展望 特刊

ZHAN WANG

accenture 埃森哲

负责任AI

负责任AI: 监管、治理与向善

联想如何打造负责任AI

亿滋专访: “正确做事”与“做正确的事”

阿里专访: AI治理的中国路径与全球对话

卢珊

埃森哲大中华区
技术服务事业部董事总经理、
数据与人工智能主管

卷首语

随着生成式AI的飞速发展，其深远影响已经渗透到经济结构、社会生活以及文化生产的方方面面。从自动生成文本、图像、音视频到编写代码，生成式AI不仅极大提升了内容创作与生产效率，也为商业模式创新、产业转型升级提供了全新动能。它正以前沿科技的姿态，引领一场跨时代的智能革命。

然而，伴随生成式AI的迅猛发展，诸如数据偏见、信息操纵、隐私泄露、知识产权争议以及环境可持续性问题也日益凸显。如何在释放AI技术潜力的同时有效规避其风险，成为当前全球共同面临的重要议题。

在这一背景下，“负责任AI”理念应运而生，并迅速成为全球科技治理的新共识。它主张在AI的全生命周期中融入伦理价值、法律规范与社会责任，确保技术发展始终以人为本，服务于广泛而持久的公共利益。

本期特刊旨在围绕生成式AI的关键议题展开系统性分析：一方面，梳理这一技术带来的新机遇与核心挑战；另一方面，深入比较国内外在AI治理方面的政策法规差异，呈现不同行业头部企业在负责任AI领域的实践路径与创新探索。同时，我们也聚焦于中国企业在本土及全球市场中推进负责任AI落地的关键考量，尝试提出可操作的政策建议与行业倡议。

我们希望通过这本特刊为决策者、技术开发者及行业实践者提供有价值的参考，共同推动构建一个以信任为基石、兼顾创新与治理的健康AI生态系统。

畅享数字化“悦”读体验，即刻扫码阅读全新《展望》

《展望》微信小程序

《展望》主页

目录

负责任AI：监管、治理与向善

文 胡定宇、徐浩

编 王若霈、王朓灵

在AI快速发展的今天，如何确保其在可控、安全、公正的框架下运行，已成为全球共同关注的焦点。各国政府正积极制定AI监管政策，这些政策不仅塑造着AI发展的外部环境，也对企业的AI应用和合规提出了更高要求。

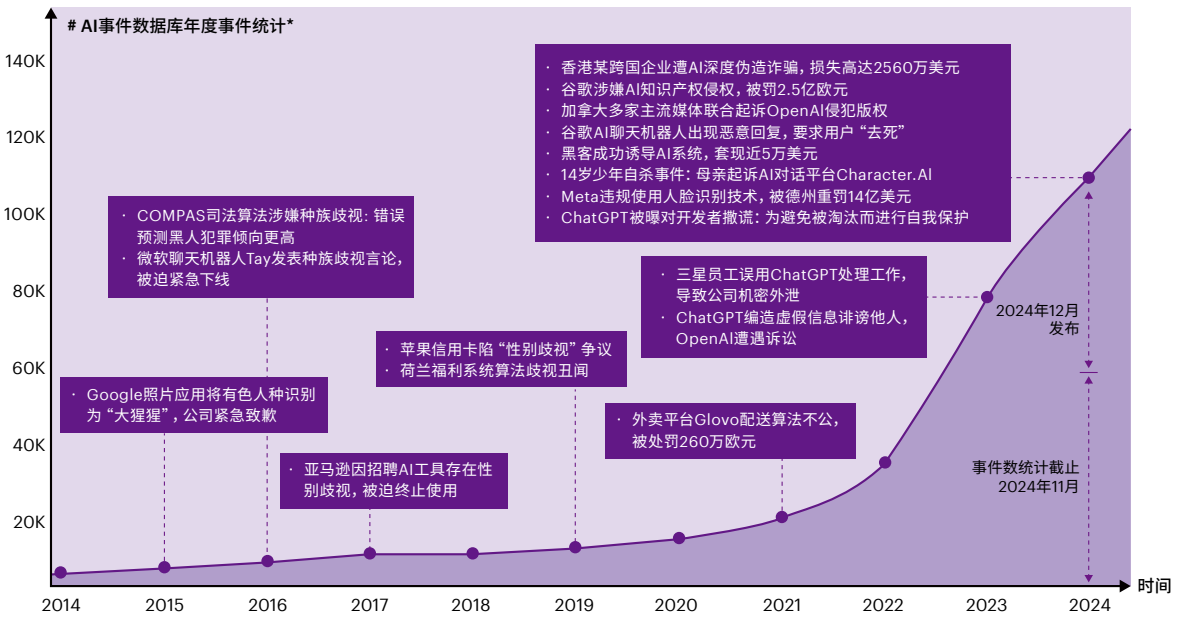
与此同时，越来越多企业开始构建自身的AI治理体系，从制定伦理准则、成立AI伦理委员会，到落实技术透明性与责任机制，力求将负责任AI理念融入各个方面。

过去两年中，生成式AI从技术突破走向广泛应用，中国作为全球AI发展的重要参与者，在政策支持、市场需求和技术积累的共同驱动下，也迎来了生成式AI领域的蓬勃发展。截至2025年1月，中国生成式AI产品的用户规模已达2.49亿人，[1] 截止到2025年6月30日已经有439款生成式AI服务在国家网信办完成备案和233款生成式AI应用或功能完成登记。[2]

然而，随着生成式AI技术的普及，AI风险事故爆发式增长（见图一）。

面对新兴技术带来的多维度挑战，各国政府正在加紧推进生成式AI的监管探索，试图在鼓励创新与治理风险之间寻找平衡点；行业领先企业也在积极响应政府监管要求，并日渐成为标准的共同制订者、生态组织者与价值领导者。

图一 AI风险事故爆发式增长



资料来源：AI事件数据库（incidentdatabase.ai）。[3]

1. 第55次《中国互联网络发展状况统计报告》，中国互联网络信息中心，2025年1月17日，https://www.cnnic.net.cn/n4/2025/0117/c88-11229.html。
2. 《关于发布生成式人工智能服务已备案信息的公告（2025年4月至6月）》，中国网信网，2025年7月11日，https://www.cac.gov.cn/2025-07/11/c_1753948489002783.htm。
3. AI事件数据库旨在系统性地记录和整理人工智能系统在实际应用过程中导致的各类已发生或潜在危害事件。

生成式AI带来新治理挑战

生成式AI的飞速发展，正突破传统技术治理的框架，带来前所未有的系统性治理挑战，主要体现在以下四个方面：

一、技术迭代远超监管响应能力

生成式AI模型更新迭代极快，训练门槛和应用门槛迅速降低，然而现有法律和相应的监管机制与技术的发展存在差距。例如，多国现行隐私保护、数据安全与内容监管体系，难以有效地应对AI生成内容“源头难溯、真假难辨”的特点。这种监管滞后使得企业的AI治理需要更具弹性和前瞻性。

二、责任边界模糊与透明度缺失

生成式AI涉及基础模型提供商、应用开发者、终端用户等多方主体，导致责任链条复杂化。一旦发生内容误导、数据泄露或版权侵犯，责任界定将成为治理难点。同时，训练数据来源不透明、模型行为不可预测等诸多挑战，也削弱了外部监管机构和公众对AI系统的可追责能力。

三、社会影响泛化与风险叠加效应

生成式AI对社会结构和经济格局的影响深度与广度远超以往技术变革，可能带来多重风险，包括加剧失业和劳动力市场两极分化，放大数据偏见，引发知识产权纠纷，增加能源消耗等。

四、全球治理体系碎片化与标准缺位

各国对生成式AI的监管目标和实施路径差异显著，导致跨国企业的合规成本剧增，也暴露出全球治理公认的统一标准的缺失。如何在技术创新、隐私保障、安全可控之间寻求平衡，仍是一个国际共识难题。

AI治理中外监管差异

随着生成式AI的全球影响力日益显现，主要经济体正加速布局AI立法和监管框架。中国、欧盟、美国以及东南亚、中东等地区，呈现出显著差异化的监管路径，反映出各自在治理理念、产业发展阶段和文化背景等层面的根本性差异。

一、中国：安全与发展并重，强调包容审慎

中国对AI的监管秉持“发展和安全并重、促进创新和依法治理相结合”的原则。早在2017年，国务院便印发《新一代AI发展规划》，提出了到2030年建成完善AI法律法规、伦理规范和政策体系的“三步走”战略目标。[4]

法律规范方面，中国的AI法律法规框架目前以垂直领域监管与高风险聚焦为核心特征，采取“敏捷治理、小切口立法”方式，逐步构建多层次监管体系。针对AI的具体应用，中国相继出台了一系列法规和标准，其中《互联网信息服务算法推荐管理规定》着重规范算法推荐服务并要求公开算法运行规则，推进算法透明度和保障用户知情权；《互联网信息服务深度合成管理规定》[5]对利用深度学习、虚拟现实等技术生成文本、图像、音频、视频等信息的服务进行了规范，特别是针对深度伪造等应用；《生成式AI服务管理暂行办法》[6]是专门针对生成式AI服务的核心法规，对生成式AI服务实行包容审慎和分类分级监管；2025年3月发布的《人工智能生成合成内容标识办法》[7]和强制性国家标准《网络安全技术——AI生成合成内容标识方法》[8]则进一步要求对AI生成内容进行显示标识和隐式标识，以保障公众知情权，确保传播内容可追溯和合法合规。这些法规共同构成了中国特色AI治理体系的基础，旨在为AI技术健康发展保驾护航。

4. 《国务院关于印发新一代人工智能发展规划的通知》，中国政府网，2017年7月20日，https：//www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm。
5. 《互联网信息服务深度合成管理规定》，司法部官网，2022年11月25日，https：//www.moj.gov.cn/pub/sfbgw/flfggz/flfggzbmgz/202307/t20230705_482071.html。
6. 《生成式人工智能服务管理暂行办法》，中国政府网，2023年7月10日，https：//www.gov.cn/zhengce/zhengceku/202307/content_6891752.htm。
7. 《关于印发“人工智能生成合成内容标识办法”的通知》，中国网信网，2025年3月14日，https：//www.cac.gov.cn/2025-03/14/c_1743654684782215.htm。
8. 《网络安全技术——人工智能生成合成内容标识方法》，全国标准信息公共服务平台，https://openstd.samr.gov.cn/bzgk/std/newGbInfo?hcno=F32EA2A561F1886CD8D606513512D547。

二、欧盟：基于风险的全面监管框架

欧盟试图在在AI监管领域延续“布鲁塞尔效应”，旨在为AI的发展方向提供指引，引导人工智能产业链的相关服务者能够以人类和社会福祉为原则，构建出符合欧洲价值观以及值得信赖的AI生态环境。

其标志性立法是《AI法案》（EU AI Act），该法案于2024年8月1日正式生效，各项条款将分阶段实施。《AI法案》的核心特点是采用"基于风险"的分类分级监管方法。它将AI系统划分为四个风险等级：不可接受风险（如操纵人类行为、社会评分系统等）、高风险（涉及关键基础设施、教育、就业、医疗、执法、司法等敏感领域的AI系统，这些系统在上市前和整个生命周期中都必须满足严格的要求，包括数据治理、技术文档、透明度、人类监督、准确性、鲁棒性和网络安全等）、有限风险（如聊天机器人、深度伪造等，这类系统需履行特定的透明度义务）以及最低或无风险（如垃圾邮件过滤器、AI游戏等，这类系统基本不受额外监管，但鼓励遵守自愿行为准则）。《AI法案》同时定义了六种AI系统参与者的角色，分别为提供者、部署者、分销者、进口者、授权代表和产品制造商，并分别规定了所需承担的义务。

《AI法案》具有域外效力，即便是欧盟以外的提供者，只要其AI系统在欧盟市场投放或其输出在欧盟境内使用，也受到管辖。跨国公司以及准备进军欧洲市场的中国投资者，需要考虑AI合规可能带来的较大的成本和技术压力。

三、美国：鼓励创新与多方监管并行

相较于欧盟的统一立法，美国在联邦层面尚未出台一部统一和全面的AI监管法案。联邦政府在监管方面采取了更加分散和灵活的方法。通过相关监管机构针对特定行业发布指导原则和政策，例如食品和药物管理局（FDA）发布了《使用AI支持药品和生物制品监管决策的考量（草案）》对使用AI在药物研发中应用提出了指导政策；卫生与公众服务部（HHS）发布了《AI战略计划》，侧重于公共卫生、医疗服务、医学研究领域促创新和安全负责任地使用AI。

此外，联邦政府强调企业自愿遵守原则性指导，例如国家标准与技术研究院（NIST）发布的AI风险管理框架，鼓励企业在开发和部署AI时考虑包容性、透明度和安全性问题。这个风险管理框架旨在为不同行业中与AI系统相关的风险管理提供标准化方法，企业可以根据企业的情形灵活适用该框架，确保AI使用的透明度和负责任，保护公民自由和隐私权。

四、东南亚与中东：新兴市场的多元实践

随着中国AI企业加快国际化步伐，东南亚和中东等新兴市场的监管环境日益受到关注。

作为东南亚地区平衡创新与治理的典范，新加坡于2019年1月发布了《AI模型治理框架》（Model AI Governance Framework），并于2020年1月对原有的框架进行了扩展。同时针对AI在金融、医疗等具体行业的深度应用，新加坡发布了FEAT原则、Veritas框架、[9] AI治理测试框架和软件工具包等软性合规指引。2024年5月，新加坡进一步推出了《生成式AI治理框架》（Model AI Governance Framework for Generative AI）专门针对大语言模型等生成式AI系统，要求开发者和部署者采取负责任的态度。

在中东地区，阿联酋于2022年12月发布了《AI伦理原则与指南》（AI Ethics Principles and Guideline），近期发布的《国家AI战略2031》（UAE National Strategy for Artificial Intelligence 2031）旨在实现阿联酋在全球AI领域的领先地位，战略明确了人才吸纳，AI/数据基础设建设，AI监管等目标。2025年，沙特阿拉伯通信、空间和技术委员会（CST）正式发布《AI中心法》（Global AI Hub Law）草案，旨在确保沙特阿拉伯作为全球数字枢纽的地位和促进AI的创新和发展。

9. 《埃森哲加入Veritas联盟，推动金融服务领域负责任的人工智能和数据分析发展》，埃森哲，2021年7月13日，https：//newsroom.accenture.sg/asia-pacific/news/2021/accenture-joins-veritas-consortium-to-drive-development-responsible-ai-and-data-analytics-in-financial-services。

总体而言，上述AI监管差异使得企业在技术开发、运营部署和战略规划中面临多维挑战，不仅需要企业构建跨部门协同的合规机制，更需要在AI架构设计阶段引入“可解释性”与“地域适配”等核心原则，从源头治理确保模型满足“合规即用、风险可控”的要求。

企业AI共建式治理

在AI领域，尤其是生成式AI，核心技术、训练数据、模型能力和社会影响主要源于企业实践，企业打造负责任AI服务有利于提高企业在市场上的竞争优势，管理和降低风险以及更能创造长期价值。因此，企业已成为AI治理体系的积极建构者，具体可从以下四个关键方面展开：

一、构建并实施内部负责任AI框架

企业需率先建立一套系统性的AI治理机制，具体在于：

组织保障：设立专责团队，如AI伦理委员会或负责任AI办公室，跨部门协调技术、法律、合规、公关等。

制度规范：制定覆盖模型开发、数据使用与管理、内容生成、员工使用AI等全环节的内部规则。

技术支撑：部署模型监控、偏见检测、内容审核等技术管控工具。

旷视科技早在2019年就成立了AI道德委员会，[10] 在2020年成立AI治理研究院并发布《AI应用准则》。其《企业负责任的人工智能实践》报告提出，[11] 企业负责任AI实践框架可从AI全生命周期的ESG理念、企业内部多层治理设计（如伦理审查、问责制）以及内外双循环的治理网络（政府监管、行业服务、企业自律、公众参与）三个层次进行设计。

二、推动行业标准与自律规范建设

企业应积极推动相关行业规则制定，创建良好的生态体系：

参与联盟共建：加入或主导行业组织（如产业联盟）。

定义关键标准：协同制定“高风险AI”界定、“安全红线”、“责任边界”等核心标准。

倡导透明实践：推动算法备案与透明披露，在法规空白期，主动试点信息透明和风险报告机制。

例如，腾讯发布大模型安全与伦理报告，[12] 将大模型生命周期解构为数据处理、训练部署、应用落地三阶段，系统性剖析了各环节风险（如数据泄露/投毒，模型篡改，AI组组件洞，平台漏洞，硬件漏洞，供应链投毒，事实性错误，隐私泄露，提示注入，对抗攻击，伦理安全，大模型滥用等）。此外确定了大模型的安全目标，包括数据安全稳定，算法可解释，模型决策可信，应用合法合规以及环境安全可靠。

为实现负责任AI即有意识地采取相关行动来设计、部署和使用AI来创造价值和建立信任，埃森哲设立了负责任AI七大原则，包括以人为本，公平性，透

10. 《负责任的AI》，旷视科技，https://www.megvii.com/Responsible_AI。
11. 《旷视与中信所联合发布<企业负责任的人工智能实践>报告》，旷视科技，2023年1月11日，https://www.megvii.com/news_detail/id/258。
12. 《腾讯发布大模型安全与伦理报告：以负责任AI引领大模型创新》，腾讯，2024年1月29日，https://cloud.tencent.com/developer/article/2384570。

明，可解释与准确性，安全性，责任制，合规，数据安全，网络安全，以及可持续发展。另外，埃森哲还与斯坦福大学合作研究，对企业负责任AI成熟度展开深入剖析，为企业发展提供指引。

三、推动生态伙伴共同履责

AI治理责任需要价值链传导：

设定合作方准入要求：通过供应链和生态体系传导，对开发者、供应商、合作伙伴明确合规要求；

建立供应链问责机制：确保数据提供商、算力平台等共同承担治理责任；

赋能中小企业：提供工具与指南，协助上下游建立基础AI使用规范。

例如，阿里巴巴强调“负责任的技术”理念，包含“守己”（坚守科技伦理和安全底线）和“利他”（促进行业发展，赋能千行百业）两层含义。在2024年发布的《大模型技术发展及治理实践报告》，阿里强调了安全性能是大模型持续发展的基本要求，并通过多元融合的治理规范，多角色全周期的治理措施，多方协同的治理环境来强化数据保护和隐私安全，提升系统可靠性和抗攻击能力，确保技术的健康发展。

联想集团早在2023年设立负责任AI委员会，探索绿色AI、普惠包容AI及负责任AI等产业实践，并在2024年联合上海交大AI研究院、平安科技、西门子医疗等共31家头部AI企业和机构共同加入并启动了《人本智能发展与治理倡议》。秉持“人本智能”价值主张，联想集团积极与合作伙伴协作，共同推动AI向善，目前已在多个行业领域推动实践案例的落地。

四、践行AI向善理念，创造社会价值

企业不仅要避免AI的潜在安全隐患，更应通过AI为社会创造积极影响，包括通过AI推动社会公平、环境保护，支持公益创新生态等。

联想在江西省修水县何市镇中心小学建成全国首个乡村小学“AI未来科技馆”，以新IT为技术底座、用普慧算力赋能乡村儿童体验AI语音作画、普慧算力应用场景及原理、AR元宇宙、AI体育及智慧教育等前沿科技成果。2024年10月，联想集团携手贝智基金，在香港爱培学校设立了香港首个AI辅助学习中心，运用AI等技术助力有特殊教育需求的儿童。

通过公益咨询服务，埃森哲运用AI和增强现实技术，为江苏太仓中德融创工场开发“埃森哲产业工人助理”（AIWA），它将易于理解的文字和图像、清晰可视化的导航、自动纠错的心智模型和多种提醒方式相结合，为心智障碍员工提供实时的零件装配引导，积极支持残障员工的学习和生产，从而提高工场的生产效率和质量，赋能特殊群体。

生成式AI的深刻变革蕴含着巨大潜能与复杂风险。构建负责任AI生态系统，需要政府、企业、学术界和公众的协同共治，在鼓励技术创新的同时，确保其发展符合人类共同价值和长远福祉。负责任AI不仅是技术发展的必然要求，更是构建可信数字未来的基石。通过持续完善政策框架、深化企业实践、创新治理工具和加强全球协作，我们有望在享受AI技术带来巨大便利的同时，有效管控其潜在风险，让AI真正成为推动社会进步、增进人类福祉的强大力量。这需要我们保持警醒，持续对话，并勇于实践，方能在日新月异的AI时代行稳致远。

作者

胡定宇
埃森哲大中华区技术服务事业部高级总监，负责任AI主管

徐浩
埃森哲负责任AI业务经理

联想如何打造负责任AI

访 胡定宇、高文胜、徐浩

编 王若霈

随着生成式AI技术的突破性发展，AI的伦理与治理逐渐成为全球关注的焦点。负责任应用人工智能的理念应运而生，成为全球科技界亟待解决的重要课题。领先科技企业纷纷制定相应的指导原则和标准，以期在高速创新与风险防控之间取得平衡，从而构建一个可持续、可信赖的AI系统。

秉承“人本智能”的理念和“Smarter AI for All”的愿景，联想集团在AI技术研发、产品与服务的设计、应用以及与外界交互中，不断强化数据安全、隐私保护和合规机制，确保技术成果惠及每一个用户。

近期，《展望》对话联想集团副总裁、中国区总法律顾问高唤栋，联想集团ESG与可持续发展负责人、联想中国平台ESG委员会秘书长王旋，以及联想集团资深法律顾问、联想中国平台AI合规委员会秘书长梁潇，深入探讨了联想在负责任AI领域的战略思考、实践经验以及未来规划。

《展望》：您如何理解负责任AI?

高唤栋：在AI快速发展的时代，负责任AI已经成为全球讨论的焦点。它涉及伦理、安全、全球治理、企业实践等多个层面。因此，我希望从人类、国家、企业、部门四个角度，探讨AI发展带来的挑战和责任。

从人类层面来看，AI的进化前所未有地挑战了人类的生存和主体性。过去，科技的进步往往以增强人类能力为目标，而AI的发展，则让人们开始思考机器是否会对人类生存构成威胁。这种担忧并非空穴来风，我们已经看到AI在某些领域比人类表现更优秀，而且AI具备自我学习和演化的能力。因此，AI伦理问题是负责任AI讨论的核心。

在这个背景下，国家层面对AI伦理的关注至关重要。AI发展需要国家层面制定伦理规范和法律框架，确保AI技术的发展始终在人类控制之下。同时，随着AI进一步加速全球化的进程，在未来，负责任AI不仅是一个国家内部的问题，而会成为一个跨国界的议题——全球AI规则的制定、数据跨境流动的管理、技术标准的统一等问题，将成为各国政府、企业和国际组织共同关注的重点。

企业层面，联想提出了“人本智能”的核心价值观，其本质是以人为本，让AI服务于人类，提升我们的生活质量，提高工作效率。这与埃森哲提出的负责任AI概念一致。AI不是为了取代人，而是为了增强人类的能力。联想推出的AIPC产品，在AI应用中强调数据安全和隐私保护，正是这一理念的体现。

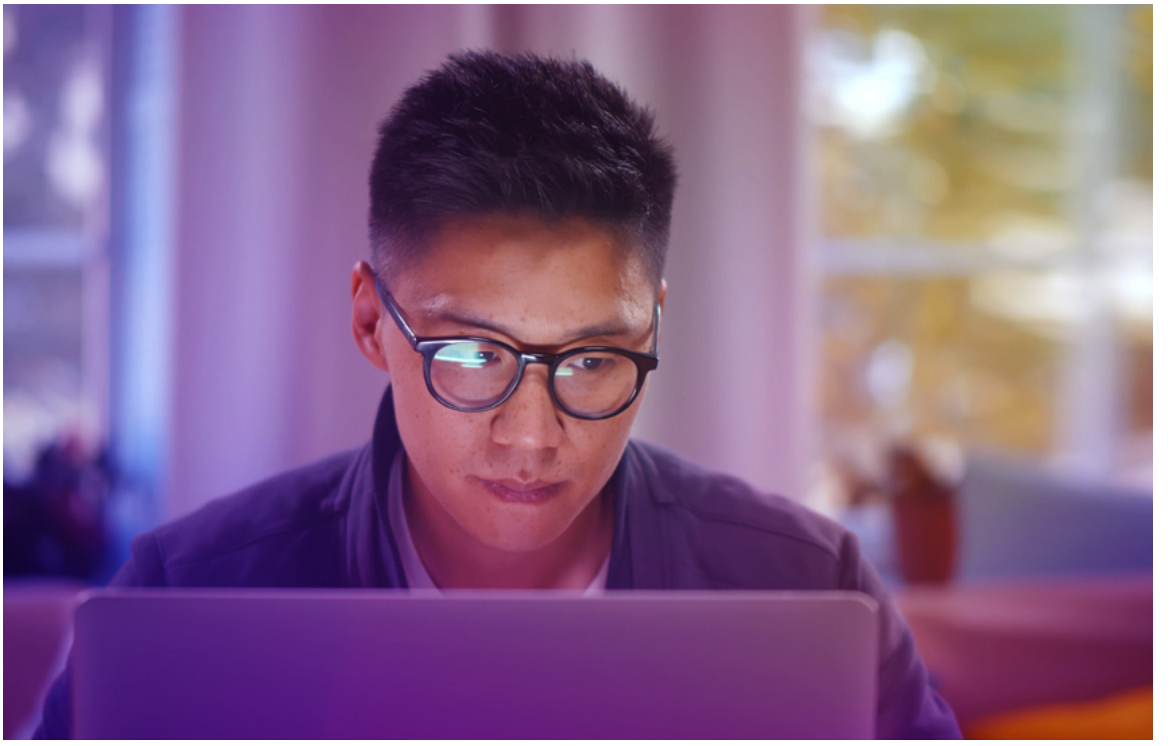
然而，在企业实践中，负责任AI的落地并不容易。如何在企业内部建立AI发展与治理体系，如何在AI发展过程中确保合规性，这是所有企业都面临的挑战。为了更好地推动负责任AI发展，联想成立了AI合规委员会，它并非传统的管理机构，而是一个无严格汇报关系的虚拟组织。

最后，在部门层面，我认为AI不是目的，而是一种提高效率、增强能力的工具。五年前我加入联想，第一项任务就是推动法务业务的数字化，推出电子签章等智能工具。生成式AI的爆发让我意识到必须加速AI在业务中的深度应用。目前，联想法务团队正在与法律技术公司合作，开发智能合同审查，让法务团队能够将更多精力投入更具有挑战性的复杂事务，而非重复性的文书工作。我们还在建设法务知识库与智能体，并探索如何将这些能力外化，为行业提供更高效的法律支持。

王旋：作为ESG（环境、社会与治理）负责人，我认为负责任AI的核心在于识别机遇和防范风险。AI可谓是一把双刃剑——它在提升生活便利性的同时，也带来了诸多潜在风险。尤其是当前AI技术的高速发展，使得这些风险的影响范围更广，迭代速度更快，因此，我们必须以动态的视角持续关注，并积极采取应对措施。

关于负责任AI，我一直在思考一个关键问题：负责任AI的边界究竟在哪里？回顾以往的“AI向善”理念，其中“善”的定义始终较为模糊，不同群体、行业或场景的理解各不相同。因此，我认为负责任AI的落地不能是一刀切的标准，而应基于行业属性和具体应用场景进行细化。我们相信，随着社会对负责任AI的理解不断深化，未来负责任AI的实践会更加精细化，形成更具针对性的规范和框架。

作为行业领先企业，联想已经建立了自身的负责任AI七大原则，并希望通过这一框架引导AI的负责任发展。同时，我们也积极呼吁行业内的专家和从业者共同探讨负责任AI的实践路径，以协作的方式应对AI带来的多维度风险，共同推动行业健康可持续发展。



《展望》：刚刚您提到联想的AI合规委员会，联想如何借助这个组织推进负责任AI，在过程中会遇到哪些挑战？

高唤栋：AI合规委员会有三大核心职能：

第一，教育与认知统一。委员会的一个重要职责是提升组织内对AI治理的共同认知，确保大家在价值观上达成一致。每次会议我们都会分享外部趋势洞察，包括国内外立法动态，并通过对比不同国家的AI治理政策，分析其背后的价值观取向，这种认知对于企业在全球化经营中制定合规策略至关重要。

第二，打破部门壁垒，提高协同效率。在AI时代之前，各业务部门往往按照垂直领域划分，形成较为独立的组织架构。然而，AI的应用打破了传统产品线的边界。例如，在AIPC的研发过程中，我们发现PC、平板和手机产品线各自采用不同的大模型，并与不同的第三方合作，技术成熟度和风控策略方面都存在较大差异。在委员会设立之前，我们的团队需要分别与每条产品线逐一沟通。而委员会成立后，我们能够快速总结PC产品线的经验，并高效复制到平板和手机产品线，大幅提升跨部门协作效率。这种模式不仅降低了成本，也加快了RAI治理的落地速度。

第三，培养复合型人才，构建负责任AI时代的人才体系。在人才培养方面，AI时代的人才需求更为复杂，我们需要的已不只是传统的“T型人才”，即具备深厚的法律专业能力，同时拥有商业认知的人才，而是“π型人才”，即在法律和商业之外，还要具备技术能力。AI治理不仅涉及法律法规，还涉及技术理解和业务应用，因此，我们需要培养能同时理解法律、技术和商业的复合型人才。

AI合规委员会的核心目标之一，就是打通法律、技术和业务这三个原本相对独立的领域。虽然这条

路仍然充满挑战，但至少，我们迈出了关键的一步，并已发挥了实质性的作用。

梁潇：在推进委员会工作的过程中，我们也遇到了不少挑战，主要体现在两个方面：

第一，跨部门协调。作为一个由法务部牵头的跨部门虚拟组织，推动负责任AI治理并不是一件容易的事。不同部门对负责任AI的理解并不完全一致，这其中需要大量沟通协调工作，尤其是在推动各部门达成一致认知方面，需要付出很大努力。而这种认知统一，是委员会能够有效运作的前提。

第二，平衡技术发展与合规要求，优化执行路径。负责任AI治理的另一个难点在于合规资源的投入，在业务端，资源往往更倾向于投入技术研发，而合规工作需要业务和技术团队额外投入精力，因此找到合适的平衡点至关重要。

以AIPC上市为例，某些关键的合规程序及技术管控措施需要在产品上市的紧迫时间表内强行落地。这对业务部门来说，无疑是增加了额外的工作。当时，我们进行了多轮沟通，召开了大量会议，最终在“确保合规底线”的前提下，达成了一个折中的解决方案——先满足最基本的合规要求，确保产品不会触碰法律红线，并采用分阶段推进的方法，按照时间表逐步完善负责任AI治理措施，而不是一次性加诸所有技术管控要求。

通过这种方式，我们既保障了产品上市进程，又在合规框架内为负责任AI治理争取到了更多落地空间。这种动态平衡的思维模式，成为了我们在AI治理过程中不断探索和优化的方法论。

比如，联想的AIPC中搭载了业界首个端侧生文大模型，按照中国监管要求，在面向消费者提供服务之前，我们需要完成大模型上线备案并落地相关合规要求，这是合规底线。当时AIPC上市的时间表非常紧张，备案又需要完成大量的申报文件的制作、技术保障措施研发落地以及安全检测工作。因此，我们跨部门的团队每日加班加点、互相配合，最终在AIPC正式发布之前，成功完成了大模型上线备案。

高唤栋：在这个备案过程中，我们的法务团队也从未如此深入地了解技术，这也再次验证了我们先前提到的π型人才的重要性，这些复合型人才不仅要具备深厚的专业知识，还能够跨领域理解和沟通，这对于推动负责任AI的落地至关重要。

《展望》：当技术部门遇到AI风险合规问题时，他们会询求法务团队的建议，我们如何引导技术团队解决问题？

高唤栋：我的核心观点是，AI带来的风险和合规问题，最终仍需要依靠技术手段来解决。这背后的逻辑在于，许多法律和合规挑战本质上是技术驱动的，而单纯依赖法律手段无法从根本上消除这些风险。

举个例子，在电商时代，精准广告投放依赖大数据分析，而这些数据往往涉及消费者的个人隐私。这种数据驱动的广告模式是电商的生存之道，直接关系到商业模式的可行性。如果完全禁止企业使用客户隐私数据，那么整个精准投放体系将无法运作，进而会影响电商业务的核心盈利能力。如何在不泄露个人隐私的前提下，仍然能实现精准投放？答案在于技术解决方案——隐私计算。隐私计算通过数据脱敏等技

术手段，去除或加密敏感信息，使得企业能够在保护用户隐私的同时，依然可以利用数据进行广告投放，既符合合规要求、又不影响商业运作，可以实现“双赢”局面。

在AI时代，企业面临的合规挑战更加复杂，例如AI模型的透明性、公平性、数据安全等问题。我始终坚信，这些问题的最佳解决方案依然来自技术本身，而非单靠法律。深入理解隐私计算这个案例，让我意识到，面对AI时代的新挑战，我们必须打破传统法务、技术和业务的边界，以跨学科的思维方式去推动企业的AI合规治理，这也是为什么我特别强调企业需要培养π型人才。只有这样，我们才能真正构建起可持续的AI治理体系，让技术既能创造商业价值，也能符合社会责任和法律合规要求。

《展望》：您如何看待负责任AI未来在中国的发展？

王旋：我认为，在当下讨论负责任AI具有非常重要的意义。从全球视角来看，目前尚未有任何一个国家或地区在负责任AI治理上占据绝对的话语权，这意味着中国有极大的机会走向领先地位。而这种可能性不仅源于技术的发展，更得益于国家政策的支持、社会需求的推动以及整体环境的成熟。

首先，从政策环境来看，中国在推进前沿技术治理方面已经展现出高度的战略前瞻性。例如，ESG和“双碳”目标已经成为国家层面的重要议题，并逐步渗透到企业经营和社会发展的方方面面。同样，负责任AI也正在成为一个不可忽视的政策方向。政府的重视、法规的完善以及企业责任意识的提升，正在为负责任AI的落地提供强有力的支撑。

其次，从AI技术发展的角度来看，我们正处于一个关键的时间节点。我还记得多年前，我曾参加北京一所高校的会议，会议的核心议题便是AI的伦理与合规问题。尽管当时大家都在讨论AI治理的必要性，但由于技术尚未成熟，我们只能等待时机。而如今，生成式AI的突破性发展已经让AI全面进入商业化和社会应用阶段，这正是负责任AI发展的最佳时机。对于头部科技公司而言，推动负责任AI不仅是技术发展的必然要求，更是企业应当承担的社会责任。

最后，从全球竞争格局来看，中国的负责任AI治理具备先发优势。相较于其他国家和地区，中国政府在AI治理上的行动速度更快、落地举措更明确，而全球尚未形成统一的负责任AI标准。这意味着，中国有望成为全球RAI治理的探索者和推动者。

基于以上因素，我对负责任AI在中国的发展前景充满信心。无论是从技术趋势、政策环境，还是全球竞争格局来看，中国都具备成为负责任AI引领者的条件。我相信，各方的共同推动下，负责任AI将在中国迎来快速发展，并为全球AI治理贡献中国智慧。

访谈者

胡定宇
埃森哲大中华区技术服务事业部高级总监，负责任AI主管

高文胜
埃森哲大中华区法务及政府事务总裁

徐浩
埃森哲负责任AI业务经理

“正确做事”与“做正确的事”

专访亿滋中国首席信息官欧庆国

访 胡定宇、徐浩

编 王朓灵、王若霈

作为全球零食行业领先企业，亿滋国际（Mondelēz International，以下简称亿滋）旗下拥有奥利奥、趣多多等家喻户晓的品牌，2024财年营收达到364.4亿美元。面对不断演进的消费需求与技术变革，亿滋秉承着“先正确做事，再做正确的事”这一关键实践原则，积极拥抱AI，并在AI应用方面构建了一套全面且系统的战略路径，涵盖风险评估、技术部署、人才发展等多个维度。

《展望》近期专访了亿滋中国首席信息官欧庆国，围绕AI如何真正赋能业务、塑造组织文化、强化合规治理等关键议题，深入探讨了亿滋AI应用实践中的前瞻思考与落地路径。

《展望》：目前亿滋尤其是中国区已经上线了哪些AI应用？

欧庆国：目前，亿滋中国已在多个内部场景中引入AI应用，主要包括两类：一是支持政策合规的“智能助手”，二是面向一线业务的职场AI工具“亿想家”。

首先是“智能助手”，主要用于内部员工对政策、法规的快速查询与合规判断，涵盖法律、人力资源等领域。它不涉及复杂的判断逻辑，也不会直接影响客户或消费者，更侧重于帮助员工高效获取信息、提升日常工作效率。

虽然这类工具对消费者来说是“隐形”的，但其价值在于操作简单、易于上手，有助于员工逐步理解AI技术的实际意义和使用场景。我们认为，如果不主动拥抱这些新工具，企业和员工都可能面临被技术浪潮淘汰的风险。因此，我们不仅提供了相关培训课程，还在组织架构层面推动员工能力提升，特别是在减少重复性、低价值工作的同时，更关注如何将AI与员工职业发展紧密结合。

另一个是“亿想家”，一款专为业务场景打造的AI助手。它的核心目标是打破信息孤岛，整合企业内的业务知识，为员工提供实用、可执行的行动建议。与“智能助手”不同，“亿想家”更加贴近具体业务流程。例如，当员工提出“拜访经销商前需要准备什么”这样的问题，系统不仅会给出基本的政策提示，还会深入到业务层面，如提醒用户检查库存水平、分销状态、历史销售数据及门店当前表现。系统还能基于数据提供精准建议，比如建议补货的SKU数量等。

此外，“亿想家”还集成了图像识别技术，可用于分析门店环境、客户画像和细分市场表现，帮助一线人员进行更全面的判断和决策。这些功能看似直观，但背后依赖大量的数据训练和持续优化，以确保信息的准确性与应用的实用性。

总的来说，这两款AI工具分别聚焦于提升组织效率和增强业务洞察，都是我们推动AI融入日常运营、释放员工潜力的重要实践。

《展望》：亿滋在规划AI用例时，如何系统评估风险？

欧庆国：在亿滋，我们始终坚持“先正确做事，再做正确的事”这一理念来规划AI用例。为了规避技术实施过程中的潜在风险，我们建立了三重风险识别与管理机制，确保AI应用与企业战略方向高度一致。

第一个机制是战略规划与治理框架。在这一层面，我们制定了清晰的AI路线图规划——每一个AI用例都会被纳入整体的数字化战略蓝图中，明确短期目标（如当年交付成效）与中长期战略价值，确保每一步都紧贴业务核心需求。另外，我们成立了由法务、HR、IT及战略部门等组成的数字化治理委员会，确保AI从概念设计到项目落地的全过程合规、透明，并通过高层承诺确保资源保障。同时，我们还鼓励区域团队反向提出本地化需求，真正实现“自上而下”与“自下而上”的融合推进。

第二个机制是深度嵌入组织运营模型。AI用例的实施不应是“技术孤岛”，而要真正融入组织的日常运营逻辑。例如，我们将AI招聘助手嵌入人才发展体系，提升匹配效率；又或是通过AI进行销售预测，动态优化各区域的资源配置。这种“深融式落地”（Deep Integration）确保AI用例具备实际业务价值。

第三是法规合规作为底线红线。采取法律风险零容忍机制——我们与法务团队建立了联合审查机制，在AI项目启动前就完成全流程的合规审查。项目团队必须对数据采集、模型训练、结果应用等环节出具清晰的合规证明，确保AI项目在“合法合规”基础上运行。另外，全球统一与本地法规双轨并行，即在满足全球合规标准（如GDPR、数据跨境传输等）的基础上，也充分考虑中国本地法律（如《数据安全法》《个人信息保护法》）的特殊要求，进行本地化策略调整。

《展望》：我们观察到一个现象：很多企业虽然在积极推进数字化，并开始使用AI产品，但往往并不清楚内部AI应用可能带来的风险。亿滋是如何看待这个问题的？

欧庆国：企业在推动数字化转型和应用AI技术的过程中，如何识别和管理内部AI风险，已经成为一个关键课题。我想从四个方面来谈谈我们在这方面的理解与实践。

第一，AI风险管理不应是事后的补救，而应成为企业组织架构和运营管理的基础性模块。如果企业高管对内部AI的使用情况缺乏了解，这本身就是一种风险。比如，在签订技术合作协议时，若未经过充分的合规与能力评估，可能导致数据违规使用、能力建设不足等问题，从而给企业带来风险。

第二，AI不仅是一项技术，更应被视为企业核心能力的一部分。早在正式接触负责任AI概念的三年前，亿滋就已在全球及中国区启动了相关的内部培训。这一过程的目标是让管理层和业务团队了解负责任AI的本质，并探索如何将其与实际业务场景相结合。此外，AI应用通常难以在短期内带来可量化的成果，因此企业必须具备长期投资的战略眼光。为此，亿滋构建了统一的AI平台和数据库体系，并将AI嵌入到数据治理框架中，打造面向未来的可持续发展模型。

第三，建立标准化的风险评估机制。我们在内部建立了一套AI应用的“风险检查清单”，帮助各团队系统性识别潜在风险。这份清单涵盖多个维度，例如：该AI应用是否明确带来了商业价值？是否符合总部在信息安全和隐私保护方面的要求？所使用的AI用例是否已有内部成功实践，是否可复用模板或优化已有设计？通过这些标准化工具，我们能更有效地在项目初期就筛查潜在风险，提升实施效率并控制可能的隐性成本。

第四，面向中国市场的本地化响应。中国市场的AI环境具有特殊性，比如大语言模型的使用受到严格限制。因此，我们在推进AI项目时必须坚持“通用性优先”的原则，确保工具具备足够的适应性和弹性。这既是一个基本要求，也是一种能力体现。亿滋在与全球总部保持战略一致的同时，也会根据中国市场的变化节奏做出快速反应——因为在中国，速度往往决定成败。但无论市场变化多快，我们在AI架构设计上的原则始终保持与总部统一，确保整体协同和治理效率。

通过这些机制，我们希望实现对AI风险的可视化、可控化与前置管理，确保AI真正服务于业务，而非成为新的负担或隐患。事实上，亿滋也在积极探讨是否将AI从现有数字化架构中进一步独立出来，作为单独的能力模块加以建设。这不仅有助于优化AI资源

管理，也有利于其在企业中的长期价值实现与可持续发展。

《展望》：在引入和应用AI的过程中，亿滋重点关注哪些关键环节？

欧庆国：在AI项目的设计与实施过程中，我们通过标准化的“检查清单”对每一个用例进行全流程监管，重点聚焦于三个风险维度：治理（Governance）、安全性（Security Risk）和架构完整性（Architecture Integrity）。核心机制包括：

治理机制确保战略对齐。在项目启动阶段，我们会明确AI用例是否已在其他市场有成功先例。如果有，我们会快速借鉴已有经验，避免重复劳动。若为新用例，则需严格审查其战略对齐程度和业务落地可行性。

架构与能力评估。我们会判断该用例是否应基于内部能力或外部支持来开发，同时明确其所需的数据、系统与人力资源配置，并制定标准化实施清单，让所有相关人员清楚AI对业务的实际价值贡献。

将AI视为业务“核心资产”而非“赠品”。我们坚信，如果一个AI项目无法被真正应用并嵌入到业务流程和组织能力中，那它就很难产生真正的影响力。推动AI落地，必须得到业务和财务双线支持，让它成为企业能力体系的一部分。

《展望》：您如何理解AI透明度在组织架构中的作用？

欧庆国：AI透明度的核心在于技术不仅要具备去除偏见的能力，更要体现出高度的诚信。这意味着AI的底层逻辑必须建立在真实、客观的数据基础之上，而这也是AI在组织中被广泛信任的前提。

需要强调的是，公平性并不是AI天生具备的能力，而是由人类在算法设计和模型训练过程中构建的。因此，组织在推进AI应用时，必须确保算法逻辑和数据来源的透明，防止因隐藏的偏见或数据不公而带来决策风险或员工信任缺失。在亿滋，我们将AI透明度的价值，具体落地在以下几个方面：

量化员工能力，实现发展可视化。借助AI，我们能够更系统地分析员工的能力结构与成长潜力，量化指标帮助组织精准识别人才发展路径。这种“能力透明化”，不仅优化了人才配置，也让员工更清晰地了解自己的职业成长方向和发展空间。

推动组织公平性，保障机会平等。通过数据驱动的决策方式，AI能够有效减少人为偏见。例如，在晋升或资源分配上，系统会基于客观表现而非主观判断，确保每一位员工的努力和成果都被公正看见。

提升运营效率与战略协同的平衡能力。AI的透明性还能帮助企业识别当前运营中的瓶颈与机会，明确哪些领域更需要提升效率，哪些板块应聚焦于收入增长。企业在开源与节流之间找到动态平衡，避免一味节流导致人才流失或组织活力下滑，从而实现真正的可持续发展。

总的来说，AI透明度不仅是技术层面的标准，更应成为组织文化的一部分。只有当技术应用足够真实、客观和公平，企业才能在“效率”与“公平”之间找到最佳结合点，实现长期价值的最大化。

《展望》：亿滋在引入大语言模型时采取了哪些训练手段?

欧庆国：在引入大语言模型时，亿滋不仅进行了基础的提示工程和模型微调，还从业务落地和人才能力两端入手，实施了更具针对性和可持续性的训练手段，主要包括：

一、本地化模型训练，贴合市场需求。我们在中国各供应链中部署了数据科学家团队，专门负责对大语言模型进行本地化的二次训练。这一做法确保模型能更深入地理解中国市场的语言习惯、数据特性以及业务场景，从而在日常运营中产生更高的精准度和业务关联性。

二、内部能力构建，减少对外依赖。亿滋坚定地认为，与其依赖外部供应商，不如构建起自身的AI能力生态。因此我们在组织战略层面持续投资数据科学人才，从职业路径设计、技能提升计划到支持资源配置，打造了一个完整的人才发展体系。我们不只是“雇佣”数据科学家，更是在构建一套支撑企业长期发展的“AI人才操作系统”。

通过这两方面的努力，亿滋不仅提升了大语言模型的实际业务适用性，也确保AI能力能持续内生化，服务于更广泛的场景与创新探索。

《展望》：您提到AI在释放员工潜力方面的应用，亿滋在人力资源管理领域如何使用生成式AI?

欧庆国：在人力资源管理领域，生成式AI正逐步重塑传统流程，不仅提升效率，更推动企业从“依赖外部人才”向“构建内部能力”的深层转型。亿滋在这方面展开了两项核心实践：

第一个是基于生成式AI的人才画像与能力量化，具体来说有以下几个具体实践：

简历筛选与能力建模：AI借助自然语言处理技术，对简历内容进行深度解析，结合岗位要求，自动生成结构化能力标签，比如“数据分析能力”、“跨文化沟通能力”或“AI工具适应能力”。这类画像不仅覆盖硬技能，还能通过行为数据挖掘出软技能，例如决策偏好或学习敏捷度，从而为人才选拔提供更立体的参考视角。

差异化能力评估：针对不同城市的岗位，AI能够识别区域市场之间的细微差异。比如苏州与上海的城市经理，即便地理相近，消费者习惯也可能大不相同，从而生成定制化的能力模型，更贴合实际业务场景。

核心能力量化与动态优化：通过分析历史绩效和市场反馈，AI还能量化像“对新技术的适应能力”这样的软性指标，生成可视化的“能力分布图”。基于这类动态评估结果，企业不仅能识别出高潜力人才，还能为其推送个性化的培训路径，帮助他们快速提升关键能力。

第二个是生成式AI驱动的全球经验萃取与个性化赋能，全球市场的成功经验转化为可落地的行动方案，同时弥补个体能力短板。具体实践包括：

跨市场经验萃取与知识图谱构建。AI从全球销售案例中提取出共通的成功策略，并将其结构化为“行动脚本”，供其他区域团队参考。更妙的是，系统还能根据地域差异，自动生成本地化的执行建议，让“借鉴”变得更加高效和实际。

个性化销售支持与晋升路径规划，比如对某位城市经理，AI可以生成“区域专属销售指南”；同时，

根据个人能力画像，帮助其设计晋升路径。举个例子，如果某人擅长数据分析但缺乏品牌管理经验，系统就会推荐“品牌营销”相关的课程知识来帮助其定向学习。

实时决策辅助与场景模拟。AI还能生成各类“假设情境模拟”，帮助员工预演复杂决策情境，增强应变能力。此外，系统会实时结合市场数据，动态调整建议内容，比如在库存告急时，自动生成“SKU补货优先清单”。

总的来说，亿滋正在通过AI推动一场从“人才筛选”走向“能力构建”的系统性转变。从静态简历到动态能力图谱，AI不再把人看成一张简历，而是一个不断进化的“能力资产”。随着数据积累和反馈闭环，匹配的准确度也在持续优化；从经验复用到场景创新，AI不仅是一个知识库，更是一个“行动引擎”，能把全球经验翻译成区域化、可执行的解决方案；从工具赋能到生态协同，未来，AI将更深地嵌入组织文化，不再只是技术工具，而是员工与管理者共同成长的“能力加速器”。

这种转型不仅提升了HR管理效率，更通过生成式AI的“共创能力”，把每位员工都变成了组织创新的参与者，而不仅仅是任务的执行者。

《展望》：亿滋如何构建AI人才文化？

欧庆国：构建AI人才文化是亿滋数字化转型的重要一环，我们采取了多维度的实践策略，旨在推动AI技术与业务的深度融合，同时激发员工的创造力和主动性。我们的策略主要体现在以下六个方面：

一、AI沙盒（AI Sandbox）——实践驱动的学习平台。在IT部门，我们设立了开放的实验环境，让员工可以自由探索如翻译工具、图像生成、合规审查等AI能力。通过亲手实践，员工不仅提升了技能，也能更直观地将AI与自身业务场景进行连接，实现能力迁移。

二、从高层做起的AI认知塑造。我们特别强调从高管层级开始普及AI教育。领导者不仅要了解技术，更要能将AI纳入到战略思考中。比如展示“五秒生成视频”的AI能力，不仅让营销部门受益，也让整个组织意识到AI对各自业务的深远影响，从而推动整个组织更积极地探讨AI相关话题。

三、组织激励。我们定期举办Digital Day与AI Day，结合教育培训与案例分享，让员工接触到来自不同企业的成功AI实践，激发他们的好奇心和创新意愿，从而主动思考如何将AI用于自己工作中。

四、黑客马拉松等创新大赛。我们已举办多场围绕“数字化”的黑客马拉松活动，并计划在未来2—3年内将重点转向AI相关主题。随着AI用例的积累和能力的提升，黑客马拉松将成为推动创新的重要平台，员工可以在真实项目中探索AI的潜力，实现“百花齐放”的创新状态。

五、基础能力建设，支撑长期成长。AI文化的打造不是一蹴而就的，它需要强大的底层支持，包括治理、信息安全、风险控制、架构设计等基础能力的完善。这些“隐形基建”是确保AI长期可用、可信、可控的根基。

六、业务价值导向的AI核心能力转化。最终，我们希望让AI真正成为推动业务增长的“发动机”，而不仅仅是一个技术工具。这包括通过AI提升客户体验、优化运营效率、强化供应链灵活性等，全面赋能企业核心能力建设。

通过这些多维度的举措，亿滋正系统性地推动AI人才文化的落地，让每一位员工都能在AI时代中找到属于自己的成长路径，也让组织在激烈的市场竞争中始终保持技术与人才的双重领先。

访谈者

胡定宇
埃森哲大中华区技术服务事业部高级总监，负责任AI主管

徐浩
埃森哲负责任AI业务经理

AI治理的中国路径与全球对话

专访阿里研究院院长袁媛

访 胡定宇、徐浩

编 王朓灵、王若霈

作为数字化领域的领军企业，阿里巴巴始终将AI治理视为重中之重，并在过去近十年间推动了一系列负责任AI发展举措。这些举措不仅涵盖提升AI安全性与技术稳定性、关注科技伦理并设立专门委员会及发布伦理准则，还有对生成式AI的全面治理。

在本辑《展望》特刊中，我们与阿里研究院院长袁媛展开对谈，听她分享对全球AI治理趋势的观察、中国企业在技术发展与伦理规范之间的平衡思路，以及中国科技企业如何在出海过程中讲好自己的技术伦理叙事。

《展望》：在过去十几年，阿里如何推动和实践负责任AI？

袁媛：阿里的AI治理工作主要通过一个虚拟组织来推动，组织成员有各自的本职工作，从而将AI治理能力融入各自的实际工作中。过去十多年来，阿里巴巴在推动和实践负责任AI方面采取了一系列措施，大致可分为三个阶段：

第一阶段（2011年至2018年）：聚焦AI安全，提升算法的可靠性与稳定性，并利用数据和算法解决电商业务及社会责任中的难题，如打假和知识产权保护。同时支持经营范畴外的算法向善应用，如开发了团圆系统帮助找回失踪儿童。

第二阶段（2019年至2022年）：随着信息茧房和“困在系统里的外卖骑手”等社会问题出现，公众对隐私保护和算法公平的关注度提高，阿里开始以科技伦理为牵引进行AI治理，并于2022年6月成立科技伦理治理委员会，同时设立外部顾问委员会提供咨询建议。委员会发布了六项科技伦理准则，并在多个AI安全领域取得成效。

第三阶段（2022年至今）：生成式AI的快速发展促使阿里意识到新一代AI技术的特点与影响范围有所不同。为实现更全面的治理，阿里将科技伦理治理委员会从首席技术官（CTO）条线调整至集团可持续发展委员会（ESG）下。这一点很重要，意味着负责任AI不仅包含守护合规底线，而且成为企业战略的一部分。同时，为了让科技伦理更好地融入AI科技研发应用的工作流程之中，相关工作主导者仍然是集团首席技术官（CTO），而非主要依赖法务或安全团队的监督。

在实践层面，我们也在推动法务、合规、技术和业务的协同机制。负责任AI不能只依赖一个部门兜底，否则它很容易被视为成本负担，而不是价值投资。我们希望构建一种共同责任的文化，促使所有相关方都能积极参与，共同推动负责任AI的发展。

《展望》：阿里是否有计划推出与负责任AI相关的培训或考试，帮助员工更好地理解负责任AI理念？

袁媛：是的，阿里确实开展了相关的培训和考试，但我们的初衷并不是通过考试来设定合格与否的门槛，更重要的是引导大家通过参与提升对负责任AI的认知，尤其是在技术与伦理交叉的复杂场景中。

我们曾经设计过一个富有启发性的问卷，其中一个问题是：如果一位医生将带有病人信息的笔记本遗忘在出租车上，责任应该由谁来承担？

我们发现，技术团队几乎一致认为医生应该负责，这反映出他们认为专业人士在此类事件中应承担主要责任。然而，这个问题的本质远比表面复杂，比如，医院为何允许医生携带敏感信息外出？出租车司机有没有归还义务？这其实是一个复杂的社会性问题。从事后归责角度来看，大多数人认为医生应承担主要责任。但真正有效的治理目标应当是降低风险概率、减少损害后果，这就需要系统性地考虑制度设计、技术防控与伦理文化等多因素的协同。

基于这些讨论，我们在培训设计中不断加入更多案例式、反思性的内容，鼓励参与者意识到：员工们不仅可能是解决方案的一部分，也可能是问题的一部分。负责任AI的关键不是强调技术人员必须负责，而是帮助他们认识到：不能依赖技术手段来解决一切问题。

《展望》：随着生成式AI的高速发展，您如何看待各国在AI风险治理上的差异与应对策略？

袁媛：各国在负责任AI的原则方面已基本达成共识，然而，这些原则在不同国家和企业的实践中却呈现出显著差异。在国家层面，即便在风险认知上趋于一致，实际面临的风险类型、优先级和应对路径却并不相同。

举例来说，版权问题就是一个典型案例。以图像生成技术为例，在GPT-4尚未具备图像生成能力时，版权争议并不显著；但随着GPT-4o版本这类多模态模型的发布，其能够生成高质量图像，使版权侵权问题迅速成为国际讨论的焦点。而各国对于图像生成所涉内容的原创性、归属权等可版权性问题，依旧存在着显著的差异。

在中国，技术发展被置于高度重要的位置。相较于担忧技术本身的风险，我们更倾向于认为，不发展技术才是最大的不安全因素。这一立场凸显了技术平权的理念，即通过降低AI技术使用门槛来提升其普及率。这种思路不仅体现在推动算力基础设施的建设和降低成本上，也体现在对低代码开发工具、开源社区建设等方面的扶持上。

事实上，斯坦福大学发布的《AI指数》报告显示，中国公众对AI技术的乐观度连续两年位居全球首位。这种乐观态度背后有三大关键因素：

一、数字化红利的切身体验：中国在移动支付、电商等领域的数字化进程迅速，显著提升了公众生活质量，从而增强了社会对新技术的接受度。

二、“技术发展即安全”的主流认知：在许多发展中国家，技术被视为弥补发展差距的重要工具。这与欧美国家，尤其是欧洲更为审慎的技术伦理观形成鲜明对比。

三、治理的系统性与执行力：中国采取了较为细致且可落地的治理路径，例如，在输入端要求使用“清洁语料”、在输出端推动“内容标识机制”，这些做法有助于增强系统的可控性和可信性。这样的治理基础，为AI技术在商业和社会层面的进一步拓展提供了坚实支撑。

值得注意的是，这种认知格局并非一成不变。比如过去一年中，欧洲地区对AI的乐观度上升了10余个百分点。这一变化可能反映了欧盟在全球技术竞争格局中重新评估自身定位，并意识到“技术脱节”带来的战略风险。这也表明，全球范围内围绕AI的认知、政策与市场格局在不断演化。

因此，在科技伦理和负责任AI的整体角度下，我们必须接受这种不确定性，并认识到我们面对的是一个复杂且动态的系统。这个系统不仅涉及技术层面的考量，还涉及经济和社会层面的判断。衡量一个国家或地区在AI治理上的成熟度，不应仅仅依赖于法规的数量，而应关注对风险的共识形成以及多方协作机制的有效性。

从全球企业的角度看，针对AI风险和治理议题，企业可以从以下两个方面积极参与并发声：

一、在合规与伦理框架的基础上，进一步讨论如何在保障安全的前提下最大化AI的社会影响力，实现“有益的创新”与“有效的风险防控”动态平衡；

二、构建全球共识与本地实践之间的桥梁：虽然各国面临的AI风险在类型上有一定相似性，但应对方式各不相同。企业应在尊重差异的前提下，探索制定一套具有普适价值的企业原则和技术底线。

《展望》：阿里研究院在进行出海业务的技术研究时，是否也充分考虑了海外的法律法规和伦理要求？

袁媛：我们在进行出海业务的技术研究时，确实有专门针对海外风险和出海相关的内容，我们会针对不同区域的风险特征制定相应策略。正如我们在负责任AI培训中强调，不能依赖单一角色解决复杂问题，面对国际化过程中的治理挑战，也同样需要系统性、多维度的应对。

具体来说，我们目前在不同市场观察到的关注点差异较大。例如，在东南亚地区，我们尚未观察到显著的伦理诉求，那里的关注点更多集中在小语种问题上，特别是如何帮助商家在低代码多语言环境中顺利出海。相比之下，在中东和欧盟地区，我们注意到在伦理和合规方面存在更为突出的挑战。

为了更好地理解和应对这些差异，我们积极参与了多个国际会议，例如在新加坡举办的中美AI治理合作会议，以及与美国国家标准与技术研究院（NIST）、英国AI安全所（AISI）交流，这些国际交流帮助我们把握全球在AI治理上的实用主义趋势，即不只是停留在原则层面，而是强调AI治理机制的可落地性和可执行性。

今年，阿里内部特别关注两个方向：一是内部反馈机制的强化，比如定向征集项目团队在落地过程中遇到的负责任AI相关问题，进而优化治理机制；二是对本地市场的深入研究，例如在中国，我们会关注数字水印的处理方式及其要求，以及水印对用户体验可能产生的影响。这些研究和讨论有助于我们更好地理解和应对出海业务中可能遇到的各种法规和市场需求。

《展望》：对于中国企业如何做好AI治理，您有哪些建议？

袁媛：我有三个建议：

第一、建立共识与透明对话机制。关键在于让所有相关方能够在一个共同的语境中进行对话。例如在透明度问题上，今年的讨论已经逐步聚焦到“开源协议”和“开源程度”上，这将有助于提升整体的透明度水平。然而，仅靠一个国际组织是难以推动这一进程的，头部企业应主动牵头建立相关机制。同时，AI评估的透明度必须成为实施标准，这要求我们在求同存异的基础上进行合作。我们应该致力于形成一种能够实现多方交互的评估模式，这种交互不仅仅是技术层面的，还应包括设计逻辑和实施过程，共同推动建立人类社会长期认可的共识。

第二、转变对AI治理的认知。我们认为，AI治理的根源问题，其实可以归结为成本与信息不对称问题。当治理成本过高、信息披露不足时，反而容易引发不安全或恶意问题。因此，降低信息成本，不仅有助于提升系统安全性，也能建立信任基础。当然，这种改善是渐进式的，不可能一劳永逸，需要企业持续投入。更重要的是，企业需要摆脱“治理等于负担”的传统认知。AI治理不是被动应对监管的成本，也不是为了提升品牌形象的营销预算，是负责任创新的基本表现。

在实践层面，我们发现“事件驱动”往往是推动共识最有效的方式。企业可以通过主动的调研和培训，在设计阶段尽可能将治理融入现有业务中，让最重要的业务部门承担责任。这样，AI治理才能真正嵌入企业运作，而非停留在制度层面。

第三、主动发声，讲好中国在AI治理中的独特故事。在国际化过程中，我们常常面临不同国家对AI的认知差异，这种差异不应被回避，而应成为交流与共识形成的契机。发出中国的声音，讲好中国的故事，既能展示我们的透明度，也有助于提升全球理解和信任。

从机制推动角度来看，我建议通过行业协会等官方/非官方渠道，主动推动负责任AI议题的行业协作。例如在美国，虽然早期的AI行政令并未延续，但作为行政令基础的七家头部企业的治理共识依然持续发挥影响；再比如美国企业较多影响着联合国AI机构和国际标准化组织议程，其他国家多由政府机构牵头。

因此，企业在AI治理上的角色，不只是规则的遵循者，更应该是规则的建设者。面对技术的高速演进，我们不能仅仅等待监管到位，而应凭借自身在专业能力、市场实践和国际理解上的积累，主动为政府与行业提供建设性意见，这也是企业参与AI全球治理的关键一步。

访谈者

胡定宇
埃森哲大中华区技术服务事业部高级总监，负责任AI主管

徐浩
埃森哲负责任AI业务经理

构建信任，生益无界

文 阿纳布·查克拉博蒂、詹妮弗·杰克逊

AI带来的种种风险正加剧企业的信任危机，而成熟的负责任AI机制能够有效化解危机、建立信任并创造价值。埃森哲调研发现，负责任AI是整个组织中推动360°价值的驱动因素，从运营效率到股东信心，从人才到合规，其影响力已延伸到了每一个角落，而它在创新方面发挥的作用尤其值得关注。

AI作为一项新兴技术无疑拥有巨大的变革潜力，但企业的应用步伐却相对谨慎。在《财富》500强企业中，超过56%的企业在2024年度报告中将AI列为了“风险因素”，而2023年该比例仅为9%。[1] 因此，企业迫切需要推进负责任AI，来弥合AI信任缺口。

埃森哲与亚马逊云科技（AWS）针对全球15个国家和地区、21个行业的1000多位高管开展调研，结果显示，负责任AI不仅能够有效降低合规风险，更能成为价值创造的新引擎。数据显示，充分部署负责任AI系统的企业，其AI相关业务收入预计将实现18%的年均增长。这意味着负责任AI先行布局者有望提前抢占市场红利。

搭建信任基石，启动良性循环

信任是所有AI战略的基础，同时也是开启价值创造的关键。对于那些想要积极拥抱AI的企业来说，首先必须使员工信任这项技术。一方面，企业要确保员工相信AI工具能提升其业务能力；另一方面，企业也应帮助员工了解AI风险，助其做好充分准备。

内部信任是企业获得外部信任的基石。当前，员工、合作伙伴、客户、股东和监管机构等利益相关方正在发出日益高涨的呼声，要求企业负责任地使用AI。企业若无法在应用AI方面展现出值得信赖的形象，就可能错失获得全新客户、细分市场和合作伙伴的机遇。

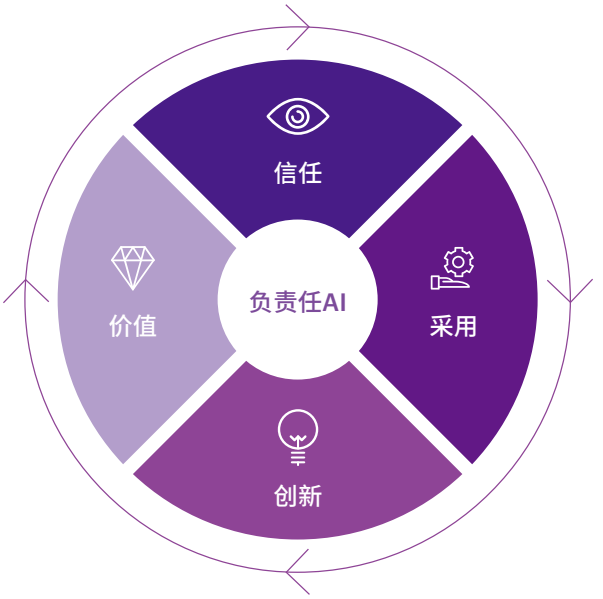
想要搭建信任的桥梁，离不开成熟的负责任AI治理框架。埃森哲调研数据显示，82%的企业相信，将自身打造负责任AI的举措告知员工，可以显著提升员工对AI的信任度，激发其创新热情。

这意味着，当企业投资于负责任AI治理框架，并确保自身的AI系统具备可解释性、公平性和公正性时，信任水平就会迅速提升，这反过来又将推动AI应用以及企业创新水准的升级，而创新正是价值创造和业务增长的关键引擎（见图一）。

由此可见，成熟的负责任AI框架，是开启企业价值创造良性循环的关键。成熟的负责任AI系统会配备专门资源，用以持续评估当前及未来的风险，并根据新的情况，主动进行风险管理。此外，成熟的负责任AI还能主动与外部利益相关方、监管机构和价值链合作伙伴积极沟通，以建立信任关系，实现协同效应。

1. 《财富500强公司标记AI风险的数量激增473.5%》，财富，2024年8月18日，https://fortune.com/2024/08/18/ai-risks-fortune-500-companies-generative-artificial-intelligence-annual-reports/。

图一 负责任AI如何构建信任 开启价值



创造360°价值

研究表明，与AI相关的治理事件如算法故障和数据泄露，数量逐年增加，2022年至2024年两年间增长32%，相比2013年以来增加了20倍。[2] 91%的企业预计，未来三年此类事件将愈发频繁。

一旦受到隐私、公平与安全问题、算法偏见，以及AI幻觉、深度造假和网络安全漏洞的威胁，企业决策就将陷入瘫痪，无法推进开发、创新等工作。员工害怕失业，客户惧怕诈骗和垃圾邮件，高管则担心运营异常、名誉受损和巨额罚款。稳健的负责任AI能力可以化解这些威胁，建立信任，并使企业可以充满信心地运用AI，进而更大规模地部署现有和全新用例。

另一方面，若把打造负责任AI仅视为一种合规要求，企业会只关注运营成本，而错失负责任AI能够带来的价值。埃森哲研究发现，负责任AI能够同时在财务绩效、体验、最优人才、可持续、多元与包容、以及风险与监管等六大关键价值维度中创造价值，使企业成功跻身AI创新领军者之列。

一、财务绩效

负责任AI可以通过多种方式，帮助企业达成财务预期，甚至实现更出色的增长。这些方式包括：

收入增长。以埃森哲为例，负责任AI合规计划不仅使各方更清楚地了解到，我们是以目标为导向、运作透明的AI合作伙伴，同时也在业界树立了负责任AI应用新标杆。该计划使埃森哲生成式AI的收入同比增长了八倍，[3] 仅2024财年，埃森哲生成式AI的新订单金额就超过了30亿美元。

效率提升。例如，罗尔斯·罗伊斯公司（Rolls-Royce，简称“罗罗公司”）凭借由AI驱动的发动机检测工具“智能孔探镜”（Intelligent Borescope），将检测用时缩短了75%，[4] 并有望在未来五年内削减约1亿

2. 《2024 AI指数报告》，斯坦福大学，2024年，https://aiindex.stanford.edu/report/。
3. 《埃森哲负责任AI蓝图》，埃森哲，https://www.accenture.com/us-en/case-studies/data-ai/blueprint-responsible-ai。
4. 《利用AI力量实现更智能的发动机检查》，罗尔斯·罗伊斯，https://www.rolls-royce.com/media/our-stories/discover/2021/intelligentengine-harnessing-the-power-of-ai-to-deliver-more-intelligent-engine-inspections.aspx。

英镑的检测成本。为了取得长期成功，罗罗公司使用负责任AI框架Aletheia来支持每项AI计划。[5] 这套行动指南共有32项原则，涵盖治理、准确性/信任和社会影响等方面，能够确保以负责任的方式构建AI系统。

积累信任和协作优势。例如，Adobe公司构建了一款运行在亚马逊云端的创意生成式AI模型Firefly。[6] 为了确保该产品以清晰透明的方式提供服务，并尊重创作者权利，Adobe创建了内容凭据技术，基于内容来源和真实性联盟（C2PA）制定的开放标准，任何人都可以在自己的工具和平台中添加凭据功能，以表明图像创建者是人类还是AI。这一创新赢得了客户信任，截至2024年9月，以Firefly为基础、利用各种Adobe工具生成的作品已超过120亿份。[7]

扩展业务。66%的受访企业预计，未来三年负责任AI将在开拓新的业务区域、行业等方面产生强大影响力，比如进军银行业和医疗保健业等监管严格的新市场。在这些市场中，负责任的做法和透明度正变得越来越重要。采取负责任且透明的AI实践做法，还可帮助企业更加广泛地吸引多样化客户群。

二、体验

打造负责任AI的企业预计，其客户忠诚度和满意度将提高25%；此外，近79%的受访企业认为，分享搭建负责任AI的相关经验将显著提升品牌认知度。

以BBH公司（Beauty by Holition）为例，其客户覆盖许多世界知名美妆品牌，如夏洛特·蒂铂丽（Charlotte Tilbury）和蜜丝佛陀（Max Factor）。BBH利用机器学习和AI技术为客户提供最先进的数字体验，如虚拟试妆解决方案、精准AI皮肤诊断等。公司的“Neoskin”肤质AI诊断工具已得到市场验证，可将购买意愿提高至五倍，并使销售转化率提高24%。BBH认为，自身解决方案的成功有赖于在设计中注入负责任要素，始终遵循着六项核心原则——公平、透明、无害、负责任、隐私和包容。通过创建合适的数据集，BBH确保AI算法训练时输出平衡结果，完整提供皮肤类型、肤色、性别和年龄等信息。BBH还使用同行评审和迭代偏差测试，发现那些潜藏的偏差问题，从而更精细地微调数据集。

三、风险与监管

伴随着AI迅猛发展，监管和合规挑战也在日益增多，世界各地纷纷推出监管AI、数据隐私和网络安全的新法规。企业需要紧跟这种发展态势，积极主动地厘清不断扩大的法规组合。

负责任AI框架会优先考虑透明度和公平性，从而帮助企业降低AI风险，避免陷入代价高昂的罚款和诉讼之中。例如，新加坡金融管理局（MAS）已为金融行业设定了合规标准。通过确立公平、道德、问责、透明四大原则（FEAT）以及Veritas和Mindforge等解决方案，新加坡金融管理局鼓励整个金融行业都引入负责任AI实践，使金融服务机构以切实负责的方式利用AI获取价值。

另一方面，网络攻击越来越多，与安全状况不佳的企业相比，拥有成熟AI网络的企业在1天内解除威胁的可能性高出两倍以上；在5天内修复威胁的可能性则高出1.5倍。据埃森哲调研显示，近50%的企业需要耗费10天以上时间来识别和接触这些威胁。因此，将负责任AI整合到网络安全实践中，对于化解风险有着重要意义，回报颇丰。

四、最优人才

近82%的受访企业认为，分享自身成熟的负责任AI经验将提高员工对AI的信任程度。此外，负责任AI的引入，还能使招聘用时缩短20%，质量提高21%，留存率提升21%。

以AWS为例，该企业优先考虑负责任AI创新，将安全、公平和隐私原则嵌入开发流程，并为员工和客户传授相关知识。AWS聘请负责任AI专家，致力于确保自身处于相关研究最前沿，同时设置严格的开发方法体系，以负责任的方式构建AI和机器学习服务。

5. 《Aletheia框架®》，罗尔斯·罗伊斯，https://www.rolls-royce.com/innovation/the-aletheia-framework.aspx。
6. 《Adobe 使用AWS走向生成式AI领导地位》，亚马逊云科技，2024年，https://aws.amazon.com/solutions/case-studies/adobe-case-study/。
7. 《使用Adobe Firefly AI生成作品达到120亿份》，PYMNTS，2024年9月13日，https://www.pymnts.com/earnings/2024/adobes-firefly-ai-hits-12-billion-generations-previews-video-creator/。

亚马逊推出了“AI就绪”（AI Ready）计划，在2025年底前为全球200万人提供免费的AI培训，其中就包括，通过亚马逊数字化学习中心提供使用负责任AI的免费培训课程。[8]

五、可持续性

随着生成式AI工具的广泛应用，能源消耗也水涨船高。企业急需引入可持续的绿色AI举措，尽量减少AI系统给环境带来的负面影响，同时最大程度提升AI开发和部署的可持续性和效率。埃森哲研究预测，通过在绿色AI全生命周期内贯彻可持续性原则，可使能源消耗和碳排放减少40%到60%。

亚马逊开发的生成式AI购物助手工具Rufus可以回答购物相关的各种问题，帮助买家做出更明智的消费决策。[9] 为了满足客户需求，Rufus需要在全球范围内低延迟地接入有着数十亿参数的大语言模型。为此，亚马逊使用Trainium芯片来提高大语言模型训练的可持续性，它是一款专门为超大模型的深度学习训练而打造的机器学习芯片，与其他加速机器学习的解决方案相比，拥有最佳的性能功耗比。凭借这些专用芯片，Rufus的性能功耗比超出其他解决方案54%，帮助团队达成了能效目标。

六、多元与包容

基于大语言模型的生成式AI具有特定的风险，若缺乏适当的化解措施，或用于构建模型的训练数据中有人为偏见数据，都将长久地影响输出结果，这将削弱模型的可信度，阻碍处理进程和降低价值。

负责任AI可以减少无意识的偏见，比如，负责任AI系统可以检测和消除客服人员的偏见语言或语气，确保所有客户都能感觉到自己被尊重和公平对待；它还可以减少招聘中无意识的偏见，受访企业预计可以通过负责任AI增强包容性，吸引更广泛的人才，使员工多元化水平提高21%。

加速负责任AI价值创造

企业应将成熟、完善的负责任AI能力置于价值创造活动的核心。这种能力越成熟，企业创造价值的速度就越快。然而，根据埃森哲调研，72%的企业开始践行负责任AI的时间不足两年（见图二）。

企业必须全力以赴地建设负责任AI，积极推进投资计划，以加快实现价值的速度。具体来说，我们建议企业采取以下行动。

一、培养价值导向型思维

负责任AI不仅能化解风险，还能创造价值。通过定期衡量负责任AI在财务绩效、体验、合规等领域的影响，企业可以展现该技术实实在在的益处。

利用清晰的衡量标准，企业能够跟踪进展并找到需要改进的环节。这种方法将使企业确信，负责任AI并非运营的成本，而是增长的关键驱动力。

为此，企业不应设置静态的负责任AI目标。负责任AI的能力越成熟，企业创造价值的步伐就越快，这样才能比竞争对手更早从中受益。压缩投资周期无疑能够加速胜出。

二、优化责任设计

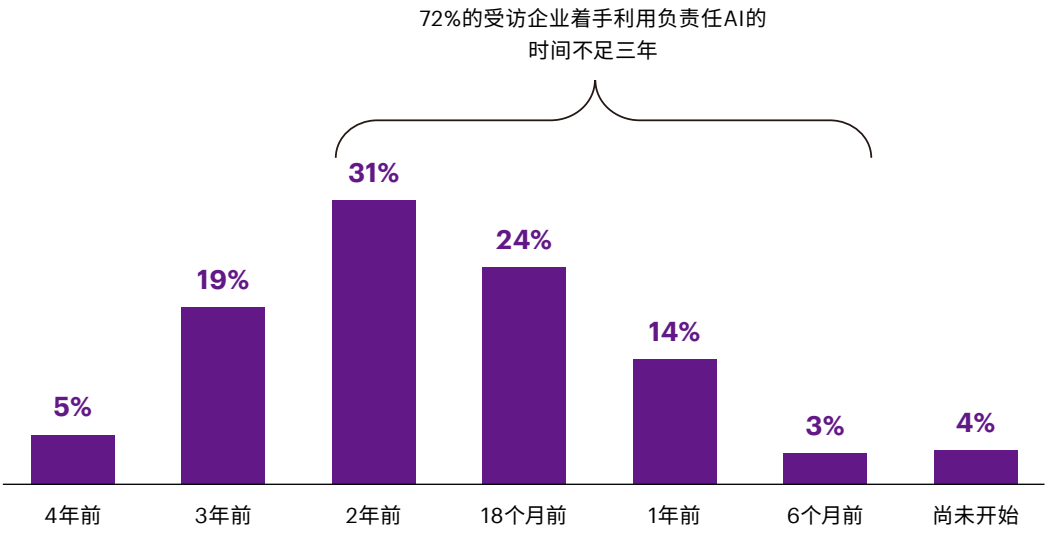
企业必须将负责任AI嵌入其业务核心，AI价值链的每一阶段均采取“通过设计切实担负责任”的方法。

若想更加积极地设计责任，每家企业面临的挑战各不相同，这取决于企业处于哪个阶段。不过，所有企业都可以通过专注于五大优先事项而获益，目标明确地构建负责任AI能力。这些优先事项包括：制定负责任AI战略和实施路线图，开展AI风险评估，系统性地实施负责任AI测试，持续监测与合规，评估AI对员工、可持续发展、隐私和安全的影响。

8.《亚马逊计划通过其新的“AI 就绪”计划承诺在2025年前为200万人提供免费的AI技能培训》，亚马逊，2023年11月20日，https://www.aboutamazon.com/news/aws/aws-free-ai-skills-training-courses。

9.《在Prime Day期间，利用超过80000个AWS Inferentia和AWS Trainium芯片扩展亚马逊生成式AI购物助手Rufus》，亚马逊云科技，2024年10月10日，https://aws.amazon.com/blogs/machine-learning/scaling-rufus-the-amazon-generative-ai-powered-conversational-shopping-assistant-with-over-80000-aws-inferentia-and-aws-trainium-chips-for-prime-day/。

图二 企业开始实施负责任AI的时间分布



三、采用平台化方法

平台化方法可以将负责任AI原则整合到所有AI计划当中，增强可扩展性、风险管理和运营效率，且通常还能提升自动化水平。

平台化策略支持企业领导通过风险筛查，衡量组织的负责任AI成熟度水平。这样做还使企业能够持续完善数据治理和管理实践，并运用预测分析和实时数据监测，更准确地了解、管理数据对AI系统的影响。

在夯实数据基础后，企业就可以利用数据隐私保护、模型可解释性和偏差测量等方法，为更加安全地开发和使用AI设定新标准。平台模式还集成了各种关键组件，如法律法规要求，帮助企业更轻松快捷地适应监管变化，准确满足合规需求。

长期以来，负责任AI都被视为一种风险缓解措施，而非价值驱动引擎。但现在，通过将负责任AI嵌入业务核心，以价值思维指导行动，并采用更迅捷的平台化方法，企业可以开辟出一条从信任构建到收获AI价值的清晰路径。除了制定AI治理框架和原则、开展AI风险评估，企业还应利用正确的工具、技术和人员，系统性地支持负责任AI。

AI不只是一项新技术。如同云计算和互联网一样，它正在重塑我们的世界。企业绝不能持观望心态，在AI征程中踟蹰不前。在机遇面前，企业应当立即行动，投入资源，实现AI的全部价值。

作者

阿纳布·查克拉博蒂（Arnab Chakraborty）
埃森哲首席负责任AI官

詹妮弗·杰克逊（Jennifer Jackson）
埃森哲亚马逊云科技事业部总裁

生成式AI加速可持续发展

文 斯蒂芬妮·贾米森、阿纳布·查克拉博蒂

生成式AI有望给可持续发展进程注入蓬勃动能，但其成效多少取决于我们是否负责任地应用生成式AI。埃森哲提出生成式AI助力加速可持续发展的四种应用模式，帮助企业制定战略决策，稳步迈向可持续发展目标。

当前，日益加剧的地缘政治紧张局势、气候变化等诸多因素都在阻碍着联合国可持续发展目标实现，[1] 按照现有速度，在2030年之前我们可能只能实现其中17%的目标。

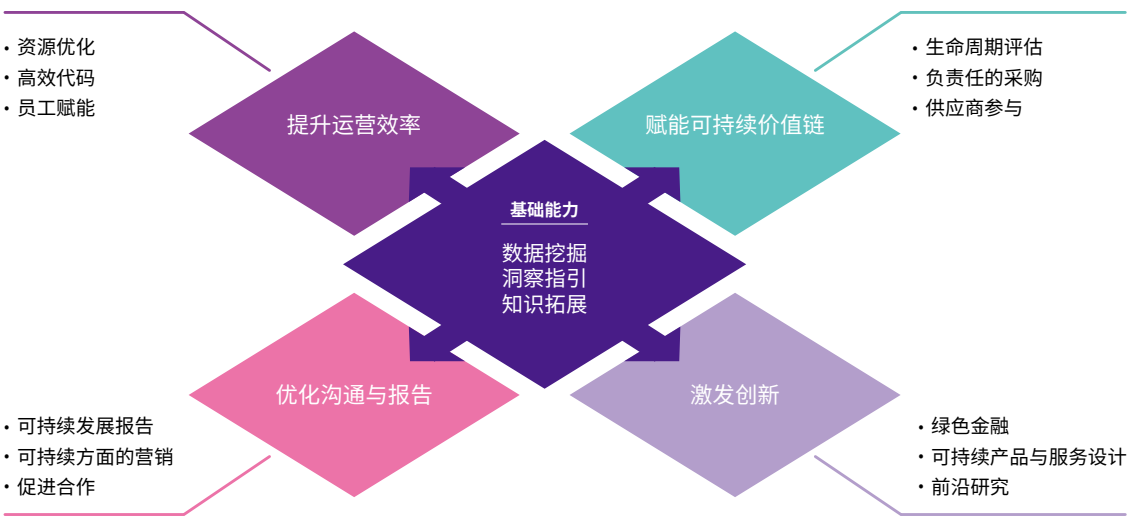
生成式AI的出现带来了新的希望。它可以支持企业获取量身定制的信息，通过跨学科思维加速创新，提高生产率，从而帮助企业消除不利因素，解决复杂难题。2024年联合国大会呼吁，抓住“安全、可靠和值得信赖的”人工智能系统带来的机遇，让人工智能给人类带来“惠益”，并以此促进可持续发展。[2]

四大应用模式，促进可持续发展

生成式AI之所以能够创造商业价值，主要因为它的三大基础能力：数据挖掘、洞察引导和知识扩展。生成式AI能够从文本、图像、音视频等非结构化的数据中提取有价值的发现，帮助员工分析复杂数据并提供建议或预测，作为功能强大、可定制的搜索引擎或是虚拟助理，多方位地增强员工能力。

这三大基础能力在与具体业务深度结合后，又进一步延展为四类促进可持续发展的应用模式（见图一）。

图一 生成式AI助力可持续发展的应用模式



1. 《联合国可持续发展目标》是一套包含17项全球目标的体系，旨在消除贫困、保护地球、确保人人享有和平与繁荣。
2. 《具有里程碑意义！联合国大会通过首个有关人工智能的决议》，央视网，2024年3月22日，https://news.cctv.com/2024/03/22/ARTIBAQ7LUapb3VTE3kkzUXb240322.shtml。

一、提升运营效率

在财务和环境双重约束下，企业需高效管理资源，实现稳定、可持续的回报。生成式AI为提升运营效率提供了新机遇，涵盖资源优化、员工赋能和高效编码等。当然，企业也需权衡引入和使用生成式AI的资源成本。

- 资源优化：企业可借助生成式AI叠加现有分析工具，实现从计算能力到运输网络等多方面资源的优化。例如，企业可利用生成式AI，将预测性分析系统升级为规范性维护系统，为员工生成指令和建议，降低成本、减轻环境负担。
- 员工赋能：员工可以将生成式AI作为专业教育工具，根据自身的岗位、语言、以及特定地区的法规或政策，个性化地学习可持续发展主题。此外，生成式AI还能帮助识别和设计与企业目标相关的可持续发展培训课程，充当助手与员工合作，提高他们的工作效率和决策能力，在创造商业价值的同时促进可持续发展。
- 高效编码：生成式AI可自动生成代码，并对其进行优化、编排，从而帮助IT团队提升运作效率。比如，利用生成式AI缩短软件开发周期，不仅能提高整体绩效，还能减少资源消耗。

案例：

西门子与微软合作在德国舍弗勒集团（Schaeffler）的一条生产线上部署了生成式AI解决方案Industrial Copilot，协助舍弗勒自动化工程师为可编程逻辑控制器（PLC）编制代码。新方案使用自然语言输入来开发代码，大大减少了编码过程的时间和出错概率。这不仅避免了员工的重复性劳动，还使工程师资源能够专注于更高价值的工作。同时，Industrial Copilot能够访问所有相关的文档、指南和手册，从而迅速甄别出潜在错误，减少生产流程所受的干扰，减少机器停工时间。得益于效率提升，西门子不仅降低了成本，还减少了整个运营过程中的碳排放。

二、赋能可持续价值链

实现行业层面的可持续发展，必须让整个供应链参与其中。然而，当前企业需耗费大量精力处理繁琐流程以获取数据并提升供应链透明度。生成式AI可高效解析非结构化数据，助力企业通过生命周期分析、负责任的采购及供应商协作，全面掌握产品和服务的环境与社会影响。

- 生命周期分析（LCA）：因为监管机构要求企业提供详尽的可持续数据和说明，对产品和服务进行详细环境影响评估的需求正与日俱增。生成式AI可大幅提升LCA效率，依据《温室气体核算体系》（Greenhouse Gas Protocol）等标准快速生成生命周期分析的准确结果，支持企业制定科学的可持续发展策略。
- 负责任的采购：采购决策会直接影响企业的环境和社会足迹（碳排放、公平用工等）。生成式AI可以分析供应商报告和采购数据，甄别合规与风

险，制定涵盖人权、劳工、环保等因素的采购决策。企业还应将此能力拓展至能源、基础设施和AI模型采购中，实现全面责任管理。

- 供应商协同：许多供应链风险无法仅靠采购解决，需依赖上下游协同。通过比对企业标准与供应商报告或审计结果，生成式AI可识别供应链风险与改进空间，强化供应商培训与沟通，帮助他们消除风险，并采纳可持续举措。

案例：

多层级供应链导航器（N-Tier Supply Chain Navigator）是埃森哲旗下覆盖多级供应链的管理类工具，它可以帮助企业改进供应链运营。其工作原理是利用生成式AI将供应链数据与业务地点和行业的指标进行交叉对比，如是否位于童工高风险地区、是否与国际公约保持一致，从而识别可持续发展和人权等方面的风险。埃森哲近期与一家大型软件企业合作，梳理了一套覆盖154个国家超过12.2万家供应商组成的庞大网络。埃森哲使用N-tier工具发现，50%至60%的碳排放热点来自二级和三级供应商。通过此举，企业可以了解复杂的全球供应链，进而做出明智的可持续采购决策。

三、促进创新

由于实现联合国可持续发展目标的时间已所剩无几，企业需采用创新型解决方案，以加速弥合愿景与现实之间的差距。生成式AI可以在绿色金融、可持续产品与服务设计、以及研究前沿等领域中，寻找创意和解决方案，帮助推动可持续发展目标的快速实现。

- 绿色金融：企业为可持续发展方案融资时往往面临复杂挑战，而生成式AI可以帮助企业充分了解金融环境，确保获取所需资源。生成式AI还能助力金融机构更有效地借鉴现有成功案例，设计出更适合当前状况的方案，比如，针对绿色项目的债券、贷款、基金和保险条款，为可持续发展开辟新的融资渠道。
- 可持续产品与服务设计：生成式AI可以帮助企业将可持续发展的理念融入产品与服务设计整个流程中——从设计开发的早期阶段，一直到确定绿色材料和选用持续制造工艺。此外，生成式AI还可助力设计师合理考量那些看似相互冲突的选择，比如，在满足产品功能性要求的同时，考虑选材、负责任采购和循环设计等可持续发展因素。
- 前沿研究：生成式AI可以广泛抓取庞杂且海量的数据，对数据相关性进行分析，识别前沿趋势，并生成最新的可持续发展解决方案，从而加速可持续发展目标的实现。比如，生成式AI可以快速访问不同数据源的资料库，找出其中各种相关线索，为可持续发展困境找到破局关键。生成式AI还可以打造更适合可持续发展项目的合成数据集，以修正样本代表性不足的问题。

案例：

2024年，雅马哈和Final Aim发布了一款紧凑型电动汽车“概念451”（Concept 451），以支持日本山区的农业生产。两家公司利用生成式AI研究人口老龄化等社会问题，并确定车辆功能需求。接下来，设计团队还利用图像生成式AI技术，研究了近2000种设计方案，并最终确定了一套兼具易用性和实用性的方案。在使用Autodesk Fusion系统绘制电动汽车三维模型的过程中，团队还利用图像生成式AI来促进沟通。通过整合生成式AI，雅马哈和Final Aim成功的展示了，企业如何利用生成式AI缩短研发周期，从而满足不断变化的社会需求。

四、优化沟通与报告

有效的可持续发展沟通与报告是当前企业的切实之需，但这些任务过多挤占了团队的时间。值得庆幸的是，生成式AI已在支持ESG报告和可持续发展营销等一系列耗时任务中取得了长足进步。同时，生成式AI还有利于打破信息孤岛，促进企业内部的沟通和协作。

- 可持续发展报告：准确而详细的ESG治理报告，不但可以满足监管合规和公众问责要求，还能向利益相关方展示积极成果。生成式AI可管理多种数据源，进而确定关键指标、凸显重点举措，并生成初步报告，加快团队的报告制作进度。
- 可持续发展营销：有效、真实地宣传可持续发展成果将帮助企业赢得消费者和投资者的长久支持。营销团队可以利用生成式AI创建内容，如广告、产品标语等，既贴合特定的受众，又与品牌调性保持一致。此外，生成式AI还可以帮助营销人员理解各项可持续发展概念，避免出现虚假的环保宣传。
- 促进协作：生成式AI可以支持可持续发展团队和其他部门之间协作沟通，从而将可持续发展的理念贯彻到整个业务活动中。例如，生成式AI可将功能术语转化成通俗易懂的语言，提升沟通效率。生成式AI还可以支持员工访问整个企业的知识库，使所有决策者都能依据统一的信息做出决策。

案例：

微软在其AI平台Copilot Studio上推出了“可持续发展洞察”（Sustainability Insights）模板，帮助企业高效地获取和分享有关可持续发展目标、进展的定制化数据与洞见。该工具使用生成式AI来分析可持续发展数据，并为所有员工提供适用于具体情境的分析支持。分析结果可以形成报告、文档和记录，在团队间共享。企业还可以将自身的数据与同行数据加以比较，为可持续发展进度设定基准，提高准确性和数据可信度。这些工具还有助于传播可持续发展知识，减少错误，由此完善整个组织中的决策。

生成式AI的可持续发展风险

尽管生成式AI拥有灿烂的发展前景，但它也带来了一些风险，如果不加以控制，这些风险可能直接抵消生成式AI对可持续发展的积极影响。

一、流程不够透明

生成式AI解决方案有望支持甚至推动企业关键决策流程，这意味着AI系统必须具备可追溯性、透明度和问责制。然而，企业需要依赖第三方来开发生成式AI技术堆栈中的基础设施、模型和数据，而外部数据源中或许包含了标签错误或侵权的数据，可能存在责任漏洞。另一方面，开发人员可能无法解释、甚至无法理解模型如何得出结论，这就可能导致用户也不清楚应该如何使用和验证生成式AI的输出结果。

二、结果缺乏准确性

同人类一样，生成式AI模型也会受到偏见和不确定性的影响。一方面，使用生成式AI可能会在无意中引入或延续偏见；另一方面，生成式AI模型有可能会产生AI幻觉。由于“自动化偏见”问题和生成式AI的大规模应用，一些用户可能会过度信任生成式AI，在没有进行充分检查的情况下就听从模型的输出意见。

三、破坏隐私和数据安全

生成式AI有可能在无意间泄露敏感的个人信息或公司机密，从而使企业面临违反某些国际数据安全协议的风险。这很有可能因为大型数据集没有经过应有的审核，或者即使进行了正确的数据筛选，但由于缺乏人工判断，造成意外披露和泄露敏感数据或机密数据的风险。

四、生成式AI能力被滥用

滥用生成式AI可能会阻碍可持续发展目标的实现，比如通过深度伪造等合成内容传播错误信息，为制造数字或实体武器（如网络攻击或自制武器）提供指导，或以其他方式支持欺骗或暴力。使用者还可能故意操纵生成式AI模型，制造有毒输出或侵犯数据隐私，从而增加上述风险。

五、消耗大量资源

生成式AI需要依靠资源密集型的数据中心进行训练和推理，这将消耗能源、水资源和计算机硬件。预计生成式AI的扩展将加大数据中心的电力消耗，从2022年到2026年，电力消耗将增加一倍以上，占比超过全球能源总产量的1.5%。

六、重塑工作岗位

生成式AI将改变各行各业的工作性质和质量，技能需求会随之产生很大的变化。埃森哲研究发现31%至47%的工作时间可能因生成式AI的发展而得到优化或实现自动化，[3] 特别是那些重复性、常规性或标准化数字任务的岗位（比如技术支持、文案写作或数据录入类工作），很有可能会在该技术应用初期就感受到影响。

七、加剧数字鸿沟和贫富差距

尽管我们努力将AI推广至计算资源匮乏的环境中，但互联网和电力等资源分布不均的现实意味着许多人仍将滞后于科技进步，长此以往，生成式AI可能扩大现有的数字、数据和创新鸿沟，加剧当前的不平等以及弱势群体的人权问题。

管理每一次生成式AI部署

负责任地成功应用生成式AI，企业要做好三个准备。

一、奠定运营基础

无论企业从头开始自行构建模型，还是向供应商购买访问权限，关键都是要创建一支由跨职能部门代表组成的负责任AI团队。在部署生成式AI之前，这一团队应逐一完成下列工作：盘点生成式AI用途、识别相关风险、定义负责任AI原则、设计风险管理框架，并且创建或升级相关政策，与该框架保持一致，由此夯实运营基础。

3. 《生成式AI时代的工作模式、工作组织、工作者》，埃森哲，2024年，https://www.accenture.com/content/dam/accenture/final/accenture-com/document-2/Accenture-Work-Can-Become-Era-Generative-AI.pdf。

二、打造生成式AI生态系统

这一过程必须统筹协调多个业务职能部门，以确定现有的能力和目标等。企业的技术堆栈可能会包括来自第三方供应商的数据、云计算和其他功能。企业应确保选取那些在资源消耗和数据隐私等风险领域中与自身价值观一致的供应商，决定使用预开发模型的企业尤其应当如此。企业还需要管理数据供应链并调整流程，确保应用程序具有代表性。在某些情况下，企业或许有必要围绕新出现的情况或代表性不足的人群，设法获得以往未曾接触的数据。

三、使员工做好准备

企业选择的生成式AI解决方案是否有效，取决于与之互动的人。要对任何分析型生成式AI应用程序进行有效的人工监督，就必须确保相关员工具备数据素养。为了推进以人为本的转型，并避免员工失去工作岗位，企业就必须为员工提供相关新技能的培训。因为这些技能非常前沿，培训很可能是确保企业能留住合适人才、进而充分发挥生成式AI潜力最实用的方法。员工还必须了解与使用生成式AI相关的重大风险、知道如何识别风险、以及解决风险的流程。这意味着，企业需要依靠上述运营基础，广泛吸引内部和外部利益相关方参与进来，并对他们进行AI伦理、法规和最佳实践等方面的培训。

在做好准备之后，企业就可以着手负责任地开发和部署生成式AI解决方案。每项生成式AI部署都独一无二，我们建议不论企业处于何种规模、行业、地区，都可以将以下步骤作为通用指南（见图二）。

与其他重大技术进步一样，生成式AI将对可持续发展产生错综复杂的影响，但我们可以通过尽职调查，发现并降低其中风险。我们对未来保持乐观，期待利用生成式AI加速实现可持续发展目标，使世界能够更好地应对挑战。

作者

斯蒂芬妮·贾米森（Stephanie Jamison）
埃森哲资深董事总经理、公用事业全球负责人兼可持续业务主管

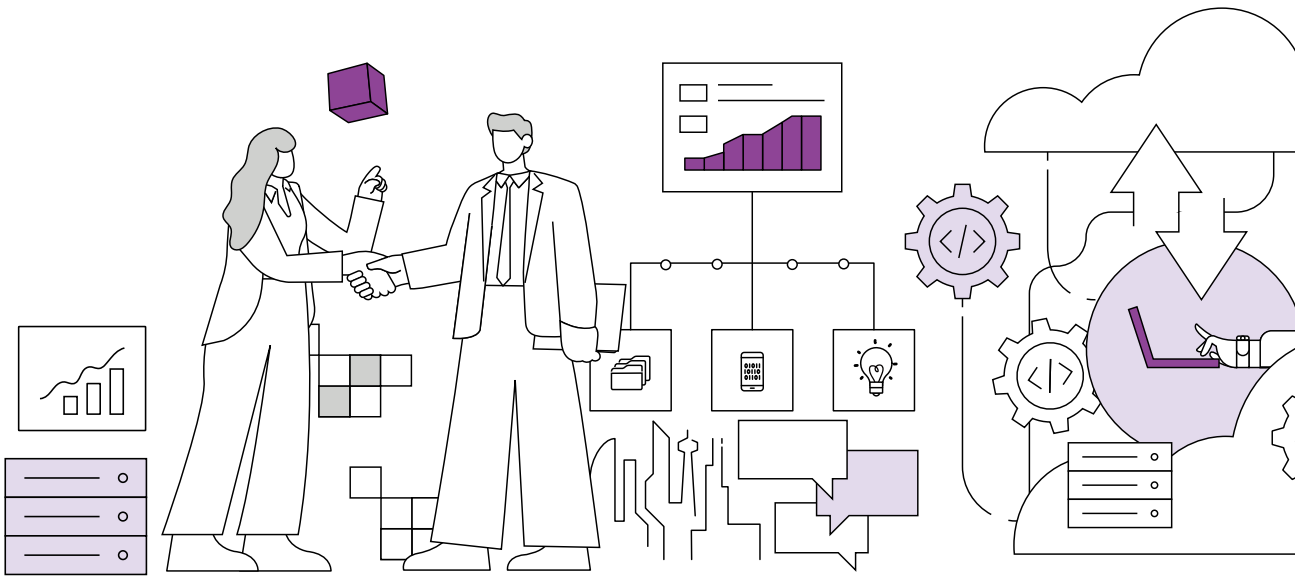
阿纳布·查克拉博蒂（Arnab Chakraborty）
埃森哲首席负责任AI官

图二 部署生成式AI来促进可持续发展的步骤

步骤	说明	生成式AI能够帮助
 承诺	企业领导层承诺支持《全球契约》，并将“十项原则”作为企业战略、文化和日常运营的组成部分，由透明的治理结构进行监督。	• 起草说明、交流要点和业务模式 • 针对特定受众进行宣传
 评估	企业应发现潜在风险，并按照概率、范围和影响程度进行分类。需要根据企业的风险承受能力，量身定制化解策略，帮助保障生成式AI的使用，防止在实现可持续发展目标的道路上出现退步。	• 持续捕捉、汇总和清理不同类型的非结构化数据 • 分析数据以识别风险与机遇 • 评判风险的重要性和公众看法
 定义	在评估的基础上，企业制定并完善针对自身运营环境的目标和衡量标准，全面跟踪AI对于业务目标和可持续发展目标的积极或消极影响。企业必须落实每项部署工作的责任，建立监控框架，列出可能导致部署终止的任何状况，并了解怎样判别应用程序的生命周期终点。	• 确定各项机遇的优先次序 • 定义目标和衡量成功的关键绩效指标 • 起草路线图并广泛宣传
 实施	企业建立核心流程并确保对其不断调整。吸引员工积极参与并对其进行培训，建设所需能力和资源，并与供应链合作伙伴携手，共同完善和实施战略。同时开通适当的报告渠道，使最终用户可以就应用程序使用过程中遇到的问题进行沟通。	• 将基础功能作为解决方案技术堆栈的组成部分 • 通过实施生成式AI来管理利益相关方和解决方案的开发
 衡量	无论部署规模大小，企业都需要对生成式AI展开渐进式的检查和投资，及时发现问题并逐步改进。企业应捕捉数据，监控和优化能消水平、对利益相关群体的影响、以及使用生成式AI的其他副作用，从而确保更高质量、更高效率、恪守伦理道德的成果，同时最大限度地减少浪费、缺陷和差异。	• 确定衡量可持续发展进程的最佳实践做法和监管要求 • 对不同来源的数据进行整合和标准化处理
 宣传	企业通报其进展情况和前瞻性战略，以此履行承诺，并与利益相关方合作来取得更大进展。	• 收集和汇总内部交流信息 • 生成视觉效果 • 与报告框架保持一致

各司其职，通力协作

成功实施生成式AI来实现业务和可持续发展目标，需要企业各部门倾力协作。虽然每家企业的具体情况不尽相同，但可以将这份指南作为起点，深入了解企业自身特点，根据建议采取行动。



领导层与战略

企业最高管理层可以对生成式AI应用程序设定明确的期望——包括如何负责任地设计和实施这些应用程序，支持生成式AI的有效使用。他们应提供资源并扫除障碍，帮助各部门完成工作。他们还需要确保制定明确的负责任AI指导原则、强有力的AI治理措施，这些措施将由AI治理委员会负责落实，并接受监督。

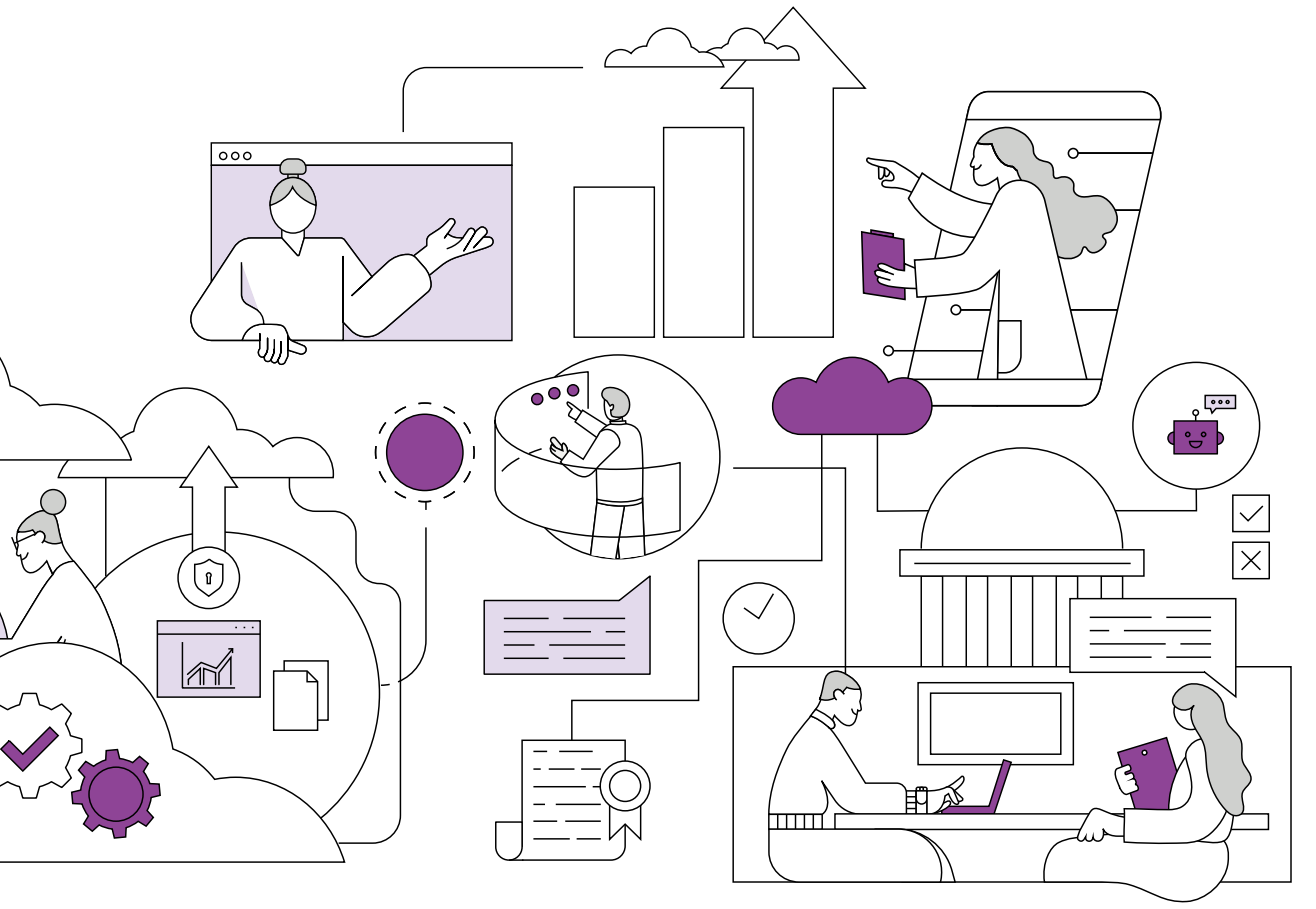
人力资源

生成式AI工具仍然需要由人来管理。人力资源部门要确保员工接受培训，负责任地使用生成式AI。适合企业且卓有成效的培训绝不会一蹴而就，因此人力资源部门需要与信息技术部门等其他部门紧密合作，充分了解部署每套生成式AI系统的时间表和需求。随着生成式AI的发展，提升技能、开展培训将持续成为优先事项，需要人力资源团队不间断地检查与管理。

可持续发展

可持续发展团队对生成式AI负有双重职责。首先，他们应当是负责任AI实践的倡导者，包括引导将负责任的生成式AI原则全面纳入企业AI、数据和可持续发展战略的对话当中。

其次，可持续发展团队应推动利用生成式AI来促进可持续发展。这包括在自己的工作活动（如编写可持续发展报告）中以身作则地使用生成式AI，并与其他部门合作，深化变革。



采购

采购部门应与信息技术部门一起，积极管理与数据和技术堆栈提供商有关的风险。与任何其他采购一样，尽职调查至关重要；采购部门必须与其他部门协调，了解需求和风险承受能力，确定最合适的供应商。

信息技术和数据

信息技术团队和数据团队负责生成式AI应用的技术落地。他们必须准备好企业数据和技术堆栈，充分利用生成式AI，同时牢记负责任的原则和面向未来的考量。一旦企业开始使用生成式AI，如果没有对数据和模型进行必要的审核，或没有提供保障和管理，就很可能面临巨大的风险。

财务和法务

生成式AI或许不是解决问题的唯一方案。盲目冒进可能会导致投资浪费和财务负债。不道德地获取第三方数据或员工滥用企业AI系统等问题，可能引发的法律后果尚不明确，因为对于使用生成式AI所造成不良后果的追责，很可能并没有先例可循。

认识到这一点后，财务和法务团队应参与生成式AI治理，严格把关，评估业务模式和法律环境，确认在推进之前所有风险都已得到妥善管控。

如何提升负责任AI的成熟度

文 阿纳布·查克拉博蒂、森蒂尔·拉马尼

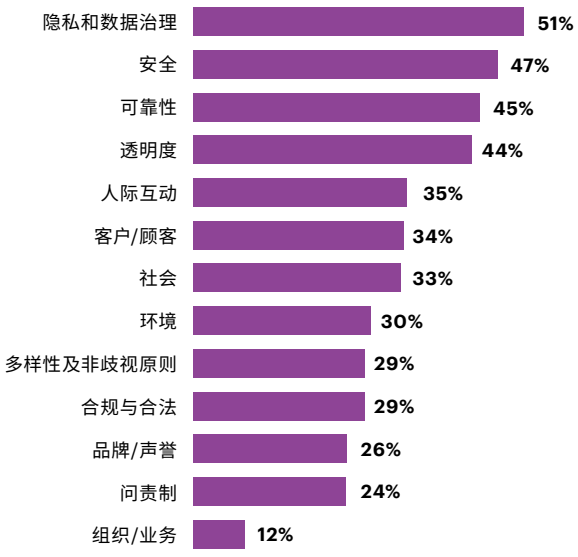
无论是主动还是被动，很多企业已开始加紧布局负责任AI，并将其设定为发展目标。但我们距离这个目标究竟有多远？如何量化我们的发展阶段？埃森哲联合斯坦福大学，共同搭建负责任AI的成熟度体系，我们将这一过程划分为四个发展阶段，并提出五大优先事项，助力企业量化发展进程，稳步实现最终目标，争做负责任AI的先锋。

AI风险正在快速演变中，生成式AI的普及更是加快了这一过程。仅过去一年，我们就见证了层出不穷的新风险，涉及AI幻觉、知识产权保护、网络安全和环境影响等诸多方面，有关数据隐秘性、可靠性和透明度等老问题也卷土重来。AI风险将在以下三个方面持续演变：

涉及领域更多，发生频率更快。埃森哲调研发现，在企业面临的诸多AI风险中，最突出的是“隐私和数据治理”相关风险，有近半数（51%）的受访者认为其所在的企业正面临此项风险（见图一）。紧随其后的是安全性和可靠性风险，后者包括输出错误、AI幻觉和模型失灵等问题。此外，生成式AI的发展还将持续影响AI风险格局，包括透明度、“黑箱”挑战（44%受访者提及）、数据中心的环境和能源需求（30%的受访者有此担忧）、问责制、知识产权保护（24%的受访者提及）等。

监管形势加速演变。随着政府加速推动AI监管政策落地，那些在打造负责任AI方面有些落后的企业将承担更多的合规性风险。目前已有超过37个国家推出了各自的AI监管措施，由于各国政府所颁布的法案不尽相同，跨国公司面临的合规要求只会越来越复杂。绝大多数受访企业都深受影响，其中77%的企业要么已面临AI监管，要么预计在未来5年内将接受监管。90%的企业预计在未来5年内将承担AI相关法律义务，包括但不限于网络安全、AI相关的法律合规、隐私数据和消费者保护等。此外，有些国家和地区的立法进程正同时在国家和地方层面推进，这无疑会使企业面临更复杂的合规性挑战。

图一 企业面临的主要AI风险



数据来源：埃森哲-斯坦福大学联合高管调研，*N*=1000位高管。

风险管理逐步延伸，贯穿整条价值链。不仅仅是风险和监管在悄然变化中，随着众多企业开始购买和开发AI模型，特别是很多企业既是购买者又是开发者，风险管理已逐渐延伸至AI业务的整条价值链。埃森哲调查发现，在获取AI模型的受访企业中，仅有43%的企业采取了监管检查和第三方审计等强有力的采购措施。这意味着面对AI价值链风险的日趋复杂，很多企业尚未做好充分准备。

埃森哲调研预测，若企业成为负责任AI先锋，其AI相关收入将平均增长18%（AI相关收入是指AI赋能产品及服务而产生的总收入），反之，仅一起与AI相关的重大事件就可导致企业市值蒸发24%。因此我们不难理解，为何企业正加紧投资开发负责任AI——42%的受访企业已将AI总体预算的10%以上用于负责任AI项目；79%的受访企业则计划在未来两年内落实这一目标。

那么，这些企业距离实现“负责任AI”的目标还有多远？

重新定义负责任AI成熟度

为了充分了解当前负责任AI的成熟度，埃森哲与斯坦福大学合作制定了一套框架，企业在该框架中所处阶段越高，表明其在成熟度方面取得的进步越大。通过分析调研1000位受访高管的回复，我们针对组织成熟度[1]和运营成熟度[2]两方面分别打分，确定了企业所在的四个阶段。[3]

1. 组织成熟度反映了企业已在多大程度上实施了负责任AI流程及其有效性。衡量指标包括高管层支持、治理、数据和AI风险识别与管理、模型开发、采购、监测与控制、网络安全和培训等。组织成熟度得分高，表明企业已做好充分准备，能够负责任地使用AI。
2. 运营成熟度反映了企业能否采用负责任AI措施，减轻特定AI相关风险，包括与隐私和数据治理、多样性及（非）歧视性、可靠性、安全性、人际互动、问责制和环境影响相关的各类风险。运营成熟度得分较高，表明企业擅于在自身所有适用的风险领域中负责任地实施AI。
3. 本次研究不包括那些尚未出台负责任AI行动计划的企业。

第一阶段：制定负责任AI原则

处于这一阶段的企业已具备一定开发AI系统的基础能力，但在负责任AI方面的工作还不太成体系，具体标准包括：

- 已制定一套符合伦理道德的AI原则和指导方针，包括以负责任且安全的方式访问和使用数据的相关政策和规则；
- 尚未确立用来治理数据质量、数据隐私、数据安全和AI模型风险的流程；
- 偶尔会开展针对数据和AI项目的风险评估审查；
- 已部署AI项目工作流程，但尚未在数据、模型和AI应用程序中系统性整合对负责任AI的评估。

第二阶段：构建负责任AI计划

在对负责任AI做出评估后，企业已采取以下举措：

- 制定负责任AI战略，包括定义明确的运营模型、数据和AI治理措施，从而推动企业将愿景转化为切实行动；
- 跨越数据、模型和AI应用程序，制定强有力的AI风险评估方法和流程；
- 针对训练数据、以及模型输入和输出信息，建立提升透明度和可审计性的流程，并做出适当决策；
- 制定跨越数据和AI管道的监测和控制框架，可在项目工作流程中予以执行；
- 所实施的各项流程仍处于早期阶段，缺乏由相应工具和技术支持的进一步系统性实施。

第三阶段：付诸负责任AI实践

企业在整个组织内系统性地实施了以下措施，确保履行相关的监管和法律义务：

- 借助运行原则、指导方针和相关流程，在整个企业范围内实施负责任AI战略；
- 跨越数据管道、AI模型和AI应用程序实施风险评估，提升整个模型生命周期的可追溯性、透明度，并通过定期审计，增强监管依从度与合规性；
- 利用隐私过滤、匿名化和验证等手段来有效控制数据来源，剔除敏感信息，并减轻自助服务工具的数据偏见风险；
- 借助模型可解释性工具促进系统性AI测试，确保其可解释性、AI模型性能偏见测试、准确性等，同时助力界定模型在给定的法律、伦理道德及运营范围内运行；
- 建立负责任AI控制平台，由控制人员对数据和AI价值链持续开展人工监测，警示意外风险或违规行为，并做出补救；
- 推出负责任AI学院来培训和赋能员工，推动负责任AI的采用。

第四阶段：成为负责任AI先锋

这一阶段的企业已通过平台化方式全面发挥负责任AI的影响力，系统性地采用面向未来的做法，真正释放AI价值。

- 负责任AI已端到端、系统化地充分达到可运营状态，由技术平台支持，重新设计相关流程，并建立适配的人才团队和文化；
- 预防性地开展负责任AI活动，部署专门的资源、流程等要素，持续评估当前和未来的各种风险；
- 随着外部技术和监管环境的演变，主动调整自身数据及AI风险管控流程；
- 不断完善和推进数据治理和管理实践，采用预测性分析和实时数据监控手段，动态化地了解和管理数据对AI系统的影响；
- 制定新的标准、方法和途径，包括数据隐私保护、模型可解释性和偏见测量、对抗性测试等，以安全开发和利用AI技术；
- 在负责任AI实践方面成为公认的领军者，积极与价值链合作伙伴、监管机构和社区等外部利益相关方合作，确保反馈过程的参与度和包容性，为前瞻性的监管合规提供支持；
- 积极与第三方AI供应商合作，推动AI改进，创建信任关系和新的合作机会，以此有效管控第三方AI风险。

调研发现，绝大多数（78%）的受访企业已构建负责任AI计划，14%的受访企业开始发挥负责任AI的实际效果，仅有8%的受访企业才刚刚开始制定负责任AI原则。值得注意的是，目前还没有一家企业步入先锋阶段。

具体来看，仅9%的企业在组织成熟度方面达到了先锋水平，在运营成熟度上达到先锋水平的企业更是凤毛麟角，不足1%（见图二）。

随着我们步入生成式AI时代，企业必须基于自身组织成熟度和运营成熟度，构建负责任AI，方能充分发挥其作为价值杠杆的效力。企业应树立面向未来的思维模式，不断改进和发展，力求跟上新技术进步和监管变革的步伐。

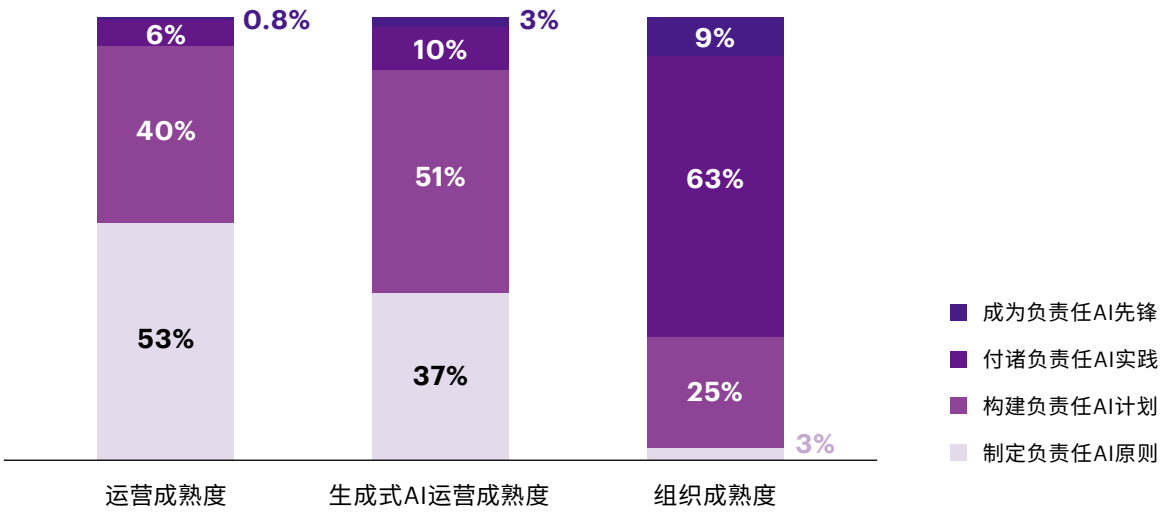
过去几年中，埃森哲一直在现有AI原则基础上，努力提升负责任AI的成熟度，目前处于成熟度的第三阶段。公司制定了一套严谨的负责任AI计划，并将其付诸实践。该计划稳健运行的关键要素包括：

- 领导层监督：埃森哲任命了首席负责任AI官，监督内部的负责任AI合规计划，埃森哲首席执行官和总法律顾问负责提供相关支持。埃森哲首席技术官、首席负责任AI官、[4]总法律顾问、数据和AI主管共同负责监督相关指导委员会。此外，董事会也定期向审计委员会报告相关进展。
- 原则和治理：埃森哲的AI解决方案的开发及部署方法建立在一套原则之上。[5]这套原则不仅适用于埃森哲自身运营，也适用于埃森哲与客户、合作伙伴和供应商之间的合作。公司组建了一支跨职能团队，成员来自法务、安全、采购和人力资源等职能部门，并且包含首席信息官和负责任AI专家。该团队负责以负责任AI原则作为制定计划的指导方针和基础，设计和领导新的负责任AI合规计划。利用治理框架来落实关键原则、策略和标准，能够有力地支持跨用例监督。
- 风险评估和解决方案措施：埃森哲根据《欧盟AI法案》和其他关键框架，采取了基于风险的方法。这意味着要花费大量的时间预先定义出“高风险用例”。过去一年里，埃森哲审查了整个公司中的数千种AI用例，进而制定出详细的风险评估和解决方案。
- 测试与赋能：在系统赋能方面，埃森哲也投入了大量时间，制定自身的AI策略、标准和控制措施，并将其嵌入技术、流程和系统中。这需要全面构建负责任AI技术能力、专业知识、工具和技术，同时开发基准测试工具。
- 人才：埃森哲还在不断培养负责任AI技能和人才，不仅强制要求与AI项目有密切关系的员工参加伦理及合规培训，还通过技商（TQ）课程让70多万名员工接受新的伦理和AI培训。

4.《埃森哲采取新措施帮助客户负责任地扩展生成式人工智能》，埃森哲，2024年5月14日，https://newsroom.accenture.com/news/2024/accenture-takes-new-steps-to-help-clients-scale-generative-ai-responsibly。

5.《埃森哲负责任AI蓝图》，埃森哲，https://www.accenture.com/us-en/case-studies/data-ai/blueprint-responsible-ai。

图二 负责任AI成熟度四个阶段的企业分布百分比（按运营、生成式AI和组织领域划分）



数据来源：埃森哲-斯坦福大学联合高管调研，*N*=1000位高管。

负责任AI先锋企业的特质

一、预判

负责任AI先锋企业能时刻跟进外部技术的发展和监管环境的变化，不断调整自身的风险监测和控制流程。他们将自身的AI活动与企业的总体战略目标和风险容忍度系统性地结合起来。先锋企业打造专门的团队和机制，持续评估和管理当前以及未来面临的各种风险，并为快速采用新的技术成果提供平台。

二、责任融于设计

先锋企业能够通过设计，将负责任AI置于自身AI总体战略的核心。这些企业制定出强有力的跨职能框架来赋能实时决策，由此信心十足地支持新技术的采用并扩展应用规模，同时大力投入未来规划和全面监测，不断完善和改进负责任AI治理结构、原则、政策和标准，从而保持竞争优势。

三、积极的合作伙伴

先锋企业通过与合作伙伴密切协作来有效管理风险，支持多利益相关方参与、反馈、包容性决策以及前瞻性监管合规。先锋企业能够打造更广泛的AI生态系统，并在其中组织和引导各利益相关方开诚布公地探讨问题，集思广益，构建信任纽带，与利益相关方共享负责任AI最佳实践，提供技术支持。

五大优先事项提升成熟度

埃森哲研究发现，尽管不同企业所处阶段不同，但所有企业都可以通过以下五大事项，提升负责任AI成熟度，收获AI红利。

事项一：确立AI治理与原则

制定负责任AI战略及路线图，包括明确的策略、指导方针和控制措施，对于负责任AI实践的成功实施、运营和治理具有至关重要的意义。

该流程的基本举措之一便是定义、采用并执行一套负责任AI原则，具体涉及所在企业的优先事项、伦理道德及价值观，以确保符合伦理地制定、部署和使用AI。埃森哲研究表明，70%的受访企业已制定负责任AI原则，其中54%已将这些原则转化为相关策略。

另一项重要举措则是实施稳健的AI治理运行模式，并以此为基础构建负责任能力及相关控制措施。令人欣喜的是，76%的受访企业已全面实施治理模式，较两年前仅31%有了大幅提升。这反映了企业并非以临时性的方式来管理风险，而是采用自上而下的跨职能方法来推进负责任AI事宜。

企业亦可更进一步，通过提供员工培训、变革管理行动计划等举措来加强这一运行模式，从而在整个企业范围内促进和维护负责任AI。

事项二：开展AI风险评估

了解企业应用AI的风险，是实施负责任AI的关键组成部分。参与本次调研的大多数企业似乎都低估了AI相关风险。例如，我们向受访者展示了13种AI相关的风险，询问企业将哪些风险视为隐患，结果显示受访者选择的风险种类平均仅有4.4种。企业对AI风险的低估，也体现在超过半数的受访企业不具备系统性的风险识别流程。

在开展风险评估时，企业必须采用系统性的方法，筛选AI用例引发的风险并加以分类。定量（情境分析、压力测试和关键风险指标等）以及定性（故障模式和影响分析、根本原因分析、专家人工判断等）工具及评估，能够帮助企业识别出应用AI的风险，包括公平性、可解释性、透明度、准确性、安全性和人为影响等因素。若能将上述这些工具及其他相关工具扩展至AI的全生命周期，贯穿整个价值链，企业就能够更加精准地识别自身面临的诸多AI风险并有效应对。

事项三：系统性地进行负责任AI测试

在全面测试负责任AI并扩展其应用规模时，企业会在AI生命周期和价值链中，广泛部署各类风险缓解措施，本次调研中，我们选取了44项风险测试和缓解措施作为指标来评估企业的成熟度。结果显示，只有19%的受访企业规模化地应用了半数以上的措施。

为了在整个企业范围内全面测试负责任AI并实现规模化应用，企业首先需要开发一个参考架构，实现客户端与其他第三方工具和服务的无缝集成，以评估AI全生命周期内的风险（包括数据、模型、应用程序）。

上述工作一旦准备就绪，企业便可在整个企业范围开展针对该架构的测试、微调、集成和部署工作。此外，企业还需要基于岗位职能提供相关培训（譬如客服专员、人力资源专员和管理岗位等），针对最新的负责任AI流程和工具，提升员工相应技能。

事项四：持续监测与合规

为了使负责任AI正确地发挥作用，企业首先应设立专门的办公室，负责规划、实施和管理自身的负责任AI行动计划。该办公室应当由来自不同学科、不同职能部门的成员组成，有明确的岗位分工以及责任划分。例如，“控制塔”人员可负责监测实时风险，其他人员则负责监测刚出现的新风险。

同时，AI监测与合规专门办公室还需配备先进的工具和方法，才能在实际工作中表现出最高水平。例如，当前某些工具已经能够监测AI模型本身，追踪性能指标并检测异常状况，漂移检测之类的方法能够审查数据，确保AI模型的准确性不会随时间推移而受到影响；用户反馈集成则有助于模型对使用者的需求作出更为妥善的响应。

事项五：关注员工、可持续发展、隐私和安全

要想成功实施负责任AI合规计划，企业应动员跨职能部门积极参与，妥善处理对全体员工队伍、可持续发展、以及隐私和安全计划的影响。

针对员工，企业必须确保员工在正确的时间、合适的地点掌握恰当的技能，并从负责任AI的角度出发，了解其具体的期望。参与本次调研的绝大多数企业（92%）都承认，员工（以及AI最终用户）在减轻风险方面发挥着重要作用。但这些企业同时指出，在提高运营成熟度方面，自身面临的最大挑战是提升人际互动能力，这也凸显了生成式AI时代对技能再培训的需求。

针对可持续发展，根据调研结果显示，只有36%的受访企业在组织层面制定了碳减排战略。未来一到两年内，新的测量及减排工具和技术将不断涌现，企业必须做好准备，采用这些工具和技术来有效管理AI碳排放，并降低云计算成本。

至于网络安全，成立专门的AI监测与合规办公室已刻不容缓。例如，71%的受访企业表示已制定以AI为重点的网络安全响应计划。但事件发生时，只有很少一部分企业（40%）指定了特定团队来执行其响应计划。尽管该领域目前仍处于起步阶段，但如果一味被动应对，企业将无法有效解决问题。

在整个组织范围内扩展负责任AI无疑是一项艰巨的任务。AI相关风险正呈爆炸式增长，以AI为焦点的全新法律法规则在各地纷纷出台，AI价值链变得日趋复杂，尤其是随着越来越多的企业同时成为AI模型的开发者和购买者，局面愈发复杂。

企业若想保持竞争优势，就必须主动出击，积极落实上述五大优先事项，将自身打造为负责任AI先锋。企业需要秉持前瞻思维，持续改进，并且超越自身边界，融入价值链上下游乃至更广泛的AI生态系统当中。

作者

阿纳布·查克拉博蒂（Arnab Chakraborty）
埃森哲首席负责任AI官

森蒂尔·拉马尼（Senthil Ramani）
埃森哲数据与人工智能全球主管

关于埃森哲

埃森哲注册于爱尔兰，是一家全球领先的专业服务公司，致力于帮助世界领先的企业、政府和其他组织构建数字核心、优化运营、加速营收增长、提升社会服务，快速且广泛地创造切实的价值。作为一家以人才和创新驱动的企业，我们拥有约79.1万名员工，为120多个国家的客户提供服务。技术是当今变革的核心，我们依托云、数据和人工智能方面的领先技术优势，凭借翘楚的行业经验、专业技能以及全球交付能力，并通过强大的生态系统关系推动全球技术变革。埃森哲战略&咨询、技术服务、智能运营、工业X和Song事业部拥有广泛的服务能力、解决方案和知识资产，我们基于共享成功的文化和创造360°价值的承诺，帮助客户重塑转型并建立长久互信的关系。我们为客户、员工、股东、合作伙伴以及社会创造360°价值，并以此为衡量自身的标准。

埃森哲在中国市场开展业务近40年，运营和办公地点分布在北京、上海、大连、成都、广州、深圳、杭州、香港和台北等城市。

了解更多，敬请访问埃森哲中文主页accenture.cn。

埃森哲大中华区主要办公室的联系方式：

上海
上海市黄浦区淮海中路381号
中环广场30层
邮编：200020
电话：86 21 2305 3333

大连
大连市甘井子区黄浦路953号
27-29号楼
邮编：116085
电话：86 411 6214 8888
传真：86 411 6214 8800

成都
成都市高新区天府大道中段
1366号天府软件园E5，9-10层
邮编：610041
电话：86 28 6555 5000
传真：86 28 6555 5288

杭州
杭州市滨江区西兴街道阡陌路
459号B楼1302-1303室
邮编：310051
电话：86 571 2883 4534

台北
台北市敦化南路二段207号16楼
电话：886 2 8722 0151
传真：886 2 8722 0099

北京
北京市朝阳区东三环中路1号
环球金融中心西楼21层1-7&16单元
邮编：100020
电话：86 10 8595 8700
传真：86 10 6563 0739

广州
广州市天河区天河北路898号
信源大厦8、13层
邮编：510898
电话：86 20 3818 3333

深圳
深圳市福田区深南中路3031号
汉国中心21层103单元
邮编：518033
电话：86 755 8270 5268

香港
香港鰂魚涌華蘭路18號
太古坊港島東中心2楼
电话：852 2249 2388
传真：852 2850 8956

敬请关注

埃森哲官方微信

《展望》微信小程序

业务联系

卢珊
埃森哲大中华区技术服务事业部董事总经理、
数据与人工智能主管
hilda.lu@accenture.cn

负责任AI