

AI 伦理道德方略

助力企业建立值得信赖的人工智能文化

IBM

如何从 IBM 获得帮助

客户可以利用 IBM 深厚的行业、职能和技术专长、企业级技术解决方案以及基于科学的研究创新来实现人工智能、分析和数据方面的潜力。

如需更多关于 IBM Consulting 的人工智能服务的信息，请访问 ibm.com/services/artificial-intelligence。

如需更多关于 IBM Software 的人工智能解决方案的信息，请访问 ibm.com/ Watson。

如需更多关于 IBM Research® 的人工智能创新的信息，请访问 research.ibm.com/artificial-intelligence。



摘要

组织可以通过战略方法来解决人工智能伦理道德问题，从而使自己区别于其他组织。

企业领导要对当今企业的人工智能伦理道德负责

非技术高管现在是人工智能伦理道德的主要拥护者，这一比例从 2018 年的 15% 增长到 3 年后的 80%，79% 的 CEO 现在准备采取行动（之前只有 20%）。

许多组织在有意义的人工智能方面取得了坚实的进展

超过一半的组织已采取措施将人工智能伦理道德纳入其现有的商业道德方法中，其中许多组织正在创建针对人工智能的伦理道德机制。

但意向与行动之间的差距仍然太大了

例如，拥有多元化和包容性的员工队伍对于减少人工智能中的偏见（68% 的组织承认这一点）很重要，但人工智能团队的多样性仍然远低于其组织的员工队伍：相比之下，人工智能团队对女性的包容性低 5.4 倍、对 LGBTQ+ 个体的包容性低 4 倍以及对种族包容性低 1.7 倍。

推进可信赖的人工智能迫在眉睫

只需浏览头条新闻，企业领导即可知道，面临各种社会、环境和政治问题的公司比比皆是。

客户、员工甚至股东会更频繁地要求组织不仅对当前关注的问题采取有原则的立场，而且还要采取有意义的行动来实现明确的结果。（参见观点“如果正确行事，做得好的公司可以做得更好。”）

今天的高管越来越受到这些力量的影响，其中许多要求超出了高管通常的舒适区。事实上，近 60% 的人表示，近期的社会和经济动荡预计将影响更广泛的社会契约。[1]

正如 IBM 商业价值研究院 (IBV) 在 2018 年进行的一项调查中首次研究的那样， 此类动荡波及到了人工智能和数据领域。[2]

对于普通人群来说，不可信人工智能的定义可能很明显：歧视、不透明、滥用以及其他方面未能达到普遍的信任预期。 然而，推进值得信赖的人工智能仍然具有挑战性，因为有时需要采取务实的平衡行动，例如，“可解释性”（理解人工智能算法结果背后的基本原理的能力）与“稳健性”（算法得出结果的准确性）之间的平衡。

如果不解决这些权衡问题和其他伦理问题，组织就不能再采用人工智能。 他们唯一的问题在于，是否有策略地、有目的地和深思熟虑地面对这些挑战。

广义的技术部门已经迈出了第一步。 这些企业包括许多数字原生代，他们拥有高增长、高利润的平台，这些平台可以补足以消费者为导向的商业模式的伦理缺失。 因此，他们组建了庞大的团队和多层流程来满足需求，并取得了不同程度的成功。

观点

如果正确行事，做得好的公司可以做得更好

公司提高股东价值的传统职权范围越来越窄。 媒体放大了消费者、公民、员工和投资者的更广泛需求，这促使组织更加关注利益相关者对业务结果所受影响的担忧。

这种增长趋势在 IBV 的可持续发展研究中有明显反映：两年前，只有 8% 的消费者愿意支付超过 100% 的溢价来支持有理念的品牌。[3] 最近，有 43% 的人表示他们愿意支付溢价来支持可持续和对环境负责的品牌。[4]

近 70% 的员工表示，他们更有可能接受他们认为对环境和社会负责的组织的工作机会，类似的动态会影响员工留存率。[5]

此外，68% 的组织客户和 62% 的员工表示，他们有能力通过越来越多地表达自己的观点来迫使企业做出改变。[6]

个人投资者也将可持续性纳入其财务决策。 有一半人表示，公司的倾向变化风险会影响其财务风险。 此群体的 92% 人预计将在未来 12 个月内根据环境因素和/或社会责任因素，决定投资、撤资，或游说基金经理改变投资组合。[7]

事实上，在重新设想传统的股东资本主义以解决对更广泛的利益相关者来说很重要的道德和伦理问题方面，势头也有所增强。[8]

然而，这并非采用最新流行咒语一类的简单操作。 借助这些市场趋势成功获益的公司，不只将可持续性视为另一个社会浪潮，更使之成为业务转型的催化剂。 惟有如此，他们才得以在整个企业中努力开展有意义的实践。[9]

做正确的事，这类企业价值观不仅仅是陈词滥调； 它们可以带来真正的财务价值。

身处这种不确定的环境中，其他行业的老牌企业为了推动盈利增长，不得不在某些业务领域尝试使用人工智能。 对他们而言，更大的难题在于他们通常没有多少资源可用于处理人工智能引发的伦理问题。

组织面临严峻的选择。

他们可以紧闭“舱门”，期待消费者、员工和投资者所参与的这场风暴消退，抑或在更熟悉的监管避风港寻求庇护。 但是，关乎人工智能治理的区域、国家甚至地方法律框架仍在不断发展。 因此，在不确定（和重叠）的监管环境中，以合规为策略可能更像是一场赌博。 不作为的危害可能会持续增大。

反之，组织可以采取战略方法来应对人工智能中的伦理问题，从而确定方向，并在众多竞争对手中脱颖而出。他们在边做边学的过程中，不可避免地要对具体行动方案进行调整。

为了解组织在处理人工智能伦理问题并将其纳入业务实践方面取得的进展，我们在 2021 年与牛津经济研究院合作，对 22 个国家/地区的 16 个业务和技术职位的 1,200 名高管进行了调查。 （请参阅第 28 页的“研究方法和方法论”。）

一种有意义的人工智能伦理规范方法可以推进值得信赖的人工智能。

意向和实施之间的差距令人堪忧

首先，让我们定义人工智能伦理。

人工智能伦理被普遍认为是一个多学科研究领域，旨在通过优先考虑人类能动性和福祉来优化人工智能的有益影响，同时降低对所有利益相关者造成不利后果的风险。[10] 有意义的人工智能伦理规范方法可以导致推进值得信赖的人工智能的结果。

在不同的专业领域，情况大同小异。 例如，医学领域致力于改善患者健康，安全、有效和值得信赖的治疗是实现这些目标的手段。 工程是一门专注于设计和建造物理基础设施的学科，安全可靠的桥梁、隧道和建筑物是预期的结果。

IBV 的其他研究表明，消费者、公民和员工将人工智能视为帮助解决人类面临的最紧迫挑战的顶级技术。[11] 因此，对人工智能伦理采取严格方法可谓至关重要。此外，超过 85% 的受访者表示，组织在解决社会问题的同时解决人工智能伦理问题是非常重要的。[12]

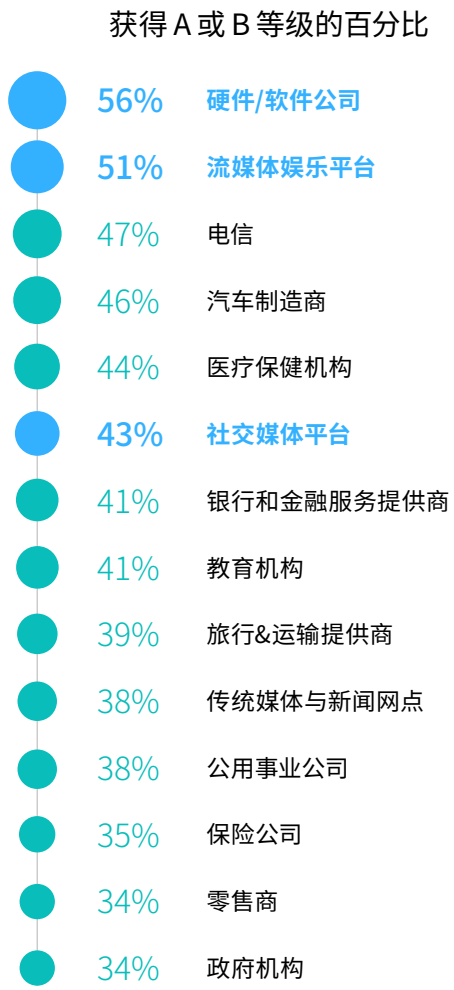
但是，这些人对银行、零售和保险等传统行业的公司在负责任地使用技术方面打分不高。 这些行业中大多数的评分甚至低于那些经常受到媒体指责的社交媒体平台的评分（见图1）[13] 此外，自 2018 年以来情况一直没有明显改善，当时 40% 的消费者相信公司在开发和实施人工智能等新技术时负责任且合乎伦理，现今各行各业的平均水平也不过如此。[14]

超过 85% 的受访消费者、公民和员工表示，组织必须解决人工智能伦理问题。

图 1

人工智能伦理标准报告卡片

消费者在负责任地使用技术方面对传统行业的评价一般



消费者对技术公司评价最高

社交媒体公司的排名甚至高于许多传统行业

问：对于以下类型的公司/组织负责任地使用人工智能等技术方面，您会给他们什么等级/评级？
资料来源：IBM 商业价值研究院对 14,526 名成年人的全球调查。2021 年 7 月。 以前未公布的数据。

但是，接受本研究调查的高管表示，他们的组织确实认为人工智能伦理现在比 三年前更重要：从我们 2018 年调查[15] 中不到一半的受访者增加到 2021 年的四分之三以上。 更具体地说，他们估计，在此期间，在训练、团队、流程、工具和其他运营能力方面支持人工智能伦理制度化的支出翻了一番。 他们预测未来 三年会有更大的投资。

那么，如何解释这些高管的观点与他们的客户和员工评估之间的差异呢？

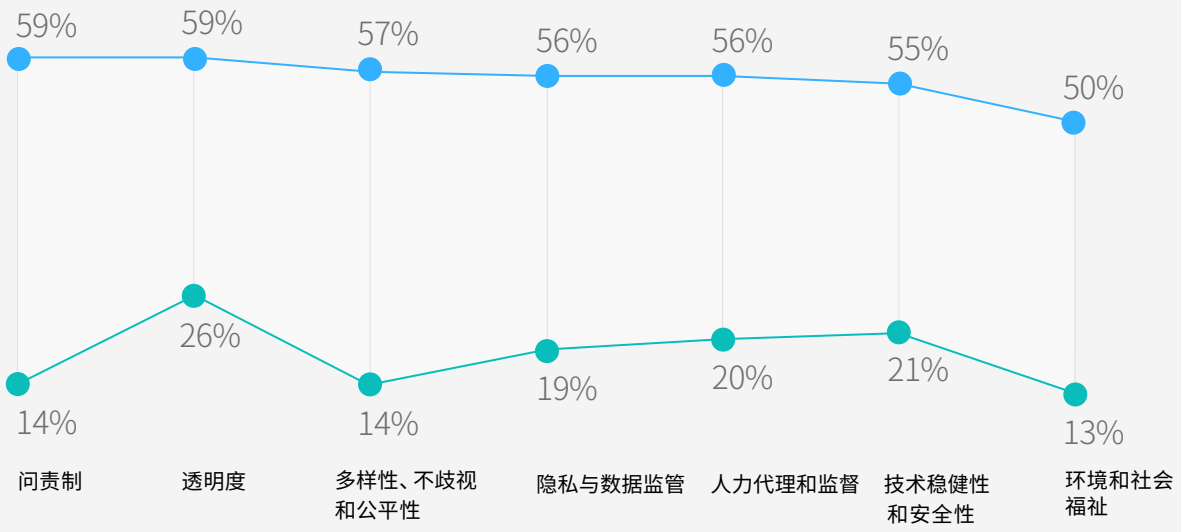
组织为了满足市场预期的紧迫感而增强的愿望似乎超出了他们将意向迅速付诸实施以安抚利益相关者的能力。 虽然超过一半的组织公开认可了人工智能伦理的共同原则，但只有不到四分之一的组织将其付诸实施（见图 2）。 不到 20% 的人强烈同意他们的组织在人工智能伦理方面的实践和行动符合（或超过）他们既定的原则和价值观。 这证实并量化了世界经济论坛 (WEF) 所称的“意向-行动”差距。[16]

不到 20% 的高管强烈支持他们的人工智能伦理行为符合或超过其规定的原则和价值观。

图 2

意向与行动之间的差距

组织已认可人工智能伦理原则，
但尚在努力实施这些原则



注：欧盟委员会人工智能高级专家组在“可信赖人工智能的伦理准则”中定义的人工智能伦理原则。 2019 年 4 月。
https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

缩小这些操作差距至关重要。 例如，为了让人工智能变得更少偏见和更值得信赖，必须实现多样性、非歧视和公平的伦理原则。 组织认识到这一点的重要性，在2021 年的人工智能工作中提升了此项目的重要性（与2018 年相比）。 在最新调查中，几乎所有领导者都认为这是重要的，而将近 6 倍的高管认为它“非常重要”。

为解决此问题，从事人工智能伦理工作的员工应该代表更广泛的受众；但是，人工智能团队的多样性仍然远低于组织的总员工群体的多样性。 我们的结果显示，人工智能团队中的女性人数比组织中少 5.4 倍，LGBTQ+人数比组织中少 4 倍，黑人、原住民和有色人种(BIPOC) 人数比组织中少 1.7 倍（见图 3）。

这些差异的根本原因各不相同。 但是，我们需要更加积极和切实地关注改善员工的招聘、雇佣和保留实践，以增加那些在人工智能设计和实施过程中未被充分代表的声音。 这种强调有助于减轻偏见并建立对人工智能结果的信任。

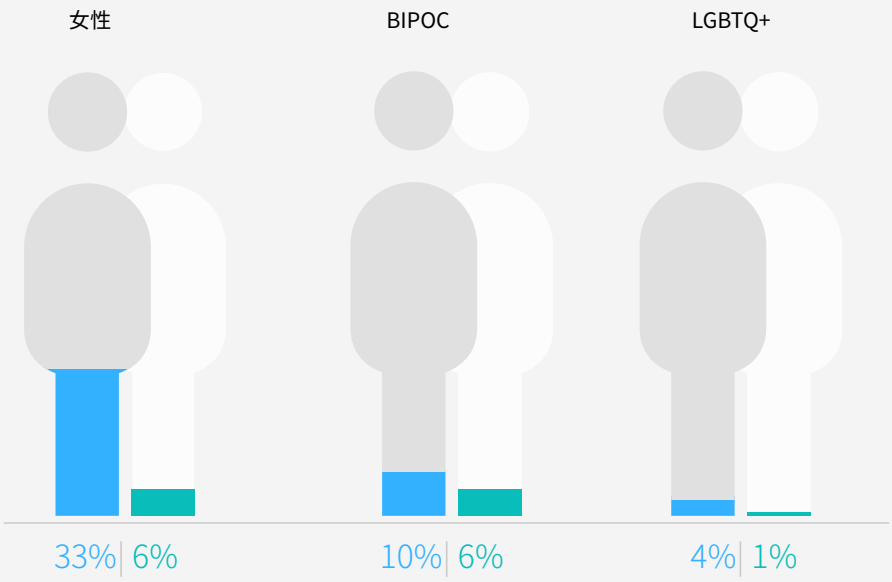
我们在调查中看到了一些证据，表明这些努力也在经济上得到了回报。 例如，员工与人工智能团队之间性别差距最小的前四分之一的企业，其人工智能项目的投资回报率更高。

为了消除其他人工智能伦理原则和实践之间的差异，我们需要做出具有切实目标和可衡量结果的类似努力。 （请参阅案例研究“Regions Bank：专注于高质量和值得信赖的人工智能”。）

图 3

多样性二分法

组织的人工智能团队远不如企业员工多样化



上述愿望和行动之间的差距与公司在环境可持续性方面的情况极为相似。 在最近的 IBV 研究中，只有 35% 的组织实施了他们的可持续发展战略，只有 37% 的组织将可持续发展目标与其业务战略保持一致。[17] 只有三分之一的企业将可持续发展目标和指标整合到他们的业务流程中。[18]

组织和高管更多地关注将人工智能伦理和其他原则性优先事项付诸实践时，他们可以更成功地解决意向与行动之间的差距。

巨变：业务主管领导，协作至关重要

随着组织努力缩小差距，在过去几年中，关于谁负责带头开展这项工作的观点发生了巨大变化。 2018 年，只有 15% 的受访者认为非技术高管是人工智能伦理的主要“拥护者”；[19]到 2021 年，80% 的受访者会这样认为。

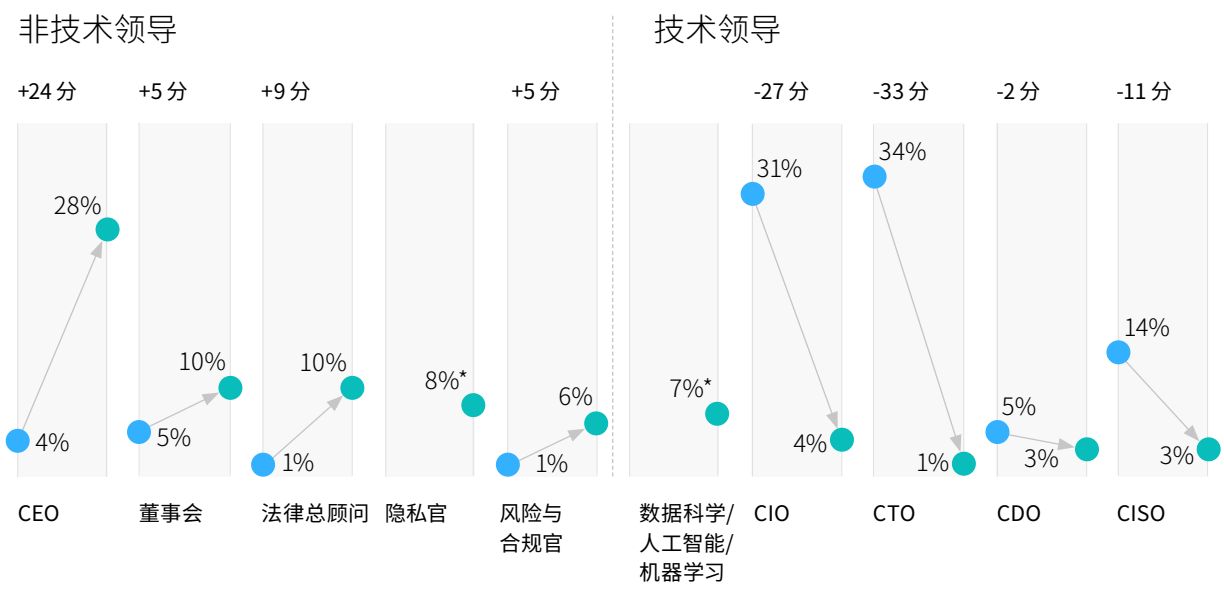
公司尤其希望首席执行官 (28%) 以及董事会成员 (10%)、总法律顾问 (10%)、隐私官 (8%) 和风险与合规官 (6%) 引领潮流，这一发现在不同行业或地区没有什么差异（见图 4）。 有了这种新的期望，CEO 可能会考虑任命一位专门的人工智能伦理主管来分担责任。[20]

高管们列举了 20 种不同的业务职能，它们都在某种程度上涉及人工智能伦理，这表明协同作用至关重要。

图 4

换岗

从 2018 年到 2021 年，主要负责人工智能伦理的人员已从技术人员转变为非技术领导



问：哪个职能部门主要负责人工智能伦理？ 2018 年调查数据来源：Goehring、Brian、Francesca Rossi 和 Dave Zaharchuk。“AI 不仅要合规，还要合乎道德：从原则到实践。” IBM 商业价值研究院。2020 年 4 月。*职位未包含在 2018 年数据中

案例研究

Regions Bank：专注于高质量和值得信赖的人工智能[21]

企业如何以可重复且可持续的方式，或许最重要的是以值得信赖的方式来实施人工智能？

Regions Bank 不得不面对这一挑战。 它的高级分析实践往往依赖于孤立的数据集、独立工作的开发团队以及不同的、有些不一致的开发方法。

首先，银行转变了它的分析功能。 创建卓越分析中心后，它将数据引入集中环境并应用了更多机器学习 (ML) 和人工智能技术。 最重要的是，它采用端到端的商业价值方法，包括人工智能质量控制。

借助这一新的基础，Regions Bank 可以更好地利用一组集成的能力和团队来帮助确保人工智能模型是公平、合乎伦理和值得信赖的。

“我们以与客户开展开放和基于信任的互动而自豪，”首席数据和分析官 Manav Misra 写道。 Misra 解释，Regions Bank 希望改善其社区、客户和员工的生活。 “这就是为什么我们将人工智能伦理监督纳入我们的开发方法论。”

他补充说，“所有这些都指向我们所说的‘负责任的人工智能’。” 此方法要求人工智能模型底层的数据能够代表用于做出决策的数据。 此外，模型必须是可解释的，因此决策过程很明确。

内部监督团队与风险管理人员、审计合作伙伴和政府监管机构一样，有助于维护解决方案的公平性、安全性及稳健性。

但是，何时才是参与人工智能伦理流程和团队的适当时机？

根据 Misra 的说法，越早参与越好。 “如果我们能够让我们的监督合作伙伴及早了解业务案例和需求，以及代码和开发人员在每个敏捷开发段中的意向，他们就可以提供更迅速的反馈，”他写道。 “这可以加速 Regions Bank 所追求的高质量、值得信赖的人工智能解决方案的开发。”

从而形成更高度信任的人工智能、机器学习和其他分析解决方案。 它们合力帮助降低风险、帮助检测欺诈、帮助商业客户，对消费者提供洞察以更好满足他们的需求，所有这些都贡献了商业价值。

也就是说，组织还认识到跨职能的协作的方法是不可或缺的。 大多数领导指定了至少 20 个不同的业务职能，这些职能在某种程度上涉及人工智能伦理。

首席信息官 (CIO)、首席技术官 (CTO)、首席数据官 (CDO) 及其团队对于实施人工智能伦理至关重要。 但这些职责不仅仅属于技术专家的范畴。 人工智能伦理的其他重要职能包括采购、产品设计、研究、公共政策和监管事务。

斯坦福大学以人为本的人工智能研究所 (HAI) 的 Rob Reich 回应上述发现时说：“组织需要密切合作才能使人工智能伦理机制成为现实。” “就像人们不能将个人伦理决策外包给其他人一样，组织追求人工智能伦理，不能通过简单任命首席伦理官，却让所有其他业务部门照常运作。 在某些基本层面上，伦理必须是每个人的责任。”[22]

例如，一家大型全球制药公司在其数据科学团队中立项开发一个概念验证技能推理模型，该模型可以帮助盘点和预测人才发展需求。 根据最初的积极成果，该公司决定在整个企业范围内扩展该项目。 但是，他们很快就发现，大规模实现人工智能不仅仅需要数据科学家的参与。

除了软件工程师、项目经理和其他 IT 专家之外，它还需要人力资源专业人士、法律团队、心理学专家及其他人才管理专业人士之间的协作努力。 这样一来，数据沿袭和起源不仅意味着可选元数据属性； 它们还是建立对结果的信任以及推动员工采用和参与的关键要素。

鉴于多个业务职能部门的参与、更高级别的敏感性和普及的动态反馈循环等情况，自上而下的参与对于促进正确团队在正确的时间进行正确的交互至关重要。

事实上，高层管理人员似乎比 2018 年准备得更充分。 也许是受到与团队、外部专家主题互动的启发，或者是受会议环境的影响，四倍于之前的 CEO 人数（从 2018[23] 年的 20% 增加到 2021 年的 79%）和更多的董事会成员（从 2018[24] 年的 45% 增加到 71% 2021）表示他们准备在人工智能伦理问题上采取行动。

我们在 2018 年发现劳动力欠缺问题。现在首席人力资源官 (CHRO) 似乎正在更好地解决这一问题。 当时，只有 37% 的组织制定了对受人工智能影响的员工进行再培训和技能重建的积极计划。[25] 2021年，这一数字跃升至 55%。 同样，为需要与人工智能进行更多互动的员工制定学习计划的组织从 2018[26] 年的 41% 增加到 70%。 随着需要直接接触人工智能的人数不断增加（根据 CHRO 的估计，未来几年将增加 20% 以上），主动解决这些教育问题对公司、政府和社会来说变得越来越重要。

“在某些基本层面上，
伦理必须是每个人的责任。”

Rob Reich, 斯坦福大学
以人为本的人工智能研究所副所长

对外部人工智能监管法规的预期可能也会提升人工智能伦理的重要性。 74% 的组织报告他们所在地区即将出台法规，64% 的组织表示他们已准备好应对这些法规。 然而，不管监管环境的发展方式如何，受访者表示第三方专业机构和其他非监管团体预计将发挥关键作用，48% 的受访者认为这些更广泛的生态系统的影响和参与是非常重要的。 （请参阅案例研究“加拿大银行财团：共同构建可信数据和人工智能”。）

总体而言，组织参与趋势似乎朝着正确的方向发展。这为公司以及政府和教育机构提供了动力，可以采取更多措施来实施人工智能伦理原则。

积极计划对受人工智能影响的员工进行再培训和技能重塑的组织数量从 2018 年的 37% 跃升至 2021 年的 55%。



案例研究

加拿大银行财团：共同构建可信数据和人工智能[27]

金融服务机构正在使用数据（包括客户、内部和第三方数据）提供传统银行解决方案的创新替代方案。 因此，银行与客户的关系可能会变得更加深入和互动。

正如加拿大国家银行综合风险管理副总裁 Mathieu Avon 所说的那样，“人工智能可以积极改变客户和员工的体验，并增强组织的核心能力。 我们也在负责任且合乎伦理地部署人工智能技术方面发挥关键作用，以维护客户和其他利益相关者的信任。”

原则上，许多金融服务组织都认同，符合伦理标准的人工智能是必要的。 但他们并不总是拥有可自行构建、执行和扩展可信数据和人工智能应用程序的资源。

为应对这一挑战，全球最大的技术专业组织电气和电子工程师协会 (IEEE) 召集的合作伙伴生态系统于 2019 年联手制定了标准指南，重点关注加拿大金融服务领域符合伦理标准的人工智能。

“数据伦理不是做生意的成本； 它们是对良好业务的投资，” CIBC 前高级副总裁兼企业数据首席信息官 Terry Hickey 表示。

该团队策划了来自业界、学术界、非政府组织和标准机构的最佳实践和路线图。 加拿大的六家主要银行、信用合作社、养老基金和金融科技公司也为该计划做出了积极贡献。

加拿大皇家银行数据使用和产品管理、数据和分析主管 William Stewart 强调了与企业原则的联系，“组织需要证明他们已经仔细考虑了人工智能的独特风险，并以符合其企业价值观并符合客户、员工和社会期望的方式做出回应。”

此生态系统继续合作，为值得信赖的机器学习操作流程和工具创建了运行手册。 这一持续努力的共同目标旨在向客户、合作伙伴和员工保证隐私受到保护，数据被负责任地使用，偏见不断减少。

行动的动力：值得信赖的人工智能可以带来商业价值

AI 伦理从原则转向实践的商业理由不仅仅是降低下行风险及满足利益相关者对负责任行事的期望。 公司可以实现潜在的额外收益——推进可持续的实践，保护信息质量，拥抱长期思维而不只顾短期利润，切实增加消费者心目中的信任。[28]

同样，我们在研究中发现，75% 的高管将伦理视为差异化竞争的来源。 同样比例的高管在人工智能领域，将优势转化为业务合作伙伴和供应商选择的务实决策标准。 超过一半的人愿意为他们认为合乎伦理的公司支付溢价。

组织在这个问题上相对于其同行的定位似乎主要受两个因素的推动：（1）人工智能的重要性，以及 （2）人工智能中伦理的重要性。

我们发现，根据他们的评估，将人工智能视为对其业务战略重要的组织，其人工智能计划的效率提高了 1.5 倍。他们的人工智能项目的投资回报率也提高了 2 倍（相对于那些认为人工智能不那么重要的公司）。 人工智能的战略重要性也与公司范围内更高收入增长和更大盈利能力相对应。

此外，那些更重视人工智能伦理的组织报告说，他们也得到了客户和员工的更大程度的信任。

因此，在务实行动的支持下，战略意图可以帮助企业从人工智能中实现更大的商业价值。

然而，组织当如何应对？

75% 的高管将伦理视为差异化竞争的源泉。

将人工智能视为对其业务战略重要的公司报告称，与那些认为人工智能不那么重要的公司相比，他们的人工智能项目实现了 2 倍的投资回报率。

图 5 显示，目前认为人工智能和人工智能伦理不太重要的人预计将在未来 3 年内移动到右上象限，在此象限中，具有前瞻性的接纳者正在推动领先实践并产生更大的影响。 根据高管的明确意图，该细分市场中的组织数量预计将在该时间段内翻倍。

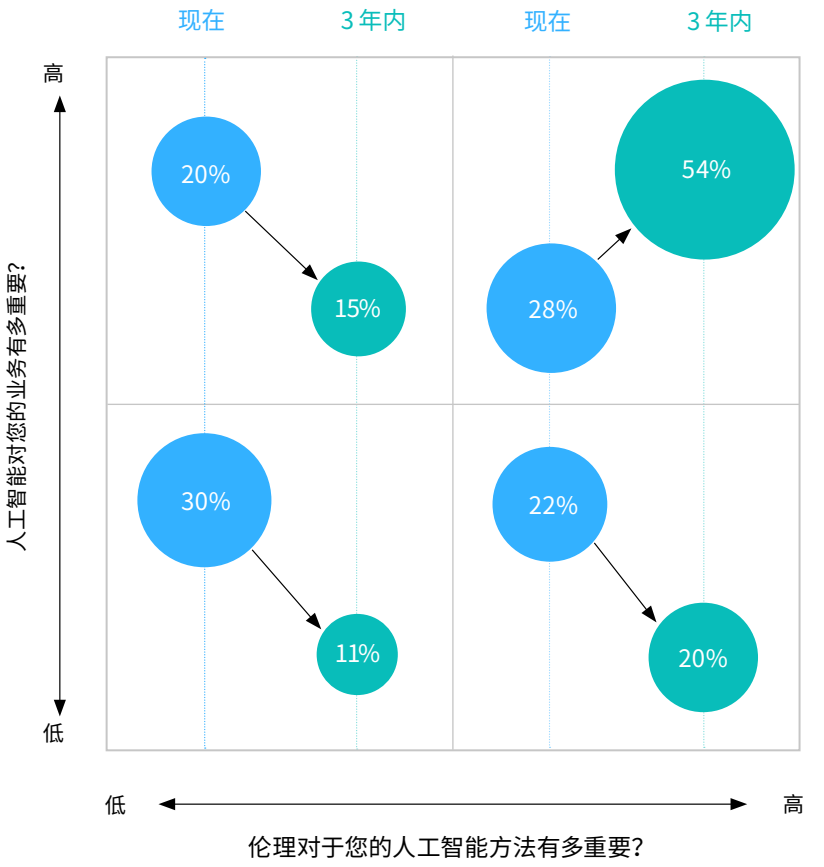
我们认为左上象限（在此象限中，人们认为人工智能很重要，但人工智能伦理不那么重要的位置）是不可持续的，特别是考虑到本报告前面提到的消费者、公民、投资者和员工动态。 组织付不起不致力于人工智能伦理的代价。 正如社交媒体创造了一个记录青少年弱点的世界一样，企业正在创建他们行为的历史记录，以供合作伙伴和潜在员工评估现在和未来。 齐心协力（以及对能力的投资）可以帮助组织走向更有活力的位置。

正如许多先前的研究表明的那样，这项研究是另一个示例，说明财务绩效的提高与人工智能的战略性和有效使用相关。[29] 现在看来，这种关系也适用于以合乎伦理的方式采用人工智能的深思熟虑且稳健的方法。 这可能就是为什么一些将人工智能和人工智能伦理作为其业务战略核心的公司正在努力在组织内扩展此项实践的原因。 （请参阅第 16 页的观点“领先实践：负责任地使用技术”。）

图 5

为进步做好准备

未来 三 年，拥抱人工智能价值和人工智能伦理的组织数量预计将增长



未来三年内，大多数公司预计人工智能和人工智能伦理在战略上都非常重要

观点

领先实践：
负责任地使用技术

世界经济论坛 (WEF) 与圣克拉拉大学的 Markkula 应用伦理中心共同推出了“负责任地使用技术案例研究”倡议。通过此项目，世界经济论坛使组织能够从其社区的见解中学习，并采取措施将伦理规范纳入其设计、开发和技术使用中。

前两 项研究侧重于 Microsoft 和 IBM，探索公司在培养伦理价值观和文化方面不断发展的努力。 对 Microsoft 的研究涵盖了其产品工程组织中的工具和实践。[30] IBM 白皮书探讨它如何转变为更深入地思考技术伦理，以及更负责任地设计、开发、部署和使用技术。[31]

通过人工智能伦理，Microsoft 专注于其研发能力，而 IBM 则在整个企业范围内进行更广泛的扩展。 两家公司都采取以下与人工智能相关的操作：

– 支持与公司价值观和外部多方利益相关者团体价值观相一致的原则

– 通过正式发布的实践和标准、设计思维技术和其他实用工具包实现操作

– 通过目标设定和绩效评估进行衡量

– 通过道德大使和培训计划在整个企业内部逐步推行，着重强调企业文化

– 通过与外部组织合作并参与跨行业、政府及科学计划进行外部扩展

– 以其社会影响为基础，在该领域进行深入的研究并持续改进。

Microsoft 的人工智能监管方法遵循中心辐射模型，帮助公司将隐私、安全和可访问性功能集成到其产品和服务中。 该中心由三个团队组成：第一个由科学和工程专家组成； 第二个侧重于政策、监管、支持和敏感使用； 第三个支持在工程团队中采用负责任的使用流程。 同时，发言人由业务专家组成，包括设计思维团队和帮助推动文化转变的“冠军”。

IBM AI 道德伦理委员会是 AI 系统伦理开发及部署的基石。 董事会（包括来自业务部门和公司职能部门的高级领导）有权实施、推动和执行其决策。 它还具有项目办公室和宣传网络，通过当地联络点共同推动整个组织的变革。

此监管方法包括 IBM 开发的人工智能和支持软件、业务合作伙伴关系、由其服务组织协助的客户部署以及内部使用（例如招聘、薪酬、教育以及多元化和包容性工作）。

两家公司发现其价值观和结果之间存在不一致时，它们均已采取措施做出响应。 他们还会本着动态反馈循环和持续改进的精神继续发展他们的方法。

其他行业的组织可能会反驳说，他们的需求与大型科技公司建立的结构和机制不同。 虽然有时确实有一些领先的实践、经验教训和其他成功因素需要考虑，并且可能需要在保留大部分稳健性的同时缩小规模。

正如世界经济论坛所说，“由于所有公司现在都是技术公司，因此所有公司都应该更仔细地思考技术伦理如何参与到他们的工作中。”[32]

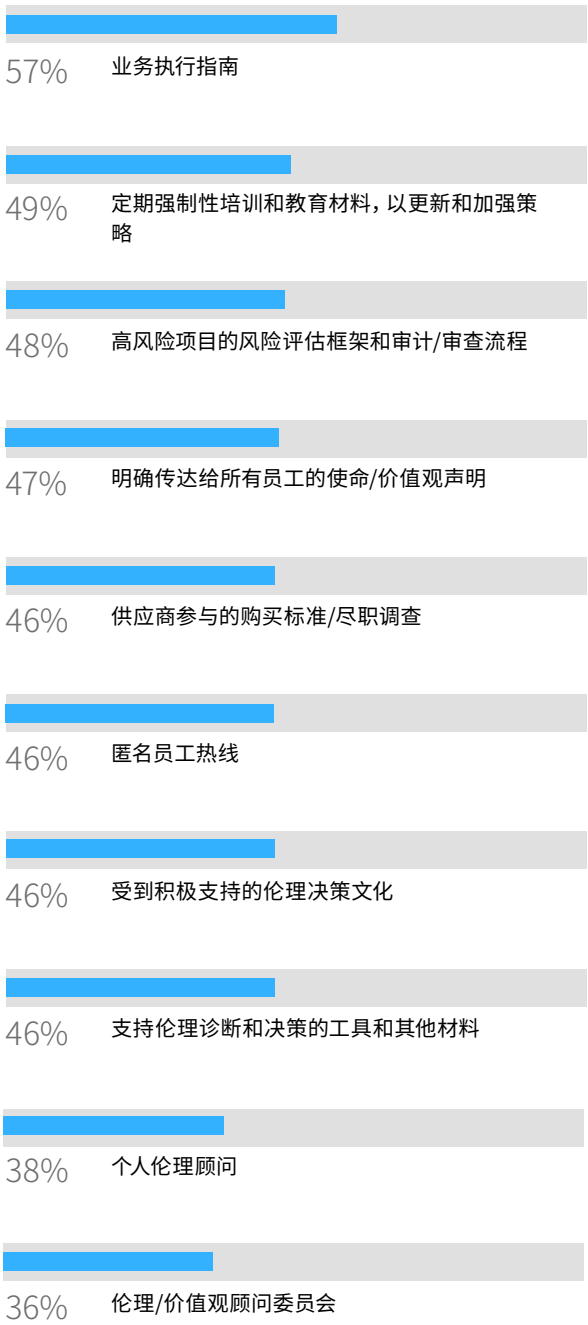
将人工智能视为其商业模式核心的公司已经采用了专门针对人工智能伦理的监管方法。 对于人工智能的战略重要性不那么突出的组织，他们也许被部署在非核心职能或非重要的业务部门，可着手将人工智能伦理原则整合到现有的商业伦理机制中（见图 6）。

将人工智能纳入业务战略与人工智能有效性之间有相关性，但两者并不一定是因果关系，组织应该警惕偏离市场的发展方向。

图 6

第一步

许多组织正在将人工智能伦理纳入现有的商业伦理机制



迎接当下：将学术理想应用于现实世界的行动

随着组织启动他们的人工智能伦理计划，他们可从丰富的资源中受益。 人工智能伦理领域（最初在跨学科研究环境中进行研究）提供了许多建设性的框架、资产和关联以供使用（见图 7）。

图 7

构筑未来

人工智能伦理领域进入更加务实的阶段

	0.0	1.0	2.0	3.0
重点	**感知**	**原则**	**策略**	**实践**
时间范围	截至 2016 年	2017-19	2020-21	2022 年及以后
方法	小型多学科学术团体	多方利益相关者参与人工智能研究社区	跨人工智能开发人员/公司的组织间参与	跨组织/组织内及全社会参与
社区	人工智能/计算机科学、哲学、法律和经济学	人工智能/科技公司、智囊团、标准机构和人工智能开发者社区	公司董事会、专业团体、政策制定者和监管机构	业务主管、职能用户、客户和认证组
工作组示例	– 波多黎各未来生活会议	– 人工智能研究社区 – Asilomar 有益人工智能会议 – 欧盟委员会 (EC) 人工智能高级专家组 – 人工智能伙伴关系 – 经济合作与发展组织 (OECD) 人工智能原则	– WEF全球人工智能联盟 – 美国国家标准与技术研究院人工智能终端用户信任倡议 – 全球人工智能伙伴关系	– 成立跨职能团队的公司 – 政府敲定法规 – 定义和发布的标准
示例输出	– 学术论文 – 会议演讲	– 组织声明的原则 – WEF Board Toolkit – 可信人工智能的 EC 伦理准则 – 美国总务管理局转型技术服务手册	– 组织声明的政策 – IEEE 可信数据和 AIS 手册 – EC 白皮书 – 个人信息保护法（中国） – 新一代人工智能伦理规范（中国）	– 组织伦理行动 – 公司培训 – 专业认证 – 欧盟人工智能法规（建议） – 中国人工智能法规（建议）

即使是大学也承认，他们的目标一直是让学术研究和抽象概念找到实用的途径，而不仅仅是在计算机科学领域。“人工智能伦理不能在扶手椅上完成，” Leverhulme 未来智能中心主任兼剑桥大学哲学系高级研究员 Stephen Cave 承认。[33] 他所在的大学从 2021 年开始试行定制研究生学位以帮助培养下一代人工智能伦理学家。

一些大学已经创建了具有类似重点的研究所（例如，牛津大学的人工智能伦理研究所）。 为其他教育环境中的学生开设的相关课程也在增加。 布鲁金斯学会和未来隐私论坛之类的智库；IEEE 等专业组织；美国国家标准与技术研究院之类的非监管政府机构；WEF、全球人工智能伙伴关系和人工智能伙伴关系之类的跨行业组织也一直处于人工智能伦理的最前沿，并创造了有用的资源。 其他基层合作（例如，美国全球服务管理局的技术转型服务）也不断做出积极贡献。

组织领导现在可以利用这些进步来帮助在那些需要协同工作的企业职能中实施人工智能伦理规范。 最终，这些工具可以帮助推进值得信赖的人工智能的交付。

实施这种人工智能伦理方法还可以表现出组织在负责任地使用技术以实现竞争优势、可持续创新甚至社会正义目标方面所做的努力。 事实上，我们的研究揭示出相关的联系：超过 67% 的组织认为人工智能和人工智能伦理很重要，他们还表示他们在可持续性、社会责任、多样性和包容性方面的表现优于同行。

“对于荷兰政府机构来说，负责任地应用算法至关重要，” 荷兰企业机构 RVO 的首席信息官 Johan Maas 指出。 “通过使用整体方法，我们的人工智能不仅为员工提供有关客户补贴申请的建议，还解释了对其建议很重要的那些参数，这为创建有效的人机协作奠定了基础。”

超过 67% 的组织认为人工智能和人工智能伦理很重要，还表示他们在可持续性、社会责任、多样性和包容性方面的表现优于同行。

他还表示：“这符合我们根据平等待遇等原则开发人工智能的政策，比如伦理和民主原则以及透明度。 因此，我们增加了员工和公民对使用人工智能的信任。”[34]

舞台搭建好了。 越来越多的企业准备拥抱利益相关者至上的观念。 消费者继续表示他们愿意为可持续的产品支付溢价。 许多采购职能越来越多地将业务责任标准整合到采购决策中。 企业伦理和宗旨被视为差异化因素。 许多人都承认人工智能对社会契约的影响越来越大。 通过正式的综合的人工智能伦理方法来构建值得信赖的人工智能绝不止是崇高抱负。

这是战略目标和社会责任。

现在是组织采取更多行动的时候了。



行动指南

后续步骤取决于您的组织在采用人工智能方面所处的位置以及您相应的战略和伦理意向。

对于那些尚未将人工智能视为业务必需要素的组织来说，密切关注竞争变化（和监管环境）对于评估如何更加积极、更合乎伦理地采用人工智能至关重要。 利用第 18 页图 6 所示现存的商业伦理机制可能是一个合理的起点。

对于在人工智能之旅中走得更远且有雄心壮志的组织而言，管理企业范围人工智能的 AI 道德伦理委员会可以成为天然催化剂，帮助采用必要实践来构建值得信赖的人工智能。 在整个人工智能生命周期（从设计到部署再到动态反馈）嵌入伦理规范对于实施这些实践至关重要。

对于那些已经在战略上接触人工智能和人工智能伦理的人，将您的方法扩展到包括供应商、客户和其他生态系统合作伙伴可以实现更大程度的“伦理互操作性”，这是由加州大学圣地亚哥分校数据科学与哲学教授 David Danks 创造的术语。[35] 这可以帮助您进一步从众多的竞争对手中脱颖而出，并保护企业抵御后门风险。

后续步骤也取决于您的职能。 CEO 和其他 C 级高管帮助确定方向，业务团队根据人工智能生成的见解做出决策，数据科学家询问数据，人工智能开发人员构建解决方案，还有许多其他人填补其他职位。

正如人工智能是团队运动，人工智能伦理领域也是如此，但它需要更多样化的团队、更大的舞台和更高的协作精神。 这不仅仅是技术挑战，而且是社会技术挑战，因此需要一个整体方法。[36] 首先考虑以下 3 个高级重点领域，然后查看针对特定角色的建议行动。

组织范围内的行动

01
战略与远景

在适当的战略背景下设置人工智能伦理实践

– 考虑构建值得信赖的人工智能对业务战略和目标的重要性：
  · 可以通过人工智能加速的关键价值创造者有哪些？
  · 如何衡量成功？
– 考虑人工智能创新在组织增长战略和方法中的作用（例如，组织被视为开拓者、快速追随者等）。
– 认可人工智能伦理的关键原则。
– 确定组织中的人机平衡。[37]

关键角色：CEO、业务和技术主管、董事会

02
方向和管理

为合乎伦理的人工智能实施建立监管方法

– 纳入利益相关者的观点（例如，领导、员工、客户和政府）。
  · 解决隐私、稳健性、公平性、可解释性、透明度和其他对您的组织重要的相关原则的维度。
  · 考虑企业和更广泛的生态系统。
– 设置人工智能和数据风险情况和设备阈值级别。
– 建立组织结构、政策、流程和监控。

关键角色：人工智能伦理“冠军”、风险和合规高管和领导、法律顾问、人力资源/人才高管和领导、多元化和包容性高管和领导、业务部门/职能高管和领导以及技术领导

03
实施和部署

将伦理融入人工智能生命周期

– 与关键利益相关者（例如，领导、员工、客户和政府）接触。
– 建立组织结构、政策、流程和监控。
– 捕获、报告和审查合规性数据。
– 推动和支持团队的教育和多元化工作。
– 定义集成的方法和工具包。

关键角色：人工智能伦理计划办公室、业务和技术领导者和团队、人力资源/人才领导者和团队、人工智能设计师、人工智能工程师、数据科学家和其他技术团队

针对角色的操作

关键角色	示例操作
CEO	– 围绕相关战略和业务目标组建执行团队以推动人工智能的创新战略。 – 适当分配资金以支持人工智能目标。 – 通过博客、主题演讲、内部沟通和其他适当的工具，传达您组织的战略观点、价值观以及在商业和人工智能伦理方面的相应行动。 – 根据人工智能的战略重要性，考虑任命首席AI伦理官来领导企业范围内的工作和/或明确当前执行角色的责任。[38] – 根据人工智能的战略重要性（以及董事会主席或合作伙伴），考虑建立AI伦理委员会。 – 向董事会和其他主要利益相关者报告进展情况。 – 批准人工智能伦理的实施方法。 – 促进培训、人工智能和数据素养以及变革管理议程。 – 赞助整体人工智能和数据风险概况并更新节奏。 – 使高管与业务部门和职能部门的人工智能伦理目标保持一致。
业务和技术管理团队	– 在企业范围的创新战略背景下，优先考虑人工智能、机器学习、分析和数据领域，从而优化业务影响。 – 确认价值观是您的企业形象和文化的一部分。 – 设计人与技术协作的方法。 – 定义宏观成功指标和实现这些指标的时间范围。 – 通过识别和参与以人工智能为重点的关键技术合作伙伴、学者、初创公司和其他业务合作伙伴，建立“合乎伦理的互操作性”，从而增强您的创新生态系统。
董事会	– 批准战略方向。 – 通过与全体董事会和关键委员会（例如，风险/审计委员会、技术委员会）的接触，让执行团队承担责任。
AI伦理“冠军”和项目办公室	– 充当向导和指导。 – 为人工智能伦理项目设定战略、方向和方法： · 从组织价值观中获得对人工智能的伦理影响。 · 确定运营和组织方法（例如，集中的卓越中心/项目管理办公室、分散的跨职能/跨业务单位倡导者、资金、内部与外部资源配置等等）。 · 为人工智能模型制定标准，例如用户同意和退出影响。 · 建立常规流程来审查和审计采购和开发的人工智能模型。 · 评估监管结构中的业务、技术和多样性平衡，并考虑轮换领导和委员会组合。 · 通过传达参与人工智能伦理机制、升级、举报人保护和缓解利益冲突的激励措施来提供透明度。 · 教育员工如何降低风险。 · 将人工智能伦理深入部署到您组织的人工智能方法中（例如，创建详细说明如何建立责任文化的剧本）。 · 与生态系统合作伙伴合作以扩大深度和范围。[39]

关键角色	示例操作
风险与合规高管和领导	– 领导和协调人工智能风险评估并推荐计划改进： · 审查相关计划的现有监管框架和文档、流程和控制，例如，风险、合规性、隐私、安全、记录管理和数据监管。 · 与训练有素的从业者合作，评估新出现的人工智能和数据伦理风险（与开放式创新文化平衡）。 · 提出增强措施以应对新出现的人工智能和数据伦理风险（例如，新的委员会、章程、流程、工具、缓解措施、风险保险等等）。 · 领导和协调建立新的人工智能风险评估计划，包括角色、职责、控制、升级协议、指标、人工智能模型审计、程序、流程和文档。 – 跟踪全球人工智能和数据伦理发展，例如，法律、监管和政策变化： · 定期提供变更摘要和趋势分析。 · 建立或整合监管变化管理例程。
首席信息安全官(CISO)	– 无论采用人工智能的程度如何，都需要开发和增强安全流程以应对特定于人工智能的威胁： · 识别和管理威胁。 · 记录并提供合规保证。
法律顾问	– 与合作伙伴和供应商生态系统合作，检查人工智能应用程序和数据使用的完整责任链。
人力资源/人才高管和领导	– 为已识别的功能风险制定缓解方法。 – 协作定义培训、人工智能和数据素养及变更管理议程。 – 制定行动计划以缩小技能差距并解决受影响的员工需求。
多样性和包容性高管及领导	– 衡量人工智能团队的多样性。 – 合作找出阻止人员流失及吸引和留住人工智能中代表性不足的群体及其他技术角色的方法。 – 公开谈论促进人工智能、数据科学和相关团队多样性的努力。 – 制定培训计划，帮助员工了解无意识偏见如何渗入数据集（和人工智能解决方案）。
业务部门和职能主管和领导	– 协作定义业务单元和功能特定的实施流程和策略。 – 公开并积极支持整个组织正在进行的人工智能伦理工作。

关键角色	示例操作
业务和技术领导及团队	– 安排日常实现： · 创建和扩充人工智能团队及流程 · 确定其他员工和合作伙伴的角色。 · 确定所需的外部专家，例如，设计师、伦理学家、心理学家和社会科学家。 – 部署人工智能实践的端到端使用： · 将伦理原则纳入业务和技术要求。 · 根据团队的反馈评估挑战和机遇并共同努力解决。 · 让团队有时间专注于发展和改进伦理实践。 · 将业务价值与关键目标结合起来，并使其对您的团队来说可以预测； 然后设定可行目标加以实现。 · 定期交流跨学科工作的好处，以创建符合伦理的人工智能实践。 · 培养自发的 “自下而上” 的想法来对 “自上而下” 的方法进行补充。 – 授权团队在不符合伦理原则时提出和解决问题。
人力资源/人才领导和团队	– 在整个员工群中策划和提供针对人工智能伦理的定制学习。 – 为需要更多使用人工智能的员工进行培训。 – 为受人工智能影响的人员实施技能再培训和技能提升计划。
设计师、建筑师、数据科学家、社会科学家、人工智能工程师和其他技术团队	– 利用方法和工具来探索潜在的伦理问题： · 编写代码之前，领导设计思维会议，与业务和其他专家一起识别和探索可能的问题。[40] · 对训练数据进行探索性数据分析。 · 设置公平阈值、衡量公平指标并减少偏见。 · 沟通并展示 “黑匣子” 透明度的技术可行性。 · 构建界面以便能够解释算法。 · 创建文档以使其他人能够评估上下文。 · 鼓励团队提出问题并解决伦理问题。 – 明确定义组件和功能需求以定义业务成功。 – 向非技术受众解释人工智能伦理的必要性。 – 自行启动关键培训（若未曾提供过，可以考虑公平性和准确性标准、数据沿袭和起源机制、不利影响分析等培训）。 – 进行人工智能影响和风险评估。 – 识别并传达说明人工智能伦理重要性的业务和用户案例。 – 在多个团队（和团队类型）之间共享知识和协作。

关于
作者

Brian Goehring

IBM 商业价值研究院
人工智能全球研究总监
goehring@us.ibm.com
linkedin.com/in/brian-c-goehring-9b5a453

Brian Goehring 是 IBM 商业价值研究院的副合伙人，他领导人工智能商业研究议程，与学者、客户和其他专家合作，以发展数据驱动的思想领导力。他拥有超过 20 年的工作经验，曾为大多数行业和业务部门的高级客户提供战略咨询。他获得了普林斯顿大学哲学学士，拥有认知研究和德语证书。

Francesca Rossi

IBM 院士，人工智能伦理全球负责人
Francesca.Rossi2@ibm.com
linkedin.com/in/francesca-rossi-34b8b95
@frossi_t

Francesca Rossi 是 IBM 院士，就职于 T.J. Watson 研究中心，并且是 IBM 人工智能伦理全球领导者。她领导提升人工智能能力的研究项目，并担任 IBM 人工智能伦理委员会的联合主席。Francesca 是人工智能合作伙伴关系董事会和人工智能全球合作伙伴关系指导委员会的成员。她是全球人工智能促进协会 (AAAI) 和欧洲人工智能协会 (EurAI) 的研究员，并被选为 AAAI 的下一任主席。在加入 IBM 之前，她是意大利帕多瓦大学的计算机科学教授。

Beth Rudden

IBM 杰出工程师，认知科学和可信 AI，
数据和技术转型，
IBM Consulting
brudden@us.ibm.com
linkedin.com/in/brudden

Beth Rudden 领导着一项全球业务，该业务由分布在各地的大型团队组成，这些团队开发基于合乎伦理地使用数据的人工智能模型，为客户提供可行的洞察。她拥有古典考古学学士学位和人类学硕士学位，还拥有多项解决方案专利，这些解决方案可提供更清晰的洞察、更好的客户理解和更快的实施。

IBM 商业价值研究院

20 年来，IBM 商业价值研究院一直是 IBM 的思想领导力智囊团。 我们提供有研究支持和技术支持的战略洞察，帮助领导者做出更明智的业务决策。

凭借我们在商业、技术和社会交叉领域的独特地位，IBV 每年都会针对成千上万高管、消费者和专家展开调研、访谈和互动，将他们的观点综合成可信赖的、振奋人心和切实可行的洞察。

要及时了解IBV最新研究成果，请在 ibm.com/ibv 上注册以接收 IBV 的电子邮件通讯。 您还可以在 Twitter 上关注 @IBMIBV 或在 LinkedIn 上的以下位置找到我们: linkedin.com/showcase/ibm-institute-for-business-value。

访问 IBM 商业价值研究院中国网站，免费下载研究报告: https://www.ibm.com/ibv/cn

关于研究洞察

研究洞察致力于为业务主管就公共和私营领域的关键问题提供基于事实的战略洞察。 洞察根据对 IBV 所做的主要研究调查的分析结果得出。 要了解更多信息，请联系 IBM 商业价值研究院: iibv@us.ibm.com。

相关报告

AI 不仅要合规，还要合乎道德
Goehring, Brian、Francesca Rossi 和 David Zaharchuk。 “AI 不仅要合规，还要合乎道德: 从原则到实践。” IBM 商业价值研究院。 2020 年 4 月。 https://www.ibm.com/downloads/cas/W3KR6KZO

人工智能的业务价值
Goehring, Brian 和 Anthony Marshall。 “人工智能的商业价值。” IBM 商业价值研究院。 2020 年 11 月。 https://www.ibm.com/downloads/cas/8VMRQ24R

可持续发展：转型之催化剂
Balta，Wayne、Manish Chawla、Jacob Karl Dencik 和 Spencer Lin。 “可持续发展: 转型之催化剂—开拓者将愿望转化为行动。” IBM 商业价值研究院。2022 年 1 月。https://www.ibm.com/downloads/cas/2GQB068O

研究方法和方法论

IBM 商业价值研究院与牛津经济研究院合作，在 2021 年 5 月到 2021 年 7 月对北美、拉丁美洲、欧洲、中东和非洲以及亚洲（包括中国和印度）的 22 个国家/地区的 1,200 名高管进行了调查。 调查范围包括来自超过 22 个行业的 16 个业务和技术角色，主要是高管，也包括人工智能专业人士。 接受调研的组织的收入/预算中位数略低于 $3B。

注释和来源

1 Goehring, Brian 和 Anthony Marshall。 “人工智能的业务价值。” IBM 商业价值研究院。 2020 年 11 月。 以前未公布的分析。 https://ibm.co/ai-value-pandemic

2 Goehring, Brian、Francesca Rossi 和 David Zaharchuk。 “以高于合规的标准推进人工智能伦理：从原则到实践。” IBM 商业价值研究院。 2020 年 4 月。 https://ibm.co/ai-ethics

3 Gupta, Sachin、Sashank Yaragudipati、Jane Cheung 和 Chris Wong。 “可持续发展的最后呼吁：消费品和零售业紧迫的增长议程。” IBM 商业价值研究院。 2021 年 8 月。 https://ibm.co/sustainability-consumer-products-retail

4 同上。

5 同上。

6 Edelman Trust Barometer 2021. Edelman. 2021 年 1 月。 https://www.edelman.com/sites/g/files/aatuss191/files/2021-01/2021-edelman-trust-barometer.pdf

7 Gupta, Sachin、Sashank Yaragudipati、Jane Cheung 和 Chris Wong。 “可持续发展的最后呼吁：消费品和零售业紧迫的增长议程。” IBM 商业价值研究院。 2021 年 8 月。 https://ibm.co/sustainability-consumer-products-retail

8 Henderson, Rebecca。 在激情澎湃的世界中重新构想资本主义。 PublicAffairs。 2020.

9 Balta, Wayne、Manish Chawla、Jacob Karl Dencik 和 Spencer Lin。 “可持续发展：转型之催化剂：开拓者将愿望转化为行动。
IBM 商业价值研究院。 2022 年 1 月。
https://ibm.co/sustainability-transformation

10 同上。

11 IBM 商业价值研究院在全球对 14,526 名成人进行了调查。 2021 年 7 月。 未公布的分析。

12 同上。

13 同上。

14 IBM 商业价值研究院对超过 14,500 名消费者的信任和透明度消费者调查。 2018 年 11 月。 以前未公布的数据。

15 Goehring, Brian、Francesca Rossi 和 David Zaharchuk。 “AI 不仅要合规，还要合乎道德：从原则到实践。” IBM 商业价值研究院。 2020 年 4 月。 https://ibm.co/ai-ethics

16 Guszcza, Jim 和 Ann Skeet。 “企业如何在科技时代创造伦理文化。”
世界经济论坛。 2020 年 1 月。
https://www.weforum.org/agenda/2020/01/how-businesses-can-create-an-ethical-culture-in-the-age-of-tech/

17 Balta, Wayne、Manish Chawla、Jacob Karl Dencik 和 Spencer Lin。 “可持续发展：转型之催化剂：开拓者将愿望转化为行动。” IBM 商业价值研究院。 2022 年 1 月。 https://ibm.co/sustainability-transformation

18 同上。

19 Goehring, Brian、Francesca Rossi 和 David Zaharchuk。 “AI 不仅要合规，还要合乎道德：从原则到实践。” IBM 商业价值研究院。 2020 年 4 月。 https://ibm.co/ai-ethics

20 Minevich, Mark 和 Francesca Rossi。 “为什么要聘请首席人工智能伦理官。” 世界经济论坛。 2021 年 9 月。 https://www.weforum.org/agenda/2021/09/artificial-intelligence-ethics-new-jobs/

21 与 Regions Bank 的 Chun Schiros and Daniel Stahl 的访谈 2021 年 11 月; Misra, Minav。 “Regions Bank：一种开发高质量和值得信赖的人工智能的方法。” IBM 博客。 2021 年 5 月 24 日。

22 与斯坦福大学以人为本的人工智能研究所副所长 Rob Reich 的访谈。 2021 年 10 月。

23 Goehring, Brian、Francesca Rossi 和 David Zaharchuk。“AI 不仅要合规，还要合乎道德：从原则到实践。” IBM 商业价值研究院。2020 年 4 月。以前未公布的数据。https://ibm.co/ai-ethics

24 同上。

25 同上。

26 同上。

27 与 IEEE 标准协会新兴技术与战略发展总监 John C.Havens 的访谈。2021 年 10 月。Abdur-Rahman, Pavel。“人工智能与伦理：我们这一代的重大挑战。” IBM 博客（加拿大）。2021 年 5 月 24 日。https://www.ibm.com/blogs/ibm-canada/2021/05/ai-ethics-the-grand-challenge-for-our-generation/

28 Edelman Trust Barometer 2021. Edelman. 2021 年 1 月。https://www.edelman.com/sites/g/files/aatuss191/files/2021-01/2021-edelman-trust-barometer.pdf

29 Goehring, Brian 和 Anthony Marshall。“人工智能的业务价值。” IBM 商业价值研究院。2020 年 11 月。https://ibm.co/ai-value-pandemic

30 “负责任地使用技术：Microsoft 案例研究。” 世界经济论坛。2021 年 2 月 25 日。https://www3.weforum.org/docs/WEF_Responsible_Use_of_Technology_2021.pdf

31 “负责任地使用技术：IBM 案例研究”。世界经济论坛。2021 年 9 月。https://www3.weforum.org/docs/WEF_Responsible_Use_of_Technology_The_IBM_Case_Study_2021.pdf

32 “负责任地使用技术：Microsoft 案例研究。” 世界经济论坛。2021 年 2 月 25 日。https://www3.weforum.org/docs/WEF_Responsible_Use_of_Technology_2021.pdf

33 采访 Leverhulme 智能未来中心主任兼剑桥大学哲学系高级研究员 Stephen Cave。2021 年 2 月。

34 采访荷兰企业局荷兰企业局 (RVO) 首席信息官 Johan Maas 和 Johan Suringa。2021 年 11 月。

35 与加州大学圣地亚哥分校数据科学与哲学教授 David Danks 的访谈。2021 年 9 月。

36 Boinodiris, Phaedra。“获得值得信赖的人工智能” VentureBeat. 2021 年 3 月。https://venturebeat.com/2021/03/14/getting-to-trustworthy-ai/

37 Mantas, Jesus.“移情人工智能可能是人类进化的下一个阶段，只要我们做对了。” 世界经济论坛。2019 年 7 月。https://www.weforum.org/agenda/2019/07/empathic-ai-could-be-the-next-stage-in-human-evolution-if-we-get-it-right/

38 Minevich, Mark 和 Francesca Rossi。“为什么要聘请首席人工智能伦理官。” 世界经济论坛。2021 年 9 月。https://www.weforum.org/agenda/2021/09/artificial-intelligence-ethics-new-jobs/

39 Lohr, Steve。“由顶级公司支持的集团采取行动打击招聘中的人工智能偏见。” 纽约时报。2021 年 12 月 8 日。https://www.nytimes.com/2021/12/08/technology/data-trust-alliance-ai-hiring-bias.html

40 “为什么将设计思维应用于人工智能。” Team Essentials for AI 课程。IBM。https://www.ibm.com/design/thinking/page/badges/ai

© Copyright IBM Corporation 2022

国际商业机器（中国）有限公司
北京市朝阳区金和东路 20 号院 3 号楼
正大中心南塔 12 层
邮编：100020

美国出品 | 2022 年 4 月

IBM、IBM 徽标、IBM.com 和 ibm Research 是 International Business Machines Corp 在全球许多司法管辖区的注册商标。其他产品和服务名称可能是 IBM 或其他公司的注册商标。以下 Web 站点上的 "Copyright and trademark information" 部分中包含了 IBM 商标的最新列表：ibm.com/legal/copytrade.shtml。

本文档为自最初公布日期起的最新版本，IBM 可能随时对其进行更改。IBM 并不一定在开展业务的所有国家或地区提供所有产品或服务。

本文档内的信息“按现状”提供，不附有任何种类的（无论是明示的还是暗示的）保证，包括对适销性、适用于特定用途的任何保证以及非侵权的任何保证或条件。IBM 产品根据提供它们时给出的协议的条款和条件进行保修。

本报告的目的仅为提供通用指南。它并不旨在代替详尽的研究或专业判断依据。由于使用本出版物对任何企业或个人所造成的损失，IBM 概不负责。

本报告中使用的数据可能源自第三方，IBM 并未对其进行独立核实、验证或审查。此类数据的使用结果均“按现状”提供，IBM 不作出任何明示或暗示的声明或保证。

4DPJK92W-CNZH-00

