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。

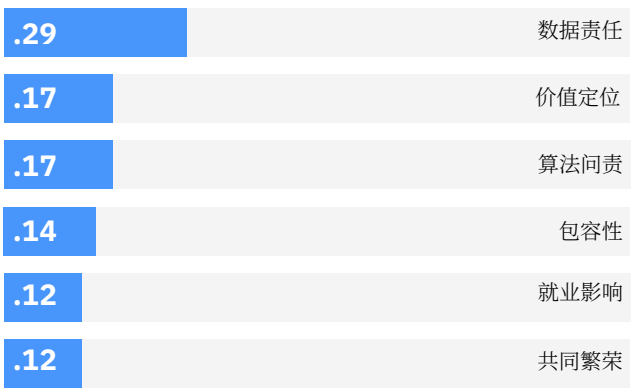
研究期间，我们还请受访高管对开发符合道德规范的 AI 系统所涉及的一系列因素的相对重要性进行评估，包括价值定位、算法问责和包容性。受访者对这些因素的重要性进行并排比较，做出权衡取舍。

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。

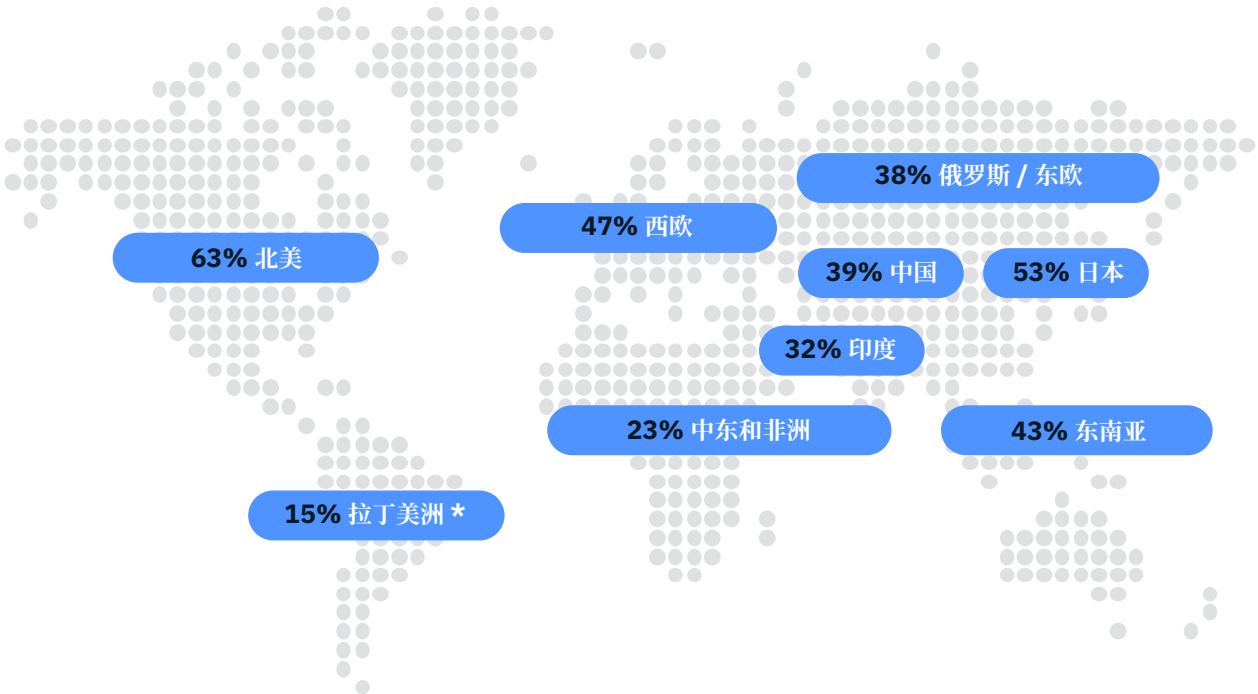
地区观点

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



** 总数不超过 20。*

来源：2018 年 IBM 商业价值研究院全球 AI 道德问题调研。问题：AI 道德在企业中的重要性，N=1,247。

仅有略超过 1/3 的 CHRO 表示组织有义务对受 AI 技术影响的员工进行再培训或重塑技能。

企业观点与民众观点

在分析消费者与高管关注的 AI 道德问题时，结果同样呈现两极分化。在这里，差异不以地域区分，而在于风险类型。例如，在数据责任风险方面，高管和消费者表达了相似的高优先级关注度。然而，在其他领域，两个群体对优先问题的认知却并不相同。

受访高管指出，他们认为 AI 对社会福祉（包括包容性、共同繁荣和就业影响）带来的影响远不及直接影响企业的因素（包括数据责任、价值定位和算法问责）。事实上，高管在考量 AI 道德问题时，共同繁荣和就业影响的重要性排名最低。当然，这些领域是全人类共同关注的焦点。

高管对社会福祉的关注度相对较低发人深思。特别是，在考察 AI 对就业和劳动力市场的影响时尤其值得注意。大量重要调研得出一致结论：AI 将对员工和技能产生巨大影响。[11] 2019 年，我们估计未来三年全球 12 大经济体将有超过 1.2 亿工作者可能需要接受再培训或重塑技能。[12] 在接下来的数月乃至数年，这一人数还可能急剧增长。

2018 年的国家或地区调研表明，2/3 的高管预计 AI 和自动化技术进步需要目前尚未出现的职位和技能。[13] 大多数全球高管 (60%) 估计，在智能自动化的影响下，未来三年最多 5% 的员工需要重塑技能或接受再培训；超过 1/3 (38%) 的高管预测这一比例可能会上升至 10%。[14] 目前，高管可能在集中研究部署 AI 会对组织和利益相关方造成的直接影响。但随着 AI 技术以及大众看法的日趋成熟，更广泛社会问题的影响必须受到重视，需时刻保持警醒。

由于许多工作岗位受到 AI 的影响，因此，如果受访高管只聚焦资方利益，则无疑是目光短浅的。仅有略超过 1/3 的受访 CHRO 表示组织有义务对受 AI 技术影响的员工进行再培训或重塑技能。从实用角度而言，投资培养技能（包括培训员工使用 AI）对于建立高素质员工队伍至关重要；从社会层面而言，倘若将 AI 可能引发的就业错位问题推给其他实体解决，或许会让涉事企业处于舆论的风口浪尖，市场信任毁于一旦。

影响 AI 道德：由谁负责？

由于 AI 道德引发种种风险、不同地域的人们观点各异、高管与消费者的看法存在分歧，人们自然会提出下一个问题：企业如何才能最有效地做出响应？为解决这一难题，我们的调研深入探讨了究竟谁该对组织中的 AI 道德问题负主要责任，以及各级领导对这个问题该有怎样的重视程度。结果表明受访者观点存在严重的不一致。

董事会

在私营组织中，主要由董事会负责 AI 道德规范事宜。根据《G20/OECD 企业治理原则》：“董事会在塑造整个公司的道德伦理形象方面发挥关键性的作用，他们不仅要身体力行，还要指派和监督关键高管和整个管理层。高水平的道德准则符合公司的长远利益，它能够在日常运营和长期合作中为企业赢得声誉和信任。”[15]

我们的调研结果与这些原则完全吻合：80% 的董事会成员告诉我们，AI 道德至少在一定程度上属于董事会层面的问题。不足之处在于：仅有 45% 的董事会成员表示已经为解决这些问题做好充分准备，二者之间的差距令人担忧。

最高层主管

另外，董事会成员希望高管团队推动实施战略，解决 AI 道德问题。但我们的研究表明，CEO 与董事会的步调不太一致：仅有不到一半的受访 CEO 至少在一定程度上将 AI 道德视为 CEO 层面的责任。（一位 CEO 认为：“这些问题应由专注于技术、数据及相关道德原则的董事委员会负责处理。”）此外，CEO 与共同负责执行董事会指令的高管团队的步调也不够一致：CEO 对 AI 道德整体重要性的平均评分远低于最高层主管的评分。

CEO 与董事会的看法之间，以及 CEO 与其高管团队的认识之间双双严重脱节，这是近期企业调研中呈现的最令人不安的结果之一。在传达企业级计划重要性的过程中，采用自上而下的领导模式非常重要，但倘若 CEO 积极性不足，将会对企业产生怎样的影响呢？董事会不得不强制解决这个问题吗？这又会造成多大的冲击呢？

在运营角色方面，同样也呈现出一些不理想的结果。在调研过程中，超过半数高管认为 CTO 和 CIO 是企业 AI 道德问题的主要负责人。换言之，人们基本上将 AI 道德问题视作一项技术责任。当被要求推选一位高管负责人时，仅有 15% 的受访者选择非技术职位。

亟需担起责任

我们的研究还表明，高管们踌躇不决，不情愿将 AI 道德视为企业问题。事实上，高管预计企业外部的多个实体也会对 AI 道德产生显著的影响（见图 3）。他们首先指出科技企业，其次是政府和客户 — 最后才是其他企业。

哈佛大学伯克曼·克莱因互联网与社会中心 (Berkman Klein Center for Internet & Society) 面向全球各大城市举办的区域性治理和包容性圆桌会议的结果证实，企业倾向于将责任推给外部参与者。[16] 该中心的副研究主任 Ryan Budish 观察得出结论：尽管这些会议是加强公私领域合作，推进 AI 道德工作的建设性步骤，“但我们发现双方在责任方面相互推诿。”[17]

图 3

高管预计多种企业外部因素会对 AI 道德问题造成极大的影响

来源：2018 年 IBM 商业价值研究院全球 AI 道德问题调研。问题：您预计 [未来三年] 下列因素会对 AI 道德造成多大程度的影响？N=1,250。

调研结果表明，亟需在董事会层面加强 AI 道德问题教育。

企业内部：采取行动

由于企业对 AI 道德问题的观点含糊不清而且严重脱节，为应对 AI 日益普及的趋势，可以考虑采取哪些步骤？

汲取过去的经验教训

汲取上世纪七十年代生物技术崛起的经验教训。当时，新的生物伦理学学科应运而生，解决了新的科学创新带来的问题。随着时间的推移，在行业支持下，美国陆续成立了多个有关生物伦理的总统委员会。[18] 近期一些有关 AI 道德原则的工作明确引用生物伦理宗旨。[19] 道德哲学家和生物伦理学家 Peter Singer 建议：“相关领域的知名专家应当加入计划。”[20] 独立专家的参与和受聘有助于加强对话、避免陷阱及增进信任。[21]

Singer 强调，AI 与生物技术的一大区别在于，“当今媒体动态激化了人们的情绪，使人们不免有些杞人忧天。”[22] 企业应伸出援手，确保对值得细致推敲的严肃问题予以充分重视 — 而不是任由社交媒体争论不休。关于道德问题的实质讨论通常无法在社交媒体中用 140（或者 280）个字符清楚表达。

另外，Singer 还指出了一个区别：非营利机构（如医院）在生物技术领域发挥着主导作用。[23] 这份观察报告强调，需要研究利润驱动的业务模式的影响，这是当今许多切实的 AI 创新的源动力。

让董事会参与进来

调研结果表明，亟需在董事会层面加强 AI 道德问题教育，邀请董事参与解决 AI 道德问题。在众多公私领域合作伙伴（包括 IBM）的共同努力下，世界经济论坛开发了 AI Board Toolkit，这是一个很好的开端。[24]

其他组织也开展了类似的工作，其中一些专注于解决 AI 道德问题。[25] 例如，普林斯顿大学人类价值观研究中心的一个团队发布了一些案例研究，深入探索 AI 伦理问题。[26] 这些案例得到众多重视 AI 道德和治理活动的教育机构和跨国公司的广泛采用。[27] 又如，案例研究纲要“AI、劳动力和经济”也是一次合作解决 AI 问题的有益尝试，该联盟由包括 IBM 在内的约 90 家合作伙伴机构组成。[28]

实现 AI 道德制度化

除董事会层面以外，还需将 AI 道德融入现有的企业机制之中，从 CEO 办公室和最高管理层一直到各级运营部门。这包括业务行为准则、价值观声明、员工培训和道德咨询委员会等。一家英国企业的法律总顾问指出，企业可以采取的一项最重要的举措是“成立由伦理学家、软件开发人员、数据工程师和法律专家组成的团队”，该观点与一位日企董事会成员和一位加拿大首席风险官的观点一致。

高管需谨防对 AI 道德的支持沦为口头行为，或者成为哈佛大学 Budish 强调的企业治理社区日渐突显的一个问题：道德洗白。[29] 在一篇关于 AI 道德原则的文章中，牛津大学哲学教授 Luciano Floridi 及其合著者 Tim Clement-Jones 强调，必须重视实质效果，而非空洞宣传，明确展示道德咨询委员会、教育计划及其他工具和方法对实际业务决策产生的影响。[30]

意愿很重要，但结果更重要。

关于商业实体是否不止要遵守现有法规、法律和行业标准，还要领先于现状，遵守其他一些规范，人们对此还存在一定的争议。倘若企业的战略目标与领导团队高瞻远瞩的愿景保持一致，势必可以更轻松地证明有关工作和投资的合理性，而更关注短期利益的传统企业则可能会拒绝这样做。但是，所有企业越来越依赖于共享和使用来自客户和其他合作伙伴的数据，因此，为了创造并保护股东及利益相关方的价值，建立信任就显得尤为重要。

周一上午企业行动手册

1 **董事会**
确保 CEO 和最高管理层团队充分了解并参与解决 AI 道德问题；监控进度。

2 **CEO**
成立内部 AI 道德委员会，履行治理、监督和建议职能；确保最高管理层团队明确职责。

3 **CHRO**
评估 AI 对技能和员工队伍的影响；负责实现成果。

4 **技术高管**
将道德治理和培训纳入所有 AI 计划。

5 **风险 / 法务职能**
确保将 AI 道德规范融入各种机制之中，实现价值观制度化。

企业外部：积极准备，迎接未来

有关 AI 道德的规则在企业内部逐步建立：在我们的道德调研中，有超过半数受访企业采用 AI 业务行为准则、价值观声明、员工培训或道德咨询委员会等方法。然而，仅凭企业自身不太可能形成完善的解决方案。

从基础开展培训教育

首先，可以加强通向企业的“管道”。一位董事会成员指出，“可以在高等院校层面开展学生教育，然后面向不同行业的现有专业人员传授适当的知识并提供指导，以此推广 AI 道德规范意识。”

商学院、法学院、计算机科学课程及技术机构（如美国人工智能协会 (AAAI) 以及电气与电子工程师学会 (IEEE)）已开始试点道德相关课程，建立认证机制，施行标准以及创建指南和工具包。[31] 科罗拉多大学波尔得分校的 Casey Fiesler 发起了一份 AI 和科技伦理大学课程的众包清单，截至 2020 年 2 月，清单中已有 250 多门课程，而且还在不断增加。[32] 从这些机构招聘毕业生和成员的企业，也可以在帮助确保开展有效的道德培训方面发挥重要作用。

制定准则和标准

其次，敦促政府开展行动。随着高管对 AI 道德问题整体表现出高度重视，相关立法也在逐步进行之中，这些都表明，监管标准将在 AI 技术的未来发展中发挥实质性的作用。当问及希望在哪些层面确立标准时，参与 AI 道德问题调研的高管普遍表示，期待在国家 / 地区、超国家 / 地区乃至全球层面制定正式准则 — 而不单纯局限于地方 / 区域层面或专业机构。

“可以在高等院校层面开展学生教育，然后面向不同行业的现有专业人员传授适当的知识…以此推广 AI 道德规范意识。”

某受访董事会成员

继 GDPR 生效近一年后，欧盟又于 2019 年 4 月颁布了《可信 AI 道德准则》(Ethics Guidelines for Trustworthy AI)。欧盟委员会任命的 AI 独立高级专家组的工作达到顶点。[33] 该专家组还于 2019 年 6 月发布了《可信 AI 政策与投资建议》(Policy and Investment Recommendations for Trustworthy AI)，并计划于 2020 年针对各领域出台进一步的具体建议版本。[34]

尽管大多数科技企业已发布自己的准则，但一些企业明确支持欧洲高级专家组颁布的指南。这些准则围绕以下七项要求定义了以人为本的“可信”AI 方法：人的能动性和监督；技术健全性和安全性；隐私和数据治理；透明度；多样性、非歧视性和公平性；社会和环境福祉；问责制。[35]

出台统一方法

AI 监管环境将继续不断发展。AI 的颠覆性质，加之采用步伐的持续加快，对许多监管机构的敏捷性带来挑战。然而，解决道德问题属于社会职能。如果充分认清这一点并致力共同承担责任，企业就可以更有效地满足受影响民众的需求。

即使准则因地区和行业而异，仍可将欧盟委员会方法的设计原则作为最佳实践：a) 成立代表多学科和多利益相关方的独立专家组；b) 就指令发表协商一致的人权宣言；以及 c) 明确具体执行建议方向。

有些工作着眼于 AI 技术的当前状态，重点在于提高效益，缓解风险。联合国 AI for Good 平台体现了另外一种方法，即重点探寻全社会的 AI 发展路线。

该平台的未来愿景以联合国的 17 个可持续发展目标及实现目标所需的方法、工具和技术创新为指导。[36] 鼓励采用混合方法，确保实施稳健、完整、全面的评估，确定当前和未来影响。

单纯推行企业教育、专业标准乃至有效监管还不够。任何政府机构、科技企业、法人团体、专业机构、相关民间组织、学术机构或其他实体都无法单独达成所需的目标。各利益相关方必须携手行动。

各种全球商业和 AI 论坛（包括 2020 年 2 月初召开的最新 AAAI/ACM 人工智能、道德与社会大会）进行了热烈而又不失礼节的讨论，上述意见以及其他各种结论屡见不鲜。[37] 此类对话绝不能局限于会议和学术环境，应广泛引入现阶段实施 AI 技术的企业的董事会、行政管理会议、IT 实验室乃至一线运营工作。人文精神是我们的共同基础：包括我们所推动的社会权利，以及所承担的道德责任。

促进可持续的未来

道德问题并不是非黑即白的简单判别。AI 的实际影响非常深刻，在处理 AI 道德问题时也应当秉承同等的严肃态度。在权衡个人隐私与商业价值、监管与创新、透明度与竞争优势的过程中，面临很多实质性问题。为此，在做出权衡时，务必本着周全、文明和积极的态度开展讨论，避免煽动性言辞。

分析利害关系的重要意义或许丝毫不亚于整体反思社会契约。

在 AI 时代，道德问题不仅仅是学者和专家的责任，企业、客户和民众同样发挥着重要的作用。倘若企业主动解决道德问题并采取有意义的措施，就有机会树立未来竞争优势 — 提高人们对 AI 的信任度。

关于作者

Brian C. Goehring

goehring@us.ibm.com
linkedin.com/in/
brian-c-goehring-9b5a453/

Brian Goehring 是 IBM 商业价值研究院副合伙人兼 AI 负责人，拥有 20 多年的战略咨询经验，高级客户遍布大多数行业和业务职能领域。他拥有普林斯顿大学哲学学士学位，并荣获认知研究和德语证书。

David Zaharchuk

david.zaharchuk@us.ibm.com
bit.ly/DaveZaharchuk
@DaveZaharchuk

Dave Zaharchuk 是 IBM 商业价值研究院的研究主任兼政府领域负责人。Dave 负责指导公共和私营领域相关问题的思想领导力调研。

Francesca Rossi 博士

Francesca.Rossi2@ibm.com
linkedin.com/in/francesca-rossi-34b8b95/
@frossi_t

Francesca Rossi 是 T.J.Watson 研究中心 IBM 院士兼 IBM AI 道德全球负责人。加入 IBM 之前，她曾在意大利帕多瓦大学担任计算机科学教授。

选对合作伙伴，驾驭多变的世界

在 IBM，我们积极与客户协作，运用业务洞察和先进的研究方法与技术，帮助他们在瞬息万变的商业环境中保持独特的竞争优势。

IBM 商业价值研究院

IBM 商业价值研究院 (IBV) 隶属于 IBM Services，致力于为全球高级商业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。

了解更多信息

欲获取 IBM 研究报告的完整目录，或者订阅我们的每月新闻稿，请访问：ibm.com/iibv

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

相关报告

Francesco Brenna、Giorgio Danesi、Glenn Finch、Brian Goehring 与 Manish Goyal 合著，“向企业级人工智能转变：勇敢面对技能和数据挑战，积极实现价值”，IBM 商业价值研究院，2018 年 9 月。
https://www.ibm.com/downloads/cas/VNZBXQAO

Annette LaPrade、Janet Mertens、Tanya Moore 与 Amy Wright 合著，“弥合技能缺口之企业指南：培养和留住高技能人才之战略”，IBM 商业价值研究院，2018 年 9 月。
https://www.ibm.com/downloads/cas/KMXOY6XM

Jesus Mantas 著，“实现 AI 之智能方法”，《NACD Directorship》杂志，2019 年 11/12 月。
ibm.com/intelligent-approaches-ai

备注和参考资料

1 “Greek Medicine.” U.S. National Library of Medicine website, accessed December 3, 2019. https://www.nlm.nih.gov/hmd/greek/greek_oath.html

2 Unpublished data from the 2018 IBM Institute for Business Value Global Consumer Study. IBM Institute for Business Value.

3 Unpublished data from the IBM Institute for Business Value survey on AI/cognitive computing in collaboration with Oxford Economics. IBM Institute for Business Value. 2018.

4 Unpublished data from the 2018 IBM Institute for Business Value Global Consumer Study. IBM Institute for Business Value.

5 Jaffe, Justin, and Laura Hautala. “What the GDPR means for Facebook, the EU and you.” CNET. May 25, 2018. https://www.cnet.com/how-to/what-gdpr-means-for-facebook-google-the-eu-us-and-you/

6 Patel, Neel V. “Why Doctors Aren’t Afraid of Better, More Efficient AI Diagnosing Cancer.” The Daily Beast. December 11, 2017. https://www.thedailybeast.com/why-doctors-arent-afraid-of-better-more-efficient-ai-diagnosing-cancer

7 Zech, John R, Marcus A. Badgeley, Manway Liu, Anthony B. Costa, Joseph J. Titano, and Eric Karl Oermann. “Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study.” PLOS Medicine. November 6, 2018. https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1002683

8 Dastin, Jeffrey. “Amazon scraps secret AI recruiting tool that showed bias against women.” Reuters. October 9, 2018. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

9 Dodds, Laurence. “Chinese businesswoman accused of jaywalking after AI camera spots her face on an advert.” The Telegraph. November 25, 2018. https://www.telegraph.co.uk/technology/2018/11/25/chinese-businesswoman-accused-jaywalking-ai-camera-spots-face/; Blier, Noah. “Stories of AI Failure and How to Avoid Similar AI Fails in 2019.” Lexalytics blog. October 30, 2019. https://www.lexalytics.com/lexablog/stories-ai-failure-avoid-ai-fails-2019; Sonnad, Nikhil. “Google Translate’s gender bias pairs ‘he’ with ‘hardworking’ and ‘she’ with lazy, and other examples.” Quartz. November 29, 2017. https://qz.com/1141122/google-translates-gender-bias-pairs-he-with-hardworking-and-she-with-lazy-and-other-examples/; Doctorow, Cory. “Two years later, Google solves ‘racist algorithm’ problem by purging ‘gorilla’ label from image classifier.” Boing Boing. January 11, 2018. https://boingboing.net/2018/01/11/gorilla-chimp-monkey-unpersone.html

10 Survey commissioned by IBM in partnership with Morning Consult: “From Roadblock to Scale: The Global Sprint Toward AI.” January 2020.

11 LaPrade, Annette, Janet Mertens, Tanya Moore, and Amy Wright. “The enterprise guide to closing the skills gap: Strategies for building and maintaining a skilled workforce.” IBM Institute for Business Value. September 2019. ibm.com/closing-skills-gap

12 2018 IBM Institute for Business Value Global Country Survey. IBM Institute for Business Value; “Labor force, total by country.” The World Bank. 2017; IBM Institute for Business Value analysis and calculations. 2019.

13 Unpublished data from the 2018 IBM Institute for Business Value Global Country Survey. IBM Institute for Business Value.

14 Ibid.

15 Principle VI:C. "G20/OECD Principles of Corporate Governance." Organisation for Economic Co-operation and Development. http://www.oecd.org/corporate/principles-corporate-governance.htm

16 Berkman Klein Center for Internet & Society at Harvard University website, accessed November 20, 2019. https://cyber.harvard.edu

17 Interview with Ryan Budish, Levin Kim, and Jenna Sherman. May 2019.

18 Gray, Bradford H. "Bioethics Commissions: What Can We Learn from Past Successes and Failures?" National Center for Biotechnology Information website, accessed November 20, 2019. https://www.ncbi.nlm. nih.gov/books/NBK231978/

19 Floridi, Luciano, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Francesca Rossi, Burkhard Schafer, Peggy Valcke, and Effy Vayena. "AI4People – An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations." Minds and Machines: Volume 28, Issue 4. December 2018. https://link.springer.com/article/10.1007/s11023-018-9482-5

20 Interview with Peter Singer. April 2019.

21 Zimmermann, Annette, and Bendert Zevenbergen. "AI Ethics: Seven Traps." Freedom to Tinker. Princeton Center for Information Technology Policy. March 25, 2019. https://freedom-to-tinker.com/2019/03/25/ai-ethics-seven-traps/

22 Interview with Peter Singer. April 2019.

23 Ibid.

24 "Empowering AI Leadership." World Economic Forum's Shaping the Future of Technology Governance: Artificial Intelligence and Machine Learning platform. https://www.weforum.org/projects/ai-board-leadership-toolkit

25 Butterfield, Kay Firth, and Ana Isabel Rollan Galindo. "AI Toolkit for Boards of Directors." Australian Institute of Company Directors. September 26, 2018. https://aicd.companydirectors.com.au/membership/membership-update/ai-toolkit-boards-directors; Else, Shani R, and Francis G.X. Pileggi. "Corporate Directors Must Consider Impact of Artificial Intelligence for Effective Corporate Governance." Business Law Today. February 12, 2019. https://businesslawtoday.org/2019/02/corporate-directors-must-consider-impact-artificial-intelligence-effective-corporate-governance/; Bethke, Anna, and Kate Schneiderman. "AI Ethics Toolkits." Intel AI blog. February 19, 2019. https://www.intel.ai/ai-ethics-toolkits/#gs.cbuub4

26 "Princeton Dialogues on AI and Ethics Case Studies." Princeton University Center for Human Values and the Center for Information Technology Policy. https://aiethics.princeton.edu/case-studies/

27 Zevenbergen, Bendert. "Princeton Dialogues of AI and Ethics: Launching case studies." Freedom to tinker website (accessed December 13, 2019). May 21, 2018. https://freedom-to-tinker.com/2018/05/21/princeton-dialogues-of-ai-and-ethics-launching-case-studies/

28 "AI, Labor, and the Economy Case Study Compendium." Partnership on AI. https://www.partnershiponai.org/compendium-synthesis/

29 Interview with Ryan Budish, Levin Kim, and Jenna Sherman. May 2019.

30 Florid, Luciano, and Lord Tim Clement-Jones. "The five principles key to any ethical framework for AI." NS Tech. March 20, 2019. https://tech.newstatesman. com/policy/ai-ethics-framework

31 Zittrain, Jonathan, and Joi Ito. "The Ethics and Governance of Artificial Intelligence." MIT Media Lab. November 16, 2017. https://www.media.mit.edu/courses/the-ethics-and-governance-of-artificial-intelligence/; Jagadish, H.V. "Data Science Ethics." Michigan Online, University of Michigan. https://online. umich.edu/courses/data-science-ethics/ and https://www.wsj.com/articles/university-is-rolling-out-certificate-focused-on-ai-ethics-11558517400?mod=djemAIPro; Sullins, John. "Responsible Innovation in the Age of AI." IEEE Xplore. April 2018. https://ieeexplore.ieee.org/courses/details/EDP496; "The Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS)." IEEE Standards Association. https://standards.ieee.org/industry-connections/ecpais.html; "Ethical Considerations in Artificial Intelligence Courses." AI Magazine. July 1, 2017. https://aaai.org/ojs/index. php/aimagazine/article/view/2731; Cutler, Adam, Milena Pribić, Lawrence Humphrey, Francesca Rossi, Anna Sekaran, Jim Spohrer, and Ryan Caruthers. "Everyday Ethics for Artificial Intelligence." IBM Design for AI. https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf; "Ethics in Action." IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. https://ethicsinaction.ieee.org/; "AI Fairness 360 Open Source Toolkit." IBM Research. http://aif360.mybluemix.net/?cm_mc_uid=46348957889315487882623&cm_mc_sid_502 00000=71493781559770063924; "AI for Good Global Summit." International Telecommunication Union. https://aiforgood.itu.int/

32 Fiesler, Casey. "Tech Ethics Curricula: A Collection of Syllabi." Post on Medium.com, accessed November 15, 2019.

33 "High-Level Expert Group on Artificial Intelligence." European Commission. May 2, 2019. https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence

34 Ibid.

35 "Artificial Intelligence at Google: Our Principles." Google AI. https://ai.google/principles/; "Microsoft AI principles." Microsoft AI. https://www.microsoft.com/en-us/ai/our-approach-to-ai; "DeepMind Ethics & Society Principles." DeepMind. https://deepmind.com/applied/deepmind-ethics-society/principles/; "Trusted AI." IBM Research. https://www.research. ibm.com/artificial-intelligence/trusted-ai/; Rende, Andrea. "Europe's Quest For Ethics In Artificial Intelligence." Forbes. April 11, 2019. https://www. forbes.com/sites/washingtonbytes/2019/04/11/europes-quest-for-ethics-in-artificial-intelligence/#5137a7a57bf9; "Ethics Guidelines on Trustworthy AI." European Commission. https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top

36 "AI for Good Global Summit." International Telecommunication Union website for AI for Good Global Summit, accessed December 12, 2019. https://aiforgood.itu.int

37 Third AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society. February 7-8, 2020. New York, NY.

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。洞察根据对自身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.

国际商业机器中国有限公司
北京朝阳区北四环中路 27 号
盘古大观写字楼 25 层
邮编：100101

美国出品
2020 年 4 月