西南财经大学
SOUTHWESTERN UNIVERSITY OF FINANCE AND ECONOMICS

课 程 论 文

论文题目：基于大语言模型的 A 股上市公司气候风险测度研究——来自 2024 年年报的证据

课程名称	人工智能大模型应用
任课教师	郑海超
姓　　名	王玉杰
学　　号	125027000004
学　　院	统计与数据科学学院
专　　业	统计学

2026 年 1 月 16 日

摘要

本研究聚焦于 2024 年《上市公司可持续发展报告指引》实施背景下的气候风险测度难题。利用 A 股全量年报数据，系统对比了增强词典法、FinBERT2 预训练模型与 DeepSeek 大语言模型在气候信息提取中的效能。研究发现：(1) 传统词典法存在严重的“长尾分布”与稀疏性缺陷，漏判了隐性风险；(2) 判别式模型 FinBERT2 陷入“二元极化”陷阱，且判定置信度显著偏低（均值仅 0.53），说明其在缺乏中性选项时被迫进行“犹豫的强制分类”；(3) DeepSeek 大模型凭借逻辑推理能力，成功剔除了 89.0% 的“漂绿”与合规口号，精准还原了实质性风险结构。本研究为大模型时代的金融文本分析提供了新的方法论证据。

关键词：气候风险；大语言模型；DeepSeek；FinBERT2；漂绿

1 引言

进入 21 世纪 20 年代，气候变化已演变为重塑全球宏观经济格局的重要力量。国际清算银行提出的“绿天鹅”概念指出，气候相关风险具有极高的不确定性，且其破坏力可能远超传统的周期性金融危机[1]。作为全球最大的发展中国家，中国不仅在政策层面明确了“3060”双碳目标，更在金融监管领域积极行动。中国人民银行与监管机构通过顶层设计推动资金向绿色领域流动，使得气候风险在金融系统中的传导机制成为学术界关注的焦点。

然而，2024 年成为了中国气候金融信息披露的关键转折点。随着沪深北交易所《上市公司可持续发展报告指引》的正式发布与实施，中国 A 股上市公司迎来了 ESG 信息披露的“爆发元年”。海量的非结构化文本数据在提升透明度的同时，也带来了前所未有的噪音与挑战。企业年报中充斥着复杂的语义表达，既有实质性的转型阵痛描述，也包含了大量避重就轻的“漂绿”话术[2] 以及形式主义的政策复述。

在这一新语境下，传统的文本分析方法面临失效的风险。既往研究多依赖“词袋法”，即构建包含“全球变暖”、“碳排放”等关键词的静态词典并统计词频。然而，这种方法在处理 2024 年的复杂文本时显得力不从心，它无法区分否定句式，也难以识别“新质生产力”等新兴概念背后的绿色内涵。更重要的是，简单的关键词匹配无法有效甄别企业是在披露实质性风险，还是在进行空洞的口号式宣传，这直接导致了气候风险测度的失真。鉴于此，如何从 2024 年海量的年报文本中精准、高效地提取气候风险信息，已成为实证金融研究亟待解决的首要技术难题。如果测度工具本身存在偏差，后续关于“气候风险影响系统性风险”的实证回归结果将失去可信度。

近年来，以 Transformer 架构为基础的大语言模型展现出了卓越的自然语言理解能力。其强大的上下文感知与零样本推理能力，为解决金融文本挖掘中的语义消歧难题提供了新的路径。基于此，本研究并未沿袭传统的“构建指标-实证回归”范式，而是聚焦于测度方法本身的有效性评估。本实验旨在 2024 年 A 股年报的新语境下，以 DeepSeek 为代表的国产大语言模型是否能够替代传统词典法，提供更准确的气候风险测度。

本研究突破了现有文献多关注“气候风险经济后果”的局限，转而深入探讨“气候风险测度方法”本身的科学性，通过对比增强词典法与大语言模型的性能差异，揭示了传统方法在处理新规下复杂披露文本时的系统性偏差。同时，研究紧扣指引发布后的最新市场变化，针对 2024 年特有的“漂绿”话术和政策术语进行了针对性的识别测试。此外，本研究还摒弃了高成本的模型微调路径，探索了基于“提示工程”的零样本推理方案，评估了 DeepSeek 等国产高性价比模型在金融文本分析中的应用潜力，为监管机构和投资者构建低成本、高精度的实时气候风险监测系统提供了实验依据。

2 文献综述

2.1 气候风险测度方法的演进与局限

早期关于气候风险的研究主要依赖碳排放量、环境评分等单一硬指标来衡量企业的风险暴露，但 Li 等人（2024）指出，这些指标往往滞后于市场定价，且难以捕捉政策不确定性带来的动态冲击[3]。随着研究的深入，基于文本分析的方法逐渐成为主流，其中“词袋法”应用最为广泛。Wang 等人（2025）在中国情境下的研究中，曾通过构建包含“节能减排”等气候关键词的静态词典，统计其在企业报告中出现的频率来量化宏观气候风险[4]。

然而，现有研究指出这种基于静态关键词统计的方法在处理中文金融文本时存在显著的内生性缺陷。首先是语境缺失导致的误判，Webersinke 等人（2022）明确指出，简单的关键词匹配无法区分企业是在描述面临的外部风险还是在展示内部的应对能力，甚至无法甄别单纯的政策复述[5]。其次是多义词混淆带来的噪音，中文词汇含义丰富，如“环境”一词可能指代自然环境，也常被用于描述商业或政策环境，导致测度结果失真。更为关键的是静态词典的滞后性，随着 2024 年可持续发展报告指引的发布，诸如“新质生产力”、“碳足迹管理”等新兴术语迅速涌现，固定词典难以敏锐捕捉这些新出现的风险点，也无法有效识别企业通过堆砌正面词汇来掩盖实质性行动缺失的“漂绿”行为[2]。

2.2 大语言模型在气候金融领域的应用进展

鉴于传统词典法的局限性，学术界开始引入基于 Transformer 架构的大语言模型以提升测度精度。LLM 凭借其自注意力机制，能够捕捉长距离的语义依赖关系，从而准确理解句子的情感色彩与逻辑结构。在宏观政策层面，Jiang 等人利用生成式模型自动识别 IMF 政策文件中的碳税与排放配额内容，构建了跨国气候政策严格度指数，证明了该方法在处理非结构化文本时的规模化优势与跨国对齐能力[6]。在微观企业层面，Lopez-Lira 与 Tang 以及 Bolton 等学者的研究表明，金融专用语言模型如 FinBERT 能够从年报中精准抽取减排策略与技术替代风险，进而计算出更为精确的企业级转型压力指数[7-8]。此外，Webersinke 等人提出的 ClimateBERT 模型通过在气候语料上的领域适应，进一步提升了对气候特定文本的分类性能[5]。针对中文语境的特殊性，姜富伟等人指出经过金融语料微调的模型能更好地区分实质性风险披露与口号式宣传，不仅能识别气候信息的数量，更能解析其质量[9]。在物理风险识别方面，多模态模型的引入为评估极端气候造成的资产损失提供了新工具，例如 CLIP 模型在识别洪水、火灾等灾害图像上的表现已得到验证[10]。

2.3 气候风险向系统性风险的传导机制

精准测度气候风险的最终目的是评估其对金融系统稳定性的影响。国际清算银行提出的“绿天鹅”概念强调了气候风险引发系统性金融危机的可能性，其破坏力可能超过传统周期性危机。现有文献主要确认了三条核心传导路径。首先是实体经济渠道，即信用风险传导，气候灾害直接破坏企业实物资产，或转型政策导致合规成本激增，削弱企业盈利与偿债能力，进而推高银行体系的违约概率。其次是资产重估渠道，即市场风险传导，市场对“双碳”目标的预期调整会导致高碳资产价格暴跌，这种波动通过金融机构间的共同持仓网络传染，极易引发流动性螺旋。最后是投资者情绪渠道，Ardia 等人发现气候新闻报道会引发投资者恐慌与非理性抛售，在中国 A 股这样散户占比较高的市场中，这种羊群效应会迅速放大风险，导致系统性风险水平短时间内飙升[11]。在量化指标上，Adrian 和 Brunnermeier 提出的 CoVaR 以及 Brownlees 和 Engle 提出的 SRISK 指标已成为衡量机构对系统尾部风险贡献的主流工具[12-13]，结合大模型提取的高频文本数据，研究者能够更动态地解释这些风险指标的时变特征及其深层来源。

2.4 文献评述与研究空间

综上所述，尽管大模型在气候金融领域的应用已取得显著进展，但现有研究仍存在明显的不足，这为本研究提供了切入点。现有文献多基于历史数据训练模型，缺乏针对 2024 年中国 ESG 披露新规下特有的标准化与形式主义并存文本的实证考察。同时，当前研究多侧重于高性能但昂贵的通用模型或需要大量算力微调的开源模型，缺乏关于低成本、无需微调的国产大模型在实际应用中的成本效益分析。此外，虽然“漂绿”概念已被广泛讨论，但鲜有研究利用大模型的语义推理能力，通过检测风险披露与财务战略行动之间的语义一致性来构建量化的“洗绿指数”。本研究将基于 2024 年 A 股全量年报数据，通过对比实验的方式，对上述尚未深入研究的领域进行探索与完善。

3 实验设计

3.1 数据来源与样本选择

本研究选取 2024 年中国 A 股上市公司作为初始研究样本，选择这一年份主要基于 2024 年作为沪深北交易所《上市公司可持续发展报告指引》实施元年的特殊背景，此时期的年报文本披露兼具高信息密度与高噪音特征，是检验自然语言处理技术有效性的理想场景。文本数据直接提取自巨潮资讯网及 Wind 金融终端，涵盖了 5405 家 A 股上市公司的 2024 年年度报告。在样本筛选过程中，为确保数据的有效性与可比性，本研究剔除了当年被实施 ST、*ST 或 PT 特别处理的公司，以及文件损坏无法解析的观测值。最终，研究构建了一个涵盖制造业、能源、交通运输等行业的全量观测面板。

3.2 文本预处理与实验语料构建

为了精准评估不同模型在处理气候风险信息时的性能差异，本研究采用了“粗筛-精选”的两阶段策略构建实验语料库。不同于以往研究直接对全量文本进行随机抽样，本研究首先利用增强词典法进行定向召回，以确保入选语料均具备基础的气候语义相关性。

3.2.1 阶段一：基于增强词典的定向召回

首先，利用前述构建的包含物理风险、转型风险及气候机遇的“增强词典”（Seed Words + Word2Vec 扩充），对 2024 年全量 A 股年报的 MD&A 章节进行扫描。匹配规则为保留所有至少包含一个扩充关键词的句子。其次去噪处理，对召回的句子进行物理去重，即去除重复语句，以及长度过滤，只保留 20-300 字的句子，并剔除全市场出现频次超过 50 次的模板化口号。

3.2.2 阶段二：分层抽样与统一测试集

经过第一阶段召回后，获得包含数十万条句子的“气候相关候选池”。为了控制计算成本并保证评估的代表性，本研究采用分层抽样法，从“物理风险”、“转型风险”及“机遇”三类候选池中各抽取数据，最终构建了一个包含 30,000 条句子的标准化测试集。

后续的三组实验均完全基于这一固定的候选测试集进行。这一设计控制了输入数据的一致性，旨在考察当文本已经具备“表面相关性”时，不同技术范式区分“实质风险”与“噪音”的能力。

3.3 测度模型构建：三组对比实验设置

为了系统评估不同技术范式在气候风险测度中的表现差异，本研究设计了包含增强词典法、FinBERT2 模型与 DeepSeek 大模型的三组对比实验。这三组实验分别代表了“关键词匹配”、“判别式语义表征”与“生成式逻辑推理”三种技术发展阶段。

3.3.1 第一组：增强词典法

第一组采用增强词典法作为基准对照。参照现有文献的主流做法，首先基于 TCFD 框架及中国“1+N”政策体系，并结合 2024 年最新的监管指引，手工筛选出一组覆盖面极广的核心种子词。具体而言，本研究构建了包含三大维度的关键词库（见表 1）：

1. **物理风险**：不仅包含台风、洪涝等直接灾害，还纳入了“供应链中断”、“资产减值”等灾害引致的财务后果词汇；

2. **转型风险**：紧扣 2024 年政策热点，覆盖了“CBAM（碳边境调节机制）”、“碳足迹”、“Scope 3（范围三排放）”及“ESG 指引”等前沿术语；

3. **气候机遇**：考虑到企业在应对气候变化中可能获得的竞争优势，本研究亦纳入了“新质生产力”、“绿色信贷”等反映绿色转型收益的词汇。

表 1: 气候风险与机遇种子词列表

维度	**关键词集合 (Seed Words)**
物理风险	台风、洪涝、干旱、极端高温、暴雨、山火、低温冰冻、内涝、地质灾害、海平面上升、气温升高、降水模式改变、限电、停工停产、供应链中断、资产减值、厂房受损、经营地受灾
转型风险	碳中和、碳达峰、碳配额、碳税、排放权交易、环保督察、碳排放双控、能耗双控、ESG 评级、可持续发展报告指引、合规成本、新能源替代、技术迭代、低碳技术研发、设备更新、落后产能、高耗能、碳边境、碳关税、CBAM、绿色贸易壁垒、高碳资产、搁浅资产、两高项目、CCER、碳足迹、碳盘查、范围三、Scope 3
气候机遇	绿色收入、技术突破、绿色信贷、绿色债券、碳汇交易、新质生产力、绿色低碳转型、能源结构优化

为克服传统静态词典覆盖面不足的缺陷，在确定上述种子词后，本研究引入了腾讯 AI Lab 开源的大规模中文词向量库（Tencent AI Lab Embedding Corpus）。该库提供了覆盖约 800 万个中文词汇的预训练词向量（200 维）。基于这一高质量的语义空间，本研究通过计算候选词与上述种子词的余弦相似度，筛选出语义高度相关的近义词以扩充词库，最终利用 Jieba 分词工具统计关键词频次并取对数得到风险指标。

3.3.2 第二组：FinBERT2 预训练模型

第二组采用在金融语料上大规模预训练的 BERT 模型——FinBERT2。本次选用的 FinBERT2 模型，其金融语料预训练规模超过 320 亿 Token。据 Xu 等人（2025）指出，在开源的中文金融领域 BERT 类模型中，这将是预训练语料规模最大、性能表现最好的模型[14]。该模型代表了自然语言处理中“预训练 + 微调”范式在特征提取层面的最新进展。由于缺乏大规模人工标注数据进行监督微调，本研究利用 FinBERT2 卓越的上下文表征能力进行零样本分类。具体而言，将气候风险定义的描述语句与待测句子分别输入模型，提取其 [CLS] 位置的句向量。通过计算待测句向量与风险定义向量之间的余弦相似度，衡量句子属于气候风险的概率。相比词典法，FinBERT2 能够捕捉深层语义信息，有效解决多义词问题，但作为判别式模型，其逻辑推理能力仍弱于生成式大模型。

3.3.3 第三组：DeepSeek 大语言模型

第三组采用国产大语言模型 DeepSeek-V3 进行生成式零样本推理，代表了当前最先进的“提示工程（Prompt Engineering）”范式。该方法摒弃了概率计算，转而模拟人

类专家的思维链（Chain-of-Thought）。提示工程指令要求模型扮演资深气候金融专家，在通读句子的基础上，依据逻辑判断文本是否包含实质性风险，并明确区分“风险暴露(-1)”、“风险防范（1）”与“不相关（0）”。这种设计利用了 Transformer 架构的自注意力机制与海量参数带来的世界知识，旨在解决前两组方法无法处理的否定句式、条件假设及“漂绿”识别难题。

3.4 变量定义与指标构建

为了对比三组方法的结果，本研究统一构建了标准化的气候风险指标。对于词典法，计算扩充关键词在文本中的出现频率。对于 FinBERT2，计算所有句子与风险定义的平均语义相似度得分。对于 DeepSeek，计算模型判定为“风险暴露”句子的加权比例。

4 结果分析

本研究首先对三组的测度指标——增强词典法指标（$Risk_{Dict}$）、FinBERT2 语义指标（$Risk_{BERT}$）以及 DeepSeek 大模型指标（$Risk_{LLM}$）进行了描述性统计分析。统计结果显示，三组指标在分布形态上呈现出显著的结构性差异，反映了不同技术范式对信息的不同敏感度。

4.1 增强词典法：机械匹配下的稀疏性与长尾分布

基于增强词典法构建的风险指标 $Risk_{Dict}$ 呈现出典型的尖峰厚尾分布特征，暴露了关键词匹配技术的内生性局限，具体表现如下：

首先，分布呈显著右偏且高度稀疏。如图 1 (A) 直方图与 (D) 经验累积分布函数图所示，绝大多数样本得分聚集于 0 值或极低区间，即 0 至 5 分，呈现出明显的零值膨胀现象。这表明，尽管已扩充专业词库，但机械式匹配依然无法有效识别未预先录入词表的、表述隐晦的风险信息，例如“原料运输因极端天气受阻”这类表述，从而导致大量的漏判。

其次，极端离群值引发长尾效应。图 1 (B) 箱线图显示存在大量高分散度的离群点，少数样本得分高达 60 至 95 分，远超过约为 2 分的中位数。这种超级长尾特征表明，该指标在本质上度量的是风险关键词出现的频次或声量，而非语义层面的风险实质。因此，指标极易受到年报文本长度及程式化表述重复的影响，使得少数文本对整体统计结果产生过度的杠杆效应。

4.2 FinBERT2 模型：二元极化与弱信号强制分类

FinBERT2 模型输出的统计结果揭示了判别式模型在零样本环境下的内生性缺陷。

第一，缺乏中性缓冲地带的二元极化。如表 2 所示，FinBERT2 将所有测试样本强制划分为“风险暴露”（64.61%）与“风险防范”（35.39%）。由于缺乏“中性”类别，大

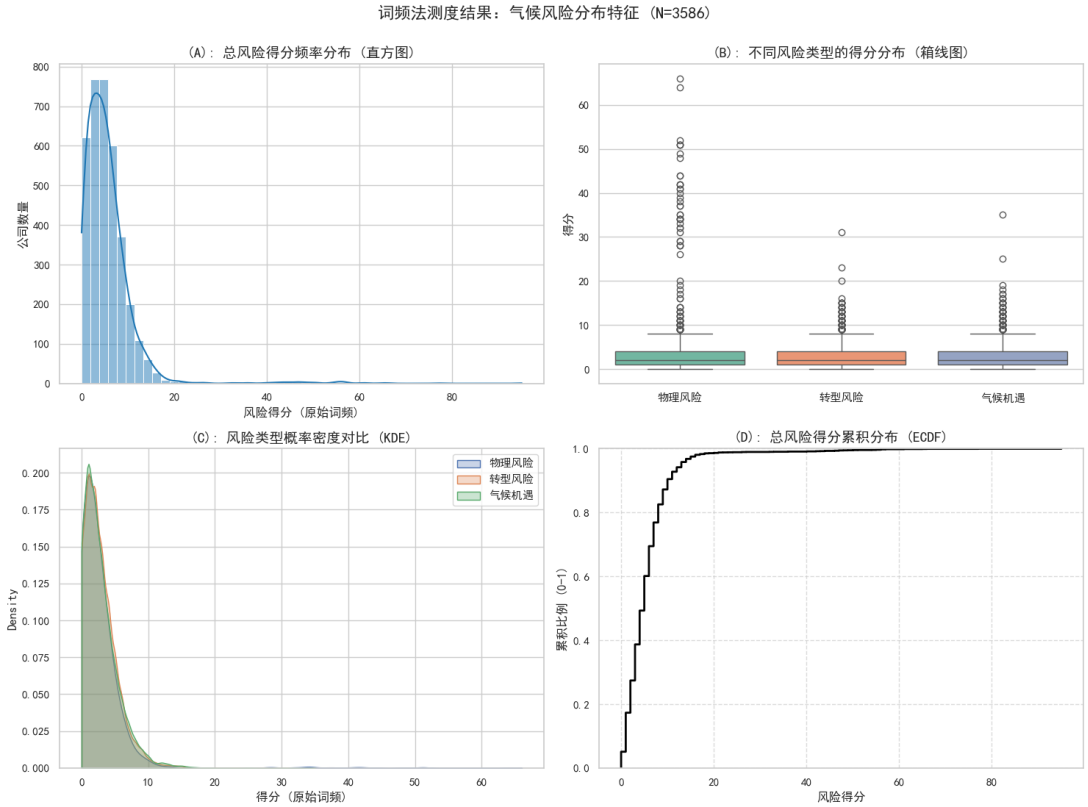

图 1: 增强词典法测度结果

表 2: FinBERT2 模型测度结果统计

统计指标	**数值**	**说明**
Panel A: 分类占比 (AI_Label_Text)		
风险暴露 (Risk)	64.61%	强制归类为主
风险防范 (Prevention)	35.39%	被动防御为辅
中性/无关 (Neutral)	0.00%	**缺乏噪音过滤机制**
Panel B: 置信度分布 (AI_Confidence)		
均值 (Mean)	0.5321	**极接近随机阈值 0.5**
中位数 (Median)	0.5269	判定信心不足
标准差 (Std)	0.0246	分布高度集中
最大值 (Max)	0.6849	缺乏高确信度样本

量原本属于背景描述或合规口号的噪音文本被强制赋予了情感极性，导致风险敞口被系统性高估。

第二，低置信度下的“犹豫”分类。置信度统计揭示了模型判定的内在脆弱性。‘AI‘的均值仅为 0.532，标准差为 0.025。这种“低均值、低方差”的分布特征表明，FinBERT2 虽然捕捉到了气候关键词的相关性，但在判断其情感倾向时并未找到强有力的语义证据。模型实际上是在大量模棱两可的文本中进行“微弱优势的猜测”。

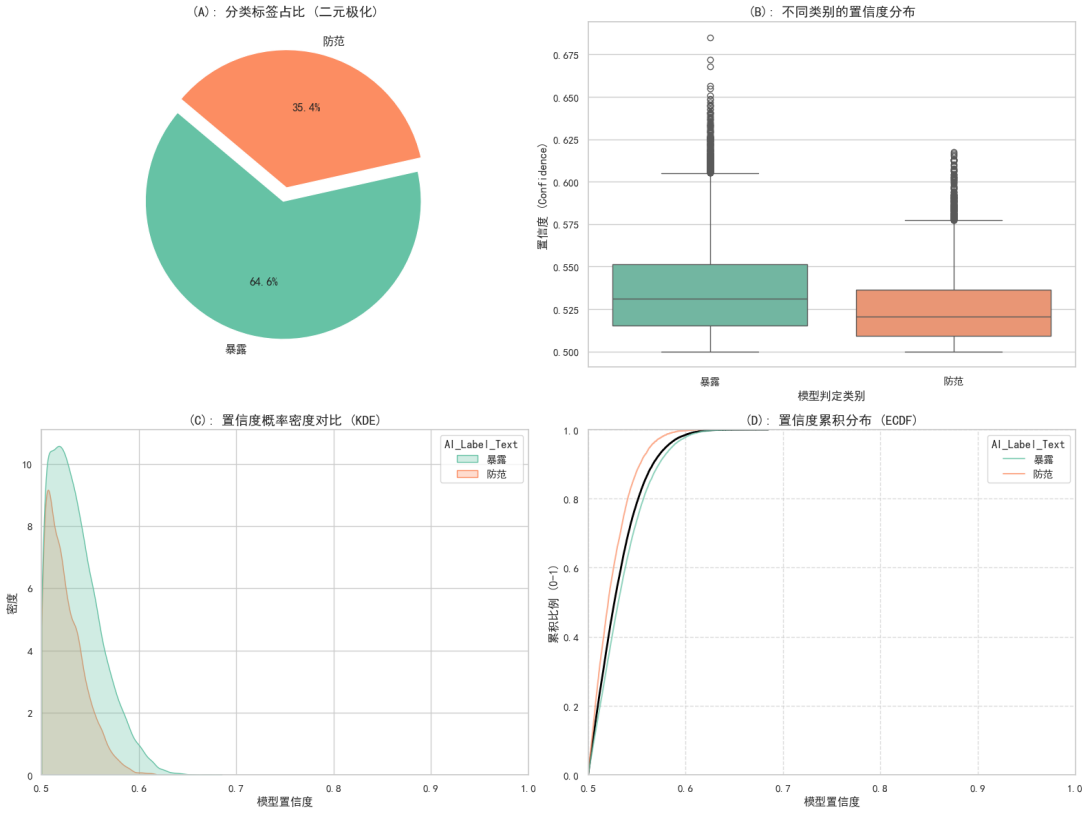

图 2: FinBERT2 模型测度结果分布

与 DeepSeek 模型能够以高置信度剔除噪音不同，FinBERT2 的这种“基于弱信号的强制分类”引入了大量随机噪声，进一步证明了在处理复杂的 ESG 年报文本时，简单的判别式模型难以替代具备逻辑推理能力的生成式大模型。

4.3 DeepSeek 大模型：测度效能的综合评估

DeepSeek 大模型生成的指标 $Risk_{LLM}$ 在多个维度上展现出更优的测度效能，具体分析如下：

第一，精准的逻辑切分与全量降噪能力。与 FinBERT2 的强制站队模式形成鲜明对比，DeepSeek 模型展现出强大的信息提纯能力。实验结果显示，模型将原始语料中高达 89.0% 的样本判定为无关或中性陈述，仅保留了 11.0% 被认定为具有实质性风险或机遇的样本。其判定的概率密度分布呈现清晰的三峰结构，如图 3 (B) 所示，特别是代表中性类别的概率峰显著隆起，表明模型是基于深度语义理解确信地剔除了大量合规声明与

空洞表述，而非随机猜测。这一机制有效解决了前述方法普遍存在的虚假阳性问题，实现了对核心风险因子的精准识别与降噪。

第二，披露策略的物理-转型二元非对称性。聚焦于经模型筛选出的实质性披露样本，即图 3 (C)，A 股上市公司的气候信息披露策略呈现出显著的结构性差异。在物理风险维度，风险暴露类陈述占据主导，这反映了台风、洪涝等实体灾害作为客观硬约束的不可抗特性；与之相对，在转型风险与机遇维度，风险防范或机遇把握类陈述的占比显著提升，例如转型风险中防范类占比约 76.0%。这揭示了企业在面对政策与市场转型压力时，更倾向于采用强调自身技术改造与管理行动的防御性披露策略，以对冲潜在合规成本带来的负面市场预期。

第三，风险信息硬度的异质性度量。模型输出的置信度分数进一步量化了不同类别风险信息的语义确定性，如图 3 (D) 所示。统计分析发现，DeepSeek 模型对物理风险暴露判定的置信度极高，其中位数接近 0.95，且分布集中，表明此类文本多由受灾损失、资产减值等客观事实构成，逻辑清晰。相比之下，对转型风险判定的置信度分布则更为分散，其数值下探至 0.70 区间，精准地捕捉了此类涉及政策预期与未来行动的信息所固有的模糊性。这证明生成式大模型不仅能够进行分类，更具备了类似专家般区分既定事实与潜在预期的深度语义理解与量化评估能力。

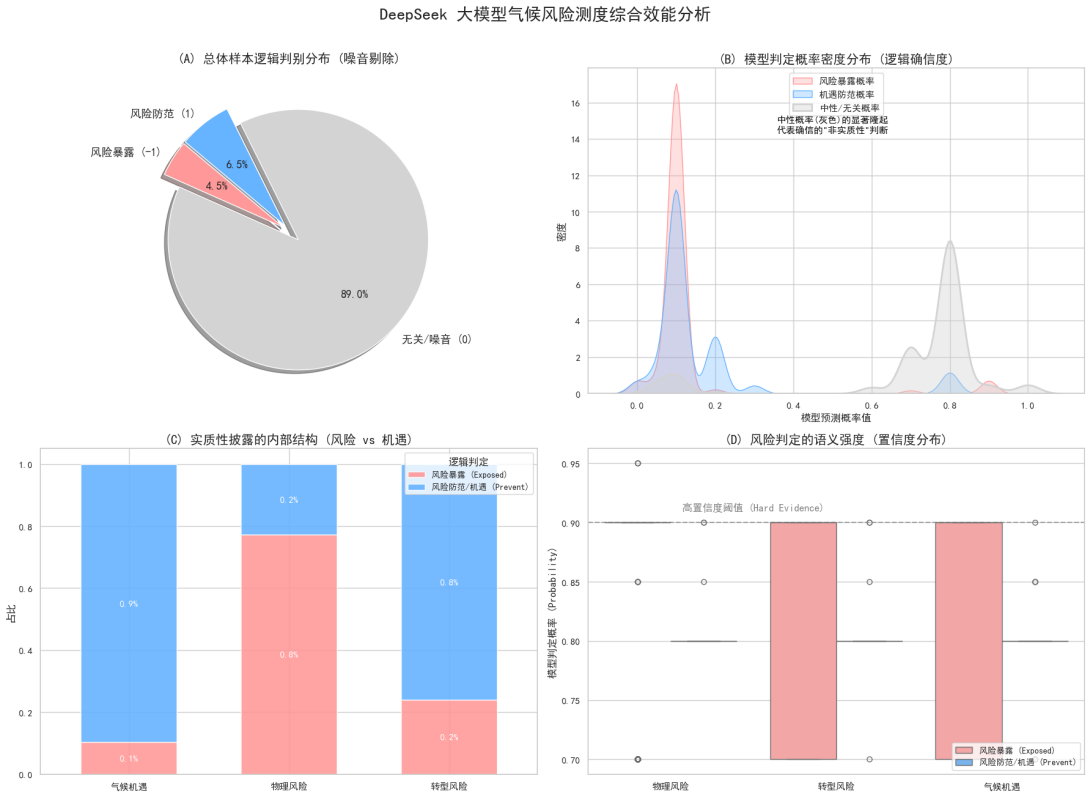

图 3: 大模型气候风险测度综合效能分析

5 研究结论与不足

5.1 研究结论

本研究聚焦于气候金融文本测度这一核心命题，利用 2024 年 A 股上市公司年报语料，系统对比了增强词典法、FinBERT2 预训练模型与 DeepSeek 大语言模型在气候风险识别中的效能差异。通过构建包含三万条样本的标准化平衡测试集，本研究得出以下主要结论：

第一，大语言模型有效解决了传统测度方法的语义鸿沟与噪音干扰难题。研究发现，传统方法均存在显著的内生性缺陷：增强词典法受制于机械匹配，导致指标呈现极端的长尾分布与稀疏性，难以捕捉隐性风险；FinBERT2 模型虽具备语义理解能力，但陷入“二元强制站队”的陷阱，无法识别中性样本（Label 0 缺失）。更关键的是，FinBERT2 的判定置信度极低（均值仅为 0.53），表明模型在无法输出“无关”选项时，实质上是在进行“微弱优势的随机猜测”，引入了大量系统性噪音。相比之下，DeepSeek 大模型展现了卓越的判别效度，通过逻辑推理成功剔除了 89.0% 的漂绿噪音与合规口号，实现了对气候风险因子的精准提纯。

第二，A 股上市公司气候信息披露存在虚假繁荣现象，实质性风险敞口被显著高估。实验数据表明，在传统方法判定为高度相关的文本中，仅包含 11.0% 的实质性风险暴露或具体防范措施。绝大多数企业年报充斥着标准化的合规声明与模糊的战略愿景。这一发现修正了过往基于词频统计的研究结论，即市场层面的气候关注度虽然在字面量上大幅提升，但实质性的净风险信息含量依然稀缺。

第三，不同类型气候风险的披露策略呈现显著的结构性异质性。DeepSeek 的测度结果揭示了企业在面对物理风险与转型风险时截然不同的披露行为逻辑。**首先，物理风险呈现高暴露、强证据特征**。企业对台风与洪涝等实体灾害的披露多为直接的损失陈述，其风险暴露类陈述占比近 80%，且模型判定置信度极高，中位数大于 0.95，表明此类信息具有硬事实属性。其次，转型风险呈现高防御、软预期特征。企业在面对双碳政策压力时，倾向于采取防御性披露策略，风险防范类陈述占比超过 75%，侧重描述技术改造等应对措施以对冲合规成本预期。同时，模型对转型风险的判定置信度分布较为发散且数值较低，反映了政策预期类信息固有的模糊性。

5.2 政策建议与管理启示

基于上述结论，本研究对监管部门、投资者及企业提出以下建议：

1. **监管部门应从量的扩充转向质的提升**。鉴于当前市场存在高达 89% 的无效披露噪音，监管机构在制定 ESG 或气候披露指引时，应降低对单纯篇幅或关键词数量的要求，转而强制要求企业披露实质性数据。建议建立类似实质性风险清单的披露模板，要求企业明确区分已发生的物理损失与预期的转型成本，减少模棱两可的口号式表述。

2. **投资者应警惕漂绿陷阱，并关注净风险因子**。投资者在使用文本数据进行资产定价或风险管理时，应摒弃传统的词频因子，转而采用具备逻辑剔除能力的大语言模型增强因子。特别需要关注模型识别出的那些置信度高、且被归类为风险暴露的硬核信息，这部分信息往往包含了未能被股价充分反映的尾部风险。

3. **企业管理层应优化披露策略，增强信息硬度**。企业应意识到，随着人工智能技术的发展，单纯堆砌关键词的印象管理策略已难以奏效。高质量的披露应包含具体的财务影响数据与可验证的减排行动。特别是在转型风险方面，增加定量数据的披露有助于消除市场对政策不确定性的担忧，从而降低权益资本成本。

5.3 研究不足与展望

尽管本研究在测度方法上取得了一定突破，但受限于数据与算力资源，仍存在以下局限性，有待未来研究进一步完善：

第一，存在样本的时间维度局限。 受限于大语言模型推理的高昂成本，本研究仅对2024 年的横截面数据进行了精细化测度，未构建长时间序列的面板数据。这使得本研究无法动态考察 DeepSeek 因子在时间维度上的演变规律及其对股价的长期预测能力。未来研究可尝试通过大模型蒸馏技术，训练轻量级的小模型来处理历史海量数据。

第二，存在模型的可解释性与黑箱问题。 虽然 DeepSeek 输出了分类理由与置信度，但作为深度神经网络，其内部的推理权重分配仍具有不可解释性。特别是对于处于置信度边界，例如 0.6 至 0.7 之间的模糊样本，模型的判断逻辑是否存在系统性偏差，仍需更细致的稳健性检验。

第三，存在提示工程的主观依赖问题。 大语言模型的输出质量在很大程度上取决于提示词的设计。尽管本研究经过多轮迭代优化了提示，但不同的提示词策略，如思维链或少样本学习，可能会导致测度结果的波动。未来研究可建立一套标准化的金融文本测度提示词库，以提高不同研究之间结果的可比性。

第四，缺乏多模态信息的融合。 本研究仅关注文本信息，忽略了年报中包含的图表与数据表格等多模态信息。实际上，很多关键的气候绩效，如碳排放量与能源消耗数据，往往以表格形式呈现。未来可引入多模态大模型，实现对文本与数值数据的联合解析，以构建更为全景式的气候风险画像。

参考文献

[1] Bank for International Settlements. Climate-related financial risks: Measurement and modelling[R]. Bank for International Settlements, 2023.

[2] BATTISTI E, BO S, GIAKOUMELOU A. Climate risk disclosure and greenwashing: Evidence from chinese a-share listed companies[J]. International Review of Economics & Finance, 2025, 103: 104568.

[3] LI Q, SHAN H, TANG Y, et al. Corporate climate risk: Measurements and responses [J/OL]. The Review of Financial Studies, 2024, 37(6): 1778-1830. DOI: 10.1093/rfs/hhad094.

[4] WANG H, BAO S, JIANG S. Low carbon policies and the systemic risk of mutual funds—the case of the "double carbon" target in china[J/OL]. Humanities and Social Sciences Communications, 2025, 12(1): 1-8. DOI: 10.1057/s41599-025-05996-1.

[5] WEBERSINKE N, KRAUS M, BINGLER J A, et al. Climatebert: A pretrained language model for climate-related text[C]//Proceedings of AAAI 2022 Fall Symposium. Arlington, VA: AAAI, 2022.

[6] JIANG W, KOROBILIS D, SEN S. Mining climate policy texts using large language models[R]. IMF Working Paper 23/207, 2023.

[7] LOPEZ-LIRA A, TANG Y. Can chatgpt forecast stock price movements?[J]. Journal of Financial Economics (forthcoming), 2023.

[8] BOLTON P, KACPERCZYK M. Do investors care about carbon risk?[J]. Journal of Financial Economics, 2021, 142(2): 517-549.

[9] 姜富伟, 刘雨𠕇, 孟令超. 大语言模型、文本情绪与金融市场[J]. 管理世界, 2024, 40(8): 42-64.

[10] RADFORD A, KIM J W, HALLACY C, et al. Learning transferable visual models from natural language supervision[C]//International Conference on Machine Learning (ICML). PMLR, 2021.

[11] ARDIA D, BLUTEAU K, KATSIAMPA P. Climate news and stock market volatility [J]. Journal of Econometrics, 2023, 233(2): 566-587.

[12] ADRIAN T, BRUNNERMEIER M K. Covar[J]. American Economic Review, 2016, 106(7): 1705-1741.

[13] BROWNLEES C, ENGLE R F. Srisk: A conditional capital shortfall measure of systemic risk[J]. The Review of Financial Studies, 2017, 30(1): 48-79.

[14] XU X, WEN F, CHU B, et al. Finbert2: A specialized bidirectional encoder for bridging the gap in finance-specific deployment of large language models[C/OL]//

Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2 (KDD '25). 2025. https://doi.org/10.1145/3711896.3737219.