华夏基金 CHINA ASSET MANAGEMENT

紫顶

AI浪潮下的ESG:
风险、机遇、负责任AI

目录

CONTENTS

执行摘要

人工智能正以前所未有的速度重塑全球经济与社会形态。从生成式大模型的爆发到智能体（Agent）走入产业一线，AI 既是新一轮生产力革命的引擎，也给环境、社会与公司治理（ESG）带来了一系列全新而真实的挑战。

本报告以“负责任 AI”为主线，沿着环境（E）、社会（S）、治理（G）三个维度，系统梳理了 AI 浪潮给 ESG 带来的机遇与风险，并着力发现企业、监管机构与投资者层面已经涌现的有效实践。以下是我们的核心判断：

环境（E）：绿色算力是确定的长期大趋势。基于我们的观察与访谈发现，头部企业已普遍开展算力能耗优化，绿色算力正从“锦上添花”变为“准入门槛”。虽处早期阶段，但趋势方向确定，行业具备长期结构性机会。

社会（S）：大规模岗位重构而非单纯岗位消失。部分岗位确实在消失，尤其是偏技术性的初级岗位，但面向客户、基于实际工作流沉淀的经验诀窍（know-how）与人际沟通能力（people skills）相关的岗位难以替代，AI 在这些领域更多是增强而非替代。同时我们也发现，这场社会转型的真实后果并非仅仅由 AI 的技术边界决定，企业的战略态度也发挥着重要作用。若企业将 AI 定位为“员工能力放大器”，则能有效推动全员赋能与岗位重构，实现劳动力价值的长期共赢。

治理（G）：企业对 AI“重应用、轻治理”，监管规则正在成形。通过对科创 50 成分股 2025 年度 ESG 报告的文本挖掘显示，AI 相关词覆盖率已达 92%，但真正指向 AI 专项治理的表达仅覆盖约两成公司，反映出“高频用 AI、低频谈治理”的落差。在监管层面，中、欧、美三大经济体走出了三条特征鲜明的监管路径——中国发展与安全并重，并通过敏捷立法快速迭代；欧盟以风险分级为核心构建较为严密的监管体系；美国则在“联邦松绑”与“州级收紧”之间拉锯。

投资者行动：“负责任 AI”正从少数先行投资者的倡议，演变为可量化、可比较的尽责管理议程。美股市场 AI 相关股东提案持续涌现，投资者关注议题也逐渐转向具体的治理架构缺口。

我们认为，AI 对 ESG 的真实影响取决于技术被如何设计、部署与治理。在不确定中锚定确定，正是本报告希望与监管机构、企业与投资者共同坚守的方向。

引 言

当 AI 浪潮遇见 ESG

2022 年底以来，以 ChatGPT 为代表的生成式人工智能掀起了新一轮技术浪潮。短短几年，AI 从实验室走向千行百业，正以基础设施般的姿态嵌入经济运行的底层。对资本市场而言，这既意味着巨大的投资机遇，也意味着一套需要重新校准的 ESG 分析坐标系。

理解 AI 与 ESG 的交汇，在当下的国际局势之中显得更加复杂。一方面，地缘政治的紧张态势使各国对“算力主权”与“技术安全”空前重视，AI 被提升到国家战略竞争的高度；另一方面，气候变化的紧迫性并未因技术热潮而削减，全球仍处在能源安全与绿色转型的双重压力之下。AI 恰好同时站在这两条主线的交点上：它既是耗电的大户，可能加剧短期的能源与排放压力；又是优化能源系统、加速清洁技术研发的有力工具。如何在“发展”与“减碳”、“创新”与“安全”之间取得平衡，正是理解 AI 与 ESG 交汇点的关键。

简单地把 AI 标签化为 ESG 的“风险源”或“解决方案”可能都失之片面。AI 对 ESG 的真实影响，取决于技术被如何设计、部署与治理——这正是“负责任 AI”概念的要义所在：在追求效率与创新的同时，兼顾安全、公平、透明、隐私与可持续，使技术发展始终服务于人的福祉，这也与 ESG“创造长期价值的同时管理好外部性”的内核高度契合。

基于这一立场，本报告探讨了 AI 带来的能耗攀升、就业冲击、数据滥用与算法歧视等真实风险，同时我们也积极发现“出路”——那些已经在企业、监管机构与投资者层面涌现的、行之有效的应对实践。

本报告分为四章。第一、二、三章分别沿环境（E）、社会（S）、治理（G）三个维度，剖析 AI 带来的风险与机遇，并辅以多家上市公司的一手案例；第四章回到投资者本位，探讨机构投资者如何以尽责管理推动被投企业的负责任 AI 实践。

CHAPTER 01

AI 与环境（E）

算力的能耗悖论与绿色解法

在 ESG 的三个维度中，AI 对环境的影响最具“双面性”，也最容易被简化为单向叙事——AI 既是不容忽视的能源与资源消耗者，也是强有力的绿色转型工具——关键在于如何管理其“净影响”。

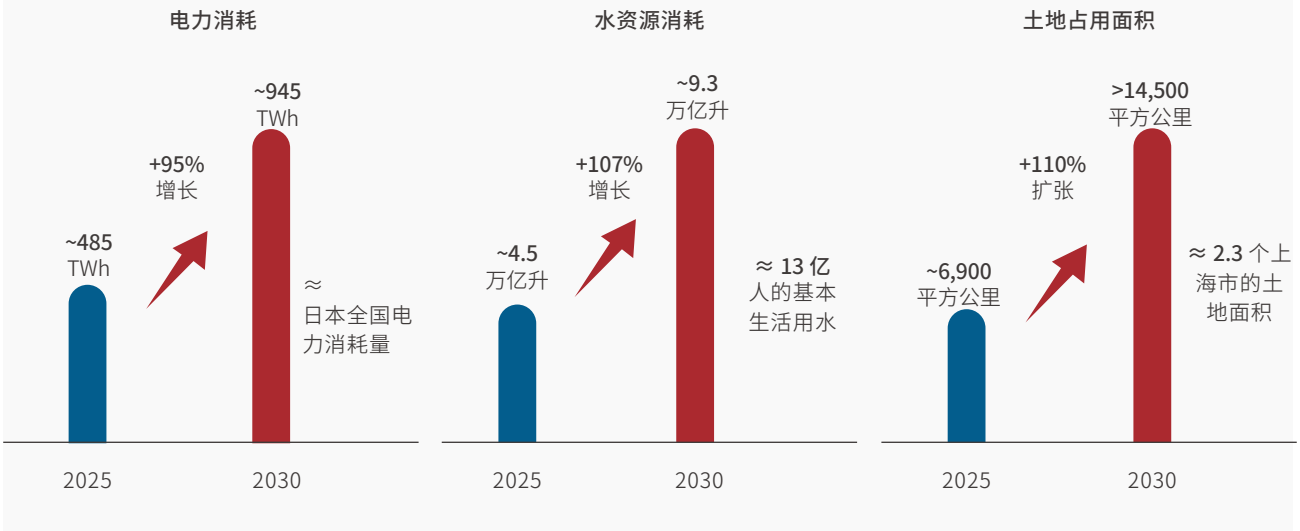
基于我们的访谈与观察，本章最明确的结论是：**绿色算力是一个方向明确、不可逆转的长期大趋势。**

1 成本重估：被重新定价的能源、水与物质足迹

AI 带来技术进步的同时，增加了对资源的消耗。在成本方面，AI 对环境的挑战，首先体现为电力需求的结构性攀升。国际能源署（IEA）2025 年发布的《Energy and AI》报告预测，全球数据中心用电量将从 2025 年的约 485 太瓦时[1] 接近翻倍增长至 2030 年的约 945 太瓦时——大致相当于日本全国用电量。根据 IEA 2026 年初的跟踪数据显示，2025 年全球数据中心用电量实际增长约 17%，AI 专用数据中心用电更是激增约 50%；支撑这轮扩张的，是大型科技公司当年合计超过 4,000 亿美元的资本开支，且这一资本开支预计 2026 年还将再增约 75%。

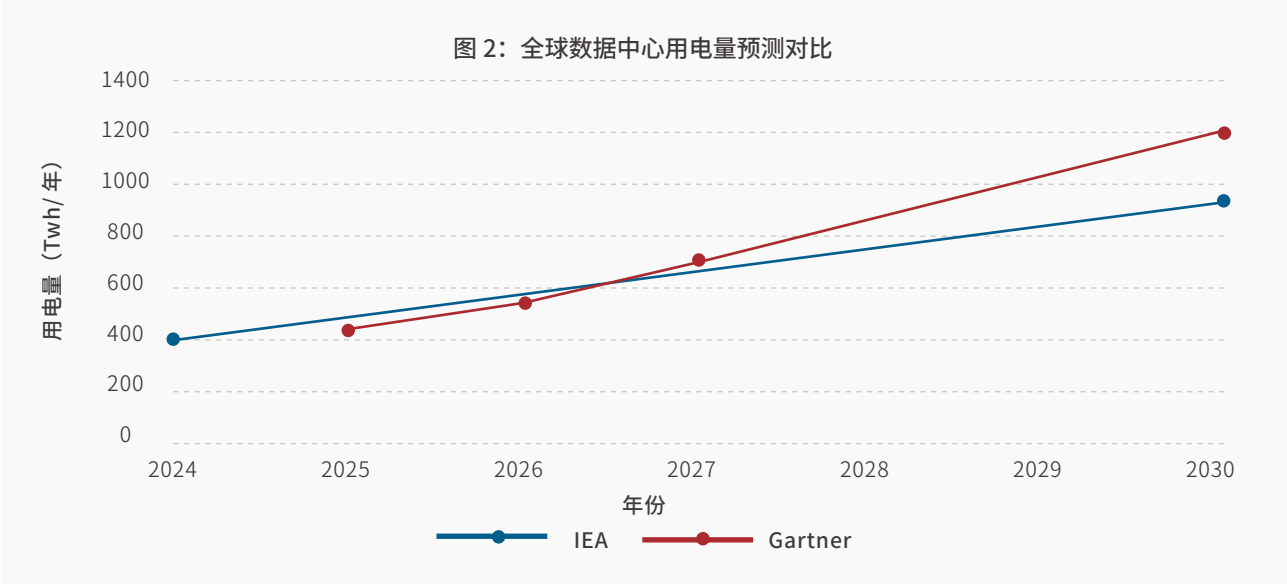
能源之外，水与物质足迹同样不可忽视。联合国 2026 年的研究估算，若按当前轨迹扩张，与数据中心用电相关的年度水足迹可能从 2025 年约 4.5 万亿升增至 2030 年约 9.3 万亿升，接近 13 亿人的基本生活用水；年二氧化碳排放升至约 3.99 亿吨，大致相当于英国 2025 年全国温室气体排放总量的 1.1 倍；占地面积则从 2025 年的约 6,900 平方公里扩张至逾 14,500 平方公里，相当于 2.3 个上海市的土地面积。

图 1：全球数据中心资源消耗预测



资料来源：UNU-INWEH (2026)

[1] 国际能源署（IEA）2025 年发布的《Energy and AI》中提及的 2024 年全球数据中心用电量为 415 太瓦时，根据 IEA 最新发布的数据，此处更新为 2025 年全球数据中心用电量 485 太瓦时，详见 https://www.iea.org/reports/key-questions-on-energy-and-ai/executive-summary



资料来源：国际能源署（IEA），Gartner

AI 对环境的影响具有鲜明的双重性：它既是快速增长的资源消耗者，也正成为推动能源系统优化和绿色技术突破的重要力量。在电网侧，AI 可用于预测可再生能源出力状况、优化调度与需求响应，提升电网对高比例可再生能源的消纳能力；IEA 指出，AI 在能源系统中的广泛应用，有望带来远超其自身耗能的减排收益。在产业侧，AI 加速了新材料、新型电池与光伏技术的研发，压缩了清洁能源的创新周期；在气候适应侧，AI 驱动的高精度气候建模，正不断提升社会应对极端天气的韧性。

总的来说，AI 的“杰文斯悖论”（Jevons Paradox）依然值得警惕：技术进步提高了单位算力的能效，但成本下降往往刺激总需求更快增长，反而推高能源的绝对消耗。这意味着，仅靠芯片与模型的效率优化并不足以自行化解 AI 的能耗难题，还须辅以电力结构的绿色化与使用侧的总量管理。在特定地区，规划不周的数据中心扩张可能与当地既有的资源压力正面相撞——“负责任选址”由此成为新议题，也是投资者评估算力资产时需要纳入的区域性风险变量。

2 破局之路：算力扩张的现实挑战与中国“绿色解法”

AI 相关的环境与资源压力正迅速从“纸面成本”转化为现实的“落地风险”。据追踪机构 Data Center Watch 统计，2026 年第一季度全美至少有 75 个数据中心项目被阻止或推迟、涉及金额约 1,300 亿美元，创下其 2023 年开始追踪以来的季度之最，已逼近 2025 年全年被阻项目总额。具体面临的压力主要有两大方面：

- **反对力量正快速组织化：**活跃的草根反对组织从 2025 年底的 396 个激增至 2026 年 3 月的 833 个，遍布 49 个州（以马里兰、俄亥俄、得克萨斯居多），部分地区甚至项目尚未正式申报、仅凭传闻即触发有组织抵制。这股反弹还呈现出一些“跨党派”的一致性——共和党人多聚焦电价上涨与电网压力，民主党人则更关注环境与水资源影响。盖洛普 2026 年 3 月的民调亦显示，约七成美国人反对在本地新建数据中心，近半数“强烈反对”。
- **监管从激励转向收紧：**2026 年前六周，各州议会已提出逾 300 项数据中心相关法案，覆盖 30 多个州，标志着从“税收激励”向“监管问责”的转向，多州酝酿建设暂停（moratorium）立法；联邦层面，参议员桑德斯与众议员奥卡西奥 - 科尔特斯于 2026 年 3 月宣布《AI 数据中心暂停法案》（AI Data Center Moratorium Act）。

对投资者而言，这意味着算力扩张的瓶颈正从“能不能供电”延伸到“社区是否接纳”。社会许可（social license）正成为评估数据中心与算力资产时不可忽视的新型政治与监管风险变量。

中国语境：“东数西算”与绿电如何重塑算力地图

对中国而言，数据中心的建设已上升到国家战略层面。中国在宏观层面对数据中心建设的空间布局进行了系统的顶层规划。首先，国家明确设立了八个国家算力枢纽节点、十个国家数据中心集群，构建了全国一体化算力网络的基础框架；其次，通过纵深推进“东数西算”工程，即在内蒙古、宁夏、甘肃、贵州等西部枢纽集中布局数据中心，有序引导对时延要求较低的算力需求向西部集聚，在宏观地理尺度上实现了“算力—电力—绿电”的深度匹配。

在优化空间布局的同时，中国也为绿色数据中心的建设提供了丰富的政策支持。例如，2026 年，国家发展改革委、国家能源局联合印发了《关于有序推动多用户绿电直连发展有关事项的通知》，将绿电直连从“单用户”点对点模式正式拓展为“多用户”集群化共享模式。这一国家级战略支持允许新能源不经过公共电网，直接依托专用线路向园区或算力集群输送，通过共享专线和变电设施降低重复投资，并提升绿电消费可追溯性。

强力支持的背后也伴随着绿色算力的严格约束。国家发改委等多部门 2024 年 7 月联合发布的《数据中心绿色低碳发展专项行动计划》，已明确将电能利用效率（PUE）和可再生能源占比纳入核心监管指标。行动计划提出到 2025 年底，新建及改扩建大型 / 超大型数据中心 PUE 不高于 1.25，国家枢纽节点项目不高于 1.2。这一举措提升了行业的准入门槛，为绿色算力确立了硬性标准。

总的来说，从顶层空间布局的重塑、支持性政策的引导，到约束性监管指标的落地，中国国家政策层面的导向形成了一套完整的闭环：全面向绿色低碳转型已成为中国算力产业发展的趋势。

顺应国家宏观战略的强力导向，中国数据中心企业正加速向绿色算力转型。头部企业万国数据的实践，为我们提供了一个极具代表性的窗口。

案例 | 万国数据——以绿色算力支撑 AI 发展

作为国内领先的第三方数据中心服务商，万国数据直面 AI 算力爆发所引发的能耗激增挑战，从战略顶层确立了“2030 年实现运营碳中和、100% 使用可再生能源”的双重核心目标。为将上述承诺落到实处，公司围绕能源结构、运维能效与空间布局，构建了系统性的实施路径：

- **能源结构多元化与源网协同：**积极拓展绿电直采、绿证交易及园区分布式光伏等手段，并将能够稳定锁定供给与成本的三年期购电长协（PPA），以及具备条件的“绿电直连”模式置于资源配置的核心位置。此外，公司在内蒙古某吉瓦级园区自投风电与光伏项目，实现了环境权益在园区内的直接闭环消纳。
- **技术赋能深挖能效潜力：**广泛部署液冷技术以适应高密度 AI 算力需求，并自主研发 AI 驱动的智能运维系统（AI Box），通过精准调控显著优化制冷与整体能效，使相关环节的节能水平达到 5% 至 10%。
- **空间布局顺应算力枢纽战略：**敏锐把握国家算力地图布局的机遇，将新建数据中心的战略重心加速向西部转移。目前，其业务版图已深度拓展至内蒙古乌兰察布及和林格尔、广东韶关、宁夏中卫等具备超大规模 AI 部署潜力的新兴算力枢纽。

万国数据的绿色转型不仅源于内生战略，更深受产业链下游诉求的强力驱动。当前，通过签署节能减排备忘录，建立起针对 PUE 改善与节电收益的共享机制，万国数据将绿色算力真正转化为产业链上下游协同共赢的商业实践。

资料来源：华夏基金 ESG 团队、紫顶研究，根据与万国数据访谈及公开信息整理

面对 AI 算力爆发带来的资源与环境挑战，中国正在探索并给出一条独具特色的“绿色解法”。这套解法并非单一维度的技术升级，而是宏观顶层空间布局（如“东数西算”工程）与微观企业创新实践的深度共振。万国数据的案例生动地表明，在国家硬性监管约束与产业链下游减排诉求的双重驱动下，绿色已经从数据中心服务商的纯粹成本项，转化为面向整条产业链的“价值主张”。全面拥抱并系统性构建绿色算力，不仅是破解 AI 能耗悖论的必由之路，更是顺应中国乃至全球产业长期高质量发展的大势所趋。

CHAPTER 02

AI 与社会（S）

岗位重构而非岗位消失

在 AI 引发的所有 ESG 议题中，就业与劳动力问题最拨动人心，也最容易引发恐慌。基于我们的访谈与观察，我们在本章给出的判断是：AI 对就业的冲击是一场大规模的“岗位重构”而非单纯的“岗位消失”——但这场重构的分布是不均匀的，且最先感受寒意的正是入门级岗位。

具体而言，我们观察到三条清晰的线索。其一，偏技术性的初级岗位确实在承压，并且冲击首先落在“入门”而非“存量”之上。其二，在人工成本占比较低的传统产业中，AI 的赋能作用远大于替代。其三，面向客户、基于实际工作流沉淀的经验诀窍（know-how）与人际沟通能力（people skills）相关的岗位相对难以替代。

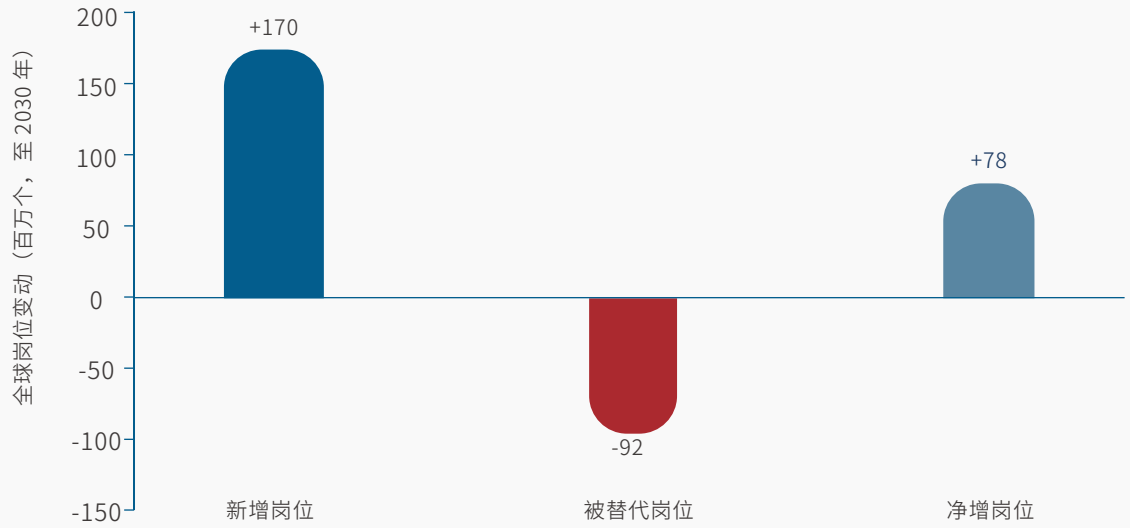
这三条线索指向同一个结论：**真正决定这场转型社会后果的，不是 AI 的技术能力边界，而是企业把 AI 定位为“员工能力放大器”还是“成本削减刀”。**

1 替代还是重塑？ AI 引发的大规模岗位重构

关于 AI 是否正在“夺走工作”，一些代表性的实证研究尝试给出了一些回答。Anthropic 于 2026 年 3 月发布的《Labor market impacts of AI: A new measure and early evidence》，基于 Claude 平台真实工作场景引入“观察暴露度”（observed exposure）指标，核心发现是：AI 对整体失业率的冲击目前尚不明显，但在高暴露职业中，22—25 岁年轻劳动者的入职率较 2022 年下降约 14%，而 25 岁以上人群则没有相应变化；这与 Brynjolfsson、Bharat Chandar 与 Ruyu Chen 发表的研究《Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence》相呼应，即“暴露职业中 22—25 岁群体就业下降 6%—16%[2]，且主要源于招聘放缓而非离职（裁员增加）”。这一早期信号清晰指向结构性分化：冲击首先落在“入门门槛”而非“存量”之上。

世界经济论坛《Future of Jobs Report 2025》基于逾千家雇主调查预测，技术变革到 2030 年将净增约 7800 万个岗位（创造约 1.7 亿、消减约 9200 万），但同时指出接近六成劳动者需要完成培训才能适应新的岗位需求——挑战不在总量，而在转换速度与公平性。

图 3：WEF 对全球劳动力市场结构转型的预测

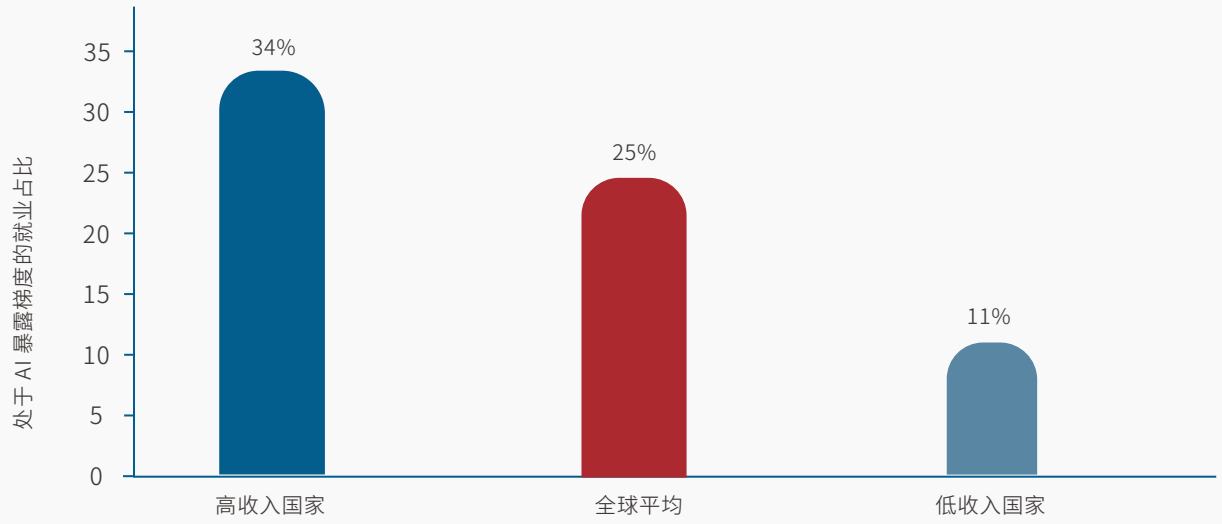


资料来源 : 世界经济论坛《Future of Jobs Report 2025》

[2] 这一范围之所以较为宽泛，是因为作者针对多种假设情形提供了估算结果。与一种假设情形（即就业增长持平）相比，降幅为 6 个百分点。而 16 个百分点的估算结果则源自另一种设计方法，该方法比较了同一家企业中从事不同职业的相似员工。

国际劳工组织（ILO）则提供更细腻的分布视角：全球约四分之一就业处于生成式 AI 的“暴露”梯度，但反复强调“暴露≠替代”——全球处于最高暴露区间的就业仅约 3.3%，大多数岗位是被 AI 增强而非取代；暴露度的不均衡同样值得警惕：在就业结构内部，文书与行政岗位以及女性从业者的暴露度相对更高；在国家和地区层面，数字基础设施与技能储备不足的发展中经济体，可能更难以分享 AI 带来的生产率红利，从而进一步扩大既有发展差距。

图 4：国际劳工组织关于 AI 暴露梯度的就业占比分析



资料来源 : 国际劳工组织（ILO）《Generative AI and Jobs》2025 年更新；“暴露”不等于被替代

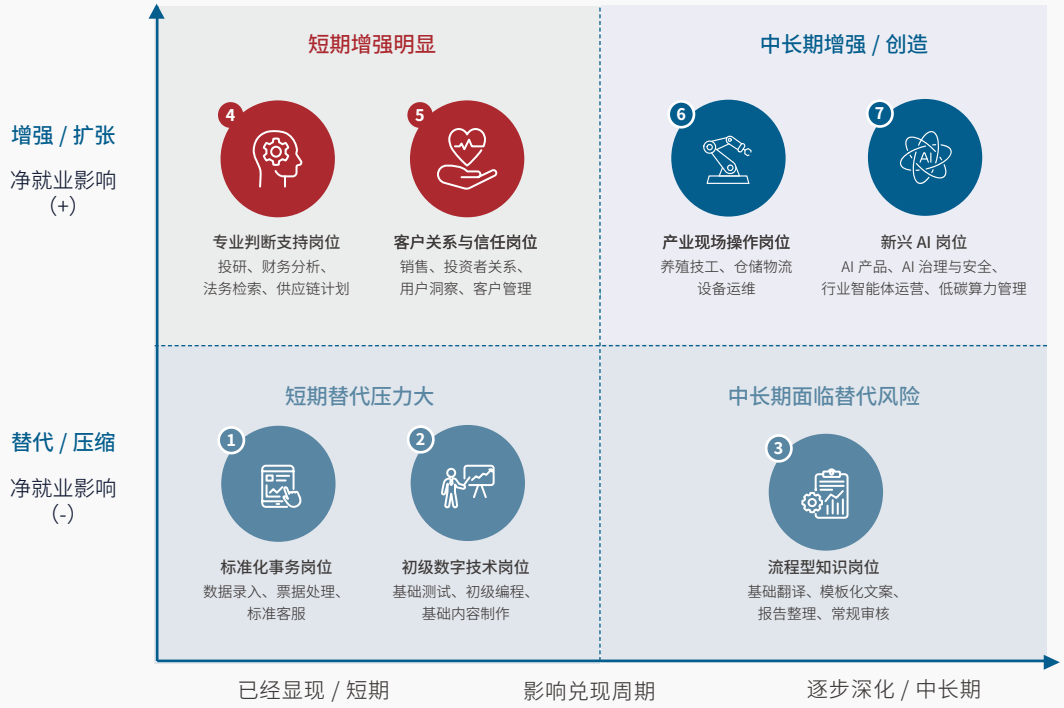
综合三方证据，我们的判断是：AI 的就业冲击当前是结构性而非总量性的，但早期信号已经出现，且最先感受寒意的正是入门级岗位。

AI 对劳动生产率的提升也呈现出一定的“拉平效应”。Brynjolfsson、Li 与 Raymond 发表于《季度经济学期刊》（QJE，2025）的研究基于逾 5,100 名客服人员的工作数据发现，使用生成式 AI 助手后，客服人员平均每小时处理的问题数量提高约 15%。其中，经验较少、原本绩效较低的员工改善最为明显，效率提升约 30%；对经验丰富、绩效较高的员工而言，AI 带来的额外帮助则较为有限。

这种拉平效应意味着，AI 有望成为缩小技能鸿沟、提升一线劳动者价值的工具，而非单向加剧不平等的技术。对于传统产业而言，这尤为关键——AI 让那些原本因资历不足、地理偏远、培训资源匮乏而处于弱势的劳动者，获得了与经验丰富同事接近的工作支撑。关键变量仍在于企业如何设计培训机制，以及能否让这种效率提升真正转化为员工的薪酬与职业发展，而非单纯归于资本端的成本压缩。

结合上述实证研究与产业走访，我们构建了一张 AI 对不同类型岗位的差异化影响的示意图，从“净就业影响”与“影响兑现周期”两个维度，对当前劳动力市场可能面临的转型进行了分类，如下所示。总体来看，我们认为，面向客户、基于实际工作流沉淀的经验诀窍（know-how）与人际沟通能力（people skills）相关的岗位相对难以替代。

图 5：AI 对不同类型岗位的差异化影响



资料来源：华夏基金 ESG 团队、紫顶研究

2 抉择的分野：是“成本削减刀”还是“能力放大器”

如果说上述研究回答的是“AI 对就业做了什么”，那么接下来的问题便是“企业选择让 AI 做什么”——后者才是真正决定社会后果的变量。同样的技术能力，嵌入不同的组织战略，会导向截然不同的结果：一家企业可以用 AI 压缩人员编制、将效率收益直接转化为短期利润；也可以用 AI 重新定义岗位内涵，把释放出的人力投向更高价值的工作，从而在提升效率的同时保留甚至扩展就业容量。

在我们的访谈与案例追踪中，中国企业的应对光谱已经出现分化。我们选取两家行业属性、人员结构与 AI 成熟度各异的公司作为观察样本，试图回答一个共同问题：企业究竟把 AI 定位为“成本削减刀”还是员工“能力放大器”？

案例：两种转型路径

案例一 金山办公——以赋能替代焦虑

除了“不裁员”“保持研发投入”等常见叙事之外，金山办公最有辨识度的特色在于把 AI 素养培训做成了一套可量化、可考核的“学—考—用”闭环。截至 2026 年一季度，公司研发人员接近 4000 人、占员工总数约 66%，保持稳定。并且，公司 2025 年组织全员 AI 素养测评与专项培训，以考试形式检验成果，全部参训人员均通过考试；AI 使用从研发部门扩展至全员，投资者关系等非研发岗位也用内部智能体处理股东名册分析等事务性工作。金山办公的逻辑简单但有力：在 AI 取代工作的焦虑叙事之外，选择相信“被 AI 赋能的员工”比“被 AI 替代的岗位”更有价值。

资料来源：华夏基金 ESG 团队、紫顶研究，根据与金山办公访谈及公开信息整理

案例二 牧原股份——从“养殖技工”到“养猪工程师”

作为一家农业企业，牧原股份最明显的特色在于它算的是一笔与科技企业截然不同的账：人工成本在养殖总成本中占比有限，简单减员意义不大；但若 AI 帮助一个表现偏弱的场线把成本降低约 1 元 / 公斤，单头猪即对应约 120 元的成本改善，相比裁员则有更深远的经济利益。

在岗位转型层面，牧原推动一线员工从传统“养殖技工”向掌握数据与智能装备的“养猪工程师”转型：2019 年非洲猪瘟后大规模引入智能环控、智能饲喂、智能巡检等装备，减少对经验的依赖；近两年在海量底层数据之上构建“养猪大模型”，为猪场批次生成诊断报告，为区域场线生成经营分析。牧原的护城河不只是模型本身，而是“自繁自养沉淀的动态数据＋数百万级智能装备＋内部生产管理体系”三者的耦合。

资料来源：华夏基金 ESG 团队、紫顶研究，根据与牧原股份访谈及公开信息整理

两个案例，两种转型逻辑，共同指向同一个结论：企业对 AI 的定位选择，直接决定了社会后果的性质。将 AI 定位为裁员工具，结果是就业存量减少；将 AI 定位为能力放大器，结果是效率提升、岗位重构，且对弱势劳动者的赋能效果往往更显著。对投资者而言，这一定位选择不只是 ESG 意义上的价值观问题，更是判断企业 AI 战略能否形成可持续竞争优势的核心变量——靠裁员省出来的成本是一次性的，靠全员赋能提升的组织能力是长期的。

CHAPTER 03

AI 与治理（G）

数据安全、监管图景与主动治理

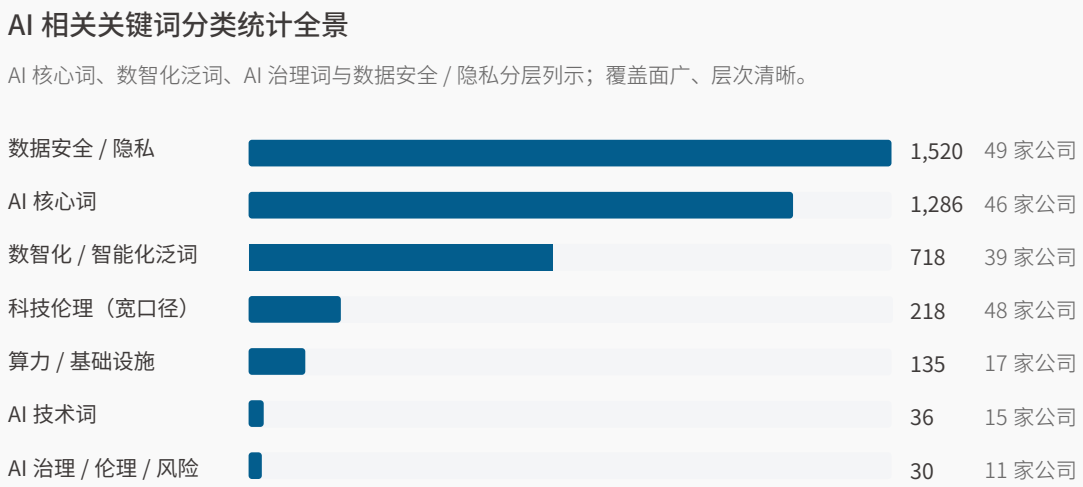
如果说环境与社会维度回答的是“AI 带来了什么后果”，那么治理维度回答的则是“谁来为后果负责”。数据与隐私是生成式 AI 的“原料”，也是治理风险最集中的领域。**我们通过研究发现，大多企业对 AI 仍处于“重应用、轻治理”的状态。**与此同时，全球监管正加强政策约束，中、欧、美三大经济体走出了特征鲜明的三条路径。

1 披露现状：“高频用 AI、低频谈治理”的落差

为了评估 AI 浪潮对上市公司 ESG 实践的真实影响，我们尝试从一个可观察、可量化的披露视角切入。为此，我们选取了科创 50 指数成分股作为观察样本，这些公司是我国硬科技与创新驱动的代表性企业，无论在研发 AI 还是应用 AI 上都走在前列。我们对其 2025 年度可持续发展（ESG）报告进行系统的文本挖掘与词频分析，试图用数据勾勒出 AI 在企业 ESG 披露中的真实位置。

分析显示，AI 相关词的覆盖率已达 92%（50 份报告中 46 家出现 AI 相关词），累计提及 1,286 次——AI 已成为硬科技企业 ESG 叙事的高频议题，而非少数先行者的独有表达。与此同时，基础合规方面相当扎实：数据安全、隐私保护、网络安全等词汇覆盖了 49 家公司（98%）、累计 1,520 次，宽口径的“科技伦理”亦覆盖 48 家（96%）。这说明，绝大多数科创企业已具备“数字责任”的基础，并正沿“数智化→ AI 化”的路径持续进阶。

图 6：科创 50 成分股公司 2025 年 ESG 报告词频分析统计

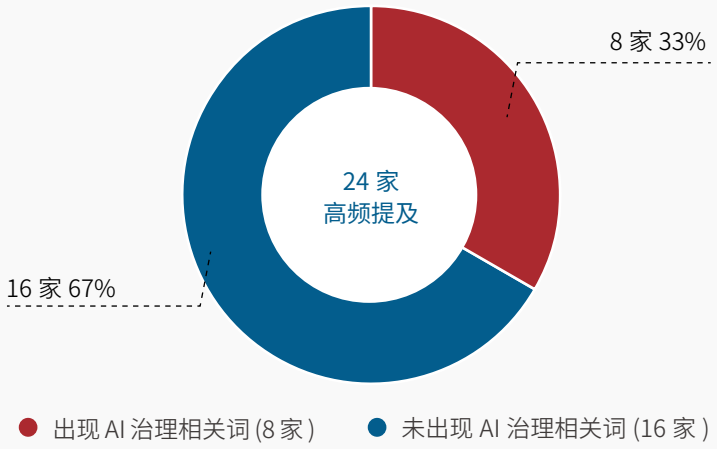


资料来源：华夏基金 ESG 团队，紫顶研究

注：“数据安全 / 隐私”包括数据安全、网络安全、隐私保护、个人信息保护、数据治理；“AI 核心词”包括生成式人工智能、生成式 AI、大语言模型、人工智能、大模型、智能体、DeepSeek、ChatGPT、AIGC、LLM、AI；“数智化 / 智能化泛词”包括智能制造、智能化、数智化、数智；“科技伦理（宽口径）”仅统计科技伦理；“算力 / 基础设施”包括智算、智能算力、算力；“AI 技术词”包括机器学习、深度学习、自然语言处理、计算机视觉、神经网络等；“AI 治理 / 伦理 / 风险”包括负责任人工智能、负责任 AI、可信人工智能、可信 AI、人工智能治理、AI 治理、人工智能伦理、AI 伦理、人工智能安全、AI 安全、AI 风险、算法安全等。

值得一提的是，真正精确到 AI 专项治理的表达仍相对低频。我们观察到，有 24 家公司 AI 相关词提及已达 10 次以上，其中却有 16 家在本次关键词统计口径下，未出现“AI 治理 / 伦理 / 风险”相关表述，呈现出“高频用 AI、低频谈治理”的明显反差。

图 7：科创 50 成分股公司 2025 年 ESG 报告词频分析之高频提及 AI 的公司分析



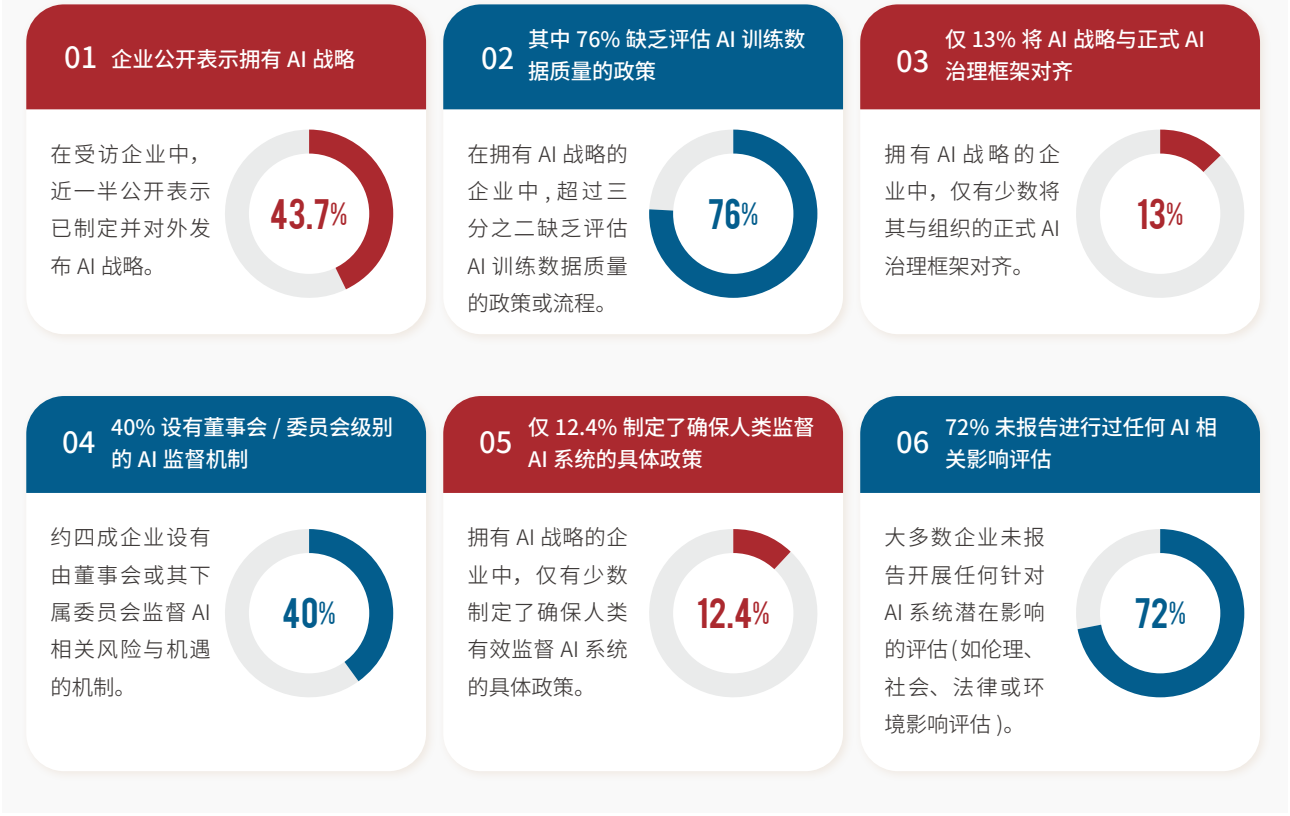
资料来源：华夏基金 ESG 团队，紫顶研究

总体来看，中国科创企业的“负责任 AI”披露正处在由通用数据合规向 AI 专项治理延伸的转变之中——只是目前仍集中于少数先行者，尚未成为普遍的常规披露。

不过，这种“高频用 AI、低频谈治理”的反差并非中国市场独有，而是一个全球性的共同挑战。汤森路透基金会与联合国教科文组织（UNESCO）于 2026 年联合发布的《Responsible AI in Practice》报告中，基于全球 2,972 家企业的公开披露数据，揭示了海外企业在 AI 治理方面同样存在显著缺口：虽然 43.7% 的企业公开表示拥有 AI 战略，但其中 76% 缺乏评估 AI 训练数据质量的政策；仅 13% 的企业将其 AI 战略与正式的 AI 治理框架对齐；在监督层面，40% 的企业报告设有董事会或委员会级别的 AI 监督机制，但仅 12.4% 制定了确保人类监督 AI 系统的具体政策；更值得关注的是，72% 的企业未报告进行过任何与 AI 相关的影响评估。

该报告的结论与我们的观察高度一致——当前负责任 AI 的核心挑战已不再是“意识”（awareness），而是“落地”（operationalisation）。

图 8：汤森路透基金会与联合国教科文组织（UNESCO）关于治理的统计



资料来源：汤森路透基金会与联合国教科文组织（UNESCO）于 2026 年联合发布的《Responsible AI in Practice》报告，基于全球 2,972 家企业的公开披露数据

因此，这道“治理缺口”是企业披露成熟度滞后于技术应用速度的客观写照。对投资者而言，这恰恰孕育着尽责管理与价值发现的机会：推动企业把已经在落实中的治理实践（如数据安全委员会、算法审查、幻觉率管控等）系统地披露出来，既能提升透明度，也能反向夯实其内部治理体系。

2 规则演进：中美欧 AI 治理的三种路径

伴随 AI 技术的加速落地，全球主要监管机构正迅速完善政策，加强治理约束。面对技术演进带来的治理挑战，中、欧、美三大核心经济体基于各自的产业禀赋，探索出了三条特征鲜明的监管路径：中国边发展边规范，以敏捷立法快速迭代；欧盟以统一规范严控高风险应用；美国则联邦与各州来回拉锯，规则分散。三者的分野，本质上是对“发展”与“安全”如何排序、以及“由谁来立规”的不同回答。

中国：发展与安全并重的“敏捷治理”

中国人工智能治理始终坚持发展与安全并重。2026 年 7 月 17 日，习近平主席在 2026 世界人工智能大会开幕式上系统阐释了中国的人工智能治理主张，提出秉持以人为本、向上向善理念，携手构建公正合理的全球人工智能治理体系。在安全治理层面，讲话强调推动构建涵盖法律法规、技术监测、风险预警和应急响应的治理体系，防范人工智能滥用恶用，确保其安全、可靠、可控；在全球治理层面，则倡导加强各国发展战略、治理规则与技术标准的协调，并通过能力建设帮助全球南方弥合数智鸿沟。由此，中国人工智能治理呈现出两条相互衔接的主线：在国内鼓励创新发展的同时持续筑牢安全底线，在国际层面推动形成兼顾安全、发展与普惠的多边治理框架。

在国内制度实践层面，中国并未以一部综合性法案规范所有人工智能场景，而是采取“小步快走、他律与自律结合”的敏捷治理路径，其制度底座是“算法备案＋大模型备案”的双备案机制。2023 年 8 月 15 日，七部门联合施行《生成式人工智能服务管理暂行办法》，此后监管框架不断细化。截至 2025 年底，我国累计已有逾 700 款生成式人工智能服务完成备案，反映出“备案制”在促进创新与守住底线之间的动态平衡。

在此基础上，规则进一步向“内容可溯”与“智能体可控”两个前沿延伸。其一，在内容层面，2025 年 3 月 14 日国家网信办等四部门发布《人工智能生成合成内容标识办法》，并于 2025 年 9 月 1 日与配套强制性国标《网络安全技术人工智能生成合成内容标识方法》（GB 45438-2025，2025 年 2 月 28 日发布）同步施行，确立“显式标识＋隐式标识”的双重标识体系——显式标识以角标、文字或语音提示等可感知方式呈现，隐式标识则嵌入文件元数据（数字水印为鼓励性而非强制要求），直接回应了“深度伪造”与虚假信息的社会关切。

其二，随着智能体（Agent）加速落地，治理触角随之前移：2026 年 4 月 10 日，中国人工智能产业发展联盟（AIIA）安全治理委员会发布《OpenClaw 类智能体部署风险管理指南》，围绕选型部署、运营维护、下线审计等环节提出“操作可信、权限可控、风险可溯”的治理原则与五十余项可落地条目，并区分应用方与提供方的不同责任。此外，2025 年修订的《网络安全法》自 2026 年 1 月 1 日起施行，新增了针对人工智能安全与治理的专门条款，进一步收紧了企业的合规义务。

欧盟：风险分级驱动的“统一严管”

相比于中国的敏捷迭代，欧盟走的是“大而全”的法典化路线。《人工智能法案》（EU AI Act）作为其制度底座，确立了以“风险分级”（不可接受—高—有限—极低四个级别）为核心的严密监管体系，

其最大特点是“防范优先”与“重罚机制”：对高风险应用设定了严格的前置合规义务，并配套了阶梯式罚则结构——违反禁止性条款最高 3,500 万欧元或全球营收 7%，违反高风险义务最高 1,500 万欧元或 3%，提供虚假信息最高 750 万欧元或 1%。

尽管受 2026 年 6 月 29 日欧盟理事会批准通过的“AI Omnibus”简化提案影响，部分高风险义务的适用时点推迟至 2027—2028 年，但欧盟这种强监管、高压力的防御型框架，已然成为跨国企业当前面临的最严峻的合规挑战。

美国：联邦与州拉锯的“分散监管”

美国的监管图景则呈现出鲜明的拉锯特征。为在激烈的全球 AI 竞赛中维持技术话语权，美国在联邦层面倾向于推动创新，试图通过减少干预来规避碎片化立法对产业发展的拖累。但这一路线并不平坦，美国至今尚未出台综合性的联邦 AI 法案，国会对强监管亦持高度谨慎态度，如参议院曾以 99:1 的压倒性投票删除了一项法案中的 10 年暂停期条款。

各州政府在监管尺度上也剧烈摇摆。例如，科罗拉多州综合 AI 法在 2026 年 5 月遭遇废止，而加利福尼亚、伊利诺伊等州却坚持保留对自动化决策工具与算法偏见审计的具体约束。

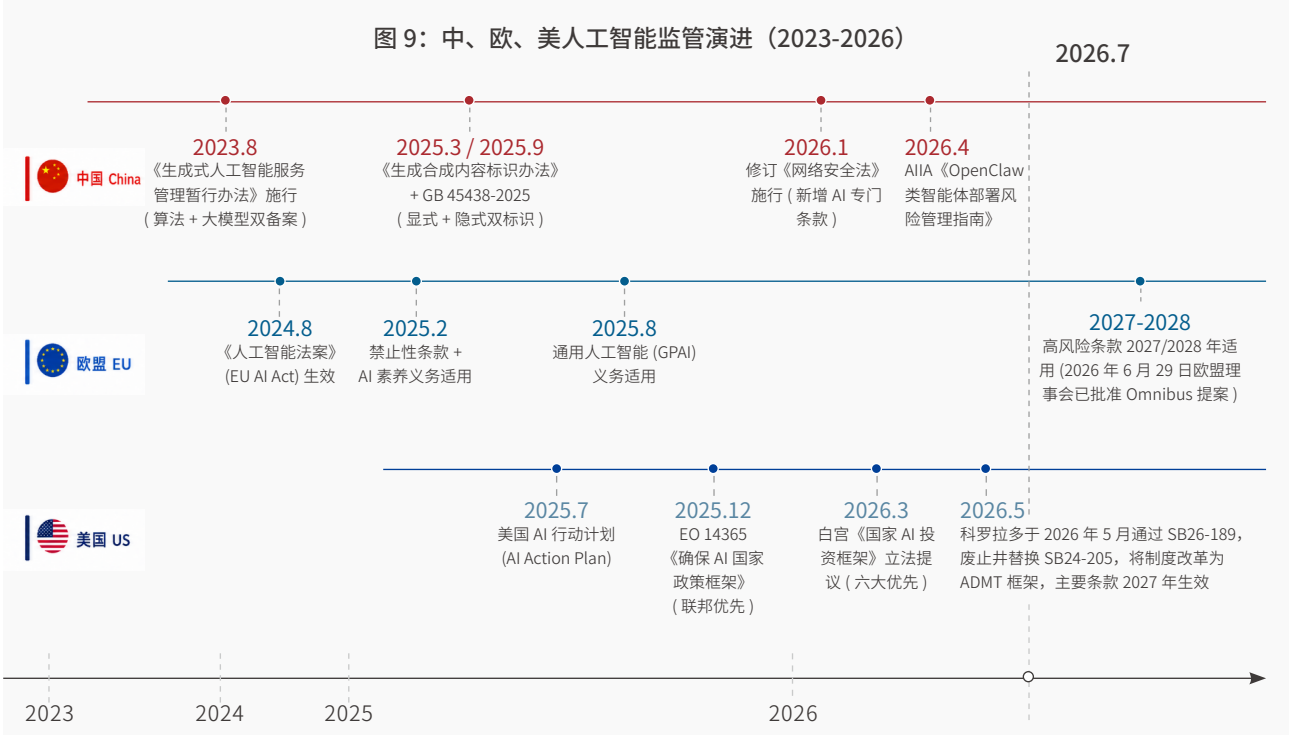
对企业而言，美国监管的挑战并不在于统一的严法，而在于应对碎片化且政策摇摆的地方法规，这无疑增加了运营的合规复杂性。

表 1：中、欧、美三大经济体 AI 监管路径对比

维度	中国	欧盟	美国
监管特点	发展与安全并重	风险分级、防范优先	联邦与州并行，多层规则共同演化
标志性规则	生成式 AI 服务管理暂行办法、内容标识办法（GB 45438-2025）	《人工智能法案》（EU AI Act）	EO 14365 联邦 AI 政策行政命令（2025.12）、《国家 AI 政策框架》立法建议（2026.3）、州级立法
企业核心合规点	备案登记、内容双重标识、数据安全	高风险系统义务、透明度、高额罚款风险	关注联邦与州规则变化、诉讼风险

资料来源：华夏基金 ESG 团队，紫顶研究

图 9：中、欧、美人工智能监管演进（2023-2026）



资料来源：华夏基金 ESG 团队，紫顶研究

3 企业实践：从被动合规走向主动治理

如果说前述监管图景勾勒的是“外部约束”，那么真正决定治理成效的，仍是企业自身能否把合规要求内化为日常的产品与组织实践。前文的词频画像已揭示出 A 股科创龙头“高频用 AI、低频谈治理”的落差：数据安全、隐私保护等传统合规词几近全覆盖，但真正指向 AI 本身的专项治理表达（负责任 AI、算法安全、合成内容标识等）仍只在少数先行者中出现。这意味着，多数企业的治理重心仍停留在“数据不出事”的底线层面，尚未上升到“模型可信、算法可控、内容可溯”的主动治理层面。

从被动合规走向主动治理，领先企业的实践大体可归纳为四个层次的协同：其一是组织层面，在董事会或专门委员会设立 AI/ 数据安全与隐私治理机制，明确问责链条；其二是数据层面，建立分级分类、最小化采集、加密存储与个人 / 企业数据隔离的全生命周期管理；其三是技术与部署层面，针对政企客户提供私有化部署、数据“不出域”，并以数字水印、内容标识回应合成内容的可溯要求；其四是产品层面，通过企业知识库约束生成范围、压降幻觉率，并嵌入人工复核与用户反馈闭环，使 AI 输出“可控、可信、可纠错”。这四层并非彼此孤立，而是构成一条“组织定责—数据筑基—技术兜底—产品落地”的闭环，缺一环都可能让治理流于纸面。

这一逻辑在数据密集型行业尤为关键。对于掌握海量用户文档、且正把 AI 深度嵌入核心产品的企业而言，数据安全与隐私既是合规红线，更是产品信任的生命线，一旦失守，损失的不只是罚单，而是用户与政企客户的长期信任。正因如此，办公软件龙头金山办公的系统性实践，提供了一个观察“主动治理”如何落地的样本。

案例|金山办公——从“Office AI”到“AI Office”的数据安全治理

金山办公旗下 WPS 用户累计上传云文档总量超 2,900 亿份，公司同步推进 “Office AI 重构升级”与原生“AI Office”探索的产品转型工作，并始终将数据安全与用户隐私保护置于最高优先级。金山办公差异化的核心治理特色，在于其“工程化治理”的落地方式——不停留于原则声明，而是把治理要求变成可执行的系统架构。

在组织层面，公司同时设立安全委员会与隐私保护委员会，配套内部章程，将问责机制落至制度层；在数据层面，对海量用户文档实施全栈加密，通过账号体系将个人数据与企业数据严格隔离；在技术与部署层面，面向党政与央国企客户提供私有化部署、数据“不出域”，面向海外业务则在欧洲、美洲、东南亚分区部署服务器、就地处理数据，以符合当地监管要求；在产品层面，公司不是简单接入大模型 API，而是以企业知识库约束生成范围、辅以数据清洗标记与场景化工程优化，并在产品端设置申诉反馈渠道，用户亦可通过客服渠道提出问题或建议，使输出“更可控、更可信”。

这套从组织、技术到产品的闭环，使数据安全从合规成本转化为面向政企客户的差异化竞争优势——这正是“负责任 AI”从理念走向可落地实践的典型注脚，也是其他企业可参照的高质量样本。

资料来源：华夏基金 ESG 团队、紫顶研究，根据与金山办公访谈及公开信息整理

如果说金山办公展示的是如何把“数据安全”做成闭环，那么当 AI 生成的画面、素材、视频本身成为企业的核心资产与风险源时，治理的重心还需从“数据不外泄”延伸到“内容可确权、可追溯”。在这一方向，以内容生产见长的世纪华通提供了另一个侧面的样本。

案例|世纪华通——为 AIGC 内容打上“可溯”的水印

作为以游戏为核心的数字娱乐龙头，世纪华通在“全员 AI”战略下将 AIGC 大量用于美术与广告等内容生产，也由此把“生成内容的权属如何确认、侵权如何阻断、过程如何追溯”推到治理前台。公司为此构建了“制度 + 技术 + 合规”三位一体的体系：以《世纪华通 AI 安全管理制度》明确从输入素材到最终产出的登记与责任规则；要求训练数据与输入素材须取得合法授权或确认进入公有领域，从源头阻断侵权；并通过数字水印与区块链存证实现内容标识与溯源，确保权属可验证——这正是对《合成内容标识办法》与 GB 45438-2025 要求的企业化落地。

在此之上，公司还设有面向版权纠纷的“三层防御”：事前以合规指引与培训禁止输入未授权素材，事中由法务 / 合规审核后方可发布，事后建立证据固定与快速响应的应急预案。

资料来源：华夏基金 ESG 团队、紫顶研究，根据与世纪华通访谈及公开信息整理

从金山办公的数据安全治理到世纪华通的内容可溯体系，两个样本共同表明，领先企业的 AI 治理，正沿着“组织定责—数据筑基—技术兜底—产品落地”的链条，从守住合规底线走向主动加强治理。而要让这类实践从少数先行者扩散为群体性的常规动作，离不开资本市场的推动——这正是下一章的主题。

CHAPTER 04

负责任 AI 与机构投资者从“倡议”到“议程”

前文从企业与监管视角剖析了 AI 对 ESG 三大维度的影响。本章回到投资者本位：作为资产的所有者与管理人，机构投资者应如何面对 AI 浪潮？我们认为，AI 既是重塑企业估值的新变量，也是责任投资与尽责管理（Stewardship）的新前沿。**全球领先的资产所有者已率先行动，将“负责任 AI”从抽象理念转化为可操作的投票立场与沟通议程。**

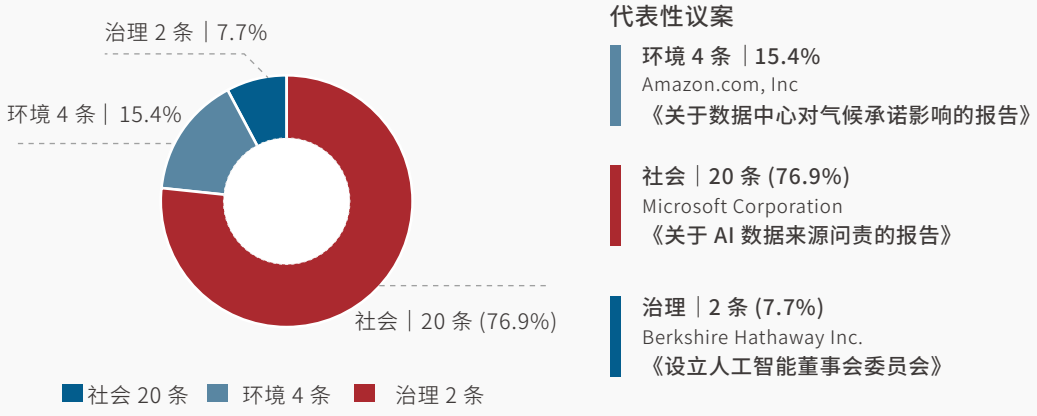
1 海外趋势：从倡议走向深度的尽责管理

在全球范围内，领先的资产所有者与资产管理人已率先将“负责任 AI”纳入尽责管理的工作范围。挪威主权财富基金（NBIM）早在 2023 年便发布负责任 AI 投资立场《Responsible artificial intelligence》，明确要求董事会承担 AI 问责、保障透明度与可解释性，以及实施涵盖隐私、安全、非歧视与人类监督的稳健风险管理；纽约州共同退休基金（NYSCRF）在其 2025 年尽责管理优先事项中，将 AI 治理与网络安全并列，要求被投企业在董事会层面建立 AI 伦理与安全准则，并着重强调人类对自动化安全系统的监督；安联全球投资（Allianz Global Investors）则在 2025 年发布立场文件，提出横跨 AI 全价值链的稳健性、道德完整性与可持续性三大支柱，明确主张负责任 AI 不是创新的束缚，而是赢得长期信任的前提。三家机构表述各异，却共同指向同一组关键词——董事会问责、透明度、稳健风险管理与人类监督。

近期，在倡议之外，我们观察到领先投资者在更多地通过投票与沟通等路径，把对企业的期望转化为实际压力。

投票方面，AI 相关股东提案方兴未艾。过去两三年，美股 AI 相关股东提案数量明显增多。第三方统计[3]显示，美股上市公司的相关提案从 2023 年的 16 项增至 2024 年的 24 项，再到 2025 年的 26 项，并且这些提案的独立股东平均支持率约为 30%，显著高于同期一般环境与社会（E&S）提案约 16%[4] 的水平。

图 10：2025 年美股上市公司 AI 相关议案分类统计情况



资料来源：华夏基金 ESG 团队，紫顶研究，ISS

[3] 数据源自 ISS STOXX 的“A Look at AI-Related Shareholder Proposals at U.S. Companies, 2022-2025”，详见 https://corpgov.law.harvard.edu/2025/12/22/a-look-at-ai-related-shareholder-proposals-at-u-s-companies-2022-2025/

[4] 引自 Morningstar 美国代理投票数据库，统计期为截至 2025 年 6 月 30 日的 12 个月。

一个较有代表性的机构投资者行动案例，是安联全球投资（AllianzGI）在 Alphabet 于 2025 年 6 月 6 日召开年度股东大会前，专门公开披露其投票意向及判断依据。AllianzGI 宣布支持三项涉及 AI 与社会风险的股东提案，其中提案 12 要求公司就 AI 驱动的定向广告开展独立第三方人权影响评估。该行动反映出机构投资者开始将 AI 风险转化为具体的投票判断，并在会前主动公开其投票逻辑。

具体地，提案 12“要求公司就 AI 驱动的定向广告开展独立第三方社会影响评估”由加拿大投资者组织 SHARE（Shareholder Association for Research and Education）代表投资者提出，其逻辑直指 Alphabet 的商业模式核心：公司约 75% 的收入来自广告，AI 是其核心驱动力；参考 Meta 因广告算法歧视被起诉的先例，如果公司 AI 广告算法存在系统性歧视风险，将面临欧盟 AI 法案等监管压力和法律诉讼，从而损害长期股东价值。

这一提案精准指向了“AI 算法＋定向广告”这一此前被泛化的“AI 伦理”声明所遮蔽的交汇盲区——广告算法的影响、平台游说行为的内在一致性、算法对未成年人认知与行为的潜在影响，都是具体的问责维度，而非抽象的价值观宣示。

沟通方面，机构投资者坚持与科技巨头开展长期对话。比单次投票更具持续性的，是机构投资者与企业之间的沟通（Engagement）。Federated Hermes 旗下专门主导尽责管理工作的 EOS 团队即是典型：早在 2019 年 6 月 Alphabet 股东大会前，EOS 便公开呼吁该公司“在负责任 AI 上引领行业”，敦促其董事会就 AI 的使用及社会影响承担问责、建立内部治理机制；此后，EOS 又通过 2022 年《数字权利原则》与 2025 年《数字治理原则》，把对 AI 治理的期望进一步系统化——涵盖披露算法系统的用途、解释其运作方式、保障用户选择权、消除非预期偏见等。这种“年复一年、就同一议题持续追问”的方式，往往比一次性投票更能推动企业改进。

总体来看，投资者对“负责任 AI”议题的关注，正从泛化的“AI 风险”转向具体的治理架构缺口——要求企业就某一特定场景（如 AI 驱动广告的社会影响）出具第三方评估或董事会监督证明。这一变化意味着，机构投资者不再满足于企业用大而化之的“我们重视 AI 伦理”来回应，而是深入追问企业具体的实践与管理机制。

2 中国实践：华夏基金构建“负责任 AI”的 ESG 评价体系

为了将“负责任 AI”从抽象的伦理声明转化为可量化、可比较的投资准则，华夏基金构建了内部 ESG 评价体系，并依据产业前沿趋势持续演进和优化。针对涉及模型研发和商业化的科技公司，ESG 团队将负责任 AI 作为一项新纳入的议题，从以下三个维度进行系统考察：

(一) 治理架构

华夏基金关注公司是否在董事会或集团层面设立科技伦理与负责任 AI 的专门治理机构，并重点考察该机构的职责边界与决策权力是否明确、是否具备对 AI 产品研发和商业化方向的实质影响力。例如治理架构是否由高层直接领导、是否覆盖跨部门决策、是否引入外部独立视角参与监督与评估，以及董事会层级的战略监督与业务层级的执行管理是否形成有效衔接，确保 AI 伦理从宏观理念转化为可执行的治理制度。

(二) 风险识别与评估

华夏基金关注公司在 AI 产品上线前是否建立系统性的风险评估机制，包括伦理评审、偏见检测与安全测试等环节，并考察其方法论是否覆盖模型训练数据、算法设计、应用场景等关键环节；同时关注产品上线后是否具备持续监测与快速响应的能力，如自动化监测与人工抽查相结合、内容鉴伪技术等，确保风险可识别、度量与干预。

(三) 行业正外部影响力

华夏基金关注领先公司如何将内部治理能力转化为对行业生态的积极辐射。具体包括：是否通过参与国家或国际标准制定、签署行业自律承诺等方式推动负责任 AI 的行业共识形成；是否将自研的安全与伦理检测能力以开源技术、开放平台等形式向行业输出，降低中小开发者践行负责任 AI 的门槛；以及在供应链管理中，是否将 AI 伦理合规纳入对供应商的准入要求，推动负责任 AI 理念向产业链上下游传导。

华夏基金自身的 AI 实践亦恪守“发展与安全并重”原则，持续将 AI 应用于智能投研、数字化管理。华夏基金认为，负责任 AI 作为一项新兴议题，其治理与正外部影响力正在成为衡量企业长期可持续竞争力的重要因子。

结 语

在不确定中锚定确定

回望这场仍在加速的 AI 浪潮，唯一确定的或许就是“不确定”本身：技术演进的速度、监管框架的走向、就业结构的重塑，都难以被精确预判。但在层层不确定之中，仍有一些确定的坐标值得锚定。

其一，方向是确定的——无论技术如何演进，“技术服务于人”始终是负责任 AI 的初心，安全、公平、透明、隐私与可持续不会过时。其二，方法是确定的——把外部性纳入定价、把治理嵌入流程、把人放在转型的中心，是穿越周期的稳健之道。其三，角色是确定的——企业是负责任实践的第一责任人，监管机构划定底线与边界，投资者则以尽责管理推动企业把“已经在做的”系统披露、把“尚未补齐的”尽快补上。

AI 与 ESG 的交汇，既不是非黑即白的判断题，也不是一劳永逸的选择题，而是一道需要持续校准的“净影响”命题。唯有以负责任的方式设计、部署与治理 AI，技术进步带来的长期价值才能真正沉淀下来。在不确定中锚定确定，正是本报告希望与监管机构、企业与同业投资者共同坚守的方向。

免责说明

本报告所提供的信息和资料可能包含预测信息，鉴于实践中存在诸多不确定性，可能导致实际结果不同于预测信息所描述结果，该等信息不应被视为向本报告接收者提出任何建议或保证。除依法应当披露的内容外，任何本报告所参考的信息（包括但不限于产品信息、企业信息、行业资讯等内容）仅为方便接收者阅读而纳入，华夏基金不对该等内容承担任何责任。接收者如使用该等参考信息，应自行确认信息的真实性、准确性、完整性、时效性及其他方面。

华夏基金与紫顶已经采取一切合理措施核实本报告中包含的信息，本报告反映的是本报告发布时华夏基金和紫顶的确信和判断，华夏基金与报告作者不对本报告中的任何内容作出任何保证和承诺，也不就该等信息的任何错误或遗漏承担任何责任。华夏基金可发出其他与本报告内容不一致的报告，但华夏基金没有义务和责任及时更新本报告或将所涉内容通知接收者。解释和使用材料的责任由接收者承担，华夏基金不对任何因使用这些材料造成的损失承担责任。本报告所包含的资料仅供一般性参考。投资者不应当将本报告作为做出投资决策的唯一参考。

关于华夏基金

华夏基金管理有限公司成立于 1998 年 4 月 9 日，是经中国证监会批准成立的首批全国性基金管理公司之一。公司超 27 年投资管理经验，管理规模持续行业领先。公司定位于综合性、全能化的资产管理公司，服务范围覆盖多个资产类别、行业和地区，构建了以公募基金和机构业务为核心，涵盖华夏香港、华夏资本、华夏财富、华夏股权投资等子公司的多元化资产管理平台。

截至 2026 年 6 月底，华夏基金（含子公司）共计在管规模约 3.15 万亿元。服务全国社保、年金、银行、保险、券商及其他大型央企等客户，覆盖欧美亚多个国家和地区的主权基金、央行、银行、资产管理公司等境外机构客户。

可持续发展和 ESG 理念是华夏基金的底层价值观。2017 年 3 月，华夏基金成为联合国负责任投资原则（UN PRI）签约方，是境内首家加入该国际组织的公募基金公司。自此，华夏基金深化践行积极股东主义，是 Climate Action 100+、FAIRR 等国际投资者倡议的第一家国内支持机构，至今已与 70 余家上市公司进行了 170 余次 ESG 深度沟通。此外，华夏基金在行业内率先建立内部代理投票数字化平台，2025 年参与股东大会 1000 余次。协助上市公司提升可持续发展水平的同时，公募基金管理人在上市公司治理的影响亦逐步深化。

关于紫顶股东服务

紫顶（北京）企业服务有限公司（“紫顶股东服务”）成立于 2016 年，是中国本土投票权管理服务的引领者，也是中国投票咨询行业“从 0 到 1”的开拓者。我们致力于为机构投资者提供专业的中国上市公司股东大会投票建议，助力投资者更好地参与公司治理，践行责任投资。

作为行业领导者，紫顶已经研究覆盖了超过 1400 家中国上市公司，为合计管理规模超过 20 万亿美元的资管机构提供服务。紫顶协助机构股东积极地与公司沟通、专业地表达投资者诉求、审慎并坚定地做出投票决策。

成立至今，紫顶持续发挥着“桥梁”的作用，既是促进机构股东与上市公司沟通的桥梁，也是弥合国际投资者与中国公司治理认知差的桥梁。我们期望通过持续高质量的专业服务，助力机构投资者深度参与公司治理，推动实现资本市场的高质量发展。

华夏基金公众号　　紫顶公众号

感谢您阅读本报告，欢迎传播和交流探讨。请关注公众号下载本报告，并获取更多相关资讯。