

从合规到信赖

采用新的思维方式，提升负责任的人工智能成熟度

>
accenture

内容

03页

执行摘要：重新思考负责的AI

第 05 页

高风险业务：当前风险格局中的三大挑战

第10页

从合规到价值：企业正认识到负责任人工智能的影响

第12页

成为重塑就绪：负责任人工智能成熟度的里程碑

第15页

责任现实检查：企业准备好负责的人工智能了吗？

第22页

准备好，开始成长：负责任人工智能的五大优先事项



执行摘要：重新思考负责的AI

随着生成式人工智能在商业和社会中的普及，使用它所带来的风险也随之增加。请考虑这样一个聊天机器人，它向客户提供了错误建议，从而给部署该聊天机器人的公司造成了责任。或者那个在使用ChatGPT后意外泄露了公司专有数据的员工。又或者那个错误地将数千人标记为欺诈的算法。

在这个新环境中，数据与人工智能的负责任开发和使用，成为组织既能最小化使用这些技术带来的风险，又能释放其诸多益处的关键推动力——从帮助个人更高效地完成任务，到让企业通过创新重塑自我，创造更多价值并获取竞争优势。

这当然说易行难，因为跨组织扩展负责任的人工智能面临的挑战是巨大的。随着生成式人工智能创造和加速数据及人工智能风险，人工智能相关的风险在持续累积。各地都在孕育新的聚焦人工智能的法律法规。而且人工智能价值链¹ 它们变得越来越复杂，尤其是随着越来越多的公司成为人工智能模型的开发者和购买者。

人工智能 101

什么是负责的AI？

负责的AI意味着采取有意行动来设计、部署和使用AI，通过防范AI的潜在风险来创造价值并建立信任。

什么是成熟的负责的AI？

已将负责任的人工智能工作作为平台全面实施，以采取更系统化、面向未来的方法，从而释放人工智能的真实价值。负责任的人工智能成熟度的北极星是成为先驱者。目前还没有公司达到这一阶段。

如何构建一个负责的AI程序？

我们的研究和为客户提供建议的工作表明，所有公司都可以通过专注于这五个重点来提高其成熟度，并开始获得人工智能带来的好处。

01

建立人工智能治理与原则

02

进行人工智能风险评估

03

系统性地支持负责任的人工智能测试

04

持续监测与合规

05

劳动力影响，可持续性，隐私，安全

我们的调查受访者估计，当一家公司成为负责任人工智能的先驱时，其人工智能相关收入将增加

18%
平均而言。

当今生成式人工智能的时代的到来，在负责任的人工智能方面提出了新的要求。要成为负责任人工智能的领导者，企业必须树立前瞻性思维，致力于持续改进，并将关注点从自身组织扩展到整个价值链和更广泛的AI生态系统。我们将这一新的成熟度水平称为“成为负责任人工智能的先驱”，目前还没有任何企业达到这个阶段。

调查1000家公司，涵盖19个行业和22个国家的高管。²

我们通过开发一个四阶段框架来评估公司负责任人工智能的成熟度——阶段越高，进步越大。然后将该框架应用于分析我们调查的1000家公司的组织成熟度和运营成熟度。

为了更好地理解公司对人工智能相关风险的立场以及他们对待负责任人工智能的方法，我们与斯坦福大学合作

值得称赞的是，大多数商业领袖都认识到负责任的AI在创造商业价值方面的重要性。

负责任的数据是负责任的人工智能旅程的有机组成部分

值得注意的是，当我们谈论负责任的AI时，负责任地使用数据是方程式的一个关键部分。在生成式AI时代，数据的重要性不可低估。大多数大型组织仍在应对围绕数据质量、可用性和治理的长期问题。在安永对2024年2000名商业领袖进行的全球调查中，48%的人表示他们的组织缺乏足够的高质量数据来实施其生成式AI计划。³ 具备高度数据准备度的公司拥有正确的数据，兼具合适的质量与数量。他们已建立可扩展的数据管理与治理实践，从而能够在流程中无缝使用数据，确保负责任的应用，并实现数据的经济价值。

[点击这里](#) 了解更多公司可以采取的关键行动，以提高他们在生成式人工智能时代的数据准备情况。

他们估计，当一家公司成为负责任人工智能的先驱时，其人工智能相关收入将平均增长18%。然而，我们的研究发现，我们调查中的绝大多数公司并没有像他们期望的那样为负责任人工智能做好准备。

多年的客户经验表明，企业可以采取行动，系统地运营负责任的AI。在本报告中，我们将探讨那些将提供组织基础设施的行动，以成为负责任的AI先驱，并释放AI的真正价值。

高风险业务：当前风险格局中的三大挑战

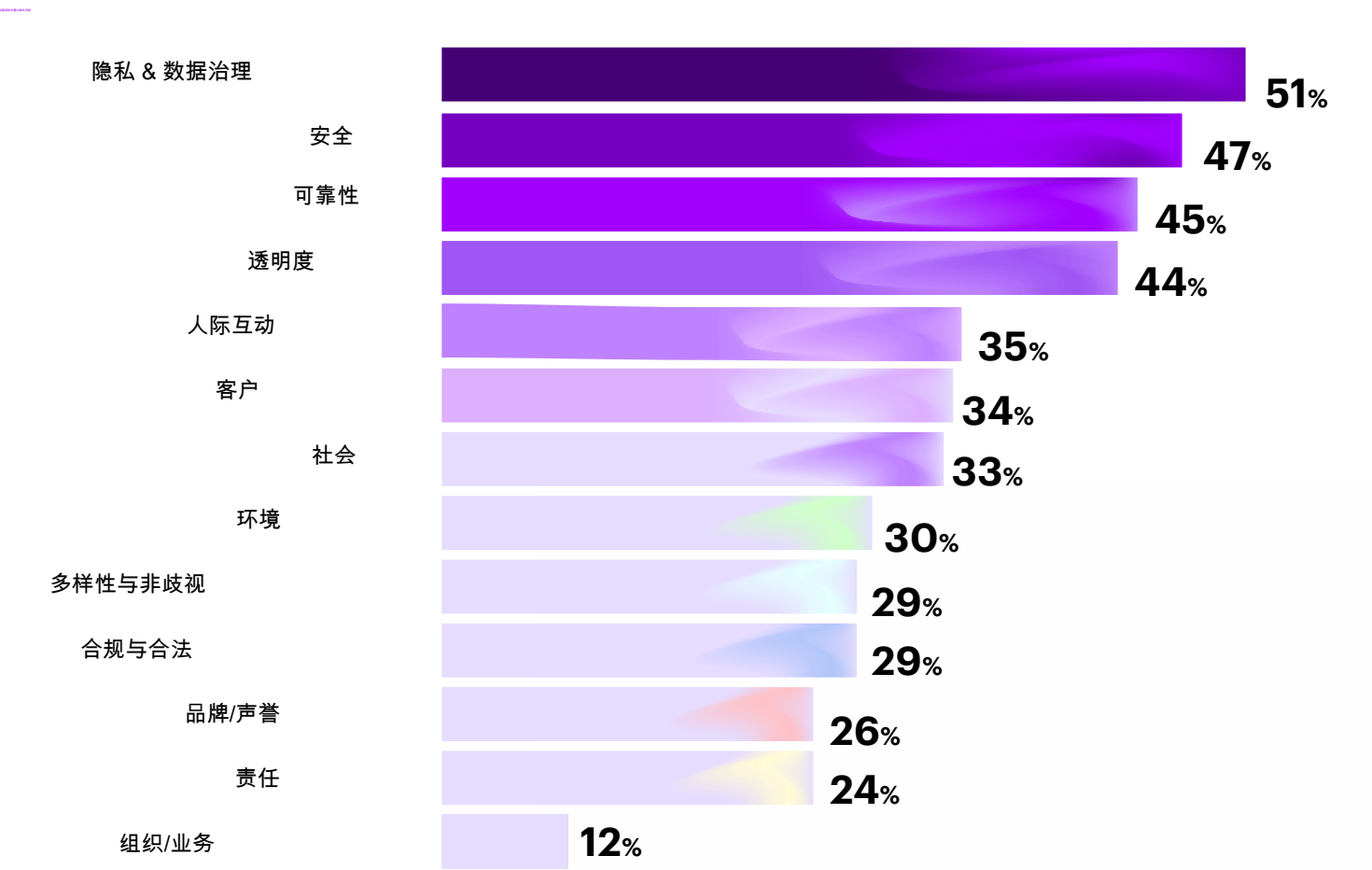
与AI相关的风险正在快速演变，尤其是随着生成式AI的广泛传播。仅在过去12个月内，我们就见证了新一波风险的涌现（例如与幻觉、知识产权（IP）、网络安全和环境影响相关的风险），而数据隐私、可靠性和透明度等更成熟的风险则再次凸显。这导致了日益复杂的风险格局。

我们发现风险格局将跨三个主要领域继续扩展和演变。

01 一个不断增加的范围和 风险频率	02 一个持续地 演化 监管 风景	03 一个扩展的 风险范围 管理 跨越价值 链
--	---	---

01 一个不断增宽的范围和 风险频率

如图1所示，最常被引用的风险涉及隐私和数据治理问题（ 51%的受访者将其列为公司面临的 风险 ）。安全性（ 47%的受访者提及 ）和可靠性风险，如输出错误、幻觉和模型故障（ 45%的受访者表示担忧 ），分别位居第二和第三。



来源：埃森哲斯坦福高管调查，N=1,000。详情请参见“关于此项研究”。

生成式人工智能对不断演变的AI风险格局的影响也体现在许多令公司最担忧的风险中。这些风险涵盖了从透明度、“黑箱”模型面临的挑战（ 44%的受访者引用 ），到环境以及数据中心能源需求（ 30%的受访者关注 ），再到责任归属以及侵犯版权和其他知识产权的担忧（ 24%的受访者引用 ）。欲了解更多关于我们评估的风险维度，请点击[此处](#)阅读我们与斯坦福大学合作撰写的报告。

近年来，随着AI风险的数量和类型急剧增加，由AI驱动的突发事件（ 偏见、深度伪造、幻觉、隐私泄露等 ）变得非常普遍，根据监测此类事件的AI事件数据库网站，2023年增加了32.3% ⁴ 这种趋势应引起企业的警惕。人工智能风险的复杂性在范围和频率上持续扩大，这得益于生成式人工智能的进步。因此，负责任的人工智能缓解策略也必须发展和变化，以便企业能够快速采用人工智能。在这种动态环境中，企业不能再对风险做出反应，它们必须学会预测风险。



02 一个不断发展的监管环境

这也确实如此，那些在负责任的人工智能方面落后的公司将会暴露自身面临日益增长的非合规风险，因为越来越多的政府开始规范人工智能。

欧盟的AI法案启动了引擎，目前已有超过37个国家正在考虑其他AI法规。随着各国政府不可避免地以不同方式选择监管AI，跨国公司合规性将变得越来越困难。这一现实已经影响到我们调查中的大多数公司，**77%要么已经面临AI监管，要么预计在接下来五年内将受到监管**。随着各国政府倾向于以不同方式监管人工智能，跨国公司可能会面临日益复杂和混乱的局面。

更重要的是，不仅人工智能监管机构需要应对。新兴生成式人工智能风险的出现正迫使政府通过和修订现有法律来采取行动。

这只会增加复杂性。实际上，几乎所有**(90%)接受调查的公司预计在未来五年内将受到与人工智能相关的法律义务，例如网络安全、与人工智能相关的法律、数据保护和消费者保护的约束**。这种立法在全国和地方各级进行，正在给企业带来进一步的合规挑战。

例如，中国、新加坡、巴西和沙特阿拉伯已经提出或引入了新的知识产权或版权法律。2023年，韩国引入了人工智能责任法，这早于欧盟计划的一项类似法律，而加利福尼亚州修订了加州隐私权法案（C PRA），赋予消费者选择退出个人数据出售和共享等行为的权利。深度伪造法已被法国、英国、澳大利亚和韩国提出或引入，许多国家正在调查新的人工智能网络安全法律。

77%

我们调查的许多公司要么已经面临人工智能监管，要么预计在未来五年内将受到监管。

90%

接受调查的公司中有预期将在未来五年内受到与人工智能相关的法律义务的约束，例如网络安全、人工智能相关法律、数据和消费者保护。

全球人工智能监管倡议（截至发布日期）

~列出的倡议并不详尽~

英国

• 人工智能监管白皮书2023 • 教育中的生成式人工智能2023 • 先锋人工智能工作组2023 • 英国人工智能安全研究所2023 • 英国人工智能安全峰会和布莱切利宣言2023 • 美英人工智能安全科学合作2024 • 英加人工智能安全科学合作2024 • 监管人工智能：信息专员办公室的战略方法2024

加拿大

• Bill C-27：数字宪章实施法案，包含人工智能和数据法案（AIDA 部分C）2022 • 人工智能和数据法案（AIDA）2023 • 2024年联邦投资240亿美元 • 加拿大人工智能安全研究所2024

法国

• 政府委员会对人工智能治理的2022年方法 • 法国人工智能委员会报告2024年 • 法国生成式人工智能计划2024年

美国

• 白宫来自15家领先的科技公司就某些安全和透明度要求的自愿承诺 2023 • 联邦选举委员会 - 关于在竞选广告中使用欺骗性人工智能的拟议法规 2023 • 国家电信和信息管理局（NTIA）关于人工智能问责制的征求建议书 2023 • 关于人工智能安全、可靠和可信赖的开发和使用的行政命令 2023 • NIST RMF 2024的GenAI草案 2024 • 人工智能安全研究所 2024 • 国家人工智能研究资源（NAIRR）2024

欧盟

• 人工智能责任指令 2022 • 数据治理法案 2022 • 数字市场法案 2023 • 欧盟人工智能法案 2024 • 数字服务法案 2024

德国

• 人工智能系统安全基础知识 2022 • 人工智能行动计划 2023 • 基于人工智能的机器人测试与认证中心 2024 • 行政和专业任务中人工智能的使用指南 2024

中国

• 个人信息保护法2022 • 互联网信息服务算法2022推荐管理规则2022 • 中国深度合成管理规定2023 • 生成式人工智能服务管理办法草案2024

日本

• 以人为本的AI社会原则 2019 • AI应用指南 2019 • AI治理指南 2022 • AI战略 2022 • AI战略委员会 2023 • 广岛进程 2023 • 日本AI安全研究所 2024 • 用于规范生成式AI的法律（提议） • 调节和使用生成式AI的国际框架 2024

新加坡

• 人工智能治理方法 + 实施自我评估指南 2020 • 人工智能治理测试框架最小可行产品（mvp）2021 • A.I. Verify 2022 • 新加坡国家人工智能战略 2.0（NAIS 2.0）2023 • 人工智能安全标签（提议）2024 • 人工智能资助计划以推动新加坡经济增长（2024） • 生成式人工智能的模型人工智能治理框架（2024） • 模型开发者和应用开发者的安全指南（2024） • 新加坡金融管理局（mas）原则以促进在人工智能方面的应用

沙特阿拉伯

• 沙特数据与人工智能管理局人工智能伦理原则（提议）2023 • 沙特数据与人工智能管理局 - 政府生成式人工智能指南 2024年1月10日 • 沙特数据与人工智能管理局公共生成式人工智能指南 2024年1月11日

印度

• 负责任的AI实施原则 2021 • 数字个人数据保护法案 2023 • 公开发布前批准AI工具 2024 • 印度AI报告 2023 • 印度AI任务框架 2024 • 通过监管沙盒鼓励创新技术的建议 2024

巴西

• 巴西人工智能战略 2021 • 人工智能法案提案 2021 • 人工智能法案第 2238 号 2023（拟议） • 2024国家数字政府战略

澳大利亚

• 安全可靠的AI（临时回应）（2024） • 政府人工智能保障国家框架（2024） • 政府负责任使用人工智能政策（2024） • 高风险环境中人工智能的强制护栏（提议）（2024）



03 价值链风险管理范围的扩大

不仅仅是风险类型和法规在变化。随着企业开始购买和发展人工智能模型，它们也需要为这两种行为带来的风险做好准备。例如，一个开发者可能会特别担心在以违反知识产权法的方式训练模型时被起诉。另一方面，一个购买者可能更担心新获得的人工智能模型能否按广告宣传的那样运行。当一家公司同时成为模型的购买者和开发者时，它面临的风险类型可能会呈指数级复杂化。

然而，今天，我们调查中近三分之一（28%）的公司同时扮演买方和开发者的角色（33%的公司只负责开发，38%的公司只负责购买）。

我们的研究也表明，许多公司尚未充分准备好应对其人工智能价值链日益增长的复杂性。例如，在被调查的获取人工智能模型的公司中，只有43%建立了完善的采购措施，例如监管检查和第三方审计。组织必须立即着手评估第三方人工智能产品或服务是否符合组织标准，并进行监控，以实现持续合规和风险管理。

随着第三方风险不断演变，企业必须超越自身的责任AI策略进行思考。现实情况是，企业必须尽职调查，并确保所有法律和监管责任都由价值链上的所有相关方达成共识并履行。对于高风险AI应用场景，仅仅以“非预期后果”来合理化结果是不够的。企业应预期在监管高风险应用场景方面，会受到客户和监管机构的追究责任。

只

43%

获取AI模型的公司都有采购措施，例如监管检查和第三方审计。

这种双向风险代表着人工智能领域又一个重大转变。直到最近，很少有公司既购买又开发人工智能模型。

从合规到价值：企业正认识到负责任人工智能的影响

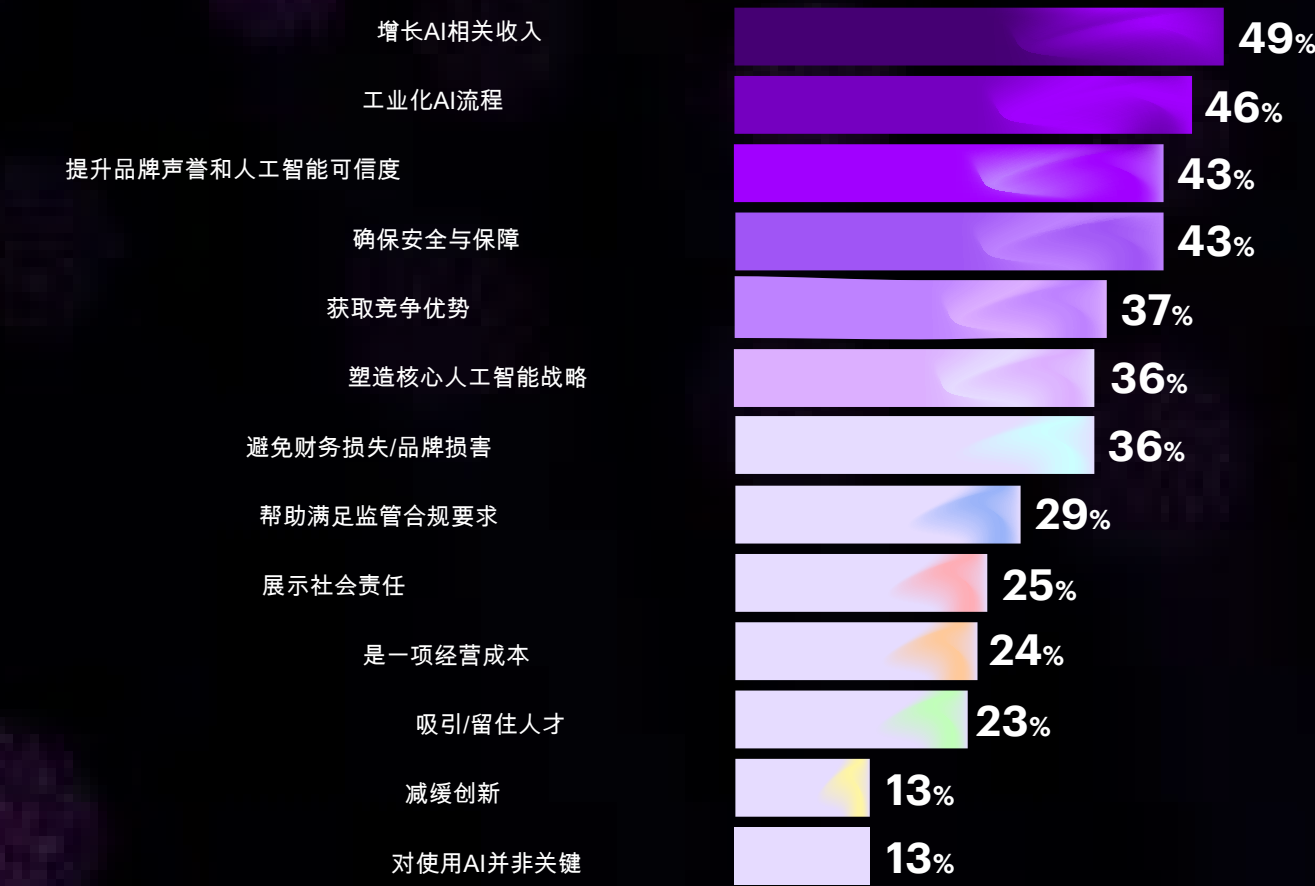
过去，企业常常将负责任的AI视为仅仅是合规问题，而非价值创造的必要贡献者。幸运的是，如今近半数受访公司并不持有这种观点。

例如，大约一半 (49%) 的我们调查的公司表示，他们认为负责任的AI是其公司人工智能相关收入增长的一个**关键因素**，而只有29%的公司表示负责任的AI主要是监管和合规问题（图2）。同样，43%的受访公司表示负责任的AI是保护其品牌价值的重要贡献者，而只有24%的公司将负责任的AI视为仅仅是“经营成本”。

49%

在我们调查的公司中，有%的公司认为负责任的AI是他们公司AI相关收入增长的关键因素。

图2：C级管理层的责任AI认知



来源：埃森哲斯坦福高管调查，N=1,000。详情请参见“关于研究”。

为了更好地了解负责任的AI能解锁多少金融价值，我们请高管们提供他们自己的估计。他们的答案表明，如今大多数公司都认为负责任的AI至关重要。例如，我们的调查受访者预测，当一个公司成为负责任的AI的先驱时，其AI相关收入将平均增加18%。（与人工智能相关的收入是由人工智能赋能的产品和服务产生的总收入。）

相反也如此。当负责任的人工智能在企业中未能良好建立时，其滥用会迅速摧毁品牌价值。而我们的调查受访者也估计

一个单一的、与AI相关的主要事件，平均将抹去其公司市值的24%。这些观点进一步证明了为什么公司正在加大负责任人工智能的投资。在我们调查的公司中，42%已经将超过10%的整体人工智能预算投入到负责任的人工智能计划中；在未来两年内，79%的公司计划达到这一稳固的支出目标。

简而言之，越来越多的公司正在优先考虑负责任的AI，并相应地进行投入。那么，他们实现负责任的AI目标还有多远呢？



成为重塑就绪：负责任人工智能成熟度的里程碑

在当今的商业环境中，持续变革是新的现实。准备好持续变革意味着你需要在企业的每个职能和每个组成部分中具备再造能力。负责任的AI也不例外。具备再造能力的企业能够保持敏捷，吸收市场上持续发生的转变，并能快速主动地应对这些转变，利用新机遇并减轻未预料到的后果。我们的研究表明，有四组企业正处于其负责任AI演进的明显里程碑上，其目标是为再造做好准备。

当我们谈到负责任的人工智能时，我们指的是什么“重塑准备”状态？有趣的是，还没有哪家公司达到那个里程碑，但率先抵达的人将成为负责任的人工智能先驱。他们将把负责任的人工智能工作完全转化为一个平台，以采取更系统化、面向未来的方法，从而释放人工智能的真正价值。

我们定义了四个负责任的人工智能成熟度里程碑，最终将使公司做好准备进行重塑。在这里，成熟度是一种持续演进的事物，而不是一个有限的状态或目的地。今天成熟的可能在未来几个月里会不同。

重新定义成熟

为了衡量企业的负责任人工智能成熟度，我们与斯坦福大学合作开发了一个四阶段框架。阶段越高，进步越大。根据我们对所调查的1000名高管的回应分析，我们将组织划分到各自阶段，并为组织成熟度和运营成熟度分别打分。

注意：没有负责任的人工智能计划的公司被排除在我们的分析之外。



负责任的AI成熟度里程碑

第一阶段	阶段 2	第3阶段	阶段4
设定负责任的AI原则	建立负责任的 AI 计划	将负责任的AI付诸实践	成为一名有责任心的AI先锋
该公司有一些开发人工智能系统的基础能力，但其负责任的 AI 努力是临时的： • 已制定一套人工智能伦理原则和指南，包括负责任和安全的数据访问和使用政策及规则 • 未建立管理数据质量、数据隐私、数据安全和人工智能模型风险管理的过程 • 偶尔对数据和人工作业项目进行风险评估审查 • 已部署人工智能项目工作流程，但在数据管道、模型管道和人工作业应用中未系统整合负责任的人工智能评估	经过负责任的AI评估，该组织已经采取了以下步骤： • 为组织建立负责任的AI策略，具有明确的运营模式和数据处理与AI管理措施，将愿景转化为行动 • 定义了跨数据管道、模型管道和AI应用的人工智能风险评估的稳健方法和流程 • 建立了创建训练数据、模型输入和输出透明度和可审计性的流程——并做出适当决策 • 设计了跨数据和AI管道的监控和控制框架，可在项目工作流程中执行 • 实施了仍处于早期阶段且没有更系统性的工具和技术支持的过程	该公司在整个组织范围内系统地实施了以下措施，以帮助满足相关的监管和法律义务： • 通过实施原则、指南和流程，将负责任的ai战略操作化，以在整个业务中实现赋能 • 在数据管道、ai模型和ai应用中实施了风险评估，以实现整个模型生命周期的可追溯性和透明度，并通过定期审计确保合规性 • 实施数据来源控制，包括隐私过滤、匿名化和验证，以移除敏感信息并降低嵌入自助工具中的数据偏见风险 • 通过模型可解释性工具实现系统性ai测试，以确保可解释性、对ai模型进行偏见、准确性等性能测试，并帮助确保模型在所需的法律法规、伦理和运营边界内运行 • 建立负责任的ai控制平面，配备人工控制，以在整个数据和ai价值链中进行持续监控，以提醒并纠正任何非预期的风险或违规行为 • 推出负责任的ai学院，用于员工培训和赋能，以推动负责任的ai应用	该公司将负责任的人工智能工作全面运营化，作为一个平台，采取更具系统性和面向未来的方法，解锁人工智能的真实价值： • 由技术平台驱动，配备合适的才能和文化，重新设计流程的端到端系统负责任AI完全运作的举措 • 采取了前瞻性的负责任AI方法，部署专用资源、流程等，以持续评估当前和未来的风险 • 随着外部技术和监管环境的扩展和演变，主动调整其数据和AI风险管理及控制流程 • 持续改进和推进数据治理和管理实践，采用预测分析和实时数据监控，动态理解和管理数据对AI系统的影响 • 塑造AI安全开发和使用的新的标准、方法和途径，包括数据隐私保护、模型可解释性和偏差测量、对抗性测试和红队测试 • 被认为是塑造负责任AI实践的领导者，并积极与外部利益相关者合作——包括价值链合作伙伴、监管机构和相关社区——以确保参与、包容性反馈和支持前瞻性监管合规 • 积极与第三方AI供应商合作，促进改进，建立信任关系和新的合作机会，以有效管理第三方AI风险



负责任的AI成熟度里程碑：示例*

第一阶段	阶段 2	第3阶段
设定负责任的AI原则	建立负责任的 AI 计划	将负责任的AI付诸实践
一家亚洲制造公司想要防范未经控制的人工智能使用带来的风险。 它建立了一项内部人工智能政策和检查清单，以确保安全、安全、公平、透明度和问责制——但需要帮助正确实施，包括培训员工。 该公司开发了一部操作手册，以支持人工智能伦理审查流程，将现场操作员完成清单所需的时间减少了一半。现在，该公司能够快速识别人工智能伦理风险，并将人工智能伦理治理融入其业务核心——并负责地实现其愿景。	一家跨国消费医疗保健公司希望制定一个明确的政策和愿景，以负责地扩展人工智能 跨企业标准化构建和部署人工智能的流程和方法。 这家公司没有高风险人工智能应用的库存，并且在与专门负责人工智能的角色和决策责任缺失作斗争。 他们与埃森哲合作，针对监管环境和草拟的人工智能原则和政策进行了全球基准测试和评估，并提出了一个负责的人工智能运营模式。公司对其所有人工智能应用进行了风险筛查。对高风险案例进行风险评估，以确保其应用符合原则和监管要求。他们还与第三方律师合作，为立法监测提供一个框架。 该公司现在对其人工智能的使用，包括风险管理和问责制，充满信心，以至于它计划发布一份关于负责任人工智能的外部立场文件。	在过去的几年里，埃森哲已经完成了自身的努力来增加负责任的人工智能成熟度，建立在其现有人工智能原则。 该公司开发了一个严格的负责 AI 项目并将其付诸实践。 鲁棒程序的关键要素是： • 领导监督： 埃森哲任命了首席负责任人工智能官，负责监督内部负责任人工智能合规计划，由埃森哲首席执行官和总法律顾问担任赞助人。该公司的首席技术官、首席负责任人工智能官⁶ 总法律顾问和数据与人工智能负责人负责监督相关指导委员会。此外，还有定期向审计委员会汇报进展的情况。 • 原则与治理： 埃森哲在开发和部署人工智能解决方案方面所采取的方法基于一套原则⁷ 那些应用于其自身运营以及其与客户、合作伙伴和供应商的合作。公司任命了一个跨职能团队（包括法务、安全、CIO、采购、人力资源和负责任的AI专家等）设计和来领导新的负责任的AI合规计划，并以负责任的AI原则作为设计和指南。 一个实施关键原则、政策和标准的治理框架支持跨用例监督。 • 风险评估和缓解： 埃森哲根据欧盟人工智能法案和其他关键框架采用基于风险的方法。这意味着在前期投入大量时间来定义高风险用例。在过去的一年里，埃森哲筛选了公司范围内的数千个人工智能用例，并完成了详细的风险评估和缓解。 • 测试和启用： 该公司还花费时间进行系统性赋能——设计其人工智能政策、标准和控制，并将它们嵌入到技术、流程和系统中。这需要构建负责任的智能技术能力、专业知识、工具和技术，并开发基准测试工具。 • 人才： 埃森哲也在提升其负责任人工智能技能和人才，为那些最直接参与人工智能的埃森哲员工提供强制性的伦理和合规培训，以及通过埃森哲技术商数（TQ）课程为其774,000名员工提供新的伦理和人工智能培训。

* 这些示例中的公司在各自阶段实现了主要要素。负责任人工智能成熟度的北星是成为先驱者。目前还没有公司达到这个阶段。



责任现实检查：企业准备好负责任的人工智能了吗？

尽管取得了良好进展，但公司可能正面临着意图与执行之间的感知差距，因为即使那些自认为已经成熟的公司也还有很长的路要走。我们的研究发现，公司可能仍然低估了自身面临的风险数量、所需措施的规模以及实施措施的完整性。如果没有一套完善的风险缓解措施，组织不仅会面临现有风险，还将难以适应新法规、预见新出现的风险以及自信地扩展人工智能机遇。

运营成熟度与组织成熟度

为了提升负责任的AI成熟度，企业必须将组织成熟度转化为一套全面的缓解措施，即运营成熟度。我们观察到，在过去两年中，组织成熟度虽然持续增长，但与运营成熟度之间存在显著的脱节。

运行成熟度

运营成熟度目前是大多数公司的一个重大弱点。运营成熟度衡量一家公司采用负责任的AI措施来减轻特定AI相关风险的程度，包括与隐私和数据治理、多样性和歧视、可靠性、安全、人机交互、问责制和环境影响相关的风险。运营成熟度高得分表明，公司在所有适用于其业务的AI风险领域都擅长负责任地实施AI。可惜，似乎只有少数公司成功实施了负责任的AI。我们发现 **在我们调查的公司中，有6%的公司已经达到运营成熟度的实践里程碑，而不到1%的公司处于开拓阶段**（图3）

负责任的生成式AI

当涉及到生成式AI时，我们分析了公司缓解开发人员（一些公司）和技术（几乎所有的公司）用户风险的能力。

对于生成式AI用户，我们评估了AI生命周期每个阶段应用的缓解措施，即供应商选择、评估、基础设施、应用、最终用户支持和监控、控制和可观察性。对于开发人员，我们关注AI基础设施、模型开发和评估、应用开发、最终用户措施以及部署后的监控、控制和可观察性。我们发现生成式AI和通用AI的操作成熟度相似地较低（图3）—— **仅有13%的公司处于实践阶段（10%的公司）或先锋阶段（3%的公司）**。

000000

组织成熟度反映了组织负责任的人工智能流程和实践的程度及有效性——由首席领导层支持、治理、数据和人工智能风险识别和管理、模型开发、采购、监控和控制、网络安全和培训衡量。组织成熟度高得分表明该公司已准备好负责任地使用人工智能。

对于我们调查的许多公司而言，组织成熟度是一种相对优势（注：分数反映公司自我报告的能力）。如图3所示，72%的公司处于实践阶段或先驱阶段（分别占63%和9%）。与此同时，只有3%的公司处于原则阶段，25%的公司处于项目阶段。与两年前相比，企业高管对负责任人工智能的支持率从50%上升到79%，而拥有完全运营的风险管理框架的公司比例从48%上升到69%。这表明公司开始认真对待负责任人工智能，并采用自上而下、全组织范围的 approaches，这是达到成熟度里程碑的基础步骤。

优先考虑负责任的AI计划对成熟度有直接影响

我们的分析表明，投资负责任的AI与在运营、生成和组织三个维度上更高的成熟度呈正相关：负责任的AI运营成熟度、负责任的AI生成成熟度。在负责任的AI上花费超过10% AI预算的公司更有可能在运营成熟度上处于实践和创新阶段，并且更有可能在负责任的AI生成成熟度上处于实践和创新阶段，可能性分别是两倍和三倍。



整体负责AI成熟度

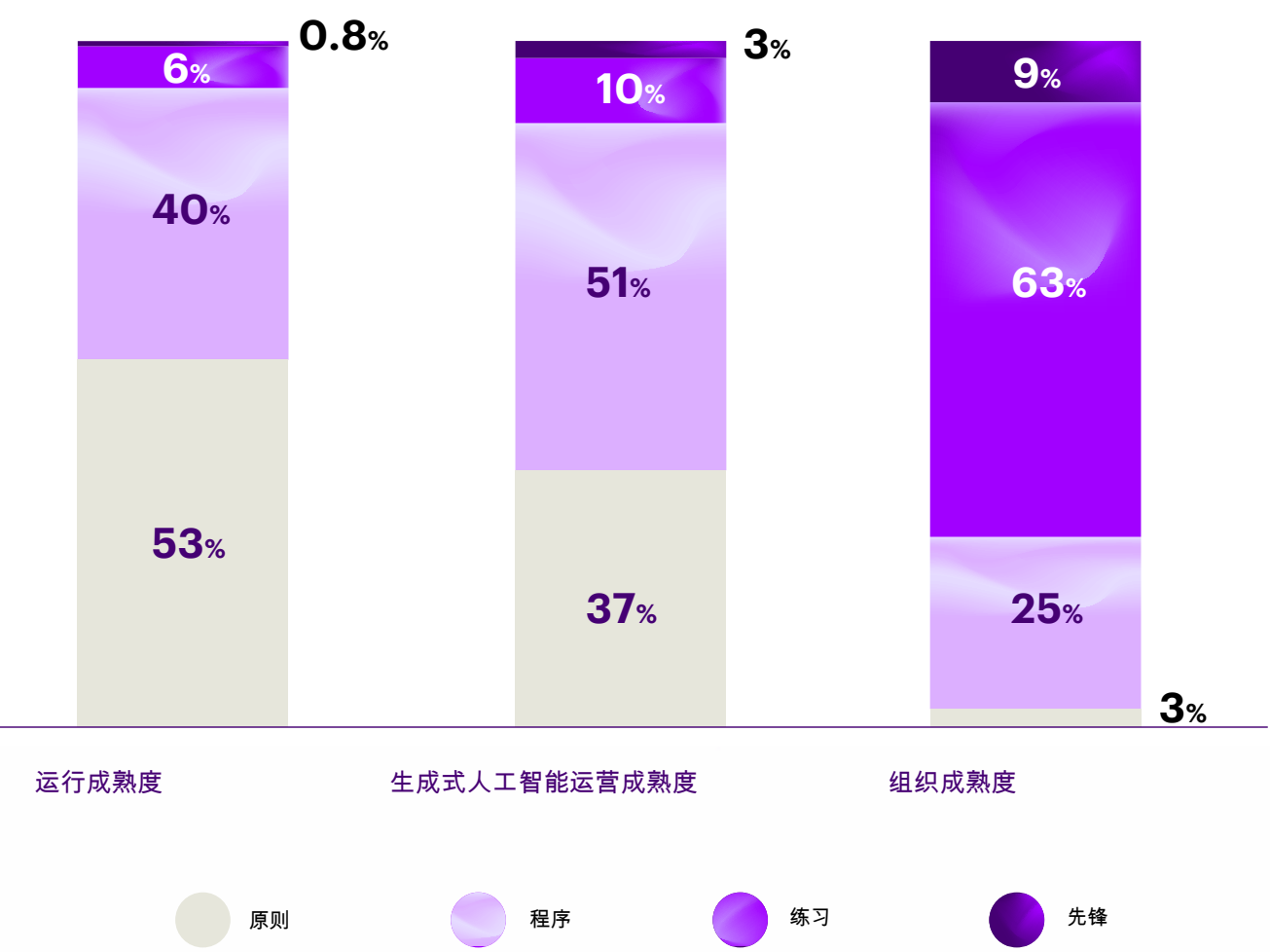
为了全面了解当前负责任人工智能成熟度状况，我们设计了一个框架，该框架采用组织成熟度、运营成熟度和生成式人工智能成熟度的综合评分。

我们的研究揭示，绝大多数（78%）公司已建立负责任的AI项目*。较小比例，14%，已将负责任的AI付诸实践，而8%正开始他们的旅程，通过制定负责任的AI原则。值得注意的是，没有一家公司已成为先驱。

然而，这种总体观点掩盖了组织成熟度与运营成熟度之间存在的显著差异（图3）。例如，尽管在组织和运营层面存在先驱者，但当您合并评分时，您将没有任何先驱者。组织需要确保他们优先考虑组织成熟度和运营成熟度。组织成熟度是自然开始的起点，因此公司在这一方面通常更先进。但他们必须记住，采取系统的方法来对待运营成熟度也是至关重要的（因为那时您才真正在组织中实施这些措施）。否则，可能会出现认知差距，这将使公司暴露且无法适应变化的风险格局。

我们的分析发现，目前不同行业在负责任人工智能成熟度方面没有显著差异，这表明各行业在负责任人工智能方面采取了一致的方法。然而，我们观察到亚洲的公司，例如印度和新加坡的公司，在负责任人工智能成熟度方面处于领先地位，与它们的全球同行相比，展示了更先进的做法。在欧洲国家中，德国脱颖而出，其公司在负责任人工智能成熟度方面比该地区其他国家表现更高。在许多情况下，这些国家已经建立了更广泛的AI政策、投资和监管，这很可能有助于推动当地采用。

图3：按运营、生成式人工智能和组织领域划分，各公司在四个负责任人工智能成熟度级别的百分比分布



来源：埃森哲斯坦福高管调查，N=1,000。详情请参见“关于此项研究”。

什么样子是好的：有责任感的AI先锋的标志

当谈到负责任的AI成熟度时，达到先驱阶段应该是新的目标。 如所注，我们调查的少数公司（9%）是组织成熟度负责任人工智能先锋；甚至更少（不到1%）是运营成熟度负责任人工智能先锋。 当结合运营和组织成熟度时，没有一家公司成为先驱。

从地理角度来看，亚洲在全球处于领先地位，拥有最多的先锋企业，这得益于新加坡、日本和印度的强劲表现。亚洲占据了组织先锋的34%和运营先锋的37%。在欧洲，德国是区域领导者，特别是在组织成熟度方面。考虑到这些国家要么已经实施，要么正在探索政府层面的监管，这与我们的预期一致。在北美洲，监管正零星地出现在各个州。

在市级或行业层面，我们看到它们在运营成熟度方面落后于亚洲，甚至落后于一些新兴地区。

从行业角度来看，通信、媒体和技术（尤其是在高科技和软件与平台领域）在成熟度方面最为先进，其中27%的公司运营成熟度上有所提升，20%的公司组织成熟度上有所提升。在金融服务领域，保险和银行业最为成熟。

这些发现并不令人意外，当你考虑到金融服务行业受到高度监管，而软件和平台玩家大体上认为自律对其未来至关重要时。他们已经不得不建立流程来监控和缓解他们正在使用的AI。最后，零售和工业设备行业显示出很强的成熟度，而消费品和服务行业则落后。

9%

有%的公司已达到组织成熟度的先驱里程碑，而对于运营成熟度，这一比例降至不到1%。

当结合组织成熟度和运营成熟度时，还没有哪家公司是先驱。

什么使负责任的AI先锋与众不同？

随着我们进入生成式人工智能时代，希望将负责任的AI作为价值杠杆的组织必须建立在它们的组织成熟度和运营成熟度之上。采用面向未来的心态，持续改进和演进以符合新技术进步和监管变化将是必不可少的。负责任的AI先驱是：

预期器

为了成为有远见者，开拓者会随着外部技术和监管环境的变化，持续调整其风险监测和控制流程。他们系统地将其人工智能活动与组织的整体战略目标和风险容忍度相结合。并且他们拥有专门的团队和内置机制，以持续评估和管理当前和未来的风险，为快速采用新技术进步搭建平台。

当其他公司在部署最新的AI进步时犹豫不决——例如“代理式”AI（表现出越来越高的自主性和独立行动能力的AI系统）——出于对风险的担忧时，先驱者可以自信地向前迈进，因为他们知道自己的负责任AI工作已经为挑战做好了准备。

设计负责

先锋者们以“设计赋予责任”为核心，将负责任的AI置于其总体AI战略的中心。他们认识到负责任的AI对于日常运营至关重要，它提供了一个强大的跨职能、组织框架，能够实时支持决策，并自信地支持新技术的采纳和扩展。

先锋被视为人工智能创新的领导者，通过不断精炼和改进其负责任的AI治理结构、原则、政策与标准，投资于未来规划和前瞻性洞察，从而保持其竞争优势。

主动合作伙伴

先驱者也与合作伙伴紧密合作，以管理风险、支持多方利益相关者的参与、反馈和包容性决策以及支持前瞻性的监管合规。例如，先驱者组织和领导其更广泛的AI生态系统中的利益相关者之间的坦诚讨论。此类集会的目的是建立信任关系，促进负责任AI最佳实践 分享，同时为需要技术支持的合作伙伴提供渠道。

通过与合作伙伴在负责任的AI领域合作，先驱者建立相互信任，增强自身的成熟度，并支持新的商业机会。

准备好，开始成长： 负责任人工智能的五大优先事项

根据我们的数据，各组织在负责任人工智能成熟度旅程上已达到非常不同的里程碑。处于负责任人工智能成熟度原则阶段的公司，在寻求提升其负责任人工智能能力时，其优先事项必然与处于项目或实践阶段的公司不同。然而，我们的研究和为客户提供建议的工作表明，所有公司都可以通过关注这五个优先事项来提升其成熟度，并开始获得人工智能带来的好处。

01 建立AI治理和原则	02 开展人工智能风险评估	03 系统启用 负责任的AI测试
04 进行中 监控和合规性	05 劳动力影响， 可持续性 隐私，安全	

建立人工智能治理与原则

制定包含清晰政策、指南和控制的负责任人工智能战略和路线图，对于负责任人工智能实践的成功实施、运营和管理至关重要。

本过程的基础步骤是制定、采纳和执行一套负责任的AI原则（具体针对贵公司的优先事项、道德和价值观），以确保AI的道德设计、部署和使用。我们的研究表明，70%的公司已经建立了负责任的AI原则，其中54%已将这些原则转化为政策。

在负责任的

构建能力和控制是另一个重要步骤。

好消息是，76%的公司已经完全实现了其治理模式——比两年前的31%有大幅增长。这反映了公司采取自上而下、跨职能的方法来负责任的AI，而不是试图以临时的方式管理风险。

企业可以更进一步，通过提供员工培训和变革管理举措来强化这一运营模式，从而在组织中推动和持续发展负责任的AI。





优先级2

进行人工智能风险评估

理解组织使用人工智能所面临的风险敞口是负责任人工智能运营的关键组成部分。如果你不知道你在哪里面临风险，你如何能保护自己？

我们调查的大多数公司似乎低估了他们面临的与人工智能相关的风险。例如，当我们向调查受访者展示13种与人工智能相关的风险，并询问他们公司的哪些风险让他们担忧（可以多选）时，公司平均只选择了4.4种风险。这种对人工智能风险的低估在另一个发现中可见：我们调查的超过50%的公司还没有建立一套系统的风险评估流程。

在实施风险评估时，采用系统化方法对人工智能用例的风险进行筛选和分类非常重要。定量（情景分析、压力测试和关键风险指标）和定性（故障模式与影响分析、根本原因分析、专家人类判断）工具和评估可以帮助突出一个组织使用人工智能的风险，包括公平性、可解释性、透明度、准确性、安全性和人类影响。将这些和其他工具扩展到人工智能的生命周期和价值链中，将帮助公司更好地识别和应对他们面临的大量人工智能风险。

优先级 3

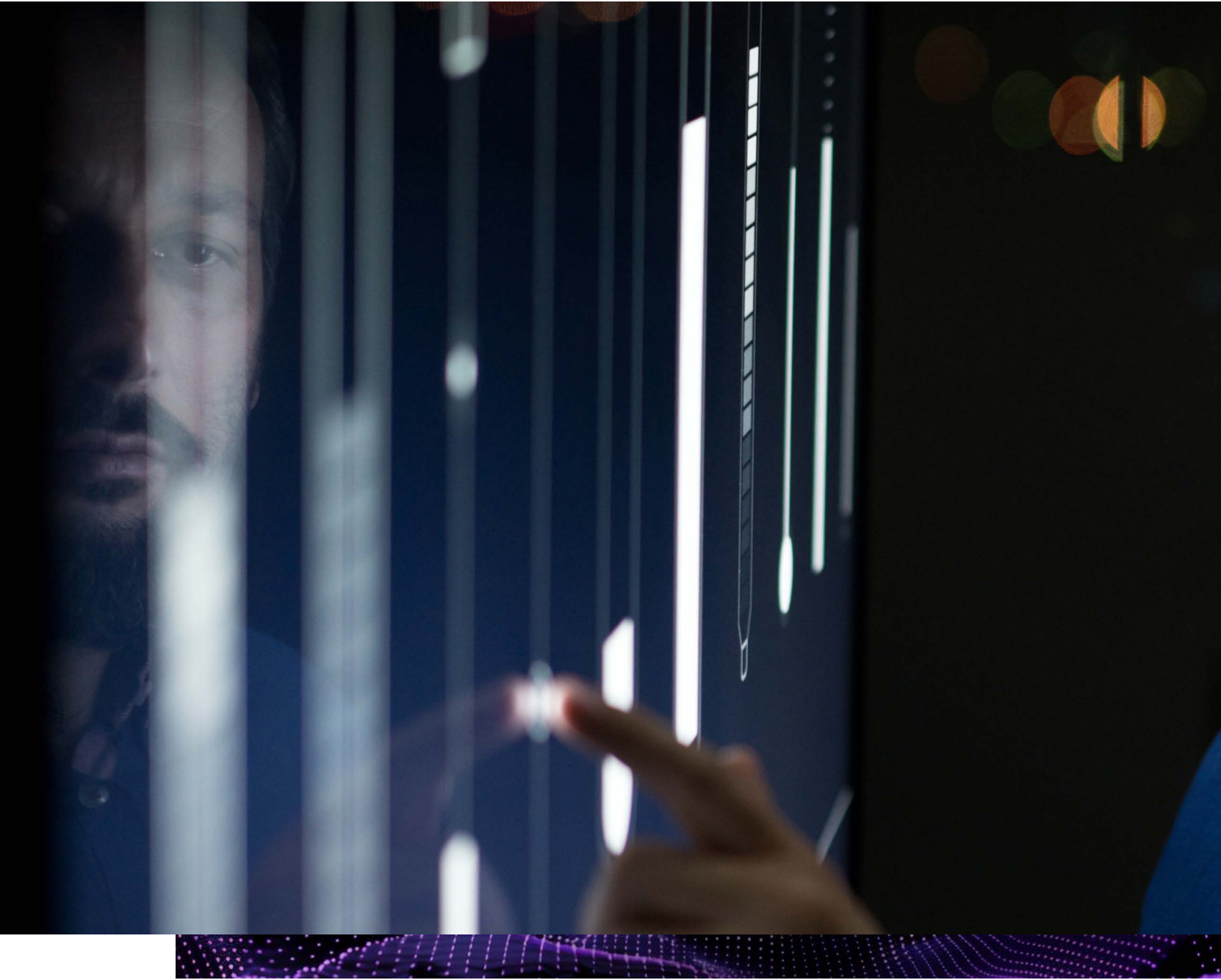
系统性地支持负责任的人工智能测试

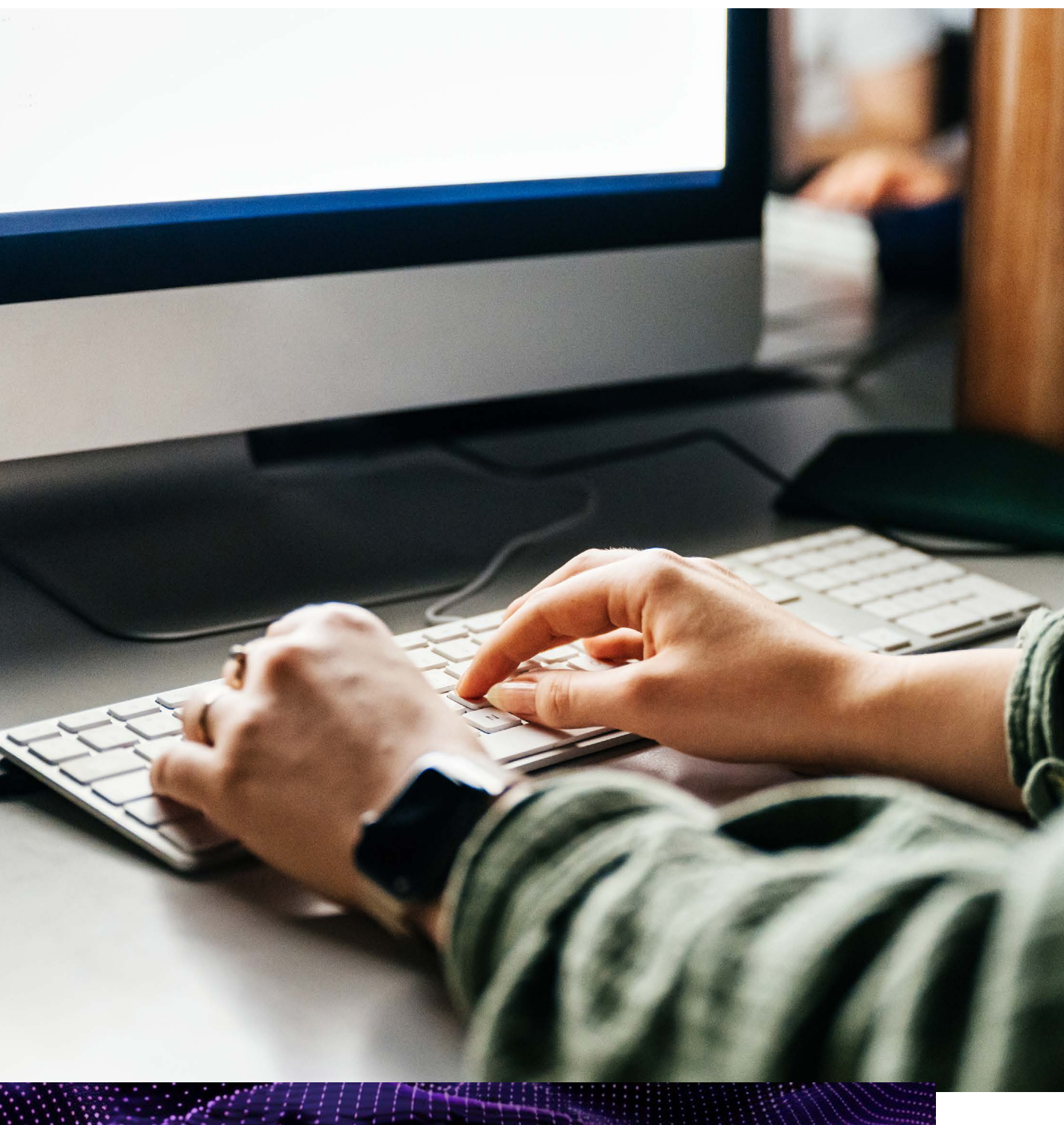
当企业全面测试和规模化负责任的AI时，它们在AI生命周期和价值链中部署了广泛的风险缓解措施（我们在调查中评估了公司在这44项措施上的成熟度）。然而 **仅有19%的受访公司已扩展了超过一半的风险测试和缓解措施** 那是我们向他们询问的。

集成客户和其他第三方工具和服务，以在整个人工智能生命周期（数据、模型、应用程序）中评估风险。

在此情况下，公司可以测试、微调、集成并跨整个组织部署该架构。此外，他们还需要提供基于角色的培训（例如针对客户服务专业人员、人力资源专业人员和管理角色），以增强员工在使用最新负责任的人工智能流程和工具方面的技能。

为了全面测试和跨组织扩展负责任人工智能，公司需要首先开发一个参考架构，以实现无缝





持续监测与合规

建立专门的AI监控与合规职能对于确保组织内AI模型的合规、合乎道德和可持续性表现至关重要。这一步骤在涉及生成式AI应用时尤为重要，因为目前数据及模型透明度较低，且幻觉、偏见以及IP或版权侵犯等事件发生的频率远更高。随着非预期和持续的后果不断出现，以及公司部署能够自主采取行动和决策的模型或代理，监控能力将变成一项必需。

尽管如此，**43%的公司尚未完全实现其监控和控制流程**，使其成为组织成熟度最薄弱的环节。此外，**52%的生成式AI用户尚未有任何监控、控制和可观察性措施。**

要这样做，公司应首先设立一个负责规划、实施和管理组织负责任人工智能计划的办公室。该办公室应由来自不同学科的跨职能团队组成，人员职责明确，责任层级清晰界定。例如，“控制塔”中的办公室人员可能负责实时监控风险，而其他人员可能负责监控未来风险。

一个专门的AI监控与合规办公室也需要配备先进的工具和方法论，以最高水平进行工作。例如，如今的某些工具可以监控AI模型本身，追踪性能指标并检测异常。像漂移检测这样的方法论可以审查数据，以确保AI模型在一段时间内保持准确；用户反馈集成可以使模型更加响应使用者的需求。

获得负责任的AI也需要有专职为此奋斗的员工。

劳动力影响，可持续性，隐私，安全

要实现一个成功的、负责任的AI合规计划，跨职能协作必须解决对员工的影响、可持续性以及整个企业范围内的隐私和安全项目。

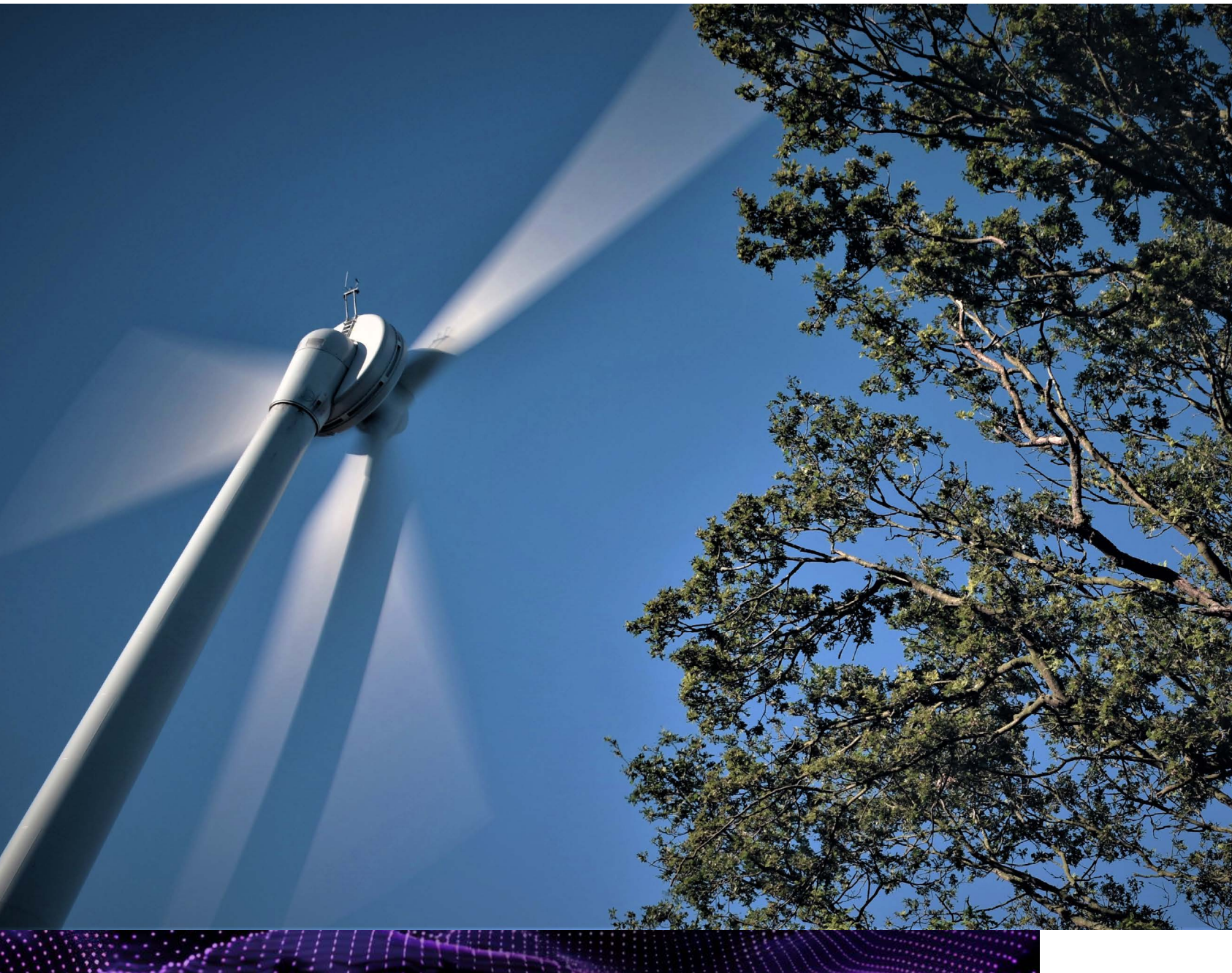
然而，这些公司报告说，开发人类交互能力是他们面临的最大的挑战，以提升其运营成熟度，突出了在生成式人工智能时代进行再培训的必要性。

劳动力

人才是全球领导者最优先考虑的事项，也是转型成功最大的阻碍之一。负责任的AI也不例外。企业必须确保其人员具备正确的技能，在正确的地方，在正确的时间，并了解从负责任的AI角度出发的期望。在我们调查的公司中，绝大部分（92%）承认员工（以及AI的最终用户）在减轻风险方面发挥着重要作用。

正如Accenture近期与MIT的合作所展示的⁸，员工是风险 mitigation 的第一道防线，但要有效防御，他们需要支持以克服潜在障碍（如对风险的意识不足或自满）。结合持续性的高质量培训和技能再培训，以及行为科学、责任设计和技术干预的跨学科方法，对于确保优化的风险 mitigation 和整体绩效至关重要。





围绕AI碳排放的意识不断增强，随着研究人员超越大型语言模型训练，着眼于推理以及更广泛的AI生命周期，以了解潜在问题所在以及可应用的解决方案。例如，训练单一多语言语言模型BLOOM产生了24.7吨⁹ 温室气体排放量相当于15次纽约与伦敦之间的往返航班。¹⁰ 因此，亚马逊、谷歌和微软都宣布计划采用核能来支持日益增长的能源需求。

我们的调查显示 **只有 36% 的公司**在组织层面制定了**碳减排策略** 随着时间的推移，在接下来的12-24个月里，新的测量和缓解工具和技术将出现，企业必须确保他们能够采用这些工具和技术来管理人工智能碳排放并降低云计算成本。

网络安全

为帮助缓解网络安全风险，设立专门的AI监控与合规办公室的需求尤为迫切。例如， **71%的受访公司**称他们已制定以**人工智能为重点的网络安全应对计划**。但极少公司（**40%的企业**）指定了专门的团队**来实施其应对计划**，以防发生事件。尽管这个空间仍然处于起步阶段，公司无法承担采取被动应对策略的代价。

结论：将AI风险转化为商业价值

如果负责的AI曾经仅仅是一个事后补救措施，那么那些日子已经一去不复返了。如今，公司知道为了最大化他们在生成式AI和其他AI技术上的投资，他们需要将负责的AI放在首位。

这当然说起来容易做起来难，因为跨组织规模化负责的AI面临的挑战是巨大的。AI相关的风险正在激增。各地都在孕育新的专注于AI的法律法规。而且，随着越来越多的公司同时成为AI模型的开发者和购买者，AI价值链的复杂性正在增加。

企业若想保持竞争力，必须拥抱上述五大优先事项，并成为负责的人工智能先驱。作为这些努力的一部分，企业必须追求前瞻性思维，致力于持续改进，并将它们的关注焦点从自身组织扩展到整个价值链和更广泛的AI生态系统。

成为负责的AI先锋的回报将相当可观：始终将AI风险转化为巨大的商业价值。

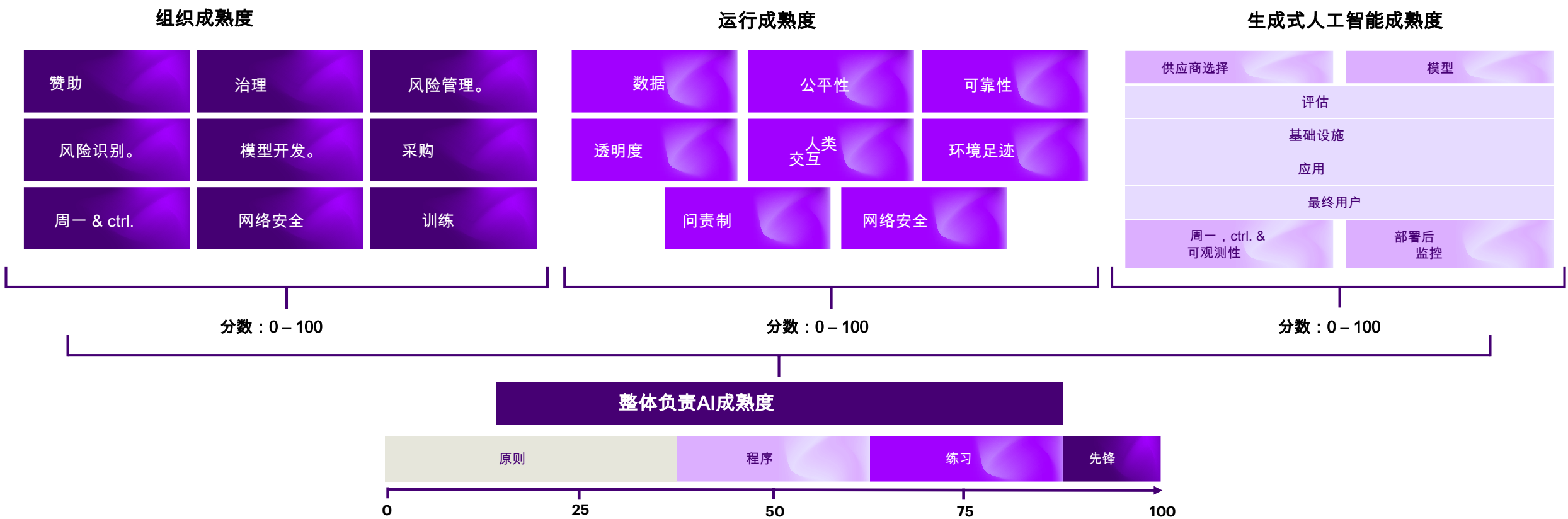
关于研究

2024年1月至3月，埃森哲对总部设在22个国家的19个行业的1000家公司进行了调查。样本包括一群多元化的商业领袖，包括首席执行官、高管、董事会成员和董事。调查问卷与斯坦福大学共同开发，从而能够采取稳健和全面的方法。负责型人工智能成熟度指数的方法论基于三个核心支柱：组织成熟度、运营成熟度和生成式人工智能成熟度。这三个支柱被赋予相等的权重，它们的综合得分构成了整体负责型人工智能成熟度指数，提供了对企业在人工智能领域准备情况的全面了解。

区域	样本
北美洲	27%
亚洲	22%
欧洲	30%
中美洲	9%
其余世界	12%
总计	100%

行业（分组）	样本
航空航天、汽车与交通	16%
通信、媒体与技术	16%
金融服务	15%
医疗保健与生命科学	10%
产品	16%
公共服务	6%
资源	21%
总计	100%

关于埃森哲-斯坦福负责任人工智能成熟度评分模型



作者



阿南布·查克拉博尔蒂，埃森哲首席负责人工智能官



卡齐克·纳尔因
集团首席执行官—技术和首席技术官，埃森哲



Senthil Ramani
全球领导者—数据与人工智能，埃森哲



致谢

研究主管

帕特里克·康诺利

研究团队

菲利普·罗西耶 普雷文·唐塔鲁基 杰库布·维亚特拉克 希克哈·蒂瓦里 德瓦贾·帕蒂尔 迪克希塔·文卡塔什 大卫·金布尔

市场团队

Sophie Burgess Erika Marshall Dries Cuyper s Micaela Soto Acebal Mark Klinge

参考文献

1. 一个“AI价值链”不仅包括公司的内部流程，还包括其合作伙伴和客户在创建、部署和维护AI系统中的流程。
2. 这19个行业是：航空航天/国防、汽车、银行、资本市场、化工、电信/媒体/娱乐、消费品和服务、能源、医疗保健、高科技、工业设备、保险、生命科学、自然资源、公共服务、零售、软件/平台、旅游/交通和公用事业。这22个国家是：阿根廷、澳大利亚、巴西、加拿大、中国、丹麦、德国、芬兰、法国、印度、意大利、日本、墨西哥、挪威、沙特阿拉伯、新加坡、南非、西班牙、瑞典、阿拉伯联合酋长国、英国和美国。
3. <https://www.accenture.com/us-en/insights/data-ai/new-data-essentials>
4. <https://incidentdatabase.ai/>
5. <https://www.comparitech.com/blog/information-security/data-breach-share-price-analysis/>
6. <https://newsroom.accenture.com/news/2024/accenture-takes-new-steps-to-help-clients-scale-generative-ai-responsibly>
7. <https://www.accenture.com/us-en/case-studies/data-ai/blueprint-responsible-ai>
8. <https://sloanreview.mit.edu/article/nudge-users-to-catch-generative-ai-errors/>
9. <https://sloanreview.mit.edu/article/how-developers-can-lower-ais-climate-impact/>
10. <https://www.jmlr.org/papers/volume24/23-0069/23-0069.pdf>

关于埃森哲

埃森哲是一家领先的全球专业服务机构公司，帮助全球顶尖组织构建其数字核心、优化运营、加速收入增长并提升服务——以速度和规模创造切实价值。我们是一家人才与创新驱动的公司，拥有77.4万名员工，为120多个国家的客户提供的服务。当今变革的核心是技术，而我们是帮助推动这一变革的世界领导者之一，并拥有强大的生态系统关系。我们将我们在技术方面的优势与在云、数据和人工智能方面的领导力相结合，并拥有无与伦比的行业经验、职能专业知识和全球交付能力。我们的战略与咨询、技术、运营、行业X和埃森哲之歌等广泛的服务、解决方案和资产，加上我们共享成功的文化和对创造360°价值的承诺，使我们能够帮助客户成功并建立信任、持久的关系。我们通过为我们的客户、彼此、股东、合作伙伴和社区创造的360°价值来衡量我们的成功。

访问我们 www.accenture.com。

关于埃森哲研究

埃森哲研究围绕组织面临的最紧迫的商业问题创建思想领导力。结合创新研究技术，如数据科学引领的分析，以及对行业和技术深刻理解，我们分布在20个国家的300名研究人员每年发表数百份报告、文章和观点。我们与世界领先组织共同开发的思想启发研究，帮助客户拥抱变革、创造价值，并实现技术和人类智慧的力量。欲了解更多信息，请访问埃森哲研究网站。 www.accenture.com/research。

免责声明：本文件中的内容反映了该文件准备时（以文件属性中的日期为准）可获得的信息，但全球形势正迅速变化，立场可能发生变化。本内容仅供一般信息目的，未考虑读者的具体情况，且不打算取代咨询我们的专业顾问。埃森哲在适用的法律允许的最大范围内，对本文档中信息的准确性和完整性以及基于此类信息所采取的任何行为或疏忽概不负责。埃森哲不提供法律、监管、审计或税务建议。读者需自行从其法律顾问或其他持证专业人士处获取此类建议。本文件涉及由第三方拥有的商标。所有此类第三方商标均为其各自所有者的财产。不打算、不表示或暗示此类商标所有者对本内容提供赞助、认可或批准。