首尔 AI 安全峰会承诺

外交部 (MFA)

2024 年 5 月

中文翻译版

首尔 AI 安全峰会承诺（Seoul AI Safety Commitment）

概述

2024 年 5 月 21-22 日，第二届 AI 安全峰会在韩国首尔举行。新加坡总理黄循财（Prime Minister Lawrence Wong）应韩国总统尹锡悦和英国首相苏纳克的邀请，参加了峰会的领导人虚拟会议。新加坡签署了《首尔 AI 安全承诺》。

峰会背景

首尔 AI 安全峰会由韩国和英国联合举办，是 2023 年 11 月 Bletchley Park AI 安全峰会的延续。峰会在布莱切利宣言（Bletchley Declaration）的基础上进一步深化承诺。

三大优先议题

1. 安全（Safety）

- 推动前沿 AI 安全评估标准的制定
- 建立 AI 系统安全测试的国际协作机制
- 分享 AI 安全测试方法论
- 加强对前沿 AI 模型潜在风险的监测

2. 创新（Innovation）

- 促进负责任的 AI 创新
- 在安全与创新之间取得平衡
- 支持有益于全人类的 AI 发展
- 推动 AI 研究和应用的国际合作

3. 包容性（Inclusivity）

- 确保 AI 发展惠及所有国家和群体
- 缩小 AI 技术的全球差距
- 支持发展中国家的 AI 能力建设
- 推动 AI 治理的多边参与

新加坡的参与

高级代表出席

- 总理黄循财参加虚拟领导人会议
- 通信及新闻部兼卫生部高级政务部长普杰立医生（Dr Janil Puthucheary）亲赴首尔出席 AISS 和 AI 全球论坛

持续国际参与

- 继 Bletchley Park 峰会后连续参与第二届峰会
- 巩固新加坡在全球 AI 治理中的积极参与者角色
- 推动将发展中国家纳入全球 AI 治理对话

与 Bletchley Declaration 的关系

首尔峰会在布莱切利宣言基础上进一步推进： - 从风险识别走向具体行动 - 扩展议题范围至创新和包容性 - 深化 AI 安全研究所之间的国际协作 - 推动更具操作性的国际合作机制

意义

新加坡连续参与两届全球 AI 安全峰会，展示了其在 AI 治理国际合作中的积极立场。作为一个小型开放经济体，新加坡在全球 AI 治理话语中发挥着超越其体量的影响力，这与其在贸易、金融等领域的国际参与策略一脉相承。