中国平安 PING AN
专业 · 价值

专业 让生活更简单

平安集团
负责任AI政策声明

2026年9月

制定目的

为规范中国平安保险（集团）股份有限公司（以下简称“平安”或“平安集团”）负责任人工智能（AI）治理工作，保障 AI 技术应用安全、可控、可靠，推动技术合规赋能主业发展，助力社会及生态可持续发展，平安特制定本政策声明。

负责任 AI 治理原则与措施

平安遵循“以人为本、人类自治、安全可控、公平公正、公开透明”的五大伦理原则，持续完善负责任 AI 治理体系，并将“AI 安全和治理”纳入集团 AI 培训体系的核心模块，夯实全员合规意识，确保安全、可控及符合伦理的 AI 技术应用。

以人为本

遵循人类的价值观和伦理道德，促进人机和谐，服务人类文明进步，以保障社会安全、尊重人类权益为前提，避免误用，禁止滥用、恶用。

人类自治

- 应用过程中确保人类知情权，告知可能产生的风险和影响，人类能够干预或监督 AI 做出的每一个决定；
- 明确 AI 应用的业务属主及技术属主，合理划分各方责任及分工，不得将应由人承担的责任转移给 AI 应用；
- 严格落实高风险 AI 应用的报审报批和监督管控，确保重要决策始终处于有效的人类控制和组织问责机制内。

安全可控

- 将人工智能风险纳入全面风险管理体系，定期开展对人工智能应用风险及管理措施的评估审查，建立健全 AI 应用全生命周期风险管控能力；
- 涉及个人信息时，在收集、存储、使用等各环节设置边界，建立制度规范，落实安全管控技术，尊重和保护个人隐私；
- 严格落实敏感权限申请审核，仅在符合授权要求及管控策略的前提下授权相关访问；
- 对违法、有害、欺诈、侵权、歧视性或其他不当内容实施合理的识别、提示、限制或拦截，并定期评估控制措施的准确性和有效性；
- 确保数据安全治理模型的落地执行，从业务系统、主机安全、网络防护等方面增强数据安全的管控技术。

公平公正

- 促进协调发展，推动各行业领域转型升级，使用 AI 赋能多行业生态圈；
- 依法开展公平竞争，防止利用数据、算法或平台优势不当排除、限制竞争，并在安全、合规的前提下推动行业合作和技术普惠；
- 通过持续提高技术水平、改善管理方式，在数据获取、算法设计、技术开发、产品研发和应用过程中消除偏见和歧视。

公开透明

- 对 AI 生成内容及结果加以充分标识，明示自动决策的内容；
- 建立完善的自动决策客户异议响应机制，为客户提供清晰明确的申诉方式，并持续监测决策应用的运行情况，不断优化 AI 决策合理底线；
- 在不损害个人隐私、商业秘密、知识产权和系统安全的前提下，按照风险程度披露 AI 系统的数据来源类别、主要能力、适用边界、已知局限及治理措施；
- 确保高风险 AI 应用可追溯、可验证并接受授权范围内的内部审计、独立验证和监管检查；
- 持续监测模型性能、数据漂移、公平性、安全性和业务适用性；发生重大变更或风险水平变化时，重新开展评估和审批。

负责任 AI 治理目标

平安从数据使用、算法研究、行业应用及低碳运营四方面制定了负责任 AI 治理目标：

数据使用

- **隐私数据充分保护：**遵循最小必要、授权同意的原则，确保个人信息的处理不超出收集个人信息时所获得授权的范围，严格控制使用场景，遵守集团数据合规与信息安全需求。
- **个人敏感信息审慎处理：**对个人敏感信息的收集处理确保经过主体明确同意，对其进行加密储存或采取更为严格的访问控制等安全保护措施。

算法研究

- **技术透明：**建立与风险等级相匹配的模型文档体系，确保 AI 应用的用途、数据、局限性和运行逻辑对内部审计和监管机构具有可审计性。
- **算法可靠：**在预定用途和合理可预见条件下，对模型准确性、稳定性、鲁棒性进行测试；重点防范模型幻觉、提示词注入、数据投毒、对抗攻击及深度伪造等新型 AI 安全风险。
- **决策可解释：**对影响客户重大权益的 AI 应用，提供与业务场景相适应的决策解释，确保重要自动化决策结果可追溯。
- **运行结果可验证：**在一定条件下可以复现算法运行产生的结果。

行业应用

- **服务人类：**应用 AI 技术的目的不应违背人类伦理道德的基本方向，在使用过程中不作恶。
- **公平无偏：**使用完备且相对中立客观的数据训练模型，并将公平性与偏见相关需求纳入模型能力测试体系，力保 AI 算法及应用不具有某些偏见或歧视。

低碳运营

将低生态足迹管理纳入 AI 相关基础设施全生命周期管理体系，在建设和运维过程中采取绿色低碳措施，以降低对生态环境的影响。

该政策由平安集团解释和修订，公司会适时根据国家政策、监管要求和行业发展进行更新。