研究洞察

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

IBM

IBM 如何提供帮助

借助 IBM 深厚的行业专业知识、技术解决方案和强大能力，客户可以充分发挥人工智能 (AI) 和分析技术的潜力，开始将智能融入几乎每一项业务决策和流程之中。IBM 的 AI 和分析服务部门能够帮助企业：为 AI 准备好数据，最终在数据推动之下，做出更明智的决策；获取深度洞察，提供更出色的客户服务；运用专注于安全、风险与合规的 AI 支持技术，建立信任，增强信心。如欲了解有关 IBM AI 解决方案的更多信息，请访问：ibm.com/services/ai。如欲了解有关 IBM AI 平台的更多信息，请访问：ibm.com/watson。

扫码关注 IBM 商业价值研究院

官网　微博　微信　微信小程序

作者：Brian Goehring、
Francesca Rossi 和
Dave Zaharchuk

谈话要点

关于 AI 道德挑战的不同观点

高管和消费者在审视 AI 风险时，侧重点各有不同：企业主要关注对组织的影响，而消费者则更关注社会问题，比如共同繁荣、包容性以及 AI 对就业造成的冲击。

CEO 似乎与 AI 道德问题脱节

尽管董事会成员认定 AI 道德是一个重要问题，也是发挥企业监督责任的关键领域，但 CEO 对此的重视程度显然不及最高管理层团队或董事会的其他成员。

解决道德困境

大多数董事会成员认为自己还没有为应对 AI 道德问题做好准备。有些企业在单独制定道德准则和实施计划，但任何组织都无法凭一己之力解决这个重大问题。道德问题纷繁复杂而且无先例可寻，必须建立合作关系和联盟，大家携手加以解决。

AI 道德构想

如果得知客户的信贷申请遭拒但并未给出明确理由，您能容忍这种决定吗？如果医生并未查阅最新医学文献，盲目建议您的至爱亲朋接受侵入性治疗，您会签字同意手术吗？当然不会。

人们总是依据适当的行为标准做出关键判断 — 特别是直接关乎他人生命和幸福的重大决策。在日常生活中，人们需要遵守公认的道德规范，同时还受到法律、法规、社会压力和公众舆论的约束。尽管道德规范可能因时代和文化而异，但自人类早期文明发祥以来，道德在指导决策方面一直发挥着关键的作用。例如，希波克拉底誓言起源于古希腊 — 无论是“首先，不要造成伤害”还是“保守秘密”，自中世纪以来始终是医疗界的金科玉律。[1]

在商业领域，道德主题同样并不陌生。传统上而言，道德规范是现代商业文化时常提及但往往被忽视的要素之一。在取得经济成功的过程中，有时会抄近道，占便宜，为获得短期效益而牺牲长期发展目标甚至价值观。作为应对之策，企业在内部建立了合规机制。组织努力构筑“合规围栏”及其他约束机制，应对有意或无意的过失。

毋庸置疑，目前，道德规范体现出前所未有的重要意义。无论是企业高管还是一线员工，无论是政府部长还是普通民众，全世界的人们发现，自己需要做出一些过去无法想象的重大决定 — 可能深刻影响同事和大众的生活。只有在伦理、道德和价值观的引领之下，很多人才会在经济效益与健康福祉之间做出看似不可想象的权衡取舍。

81% 的消费者表示比上一年更关注企业如何利用他们的个人数据，75% 的消费者对企业处理个人信息的信任度有所降低。[2]

80% 的董事会成员认为，AI 道德问题至少在一定程度上属于董事会层面的职责，但仅有一半的 CEO 将此视为 CEO 层面的职责。

超过半数的高管认为 CTO 和 CIO 是 AI 道德问题的主要负责人。

高管预计，科技企业将对 AI 道德问题产生最主要的影响，政府和客户次之 — 最后才是其他企业。

然而，现有体系不够完善，无法应对即将到来的挑战。最近几年，商业环境飞速变化，越来越多的决策在人工智能 (AI) 的推动下做出。现在，企业纷纷运用 AI 技术，帮助筛选人才、处理保险理赔、提供客户服务以及实施大量其他重要工作流程。不过，有关 AI 的道德参数仍然模糊不清，难以捉摸；在某些情况下，人们认为道德因素妨碍业务发展而将其搁置 — 全然不顾可能产生的短期和长期影响。

AI 道德：企业界的形势

与此同时，AI 的采用率预计将继续迅速攀升。事实上，IBM 商业价值研究院的上一次全球高管调研结果表明，未来三年的 AI 平均支出预计将增加一倍以上。[3] 随着 AI 的使用达到新的高度，数据责任、包容性和算法问责等领域的风险也与日俱增。

为了更深入地了解高管如何看待随着 AI 在企业中的作用日益重要，会带来哪些后果以及道德问题，我们对不同地区从事不同行业的 1,250 位高管开展了调研。（有关这项研究的更多信息，请参阅侧边栏“洞察：调研方法”。）我们的研究表明，未来三年 AI 在组织战略中的重要性可能会翻倍，因而迫切需要解决道德问题。

尽管几乎所有受访高管均表示自己的组织严格践行道德规范，但他们普遍担忧：一旦 AI 技术应用不当，很可能会导致严重的后果。因此我们很自然地得出以下结论：在讨论商业环境中的 AI 系统时，必须考量道德规范。

从本质上而言，人机之间的相互认知理解水平必然不如人与人之间的相互理解，而后者数百年来一直与伦理道德形影不离。由于 AI 离不开庞大的计算能力，它可以透过海量数据发掘洞察，挑战人类认知。单纯依靠传统道德方法约束决策，可能不足以做出基于 AI 的决策。

事实上，在接受调研的高管中，有一半以上表示 AI 可以提高企业的道德决策水平。（不到 10% 的受访高管担心产生负面影响。）此外，大多数受访高管表示可以利用 AI 推进社会公益事业，而不仅限于营造良好的商业环境。我们曾问一位受访首席人力资源官，提到 AI，首先想到的是什么？他回答说：“技术进步可以改善人类生活。”其他受访者也赞同这一观点。

排外现象、厌女倾向以及其他一些偏见（即使是无心之失）可能深藏在人性之中，但调研受访者认为，AI 可以做到公平、透明甚至共情。通过检测和纠正 AI 的偏见（训练技术更有效地认识人性），或许可以增强企业协作能力，实现更理想的成果。

AI 可以在出现意外结果时诊断根本原因，从本质上消除偏见。这样，就能够更深入地理解过去的缺点并加以改进，从而实现社会目标。但是，首先必须确立正确的道德框架。

首要问题：解决 AI 道德问题的因素

要能够在合乎道德规范的前提下发挥 AI 技术的威力，最重要的因素是什么？由谁负责确保将道德规范整合到企业内外的 AI 系统之中？社会如何善用 AI 造福人类？

以上是我们希望在研究中寻找答案的三大主要问题，本报告将逐一进行说明。迄今为止，关于 AI 道德的对话主要集中于媒体、科技企业、咨询公司、学术界以及某些政府机构。而真正采用 AI 技术的企业贡献的洞察少之又少。本次调研的目的在于，为这些被忽视的群体（包括银行、医疗机构、零售商等）提供同等的发声机会。

洞察：调研方法

2018 年末，IBM 商业价值研究院与牛津经济研究院合作，采访了 1,250 位全球高管。受访者代表 20 个行业，来自 6 大洲 26 个国家 / 地区，包括董事会成员、首席执行官 (CEO)、首席信息官 (CIO)、首席技术官 (CTO)、首席数据官 (CDO)、首席人力资源官 (CHRO)、首席风险官 (CRO)、法律总顾问和政府政策官员。

在接受调研的高管中，有一半以上表示 AI 可以提高企业的道德决策水平。

毫无疑问，AI 已成为核心技术并引发广泛讨论。在受访高管中，有近 2/3 至少在一定程度上将“AI 道德”视为重要的业务主题，哪怕目前不打算采用 AI 技术的企业高管也不例外。现阶段正在采用 AI 技术的企业的受访高管中持此观点的比例接近 90%。几乎所有受访高管均表示，至少在一定程度上正式考虑将道德规范体现在 AI 计划中。

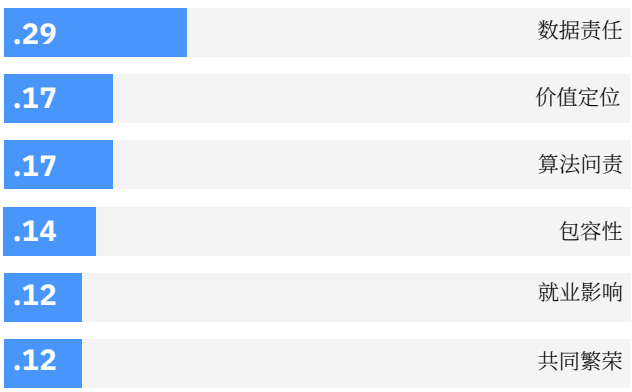
研究期间，我们还请受访高管对开发符合道德规范的 AI 系统所涉及的一系列因素的相对重要性进行评估，包括价值定位、算法问责和包容性。受访者对这些因素的重要性进行并排比较，做出权衡取舍。

我们发现，以下三个领域成为组织和高管研究道德风险时的关注焦点：数据责任、价值定位和算法问责（见图 1）。下图中的值表示各项因素相互比较得出的相对重要性。

图 1

高管普遍认为，在开发符合道德规范的 AI 系统时，数据责任是最重要的因素之一。

让 AI 符合道德规范的各种因素的相对重要性



来源：2018 年 IBM 商业价值研究院全球 AI 道德问题调研。
问题：思考 AI 道德问题时，下列哪些因素相对而言更为重要？
N=1,250。

数据责任

首要风险领域是数据责任，包括数据所有权、存储、使用和共享。可以看出，数据责任的得分差不多是排名第二的 AI 道德因素的两倍。毫无疑问，媒体大肆报道的数据泄露和滥用事件对此产生了较大的影响。

客户影响十分明显：IBM 商业价值研究院差不多同期开展的一项全球消费者调研表明，在过去的一年中，81% 的消费者更加担心企业如何使用他们的个人数据，同时 75% 的消费者表示对企业处理个人信息的信任度有所降低。[4] 在消费者担忧程度上升的同时，企业面临的监管压力也越来越大，欧盟 (EU) 的《通用数据保护条例》(GDPR) 正式生效，美国国会也举办了数轮数据隐私听证会。[5]

AI 领域已经开始感受到数据相关风险带来的重大影响。40% 的受访高管表达了不同程度的担忧，担心数据信任、隐私和透明度成为妨碍 AI 技术采用的拦路虎。除非解决信任问题，否则不可能实现 AI 承诺，因为 AI 的威力几乎完全取决于底层数据。仅有半数 (54%) 高管表示对自己的业务数据具有高度信心。为顺利采用 AI 技术，增强信心至关重要。

价值定位和算法问责

高管提及最多的另外两项 AI 道德风险因素几乎并列，比率均为 17%，或许是因为二者息息相关。价值定位是指 AI 算法按预期运行的能力 — 能否产生体现相应价值、限制和过程的决策。算法问责是指确定谁对 AI 算法的输出负责 — 这也从另一个方面说明如何在共识和争议情况下做出决策。2/3 的最高层技术主管将算法问责视为重要的优先任务。这充分彰显了他们的未来预期：消费者对于透明度和可解释性的要求不断提升。

为深入认识这两种相互交织的风险，请思考几个示例，设想一下基于 AI 的系统采取的行动或决策可能造成的不完美结果或意外后果。在一项评估 AI 识别癌性病变准确性的研究中，研究人员指出，AI 系统倾向于使用尺子或其他显而易见的测量指标来标记病变照片。如果在初步评估后，皮肤科医生怀疑存在恶性肿瘤，往往会使用此类工具。借助这些工具，AI 代理可以“学到”影像中呈现的病变很可能是癌变 — 但对成因却一无所知。[6]

在评估 AI 能否通过射线影像诊断肺炎时，也遇到了类似的问题：AI 代理将高概率患者与收治重症患者的专科医院进行关联；它遵循的是相关性原则，但并不确定甚至根本没有探寻根本原因。[7]

再看一个现实世界的典型例子：由于一个基于 AI 的招聘引擎似乎对女性存有偏见，因此亚马逊将其束之高阁。鉴于过去十年的行业招聘实践，形成了以男性为主的员工队伍，因此 AI 算法从历史数据中“学习”（过去以男性为主导的）成功招聘惯例时，得出的结论是需要筛选出包含“女子曲棍球队队长”等经历的简历。[8]

亚马逊并非唯一面临此类 AI 挑战的企业；执法、客户互动、翻译和图像解读等领域也存在类似问题。[9] 一旦 AI 对预期任务产生误解，价值错位问题往往接踵而至，届时人们不禁要问究竟由谁为算法担责。确保 AI 系统按预期运行并且可以解释相关原因，这一点非常重要。

洞察：AI 道德定义

AI 道德是一个多学科研究领域，主要目标在于为所有利益相关方最大程度发挥 AI 的积极影响，减少风险和负面结果，因此，要优先考虑人的能动性和福祉及环境的可持续发展。为此，AI 道德研究主要探讨如何设计和构建一个了解部署场景中所遵循的价值观和原则的 AI 系统。此外，还要确定和研究因社会生活中普遍采用 AI 技术所引发的道德问题，并提出技术和非技术解决方案。例如，数据责任与隐私、公平性、包容性、道德行为能力、价值定位、问责制、透明度、信任度和技术误用均属于此类问题。

重要问题：谁在关注 AI 道德问题 — 为什么

在群体研究中，我们发现了一个令人惊讶的结果：不同地域的群体关注的焦点差异极大。更令人惊讶的是，高管与消费者对重点问题实质的认知存在分歧。

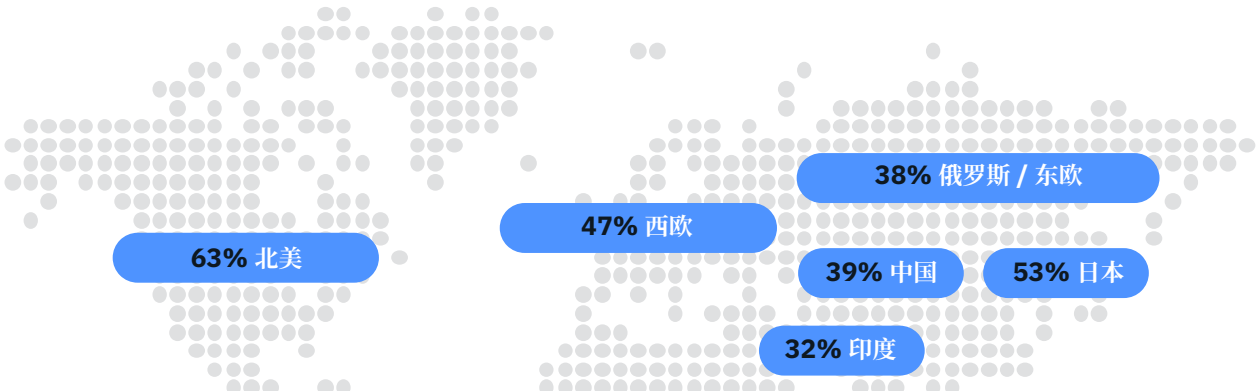
地区观点

从表面上看，不同地区对于 AI 道德问题的看法存在差异，这一点并不意外。事实上，由于文化不同，观点不太可能完全一致，观点一致反倒令人惊讶。但是，调查结果的地理分布引人深思（见图 2）。我们的调研表明：在北美、日本和西欧等一些成熟发达市场中，约有半数高管表示 AI 道德对于企业非常重要。

同时还表明：在发展中市场（如拉美、非洲和东南亚）和高速发展的国家 / 地区（比如印度和中国）中，仅有不足一半的高管认识到同样的（AI 道德）重要性。概括而言，欠发达经济体企业对于经济发展所依赖的 AI 技术的道德影响关注度较低。但是，二者之间的认识差距可能会不断缩小。特别是，当采用 AI 技术的 IT 决策者思考信任、透明和公平问题时，必然会考虑 AI 道德问题 — 2019 年末 IBM 委托开展的一项调研的结果印证了这一点。[10]

图 2

企业对 AI 道德的重视程度因地区而异



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_46728