

ESG 体系下的 AI 研究 (一): 多维投资增效, 防范伦理风险

核心观点

随着大模型加速迭代, AI 有望快速实现大规模应用。面对 ESG 投资生态中的监管追踪难、企业合规成本高、投资端数据获取分析困难等痛点, AI 能在多个环节帮助不同参与者降本增效, 通过智能、高效、创造性输出等特点, 助力 ESG 投资打通生态链条。目前 AI 正面影响与负面争议作用于多个行业, 并与多方面的 ESG 议题相关, 但对 AI 治理的直接评估并不直接体现在 ESG 框架中, 我们整理了全球实践情况, 发现将负责任 AI 原则与 ESG 框架相融合, 可以建立 AI 伦理风险分析体系, 为管控相关投资风险提供思路。

□ AI 有望大规模应用, 帮助完善 ESG 投资基础设施。

近期, AI 大模型迭代速度加快, 模型热度在周度统计中都发生了较大变化, 同时算法优化成本降低, 有望进一步加速大规模应用。AI 有助于解决 ESG 生态中的多方痛点。对监管与国际机构而言, 过去 ESG 政策追踪困难、追踪成本高, AI 通过广泛整合信息的动态化监测, 以及智能对比逻辑关系的交叉验证体系, 有望帮助监管端降低追踪成本, 助力 ESG 政策与倡议的顺利施行; 对企业而言, 面对多样的 ESG 政策, 企业学习理解与合规保本的成本都较高, AI 提供了知识图谱等工具, 助力企业政策分析; 同时生成式 AI 能够帮助企业进行合规报备和 ESG 信息披露。合规成本的降低将使企业更有动力参与 ESG 实践。

□ AI 助力投资端优化策略与营销。

在投资层面, 传统 ESG 数据存在来源多、收集处理困难、更新频率低、滞后性强等问题。AI 首先可以智能化获取包含文本、图片在内的多模态海量数据, 同时能够智能化将数据整理成为“终端友好”的形式, 为投资者分析 ESG 变化提供支撑。同时, AI 可以实现争议事件的实时抓取和分析, 帮助投资者及时调整估值。在投资策略方面, 机器学习算法能够在因子策略构造与因子筛选方面提供帮助, 同时语言分析、识别与预测模型也能够成为构建新型量化策略的核心, 帮助投资人优化风险-收益情况。在营销端, 生成式 AI 通过生成营销方案、营销内容, 大大降低了营销成本。

□ 负责任 AI 成为共识, 整合 ESG 框架的 RAI 体系助力伦理风险管理。

当前 AI 立法尚不完善, 而 AI 的争议性影响在环境、社会、治理层面广泛存在, 例如: 在环境维度, AI 既能够通过智能化系统助力降低能效, 但 AI 所需的算力耗能巨大; 社会维度, AI 既提高了生产效率, 也造成了就业冲击、隐私泄露等社会问题甚至伦理争议; 公司治理维度, AI 赋能了公司治理也带来了监管挑战。面对缺少法律框架参考但仍会受到监管督导、社会舆论压力等因素冲击的 AI 行业, 整合 ESG 框架与负责任 AI 原则, 能够帮助投资者识别有 AI 伦理风险的企业。在投资实践中, 主要有公司治理维度关注 AI 与 ESG 框架整体整合两种方式。投资者也可以通过与企业的进一步沟通获得信息, 为管理有关企业的投资风险提供参考。

□ 风险提示

经济修复不及预期; AI 模型不稳定风险; 市场情绪与偏好波动风险。

分析师: 祁星
执业证书号: S1230524120002
qixing@stocke.com.cn

相关报告

- 1 《尽责管理撬动资本市场可持续发展, 气候适应与转型金融受多方关注》 2025.06.01
- 2 《公募基金参与上市公司治理新规发布, “以披促优” 加速资本市场可持续发展》 2025.05.12
- 3 《国家主席习近平致辞为应对气候变化注入强心针, ESG 基金与债券市场热度分化呈现短期震荡》 2025.05.10

正文目录

| | |
|--|-----------|
| 1 AI 将有助于完善 ESG 基础设施、疏通 ESG 投资链条 | 5 |
| 1.1 ESG 投资中的信息流通过程存在难点 | 5 |
| 1.2 AI 拐点已至，性能提升+成本降低有望推动在 ESG 事务中的使用 | 6 |
| 1.3 监管机构端：AI 助力监管有效监督政策落实情况 | 9 |
| 1.3.1 ESG 监管痛点：难验证、难追溯，企业或主动漂绿 | 9 |
| 1.3.2 AI 方案：动态监管与多模态交叉验证 | 10 |
| 1.4 企业端：AI 助力企业多方位降低合规成本 | 11 |
| 1.4.1 合规成本高成为企业参与 ESG 实践的重要阻碍 | 11 |
| 1.4.2 AI 政策智能解析：从被动应对到主动适应 | 12 |
| 1.4.3 AI 帮助报告自动化：从数据孤岛到智能洞察 | 13 |
| 2 AI 技术赋能，为资管机构提供了 ESG 投资的丰富工具 | 14 |
| 2.1 投研与风控：AI 带来更加丰富 ESG 投资手段 | 14 |
| 2.1.1 数据收集与处理：AI 助力多模态数据的敏捷挖掘和快速处理 | 14 |
| 2.1.2 AI 助力优化投资策略 | 17 |
| 2.2 AI 助力资管企业与投资人进行沟通 | 20 |
| 3 AI 产生广泛外溢影响，ESG+RAI 框架助力分析风险与机遇 | 21 |
| 3.1 ESG 议题角度下，AI 的效率提升与伦理争议 | 21 |
| 3.1.1 环境：提升环保效率与自身消耗资源的两面性 | 21 |
| 3.1.2 社会：促进社会发展与产生伦理争议的两面性 | 25 |
| 3.1.3 公司治理：机遇与挑战并存 | 28 |
| 3.2 AI 面对强监管风险，AI 伦理与负责任 AI 实践成为趋势 | 29 |
| 3.2.1 强隐性监管背景下，企业人工智能风险凸显 | 29 |
| 3.2.2 AI 伦理逐步成为道德共识，负责任 AI 成为行动规范 | 30 |
| 3.3 ESG+RAI 框架整合助力防范投资中的 AI 风险暴露 | 31 |
| 3.3.1 挪威银行资管 NBIM：关注公司治理中对 AI 的治理措施和机制建设 | 32 |
| 3.3.2 UBS 资管：多层级 AI 公司治理分析模型 | 33 |
| 3.3.3 CRISO 的 ESG-AI 全面整合框架 | 34 |
| 4 风险提示 | 36 |

图表目录

| | |
|---|----|
| 图 1: ESG 投资生态模型 | 5 |
| 图 2: 国内上市公司可持续报告编制负责人 | 6 |
| 图 3: 国内上市公司可持续报告编制咨询成本 | 6 |
| 图 4: 人工智能系统模型 | 7 |
| 图 5: 具备影响力的人工智能模型发布时间、行业及数据量 | 8 |
| 图 6: 人工智能在多个能力测试上已经超过人类基准水平 | 8 |
| 图 7: 2025 年后具备推理能力的大语言模型收到用户信息占比 | 8 |
| 图 8: Deepseek-R1 定价较 OpenAI o1 显著降低 | 8 |
| 图 9: 2017-2024 全球各类机构应用 AI 与生成式 AI 的比例变化 | 9 |
| 图 10: 2023-2024 分地区机构应用 AI 的比例变化 | 9 |
| 图 11: 2021、2022 年全国电力消费平均二氧化碳排放因子 | 10 |
| 图 12: 2021、2022 年各区域电力消费平均二氧化碳排放因子 | 10 |
| 图 13: ESG 合规是参与 ESG 活动的主要驱动 | 11 |
| 图 14: 国内上市公司可持续报告披露成本占比最高的部分 | 11 |
| 图 15: 知识图谱概念界定 | 12 |
| 图 16: 政策知识图谱分析的本体模型 | 12 |
| 图 17: 使用神经网络模型进行政策情感分析流程 | 13 |
| 图 18: 针对国内文献对“双碳”政策研究的关键词聚类 | 13 |
| 图 19: 多元异构数据的融合处理与 AI 分析流程图 | 14 |
| 图 20: 大数据“4V”特性 | 15 |
| 图 21: 风云四实时卫星图像 | 16 |
| 图 22: 台风监测-大气海平面气压实时情况 | 16 |
| 图 23: Wind 跟踪的中国平安近半年的争议事件信息 | 16 |
| 图 24: 传统数据处理方式与结合大模型的数据处理方式对比 | 17 |
| 图 25: 两层决策树结构助力 ESG 因子筛选 | 18 |
| 图 26: ESG 观点值与国家 ESG 相关政策变化相关性 | 19 |
| 图 27: GEO 五分位对应企业远期碳强度同比变化 | 20 |
| 图 28: GEO 五分位对应企业致力于科学碳目标和绿色专利比例 | 20 |
| 图 29: 大数据与 AI 分析在能源网络中的应用 | 21 |
| 图 30: 机器学习技术如何推动建筑行业减碳 | 23 |
| 图 31: 人工智能模型在训练过程中产生的碳排放量估计 | 24 |
| 图 32: 2024 年发表的领域内论文中由 AI 驱动研究的比例 | 25 |
| 图 33: 提及 AI 应用的临床实验数量 (单位: 个) | 25 |
| 图 34: AI 算法中可能存在偏见 | 26 |
| 图 35: 2012-2024 年有记录的 AI 伦理争议事件数量 (单位: 件) | 26 |
| 图 36: ChatGPT 发布后写作类岗位发布数量锐减 | 27 |
| 图 37: 图像生成 AI 发布后制图相关岗位需求减少 | 27 |
| 图 38: 2012-2023 全球工业机器人存量年度变化 (单位: 千台) | 27 |
| 图 39: 高收入人群岗位也将受到 GenAI 的冲击 | 27 |
| 图 40: 欧盟《人工智能法案》分类方法及监管措施 | 29 |
| 图 41: 2024-2025 企业对负责任 AI 相关风险的关注度 | 31 |
| 图 42: 企业在负责任 AI 实践中受到相关法律影响的调查 | 31 |

| | |
|---|----|
| 图 43: ESG 议题与负责任 AI 议题在范畴上相互交织 | 32 |
| 图 44: NBIM 公司负责任 AI 问责三大要素 | 33 |
| 图 45: ESG-AI 分析步骤 | 34 |
| 表 1: 人工智能发展历史 | 7 |
| 表 2: 欧盟碳边境调节机制 (CBAM) 政策内容概述 | 10 |
| 表 3: 欧盟可持续披露规则适用范围和适用内容 | 12 |
| 表 4: 基于决策树的 ESG 因子策略表现 | 18 |
| 表 5: 易方达、华夏、嘉实三家基金智能客服对 ESG 有关提问的回复 | 20 |
| 表 6: 人工智能模型在电力系统中的应用 | 22 |
| 表 7: AI 助力工业脱碳的应用场景与作用机制 | 22 |
| 表 8: AI 大模型发展的不同阶段及对应碳排放强度变化 | 24 |
| 表 9: AI 大模型碳排放的主要来源 | 25 |
| 表 10: AI 应用到公司治理面临的问题 | 28 |
| 表 11: 我国人工智能有关的主要部门规章 | 29 |
| 表 12: 科技部《新一代人工智能伦理规范》基本规范内容 | 30 |
| 表 13: UBS 资管对公司 AI 治理进行评估的三层框架 | 34 |
| 表 14: ESG-AI 框架的 AI 用例及对应用例的风险 | 34 |
| 表 15: RAI 治理指标分析问题 | 35 |

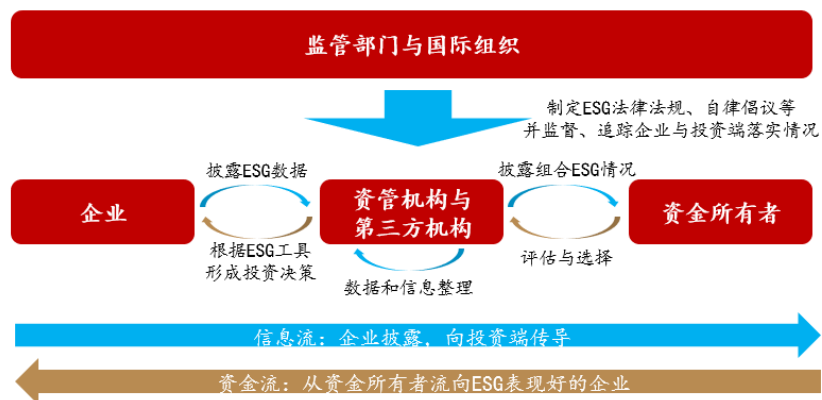
1 AI 将有助于完善 ESG 基础设施、疏通 ESG 投资链条

随着全球可持续发展目标的推进，ESG 投资已成为重塑资产配置逻辑的重要工具。ESG 投资的核心理念在于将环境、社会和治理因素纳入投资决策，帮助投资者识别企业的长期价值潜力和潜在风险，推动资本向可持续发展领域倾斜。随着近期 AI 大模型的迭代加速，我们认为，AI 应用密集落地后，ESG 投资中的痛点有望被解决，ESG 能够更好地发挥作用，助力全球的可持续发展。

1.1 ESG 投资中的信息流通过程存在难点

我们构建了 ESG 投资的简化生态链模型来梳理 ESG 投资的主要参与者和主要活动。ESG 投资的参与者大体可以分为监管部门与国际组织、企业、资管机构与第三方组织、资金所有者四类。我们认为监管部门与国际组织是 ESG 投资的主要倡导者，他们通过立法/倡议的方式设置合规要求，并追踪企业和投资端的落实情况。企业出于法律压力或者利益相关者约束执行 ESG 倡议，并将对应的信息披露给社会公众。资管机构与三方机构分析企业的 ESG 信息，并以此为依据形成 ESG 评级等投资工具，同时资管机构也会进一步拟定 ESG 投资策略，而资金所有者根据投资倾向选择资管机构。

图1：ESG 投资生态模型



资料来源：浙商证券研究所绘制

提升信息的真实可靠性，降低信息产生与分析成本是 ESG 投资的内在需求。从信息流视角，我们认为政策合规信息、企业披露信息等都是 ESG 投资中的基础信息，或者说原始信息。企业是 ESG 信息的主要生产者。而资管机构与第三方机构则是信息的加工应用者，他们将原始信息加工成更高维度、标准化的 ESG 评级等投资工具，应用在投资实践中。资金所有者又根据投资组合层次的 ESG 信息来选择资管机构，资管机构也通过 ESG 表现选择企业，使得资金流向符合 ESG 观念的企业实体。可以看出，信息的真实性和可靠性是 ESG 生态良好运行的基础，而不断降低信息的生成成本、分析成本也是所有 ESG 投资参与者的一致需求。

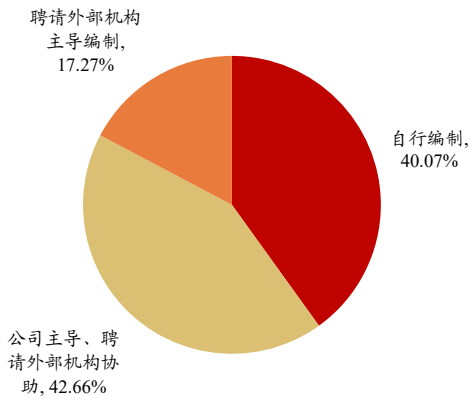
具体而言，不同的类型的参与者可能会面临如下问题：

监管部门与国际组织痛点：如何实现有效的合规跟踪。监管部门与国际组织具备解决环境、社会等层次的公共问题的战略目标，并以 ESG 为主要手段，希望缓解经济活动中产生的外部性问题。因此，监管部门与国际组织迫切需要企业完善执行 ESG 相关规定，并希望以有效手段进行跟踪和监督，从而确保政策的有效性。但政策执行情况的跟踪往往较难。一

是数据来源复杂，尤其是很多合规标准涉及复杂的供应链信息，使得信息难以获取。二是企业、资管机构都希望降低成本，催生了虚假报告的“漂绿”行为。这些都对相关法规、政策、倡议等的实现形成了挑战。

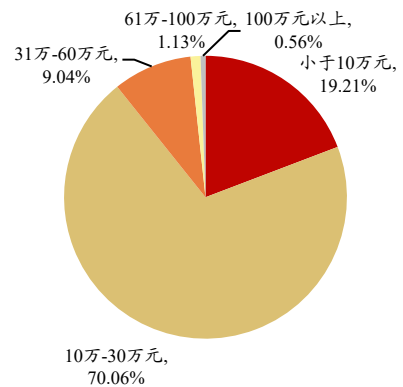
企业痛点：如何在低成本的前提下实现合规要求。在 ESG 相关政策推进的过程中，许多企业原来并无专门从事 ESG 工作的组织架构。因此无论是政策解读、政策执行还是后续的信息披露，都需要企业额外投入成本。以上市公司可持续报告的披露为例，上市公司寻求第三方机构帮助编制或咨询的费用一般在 10-30 万元，部分可能达到 100 万元以上。企业仍不断寻求方案在满足合规要求的前提下降低成本。

图2：国内上市公司可持续报告编制负责人



资料来源：上市公司协会，浙商证券研究所

图3：国内上市公司可持续报告编制咨询成本



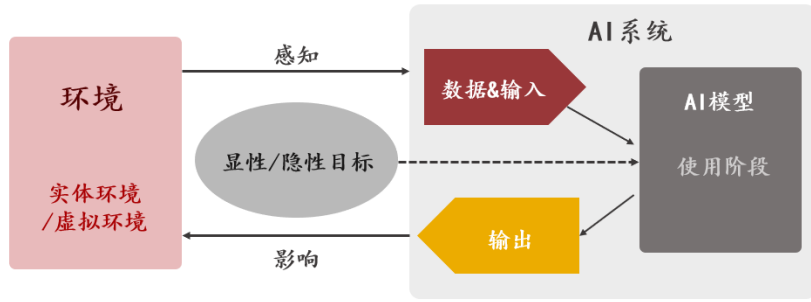
资料来源：上市公司协会，浙商证券研究所

资管机构与第三方机构痛点：如何实现快速、迅捷的信息获取和分析。资管机构与第三方机构，无论是支持投资还是提供第三方信息，都首先需要大量的基础数据做支撑。数据来源一方面依靠企业的披露，另一方面也依靠机构积极挖掘其他相关数据。其次，ESG 分析也依靠海量的非结构化文本数据，能不能低成本、快速敏捷地分析海量数据，是资管机构与第三方机构的深层次需求。

1.2 AI 拐点已至，性能提升+成本降低有望推动在 ESG 事务中的使用

AI（人工智能）是一种从输入信息中推断并输出的系统。根据经合组织的定义，AI 系统是一种基于机器的系统，对于一组显式或隐式目标，它可以从收到的输入中推断如何生成输出，例如预测、内容、推荐或决策，从而影响物理或虚拟环境。IBM 则定义 AI 是使计算机和机器能够模拟人类学习、理解、解决问题、决策、创造力和自主性的技术。**简言之，AI 系统能够智能化处理输入的信息，并根据模型输出相应内容。**

图4：人工智能系统模型



资料来源：OECD，浙商证券研究所

世界进入“生成式 AI (GenAI)”时代，AI 的交互与创造能力进一步加强。AI 技术的发展大致可以分为四大阶段。1956 年达特茅斯会议提出“人工智能”这一概念后，人工智能首先经历了弱人工智能、机器学习、深度学习三个技术迭代期。但在这三个时期人工智能形态均为“决策式 AI”，即主要用于洞察大量数据中的复杂模式和关系。在 2018 年后，以大语言模型为代表的生成式 AI 则能够通过模型训练生成新颖、有意义的内容，包含文本、图像、代码或视频等多模态。生成式 AI 颠覆了原有的编程语言交互模式，用户通过自然语言即可与 AI 系统进行交互，并生成更加多元化、具备创造性的内容。

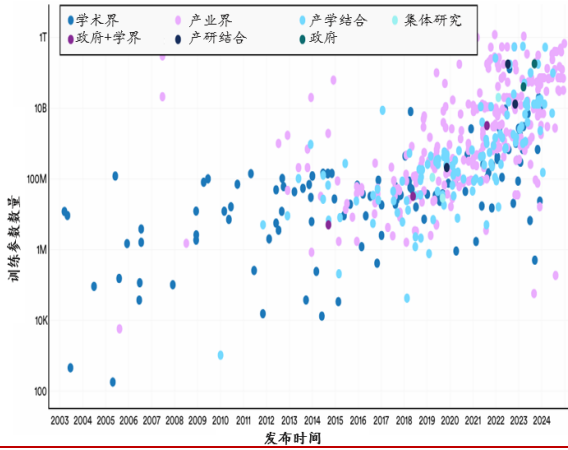
表1：人工智能发展历史

| 阶段 | 时间 | 代表性技术成果 | 数据规模 | 技术栈 |
|--------|-----------------|--|----------------|--------------------------------------|
| 弱人工智能 | 1950 年 - 1990 年 | 基于人工设计的规则系统 | 数百规则集 | 基于专家知识和规则的系统 |
| 统计机器学习 | 1990 年 - 2012 年 | HMM, CTF, SVM, 反向传播, 卷积网络 | 百万级标注数据 | 统计机器学习算法+算法包 (scikit-learn, XGBoost) |
| 深度学习 | 2013 年 - 2018 年 | ImageNet, ResNet, Word2vec, Attention, Transformer | 十亿级标注数据 | 深度神经网络+开发框架 (TensorFlow, PyTorch) |
| 生成式 AI | 2018 年 - 至今 | BERT, PaLM, LLaMA, GPT-4, GLM | 全网万亿数据, 海量用户反馈 | 预训练+微调+开源社区 |

资料来源：IBM，中关村科技会展，浙商证券研究所

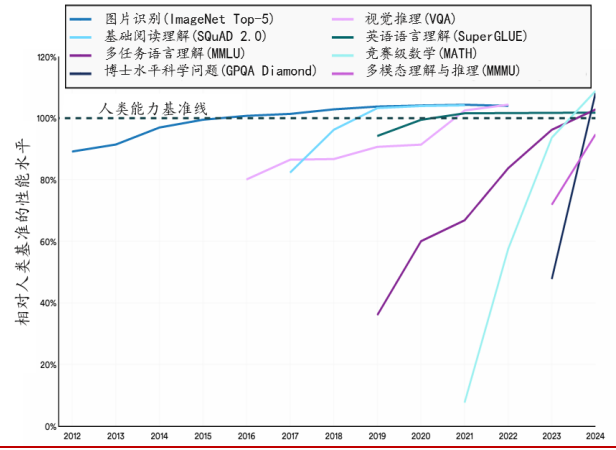
模型快速迭代，AI 功能逐渐强大，达到较高的产业化水平。早期的大语言模型需要海量的数据和成本高昂的算力进行训练。从长期趋势来看，近几年有影响力的 AI 模型成果推出更加密集，2024 年具备影响力的模型中 90% 以上的模型均为产业模型，AI 的商业化程度已经较高。2025 年以来，模型同样保持了高速迭代趋势，根据 GEO 统计数据，推理大语言模型的热度以周度统计频率仍能观察到较大变化。性能方面，在深度学习时代，AI 突破了图像分类的人类极限。随着大模型的出现，AI 迅速在阅读理解、图片推理等基础上超出了人类基准水平，已经可以胜任更加复杂的任务。

图5: 具备影响力的人工智能模型发布时间、行业及数据量



资料来源: Stanford HAI, 浙商证券研究所

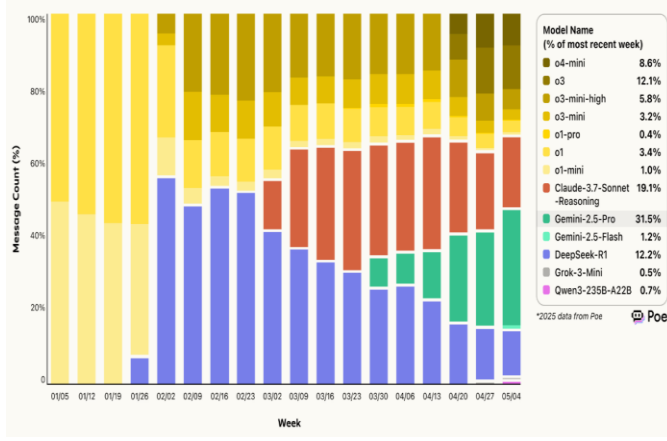
图6: 人工智能在多个能力测试上已经超过人类基准水平



资料来源: Stanford HAI, 浙商证券研究所

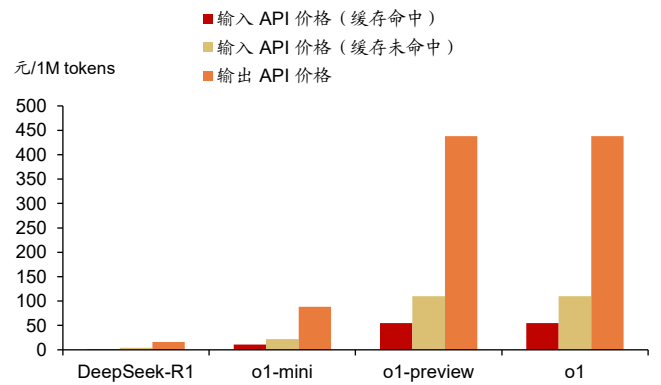
推理成本进一步降低, 使用端费用大幅下降促广泛应用。根据 Epoch AI 的统计, 根据工作任务的不同, 所有类型的大语言模型的推理成本均大幅下降, 每年下降到最初版本的 1/9~1/900 例如, 以 Deepseek 为代表的模型实现了低成本对标 OpenAI o1 (Deepseek R1 版本)、GPT4.5 (Deepseek V3 版本), 用户端的使用费用下降数十倍。随着 AI 大模型进一步快速迭代带来的性能提升+降本, 使用 AI 的学习门槛和成本门槛将显著降低, 在商业端进一步渗透, 有望在更多事务性的商业活动中发挥作用。2024 年 AI 全球各类组织应用 AI 的范围大幅提升, 根据 Stanford HAI 援引的麦肯锡调查数据, 2024 年全球各类机构使用 AI 的比例达 78%, 较上年提高 23%, 中国机构使用率为 75%, 较上年提高 27%, 是全球增速最快的地区之一。

图7: 2025 年后具备推理能力的大语言模型收到用户信息占比



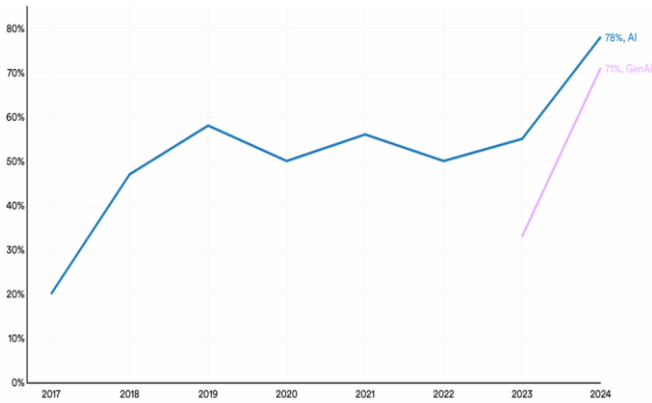
资料来源: POE, 浙商证券研究所

图8: Deepseek-R1 定价较 OpenAI o1 显著降低



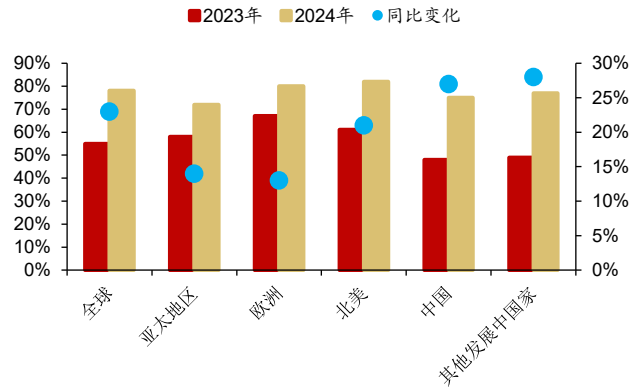
资料来源: Deepseek, 浙商证券研究所

图9：2017-2024 全球各类机构应用 AI 与生成式 AI 的比例变化



资料来源：Stanford HAI，浙商证券研究所

图10：2023-2024 分地区机构应用 AI 的比例变化



资料来源：Stanford HAI，麦肯锡，浙商证券研究所

我们认为 AI 将在各个环节帮助 ESG 投资降本增效。在 AI 持续迭代的大背景下，ESG 投资也将从中受益。在章节 1.3 和 1.4 中，我们将集中分析监管端、企业端如何使用 AI 降低合规成本，在第 2 章中，我们则着重分析资管机构如何应用 AI 赋能 ESG 投资。由于生成式 AI 同样可以通过生成代码降低机器学习、深度学习等决策式 AI 方案的使用成本，其他的 AI 分析方法也将随 AI 技术发展被更加广泛地应用到 ESG 投资领域。因此，我们在下文的分析中也不仅局限于生成式 AI 的应用场景，而是包含所有类型的人工智能。

1.3 监管机构端：AI 助力监管有效监督政策落实情况

AI 有望帮助监管端和企业端建设好 ESG 投资的“基础设施”。由监管部门、国际组织等机构建立相应的要求和倡议，企业遵循相应规则运营，并反馈/披露相关执行信息，监管端再进行相应的反馈（正面反馈如提供良好评价、提供对应标签，负面反馈如罚款、降低评价等）是 ESG 政策和倡议实施的基础模式。只有该模式良好运营，投资端才能依据披露的信息进行 ESG 实践。我们认为，AI 将帮助监管端提高监管能力，有效防范“漂绿”等行为；同时 AI 将帮助企业端降低合规成本，让企业更好进行 ESG 实践与信息披露。

1.3.1 ESG 监管痛点：难验证、难追溯，企业或主动漂绿

ESG 披露中非结构化数据较多，导致验证成本较高。企业 ESG 披露中大量数据为年报文本、供应链记录等非结构化数据，而抽样审计对非结构化数据的覆盖率低，传统的自动审阅也难以对文本进行分析。这种验证不足的情况，不仅使数据可信度遭受质疑，更催生了漂绿的投机行为，对市场信心造成了打击。

组织间数据孤岛阻碍 ESG 披露责任追溯。ESG 数据的收集和汇总需经多个部门多次转手，过程中缺乏统一的数据治理标准并存在数据壁垒。监管端一方面难以触达企业的原始数据，另一方面也难以跨越多个组织进行“穿透式”核查。

部分企业可能会进行“漂绿”行为。首先，企业选择性披露部分数据，例如企业可能会仅报告直接碳排放而忽略供应链等间接排放，造成碳排放低估；其次，企业或许采用有利的计算方法，比如使用强度指标而非绝对量来掩盖总量增长；最后，企业或许会利用数据的时间差等，让展现出的结果更为有利。这些操作都会导致 ESG 相关数据失真，使企业的 ESG 表现看起来优于实际情况，从而误导投资者和监管机构。以碳排放数据为例，BCG 的相关

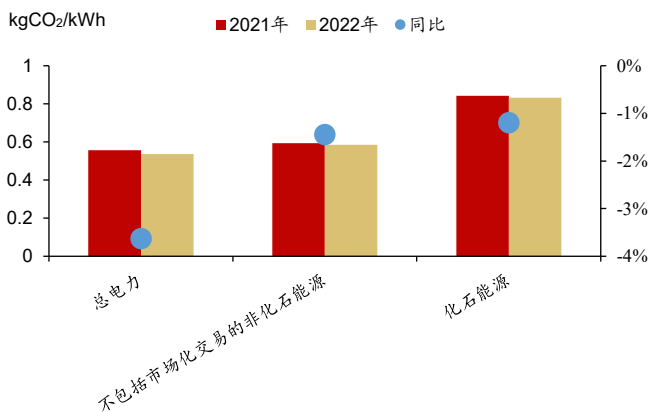
调查显示，企业在排放量估算时的平均误差率达到 30-40%，这些都影响了 ESG 数据的可信度。

1.3.2 AI 方案：动态监管与多模态交叉验证

AI 可以通过多种方式帮助提高 ESG 数据的准确性并加强对“漂绿”行为的监管。总结来看，AI 可以通过其处理信息快速的特点赋能动态监管，而生成式 AI 的发展也给融合图像、文字、数据的“多模态”验证提供了支持。仍以碳排放数据为例，AI 提升碳数据准确性的监管技术路径展望如下：

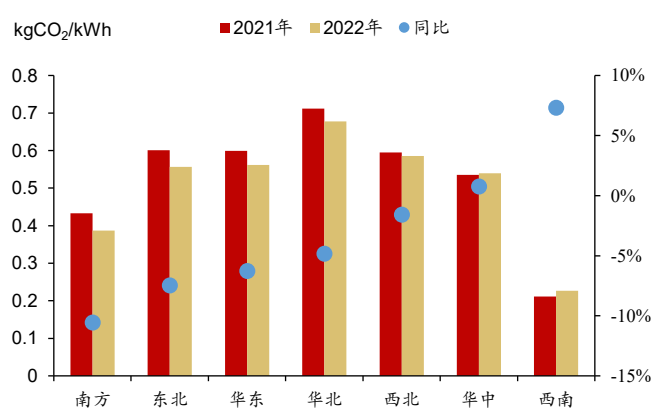
AI 帮助构建动态碳排放因子库与标准化披露监管。碳排放因子库指衡量一定量的某种活动产生温室气体排放的系数，是目前企业进行碳排放核算的重要工具，一般由监管端提供，但碳排放因子库在时间和空间上都可能发生变化，监管中可能产生数据滞后等问题。AI 首先可以帮助碳排放因子底层计算数据的收集。目前，AI 已经可以结合遥感技术的“智能延伸”已经在我国农产品碳足迹因子基准数据库助力提高了因子核算水平。其次，AI 可以自动抓取并整合 IPCC、IEA、各国生态环境部门等权威数据源，构建全球实时排放因子库，为监管提供参考。同时，监管方可以运用 AI 模型分析历史数据，预测因子变化趋势，提前预警监管机构可能的数据滞后风险，实现标准化流程监管。

图11：2021、2022 年全国电力消费平均二氧化碳排放因子



资料来源：国家温室气体排放因子数据库，浙商证券研究所

图12：2021、2022 年各区域电力消费平均二氧化碳排放因子



资料来源：国家温室气体排放因子数据库，浙商证券研究所

AI 利用多模态数据进行交叉核验。监管端可以利用自然语言处理等技术解析企业 ESG 报告，并提取报告中的碳排放相关数据，然后与监管数据库自动比对，标记矛盾点，帮助核验审查数据的准确性、完整性和可信度。这种核验方式有助于解决 ESG 投资中的数据漂绿风险，并为监管机构、投资者提供可依赖的碳排放信息。例如，根据欧盟 CBAM 法规规定，进口商须通过 CBAM 电子注册系统提交申请文件及完整的产品碳足迹数据，并接受系统自动化评估。该机制采用电子化核查方式，对申报数据的完整性和逻辑一致性进行自动验证，以确保符合欧盟技术标准要求。在此过程中，AI 验证工具通过智能算法能显著提升了数据核验的准确性和效率。

表2：欧盟碳边境调节机制（CBAM）政策内容概述

| 类别 | 政策内容 |
|------|--|
| 覆盖范围 | 商品范围：覆盖铁、钢、水泥、化肥、铝、电力和氢气，主要是具备高碳排放和碳泄漏风险商品。 国家范围：CBAM 适用于从非欧盟国家进口到欧盟的商品，欧盟内部贸易不受影响。 |
| 费用计算 | 碳价差异：费用基于 EU ETS 碳价与商品生产国已支付碳价的差额，已缴纳的碳价可扣 |

| | |
|------|---|
| | 除。 |
| | 嵌入排放: CBAM 费用取决于商品生产过程中的碳排放量, 碳足迹越高, 费用越高。 |
| 实施步骤 | 过渡期: 自 2023 年 10 月 1 日起至 2025 年 12 月 31 日, 主要进行数据收集和报告。 |
| | 逐步实施: CBAM 费用随免费配额的逐步取消而逐步引入, 影响逐渐显现。 |

资料来源: OECD, 欧盟官方公告, 浙商证券研究所

1.4 企业端: AI 助力企业多方位降低合规成本

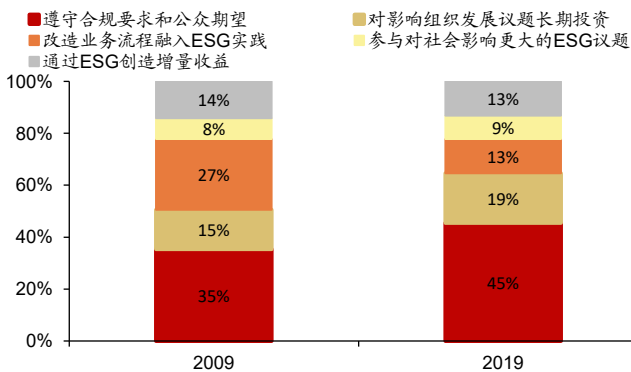
企业是履行 ESG 理念的主要主体。然而, 企业在 ESG 合规方面面临诸多挑战, 使得企业面临着较高的合规成本。企业在 ESG 合规方面面临的主要挑战包括数据收集和处理的复杂性、监管要求的不统一以及合规流程的繁琐性等。这些因素导致企业在 ESG 合规方面投入大量人力和财力, 增加了运营成本, 阻碍了企业进行 ESG 实践的积极性。但同时 ESG 合规又与 ESG 的应用非常契合, 德勤发布的报告表示, 78%的企业计划在 2025 年增加 AI 技术投入, 其中 42%的资金将流向生成式 AI 技术, 重点应用于供应链管理、合规审核等与 ESG 紧密相关的领域。

1.4.1 合规成本高成为企业参与 ESG 实践的重要阻碍

标准碎片化使得企业学习适应 ESG 规则的成本较高。当前全球 ESG 领域存在逾 600 份报告要求文件和框架 (GRI/SASB/TCFD 等), 不同监管体系的要求差异显著。这种碎片化导致企业每年需投入额外成本进行标准适配, 在多重框架并行的情况下, 企业往往难以全部理解不同标准的执行细节。跨国、跨地区企业更需要适配不同地区的标准。这导致企业需花费较高的成本理解不同的 ESG 标准, 并对企业的战略设定、合规执行、信息披露都造成了挑战。

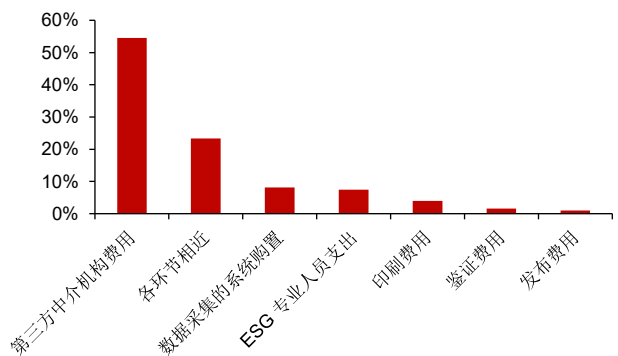
ESG 合规要求快速提升, 企业难以跟上, 可能造成战略和执行上的滞后。近年来, ESG 相关的合规要求不断提升, 但非强监管行业企业普遍缺乏应对复杂 ESG 要求的经验。BCG 调查显示仅有 2%的公司认为自身在面对 ESG 合规复杂性时的管理措施是完全有效的, 30%的企业认为相应的措施中等有效, 大部分企业合规能力存在明显短板。同时, ESG 执行的全流程需要多个部门协同参与。部分企业缺乏跨部门的 ESG 治理架构, 难以进行内部协调。

图 13: ESG 合规是参与 ESG 活动的主要驱动



资料来源: 麦肯锡, 浙商证券研究所

图 14: 国内上市公司可持续报告披露成本占比最高的部分



资料来源: 上市公司协会, 浙商证券研究所

ESG 信息披露报告通常涉及大量的数据收集、整理和分析工作, 形成了额外成本。企业需要从多个部门和外部供应商处获取数据, 进行复杂的计算和验证, 同时在披露文本上配

合监管的相关规范，最终才能生成符合监管要求的报告。以欧盟企业可持续发展报告要求为例，企业需要报告六大项下的数十个细分信息，才能够满足监管要求的披露标准。

表3: 欧盟可持续披露规则适用范围和适用内容

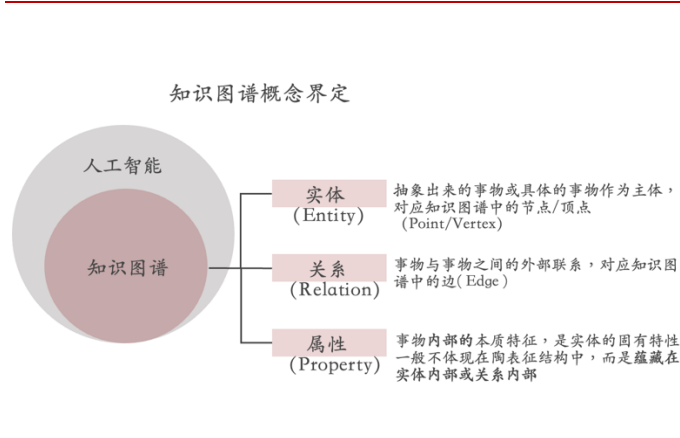
| 类别 | 对象 | 内容 |
|-------|---------------|--|
| 适用范围 | 大型企业 | 所有大型企业（非微型）及上市企业需报告。 |
| | 非欧盟企业 | 年营业额>1.5 亿欧元且在欧盟有子公司（大型/上市）或分支机构（年营业额>4000 万欧元）。 |
| | 中小企业（SMEs） | 上市中小企业从 2026 年起报告，2026-2028 年可暂不报告（需声明原因）。 |
| 报告内容 | 商业模式与战略 | 包括气候韧性、转型计划（如 1.5°C 目标）、利益相关方影响。 |
| | 目标与进展 | 时间绑定目标（如 2030/2050 减排）、科学依据、进展披露。 |
| | 治理 | 管理层角色、专业知识、激励机制（如与 ESG 挂钩的薪酬）。 |
| | 尽职调查 | 覆盖价值链（运营、产品、供应链），遵循 UN/OECD 标准。 |
| | 风险与指标 | 气候、生物多样性等风险；量化指标（如碳排放、性别薪酬差距）。 |
| | 价值链信息 | 全价值链披露，前三年允许部分缺失（需说明原因）。 |
| 报告标准 | 通用标准 | 2023 年 6 月 30 日前通过，覆盖所有行业。 |
| | 行业特定标准 | 2024 年 6 月 30 日前制定，针对高风险行业（如能源、制造业）。 |
| | 中小企业简化标准 | 2024 年 6 月 30 日前制定，比例适度。 |
| 审计与保证 | 有限保证（2026 年前） | 法定审计师或独立机构出具意见。 |
| | 合理保证（2028 年后） | 可能过渡，需评估可行性。 |
| | 审计内容 | 包括标准符合性、数据标记、欧盟分类法第 8 条披露。 |

资料来源：CSRD 摘要，浙商证券研究所

1.4.2 AI 政策智能解析：从被动应对到主动适应

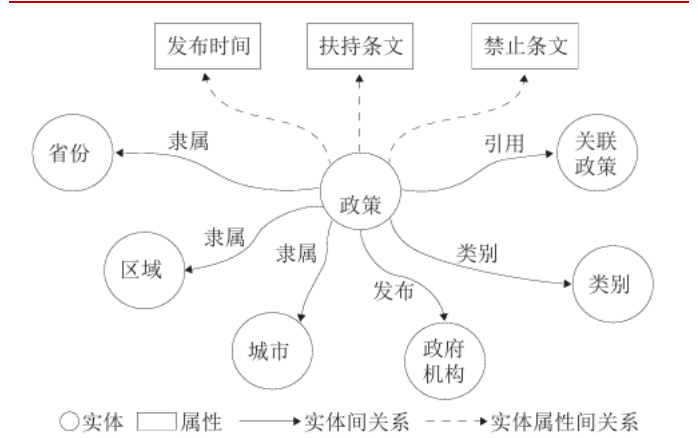
自然语言处理+知识图谱，帮助企业迅速理解合规要求。基于自然语言处理技术，AI 能够深度理解政策文本的复杂语义，准确识别关键条款和合规要求。其次，通过知识图谱技术，AI 可以建立政策条款与企业实际业务之间的智能关联，实现精准映射，帮助企业对与自身相关的 ESG 合规政策做出反应。知识图谱通过三元组即“实体×关系×属性”集合的形式来描述事物之间的关系，将非线性世界中的知识信息结构化、可视化，辅助人类进行信息的推理和归类。目前在政策分析中已经有成熟的应用。

图15: 知识图谱概念界定



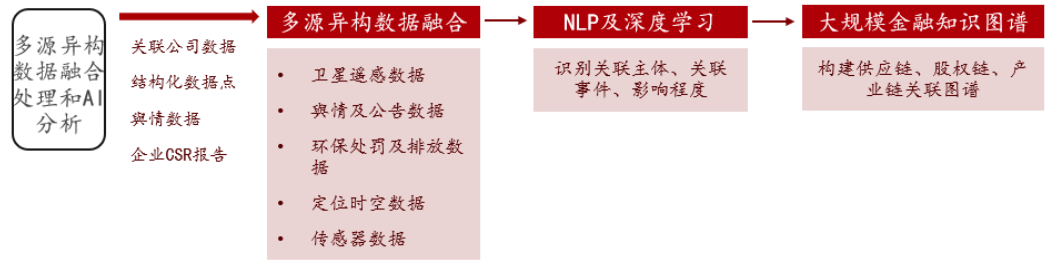
资料来源：艾瑞咨询，浙商证券研究所

图16: 政策知识图谱分析的本体模型



资料来源：张雨等《科技政策知识图谱构建研究》，浙商证券研究所

图19: 多元异构数据的融合处理与 AI 分析流程图



资料来源:《AI 驱动的 ESG 投资链接可持续未来》, 浙商证券研究所

文本生成: 从数据到文稿, 跨越 ESG 报告编写的“最后一公里”。信息披露文件的起草与审核是 ESG 合规工作的基础性任务之一。传统方式下撰写相关文书耗时耗力。有了人工智能特别是大语言模型 (LLM) 的助力, ESG 合规的“最后一公里”正在逐步实现自动化。生成式 AI, 如 GPT-4 可根据输入数据自动生成符合 GRI、SASB 等标准的报告初稿, 并支持多语言输出。企业甚至可以考虑训练专属于所在行业的垂直大模型, 更加具备专业性。AI 除了生成文书, 对文书进行自动校验, 包括语言逻辑一致性、合规规范性等都可以通过 AI 进行检查。同时, 随着语言模型的能力逐渐增强, AI 在多语言翻译方面也可以提供更加精准的支持。

2 AI 技术赋能, 为资管机构提供了 ESG 投资的丰富工具

AI 的低成本和便捷性为投资端也提供了工具包。我们认为, AI 对投资端的影响总体体现在两个方面, 一是在投资与风控端, AI 工具从数据收集、数据处理到后续的跟踪监控都提供了自动化的解决方案, 同时大大扩展了 ESG 投资的信息处理范围, 从结构化数据迈向多模态数据。二是生成式 AI 可以通过快速生成图文乃至视频, 以更快速、可视化、易理解的方式与广泛的资管产品投资者互动。

2.1 投研与风控: AI 带来更加丰富 ESG 投资手段

2.1.1 数据收集与处理: AI 助力多模态数据的敏捷挖掘和快速处理

ESG 投研过程中面临着信息收集和处理的难题。一是 ESG 相关信息量大, 且来源分散。ESG 信息涵盖环境、社会和公司治理三方面, 环境信息既包括公开披露的企业环境战略信息, 也包括企业的环保处罚记录、碳排放数据及资源能耗数据等; 社会信息包括企业相关事件、供应链信息等, 治理信息包括董事会架构、公司战略目标、公司董监高变动等。ESG 评级体系一般都包括几百甚至上千数据点。在信息量巨大的同时, ESG 信息来源也分布多个渠道, 除年报、ESG 报告 (包含社会责任报告、可持续发展报告等) 由企业主动披露的信息外, 还包括由新闻媒体、社交媒体、数据媒体等公共平台发布的信息, 以及由发监管政策部门所发布的政策信息。

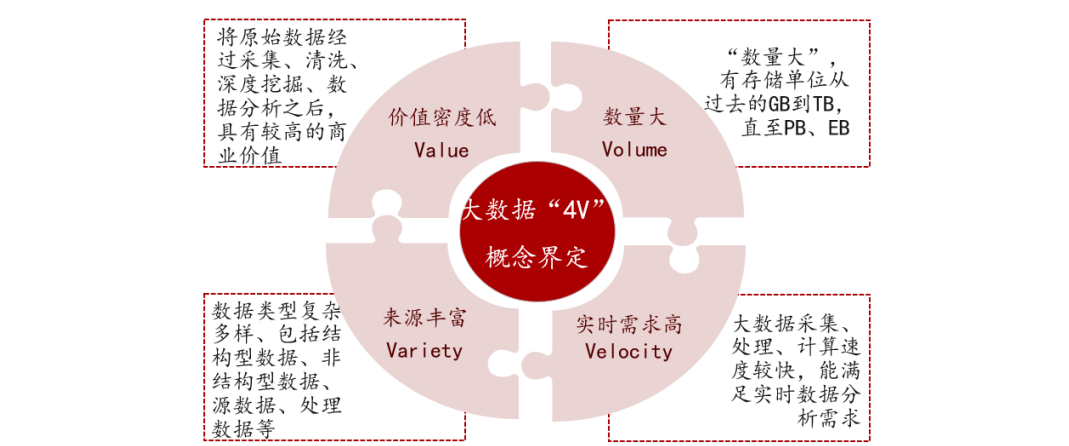
二是非结构化数据多。除了少数数据是企业披露的结构化数据之外，大部分 ESG 数据都是非结构化数据，主要以文本的形式分布在企业 ESG 相关报告、企业年报、新闻报导和相关信息平台中。传统上需要人力对不同的非结构化数据进行打分，人力成本高昂。

三是实时性需求无法满足。以 ESG 评级为例，多数数据都来源于企业的定期报告和社会责任报告。这些报告频率大多数在一年，时效性较为滞后，难以满足投资要求。但和 ESG 相关的信息可能实时反映在股价中，如重大的环境处罚会在公告后对企业估值快速产生实质性影响。对争议性事件的实时监测是 ESG 投资的重点需求。

四是数据颗粒度无法满足更深入的研究需求。当前，气候风险越来越引发投资人的关注。但目前的 ESG 数据完全无法满足更深入的 ESG 研究需求。例如，在物理风险测量方面，投资者需要衡量企业所有资产面临的气候变化物理影响，但大型企业的资产往往分布在全球各地，不同地区的资产实际情况以及气候情况较难整合，已经成为气候风险量化过程中的主要挑战。

大数据+AI 在解决上述问题中逐渐扮演着越来越关键的作用。投资端对 ESG 数据需求本质上是一种大数据挖掘+处理的需求，具备大数据的“4V”特征：数量大（volume）、来源丰富（variety）、实时性需求高（velocity）、价值量低需要处理（value）。因此，大数据处理手段也可以同样应用在 ESG 数据处理的过程中，同时利用 AI 更好赋能数据分析的全流程。

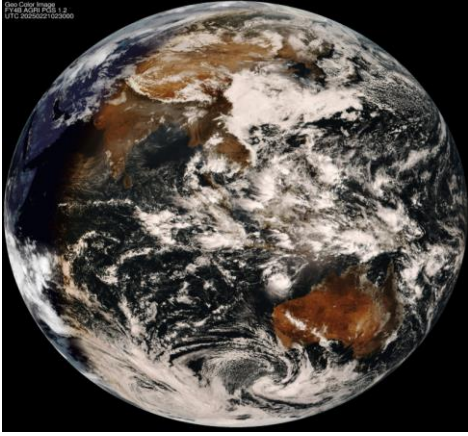
图20： 大数据“4V”特性



资料来源：政府门户网站，浙商证券研究所

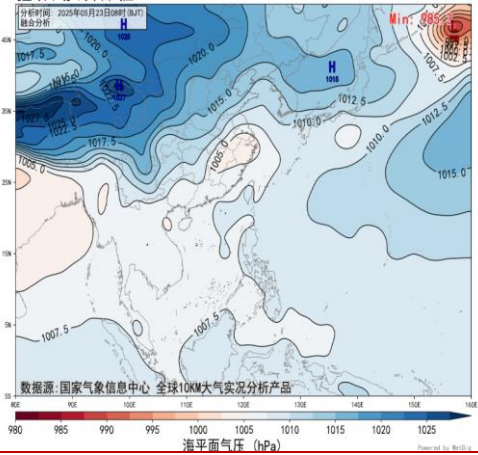
AI 可使 ESG 数据收集的范围更广。通过大语言模型、机器视觉、机器学习模式识别等工具，AI 将数据收集的范围从结构化数字扩展到非结构化的文本、图片、PDF 等多种来源。例如，国内数据公司已将大语言模型融入了对 ESG 报告数据提取的流程中，通过语义检索，能够同时在报告中提取定量结构数据和定性的非结构数据；图像信息方面，卫星图像、气候图像数据也是 ESG 投资者重点关注的另类数据。英国 Cervest 公司开发了 Earth-Scan 平台，利用人工智能技术获取卫星图像和投影中的气候数据，创建了动态的地球气候模型，并将进一步的环境数据（水、大气、生物多样性及人口要素等）纳入其中，为气候物理风险的衡量提供了数据支撑。

图21: 风云四实时卫星图像



资料来源: 国家气象信息中心, 浙商证券研究所

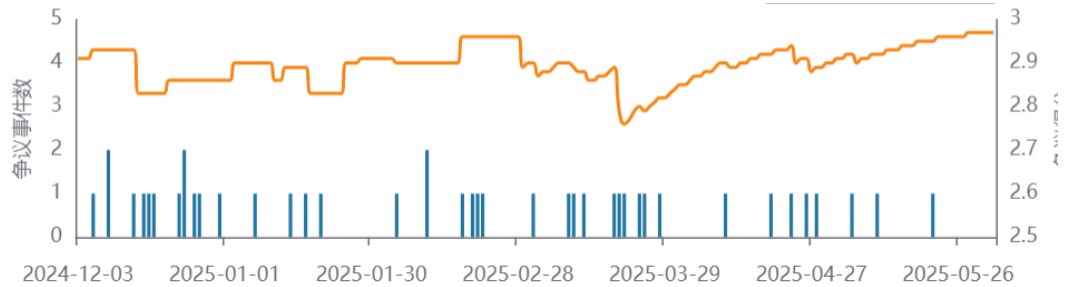
图22: 台风监测-大气海平面气压实时情况



资料来源: 国家气象信息中心, 浙商证券研究所

AI 帮助投资者实时跟踪 ESG 数据, 控制突发的投资风险。 AI 的实时跟踪主要体现在对争议事件的高频更新中。数据商往往通过大语言模型、图片识别等技术实时提取信息。一些机构也是用神经网络进行信息筛选, 大大减少了人工流程。目前, MSCI、Wind 等数据商均提供日度更新的争议事件数据, 以便投资者及时追踪全媒体 ESG 信息。以 MSCI 为例, 该数据商通过密切跟踪数千家媒体的新闻报道, 提取包括商业道德、环境污染、消费者诉讼、环保公告、监管调查约谈等信息事件。Wind 的争议事件则分类为新闻舆情、监管处罚、法律诉讼三大类。

图23: Wind 跟踪的中国平安近半年的争议事件信息



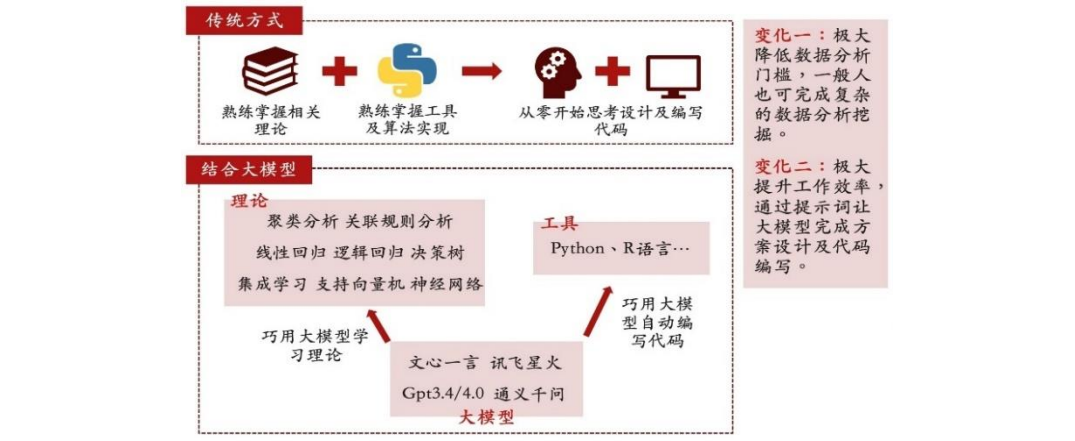
资料来源: Wind, 浙商证券研究所

注: 数据截至 2025 年 6 月 3 日

AI 帮助投资机构和第三方机构将原始信息转化为“终端友好”数据。 即人工智能可以将非结构化的零散数据处理为便于 ESG 投资的数字型数据或可视化关系图等。例如, 面对 ESG 评级中定性描述转定量指标的工作, 研究人员可以利用历史数据对 AI 模型进行有监督训练, AI 模型在后续即可帮助研究人员智能化打分。针对争议事件文本, 投资人可以利用大语言模型分析事件主体, 并对监管处罚的严重程度进行分类; 也可以通过语言情绪分析, 识别监管机构的态度。此外还可以利用知识图谱梳理不同的争议事件与利益相关方的关联。面对数据中的缺省值问题, 投资机构可以采用聚类等相似性分析方案, 实现更精确的估计和填充。

除了本身的技术提升外，AI 还大大提高了数据分析的工作效率。传统数据分析工作需要研究人员掌握数据分析的理论和相关编程工具。但当下结合 AI 大模型，无论是否具备相关知识，都可以通过人工智能搜索相关理论，同时利用 AI 生成代码来完成数据分析的编程工作。内部集合 AI 的数据分析工具也能帮助投资者一键实现相关任务。例如，Scrape Graph AI 结合了大型语言模型和图逻辑，只需通过简单描述就能自动提取网页信息，无需编写复杂的数据挖掘代码，同时支持标准化输出格式，大大简便了数据挖掘的步骤。

图24：传统数据处理方式与结合大模型的数据处理方式对比



资料来源：数据 STUDIO，浙商证券研究所

2.1.2 AI 助力优化投资策略

目前 AI 优化投资策略有三个主要的研究方向。一是利用机器学习模型的优化能力筛选出重要的 ESG 因子，并用这些因子对原始投资组合进行增强；二是利用语言模型将 ESG 信息作为主要的投资策略；三是利用 AI 模型的预测能力，进行某个方面的数据预测，并进行该方向的主题投资。

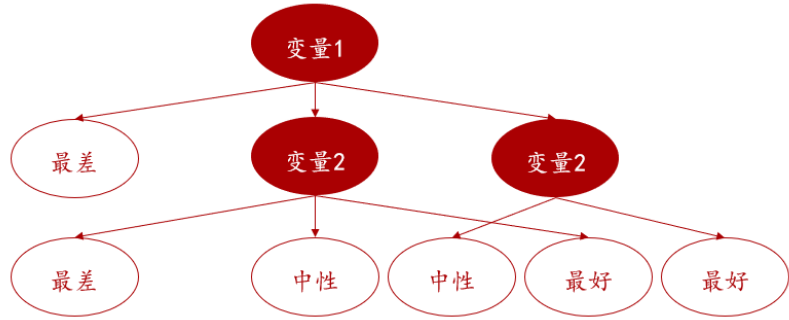
机器学习模型 ESG 因子挖掘

机器学习模型的 ESG 因子策略具备比较强的适用性，有望更多运用在指数增强产品中。我们以 Enrico Bernardini 等人 (2023)¹ 的论文为例叙述该策略的主要思路。该论文使用的初始因子数据包括 REFINITIV ESG 数据库的 100 个 ESG 变量、MSCI 数据库的 ESG 支柱项得分和关键议题得分、以及少数基本面因子，共计 217 个 ESG 因子构建了初步的因子库。研究的股票是 EURO STOXX 300 指数内的公司。为了排除其他影响，作者剔除了退出和新加入指数的公司，最后共计 252 只股票在研究范围。研究的时间范围包含从 2006 年 12 月 31 日至 2019 年 4 月 30 日的时间序列。

论文主要选取了机器学习中的决策树模型进行因子筛选。论文依次将最大化收益、最小化波动、最大化夏普比率设置为优化目标，以 ESG 因子构建多个两层决策树。具体来说，是选用一个 ESG 因子作为划分决策树的阈值（例如碳排放量最高 25% 公司，中间 50%，最低 25%），计算分割后这些公司的绝对收益（或者波动率和夏普比率）的表现如何，并选用划分后 ESG 绩效最好的组绝对收益提升最大的划分阈值。然后再用另一个 ESG 因子构建第二层决策树，最优化绝对收益。在所有的因子组合中构建完这样的两层决策树后，选择其中 ESG 因子对绝对收益提升最大的 10 个因子构建组合，并用因子大小加权。

¹ Lanza, Ariel & Bernardini, Enrico & Faiella, Ivan. (2023). Machine Learning, ESG Indicators, and Sustainable Investment. 10.1007/978-3-031-33882-3_10.

图25： 两层决策树结构助力 ESG 因子筛选



资料来源：《Machine Learning, ESG Indicators, and Sustainable Investment》，浙商证券研究所

该策略使组合表现明显优化。论文中的结果显示，在整合该 ESG 投资策略后，组合表现有明显改善。优化回报率、波动、夏普的三大模型均在对应的优化目标上表现良好，且在 Fama-French 模型和 BIRR 模型下都录得明显的正向阿尔法。

表4： 基于决策树的 ESG 因子策略表现

| 模型 | 绝对收益 | | 波动 | | 夏普比率 | |
|------------------|---------|---------|---------|---------|---------|----------|
| | 训练集 | 测试集 | 训练集 | 测试集 | 训练集 | 测试集 |
| 回报率最优 | 4.50% | 1.20% | 1.20% | -0.60% | 2.40% | 0.50% |
| 波动率最优 | 0.01% | -0.02% | -0.12% | -0.09% | -0.18% | -0.09% |
| 夏普比率最优 | 0.07393 | 0.03856 | 0.02661 | 0.02058 | 0.04853 | 0.046937 |
| FF 模型 α | 3.66% | | 0.81% | | 1.70% | |
| BIRR 模型 α | 3.28% | | 0.23% | | 1.07% | |

资料来源：《Machine Learning, ESG Indicators, and Sustainable Investment》，浙商证券研究所

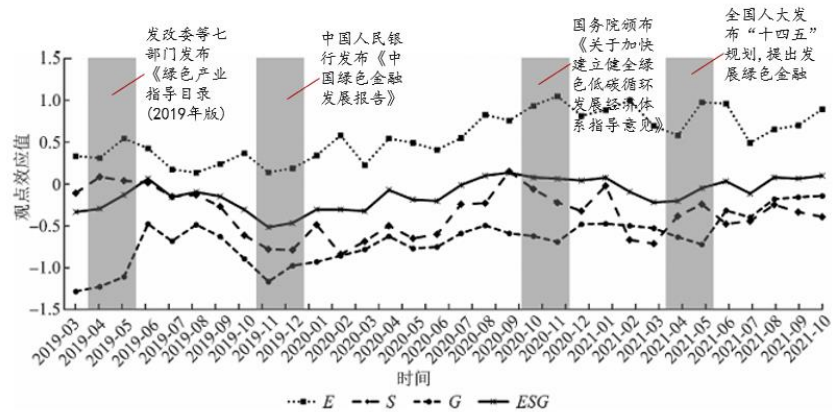
除了决策树外，仍有其他方法在因子筛选上具备优势。上述论文选择了决策树模型作为筛选有效 ESG 变量的方案。论文构建的两层决策树结构简单，利于直观理解变量的相互作用，能够解决机器学习中经济解释能力弱的问题。但仍有其他模型在筛选因子上表现较好，例如线性模型中的 LASSO 方法和 SCAD 方法等，神经网络则是非线性优化的主要方式。AI 模型能够在难以观察到的特征中建立联系，但随着模型越复杂，算法越来越成为“黑箱”，得到结果的经济机理难以解释。投资人有时不得不在算法有效性和可解释性上进行选择和权衡。

ESG 语言模型分析

利用语言模型进行文本分析的主要依据是金融文本与金融市场的相关性。文本信息（包含企业披露信息、新闻媒体、网络讨论信息等）随着互联网的发展快速增长，拓展了金融市场可预测性的分析途径。互联网传播改变了传统媒体向投资者传递信息的方式，投资者对信息的反应效率和反应程度也在不断改变。交易行为和文本情绪相互影响，使得金融交易数据可以与大数据下的网络文本信息相匹配。自然语言处理等技术手段的兴起也使投资者能够通过科学方法分析文本数据。

国内市场研究指出文本分析获得的 ESG 情绪能够被市场定价。徐凤敏等 (2023)², 利用 2018—2021 年 40 余万篇企业信息披露与新闻媒体文本数据, 通过自然语言处理技术分析企业信息披露与新闻媒体中的情绪倾向, 并将所得结果量化为时间序列上的情绪观点指数。该情绪指数的变动与国家政策变化相关, 能够较好反映文本信息。通过回归发现, 文本中隐含的 ESG 观点被股票市场定价, 且体现出投资者的部分投资偏好。“双碳”目标提出后, ESG 上述观点对股票收益率的影响得到加强。ESG 观点对股票收益率的影响存在月度反转效应, 收益在政策的下个月有部分修正。

图26: ESG 观点值与国家 ESG 相关政策变化相关性



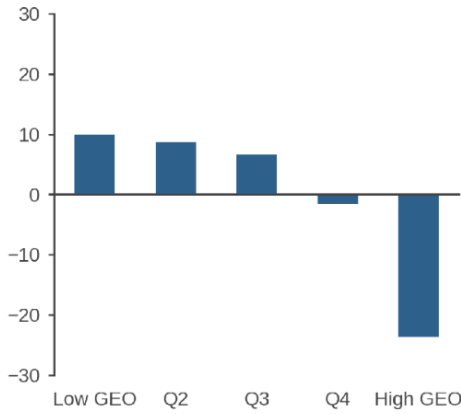
资料来源:《ESG 观点与股票市场定价》, 浙商证券研究所

人工智能算法预测

机器学习和深度学习算法能够在训练集上的训练, 可以发现高维的特征, 并根据特征进行识别和预测。在投资中, 这种能力可以帮助投资者识别具备特定特征的股票, 并且在时间上具有前瞻性。海外资管机构 ACADIAN 开发了一种基于自然语言处理技术的机器学习模型, 通过分析财报电话会议记录、监管文件和年度股东大会的信息, 衡量公司是否符合能源转型企业的定义。该系统被称作 GEO (绿色经济机会) 系统预测信号。该信号在能源转型方面具备前瞻性, 与企业远期碳排放密度的下降水平、科学碳目标执行、绿色专利比例等数据正相关。根据 GEO 数据, ACADIAN 开发了一系列产品在不同程度上提升 GEO 敞口, 为投资者提供投资转型相关企业的工具。

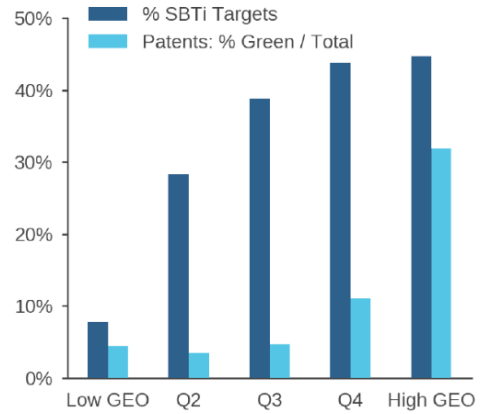
² 徐凤敏, 马杰傲, 景奎. ESG 观点与股票市场定价——来自 AI 语言模型和新闻文本的证据[J]. 当代经济科学, 2023, 45(6):29-43.

图27: GEO 五分位对应企业远期碳强度同比变化



资料来源: ACADIAN, 浙商证券研究所

图28: GEO 五分位对应企业致力于科学碳目标和绿色专利比例



资料来源: ACADIAN, 浙商证券研究所

2.2 AI 助力资管企业与投资人进行沟通

人工智能重构基金营销生态。业内人士认为,生成式 AI 推动了基金营销的转型。在营销过程中,过去工作流程繁琐、消耗工作人员大量精力的营销内容制作(包含文案、图片、视频等形式)、客户画像构建和营销策划等事务性任务,通过 AI 大模型的加持,生产效率和质量显著提升。南方基金推出了基金的 AI 宣传片,文案、画面、配音均由 AI 生成。自动化视频制作能技术大大降低了宣发内容创作成本,有助于更多中小基金公司进行品牌营销。

人工智能已经可以部分替代沟通职能,有助于普及 ESG 产品信息和 ESG 投资理念。目前已经有一部分基金公司上限了智能客服服务,通过与智能客服的在线沟通,投资者可以了解详细的产品信息和投资信息,这有助于进行投资者教育,更好普及 ESG 投资理念。我们测试了华夏基金、易方达基金、嘉实基金的智能客服,键入“ESG 基金产品”信息,智能客服可以反馈对应的产品名称或产品代码,继续追问也可以获得基金投资范围等详细信息。

表5: 易方达、华夏、嘉实三家基金智能客服对 ESG 有关提问的回复

| 输入问题 | 易方达基金 | 华夏基金 | 嘉实基金 |
|--------------|------------------------------|---|---|
| ESG 基金产品有哪些? | 易方达 ESG 责任投资股票基金相关问题 | 华夏 ESG 可持续投资一年持有混合 A 基金网页 | “目前嘉实涉及新能源主题相关产品如下,您可以点选查看 ★主动精选(基金经理主动选股配置) ★被动精选(被动跟踪相关指数)” |
| 某基金产品代码是什么? | “易方达 ESG 责任投资的基金代码为: 007548” | “华夏 ESG 可持续投资一年持有混合 A 类代码为 014922, C 类代码为 014923” | 点击“被动精选”,输出基金名称及对应基金产品代码 |
| 某基金经理是谁? | “易方达 ESG 责任投资的基金经理是: 郭杰” | “华夏 ESG 可持续投资一年持有混合基金(基金代码 014922/014923)的现任基金经理是: 潘中宁” | 询问特定基金的基金经理,输出对应基金经理简历 |

资料来源: 基金官网智能客服, 浙商证券研究所

注: 引号内的内容为原回复的应用, 非引号的内容为对应回答内容的概括。表格内涉及的内容仅作为基金公司智能客服情况的展示, 不涉及对任何具体产品的推荐。

3 AI 产生广泛外溢影响，ESG+RAI 框架助力分析风险与机遇

AI 与 ESG 的融合正在重构责任投资的基础架构和运行逻辑。一方面，AI 不仅驱动 ESG 基础设施的智能化转型，更通过打通 ESG 投资全链条实现投资效率的跃升；与此同时，AI 的高速发展也产生了伦理问题和社会争议，AI 外溢的正面或负面影响与 ESG 议题相交织。但相关法规尚不完善，监管也仍会就出现的争议对相关企业提出要求。为了有效管理 AI 外溢至多个方面的风险和机遇，融合覆盖公司广泛非财务信息的 ESG 框架，是一个值得探讨的思路。我们认为，ESG+负责任 AI 框架或将成为投资者识别和管理 AI 相关风险的手段。

3.1 ESG 议题角度下，AI 的效率提升与伦理争议

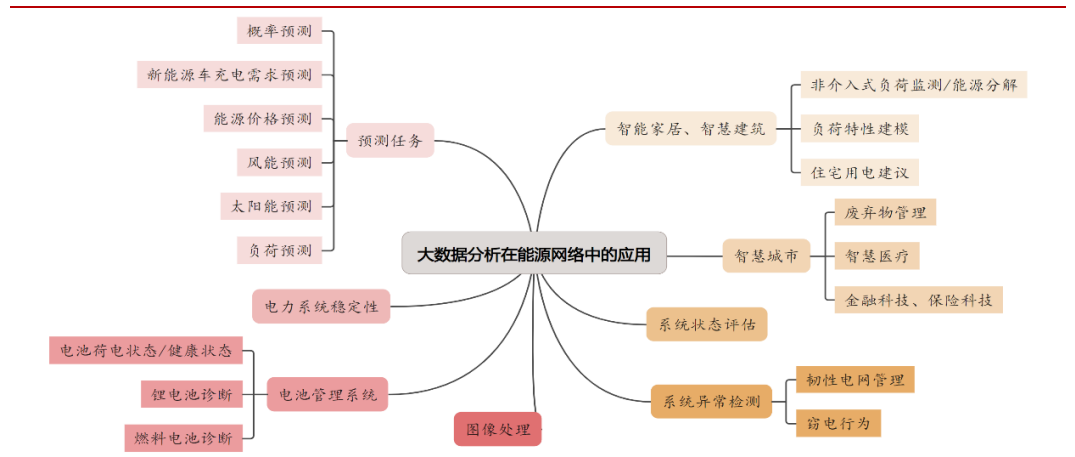
AI 在 ESG 议题中表现出影响的两面性，即具备在系统中提高效率的能力，同时也产生了诸多挑战与争议，并且争议也随着 AI 应用的快速推广而放大。在环境层面，AI 能够提升环境管理系统的效能，但 AI 训练、运营所需的算力也产生了大量碳排放；在社会层面，AI 即降低了信息获得的成本，并赋能了科研和医疗，但也潜藏着传播偏见、冲击就业的隐患；在公司治理层面，AI 可以帮助决策，但也冲击着现存的监管与法律。

3.1.1 环境：提升环保效率与自身消耗资源的两面性

AI 助力提升能效

在环境方面，AI 推动系统性环境治理创新。在检测端，AI 通过结合无人机、传感器等硬件设施，可以实现环境数据的智能检测；在管理系统端，AI 则可以整合环境数据，实现从单点优化到全局治理的跨越，通过数字化重构能源系统和智能化升级治理模式，AI 可以系统性破解传统环境治理中效率低下、监管滞后和基础设施不完善等结构性难题。具体而言，AI 可以应用在环境保护、自然资源开发、能源应用、生物多样性等多个环境议题，推动系统性的治理提升。我们以能源为例具体分析。

图29：大数据与 AI 分析在能源网络中的应用



资料来源：《AI-powered Energy Internet Towards Carbon Neutrality: Challenges and Opportunities》，浙商证券研究所

能源供应端角度：AI 帮助优化可再生能源系统。随着风光新能源在电网中的占比不断提高，新能源发电出力的随机性、波动性和间歇性，正成为威胁电力系统安全、稳定、高效运行的核心挑战。传统上，电站主要根据物理模型和统计方法对新能源发电功率进行预测，以提前应对电网波动。但面对复杂多变的气候条件和不断增长的新能源装机规模，传统的预测方式逐难以满足精准预测的需求。根据中国能源网的报导，华为推出了 AI 解决方案。利用大模型技术，使试点场站超短期 15 分钟预测准确度达到 97.24%，4 小时预测准确度达到

91.72%，24 小时内的中短期预测整体准确度超 90%，显著优于传统超算预测模式。同时，该方案使风电场和光伏电站的考核费用分别降低了约 27%和 15%，有效削减了电站运营成本。

能源网络端：AI 技术能为大数据分析提供强大的支持，从单点应用向综合大模型跨越。

大数据分析在能源互联网中的最新应用借助 AI 技术得以更好实现，具体应用包括预测、监测、感知、控制等领域。在早期实践中，机器学习算法主要为小范围能源网络提供预测服务，但近期，AI 应用呈现大数据、系统性的趋势。特别是以大模型为代表的深层次人工智能技术，充分利用海量多模态数据，为新型电力系统网络资源精准高效配置、自然灵活调节供需提供了综合解决方案。国家电网在 2024 年 12 月发布了光明电力大模型，这也是国内首个千亿级多模态电力行业大模型，能够参与电网规划、运维、运行等诸多环节；而南方电网“大瓦特”大模型也在 2024 年推出，在 2025 年初又接入 DeepSeek 进一步升级。

表6：人工智能模型在电力系统中的应用

| 任务名称 | 年份 | 方法 | 数据来源 |
|--------------------|------|----------------|----------------------------------|
| 光伏发电预测 | 2014 | 机器学习 | 美国数十个光伏电站相关天气、传感器、卫星等数据 |
| 电力盗窃检测 | 2015 | 机器学习(聚类+SVM) | 法国电力客户行为试验 |
| 光伏发电预测 | 2016 | 深度学习(AutoLSTM) | 德国太阳能发电场数据 |
| 住宅电力负载预测 | 2017 | 深度学习(LSTM) | 澳大利亚智能城市项目：住宅用电负载 |
| 电力价格预测 | 2017 | 深度学习(WASENN) | ISO 新英格兰电力价格数据 |
| 概率网络负载预测 | 2019 | 深度学习(贝叶斯 LSTM) | 美国太阳能和负载数据 |
| 电动汽车充电需求预测 | 2020 | 深度学习(ANN) | 国家公路旅行调查 |
| Netflix 电网故障监测智能系统 | 2020 | 机器学习系统 | 爱沙尼亚电网智能电表数据 |
| 一小时风功率预测 | 2020 | 深度学习(深度残差网络) | ERCOT 间歇性可再生能源资源数据 |
| “大瓦特”电力大模型 | 2024 | AI 大模型 | 南方电网多场景多模态数据 |
| 虚拟电厂+AI | 2025 | 深度学习/AI 大模型 | 中国电力市场交易政策、负荷特性分析和交易辅助决策专家经验文本数据 |

资料来源：《电力系统自动化》期刊，《AI-powered Energy Internet Towards Carbon Neutrality: Challenges and Opportunities》，南方电网，国家核安全局信息门户，AI 能源公众号，浙商证券研究所

能源消费端：助力重点行业节能减排。工业、建筑与交通减排是用能刚性需求最强、技术替代路径最明确的突破口。AI 技术可以通过多种途径在不同的工业减排场景中得到应用。具体而言，多种 AI 算法通过助力全方位节能降耗、加快供能结构低碳转型、推动工艺流程清洁化和助力负排放技术部署四类方式，直接或间接地推动工业领域实现“双碳”目标。

表7：AI 助力工业脱碳的应用场景与作用机制

| | 机器学习 | 强化学习与智能控制 | 时间序列分析 | 机器视觉与智能传感 | 智能通信 | 大模型与自然语言处理 |
|--------------------|-----------------------|-------------------|----------------|------------------------|---------------------------------|-----------------------------|
| 1 助力全方位节能降耗 | | | | | | |
| 1.1 促进能效提升 | 助力节能技术、高性能材料研发 | 实现工艺过程的动态优化和控制 | 精准预测工艺能耗曲线 | 实时提供工艺过程数据；实现智能检测与安全运营 | 安全高效地提供数据调度、集成、运算等基础服务；支持生命周期管理 | 整合专家知识，辅助相关技术研发和工艺设计 |
| 1.2 降低非必要能源需求 | 助力回收再利用技术研发；高效识别高能耗环节 | 提高自动化水平，减少非必要能源损耗 | 优化库存管理与供应链管理过程 | 支持产品生命周期管理 | | 优化运营管理决策；支持生命周期管理；助力需求端行为改变 |

2 加快供能结构低碳转型

| | | | | | | |
|---------------|---|----------------|-------------------------------|--------------|---------------------------------|--|
| 2.1 提高清洁能源渗透率 | 助力相关材料和技术研发 优化多能协同供应策略和相关部署决策；优化设施布局 | 实现动态的多能管理和综合调度 | 精准预测可再生能源供应曲线；提高能源供需匹配度与技术协同度 | 提供辅助能源调度的数据集 | 安全高效地提供数据调度、集成、运算等基础服务；支持生命周期管理 | 整合专家知识，辅助相关技术研发和工艺设计 减少各环节信息不对称性；优化运营管理决策 |
|---------------|---|----------------|-------------------------------|--------------|---------------------------------|--|

3 推动工艺流程清洁化

| | | | | | | |
|----------------|----------------------|----------------|-----------------|-------------|---------------------------------|--|
| 3.1 助力低碳新工艺应用 | 助力工艺过程革新；辅助相关材料和技术研发 | 实现工艺过程的动态优化和控制 | 精准预测工态参数，降低检测时滞 | 实现智能检测与安全运营 | 安全高效地提供数据调度、集成、运算等基础服务；支持生命周期管理 | 整合专家知识，辅助相关技术研发和工艺设计 减少各环节信息不对称性；优化技术升级相关决策 |
| 3.2 优化传统高碳工艺流程 | 辅助优化水泥配方、化学反应条件等 | | | | | |

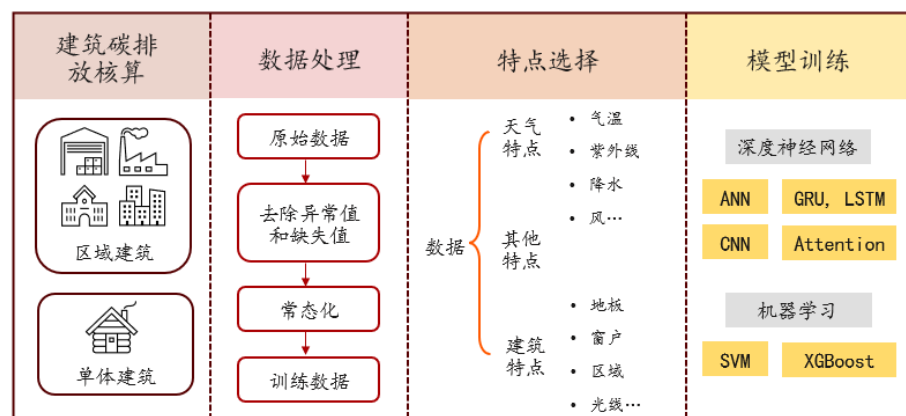
4 助力负排放技术部署

| | | | | | | |
|-----------------------|----------------------|----------------|--|-------|--|----------------------|
| 4.1 助力 DAC、CCS 等技术的应用 | 助力相关材料和技术研发；识别优质碳封存区 | 优化碳封存的注入策略 | | 封存点监测 | | 整合专家知识，辅助相关技术研发和工艺设计 |
| 4.2 助力各类碳利用技术的应用 | 辅助碳利用相关工艺和技术研发 | 实现工艺过程的动态优化和控制 | | | | |

资料来源：《学术前沿》期刊，浙商证券研究所

人工智能在推动建筑行业碳减排方面发挥着关键作用，主要体现在智能能耗管理和低碳优化两大领域。在能耗监测与预测方面，AI 融合传统统计模型（如 ARIMA）和深度学习技术（如 LSTM），实现了建筑能耗的精准预测，为能源管理决策提供科学依据，使管理者能够优化能源调度，有效减少能源浪费。在低碳建筑智能化方面，AI 通过电力系统智能调控、HVAC 和照明设备自动化管理、可再生能源协同优化以及电动汽车充电智能调度等多维度的创新应用，显著提升了建筑能源使用效率。这些智能解决方案不仅降低了建筑运营的碳排放，更为建筑行业绿色转型提供了切实可行的技术路径，推动整个行业向更可持续的未来发展。

图30：机器学习技术如何推动建筑行业减碳



资料来源：《AI and Big Data-Empowered Low-Carbon Buildings: Challenges and Prospects》，浙商证券研究所

人工智能可以优化交通效率和交通结构实现碳减排。碳减排可通过效率导向策略（优化交通系统性能）和结构导向策略（推广公共交通与非机动出行）实现，其中后者对减少城市交通能源消耗的影响更显著。效率导向策略主要指 AI 在交通技术中通过提升流量预测准确性（如记忆网络、时空 GIS 分析）、优化出行路径（如匹配框架、轨迹预测）以及分析个体

差异（如路网分类、行为预测）来弥补传统技术在指导交通效率上的不足；同时，AI 在优化交通结构方面也发挥了关键作用，包括提高轨道交通能源效率、监测公交运营效率、规划共享单车布局，以及改善步行环境，从而推动更低碳、高效的交通系统发展。

AI 算力消耗大量能源并排放大量温室气体

AI 尤其是大模型的发展，极度依赖高性能计算资源。训练一个大模型（如 GPT、BERT、DALL·E）往往需要数百甚至上千 GPU 在数周内并行运算。其计算背后的电力消耗巨大，而这些电力在许多地区依然主要来自煤炭、天然气等高碳能源。此外，AI 的能耗并非仅限于训练阶段，部署后的模型推理、数据传输与存储也会持续消耗电力，并累积碳足迹。

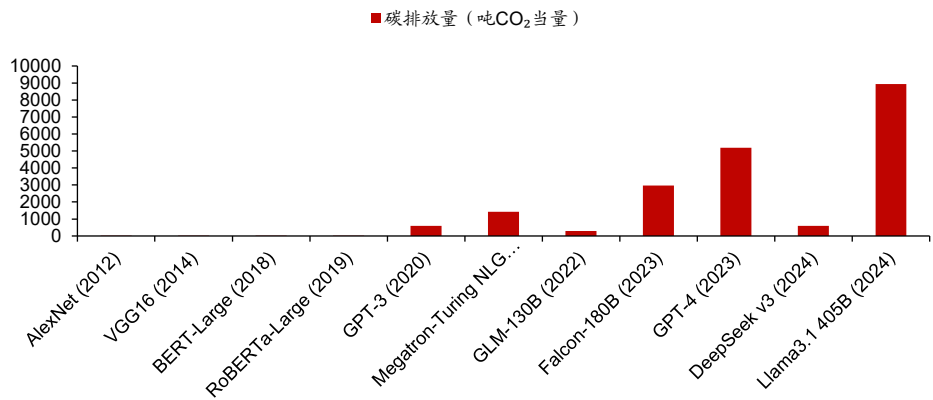
表8：AI 大模型发展的不同阶段及对应碳排放强度变化

| 应用阶段 | 碳排强度 | 具体情况 |
|---------|------|-------------------------|
| 模型训练阶段 | 高 | 模型规模越大，训练越久，排放越大。 |
| 推理阶段 | 较高 | 多数模型部署后在终端频繁调用也会产生持续能耗。 |
| 数据存储/传输 | 中等 | 特别是分布式存储和重复训练任务会叠加。 |

资料来源：《eco2AI: Carbon Emissions Tracking of Machine Learning Models as the First Step Towards Sustainable AI》，浙商证券研究所

随着模型训练数据量的增加，碳排放量也随模型迭代而增加。随着 AI 大模型的发展，碳排放量和模型规模都有所扩张。根据 Stanford HAI 的估计，2018 年面世的预训练模型 BERT-Large 参数量在数亿，训练阶段的温室气体排放量达到了 2.6 吨二氧化碳当量。2023 年发布的 GPT-4 训练参数达到万亿级别，碳排放量也上升到 5184 吨。而 AI 在运行过程中也会排碳。根据 IEA 的估算，在 2023 年数据中心和数据传输系统就分别占到了全球用电量的 1-1.5%，随着 AI 的广泛应用，对数据中心的需求也会增加，产生了与发电相关的碳足迹。

图31：人工智能模型在训练过程中产生的碳排放量估计



资料来源：Stanford HAI，浙商证券研究所

表9: AI大模型碳排放的主要来源

| 碳排放的主要来源 | |
|----------|---|
| 硬件能耗 | GPU 和 CPU 是主要耗能部件，尤其是训练大规模模型时。 |
| 数据中心效率 | PUE（电能使用效率）直接影响总碳排放（如 PUE=1.1 表示 10% 的额外能耗用于冷却等）。 |
| 区域电力结构 | 使用高比例可再生能源的地区（如法国，67.53kg CO ₂ /MWh）碳排放更低。依赖化石燃料的地区（如印度，625.57kg CO ₂ /MWh）碳排放显著更高。 |

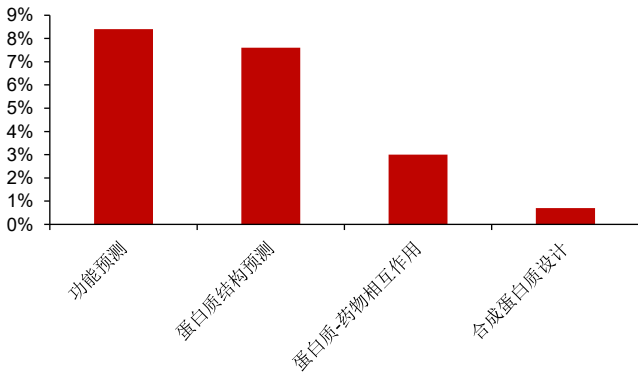
资料来源:《eco2AI: Carbon Emissions Tracking of Machine Learning Models as the First Step Towards Sustainable AI》, 浙商证券研究所

3.1.2 社会: 促进社会发展与产生伦理争议的两面性

AI 促进社会发展

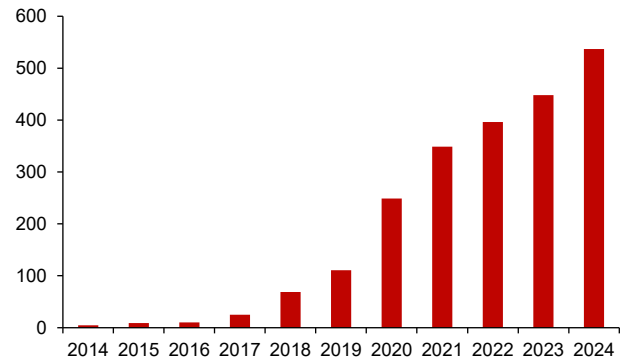
以 Deepseek 为代表的 AI 降本正在实现“智能平权”，并在科研、医疗等促进社会发展的方向发挥作用。随着 AI 大模型的加速迭代，一方面新一代大模型本身有更低的成本、更好的性能，给普通人使用 AI 更低的门槛，大众也可以通过 AI 掌握知识和技能、处理更多任务。另一方面大模型开始采用开源形式，使得更多的 C 端和 B 端客户可以直接参与其中。借助开源模型，用户经过模型微调即可用于自身的场景中。在研究领域，AI 在蛋白质研究方面帮助科研人员做出了巨大贡献，并驱动了诸多科研结果的生成。此外，AI 数据分析手段在临床医学中正从被动干预转向主动的决策支持。Stanford HAI 的研究表明 AI 在癌症诊断方面的准确度已经超过了医生，AI 与医生结合的诊断方式可以提高临床医疗的效果。

图32: 2024 年发表的领域内论文中由 AI 驱动研究的比例



资料来源: Stanford HAI, 浙商证券研究所

图33: 提及 AI 应用的临床实验数量 (单位: 个)



资料来源: Stanford HAI, 浙商证券研究所

AI 产生社会争议

AI 的部分问题由于算法本身的不足。算法偏见是指算法系统可能产生带有社会偏见的结果，从而加剧社会中的不平等的现象。具体有如下四种形式：输入数据偏见：指训练数据本身就存在偏见，而这一部分有偏见的“垃圾数据”又重新输出了有偏见的结果，在人工智能中被称为 GIGO (Garbage in, Garbage out)；模型设计偏见：指特征选择或目标函数设计不当导致模型的优化目标与社会伦理背离，而这些算法错误又往往在黑匣子中难以被发现；反馈循环偏见：有偏输出被重新输入学习将强化偏见；开发人员偏见：开发人员可能没有足够的社会科学知识，并且不太关注公平性问题。

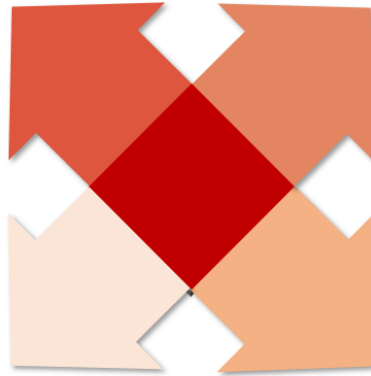
图34: AI 算法中可能存在偏见

输入数据偏见

AI 模型中使用的数据的过度和不足可能会导致少数群体偏见，导致有偏差的输出

反馈循环偏见

据收集和训练过程产生的偏见可以在反馈循环中得到加强



模型设计偏见

AI 模型使用的算法也可能存在问题。开发人员和用户很难找到AI算法中的所有错误和偏差，而这些错误和偏差隐藏于机器学习算法的“黑匣子”中

开发人员偏见

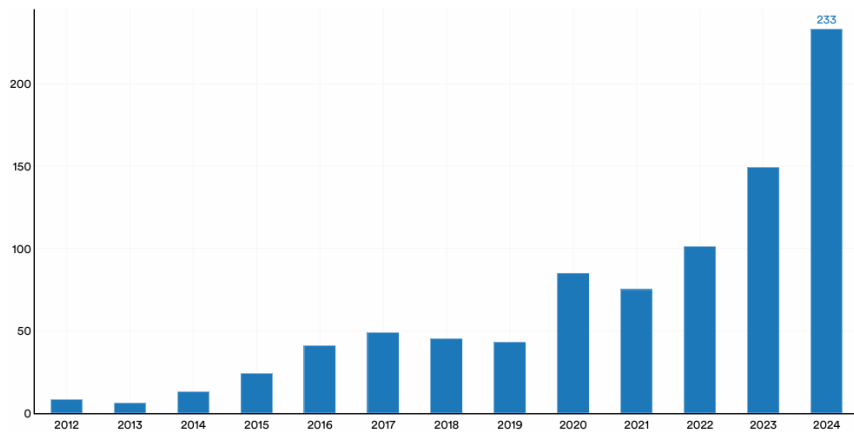
开发人员在标记数据时可能因为主观因素产生不道德行为，或者开发人员可能没有足够的社会科学知识，也会使结果有偏。

资料来源:《Societal impacts of artificial intelligence: Ethical, legal, and governance issues》，浙商证券研究所

有偏见的算法会影响社会公平，并在反馈中不断加深影响。当算法包含输入数据偏见时，它会自动学习并复制这些偏见。这种机制可能会导致弱势群体遭受不公正待遇——例如亚马逊招聘算法因训练数据中男性管理者占比过高，自动降低女性求职者的评分。算法偏见另一危险的特性在于其自我强化的反馈循环机制。当初始有偏的输出被重新输入系统迭代训练时，偏见会呈指数级放大。海外相关研究中的一个典型例证是：预测性警务工具。它的机制如下：算法基于历史逮捕数据将低收入社区标记为“高风险”→警方增加巡逻→发现更多轻微违规行为→新数据进一步“验证”算法预测→形成无法打破的监控闭环。这种循环使得特定人群被永久锁定在“高犯罪率”标签中，最终使得偏见进一步加深。

相较于传统人工智能，生成式人工智能还会产生其他问题。一是生成虚假信息：生成式AI可以生成看似可信，但事实不正确或具有误导性的内容。这可能导致错误信息或虚假信息的传播。二是隐私泄露风险：使用GenAI工具处理和分析个人数据时，AI模型可能会抓取这些内容，造成隐私泄露。三是知识产权问题，GenAI可以创建新内容，例如文本、图像或音乐，但新内容可能会无意中侵犯现有的知识产权。四是违反合约问题，如果有组织将来自第三方的信息提交给GenAI处理，由于多数情况下这类数据可能存在限制泄露的合同，将其用作GenAI的输入通常会造成信息未经授权的违约。在GenAI被广泛使用的2023-2024年，AI相关的伦理争议事件发生也更加频繁。

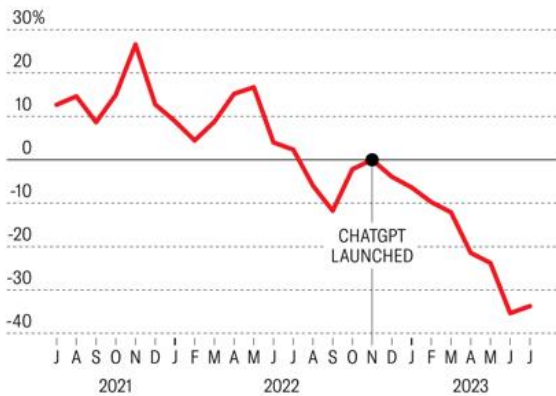
图35: 2012-2024 年有记录的 AI 伦理争议事件数量 (单位: 件)



资料来源: Stanford HAI, 浙商证券研究所

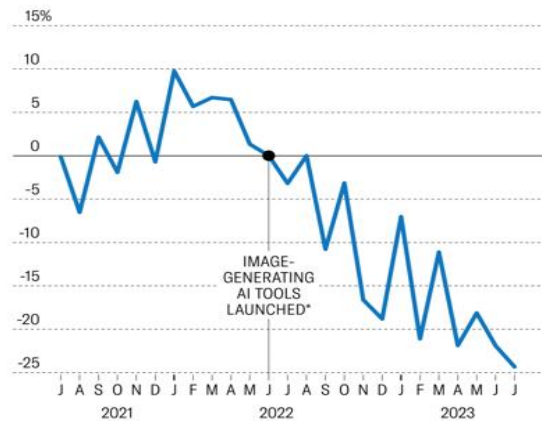
AI的发展可能引发结构性失业。近期首当其冲的是与生成式AI生成内容相关的行业。ChatGPT和图像生成AI这两种智能工具的推出显著影响了就业市场，特别是容易被AI替代的工作岗位。在ChatGPT上线后，容易受到GenAI替代的岗位每周发布量相比劳动密集型工作下降了21%，其中写作类岗位受影响最为严重，降幅达30.37%，软件应用/网页开发类岗位分别下降20.62%和10.42%。同时，在Midjourney、Stable Diffusion和DALL-E 2等图像生成AI工具推出后的一年内，市场对平面设计和3D建模技能的需求也下降了17.01%。

图36: ChatGPT发布后写作类岗位发布数量锐减



资料来源: Harvard Business Review, 浙商证券研究所

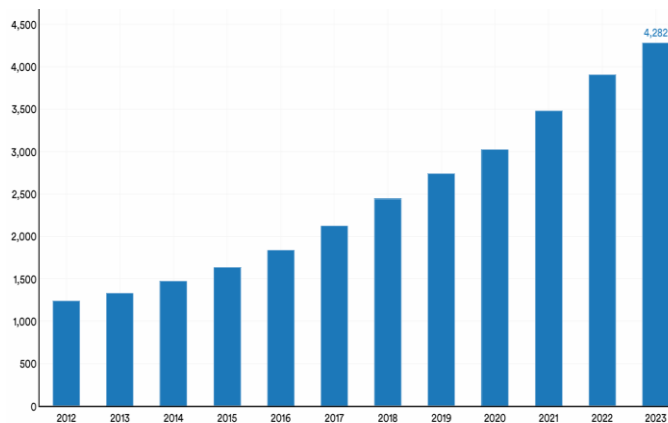
图37: 图像生成AI发布后制图相关岗位需求减少



资料来源: Harvard Business Review, 浙商证券研究所

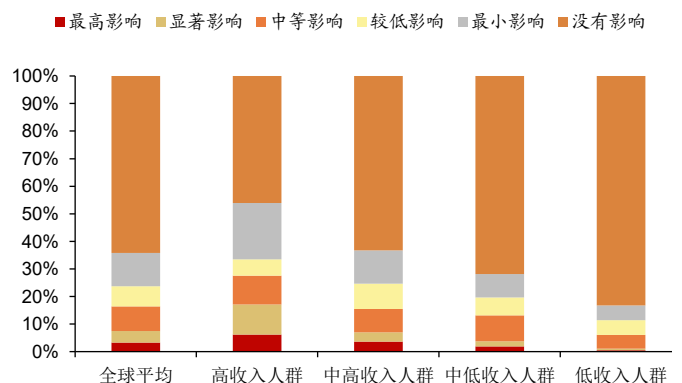
结构性失业带来更深层次的社会公平问题。一是冲击范围较广，并将在总体上挤出了全社会劳动收入份额。基于任务型人工智能软件编程的机器人已经可以替代诸多重复的机械劳动和服务型工作。2023年全球运行的工业机器人已经达到428万台，存量保持增长势头。而生成式AI可以进一步影响白领工作。国际劳工组织(ILO)与近期的研究则表明，全球35.8%的就业岗位可能受到生成式人工智能影响。需要人工输入的文员工作受到冲击最大，因此高收入国家就业岗位受到GenAI冲击比例更高，达到53.9%。此外，虽然人工智能的应用能够在一定程度上提升生产效率，但这将进一步提高资本收入，并推进资本积累，因此，布莱恩杰尔夫森(Brynjolfsson, 2014)的研究认为，人工智能能提高资本收入份额，并降低劳动收入份额。

图38: 2012-2023全球工业机器人存量年度变化(单位:千台)



资料来源: Stanford HAI, 浙商证券研究所

图39: 高收入人群岗位也将受到GenAI的冲击



资料来源: ILO, 浙商证券研究所

二是人工智能在就业上的影响的深度并不均衡，对低收入群体影响更大。高薪职业人群的工作任务也有较高的人工智能替代风险，但这些人群众体的工作往往含有多元化的任务组合，他们反而可以利用人工智能节省的时间聚焦于更具创造性和人际互动性的任务中。同时，AI提高的生产率也为这些业务带来的规模扩张，产生了新的岗位需求。相反，同时餐饮服务、低薪职业反而面临更大困境，AI（如自动化点餐系统）在这些行业中的使用主要目的是削减人力成本而非扩大经营规模，导致就业岗位净减少。

3.1.3 公司治理：机遇与挑战并存

AI 赋能 ESG 治理升级，以数据智能驱动风险识别与决策优化的治理新生态。人工智能通过数据驱动的精准分析、自动化流程和智能决策支持，显著提升了公司治理的效能。应用主要集中在四个关键领域：

- 财务风险预警：**人工智能采用深度学习、人工神经网络等技术，能够以超过 93% 的准确率预测企业财务困境，这一表现远优于传统统计方法。这些技术已成功应用于制造业、矿业和网络零售等多个行业，并在中国、欧洲和北美等不同地区展现出显著效果。
- 违规行为识别：**人工智能通过机器学习、人工神经网络等技术，有效发现财务报告造假和会计舞弊行为，在证券市场监管方面表现尤为突出。例如，人工神经网络与分类回归树模型的组合在识别财务舞弊方面展现出卓越性能。但由于不同国家和行业的违规行为特征存在差异，需要开发针对性的识别模型。
- 董事会效能提升：**人工智能主要通过机器学习算法和智能辅助系统发挥作用。它能够帮助优化董事提名流程，比如通过智能算法筛选更合适的董事候选人。人工智能在董事会工作中可以承担五种角色：日常事务辅助、深度分析支持、部分决策参与、正式董事成员以及完全自动化治理（目前仅停留在理论探讨阶段）。不过，由于技术限制和伦理考量，人工智能完全取代人类决策仍面临诸多挑战。
- 企业可持续发展：**人工智能利用机器学习、自然语言理解等技术预测环境社会治理评分，优化企业可持续发展规划。以智能对话系统为代表的工具还能提升社会责任信息披露的透明度和利益相关方参与度。

将 AI 应用到公司治理中会面临挑战。当 AI 系统处理大量敏感信息时，会出现数据隐私和安全隐患。由于深度学习算法属于一种“黑箱”，确保算法的透明度和问责制对于避免偏见和不道德的决定至关重要。同时，缺乏明确的监管框架使合规性和标准化工作变得复杂。习惯于传统治理方法的利益相关者对变革的抵制也会阻碍采用。在将 AI 无缝集成到公司治理中时，平衡创新与伦理问题和监管要求仍然困难。

表10： AI 应用到公司治理面临的问题

| 问题 | 原因 |
|------------|------------------------------------|
| 数据隐私和安全 | AI 依赖于大型数据集，引发了对数据隐私和敏感信息安全性的担忧。 |
| 道德和偏见问题 | 如果模型没有经过精心设计和测试，AI 可能会无意中在决策中嵌入偏见。 |
| AI 模型的不透明性 | 董事会可能难以理解复杂的 AI 算法，因此很难完全信任自动推荐。 |
| 法规适应性 | 企业需要确保 AI 解决方案满足相关法规的变化情况。 |
| 变更阻力 | 实施 AI 可能会面临来自关注工作角色和传统流程的员工和高管的阻力。 |

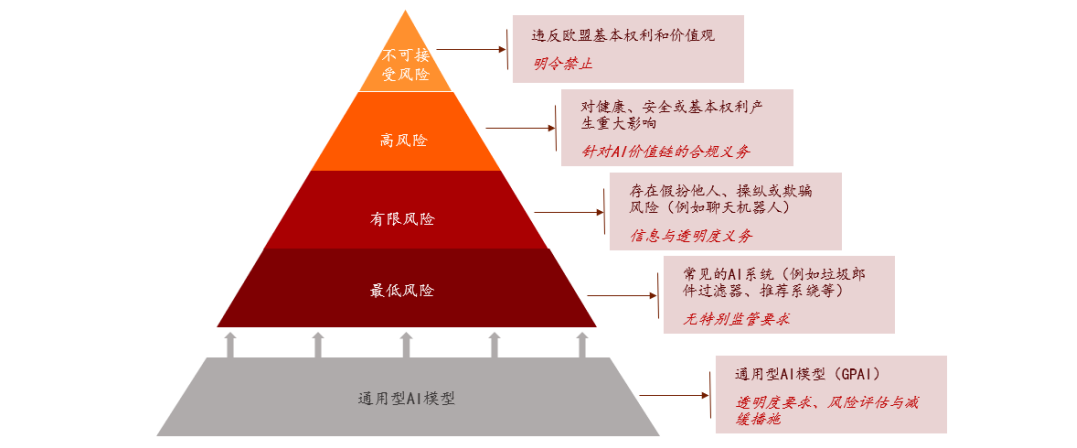
资料来源：Trust Cloud，浙商证券研究所

3.2 AI 面对强监管风险，AI 伦理与责任 AI 实践成为趋势

3.2.1 强隐性监管背景下，企业人工智能风险凸显

2024 年欧盟推出了全球世界首个全面人工智能法律框架。2024 年 3 月欧盟推出了人工智能法律框架《人工智能法案》(AI Act)，该法案适用主体包括了整个人工智能产业链，包括与欧盟市场有连接点的人工智能系统提供商、使用商、进口商、分销商和产品制造商。法案基于系统对用户和社会的潜在影响程度将 AI 分为 4 类分别监管，不同类别适用不同程度的监管要求。高风险的 AI 相关主题必须履行更严格的合规义务，而被判定为“不可接受风险”的系统则会被明令禁止。

图40： 欧盟《人工智能法案》分类方法及监管措施



资料来源：安永，搜狐网，浙商证券研究所

我国对人工智能的立法体系更为灵活，在部门规章层次提出更细节的规定。在围绕人工智能的法律体系建设方面，我国形成了以《网络安全法》、《数据安全法》和《个人信息保护法》为主体的“三驾马车”体系，分别从网络、数据、个人信息的角度对人工智能进行规范与治理。而面对人工智能领域的多种新技术，我国主要以部门规章做出更具体的规定。例如 2025 年 3 月由国家网信办等四部门联合发布的《人工智能生成合成内容标识办法》，要求相关企业通过显式标识（如文字提示、角标）和隐式标识（如元数据嵌入），确保用户能够清晰区分 AI 生成内容与真实内容，减少误导风险。

表11： 我国人工智能有关的主要部门规章

| 政策 | 发布时间 | 发布部门 | 主要内容 |
|-------------------|-------------|-----------|--|
| 《互联网信息服务算法推荐管理规定》 | 2021 年 12 月 | 国家网信办等四部门 | 对互联网信息服务的推荐算法在推荐内容、模型数据、页面生态方面进行了规定规定，明确了用户的选择权、知情权等有关权益。 |
| 《互联网信息服务深度合成管理规定》 | 2022 年 11 月 | 国家网信办等三部门 | 要求深度合成技术应用要保护用户信息、评估生成式算法机制机理，并对易混淆的内容添加明显标识。 |
| 《生成式人工智能服务管理暂行办法》 | 2023 年 7 月 | 国家网信办等四部门 | 禁止生成式人工智能生成违法内容、禁止算法设计中的歧视行为，提升内容透明度与准确性，保护数据安全与用户隐私。 |
| 《人工智能生成合成内容标识办法》 | 2025 年 3 月 | 国家网信办等四部门 | 明确了深度合成服务提供者对生成内容需添加不影响用户使用的标识（隐式标识）及在可能误导公众时对内容作出显著标识（显式标识）的义务。 |

资料来源：政府政务信息平台，浙商证券研究所

人工智能算法会造成广泛的社会影响，监管部门对负面影响也会持续督导，形成隐性监管格局。近年来，网信办持续部署开展“清朗”系列专项行动，集中时间、集中力量打击网上各类乱象问题，从严处置违规平台和账号。网信办在 2024 年 11 月起实行了“清朗·网络平台算法典型问题治理”，对推送算法同质化营造“信息茧房”、算法侵犯新就业形态劳动者权益、大数据“杀熟”、算法向上向善服务缺失等问题进行了督导整改，督促指导重点平台针对性优化信息推荐算法功能、调整信息推荐算法规则。

影响的外溢性使得 AI 面临的监管风险来源广泛。2024 年 5 月发布的《国务院 2025 年度立法工作计划》将“提请全国人大常委会审议名为‘人工智能法’的法律草案”的相关表述改为了“推进人工智能健康发展立法工作”。我们认为，当前有关部门对人工智能全面法律的立法态度已经较为温和，未来或将延续以部门规章为主的法律框架，以降低 AI 企业的合规成本，确保 AI 产业的活力。但同时，AI 对社会影响的负责性、广泛性使得监管会采取督导等相关措施灵活监管。因此我们认为企业在 AI 方面的风险暴露较为广泛，不仅考虑相关法律法规，同时应当充分考虑 AI 的伦理争议和社会影响。投资者在关注企业 AI 的过程中，也应当考虑类似 ESG 的高于法律需求的风险分析框架。

3.2.2 AI 伦理逐步成为道德共识，负责任 AI 成为行动规范

人工智能伦理 (AI ethics) 是发展人工智能的主要自律性质共识。受限于人工智能的快速发展以及国际间人工智能技术的激烈竞争，现阶段对人工智能的立法管制较难实施。相比之下，人工智能伦理则作为一种软性约束，同时覆盖范围更广、实施更灵活，成为目前全球指导人工智能治理的主要共识。

AI 伦理关注点在于确保 AI 系统的设计和使用符合道德规范和社会责任。如我们在前文所述，人工智能在社会的多个部门起到了提高效率的作用，但与此同时也产生了广泛的社会争议和伦理问题。这就需要人工智能伦理作为道德保障。2021 年 11 月，联合国教科文组织制定了有史以来第一份全球人工智能伦理标准——《人工智能伦理问题建议书》。该框架由 193 个会员国一致通过。该建议书的核心理念是：在推进透明和公平等根本原则的基础上，保护人们的权利和尊严，并始终牢记人类监督人工智能系统的重要性。

具体来说，AI 伦理从多方面对人工智能活动提出了要求。目前，多个国家的有关部门、产业链主要企业都阐述了自身关于 AI 伦理的有关界定。总体而言，AI 伦理的公约数包括以人文本、保障隐私、算法透明可监督、算法可靠可控。以中国科技部发布的《新一代人工智能伦理规范》为标准，该文件以促进公平、公正、和谐、安全，避免偏见、歧视、隐私和信息泄露等问题为起点，提出了增进人类福祉、促进公平公正、保护隐私安全、确保可控可信、强化责任担当、提升伦理素养等 6 项基本伦理要求。同时，提出人工智能管理、研发、供应、使用等特定活动的 18 项具体伦理要求。

表12: 科技部《新一代人工智能伦理规范》基本规范内容

| 规范类别 | 规范内容 |
|--------|---|
| 增进人类福祉 | 坚持以人为本，遵循人类共同价值观，尊重人权和人类根本利益诉求，遵守国家或地区伦理道德。 坚持公共利益优先，促进人机和谐友好，改善民生，增强获得感幸福感，推动经济、社会及生态可持续发展，共建人类命运共同体。 |
| 促进公平公正 | 坚持普惠性和包容性，切实保护各相关主体合法权益，推动全社会公平共享人工智能带来的益处，促进社会公平正义和机会均等。 在提供人工智能产品和服务时，应充分尊重和帮助弱势群体、特殊群体，并根据需要提供相应替代方案。 |
| 保护隐私安全 | 充分尊重个人信息知情、同意等权利，依照合法、正当、必要和诚信原则处理个人信息，保障个人隐私与数据安全，不得损害个人合法数据权益，不得以窃取、篡改、泄露等方式非法收集利用个人信息，不得侵害个人隐私权。 |

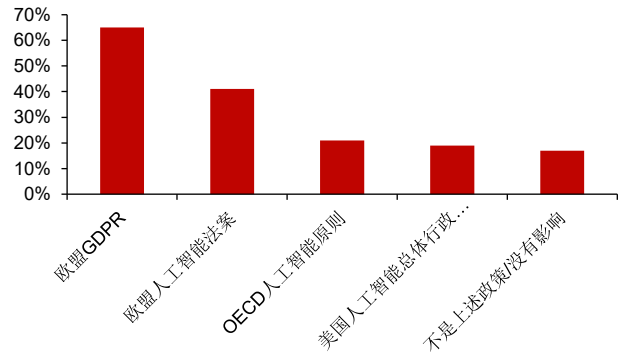
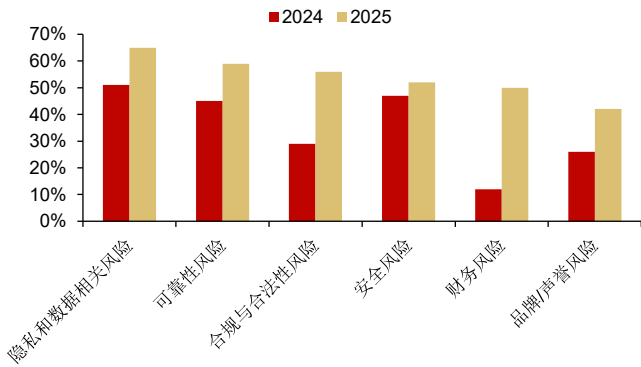
| | |
|--------|--|
| 确保可控可信 | 保障人类拥有充分自主决策权，有权选择是否接受人工智能提供的服务，有权随时退出与人工智能的交互，有权随时中止人工智能系统的运行，确保人工智能始终处于人类控制之下。 |
| 强化责任担当 | 坚持人类是最终责任主体，明确利益相关者的责任，全面增强责任意识，在人工智能全生命周期各环节自省自律，建立人工智能问责机制，不回避责任审查，不逃避应负责任 |
| 提升伦理素养 | 积极学习和普及人工智能伦理知识，客观认识伦理问题，不低估不夸大伦理风险。主动开展或参与人工智能伦理问题讨论，深入推动人工智能伦理治理实践，提升应对能力。 |

资料来源：科技部门户网站，浙商证券研究所

头部企业提出符合 AI 伦理的内控要求，并积极进行实践负责任 AI (RAI) 实践。无独有偶，与负责任投资类似，重视社会影响、以人为本、关注长期发展的 AI 实践也被成为“负责任 AI”。负责任 AI 也可以看成是 AI 伦理的具象化和规则化。随着 AI 运用更加广泛，企业对负责任 AI 相关风险的也更加重视。根据 Accenture 和斯坦福大学的联合调查，企业与 AI 相关的财务风险、合规风险、隐私和数据风险等议题的关注度均有所增加。为了防止 AI 相关风险产生的负面影响，许多企业并非由于法律要求，而是更多自发接受负责任 AI 的框架作为软约束。谷歌 TensorFlow 建立了定义问题-准备数据-训练模型-评估模型-部署和监控全流程的 RAI 工具；阿里云也建立了 RAI 下的内容安全监测、公平性分析、错误分析流程。

图41： 2024-2025 企业对负责任 AI 相关风险的关注度

图42： 企业在负责任 AI 实践中受到相关法律影响的调查



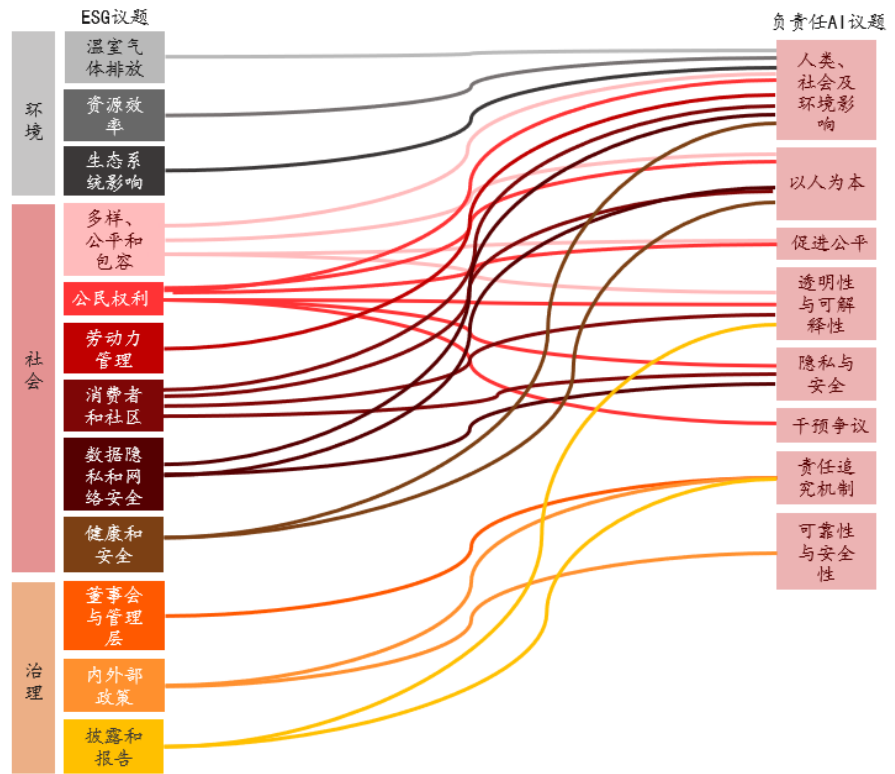
资料来源：Stanford HAI，浙商证券研究所

资料来源：Stanford HAI，浙商证券研究所

3.3 ESG+RAI 框架整合助力防范投资中的 AI 风险暴露

AI 的伦理问题实质上会外溢影响企业 ESG 表现，并与 ESG 议题相重合。虽然目前大部分 ESG 准则并不直接包含 AI 的相关内容，但随着 AI 对社会的影响范围扩大、影响程度加深，AI 的相关影响正外溢到环境（例如 AI 算力高耗能增加碳排放）、社会（如 AI 冲击就业结构造成失业）、公司治理（如 AI 数据治理问题）的多个维度，影响着企业的 ESG 表现。负责任 AI 相关议题和 ESG 议题出现范围上的交叉重合。参照 CRISO 的 ESG 议题与负责任 AI 议题对比，负责任 AI 中的人类、环境和社会影响与 ESG 框架下的环境、社会议题交叉重合。以人为本、促进公平、隐私与安全、争议可干预性与社会议题有关，责任追究机制、可靠性与安全性则与治理议题交叉，透明性与可解释性则横跨社会议题和公司治理议题。

图43: ESG议题与负责任AI议题在范畴上相互交织



资料来源: CSIRO, 浙商证券研究所

投资端, ESG+RAI 框架或可助力识别 AI 相关风险和机遇。在投资中,随着 AI 的影响日益深厚,与 AI 有关的技术风险、执行风险、估值风险、监管风险和机遇也需要被纳入投资考量中。ESG 体系中目前尚没有与企业 AI 治理直接相关的议题设定,但仍有不少投资研究试图弥合 ESG 与负责任 AI 两大体系。根据我们对资管机构相关实践的梳理,我们认为主要有关关注 AI 相关公司治理与 ESG 框架的全面整合两种思路来识别和管理 ESG 风险与机遇。

3.3.1 挪威银行资管 NBIM: 关注公司治理中对 AI 的治理措施和机制建设

挪威银行资管 (NBIM) 认为 AI 产业应当同样重视负责的商业行为。NBIM 是规模挪威政府全球养老基金的管理公司,管理规模达到 1.5 万亿美元。即随着人工智能越来越多地融入各行业和市场的公司运营,其使用必须与负责的商业行为保持一致,其中包括重视商业道德、以人为本、重视股东价值等等。NBIM 认为,人工智能有巨大的机遇,但也包含着对被投资公司和利益相关者的巨大风险。除了对公司造成监管、运营和声誉风险外,人工智能系统的开发和使用还会产生广泛的社会影响,增加错误信息传播、欺诈和操纵的风险。

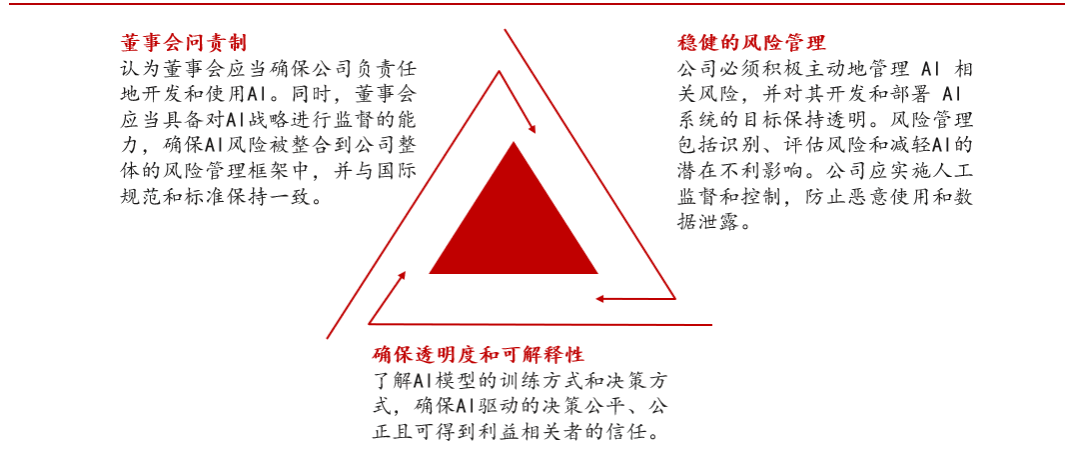
NBIM 通过治理过程中的三大要素衡量公司负责任 AI 治理能力。

一是董事会问责制。董事会对公司负责任地开发和 AI 负主要责任。NBIM 认为,董事会需要在公司治理过程中平衡 AI 产生的社会风险与机遇。这需要公司董事会具备相关的专业知识。同时,董事会应当监督公司在相应的法规和指南下进行业务。包括地区负责任 AI 法规、国际组织有关指南等等 (如经合组织 AI 原则和联合国教科文组织的 AI 伦理建议书等)。同时董事会应当推动企业应当构建符合自身的 AI 治理结构。

二是确保透明度和可解释性。NBIM 认为公司应当能够解释开发或使用的 AI 系统的设计、训练和测试流程，并说明 AI 如何与人类价值观对齐。公司应使利益相关者能够评估 AI 系统的潜在影响，并了解系统的准确性、效率和可靠性。公司还应向相关且受信任的第三方（例如审计机构或监管机构）提供信息，以便他们验证 AI 系统并评估其风险。

三是稳健的 AI 风险管理措施。公司必须积极主动地管理 AI 相关风险，风险管理流程应稳健且与公司的风险敞口相称，并寻求识别、评估和减轻 AI 对业务、人员和社会的风险。除了确保业务弹性外，AI 风险管理流程还应关注广泛的社会影响，包括隐私保护、数据安全和避免歧视等等，并确保有效的人工监督和控制。在风险措施方面，应包括对系统限制和系统故障潜在后果的评估内容。开发或部署可能对内部人员、社会或业务成果构成严重风险的 AI 系统应受到额外控制。风险管理流程还应有适当的保护措施来管理错误信息、欺骗等风险。在长期管理上，AI 系统、内部指南和风险管理流程应该经过独立验证，并随着时间的推移定期审计。

图44: NBIM 公司负责任 AI 问责三大要素



资料来源: NBIM, WEF, 浙商证券研究所

3.3.2 UBS 资管: 多层级 AI 公司治理分析模型

UBSAM 建立了一个公司分层次的 AI 治理框架。UBS 资管关注董事会和管理团队如何优先治理 RAI，同时了在 AI 治理过程中内部问责和跨业务职能沟通的重要性。因此，UBS 建立了一个公司内部的三层框架，包含集团、业务线、应用三大层次，每个层次都包含三个类型的风险。该框架将概念性 AI 治理模型（例如数据、模型和应用程序层）转换为公司治理模型，并对标到具体的公司层次和公司部门之中。

表13: UBS 资管对公司 AI 治理进行评估的三层框架

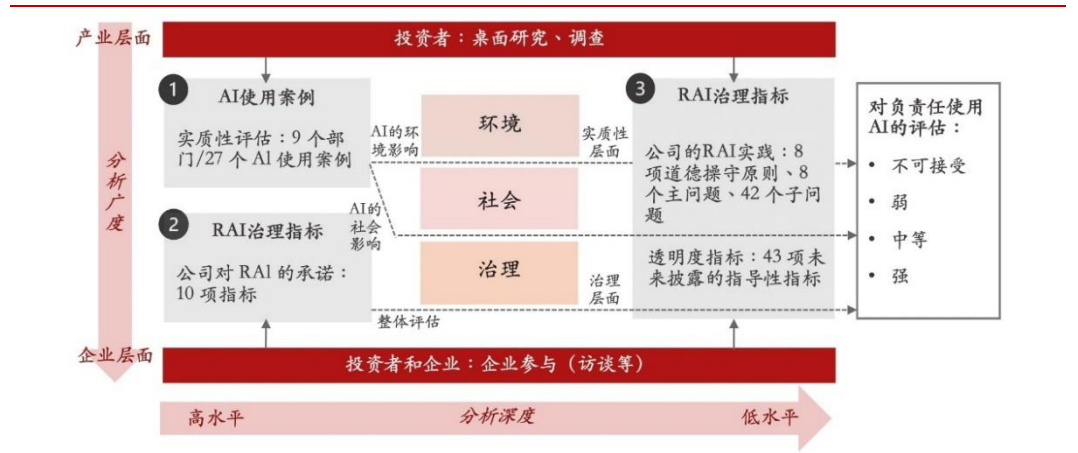
| 层级 | 确定性风险管理 | 预期风险管理 | 意外风险管理 |
|---|---|------------------------------------|---|
| 集团层面: 为整个组织的 AI 应用奠定基础。包括确立 AI 规则、建立治理架构、明确政策、监督标准执行 | 需要健全的治理框架和原则、全面的 AI 政策和标准以及集中的技术能力。 | 为 AI 风险做准备, 建立战略风险管理的流程、学习和改进周期 | 形成敏捷、诚信和有韧性的 AI 风险管理文化, 并维持外部合作伙伴关系, 保证面对风险时的快速响应和恢复。 |
| 业务线层面: 在部门 (如人力资源、财务) 或业务单位中应用 AI 治理。使 AI 计划与更广泛的组织标准一致, 通过评估管理风险并为应急情况做好准备。 | 确保 AI 计划符合组织层面的治理、合规和风险管理标准, 并在特定业务单位内对齐 AI 计划。 | 对关键 AI 系统进行风险评估和影响分析, 制定应急计划 | 制定业务连续性计划、紧急程序和危机管理计划 |
| 应用层面: 确保每个 AI 系统遵循既定标准, 监控新风险并维护事件响应机制。 | 遵守 AI 开发和部署标准, 进行彻底的测试和验证 | 监控 AI 系统以发现新风险和性能问题, 建立实时风险检测和缓解机制 | 制定快速事件响应协议, 针对不可预见问题的 AI 系统能够快速回滚或关闭程序。 |

资料来源: WEF, UBS AM, 浙商证券研究所

3.3.3 CRISO 的 ESG-AI 全面整合框架

澳大利亚联邦科学与工业研究组织 (CSIRO) 的数据部门与 Alpha 投资管理公司合作开发出了一个更为全面的 ESG-AI 分析框架, 该框架以 ESG 分析思路为基础, 构建了 AI 用例分析-RAI 指标分析-RAI 深度分析的三步 AI 风险机遇风险体系。

图45: ESG-AI 分析步骤



资料来源: CRISO, CSDN, 浙商证券研究所

该框架的第一步是 AI 用例分析, 相当于分析 AI 对哪些企业属于实质性议题。该框架对 9 个重点一级行业设置了共 27 个关键用例, 并标注了这些用例的相关风险。其中, 监管风险主要依据欧盟《人工智能法案》的 AI 风险分类标注; 环境和社会影响用 ESG 体系下的二级议题分别评估, 而影响范围则区分行业内影响和系统性影响两大类。AI 用例分析有助于投资者快速筛选出业务有 AI 风险暴露的公司, 相当于以用例模式判断 AI 对哪些公司有实质性影响。

表14: ESG-AI 框架的 AI 用例及对应用例的风险

| 行业 | AI 用例 | 监管风险 | E&S 影响水平 | 影响范围 | 重要性 |
|------|-------|-------|----------|------|-----|
| 信息技术 | 产品开发 | 无确定监管 | 高 | 行业 | 中等 |
| | 自动化 | 无确定监管 | 中等 | 行业 | 中等 |
| | 风险管理 | 中等风险 | 高 | 系统性 | 高 |

| | | | | | |
|------|--------------------------|------|----|-----|----|
| 医疗保健 | 健康研究 / 测试 | 高风险 | 中等 | 行业 | 中等 |
| | 临床护理 | 高风险 | 中等 | 系统性 | 高 |
| | 产品开发 | 中等风险 | 中等 | 系统性 | 高 |
| 金融 | 情景测试 / 预测建模和保费定价 (保险公司) | 中等风险 | 中等 | 行业 | 中等 |
| | 欺诈检测 | 中等风险 | 中等 | 行业 | 中等 |
| | 信用评分和预申请筛选 | 高风险 | 中等 | 系统性 | 高 |
| 消费品 | 区块链、供应链和可追溯性 | 中等风险 | 中等 | 行业 | 中等 |
| | 个性化客户推荐 (例如建议) | 中等风险 | 中等 | 行业 | 中等 |
| | 店内监控 | 高风险 | 中等 | 行业 | 中等 |
| 工业 | 流程改进和自动化 | 中等风险 | 中等 | 行业 | 中等 |
| | 资产监控 / 预测性维护 | 中等风险 | 中等 | 行业 | 中等 |
| | 物流和供应链管理 | 中等风险 | 中等 | 行业 | 中等 |
| 能源 | 建筑和住宅的能源效率 | 中等风险 | 中等 | 行业 | 中等 |
| | 预测性基础设施维护 | 高风险 | 中等 | 系统性 | 高 |
| | 电网管理和能源优化 | 高风险 | 中等 | 系统性 | 高 |
| 房地产 | 房地产评估 | 中等风险 | 中等 | 系统性 | 高 |
| | 设施管理 | 中等风险 | 中等 | 行业 | 中等 |
| | 客户服务 | 中等风险 | 中等 | 行业 | 中等 |
| 材料 | 材料发现与创新 | 中等风险 | 高 | 行业 | 中等 |
| | 探索和资源识别 | 中等风险 | 高 | 行业 | 中等 |
| | 健康与安全 | 高风险 | 中等 | 系统性 | 高 |
| 电信 | 资产监控 / 预测性维护和网络管理 / 系统优化 | 中等风险 | 中等 | 行业 | 中等 |
| | 客户服务 | 中等风险 | 中等 | 行业 | 中等 |
| | 欺诈检测 | 中等风险 | 中等 | 行业 | 中等 |

资料来源: CRISO, 浙商证券研究所

框架的第二步是 RAI 治理指标分析, 融合了 ESG 评级的思路。这一步主要通过指标分析公司在 RAI 方面的治理能力。指标包括董事会监督、RAI 政策、RAI 承诺、RAI 披露措施四大方面的共 10 个指标, 并依据打分卡模式在 0-10 分范围内进行打分。鉴于目前于 RAI 相关的披露较少, 该框架也建议在披露信息的基础上, 与被投资公司建立交流渠道。

表15: RAI 治理指标分析问题

| 类别 | 指标 | 描述 |
|--------|----------------|---|
| 董事会监督 | 1. 董事会问责制 | RAI 被明确提及为董事会或下设委员会 (例如风险委员会或 ESG 委员会) 的责任。董事会至少每年接受一次结构化的 RAI 报告, 并根据情况提高频率。 |
| | 2. 董事会能力 | 至少有一名具备强技术经验的董事。 |
| RAI 承诺 | 3. 公开的 RAI 措施 | 政策应与相关的行业标准一致。RAI 政策应考虑伦理、公司价值观、AI 测试和透明度。 |
| | 4. 敏感用例 | 敏感的高风险用例 (如面部识别) 在 RAI 政策中进行风险管控。例如敏感用例需要额外的监督和批准等。 |
| RAI 实践 | 5. 人工智能道德目标 | RAI 政策或承诺有明确的目标 (例如受培训的员工百分比、人工智能伦理事件减少量) 作为支持。 |
| | 6. 专门的 RAI 负责人 | RAI 监督可以是专门的, 也可以是另一个角色或职能的一部分。 |
| | 7. 员工意识 | 有专门的计划来提高员工对人工智能、AI 伦理和 ESG 的意识。 |
| | 8. 系统整合 | 将 RAI 政策整合到现有的业务流程中, 包括风险管理、产品开发、采购和 ESG。 |

| | | |
|-----------|--------------------------|--------------------------------------|
| 9. 人工智能事件 | 与 RAI 相关的问题和事件被跟踪并在内部报告。 | |
| RAI 披露 | 10. RAI 指标披露 | 与 RAI 措施相关的指标（如人工智能的使用）被识别并向利益相关者报告。 |

资料来源：CRISO，浙商证券研究所

RAI 深度研究是该框架的第三步。在前两步的基础上，第三步投资者可以根据对 AI 密切关注的问题与公司进行更深度的沟通。CRISO 借鉴了欧盟《人工智能法案》、NIST AI 风险管理框架、ISOAI 标准等准则作为参考，提供了 RAI 原则下 8 个议题的共 42 个问题，每个问题下有 27 个指标，供投资者进一步加深对公司 RAI 情况的了解。我们认为 RAISO 的方法论借鉴了 ESG 分析的主要思路，采用识别-深入研究的模式，在可行较高的情况下有助于投资者关注投资标的的 AI 风险和机遇，可以作为分析 AI 相关 ESG 风险的参考框架。

4 风险提示

- 1) 经济修复不及预期：宏观经济发展不及预期，导致企业盈利受挫从而削减可持续发展投入，消费者也更关注价格而忽视绿色投资价值，不利于 ESG 投资发展；
- 2) AI 模型不稳定风险：AI 模型在不同的数据集上表现不同，实际应用中出现的结果可能与本文结果有偏差；
- 3) 市场情绪与偏好波动风险：市场情绪和投资者偏好多变，ESG 投资容易因市场转向而在短期失去关注，面临资金撤离、发展桎梏等风险。

股票投资评级说明

以报告日后的6个月内，证券相对于沪深300指数的涨跌幅为标准，定义如下：

1. 买入：相对于沪深300指数表现+20%以上；
2. 增持：相对于沪深300指数表现+10%~+20%；
3. 中性：相对于沪深300指数表现-10%~+10%之间波动；
4. 减持：相对于沪深300指数表现-10%以下。

行业的投资评级：

以报告日后的6个月内，行业指数相对于沪深300指数的涨跌幅为标准，定义如下：

1. 看好：行业指数相对于沪深300指数表现+10%以上；
2. 中性：行业指数相对于沪深300指数表现-10%~+10%以上；
3. 看淡：行业指数相对于沪深300指数表现-10%以下。

我们在此提醒您，不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系，表示投资的相对比重。

建议：投资者买入或者卖出证券的决定取决于个人的实际情况，比如当前的持仓结构以及其他需要考虑的因素。投资者不应仅仅依靠投资评级来推断结论。

法律声明及风险提示

本报告由浙商证券股份有限公司（已具备中国证监会批复的证券投资咨询业务资格，经营许可证编号为：Z39833000）制作。本报告中的信息均来源于我们认为可靠的已公开资料，但浙商证券股份有限公司及其关联机构（以下统称“本公司”）对这些信息的真实性、准确性及完整性不作任何保证，也不保证所包含的信息和建议不发生任何变更。本公司没有将变更的信息和建议向报告所有接收者进行更新的义务。

本报告仅供本公司的客户作参考之用。本公司不会因接收人收到本报告而视其为本公司的当然客户。

本报告仅反映报告作者的出具日的观点和判断，在任何情况下，本报告中的信息或所表述的意见均不构成对任何人的投资建议，投资者应当对本报告中的信息和意见进行独立评估，并应同时考量各自的投资目的、财务状况和特定需求。对依据或者使用本报告所造成的一切后果，本公司及/或其关联人员均不承担任何法律责任。

本公司的交易人员以及其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。本公司没有将此意见及建议向报告所有接收者进行更新的义务。本公司的资产管理公司、自营部门以及其他投资业务部门可能独立做出与本报告中的意见或建议不一致的投资决策。

本报告版权均归本公司所有，未经本公司事先书面授权，任何机构或个人不得以任何形式复制、发布、传播本报告的全部或部分内容。经授权刊载、转发本报告或者摘要的，应当注明本报告发布人和发布日期，并提示使用本报告的风险。未经授权或未按要求刊载、转发本报告的，应当承担相应的法律责任。本公司将保留向其追究法律责任的权利。

浙商证券研究所

上海总部地址：杨高南路729号陆家嘴世纪金融广场1号楼25层

北京地址：北京市东城区朝阳门北大街8号富华大厦E座4层

深圳地址：广东省深圳市福田区广电金融中心33层

上海总部邮政编码：200127

上海总部电话：(8621) 80108518

上海总部传真：(8621) 80106010

浙商证券研究所：<https://www.stocke.com.cn>