



# AI 伦理道德方略

助力企业建立值得信赖的  
人工智能文化

## 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](https://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](https://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](https://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。<sup>1</sup>

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。<sup>2</sup>

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。

## 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。<sup>3</sup> 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。<sup>4</sup>

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。<sup>5</sup>

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。<sup>6</sup>

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。<sup>7</sup>

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。<sup>8</sup>

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。<sup>9</sup>

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。<sup>10</sup> 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。<sup>11</sup> 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。<sup>12</sup>

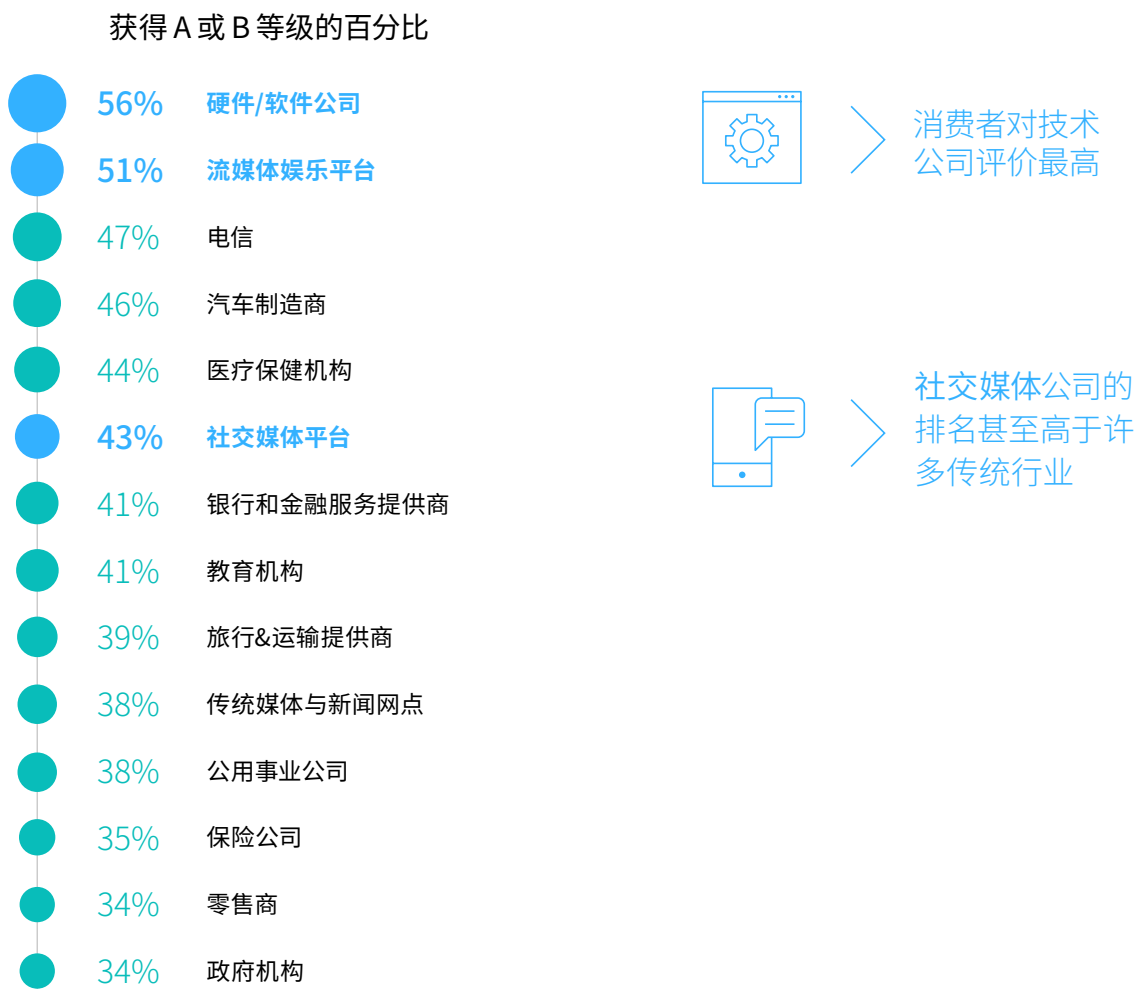
但是，这些人对银行、零售和保险等传统行业的公司在负责任地使用技术方面打分不高。这些行业中大多数的评分甚至低于那些经常受到媒体指责的社交媒体平台的评分（见图 1）<sup>13</sup> 此外，自 2018 年以来情况一直没有明显改善，当时 40% 的消费者相信公司在开发和实施人工智能等新技术时负责任且合乎伦理，现今各行各业的平均水平也不过如此。<sup>14</sup>

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



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<sup>15</sup>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。

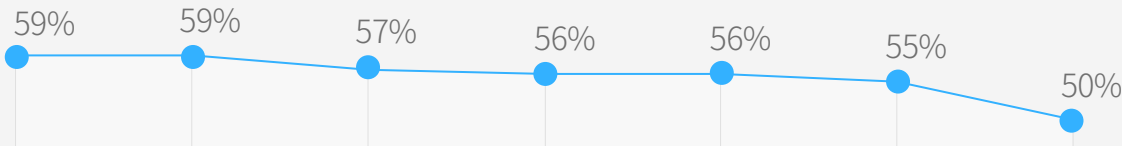
那么，如何解释这些高管的观点与他们的客户和员工评估之间的差异呢？

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。<sup>16</sup>

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



预览已结束，完整报告链接和二维码如下：

[https://www.yunbaogao.cn/report/index/report?reportId=1\\_46719](https://www.yunbaogao.cn/report/index/report?reportId=1_46719)

