

# AI 治理模型框架

IMDA / PDPC

2019 年 1 月

中文翻译版

# AI 治理模型框架 (Model AI Governance Framework)

---

## 概述

2019 年 1 月，新加坡在瑞士达沃斯世界经济论坛上发布了《AI 治理模型框架》第二版，这是亚洲首个全面的 AI 治理框架，由新加坡资讯通信媒体发展局（IMDA）和个人数据保护委员会（Personal Data Protection Commission, PDPC）联合制定。

## 四大核心原则

### 1. 内部治理结构与措施 (Internal Governance Structures and Measures)

- 建立明确的 AI 治理组织架构
- 设置 AI 伦理委员会或指定负责人
- 制定 AI 使用政策和操作指南
- 建立风险管理与合规流程

### 2. 决策中的人类参与 (Human Involvement in AI-Augmented Decision-Making)

- 确定 AI 决策的不同风险等级
- 高风险决策需要人类审核和批准
- 建立人类否决 AI 决策的机制
- 确保人类对关键决策的最终控制权

### 3. 运营管理 (Operations Management)

- AI 系统开发和部署的最佳实践
- 数据管理和质量保证
- 模型监控和持续优化

- 安全性测试和漏洞管理

## 4. 利益相关方互动与沟通（Stakeholder Interaction and Communication）

- 与用户和公众进行透明沟通
- 提供 AI 系统工作原理的说明
- 建立反馈和投诉处理渠道
- 定期发布 AI 治理报告

## 实施指南

框架提供了详细的实施指南，包括：  
- 组织 AI 治理成熟度自评工具  
- 具体行业的实施案例  
- 常见问题和最佳实践

## 国际影响

- 被经济合作与发展组织（Organisation for Economic Co-operation and Development, OECD）引用为最佳实践
- 影响了多个国家和地区的 AI 治理政策
- 为后续的 AI Verify 测试框架奠定了理论基础

## 与其他框架的关系

AI 治理模型框架是新加坡 AI 治理体系的基石：  
- **AI Verify** (2022) 将其原则转化为可测试工具  
- **生成式 AI 治理框架** (2024) 在其基础上应对新技术挑战  
- **Agentic AI 治理框架** (2026) 进一步扩展至自主 AI

## 意义

作为亚洲首个 AI 治理框架，该文件确立了“以人为本、透明可解释、公平负责”的治理理念，为全球 AI 治理话语体系贡献了亚洲视角。其实用性导向和多利益相关方参与的方法论，至今仍是新加坡 AI 治理政策的核心特征。