2026

奇瑞人工智能合规白皮书

Chery Artificial Intelligence Compliance White Paper

寄语

今天，当一辆奇瑞汽车能够听懂乡音、预判弯道、自主泊入车位，在察觉驾驶者疲惫后主动建议休息时，我们深知，驱动这一切的，远不止是参数与算力，更是对人的理解、对生命的敬畏、对责任的自觉。

人工智能正以史无前例的深度融入出行的每一个瞬间。它赋予机器感知与思考的能力，也提出了一系列亟待明确回答的问题：如何跨越生命安全的首要验证关？如何建立符合人类价值观的 AI 治理架构？如何确保 AI 系统在全生命周期的稳定、可靠与可追溯？这些问题不再只是技术层面的注脚，而已成为关乎用户安全、尊严与幸福的根本命题。

面对 AI 技术的快速演进与全球监管格局的深刻变化，奇瑞始终将“以人为本”作为智能创新的根本原则。我们清醒地认识到，AI 是迷人的武器，要让它忠实地为人类服务，不能成为脱缰的野马。正因如此，我们积极构建面向未来的 AI 治理与合规框架，致力于在研发、产品与服务中融入算法透明、信息安全、隐私保护与伦理考量，使我们的实践既与国内外监管要求相契合，也推动建立出行领域更高的可信赖 AI 标准。

奇瑞坚持，AI 要向新，更要向善。让技术普惠大众，让产业发展更加繁荣，让未来社会更加富有、更加公平、更有温度——这是 AI 的终极意义，也是奇瑞不变的初心。

正是在这样的理念下，本白皮书得以形成，它是我们观全球监管之势、鉴行业先行之智、省自身合规之道的成果，更是奇瑞向用户、向社会、向未来立下的一份郑重承诺。

奇瑞汽车股份有限公司董事长

目录 CONTENT

Trend
趋势

CHERY 2026
DATA COMPLIANCE

AI 进入制度化监管时代

一、AI 作为新一代关键技术

1956 年，人工智能（Artificial Intelligence, AI）作为研究领域被正式提出，开启了以计算方法实现机器智能的探索。此后数十年间，AI 沿着规则与知识驱动、数据驱动等多条技术路线持续演进，并经历了多次技术浪潮与低谷。

自 2010 年起，深度神经网络在海量数据与算力的加持下取得关键突破，图像识别、语音处理等能力快速进步，推动了 AI 的规模化应用。

2017 年，Transformer 架构提出后，大规模预训练模型进入快速发展阶段，并在规模扩展过程中呈现出“涌现”（emergence）现象，即模型规模突破临界点后涌现出意料之外的新能力。而 2022 年起以 ChatGPT 等产品为代表的生成式人工智能（Generative AI）通过对话式交互迅速普及；近年来多模态能力与智能体化趋势进一步推动了 AI 应用的广泛渗透。

伴随技术潜力的释放，AI 应用所衍生的风险也持续上升，从数据来源侵权、算法偏见到系统安全与可靠性等诸多挑战日益凸显。这使得 AI 治理与合规成为与技术创新同等重要、须并行推进的核心议题。

二、AI 应用的常见风险

AI 应用的风险贯穿其全生命周期。我们归纳出以下几类常见风险，以较多国家规制的“AI系统”为例：

1. 数据来源侵权风险：AI 系统的训练高度依赖数据投入。若训练数据的采集、处理与使用缺乏合法性，可能引发个人信息违法处理或知识产权侵权。

2. 算法偏见风险：AI 系统可能吸收并继承训练数据中的历史性偏差，导致在影响个人权益的决策场景（如招聘、信贷、营销等）中产生不公平、歧视性输出，甚至损害特定群体的合法权益和社会公平。

3. 透明度与可解释性风险：部分 AI 系统具有“黑箱”属性，再加上服务提供者信息披露不足，易造成信息不对称。用户可能无法识别交互对象身份、不理解决策逻辑或难以辨识内容真伪，导致知情权与自主选择权受限，并可能受到虚假信息误导。

4. 准确性与可靠性风险：AI 系统通常基于概率预测机制运行，可能产生“幻觉”（hallucination），例如生成不实或虚构内容，或在长尾场景下出现模型漂移（model drift）导致性能下降。在交通、医疗等低容错领域，输出结果一旦发生偏差，可能造成人身伤亡或重大财产损失。

5. 系统安全风险：AI 系统的输入敏感性与数据依赖性可能被恶意第三方利用。例如，通过对抗性攻击（adversarial attack）诱导系统误判，或通过数据投毒（data poisoning）篡改模型预期行为，进而导致系统失控、关键业务信息泄露或安全防线失效等后果。

6. 问责困难：AI 系统的自动化决策、“黑箱”属性以及上线后的持续迭代，使得人类意图、系统决策与损害后果之间的因果链条难以有效厘清。若缺乏覆盖全生命周期的可追溯机制（比如日志记录、关键文件留存），损害发生后可能因证据不足、决策过程不可复现而难以有效审计或归责。

AI 进入制度化监管时代

三、AI 的治理路径演进

各国政府与国际组织对 AI 应用潜在风险的关注由来已久。回溯近十年的治理实践，全球对 AI 的治理路径经历了从伦理准则向制度化监管的演变：

1. 伦理准则阶段（2010 年代中后期）

2010 年代中后期，机器学习技术取得显著突破，社会各界对 AI 应用风险的关注也日益升温。在此背景下，包括经济合作与发展组织（OECD）和欧盟在内的多国政府及国际组织，相继出台了一系列 AI 伦理准则，旨在为 AI 的开发与应用提供规范性指引。这些准则通过提出一系列价值导向与行为原则，如倡导以人为本与尊重人权、强调公平与非歧视、要求透明可解释等，为后续政府制定 AI 相关政策、企业开展自律治理提供框架指引。

然而，单纯的伦理准则逐渐暴露出其局限性——高层次原则难以直接转化为可操作的具体要求，且在商业利益驱动下，仅依靠企业自律无法有效应对系统性风险。这一认识在政策文件中逐步得到体现：欧盟《人工智能白皮书》（2020年）与联合国教科文组织《人工智能伦理问题建议书》（2021年）等重要文件，均明确倡导通过具体监管政策将伦理原则落地。

2. 制度化监管阶段（2020 年起）

自 2020 年以来，全球 AI 治理加速迈入制度化监管的新阶段。根据 OECD 人工智能政策观察站（OECD.AI Policy Navigator）的统计，截至 2026 年 2 月，全球已经有 80 多个司法管辖区及国际组织颁布了 2214 项 AI 相关政策与监管举措。当前，一个覆盖 AI 全生命周期、关注风险治理、持续演进的全球监管格局已初具雏形。

以下章节将从宏观态势与具体规定两个视角，对全球 AI 监管格局进行系统扫描，呈现我们对各国监管思路的观察，解析全球典型的 AI 监管规则的核心要求。

Observation
观察

AI 的全球监管格局

一、宏观视野：全球 AI 监管态势观察

1. 特征一：主要法域呈现“规则密集推出”与“去监管化反思”的双重态势。

- **规则密集推出：**2022 至 2023 年间，中国采取“小步快走”的方式，相继颁布了算法推荐、深度合成和生成式人工智能等专项规则；2024 年，欧盟《人工智能法案》（Artificial Intelligence Act）正式生效；随后，韩国通过了《关于人工智能发展和构建信赖基础的基本法》（Basic Act on the Development of Artificial Intelligence and the Establishment of Foundation for Trustworthiness）。美国、英国、巴西等国家和地区也在持续推动 AI 相关法律及政策草案的制定与讨论。

- **去监管化反思：**在全球监管规则密集推出的同时，部分国家也在反思过度监管带来的问题，开始在风险防控与科技创新之间寻求新的平衡。欧盟作为 AI 监管的先行者，其《人工智能法案》的实施过程中出现了调整迹象。2025 年 11 月，欧盟委员会发布了《数字综合简化提案》（Digital Omnibus Package），提议对部分义务要求及其实施时间进行简化和延期，以缓解合规负担。澳大利亚、加拿大等国政府也强调过度监管对创新的寒蝉效应；美国则通过行政命令推进 AI 发展的去监管化改革。然而，这些调整并未改变全球 AI 监管持续推进的总体趋势。

2. 特征二：主要法域的监管思路、治理范围与核心监管要素趋同。

- **风险导向思路：**基于 AI 系统或应用场景的风险水平来决定监管介入强度与规则配置。例如，欧盟、韩国等法域以“高风险”或“高影响”AI 系统及生成式 AI 为重点监管对象；中国则将算法推荐、深度合成、生成式 AI 等特定技术形态的应用与服务作为监管重点。

- **全生命周期治理：**关注 AI 系统设计、开发、部署及运营全生命周期，要求在每个阶段持续识别、评估并管控风险。

- **核心监管要素趋同：**常见监管要素集中在落实数据治理要求、保障系统的可靠性与安全性、强化信息披露与透明度、健全组织治理与风险管理机制以及保护用户权利等核心维度。

3. 特征三：主要法域监管工具与立法结构的差异仍未收敛。

- **硬法约束与软法引导并存：**欧盟、中国、韩国等法域以具有法律约束力的立法或行政规范（“**硬法**”）构建监管基线，明确义务、设定罚则；英国、新加坡、澳大利亚等法域目前以原则性指引、行业指南和自愿性框架（“**软法**”）为主，引导组织自律。

- **横向统一监管与纵向细分监管分途演进：**欧盟、巴西等法域通过综合性 AI 立法确立横向统一义务框架（“**横向统一监管**”）；另一方面，英国、澳大利亚依托既有行业监管分散治理，中国则针对算法推荐、深度合成、生成式 AI 等技术形态实施纵向专项监管（“**纵向细分监管**”）。

值得注意的是，上述区分更多是监管路径的阶段性差异，而非最终形态。多数法域正在不断调整监管工具与立法结构，未来软硬法结构与横纵向监管模式均存在相互借鉴与转型的空间。全球 AI 监管格局仍在持续演进。

AI 的全球监管格局

二、全景透视：全球 AI 监管要求纵览

以下章节中，我们对全球主要经济体的 AI 监管政策开展了系统性梳理，重点关注已正式出台或已进入实质性立法进程、对企业具有适用效力的 AI 专门规则，包括“硬法”以及具备广泛影响力的“软法”。基于上述梳理，我们就全球 AI 监管要求编制了一览表，以清晰呈现各主要法域在 AI 治理方面的特点。

图例：✅ 已通过/生效（硬法） ➖ 立法进程中（硬法） ⭕ 已通过（软法）

法域[1]	通用监管要求																				
	文件名称	法律效力	规制对象[2]	数据治理		系统准确性、可靠性与网络安全		透明度与信息披露要求							组织治理与风险管理					用户权利		禁止性要求
				数据来源合法	数据质量保障	准确性与可靠性	网络安全	事先告知	决策解释	内容标识	公示披露	下游信息披露	监管报备	事故报告	治理框架	开展评估	人类监督	可追溯性	人员素养	申诉权	不受歧视的权利	明确禁止事项
中国	《互联网信息服务算法推荐管理规定》	✅	算法推荐服务	○	○	●	●	●	●	●	●	○	●	●	●	●	○	●	○	●	●	●
	《互联网信息服务深度合成管理规定》	✅	深度合成服务	●	○	●	●	○	○	●	●	○	●	●	●	●	○	●	○	●	○	●
	《生成式人工智能服务管理暂行办法》	✅	生成式 AI	●	●	●	●	○	○	●	●	○	●	●	○	●	○	●	●	●	●	●
欧盟	《人工智能法案》	✅	AI 系统	●	●	●	●	●	●	●	○	●	●	●	●	●	●	●	●	○	●	●
美国	加利福尼亚州《生成式人工智能：训练数据透明度法案》	✅	生成式 AI	○	○	○	○	○	○	○	●	○	○	○	○	○	○	○	○	○	○	○
	加利福尼亚州《人工智能透明度法案》	✅	生成式 AI	○	○	●	○	○	○	●	○	○	○	○	○	○	○	○	○	○	○	○
	纽约州《负责任人工智能安全与教育法》	✅	AI 系统	○	○	○	●	○	○	○	●	○	●	●	●	●	○	●	○	○	○	●
	科罗拉多州《人工智能法案》	✅	高风险 AI 系统	○	●	○	○	●	●	○	●	●	○	●	●	●	○	●	○	●	●	○
韩国	《关于人工智能发展和构建信赖基础的基本法》	✅	高影响 AI 系统	○	○	●	●	●	●	●	○	○	○	○	●	●	●	●	○	○	○	○
越南	《人工智能法》	✅	AI 系统	●	●	●	●	●	●	●	●	○	●	●	●	●	●	●	○	○	●	●
哈萨克斯坦	《人工智能法》	✅	AI 系统	●	●	●	●	●	●	●	○	○	○	○	●	●	○	●	○	●	●	●
萨尔瓦多	《人工智能与技术促进法》	✅	AI 系统	●	○	○	○	●	●	○	○	○	○	○	○	○	○	○	○	●	●	○
土耳其	《人工智能法案》	➖	AI 系统	●	○	●	○	○	○	○	●	○	●	○	●	●	○	○	○	○	●	●
巴西	《人工智能使用法案》	➖	AI 系统	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	○	●	●	●
阿根廷	《人工智能负责任使用适用法律制度》	➖	AI 系统	●	●	●	●	●	○	●	●	○	●	●	●	●	○	●	●	○	●	●
加拿大	《先进生成式人工智能系统负责任开发与管理自愿行为准则》	⭕	生成式 AI	●	●	●	●	○	○	●	●	●	○	○	●	●	●	●	●	○	●	○
澳大利亚	《人工智能应用实施指南》	⭕	AI 系统	●	○	●	●	●	●	●	○	●	○	●	●	●	●	●	●	●	●	○
日本	《人工智能技术研发促进与利用法案》及其适用指南	⭕	AI 系统	●	○	●	●	○	●	●	●	○	○	○	●	●	○	○	●	○	●	○
新加坡	《生成式人工智能模型管理框架》	⭕	生成式 AI	●	●	●	●	○	○	●	○	●	○	●	○	●	○	○	●	○	●	○
中国香港	《香港生成式人工智能技术及应用指引》	⭕	生成式 AI	●	●	●	●	○	○	●	●	●	○	○	●	●	●	●	○	○	●	●

[1] 地理范围：除一览表中所列法域外，我们还对瑞士、印度等另外 19 个法域的 AI 监管现状进行了研究。鉴于这些法域尚未制定正式的 AI 立法，或立法进程仍存在重大不确定性，且缺乏具有广泛约束力的规范性文件，故未纳入本表。

[2] 规制对象：本一览表所列规制对象 AI 系统泛指 AI 系统、模型、产品和服务。表中多数监管要求仅适用于特定类别的 AI 系统（如高风险 AI 系统、高影响 AI 或生成式 AI），具体请参照各国规范性文件的相关规定。

AI 的全球监管格局

二、全景透视：全球 AI 监管要求纵览

我们将一览表所涵盖的监管要求划分为数据治理，系统准确性、可靠性与网络安全，透明度，组织治理与风险管理，用户权利以及禁止性要求等六大核心模块。

数据治理

聚焦 AI 数据风险管控，核心是保障训练数据来源合法、质量合格，防范数据使用违规及结果偏差风险。

组织治理与风险管理

建立 AI 全生命周期内控与风险管理体系，通过内控框架、风险评估、人员管理等关键举措实现系统可控。

系统准确性、可靠性与网络安全

AI 系统须在生命周期内维持适当的准确水平，防范运行环境偏差及外部攻击。

用户权利

明确 AI 系统用户享有对不利决策提出申诉、要求人工复核，以及不受算法歧视的权利。

透明度与信息披露要求

通过透明度义务缓解信息不对称，需向相关方履行告知、披露与报备等义务。

禁止性要求

划定 AI 不可接受风险边界，严禁开发使用危害公共利益及损害合法权益的 AI 系统。

1. 数据治理

数据治理是管控 AI 数据风险的核心要求，主要关注点包括：

- **数据来源合法：**要求 AI 系统的训练、验证和测试数据取得及使用具备合法基础，符合个人数据保护与知识产权保护规则。

例如，欧盟《人工智能法案》第 10 条、第 53 条规定，高风险人工智能系统的训练、验证和测试数据应纳入数据治理与管理实践，尤其须关注数据来源；若涉及个人数据，还应关注其原始收集目的。通用人工智能模型的提供者也必须制定相关政策，以遵守欧盟版权规则。

- **数据质量保障：**要求训练、验证和测试数据具备相关性、代表性，尽可能完整、无错误，并能够识别与减少可能导致歧视性结果的偏差。

例如，中国《生成式人工智能服务管理暂行办法》第 7 条要求服务提供者应依法处理训练数据，提高数据质量，增强其真实性、准确性、客观性与多样性。

2. 系统准确性、可靠性与网络安全

要求 AI 系统在生命周期内维持适当的准确水平，防范运行环境偏差及外部攻击。

- **准确性与可靠性：**要求系统在预期用途下达到适当准确水平，且有效应对运行环境中可能出现的错误、干扰或异常情况，通过技术与组织措施确保稳定运行。

例如，欧盟《人工智能法案》第 15 条要求高风险 AI 系统应在生命周期内保持适当的准确性、稳健性和网络安全水平。

- **网络安全：**要求采取技术措施抵御对抗性攻击、数据投毒等利用系统漏洞的恶意行为。

例如，澳大利亚《人工智能应用实施指南》（Guidance for AI Adoption: Implementation practices）第 5.4 条规定，AI 系统需实施完善的数据和网络安全措施以应对 AI 特定风险。

AI 的全球监管格局

3. 透明度与信息披露要求

透明度义务旨在缓解 AI 提供者与用户间信息不对称，目前监管侧重于以下维度：

面向用户：

- **事先告知：**在用户与 AI 系统直接交互时，应以适当方式告知其交互对象为 AI。

例如，韩国《关于人工智能发展和构建信赖基础的基本法》第 31 条规定，在提供生成式 AI 产品和服务时，AI 业务运营者应提前通知用户该产品和服务是由 AI 生成的。

- **决策解释：**针对高风险或影响用户权益的决策，提供解释及 AI 贡献度说明。

例如，欧盟《人工智能法案》第 86 条规定，受高风险 AI 系统决策影响的自然人有权要求部署者提供清晰、有意义的解释。

- **内容标识：**对 AI 生成内容添加显式或隐式标识。

例如，美国加利福尼亚州《人工智能透明度法案》（AI Transparency Act）第 22757.3 条规定，生成式 AI 提供商应在生成内容中嵌入隐式标识，并向用户提供添加显式标识的选项。

- **公示披露：**在官方渠道披露模型技术原理、训练数据、算法歧视风险或注册备案号相关的信息。

例如，中国《互联网信息服务算法推荐管理规定》第 7、16、26 条规定，算法推荐服务提供者应公开算法推荐服务相关规则、基本原理、目的意图和主要运行机制以及备案编号等信息。

面向部署者：

- **下游信息提供：**AI 系统开发者应向下游部署者提供系统预期用途、系统性能限制、人类监督条件等关键信息。

例如，新加坡《生成式人工智能治理模型框架》（Model AI Governance Framework for Generative AI）要求 AI 模型开发者向下游部署者披露训练数据、基础设施、评估结果等关键信息。

面向监管：

- **监管报备：**要求就相关 AI 系统或服务（如高风险 AI 系统、具有舆论影响力的生成式 AI 服务等）履行向监管备案、注册登记，或提交评估报告等资料。

例如，中国《互联网信息服务算法推荐管理规定》、《互联网信息服务深度合成管理规定》、《生成式人工智能服务管理暂行办法》规定，算法推荐、深度合成、生成式 AI 服务具有舆论属性或社会动员能力的，应履行相应的备案义务。欧盟、越南等地区法律明确，高风险 AI 系统需履行数据库注册及信息更新义务。

- **事故报告：**发生事故（比如安全事件、算法歧视等）后向监管机关报告。

例如，美国纽约州《负责任人工智能安全与教育法》（Responsible AI Safety and Education Act）第 1422 条要求大型 AI 开发者在知晓安全事件发生后的 72 小时内履行报告义务。

AI 的全球监管格局

4. 组织治理与风险管理

要求企业搭建内控体系，确保 AI 系统全生命周期可管可控。

- **治理框架：**应针对 AI 系统制定内部治理计划，比如质量管理体系、风险管理框架等。

例如，中国《互联网信息服务深度合成管理规定》第 7 条要求，深度合成服务提供者应建立全面的信息安全管理等制度，覆盖科技伦理、数据安全、个人信息等各方面。欧盟《人工智能法案》第 17 条则要求，高风险 AI 系统提供商应建立覆盖设计与验证、开发与质保、数据管理、风险管理等环节的质量管理体系。

- **开展评估：**AI 系统应实施风险评估、影响评估或安全评估。

例如，韩国《关于人工智能发展和构建信赖基础的基本法》第 32 条、第 35 条规定，AI 提供者应对达到标准的 AI 系统开展全生命周期安全风险评估，高影响力 AI 产品或服务应事前开展基本人权影响评估。

- **人类监督：**设置人类监督机制，使部署者能够理解系统运行并在必要时进行干预或接管。

例如，巴西《人工智能使用法案》（Bill on the use of Artificial Intelligence）第 20 条规定了应确保人类监督者能够有效理解、干预并控制高风险 AI 系统。

- **可追溯性：**要求系统具备日志记录功能，组织应留存事件记录、技术文档等关键证明。

例如，欧盟《人工智能法案》第 11 条、第 12 条对高风险 AI 系统的提供者、进口商到部署者设置了严格的技术文档留存及日志记录义务。

- **人员素养：**要求对工作人员进行培训，提升其 AI 知识水平。

例如，欧盟《人工智能法案》第 4 条要求 AI 系统提供者和部署者确保其工作人员具备与其职责匹配的 AI 素养。

5. 用户权利

明确 AI 系统用户的权利，包括对 AI 决策提出申诉，要求人工复核的权利以及不受歧视的权利。

- **申诉权：**用户有权对不利决策提出异议并要求人工复核。少数法域允许用户选择退出 AI 决策。

例如，澳大利亚《人工智能应用实施指南》第 2.2 条规定，受 AI 系统影响的个人或相关方，享有对 AI 决策及使用行为提出质疑、进行申诉并获得人工复核与补救的权利，相关流程应便捷可及、清晰易懂。

- **不受歧视的权利：**采取措施降低基于种族、性别等受保护特征的算法歧视。

例如，欧盟《人工智能法案》第 5 条规定 AI 系统不得针对种族、宗教信仰、性取向等受保护特征实施算法歧视，保障用户不受歧视的权利。

6. 禁止性要求

通过禁止性条款明确不可接受的风险边界，比如禁止潜意识操纵、教育和公共场所情绪识别、利用弱势群体脆弱性以及具有明显社会危害等特征的 AI 系统。

例如，中国《互联网信息服务深度合成管理规定》第 6 条明确禁止危害国家、社会公共利益以及个人合法权益的深度合成服务。

除上述针对 AI 系统自身的通用要求外，主要法域在自动驾驶汽车、金融、医疗等关键领域已出台基于应用场景的细分监管规则，相关行业的企业须同步符合该等规定。

AI 的全球监管格局

三、违规后果

为确保监管要求落地，具有法律约束力的硬法通常设有明确罚则。例如，欧盟《人工智能法案》规定，对于违反禁止性规定的组织，最高可处以 3500万欧元或全球年营业额 7% 的罚款（以金额较高者为准）；对于违反高风险 AI 系统要求等大多数其他违规行为，最高可被处以 1500 万欧元或全球年营业额的 3% 的罚款（以金额较高者为准）。中国《生成式人工智能服务管理暂行办法》亦明确，生成式 AI 服务提供者违反该办法规定的，由相关主管部门依照网络安全、数据安全等法律、行政法规规定予以处罚；法律、行政法规没有规定的，由有关主管部门依据职责予以警告、通报批评，责令限期改正；拒不改正或者情节严重的，责令暂停提供相关服务。违反软法虽一般不会直接触发行政处罚，但可能被视为偏离行业标准和最佳实践。

对于企业而言，这些监管要求与违规后果已构成明确的合规压力。因此，建立完善的 AI 治理与合规体系，成为企业的必然选择。

Practice
实践

企业的 AI 合规实践

一、AI 合规工作的实践落点

AI 合规虽然是一个新的合规领域，但企业可以在现有产品开发、安全合规、运营运维基础上进行拓展与衔接。尤其是企业在前期数据合规工作中所形成的评估机制、透明度设计、内控流程及用户权利响应机制等成果，为统筹推进 AI 合规提供了坚实的基础。

我们总结了企业高效开展 AI 合规工作的七个关键落脚点，以切实履行 AI 监管要求：

1. 数据治理：建立训练、验证与测试数据的准入评估机制。审查数据来源合法性及个人数据处理合规性；通过评估数据的相关性、代表性、是否完整、无误，识别并纠正偏差，确保数据质量达标。

2. 系统可靠性与安全防护：制定并验证模型准确性及可靠性标准。开展 AI 系统专项安全测试（如红队演练以及渗透测试），实施针对性防护并持续监测运行安全。

3. 透明度及信息披露义务：满足面向用户、部署者与监管机构的信息披露义务，落实用户告知与内容标识、向部署者提供必要的技术与合规文档，以及履行相关的监管报备义务。

4. 建立内部治理体系：设立 AI 治理框架与内控流程，执行风险评估制度、人工监督、可追溯机制，并提升员工的 AI 素养。

5. 用户权利救济保障：健全用户权利响应与救济渠道，同时通过产品设计防止 AI 系统决策对用户造成歧视。

6. 审查系统用途：制定内部的 AI 限制或禁止用途规则。新系统须经用途审查方可上线，已有系统定期排查，对触及禁止性规定的系统执行整改或关停。

7. 监测监管动向：追踪全球立法与执法动态，研判监管趋势，调整企业的产品设计与合规策略。

企业的 AI 合规实践

二、领先企业的 AI 合规实践

1. 领先企业的 AI 治理体系建设历程

截至目前，已有领先企业构建了较为完善的 AI 治理体系。以美国某大型科技企业为例，其 2018 年开启了“负责任 AI”体系建设历程，现已被业界广泛视为 AI 合规的标杆。

领先实践：某大型科技企业的“负责任 AI”体系建设之路

该企业自 2018 年起，系统化推进“负责任 AI”项目，其历程包括三个阶段：

第一阶段：原则确立（2018 年）

率先发布公平、可靠、隐私、包容等六项 AI 原则，为后续所有工作奠定了价值基石。

某领先科技企业 AI 治理架构

第二阶段：治理体系建设（2018 年至今）

围绕原则，构建了一个贯穿组织、流程与工具的多层次治理体系：

- **制度与框架：**出台公司级《负责任 AI 标准》，明确全生命周期要求；其风险管理框架与美国国家标准与技术研究院（NIST）《AI 风险管理框架》（AI RMF）保持对标，以管理生成式 AI 风险。
- **治理架构：**形成由董事会监督、跨职能委员会协调、负责任 AI 办公室作为核心执行机构的治理架构。
- **工具建设：**通过《透明度说明》披露服务信息，并向客户提供合规工具。
- **人才与社区：**内部建立了数百人的负责任 AI 社区，AI 素养必修课程完成率高达 99%。
- **社会影响：**持续资助前沿研究，积极参与全球多方治理对话，从行业生态层面推动 AI 治理实践。

第三阶段：合规工作启动（2024 年起至今）

面对欧盟《人工智能法案》等新规，其治理体系具备敏捷的适应性：

- **针对禁止性实践：**开展存量系统筛查，同步更新内部政策、营销指引与合同条款，确保业务全线合规。
- **提高 AI 素养：**构建 AI 知识库，并为员工定制课程、向公众开放培训资源。
- **通用 AI 模型（GPAI）合规准备：**在开发流程中增设模型级政策，构建自动化文档流程，并与欧盟监管机构沟通，为 GPAI 规则生效做准备。

* 本案例信息均整合自该企业公开的《2024 年负责任 AI 透明度报告》及《2025 年负责任 AI 透明度报告》

企业的 AI 合规实践

2. 领先企业 AI 合规实践经验

领先的 AI 治理与合规实践为尚处摸索期的企业提供了宝贵经验，包括：

第一，以全生命周期风险管理为治理理念的锚点。

将风险管理融入 AI 系统的设计、开发、部署及运营全生命周期，构建起一套稳定的合规内核。这一策略精准契合了全球不同法域监管对 AI 全生命周期风险管理的共识性要求，使企业能够以这套通用机制为锚点，满足各法域监管的核心要求。

第二，以制度化治理架构保障理念落地。

通过董事会监督、跨职能委员会协调、负责任 AI 办公室执行的治理架构，将权责链条与关键决策节点明确固化，以实现 AI 系统研发与运营全流程的风险管控。此举以制度化和组织化的方式，保障了治理理念得以有效贯彻。

第三，以工程化工具与组织进化支撑持续运转。

通过开发自动化工具、标准化文档，领先企业将 AI 合规要求“工程化”，提升了效率与可追溯性。在工具之外，组织的进化能力同样重要：内部 AI 社区的持续运转、员工素养的提升、与监管机构的有效沟通——这些工作让治理体系具备了持续运转的适应性。

领先企业的 AI 治理体系，并非依照预设蓝图一蹴而就，而是在持续解决问题的过程中逐步“生长”出来。从发布治理原则、构建治理框架，到启动合规响应，领先企业经历多年迭代与沉淀，逐步实现三个关键转变：治理要求嵌入开发流程，治理责任融入组织体系，责任意识内化为员工素养。“负责任 AI”由此从阶段性项目演变为组织的运转机能。这一过程没有捷径，需要持续投入与试错，方能沉淀出真正契合企业自身逻辑的治理体系。

企业的 AI 合规实践

三、奇瑞的 AI 合规实践

奇瑞 NEXTAI 智能研究院成立于 2024 年 6 月，是奇瑞设立的智能化核心科研机构，愿景为“AI 重塑高效未来”，定位为“致力于驱动集团全面智能化转型，以创新和全球信赖为驱动力的研究院”。研究院聚焦人工智能前沿探索、创新性技术及企业全场景数字化应用研究，始终致力于将 AI 合规理念融入 AI 全生命周期，并已取得了一系列阶段性实践成果。

案例 1 **NEXTAI-Coder 智能编程助手：**选用高标准开源数据集进行训练，严格规避知识产权侵权风险，从源头保障训练数据的质量与合规性。

案例 2 **零件 BOM 智能审核平台：**以奇瑞内部自主沉淀的 3D 数模与 BOM 清单作为平台训练数据，确保底层数据来源的自主性与合法性。

案例 3 **NEXT-Ada 高频任务深度 AI 化平台：**严格实施权限控制与数据隔离，确保员工调用 Agent 处理会议、文档、代码等高频任务时，内部数据不越权访问、不外泄。

案例 4 **零件成本深度寻金 AI 平台：**基于大模型与加密数据库，安全拆解黑盒及灰盒零件的成本。对核心数据及采购核价逻辑进行隔离，仅在内网运算，严防机密数据泄露。

案例 5 **智能人才发现助手：**坚持以能力图谱、项目经验及绩效数据为核心维度进行人岗匹配，严格排除年龄、性别、婚姻状况等非职业竞争要素，保障招聘公平。

案例 6 **AI for CAE 物理仿真平台：**为仿真 AI 模型建立版本溯源机制，确保白车身刚度、风噪等物理预测数据的所有修改均有留痕且不可篡改，实现全过程可追溯。

案例 7 **DevOps 智能监控平台：**提供生产线全流程智能监控，针对故障自愈、配置变更等高风险环节设置人工复核节点，从而实现对 AI 运维活动的有效监督。

案例 8 **AI 赋能培训：**开展 AI 赋能培训 40 余场，覆盖各核心事业部与职能部门，参训人次上万；联合奇瑞大学启动 AI 布道师培养计划，深化全员 AI 素养。

案例 9 **AI 合规准入评估流程：**创设 AI 模型与应用合规准入机制，对训练数据来源合法性、模型伦理风险、算法备案义务等方面开展评估，保障 AI 模型与应用合规上线。

企业的 AI 合规实践

四、AI 合规实践的未来挑战

我们预计，未来 AI 合规工作需应对以下核心挑战：

1. 监管环境的复杂性

全球化企业面临“横向统一立法、纵向行业细分、属地司法管辖”交织的立体监管矩阵。不同法域规则可能存在显著差异，例如同一模型的训练数据在 A 法域符合版权豁免，在 B 法域可能构成侵权。加之全球监管规则处于高频演变期，执法尺度动态调整，提升了企业 AI 合规工作的难度。

2. 新技术范式的规则不确定性

AI 技术范式的迭代速度快于监管规则更新，导致新技术应用常处于“标准模糊”状态。这要求企业在 AI 技术创新性应用时必须自主研判合规风险，建立一套能够动态适应技术变化的内控机制。

3. 全生命周期的风险管理难度

将风险管理嵌入产品全生命周期，意味着需要对现有研发运维体系进行流程改造与重构，匹配相应的工具，将合规检查点嵌入数据采集、模型训练、部署上线及持续监控的各个环节。核心难点在于确保风险可管、可控的同时，避免合规流程阻碍业务敏捷迭代。

4. 专业资源的供给短缺

上述所有挑战的解决，依赖于懂规则又懂技术的复合型人才以及成熟的解决方案。然而，当前市场上专家级人才稀缺，适配复杂业务场景的解决方案不足，同时企业内部 AI 素养提升需要时间和资源持续投入。这成为制约企业 AI 合规战略有效实施的挑战。

Envision
展望

奇瑞 AI 合规的未来展望

AI 已迈入制度化监管时代。面对技术快速演进所带来的复杂风险与深远影响，各主要法域正加速构建监管政策体系。

在此背景下，企业亟需构建系统性的 AI 治理与合规能力，需要将风险管理嵌入 AI 系统的全生命周期之中，在数据治理、系统准确性与可靠性、网络安全、透明度建设、组织治理与风险管理、用户权利保障以及禁止性规定等方面，切实回应各主要法域的监管要求。

自奇瑞启动各类 AI 项目以来，我们的目标不止于保障产品满足当下的监管合规要求。更重要的是，通过体系化建设将 AI 治理逐步转化为组织的基础机能，使 AI 等新技术的合规最终内化为企业的基因，沉淀为用户信任与品牌口碑。

AI 合规治理是一个需要持续投入、不断迭代与长期沉淀的过程。作为行业参与者，我们愿与业界共同探索、互鉴经验，促进 AI 治理实践不断走向成熟，成为推动行业向善的坚定力量。

Appendix
附录

CHERY 2026
DATA COMPLIANCE

附录：全球主流法域 AI 监管规范

中国

《互联网信息服务算法推荐管理规定》

《互联网信息服务深度合成管理规定》

《生成式人工智能服务管理暂行办法》

《人工智能生成合成内容标识办法》

《网络安全技术 人工智能生成合成内容标识方法》

《网络安全技术 生成式人工智能服务安全基本要求》

中国香港

《生成式人工智能技术及应用指引》

中国台湾

《人工智能基本法》

美国

加利福尼亚州

《人工智能透明度法案》

《前沿人工智能透明度法案》

《生成式人工智能：训练数据透明度法案》

《陪伴聊天机器人法》

纽约州

《负责任人工智能安全与教育法》

科罗拉多州

《人工智能法案》

德克萨斯州

《负责任人工智能治理法案》

欧盟

《人工智能法案》

《人工智能系统定义指南》

《禁止人工智能行为指南》

《通用人工智能模型提供者指南》

《通用人工智能行为准则》

英国

《促进创新的人工智能监管路径》

附录：全球主流法域 AI 监管规范

其他地区

日本

《人工智能技术研发促进与利用法案》

《确保人工智能相关技术研发与应用适当性的指南》

韩国

《关于人工智能发展和构建信赖基础的基本法》

《生成式人工智能开发与应用个人信息处理指南》

新加坡

《生成式人工智能治理模型框架》

越南

《人工智能法》

印度

《人工智能治理指南》

哈萨克斯坦

《人工智能法》

沙特阿拉伯

《公众生成式人工智能指南》

阿拉伯联合酋长国

《人工智能伦理原则与指南》

卡塔尔

《人工智能伦理使用的原则与指南》

土耳其

《人工智能法案》

澳大利亚

《人工智能应用实施指南》

新西兰

《负责任的人工智能企业指南》

加拿大

《先进生成式人工智能系统负责任开发与管理自愿行为准则》

巴西

《人工智能使用法案》

阿根廷

《人工智能伦理原则与指南》

萨尔瓦多

《人工智能与技术促进法》

声明

本文件的信息来源于本次调研所收集的数据以及公开的资料，我们对信息的完整性、准确性或及时性概不做出任何保证或担保，也不就业绩、适销性和适用于特定用途等方面提供任何明示或暗示的担保，在不同时期可能会得出与本文件不一致的观点。

本文件仅供一般参考使用，不构成具体事项和咨询意见，不构成提供任何形式的法律咨询、会计服务、投资建议或专业咨询，本文件所提供的信息不能取代专业税收、会计、法律咨询或其他相关专业咨询建议。奇瑞不对本文件内容承担审慎责任。奇瑞不就本文件内容向任何人士承担任何责任或义务，也不向任何人士承担因本文件所引起的或与本文件有关的任何责任或义务。读者不应依赖本文件内容做出投资或其他商业决定。如需具体意见，请咨询专业顾问。

DIT | NEXT AI

数 智 化 赋 能 业 务 成 功

主编人员

戴闯、莫达琳、李怡 、邹家琦、丁言中

https://m.weibo.cn/u/2005342162

https://www.facebook.com/cheryinternational

https://www.youtube.com/@cheryinternational_official