CSMAR

CSMAR数智科研：

数据赋能 + 工具进阶

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

01 CSMAR数据库简介

- 数据库概况
- 主要功能

02 CSMAR数据库应用

- 基础应用

  通过权威范文案例，展示如何使用CSMAR的单表查询及跨表查询（包括智能查询、常用变量查询及全变量查询）进行实证研究的数据收集。

- 创新探索

  梳理近期研究趋势及热点主题，并选取部分创新热点进行解析（包括数字化发展、人工智能、气候与环境、新质生产力以及大语言模型应用），提供创新选题思路。

03 CSMAR最新数据

- 最新数据库概览
- 新库资讯获取

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
- 主要功能

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

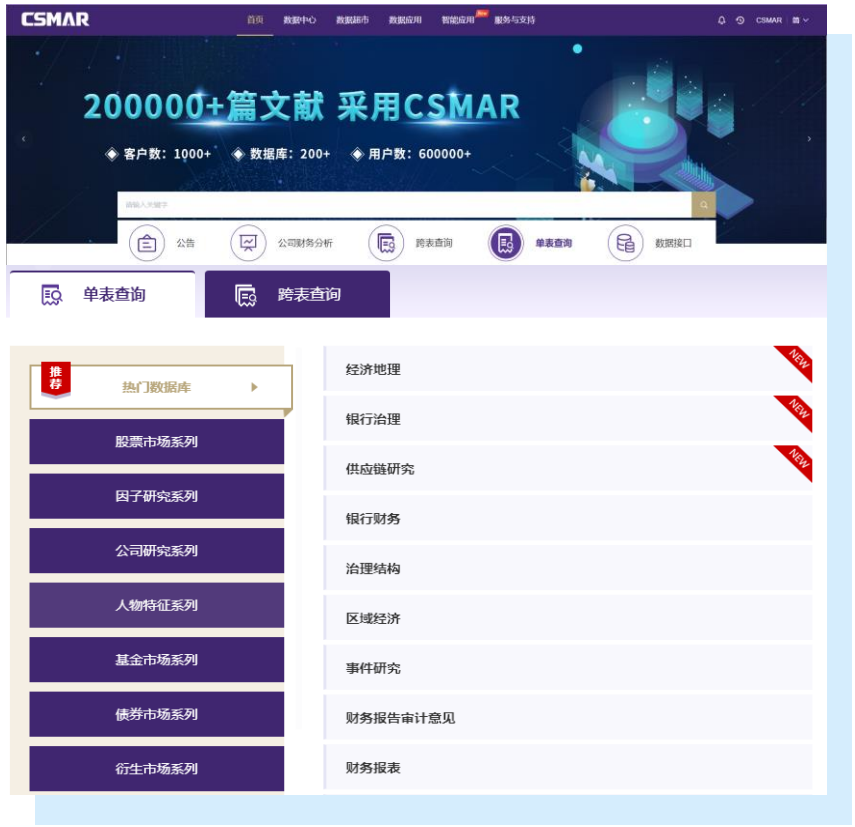
定　　位：**研究型精准**数据库

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

- 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

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	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、业绩说明会文本、税收激励与研发创新、风险监测
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、人口老龄化、普惠金融、经济内循环等，以及常用的宏观、区域经济指标	宏观经济、区域经济	税收条约
专题研究系列	聚焦重要研究主题，提供特色数据指标，包括：数据要素、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		人工智能产业链、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、AI事件、AI情绪因子
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等		央媒标签、报道热度
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、区域热辐射、企业可持续增长CSG、上市公司高管团队战略领导力、各区县政府工作报告文本、新质生产力政府关注度

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。

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

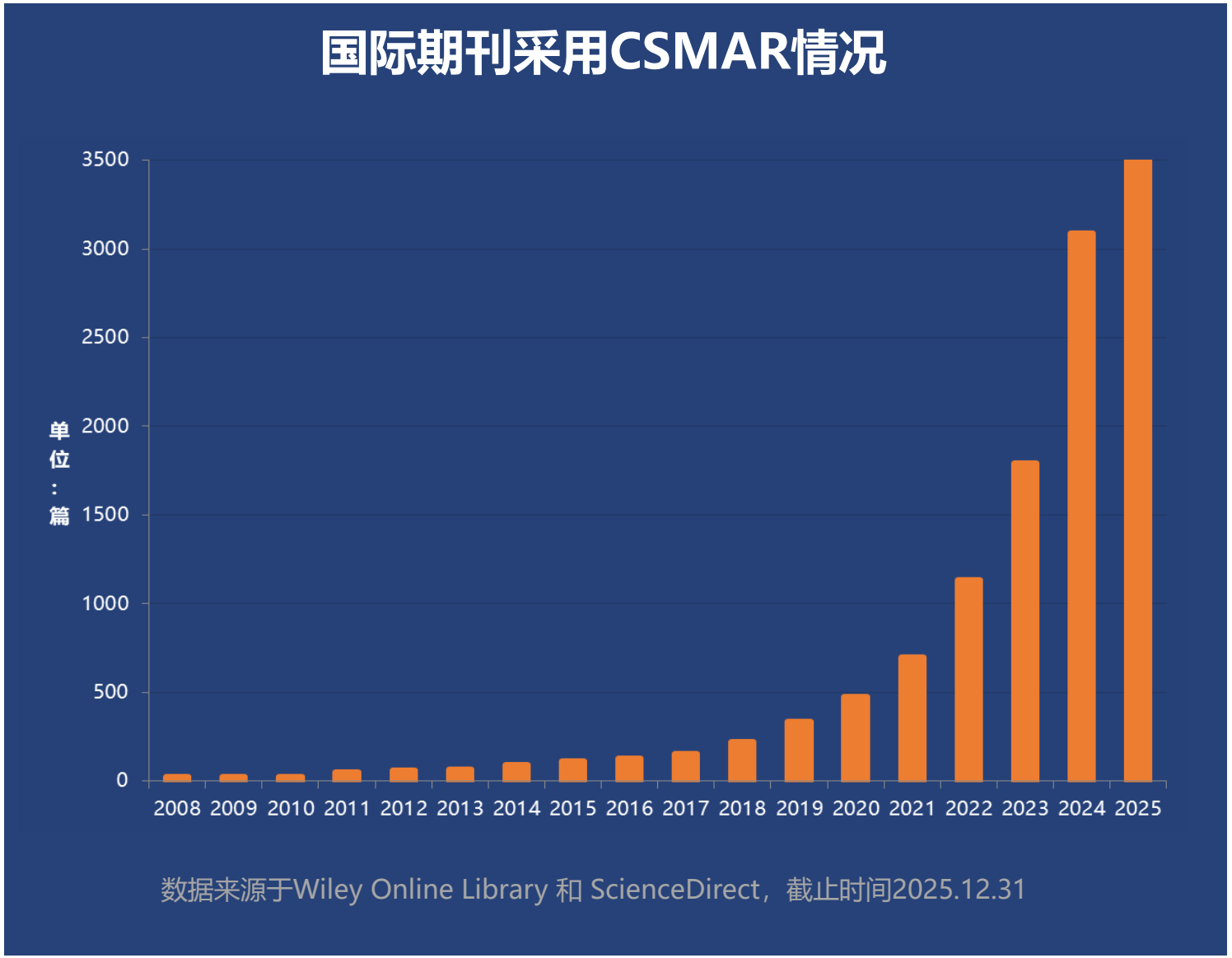
- CSMAR数据被**摩根斯坦利**选用，作为编制MSCI-A股指数的**基础**。

03

引用

- 截至2025年12月31日，国内外使用**CSMAR**数据发表的论文超过20万篇。

CSMAR数据库简介：数据库概况



截止时间2025.12.31 ， 国际顶刊累计发表量

CSMAR数据库简介：主要功能

数据查询

关键字搜索、单表查询、跨表查询（智能查询、常用变量查询、全变量查询）、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

图谱中心、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……

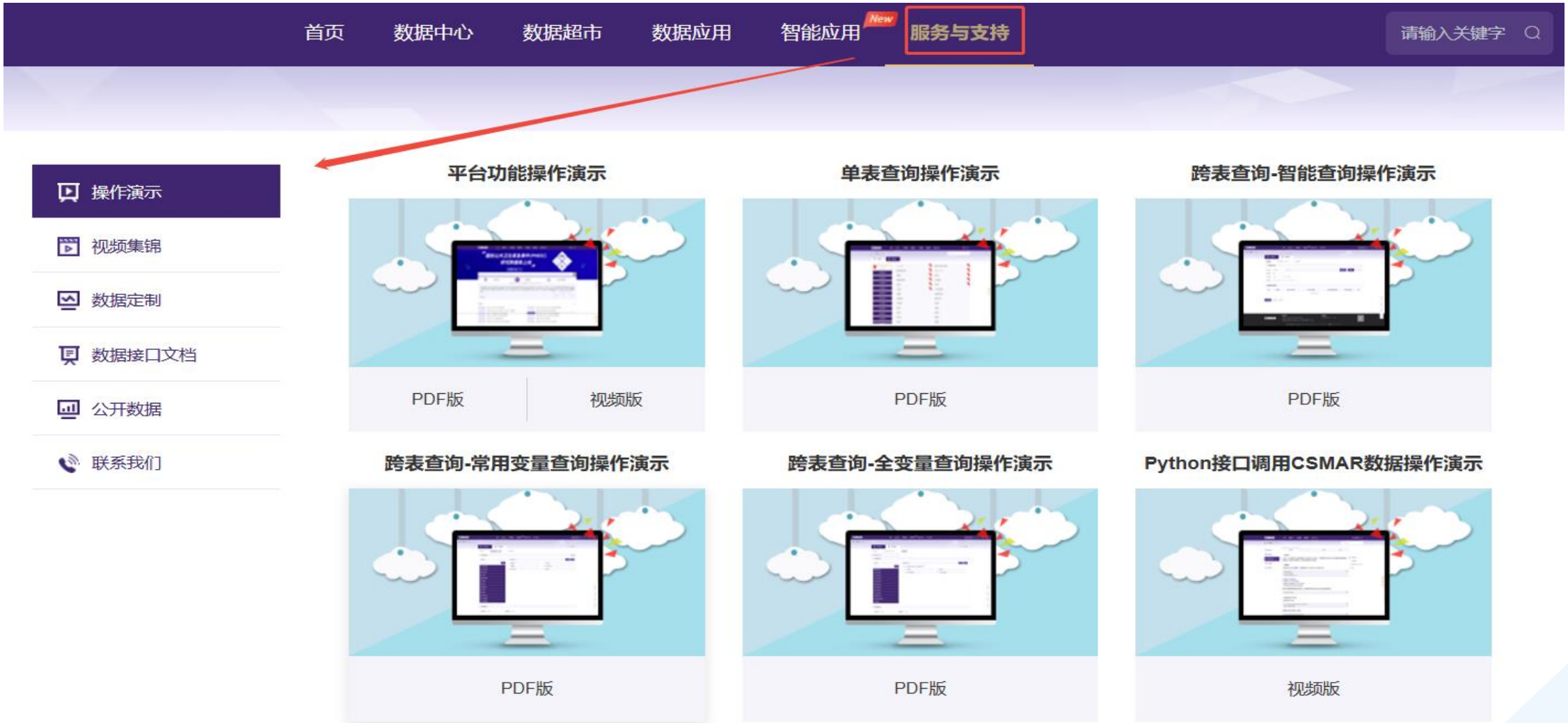
CSMAR Solution

主要功能

http://data.csmar.com/

CSMAR数据库简介：主要功能

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

- 基础应用
- 创新探索

CSMAR数据库应用：基础应用

实证研究

研究设计 → 样本选择 → 数据收集 → 数据分析 → 结果解释

案例：人工智能技术水平与企业投资效率[J].经济管理, 2025.

作为引领新一轮科技革命和产业变革的战略性力量，人工智能技术成为企业投资决策的重要支撑。文章以2007—2022年中国沪深A股上市公司为样本，实证检验人工智能技术水平对企业投资效率的影响及其内在机制。研究表明：人工智能技术能够显著提升企业投资效率。

$$INEFF_{i,t} = \beta_0 + \beta_1 LNAIP_{i,t} + \beta_i X_{i,t} + \sum FIRM + \sum YEAR \times PRO + \varepsilon_{i,t}$$

变量名称	变量符号	变量定义
企业投资效率	*INEFF_R*	基于Richardson(2006)[45]计算的投资效率，数值越大代表企业投资效率越低
企业投资效率	*INEFF_B*	基于Biddle等(2009)[16]计算的投资效率，数值越大代表企业投资效率越低
人工智能技术水平	*LNAIP*	申请的人工智能专利数量加1取自然对数
公司规模	*SIZE*	总资产的自然对数
资产负债率	*LEV*	总负债/总资产
资产报酬率	*ROA*	净利润/总资产
营业收入增长率	*GROWTH*	本年营业收入/上一年营业收入-1
产权性质	*SOE*	国有控股企业取值为1，否则为0
公司年龄	*FIRMAGE*	公司成立年限加1取自然对数
董事会规模	*BOARD*	董事会人数取自然对数
独立董事比例	*INDEP*	独立董事人数/董事会人数
股权制衡度	*BALANCE*	第二大股东持股比例除以第一大股东持股比例

Richardson模型

$$Invest_t = \beta_0 + \beta_1 Growth_{t-1} + \beta_2 Lev_{t-1} + \beta_3 Cash_{t-1} + \beta_4 Age_{t-1} + \beta_5 Size_{t-1} + \beta_6 Returns_{t-1} + \beta_7 Invest_{t-1} + \xi$$

Invest为新增投资，Growth 为公司的成长机会，Lever为公司的资产负债率，Cash等于现金及现金等价物除以总资产，Age为公司的上市年限，Size为公司规模，Return为公司股票年度回报。

Biddle模型

$$Invest_t = \beta_0 + \beta_1 Growth_{t-1} + \xi_t$$

单表查询：经营困境、人工智能专利、税收激励与研发创新、财务报表、财务指标分析、治理结构、上市公司基本信息、股票市场交易、……

跨表查询

需注册个人账户

- **智能查询：**CSMAR智能解析并推荐相关变量
- **常用变量查询：**快速整合常规控制变量
- **全变量查询：**自由组合更多变量

CSMAR数据库应用：创新探索

研究趋势

位居前10位的热点前沿体现了经济社会向**数字化、绿色化转型发展**的趋势。

2023

2024

位居前10位的热点前沿体现出**数智技术与经济学、心理学及其他社会学的交叉融合，关注人类健康和气候环境主题**等研究趋势。

2025

位居前10位的热点前沿呈现三大核心趋势：
一是**数字技术变革推动社会经济转型，重塑产业形态、就业和消费市场**；
二是**可持续发展相关研究成为核心议题，聚焦绿色创新、气候风险与能源转型**；
三是**人类健康优化与认知管理更趋技术化与个性化。**

序号	热点前沿
1	元宇宙背景下文旅数字化转型发展研究
2	科技金融与绿色创新
3	机器人发展与就业转型
4	个性化与数据驱动的心理治疗研究
5	大规模神经影像在个体行为差异研究中的应用
6	能源转型与碳排放
7	基于分位数连通性分析的风险管理研究
8	气候风险与 ESG 因素对市场的影响
9	数字化时代认知负荷理论的应用拓展
10	直播电商平台对消费者的影响分析

资料来源：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

热点主题

2022

1. 数字经济的深化研究
2. 全面推进共同富裕研究
3. 在社会主义市场经济条件下规范和引导资本健康发展
4. 产业链供应链韧性与安全水平提升
5. 国际冲突与国际经贸合作

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

2024

1. 新质生产力研究
2. 生态文明建设与高质量发展
3. 数据要素的创新驱动作用研究
4. 人工智能的产业赋能
5. 金融服务实体经济高质量发展

资料来源：经济学研究热点课题组

CSMAR数据库应用：创新探索

创新热点解析

1. 数字化发展
2. 人工智能
3. 气候与环境
4. 新质生产力
5. 大语言模型应用

CSMAR数据库应用：创新探索

案例一：数字化发展

数字经济发展阶段

第一阶段：数据数字化 → 第二阶段：数据要素化 → 第三阶段：数据驱动化

数字基础设施建设与发展

数据要素

宏观	微观
构建统计指标体系	文本分析挖掘年报中的关键词，基于词频构建企业数字化转型水平；数字创新；数字竞争优势等

宏观	微观
国家数据交易平台试点政策、地级市政府上线公共数据开放平台、大数据综合试验区的设立等	基于企业数据要素利用相关关键词词频构建

关联CSMAR子库

数字经济：提供经济发展总体情况（国民经济核算、人力投入、科技创新、固定资产投资）、数字产业化（通信业、互联网和相关服务、软件和信息技术服务业、电子信息制造业、共享经济）、产业数字化（农业数字化、工业数字化、数字金融、电子商务）、政策信息及上市公司数字化数据。

企业数字化转型：提供上市公司层面的战略引领、技术驱动、组织赋能，数字化成果及应用，中宏观层面的环境支撑，以及构建了企业数字化转型评价指数。

关联CSMAR子库

数据要素：收集与整理宏观方面的地区大数据发展指数、全国数据资源调查统计；微观层面的上市公司数据要素披露情况以及数据要素开发利用指数情况等内容。

CSMAR数据库应用：创新探索

案例一：数字化发展

[1]数字经济发展、资本跨区域非对称流动与异地分类投资[J].经济研究, 2025.	宏观数字经济发展
[2]农村数字化发展水平测度及演化趋势分析[J].统计与决策, 2025.	
[3]算力基础设施建设与企业数字化转型[J].统计研究, 2025.	企业数字化转型
[4]数据要素利用与企业金融资产配置——基于机器学习和文本分析的证据[J].会计研究, 2025.	企业数据要素利用
[5]数据要素对企业绿色创新的影响研究[J].科研管理，2025.	
[6]数据要素市场化如何驱动企业新质生产力发展：乘数效应分析视角[J].现代财经，2025.	宏观数据要素
[7]数据要素赋能城市创业活动的乘数效应：来自中国工商企业注册信息的微观证据[J].系统工程理论与实践，2025.	
[8]数据要素集聚与企业债务融资——基于国家大数据综合试验区的准自然实验[J].经济管理，2025.	

References

CSMAR数据库应用：创新探索

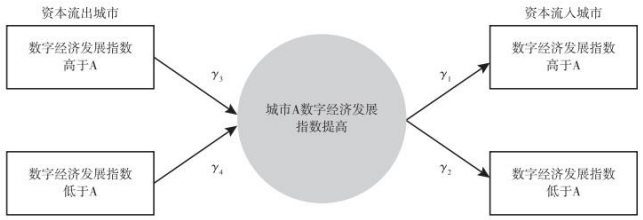
相关研究1：数字经济发展、资本跨区域非对称流动与异地分类投资[J].经济研究, 2025.

$$invest_{ijt} = \beta_0 + \beta_1 \cdot digit_{it} + \beta_2 \cdot digit_{jt} + \beta_3 \cdot X_{it} + \beta_4 \cdot X_{jt} + \beta_5 \cdot X_{ijt} + \rho_{ij} + \theta_{pt} + \theta_{qt} + \varepsilon_{ijt}$$

被解释变量	关键解释变量
若城市i的上市公司在第t年在城市j设立了子公司，则invest取值为1，否则取0。	数字经济发展水平指标体系

一级指标	二级指标	三级指标	数据来源
数字经济发展水平	数字产业化	(1)电信业务总量/总人口	《中国城市统计年鉴》
		(2)信息传输、计算机服务和软件就业人数/就业人口	《中国城市统计年鉴》
		(3)软件收入/总人口	《中国统计年鉴》
		(4)互联网上网人数/总人口	《中国统计年鉴》
	产业数字化	(5)开通互联网宽带业务的行政村比重	中国工业和信息化部网站及各省份统计年鉴
		(6)有电子商务交易活动的企业数比重	《中国电子商务年鉴》
		(7)电子商务交易额/GDP	《中国统计年鉴》
		(8)数字普惠金融指数	北京大学数字普惠金融指数
	数字基础设施建设	(9)移动电话年末用户数	《中国城市统计年鉴》
		(10)互联网宽带接入用户数	《中国城市统计年鉴》
		(11)每千人拥有域名数	《中国统计年鉴》
		(12)每千人拥有网页数	《中国统计年鉴》
		(13)电子信息产业资产总计	《中国统计年鉴》
		(14)移动电话基站数	《中国统计年鉴》
		(15)IPV4地址数	《中国统计年鉴》
	数字技术创新	(16)数字经济发明专利授权量对数	国家知识产权局专利检索和信息分析系统
		(17)数字经济实用新型授权量对数	国家知识产权局专利检索和信息分析系统
	数字化治理	(18)数字经济政策词频	中国研究数据服务平台(CNRDS)
	数据价值化	(19)是否有数据交易中心	中国工业和信息化部网站

文章利用上市公司异地投资数据以及母子公司生产关系信息，从规模、模式和方向三个维度分析了数字经济发展对资本跨城市非对称流动的影响。研究表明，数字经济发展的作用包括：第一，显著促进本城市企业的对外投资和外地资本的进入，但对资本流出的促进作用相对更大；第二，同时促进市场扩张型、跨行业型以及分工型资本的非对称流动，对生产型、销售型和研发型的资本流动存在积极作用，但会限制加工型投资的流入；第三，对资本跨区域流动的影响呈现出明显的“数字鸿沟”现象——本城市与数字经济水平更高的城市之间资本互动的增加明显超过与数字经济水平较低的城市之间的联动。



变量	OLS				2SLS
	(1)	(2)	(3)	(4)	(5)
流出城市数字经济发展指数	0.3368*** (0.0710)	0.2007*** (0.0479)	0.2015*** (0.0469)	0.1637*** (0.0427)	0.4861*** (0.1176)
流入城市数字经济发展指数	0.1294*** (0.0412)	0.1132*** (0.0238)	0.1082*** (0.0233)	0.1046*** (0.0221)	0.3280*** (0.0689)

CSMAR数据库应用：创新探索

相关研究2：农村数字化发展水平测度及演化趋势分析[J].统计与决策, 2025.

一级指标	二级指标	指标说明
农村数字化基础G	农村互联网普及率G1	农村地区网民数量占地区常住人口的比重
	农村移动电话覆盖情况G2	每百户拥有的移动电话数
	农村光缆线路覆盖率G3	每平方千米光缆线路长度
	农村社会数字产业固定投资G4	信息传输、计算机服务和软件业固定资产投资额
	农村社会数字服务固定投资G5	交通运输、仓储和邮政业固定资产投资额
	农村数字化人才拥有量G6	信息技术从业人员数
农业数字化A	农业气象监测网络覆盖程度A1	环境与农业气象观测业务站点个数
	农业电力投入产出情况A2	农、林、牧、渔业增加值/农村用电量
	农业数字化规模A3	第一产业中数字经济增加值
	农业技术水平A4	由C-D生产函数计算得出的农业技术增产进步率
	农业数字防灾率A5	农作物病虫疫情田间监测面积/耕地面积
	农业现代化水平A6	设施农业占地面积
经营数字化M	农村企业网站数M1	每百家企业拥有网站数
	农村企业电子商务活跃度M2	有电子商务交易活动的企业占比
	农村电子商务销售额M3	基于网络订单销售的商品和服务总额
	农村电子商务采购额M4	基于网络订单采购的商品和服务总额
	农村数字经验示范县建设规模M5	电子商务进农村综合示范县数量
	农村数字化基地M6	淘宝村数量
流通数字化C	农村邮政通信服务水平C1	农村平均每一邮政营业网点服务人口
	农村消费品零售水平C2	农村消费品零售额/全社会消费品零售额
	农村投递线路(物流)C3	投递段道上向农村用户投交的路线长度
	已通邮的行政村占比C4	已通邮的行政村占全部行政村的比重
生活服务数字化L	农村网络投资数量及规模L1	数字普惠金融县域投资指数
	农村网络支付数量及规模L2	数字普惠金融县域移动支付指数
	农村交通通信支出水平L3	农村居民人均交通通信支出占比
	农村数字消费水平L4	家庭使用各类数字化产品和服务的消费支出占总支出的比重

文章基于农村数字化发展内涵，选取2011—2022年中国30个省份的面板数据，构建农村数字化发展水平评价指标体系，并采用核密度估计法、Dagum基尼系数与标准差椭圆对中国农村数字化发展水平的时空演化趋势进行分析。结果表明：（1）中国农村数字化发展水平整体呈稳步上升趋势，且呈现“东部地区高，中部地区次之，西部、东北地区低”的分布格局。（2）各省份间的农村数字化发展水平差异不断扩大，两极分化现象不明显或不存在。（3）各区域间存在显著差异，且总体差异主要来源于区域间差异。（4）中国农村数字化发展水平分布重心向南移动，表现出“集聚—扩散—再集聚”的空间集聚特征。

2011—2022年农村数字化发展水平变化趋势

数据主要来源于《中国农村统计年鉴》、《中国信息年鉴》、《中国数字经济发展白皮书》、中国淘宝村研究报告、北京大学数字普惠金融指数及国家统计局网站。

农村数字化发展各子维度指数测度结果

年份	农村数字化基础指数	农业数字化指数	经营数字化指数	流通数字化指数	生活服务数字化指数	农村数字化发展水平
2011	0.090	0.175	0.027	0.414	0.278	0.099
2012	0.105	0.160	0.028	0.433	0.278	0.101
2013	0.122	0.184	0.026	0.444	0.329	0.112
2014	0.142	0.200	0.045	0.458	0.274	0.126
2015	0.166	0.216	0.078	0.470	0.398	0.158
2016	0.190	0.213	0.094	0.469	0.418	0.171
2017	0.215	0.211	0.105	0.480	0.502	0.186
2018	0.246	0.212	0.115	0.500	0.510	0.199
2019	0.264	0.228	0.115	0.511	0.516	0.207
2020	0.287	0.233	0.135	0.507	0.526	0.223
2021	0.305	0.238	0.147	0.512	0.573	0.236
2022	0.331	0.249	0.168	0.518	0.627	0.257

CSMAR数据库应用：创新探索

相关研究3：算力基础设施建设与企业数字化转型[J].统计研究, 2025.

$$DT_{i,t} = \alpha_0 + \alpha_1 CI_{c,t} + \alpha_2 X_{i,t} + \alpha_3 Z_{c,t} + \mu_i + \gamma_t + \varepsilon_{i,t}$$

被解释变量	关键解释变量
采用企业数字化转型指数衡量。相较于以数字化相关词频占管理层讨论与分析中总词频的比值测度的企业数字化转型水平，该指数涵盖的信息更为全面，通过综合考量多个维度能够准确反映企业在数字化转型过程中的整体表现。	企业所在地级市已投入运营的超大型及大型数据中心数量加一取自然对数。 *通过查阅各地政府网站及网络搜索引擎，交叉比对验证后获取各地已投入运营的数据中心名单。

②该指数来自CSMAR数据库，包含战略引领、技术驱动、组织赋能、环境支撑、数字化成果及数字化应用6个层面，具体网址为 https://data.csmar.com/。

文章以2011—2022年沪深A股上市公司为研究样本，实证检验算力基础设施建设水平对企业数字化转型的影响。研究发现，地区算力基础设施建设水平的提升显著促进企业数字化转型。机制分析表明，地区算力基础设施建设能够发挥资源效应与治理效应，并通过提升企业长期价值和改变企业信息披露行为推动企业数字化转型，即企业数字化转型兼具策略性和实质性。

基准回归分析

	DT （1）	*DT* （2）
CI	0.581** （0.264）	0.540** （0.235）

算力基础设施建设对企业数字化转型的影响机制分析：实质性转型还是策略性转型

	RD （1）	*DT* （2）	*DIQ* （3）	*DT* （4）	*IDQ* （5）	*DT* （6）
CI	0.554**	−0.212	0.048*	0.339	0.114***	0.359
	（0.222）	（0.261）	（0.026）	（0.206）	（0.028）	（0.244）
RD		0.008				
		（0.016）				
RD×CI		0.042**				
		（0.017）				
DIQ				0.570***		
				（0.092）		
DIQ×CI				0.109*		
				（0.064）		
IDQ						−0.035
						（0.026）
IDQ×CI						0.033**

CSMAR数据库应用：创新探索

相关研究4：数据要素利用与企业金融资产配置——基于机器学习和文本分析的证据[J].会计研究，2025.

$$Fin_{i,t}=\alpha_0+\alpha_1 Lnda_{i,t}+\alpha_2 Control_{i,t}+\mu_i+\lambda_t+\varepsilon_{i,t}$$

被解释变量	关键解释变量
金融资产配置 (Fin)：（货币资金+交易性金融资产+投资性房地产 +应收股利+应收利息+可供出售金融资产+持有至到期投资+金融类长期股权投资+衍生性金融资产+短期投资净额) /企业总资产	数据要素利用水平（Lnda)：企业数据要素利用词频汇总后加1取自然对数

文章利用机器学习和文本分析方法构建企业数据要素利用指标，基于2015—2022年我国A股上市公司样本，考察数据要素利用对企业金融资产配置的影响。结果显示，数据要素利用能够降低企业金融资产配置，促进企业实体投资。机制检验表明，数据要素利用可以通过缓解企业现金流不确定性、提高企业产能利用率以及减少影子银行业务来抑制不同动机下的金融投资。

变量	(1)	(2)	(3)	(4)	(5)	(6)
	Fin	*Fin*	*Fin*	*Fin*	*Inn*	*Inn*
Lnda	−0.012***	−0.012***	−0.011***	−0.075***	0.003**	0.002**
	(0.003)	(0.003)	(0.003)	(0.022)	(0.001)	(0.001)
Fin	–	–	–	–	–	−0.049***
						(0.005)

选取企业购买固定资产、无形资产及其他长期资产支付 的现金占总资产的比例来衡量企业的实体投资(Inn)， 探究数据要素利用能否促进企业实体投资。

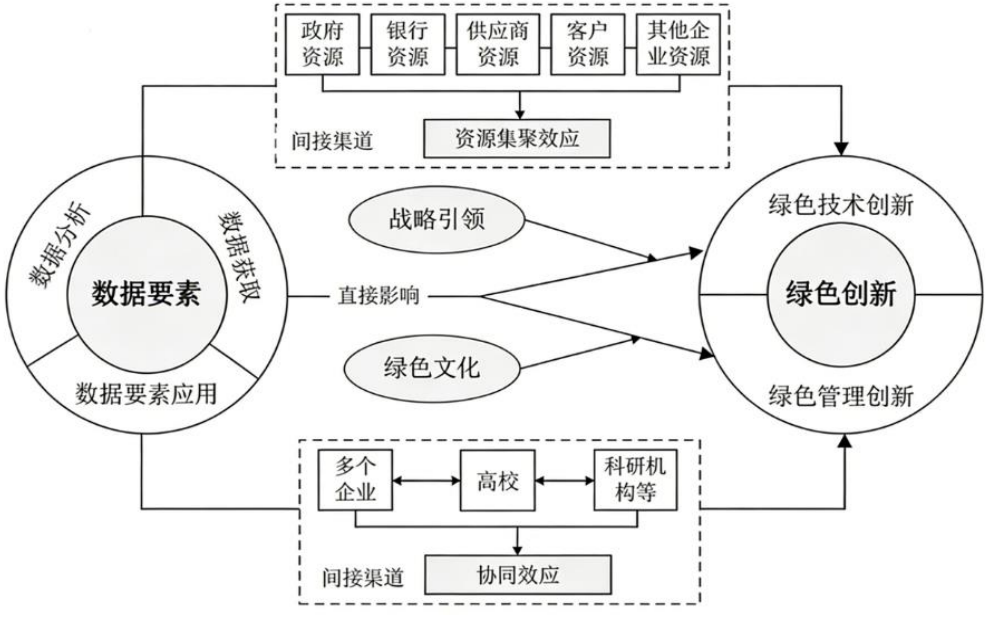
同时检验数据要素利用、企业金融资产配置对企业实体投资的影响。

CSMAR数据库应用：创新探索

相关研究5：数据要素对企业绿色创新的影响研究[J].科研管理，2025.

$$E(GGI_{i,t} \mid X_{i,t}, \varepsilon_{j,t}) = exp(\alpha_0 + \alpha_1 Data_{i,t} + \alpha_2 X_{i,t} + \delta_t + \delta_j + \varepsilon_{j,t})$$

$$Logit(p_i = P[GMI_{i,t} < \omega \mid (\beta_0 + \beta_1 Data_{i,t} + \beta_2 X_{i,t} + \delta_t + \delta_j + \varepsilon_{i,t})] = \frac{\exp(\beta_0 + \beta_1 Data_{i,t} + \beta_2 X_{i,t} + \delta_t + \delta_j + \varepsilon_{i,t})}{1 + \beta_0 + \beta_1 Data_{i,t} + \beta_2 X_{i,t} + \delta_t + \delta_j + \varepsilon_{i,t}}$$



文章基于2012-2023年A股上市公司数据，实证探究了数据要素对企业绿色技术创新和管理创新的影响效应与机制。结果表明，数据要素显著提升了企业绿色技术创新和绿色管理创新绩效，战略引导和绿色文化强化了数据要素与绿色创新之间的关系。机制分析表明，数据要素通过“资源集聚效应”和“协同效应”提升企业绿色创新绩效。

变量类型	变量名称	符号	变量定义
解释变量	数据要素	*Data*	上市公司年报中数据要素关键词数量加 1 的自然对数
	绿色技术创新	*GGI*	绿色发明申请专利数量
被解释变量	绿色管理创新	*GMI*	对五项指标进行赋值加和

根据企业对外公布环保管理制度、实施环保教育培训、开展环保专项 行动的情况来评估企业环保管理的规范度、重视程度以及实际效果。此外，结合企业环境监 管与认证披露情况表中ISO9001和ISO14001认证的通过情况，综合使用五项指标全面评估绿色管理创新绩效，并对五项指标进行二元赋值后加和，以此构建绿色管理创新指标（GMI）。

	Panel A 绿色技术创新			Panel B 绿色管理创新		
	（1）	（2）	（3）	（4）	（5）	（6）
	GGI	*GGI*	*GGI*	*GMI*	*GMI*	*GMI*
Data	0.332***	0.364***	0.291***	0.036***	0.039***	0.069***
	(29.285)	(42.330)	(11.095)	(5.216)	(5.483)	(3.783)

CSMAR数据库应用：创新探索

相关研究6：数据要素市场化如何驱动企业新质生产力发展：乘数效应分析视角[J].现代财经，2025.

文章基于2011-2022年A股上市公司数据，将数据交易平台试点政策作为一项准自然实验，实证考察了数据要素市场化对企业新质生产力的影响及其乘数效应机制。研究发现，数据要素市场化能够显著促进企业新质生产力发展。

被解释变量：企业新质生产力水平
核心解释变量：国家数据交易平台试点政策，将国家数据要素交易平台纳入试点的城市作为实验组

相关研究7：数据要素赋能城市创业活动的乘数效应：来自中国工商企业注册信息的微观证据[J].系统工程理论与实践，2025.

文章利用2008-2021年中国工商企业注册数据，基于地级市政府上线公共数据开放平台这一准自然实验，探讨了数据要素对城市创业活动的影响和经济机制。研究发现，公共数据开放显著增加了城市创业活动

被解释变量：城市创业活动，以城市户籍人口数量作为标准化基数，以城市每百万人新注册企业数刻画城市创业活动
核心解释变量：公共数据开放，以地级市政府上线公共数据开放平台为准自然实验

相关研究8：数据要素集聚与企业债务融资——基于国家大数据综合试验区的准自然实验[J].经济管理，2025.

文章基于国家大数据综合试验区设立这一外生冲击，系统考察数据要素集聚能否以及如何影响企业债务融资能力。研究发现，数据要素集聚能够有效增加企业债务融资规模，并降低企业债务融资成本。

被解释变量：长短期负债总和与期末总资产的比值衡量债务融资规模；债务融资成本的计算公式为：（利息支出+手续费+其他财务费用）/长短期负债总和
核心解释变量：将大数据综合试验区设立作为反映数据要素集聚的一项准自然实验，并构建数据要素集聚变量（DEC）。倘若企业的注册地址位于大数据综合试验区内，则在试验区设立的当年及后续年份，DEC 取值为 1，否则为 0。

CSMAR数据库应用：创新探索

案例二：人工智能

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创新水平

人工智能技术创新相关的专利和软件著作权

AI建设项目

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

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

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

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

* 计算上市公司i的生产部门员工数量占比，将该比例与全行业所有上市公司生产部门员工占比中位数相除，得到相应权重，基于该权重可将行业层面机器人渗透度分解至**企业层面**

综合评价指标体系

搭建多维度人工智能评价指标体系，使用熵权法/熵值法测算

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、工业机器人安装量等方面统计了人工智能相关数据。

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。

数字经济 - 上市公司数字化转型程度(年)
提供经济发展总体情况、数字产业化、产业数字化、政策信息以及上市公司数字化数据。

税收激励与研发创新 - 企业研发技术能力情况表/软件著作权注册情况表
统计了税收对研发激励政策、上市公司适用的税收优惠政策、企业税收指标、税费返还明细、企业税负情况、研发加计扣除程度、企业研发技术能力情况等。

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

CSMAR数据库应用：创新探索

案例二：人工智能

[1]人工智能对员工收入和就业的影响研究[J].统计与信息论坛, 2026.	AI关键词词频
[2]人工智能概念炒作与分析师盈余预测[J].广东财经大学学报, 2026.	AI投资水平
[3]人工智能技术创新如何影响制造业碳减排？[J].中国人口·资源与环境, 2025.	AI创新水平
[4]人工智能应用的绿色创新效应研究[J].经济学动态, 2025.	AI建设项目
[5] Impact of regional artificial intelligence development on corporate environmental information[J]. Finance Research Letters, 2025.	地区AI发展水平
[6]工业机器人使用与服务型制造[J].经济学动态, 2025.	机器人应用水平（企业）
[7]工业机器人应用、制造业就业转移与劳动者健康[J].经济学动态, 2025.	机器人应用水平（行业）
[8]工业机器人的应用如何影响劳动时长？——来自微观个体的证据[J].财经研究, 2025.	
[9]智能经济对新质生产力发展的影响机制研究[J].中国工业经济, 2025.	指标体系
[10]企业人工智能技术水平的分层测度及其创新效应 ——基于全球知识网络视角[J].中国工业经济, 2025.	其他测度方法
[11]人工智能与产业技术融合对制造业企业产品创新绩效的影响研究[J].中国工业经济, 2025.	
[12] “争先恐后”还是“后发制人”？——企业人工智能采用速度与竞争优势[J].外国经济与管理, 2026.	

References

CSMAR数据库应用：创新探索

相关研究1：人工智能对员工收入和就业的影响研究[J].统计与信息论坛, 2026.

$$Li1_{it} = \alpha_0 + \beta_1 Ai_{it} + \sum \beta_j Controls + \lambda_i + \mu_t + \varepsilon_{it}$$

被解释变量	关键解释变量
劳动收入：企业当期为职工支付的现金 / 营业总收入	人工智能水平：统计上市公司年报“人工智能”相关词频数取对数

变量	(1) Li1	(2) Li1	(3) Li1	(4) Li1
Ai	0.012***	0.002**	0.004***	0.004***
	(33.616)	(2.393)	(5.597)	(5.456)

随着大语言模型技术重塑人机协作模式，人工智能成为加速劳动者收入增长的引擎，还是颠覆就业结构的催化剂？基于2009-2023年中国沪深-股上市公司数据，结合2021年大语言模型的普及，采用双重差分模型进行实证研究。研究结果显示，企业人工智能应用水平对劳动收入存在显著的正向促进效应，企业数字化转型是人工智能发挥该效应的重要渠道。

人工智能水平对就业的影响

变量	(1) EMP	(2)高人工智能水平 EMP	(3)低人工智能水平 EMP	(4)劳动密集高人工智能水平 EMP	(5)劳动密集低人工智能水平 EMP
Ai	0.015**	0.023***	−0.006	0.033**	0.002
	(2.102)	(2.589)	(−0.491)	(2.163)	(0.084)

人工智能对高管收入影响

变量	(1) MLi	(2)低人工智能水平 MLi	(3)高人工智能水平 MLi	(4) ManageSumSalary	(5) TOP3ManageSumSalary
Ai	0.005	0.012	−0.001	0.038***	0.029***
	(1.295)	(1.243)	(−0.179)	(4.775)	(3.931)

CSMAR数据库应用：创新探索

相关研究2：人工智能概念炒作与分析师盈余预测[J].广东财经大学学报, 2026.

$$Ferror_{i,t+1} = \alpha + \beta AIwashing_{i,t} + \sum_{k=1}^{11} \gamma_k X_{i,t} + Year + Firm + \varepsilon_{i,t}$$

被解释变量	关键解释变量
分析师预测偏差： $Ferror_{i,t} = \lvert Mean(Feps_{i,j,t} - eps_{i,t}) \rvert / eps_{i,t}$	人工智能概念炒作： $AItext_{i,t} = \beta_0 + \beta_1 AIpat_{i,t} + \beta_2 AIinv_{i,t} + \beta_3 Ind_AItext_{i,t} + \sum \gamma Controls_{i,t} + Ind + Year + \varepsilon_{i,t}$

式(1)中 *AItext* 是企业年报中人工智能披露强度。参考姚加权等的做法[31]，使用 Python 爬虫整理上市公司年报，并计算出人工智能相关词频数量；*AIpat* 是指企业人工智能专利申请数量；*AIinv* 是企业人工智能软件和硬件的投入金额，数据来自 CSMAR 数据库；*Ind_AItext* 代表同年同行业其他企业年报人工智能披露强度均值；其余控制变量包括企业规模、资产负债率、总资产利润率、上市年龄、所有权性质、第一大股东持股比例；*Ind* 和 *Year* 分别表示行业和年份固定效应。残差 ε 即为人工智能概念炒作衡量指标(*AIwashing*)，*AIwashing* 越大，说明企业在人工智能领域的概念炒作越严重。

上市公司的人工智能活动宣传与实践“言行不一”应该如何识别并治理，是当前人工智能产业高质量发展的重要问题。基于2007-2023年上市公司年报文本和人工智能创新与应用数据，构造企业人工智能概念炒作测度指标，研究发现企业人工智能概念炒作会显著加剧分析师预测偏差，但未带来相关实质性创新和应用。该效应是通过推高分析师和投资者的乐观情绪来发挥作用。

变量	(1)	(2)	(3)	(4)
	全样本		概念炒作组	非概念炒作组
	Ferror	*Ferror*	*Ferror*	*Ferror*
AIwashing	0.049**	0.042**	0.074*	0.016
	(0.019)	(0.019)	(0.041)	(0.048)

机制分析

变量	(1)	(2)	(3)	(4)
	分析师乐观情绪		投资者乐观情绪	
	Analyst_opt1	*Analyst_opt2*	*Market_opt1*	*Market_opt2*
AIwashing	0.050**	0.009**	0.010**	0.004**
	(0.019)	(0.004)	(0.005)	(0.002)

CSMAR数据库应用：创新探索

相关研究3： 人工智能技术创新如何影响制造业碳减排？[J].中国人口·资源与环境, 2025.

宏观省级： $E_{1pt} = \alpha + \beta_0 A_{1pt} + \beta_1 C + \gamma_p + v_t + \varepsilon_{pt}$

被解释变量	关键解释变量
中国30个省份的制造业碳排放总量占生产总值比例的自然对数	省级人工智能专利数量

文章从人工智能技术创新的角度出发，基于中国30个省份2010—2020年的制造业碳排放数据，从宏观的省级层面考察人工智能技术创新对碳排放的影响，从微观的企业层面评估人工智能技术创新在制造业企业中的碳减排效应及其内在机制。研究发现：人工智能技术创新显著降低了省级层面和企业层面的制造业碳排放强度。机制检验发现，人工智能技术创新通过技术水平提升效应、管理效率优化效应以及融资约束削减效应3个途径降低制造业企业的碳排放强度。

微观企业： $E_{2it} = \alpha + \beta_0 A_{2it} + \beta_1 S + f_i + v_t + \varepsilon_{it}$

被解释变量	关键解释变量
企业碳排放总量占总产值比例的自然对数值	人工智能技术创新相关的专利和软件著作权

变量	(1)	(2)	(3)
A_1	-0.198^{***}	-0.181^{***}	-0.166^{***}
	(−4.89)	(−4.71)	(−4.19)

变量	(1)	(2)	(3)
A_2	-0.008^{**}	-0.008^{**}	-0.007^{**}
	(−2.34)	(−2.17)	(−2.04)

CSMAR数据库应用：创新探索

相关研究4：人工智能应用的绿色创新效应研究[J].经济学动态，2025.

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

被解释变量	关键解释变量
绿色发明专利申请数和绿色实用新型专利申请数	是否属于引入人工智能建设项目的企业
绿色专利	人工智能 / 财务报表附注

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。

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

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。

长期资产 — 在建工程基本情况表（FN_FN017A）

样本数据

证券代码	统计截止日期	数据来源	报表类型	项目	期初账面余额	期初减值准备
002242	2019-12-31	0	1	智能工厂项目	5661206.9	
002242	2019-12-31	0	1	零星工程	3863809.05	
002242	2019-12-31	0	1	合计	9525015.95	
002800	2019-12-31	0	1	天顺哈密物流...	7139256.56	
002800	2019-12-31	0	1	金蝶EAS软件	108465.52	
002800	2019-12-31	0	1	天顺物流信息...	174339.62	
002800	2019-12-31	0	1	合计	7422061.7	
300223	2019-12-31	0	1	君正集成电路...	35139195.18	
300223	2019-12-31	0	1	合计	35139195.18	
000737	2019-12-31	0	1	燃煤锅炉淘汰...		
000737	2019-12-31	0	1	生物质锅炉		

CSMAR数据库应用：创新探索

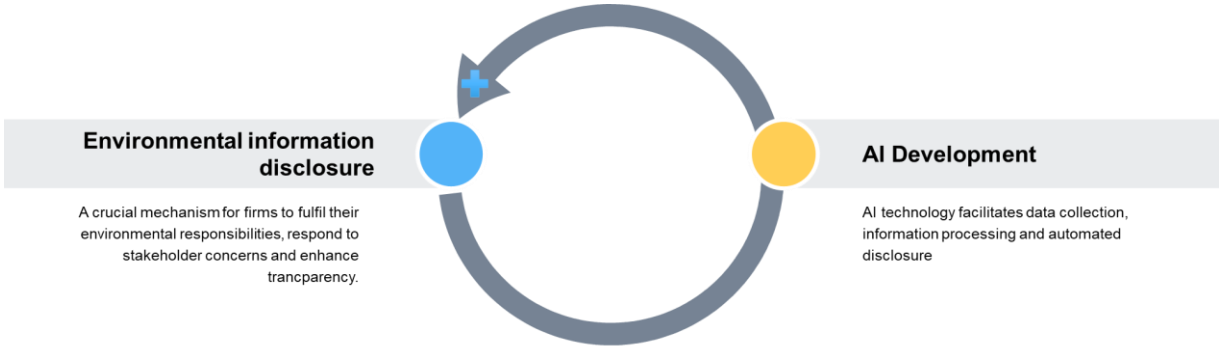
相关研究5： Impact of regional artificial intelligence development on corporate environmental information[J]. Finance Research Letters, 2025.

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

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

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

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.



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)

CSMAR数据库应用：创新探索

相关研究6：工业机器人使用与服务型制造[J].经济学动态, 2025.

$$SER_{it} = \alpha_0 + \alpha_1 \ln rob_{it} + \Phi \cdot X + \mu_i + \lambda_t + \varepsilon_{it}$$

被解释变量	关键解释变量
服务型制造：从上市公司年报中搜集整理出服务业务收入的相关数据，将其占主营业务收入的比重作为衡量企业服务型制造发展水平的指标	工业机器人使用：企业层面的工业机器人渗透度

$$rob_{jit} = \frac{R_{jt}}{L_{jt=2010}} \times \frac{E_{jit=2010}}{medTE_{t=2010}}$$

其中，等号左边的 rob_{jit} 表示行业 j 的上市公司 i 在 t 年的工业机器人渗透度。等号右边第一项 $\frac{R_{jt}}{L_{jt=2010}}$ 表示行业 j 在 t 年的工业机器人渗透度，反映了行业生产技术特征，采用行业 j 在 t 年的工业机器人使用总量（R_{jt}）与该行业在2010年（基期）的从业人数（$L_{jt=2010}$）之比衡量。等号右边第二项 $\frac{E_{jit=2010}}{medTE_{t=2010}}$ 则表示2010年（基期）上市公司 i 的生产部门员工占比（$E_{jit=2010}$）与全行业上市公司同年生产部门员工占比中位数（$medTE_{t=2010}$）之比，该项作为一个分解权重，反映了企业对工业机器人应用的敏感程度。权重越高，说明该企业生产部门规模越大，对工业机器人的潜在使用需求越高。

变量	(1)	(2)	(3)
$\ln rob$	0.108*** (0.010)	0.112*** (0.027)	0.111*** (0.027)

作为人工智能技术在制造业生产领域的重要载体，工业机器人的广泛使用催生出以服务增值为导向的新型制造模式，对“两业融合”发展产生重要影响。文章基于Bartik工具变量构造思想，将行业机器人使用数据匹配至制造业上市公司，从企业层面详细考察工业机器人对服务型制造的效应。研究发现，工业机器人使用显著提高了服务业务收入在企业主营业务收入中的比重，有助于服务型制造的发展。

技术创新效应机制检验

变量	(1)	(2)	(3)	(4)
	发明专利授权数	服务型制造	创新效率	服务型制造
$\ln rob$	0.866*** (0.302)		0.044** (0.019)	
机制变量预测值		0.074** (0.032)		1.936*** (0.548)

人力资本结构优化效应机制检验

变量	(1)	(2)	(3)	(4)
	硕士及以上学历人员占比	服务型制造	非生产人员占比	服务型制造
$\ln rob$	1.208** (0.590)		0.092*** (0.022)	
机制变量预测值		0.044*** (0.016)		0.981*** (0.261)

产业集聚效应机制检验

变量	(1)	(2)	(3)	(4)
	就业密度	服务型制造	区位熵	服务型制造
$\ln rob$	0.121*** (0.021)		0.013* (0.007)	
机制变量预测值		0.092* (0.055)		1.618*** (0.378)

CSMAR数据库应用：创新探索

相关研究7：工业机器人应用、制造业就业转移与劳动者健康[J].经济学动态, 2025.

$$m_{ict,t+2} = \alpha + \beta Robot_{ct} + \gamma X_{ict} + \eta Z_{ct-1} + \delta_i + \mu_t + \varepsilon_{ict}$$

被解释变量	关键解释变量
劳动者就业转移情况：：制造业部门劳动者下一期是否依然在制造业部门就业（是=1，否=0）、是否转向非制造业部门从事非农就业（是=1，否=0）、是否转向失业或退出非农劳动力市场（是=1，否=0）	工业机器人渗透度：$Robot_{ct} = \sum_s \frac{employ_{sct_0}}{employ_{ct_0}} \frac{robot_{st}}{L_{st_0}}$

$$h_{ict-2,t} = \alpha + \beta m_{ict-2,t} + \gamma X_{ict} + \eta Z_{ct-2} + \delta_i + \mu_t + \varepsilon_{ict}$$

被解释变量	关键解释变量
个体相邻两期的健康变化	劳动者就业转换类型

文章的核心微观数据来源于2010年至2020年的中国家庭追踪调查（CFPS），该调查是一项全国范围的追踪调查项目，自2010年启动基线调查，遵循每两年一轮的频率进行追踪，调查内容广泛涵盖就业状况、健康状况等。

基于中国家庭追踪调查微观数据和国际机器人联合会提供的工业机器人数据，文章考察了工业机器人应用下制造业部门劳动者被动就业转移的结果及其对劳动者健康的影响。研究发现，随着工业机器人的广泛应用，原制造业部门的就业群体维持本部门就业的概率显著降低，且并未顺利地转向其他非农就业部门，而是更多地遭遇了非农失业或转向了退出非农就业的状态。服务业就业吸纳能力、个体失业保障情况及地区劳动力市场基础，是影响劳动者就业转移方向的关键因素。相较于能够维持制造业就业的劳动者，那些因就业挤出而转向失业状态或彻底退出非农就业市场的劳动者，其健康状况出现了显著的恶化趋势；成功转移至其他非农就业领域的劳动者也面临着健康状况下滑的潜在风险。这主要源于家庭经济状况、个体主观感受和就业质量的下降。

变量	维持制造业就业	转向其他非农就业	非农失业或退出
	(1)	(2)	(3)
工业机器人渗透度	−0.004**	0.000	0.004**
	(0.002)	(0.001)	(0.002)

变量	自评健康	客观健康	他评健康	心理健康
	(1)	(2)	(3)	(4)
转向其他非农就业	−0.031*	0.011	−0.052	−0.015
	(0.016)	(0.010)	(0.055)	(0.021)
非农失业或退出	−0.067***	0.062***	−0.208***	−0.027
	(0.018)	(0.014)	(0.066)	(0.028)

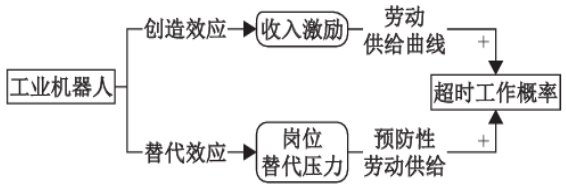
CSMAR数据库应用：创新探索

相关研究8：工业机器人的应用如何影响劳动时长？——来自微观个体的证据[J].财经研究, 2025.

$$WT_{ipt} = \beta_0 + \beta_1 RB_{pt} + \xi Z_{ipt} + \lambda_p + u_t + \varepsilon_{ipt}$$

被解释变量	关键解释变量
是否超时工作：以《劳动法》规定的非特殊情形下的最大周劳动时长为阈值，采用每周工作6天，每天加班1个小时计算所得的周劳动时长50小时作为超时工作的衡量标准	工业机器人安装密度：$RB_{pt} = \sum_{j=1}^{n} \frac{R_{jt}}{L_{p,t=2008}} \times \frac{L_{p,j,t=2008}}{L_{j,t=2008}}$

文章使用中国家庭追踪调查数据（CFPS）研究居民劳动供给。



	(1) WT	(2) WT
RB	0.233*(0.133)	0.428***(0.158)

中国劳动者平均劳动时长处于较高水平，工业机器人的应用对于这一现象是雪中送炭还是雪上加霜呢？文章基于2010—2018年中国家庭追踪调查数据进行了实证分析。研究结果发现，工业机器人的应用显著提高了中国劳动者的超时工作概率。机制分析表明，工业机器人的应用通过收入激励机制提高了劳动者超时工作概率，尤其对低收入劳动者的影响更为显著，而提高最低工资标准能有效缓解工业机器人的超时工作效应；此外，工业机器人应用所引发的岗位替代压力显著提高了劳动者超时工作概率，岗位被替代可能性较高以及社会保障水平较低的劳动者受到的影响更为显著。

工业机器人对劳动时长差距的影响

	(1) WT	(2) WT	(3) WT
RB	0.478***(0.153)	0.880***(0.268)	0.757**(0.310)
EG	−0.786***(0.035)		
RB×EG	−0.447***(0.156)		
EMAX		−0.184***(0.015)	
RB×EMAX		−0.142**(0.058)	
EY			−0.046***(0.005)
RB×EY			−0.039*(0.022)

CSMAR数据库应用：创新探索

相关研究9：智能经济对新质生产力发展的影响机制研究[J].中国工业经济, 2025.

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

被解释变量	关键解释变量
新质生产力：从新质生产力“高科技、高效能、高质量”的特征维度出发，结合数据可得性，设计新质生产力的评价指标体系	智能经济：将智能经济分为智能产业化、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

目标	一级指标	二级指标	三级指标	测度方法或数据来源	指标属性
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	正向
			集成电路水平	集成电路年销售额	正向
			算力规模	显卡进口金额对数	正向
		智能创新	智能创新力	人工智能研发水平	正向
	产业智能化	智能企业	人工智能企业	人工智能企业数	正向
		智能制造	机器人应用水平	工业机器人安装量	正向
			两化融合水平	工信部两化融合发展指数	正向

文章针对智能经济与新质生产力的相互关系，在构建新质生产力和智能经济测度指标体系的基础上，基于2011—2023年中国31个省份的面板数据，通过中介效应检验、调节效应检验、门限效应检验等，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；在智能经济对新质生产力的作用过程中，营商环境具有显著的调节作用和门限效应。

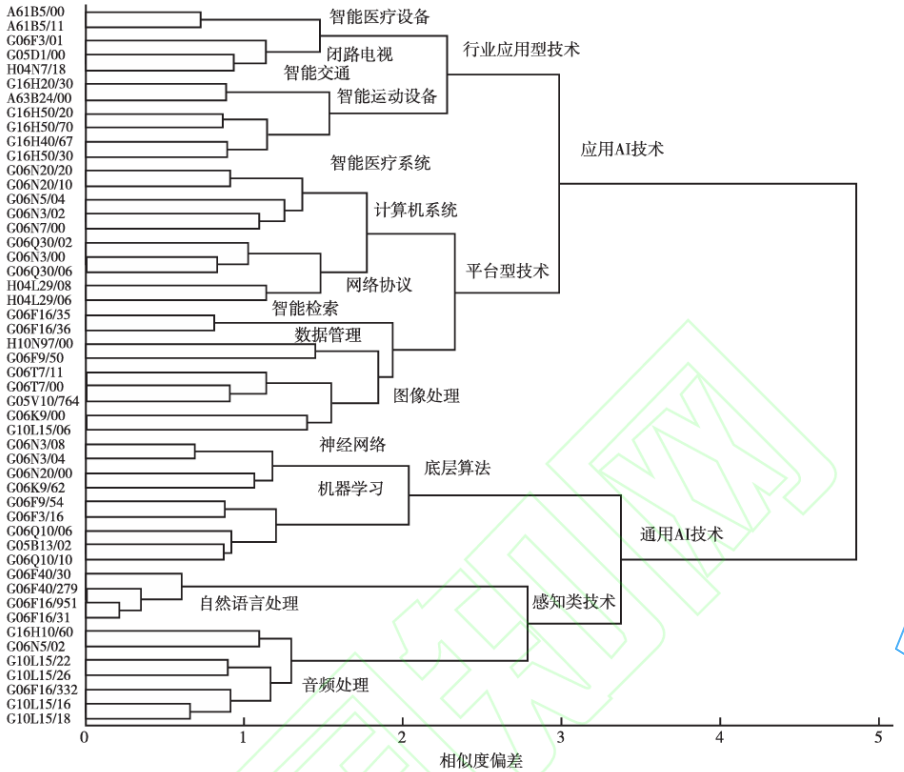
变量	New_qp	
	(1)	(2)
Int_e	0.378 3*** (17.93)	0.154 8*** (5.91)

中介效应检验结果

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			0.002 2** (2.03)	
Adv_is				0.109 6** (2.32)

CSMAR数据库应用：创新探索

相关研究10：企业人工智能技术水平的分层测度及其创新效应 ——基于全球知识网络视角[J].中国工业经济, 2025.



基于 Jaccard 相似度的 AI 技术层次聚类

文章基于复杂经济学视域，将AI凝练为复杂嵌套的层级知识系统，构建了AI领域的全球知识共现分层网络，创造性地提出了企业AI技术水平的分层测度方法；根据Jaccard相似度，将AI技术层次划分为通用AI与应用AI两大社区以及底层算法、感知类、平台型、行业应用型技术四大聚类。基于中国制造业上市公司的实证分析发现：通用AI在企业间呈现较强的幂律特征，企业AI技术构成的重心正由平台型技术向底层算法迁移。

$$ln_Innov_{i,t} = \beta_0 + \beta_1 ln_AI_{i,t} + \boldsymbol{\beta X} + \gamma_i + \delta_t + \varepsilon_{i,t}$$

变量	(1)	(2)	(3)	(4)
	ln_Innov	*ln_Innov*	*ln_Innov*	*ln_Innov*
ln_AI	2.6339***	1.9666***	1.9731***	0.6529***
	(0.2316)	(0.1679)	(0.1662)	(0.1073)

企业AI技术水平的测度逻辑在于识别企业知识结构与AI知识系统之间的结构性耦合程度。为此，文章提取中国制造业上市公司累积专利形成的知识库与AI知识系统的耦合关系，并通过加权求和的方式，将该耦合关系在一维坐标系中进行投影与量化，从而构建出反映企业AI技术嵌入深度的动态指标。

AI知识网络中底层枢纽知识节点不仅在网络中具有领先的加权PageRank值，而且向其他层辐射扩散。因此，文章立足底层枢纽知识节点，将其作为衡量企业AI技术水平的得分来源。受文本分析的词频统计启发，文章提取了企业所有专利中AI领域底层枢纽知识节点的出现频次。同时，逐一解析上市企业所涉年份的所有申请或授权的发明IPC分类号的小组信息，并按照枢纽知识节点的加权PageRank值计算上市公司AI技术水平得分。例如，上市公司申请的专利A的IPC分类号中出现了G06N20/00知识，则该企业的AI水平得分增加其对应的加权PageRank值，即得分增加0.028。最后，将上市公司特定年份所有发明专利的AI相关技术得分进行汇总，得到其相关年份的AI技术水平。进一步地，本文依据Jaccard相似度对AI技术网络进行四类聚类划分，并分别计算每个聚类中企业的AI得分，从而实现企业AI技术水平的分层测度。在处理时，综合考虑三个方面：①知识时效性。AI 技术迭代快，旧知识价值随时间降低，衰减前一年得分能突出当年新获取知识的重要性。②知识存续性。企业过往积累的 AI 知识仍有价值，保留前一年得分一定比例可体现技术发展连贯性。③平滑数据波动。企业 AI 技术发展受多种因素影响易产生数据波动，此种处理能减少偶然因素干扰，使测度结果更稳定可靠。因而，在处理时将前一年得分经衰减30%处理后加上当年的得分值，得到企业当年AI技术水平。

CSMAR数据库应用：创新探索

相关研究11： 人工智能与产业技术融合对制造业企业产品创新绩效的影响研究[J].中国工业经济, 2025.

$$ProductInno_{it+1} = \alpha_0 + \alpha_1 AITC_{it} + \sum \alpha_k Controls_{it} + \mu_i + \gamma_t + \varepsilon_{it}$$

被解释变量	关键解释变量
企业产品创新绩效：采用产品商标数据衡量。通过统计企业已注册且非营销用途的产品商标数量来衡量企业产品创新绩效。商标数据来源于国家知识产权局商标局	人工智能与产业技术融合：基于专利引用与专利IPC共现信息，反映人工智能技术知识在制造业技术体系中的转移与整合程度，衡量人工智能与产业技术融合

文章以中国制造业上市公司为研究对象，将人工智能与产业技术融合界定为企业跨越技术边界进行技术整合的行为，检验其对企业产品创新绩效的影响及作用机制。研究发现，人工智能与产业技术融合提升了制造业企业产品创新绩效。既有人工智能技术与产业技术融合对产品创新绩效的促进作用强于新引入人工智能技术与产业技术融合；企业人工智能技术关联度对人工智能与产业技术融合和产品创新绩效之间的关系有正向调节作用。

	(1)	(2)	(3)	(4)
	ProductInno	*ProductInno*	*ProductInno*	*ProductInno*
AITC	0.1253***			
	(0.0211)			
NAITC		0.1251***		0.0719***
		(0.0249)		(0.0276)
EAITC			0.1279***	0.0983***
			(0.0226)	(0.0251)

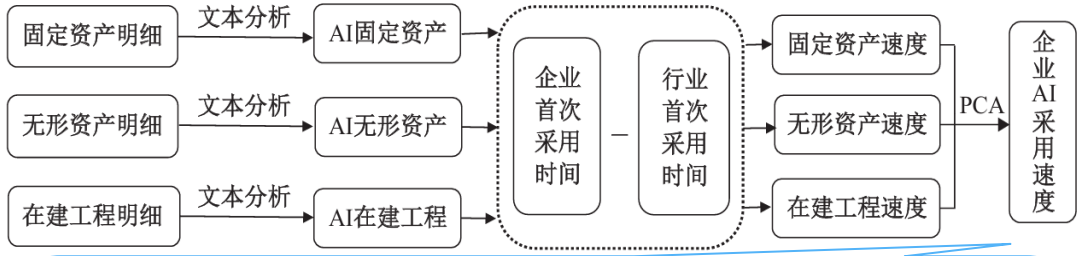
①剔除人工智能制造业企业，以反映技术融合的“跨技术领域”特征。基于年报经营范围，根据世界知识产权组织（WIPO）发布的《PATENTSCOPE Artificial Intelligence Index》，构建人工智能关键词表，利用BERT语言模型识别并剔除包含人工智能业务的制造业企业。
②构建人工智能与制造业技术专利IPC共现指标。依据《国际专利分类与国民经济行业分类参照关系表（2018）》，界定制造业技术的专利IPC集合；结合《PATENTSCOPE Artificial Intelligence Index》《战略性新兴产业分类与国际专利分类参照关系表（2021）》《关键数字技术专利 分类体系（2023）》，界定人工智能技术的专利IPC集合。若制造业企业所申请的专利同时包含以上两类专利IPC，标记为技术耦合专利。
③识别企业是否引用人工智能技术专利。若技术耦合专利的引用中包含人工智能专利，视为1次人工智能与产业技术融合行为。将专利层面的数据汇总至企业层面并取自然对数，构建人工智能与产业技术融合变量（AZTC）。
文章将人工智能与产业技术融合划分为NAITC与EAITC两类：基于四位数专利IPC， 使用5年期滚动窗口构建企业引用人工智能专利的对比池（AIIPC5Y）。若t年融合专利所引用的人工智能专利IPC（AIIPC）全部包含于AIIPC5Y中，界定为既有人工智能技术与产业技术融合专利（EAITC）；若至少有1个IPC未包含于AIIPC5Y中，则界定为新引入人工智能技术与产业技术融合专利（NAITC）。将以上两类专利数量在企业层面加总并取对数，得到EAITC与NAITC。

CSMAR数据库应用：创新探索

相关研究12： “争先恐后”还是“后发制人”？——企业人工智能采用速度与竞争优势[J].外国经济与管理, 2026.

$$Epcm_{i,t} = \beta_0 + \beta_1 AIspeed_{i,t}^2 + \beta_2 AIspeed_{i,t} + \beta_3 Control_{i,t} + Year + Pro + Ind + \varepsilon_{i,t}$$

被解释变量	关键解释变量
企业竞争优势：经行业中位数调整的勒纳指数 $LI_{i,j,t} = \frac{营业收入 - 营业成本 - 销售费用 - 管理费用}{营业收入}$ $Epcm_{i,j,t} = LI_{i,j,t} - Median(\{LI_{k,t} \mid k \in I_{j,t}\})$，其中$I_{j,t} = \{i企业i在年份t属于行业j\}$	AI采用速度：识别出与AI相关的固定资产、无形资产和在建工程，按时间顺序将首次金额不为0的年份确定为企业采用AI的时间。



不妨考虑在专用设备制造业（C35）中无形资产明细中AI技术的采用，赛象科技（002337）于2009年最先报告采用了AI技术，因此它的采用速度是0。而天桥起重（002523）于2010年在无形资产中报告采用了AI技术，较行业领先者落后了1年，因此采用速度是−1。尽管赛象科技在无形资产报告中采用了AI技术，但是它在2009年的其他会计科目中没有报告采用了AI，因此其采用速度向量为[0，−23，−23]。综合指数（AIspeed）的PCA得分为2.936。天桥起重（002523）在2008年的在建工程中报告了采用了AI技术，它的速度向量为[−23，−2，−23]，其PCA得分为2.657。该分数表明，天桥起重比赛象科技较慢地采用了AI技术。

文章基于复杂经济学视域，将AI凝练为复杂嵌套的层级知识系统，构建了AI领域的全球知识共现分层网络，创造性地提出了企业AI技术水平的分层测度方法；根据Jaccard相似度，将AI技术层次划分为通用AI与应用AI两大社区以及底层算法、感知类、平台型、行业应用型技术四大聚类。基于中国制造业上市公司的实证分析发现：通用AI在企业间呈现较强的幂律特征，企业AI技术构成的重心正由平台型技术向底层算法迁移。

变量	(1) Epmc	(2) ROIC	(3) Epmc	(4) ROIC
$AIspeed^2$	−0.0095***	−0.0042**	−0.0084***	−0.0056***
	(0.0017)	(0.0017)	(0.0016)	(0.0016)
$AIspeed$	0.0212***	0.0088**	0.0191***	0.0134***
	(0.0034)	(0.0037)	(0.0031)	(0.0036)
IT	0.0088*	0.0028		
	(0.0046)	(0.0068)		
$AIspeed^2 \times IT$	0.0176***	0.0304***		
	(0.0058)	(0.0097)		
$AIspeed \times IT$	−0.0360***	−0.0713***		
	(0.0129)	(0.0219)		
$Digital$			0.0012***	−0.0021***
			(0.0001)	(0.0006)
$AIspeed^2 \times Digital$			0.0009***	0.0034***
			(0.0002)	(0.0011)
$AIspeed \times Digital$			−0.0026***	−0.0087***
			(0.0005)	(0.0024)

CSMAR数据库应用：创新探索

案例三：气候与环境

指标类型

统计指标

统计指标：ESG、绿色创新、绿色投资、环境信息披露、碳排放、……

1

文本指标

气候关注、环境关注、……

2

关联CSMAR子库
商道融绿ESG评级：收录了沪深300成分股公司基本信息、ESG评级统计、行业评级统计、股票与财务方面的数据。
绿色专利：提供绿色产业、绿色专利（包括授权、引用及被引用）及相关上市公司的数据。
环境研究：主要包括上市公司环境绩效、环境披露、环境投资、企业排污许可等方面的数据。
碳中和：包含环境概况、碳排放、碳交易、碳治理、概念股、碳债券、碳基金、碳指数及碳信贷，共九部分内容，提供了宏观层面的污染排放情况，地区及行业二氧化碳排放量，全国碳排放权交易信息，上市公司层面的碳排放及碳减排信息，碳中和概念股，碳中和及低碳相关债券，碳中和及低碳主题基金，银行低碳信用卡及绿色贷款等内容。
绿色供应链：包含了绿色供应链、绿色转型、碳信息披露、绿色制造名单四大部分内容。
气候风险：包含了各省份气候风险、上市公司气候风险、自然灾害及突发事件三大方面内容。数据来源主要来自ISETS能源金融网络、中国环境统计年鉴、中国统计年鉴、上市公司定期报告。
环境注意力：收录了国务院、省级、地级市政府工作报告。库中不仅涵盖国务院、省级、地级市政府工作报告基本信息表，还设有环境注意力关键词词频句频统计表。其字段包含报告内容、关键词名称、关键词类别名称、关键词类别编码、词频、句频等。

CSMAR数据库应用：创新探索

案例三：气候与环境

[1] ESG评级分歧对企业绿色并购的影响研究[J].会计研究, 2025.	ESG
[2]低碳试点、供应链传导与企业绿色创新[J].经济管理, 2025.	绿色创新
[3]挤出抑或挤入：地方政府债务何以影响企业绿色投资[J].金融经济学研究, 2025.	绿色投资
[4]会计师事务所环境专长与企业环境信息披露[J].审计研究, 2025.	环境信息披露
[5]数字经济发展如何影响区域碳排放强度？——基于省级面板数据的实证分析[J].统计研究, 2026.	碳排放（区域）
[6]工业机器人、ESG信息披露与企业碳减排——来自专精特新上市公司的实证[J].中国人口·资源与环境 , 2025.	碳排放（企业）
[7]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.	气候关注
[8]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.	环境关注

References

CSMAR数据库应用：创新探索

相关研究1：ESG评级分歧对企业绿色并购的影响研究[J].会计研究, 2025.

$$GreenMA_{i,t+1} = \beta_0 + \beta_1 ESGdif_{i,t} + \sum Controls_{i,t} + \sum Year_i + \sum Industry_{i,t} + \varepsilon_{i,t}$$

被解释变量	关键解释变量
手工搜集并整理上市公司的并购公告，采用内容分析法，对其并购背景与目的、收购方公司及标的公司的业务范畴以及并购行为对收购方产生的潜在影响进行分析，设置虚变量GreenMA，若收购方公司在当年实施绿色并购，则取值为1，否则为 0	选取Wind、华证、富时罗素、商道融绿、盟浪 FIN-ESG以及彭博六家机构的ESG评级结果，计算各机构 ESG 评级分值之间的标准差，从而得到企业的ESG评级分歧

	(1)	(2)
	GreenMA	*GreenMA*
ESGdif	0.552***	0.578***
	(4.063)	(4.153)

文章基于风险感知理论，实证考察ESG评级分歧对企业绿色并购行为的影响。研究发现，ESG评级分歧显著提高绿色并购的概率，主要通过增强绿色监管问询、强化媒体ESG监督及提高投资者关注发挥作用。当绿色驱动因素较强、资源能力较强时，ESG评级分歧对绿色并购的影响更显著。进一步分析发现，管理层实施绿色并购仅是一种应激反应，并未带来实质性绿色技术突破，反而损害了绿色绩效。

项目	字段内容	数据起始时间	更新频率
上市公司基本信息年度表	公司全称、行业名称、行业代码、注册具体地址、公司成立日期、首次上市日期、所属省份、所属城市、经营范围等内容	2014-	年
ESG个股评级统计表	记录个股ESG评分及评级，包括评级年份、评级日期、股票代码、股票简称、公司全称、所属指数名称、成分股标识、证监会行业、融绿一级行业融绿二级行业、评级、ESG总评分、E评分、S评分、G评分、ESG管理评分、ESG风险评估评分、特殊说明	2015-	年
行业评级统计表	评级年份、行业分类标准、行业名称、评级总数量、A+评级数量、A评级数量、A-评级数量、B+评级数量、B评级数量、B-评级数量、C+评级数量、C评级数量、C-评级数量、D评级数量、ESG最高分、ESG最低分、ESG平均分、E平均分、S平均分、G平均分、ESG管理平均分、ESG风险评估平均分	2015-	年

CSMAR数据库应用：创新探索

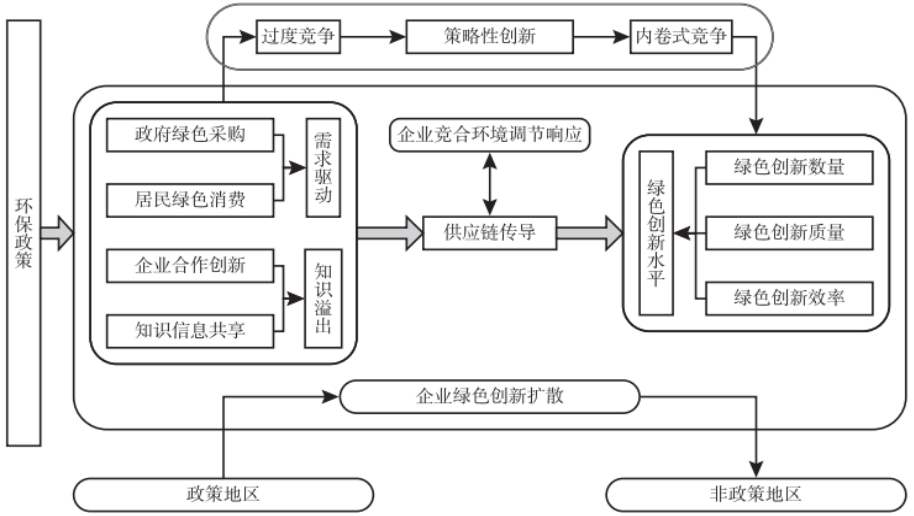
相关研究2：低碳试点、供应链传导与企业绿色创新[J].经济管理, 2025.

$$inva_{i,t} = \alpha_0 + \alpha_1 did_{i,t} + \alpha_2 Ctr_{i,t} + \mu_t + \gamma_i + \varepsilon_{i,t}$$

被解释变量	关键解释变量
绿色创新数量：上市公司获得的绿色发明专利数量的对数 绿色创新质量：根据绿色发明专利分类号计算出的知识宽度 绿色创新效率： ln(绿色发明专利获得数量 + 1)/ln(研发支出 + 1)	收集低碳省区和低碳城市试点名录及其确立时间，结合上市公司供应商和客户所在地信息，生成政策实施虚拟变量did，当上市公司的供应商和客户所在地被设立为政策试点地区时，did 及其随后年份为 1

变量	供应商			客户		
	amount	*quality*	*GIE*	*amount*	*quality*	*GIE*
	(1)	(2)	(3)	(4)	(5)	(6)
did	0.1401***	−0.0250**	0.0080***	0.0531***	0.0033	0.0016
	(2.9921)	(−2.1204)	(3.7397)	(3.5012)	(1.3577)	(0.6044)

文章探讨环保政策实施驱动企业绿色创新扩散的作用机理与实施效果，基于中国A股上市公司数据，手工匹配供应商和客户信息，从绿色创新数量、质量和效率三个维度衡量企业绿色创新水平，运用DID模型进行实证分析。研究发现：环保政策实施存在显著的供应链传导效应，能够影响非政策试点地区供应商和客户的绿色创新水平，且对上游供应商的影响明显强于下游客户；在过度竞争的市场环境下，非政策试点地区供应商企业倾向于采取策略性创新活动，导致企业绿色创新数量与效率提升，而创新质量受到抑制；环保政策的供应链传导效应主要通过需求驱动和知识溢出机制实现。



CSMAR数据库应用：创新探索

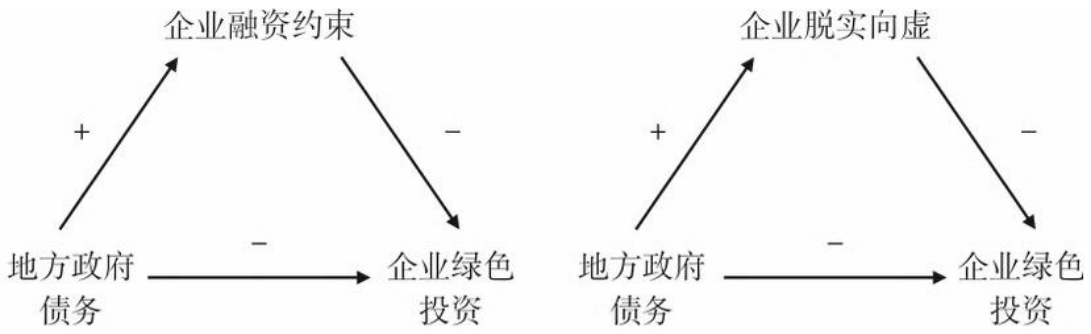
相关研究3：挤出抑或挤入：地方政府债务何以影响企业绿色投资[J].金融经济学研究, 2025.

$$EPI_{i,j,t} = \alpha_1 + \beta_1 Debt_{j,t} + + \theta_1 X_{i,j,t} + \delta_1 W_{j,t} + \mu_i + \lambda_t + \varepsilon_{i,j,t}$$

被解释变量	关键解释变量
企业绿色投资：（“在建工程”的环保类明细项目年度增加额）/总资产×100	地方政府债务负债率：各地融资平台公司债务余额/GDP

被解释变量	(1)	(2)	(3)
	绿色投资	绿色投资	绿色投资
debt_gdp	−0.929**	−0.904**	−1.011***
	(0.014)	(0.016)	(0.006)

文章基于地方融资平台债务数据与上市公司绿色投资数据，实证检验地方政府债务对绿色投资的影响。研究发现，地方政府债务对绿色投资具有显著的负向影响，即存在挤出效应。地方政府债务结构分解发现，非公开发行地方债显著抑制了企业绿色投资，存在挤出效应，而以城投债为主的公开发行债务不存在挤出效应。进一步城投债结构分解显示，绿色城投债促进了企业绿色投资，即存在挤入效应，绿色城投债占比越高，城投债绿色投资带动效应越显著。



CSMAR数据库应用：创新探索

相关研究4：会计师事务所环境专长与企业环境信息披露[J].审计研究, 2025.

$$Eidq_{i,t}=\alpha_0+\alpha_1 Aee_{i,t}+\Sigma Controls_{i,t}+\Sigma Year+\Sigma Ind+\varepsilon_{i,t}$$

被解释变量	关键解释变量
企业环境信息披露质量：将环境信息披露项目分为货币化和非货币化两类，分项赋值并取自然对数	会计师事务所环境专长：会计师事务所所审计的上市公司的排污当量数总和与当年所有上市公司的排污当量数总和的比值；当上述比值大于等于10%时，定义具有环境专长，取值为1，否则为0

基于CSMAR数据库中的环境研究数据库，并参考孔东民等（2021）、刘亦文等（2022）的做法，将企业在年报、社会责任报告及环境报告书中披露的环境信息分为货币化和非货币化两类，将不同类别项目得分总数加1取自然对数后得到企业环境信息披露质量的衡量指标。具体来说，与环境负债和环境业绩与治理相关的信息属于货币化信息，若企业进行了定量和定性结合披露赋值为2，若仅进行定性披露赋值为1，若没有披露则赋值为0；环境管理披露、环境认证披露和环境信息披露载体属于非货币化信息，若有披露赋值为2，否则赋值为0。

文章以2008-2022年我国A股非金融上市公司为研究样本，实证检验了会计师事务所环境专长对企业环境信息披露的影响。研究发现，会计师事务所环境专长对被审计单位环境信息披露质量具有显著的提升效应。披露载体差异检验发现，环境专长主要对年报中披露的环境信息发挥作用。此外，环境专长的作用在非国有企业、位于政府环境规制力度强、公众环境关注度高地区的企业中更为显著，会计师事务所环境专长引致的企业环境信息披露质量改善能进一步缓解企业融资约束并提升企业ESG表现。

变量	（1）	（2）	（3）	（4）	（5）	（6）
	Eidq	Eidq	Eidq	Eidq		Eidq
Aee1_continuous	0.923***(7.893)	0.192*(1.883)	0.310***(2.811)			
Aee2_dummy				0.087***(7.142)	0.044***(4.089)	0.044***(3.992)

分组检验：披露载体差异

变量	（1）	（2）	（3）	（4）	（5）	（6）
	仅年报中	既在年报又在社会责任报告/环境报告中	仅在社会责任报告/环境报告中	仅年报中	既在年报又在社会责任报告/环境报告中	仅在社会责任报告/环境报告中
Aee1_continuous	0.556***(3.413)	0.305(1.567)	0.161(0.236)			
Aee2_dummy				0.067***(4.073)	0.042**(2.159)	0.026(0.380)
Aee1_continuous diff	0.058	0.051	0.109	0.008	0.003	0.011

CSMAR数据库应用：创新探索

相关研究5：数字经济发展如何影响区域碳排放强度？——基于省级面板数据的实证分析[J].统计研究, 2026.

$$\ln(co_2)_{it} = \alpha_0 + \alpha_1 develop_{it} + \alpha_2 struct_{it} + \alpha_3 coal_{it} + \alpha_4 \ln(light)_{it} + \alpha_5 fdi_{it} + \alpha_6 credit_{it} + \alpha_7 urbanization_{it} + \mu_i + \lambda_t + \varepsilon_{it}$$

文章采用2003—2021年省级面板数据评估数字经济及其两大构成部分对区域碳排放强度的影响效应并探究其作用渠道。研究发现，数字经济有助于降低区域碳排放强度，但碳减排作用随数字经济发展阶段而变化；数字产业化和产业数字化均能降低碳排放强度，但产业数字化的碳减排作用小于数字产业化。

被解释变量	关键解释变量
碳排放强度：每万元地区生产总值产生的碳排放吨数	数字经济发展水平：构建测度数字产业化和产业数字化的指标体系并进行面板熵值法评分

一级指标	二级指标	测度指标
数字产业化	互联网普及率（*internet*）	每百人互联网用户数（户）
	互联网相关从业人数（*employee*）	计算机服务和软件从业人员占比（%）
	互联网相关产出（ln*output*）	人均电信业务总量（元）
	移动互联网普及率（*mobile*）	每百人移动电话用户数（部）
产业数字化	工业企业数字化设备（*computer*）	每百人使用计算机数（台）
	工业企业数字化应用（*web*）	每百家企业拥有网站数（个）
	电子商务普及率（*firm*）	开展电子商务交易活动的企业比重（%）
	电子商务销售规模（ln*sale*）	电子商务销售额（亿元）
	电子商务采购规模（ln*shop*）	电子商务采购额（亿元）
	数字普惠金融发展水平（*fintech*）	数字普惠金融指数

变量	(1)	(2)	(3)	(4)
develop	−0.015***（0.001）	−0.015***（0.002）	−0.009***（0.001）	−0.008***（0.002）
控制变量	否	否	是	是
省份效应	是	是	是	是
时间效应	否	是	否	是

变量	*digital*（1）	*internet*（2）	*employee*（3）	ln(*output*)（4）	*mobile*（5）
digital	−0.010***（0.002）	−0.002（0.001）	−0.041***（0.009）	−0.106***（0.034）	−0.003***（0.001）

变量	*industry*（1）	*computer*（2）	*web*（3）	*firm*（4）	ln(*sale*)（5）	ln(*shop*)（6）	*fintech*（7）
industry	−0.005***（0.002）	0.014***（0.003）	−0.002（0.002）	−0.006（0.005）	0.030（0.022）	0.017（0.020）	−0.007***（0.001）

CSMAR数据库应用：创新探索

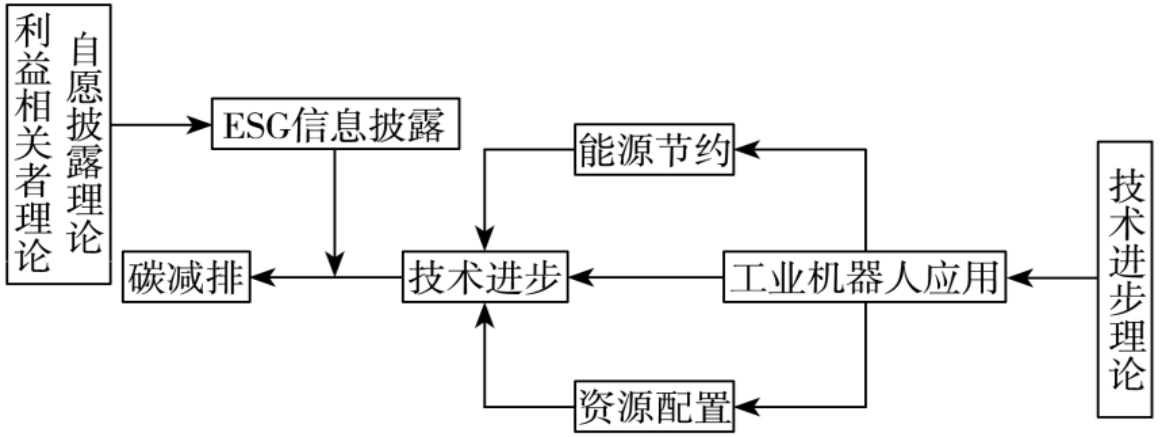
相关研究6：工业机器人、ESG信息披露与企业碳减排——来自专精特新上市公司的实证[J].中国人口·资源与环境，2025.

$$Y_{i,t} = \alpha_0 + \alpha_1 X_{i,t} + \alpha_2 C_{i,t} + \lambda_i + \gamma_t + \varepsilon_{i,t}$$

被解释变量	关键解释变量
碳排放水平：烟尘排放量 烟尘是企业排放的主要污染物之一，并且与企业碳排放具有协同排放效果，其排放量能够在一定程度上反映碳排放水平的高低	工业机器人应用：企业层面工业机器人渗透度

变量	Y	
	(1)	(2)
X	−0.004*	−0.004*
	(0.002)	(0.002)

文章基于2014—2022年专精特新上市公司数据，实证考察工业机器人应用对企业碳排放的影响机制。研究发现：(1)专精特新上市公司工业机器人应用程度的提升能够显著减少碳排放量。(2)从作用机制来看，专精特新企业通过提高工业机器人渗透度能够显著减少能源消耗、提高企业资源配置效率，进而助力企业绿色发展。(3)企业ESG信息披露水平的提高对工业机器人应用的碳减排效应具有正向调节作用。



CSMAR数据库应用：创新探索

相关研究7：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.

管理层对气候变化关注

E	S	G
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。	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度和信任度，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。	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。

企业ESG表现提升

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。

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

被解释变量	关键解释变量
华证ESG得分	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

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

上市公司气候风险 — 上市公司气候风险信息披露指数（CPR_ClimRiskDiscIndex）

样本数据

统计截止日期	上市公司ID	证券代码	证券简称	气候风险词数1	气候风险指数1	气候风险词数
2023-12-31	101704	000001	平安银行	336	0.6862	417
2023-12-31	101775	000002	万科A	128	0.316	109
2023-12-31	102294	000004	国华网安	74	0.1723	356
2023-12-31	102296	000006	深振业A	54	0.1938	66
2023-12-31	102313	000007	*ST 全新	45	0.142	51
2023-12-31	102314	000008	神州高铁	71	0.2412	219
2023-12-31	102336	000009	中国宝安	447	0.9581	1314
2023-12-31	102338	000010	美丽生态	218	0.6292	275
2023-12-31	101859	000011	深物业A	133	0.3265	211
2023-12-31	101860	000012	南玻A	529	1.5821	883
2023-12-31	101621	000014	沙河股份	19	0.0918	24
2023-12-31	101623	000016	深康佳A	279	0.6677	960
2023-12-31	101711	000017	深中华A	64	0.1768	122
2023-12-31	101629	000019	深粮控股	67	0.1759	154

CSMAR数据库应用：创新探索

相关研究8：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.

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

被解释变量	关键解释变量
华证ESG得分	环境相关词汇在政府工作报告总词数中所占比重

变量	(1)	(2)	(3)
	ESG		
GEA	23.661**	14.122*	24.216**
	(0.026)	(0.074)	(0.021)

变量	单因子分析		
	E	S	G
GEA	1.818	80.454***	19.490
	(0.899)	(0.000)	(0.204)

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。

环境注意力词频明细表 — 省级环境注意力关键词词频明细表（REA_ProvWordFreqDet）

样本数据

省份名称	报告年份	关键词名称	关键词类别编码	省份代码	关键词类别名称	词频
上海市	2009	污水治理	1	310000	环境保护类	1
上海市	2009	环境质量	1	310000	环境保护类	1
上海市	2009	空气质量	1	310000	环境保护类	1
上海市	2009	落后产能	1	310000	环境保护类	1
上海市	2009	水源	1	310000	环境保护类	1
上海市	2009	清洁	1	310000	环境保护类	3
上海市	2009	生态环境	1	310000	环境保护类	2
上海市	2009	环境治理	1	310000	环境保护类	2
上海市	2009	节能减排	1	310000	环境保护类	2
上海市	2009	污染	2	310000	环境污染类	6
上海市	2009	化学需氧量	2	310000	环境污染类	2
上海市	2009	二氧化硫	2	310000	环境污染类	2
上海市	2009	污染物	2	310000	环境污染类	1
上海市	2009	废物	2	310000	环境污染类	1

CSMAR数据库应用：创新探索

案例四：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

新质生产力测度

指标体系	文本分析
关联CSMAR子库	**关联CSMAR子库**
新质生产力：从企业新质生产力和地区新质生产力两个角度出发，参考相关文献，构建企业和地区新质生产力评价指标体系，测算出我国上市公司和各省份的新质生产力水平。	**新质生产力政府关注度**：提供中国省市县新质生产力政府关注度关键词数据，共3张表。数据库最早起始时间为2000年。数据来源为省市县政府工作报告。

CSMAR数据库应用：创新探索

案例四：新质生产力

[1]税收激励与企业新质生产力——基于双重机器学习的因果推断[J].财政科学, 2025.	指标体系（企业）
[2]人工智能政策何以赋能新质生产力？——来自国家新一代人工智能创新发展试验区的证据[J].科学学研究, 2026.	指标体系（地区）
[3]政府新质生产力关注与城乡收入差距[J].经济评论, 2025.	文本分析

References

CSMAR数据库应用：创新探索

相关研究1：税收激励与企业新质生产力——基于双重机器学习的因果推断[J].财政科学, 2025.

$$Npro_{it} = \theta_0 Tax_{it} + g(Control_{jit}) + \varepsilon_{it}$$

$$E(\varepsilon_{it} \mid Npro_{it}, Control_{jit}) = 0$$

文章以2012—2023年中国沪深A股上市公司为研究样本，实证分析税收激励对企业新质生产力的影响及作用机制。结果表明：税收激励显著促进企业新质生产力的发展。机制分析表明，缓解融资约束和提升综合创新水平是税收激励促进企业新质生产力发展的作用渠道。

被解释变量	关键解释变量
企业新质生产力指标：选取新型劳动者、新型劳动资料、新型劳动对象、要素优化四方面共14个指标，采用熵值法测度	企业所得税税率与实际税率的差值

一级指标	二级指标	指标说明
新型劳动者	高学历人员占比	本科以上人数 / 员工人数
	研发人员占比	研发人员数 / 员工人数
	研发薪酬占比	（研发费用 – 职工薪酬）/ 营业收入
	培训费用占比	员工培训费用 / 营业收入
新型劳动资料	研发费用占比	研发费用(剔除研发薪酬) / 营业收入
	数字无形资产占比	数字资产 / 无形资产
	固定资产比重	固定资产 / 资产总额
	人工智能采纳程度	机器设备账面价值 / 员工数量
	工业机器人渗透度	工业机器人数量 / 员工数量
新型劳动对象	数据资产占比	（资产市值 – 固定资产 – 金融资产 – 无形资产）/ 总资产
	数据要素利用水平	Ln(1+ 数据要素利用词频)
要素优化组合	产学研合作	Ln(1+ 产学研合作专利数)
	数字化转型	企业数字化转型指数
	企业 ESG 表现	企业 ESG 得分

变量名	（1）新质生产力	（2）新质生产力	（3）新质生产力
税收激励	0.375***	0.359***	0.404***
	（0.083）	（0.086）	（0.086）

项目	表/字段内容	数据起始时间	更新频率
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。	2015-	年
企业新质生产力三级指标表（实质性要素、渗透性要素）	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、数字技术投资、企业总污染当量、研发折旧摊销占比、研发租赁费占比、研发直接投入占比、发明专利申请数量、实用新型申请数量、生产数字化、组织数字化、生产组织绿色化_上市公司当年申请的绿色发明专利数量、生产组织绿色化_上市公司当年申请的绿色实用新型专利数量、企业数据要素。	2010-	年

CSMAR数据库应用：创新探索

相关研究2：人工智能政策何以赋能新质生产力？——来自国家新一代人工智能创新发展试验区的证据[J].科学学研究, 2026.

维度	指标	解释与来源	方向
新劳动者	战略新兴行业从业人数（人）	战略新兴产业和未来产业上市公司的总员工数，数据来源于企业年报	+
	在岗职工平均工资（元）	反映当地人才的平均能力，数据来源于各市统计年鉴	+
	普通高等学校学校数（所）	数量越多，提升劳动力素质的重要院校越多，反映当地高素质教育能力，数据来源于各市统计年鉴	+
新劳动资料	互联网宽带接入用户（千户）	数据来源于各市统计年鉴	+
	电信业务总量（亿元）	数据来源于中国城市统计年鉴	+
新劳动对象	机器人安装密度	城市层面的机器人安装密度，数据来自国际机器人联合会（IFR）	+
	环境污染治理投资（亿元）	数据来源于各市环境统计年鉴	+
	碳交易、用能权交易以及排污权交易金额（亿元）	数据根据各地级市门户网站披露整理	+
	生活垃圾无害化处理率	数据来源于中国城市统计年鉴	+
	人工智能企业数量（家）	数据来源于天眼查	+
新技术和新生产组织	科学支出占地方财政支出的比重（%）	数据来源于中国城市统计年鉴	+
	当年申请的发明数量（项）	数据来源于国家专利总局	+
	当年申请的实用新型数量（项）	数据来源于国家专利总局	+
	当年申请的绿色发明数量（项）	数据来源于国家专利总局	+
	当年申请的绿色实用新型数量（项）	数据来源于国家专利总局	+
数据要素	数据要素利用水平	采用城市层面上市公司数据资产相关词频来衡量，对该指标汇总后 + 1 再取对数	+
	有无数据交易平台（取值为 1 代表有）	数据来自于各市数据交易所设置情况	+

文章以2019年国家新一代人工智能创新发展试验区渐进推行作为准实验，定量考察人工智能政策对包含新劳动者、新劳动资料、新劳动对象、新技术新生产组织及新数据要素的多维新质生产力的影响。交叠双重差分估计结果表明，人工智能政策至少促进了5.84%的新质生产力水平提升。人工智能政策通过技术创新赋能、要素配置优化和产业升级催化促进新质生产力。

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。	2010-	年
地区新质生产力三级指标表(劳动者、劳动对象、劳动资料)	统计年度、地区代码、地区名称、人均GDP、在岗职工平均工资、第三产业就业比重、人均受教育年限、教育经费强度、在校学生结构、创新人力投入、创业活跃度、企业信息化水平、机器人安装密度、绿色资源、环境保护力度、污染防治质量_单位GDP化学需氧量排放强度、污染防治质量_单位GDP二氧化硫排放强度、绿色发明成果、传统基础设施_公路里程、传统基础设施_铁路里程、数字基础设施_光纤长度、数字基础设施_人均互联网宽带接入端口数、能源强度、能源消费结构低碳化指数、污染防治潜力、人均专利数量、新产品经济投入、数字经济、企业数字化。	2010-	年

CSMAR数据库应用：创新探索

相关研究3：政府新质生产力关注与城乡收入差距[J].经济评论，2025.

$$Theil_{it} = \alpha_0 + \alpha_1 Attn_{it} + \alpha_2 Attn_{it}^2 + \alpha_3 \sum Control_{it} + \mu_i + \delta_t + \varepsilon_{it}$$

文章通过构建城乡二元结构的经济增长模型，考察了政府新质生产力关注对城乡收入差距的非线性影响，并利用2009-2021年280个城市的面板数据进行实证检验。 研究发现，政府新质生产力关注对城乡收入差距具有显著的正U型影响。

被解释变量	关键解释变量
城乡收入差距(Theil)：泰尔指数	政府新质生产力关注（Attn）：对政府工作报告进行关键词提取，采用关键词频数和对应字数的乘积除以报告总字数得到新质生产力关注的比重

变量	(1)	(2)	(3)	(4)	(5)
$Attn$	−0.019***	−0.035***	−0.027***	−0.014***	−0.011***
	(0.005)	(0.009)	(0.009)	(0.004)	(0.004)
$Attn^2$	0.003***	0.005***	0.004***	0.002***	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

年份	省份代码	城市代码	关键词	省份	城市	关键词数量
2022	370000	370500	高效	山东省	东营市	3
2023	370000	370500	高效	山东省	东营市	5
2024	370000	370500	高效	山东省	东营市	2
2004	640000	640500	高效	宁夏回族自治区	中卫市	2
2005	640000	640500	高效	宁夏回族自治区	中卫市	3
2006	640000	640500	高效	宁夏回族自治区	中卫市	2
2007	640000	640500	高效	宁夏回族自治区	中卫市	4
2008	640000	640500	高效	宁夏回族自治区	中卫市	1
2009	640000	640500	高效	宁夏回族自治区	中卫市	1
2010	640000	640500	高效	宁夏回族自治区	中卫市	6
2011	640000	640500	高效	宁夏回族自治区	中卫市	8
2012	640000	640500	高效	宁夏回族自治区	中卫市	5
2013	640000	640500	高效	宁夏回族自治区	中卫市	7
2014	640000	640500	高效	宁夏回族自治区	中卫市	7

CSMAR数据库应用：创新探索

案例五：大语言模型应用

已有指标重新测度

投资者情绪、管理层语调、企业数字化转型程度、企业数智技术指标、人工智能创新、ESG、……

创新指标挖掘构建

数字技术风险暴露、企业劳动力需求结构、 业财融合、网络新闻媒体报道、基金经理语调、……

BERT、GPT、ERNIE、……

CSMAR数据库应用：创新探索

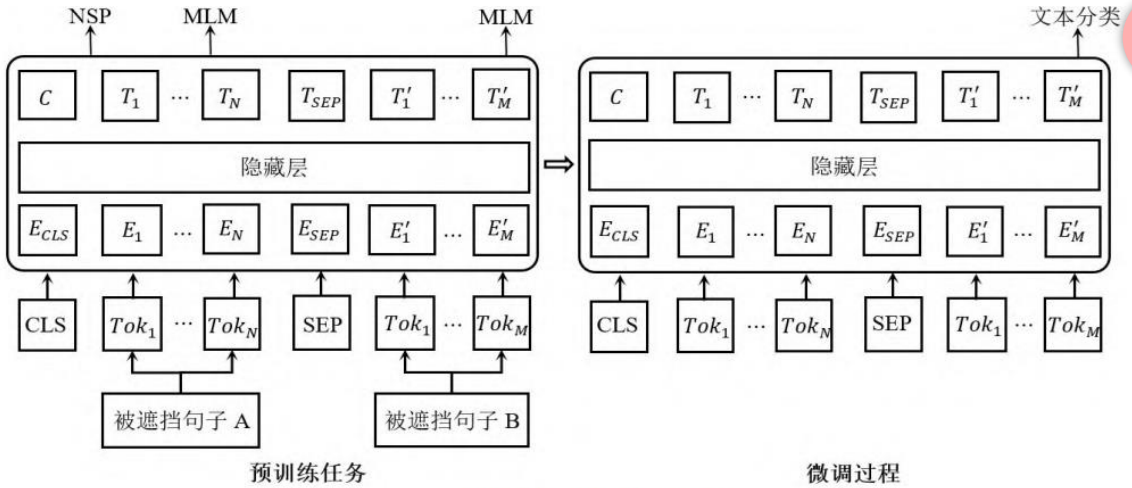
案例五：大语言模型应用

[1]大语言模型、文本情绪与金融市场[J]. 管理世界, 2024.	已有指标重新测度
[2]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.	
[3]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.	
[4]数智技术对专精特新企业韧性的提升作用——来自大语言模型的证据[J].华东经济管理, 2026.	
[5]人工智能创新如何降低供应链中断风险[J].中国流通经济, 2025.	
[6]人工智能技术创新如何赋能企业成长——来自BERT模型的证据[J].财经问题研究, 2026.	
[7]象征性还是实质性：合法性压力下的企业ESG行动策略研究[J].南开管理评论, 2025.	
[8]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.	创新指标挖掘构建
[9]外部技术断供、工作任务与企业劳动力需求结构[J].数量经济技术经济研究, 2025.	
[10]业财融合能否推动企业可持续发展？——基于大语言模型的研究[J].证券市场导报, 2025.	
[11]网络新闻媒体报道与企业环保投资行为——来自大语言模型文本分析的证据[J].证券市场导报, 2025.	

References

CSMAR数据库应用：创新探索

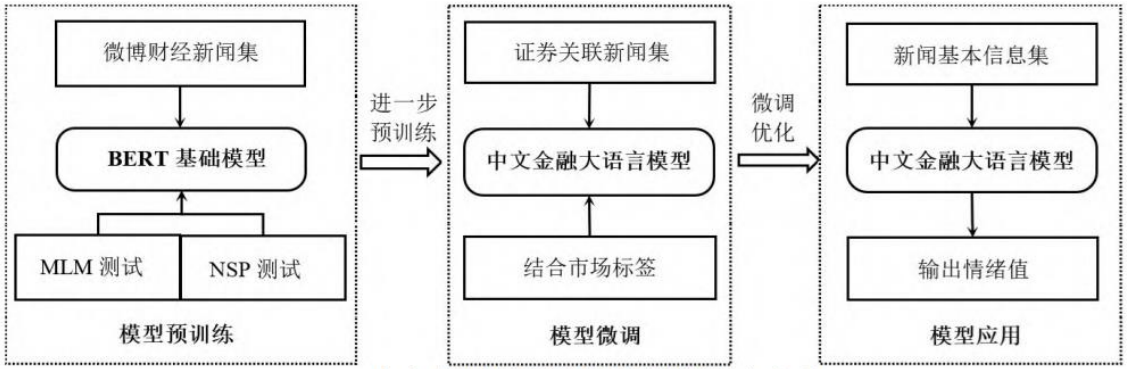
相关研究1：大语言模型、文本情绪与金融市场[J]. 管理世界, 2024.



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_{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。

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。



中文金融大语言模型总体应用流程图

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。

CSMAR数据库应用：创新探索

相关研究2：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.

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)

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。

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

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. R^2	0.132	0.129	0.126	0.128

CSMAR数据库应用：创新探索

相关研究3：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.

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

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)

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。
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
第四步：采用有监督的机器学习方法训练模型

CSMAR数据库应用：创新探索

相关研究4：数智技术对专精特新企业韧性的提升作用——来自大语言模型的证据[J].华东经济管理, 2026.

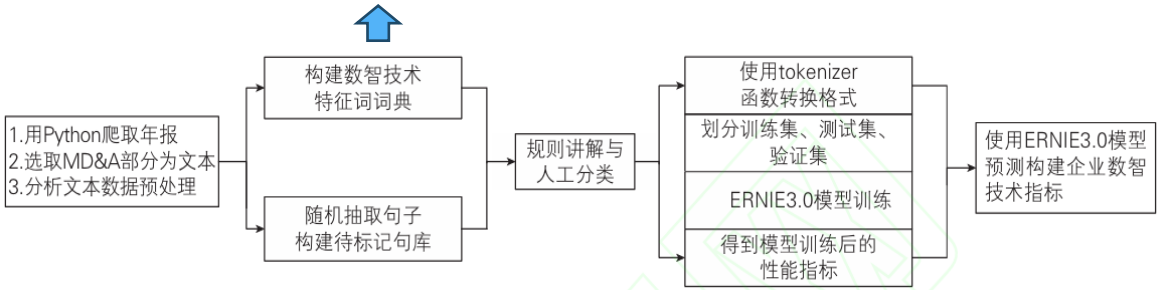
$$\mathrm{Res}_{it} = \alpha_0 + \alpha_1 \mathrm{DITech}_{it} + \alpha_2 X_{it} + \mu_t + \varphi_i + \varepsilon_{it}$$

被解释变量	关键解释变量
专精特新企业韧性：以专精特新企业三年内累计销售收入的增长额作为衡量其韧性的指标，增长额越高，表明企业在应对外部冲击时展现出更强的韧性	数智技术：使用大语言模型ERNIE3.0完成文本分类任务，进而测度企业数智技术指标

变量	(1)	(2)	(3)	(4)
	基准回归		工具变量法	
	Res	Res	第一阶段	第二阶段
DITech	0.077 7*** (4.32)	0.070 3*** (4.48)		1.088 3** (2.16)
Communication			0.002 2*** (6.15)	

文章以2013—2023年中国A股及新三板上市的专精特新企业为样本，运用ERNIE3.0大语言模型构造数智技术指数，实证检验数智技术应用对专精特新企业韧性的影响及内在机制。研究发现：数智技术显著提升了专精特新企业韧性。

将数智技术分为数字化技术3类（大数据、云计算、区块链）、智能化技术3类（人工智能、物联网、互联网+），并结合已有文献、研究报告等，最终整理出包含388个数智技术特征词的词典



ERNIE3.0大语言模型文本分析

CSMAR数据库应用：创新探索

相关研究5：人工智能创新如何降低供应链中断风险[J].中国流通经济, 2025.

$$SCDRisk_{it} = \alpha_0 + \alpha_1 AIINNO_{it} + \alpha_2 Control_{it} + \mu_i + \lambda_t + \varepsilon_{it}$$

被解释变量	关键解释变量
供应链中断风险：基于词嵌入模型计算供应链中断风险指标。）基于供应链中断风险的定义与来源，将“供应链”、“供应商”、“中断”等词汇作为供应链中断风险种子词	人工智能创新：基于深度语义建模方法，采用SBERT架构的大语言中文模型（BAAI/bge-large-zh-v1.5）将企业211万条人工智能专利文本向量化，并计算其与过去三年历史技术语料的语义相似度，从而识别专利的技术创新性程度。进一步以每项人工智能专利的创新性为权重进行加权聚合，构建企业人工智能创新指标

	(1)	(2)	(3)	(4)	(5)	(6)
人工智能创新(*AIINNO*)	−0.042 4*** (0.010 8)			−0.031 2*** (0.011 2)		
企业内创新(*AIINNO_IN*)		−0.058 8*** (0.014 1)			−0.043 8*** (0.014 7)	
企业外创新(*AIINNO_OUT*)			−0.042 3*** (0.010 8)			−0.031 2*** (0.011 2)

文章利用2013—2023年我国A股上市公司数据，通过SBERT架构的大语言模型构建企业层面人工智能创新指标，系统考察人工智能创新对企业供应链中断风险的影响以及作用机制。研究发现，人工智能创新能够增强企业需求预测与规划能力，减少供应链上下游企业的信息壁垒，从而显著降低企业供应链中断风险。人工智能创新主要通过提升企业声誉、提高企业智能化业务水平、降低代理成本三条路径降低供应链中断风险。

引入SBERT架构的大语言中文模型（BAAI/ bge-large-zh-v1.5），作为专利语义嵌入的基础模型，从根本上提升文本表示质量。

首先，利用BGE模型将企业所有专利的摘要文本编码为归一化语义向量。其次，参考黄先海等的研究，人工智能专利创新度（INNO）被定义为其与前三年内所有专利的平均语义相似度取倒数后的自然对数值：

$$INNO_{it} = \ln\left(\frac{1}{\frac{\sum_{j \in I} \rho_{ij}}{N_{I_t}} + 0.1}\right)$$

其中，i、j代表专利，t代表年份，I代表前三年内的所有发明专利集合，N表示专利数量，ρ表示专 利相似度。INNO表示人工智能专利创新度，即该值越大，说明一项人工智能专利技术中的文本内容中对已有专利的重复较少，而原创性的创新内容较多，即创新度较高。考虑到可能存在相似度 近似为0的情形，使结果严重左偏，在分母加入调整因子（常数0.1），以避免极端值问题。

最终，以人工智能专利创新度作为权重，将人工智能专利进行加权汇总到企业层面，作为人工智能创新（AIINNO）的代理变量。

该指标不仅从规模上衡量人工智能发展程度，也从质量上体现人工智能创新程度，可以较准确地测度企业人工智能应用的真实水平。

CSMAR数据库应用：创新探索

相关研究6：人工智能技术创新如何赋能企业成长——来自BERT模型的证据[J].财经问题研究, 2026.

$$growth_{it} = \alpha_0 + \alpha_1 ai_patent_{it} + \alpha_2 Controls_{it} + \mu_i + \varphi_t + \varepsilon_{it}$$

被解释变量	关键解释变量
企业成长：企业成长能力和企业规模。其中，企业成长能力维度包括总资产增长率、营业收入增长率和净资产收益率增长率，以捕捉企业的动态扩张潜力；企业规模维度包括总资产、营业收入和员工人数的自然对数，以反映企业的静态资源禀赋	人工智能技术创新：基于训练好的BERT模型，对上市公司在样本期内申请的专利进行批量自动分类，以判断该项专利是否属于人工智能专利

变　量	(1)	(2)
ai_patent	0.045^{***} (0.013)	0.038^{***} (0.011)

文章利用双向编码器表示模型识别人工智能专利，以刻画企业人工智能技术创新水平，并基于2012—2022年中国A股上市公司数据，利用双向固定效应模型实证检验了人工智能技术创新对企业成长的影响及作用机制。研究结果显示：人工智能技术创新能够促进企业成长。

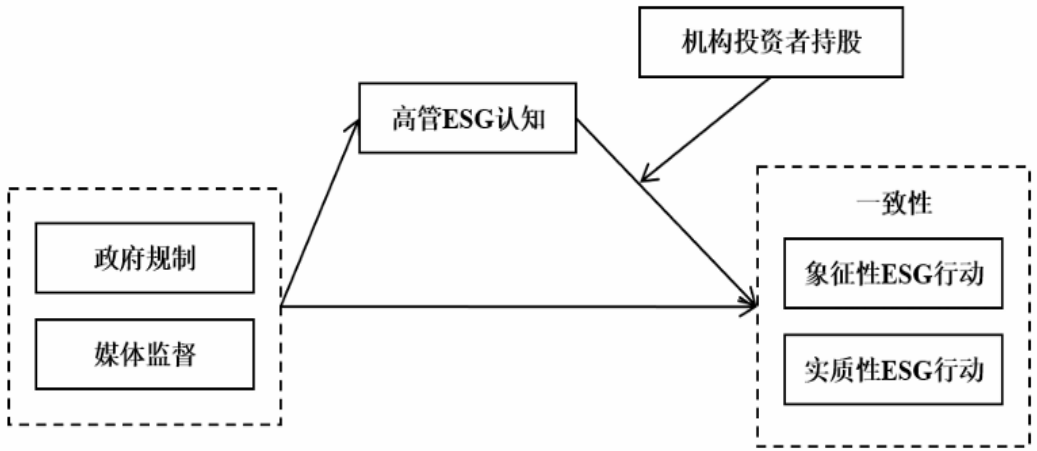
现有研究通过IPC识别人工智能专利，即将归属特定IPC类别的所有专利认定为人工智能专利，或采用文本分析法对人工智能专利的摘要分词处理后，通过构建关键词词典来匹配。

这两种方法虽能在一定程度上识别人工智能专利，但随着人工智能的不断更新迭代，原本的IPC分类或关键词词典构建不能很好地识别新的人工智能技术名词，导致估计结果的偏误。因此，应基于专利摘要的文本内容识别人工智能专利。但是，由于专利数量巨大，仅依赖人工逐一识别是一项浩大且漫长的工程。基于此，文章利用BERT模型识别人工智能专利。

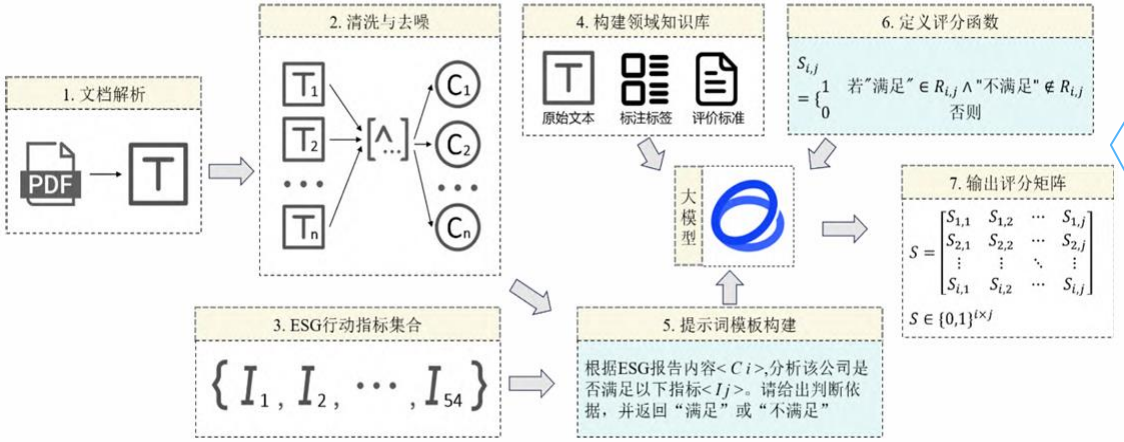
相关的技术过程如下：从专利样本中随机抽取一定数量的专利，通过人工识别的方式对这些专利的摘要文本进行分类，然后利用BERT模型对人工标注的专利进行学习，最后使用训练好的BERT模型对所有专利进行分类。

CSMAR数据库应用：创新探索

相关研究7：象征性还是实质性：合法性压力下的企业ESG行动策略研究[J].南开管理评论, 2025.



文章基于合法性理论，深入剖析了企业ESG行动的内涵，并将其解构为象征性与实质性行动，探讨企业如何动态调整ESG行动策略组合，以应对来自政府、媒体和投资者的多维合法性压力。研究开发了一种基于大语言模型的ESG行动测量方法，并借助方向性和非方向性差异（DNDD）框架和响应面分析，创新性地检验了企业象征性与实质性ESG行动之间的一致性及其影响。研究结果表明，激励性和惩罚性的政府规制均可通过提升高管的ESG认知，显著增强企业ESG行动，尤其对实质性行动起到更强的促进作用，从而规避企业采取更为隐蔽的“漂绿”行为。这一发现凸显了政府在引导企业采取实质性ESG行动中的关键作用。



基于大语言模型的 ESG 行动度量方法工作流程

文章将大语言模型应用于企业社会责任报告文本的分析，以实现多维度ESG行动的精细化度量和量化评分。

具体而言，该方法结合预设的ESG行动维度知识库，并通过指令优化提升模型在ESG行动提取任务中的表现，使其能够更有效地解析和理解长篇报告中的复杂语法结构与上下文依赖。在评价过程中，模型将报告内容与评分维 度整合为输入提示，依托其大规模参数容量生成针对性的评价依据，并据此严格区分“满足”与“不满足”的条件，从而实现对企业社会责任报告中真实ESG行动的精准提取。

文章究以企业社会责任报告为唯一评价依据，结合成熟的量表和第三方评级标准，构建了包含54个维度的评价体系。在数据准备阶段，作者收集了A股上市公司2007-2024年间发布的10992份社会责任报告的PDF文件链接，并通过随机抽样方法抽取其中5%的样本，参考金星晔等的方式进行人工标注，构建了包含原始文本、标注标签与维度评价标准的领域知识库。文章基于glm-4-flash 模型进行维度提取与评分。

CSMAR数据库应用：创新探索

相关研究8：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.

企业数字技术风险暴露变量的具体构造方式

1. 抓取数字技术风险相关的文本论述。
2. 标注训练集。
3. 模型预训练和预测识别。
4. 构建数字技术风险暴露程度指标

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、僵木蠕、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	

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。

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

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：创新探索

相关研究9：外部技术断供、工作任务与企业劳动力需求结构[J].数量经济技术经济研究, 2025.

$$LDS_{it} = \beta_0 + \beta_1 Blockade_{it} + X'_{i,jt}\gamma + \mu_i + \theta_t + \varepsilon_{it}$$

被解释变量	关键解释变量
企业劳动力需求结构。所使用的上市公司招聘大数据收集自各大知名互联网招聘信息平台。	企业受到外部技术断供冲击的政策虚拟变量：企业在某年开始被列入某项美国制裁清单则将该企业在当年及之后年份的核心解释变量Blockade均赋值为1，否则为0。

劳动力类型	职业类别	《大典》中类职业示例	典型标准职业
非常规劳动力	非常规分析型职业	科学研究人员、工程技术人员	数学研究人员、人工智能工程技术人员
	非常规互动型职业	经济和金融专业人员	会计专业人员、知识产权分析专业人员
常规劳动力	非常规操作型职业	居民服务人员	中式烹调师、家政服务员
	常规认知型职业	行政办事及辅助人员	秘书、营销员
	常规操作型职业	机械制造基础加工人员	车工、机床装调维修工

变量	(1)	(2)	(3)	(4)	(5)
	LDS	*LDS*	*LDS*	*LDS*	*LDS*
Blockade	0.420***	0.418***	0.422***	0.428***	0.428***
	(0.098)	(0.097)	(0.097)	(0.097)	(0.097)

文章以美国对华技术断供制裁清单的典型特征为切入点，采用两类人工智能领域较为前沿的Transformer架构大语言模型GPT和BERT，对互联网招聘大数据实施文本挖掘，实证检验外部技术断供对中国企业劳动力需求结构的影响及其作用机制。结果表明，整体上，外部技术断供通过正向任务配置效应，显著增加中国企业对非常规劳动力的相对雇佣需求，推动企业劳动力需求结构转型升级，然而，任务替代效应并未对企业劳动力需求产生明显影响。

首先，利用GPT架构的摘要提取优势，文章将非结构化招聘文本与《中华人民共和国职业 分类大典（2022年版）》（以下简称《大典》）中的标准化工作任务文本示例进行数据融合，通过 ChatGLM2 模型实施结构化转换。要求大模型解析原始招聘文本中的工作内容、技能要求等信息，并按照《大典》既定的工作任务语义框架和格式规范进行重新编码，生成具备计算价值的半结构化文本数据。

其次，将ChatGLM2生成的半结构化文本和《大典》中1639个标准职业的主要工作任务描述文本共同输入Sentence-BERT（SBERT）模型中，利用文本语义相似度算法 逐一比对招聘岗位与标准职业的工作任务相似度。遵循Imbert等（2022）的建议，文章筛选出各招聘岗位任务相似度最高的标准职业，继而依据经验法则将任务相似度超过0.8的“岗位—职业”组合予以保留（文本相似度的数值范围在0~1之间）。

最后，通过信息映射对筛选后的匹配组合进行处理，实现企业招聘岗位与《大典》标准职业的精准匹配。

CSMAR数据库应用：创新探索

相关研究10：业财融合能否推动企业可持续发展？——基于大语言模型的研究[J].证券市场导报, 2025.

$$TDP_{i,t}=\alpha_0+\alpha_1 BFII_{i,t}+\alpha_2 X_{i,t}+\sum Ind+\sum Year+\varepsilon_{i,t}$$

被解释变量	关键解释变量
可持续发展绩效：基于熵权TOPSIS法，使用ROA（财务绩效）与彭博ESG评分（非财务绩效）进行构建	业财融合：ln(业财融合总词频数+1)

	(1)	(2)	(3)
BFII	0.163***	0.070***	0.062***
	(33.87)	(15.87)	(14.25)

作为财务管理和业务流程深度融合的管理模式，业财融合能够提高企业经营效率、防控运营风险，但也有研究指出，业财融合对企业组织边界、业务发展和成本费用存在潜在负面影响。文章基于大语言模型，创新性地构建较为完备的业财融合词库，实证检验业财融合对企业可持续发展的影响及作用机制。研究发现，业财融合能够推动企业可持续发展。

第一，采集与业财融合相关的政策法规及政策解读、答记者问和学术文献等文本构建语料库。文章考虑到我国政策法规中直接提及业财融合概念的文本较少，仅以政策法规作为语料库难以全面涵盖业财融合的相关概念，因此进一步采集了对业财融合的本质进行界定或分析的学术文献、所选政策的相关解读、答记者问等文本纳入语料库，最终形成多元的语料库，提高指标的有效性。

第二，采用 Word2Vec 神经网络模型和深度学习技术构建基本词库。基于 Word2Vec 神经网络模型，对语料库中每个词的上下文进行分析，通过 Skip-gram 模型训练生成 100 维词向量空间，从而计算词语之间的相似度。筛选出与业财融合关键词相似度较高的词语，从而初步构建业财融合基本词库。

第三，基于 Ernie-bot 模型实现词库扩充与优化。鉴于 Word2Vec 模型生成的基本词库受限于语料库文本内容，其词汇覆盖范围存在局限性，影响指标的准确性，本文进一步使用 Ernie-bot 模型，将基本词库导入到大语言模型中，并设计提示指导模型通过其深度语义理解能力对基本词库中的每个词进行同义词和近义词扩展。经过去重处理后，最终构建了一个包含 262 个词汇的业财融合词典。

第四，将我国 A 股上市公司年报转化为文本文件，并将业财融合词典与年报文本进行匹配和词频统计，得到年报文本的词频总数和匹配的总词频，再加 1 后取自然对数，构建业财融合指标 (BFII)。BFII 数值越大，表明企业的业财融合程度越高。

CSMAR数据库应用：创新探索

相关研究11：网络新闻媒体报道与企业环保投资行为——来自大语言模型文本分析的证据[J].证券市场导报, 2025.

$$GI_{i,t}(GI_pre_{i,t},GI_end_{i,})=\alpha_0+\alpha_1 Media_{i,t}+\sum_j \alpha_j X_{i,t,j}+\mu_i+\gamma_t+\epsilon_{i,t}$$

被解释变量	关键解释变量
企业环保投资水平：从上市公司年报中的“在建工程”和“管理费用”科目获得与环保及绿色发展相关的投资金额总额，以此衡量企业的环保投资水平。	网络新闻媒体报道：采用企业当年在网络新闻媒体上关于环保主题的负面报道总次数来表征网络新闻媒体报道

	(1)	(2)	(3)
BFII	0.163*** (33.87)	0.070*** (15.87)	0.062*** (14.25)

文章运用GPT大语言模型对上市公司的财务报表和新闻报道进行文本分析，实证检验网络新闻媒体报道对企业前端预防型与末端治理型环保投资行为的影响及作用机理。研究发现，网络新闻媒体报道能显著提升企业整体环保投资水平，且主要表现为对末端治理型环保投资的驱动作用。

环保投资类型处理：
通过 Python 调用 OpenAI 提供的 GPT API 服务，对上市企业年报内容进行判断。若上市企业年报中的项目属于前端预防型环保投资，则赋值 1；若上市企业年报中的项目属于末端治理型环保投资，则赋值 0。
前端预防型环保投资水平 (GI_pre)：具体包括绿色产业项目、清洁和可再生项目、能源低碳高效利用项目、水资源高效利用项目、资源综合利用项目等从生产阶段进行的投资，以上述项目金额之和来衡量前端预防型环保投资水平。
末端治理型环保投资水平 (GI_end)：具体包括废弃治理项目、废水治理项目、脱硫项目、脱销项目、固废处理项目以及生产修复等对污染物生化处理的投资，以上述项目金额之和来衡量末端治理型环保投资水平。）

网络新闻媒体报道处理：
通过 Python 调用 OpenAI 提供的 GPT API 服务，对媒体报道内容进行判断。若新闻报道内容属于正面报道，则赋值 1；若新闻报道内容属于负面报道，则赋值 - 1；若新闻报道内容属于中立报道，则赋值 0。

03

CSMAR最新数据

- 最新数据库概览
- 新库资讯获取

最新数据资源

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
- 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
- 风险监测
- 数据要素
- 上市公司税收
- 上市公司高管团队领导力
- 税收激励与研发创新
- 高新技术企业

2026

- 绿色供应链
- ……

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。

	主题	讲师
2025	人工智能金融学前沿方向	陈坚 厦门大学
	内生性“能好怎”？战略管理实证研究的内生性问题初探	吴哲颖 复旦大学
	AI工具在会计研究中的应用	吴璇 哈尔滨工业大学（深圳）
	Implied Cost of Capital: What, Why and How?	顾俊 深圳大学
	数字化绿色化协同转型对出口企业绿色低碳创新的影响	王智新 河北大学
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

✓ 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
- ESG前沿研究速递&CSMAR气候风险数据
- 数字政府前沿研究速递&CSMAR各区县政府工作报告数据
- 权威高管团队数据助力企业管理研究
- 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
- 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

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/