国沪深两市A股重污染行业公司的样本，深入考察了投资者绿色关注对企业绿色创新行为的影响以及绿色产业政策的调节作用。研究发现，投资者绿色关注会导致企业绿色创新短视行为，即投资者绿色关注越多，企业的绿色创新数量越多，而绿色创新质量并没有得到显著提升；绿色产业政策能够缓解投资者绿色关注所带来的绿色创新短视行为。机制分析发现，投资者绿色关注主要通过增加外部压力来影响企业的绿色创新短视行为；绿色产业政策主要通过减轻融资约束和降低管理层短视主义来缓解投资者绿色关注所导致的绿色创新短视行为。

$$Gpatent1(Gpatent2)_{i,t+1} = \alpha_0 + \beta_1 Gattention_{i,t} + \beta_2 Controls_{i,t} + \Sigma Year + \Sigma Firm + \varepsilon_{i,t}$$

$$Gpatent1(Gpatent2)_{i,t+1} = \alpha_0 + \beta_1 Gattention_{i,t} + \beta_2 Gattention_{i,t} \times IP_{i,t} + \beta_3 IP_{i,t} + \beta_4 Controls_{i,t} + \Sigma Year + \Sigma Firm + \varepsilon_{i,t}$$

变量	(1)	(2)	(3)	(4)
	<i>Gpatent1</i>	<i>Gpatent2</i>	<i>Gpatent1</i>	<i>Gpatent2</i>
<i>Gattention</i>	0.0527*** (2.9160)	0.0137 (1.1062)	0.0460** (2.5235)	0.0100 (0.8089)
<i>Gattention*IP</i>			0.1105*** (3.4805)	0.0618** (2.4498)
<i>IP</i>			-0.0413 (-0.6928)	-0.0296 (-0.6310)
Cons	-5.6062*** (-4.7757)	-1.7816** (-2.2771)	-5.4706*** (-4.6550)	-1.6974** (-2.1473)
Controls	Yes	Yes	Yes	Yes
Year/Firm	Yes	Yes	Yes	Yes
N	5202	5202	5202	5202
Within R ²	0.1707	0.0321	0.1746	0.0343

CSMAR数据库应用：企业经济领域

相关研究6：人工智能、企业数字责任对企业社会绩效的影响—基于中国上市企业ESG表现的证据[J].西安财经大学学报, 2024.

关键解释变量	调节变量	被解释变量
人工智能技术： 人工智能专利 企业智能化转型：年报中与人工智能相关的关键词词频	数字责任相关词 汇词频	华证 ESG 评级 企业社会责任
人工智能/人工智能专利	Wingo深度学习数据库	商道融绿ESG评级/社会责任

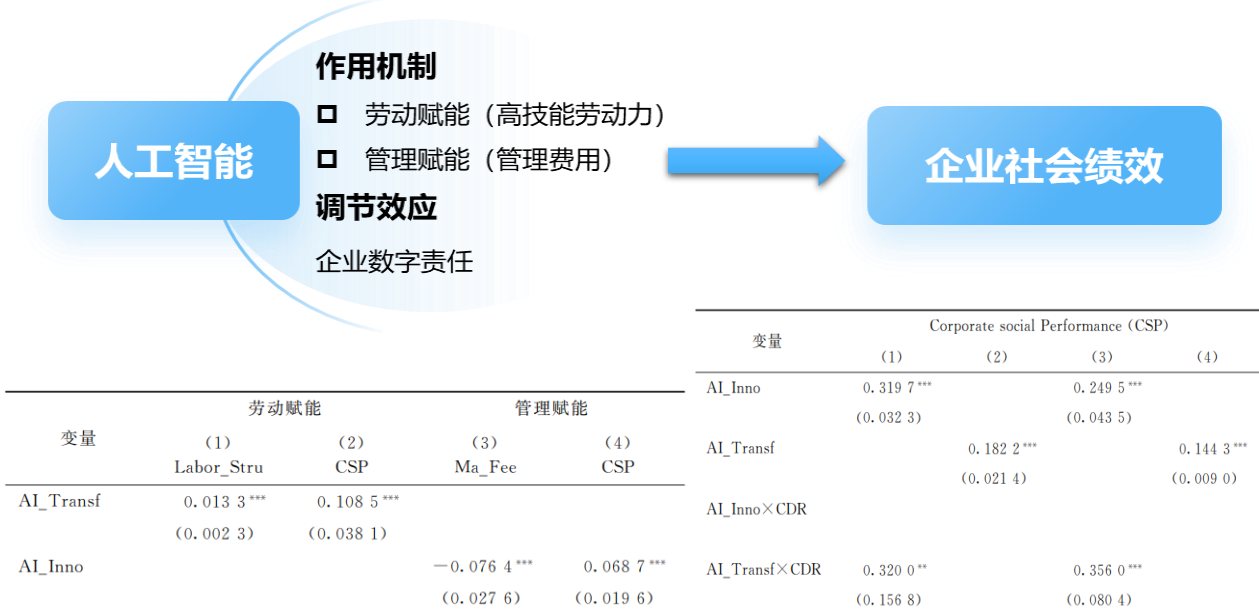
$$CSP_{it} = b_0 + b_1 AI_Inno_{it} (AI_Transf_{it}) + Controls + Year_t + City_r + Firm_i + Ind_r + \epsilon_{it}$$

Corporate social Performance (CSP)						
变量	(1)	(2)	(3)	(4)	(5)	(6)
AI_Inno		0.231 8 *** (0.013 8)			0.174 4 *** (0.018 8)	
AI_Transf			0.027 0 *** (0.007 2)			0.025 1 *** (0.004 8)

其他CSMAR关联子库：上市公司研发创新、财务报表、财务指标分析、上市公司基本信息、.....



文章以2009—2022年中国沪深A股上市企业为样本，探究了企业数字责任和人工智能对企业社会绩效的影响机理和路径，并尝试从企业发展人工智能和ESG的竞争性动机和合法性动机解释企业数字责任和人工智能影响企业社会绩效的路径。研究结果表明，人工智能技术优于企业智能化转型对社会绩效的正向影响。企业智能化转型满足了政治战略需求，通过管理赋能提升企业社会绩效；人工智能技术则体现了企业竞争战略，通过劳动赋能促进企业社会绩效。



CSMAR数据库应用：企业经济领域

相关研究7： Impact of regional artificial intelligence development on corporate environmental information[J]. Finance Research Letters, 2025.

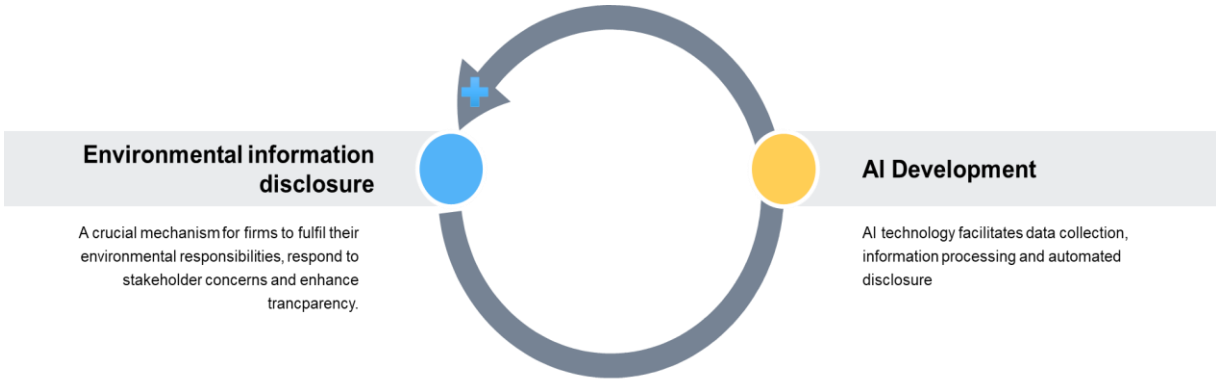
Variables	Measurement
Dependent variable: EID	Using index system to measure the extent to which listed companies disclose environmental information, including environmental management, environmental performance and governance, environmental investment, etc.
Independent variable: Reginal AI development	The number of AI companies in each city Industrial robot adoption at the city level

$$EID_{i,t} = \alpha + \beta * AI + \lambda \sum X_{i,t} + firm_i + year_t + \varepsilon_{i,t}$$

Primary indicator	Secondary indicator	Primary indicator	Secondary indicator	
Environmental management	Environmental Policies and Concepts, Environmental Goals, Environmental Management System, Environmental Education and Training, Special Environmental Actions, Environmental Incident Response Mechanism, Environmental Honors or Awards, “Three Simultaneities” System	Environmental Regulation and Certification	Key Pollution Monitoring Units, Compliance with Pollutant Emission Standards, Environmental Emergencies, Environmental Violations, Environmental Complaints, ISO 14001 Certification, ISO 9001 Certification	
Environmental performance and governance	Emissions Reduction and Control of Waste Gases, Emissions Reduction and Control of Wastewater, Dust and Smoke Control, Solid Waste Utilization and Disposal, Control of Noise, Light Pollution, Radiation, etc. Implementation of Cleaner Production	Environmental Liabilities	Wastewater Discharge Volume, COD Emissions, SO ₂ Emissions, CO ₂ Emissions, Smoke and Dust Emissions, Industrial Solid Waste Generation	
Environmental investment	Environmental Protection Investment, Pollution Discharge Fees	Environmental Information Disclosure Platforms	Annual Report, Social Responsibility Report, Environmental Report	
Baseline regression.				
	(1)	(2)	(3)	(4)
	EID1	EID1	EID2	EID2
AI	0.041**	0.034**	0.027***	0.025***
	(2.204)	(1.980)	(3.743)	(3.594)



Using data from Shanghai and Shenzhen A-share listed companies between 2008 and 2022, this study investigates the impact of regional artificial intelligence(AI) development on the quality of corporate environmental information disclosure(EID). The findings indicate that regional AI development significantly enhances the quality of corporate EID.



Related CSMAR sub-databases： Artificial Intelligence / Environmental Research / Financial Statements / Financial Indicators / Corporate Governance / China Listed Firm’s Basic Information / ...

CSMAR数据库应用：企业经济领域

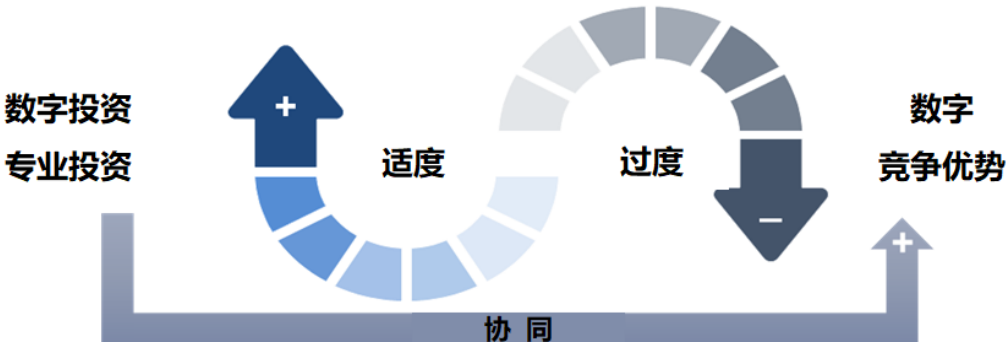
相关研究8：如何实现数字竞争优势：“专精特新”企业数字投资与专业投资的协同逻辑 [J].中国工业经济, 2025.

关键解释变量	被解释变量
企业数字投资：数字投资总额除以营业收入 专业投资：专业投资总额除以营业收入 数字投资和专业投资协同：度数字投资和专业投资的耦合协调度	数字竞争优势：①价值性（数字专利被引用次数）②稀缺性（Saidin指数）③难以模仿性（基于网络多样性的方法对IPC四位分类的技术复杂性进行测度）④不可替代性（以数字发明专利申请专利的授权公告中权利说明的数量衡量）
企业数字化转型/人工智能/财务报表/财务报表附注	incoPat平台/国家知识产权局网站

变量	(1) <i>Digpat</i>	(2) <i>Digpat</i>	(3) <i>Digpat</i>
<i>DI</i>	0.9138*** (3.1024)		
<i>DI²</i>	-0.5426*** (-3.2084)		
<i>PI</i>		0.0487*** (2.7929)	
<i>PI²</i>		-0.0034*** (-3.1549)	
<i>Syn</i>			0.0863*** (2.9293)



文章基于2012—2023年1421家中国“专精特新”企业的面板数据，探讨数字投资、专业投资及二者协同对数字竞争优势的影响效应。研究发现，数字投资、专业投资分别对数字竞争优势具有“倒U型”影响，适度的数字投资或专业投资能够为企业带来数字竞争优势，但过度投资不利于数字竞争优势的构建，数字投资与专业投资的协同能够推动企业数字竞争优势积极增长。



其他CSMAR关联子库：专精特新企业、治理结构、上市公司基本信息、上市公司研发创新、数字经济、...

CSMAR数据库应用：企业经济领域

相关研究9：数字化转型的持续性创新效应研究[J].数量经济技术经济研究, 2025.

关键解释变量

得到企业年报数字化转型词频后，基于机器学习中词频—逆文本频率（Term Frequency - Inverse Document Frequency, TF - IDF）方法计算企业数字化转型指数

数字经济 / 企业数字化转型 / 人工智能

被解释变量

采用企业研发投入和专利申请量当期与前期和的比值，分别测度企业持续性创新投入与持续性创新产出

上市公司研发创新

$$dt_{it} = \sum_w \underbrace{\{\ln[tf_{it}(w)+1]\}}_{TF} \times \underbrace{\{\ln[N_t / (n_t(w)+1)]\}}_{IDF}$$

$$rds_{it} = \ln\left[\frac{rd_{it} + rd_{it-1}}{rd_{it-1} + rd_{it-2}} \times (rd_{it} + rd_{it-1}) + 1\right]$$
$$pts_{it} = \ln\left[\frac{pt_{it} + pt_{it-1}}{pt_{it-1} + pt_{it-2}} \times (pt_{it} + pt_{it-1}) + 1\right]$$

变量	<i>rds</i> (1)	<i>rds</i> (2)	<i>pts</i> (3)	<i>pts</i> (4)	<i>rds</i> (5)	<i>pts</i> (6)
<i>L.dt</i>	0.008*** (0.001)	0.007*** (0.001)	0.003** (0.001)	0.003** (0.001)		
<i>L2.dt</i>					0.005*** (0.001)	0.003** (0.001)



持续性创新是企业发展新质生产力的重要抓手。文章以2012—2021年中国A股上市公司为样本数字化转型对企业持续性创新的影响及作用机制。研究发现，企业数字化转型显著促进持续性创新投入和持续性创新产出。机制检验表明，数字化转型通过供应链集中度降低效应、融资约束缓解效应、企业韧性增强效应作用于持续性创新投入，通过创新要素协同效应和知识协同效应作用于持续性创新产出。

数字化转型

供应链集中度降低效应、融资约束缓解效应、企业韧性增强效应

持续性创新投入

创新要素协同效应、知识协同效应

持续性创新产出

其他CSMAR关联子库：财务报表、财务指标分析、治理结构、上市公司基本信息、经营困境、...

CSMAR数据库应用：企业经济领域

相关研究10：企业管理层气候关注与ESG表现——来自中国A股上市公司的经验证据[J].南方金融, 2025.

关键解释变量

基于上市公司年报，运用文本分析和机器学习技术，构建中国企业层面管理层气候关注指数

气候风险

被解释变量

华证ESG得分

商道融绿ESG

$$\ln ESG_{i,t} = \beta_0 + \beta_1 MAC_raw_{i,t} + \gamma Controls_{i,t} + \eta_i + \nu_t + \varepsilon_{i,t}$$

变量	(1)	(2)	(3)	(4)	(5)	(6)
	lnESG	lnESG	lnESG	lnE	lnS	lnG
MCA_raw	0.422*** (6.036)	0.270*** (3.851)	0.313*** (3.954)	0.649*** (4.697)	0.609*** (3.786)	0.218* (1.829)

变量	(1)	(2)	(3)
	Environmental Regulation	Public Concern	Brokerage Report
MCA_raw	0.303*** (3.740)	0.185** (2.070)	0.330*** (4.140)
En_reg	-0.000 (-0.197)		
MCA_raw×En_reg	0.188* (1.746)		
IndexS_PH		-0.001 (-0.450)	
MCA_raw×IndexS_PH		0.063 (0.964)	
1ZR_num			0.001 (1.108)
MCA_raw×1ZR_num			0.035** (2.571)



文章基于沪深A股上市公司数据，实证检验了管理层气候关注如何影响企业的ESG表现。研究结果表明：管理层气候关注能够显著提升企业ESG表现；环境规制与券商公众号报道能够显著增强管理层气候关注对企业ESG表现的提升作用；管理层气候关注既可降低企业的经营风险和违约风险，又能激发企业的绿色创新活力。

E

更倾向于将环境可持续性纳入企业的核心战略，积极投资于清洁能源、节能减排技术以及绿色产品开发等领域，从而有效降低企业的碳足迹。

S

体现了企业对社会福祉的重视。通过积极参与应对气候变化的行动，企业不仅履行了自身的社会责任，还能够提升社会公众对企业的好感和信任度，增强企业的社会形象和品牌价值，为企业的长期发展奠定良好的社会基础。

G

更可能推动企业治理结构优化，例如通过设立环境与社会责任委员会等机构，明确各部门在气候应对中的职责与权限，并强化企业内部的环境风险监控机制。治理优化有助于企业更高效地识别、评估和应对气候相关风险与机遇，确保决策科学、合理。

管理层对气候
变化关注

企业ESG表现
提升

其他CSMAR关联子库：财务报表、财务指标分析、治理结构、上市公司基本信息、...

CSMAR数据库应用：企业经济领域

相关研究11：政府环境关注度与企业ESG表现——高管团队环保意识与企业创新的机制作用[J].科技进步与对策, 2025.

关键解释变量

环境相关词汇在政府工作报告总词数中所占比重

环境注意力

被解释变量

华证ESG得分

商道融绿ESG

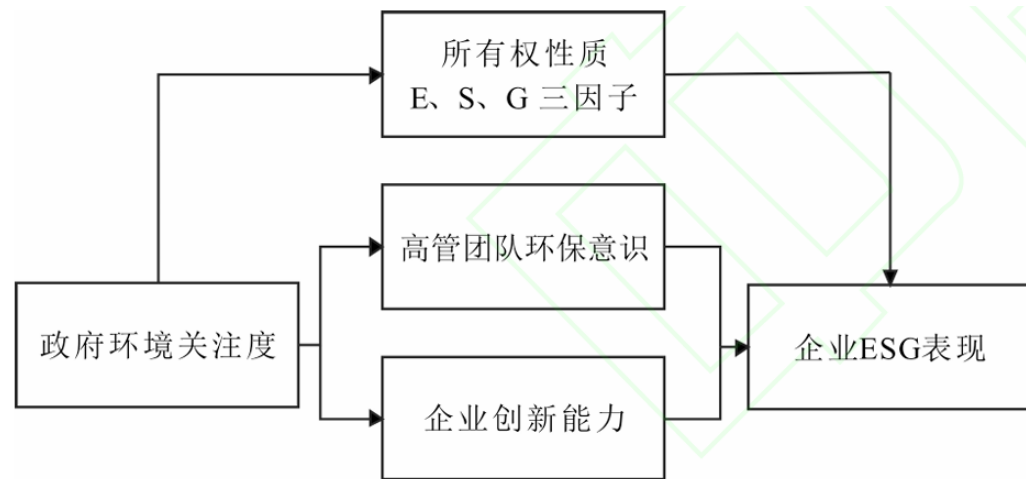
$$ESG_{it} = \alpha_1 + \beta_1 GEA_{it} + \sum_j \beta_j Controls_{it} + \theta_i + \delta_t + \epsilon_{it}$$

变量	(1)	(2)	(3)
	ESG		
GEA	23.661 ** (0.026)	14.122 * (0.074)	24.216 ** (0.021)
变量	单因子分析		
	E	S	G
GEA	1.818 (0.899)	80.454 *** (0.000)	19.490 (0.204)

其他CSMAR关联子库：财务报表、财务指标分析、治理结构、上市公司基本信息、上市公司研发创新、...



文章以2015—2022年中国A股上市公司为研究样本，探讨政府环境关注度对企业ESG表现的影响，以及高管团队环保意识和企业创新的作用机制。研究发现：政府环境关注度能够促进企业ESG表现改善，尤其对社会责任方面的改善更为显著。机制检验结果表明，政府环境关注度能够影响高管团队环保意识和企业创新文化，进而提升企业ESG表现。



CSMAR数据库应用：企业经济领域

相关研究1：基于人工智能技术的管理层语调识别及其有效性——来自对企业长期投资预效果的证据[J].财经理论与实践, 2025. 文章以上市公司管理层讨论与分析文本为研究样本，研究谷歌公司的BERT人工智能模型能否识别上市公司高管的真情实感。研究发现，模型度量下的上市公司管理层语调能反映管理层对于企业未来发展前景的态度，管理层预期任期越长，持股比例越高，内部控制水平越高时，其语调积极程度对企业长期投资的预测效果越强。

管理层语调度量



(1)爬取文本数据

使用Python从巨潮资讯网爬取本文样本期间内所有A股上市公司的年报。调用PyMuPDF库，句号、感叹号和问号设置正则表达式，选取“管理层讨论与分析”内容。



(2)训练BERT模型

①从GitHub下载谷歌 团队开源的BERT-Base (Chinese)模型。②安装PyTorch和CUDA用以准备深度学习训练，其中本文使用的PyTorch版本号为2.2.1,CUDA版本号为12. 1, Python版本号为3.10。③从年报中人为随机选择语句并标注积极/消极，其中积极语调共16567条,消极语调共18615条。④将训练样本合并,按8:2分配训练集与验证集。⑤在训练集中指定GPU进行本地训练,训练设备配置如下:GPU为 RTX 3070 Ti, CPU为Core i7-12700H,内存为16GB.⑥保存模型。



(3)计算积极程度

针对每条用正则表达式切割后的语句，采用BERT模型判断其情感积极程度。管理层语调积极程度(Tone)=(积极语句数-消极语句数)/(积极语句数十消极语句数)

$$Capex_{i,t+1} = \alpha_0 + \alpha_1 Tone_{i,t} + \alpha_i Controls_{i,t} + \epsilon_{i,t}$$

变量符号	变量名称	
Capex	企业长期投资	资本性支出/年末总资产
Tone	管理层语调	管理层语调积极程度

管理层语调积极程度与企业长期投资的回归结果	
变量	Capex
Tone	0.017*** (14.26)

管理层语调不同度量方法的比较				
变量	(1) Capex	(2) Capex	(3) Capex	(4) Capex
Diff1	0.017*** (14.33)			
Diff2		0.015*** (12.42)		
SnowTone			0.012 (1.52)	
THUTone				0.018* (1.77)
其他控制变量	控制	控制	控制	控制
N	25043	25043	25043	25043
Adj. R²	0.132	0.129	0.126	0.128

朴素贝叶斯算法分类器法：通过调用Python的SnowNLP库实现,具体步骤(1)从文本的预处理开始,将中文文本分割成单独的词汇单元。(2)从分词结果中提取特征,在提取过程中去除停用词以减少噪声,即去除“的、是、地和得”等没有实际含义的词。该类词虽然在语法结构中频繁出现,但对于捕获情感倾向贡献有限。(3)根据提取的特征,利用预先训练好的贝叶斯分类器对其进行分 类来计算文本属于积极语调的概率。(4)通 过计算某文本分词后的算术平均语调值来代表该样本管理层语调的积极程度(SnowTone)。BERT模型相较于朴素贝叶斯算法分类器法的差异(Diff1)—BERT模型管理层语调积极程度(Tone)-朴素贝叶斯算法分类器法管理层语调积极程度(SnowTone)。

词袋法：使用清华大学李军中文褒贬义词典,在对管理层讨论段落使用Python的Jieba库进行分词后,若某词在褒义词典中出现,则判断为积极词;若某词在贬义词典中出现,则判断为消极词,进而计算词袋法下的管理层语调积极程度(T HUTone)。BERT模型相较于词袋法的差异(Diff2)-BERT模型管理层语调积极程度(Tone)一词袋法管理层语调积极程度(THUTone)。

CSMAR数据库应用：企业经济领域

相关研究2：企业数字化转型的测度难题：基于大语言模型的新方法与新发现[J].经济研究, 2024.

文章运用机器学习和大语言模型构造一套新的企业数字化转型指标。作者首先对2006—2020年上市公司年报中的句子进行人工打标签，然后用标记结果训练和微调包括大语言模型在内的多个机器学习模型，选择其中分类效果最好的ERNIE模型作为句子分类模型来预测全部文本中句子的标签，最终构造了企业数字化转型指标。在此基础上，文章实证检验企业数字化转型对财务绩效的影响。研究发现：第一，企业数字化转型能够显著提高财务绩效，其中，大数据、人工智能、移动互联、云计算和物联网均有明显作用，但区块链并没有明显的作用；第二，只有在财务绩效较差的企业中，数字化转型才能够显著提高财务绩效；第三，企业数字化转型提高财务绩效的主要渠道包括改善效率和降低成本。

使用ERNIE模型测度企业数字化转型

第一步：确定文本分析对象

通过爬虫和人工收集这两个途径，收集了上市公司年报，最终得到2006—2020年4181家公司的39175份年报文本。

1



第二步：构建待预测句库和待标记句库

2



第三步：对待标记句库的句子进行人工标注

3



第四步：采用有监督的机器学习方法训练模型

4



第五步：采用ERNIE模型构造数字化转型指标

公司在当年只要使用大数据、人工智能、移动互联、云计算、区块链和物联网中的任意技术，则指标赋值为1，反之为0

5



$$Y_{i,t} = \alpha + \beta \cdot DT_{i,t} + \sum_n \chi_n \cdot Controls_{i,t} + \lambda_i + \mu_i + \varepsilon_{i,t}$$

变量名称	变量定义	观测值
A 栏：被解释变量		
ROA	净利润/总资产	25107
ROE	净利润/权益	25107
B 栏：解释变量		
DigiTech	是否数字化转型	25107
BD	是否使用大数据	25107
AI	是否使用人工智能	25107
MI	是否使用移动互联	25107
CC	是否使用云计算	25107
Iot	是否使用物联网	25107
BC	是否使用区块链	25107

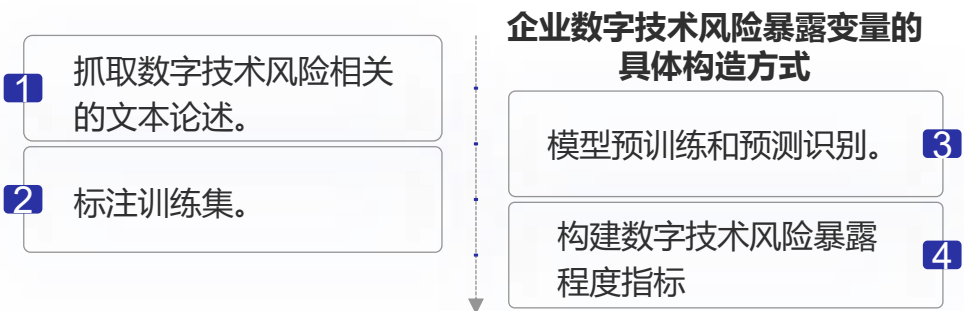
企业数字化转型与企业财务绩效的基准回归结果

变量	(1)	(2)	(3)	(4)
	ROA	ROE	F.ROA	F.ROE
DigiTech	0.398*** (0.0896)	0.740*** (0.208)	0.183* (0.0968)	0.435*** (0.218)

CSMAR数据库应用：企业经济领域

相关研究3：中国企业数字技术风险暴露对企业价值的影响——来自大语言模型的文本分析证据[J].经济研究, 2025.

文章利用FinBERT和GPT大语言模型对A股上市公司年报管理层讨论与分析部分进行文本分析，构建多层次数字技术风险暴露程度的衡量指标，探讨我国企业数字技术风险的暴露情况及对企业市场价值的影响。研究发现：年龄较大、杠杆率较高、资产收益率较低、公司治理水平较差的企业面临更严重的数字技术风险暴露。此外，数字技术风险暴露会显著降低公司的长期市场价值。其中，数据安全风险和网络安全风险均有明显负面影响。



数字技术风险相关关键词及类别

	分类	子类	关键词
数字技术风险	数据风险	数据泄露	数据安全/风险、信息安全/风险、代码安全/风险、数据丢失、信息丢失、信息泄露、数据破坏、信息毁损、数据窃取、信息窃取、窃取网络数据、隐私保护、信息保护、数据保护、非法访问、未经授权访问
		数据篡改	篡改网页、篡改网站、数据假冒、信息假冒、仿冒页面、虚假页面
		数据滥用	信息滥用
		违规传输	数据非法交易、流量异常、流量规模异常、流量内容异常
	网络风险	有害程序	网络安全/风险、互联网安全/风险、计算机病毒、脚本病毒、勒索病毒、网络感染、脚本攻击、杀毒软件、僵尸程序、蠕虫、木马、僵尸木马、勒索软件、内部威胁、程序攻击、恶意程序、恶意攻击、恶意网址、恶意软件、网络侵入、权限绕过、越权存取、越权操作、黑客、非法入侵
		网络攻击	拒绝服务攻击、后门攻击、拒绝服务、漏洞攻击、弱口令、网络扫描窃听、网络扫描、网络钓鱼、钓鱼网站、网络干扰、系统安全/风险、攻击监测、攻击资源、攻击流量、网络威胁
		设备设施故障	硬件故障、防火墙、外围保障设施故障、信息基础设施安全、设备破坏、设备故障、设施故障
		灾害性事件	断网、停电、断电、自然灾害
	具体技术风险	数字安全/风险、人工智能安全/风险、AI安全/风险、物联网安全/风险、机器学习安全/风险、云计算安全/风险、区块链安全/风险、5G安全/风险、信息技术安全/风险、自动化安全/风险	

$$Prob(Digi_Risk_Dummy_{i,k,p,t+1} = 1|X_{ikpt}, Ind_k, Year_t, Firm_i) = \frac{exp(\beta_0 + \beta_1 X_{ikpt} + \delta_{kt})}{1 + exp(\beta_0 + \beta_1 X_{ikpt} + \delta_{kt})}$$
$$Prob(Digi_Risk_Dummy_{i,k,p,t+1} = 1|X_{ikpt}, Ind_k, Year_t, Firm_i) = \Phi(\alpha_0 + \alpha_1 X_{ikpt} + \delta_{kt})$$

公司规模 (Ln_Asset)、年龄 (Ln_Age)、杠杆率 (Leverage)、有形资产比率 (PPE/TA)、营业收入增长率 (Sales_Growth)、资产收益率 (ROA)、独立董事占比 (Out_Ratio) 和第一大股东持股比例 (Top1_Sh)

$$Y_{i,k,t+j} = \gamma_0 + \gamma_1 Digi_Risk_Exposure_{ikt} + \theta Controls_{ikt} + \delta_{kt} + \mu_i + \xi_{ikt}$$

变量	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable	Tobin's Q					
Period	t+1	t+1	t+1	t+1	t+2	t+3
Digi_Risk_Exposure	-0.153*** (0.0179)	-0.136*** (0.0176)			-0.111*** (0.0188)	-0.110*** (0.0197)
Data_Risk_Exposure			-0.155*** (0.0177)			
Cyber_Risk_Exposure				-0.126*** (0.0283)		

CSMAR数据库应用:宏观经济领域

小模型范式

- [1] 智能经济对新质生产力发展的影响机制研究[J].华东经济管理, 2025.
- [2] 人工智能技术促进“双碳”目标实现机制研究[J].科技进步与对策, 2025.
- [3] 人工智能嵌入对中国“卡脖子”创新链布局重构及价值延伸的影响研究[J].现代财经(天津财经大学学报), 2024
- [4] 数据要素与人工智能技术融合对区域经济差距的影响——基于30个省份的实证研究[J].科技进步与对策, 2025.
- [5] 机器人应用、职业伤害减少与生命健康收益[J].劳动经济研究, 2024.
- [6] 东道国数字基础设施对中国高技术制造业对外直接投资的影响——基于“一带一路”沿线55国的考察[J].经济经纬, 2025.

大模型应用

- [1] 企业人工智能技术水平与劳动力需求结构变化: 基于大语言模型的新方法和新发现[J].技术经济, 2025.
- [2] Topic-sentiment analysis of citizen environmental complaints in China: Using a Stacking-BERT model[J]. Journal of Environmental Management, 2024.



宏观经济

CSMAR数据库应用:宏观经济领域

相关研究1: 智能经济对新质生产力发展的影响机制研究[J].华东经济管理, 2025.

关键解释变量

分为智能产业化、产业智能化2个一级指标，智能设施、智能创新、智能企业、智能制造等4个二级指标以及7个三级指标构建评价指标体系

人工智能/数字经济

被解释变量

从新质生产力“高科技、高效能、高质量”的特征维度出发设计评价指标体系

新质生产力



伴随新一代人工智能技术的发展，智能经济作为数字经济基础上的典型经济形态，对智能时代新质生产力的形成和发展发挥积极的推动作用。文章基于2011—2023年中国31个省份的面板数据，实证分析智能经济驱动新质生产力发展的作用机制。研究发现：智能经济对新质生产力的形成和发展具有显著的促进作用；智能经济通过数据要素创新性配置和产业深度转型升级，推动新质生产力的形成和发展。

$$Intv_{it} = \theta_0 + \theta_1 Int_e_{it} + \theta_2 Con_{it} + v_i + \sigma_t + \varepsilon_{it}$$

$$New_qp_{it} = \beta_0 + \beta_1 Int_e_{it} + \beta_2 Intv_{it} +$$

$$\beta_3 Con_{it} + v_i + \sigma_t + \varepsilon_{it}$$

$$New_qp_{it} = \alpha_0 + \alpha_1 Int_e_{it} + \alpha_2 Con_{it} +$$

$$v_i + \sigma_t + \varepsilon_{it}$$

变量	New_qp	
	(1)	(2)
Int_e	0.378 3*** (17.93)	0.154 8*** (5.91)

变量	Dig_ae	Adv_is	New_qp	New_qp
	(1)	(2)	(3)	(4)
Int_e	2.123 5* (1.71)	-0.050 3* (-1.74)	0.150 1*** (5.73)	0.160 3*** (6.13)
Dig_ae	软件业务销售收入		0.002 2** (2.03)	
Adv_is	产业结构层次			0.109 6** (2.32)

其他CSMAR关联子库：专利、企业创新、宏观经济、区域经济、经济内循环...

目标	一级指标	二级指标	三级指标	测度方法或数据来源
智能经济	智能产业化	智能设施	光缆密度	单位面积长途光缆长度
			集成电路水平	集成电路年销售额
		智能制造	算力规模	显卡进口金额对数
			智能创新力	人工智能研发水平
	产业智能化	智能企业	人工智能企业	人工智能企业数
			机器人应用水平	工业机器人安装量
		智能制造	两化融合水平	工信部两化融合发展指数

一级指标	二级指标	三级指标	测度方法或数据来源
高科技	科技水平	高技术发明专利	高技术发明专利申请数
		高技术研发机构	高技术研发机构数
	高效能	全要素生产率	劳动力投入、资本投入与经济产出比值
		研发强度	规模以上工业企业研发投入/GDP
高质量	创新发展	创新指数	中国区域创新能力评价报告
		城乡协调水平	城镇化率
		产业协调水平	第三产业占比
		单位GDP能耗	标准煤/GDP
	绿色发展	废物利用率	工业固体废物综合利用量/产生量
		对外贸易依存度	进出口总额/GDP
	开放发展	外商投资占比	外商投资/GDP
		城乡收入差距	城乡居民人均可支配收入比
共享发展		城乡消费差距	城乡居民人均消费支出比

CSMAR数据库应用:宏观经济领域

相关研究2: 人工智能技术促进 “双碳” 目标实现机制研究[J].科技进步与对策, 2025.

关键解释变量

人工智能专利申请量

人工智能专利

被解释变量

从低碳状态(LCS)和低碳潜力(LCP)两个维度出发，构建城市“双碳”践行绩效综合指标体系，并通过熵值法测度

碳中和/城乡建设/绿色专利/宏观经济

$$DCP_{it}=\beta_0+\beta_1AIT_{it}+\beta_2AIT_{it}^2+\beta_3Con_{it}+\lambda_i+\mu_t+\epsilon_{it}$$

变量	(1)	(2)	(3)	(4)	(5)	(6)
	DCP	DCP	DCP	DCP	LCS	LCP
lnAIT	-0.001 (0.001)	0.001 (0.001)	-0.042*** (0.004)	-0.031*** (0.004)	-0.031*** (0.006)	-0.031*** (0.004)
lnAIT ²			0.003*** (0.000)	0.002*** (0.000)	0.002*** (0.000)	0.002*** (0.000)

其他CSMAR关联子库：企业创新、宏观经济、经济内循环、数字经济...



文章基于2006—2022年中国城市面板数据, 考察人工智能技术对 “双碳” 践行绩效的影响, 以期为城市实现 “双碳” 目标寻找有效路径。研究发现: 人工智能技术与 “双碳” 践行绩效之间存在U型关系, 当人工智能技术达到一定水平后, 其对 “双碳” 践行绩效的影响由抑制效应转变为促进效应。分维度看, 人工智能技术对低碳潜力维度的促进效应率先凸显。中介效应分析表明, 要素升级、产业升级和动力升级发挥着重要的中介作用。调节效应分析表明,数字基础设施、数字产业发展、数字创新能力均正向调节二者间的U型关系, 并且使拐点左移。

一级指标	二级指标	三级指标	测度指标
低碳状态	能源消费	全社会电力消耗强度(千瓦时/万元)	全社会电力消耗量/GDP
		能源消费结构(%)	煤炭消费量/能源消耗总量
	经济社会结构	产业结构(%)	第二产业增加值/GDP
		绿色全要素能源效率	SBM—DEA 模型
	碳排放水平	生活垃圾无害化处理率(%)	—
		碳排放总量(万吨)	—
低碳潜力	排放趋势	GDP 碳排放强度(吨/万元)	碳排放量/GDP
		碳排放脱钩指数	碳排放增长率/GDP 增长率
	承压力	碳排放量变化率	当年相对上一年碳排放增加量/上一年碳排放量
		GDP 能源强度(吨标准煤/万元)	能源消耗总量/GDP
	能力建设	能源碳强度(万吨/万吨标准煤)	二氧化碳排放量/能源消耗总量
		服务业比重(%)	第三产业增加值/GDP
绿色生活	政策支持	科教支出强度(%)	科教支出/地方财政一般预算支出
		技术创新投资强度(%)	研发支出/主营业务收入
	生态建设	低碳政府财政支持(%)	环保支出/政府财政一般公共预算支出
		低碳政策体系完善水平	低碳政策数量
	绿色生活	人均生活用电量(千瓦时/人)	城镇生活消费用电量/常住人口数
		每万人口绿色专利拥有量(个/万人)	绿色专利申请量/常住人口数
生态建设		生活交通脱碳(辆/万人)	公共交通工具数/常住人口数
		绿化覆盖率(%)	建成区绿化覆盖面积/城市建成区面积
		人均公园面积(平方米/人)	城市公园绿地面积/常住人口数

CSMAR数据库应用:宏观经济领域

相关研究3: 人工智能嵌入对中国 “卡脖子” 创新链布局重构及价值延伸的影响研究[J].现代财经(天津财经大学学报), 2024.

关键解释变量

被解释变量

构建指标体系测度

创新链前端、中介与后端
布局重构效能的平均数

人工智能/人工智能产业链

《国家创新能力评价指标
体系》《中国科技统计年
鉴》等



文章匹配2013—2023年62个数智创新集群35个数字核心产业的三维面板数据，构建人工智能嵌入程度和创新链布局重构效能等指标，研究人工智能嵌入对创新链布局重构的影响效应和作用机理，追溯人工智能嵌入推动创新链布局重构的“足迹”，并进一步考察人工智能嵌入对创新链价值延伸的影响。结果表明：人工智能嵌入促进了创新链布局重构及其价值延伸，创新要素空间融合与创新产业技术协同发挥部分中介作用。

人工智能
嵌入程度

直接作用

- 网络连接
- 渗透赋能
- 价值创造

创新链布局
重构效能

间接作用

- 创新集聚空间融合（指标计算：基于各个集群与行业层面关键核心技术产业的创新链前端关键要素产量与后端高技术产成品）
- 创新产业技术协同（指标计算：将创新链前端关键要素产量与后端高技术产成品相加,除以跨境技术交互成本并乘以创新链中介的产业技术协同效率）

其他CSMAR关联子库：宏观经济、数字经济、经济内循环、.....

一级指标	二级指标
数智产业发展	人工智能产业业务收入(万元)
	人工智能企业总数(个)
	全国人工智能产业园区产值(亿元)
数字技术创新	协作网络密度(%)
	人工智能万人发明专利拥有量(个/万人)
	人工智能技术专利授权量(件)
	PCT 授权的人工智能专利所占比重(%)
数智企业发展	人工智能上市企业研发投入(万元)
	年新增人工智能企业数(个)
	年新增上市人工智能企业数(个)
	平均每家人工智能企业资产(万元)
数智环境发展	平均每家人工智能企业营业收入(万元)
	政府支持项目数(个)
	政府支持项目总金额(万元)
	国家级和省部级政策支持文件数(个)
数智投融资	人工智能产业领域行业标准数量(个)
	获得投资的企业数量(个)
	投资强度(万元/公顷)
	人工智能产业投融资交易额(万元)
国际影响力	人工智能产业投融资交易数量(笔)
	人工智能产业规模占全球比重(%)
	SCI&SSCI 收录占比(%)
	人工智能领域对外投资额(万元)
	本国主办的人工智能产业各类活动占全球比重(%)

$$ELRI_{F,ijt} = \frac{Pe}{dv} \sum_{i,j=1}^n \left[\sum_{k,t=1}^M \Theta_{F,ijkt} (Y_{M,ijkt} + Y_{I,ijkt}) \right]$$

$$ELRI_{I,ijt} = \frac{a\theta}{dv} \sum_{i,j=1}^n \left[\sum_{k,t=1}^M \frac{\Theta_{F,ijkt} \Theta_{B,ijkt}}{|\Theta_{F,ijkt} - \Theta_{B,ijkt}|} (Y_{M,ijkt} + Y_{I,ijkt}) \right]$$

$$ELRI_{B,ijt} = \frac{Nc}{dv} \sum_{i,j=1}^n \left[\sum_{k,t=1}^M \Theta_{B,ijkt} (Y_{M,ijkt} + Y_{I,ijkt}) \right]$$

$$ERL_{ijt} = \alpha_0 + \alpha_1 AIE_{ijt} + \theta X_{ijt} + \lambda_i + \rho_j + \vartheta_t + \varepsilon_{ijt}$$

变量	双向固定效应回归				分位数回归			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AIE	1.007 2*** (0.130 4)	1.174 7*** (0.139 1)	1.209 8*** (0.141 3)	0.988 9*** (0.129 4)	0.938 3*** (0.156 2)	0.997 1*** (0.121 0)	1.004 1*** (0.173 3)	1.129 3*** (0.134 5)
变量	创新要素空间融合		创新产业技术协同					
	FSIE _{net}		ITSE _{net}					
	(1)		(2)		(3)			
AIE	6.737 2*** (3.387 5)		1.767 7*** (0.074 4)		0.964 8*** (0.188 9)			

CSMAR数据库应用:宏观经济领域

相关研究4: 数据要素与人工智能技术融合对区域经济差距的影响——基于30个省份的实证研究[J].科技进步与对策, 2025.

关键解释变量

利用熵值法确定数据要素与人工智能技术权重，再计算耦合协调度

数字经济/经济内循环/人工智能/人工智能专利

被解释变量

$$Egap_{i,t} = \left(\frac{1}{n-1}\right) \sum_{i \neq j=1}^n |Pgap_{i,t} - Pgap_{j,t}|$$

宏观经济/区域经济

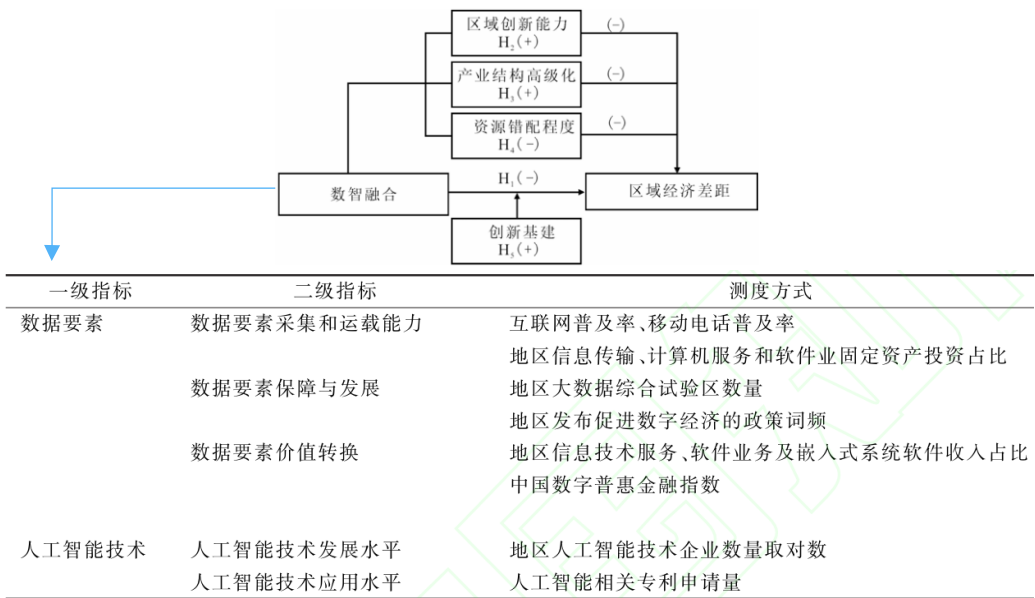
$$egap_{i,t} = \alpha_0 + \alpha_1 x_{i,t} + \delta Controls_{i,t} + \gamma_i + \lambda_t + \epsilon_{i,t}$$

变量	(1)	(2)	(3)	(4)	(5)
	<i>egap</i>	<i>egap</i>	<i>egap</i>	<i>egap</i>	<i>egap</i>
<i>x</i>	-1.621 *	-3.571 ***	-3.541 ***		
	(-1.84)	(-3.56)	(-3.47)		
<i>De</i>				-18.915	
				(-1.22)	
<i>AI</i>					-2.563
					(-1.64)

其他CSMAR关联子库: 专利、企业创新、普惠金融、.....



数据作为战略性生产要素，通过与人工智能技术融合，产生价值倍增效应，为缩小地区经济差距提供了新思路。文章利用2012-2023年中国省级面板数据，研究数智融合对区域经济差距的影响。研究发现：数智融合度越高，越能显著缩小区域经济差距；区域创新能力、产业结构高级化和资源配置效率在其中发挥中介作用，创新基建发挥调节效应。



CSMAR数据库应用:宏观经济领域

相关研究5: 机器人应用、职业伤害减少与生命健康收益[J].劳动经济研究, 2024.

关键解释变量

被解释变量

机器人渗透度

劳动者的生命健康状况：
城市年度安全生产事故事件数和死亡人数

人工智能

事件研究

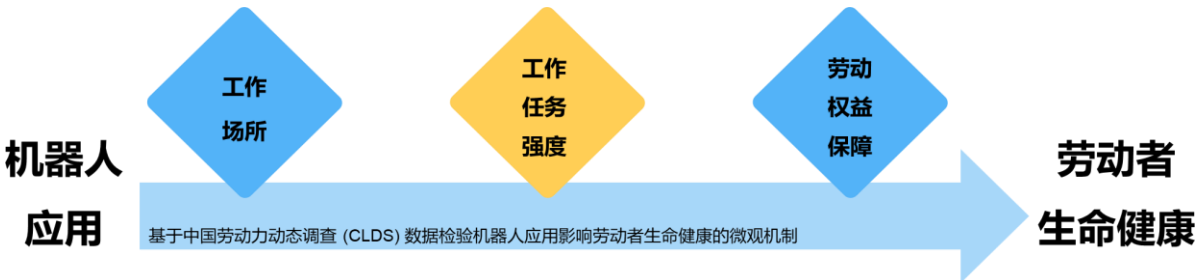
$$y_{ct} = \alpha + \beta Robot_{ct} + \gamma X_{ct-1} + \delta_c + \theta_t + \varepsilon_{ct}$$

	固定效应估计		固定效应工具变量估计	
	事故事件数	死亡人数	事故事件数	死亡人数
	(1)	(2)	(3)	(4)
机器人渗透度	-0.151 *** (0.038)	-0.555 *** (0.204)	-0.396 ** (0.193)	-2.329 ** (1.044)

	户外工作场所		繁重的体力劳动		长期合同	
	固定效应估计	工具变量估计	固定效应估计	工具变量估计	固定效应估计	工具变量估计
	(1)	(2)	(3)	(4)	(5)	(6)
机器人渗透度	-0.007 ** (0.003)	-0.019 *** (0.006)	-0.035 *** (0.012)	-0.074 ** (0.031)	0.009 ** (0.004)	0.021 *** (0.008)



文章实证考察了工业机器人应用对劳动者职业伤害和生命损失的影响。研究表明，工业机器人的应用显著减少了安全生产事故的发生及其导致的死亡人数。机制分析发现，工业机器人主要通过改善工作场所、减轻工作强度、增强劳动者权益保障来减少职业伤害。



其他CSMAR关联子库：宏观经济、区域经济、.....

CSMAR数据库应用:宏观经济领域

相关研究6: 东道国数字基础设施对中国高技术制造业对外直接投资的影响——基于“一带一路”沿线55国的考察[J].经济经纬, 2025.

关键解释变量

构建数字基础设施发展水平测度指标体系

一带一路

被解释变量

中国A股上市高技术制造业企业在“一带一路”沿线国家的直接投资，以投资事件数表示

海外直接投资

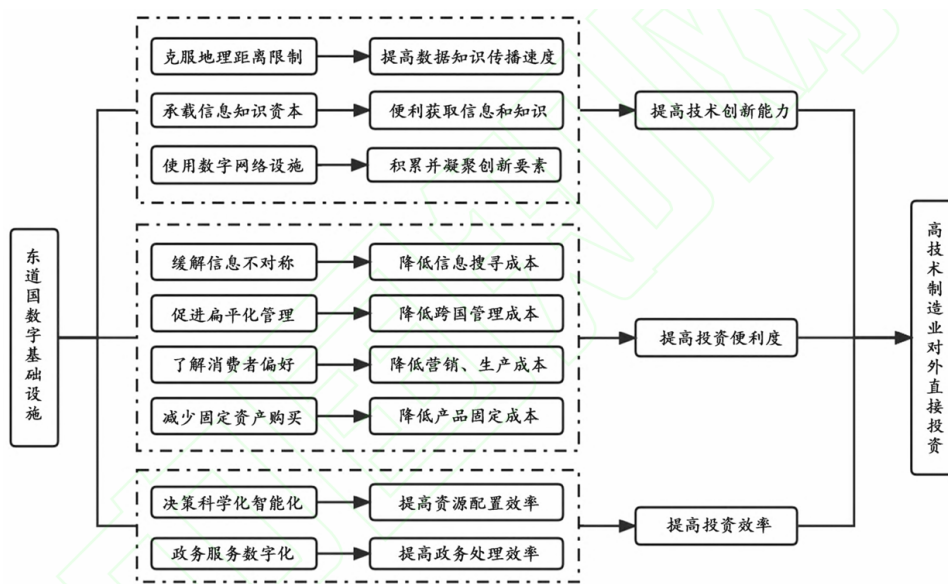


文章以“一带一路”沿线国家为样本，考察沿线国家数字基础设施发展促进中国高技术制造业对其投资的影响机制与效果。研究发现：沿线国家的数字基础设施发展促进了中国高技术制造业对其投资，东道国“数字区位优势”的投资促进效应显著。东道国数字基础设施通过提升其技术创新能力、投资效率和投资便利度三种机制，正向促进中国高技术制造业直接投资。

一级指标	二级指标
数字基础设施	移动网络覆盖人口比例
	移动宽带用户数(每百人)
	移动电话用户数(每百人)
	固定宽带用户数(每百人)
	人均互联网带宽(kbit/s)
	使用互联网的人数(占总人口的比例)
	安全服务器数(每百万人)
	高等教育入学率
	非居民专利申请数
	ICT产品出口占比

$$ofdi_{it} = \beta_0 + \beta_1 DI_{i,t-1} + \sum_{j=2}^8 \beta_j Control_{jit} + \mu_j + r_t + \varepsilon_{it}$$

变量	(1)	(2)
	<i>ofdi</i>	<i>ofdi</i>
L. DI	2.855*** (0.244)	1.175*** (0.239)



其他CSMAR关联子库：世界经济、世界经济景气指数、国际宏观.....

CSMAR数据库应用:宏观经济领域

相关研究1: 企业人工智能技术水平与劳动力需求结构变化: 基于大语言模型的新方法和新发现[J].技术经济, 2025.

文章通过优化Chinese-BERT-wwm大模型, 将2013—2019 年上市企业的招聘信息区分为非常规型和常规型工作, 测试集分类准确率近93%。同时, 利用GLM4大模型根据岗位名称和职责描述, 将其匹配到《中华人民共和国职业分类大典(2022年版)》的小类标准职业名称以识别数字职业, 分析人工智能技术(以年报中出现的相关人工智能关键词词频测度)对劳动力需求结构的影响。实证结果表明, 企业人工智能技术水平提高显著增加了对非常规岗位的需求, 减少了对常规岗位的需求, 且这一效应在非国有企业、高科技行业和制造业中尤为显著。

利用大语言模型进行工作类型分类的新方法

相较于传统机器学习方法,大语言模型在大规模文本处理上具有显著优势,它能够自动从岗位名称和职责中提取重要特征,减少手动设计的复杂性,同时提升效率和准确性。此外,大语言模型在语义理解和上下文处理方面更加有效,能够捕捉到传统方法难以识别的细微差异和复杂关系。通过预训练和迁移学习,大模型能够有效减少新任务的训练时间和数据需求,特别适合处理大规模的招聘数据。

文章利用Chinese-BERT-wwm大语言模型进行工作类型分类。首先,对原始招聘数据进行清洗,去除乱码、无关信息及缺失值,提取岗位名称和岗位职责部分的内容,同时剔除临时工、兼职和实习岗位的招聘信息,以及当年被ST(special treatment)的企业招聘信息,最终得到1568624条招聘数据。其次,从中随机抽取2000条数据构建训练集和测试集。在构建训练集和测试集时,基于Autor定义的分类原则,从两个维度结合进行工作分类,将工作划分为非常规分析型、非常规互动型、常规认知型、常规操作型、非常规操作型五类,并为每一类工作提供相应的分类规则和示例。向ChatGPT4大模型输入分类规则和已完成的分类示例,优化提示词,让大模型智能对每条招聘信息进行分类。为提高分类的准确性,本文还结合两个中文大模型(ChatGLM4和Kimi)的结果综合判断,最终确定招聘信息的工作类型。最后,人工审核分类结果。

文章的改进点在于采用混合标注策略,结合大模型标注的高效率、一致性和人工的准确性。

完成数据预处理和标注后,80%的数据用于训练,20%用于测试,利用完成标注的数据训练Chinese-BERT-wwm大模型。该模型是基于BERT模型的改进版本,针对中文进行优化,采用全词遮蔽技术,能够更好地捕捉中文的词语、语义和上下文信息。这种改进使得模型在中文NLP任务中的表现大幅提升。由于招聘信息中非常规型工作的占比较高,常规型工作的占比较低,数据存在不平衡问题,因此文章采用过采样方法平衡不同工作类别的数据量,提升模型在少数类样本上的表现。

训练微调后的Chinese-BERT-wwm大模型在测试集上的准确率达到近93%。

$$labor_{it} = \beta_0 + \beta_1 AI_{it} + \beta_2 controls_{it} + \eta_t + \delta_i + \varepsilon_{it}$$

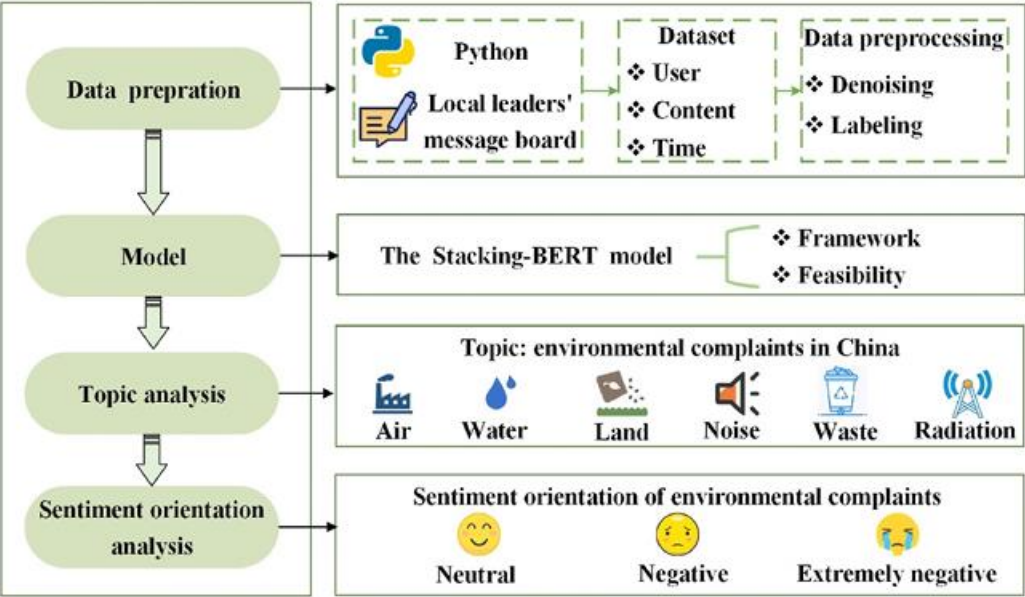
变量名
非常规员工招聘数量
常规员工招聘数量
非常规分析员工招聘数量
非常规互动员工招聘数量
常规认知员工招聘数量
常规操作员工招聘数量
非常规操作员工招聘数量
年报中人工智能关键词数(AI)
MD&A 部分人工智能关键词数[AI(MD&A)]
资产回报率(ROA)
赫芬达尔指数
资产负债率
企业成立年限
平均工资(元)
独立董事占比
本科人数占比

基准回归结果							
变量	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	非常规	常规	非常规分析	非常规互动	常规认知	常规操作	非常规操作
AI	0.0468 ** (2.137)	-0.0293 ** (-2.591)	0.0435 ** (2.663)	0.0491 ** (2.344)	-0.0262 * (-1.952)	-0.0568 ** (-2.944)	0.0359 (1.393)

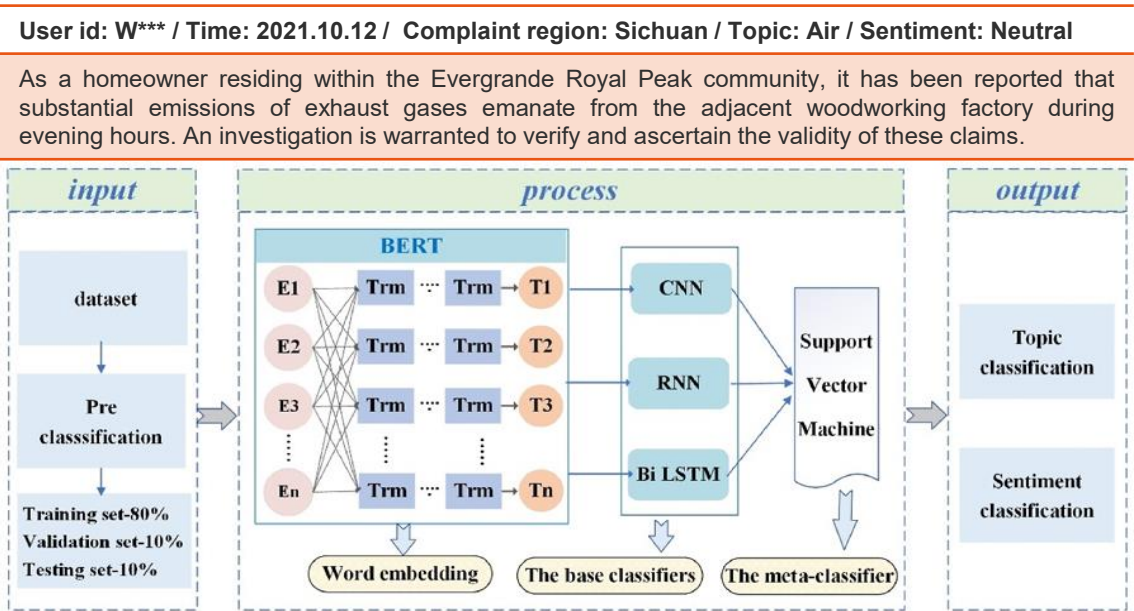
CSMAR数据库应用:宏观经济领域

相关研究2：Topic-sentiment analysis of citizen environmental complaints in China: Using a Stacking-BERT model[J]. Journal of Environmental Management, 2024.

This study analyzes 102,782 environmental complaint texts from China's e-government platform between 2016 and 2022, firstly employing an ensemble machine learning model (Stacking-BERT) to examine the hot topics and sentiment characteristics of these environmental complaints. The results indicate that: (1) The number of environmental complaints exhibits an "M-shaped" fluctuation, with the proportion of complaints related to noise, waste, and radiation continuously rising. The sentiment orientation of these complaints has shifted from predominantly negative (2016–2020) to more neutral (2021–2022), indicating a positive trend. (2) Complaints regarding air, water, land, noise, and waste demonstrate significant seasonal cyclical fluctuations, characterized by a homogenized pattern. (3) Approximately 70.41% of environmental complaints express negative emotions. Noise complaints are the dominant topic of negative emotions, while air and radiation complaints are the core themes of extreme negative emotions. (4) There is a high overlap between major complaint regions and hotspots of negative emotions.



Research framework



Framework of the Stacking-BERT model.

CSMAR数据库应用

小模型范式

- [1] 新质生产力赋能金融高质量发展：理论机制与实证检验[J].统计与决策, 2025.
- [2] 科技金融、科技创新与新质生产力——基于企业家精神的视角[J].华东经济管理, 2025.
- [3] 绿色债券与企业绿色创新：内部治理与监督效应[J].经济学动态, 2025.
- [4] 普惠金融竞争下的中小银行风险重构——基于大型银行业务下沉的实证发现[J].中国工业经济, 2025.
- [5] 中国养老金融发展的规模测算与实施路径研究[J].金融监管研究, 2025.
- [6] 数字金融影响城市绿色发展：与邻为善还是以邻为壑[J].南方金融, 2025.

大模型应用

- [1] 大语言模型、文本情绪与金融市场[J]. 管理世界, 2024.
- [2] 基于情感分析大模型的股票预测：结合GRU和ALBERT的预测模型[J]. 东岳论丛, 2024.
- [3] 基金经理语调、基金收益与投资者行为——基于BERT人工智能模型的检验[J]. 现代金融研究, 2025.

金融

CSMAR数据库应用:金融领域

相关研究1: 新质生产力赋能金融高质量发展: 理论机制与实证检验[J].统计与决策, 2025.

关键解释变量

构建包含要素基础与具化表征的新质生产力评价指标体系, 利用熵权法测算

新质生产力

被解释变量

构建金融高质量发展评价指标体系, 利用熵权法测算

宏观经济/金融科技/银行财务

$$fin_{it} = \alpha_0 + \alpha_1 npro_{it} + \alpha_2 \sum control_{it} + \delta_i + \varphi_t + \varepsilon_{it}$$

	(1)	(2)	(3)	(4)
$npro$	0.7715*** (14.26)	0.4286*** (7.15)	0.4705*** (11.54)	0.3740*** (5.09)
		(1)	(2)	
		$data$	fin	
$npro$		0.4215*** (3.43)	0.2322*** (3.82)	
$data$			0.3364*** (4.18)	
		(1)	(2)	
		静态门槛	动态门槛	
fin_{t-1}			0.6679*** (8.90)	
$npro \times I(open < 0.2425)$		0.1596** (2.16)		
$npro \times I(open > 0.2425)$		0.5583*** (4.29)		
$npro \times I(open < 0.2876)$			0.2237** (2.01)	
$npro \times I(open > 0.2876)$			0.5985*** (4.41)	

一级指标	二级指标	三级指标	指标说明(单位)
金融规模	金融产出规模	金融业增加值占比	金融业增加值与GDP的比值(%)
		银行资产占比	银行业资产总额与GDP的比值(%)
	金融资产规模	证券市场资产占比	股票总市值及债券存量之和与GDP的比值(%)
		保险收入占比	保险收入与GDP的比值(%)
金融结构	投资结构	研发领域支持	R&D经费支出与金融业增加值的比值(%)
		绿色领域支持	绿色行业上市公司融资额与股市融资额的比值(%)
		涉农领域支持	涉农贷款与全部贷款总额的比值(%)
	融资结构	股市融资	股市融资额与银行贷款余额的比值(%)
金融效率	机构结构	直接融资	直接融资在社会融资中的占比(%)
		金融机构	银行机构数量与各类金融机构数量的比值(%)
	银行规模		大型商业银行资产总额与各类银行资产总额的比值(%)
	金融科技基础	数字金融	北京大学数字普惠金融指数
金融安全		金融科技	金融科技企业数量(个)
	金融转化效率	金融杠杆率	金融机构贷款余额与GDP的比值(%)
		储蓄投资转化率	全社会固定资产投资与银行存款余额的比值(%)
	金融劳动效率	股票市场资本转化率	股市融资额与股票总市值的比值(%)
金融安全		证券化率	股票总市值与GDP的比值
	监管力度	劳动效率	金融业增加值与金融业从业人数的比值(%)
	风险水平	金融监管	金融监管支出与财政支出的比值(%)
		银行业风险	银行不良贷款率(%)
		股市风险	平均滚动市盈率(%)



文章选取我国2013—2023年30个省份的面板数据作为研究样本, 研究了新质生产力对金融高质量发展的影响效应与作用机制。结果表明, 新质生产力对金融高质量发展具有正向影响。中介效应检验结果显示, 新质生产力可通过数据要素应用赋能金融高质量发展。门槛效应分析结果证实, 随着对外开放水平的提升, 新质生产力对金融高质量发展的影响具有边际效应递增的非线性特点。

新质
生产力

直接影响

伴随对外开放水平的不断提升,
呈现边际效应递增的非线性特征

金融
高质量发展

间接影响 (数据要素)

其他CSMAR关联子库: 数字经济、经济内循环、.....

CSMAR数据库应用:金融领域

相关研究2: 科技金融、科技创新与新质生产力——基于企业家精神的视角[J].华东经济管理, 2025.

关键解释变量

从政府科技金融、市场科技金融两个维度构建科技金融指标体系,采用熵值法测度

数字经济/经济内循环/宏观经济

政府科技金融 (GTF) 采用人均财政科技支出、科技支出强度和政府研发支出进行衡量, 人均财政科技支出采用政府科技支出与区域总人口比值衡量, 财政科技支出强度采用政府科技支出占政府财政总支出的比值进行衡量, 政府研发支出采用政府R&D支出与GDP的比值衡量; 市场科技金融 (MTF) 用金融机构科技贷款、科技资本市场投入与创业风险投资进行表征, 金融机构科技贷款采用金融机构科技贷款与地区GDP比值衡量。

被解释变量

从技术革命突破、要素创新配置和产业深度升级三个维度构建新质生产力指标体系, 利用熵值法测算

新质生产力/人工智能/数字经济

$$NQPF_{it} = \beta_0 + \beta_1 TF_{it} + \beta_2 Ctrl_{it} + \mu_i + v_t + \varepsilon_{it}$$

变量	NQPF (1)	NQPF (2)	TRB (3)	RAE (4)	DUI (5)	
TF	0.529*** (0.022)	0.434*** (0.022)	0.664*** (0.033)	0.255*** (0.024)	0.317*** (0.018)	
变量	NQPF (1)	TECH (2)	NQPF (3)	NQPF (4)	TECH (5)	NQPF (6)
TF	0.434*** (0.022)	7.444*** (0.657)	0.403*** (0.023)	0.433*** (0.022)	7.419*** (0.657)	0.232*** (0.019)
TECH			0.004*** (0.001)			0.004*** (0.001)
ENTRE				0.024*** (0.003)	0.533*** (0.078)	0.514*** (0.013)
TF*ENTRE				0.041*** (0.006)	0.862*** (0.164)	
TECH*ENTRE						0.042*** (0.001)
变量	NQPF (1)	NQPF (2)	NQPF (3)	NQPF (4)		
TF (MAR _{it} ≤2.737)	0.241*** (0.081)					
TF (2.737<MAR _{it})	0.657*** (0.118)					
TF (TECH _{it} ≤9.972)		0.259*** (0.093)				
TF (9.972<TECH _{it})		0.329*** (0.104)				
TF*ENTRE (MAR _{it} ≤2.663)			0.019*** (0.024)			
TF*ENTRE (2.663<MAR _{it})			0.215*** (0.090)			
TF*ENTRE (TECH _{it} ≤10.795)				0.042 (0.002)		
TF*ENTRE (10.795<TECH _{it})				0.463*** (0.107)		



文章基于2013—2023年中国287个城市的面板数据, 实证分析了科技金融对新质生产力的影响。研究结果表明: 科技金融对新质生产力发展具有显著促进作用; 科技创新在科技金融与新质生产力之间发挥了显著的中介作用; 企业家精神在科技金融通过科技创新对新质生产力的积极影响中发挥了正向的调节作用。当制度环境达到一定水平后, 科技金融对新质生产力的促进作用显著增强。



其他CSMAR关联子库: 专利、治理结构、财务报表、.....

CSMAR数据库应用:金融领域

相关研究3: 绿色债券与企业绿色创新: 内部治理与监督效应[J].经济学动态, 2025.

关键解释变量

Treat×Post。Treat是个体虚拟变量，当企业是绿色债券发行主体时，Treat 取值为1。Post为时间虚拟变量，企业在样本期间首次发行绿色债券之后取值为1，首次发行绿色债券之前取值为0

绿色金融

被解释变量

企业绿色发明专利以及企业绿色实用新型专利申请数。

绿色专利

$$Y_{it+1} = \beta_0 + \beta_1 Treat_i \times Post_{it} + \gamma_1 X_{it} + \mu_i + \delta_t + \epsilon_{it+1}$$

变量	(1)	(2)	(3)	(4)	(5)	(6)
	<i>patt</i>	<i>pati</i>	<i>patu</i>	<i>patt</i>	<i>pati</i>	<i>patu</i>
<i>Treat × Post</i>	8.596*** (2.978)	1.931 (1.176)	6.890** (1.922)	9.109*** (3.064)	1.681 (1.230)	7.655*** (1.907)



绿色债券是加快经济发展方式绿色转型的重要金融工具。文章探究了发行绿色债券对上市公司绿色创新的影响及作用机制。研究发现：企业存在策略性创新行为，即发行绿色债券后只增加了绿色实用新型专利而非绿色发明专利的申请数量，并且对企业的绿色项目投资、研发投入、企业价值和环境绩效都没有显著的促进作用；绿色债券额度的增加会增强企业的策略性表现，并且国有企业和非高科技行业企业更可能进行策略性创新；企业发行绿色债券后的绿色创新行为受到内部治理水平、第三方认证和外部环境规制力度等治理手段的影响。

绿色
债券

企业内部治理水平
绿色债券第三方认证
环境规制强度

绿色
创新

实质性
创新 策略性
创新

其他CSMAR关联子库：上市公司基本信息、治理结构、财务报表、财务指标分析、环境研究.....

CSMAR数据库应用:金融领域

相关研究4：普惠金融竞争下的中小银行风险重构——基于大型银行业务下沉的实证发现[J].中国工业经济, 2025.

关键解释变量	被解释变量
大型银行普惠金融业务下沉：Intensity×Post。 Intensity 为强度变量，采用政策前1年（2016年）地区大型银行的普惠性小微支行数量的相反数度量； Post为政策实施的时间虚拟变量，当年份大于或等于2017年取值1。	中小银行风险：采用能够综合反映银行风险的指标Z-score测度

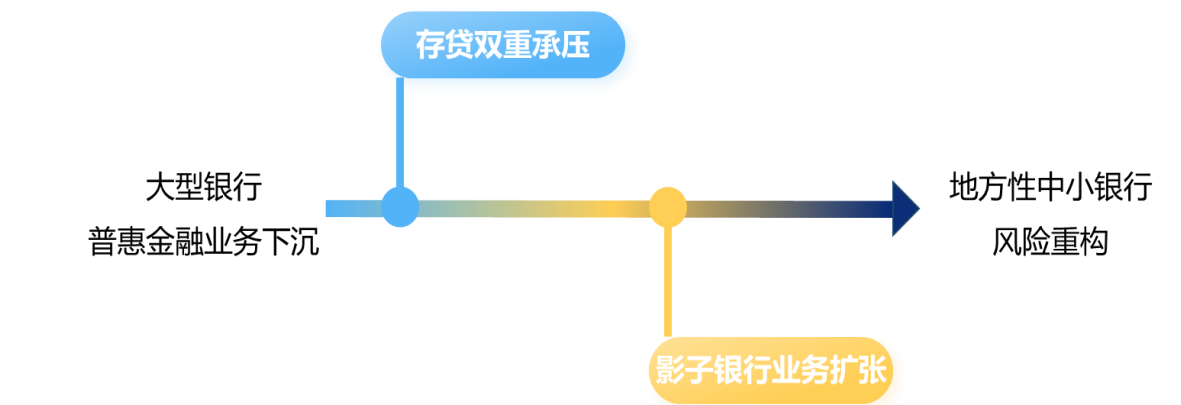
金融机构分支机构	银行财务
----------	------

$$\ln Risk_{i,j,t} = \alpha_0 + \beta_1 Intensity_{j,2016} \times Post_t + \gamma Controls + \varphi_i + \mu_t + \varepsilon_{i,j,t}$$

	(1)	(2)	(3)	(4)	(5)	(6)
	lnRisk	lnRisk	lnRisk	lnRisk	lnRisk	lnRisk
Intensity×Post	1.6620** (0.6756)	1.0313** (0.4376)	0.4967*** (0.1735)	0.9012*** (0.2411)	0.5564*** (0.1656)	0.8485*** (0.2334)



2017年11部门共同发文要求大型银行成立普惠金融事业部，发力普惠金融、缓解小微企业“融资难、融资贵”问题。那么，这种政策推动的大型银行普惠金融业务下沉是否会影响地方性中小银行的经营与风险？文章以2010—2021年地方性中小银行为样本，研究发现，大型银行普惠金融业务下沉会推升中小银行风险，引致城市商业银行和农村商业银行的风险分化，表现为城市商业银行的风险提升更为明显。影响机制显示，大型银行普惠金融业务下沉会引致中小银行存贷双重承压、影子银行业务扩张，进而抬升其风险。



其他CSMAR关联子库：宏观经济、区域经济、.....

CSMAR数据库应用:金融领域

相关研究5：中国养老金融发展的规模测算与实施路径研究[J].金融监管研究, 2025.

1. 养老资金总需求测算模型

本文从微观主体视角分析居民养老支出行为。第一步，参考刘方涛等（2022）对个人养老金潜在规模的测算模型，按资金性质将单个老年人的养老资金需求端（即退休后支出）分解为基本生活支出、医疗支出，护理和娱乐支出三项，如式（1）所示：

$$Demand_{ipt} = BasicExp_{ipt} + Health_{ipt} + Care_{ipt} \quad (1)$$

式中， $BasicExp$ 表示老年人的基本生活支出， $Health$ 表示医疗支出， $Care$ 表示护理和娱乐支出，个人养老资金需求因收入水平、教育水平、年龄、所在地区、所处经济环境等因素的不同存在显著差异。下标 i 、 p 和 t 分别表示个人、地区和时间。

第二步，将式（1）链接到基期值和经济变量，测算个人层面养老资金需求：

$$Demand_{ipt} = BasicExp_{ipt_0} \times (1 + g_e)^{t-t_0} + Health_{ipt_0} \times (1 + g_h)^{t-t_0} + Care_{ipt_0} \times (1 + g_c)^{t-t_0} \quad (2)$$

式中， t_0 表示基期，即当前年份。 g_e 、 g_h 和 g_c 分别是根据历史数据估计的基本生活支出（ $BasicExp$ ）、医疗支出（ $Health$ ）、护理和娱乐支出（ $Care$ ）的年均增长率。

第三步，测算全社会养老资金总需求。将式（2）测算的个人层面养老资金需求的平均值乘以 t 时点的退休人口总数，即可得到全社会的养老资金总需求：

$$Demand_t = \overline{Demand_t} \times RetireSize_t \quad (3)$$

式中， $\overline{Demand_t}$ 表示老年人平均养老资金需求， $RetireSize_t$ 表示 t 时点的退休人口总额。

2. 养老总资产测算方法

与退休后支出构成的稳定性不同，我国居民的养老资产配置经验尚不成熟，配置比例有较大的提升空间。本文综合考虑了居民财富增长趋势和财富结构变迁趋势，以当前居民总财富为基准，通过估计居民财富向养老资产转移的比例来测算居民养老总资产规模。由该方法得出的测算结果弥补了以往养老金缺口研究未能考虑到居民资产配置结构存在巨大优化空间的不足。



推动养老金融高质量发展是积极应对人口老龄化和建设金融强国的关键路径。文章通过分解老年人退休后养老支出和动态估计居民财富中养老资产的配置比例，分别构建了测算全社会养老资金总需求和居民养老总资产规模的理论模型，并基于模型对我国不同情境下未来居民养老资金总需求和养老总资产规模进行了测算。

测算结果：

1. 养老资金总需求：

基于中国健康与养老追踪调查数据库（China Health and Retirement Longitudinal Study，简称CHARLS）在2011—2020年间开展的五次问卷调查数据测算，预计到2049年，养老资金总需求将达到 51.32万亿元。

2. 养老总资产：

2035和2049年居民养老总资产测算（单位：万亿元）					
养老资产占比	年份	居民财富增长率			
		保守情境	基本情境	乐观情境	
		1%	3%	5%	
保守情境	5%	2035	24.21	30.64	38.59
		2049	27.83	46.34	76.40
基本情境	10%	2035	48.43	61.27	77.18
		2049	55.66	92.68	152.81
乐观情境	15%	2035	72.64	91.91	115.77
		2049	83.50	139.02	229.21

CSMAR关联子库：宏观经济、人口老龄化、.....

CSMAR数据库应用:金融领域

相关研究6: 数字金融影响城市绿色发展: 与邻为善还是以邻为壑[J].南方金融, 2025.

关键解释变量

北京大学数字普惠金融指数

北京大学数字金融研究中心

被解释变量

从生活、生产和政府治理三个维度构建城市绿色发展多维评价指标体系

经济内循环/绿色专利/城乡建设

$$ugd_{it} = \beta_0 + \beta_1 dfi_{it} + \beta_2 Control_{it} + \theta_i + \gamma_t + \varepsilon_{it}$$

$$ugd_{it} = \beta_0 + \rho_1 Wugd_{it} + \beta_1 dfi_{it} + \beta_2 Control_{it} + \theta_i + \gamma_t + \varepsilon_{it}$$

$$ugd_{it} = \beta_0 + \beta_1 dfi_{it} + \beta_2 Control_{it} + \theta_i + \gamma_t + \mu_{it}$$

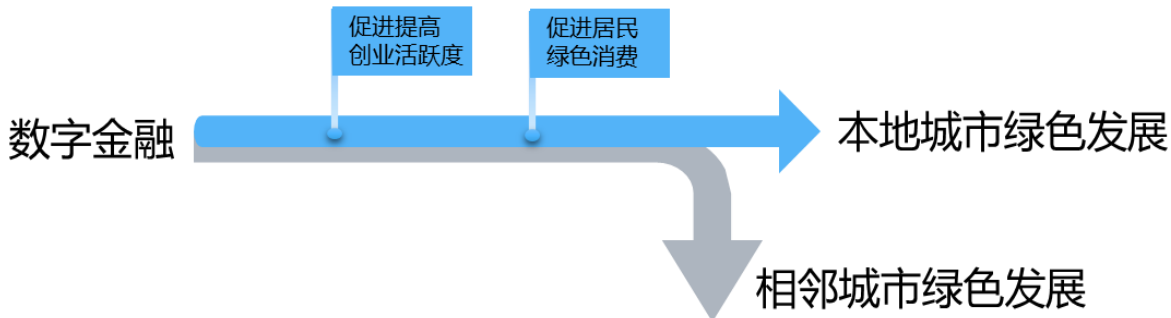
$$\mu_{it} = \lambda W \mu_{it} + \varepsilon_{it}$$

$$ugd_{it} = \beta_0 + \rho_1 Wugd_{it} + \beta_1 dfi_{it} + \xi_1 Wdfi_{it} + \beta_2 Control_{it} + \xi_2 WControl_{it} + \theta_i + \gamma_t + \varepsilon_{it}$$

变量	OLS	地理邻接矩阵	地理距离矩阵	经济距离矩阵
	(1)	(2)	(3)	(4)
$w \times ugd$		-0.3073* (-1.8169)	-0.2537* (-1.6887)	-0.0485 (-1.1352)
dfi	0.0002* (1.91)	0.0002*** (3.8497)	0.0002*** (3.8138)	0.0002*** (3.8284)
$w \times dfi$		-0.0002*** (-3.4860)	-0.0001*** (-2.2810)	-0.0002*** (-3.3489)



文章基于2011-2022年国内226个地级及以上市面板数据, 采用空间杜宾模型从供需两侧分析数字金融对城市绿色发展的影响效应和作用机制。研究表明: 第一, 数字金融能够显著促进城市绿色发展, 提升创业活跃度和绿色消费水平是数字金融赋能城市绿色发展的两条重要作用渠道; 第二, 一个城市数字金融对相邻城市绿色发展具有负向溢出效应, 表现出“以邻为壑”的特征; 第三, 一个城市数字金融通过降低相邻城市创业活跃度和绿色消费水平抑制相邻城市绿色发展。

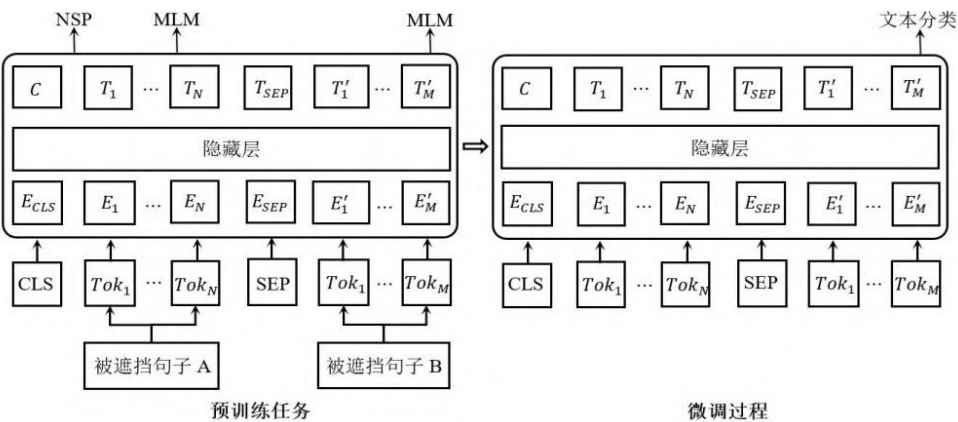


其他CSMAR关联子库: 宏观经济、区域经济、.....

CSMAR数据库应用:金融领域

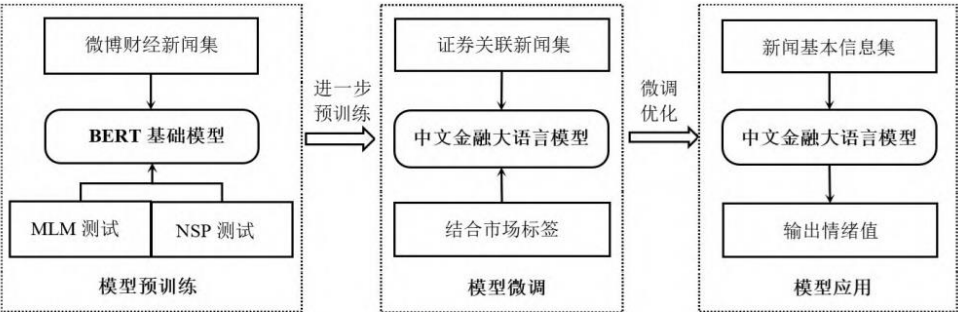
相关研究1：大语言模型、文本情绪与金融市场[J]. 管理世界, 2024.

文章创新性地融合结构化金融市场数据和非结构化金融文本大数据，并结合中国特色金融市场的独特特征，训练了一个更适用于我国金融领域的中文金融大语言模型，并开展金融市场情绪测度和资产价格风险预测。研究发现，与传统字典法相比，使用中文金融大语言模型构建的大模型情绪在金融市场回报预测方面表现显著更佳。大模型情绪对很多宏观经济变量也有显著预测能力，能够捕捉非理性情绪冲击对宏观经济基本面的影响。大模型情绪在经济下行和极端风险事件期间的预测效果更强，契合了金融理论中非理性情绪对金融市场和宏观经济会产生非对称与非线性影响的结果。综上，本研究展现了“人工智能+”行动在我国金融领域应用落地的潜在技术路径和理论逻辑。



BERT 模型框架示意图

注: CLS 是用于分类任务的标记符号; SEP 是分句标记符号, 用于断开语料; Tok 是 Token 的缩写, 代表 BERT 模型识别的文本分割单元, Tok_i 是第 i 个文本分割单元; $E_{CLS}, E_{CLS}, E_{SEP}, E_i$ 分别表示对应的词嵌入向量, C, T_{SEP}, T_i 分别表示对应经过 BERT 模型神经网络计算得到的特征向量; 箭头示意运算方向。



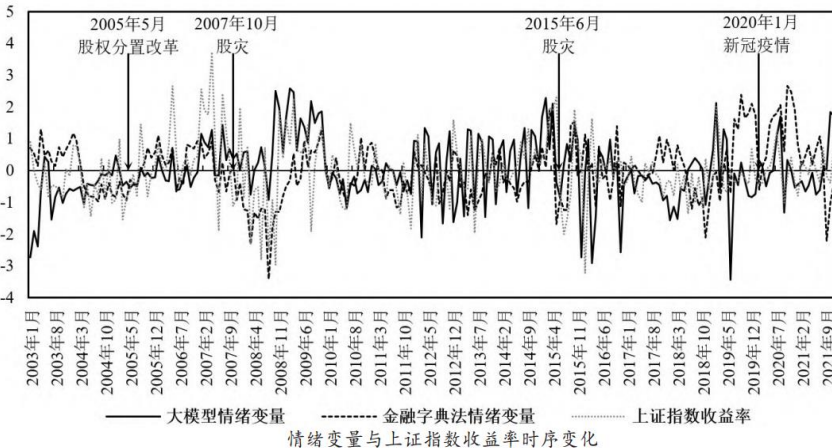
中文金融大语言模型总体应用流程图

研究数据

文章使用微博财经新闻数据集作为中文金融大语言模型针对金融领域进行特质化训练的语料。微博财经新闻数据集经过清洗后共 30 余万条有效数据。由于微博没有统一的财经新闻入口，本文首先人工列举出了“21 世纪经济报道”、“东方财富网”、“央视财经”等 21 家具有微博账户的权威新闻媒体列表，然后爬取这些账户的微博内容作为微博财经新闻数据集。尽管该数据来自微博，但由于爬取账户均为权威新闻媒体，获取的财经新闻仍然是标准的正式新闻。文章对模型预测能力进行实证检验所采用的金融市场数据与宏观经济数据选自 CSMAR 数据库。

情绪变量合成

$$S_i^k = \frac{\sum_{i=1}^{N_i} A_i^k}{N_i}, k = llm \text{ or } dic$$
$$A_i^{llm} = \begin{cases} 0, negative \\ 1, positive \end{cases}; A_i^{dic} = \frac{\sum_{j=1}^{n_i} word_{ij}}{n_i}, word_{ij} = \begin{cases} -1, negative \\ 1, positive \end{cases}$$



CSMAR数据库应用:金融领域

相关研究1：大语言模型、文本情绪与金融市场[J]. 管理世界, 2024.

文章创新性地融合结构化金融市场数据和非结构化金融文本大数据，并结合中国特色金融市场的独特特征，训练了一个更适用于我国金融领域的中文金融大语言模型，并开展金融市场情绪测度和资产价格风险预测。研究发现，与传统字典法相比，使用中文金融大语言模型构建的大模型情绪在金融市场回报预测方面表现显著更佳。大模型情绪对很多宏观经济变量也有显著预测能力，能够捕捉非理性情绪冲击对宏观经济基本面的影响。大模型情绪在经济下行和极端风险事件期间的预测效果更强，契合了金融理论中非理性情绪对金融市场和宏观经济会产生非对称与非线性影响的结果。综上，本研究展现了“人工智能+”行动在我国金融领域应用落地的潜在技术路径和理论逻辑。

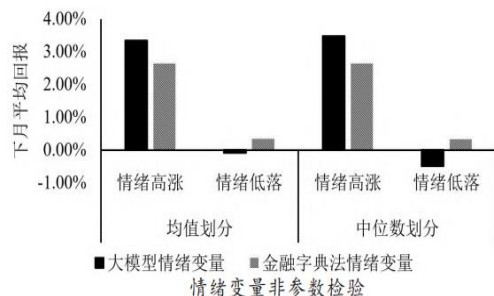
非参数检验

样本内预测能力检验

样本外预测能力检验

资产配置检验

按照各期情绪指标是否大于全样本平均值与中位数将样本划分为情绪高涨时期与情绪低落时期，然后分别计算下月等权上证指数收益率的均值。



$$r_t^m = \alpha + \beta SENTIMENT_{i,t-1} + \varepsilon_{i,t}, i = 1, 2 \quad \hat{r}_t = \hat{\alpha}_{i,t-1} + \hat{\beta}_{i,t-1} SENTIMENT_{i,t-1}, i = 1, 2$$

$$R_c^2 = 1 - \frac{\sum_{t=1}^T I_t^c (\hat{\varepsilon}_{i,t})^2}{\sum_{t=1}^T I_t^c (R_t^m - R_t^i)^2}, c = up, down \quad R_{i,OS}^2 = 1 - \frac{\sum_{t=P+1}^T (r_t^m - \hat{r}_t^m)^2}{\sum_{t=P+1}^T (r_t^m - \bar{r}_t^m)^2}, i = 1, 2$$

月度情绪与整体股票市场关系

	(1)	(2)	(3)	(4)
	等权重	等权重	市值加权	市值加权
S_{lim}	0.330*** (5.26)		0.254*** (3.95)	
S_{dic}		0.254*** (3.28)		0.235*** (3.61)
Constant	-0.004 (-0.07)	-0.002 (-0.04)	-0.005 (-0.08)	-0.003 (-0.05)
R^2 (%)	10.95	4.55	6.49	5.47
R_{up}^2 (%)	5.88	6.22	3.55	8.25
R_{down}^2 (%)	17.19	2.51	9.99	2.18

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

样本外预测结果

	区间一 (%)	区间二 (%)	区间三 (%)
S_{lim}	9.97	9.42	16.72
S_{dic}	3.85	5.20	0.66

文章检验了均值方差投资者分别利用大模型情绪变量和金融字典法情绪变量在市场组合与无风险资产间进行资产配置时所能获得的收益。根据均值方差理论，投资者投资于市场组合的最优权重为：

$$\omega_{t-1} = \frac{1}{\gamma} \frac{\hat{R}_t^m}{\hat{\sigma}_t^2}$$

投资者在第t期的组合收益率为：

$$R_t^{portfolio} = \omega_{t-1} R_t^m + R_t^f$$

确定性等价收益为：

$$CER_p = \hat{\mu}_p - \frac{1}{2} \gamma \hat{\sigma}_p^2$$

资产配置检验结果

	无交易费用		交易费用(50BP)	
情绪变量	CER gain (%)	Sharpe Ratio	CER gain (%)	Sharpe Ratio

Panel A: $\gamma=3$

	区间一			
S_{lim}	0.8834	0.2919	0.5614	0.2517
S_{dic}	0.6410	0.2633	0.4178	0.2323
	区间二			
S_{lim}	0.7973	0.2634	0.4644	0.2190
S_{dic}	0.5582	0.2309	0.3375	0.1983
	区间三			
S_{lim}	0.7153	0.2069	0.4341	0.1586
S_{dic}	0.0390	0.1003	-0.1761	0.0652

Panel B: $\gamma=5$

	区间一			
S_{lim}	0.5770	0.3050	0.3378	0.2686
S_{dic}	0.3578	0.2557	0.1485	0.2226
	区间二			
S_{lim}	0.4803	0.2702	0.2421	0.2301
S_{dic}	0.2852	0.2250	0.0775	0.1913
	区间三			
S_{lim}	0.3931	0.2017	0.1826	0.1573
S_{dic}	-0.1475	0.1004	-0.3848	0.0551

Panel C: $\gamma=7$

	区间一			
S_{lim}	0.3092	0.3005	0.1281	0.2660
S_{dic}	0.1273	0.2436	-0.0532	0.2129
	区间二			
S_{lim}	0.2128	0.2606	0.0382	0.2242
S_{dic}	0.0654	0.2174	-0.1155	0.1865
	区间三			
S_{lim}	0.2020	0.1897	0.0352	0.1457
S_{dic}	-0.2643	0.0981	-0.4766	0.0524

CSMAR数据库应用:金融领域

相关研究1：大语言模型、文本情绪与金融市场[J]. 管理世界, 2024.

文章创新性地融合结构化金融市场数据和非结构化金融文本大数据，并结合中国特色金融市场的独特特征，训练了一个更适用于我国金融领域的中文金融大语言模型，并开展金融市场情绪测度和资产价格风险预测。研究发现，与传统字典法相比，使用中文金融大语言模型构建的大模型情绪在金融市场回报预测方面表现显著更佳。大模型情绪对很多宏观经济变量也有显著预测能力，能够捕捉非理性情绪冲击对宏观经济基本面的影响。大模型情绪在经济下行和极端风险事件期间的预测效果更强，契合了金融理论中非理性情绪对金融市场和宏观经济会产生非对称与非线性影响的结果。综上，本研究展现了“人工智能+”行动在我国金融领域应用落地的潜在技术路径和理论逻辑。

经济变量预测分析

极端事件预测分析

文本特征预测分析

$$Econ_i = \alpha_i + \beta_i SENTIMENT_{i,t-1} + \varphi_i Econ_{i,t-1} + \varepsilon_{i,t}, i = 1, 2$$

$$r_t^m = \alpha + \beta SENTIMENT_{i,t-1} + \varphi_j Control_{i,t-1} + \varepsilon_{i,t}, i = 1, 2$$

月度新闻标题情绪与整体股票
市场收益

变量	(1)	(2)	(3)	(4)
	等权重	等权重	市值加权	市值加权
S_{lm}	0.214*** (3.29)		0.200*** (3.07)	
S_{lw}		0.130** (1.96)		0.136** (2.05)
Constant	-0.004 (-0.05)	-0.003 (-0.05)	-0.004 (-0.07)	-0.004 (-0.06)
R^2 (%)	4.59	1.68	4.03	1.83

情绪变量与经济变量预测

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: 经济上行期大模型情绪变量										
变量	PMI	CPI	M2	SF	F_{sw}	inv	house	PE	Turnover	LVOL
S_{lm}	0.15* (1.85)	0.02 (0.87)	0.001 (0.95)	0.05 (0.71)	0.04 (0.67)	-0.01 (-0.99)	-0.004 (-0.50)	0.05 (1.11)	0.16** (2.54)	0.04 (0.54)
Econ	0.68*** (9.47)	0.94*** (34.07)	1.01*** (884.61)	0.98*** (12.42)	0.74*** (11.98)	1.10*** (106.37)	1.11*** (113.79)	0.93*** (19.84)	0.68*** (10.98)	0.69*** (9.77)
Constant	-0.05 (-0.73)	0.03 (1.31)	0.01*** (10.23)	-0.08 (-1.03)	0.01 (0.24)	0.16*** (17.53)	0.17*** (19.45)	-0.05 (-1.20)	-0.01 (-0.17)	0.03 (0.38)
R^2 (%)	52.11	91.32	99.98	82.39	54.01	99.13	99.24	86.45	57.87	47.64

Panel B: 经济下行期大模型情绪变量

变量	PMI	CPI	M2	SF	F_{sw}	inv	house	PE	Turnover	LVOL
S_{lm}	0.06 (0.71)	-0.09** (-2.16)	0.01*** (3.35)	0.09* (1.70)	0.27** (2.58)	0.14* (1.81)	0.13* (1.81)	0.13*** (3.66)	0.27** (3.70)	-0.01 (-0.10)
Econ	0.52*** (5.69)	0.88*** (20.27)	1.01*** (430.32)	0.99*** (17.52)	0.64*** (5.32)	0.71*** (8.80)	0.78*** (10.30)	0.93*** (26.31)	0.64*** (8.82)	0.61*** (7.28)
Constant	0.04 (0.43)	-0.06 (-1.39)	0.02*** (10.55)	0.09 (1.64)	0.05 (0.51)	0.03 (0.37)	0.04 (0.52)	0.05 (1.30)	0.03 (0.42)	-0.05 (-0.62)
R^2 (%)	26.74	84.54	99.95	91.41	24.77	47.14	54.96	88.87	52.11	36.87

Panel C: 经济上行期金融字典法情绪变量

变量	PMI	CPI	M2	SF	F_{sw}	inv	house	PE	Turnover	LVOL
S_{lw}	0.01 (0.15)	0.004 (0.12)	0.0004 (0.30)	-0.001 (-0.01)	0.08 (1.30)	-0.01 (-1.10)	0.004 (0.40)	0.04 (0.96)	0.13** (2.23)	-0.17** (-2.15)
Econ	0.72*** (10.14)	0.95*** (29.03)	1.01*** (852.78)	0.97*** (12.29)	0.72*** (11.48)	1.10*** (104.50)	1.11*** (109.13)	0.96*** (22.10)	0.72*** (12.15)	0.62*** (7.88)
Constant	-0.03 (-0.44)	0.03 (1.31)	0.01*** (10.20)	-0.08 (-1.05)	0.01 (0.24)	0.16*** (17.51)	0.17*** (19.44)	-0.04 (-1.03)	-0.01 (-0.19)	0.03 (0.43)
R^2 (%)	50.51	91.26	99.98	82.12	54.47	99.13	99.23	86.40	57.36	49.45

Panel D: 经济下行期金融字典法情绪变量

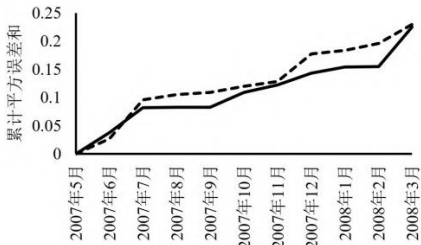
变量	PMI	CPI	M2	SF	F_{sw}	inv	house	PE	Turnover	LVOL
S_{lw}	-0.03 (-0.38)	0.04 (0.89)	0.002 (1.03)	0.02 (0.29)	0.07 (0.77)	-0.04 (-0.56)	-0.05 (-0.69)	0.09** (2.33)	0.09 (1.13)	0.03 (0.34)
Econ	0.52*** (5.61)	0.90*** (21.45)	1.00*** (415.34)	0.96*** (15.80)	0.49*** (4.51)	0.68*** (8.50)	0.76*** (10.00)	0.92*** (24.89)	0.67*** (8.68)	0.61*** (7.22)
Constant	0.04 (0.42)	-0.06 (-1.37)	0.02*** (10.04)	0.09 (1.61)	0.04 (0.45)	0.03 (0.36)	0.04 (0.50)	0.05 (1.37)	0.03 (0.42)	-0.05 (-0.61)
R^2 (%)	26.44	83.89	99.95	90.58	19.50	45.35	53.51	87.94	45.67	36.94

极端事件期间情绪变量与市场收益预测

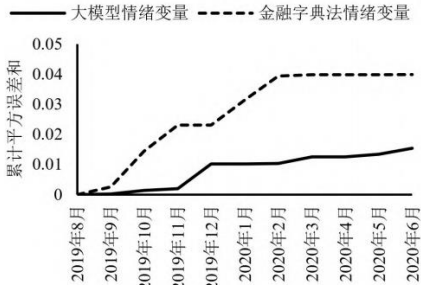
变量	(1)	(2)	(3)	(4)
Panel A: 2007 年股灾期间				
S_{lm}	0.014** (2.10)		0.012* (1.91)	
S_{lw}		0.259* (1.85)		0.227 (1.60)
Constant	Yes	Yes	Yes	Yes
Control	No	No	Yes	Yes
R^2 (%)	2.18	1.69	3.63	3.11

Panel B: 2020 年新冠疫情期间

S_{lm}	0.005* (1.87)		0.005* (1.89)	
S_{lw}		0.057 (1.43)		0.069 (1.67)
Constant	Yes	Yes	Yes	Yes
Control	No	No	Yes	Yes
R^2 (%)	1.74	1.03	1.83	1.30

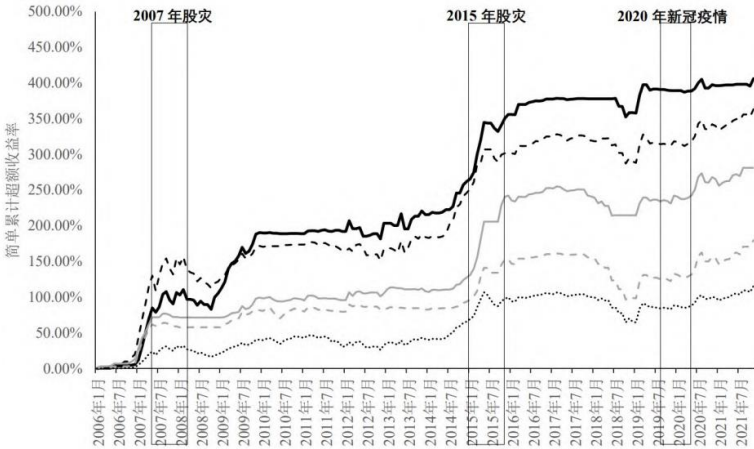


2007 年股灾期间



2020 年新冠疫情期间

—— 大模型情绪变量 —— 金融字典法情绪变量
极端事件期间累计平方误差和变动趋势



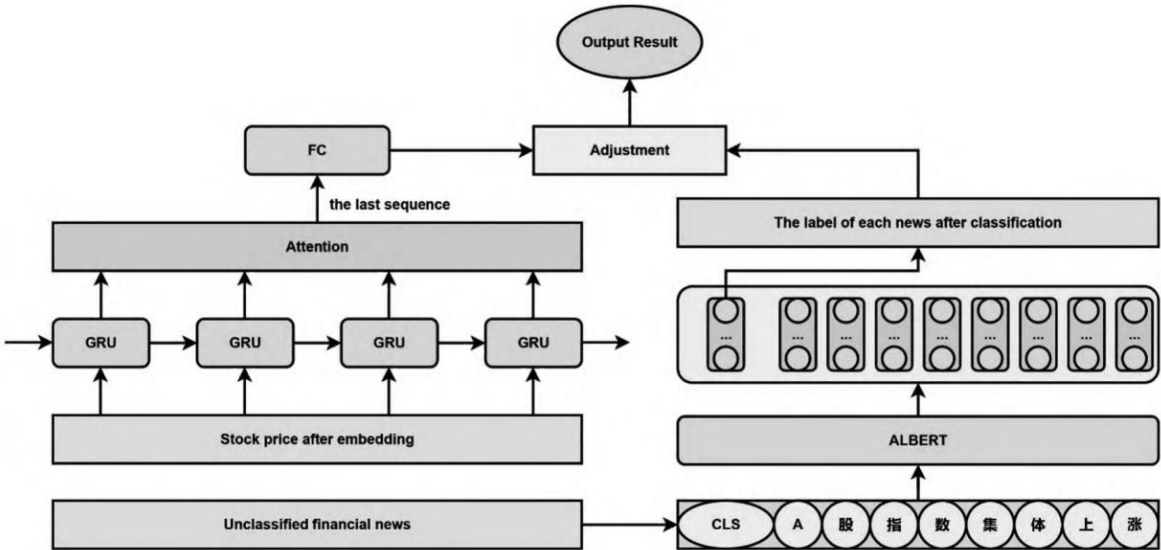
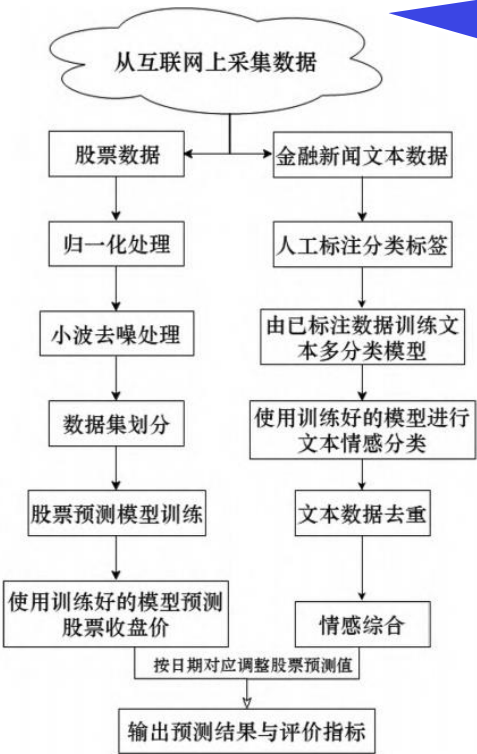
—— 大模型情绪变量(短) —— 大模型情绪变量(长) —— 金融字典法情绪变量(短)
—— 金融字典法情绪变量(长) 基准预测
文本情绪资产组合与基准预测资产组合的简单累计超额收益率时序变动

CSMAR数据库应用:金融领域

相关研究2：基于情感分析大模型的股票预测：结合GRU和ALBERT的预测模型[J]. 东岳论丛, 2024.

文章提出了一种新的股票价格预测方法，该方法利用GRU深度学习网络，根据股票历史数据进行股价的预测，并构建出基于GRU框架的预测模型。同时，采用轻量级大模型 ALBERT，对媒体新闻进行情感提取，获取情感特征。最后，将情感特征输入至于GRU预测模型中，并适应性调整GRU的输出结果，以获得最终的股票预测。本文最终选择了GRU-attention与Google Chinese ALBERT-Large结合的方案，该方案充分利用了客观特征以及情感特征包含的信息，使得输入模型的特征更加全面，并且在预测精度上相对其他方案有明显的优势。从实际应用上来看，本文搭建的股票预测模型能够为投资者提供更为准确的股票价格预测结果，帮助投资者制订更加合理的投资决策，并可以预警金融市场中可能发生的风险，帮助投资者规避损失。

研究数据：股票数据部分，采用宁德时代2021年7月1日至2022年8月3日的股票数据信息，包括开盘价、收盘价、最高价、最低价等；金融新闻文本数据部分，爬取了云财经、新浪财经、东方财富、同花顺等金融新闻数据。相关数据可从CSMAR股票市场交易等子库获取。



模型结构

不同模型的预测结果评估指标

model	accuracy_value	f1_trend	final_score
ResNet34	0.7551	0.5306	0.6878
ResNet101	0.6122	0.5306	0.5878
ResNet152	0.6531	0.5769	0.6302
DNN	0.6327	0.5098	0.5958
CNN	0.6531	0.5098	0.6101
LSTM	0.8163	0.5417	0.7339
LSTM-attention	0.5918	0.6415	0.6067
GRU	0.7959	0.4800	0.7011
GRU-attention	0.8776	0.5385	0.7758
GRU-attention-sum	0.7551	0.4348	0.6590
GRU-multi-head-attention	0.7755	0.5882	0.7193

不同文本分类模型的训练效果

model	accuracy	f ₁ _trend
HFL Chinese BERT-wwm	0.7025	0.6403
Google Chinese BERT-Base	0.7113	0.6536
Google Chinese ALBERT-Large	0.7313	0.6724
HFL Chinese RoBERTa-wwm-ext	0.7038	0.6483
HFL Chinese RoBERTa-wwm-large-ext	0.7013	0.6444
HFL Chinese BERT-wwm-ext	0.7063	0.6427
Google Chinese ALBERT-Base	0.6775	0.5859

CSMAR数据库应用:金融领域

相关研究3: 基金经理语调、基金收益与投资者行为——基于BERT人工智能模型的检验[J]. 现代金融研究, 2025.

文章以2014-2023年我国基金年度报告“展望”章节文本为样本,运用BERT人工智能模型和传统的词袋法,分别构建“词藻堆砌”与“真情实感”变量,对比基金经理语调对基金未来收益的预测效果及其对个人投资者行为的影响。研究发现:(1)相比传统的词袋法,BERT人工智能模型度量的基金经理语调更能识别基金经理的真情实感,对基金未来收益、顺境中的基金业绩持续性及逆境中的基金崩盘风险的预测效果更好,且上述识别优势在复杂度较高、可读性较差的“展望”文本中更显著;(2)个人投资者的“有限理性”特征和公募基金存在的委托代理问题导致个人投资者更容易被基金经理“词藻堆砌”的文本所吸引,进而导致投资收益率降低。

语调变量的度量

1.“真情实感”程度

文章使用BERT语言模型构建基金经理语调“真情实感”程度指标的过程如下。(1)年报文本解析。从聚源数据库中提取我国公募基金年度报告文本,节选“展望”章节文本内容,并基于句号、问号、感叹号切分成语句。(2)模型微调。①从Github下载熵简科技基于中文金融语料进行预训练的开源FinBERT (Chinese) 模型①。②从“展望”文本中随机选择语句,并人工标记“积极”“中性”或“消极”的标签形成调整样本,共2650条。③将训练样本合并,按8:2分配训练集与验证集。④进行本地训练。⑤保存模型并进行模型评估。(3)判断语调积极程度。针对每条语句,采用FinBERT (Chinese) 语言模型判断其情感积极程度。参照谢德仁和林乐(2015)、曾庆生等(2018)的度量方法,基金经理语调“真情实感”程度(Emotion) = (积极语句数-消极语句数) / (积极语句数+消极语句数)。Emotion越大,表示基金经理语调越积极。

2.“词藻堆砌”程度

面对基金经理有意或无意的“词藻堆砌”行为,虽然“文本分词技术+词频统计”的方法难以识别其“真情实感”,但是对积极与消极词“词频统计”所得到的语调积极程度(即“积极词语的频率”)提供了一种识别“词藻堆砌”行为的方法,即通过BERT模型和传统的词袋法对基金经理语调积极程度的度量差异衡量“词藻堆砌”程度(Diff),以反映真情实感不足却使用了大量积极“词藻”的“展望”文本,具体过程如下。(1)构建词袋法度量的语调指标。本文采用姚加权等(2021)提出的《金融情绪词典(正式文本)》,构建了传统的词袋法度量的基金经理语调积极程度指标(Words)。(2)对语调指标标准化处理。本文对BERT模型和传统的词袋法构建的基金经理语调积极程度指标分别进行减去自身均值,并除以自身标准差的标准化处理,使不同度量方法的语调指标具有可比性。(3)构建“词藻堆砌”程度变量(Diff)。“词藻堆砌”程度变量(Diff) = 标准化后的语调积极程度(词袋法) - 标准化后的语调积极程度(BERT)。“词藻堆砌”程度变量(Diff)越大,表示传统的词袋法度量的基金经理语调相比基于BERT度量的语调更积极,即“词藻堆砌”程度越高。

H1: 相比词袋法, BERT模型更能识别基金经理语调的“真情实感”, 对未来基金收益的预测效果更好。

H2: 个人投资者无法识别基金经理语调的“真情实感”, 却会被“词藻堆砌”的基金经理语调所吸引, 并反映在当期的投资决策上。

$$Performance_{i,t+1} = \alpha_0 + \alpha_1 Words_{i,t} + \alpha_i Controls_{i,t} + \varepsilon_{i,t}$$

$$Performance_{i,t+1} = \alpha_0 + \alpha_1 Emotion_{i,t} + \alpha_i Controls_{i,t} + \varepsilon_{i,t}$$

	(1) Performance	(2) Performance	(3) Performance	(4) Performance
Words	-0.050 (-0.09)		-0.269 (-0.51)	
Emotion		0.850*** (4.20)		0.774*** (3.88)

$$Individual_{i,t+1} = \alpha_0 + \alpha_1 Emotion_{i,t} + \alpha_i Controls_{i,t} + \varepsilon_{i,t}$$

$$Individual_{i,t+1} = \alpha_0 + \alpha_1 Diff_{i,t} + \alpha_i Controls_{i,t} + \varepsilon_{i,t}$$

Panel A:个人投资者对基金经理语调“真情实感”的识别效果

	(1) Individual1	(2) Individual2	(3) Individual1	(4) Individual2
Emotion	-2.909*** (-7.75)	-0.603*** (-5.02)	-2.817*** (-7.55)	-0.628*** (-5.26)

Panel B:基金经理语调“词藻堆砌”对个人投资者的吸引效果

	(1) Individual1	(2) Individual2	(3) Individual1	(4) Individual2
Diff	1.521*** (7.84)	0.360*** (6.15)	1.509*** (7.82)	0.362*** (6.23)

03 CSMAR最新数据资源

- 最新数据资源
- 新库资讯获取



最新数据资源

2021

- 上市公司环境研究
- 供应链研究
- 银行治理研究
- 绿色专利研究
- 经济内循环研究
- 经济地理研究
- 润灵环球ESG评级
- 商道融绿ESG评级
- 金融科技
- 数字经济
- 经营困境

2022

- 碳中和
- 政府审计
- 企业共同富裕
- 美股文本系列
- 港股市场交易
- 普惠金融
- 县域二氧化碳排放
- 人口老龄化
- 资本市场监管
- 腾景全口径
- 会计差错更正
- 上市公司数字化转型
- 智能研究系列
- 龙头股
- 分析师一致预测

2023

- 城乡建设
- 美股系列
- 融资租赁
- 工业企业
- 上市公司与关联公司专利
- QFII持股
- 员工持股计划
- 新三板系列
- 房地产风险
- 专精特新企业
- 区域热辐射

2024

- 中汉指数
- 债券财务系列
 - 债券财务指标分析
 - 债券财务衍生报表
- 破产重整
- 企业碳信息披露指数
- 上市公司联盟
- 企业可持续增长CSG
- 市场竞争
- 新质生产力
- 董事长致辞
- 市场分析报告文本

2025

- 人工智能
- 人工智能产业链
- 人工智能专利
- 证券市场中介机构
- 宏观税收系列
- 证券市场监管公告
- 业绩说明会文本
-

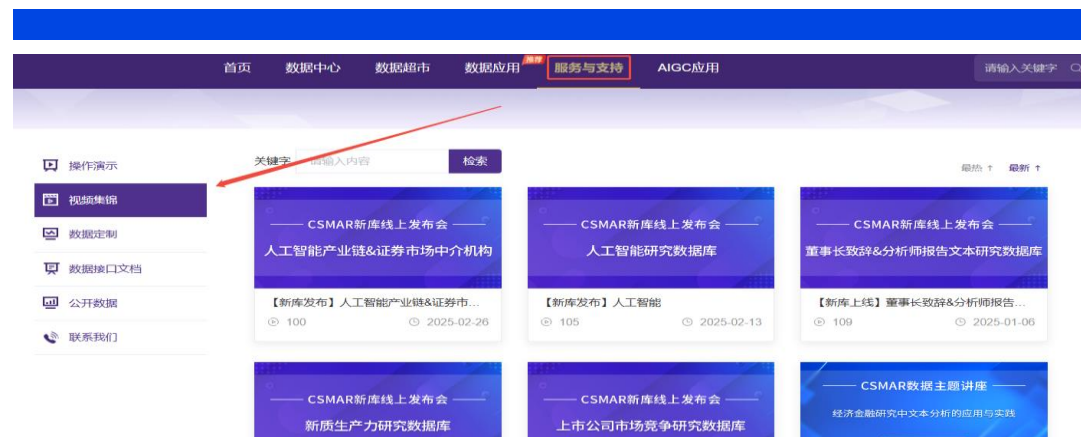
[更多新库即将上线!](#)

新库资讯获取



新库上线动态

访问CSMAR主页 (<http://data.csmar.com/>)，在“产品动态”模块即可了解CSMAR新库上线动态，点击文章标题可获取新库详细介绍。



新库发布会视频

访问CSMAR主页 (<http://data.csmar.com/>)，在“服务与支持-视频集锦”模块即可观看往期CSMAR新库发布会视频。

04 CSMAR科研资讯

- 讲座资源
- 科研频道



讲座资源

CSMAR实证研究与论文写作学习资源

为帮助广大学子与科研工作者系统解决在实证研究与论文写作当中的各种难题，CSMAR联合各大高校，广邀名师，定期开展**CSMAR实证论文复刻**线上讲座，以精彩纷呈的主题讲座，与广大学子共攀学术高峰。

CSMAR实证论文复刻讲座		
	主题	讲师
2024	社会关系网络与金融机构经济行为研究	余峰燕 天津大学
	财政与金融的绿色政策协调效果研究	洪祥骏 厦门国家会计学院
	方言距离与VCF风险承担	李建军 华东理工大学
	Does Good News Cover Bad News?	孟庆斌 中国人民大学
2023	艺高人胆大？标的公司创新能力与签订业绩补偿承诺意愿	李哲 中央财经大学
	国内权威期刊发文经验分享——立足中国情景研究本土问题	连燕玲 华东师范大学
	一篇SSCI是怎样炼成的？	许元镒 上海商学院

- ✓ 关注**CSMAR**微信视频号
- ✓ 访问CSMAR主页“服务与支持-视频集锦”模块 (<http://data.csmar.com/>)

即可免费回看往期精彩讲座视频！



科研频道

CSMAR
官方公众号



数据资源上新



权威文献解析



科研活动资讯



热门文章推荐

- ◆ 人工智能深度研究系列（一）：人工智能研究数据库的科研应用
- ◆ 人工智能深度研究系列（二）：人工智能产业链研究数据库的科研应用
- ◆ 人工智能深度研究系列（三）：人工智能专利数据库的科研应用
- ◆ 业绩说明会文本数据的科研应用
- ◆ 分析师深度研究报告赋能文本分析
- ◆ 董事长致辞：年报文本分析新视角
- ◆ 研究新质生产力，赋能高质量发展
- ◆ 市场竞争研究新视点
- ◆ 论员工持股的利与弊
- ◆ QFII持股的研究与思考
- ◆ CSMAR专利数据与创新领域研究深化
- ◆ 新型工业化背景下的工业企业研究
- ◆ 企业数字化转型研究探索
- ◆ 董监高责任险与企业可持续发展
- ◆ 数字经济时代的会计差错更正研究



CSMAR

感谢参与

欢迎关注CSMAR官方公众号

获取更多学术科研信息！



深圳希施玛数据科技有限公司

<https://data.csmar.com/>