watsonx.governance™

AI 治理

↪ 面向企业

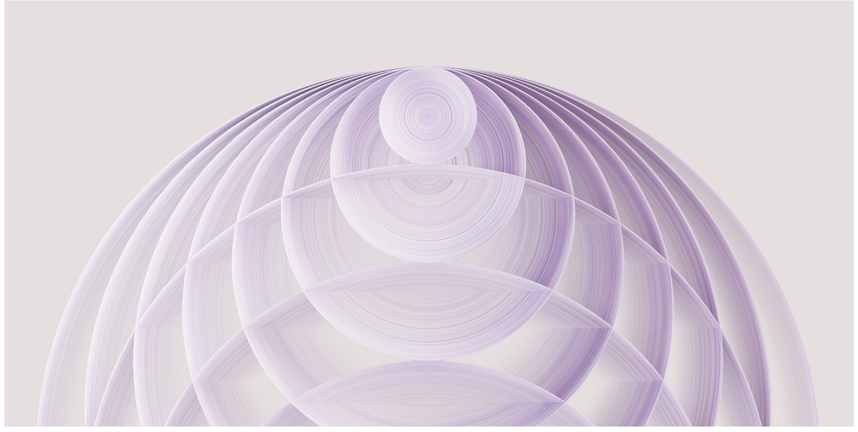
IBM

目录

01

简介

治理令 AI 在企业层面变得实用



您的同事是否在推动有效运行 AI？
这个进程令人振奋，大家都感到非常激动。

《哈佛商业评论》报道[1]：“将生成式 AI 称为重大变革并不夸张。它潜力巨大，针对任何涉及认知任务的功能，均可提高其工作效率。”

当然，AI 的美好前景不可否认。同样可以肯定的是，AI 的风险也是真实存在的。经过深思熟虑的治理方法可支持每个人继续前进。

有了治理作为安全网，就没有理由在 AI 变革方面退缩。

推动企业走上快车道。

请继续阅读完整案例，或免费试用 watsonx.governance。

生成式 AI 的市场规模预计将呈现 24.40% 的年增长率。[2]

02

扩展 AI 的挑战

随着组织领导者在几乎所有行业部署此技术，AI 的影响力也在呈指数级迅猛增长。

同时，很多这类组织的员工和领导者在实施 AI 的以下方面面临困难。

难以从容自信地有效运行 AI

用于 AI 治理的工具种类繁多，但模型的构建往往没有达到良好的清晰度、监控或编目要求。如果没有使用自动化流程进行端到端 AI 生命周期跟踪，可扩展性和透明流程就会受到阻碍。可解释的结果也会变得难以捉摸。

您可能听说过“黑匣模型”，这是 AI 利益相关者日益关注的问题。AI 模型历经构建和部署而生成，但跟踪决策制定的方式和原因并不总是那么容易，即使对于创建它们的数据科学家也是如此。这些挑战会造成效率低下，导致范围偏移、模型延迟或从未投入生产，或是质量水平不一致且风险未被察觉。

阅读对全球 IT 高级决策者进行的关于 AI 采用速度的调查的关键要点。

《2022 年 IBM 全球 AI 采用率指数》→

难以管理风险和声誉

您已经看到了各种头条新闻：生产中的不公平、无法解释或带有偏见的 AI 模型。由此产生的错误假设和决策可能会影响客户并损害您的品牌。

可解释的流程和结果有助于审计员和客户了解具体的分析结果的取得方式。这些流程有助于确保结果不会反映出对种族、性别、年龄或其他关键因素的偏见。这些流程对于患者诊断和治疗方案、审查标记为可疑的交易以及已拒绝的贷款申请至关重要。

采取行动，构建透明、可解释、公平且包容的 AI 系统。您将帮助保障隐私、安全、客户忠诚度和信任等要素。

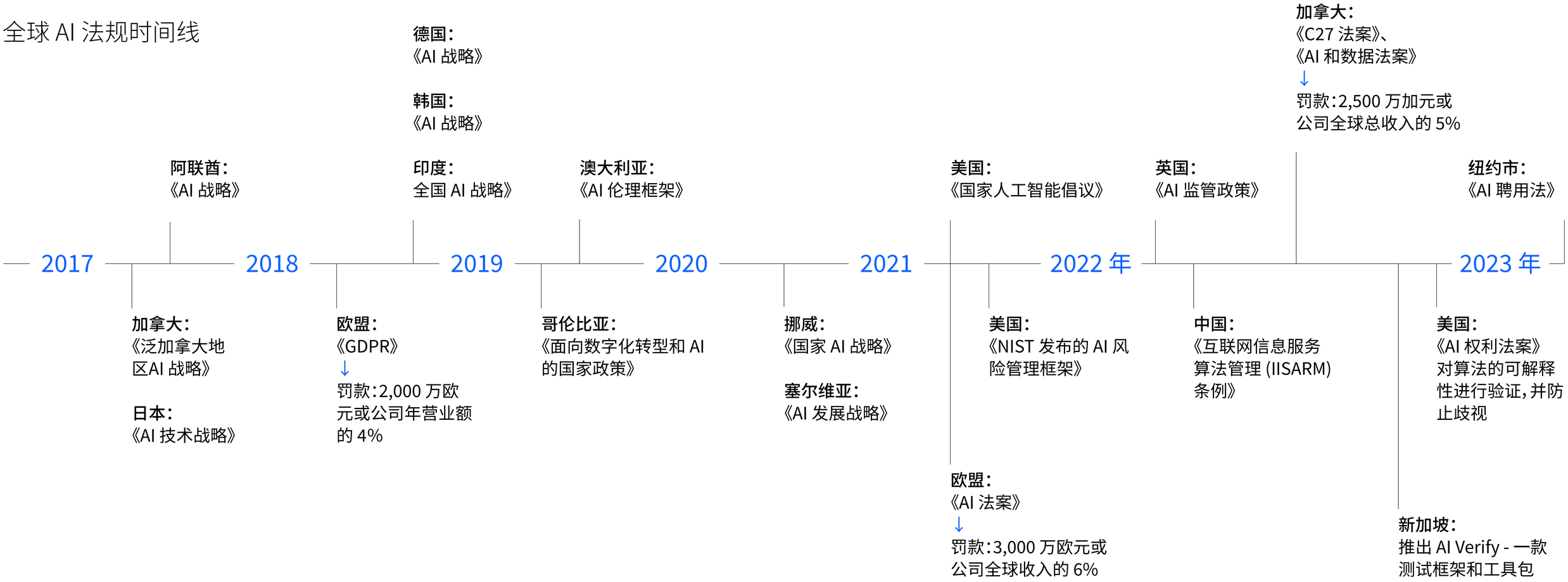
AI 法规在不断变化

成功的 AI 需要遵守地方、区域和国家/地区的法律法规，这些法律法规正在迅速发展。违规可能会导致组织遭受数千万美元的罚款，目前全球争议的一些最严格的AI法规（例如拟议的《EU AI 法案》）就证实了这一点。当前的《EU AI 法案》草案考虑处以最高 3,000 万欧元的罚款，或公司全球收入的 6%。

模型记录至关重要，对于时间紧迫且组织缺乏明确要求的数据科学家而言，这是一个容易遗漏的领域。

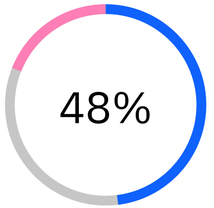
请勿忽视这一步：新法规将需要元数据和沿袭的模型记录。

全球 AI 法规时间线

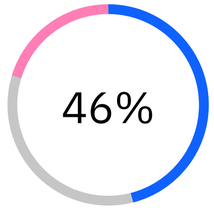

2017
2018
2019
2020
2021
2022 年
2023 年
加拿大：
《泛加拿大地区AI 战略》
日本：
《AI 技术战略》
阿联酋：
《AI 战略》
欧盟：
《GDPR》
↓
罚款：2,000 万欧元或公司年营业额的 4%
德国：
《AI 战略》
韩国：
《AI 战略》
印度：
全国 AI 战略》
哥伦比亚：
《面向数字化转型和 AI 的国家政策》
澳大利亚：
《AI 伦理框架》
挪威：
《国家 AI 战略》
塞尔维亚：
《AI 发展战略》
美国：
《国家人工智能倡议》
欧盟：
《AI 法案》
↓
罚款：3,000 万欧元或公司全球收入的 6%
美国：
《NIST 发布的 AI 风险管理框架》
英国：
《AI 监管政策》
中国：
《互联网信息服务算法管理 (IISARM) 条例》
加拿大：
《C27 法案》、
《AI 和数据法案》
↓
罚款：2,500 万加元或公司全球总收入的 5%
新加坡：
推出 AI Verify - 一款测试框架和工具包
纽约市：
《AI 聘用法》
美国：
《AI 权利法案》
对算法的可解释性进行验证，并防止歧视

80% 的企业领导者认为这些伦理问题中至少有一个是主要问题[3]

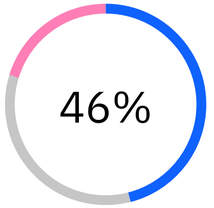
同意 中立 不同意



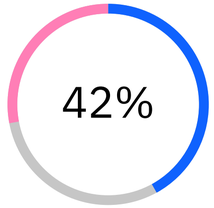
可解释性
认为生成式 AI 做出的决定并非完全可解释。



伦理
关注生成式 AI 的安全和伦理方面。



偏见
认为生成式 AI 会传播的固有偏见。



信任
认为生成式 AI 不可信。

03

需要治理的所有模型

AI 模型的创建方式各有不同。但是所有模型必须得到治理。截至 2023 年，大部分组织均采用传统式机器学习，而其领导者开始采用生成式 AI。

机器学习模型

ML 模型使用预测性分析来识别数据中的趋势和模式。这些模型通过其经历来学习，这样既可以提高技能，还能做出更准确的分析决策。这些模型是根据使用分类数据、未分类数据或混合数据训练的算法创建的。ML 让模型能够自动学习，无需人工干预。

不同的机器学习算法适用于不同的目标，例如分类或预测建模，以便数据科学家使用不同的算法作为不同模型的基础。当数据被引入特定算法时，会经过修改以更好地管理特定任务，并成为机器学习模型。

生成式模型
此类 AI 模型包括基础模型 (FM) 和大型语言模型 (LLM)。这些模型有可能释放数万亿美元的经济价值，因为其卓越的性能可以提高工作效率，并且可以扩展到分布广泛的各类任务。

此类模型具备高度可定制性、可扩展性和成本效益。它们可以查询极大的数据量，并且可一直保持学习状态。“现成的”生成式应用程序几乎不需要专业知识，并且有可能消除许多乏味、耗时的任务。

多年来，统计学中一直使用生成式模型来分析数值数据。最近，深度学习实现了将这些模型扩展到生成图像、音乐、语音、视频、文本甚至代码。用例囊括营销、客户服务、零售和教育等领域。

虽然生成式模型已经将 AI 推上了大多数企业领导者的议程，但其功能也产生了新的复杂性，可能给组织和社会造成风险。

了解如何负责任地扩展 AI。

阅读博客→

04

全面的 AI 治理

像任何其他举措一样，成功的 AI 治理取决于人员、流程和技术的交互。

若要正确实施 AI，您需要一个强大的跨职能团队。对于许多领导者而言，AI 在很大程度上是一项战略要务，而且利益相关者的名单每天都在变得更长。这些人中一些是新接触 AI 生命周期概念，而另一些则是有各种新理由参与 AI 工作。

尽量满足所有这些群体的需求，不给数据科学家造成过重的负担，因为他们几乎没有时间发送或管理审批和信息请求。

首先，让您的利益相关者保持协同一致。获得相关各方的支持，并鼓励他们参与构思、协调结果并采用负责任的 AI。然后，采取措施，确保根据公司的业务管控和法规来定义正确的指标、KPI 和目标集。您还需要监控为 AI 模型确定的特定指标。

了解如何构建全面的 AI 治理方法

阅读博客→

AI 生命周期中的角色

鼓励与主要利益相关者合作，并了解他们最关心的问题：

- CFO，盈利风险
- CMO，品牌风险
- CRO，企业风险
- CDO，高效的数据运营
- CHRO，潜在的人才影响
- CEO，组织问责制
- CPO，监管问责制

进程

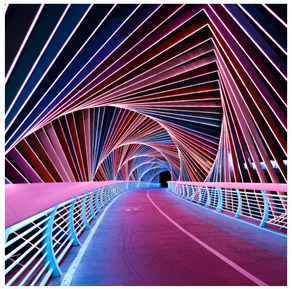
AI 治理可跟踪并记录数据的来源、关联的模型和元数据，以及用于审计的整体数据管道。文档记录应包括训练每个模型的技术、使用的超参数以及测试阶段的指标。这可以提高相对应的利益相关各方在整个生命周期中对模型行为的透明度可视化参与及管理，包括对其开发有影响的数据和模型的潜在风险。

您首先需要对组织当前的 AI 技术和流程进行基准测试和评估。部分流程和利益相关者可能已协调一致并可以扩展，而其他部分可能需要更换。然后创建一组符合合规要求的自动化治理工作流程。新开发的以及现有的 AI 模型均可以采用这些工作流程，其设计应避免上述流程延迟。最后，设置一个框架，以便在模型的指标超过可接受的阈值时向所有者和用户发出警报。

科技

构建计划周密、执行顺畅、控制良好的 AI 需要特定的技术构建块。寻找治理端到端 AI 生命周期并具有以下功能的解决方案：

- 跨不同部署集成多种类型和来源的数据
- 开放、灵活，可与您选择的现有工具配合使用
- 提供具备隐私控制的自助服务访问和跟踪沿袭的方法
- 自动执行模型构建、部署、扩展、训练和监控
- 通过可定制的工作流程连接多个利益相关者
- 支持使用治理元数据为不同的角色构建自定义工作流程



负责任、受治理的 AI 框架

	从容自信地有效运行	管理风险和声誉	加强合规	满足利益相关者的需求
计划	为整个组织的 AI 使用情况定义可衡量的性能指标	审核监控公平性和可解释性的现有流程	根据当前和潜在的 AI 法规进行差距分析	审核负责任的 AI 的现有技能和需求，并与业务目标保持一致
构建	构建当前流程的可追溯性和可审计性	在整个 AI 生命周期中有效运行经更新的流程和检查点	确保可以访问模型文档记录	指定实施负责任的 AI 所需的新角色、技能和学习议程
创建	创建模型沿袭和元数据的自动文档记录	启用公平、可解释和高质量的 AI 模型，最大限度地减少漂移并定期进行政策审核	采取行动，在不产生开销的情况下加强数据科学团队的法规一致性	建立可重复的端到端工作流程，并内置利益相关者审批程序，确保能降低风险并扩大规模

05

watsonx.governance 可实现负责任、透明且可解释的 AI。

了解 AI 治理工具包。IBM® watsonx.governance™ 方法可以帮助您指导、管理和监控组织的各种 AI 活动。

此工具包基于 IBM® watsonx™ AI 和数据平台构建，采用软件自动化来增强您满足监管要求和应对伦理问题的能力。您可以获得全面的 AI 治理能力，而无需从当前数据科学平台切换的过高成本。

模型投入生产之前已经过验证，可以评估业务风险。模型上线后，将对其公平性、质量和漂移进行持续监控。监管机构和审计师可以获得访问提供模型行为解释和预测的文档记录。

您可以了解模型的工作原理，以及模型接收了哪些过程和训练。watsonx.governance 跨越整个生命周期，您的团队在设计、构建、部署、监控和集中事实以实现 AI 可解释性时均可获得帮助。

借助这个治理工具包，审计可以变得更加轻松。跟踪并记录数据的来源、模型及其关联元数据和管道。

文档记录将包括训练每个模型的技术、使用的超参数以及测试阶段的指标。

期望提高每个模型在其整个生命周期中行为的透明度，了解对其开发有影响的数据，以及确定可能风险的能力。

IBM 负责任的 AI 原则

AI 的目的是增强人类智慧

数据和洞察属于其创建者

AI 系统必须是透明的、可解释的

请考虑以下组件：

法规一致性

管理 AI 以满足全球即将出台的安全和透明度法规和政策，即针对 AI 的“营养成分标签”。

- 将外部 AI 法规转换为政策以实现自动化执行
- 加强对审计和合规性相关法规的遵守
- 使用动态仪表板实现跨政策和法规的合规性

自动化元数据
通过 Python Notebook 进行数据转换和沿袭捕获。

风险管理

主动检测和降低风险，监控公平性、偏见、漂移和新的 LLM 指标。

- 自动执行资料和工作流程以符合业务标准
- 大规模识别、管理、监控和报告风险和合规情况
- 使用动态仪表板获得清晰、简洁、可定制的结果
- 加强跨多个地区和地域的协作

公开
支持在第三方工具中构建和部署的模型治理。

生命周期治理

管理、监控和治理来自 IBM、开源社区和其他模型提供商的 AI 模型。

- 从 AI 模型驻留的位置对其进行监控、编目和治理
- 自动捕获模型元数据
- 提高预测准确性，确定 AI 的使用方式和滞后位置

全面
治理端到端的 AI 生命周期。

06

AI 治理实际应用

IBM 首席隐私官↷

扩展自动化以满足 AI 监管要求
IBM 的首席隐私办公室 (CPO) 已在公司的 AI 框架基础上展开要素构建，从而达到满足 AI 监管的要求。具体而言，CPO 通过采取重要措施，将隐私、安全、AI 治理、伦理、流程、技术和工具等要素纳入强有力的合规组合，并在此组合基础上，将 AI 和数据行业领先的各项功能付诸实践。

在 IBM AI 伦理委员会的支持下，IBM CPO 制定了一整套增强型流程，可以更详细地跟踪针对现有标准和适用法律要求的遵守情况。

使用 IBM 的集成治理框架和流程来管理和监控整个公司的 AI 开发和运用，以便团队能够：

- 使用 IBM 工具创建稳健的工作流程，以收集、整合、显示和监控工作流程
- 自动捕获和集成 AI 生命周期中的资料，以加速全局 AI 清单的维护

了解更多 →

07

后续步骤

了解使用 watsonx.governance 工具包创建负责任、透明且可解释的 AI 工作流程的速度，而无需从当前数据科学平台切换的成本。

- 有效实现 AI 治理
- 管理风险和声誉
- 支持法规一致性

开始使用

预约演示 →
了解关于此治理工具的更多信息 →
亲自免费试用 →

1. 《如何利用生成式 AI》，《哈佛商业评论》，2023 年。

2. 《全球生成式 AI》，Statista，2023 年。

3. 《生成式 AI：市场现状》，IBM 商业价值研究院，2023 年。

国际商业机器（中国）有限公司
了解更多信息，欢迎访问我们的中文官网：
https://www.ibm.com/cn-zh
IBM Corporation
New Orchard Road
Armonk, NY 10504

美国出品
2023 年 11 月