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：杨曼莎

深圳希施玛数据科技有限公司

https://data.csmar.com/

内容大纲

1 **CSMAR数据库简介**
- 数据库概况
- 基础操作指引

2 **CSMAR数据库应用**
- CSMAR与热门主题
- CSMAR与前沿主题

3 **CSMAR最新数据资源**
- 最新数据库概览
- 新库资讯获取

4 **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

数据库概况

全　　称：China Stock Market & Accounting Research Database

中国经济金融研究数据库

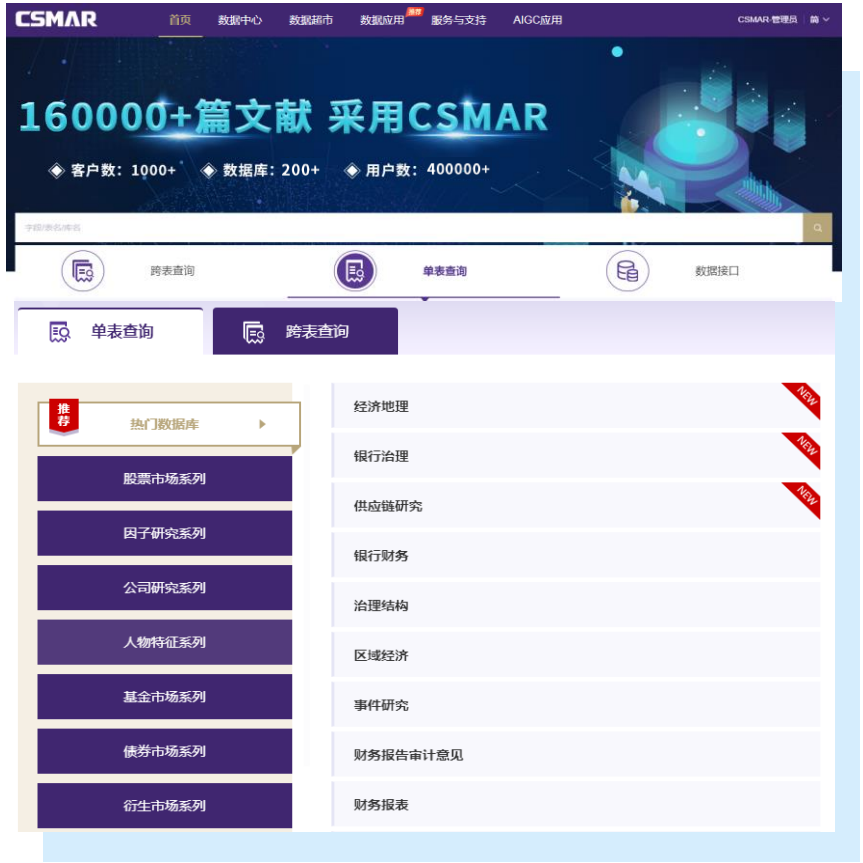
定　　位：**研究型精准**数据库

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/



数据库概况

18+1
系列

200+
数据库

70000+
字段

- 公司研究系列
- 股票市场系列
- 债券市场系列
- 人物特征系列
- 银行研究系列
- 基金市场系列
- 经济研究系列
- 专题研究系列
- 商品市场研究系列
- 海外研究系列
- 行业研究系列
- 智能研究系列
- 绿色经济系列
- 市场资讯系列
- 货币市场系列
- 因子研究系列
- 历史数据
- 衍生市场系列
- 合作数据

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	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、市场分析报告及公告		
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

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。

01

权威

- **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

02

基础

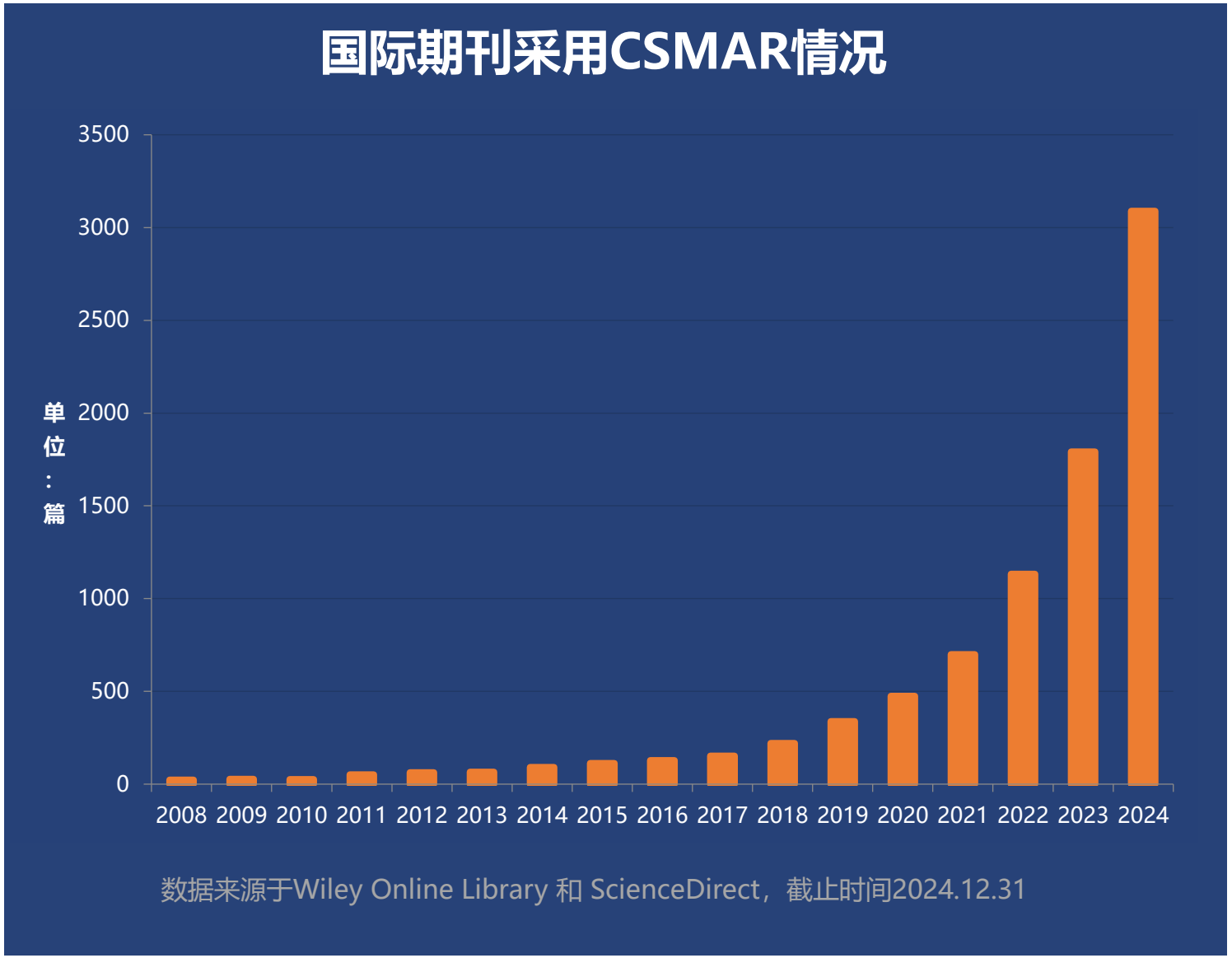
- CSMAR数据被**摩根斯坦利**选用，作为编制MSCI-A股指数的**基础**。

03

引用

- 截至2024年12月31日，国内外使用**CSMAR**数据发表的论文超过16万篇。

数据库概况



截止时间2024.12.31 ， 国际顶刊累计发表量

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、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/

02 CSMAR数据库应用

- CSMAR与热门主题
- CSMAR与前沿主题

CSMAR与热门主题

热门主题：采用CSMAR数据发文较多的研究主题，展现CSMAR数据适用的经典领域。

企业经济	金融	宏观经济
公司治理	股票流动性	经济增长
融资约束	股价崩盘风险	全球价值链
内部控制	金融稳定	环境规制
……	……	……

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。

参考资料：2024研究前沿. 中国科学院科技战略咨询研究院，中国科学院文献情报中心，科睿唯安,2024.

CSMAR与热门主题：（一）企业经济

方向	主题	参考范文	关联CSMAR数据库
企业经济	公司治理	高管环保背景如何促进提高企业ESG表现？——来自中国A股上市公司的经验证据 地区新质生产力对企业全要素生产率的影响与机制	财务报告审计意见、上市公司研发创新、财务指标分析、上市公司人物特征、机构投资者、行业财务指标、上市公司基本信息、财务报表附注、财务报表、海外直接投资、股东、资质认定、民营上市公司、股权性质、治理结构、审计研究、股权质押、分析师预测、一致预测、经营困境、家族企业、社会责任、环境研究、政府审计、企业创新、会计信息质量、会计差错更正、数字经济、企业数字化转型、供应链研究、资本市场监管、企业共同富裕、董监高责任险、龙头股、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、董事长致辞、市场分析报告文本、新质生产力、人工智能、人工智能专利
	融资约束	数字经济、融资约束与企业创新质量——来自高新技术上市公司的经验证据 融资约束与企业创新：基于面板数据模型的实证分析	
	内部控制	互动平台信息交流能否提升上市公司内部控制质量？——来自“互动易”和“e互动”的数据证据 控股股东股权质押与企业违规行为——基于内部控制与外部监督视角	
	盈余管理	双碳目标背景下高碳排放企业低碳转型会提升真实盈余管理吗 企业盈余管理程度对社会责任报告可读性的影响研究	
	投资效率	企业生产网络内的差异化税率与投资效率损失 气候政策不确定性、数字化转型与企业投资效率	
	会计信息质量	财务共享对会计信息质量的影响：治理还是代理？ 董事高管责任保险与会计信息质量——基于党组织参与治理的替代或互补效应	
	企业创新	“防微杜渐”还是“推波助澜”？散户“在线发声”与企业策略性创新信息披露 人工智能技术应用如何影响企业创新 企业体育健康投资的创新效应——基于企业开展体育运动的经验证据	
	企业社会责任	数字经济发展驱动企业社会责任履行的内在机理与统计检验 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	
	审计质量	企业气候变化风险与关键审计事项决策 企业披露数据资源会影响审计质量吗？——基于审计师的多维学习视角	
	股权质押	机构投资者相机抉择与控股股东股权质押及其公司治理效应 大股东支持行为的融资效应研究：从控股股东为上市公司进行股权质押到监督型资本	
	家族企业	家族企业代际传承与年报文本信息可读性 儒家文化中的创新因子——来自中国家族企业的证据	
	混合所有制改革	混合所有制改革与企业ESG表现：国企与民企的比较研究 混合所有制改革能否促进国有企业数字化转型？	
	企业金融化	“新瓶装旧酒”还是“旧貌换新颜”：地方金融监管与企业金融化 企业数字化转型能提高劳动收入份额吗?——基于企业金融化视角的经验证据	
	分析师	分析师社交圈、利益冲突与研究报告文本信息 上市公司产品竞争与分析师研究报告文本信息	

CSMAR与热门主题：（一）企业经济

相关研究： 人工智能技术应用如何影响企业创新[J].中国工业经济, 2024.

关键解释变量	被解释变量
AI技术应用：上市企业年报中AI技术关键词数量加1取对数	企业创新：上市企业当年发明专利申请量加1取对数
人工智能	上市公司研发创新/上市公司与关联公司专利/专利被引用

$$Innov_{it} = \alpha_0 + \alpha_1 AI_utilize_{it} + \delta X + \lambda_i + \mu_t + \varepsilon_{it}$$

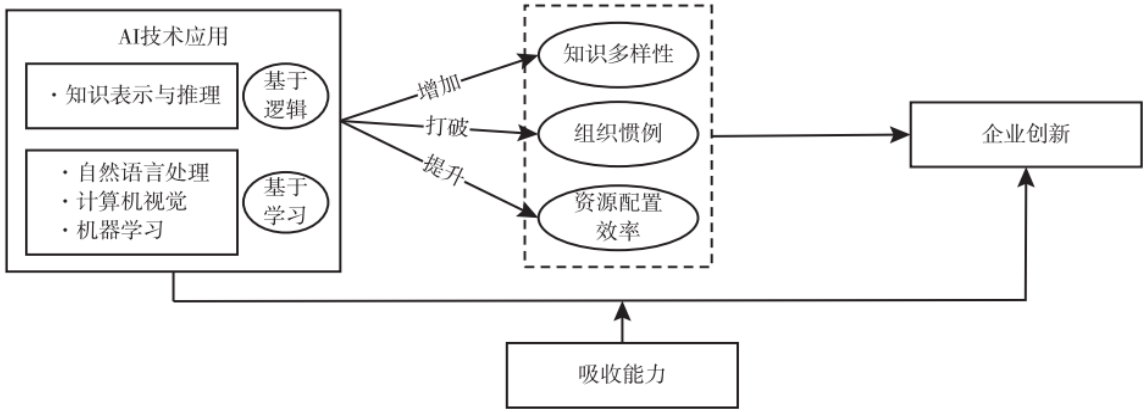
变量	(1)	(2)	(3)	(4)
	Innov	Innov	Innov	Innov
AI_utilize	0.4480***	0.2097***	0.1999***	0.0671***
	(0.0317)	(0.0293)	(0.0286)	(0.0207)

➢ **进一步研究：**

不同 AI 技术应用对企业创新的影响：不同 AI 技术（自然语言处理、计算机视觉、机器学习）得到的关键词词频数加 1 取对数

作用机制检验：知识多样性（计算专利知识宽度）、组织惯例（年报文本分析获取打破组织惯例相关的高频词汇）、资源配置效率（计算过度投资程度）、吸收能力（企业产学研合作情况-是否拥有与高校和科研机构联合申请的产学研合作专利、研发强度-研发投入占营业收入的比重）

文章以2009—2022年A股上市公司为样本，运用机器学习方法生成AI词典，进行年报文本分析，测度AI技术应用水平，研究AI技术应用对企业创新的影响。研究发现，AI技术应用通过增加知识多样性、打破组织惯例、提高资源配置效率三个作用机制促进企业创新。



其他CSMAR关联子库：经营困境、上市公司联盟、财务报表、财务指标分析、上市公司基本信息、……

CSMAR与热门主题：（一）企业经济

相关研究： 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

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。

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

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

人工智能

作用机制

- 劳动赋能（高技能劳动力）
- 管理赋能（管理费用）

调节效应

企业数字责任

→ **企业社会绩效**

变量	劳动赋能		管理赋能	
	(1) Labor_Stru	(2) CSP	(3) Ma_Fee	(4) CSP
AI_Transf	0.013 3***	0.108 5***		
	(0.002 3)	(0.038 1)		
AI_Inno			−0.076 4***	0.068 7***
			(0.027 6)	(0.019 6)

变量	Corporate social Performance (CSP)			
	(1)	(2)	(3)	(4)
AI_Inno	0.319 7***		0.249 5***	
	(0.032 3)		(0.043 5)	
AI_Transf		0.182 2***		0.144 3***
		(0.021 4)		(0.009 0)
AI_Inno×CDR				
AI_Transf×CDR	0.320 0**		0.356 0***	
	(0.156 8)		(0.080 4)	

CSMAR与热门主题：（一）企业经济

相关研究： 气候政策不确定性、数字化转型与企业投资效率[J].上海财经大学学报，2024.

关键解释变量	调节变量	被解释变量
中国省级气候政策不确定性指数	年报中数字化转型词频	基于Richardson（2006）提出的理论模型测算
以《人民日报》等6家主流报纸作为数据主要来源	数字经济/企业数字化转型	经营困境

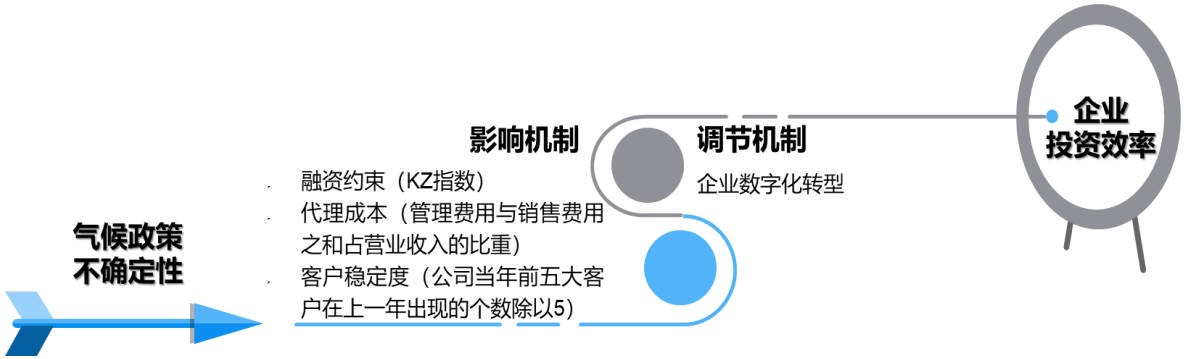
文章利用2005—2022年中国A股上市公司数据，实证检验气候政策不确定性对企业投资效率的影响及作用机制。研究表明：气候政策不确定性显著降低企业投资效率，而数字化转型能有效缓解气候政策不确定性对企业投资效率提升的抑制作用。机制识别显示,融资约束、代理成本和客户稳定度是气候政策不确定性影响企业投资效率的重要传导渠道。

$$Invest_{i,t} = \alpha_0 + \alpha_1 CCPU_{i,t-1} + \alpha_2 Control_{i,t} + Firm + Year + Province + Industry + \mu_{i,t}$$

变量	（1）	（2）	（3）
	Invest	*Overinvest*	*Underinvest*
CCPU	0.0035***（0.0010）	0.0057***（0.0021）	0.0019**（0.0009）

变量	（1）	（2）	（3）	（4）	（5）	（6）
	i=1	*i*=2	*i*=3	*i*=4	*i*=5	*i*=6
CCPU	0.0033***	0.0033***	0.0034***	0.0034***	0.0034***	0.0034***
	（0.0010）	（0.0010）	（0.0010）	（0.0010）	（0.0010）	（0.0010）
Digital$_i$×*CCPU*	−0.0134***	−0.0307*	−0.0421***	−0.0187**	−0.2253**	−0.0353***
	（0.0028）	（0.0180）	（0.0110）	（0.0090）	（0.1141）	（0.0068）

变量	融资约束机制		代理成本机制		客户稳定度机制	
	KZ	*Invest*	*AC*	*Invest*	*Customer*	*Invest*
CCPU	0.0395**		0.0027**		−0.0296**	
	（0.0199）		（0.0014）		（0.0145）	
$\widehat{KZ}$		0.0873***				
		（0.0244）				
$\widehat{AC}$				1.2736***		
				（0.3565）		
$\widehat{Customer}$						−0.1168***
						（0.0327）



其他CSMAR关联子库： 供应链研究、财务报表、财务指标分析、上市公司基本信息、……

CSMAR与热门主题：（一）企业经济

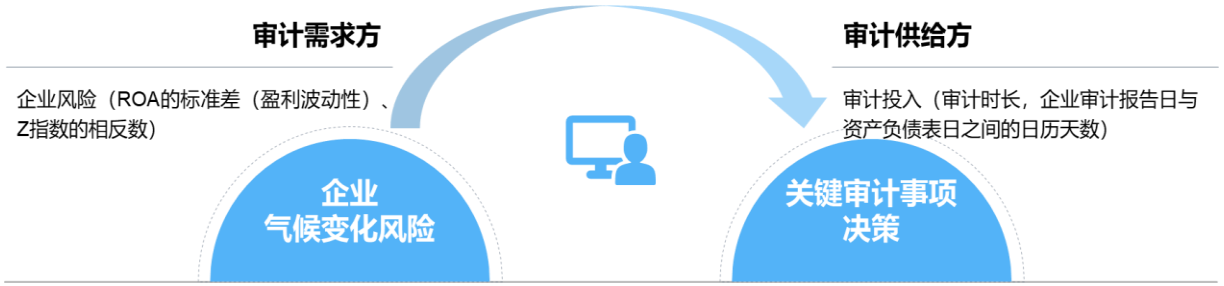
相关研究： 企业气候变化风险与关键审计事项决策[J].上海财经大学学报，2024.

关键解释变量	被解释变量
企业年报中与气候变化相关的文本词频，从严重风险、慢性风险和转型风险三个维度	与气候变化相关的关键审计事项强度 与风险匹配的气候变化业务关键审计事项强度
通过Python软件中的jieba分词对公司年报内容进行分词处理	财务报告审计意见

文章以2016—2022年中国A股上市公司为样本，探究企业气候变化风险对关键审计事项决策的影响。研究表明，年报审计师在气候变化程度较高企业中的关键审计事项决策更加谨慎，表现为更易于出具与气候变化相关的关键审计事项和与气候业务风险匹配的关键审计事项，并且这种关键审计事项的风险控制效用主要体现在转型气候风险方面。作用机制检验表明，气候变化会强化企业风险因素传导和增加审计投入力度，进而增进关键审计事项披露的谨慎性。

$$CKAM_{i,t}/RCKAM_{i,t} = \alpha_0 + \alpha_1 Climate_{i,t} + \alpha_m Controls_{i,t} + Year + Firm + \varepsilon_{i,t}$$

	整体气候变化风险		物理风险				转型风险	
			严重风险		慢性风险			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	CKAM	*RCKAM*	*CKAM*	*RCKAM*	*CKAM*	*RCKAM*	*CKAM*	*RCKAM*
Climate	0.129***	0.078**						
	(2.90)	(2.03)						
Climate_se			1.150	−0.297				
			(1.10)	(−0.33)				
Climate_ch					1.224	0.548		
					(1.50)	(0.96)		
Climate_tr							0.123***	0.078**
							(2.73)	(2.01)



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

CSMAR与热门主题：（一）企业经济

相关研究： 企业体育健康投资的创新效应—基于企业开展体育运动的经验证据[J].当代财经，2022.

关键解释变量	被解释变量
采用上市公司每年是否开展体育运动作为企业体育健康投资的度量指标	企业创新产出：专利申请和授权数量
借助Python爬取企业官方网站各年度开展体育运动的所有相关新闻资讯	上市公司研发创新/上市公司与关联公司专利/专利被引用

$$InnovationX_{i,t+1} = \beta_0 + \beta_1 Sports_{i,t} + \sum_k \beta_k Controls_{k,i,t} + \sum Industry + \sum Year + \varepsilon_{i,t}$$

	(1) *Apply*	(2) *Grant*
Sports	0.246***(3.02)	0.219***(2.85)

	健康机制检验			畏难情绪机制检验		
	(1) *Health*	(2) *Apply*	(3) *Grant*	(4) *Mood*	(5) *Apply*	(6) *Grant*
Sports	0.011**(2.06)	0.227***(4.36)	0.196***(4.09)	−0.005**(−1.98)	0.228***(4.37)	0.198***(4.11)
Health		0.508***(5.25)	0.607***(6.80)			
Mood					−1.121***(−5.40)	−1.084***(−5.66)

文章以沪深两市2007—2018年的上市公司为研究样本，检验企业体育健康投资对企业创新的影响。结果显示，企业体育健康投资与创新产出正相关。机制检验结果表明，企业体育健康投资的创新效应路径主要是增强员工的身心健康水平和改善员工畏难情绪。

企业体育健康投资

健康

员工健康水平（基于中国家庭追踪调查 （CFPS） 数据测度）

畏难情绪

员工畏难情绪（基于中国家庭追踪调查 （CFPS） 数据测度）

企业创新

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

CSMAR与热门主题：（二）金融

方向	主题	参考范文	关联CSMAR数据库
金融	股票流动性	数实产业技术融合能提升企业股票流动性吗？ 企业ESG与资本市场表现——来自股票流动性的证据	行为金融、市场指数、股票市场交易、财务指标分析、股票市场衍生指标、机构投资者、股票流动性、已实现指标、宏观经济、分析师预测、一致预测、资本市场监管、金融科技、事件研究、投资者情绪、社交媒体、普惠金融、绿色金融、碳中和、货币市场与政策工具、会计差错更正、智能研报基本信息、智能风险感知、智能财务诊断、智能舆情分析、智能股价预测、龙头股、新三板系列
	股价崩盘风险	基于深度学习的投资者情绪挖掘及其对股价崩盘风险的影响 环境立法与投资者利益保护——基于股票市场异常波动的视角	
	金融稳定	调预期、防风险：央行金融稳定沟通效果研究——基于文本分析的视角 数字基础设施建设与金融市场稳定	
	数字金融	产业数字金融何以促进“专精特新”企业新质生产力发展 数字金融发展、金融错配度与企业全要素生产率	
	绿色金融	走向绿色之路：财政环境支出与企业气候风险 绿色银行借贷关系存在绿色价值溢出效应吗？——基于企业债券融资成本的研究	
	货币政策	货币政策不确定性如何影响企业投资效率——来自中国上市企业的经验证据 结构性货币政策、预算软约束与企业杠杆率——基于DSGE模型的模拟分析 人口老龄化对中国货币政策资产价格渠道的影响分析	

CSMAR与热门主题：（二）金融

相关研究： 数实产业技术融合能提升企业股票流动性吗？[J]. 财经论丛, 2024.

关键解释变量	被解释变量
数字产业技术专利引用次数	资本市场非流动性指标
中国国家知识产权局的专利数据库	股票流动性

$$LIQ_{it} = \alpha_0 + \alpha_1 TC_{it} + \sum \alpha_m Control_{it} + \mu_t + \sigma_k + \varepsilon_{it}$$

变量	(1)*LIQ*	(2)*LIQ*
TC	0.0043***(0.0011)	0.0041***(0.0011)

变量	(1)*NPR*	(2)*LIQ*	(3)*TFP*	(4)*LIQ*
TC	0.0363**(0.0166)	0.0021*(0.0011)	0.0791***(0.0083)	0.0024***(0.0007)
NPR		0.0074***(0.0025)		
TFP				0.0112***(0.0008)

变量	(1)*OPM*	(2)*LIQ*
TC	0.0037**(0.0016)	0.0031***(0.0011)
OPM		0.0220***(0.0052)

变量	(1)*KV*	(2)*LIQ*	(3)*AF*	(4)*LIQ*
TC	-0.0034*(0.0018)	0.0037***(0.0007)	0.1090***(0.0157)	0.0030***(0.0010)
KV		-0.0857***(0.0033)		
AF				0.0094***(0.0005)

文章基于 2008—2022 年沪深 A 股上市企业数据，考察数实产业技术融合对股票流动性的影响。结果发现，数实产业技术融合能够显著提升企业股票流动性。机制研究发现，从创新效应视角来看，数实产业技术融合通过提高创新质量提升企业股票流动性；从治理效应视角来看，数实产业技术融合通过增强企业财务稳定性提升股票流动性；从信息效应视角来看，数实产业技术融合通过提高信息披露质量和分析师关注提升企业股票流动性。

数实产业技术融合 → 股票流动性

- **创新效应：** 创新质量（新产品销售收入、全要素生产率）
- **治理效应：** 财务稳定性（企业营业利润率）
- **信息效应：** 衡量信息披露质量（KV 指数）、分析师关注度（的分析师跟踪人数）

其他CSMAR关联子库： 市场竞争、上市公司人物特征、……

CSMAR与热门主题：（二）金融

相关研究： 环境立法与投资者利益保护——基于股票市场异常波动的视角[J].金融研究, 2024.

自然实验	被解释变量
2015年《环境保护法》实施和2005年至2016年14次地方环保条例颁布或修订	股价崩盘风险：负收益偏态系数(NCSKEW) 和上下波动比率(DUVOL)
	行为金融

文章以沪深两市上市公司为研究对象，以 2015 年《环境保护法》实施和 2005—2016年地方环保法规颁布或修订作为自然实验，从股票市场异常波动角度，考察环境立法对投资者利益保护的影响及其内在机制。研究发现，相对于非重污染企业，环境立法显著增加了重污染企业的行为约束，降低了其股价崩盘风险。从作用机制看，环境立法促使重污染企业增加环保投资，提升环境信息披露质量，且环保投资更多、环境信息披露质量更好的重污染企业股价崩盘风险下降更明显。

$$CrashRish_{i,t+1} = \beta_1 Treat_i \times Post_t + \mu_i + \tau_t + \gamma X_{i,t} + \varepsilon_{i,t}$$

$$CrashRish_{i,t+1} = \beta_1 Treat_i \times Post_{i,t} + \beta_2 Post_{i,t} + \mu_i + \tau_t + \gamma X_{i,t} + \varepsilon_{i,t}$$

变量	$NCSKEW_{t+1}$			
	中央法		地方法	
	(1)	(2)	(3)	(4)
$Treat_i \times Post_{(i,)t}$	−0.2834***	−0.2374***	−0.1114**	−0.0987*
	(0.0606)	(0.0683)	(0.0553)	(0.0568)

其他CSMAR关联子库： 股票市场交易、股票流动性、环境研究、分析师预测、财务报表、财务指标分析、……

降低股价崩盘风险

中介作用

- 企业环境信息披露的详细程度
- 环境保护有关的项目(如回收利用、污废治理、生态保护等) 当年变动额占总资产的比重

调节作用

- 审计质量（审计费用）和分析师关注度（发布公司 EPS预测的分析师数量）

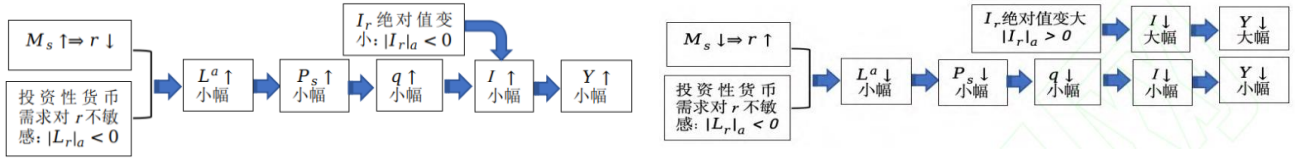
环境立法

CSMAR与热门主题：（二）金融

相关研究： 人口老龄化对中国货币政策资产价格渠道的影响分析[J].学术论坛，2024.

关键解释变量	货币政策代理变量	货币政策资产价格渠道变量
老年人口抚养比：各省65 岁以上人口占15至65岁人口之比	货币供应量：中国广义货币M2	财富变量：各省份商品房平均销售价格 托宾Q值 利率：中国银行间同业拆借利率
人口老龄化	宏观经济	宏观经济/银行体系/财务指标分析

文章聚焦人口老龄化与货币政策资产价格渠道之间的潜在关联，通过把人口老龄化因素纳入消费、投资和货币需求函数的分析发现，人口老龄化倾向于降低投资（投机）性货币需求对利率的敏感度，使实际投资的利率敏感度和交易性货币需求的收入敏感度在松紧环境下具有不对称性，因而弱化扩张性货币政策资产价格渠道传导的q效应和财富效应。基于2000—2022年我国31个省级面板数据的回归分析发现，现阶段人口老龄化弱化了我国货币政策资产价格渠道的传导效果。



人口老龄化下扩张性/紧缩性货币政策资产价格渠道传导的q效应

托宾q效应渠道

$$lnGDP_{it} = \alpha_0 + \alpha_1 q_{it} + \alpha_2 OLD_{it} + \alpha_3 lnM2 + \alpha_4 RATE_{it} + \mu_i + \mu_t + \varepsilon_{it}$$

$$lnGDP_{it} = \alpha_0 + \alpha_1 q_{it} + \alpha_2 OLD_{it} + \alpha_3 lnM2 + \alpha_4 RATE_{it} + \alpha_5 OLD_{it} \times lnM2 + \mu_i + \mu_t + \varepsilon_{it}$$

$$lnGDP_{it} = \alpha_0 + \alpha_1 q_{it} + \alpha_2 OLD_{it} + \alpha_3 lnM2 + \alpha_4 RATE_{it} + \alpha_5 OLD_{it} \times lnM2 \times q_{it} + \mu_i + \mu_t + \varepsilon_{it}$$

变量	模型(1) lnGDP	模型(2) lnGDP	模型(3) lnGDP
q	0.0113***(2.47)	0.0114**(2.47)	0.0130**(2.37)
lnM2	0.6156***(4.31)	0.6162***(4.30)	0.6169***(4.31)
OLD	-0.9116***(-3.12)	-1.0649***(-3.17)	-0.9146***(-3.13)
RATE	-0.2525*(-1.79)	-0.2506*(-1.75)	-0.2514*(-1.78)
OLD×lnM2		-0.0108**(-2.38)	
OLD×lnM2×q			-0.0001**(-2.47)

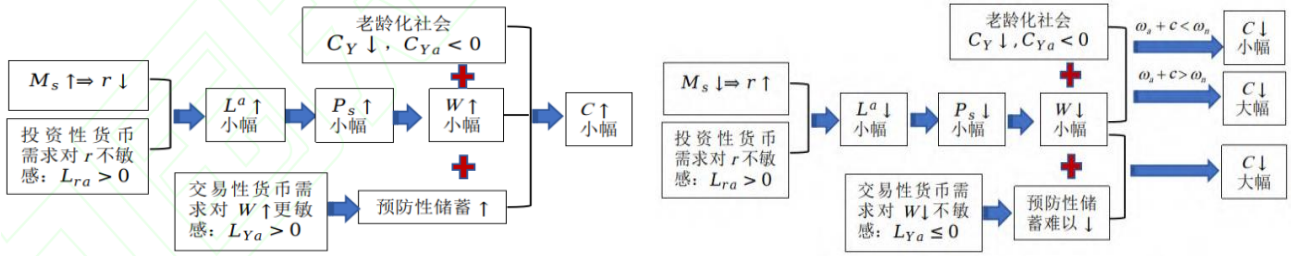
财富效应渠道

$$lnGDP_{it} = \alpha_0 + \alpha_1 lnPrice_{it} + \alpha_2 OLD_{it} + \alpha_3 lnM2 + \alpha_4 RATE_{it} + \mu_i + \mu_t + \varepsilon_{it}$$

$$lnGDP_{it} = \alpha_0 + \alpha_1 lnPrice_{it} + \alpha_2 OLD_{it} + \alpha_3 lnM2 + \alpha_4 RATE_{it} + \alpha_5 OLD_{it} \times lnM2 + \mu_i + \mu_t + \varepsilon_{it}$$

$$lnGDP_{it} = \alpha_0 + \alpha_1 lnPrice_{it} + +\alpha_2 OLD_{it} + \alpha_3 lnM2 + \alpha_4 RATE_{it} \quad \alpha_5 OLD_{it} \times lnM2 \times lnPrice_{it} + \mu_i + \mu_t + \varepsilon_{it}$$

变量	模型(4) lnGDP	模型(5) lnGDP	模型(6) lnGDP
lnPrice	0.1401***(3.47)	0.1447***(3.55)	0.1966***(4.18)
lnM2	0.4858***(3.32)	0.4779***(3.26)	0.4435***(3.02)
OLD	-0.8016**(-2.76)	0.3004**(2.16)	1.5488**(2.44)
RATE	-0.2905**(-2.07)	-0.3048**(-2.14)	-0.3346**(-2.37)
OLD ×lnM2		-0.0775**(-2.30)	
OLD ×lnM2×lnPrice			-0.0188**(-2.27)



人口老龄化下扩张性/紧缩性货币政策资产价格渠道传导的财富效应

CSMAR与热门主题：（三）宏观经济

方向	主题	参考范文	关联CSMAR数据库
宏观经济	经济增长	赋能中的分化：数字技术扩散与中国南北经济增长“失衡” 气候风险、经济增长与城市内涵式发展——基于暴雨冲击的经验证据 机器人应用、职业伤害减少与生命健康收益 人口老龄化趋势下卫生资源配置结构性失衡与社会福利损失	宏观经济、区域经济、国际宏观、经济内循环、经济地理、人口老龄化、供应链研究、数字经济、碳中和、环境研究、政府审计、一带一路、企业创新、上市公司研发创新、专利、新质生产力、人工智能产业链
	全球价值链	供应链信息披露如何影响中国企业参与全球价值链分工？ 中国企业“双循环”下本土供应链与全球价值链协同创新效应研究	
	环境规制	碳排放权交易试点政策对经济高质量发展的多维影响研究 流通业智能化对城市绿色经济效率的影响	
	”一带一路“	“一带一路”背景下数字化转型对制造业高质量发展影响研究——基于产业结构升级与数字贸易的视角 中国“一带一路”倡议投资对降低共建国家气候风险的影响	
	创新	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 人口年龄结构变动会影响创新吗？	
	产业政策	分类施策：产业政策影响工业智能化的经验证据 国家大基金对中国人工智能芯片产业创新绩效的影响	

CSMAR与热门主题：（三）宏观经济

相关研究： 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/人工智能专利	《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。

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)
	协作网络密度(%)
数字技术创新	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(%)
	人工智能上市企业研发投入(万元)
数智企业发展	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)
	平均每家人工智能企业营业收入(万元)
数智环境发展	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)
	人工智能产业领域行业标准数量(个)
数智投融资	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)
	人工智能产业投融资交易数量(笔)
国际影响力	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			
					Q=0.2	Q=0.4	Q=0.6	Q=0.8
	(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 5)	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	
	$FSIE_{it}$	$ITSE_{w(y)}$	$ITSE_{w(x)}$
	(1)	(2)	(3)
AIE	6.737 2** (3.387 5)	1.767 7*** (0.074 4)	0.964 8*** (0.188 9)

人工智能嵌入程度

直接作用
- 网络连接
- 渗透赋能
- 价值创造

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）

创新链布局重构效能

其他CSMAR关联子库： 宏观经济、数字经济、经济内循环、……

CSMAR与热门主题：（三）宏观经济

相关研究： 中国“一带一路”倡议投资对降低共建国家气候风险的影响[J].中国人口·资源与环境, 2024.

关键解释变量	被解释变量
中国直接投资	气候变化脆弱性风险
一带一路	圣母大学构建的ND-GAIN脆弱性指数

文章基于2004—2020年65个共建国家的面板数据，评估中国对外直接投资对东道国应对气候变化风险能力的影响。研究发现：中国对“一带一路”倡议地区的直接投资显著减轻了共建国家的气候变化脆弱性风险；在全球气候变化危机加剧的背景下，提高了共建国家气候风险应对能力。

$$VULNER_{i,t} = \beta_0 + \beta_1 CFDI_{i,t-1} + \beta_2 LGDP_{i,t-1} + \beta_3 GPD2_{i,t-1} + \beta_4 Popdensity_{i,t-1} + \beta_5 Urbanrate_{i,t-1} + \beta_6 Forestrate_{i,t-1} + \beta_7 GovernE_{i,t-1} + \beta_8 Agricul_{i,t-1} + \beta_9 Disasters_{i,t-1} + Nation_i + Year_t + \varepsilon_{it}$$

变量	VULNER			DFVULNER
	控制固定效应	引入控制变量	回归结果	
CFDI	-0.139**	-0.115**	-0.094**	-0.140***
	(0.063)	(0.047)	(0.041)	(0.044)

$$Mediators^j_{i,t} = \beta_0 + \beta_1 CFDI_{i,t-1} + \beta_2 LGDP_{i,t-1} + \beta_3 GPD2_{i,t-1} + \beta_4 Popdensity_{i,t-1} + \beta_5 Urbanrate_{i,t-1} + \beta_6 Forestrate_{i,t-1} + \beta_7 GovernE_{i,t-1} + \beta_8 Agricul_{i,t-1} + \beta_9 Disasters_{i,t-1} + Nation_i + Year_t + \varepsilon_{it}$$

$$VULNERABILITY_{i,t} = \beta_0 + \beta_1 CFDI_{i,t-1} + \beta_2 LGDP_{i,t-1} + \beta_3 GPD2_{i,t-1} + \beta_4 Popdensity_{i,t-1} + \beta_5 Urbanrate_{i,t-1} + \beta_6 Forestrate_{i,t-1} + \beta_7 GovernE_{i,t-1} + \beta_8 Agricul_{i,t-1} + \beta_9 Disasters_{i,t-1} + \gamma^j_1 Mediators_{it} + Nation_i + Year_t + \varepsilon_{it}$$

变量	收入增长效应		政治稳定效应	
	Income	VULNER	PoliticS	VULNER
CFDI	0.012***	-0.077**	0.040***	-0.084**
	(0.005)	(0.036)	(0.011)	(0.041)
Income		-0.845***		
		(0.133)		
PoliticS				-0.269***
				(0.102)

其他CSMAR关联子库： 宏观经济、世界经济、……

CSMAR与热门主题：（三）宏观经济

相关研究： 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.

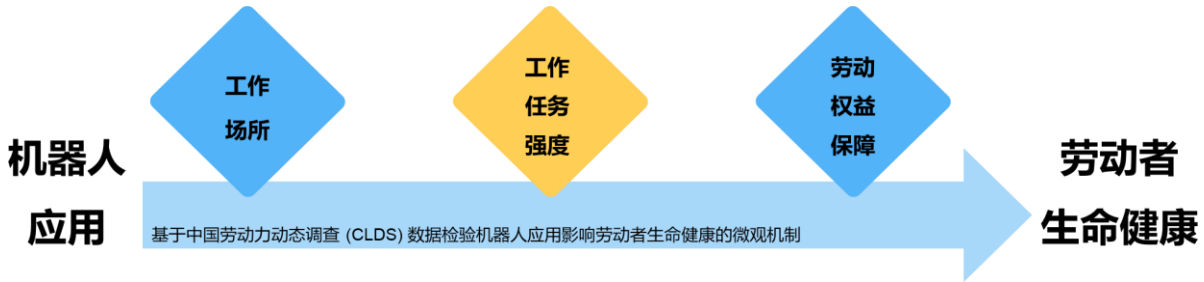
关键解释变量	被解释变量
机器人渗透度	劳动者的生命健康状况：城市年度安全生产事故事件数和死亡人数
人工智能	事件研究

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

	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***	-0.555***	-0.396**	-2.329**
	(0.038)	(0.204)	(0.193)	(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**	-0.019***	-0.035***	-0.074**	0.009**	0.021***
	(0.003)	(0.006)	(0.012)	(0.031)	(0.004)	(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与前沿主题

基于2023-2024年科学研究呈现出的趋势特点，权威研究遴选了近期经济管理学领域的重点**前沿主题**如下：

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。

中国科学院科技战略咨询研究院 - 2024

1. 政策变化评估及特定事件影响研究
2. COVID-19 对不同人群的心理健康影响
3. 气候变化与投资决策
4. 人类活动对气候变化、生物多样性等影响
5. 社会结构因素与健康差异关系研究
6. 脑电图（EEG）数据分析在神经科学、心理学、认知科学中的应用
7. 深度学习算法与金融预测、资产定价
8. 超加工食品消费问题
9. 零工经济的算法、人力资源及平台管理
10. 人工智能伦理准则

《经济学动态》经济学研究热点课题组 - 2023

1. 面向高水平科技自立自强的国家创新体系研究
2. 加快发展数字经济研究
3. 金融供给侧结构性改革与现代金融监管
4. “一带一路”十载理论与实践
5. 建设农业强国
6. 提升产业链供应链现代化水平
7. 区域协调发展研究
8. 以扩大内需增强中国经济发展内生动力
9. 建构中国自主的经济学知识体系

参考资料：
[1] 2024研究前沿. 中国科学院科技战略咨询研究院，中国科学院文献情报中心，科睿唯安,2024.
[2] 2023年中国经济学研究热点分析[J]. 经济学动态,2024.

CSMAR与前沿主题：（一）深度学习算法与金融预测、资产定价

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.

已有研究：已有研究当中情绪测度主要分为基于市场交易指标构建和基于词典法构建两类。市场交易指标法数据可得性较强，但市场交易指标还受其他因素影响，该方法难以准确反应情绪变动。文本字典法是简单的线性方法，难以刻画非线性与非对称的市场情绪动态变化，并受字典设定质量影响较大，还忽略了上下文等其他维度的文本信息。

模型改进：文章引入了以 BERT（Bidirectional Encoder Represen-tations from Transformers）为代表的深度学习和大语言模型，将市场回报作为情绪标签变量引入到大模型训练中，创新融合结构化市场数据和非结构化文本数据，训练了一个更具有中国市场特色的中文金融大语言模型，构建能反映我国金融市场非线性与非对称特征的大模型情绪测度。

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.

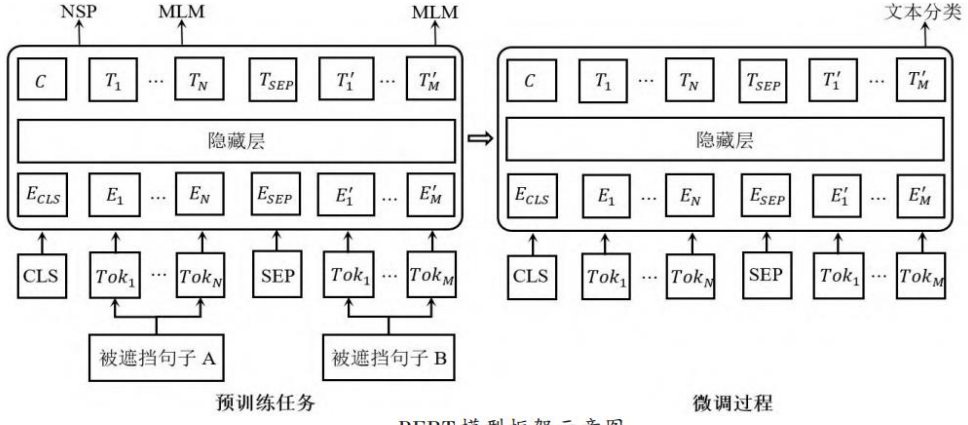
已有研究：大多数的研究都是基于传统方法，在分析非结构化的文本数据时优势不足，同时在对文本情感进行处理时，基于机器学习的传统方法仍是主流。此外，基于卷积神经网络或循环神经网络的情感分析模型均暴露出了一些问题：卷积神经网络无法获取文本的远距离依赖信息不适用于处理自然语言，而循环神经网络无法同时获取双向语义信息且并行能力较差，参数规模较小无法较好拟合大规模数据。

模型改进：采用轻量级大模型ALBERT，对文本信息具有更强的处理能力。ALBERT模型是在BERT模型的基础上进行改进的，它设计了参数减少的方法，用来降低内存消耗， 同时加快BERT的训练速度。

CSMAR与前沿主题：（一）深度学习算法与金融预测、资产定价

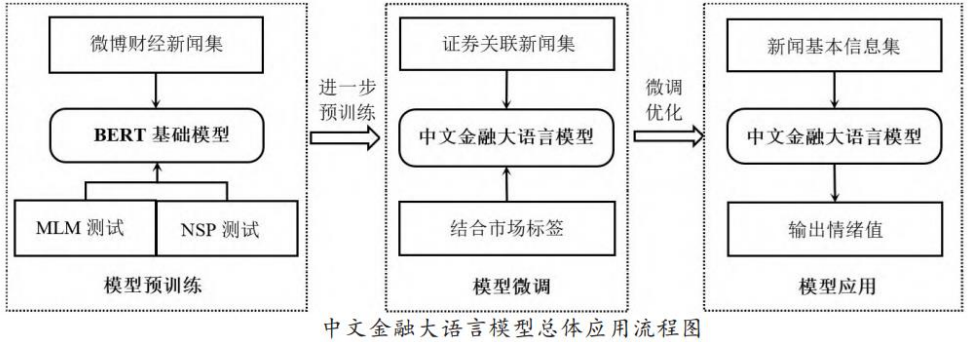
相关研究1：大语言模型、文本情绪与金融市场[J]. 管理世界, 2024.

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

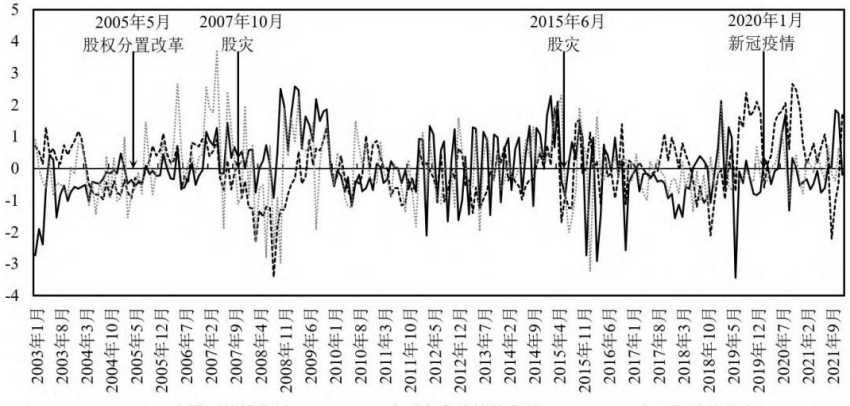
文章使用微博财经新闻数据集作为中文金融大语言模型针对金融领域进行特质化训练的语料。微博财经新闻数据集经过清洗后共 30 余万条有效数据。由于微博没有统一的财经新闻入口，本文首先人工列举出了“21 世纪经济报道”、“东方财富网”、“央视财经”等 21 家具有微博账户的权威新闻媒体列表，然后爬取这些账户的微博内容作为微博财经新闻数据集。尽管该数据来自微博，但由于爬取账户均为权威新闻媒体，获取的财经新闻仍然是标准的正式新闻。

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

$$S_t^k = \frac{\sum_{i=1}^{N_t} A_i^k}{N_t}, k = llm \text{ or } dic$$

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



情绪变量与上证指数收益率时序变化

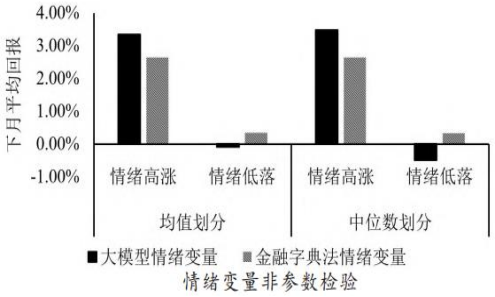
CSMAR与前沿主题：（一）深度学习算法与金融预测、资产定价

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

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。



情绪变量非参数检验

样本内预测能力检验

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

$$R_c^2 = 1 - \frac{\sum_{t=1}^{T} I_t^c \left(\hat{\varepsilon}_{i,t}\right)^2}{\sum_{t=1}^{T} I_t^c \left(R_t^m - R_t^{\bar{m}}\right)^2}, c = up, down$$

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_{llm}	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^2_{up}(%)	5.88	6.22	3.55	8.25
R^2_{down}(%)	17.19	2.51	9.99	2.18

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

样本外预测能力检验

$$\hat{r}_t = \hat{\alpha}_{i,t-1} + \hat{\beta}_{i,t-1} SENTIMENT_{i,t-1}, i = 1,2$$

$$R_{i,OS}^2 = 1 - \frac{\sum_{T-P+1}^{T} \left(r_t^m - \hat{r}_t^m\right)^2}{\sum_{T-P+1}^{T} \left(r_t^m - \bar{r}_t^m\right)^2}, i = 1,2$$

样本外预测结果

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

资产配置检验

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：γ=3				
	区间一			
S_{llm}	0.8834	0.2919	0.5614	0.2517
S_{dic}	0.6410	0.2633	0.4178	0.2323
	区间二			
S_{llm}	0.7973	0.2634	0.4644	0.2190
S_{dic}	0.5582	0.2309	0.3375	0.1983
	区间三			
S_{llm}	0.7153	0.2069	0.4341	0.1586
S_{dic}	0.0390	0.1003	−0.1761	0.0652
Panel B：γ=5				
	区间一			
S_{llm}	0.5770	0.3050	0.3378	0.2686
S_{dic}	0.3578	0.2557	0.1485	0.2226
	区间二			
S_{llm}	0.4803	0.2702	0.2421	0.2301
S_{dic}	0.2852	0.2250	0.0775	0.1913
	区间三			
S_{llm}	0.3931	0.2017	0.1826	0.1573
S_{dic}	−0.1475	0.1004	−0.3848	0.0551
Panel C：γ=7				
	区间一			
S_{llm}	0.3092	0.3005	0.1281	0.2660
S_{dic}	0.1273	0.2436	−0.0532	0.2129
	区间二			
S_{llm}	0.2128	0.2606	0.0382	0.2242
S_{dic}	0.0654	0.2174	−0.1155	0.1865
	区间三			
S_{llm}	0.2020	0.1897	0.0352	0.1457
S_{dic}	−0.2643	0.0981	−0.4766	0.0524

CSMAR与前沿主题：（一）深度学习算法与金融预测、资产定价

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

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

情绪变量与经济变量预测

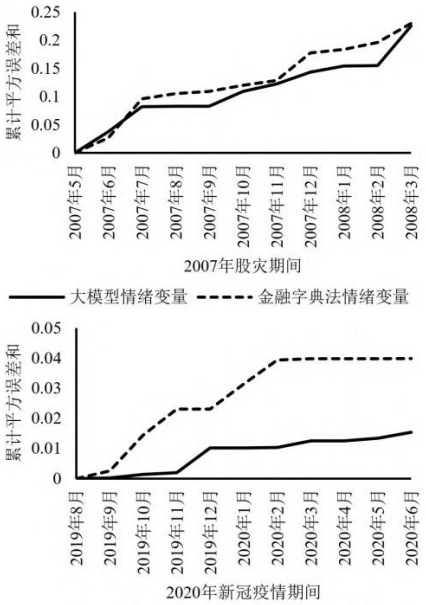
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Panel A：经济上行期大模型情绪变量									
变量	*PMI*	*CPI*	*M2*	*SF*	F_{out}	*inv*	*house*	*PE*	*Turnover*	*LVOL*
S_{llm}	0.15*	0.02	0.001	0.05	0.04	−0.01	−0.004	0.05	0.16**	0.04
	(1.85)	(0.87)	(0.95)	(0.71)	(0.67)	(−0.99)	(−0.50)	(1.11)	(2.54)	(0.54)
Econ	0.68***	0.94***	1.01***	0.98***	0.74***	1.10***	1.11***	0.93***	0.68***	0.69***
	(9.47)	(34.07)	(884.61)	(12.42)	(11.98)	(106.37)	(113.79)	(19.84)	(10.98)	(9.77)
Constant	−0.05	0.03	0.01***	−0.08	0.01	0.16***	0.17***	−0.05	−0.01	0.03
	(−0.73)	(1.31)	(10.23)	(−1.03)	(0.24)	(17.53)	(19.45)	(−1.20)	(−0.17)	(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_{out}	*inv*	*house*	*PE*	*Turnover*	*LVOL*
S_{llm}	0.06	−0.09**	0.01***	0.09*	0.27**	0.14*	0.13*	0.13***	0.27***	−0.01
	(0.71)	(−2.16)	(3.35)	(1.70)	(2.58)	(1.81)	(1.81)	(3.66)	(3.70)	(−0.10)
Econ	0.52***	0.88***	1.01***	0.99***	0.64***	0.71***	0.78***	0.93***	0.64***	0.61***
	(5.69)	(20.27)	(430.32)	(17.52)	(5.32)	(8.80)	(10.30)	(26.31)	(8.82)	(7.28)
Constant	0.04	−0.06	0.02***	0.09	0.05	0.03	0.04	0.05	0.03	−0.05
	(0.43)	(−1.39)	(10.55)	(1.64)	(0.51)	(0.37)	(0.52)	(1.30)	(0.42)	(−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_{out}	*inv*	*house*	*PE*	*Turnover*	*LVOL*
S_{dic}	0.01	0.004	0.0004	−0.001	0.08	−0.01	0.004	0.04	0.13**	−0.17**
	(0.15)	(0.12)	(0.30)	(−0.01)	(1.30)	(−1.10)	(0.40)	(0.96)	(2.23)	(−2.15)
Econ	0.72***	0.95***	1.01***	0.97***	0.72***	1.10***	1.11***	0.96***	0.72***	0.62***
	(10.14)	(29.03)	(852.78)	(12.29)	(11.48)	(104.50)	(109.13)	(22.10)	(12.15)	(7.88)
Constant	−0.03	0.03	0.01***	−0.08	0.01	0.16***	0.17***	−0.04	−0.01	0.03
	(−0.44)	(1.31)	(10.20)	(−1.05)	(0.24)	(17.51)	(19.44)	(−1.03)	(−0.19)	(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_{out}	*inv*	*house*	*PE*	*Turnover*	*LVOL*
S_{dic}	−0.03	0.04	0.002	0.02	0.07	−0.04	−0.05	0.09**	0.09	0.03
	(−0.38)	(0.89)	(1.03)	(0.29)	(0.77)	(−0.56)	(−0.69)	(2.33)	(1.13)	(0.34)
Econ	0.52***	0.90***	1.00***	0.96***	0.49***	0.68***	0.76***	0.92***	0.67***	0.61***
	(5.61)	(21.45)	(415.34)	(15.80)	(4.51)	(8.50)	(10.00)	(24.89)	(8.68)	(7.22)
Constant	0.04	−0.06	0.02***	0.09	0.04	0.03	0.04	0.05	0.03	−0.05
	(0.42)	(−1.37)	(10.04)	(1.61)	(0.45)	(0.36)	(0.50)	(1.37)	(0.42)	(−0.61)
R^2(%)	26.44	83.89	99.95	90.58	19.50	45.35	53.51	87.94	45.67	36.94

极端事件预测分析

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

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_{llm}	0.014**		0.012*	
	(2.10)		(1.91)	
S_{dic}		0.259*		0.227
		(1.85)		(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_{llm}	0.005*		0.005*	
	(1.87)		(1.89)	
S_{dic}		0.057		0.069
		(1.43)		(1.57)
Constant	Yes	Yes	Yes	Yes
Control	No	No	Yes	Yes
R^2(%)	1.74	1.03	1.83	1.30

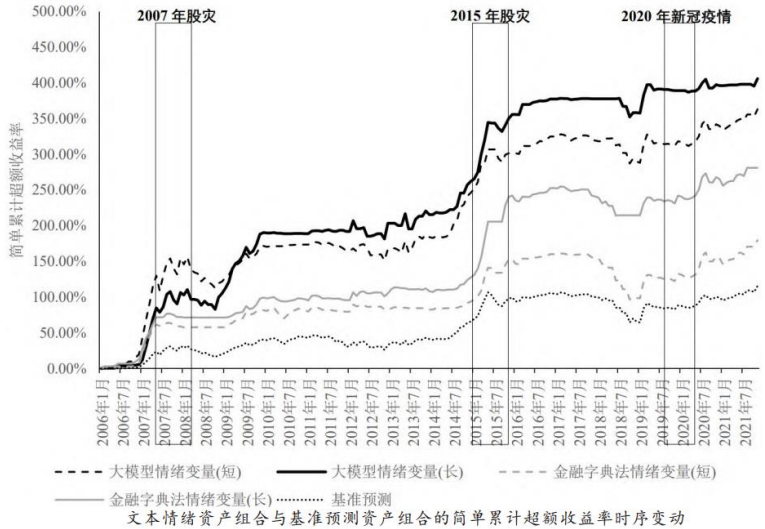


极端事件期间累计平方误差和变动趋势

文本特征预测分析

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_{llm}	0.214***		0.200***	
	(3.29)		(3.07)	
S_{dic}		0.130**		0.136**
		(1.96)		(2.05)
Constant	−0.004	−0.003	−0.004	−0.004
	(−0.05)	(−0.05)	(−0.07)	(−0.06)
R^2(%)	4.59	1.68	4.03	1.83



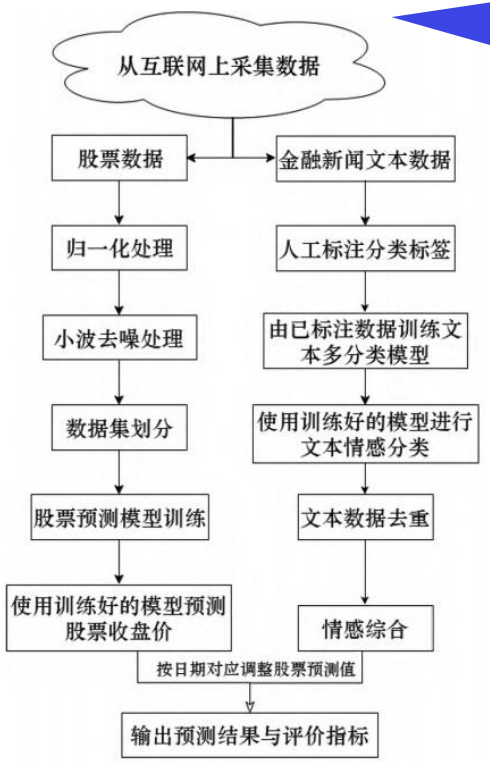
文本情绪资产组合与基准预测资产组合的简单累计超额收益率时序变动

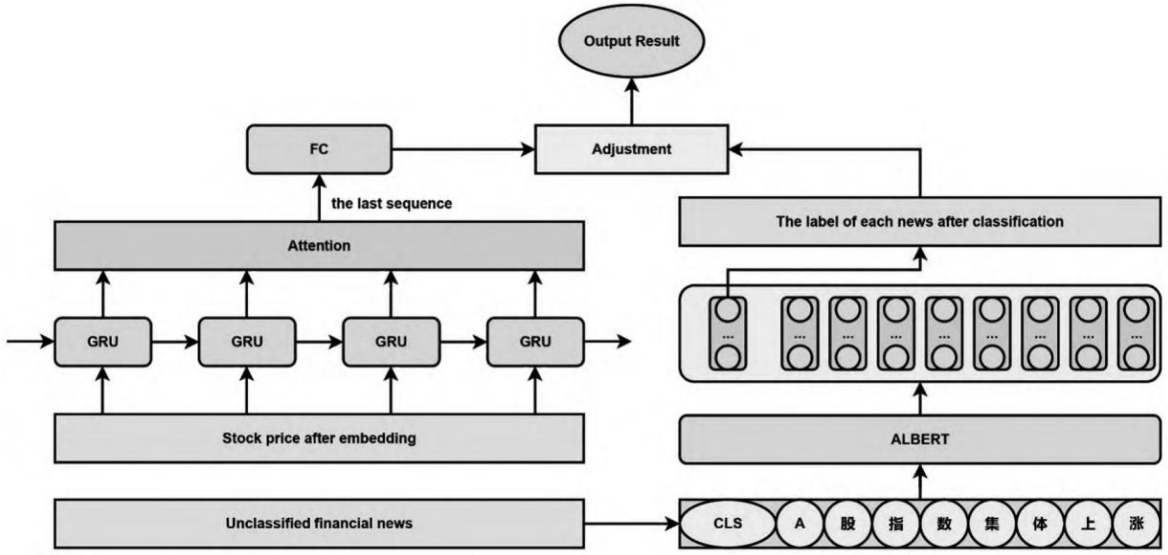
CSMAR与前沿主题：（一）深度学习算法与金融预测、资产定价

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、东方财经、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与前沿主题：（一）深度学习算法与金融预测、资产定价

会计	**大型语言模型对会计未来发展的影响[J].财务研究, 2023.** 文章简要总结了人工智能的演进和大型语言模型的技术框架，研究其对会计的影响。本文认为：大型语言模型的强大功能源自其巨量数据和大量参数，但具体实现原理目前尚不清楚，通用人工智能尚任重道远。通用大型语言模型本身难以对会计产生直接影响，但人工智能发展将使会计环境和具体对象发生变化，会计工作内容、会计人员能力框架和会计服务市场将被重塑。
审计	**生成式人工智能在金融审计中的应用——以Kimi为例[J].审计研究, 2025.** 文章对国产生成式人工智能大模型Kimi进行测试，探讨审计应用场景，评估Kimi在审前调查、审计实施和报告编制等审计工作环节的辅助作用。结果显示，Kimi在法规梳理、内容检索、文件比对、财务报告分析等方面能显著提升工作效率，尤其在处理大量文本信息和数据时表现出优势。但Kimi在文件比对的完整性和准确性、文本内容生成深度等方面存在不足，可能会影响审计结论的准确性。且生成式人工智能的应用也带来了隐私泄露等风险。
统计	**大模型统计学:统计学与大模型前沿问题研究[J].统计研究, 2024.** 大模型统计已经开始在社会学研究中发挥作用，一是进行社交网络分析，通过分析社交媒体上的大规模社会网络数据，揭示人们之间的互动模式、社会影响力的分布以及信息的传播路径。二是文化趋势分析，通过分析大量的文本数据，如新闻报道、文学作品、在线评论等，大模型可以捕捉到文化趋势和社会观念的变迁。三是人口统计学分析，通过处理和分析人口普查数据、健康记录和迁移记录等，为人口统计学研究提供支持。四是社会政策评估，通过模拟政策变化对社会经济指标的影响，帮助政策制定者预测政策的潜在结果，从而做出更加明智的决策。
	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。

CSMAR与前沿主题：（二）面向高水平科技自立自强的国家创新体系研究

创新测度			
类型			参考范文
创新数量	专利申请数量、发明专利申请数量 、非发明专利申请数量		公共性发展金融与企业创新[J].经济学(季刊), 2023.
创新质量	专利知识宽度	基于分类号计算赫芬达尔指数	专利质量对企业出口竞争力的影响机制：基于知识宽度视角的探究[J].世界经济研究, 2021.
	专利被引用	专利平均被引次数、专利首次被引所需时间平均间隔、国内被引、国外被引	独占还是共享? 研发国际化与企业创新价值获取——来自上市企业专利引用数据的证据[J].国际贸易问题, 2022.
	专利CD指数	CD指数（Consolidating/Disruptive index）是衡量专利“颠覆性”的指标，范围从-1（颠覆性最小）到1（最具颠覆性）。	创新竞争与企业高质量创新模式选择[J].经济研究, 2023.
	专利价值	战略价值（独立权利要求数量、无效情况、异议情况、许可频次等）、市场价值（市场规模、剩余有效期、布局国家等）、经济价值（同类技术近 3 年专利增长率、许可或转移记录等）、技术价值（引用情况、发明人数量、技术独立性指数等）和法律价值（专利法律状态、专利法律事件、权力要求数量等）	新型举国体制下重大工程创新生态系统的资源配置模式[J].管理世界, 2024.
创新信息披露	从公司年报中获取有关研发创新的关键词频构建创新信息披露水平。		“防微杜渐”还是“推波助澜”？散户“在线发声”与企业策略性创新信息披露[J].南开管理评论, 2024.
创新成本	研发强度	研发投入占营业收入的比重	人工智能技术应用如何影响企业创新[J].中国工业经济, 2024.
	研发成本粘性	$Recost_{jit} = log\left(\frac{\Delta cost}{\Delta sale}\right)_{jia} - log\left(\frac{\Delta cost}{\Delta sale}\right)_{jib}$ 研发成本粘性越大，代表企业对研发投入和管理决策的动态调整能力越低。	人工智能对NEV产业新质生产力的影响研究[J].科研管理, 2025.
创新效率	$Eff_{jit} = \frac{ln(1+Patent_{jit})}{ln(1+Research_{jit})}$ 其中 $Patent_{jit}$ 为企业年度专利授权数，$Research_{jit}$ 为企业年度研发投入。		人工智能对NEV产业新质生产力的影响研究[J].科研管理, 2025.

CSMAR与前沿主题：（二）面向高水平科技自立自强的国家创新体系研究

创新测度		
类型		参考范文
绿色创新	绿色创新数量（绿色专利申请数）、绿色创新质量（专利被引用数量从授权年开始以后作移动平均）、绿色创新效率（创新质量中的所有专利的年平均被引用次数之和除以企业近三年研发投入均值在资产中的占比后取自然对数）	数字化转型与绿色创新：基于信息的双重效应识别[J].改革, 2023.
数字创新	依据数字技术关键词对专利申请文件的内容进行文本分析，计算得到上市公司各年度内的数字专利申请数量。	数字技术创新与中国企业高质量发展——来自企业数字专利的证据[J].经济研究, 2023.
人工智能创新	人工智能专利。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上市公司的人工智能专利。	人工智能专利网络对企业智能化发展的影响[J].科研管理, 2024.

CSMAR关联子库

上市公司研发创新、上市公司与关联公司专利、专利被引用、企业数字化转型、企业创新、绿色专利、专利、数字经济、经济内循环、董事长致辞、市场分析报告文本、大为专利检索平台、……

综合创新

CSMAR与前沿主题：（二）面向高水平科技自立自强的国家创新体系研究

相关研究1：客户战略联盟如何激发企业创新？——基于文本分析的经验证据[J].南开管理评论，2023.

文章以2003—2021年沪深A股上市公司为样本实证检验了客户战略联盟与企业创新之间的关系。研究结果表明，客户战略联盟可以显著促进企业创新，包括创新的数量和质量。经济机制分析表明，客户战略联盟通过提高专用性资产的经济价值来激励管理层增加创新投资、通过客户信息溢出为创新提供市场和技术信息、通过缓解融资约束为创新提供资金等渠道促进企业创新。

H 1：企业与客户战略联盟关系的建立促进了企业创新。

H 2：企业与客户战略联盟关系的建立，通过管理层激励机制，促进了企业创新。

H 3：企业与客户战略联盟关系的建立，通过信息溢出机制，促进了企业创新。

H 4：企业与客户战略联盟关系的建立，通过融资约束机制，促进了企业创新。

H 5：客户战略联盟对企业创新的促进作用在高科技行业更为显著。

H 6：客户战略联盟对创新的促进作用在客户地理距离较远的企业中更为显著。

H 7：客户战略联盟对创新的促进作用在客户集中度较高的企业中更为明显。

$$Innovation_{i,t+1} = \alpha + \beta CSA_{i,t} + \gamma Controls_{i,t} + Firm_i + Year_t + \varepsilon_{i,t}$$

变量	构建方法
Panel A：企业创新变量	
LnApply	专利申请数量加 1 取对数
LnApplyI	发明专利申请数量加 1 取对数
LnApplyUD	非发明专利（实用新型、外观设计专利）申请数量加 1 取对数
LnGrant	专利授权数量加 1 取对数
LnGrantI	发明专利授权数量加 1 取对数
LnGrantUD	非发明专利（实用新型、外观设计专利）授权数量加 1 取对数
LnCite	发明专利未来三年引用数量加 1 取对数
Panel B：客户战略联盟变量	
CSA	客户战略联盟相关语句在年报文本中数量

供应链联盟情况表

证券代码、统计截止日期、业务关系、排名、业务往来公司名称、是否上市公司、公司证券代码、是否与上市公司同行业、是否战略合作联盟伙伴、业务额等

关联CSMAR子库：上市公司联盟 / 财务报表 / 财务报表附注 / 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 / 专利被引用 / 经营困境 / 供应链研究 / …

客户战略联盟对企业创新的影响

变量	(1)	(2)	(3)	(4)	(5)	(6)
	$LnApply_{t+1}$	$LnApply_{t+1}$	$LnApplyI_{t+1}$	$LnApplyI_{t+1}$	$LnApplyUD_{t+1}$	$LnApplyUD_{t+1}$
CSA	0.072***	0.047***	0.081***	0.057***	0.066***	0.045***
	(4.19)	(2.79)	(4.99)	(3.58)	(3.87)	(2.66)

客户战略联盟对创新投入价值和创新投入的影响

变量	(1)	(2)
	TobinQ	*RD*
RD	1.521**	
	(2.09)	
RD×CSA	1.909***	
	(3.16)	
CSA	-0.096***	0.001**
	(-4.07)	(2.52)

客户战略联盟的信息溢出效应

变量	(1)	(2)	(3)
	$LnApply_{t+1}$	$LnApplyI_{t+1}$	$LnApplyUD_{t+1}$
CusApplyI	0.046***	0.045**	-0.013
	(2.99)	(2.30)	(-0.52)
CusApplyI×CSA	0.019***	0.017**	0.045***
	(2.80)	(2.34)	(6.23)
CSA	-0.033	-0.078	-0.016
	(-0.61)	(-1.58)	(-0.30)

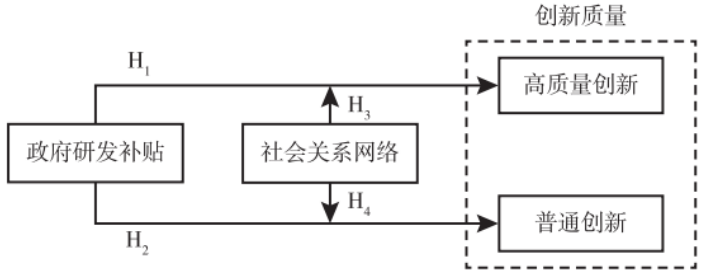
客户战略联盟对企业经营风险和融资约束的影响

变量	(1)	(2)	(3)
	ROASD	*KZ*	*SA*
CSA	-0.017**	-0.033**	-0.004***
	(-2.28)	(-2.04)	(-3.19)

CSMAR与前沿主题：（二）面向高水平科技自立自强的国家创新体系研究

相关研究2：政府研发补贴、社会关系网络与企业创新质量[J].经济管理，2024.

文章基于2013-2020年高端装备制造业上市公司数据，在国家知识产权局界定的高价值发明专利标准基础上，将企业创新区分为高质量创新与普通创新，考察了政府研发补贴对二者的影响以及社会关系网络的调节效应。研究发现：政府研发补贴有助于促进企业高质量创新，也会促进普通创新；社会关系网络负向调节政府研发补贴与高质量创新、普通创新之间的关系。



高价值发明专利数据来源于**中国知网专利检索与分析数据库**。国家知识产权局提出满足以下条件之一即可归为高价值发明专利：①战略性新兴产业；②在海外有同族专利权；③维持年限超过10年；④实现较高质押融资金额的发明专利；⑤获得国家科学技术奖或中国专利奖。其中，条件①②⑤侧重于专利的技术价值测度，条件③④侧重于专利的经济价值测度。本研究将具有较高技术价值或经济价值的创新视作高质量创新，与国家知识产权局提出的高价值发明专利的理念相致。

H1：政府研发补贴会促进企业高质量创新

H2：政府研发补贴会促进企业普通创新

$$HI_{i,t+1} = \beta_0 + \beta_1 Sub_{it} + \sum \beta_k Controls_{it} + \mu_{it} + v_{it} + \varepsilon_{it}$$

$$OI_{i,t+1} = \beta_0 + \beta_1 Sub_{it} + \sum \beta_k Controls_{it} + \mu_{it} + v_{it} + \varepsilon_{it}$$

变量类型	变量名称	变量符号	变量定义
被解释变量	综合创新	*Inn*	ln(发明专利数量+1)
	高质量创新	*HI*	ln(高价值发明专利数量+1)
	普通创新	*OI*	ln(发明专利数量-高价值发明专利数量+1)
解释变量	政府研发补贴	*Sub*	政府研发补贴总额
调节变量	社会关系网络	*Net*	标准化特征向量中心度
控制变量	企业规模	*Size*	ln(总资产)
	总资产收益率	*Roa*	净利润/总资产
	股权集中度	*Share*	企业第一大股东持股比率
	企业年龄	*Age*	ln(公司自成立年份起的年数)
	高管薪酬	*EC*	ln(董事、监事和高管前3名年薪总额)
	资产负债率	*Lev*	总负债/总资产
	现金净流量	*CF*	经营活动产生的现金净流量/总资产

基准回归结果

变量	(1)	(2)	(3)	(4)	(5)	(6)
	Inn	*Inn*	*HI*	*HI*	*OI*	*OI*
Sub	0.011**	0.011*	0.018***	0.020**	0.015***	0.013***
	(2.395)	(1.897)	(2.710)	(2.251)	(3.153)	(2.789)

调节效应回归结果

变量	(1)	(2)
	HI	*OI*
Sub	0.036***	0.025***
	(2.768)	(3.197)
Net	0.089	0.051
	(0.200)	(0.145)
Sub×Net	-0.098*	-0.073**
	(-1.779)	(-2.446)

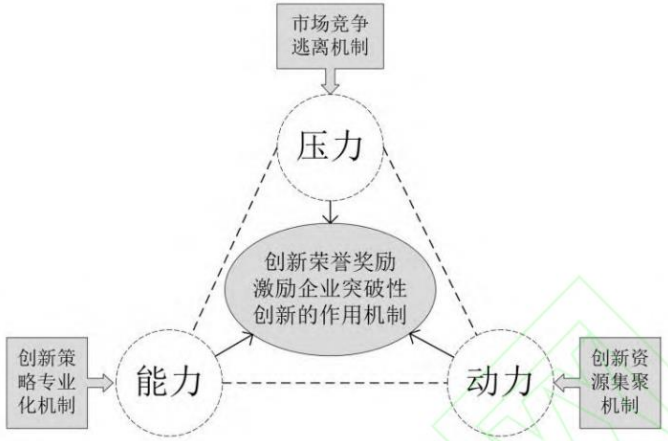
关联CSMAR子库：财务报表 /财务报表附注 / 财务指标分析 / 治理结构 / 上市公司基本信息 / 上市公司研发创新 / 专利被引用 / …

综合创新

CSMAR与前沿主题：（二）面向高水平科技自立自强的国家创新体系研究

相关研究3：创新荣誉奖励能否激励企业突破性创新——来自中国专利奖的证据[J].南开管理评论，2024.

文章以国家级权威奖项“中国专利奖”的评选作为一项准自然实验，探讨了创新荣誉奖励对企业突破性创新的影响及其作用机制。研究结果表明，创新荣誉奖励能够显著提升获奖企业的突破性创新水平，创新荣誉奖励主要通过创新策略专业化、创新资源集聚和市场竞争逃离三个机制，来促进企业突破性创新水平的提升。



创新荣誉奖励激励企业突破性创新的“三力”作用机制图

$$InvBreak_{i,t} = \beta_0 + \beta_1 Awards_{i,t} + \beta_2 \sum Controls_{i,t} + Firm_i + Year_t + \varepsilon_{i,t}$$

变量	变量定义
InvBreak	突破性创新
Award	创新荣誉奖励
Size	企业规模
TobinQ	企业价值
RDintensity	研发强度
Lev	资产负债率
Cfo	现金流水平
Roa	资产收益率
Growth	企业成长性
Market	市场势力
Density	资本密集度
Owncon	股权集中度

通过显性的专利前向引用数据来识别企业的突破性创新行为。具体而言，文章统计申请期在5年内的专利，将其中被引用总数在所属技术类别中处于前5%的专利定义为突破性创新专利。然后，在专利申请年份统计企业的突破性创新专利数量。为了更准确地刻画企业的突破性创新水平，文章采用突破性创新专利数量占企业专利申请总数的比重（×100%）来表示企业的突破性创新水平。

基准回归结果

变量	(1) *InvBreak*	(2) *InvBreak*	(3) *InvBreak*	(4) *InvBreak*
Award	0.506***	0.506***	0.572***	0.595***
	(0.149)	(0.149)	(0.149)	(0.148)

关联CSMAR子库：财务报表 /财务报表附注 / 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 / 专利被引用 / 经营困境 / …

创新策略专业化机制检验

变量	(1) *技术类别*	(2) *技术类别_新增*	(3) *技术类别_已有*	(4) *创新集中度*	(5) *技术空间相似度*
Award	-0.049**	-0.159***	0.059*	0.030***	-0.033***
	(0.025)	(0.033)	(0.033)	(0.006)	(0.010)

$$创新集中度_{i,t} = \sum\left(\frac{IPC4\ apply}{Total\ apply}\right)^2$$

$$技术空间相似度_{i,t} = \frac{X_{i,t}X'_{i,t-1}}{(X_{i,t}X'_{i,t})^{0.5}(X_{i,t-1}X'_{i,t-1})^{0.5}}$$

创新资源集聚机制检验

变量	(1) *战略合作*	(2) *产学研合作*	(3) *融资约束*	(4) *税收优惠*	(5) *政府补助*
Award	0.180***	0.029*	-0.024***	0.011*	-0.087
	(0.045)	(0.017)	(0.005)	(0.006)	(0.075)

市场竞争逃离机制检验

变量	(1) *媒体关注度*	(2) *媒体态度_正面*	(3) *行业技术竞争强度_数量*	(4) *行业技术竞争强度_获奖HHI*
Award	0.166***	0.196***	0.049**	-0.079***
	(0.061)	(0.045)	(0.020)	(0.030)

$$行业技术竞争强度_获奖HHI_t = \sum\left(\frac{Award\ apply_{i,t}}{Total\ apply_Industry_t}\right)^2$$

综合创新

CSMAR与前沿主题：（二）面向高水平科技自立自强的国家创新体系研究

相关研究4： “防微杜渐”还是“推波助澜”？散户“在线发声”与企业策略性创新信息披露[J].南开管理评论，2024.

文章基于策略性创新信息披露视角，检验散户利用股吧“在线发声”的治理效应。研究发现散户股吧发帖量越多，企业创新信息夸大披露水平越低、披露语调的积极程度越低。机制检验表明，散户“在线发声”的这一影响主要通过提高企业声誉成本和影响股价波动渠道来实现。进一步研究发现，股吧评论跟帖越多、发帖情绪越消极，以及有更多发帖活跃的散户，企业策略性创新信息披露水平及语调积极程度越低，但发帖的积极情绪会产生相反作用；散户股吧发声会降低前瞻性创新信息披露、抑制创新“多言寡行”行为。

$$INDISC_R_{i,t}/TONE_{i,t}=\alpha_0+\alpha_1 POST_{i,t}+\alpha_j Controls_{i,t}+\sum Year+\sum Industry+\varepsilon_{i,t}$$

变量类型	变量符号	变量名称	变量定义
被解释变量	*INDISC_R*	策略性创新信息披露水平	模型（1）的残差
	TONE	策略性创新信息披露语调	（积极词汇-消极词汇）/（积极词汇+消极词汇）
解释变量	*POST*	散户发帖数	发帖数加 1 取自然对数
控制变量	*SIZE*	企业规模	总资产取自然对数
	LEV	资产负债率	总负债/总资产
	ROA	总资产净利润率	净利润/平均资产总额
	SOE	产权性质	国有企业赋值 1，否则赋值 0
	TOP1	第一大股东持股比例	第一大股东持股股数/总股数
	CLRATIO	机构投资者持股比例	机构投资者持股股数/总股数
	QUA	两职合一	董事长与总经理同一人赋值 1，否则赋值 0
	BOARD	董事会规模	董事会总人数取自然对数
	INDR	董事会独立性	独立董事人数/董事会规模
	GROWTH	成长性	（本年营业收入-上年营业收入）/上年营业收入
	CASH	经营性现金流	经营活动产生的现金流量净额/总资产
	QUICK	速动比率	（流动资产-存货）/流动负债
	KZ	融资约束	KZ 指数
	IPP	知识产权保护	《全国知识产权发展状况报告》披露的知识产权保护指数除以 100
	HIGHTECH	行业属性	高科技企业赋值 1，否则赋值 0
	HHI	行业竞争	基于主营业务收入计算的行业赫芬达尔指数

$$INDISC_{i,t}=\alpha_0+\alpha_1 INDISC_M_{i,t}+\alpha_2 R\&D_{i,t}+\alpha_3 INN_{i,t}+\alpha_4 INV_P_{i,t}+\alpha_5 SIZE_{i,t}+\alpha_6 LEV_{i,t}+\alpha_7 GROWTH_{i,t}+\alpha_8 TOP1_{i,t}+\alpha_9 IC_{i,t}+\alpha_{10} QUICK_{i,t}+\alpha_{11} WAGE_{i,t}+\alpha_{12} IPP_{i,t}+\alpha_{13} MARKET_{i,t}+\sum Year+\sum Industry+\varepsilon_{i,t}$$

被解释变量为创新信息披露水平（INDISC）。解释变量包括：年度行业内其他 企业创新信息披露水平均值（INDISC_M）、研发投入金额（R&D）、申请发明专利数量（INN）、 研发人员数量（INV_P）、企业规模（SIZE）、资产负债率（LEV）、营业收入增长率（GROWTH）、 第一大股东持股比例（TOP1）、内部控制（IC）、速动比率（QUICK）、高管薪酬水平（WAGE）、 知识产权保护（IPP）、企业所处地区的市场化指数（MARKET）。另外还控制了年度和行 业固定效应。进行回归后的残差值表示异常披露水平，该值越大表明创新信息夸大披露程度越高。

散户“在线发声”与企业策略性创新信息披露

	INDISC_R	*TONE*
	(1)	(2)
POST	-0.0152***	-0.0025**
	(-3.2313)	(-2.3239)

《财富》杂志 2010-2022 年评选的最受尊敬中国企业榜单

股票月度收益率的标准差

机制检验：声誉成本

	INDISC_R	*TONE*	*INDISC_R*	*TONE*
	(1)	(2)	(3)	(4)
POST	-0.0142***	-0.0026**	-0.0129***	-0.0022*
	(-2.9130)	(-2.2532)	(-2.6393)	(-1.9297)
REP	-0.0820***	-0.0124*	0.4433	0.1341**
	(-3.0742)	(-1.7289)	(1.6387)	(2.0069)
POST×REP			-0.0537**	-0.0150**
			(-1.9753)	(-2.1806)

机制检验：股价波动

	INDISC_R	*TONE*	*INDISC_R*	*TONE*
	(1)	(2)	(3)	(4)
POST	-0.0246***	-0.0028**	0.0079	-0.0001
	(-4.6630)	(-2.5078)	(1.0636)	(-0.0813)
SPV	0.2668***	0.0092	2.9277***	0.2318***
	(3.4912)	(0.8777)	(7.0199)	(3.3356)
POST×SPV			-0.2670***	-0.0223***
			(-6.2431)	(-3.3462)

关联CSMAR子库： 社交媒体 / 董事长致辞 / 市场分析报告文本 / 财务报表 / 治理结构 / 上市公司基本信息 / 上市公司研发创新 / 专利被引用 / 股票市场交易 / …

绿色创新

CSMAR与前沿主题：（二）面向高水平科技自立自强的国家创新体系研究

相关研究5：机器人应用、CEO绿色经历与企业绿色技术创新[J]. 经济管理, 2024.

文章利用2006-2019年中国工业机器人数据与沪深A股上市企业绿色技术创新数据，实证检验了机器人应用、CEO绿色经历和企业绿色技术创新之间的关系。研究发现，CEO具有绿色经历的企业，机器人应用促进绿色技术创新能力提升的作用越强。机器人应用通过提升企业效率和优化人力资本结构推动企业技术创新，且CEO绿色经历促使机器人应用引发的技术创新偏向于绿色技术创新。

$$green_innov_{it} = \beta_0 + \beta_1 lnrobot_{jt} + \sum a_m x_{itm} + \sum \sigma_n z_{jtn} + \mu_i + \nu_t + \varepsilon_{ijt}$$

$$green_innov_{it} = \gamma_0 + \gamma_1 lnrobot_{jt} + \gamma_2 green_{it} + \gamma_3 lnrobot_{jt} \times green_{it} + \sum a_m x_{itm} + \sum \sigma_n z_{jtn} + \mu_i + \nu_t + \varepsilon_{ijt}$$

变量性质	变量名	符号	变量测度
被解释变量	绿色技术创新水平	*envpat_total*	企业绿色专利申请量加1后取对数
	实质性绿色技术创新	*envpat_inv*	企业绿色发明专利申请量加1后取对数
	策略性绿色技术创新	*envpat_uty*	企业绿色实用新型专利申请量加1后取对数
解释变量	机器人应用水平	*lnrobot*	机器人渗透度取对数
调节变量	CEO绿色经历	*green*	CEO是否接受过绿色相关教育或从事过绿色相关工作
控制变量	经济发展水平	*lngdp*	GDP平减后取对数
	外商投资水平	*lnopen*	外商直接投资总额与GDP的比值
	行业竞争程度	*hhi*	行业内的每家公司总资产与行业总资产合计的比值的平方累加
	第一大股东持股比例	*top1*	第一大股东持股份额占总股本的比例
	企业年龄	*lnage*	观测年度减去企业上市时间加1后取对数
	企业总资产收益率	*ROA*	净利润/总资产
	企业财务健康状况	*leverage*	总负债/总资产取对数
	独立董事比例	*independent*	独立董事人数/董事会总人数

基准回归结果

变量	*envpat_total*		*envpat_inv*		*envpat_uty*	
	(1)	(2)	(3)	(4)	(5)	(6)
lnrobot	0.0464***	0.0450***	0.0324***	0.0088	0.0316***	0.0304***
	(0.0070)	(0.0070)	(0.0055)	(0.0066)	(0.0054)	(0.0054)
green		0.1456**		0.0922**		0.0862*
		(0.0572)		(0.0452)		(0.0447)
lnrobot×green		0.0751*		0.0516*		0.0749**
		(0.0398)		(0.0305)		(0.0319)

$$M_{it} = \theta_0 + \theta_1 lnrobot_{jt} + \sum a_m x_{itm} + \sum \sigma_n z_{jtn} + \mu_i + \nu_t + \varepsilon_{ijt}$$

$$innov_{it} = \delta_0 + \delta_1 lnrobot_{jt} + \delta_2 M_{it} + \sum a_m x_{itm} + \sum \sigma_n z_{jtn} + \mu_i + \nu_t + \varepsilon_{ijt}$$

$$green_innovper_{it} = \tau_0 + \tau_1 lnrobot_{jt} + \tau_2 green_{it} + \tau_3 lnrobot_{jt} * green_{it} + \sum a_m x_{itm} + \sum \sigma_n z_{jtn} + \mu_i + \nu_t + \varepsilon_{ijt}$$

机制变量

企业效率：全要素生产率

人力资本：硕士及以上员工人数占比

影响机制估计结果

变量	*tfp*	*highlabor*	*pat_total*	*pat_inv*	*pat_uty*	*pat_tper*	*pat_iper*	*pat_uper*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
lnrobot	0.1517***	0.0042***	0.0969***	0.0688***	0.1001***	0.0062*	0.0002	0.0036*
	(0.0087)	(0.0005)	(0.0187)	(0.0156)	(0.0160)	(0.0035)	(0.0022)	(0.0020)
tfp			0.0422**	0.0564***	0.0152			
			(0.0211)	(0.0172)	(0.0175)			
highlabor			0.8432***	1.2389***	0.3029			
			(0.2607)	(0.2395)	(0.2263)			
green						0.0189	−0.0037	0.0193
						(0.0374)	(0.0221)	(0.0201)
lnrobot×green						0.0301	0.0246*	0.0091
						(0.0208)	(0.0140)	(0.0121)

关联CSMAR子库：绿色专利 / 人工智能 / 上市公司人物特征 / 财务报表 / 财务报表附注 / 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 / 专利被引用 / …

绿色创新

CSMAR与前沿主题：（二）面向高水平科技自立自强的国家创新体系研究

相关研究6：绿色创新的供应链溢出——基于核心企业与供应商协同发展的视角[J].经济管理，2024.

文章采用2008-2021年沪深两市A股上市公司的供应链关系数据，检验核心企业绿色创新对供应商绿色创新的影响。研究发现，核心企业绿色创新对供应商绿色创新具有正向溢出效应，其中溢出效应主要体现在供应商绿色发明专利的增加，而对绿色实用新型专利没有显著影响。机制分析表明，核心企业通过信息反馈、提高供应商环境治理水平、缓解供应商融资约束来实现对供应商绿色创新的溢出效应。

$$PatentS_{i,t}/InvPatentS_{i,t}/UPatentS_{i,t} = \alpha + \beta_1 Patent_{j,t-1} + \beta_2 Controls_{i,t} + year_t + d_i + \varepsilon_{ijt}$$

变量符号	变量定义
Patent	核心企业绿色创新
PatentS	供应商绿色创新
InvPatentS	供应商绿色发明创新
UPatentS	供应商绿色实用新型创新
AgeS	供应商年龄
AssetS	供应商资产规模
LevS	供应商资产负债率
ROAS	供应商资产收益率
OpicrtS	供应商营业收入增长率
ShrHolder1S	供应商第一大股东持股比例
InstS	供应商机构投资者持股比例
TobinQS	供应商的托宾Q值
BMS	供应商账面市值比
CashratioS	供应商现金比率

基准回归结果

变量	(1)	(2)	(3)	(4)
	PatentS	*PatentS*	*InvPatentS*	*UPatentS*
Patent	0.0855*** (2.6974)	0.0913*** (2.8096)	0.0628** (2.0496)	0.0693 (1.3536)

信息反馈机制

变量	(1)	(2)	(3)	(4)
	PatentS	*PatentS*	*InvPatentS*	*UPatentS*
Patent	0.1407*** (3.1772)	0.1383*** (2.8316)	0.1301*** (2.6462)	0.0911** (2.2883)
feedback	0.0362 (1.2770)	0.0156 (0.4984)	-0.0058 (-0.4307)	0.0705** (2.0794)
*Patent*feedback*	-0.0854*** (-3.0434)	-0.0709** (-2.0036)	-0.1007* (-1.8180)	-0.0282 (-0.8945)

改善环境治理水平机制

变量	(1)	(2)	(3)	(4)
	PatentS	*PatentS*	*InvPatentS*	*UPatentS*
Patent	0.0970*** (2.8943)	0.1018*** (2.8208)	0.0851** (2.2802)	0.0534 (1.0129)
diff	-0.0496 (-0.8686)	-0.0149 (-0.1939)	-0.0069 (-0.1242)	-0.0357 (-0.9918)
*Patent*diff*	-0.0369 (-1.3814)	-0.0308 (-0.8205)	-0.0656** (-2.3989)	0.0515 (1.2023)

缓解融资约束机制

变量	KZ指数			SA指数		
	(1)	(2)	(3)	(4)	(5)	(6)
	PatentS	*InvPatentS*	*UPatentS*	*PatentS*	*InvPatentS*	*UPatentS*
Patent	0.1417*** (3.5952)	0.1029** (2.1070)	0.1140** (2.3453)	-0.2151 (-1.3670)	-0.1712 (-1.1619)	-0.2891*** (-2.9472)
constraint	-0.0025 (-0.3439)	-0.0038 (-0.6201)	-0.0067 (-0.6276)	1.2272** (2.4009)	1.0097*** (2.6914)	1.1415*** (3.9772)
*Patent*constraint*	-0.0257*** (-2.6732)	-0.0226** (-2.0835)	-0.0200*** (-4.4339)	-0.0855* (-1.7047)	-0.0648 (-1.3934)	-0.1020*** (-3.8650)

关联CSMAR子库：绿色专利 / 供应链研究 / 财务报表 / 财务报表附注 / 财务指标分析 治理结构 / 上市公司基本信息 / 经营困境 / 环境研究 / 商道融绿ESG / …

CSMAR与前沿主题：（二）面向高水平科技自立自强的国家创新体系研究

相关研究7：内部控制对企业数字技术创新的影响研究[J].审计研究，2024.

文章以2009-2022年我国 A 股上市公司为样本，对内部控制质量如何影响企业数字技术创新进行了实证分析。结果显示，高质量的内部控制显著推动了企业的数字技术创新，尤其在资产安全、经营和战略方面对创新数量具有明显的正向作用；同时，报告、资产安全、经营和战略也显著提升了创新质量。

$$INNOVATE_{it}=\alpha_0+\alpha_1 ICQ_{it}+\alpha_k X_{kit}+\mu_i+\lambda_t+\varepsilon_{it}$$

$$M_{it}=\beta_0+\beta_1 ICQ_{it}+\beta_k X_{kit}+\mu_i+\lambda_t+\varepsilon_{it}$$

$$INNOVATE_{it}=\varphi_0+\varphi_1 M_{it}+\varphi_2 ICQ_{it}+\beta_k X_{kit}+\mu_i+\lambda_t+\varepsilon_{it}$$

变量符号	变量名称	变量定义
ICQ	内部控制质量	迪博内部控制与风险管理数据库的内部控制指数取自然对数
DTI1	数字技术创新数量	根据专利文本信息识别的企业层面的数字技术创新水平
DTI2	数字技术创新质量	数字技术专利知识宽度
SIZE	企业规模	总资产账面价值（万元）取对数
LEV	企业偿债能力	资产负债率，负债合计同资产合计的比例
ROE	盈利能力	净资产收益率
GROWTH	企业成长性	企业主营业务增长率
AGE	上市年限	企业成立年限取自然对数
Tobin Q	企业价值	市值 / 资产总计
TOP10	股权集中度	前十大股东持股比例合计
IND	董事会独立性	独立董事占董事会总人数的比重
SOE	产权性质	虚拟变量，1 代表国有企业，0 代表非国有企业
HHI	企业所处行业的竞争度	使用赫芬达尔指数，公式为：HHI=sum[（Xi/X）^2]
MARKET	市场化相对进程	根据樊纲和王小鲁（2003）编制的各地区市场化进程中各地区市场化相对进程得分

变量	被解释变量	
	DTI1	DTI2
ICQ	0.231*(1.842)	0.318**(2.269)

稳健性检验中，使用内部控制缺陷（ICW）替换解释变量，若内控自评报告或内控审计报告中披露上市公司存在内部控制缺陷信息，则ICW为 1，若未披露或者披露不存在内控缺陷，则 ICW为 0。

内部控制质量分项检验结果

被解释变量：DTI1						被解释变量：DTI2					
变量	合法合规	报告	资产安全	经营	战略	变量	合法合规	报告	资产安全	经营	战略
ICQ	0.0001 (0.007)	0.0002 (0.007)	0.501*** (5.206)	0.597*** (6.896)	0.111* (1.769)	ICQ	0.006 (0.168)	0.158** (2.263)	0.500*** (4.649)	0.631*** (6.530)	0.146** (2.260)

企业资源配置效率

自变量	中介变量：Ineff		
	Ineff	DTI1	DTI2
ICQ	−0.026*** (−2.729)	0.222* (1.731)	0.312** (2.180)
Ineff		−0.699*** (−5.399)	−0.666*** (−4.612)

融资约束

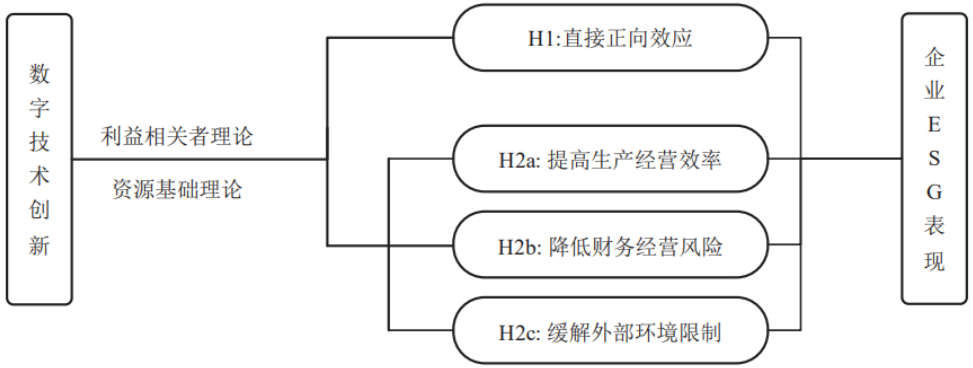
自变量	中介变量：WW		
	WW	DTI1	DTI2
ICQ	−0.325*** (−10.075)	0.199* (1.890)	0.288* (1.912)
WW		−2.483*** (−5.490)	−2.739*** (−5.434)

关联CSMAR子库：企业数字化转型 / 上市公司与关联公司专利 / 内部控制 / 财务报表 / 财务报表附注 / 财务指标分析 / 治理结构 / 上市公司基本信息 / 市场竞争 / 融资约束 / …

CSMAR与前沿主题：（二）面向高水平科技自立自强的国家创新体系研究

相关研究8：数字技术创新与企业 ESG 表现——基于专利文本与机器学习[J].研究与发展管理, 2025.

文章以2009—2022年A股上市公司为样本，研究数字技术创新对ESG表现的影响效应及作用机制。研究发现，数字技术创新能够显著提升企业ESG表现。机制分析表明，数字技术创新能够通过提高生产经营效率、降低财务经营风险和缓解外部环境限制来提升ESG表现。



$$ESG_{it} = \alpha_0 + \beta_1 \ln DTI_{it} + \sum Controls + \sum Year + \sum Firm + \sum City + \varepsilon_{it}$$

基准回归结果

变量	(1) *ESG*	(2) *ESG*	(3) *ESG*	(4) *ESG*
ln *DTI*	0.532***(0.045)	0.322***(0.044)	0.283***(0.045)	0.122***(0.044)

机制检验

变量	(1)*TFP-OP*	(2)*TFP-LP*	(3)*EBIT*	(4)*DD*	(5)*Asset*	(6)*FC*
ln *DTI*	0.013***(0.004)	0.008**(0.004)	−0.002***(0.001)	0.040*(0.022)	−0.002***(0.000)	−0.003**(0.001)

优化生产经营效率：采用 OP 法和 LP 法两种方法测算的企业全要素生产率（TFP）来衡量企业生产经营效率

降低财务经营风险：采用息税前利润波动性（EBIT）即企业相应年度的息税前利润与平均总资产的比值，以企业t-2年至t+2年为一个周期，经过同年份和同行业调整后的值度量利润波动性；采用 Merton 的KMV模型测度的债务违约距离衡量企业债务违约风险

缓解外部环境限制：资产专用性（Asset）度量企业外部交易成本；采用FC指数度量企业融资约束

关键解释变量	被解释变量
上市公司各年度的数字技术创新专利	华证ESG

关联CSMAR子库：企业数字化转型 / 商道融绿ESG / 财务报表 / 财务报表附注 / 财务指标分析 / 治理结构 / 上市公司基本信息 / 市场竞争 / 融资约束 / …

CSMAR与前沿主题：（二）面向高水平科技自立自强的国家创新体系研究

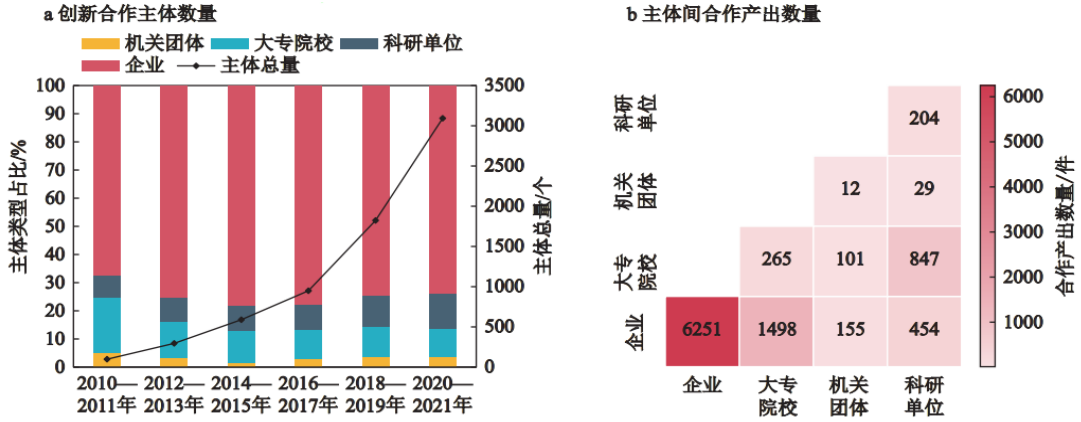
相关研究9：中国人工智能创新合作网络的时空格局及其影响因素分析[J].地理科学, 2025.

文章基于2010—2021年中国人工智能产业发明专利申请数据，采用社会网络分析、加权指数随机图模型等方法，对该产业创新合作网络的时空格局进行分析，并从网络结构、多维邻近性与城市特征 3 个角度分析网络形成与演化的影响因素。研究发现：中国城市间人工智能创新合作网络形成了以北京为核心，由京津冀、长三角、珠三角和成渝城市群构成的“菱形”格局，网络呈现典型的异配性、小世界属性和“核心–边缘”结构。人工智能对技术、产业的依赖性导致其创新合作网络受网络自组织性、多维邻近性与城市创新合作环境影响：择优链接与中介效应促进网络发育，创新合作由互补创新转向互补与协同创新并存，地理距离、省际壁垒不再是阻碍创新合作的因素；城市经济实力和人工智能产业发展水平促进创新联系的强度与广度，但城市间发展水平差距一定程度上也阻碍合作建立。

研究方法

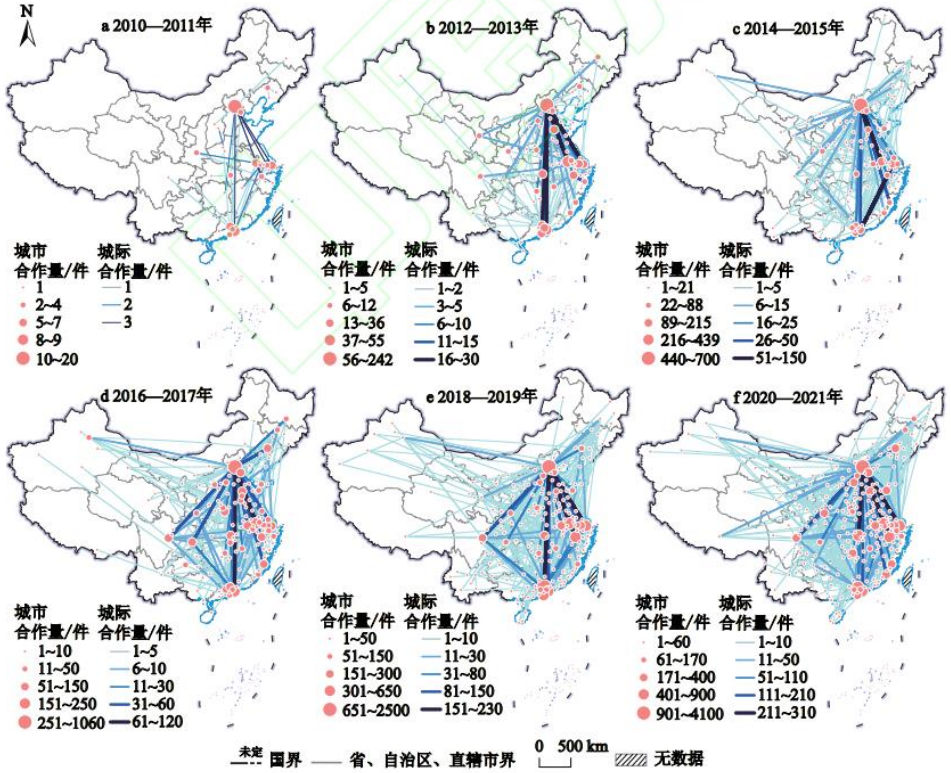
社会网络分析方法：采用节点度、平均度、度相关性系数、平均路径长度、平均集聚系数等指标考察创新合作网络的拓扑结构特征，并生成随机网络判断现实网络是否为小世界网络。

加权指数随机图模型：城市创新合作数量为非负整数且存在过度分散，学者们更倾向于采用网络计量模型， 如多元回归二次指派模型、随机指数图模型（ERGM）等分析。ERGM 通过概率函数判别真实网络相比随机网络出现可能性，定量估计网络内生结构依赖和外生机制，被广泛应用 于网络演化机制研究。加权 ERGM 模型可对加权网络回归。文章采用加权 ERGM 模型分析中国人工智能产业创新网络演化的动力机制。



港澳台数据暂缺

2010—2021 年中国城市间人工智能创新合作主体与合作产出数量



基于自然资源部地图技术审查中心标准地区（审图号：GS(2023)2763 号）绘制，底图无修改；港澳台数据暂缺

2010—2021 年中国人工智能创新合作网络

CSMAR与前沿主题：（二）面向高水平科技自立自强的国家创新体系研究

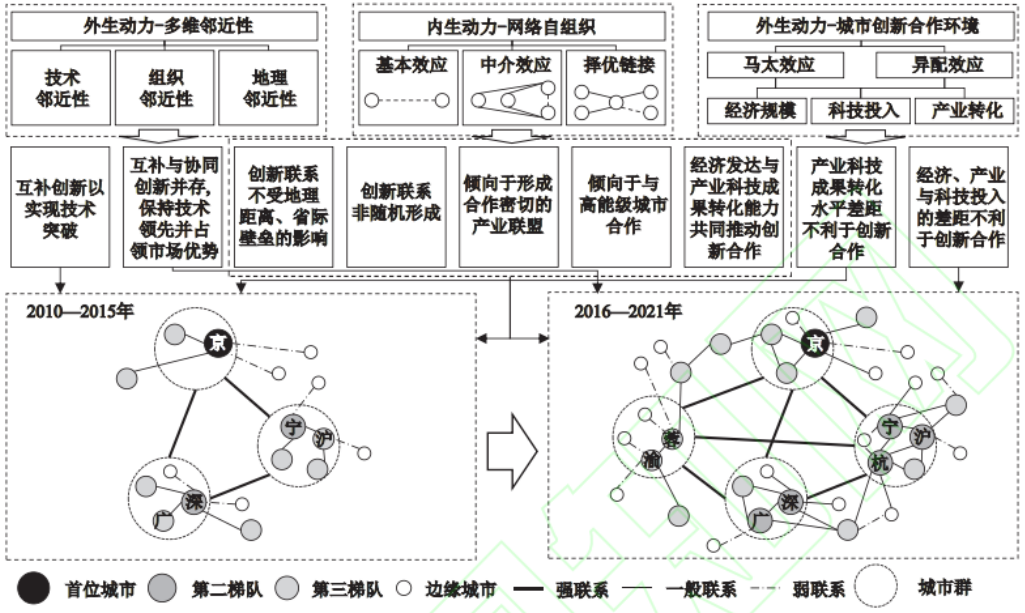
相关研究9：中国人工智能创新合作网络的时空格局及其影响因素分析[J].地理科学, 2025.

文章基于2010—2021年中国人工智能产业发明专利申请数据，采用社会网络分析、加权指数随机图模型等方法，对该产业创新合作网络的时空格局进行分析，并从网络结构、多维邻近性与城市特征 3 个角度分析网络形成与演化的影响因素。研究发现：中国城市间人工智能创新合作网络形成了以北京为核心，由京津冀、长三角、珠三角和成渝城市群构成的“菱形”格局，网络呈现典型的异配性、小世界属性和“核心–边缘”结构。人工智能对技术、产业的依赖性导致其创新合作网络受网络自组织性、多维邻近性与城市创新合作环境影响：择优链接与中介效应促进网络发育，创新合作由互补创新转向互补与协同创新并存，地理距离、省际壁垒不再是阻碍创新合作的因素；城市经济实力和人工智能产业发展水平促进创新联系的强度与广度，但城市间发展水平差距一定程度上也阻碍合作建立。

加权边
(sum)
K 星构型
(node)
三元组
(transi)
技术邻近性
(tech)
地理邻近性
(geo)
组织邻近性
(insti)
城市经济规模
(gdp)
城市科技发展
(sci)
人工智能产业
发展水平
(comp)

加权指数随机图模型结果

模型变量	2010—2015年				2016—2021年			
	模型1	模型2	模型3	模型4	模型5	模型6	模型7	模型8
网络自组织性								
sum	−0.3447*** (0.1065)	−1.0998*** (0.1481)	−1.9308*** (0.2256)	−1.9058*** (0.2306)	−0.2445*** (0.0547)	−0.6816*** (0.0828)	−1.9930*** (0.1415)	−2.0667*** (0.1501)
node		1.8024*** (0.1077)		1.8604*** (0.1768)		1.3157*** (0.0723)		0.9747*** (0.0680)
transi		0.2176** (0.0890)		0.1279 (0.1020)		0.0356 (0.0657)		0.1828** (0.0785)
多维邻近性								
geo			0.0014 (0.0020)	0.0018 (0.0022)			0.0014 (0.0013)	0.0020 (0.0013)
insti			0.0342 (0.2841)	0.2186 (0.2748)			0.0679 (0.1273)	0.0902 (0.1317)
tech			−0.0922*** (0.0296)	−0.0902*** (0.0329)			0.0211 (0.0272)	0.0236 (0.0276)
城市创新合作环境(马太效应)								
hom(gdp)			0.0154*** (0.0046)	0.0188*** (0.0044)			0.0307*** (0.0030)	0.0281*** (0.0029)
hom(sci)			−0.0090 (0.0104)	−0.0138 (0.0111)			0.0146** (0.0064)	0.0093 (0.0060)
hom(comp)			0.0112*** (0.0017)	0.0103*** (0.0015)			0.0020*** (0.0003)	0.0021*** (0.0003)
城市创新合作环境(异配效应)								
het(gdp)			0.0002 (0.0051)	−0.0003 (0.0051)			−0.0057* (0.0030)	−0.0108*** (0.0031)
het(sci)			0.0074 (0.0105)	−0.0009 (0.0112)			−0.0220*** (0.0062)	−0.0251*** (0.0061)
het(comp)			−0.0067*** (0.0016)	−0.0086*** (0.0015)			−0.0013*** (0.0003)	−0.0010*** (0.0003)
nonzero	−4.8835*** (0.1820)	−4.6539** (0.1820)	−3.7002*** (0.2011)	−4.1386*** (0.2010)	−5.1208*** (0.0970)	−4.9649*** (0.0959)	−4.1232*** (0.1113)	−4.3651*** (0.1040)
AIC	−27871	−28958	−28034	−28232	−111066	−111418	−111780	−112130
BIC	−27857	−28931	−27958	−28142	−111049	−111385	−111689	−112022



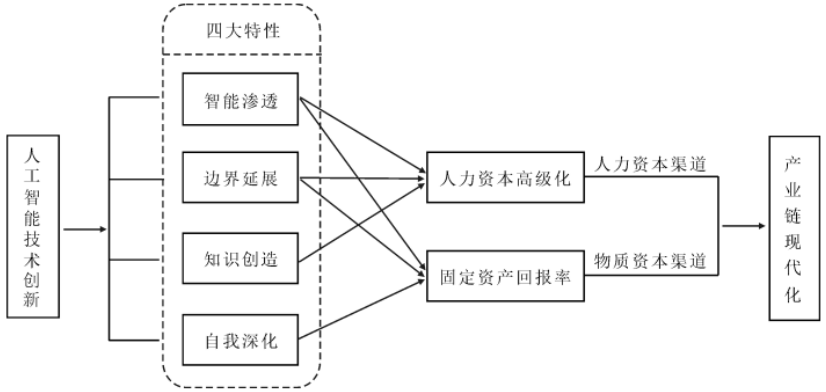
中国人工智能创新合作网络演化机制总结

关联CSMAR子库：人工智能专利 / 人工智能产业链 / 经济地理 / 宏观经济 / 区域经济 / …

CSMAR与前沿主题：（二）面向高水平科技自立自强的国家创新体系研究

相关研究10：人工智能技术创新对产业链现代化的影响效应及作用机制[J].科技进步与对策, 2025.

文章利用2003—2022年中国内地30个省份面板数据，实证检验人工智能技术创新对于产业链现代化的影响效应和作用机制。研究结果表明：人工智能技术创新能够显著促进产业链现代化。异质性分析显示，人工智能技术创新的赋能效应在高产业链现代化水平地区、高对外贸易水平地区以及东部地区更强。机制分析发现，人工智能技术创新可以通过推动人力资本高级化和提升固定资产回报率提高产业链现代化水平。进一步研究表明，随着人工智能技术与产业链的融合发展，人工智能技术创新对产业链现代化水平的影响在时间上呈先下降后上升的阶段性差异，从长远分析，其正向作用远远大于创新初期的负面影响。



$$MIC_{it} = \alpha + \beta_0 AI_{it} + \beta_1 Controls_{it} + \gamma_i + \delta_t + \varepsilon_{it}$$

关键解释变量	被解释变量
人工智能创新：人工智能专利，也可用工业机器人数据	产业链现代化水平：从产业链基础、产业链创新、产业链韧性、产业链协同等六大维度构建指标体系测度

基准回归结果

变量	(1)	(2)	(3)	(4)
AI	1.109***(0.112)	0.593***(0.106)	0.520***(0.101)	0.538***(0.103)

机制检验回归结果

变量	(1) *Labor*	(2) *H_labor*	(3) *L_labor*	(4) *R*
AI	0.971**	0.173**	−0.221**	0.463***
	(0.452)	(0.078)	(0.082)	(0.125)

人力资本：平均受教育年限
固定资产回报率：各省份工业企业利润总额与固定 资产净值之比

短期影响趋势分析结果

变量	(1)	(2)	(3)	(4)
AI	−0.290***	−4.406***	−0.247**	−3.891**
	(0.070)	(1.110)	(0.081)	(1.354)
L.AI	0.436***	5.340***	0.429***	5.190***
	(0.060)	(0.810)	(0.061)	(0.821)
L2.AI	0.524***	6.069***	0.485***	5.729***
	(0.068)	(0.950)	(0.069)	(1.011)

长期效应分析结果

变量	(1)	(2)	(3)	(4)
AI	0.669***	7.003***	0.668***	7.029***
	(0.107)	(1.598)	(0.113)	(1.675)
L.AI-AI	0.436***	5.340***	0.429***	5.190***
	(0.060)	(0.810)	(0.061)	(0.821)
L2.AI-AI	0.524***	6.069***	0.485***	5.729***
	(0.068)	(0.950)	(0.069)	(1.011)

关联CSMAR子库：人工智能专利 / 人工智能 / 数字经济 / 宏观经济 / 区域经济 / …

CSMAR与前沿主题：（三）加快发展数字经济研究

	类型	数据获取	文献
数字化水平测度	宏观：构建统计指标体系（数字经济规模、数字基础设施、数字化转型、数字研发创新……）	CSMAR数字经济/经济内循环	我国数字经济发展新特征：非均衡性与趋势演进[J].经济问题探索,2023.
	微观：**数字化战略**（文本分析挖掘年报中的关键词）、**数字化投资**（以企业年末无形资产明细中与数字经济相关部分的金额占无形资产总额的比例来度量）以及**数字化管理**（以上市公司高管团队中是否设立首席信息官（CIO）、首席数据官（CDO）职位，或者关联主营业务与数字经济相关的企业度量）	CSMAR数字经济/企业数字化转型	企业数字化的产业链联动效应研究[J].中国工业经济,2023.
	数字产业集群发展水平：构建统计指标体系（数字企业密度、数字产业上市公司客户供应商集中度、数字经济发明专利授权量……）	CSMAR数字经济/企业数字化转型/供应链研究	数字产业集群发展水平评价及其时空演变[J].经济地理, 2024.
	数字新质生产力：从数字劳动力、 数字劳动对象、 数字生产资料三方面构建评价指标体系, 使用 TOPSIS 法计算	CSMAR数字经济/企业数字化转型/宏观经济/经济内循环/专利	中国数字新质生产力统计测度及影响因素研究[J].经济问题探索, 2024.

	主题	文献
数字经济与实体经济深度融合发展	企业层面/产业层面/社会经济运行层面	企业数字化与ESG履责[J].会计研究, 2024. 产业数字化、虚拟集聚与新质生产力发展[J].科技进步与对策, 2025. 数字产业化能否增强城市产业链韧性？[J].现代经济探讨, 2024. 中国式现代化视域下数字经济的共同富裕效应：方法与证据[J].经济研究, 2024.

	主题	文献
数字化进程中的风险与治理	市场波动性风险/发展结构性风险/收入差距性风险	“多言寡行”的数字化转型披露能否被投资者识别？——基于股票错误定价的证据[J].外国经济与管理, 2025. 数字金融与企业金融化：“脱实向虚”还是“脱虚向实”？[J]. 财经理论与实践, 2024. 数字技术使用更有助于灵活就业人员收入增长吗?——基于二级数字鸿沟视角[J].软科学, 2024.

CSMAR与前沿主题：（三）加快发展数字经济研究

相关研究1：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了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

2

第二步：构建待预测句库和待标记句库

第三步：对待标记句库的句子进行人工标注

3

4

第四步：采用有监督的机器学习方法训练模型

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

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

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

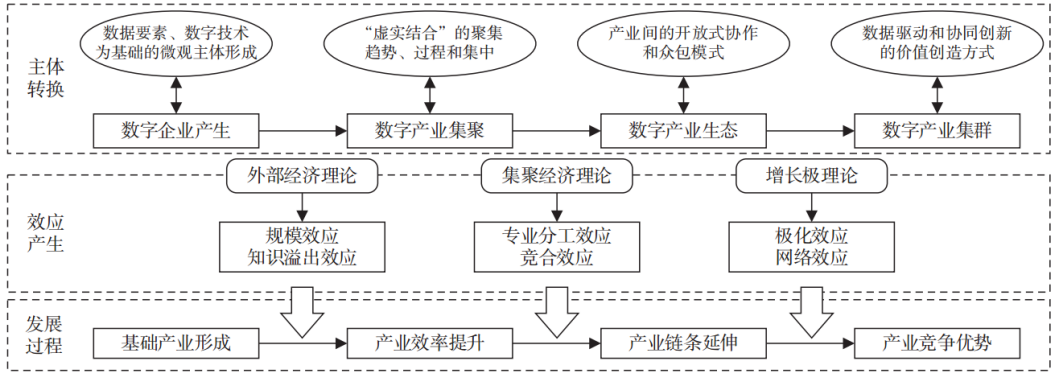
变量	(1)	(2)	(3)	(4)
	ROA	*ROE*	*F.ROA*	*F.ROE*
DigiTech	0.398***	0.740***	0.183*	0.435**
	(0.0896)	(0.208)	(0.0968)	(0.218)

关联CSMAR子库：数字经济 / 企业数字化转型 / 财务报表 / 财务指标分析 / …

CSMAR与前沿主题：（三）加快发展数字经济研究

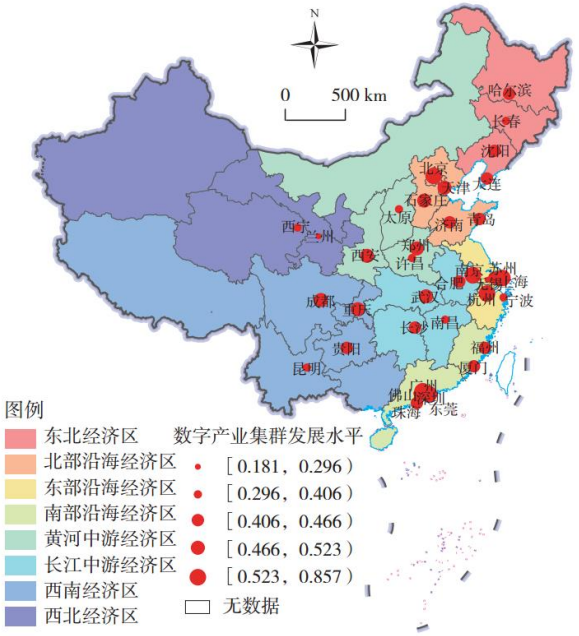
相关研究2：数字产业集群发展水平评价及其时空演变[J].经济地理，2024.

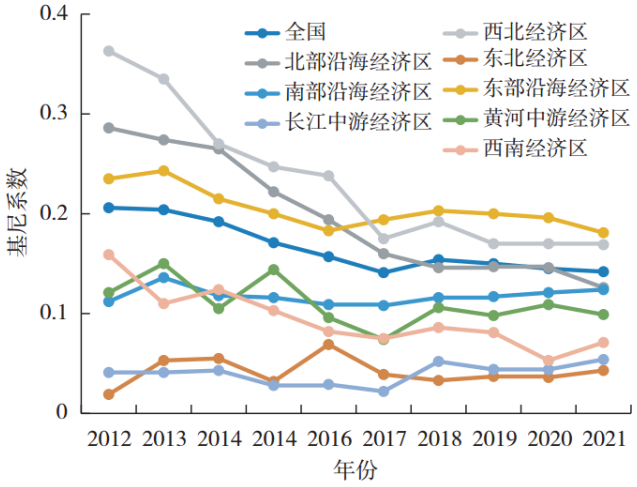
文章从数字产业集群演化视角出发，结合区位商和社会网络分析法从中国302个城市中识别出36个数字产业集群；同时综合运用动态改进理想解法首次测算了数字产业集群发展水平，并通过整体与维度分析探究其区域差异及来源和分布动态。结果表明：①中国数字产业集群发展水平呈逐年增长趋势。②支持度及其增长率最高，是数字产业集群发展的“压舱石”；关联度增长率最低，是最具潜力的发展点。③区域差异呈微弱缩小趋势，主要源自区域间差异和超变密度。④数字产业集群发展具有稳定性和俱乐部趋同特征。



数字产业集群的形成与演进过程

一级指标	二级指标	三级指标	指标属性	数据来源
集中度	企业密度	数字企业数量/地理面积	正向	企查查数据库、《中国城市统计年鉴》
	人力集聚	区位商	正向	EPS数据库、《中国城市统计年鉴》
关联度	产业链分工	数字产业上市公司供应商公司集中度	负向	CSMAR数据库
		数字产业上市公司客户公司集中度	负向	
	创新链协同	集群内联合发明专利申请强度	正向	Incopat专利数据库
		集群间联合发明专利申请强度	正向	
贡献度	经济效益	邮电业务贡献比	正向	CNRDS数据库
		数字产业上市公司纳税额	正向	
	创新成果	数字经济发明专利授权量/总发明专利授权量	正向	
	社会影响	就业贡献量(信息传输、计算机服务和软件业从业人员)[25]	正向	《中国城市统计年鉴》
支持度	基础网络	政府数字化注意力	正向	北大法宝以及各地方政府门户网站
	交易网络	市场化程度[26]	正向	各省份统计年鉴以及地级市统计公报
	资本网络	金融网点数量	正向	国家金融监督管理总局
		数字普惠金融指数[27]	正向	北京大学普惠金融指数





中国及八大经济区数字产业集群发展水平的组内基尼系数演变趋势

关联CSMAR子库：宏观经济 / 数字经济 / 经济内循环 / 金融机构分支机构 / 上市公司联盟 / 供应链研究 /…

测度

CSMAR与前沿主题：（三）加快发展数字经济研究

相关研究3：中国数字新质生产力统计测度及影响因素研究[J].经济问题探索，2024.

文章从数字产业集群演化视角出发，结合区位商和社会网络分析法从中国302个城市中识别出36个数字产业集群；同时综合运用动态改进理想解法首次测算了数字产业集群发展水平，并通过整体与维度分析探究其区域差异及来源和分布动态。结果表明：①中国数字产业集群发展水平呈逐年增长趋势。②支持度及其增长率最高，是数字产业集群发展的“压舱石”；关联度增长率最低，是最具潜力的发展点。③区域差异呈微弱缩小趋势，主要源自区域间差异和超变密度。④数字产业集群发展具有稳定性和俱乐部趋同特征。

中国数字新质生产力指标体系

	一级指标	二级指标解释变量	单位	属性	资料来源
数字新质生产力	数字劳动力	高被引学者人数	个	正	国家统计局网站、《中国劳动统计年鉴》
		不同省份计算机、软件、信息技术相关专业的硕士人数	个	正	
		信息传输、软件和信息技术服务业城镇单位就业人数	个	正	
		信息行业人员从业人数占就业人口比例	%	正	
		人均受教育年限	年	正	
		数字企业间高管人员流动数量	个	正	
	数字劳动对象	机器人数量占总体人口数量比重	%	正	国家统计局网站、《中国劳动统计年鉴》《中国环境统计年鉴》
		战略性新兴产业增加值占国内生产总值比重	%	正	
		工业互联网产业增加值废水排放总量	吨	负	
		工业互联网产业增加值二氧化硫排放总量	吨	负	
		工业互联网产业增加值固体废弃物排放总量	吨	负	
		工业互联网废水治理设施处理能力	万吨/天	正	
		工业互联网固体废物综合利用率	%	正	
		生活垃圾无害化处理率	%	正	
数字新质生产力	数字生产资料	人均互联网宽带接入端口数量	个	正	国家统计局网站、《中国劳动统计年鉴》《中国科技统计年鉴》《中国能源统计年鉴》
		光缆线路长度	公里	正	
		国内专利申请授权总体数量和规模以上工业企业 R&D 人员全时当量占比	%	正	
		数字经济指数	-	正	
		技术引进合同总体金额	万元	正	
		数字产业专利转让数量	个	正	
		R&D 经费支出占国内生产总值比例	%	正	
		能源消耗占国内生产总值比例	吨/万元	负	
		可再生能源电力消耗量占全社会用电量比例	%	正	
		人均网站数量	个	正	
		政企宽带接入用户数量	万户	正	
		移动电话基站密度	万个	正	

$$DNQP_{it} = \beta_0 + \sum_{j=1}^{5} \alpha_j Y_{it} + \mu_i + \lambda_t + \varepsilon_{ij}$$

经济水平、技术创新投入、金融发展水平、对外依存度、人力资本

基准回归结果

变量	静态面板			动态面板	
	Fe			差分 *GMM*	系统 *GMM*
	(1)	(2)	(3)	(4)	(5)
L1. DNQP	–	–	–	0.533*** (0.138)	0.923*** (0.129)
EL	0.793*** (0.239)	0.684** (0.307)	0.943** (0.449)	0.424*** (0.089)	0.537*** (0.057)
TII	0.475*** (0.099)	0.318* (0.182)	0.312 (0.325)	0.321*** (0.068)	0.435*** (0.128)
FDL	0.319*** (0.093)	0.459*** (0.098)	0.483*** (0.141)	0.332*** (0.077)	−0.093** (0.042)
FTD	−0.013 (0.576)	−0.556 (0.591)	−0.558 (0.615)	−0.071 (0.349)	1.288*** (0.428)
HC	−0.378* (0.229)	−0.359 (0.214)	−0.317 (0.234)	0.422** (0.172)	0.272 (0.183)

关联CSMAR子库：新质生产力 / 人工智能 / 数字经济 / 宏观经济 / 区域经济 /…

数实融合

CSMAR与前沿主题：（三）加快发展数字经济研究

相关研究4：企业数字化与ESG履责[J].会计研究, 2024.

文章基于ESG理论框架，以2011-2021年中国A股上市公司为样本，探究企业数字化对ESG履责的赋能影响。研究表明，企业数字化能够显著促进ESG履责表现。机制检验发现，企业数字化通过改善绿色创新能力、社会责任契约履行和公司治理效能来有效赋能ESG履责。

$$ESG_{i,t+1}=\alpha_0+\alpha_1 LnDT_{i,t}/PerDT_{i,t}+\alpha_m Control_{i,t}+Year+Industry+\mu_{i,t}$$

变量名称	变量符号	变量定义
ESG 履责	*ESG*	采用华证 ESG 评级来度量，具体对该评级数据由高至低（“AAA-C”）分别赋值为 9-1 分，并将每年四个季度的赋值分取平均值
企业数字化	*LnDT*	公司年报文本内容中数字化总词频数加 1 的自然对数
	PerDT	公司年报文本内容中数字化总词频数除以年报文本词汇总数量×100
公司规模	*LnSize*	年末总资产的自然对数
总资产收益率	*ROA*	净利润/年末总资产
负债水平	*Debt*	年末总负债/年末总资产
成长性	*Growth*	营业收入增长率
股权集中度	*Fshare*	第一大股东持股比例
股权制衡度	*NCLS*	第二到第五位大股东持股比例之和/第一大股东持股比例
管理层持股	*MShare*	高管持股比例
董事会规模	*DirSize*	董事会人数的自然对数
独立董事比例	*Indep*	独立董事占董事会人数的比例
两职兼任	*Dual*	若 CEO 兼任董事长取值为 1，否则为 0
经营现金流	*CFO*	经营活动现金流量净额/年末总资产
亏损状况	*Loss*	若公司净利润为负则取值为 1，否则为 0
产权性质	*SOE*	若公司所有权性质为国有企业则取值为 1，否则为 0
公司年龄	*LnAge*	公司上市年限加 1 的自然对数
事务所类型	*Big4*	若公司年报审计的会计师事务所为国际“四大”则取值为 1，否则为 0
违规处罚	*Punish*	若公司被监管处罚则取值为 1，否则为 0
诉讼风险	*Litigation*	若公司有诉讼仲裁情况则取值为 1，否则为 0
信息透明度	*Abs_DA*	为业绩调整的 Jones 模型估计残差的绝对值（Kothari 等，2005）

关联CSMAR子库：商道融绿ESG / 企业数字化转型 / 数字经济 / 财务报表 / 财务报表附注 / 财务指标分析 / 治理结构 / …

企业数字化对 ESG 履责的回归结果

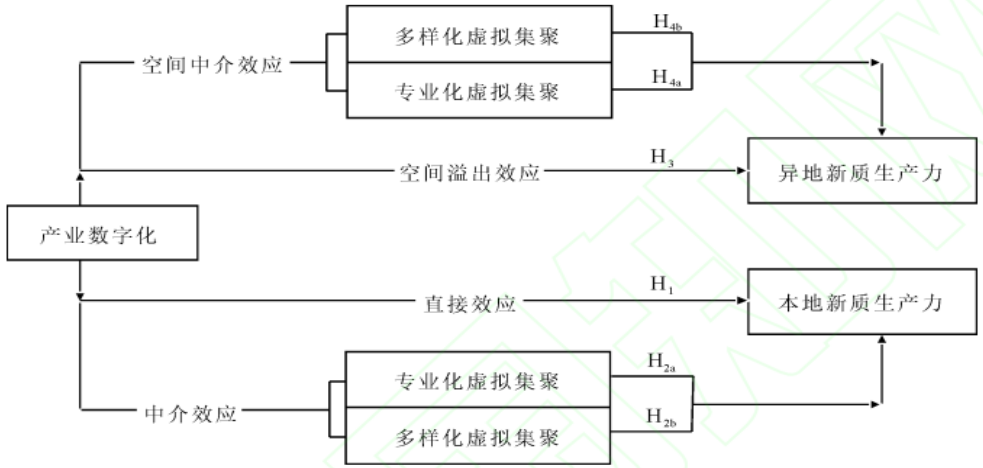
Variable	(1)	(2)	(3)	(4)
	ESG	*ESG*	*ESG*	*ESG*
LnDT	0.027***		0.015***	
	(4.48)		(2.95)	
PerDT		0.652***		0.648***
		(2.61)		(3.10)
Panel A：基于绿色创新的机制检验				
Variable	(1)	(2)	(3)	(4)
	Ginnov	*ESG*	*Ginnov*	*ESG*
LnDT	0.077***	0.011**		
	(16.79)	(2.11)		
Ginnov		0.070***		0.070***
		(8.94)		(8.94)
PerDT			2.837***	0.496**
			(14.12)	(2.35)
Panel B：基于社会责任的机制检验				
	(5)	(6)	(7)	(8)
	CSR	*ESG*	*CSR*	*ESG*
LnDT	0.008***	0.010*		
	(3.01)	(1.92)		
CSR		0.568***		0.568***
		(43.25)		(43.25)
PerDT			0.345***	0.415**
			(3.35)	(2.00)
Panel C：基于公司治理的机制检验				
	(9)	(10)	(11)	(12)
	Gove	*ESG*	*Gove*	*ESG*
LnDT	0.021***	0.014***		
	(4.54)	(2.77)		
Gove		0.036***		0.036***
		(4.85)		(4.89)
PerDT			0.478**	0.619***
			(2.47)	(2.96)

数实
融合

CSMAR与前沿主题：（三）加快发展数字经济研究

相关研究5：产业数字化、虚拟集聚与新质生产力发展[J].科技进步与对策, 2025.

文章基于新质生产力理论和虚拟集聚理论，选取2015—2022年中国内地30省份数据，研究产业数字化对新质生产力的影响效应、作用机制与空间溢出效应。研究发现：产业数字化能够显著驱动本地新质生产力发展；多样化虚拟集聚和专业化虚拟集聚是产业数字化驱动本地新质生产力发展的重要作用机制；产业数字化对驱动新质生产力发展发挥空间溢出效应，多样化虚拟集聚是重要的空间作用机制；④产业数字化对其它地区专业化虚拟集聚的影响不显著。



理论框架

虚拟集聚是指产业链供应链上下游企业由传统地理集聚转向空间虚拟集聚的模式,能够有效缩短链上 各环节间的信息交换距离并降低虚拟要素交易成本, 具有跨产业、跨地区和泛边界特征。空间虚拟性和贸易数字化是虚拟集聚的本质特征，电子商务贸易可映射地区虚拟集聚水 平。文章使用区位熵方法测度专业化虚拟集聚指数(MAR),使用 HHI指数的倒数测度多样化虚拟集聚指数(JAC)：

$$MAR_{jkt} = \sum_j \sum_k \frac{(q_{jkt} \times \delta)/q_{kt}}{Q_{jt}/Q_t} \qquad JAC_{jkt} = \frac{1}{\sum_j \sum_k |(q_{jkt} \times \delta)/q_{kt} - Q_{jt}/Q_t|}$$

$$NQF_{it} = \alpha + \beta ID_{it} + \gamma \sum X_{i,t} + \delta_i + \mu_t + \varepsilon_{i,t}$$

变量	变量名称
NQF	新质生产力
ID	产业数字化
MAR	专业化虚拟集聚
JAC	多样化虚拟集聚
FAG	金融集聚
DA	数字关注
OPEN	对外开放
ER	环境规制
GOV	政府干预

目标层	指标层	衡量方式
产业数字化	智慧农业	农村宽带用户占比
		开通互联网业务的行政村比重
	智能制造	新产品销售收入/规模以上工业企业主营业务收入
		规上工业企业当年累计 R&D 项目数
		规上工业企业 R&D 经费
		两化融合指数
	智能交通	省道拥挤度增加量
	智慧物流	快递量/装卸搬运和仓储业就业人员数
		快递业务收入/装卸搬运和仓储业就业人员数
	智慧服务	每百家企业拥有网站数
	数字治理	政务一体化指数

基准回归结果

变量	新质生产力		
	(1)	(2)	(3)
ID	0.697***	0.367***	0.202**
	(20.86)	(3.29)	(2.05)

虚拟集聚机制检验结果

变量	专业化虚拟集聚	新质生产力	多样化虚拟集聚	新质生产力
	(1)	(2)	(3)	(4)
ID	2.785***		2.431**	
	(4.27)		(2.23)	
MAR		0.091***		
		(10.64)		
JAC				0.022***
				(3.63)

关联CSMAR子库：新质生产力 / 人工智能 / 数字经济 / 宏观经济 / 区域经济 /…

数实融合

CSMAR与前沿主题：（三）加快发展数字经济研究

相关研究6：数字产业化能否增强城市产业链韧性？[J].现代经济探讨，2024.

文章基于283个地级市2013-2020年面板数据,实证研究数字产业化对城市产业链韧性的影响效应及其内在机制。研究发现：数字产业化可以显著提高城市产业链韧性，城市创新能力、产业结构高级化是数字产业化提升产业链韧性的作用机制。

关联CSMAR子库：数字经济 / 人工智能 / 人工智能产业链 / 宏观经济 / 企业创新 / 经济内循环 / 专利 / …

被解释变量：城市产业链韧性		一是产业链应对外部冲击和抗干扰能力：采用赫芬达尔指数衡量产业的多样化水平 二是产业链遭遇外部冲击后的恢复能力：采用专利授权数量衡量产业链恢复能力 采用熵权法将上述两个指标标准化后降维处理得到产业链韧性指数
解释变量：数字产业化水平		从数字基础发展水平、数字产品服务业发展水平和数字产品制造业发展水平三个维度来衡量数字产业化（用每百人互联网宽带接入用户数和每百人移动电话用户数表征数字基础发展水平；用人均电信业务总量表征数字产品服务业发展水平；用计算机服务和软件业从业人员占城镇单位从业人员比重表征数字产品制造业发展水平）
中介变量	城市创新能力	科技投入占财政支出的比重
	产业结构高级化	高技术产业就业人数占总就业人数的比重

变量	(1)	(2)	(3)
	Chain_re	*Chain_re*	*Chain_re*
Dig_ind	0.283*** (0.035)	0.220** (0.072)	0.135*** (0.034)

	(1)	(2)	(3)
Plan A:城市创新能力			
变量	*Chain_re*	*Innov*	*Chain_re*
Innov			0.363** (0.123)
Dig_ind	0.135*** (0.034)	0.057*** (0.005)	0.001 (0.038)
Bootstrap 检验	1.99 [0.008 0.116]		
效应占比	完全中介(100%)		
Plan B:产业结构高级化			
变量	*Chain_re*	*High_te*	*Chain_re*
High_te			3.151*** (0.068)
Dig_ind	0.135*** (0.034)	0.060*** (0.011)	0.032 (0.035)

CSMAR与前沿主题：（三）加快发展数字经济研究

相关研究7： “多言寡行”的数字化转型披露能否被投资者识别？——基于股票错误定价的证据[J].外国经济与管理，2025.

文章利用2006—2021年A股企业数据，探讨了“多言寡行”的数字化转型披露对股票错误定价的影响。研究发现，“多言寡行”的数字化转型披露体现了企业的市值操纵动机。这种披露策略在短期内难以被投资者识别，提高了投资者对企业股票的错误定价，使股价向上偏离基本面价值。机制检验表明，“多言寡行”的数字化转型披露策略会使投资者情绪高涨，进而导致错误定价。

$$Digital_Text_{i,t} = \beta_0 + \beta_1 Digital_Invest_{i,t} + \beta_2 Digital_Invest_{i,t-1} + \beta_3 Ind_Digital_Text_{i,t} + \sum Firm_controls_{i,t} + Industry_{i,t} + Province_{i,t} + Year_{i,t} + \varepsilon_{i,t}$$

变量类别	变量名称	变量符号	变量定义
被解释变量	错误定价	*Deviation*	使用剩余收益估值法(RIM)计算得到
解释变量	多言寡行的数字化转型披露	*Digital_ab*	取式(2)的回归残差。*Digital_ab*越大说明企业数字化转型披露“多言寡行”越严重。
与解释变量相关变量	年报中数字化文本披露	*Digital_Text*	年报中的“言”，即数字化相关关键词在年报中披露的次数，加1取对数处理。
	年报中数字化实际投入	*Digital_Invest*	年报中的“行”，即数字化资产在企业总资产池中的占比
控制变量	公司规模	*Size*	年总资产的自然对数
	资产负债率	*Lev*	年末总负债除以年末总资产
	净资产收益率	*ROE*	净利润/股东权益平均余额
	是否国企	*SOE*	国有控股企业取值为1，其他为0
	账面市值比	*BM*	账面价值/总市值
	营业收入增长率	*Growth*	本年营业收入/上一年营业收入-1
	是否亏损	*Loss*	当年净利润小于0取 1，否则取0
	上市年限	*ListAge*	*ln*(当年年份-上市年份+1)
	月均超额换手率	*Dturn*	当年股票月均换手率—去年股票月均换手率
	机构投资者持股比例	*INST*	机构投资者持股总数除以流通股本
	第一大股东持股比例	*Top1*	第一大股东持股数量/总股数
	独立董事比例	*Indep*	独立董事占比
	是否四大	*Big4*	公司经由四大(普华永道、德勤、毕马威、安永)审计为1，否则为0。
固定效应	行业	*Industry*	证监会2012年行业分类，制造业取两位代码，其他行业用大类代码
	省份	*Province*	企业注册地所在省份
	年份	*Year*	年份固定效应

关联CSMAR子库： 数字经济 / 企业数字化转型 / 财务报表 / 财务报表附注 / 财务指标分析 / 股票市场交易 / 治理结构 / 投资者情绪/机构投资者 / …

$$Digital_Text_{i,t} = \beta_0 + \beta_1 Digital_Invest_{i,t} + \beta_2 Digital_Invest_{i,t-1} + \beta_3 Ind_Digital_Text_{i,t} + \sum Firm_controls_{i,t} + Industry_{i,t} + Province_{i,t} + Year_{i,t} + \varepsilon_{i,t}$$

Digital_Text是企业数字化文本披露强度（“言”），用年报当中相关关键词词频表示；数字化实际投入强度Digital_Invest（“行”）通过查阅企业资产负债 表明细项，手工统计企业财务报表的资产中与数字化有关的投资进行测度。

基准回归结果

	(1) *Deviation*	(2) *Deviation*	(3) *Deviation*	(4) *Deviation*
Digital_ab	1.939***	1.986***	1.982***	1.991***
	(6.328)	(6.665)	(6.740)	(6.803)

投资者情绪的中介机制

	Panel A 中介机制		Panel B 中介机制稳健性	
	(1) *Sentiment*	(2) *Sentiment*	(3) *Deviation*	(4) *Deviation*
Digital_ab	2.219***	1.451***		
	(7.361)	(4.995)		
Sentiment			0.286***	0.175***
			(32.084)	(27.435)

网络检索强度和媒体正面报道

	Panel A 中介机制		Panel B 中介机制稳健性		Panel C 中介机制		Panel D 中介机制稳健性	
	(1) *baidu_index*	(2) *baidu_index2*	(3) *Deviation*	(4) *Deviation*	(5) *PNC*	(6) *OriginalPNC*	(7) *Deviation*	(8) *Deviation*
Digital_ab	4.906***	4.659***			6.506***	6.442***		
	(7.400)	(10.182)			(11.914)	(13.831)		
baidu_index			0.024***					
			(7.205)					
baidu_index2				0.048***				
				(9.892)				
PNC							0.039***	
							(11.232)	
OriginalPNC								0.044***
								(11.080)

CSMAR与前沿主题：（三）加快发展数字经济研究

相关研究8：数字金融与企业金融化：“脱实向虚”还是“脱虚向实”？[J]. 财经理论与实践, 2024.

文章2011—2021年沪深A股上市公司面板数据,构建双向固定效应模型考察了数字金融对企业金融化的影响及机制。结果显示：数字金融增强了逐利性动机和预防性动机、降低了期间费用和融资约束，总体上促进了企业金融化，尤其对非国有、非制造业、东部地区与中部地区、金融监管前的企业而言，这种促进作用更为突出。

$$Financialization_{i,t} = \delta_0 + \alpha_1 Finance_{i,t} + \sum \lambda_i X_{i,t} + \varepsilon_{i,t}$$

关键解释变量	被解释变量
数字金融：北京大学数字普惠金融指数	企业金融化：金融资产投资占比衡量企业金融化

$$Financialization_{i,t} = \theta_0 + \theta_1 Finance_{i,t} + \sum \theta_i X_{i,t} + \varepsilon_{i,t}$$

$$M_{i,t} = \beta_0 + \beta_1 Finance_{i,t} + \sum \beta_i X_{i,t} + \delta_{i,t}$$

$$Financialization_{i,t} = \gamma_0 + \gamma_1 Finance_{i,t} + \gamma_2 M_{i,t} + \sum \gamma_i X_{i,t} + \mu_{i,t}$$

预防性动机(Flow)：经营性现金净流量 与企业总资产的比值
逐利性动机 (Profit)：净资产收益率
期间费用(PE)：管理费用、财务费用以及销售费用三者之和
融资约束(FC)：SA 指数

基准回归、稳健性检验及 MQ 指标

变量	(1) Financialization	(2) Financialization	(3) MQ	(4) IV Financialization	(5)L1. Financialization	(6)L2. Financialization	(7) 面板交互 Financialization
Finance	1.820***	1.380***	25.44	0.691**			1.350***
	(7.94)	(6.17)		(2.19)			(5.87)

机制检验

变量	预防性动机、逐利性动机					期间费用、融资约束			
	(1) Financialization	(2) Flow	(3) Financialization	(4) Profit	(5) Financialization	(6) PE	(7) Financialization	(8) FC	(9) Financialization
Finance	1.380***	0.007***	1.363***	0.007***	1.369***	−0.011**	1.400***	−0.057***	1.407***
	(6.17)	(4.07)	(6.08)	(2.97)	(6.11)	(−2.14)	(6.25)	(−7.66)	(6.25)
Flow			2.315**						
			(2.14)						
Profit					1.612*				
					(1.94)				
PE							1.808		
							(1.58)		
FC									0.482*
									(1.75)

关联CSMAR子库：数字经济 / 财务报表 / 财务指标分析 / 经营困境 / 治理结构 /…

CSMAR与前沿主题：（三）加快发展数字经济研究

相关研究9：数字技术使用更有助于灵活就业人员收入增长吗?——基于二级数字鸿沟视角[J].软科学, 2024.

文章依据中国家庭追踪调查（CFPS）2016、2018、2020 三年数据，考察了二级数字鸿沟视角下的数字技术使用对灵活就业与正规就业群体收入、收入差距产生的影响及存在的作用机制。结果显示，数字技术使用对两群体收入均有正向影响，且对正规就业的影响大于灵活就业，扩大了二者的收入差距；随着收入水平的提高，数字技术对二者收入差距的影响逐渐由扩大转为缩小；机制分析表明，就业结构的改变与数字技术使用偏好是产生收入差距的主要原因。

$$\mathrm{Ln}income_{it} = \alpha_1 + \beta_1 Digit_{it} + \eta_1 Flex_{it} + \gamma_1 X_{it} + vIMR_{it} + \pi_1 year_t + \varepsilon_1$$

变量类型	变量名称	英文代号	变量说明
被解释变量	个人收入对数	income	个体年收入（元）的对数
解释变量	数字技术使用	digit	K－means 聚类划分为 4 级
控制变量	受教育年限	edu	已完成的受教育年数
	是否灵活就业	flex	灵活就业＝1、非灵活就业＝0
	工作经验	exp	根据年龄和受教育年限测算
	性别	gender	女性＝1、男性＝0
	户籍	urban	非农业＝1、农业＝0
	婚姻状况	marriage	未婚＝0、已婚＝1
	健康状况	health	健康程度由低到高赋值 1～5
	语言使用	language	使用普通话＝1、未使用＝0
	政治面貌	party	党员＝1、非党员＝0

①上网时间，根据 CFPS问 卷中“每周上网时间”得出。②各类数字技术使用能力，根据CFPS问卷中“是否使用互联网学习／娱乐／购物／社交”得出。③各类数字技术使用频率。根据CFPS问卷中“每周互联网学习的频率”“是否每天互联网学习”等得出。

基准回归与处理效应模型结果

	OLS 回归			处理效应模型		
	(1)	(2)	(3)	(4)	(5)	(6)
			income			flex
digit	0.101***	0.053***	0.059***	0.056***	0.062***	
	(0.005)	(0.005)	(0.006)	(0.005)	(0.006)	
flex		−0.177***	−0.147***	−0.186***	−0.155***	
		(0.009)	(0.019)	(0.009)	(0.020)	
digit × flex			−0.014*		−0.014*	
			(0.009)		(0.009)	

数字技术使用对就业结构的影响

	(1)	(2)	(3)	(4)	(5)	(6)
	总体	正规就业		灵活就业		
		创业	一般受雇	创业	一般受雇	散工
digit	−0.081***	0.148	0.125**	0.116***	−0.013	−0.058**
	(0.011)	(0.124)	(0.056)	(0.032)	(0.023)	(0.027)

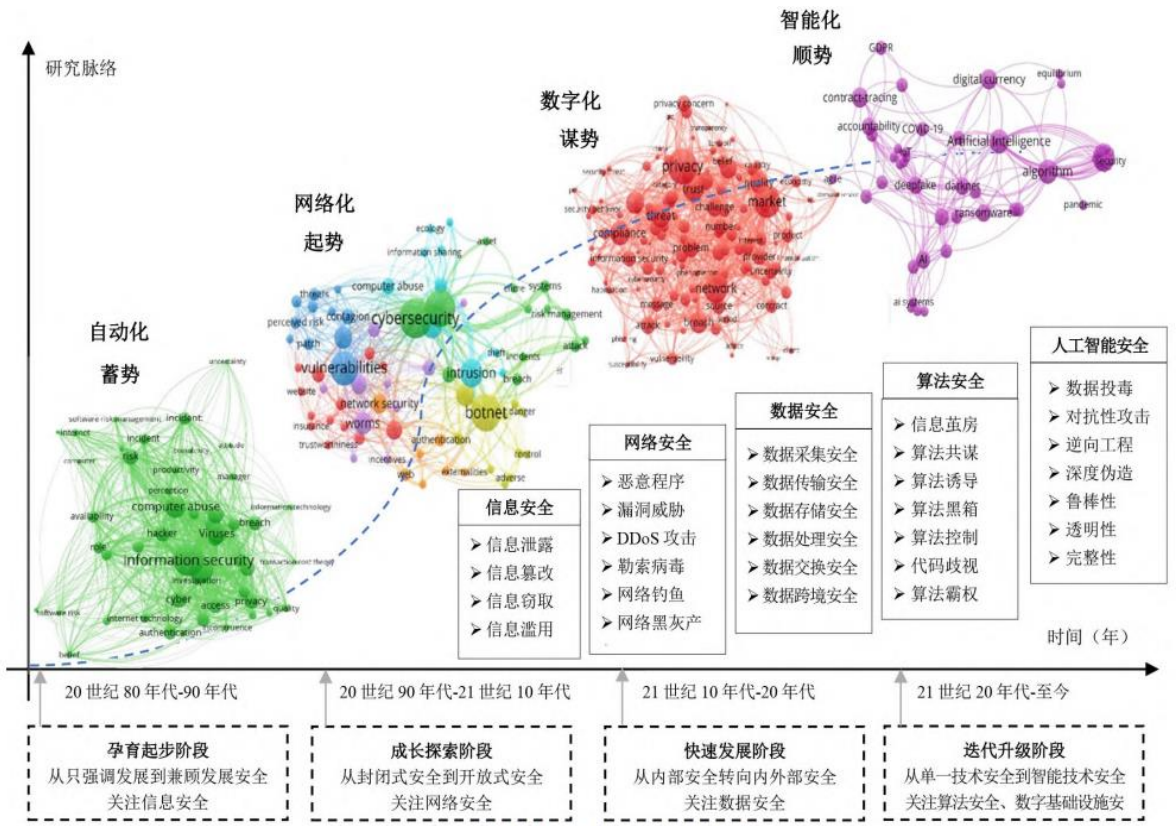
数字技术使用不同偏好对收入差距的影响

变量	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				income				
flex		−0.169***		−0.122***		−0.186***		−0.137***
		(0.011)		(0.018)		(0.016)		(0.013)
stu	0.092***	0.090***						
	(0.010)	(0.012)						
stu × flex		−0.021						
		(0.018)						
soc			0.134***	0.161***				
			(0.011)	(0.016)				
soc × flex				−0.067***				
				(0.020)				
pla					0.081***	0.064***		
					(0.010)	(0.015)		
pla × flex						0.010		
						(0.019)		
bus							0.118***	0.137***
							(0.010)	(0.012)
bus × flex								−0.065***
								(0.017)

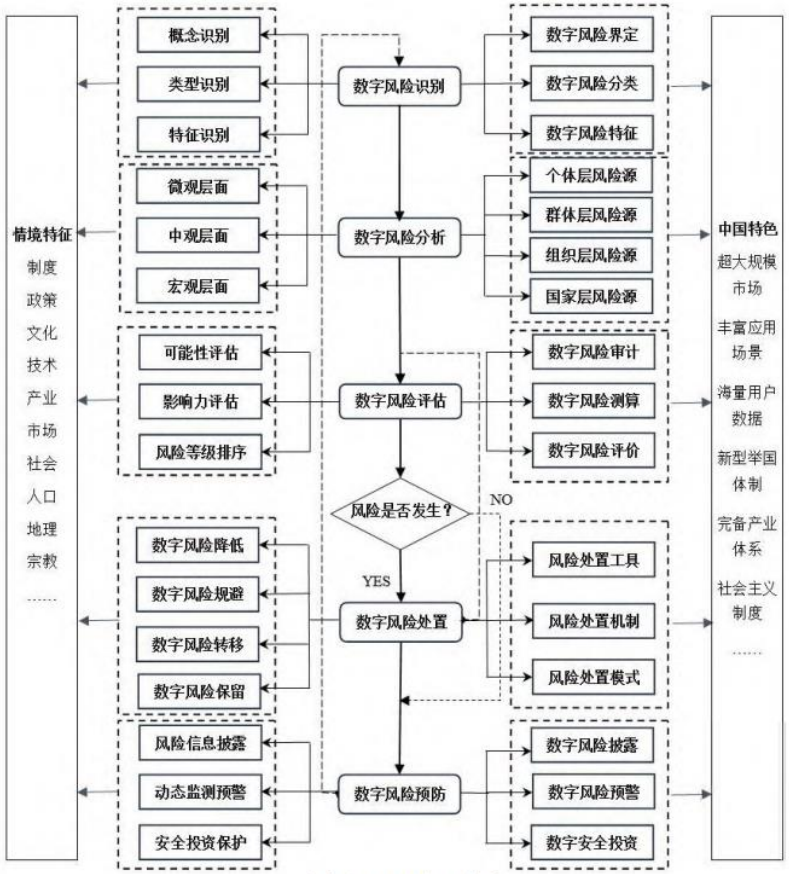
CSMAR与前沿主题：（三）加快发展数字经济研究

相关研究10：数字风险治理：研究脉络、理论框架及未来展望[J]. 管理世界, 2024.

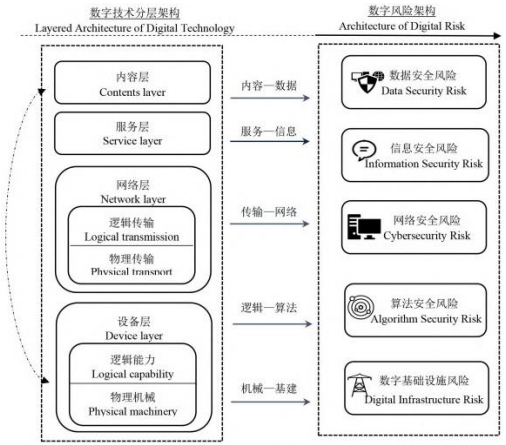
文章基于数字技术视角和流行病学模型识别数字风险的内涵特征，构建结构化、系统化的理论框架以增进学术对话，为我国数字经济持续健康发展提供理论指引可以预警金融市场中可能发生的风险，帮助投资者规避损失。



数字风险治理研究的知识图谱



数字风险治理的理论框架



传统风险与数字风险特征比较

	传统风险	数字风险
传播性质	集中式和中心化	分布式和去中心化
传播路径	单边外部效应 ·低速度扩散 ·轻量级增长 ·线性传播 ·稳态与独立	网络外部效应 ·高速率蔓延 ·指数级增长 ·非线性传播 ·变异与耦合
传播方式	人工发起	智能化攻击
传播可控性	不可逆性	持续自适应

CSMAR与前沿主题：（三）加快发展数字经济研究

相关研究11：ESG 视角下人工智能大模型风险识别与治理模型[J]. 中国科学院院刊, 2024.

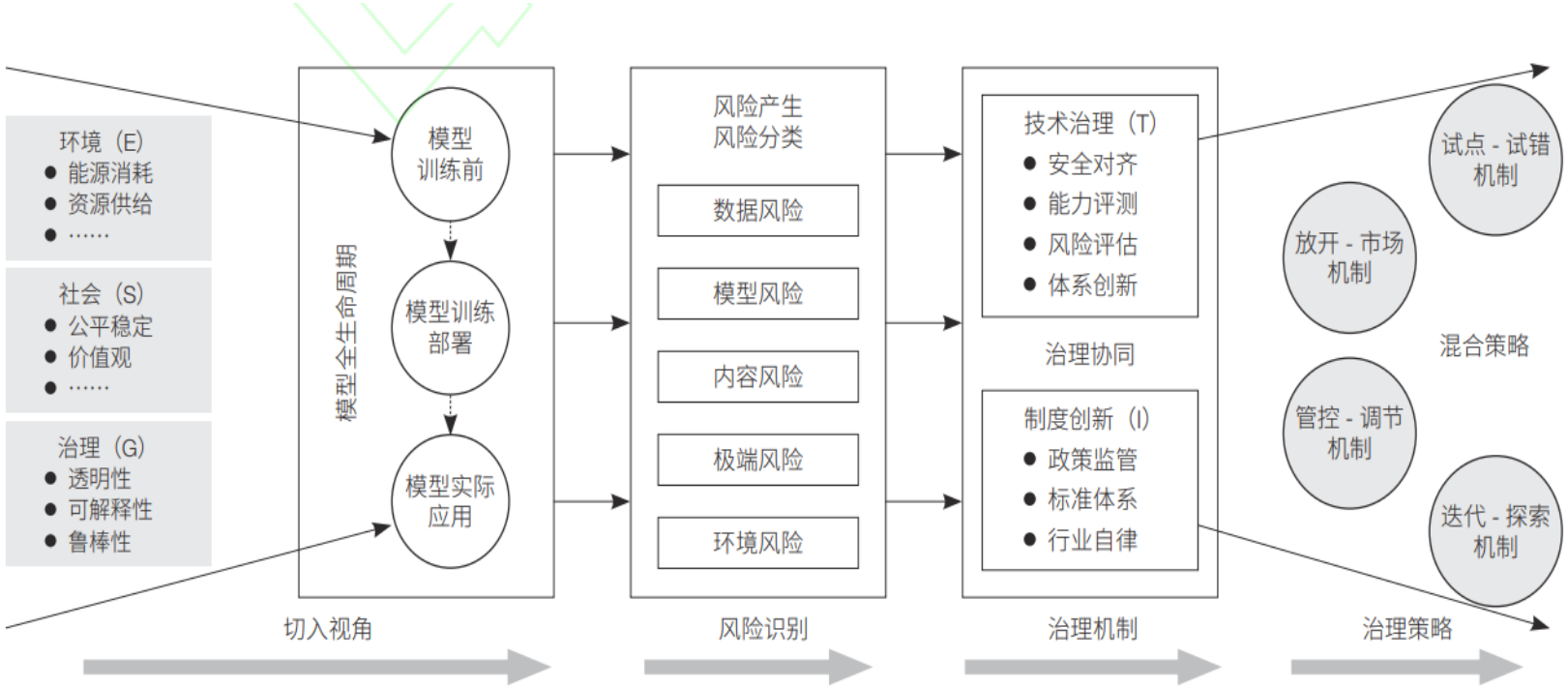
文章基于ESG（环境—社会—治理）和人工智能治理相关理论，分析ESG视角下人工智能大模型发展效益、典型风险，进而构建人工智能大模型风险治理框架和实施策略。

ESG视角下大模型发展效益

层面	维度	内容
环境	提高能源效率	自动化处理大量能源数据
		人工智能驱动的能源解决方案
	环境模拟预测	气候监测、分析和预测
		增强能源决策能力
	跟踪能源审计	人工智能辅助的能源绩效评级
		能源资源分配和调整
社会	提高生产效率	减少任务处理的时间
		替代重复性任务，使人类专注更具创造性的工作
	赋能组织创新	决策范式：基于大量客观数据发掘新洞见
		设计范式：人工智能驱动的设计模型
		产品开发：人工智能增强的“钻石模型”
	促进社会变革	改变知识重组和知识传播方式
治理	促进跨部门协同	打破部门间信息壁垒
	监控和合规自动化	大模型赋能组织合规体系建设
	治理体系革新	推动政府、企业等组织深入数字化转型

ESG视角下大模型的负面影响

层面	维度	内容
环境	研发训练	数据中心用水需求增长
		大量算力造成的电力资源消耗
		碳排放
	推理应用	在大量场景应用、大规模用户使用中确保持续运转的资源消耗
社会	数据隐私	大模型数据集获取处理可能存在恶意使用或泄漏风险
	偏见和歧视	大模型数据包容性不足，存在社会偏见和歧视问题
	安全风险和攻击	前沿模型带来的社会风险，如特定虚假信息误导
		前沿模型带来的产业风险，如在国计民生的领域应用
		利用大模型的安全漏洞来获取敏感信息
	可解释性和透明度	大模型自身的可解释性和透明度不足，决策机理和实现路径有待明晰
治理	治理主体	各利益攸关方对人工智能的危险和益处有不同的看法
		人工智能政策与其他领域相互关联交叉
	治理周期	人工智能技术周期与政策周期面临匹配挑战
	治理方式	面临从道德走向政策制定的挑战
		面临从理念走向技术治理实践的挑战



人工智能大模型风险治理模型

CSMAR与前沿主题：（三）加快发展数字经济研究

相关研究12：人工智能模型训练中合成数据的应用风险及其治理路径[J]. 情报理论与实践, 2025.

合成数据是人工智能训练数据的重要类型，随着其在人工智能模型训练中的深度应用催生了多重风险，严重影响人工智能模型的健康发展，因此有必要从治理现状入手探索应用风险的治理路径。通过综合运用文献分析、案例分析、定性分析等方法，阐明合成数据在人工智能训练中的应用风险样态以及现有治理体系的不足，并在转变治理模式的基础上提出具体的治理路径。当前，合成数据在人工智能模型训练中面临质量风险、安全风险和滥用风险，而命令控制模式下的治理体系在制度构建、技术治理、主体协同三大维度上存在明显不足。对此，亟须转向敏捷治理模式，在新模式下通过构建软硬兼施的制度体系、引入技术手段推动灵活治理、优化多元主体协同治理机制，从而形成系统性的合成数据治理体系。

合成数据在人工智能模型训练中的应用价值

合成数据凭借其独特的属性和优势已被广泛运用于人工智能模型训练当中，具有超越真实数据的诸多价值。

- 第一，即时生成海量数据，破解数据枯竭难题。
- 第二，简化数据利用 流程，降低数据处理成本。
- 第三，强化隐私保护机制，解决隐私合规问题。
- 第四，矫正数据集不平衡，增 强数据多样性。

VS

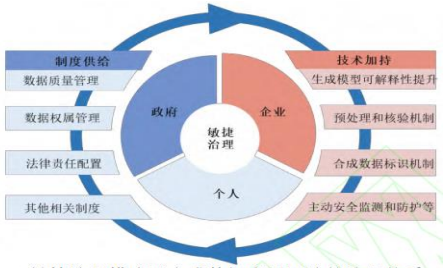
合成数据应用于人工智能模型训练的风险样态

随着合成数据在人工智能模型训练当中的深入应用，多元化的风险也日益显现。这些风险不仅会影响到人工智能模型的训练效果，还可能侵害到人身财产权益并对社会安全构成潜在威胁。

- 合成数据质量风险：数据准确性不高、数据完整性缺失、数据偏见扩散
- 合成数据安全风险：多轮次使用合成数据，可能引发模型崩溃；委托第三方生成合成数据的行为，将引发数据泄露；合成数据所涉利益主体多元，进而引发权属争议
- 合成数据滥用风险：利用合成数据训练违法模型；利用合成数据技术间接复制真实数据集的行为，将引发不正当竞争；，利用隐含错误信息的合成数据攻击其他模型

现有合成数据综合治理体系

层面	治理领域	治理工具
数据治理	数据安全管理、数据隐私保护、网络数据管理、数据知识产权保护、数据流通和交易管理、数据跨境流动管理、数据侵权救济与刑事保护、个人数据管理	法律责任、监管机制、分类分级、技术标准、监测预警机制、访问控制与权限管理、安全审计、数据加密与脱敏、匿名化和去标识化等
算法治理	算法透明度和公平性管理、算法合规性评估、算法偏差矫正、算法自动化决策管理	法律责任、监管机制、算法备案制度、算法分类机制、算法风险识别与预警系统、算法安全评估机制等
人工智能治理	人工智能伦理管理、人工智能研发和应用合规、人工智能安全管理、深度合成管理、技术滥用防控	法律责任、监管机制、人工智能伦理指南、模型风险评估框架、模型合规审查机制、模型溯源管理、监管沙盒机制、合成数据内容标识等



敏捷治理模式下合成数据应用风险的治理体系

CSMAR与前沿主题：（三）加快发展数字经济研究

相关研究13：DeepSeek的破局与变局：论生成式人工智能的监管方向[J].新疆师范大学学报（哲学社会科学版），2025.

随着生成式人工智能的快速发展，DeepSeek的问世与开源备受关注。DeepSeek凭借对算法模式、知识蒸馏以及思维链技术的破局，实现了低成本训练与高效输出。技术的变革对生成式人工智能的监管具有两面性：一方面，思维链和模型开源缓解了“算法黑箱”困境，为监管审查提供便利；另一方面，对隐私保护和数据合规、知识产权、“幻觉问题”以及模型安全性隐患等带来新挑战。DeepSeek 的爆火再度说明，生成式人工智能的监管应面向未来，将隐私保护与数据安全、技术创新与保护平衡、模型提示与反馈机制、模型安全性作为监管重点，避免技术滥用与恶意攻击，保障并促进生成式人工智能安全、良性发展。

DeepSeek技术变革带来的新挑战

1. **隐私保护和数据合规挑战加剧。**生成式人工智能的数据合规主要包括数据来源、数据处理、数据存储和传输等环节。DeepSeek的思维链和知识蒸馏技术为模型性能提升带来突破，但也加剧了隐私保护和数据合规的挑战。
2. **知识产权方面的争议。**知识蒸馏技术在实现低成本高效训练的同时，或将引发一系列知识产权争议。知识蒸馏的本质是利用 “教师模型”的训练效果。有观点认为，将蒸馏技术 用于构建直接竞争的大模型产品，属于违反其服务条款的行为；也有观点认为，利用先进大模型的输出进 行二次开发属于行业惯例，并不存在争议。
3. **“幻觉问题”加重。**大模型的“幻觉问题”是所有模型的固有弊端，“幻觉问题”指模型生成的内容看似准确，但实际上是其编造的或者缺乏数据支撑，模型的“幻觉问题”严重影响模型输出内容的可靠性，进而影响社会信任。随着 DeepSeek模型蒸馏技术及思维 链技术的应用，虽然逻辑推理能力得到提升，但 “幻觉问题”逐渐凸显并加重。例如，在实际使用过程中，有使用者发现DeepSeek大模型在思维链 推导过程和最后结果中输出虚假和编造的文件、页码、专业名词等，且模型表达内容、逻辑推演实施以及细节过于完善，普通使用者的知识储备难以纠错，即使使用者指出具体错误，模型仍会对相关内容进行再“编造”。
4. **模型安全性问题。**DeepSeek模型发布后，曾遭大量境外网络攻击，直接影响系统的注册、访问和正常使用。此次攻击是生成式人工智能模型问世以来罕见的大规模跨境攻击，此次攻击事件暴露出一系列典型的网络安全问题，包括攻击手段复杂化、跨境攻击、网络安全防护技术漏洞、应急响应能力不足、数据安全风险以及供应链安全风险等等。

生成式人工智能监管的重点方向

生成式人工智能监管问题的困境之一是监管措施要不断适应技术的飞速发展。有学者认为，隐私保护、模型安全、技术标准、开源协议等更宜采用灵活多样、合作试验、跨国适用的行业标准等“软 法”加以规制，监管机关宜与企业、行业组织或企业合作联盟等开展广泛合作，通过多种机制促进 “软法”与“硬法”结合，规范生成式人工智能的发展。

CSMAR与前沿主题：（四）新质生产力

新质生产力：创新起主导作用，摆脱传统经济增长方式、生产力发展路径，具有高科技、高效能、高质量特征，符合新发展理念的先进生产力质态。

与传统生产力的**区别**

新：创新属性强，不断裂变催生新产业、新业态、新模式

质：先进生产力质态，摆脱传统生产力发展中的粗放模式，以绿色发展为普遍形态

影响因素

数字化
金融供给地理结构
人力资本
……

新质生产力

测度：宏观层面、微观层面

高质量发展

低碳经济
产业链韧性
共同富裕
……

CSMAR与前沿主题：（四）新质生产力

相关研究1：新质生产力发展水平、区域差异与提升路径[J]. 重庆大学学报(社会科学版), 2024.

文章基于新质生产力的理论内涵与现实价值，从科技生产力、绿色生产力和数字生产力3个一级指标构建了新质生产力的综合评价体系。

一级	二级	三级	解释
科技生产力	创新生产力	创新研发	国内专利授予数
		创新产业	高技术产业业务收入
		创新产品	规上工业企业产业创新经费
	技术生产力	技术效率	规上工业企业劳动生产率
		技术研发	规上工业企业 R&D 人员全时当量
		技术生产	机器人安装原始密度
绿色生产力	资源节约型生产力	能源强度	能源消费量/国内生产总值
		能源结构	化石能源消费量/国内生产总值
		用水强度	工业用水量/国内生产总值
	环境友好型生产力	废物利用	工业固废物综合利用量/产生量
		废水排放	工业废水排放/国内生产总值
		废气排放	工业 SO2排放/国内生产总值
数字生产力	数字产业生产力	电子信息制造	集成电路产量
		电信业务通讯	电信业务总量
	产业数字生产力	网络普及率	互联网宽带接入端口数
		软件服务	软件业务收入
		数字信息	光缆线路长度/地区面积
		电子商务	电子商务销售额

➢ **CSMAR-经济研究系列-新质生产力**

地区新质生产力三级指标表(科技、绿色、数字生产力角度)	统计年度、地区代码、地区名称、创新研发专利数、创新产业业务收入、创新产品经费、技术效率、技术研发人数、技术生产、能源强度、能源结构、用水强度、废物利用率、废水排放、废气排放、电子信息制造、电信业务通讯、网络普及率、软件服务费用、数字信息、电子商务。
地区新质生产力二级指标表(科技、绿色、数字生产力角度)	统计年度、地区代码、地区名称、创新生产力、技术生产力、资源节约型生产力、环境友好型生产力、数字产业生产力、产业数字生产力。
地区新质生产力水平表(科技、绿色、数字生产力角度)	统计年度、地区代码、地区名称、科技生产力、绿色生产力、数字生产力、地区新质生产力水平。

CSMAR与前沿主题：（四）新质生产力

相关研究2： ESG发展对企业新质生产力影响的研究——来自中国A股上市企业的经验证据[J].当代经济管理, 2024.

文章基于 2015—2022 年上市企业财务报表数据，实证研究了 ESG 发展对企业新质生产力的影响，结果显示：ESG 发展对企业新质生产力水平的提升有显著促进作用。

因素	子因素	指标	指标取值说明
劳动力	活劳动	研发人员薪资占比	研发费用-工资薪酬/营业收入
		研发人员占比	研发人员数/员工人数
		高学历人员占比	本科以上人数/员工人数
	物化劳动（劳动对象)	制造费用占比	（经营活动现金流出小计+固定资产折旧+无形资产 摊销+减值准备-购买商品接受劳务支付的现金-支付给职工以及为职工支付的工资）/（经营活动现金 流出小计+固定资产折旧+无形资产摊销+减值准备）
生产工具	硬科技	研发折旧摊销占比	研发费用-折旧摊销/营业收入
		研发租赁费占比	研发费用-租赁费/营业收入
		研发直接投入占比	研发费用-直接投入/营业收入
		无形资产占比	无形资产/资产总额
	软科技	总资产周转率	营业收入/平均资产总额
		权益乘数倒数	所有者权益/资产总额

➢ **CSMAR-经济研究系列-新质生产力**

企业新质生产力三级指标表(劳动力、生产工具角度)	统计年度、证券代码、证券简称、公司ID、研发人员薪资占比、研发人员占比、高学历人员占比、固定资产占比、制造费用占比、研发折旧摊销占比、研发租赁费占比、研发直接投入占比、无形资产占比、总资产周转率、权益乘数倒数。
企业新质生产力二级指标表(劳动力、生产工具角度)	统计年度、证券代码、证券简称、公司ID、活劳动、物化劳动(劳动对象)、硬科技、软科技。
企业新质生产力水平表(劳动力、生产工具角度)	统计年度、证券代码、证券简称、公司ID、行业代码、行业名称、所属省份、所属城市、劳动力、生产工具、企业新质生产力水平。

CSMAR与前沿主题：（四）新质生产力

范文	内容提要	关键变量	关联CSMAR子库
数字化转型对制造企业新质生产力影响机理研究[J].软科学, 2024.	利用2018-2022年A股上市公司数据，讨论数字化转型对制造企业新质生产力的影响机理及其异质性特征，并针对企业创新程度的中介作用和融资约束的调节效应进行检验分析。研究发现：数字化转型对制造企业新质生产力存在显著正向影响。	**数字化转型：**企业数字化转型关键词词频 **新质生产力：**劳动者、劳动对象和劳动资料的熵值和	新质生产力 数字经济 企业数字化转型 上市公司研发创新 经营困境 …
金融供给地理结构如何赋能企业新质生产力？[J].兰州学刊, 2024.	文章以2011-2022年中国A股上市公司数据为研究样本，实证检验金融供给地理结构对企业新质生产力的影响程度、作用机制、调节效应和距离效应。研究发现：金融供给地理结构优化对企业新质生产力具有正向影响。	**金融供给地理结构：**企业5km范围内金融机构数量 **新质生产力：**基于指标体系测算	新质生产力 金融机构分支机构 内部控制 上市公司研发创新 …
人力资本对新质生产力的影响效应及机制分析[J].调研世界, 2024.	文章基于2008—2019年北京市中关村海淀科技园企业数据，从人力资本的学历、技能、工龄三个维度分别讨论了人力资本对企业新质生产力的影响效应与影响机制。实证结果显示：各维度人力资本对新质生产力有显著影响，员工受教育水平越高、工龄越长，将越有利于企业新质生产力的提高。	**共同富裕：**学历人力资本；工龄人力资本；技能人力资本 **新质生产力：**基于指标体系测算	新质生产力 上市公司基本信息 财务报表 财务指标分析 上市公司研发创新 ……

CSMAR与前沿主题：（四）新质生产力

范文	内容提要	关键变量	关联CSMAR子库
绿色技术创新、新质生产力与低碳经济高质量发展[J].统计与决策,2024.	文章选取2010—2022年我国31个省份的面板数据，实证分析绿色技术创新对低碳经济高质量发展的影响机制。研究发现：绿色技术创新对低碳经济高质量发展具有正向赋能效应。中介效应检验结果显示，绿色技术创新可加速新质生产力发展，继而促进低碳经济高质量发展。	**低碳经济高质量发展水平：**从低碳产业、低碳创新、低碳环境和低碳社会四个维度测度 **绿色技术创新水平：**绿色发明专利授权数 **新质生产力：**从新型劳动者、新型劳动资料和新型劳动对象三个维度测度	新质生产力 碳中和 绿色专利 宏观经济 区域经济 数字经济 经济内循环 ……
新质生产力如何影响产业链韧性：理论分析与经验证据[J].统计与决策,2024.	文章基于2014—2022年我国30个省份的面板数据，探讨新质生产力与产业链韧性二者之间的内在联系。研究结果表明，新质生产力对产业链韧性具有显著促进作用。作用机制检验表明，新质生产力能够通过提升技术创新水平、推动产业结构转型升级作用于 产业链韧性。	**产业链韧性：**产业链基础能力（基础设施/劳动力）、抵御能力（产业链体系/产业链稳定性）和创新能力（创新投入/创新产出） **新质生产力：**从劳动者、劳动对象和生产资料三个方面测度	新质生产力 宏观经济 区域经济 企业创新 专利 经济内循环 ……
新质生产力、国内国际双循环与共同富裕[J].统计与决策,2024.	文章基于2011—2022年我国272个地级市的面板数据，构建多种计量模型实证考察新质生产力对共同富裕的影响。研究结果显示，新质生产力能够有效促进共同富裕。机制检验结果可知，新质生产力可通过经济内循环及经济外循环渠道间接赋能共同富裕。	**共同富裕：**从物质生活富裕、精神生 活充沛、发展共建共享、公共服务改善、生态环境和谐五个 维度测度 **经济内循环与经济外循环：**供给侧/需求侧、引进来/走出去 **新质生产力：**从新业态、新技术、新要素、新质态以及新质量五个维度测度	新质生产力 宏观经济 区域经济 数字经济 企业创新 专利 经济内循环 碳中和

03 CSMAR数据库应用

- 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
- 上司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
- 证券市场中介机构
- ……

更多新库即将上线！

最新数据资源

人工智能研究数据库及应用案例

包含12张表，160个字段。数据为年度、月度更新，最早数据区间开始于1996年。

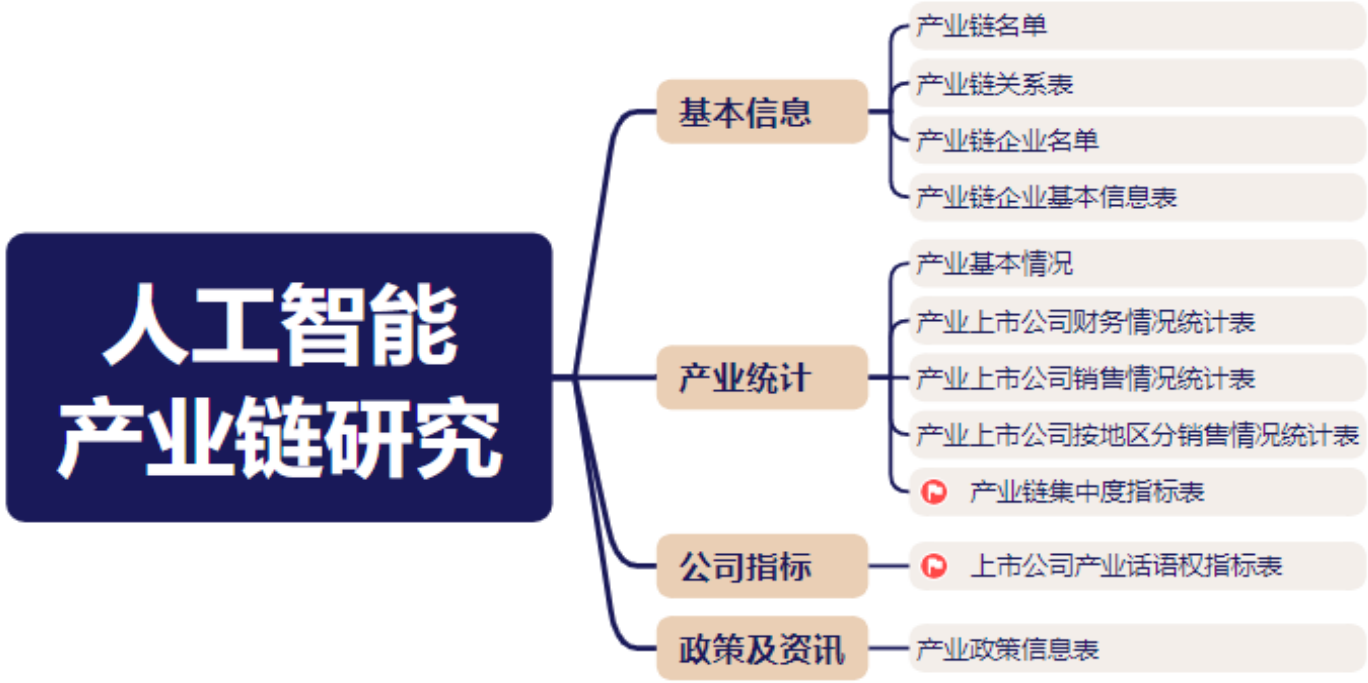
数据来源：上市公司年报、中国统计月报、各政府网站、国际机器人联合会、互联网信息服务算法备案系统。

最新数据资源

范文	内容提要	关键变量	关联CSMAR子库
产业合作复杂网络结构演化与资源配置研究：来自中国人工智能企业大数据的实证分析[J].中国软科学, 2025.	文章提出了产业合作指数、产业网络结构集中度、产业合作垄断指数、产业组织复杂性指数等指标，应用二阶段回归模型(IV2SLS) 实证分析了企业网络中心性、市场地位、集聚系数与市场绩效的相关性。	人工智能产业上市企业盈利指数	人工智能 上市公司研发创新 上市公司与关联公司专利 ……
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.	文章以2007—2018年沪深A股上市公司为样本，研究了人工智能技术如何影响生产效率和劳动力技能结构。研究发现，人工智能显著提升了中国上市公司的生产率。在影响机制方面，人工智能通过促使企业减少常规低技能劳动力需求、增加非常规高技能劳动力需求的方式提升企业的生产率，这体现了企业劳动力技能结构的调整。	年报人工智能关键词词频 人工智能专利数量	人工智能 上市公司研发创新 上市公司人物特征 ……
人工智能发展对微观主体全要素生产率的驱动效应研究——方向识别与影响机制[J].经济问题探索, 2025.	文章以2005—2021年中国 A 股非金融石油石化上市企业为样本，研究了人工智能发展影响全要素生产率的作用机理。结果显示：人工智能发展资本投入与劳动投入对微观主体全要素生产率提升具有正向促进作用；人工智能发展通过推动科技创新与产业结构高级化促进微观主体全要素生产率提升。	人工智能资本投入	人工智能 财务报表 财务指标分析 上市公司研发创新 ……
人工智能与企业客户稳定性——基于中国工业机器人应用的证据[J].经济问题, 2024.	文章基于2011—2019年中国A股制造业上市公司数据,实证检验企业层面工业机器人应用对客户稳定性的影响。结果表明，工业机器人应用会显著提高企业客户稳定性；从影响机制看，工业机器人应用主要通过“技术进步效应”和“风险治理效应”增进客户稳定性。	企业工业机器人渗透度	人工智能 供应链研究 财务报表 财务指标分析 ……

最新数据资源

中国人工智能产业链研究数据库及应用案例



包含共11张表，158个字段，主要是年度更新，数据起始时间为2020年。

数据来源：上市公司定期报告、统计局、国务院政策文件。

最新数据资源

范文	内容提要	关键变量	关联CSMAR子库
人工智能产业高质量发展体系与测度[J].宏观经济管理, 2024.	文章结合我国人工智能产业现状，紧密围绕基础研究、技术研发、产品应用 3 大核心环节，强化人才发展、安全与治理两大动力支撑，建议构建“三核心 + 两动力”的人工智能产业高质量发展体系架构。在此基础上，探索了该产业高质量发展测度指标体系，以全面评估产业发展态势和特征，为推动产业高质量发展提供科学指导与支撑。	人工智能产业统计及相关政策法规数量	人工智能产业链 宏观经济 经济内循环 ……
产业链中的话语权与非金融企业金融投资——基于产业链中商业信用水平的视角[J].会计研究, 2022.	文章选择2007-2017年我国沪深上市非金融企业的样本数据，以产业链中商业信用水平表征产业链中的话语权，实证分析了非金融企业在产业链上下游的话语权和整体话语权对其金融投资行为的影响及作用路径。研究发现，非金融企业对上游企业具有话语权时，能够抑制其金融投资；而对下游企业和在产业链整体中具有话语权时，则促进了企业金融投资。企业对上下游的话语权与金融投资之间的关系具有非对称性。	企业在产业链中对上游、下游和整个产业链的话语权	人工智能产业链 人工智能 治理结构……
不确定环境下我国人工智能产业政策的演化形成研究[J].现代情报, 2024.	文章结合文献计量、网络分析、共现分析和主题结构演化等方法，分别从人工智能产业政策数量、联合发布机构和主题结构演化3个角度进行分析。研究发现，我国人工智能产业政策发文数量逐年提高；中央地方政府协调统一；政策主题日益丰富，但是还存在人工智能技术创新性较低、政策平衡性欠缺、缺乏国际合作等不足。未来人工智能产业政策需要在技术创新、教育体系、就业问题和国际合作等方面进一步完善优化。	人工智能产业政策	人工智能产业链 ……

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?	许元镫 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
- 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
- 区域热辐射数据的科研应用
- 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/