负责任人工智能治理政策

1. 目的

USI（以下称「本公司」）致力于以负责任、安全、透明及符合伦理的方式，在其营运、产品、服务及支持性业务流程中开发、取得、部署及使用人工智能（Artificial Intelligence，AI）。

本政策建立本公司对 AI 的高层级治理要求，以确保创新能在适当的风险管理、人员问责、信息安全、隐私保护、法律与伦理合规，以及持续监督的基础下推行。其目的在于促进值得信赖的 AI 成果，降低误用或未受控部署的可能性，并为内部治理、稽核审查及利害关系人信任提供一致性的依据。

2. 适用范围

本政策适用于所有董事、主管、高阶主管、员工、承包商、临时人员及其他经授权代表本公司使用、管理、监督、开发、采购、设定、整合或支持 AI 的人员。

本政策涵盖：

- 内部自行开发之 AI
- 外部采购之 AI 服务
- 通用型 AI 工具
- 代理型 AI（Agent-based AI）能力
- 企业系统内嵌之 AI 功能

本政策适用于 AI 全生命周期，包括：

- 评估（Evaluation）
- 设计（Design）
- 采购（Acquisition）

- 导入（Implementation）
- 营运（Operation）
- 监控（Monitoring）
- 审查（Review）
- 退役（Retirement）

3. 治理与监督

AI 治理委员会（AI Governance Council）应作为公司 AI 活动之集中治理、监督与跨部门协调机制。其职责包括：

- 建立 AI 治理要求
- 审查重大 AI 风险
- 推动控制措施一致性落实
- 监督重大 AI 相关议题，包括高风险 AI 使用案例、事件及例外情况

AI 治理委员会应由相关控制与支持职能单位提供支持，包括治理、风险与合规（GRC）、资安架构、隐私法务、知识产权法务、企业架构，以及其他适当的业务或技术利害关系人。上述单位应提供专业意见，于必要时提出质疑与执行审查，并协助各自责任范围内治理要求之落实。

重大 AI 风险、重大控制缺失、重大事件或高影响力 AI 使用案例，应依公司既有治理与报告机制向高阶管理阶层提报，包括信息安全委员会（ISSC），必要时亦应提交董事会层级监督。本政策应依公司企业层级政策之正式治理程序，经董事会层级核准与支持。

USI 实施方式

USI 于 2026 年 6 月成立 AI 治理委员会，作为跨职能 AI 治理与监督机制。该委员会向信息安全委员会（ISSC）报告，并每月定期召开会议，以检视 USI 内部 AI 使用的最新状况、追踪重大治理发展，并审查需要额外监督、质疑或决策之高风险 AI 使用案例。

透过 ISSC，AI 治理委员会亦可定期向董事会层级报告 AI 相关治理与管理事项之状况。相关会议治理纪录及审查结果，均应依公司治理程序保存。

4. 负责任 AI 原则

4.1 可靠性与安全性（Reliability and Safety）

AI 系统于使用或部署前，应依其预期用途与风险程度进行适当程度的验证与测试。

公司应采取合理措施，持续监控 AI 的效能、可靠性及输出质量，包括识别异常行为、效能衰退、误用或模型漂移（Model Drift）。如发现重大效能疑虑，应视情况采取审查、重新校准、再训练或退役等措施。

如发现重大可靠性或安全性疑虑，公司应视情况及时采取更正、限制或停止使用等措施。

USI 实施方式

USI 对导入之 AI 解决方案建立审查流程，依使用案例之性质与风险，于导入前或导入期间提交 AI 治理委员会审查。AI 治理委员会的跨职能组成，旨在确保审查涵盖可靠性、安全性、法规遵循、信息安全、隐私、架构及法律等相关专业领域。

USI 亦运用 Microsoft Purview 等支持工具，强化第三方 AI 服务的合规监督，并使用监控能力持续追踪 AI 使用及 AI 输出之稳定性与可靠性。相关审查、监控活动及重大发现之纪录，均应依公司治理程序保存。

4.2 信息安全（Security）

AI 系统及相关数据应采用适当信息安全控制措施保护，包括存取管理、日志记录、监控、安全组态设定，以及适用时的加密措施。

安全评估与测试活动应依 AI 系统的风险、关键性及暴露程度执行，并于相关情况下考虑 AI 特有威胁。相关活动可依使用案例风险，包括弱点评估、渗透测试或等效安全测试，以及正式上线前

与重大变更后之安全组态审查。

AI 相关安全事件与弱点，应透过公司既有的资安事件管理及韧性程序处理。

USI 实施方式

USI 透过治理审查、技术安全控制与持续监控的组合落实本要求。针对经核准的 AI 解决方案，AI 治理委员会及相关支持职能在执行审查时，会将安全要求纳入考虑，并特别关注访问控制、安全组态、数据保护及 AI 特有安全风险。

在适用情况下，USI 对 AI 相关环境采用身分识别与存取管理、最小权限存取、日志记录与监控，以及其他既有信息安全控制。对于第三方或外部提供的 AI 服务，则使用支持工具与既有安全治理机制，强化对安全及合规风险之监督。

此外，AI 使用相关的安全事件、异常活动或已识别弱点，均透过公司既有的安全监控、事件管理与改善程序处理。相关审查、监控活动及重大安全发现之纪录，均应依公司治理程序保存。

4.3 隐私保护（Privacy）

AI 使用应符合公司的数据分类、数据最小化、目的限制、保存期限及安全处理要求。

涉及个人资料、敏感数据或较高隐私影响之使用案例，应接受适当的隐私审查与治理。

隐私考虑应整合至 AI 全生命周期，以支持合法、适度且透明的信息处理。

USI 实施方式

USI 透过数据最小化检查、限制于外部 AI 环境使用机密或个人资料、在适当情况下进行资料屏蔽或去识别化、审查个人资料处理的合法依据与目的、执行保存及删除控制，以及针对较高风险使用案例咨询指定之隐私或数据保护职能，落实本要求。

USI 亦维持集团层级数据保护长（DPO）监督，以及一套隐私保护机制，以支持符合欧盟、亚洲及北美洲适用的隐私要求，包括于相关情况下执行数据保护影响评估（DPIA）等隐私审查与评估程序。

此外，公司每年提供隐私保护教育训练，协助员工了解个人资料，以及在业务活动中，包括使用 AI 时，应如何保护个人资料。

4.4 公平性（Fairness）

在与使用案例相关的情况下，尤其针对较高风险 AI，公司应考虑是否可能产生偏见、不公平待遇或重大差异化结果。

公司应每半年维持合理的监控、审查、质疑或公平性稽核机制，以识别偏见、不公平待遇或不适当输出的迹象。针对较高风险使用案例，应依适用治理程序，于部署前、重大变更后及后续定期执行此类审查。

如发现重大公平性疑虑，应依其情境与风险，考虑采取改善或额外控制措施。

USI 实施方式

USI 透过 AI 治理委员会对导入之 AI 解决方案进行跨职能审查，并特别关注预期用途、受影响的利害关系人、决策情境，以及使用案例是否可能造成偏见、不公平待遇或重大差异化结果。

AI 治理委员会的组成，有助于确保治理、隐私、法务、信息安全、架构及业务等相关观点均纳入审查。对于较高风险或敏感使用案例，可能要求于允许使用或继续使用前，执行额外审查、质疑或人工监督。

如发现公平性疑虑、偏见指标或重大审查发现，USI 可依情况要求额外防护措施、使用限制、改善行动或进一步评估。此外，USI 透过使用者训练强调，使用者必须主动检查 AI 产出，并对其使用、依赖或传达的内容负责。相关审查、公平性评估、训练重点及重大决策纪录，均应依公司治理程序保存。

4.5 人工监督（Human Oversight）

人工监督程度应与使用案例的重要性与风险相符，尤其是 AI 可能重大影响决策、行动或外部结果时。人工监督应依风险程度与使用案例，透过适当控制模式确保，包括 Human-in-Command

（人员保有最终决策权）、Human-on-the-Loop（人员持续监督）或 Human-in-the-Loop（须经人工核准）。

应记录并保存必要的人工审查、监督行动、核准、介入、推翻决定、例外处理及重大决策，并符合公司的治理与纪录保存要求。此类纪录应能证明具实质意义的人员参与，并支持问责性、可追溯性、可稽核性，以及符合适用之法律、法规与治理义务。

除非依适用治理及法律要求明确授权，否则不得将 AI 视为敏感、高影响或其他受限制事项中的唯一决策者。

USI 实施方式

USI 透过治理机制，要求在 AI 的审查、核准及使用过程中保留具实质意义的人员参与，特别是针对较高风险或可能产生重大影响的使用案例。AI 治理委员会以及相关业务或流程负责人，协助确保 AI 不被视为敏感事项中的自主决策者，并确保适当的人类判断持续纳入决策流程。

实务上，包括针对较高风险使用案例设置核准检查点、AI 产出在用于重大业务行动或对外沟通前进行人工审查，以及明确指定人员具有在必要时质疑、推翻、限制、暂停或停止 AI 支持活动之权限。

USI 亦透过训练与治理要求强调，使用者必须主动审查 AI 产出、运用专业判断，并对其所依赖、传达或执行的决策或内容负责。必要的人工审查、监督行动、例外处理及重大决策纪录，均应依公司治理程序保存。

4.6 透明性（Transparency）

公司应在符合情境、法律、合约义务或利害关系人期待的情况下，提升 AI 在相关业务流程、服务或产出中的透明度。

针对相关使用案例，公司应寻求确保适当使用者或审查人员，能以足以支持负责任使用、审查及有效人工监督的程度，理解下列信息：

- AI 系统的预期用途

- 主要能力与限制
- 所使用的数据源或数据类型（于适当且允许情况下）
- 相关风险与不确定性
- 在合理可行范围内，AI 支持产出所依据的基础或理由

当法律、合约、治理要求规定，或未揭露可能合理造成内容系由 AI 或由人类作者或审查者产生之误解时，应对 AI 生成或实质由 AI 协助产生的内容进行适当标示、揭露或其他方式之识别。

USI 实施方式

USI 透过治理及沟通措施，使 AI 在相关业务活动中的角色可被看见、理解并获得适当揭露。对已导入之 AI 解决方案，其预期用途、AI 在流程中的角色、关键限制及重大假设，均于治理流程中接受审查与记录，使相关用户及审查者了解 AI 如何被使用及其适用限制。

当 AI 产出或实质由 AI 协助产出的内容用于商务沟通、交付成果或其他相关情境时，若未揭露可能造成对其来源、性质或人员参与程度之误解，USI 要求进行适当标示、揭露或说明。

USI 亦透过训练及治理要求强调，使用者应了解 AI 产出内容的限制、于使用前验证产出，并避免以误导或不透明的方式呈现 AI 支持结果。相关揭露、治理审查及重大决策纪录，均应依公司治理程序保存。

4.7 问责性（Accountability）

AI 相关治理、营运、审查、提报及改善责任，应透过公司的治理架构及职责分工明确指定。依适用情况，相关责任应涵盖因 AI 使用所产生的错误、营运损害、合规违规、隐私影响、资安事件，以及法律或合约后果。

涉及 AI 的重大事件、控制失效或治理疑虑，应透过适当的管理程序进行调查。

AI 相关审查、评估、事件或稽核所产生的矫正行动与改善措施，应依适当方式追踪至结案。

USI 实施方式

USI 透过为每个经核准的 AI 使用案例指定并记录负责人、保存核准及审查纪录、记录例外与风险接受、定期向管理阶层报告重大议题，以及保留足以支持内部审查、外部确信或适用稽核的证据，落实本要求。

AI 治理委员会负责监督治理一致性及重大例外事项；业务或流程负责人对其责任范围内经核准 AI 的使用负责；信息安全、隐私、法务、IT 与 GRC 等支持职能，则依各自职责负责审查、调查或控制措施。

此外，相关 AI 系统应保有可记录关键操作及输出或结果的日志，使重大活动能依公司治理及安全程序，在必要时被追溯、审查及调查。

4.8 环境责任（Environmental Responsibility）

在相关且合理可行之情况下，公司应于 AI 能力的设计、选择及运作过程中，考虑资源效率、适度性及更广泛的环境影响。

对于部署于公司自有或外部提供之数据中心的 AI 系统，公司亦必须考虑能源效率、碳足迹及永续基础设施作法，包括运算资源优化、冷却效率，以及可行时使用再生能源。

USI 实施方式

USI 在选择或营运重要 AI 能力时，会考虑能源消耗、资源使用效率，以及相关情况下更广泛的环境因素。对于内部托管的 AI 系统，USI 将环境考虑纳入数据中心营运，包括监控能源使用、优化运算工作负载，以及在适用情况下与企业永续目标对齐。

若使用第三方或外部提供的 AI，例如 Microsoft Copilot 服务，USI 亦将环境与永续相关因素纳入供货商选择及尽职调查。这包括评估供货商公开揭露的永续作法，例如 Microsoft 所公布之提升数据中心能源效率、增加再生能源使用，以及优化 AI 工作负载效率的方法。

Microsoft 的永续措施包括持续投资于节能数据中心营运、使用无碳电力，以及优化 AI 能源消耗与基础设施效率的创新，这些措施支持了 USI 对环境负责任 AI 供货商之评估。请参考微软官网说明 [https://learn.microsoft.com/en-us/industry/sustainability/advance-sustainability-ai]

此外，公司亦将供货商在信息安全、机密性、透明度、授权，以及限制公司数据二次利用等方面的控制措施，纳入供货商治理与审查，以确保符合 USI 的永续、风险管理及负责任 AI 目标。

5. 负责任使用（Responsible Use）

仅限经公司核准、授权或以其他方式允许之 AI 系统、模型、工具、服务或启用功能，可用于公司业务。

本公司不会开发或部署任何属于《欧盟人工智能法案（EU AI Act）》所定义之禁止类别 AI 系统，包括操控型系统、利用弱势族群、社会评分或未经授权之生物辨识监控。

AI 仅能用于合法业务目的，且应符合使用者被赋予之职责范围及授权访问权限。

使用者仍须运用专业判断，并应依情境以适当程度验证 AI 产出，方可将其用于业务执行、内部决策或对外沟通。

此外，AI 系统应受明确使用边界管理，包括授权范围、预期用途及能力限制，以确保仅在核准情境内使用，且不超出其预期应用。

USI 实施方式

在 USI，本要求透过核准 AI 清单（Approved AI List）及相关治理机制来实施，要求员工仅能在已定义的使用情境及职务责任范围内，为合法的业务目的使用经授权的 AI 解决方案。

所有新导入的 AI 解决方案均须依公司的 AI 治理程序，在导入前或导入过程中接受审查。审查过程中会明确定义并记录各使用案例的预期用途、授权使用范围及主要限制条件，藉此建立清楚的 AI 使用界线。

经核准的使用情境可确保使用者了解哪些类型的业务活动允许使用 AI，以及 AI 系统不得应用于超出其预期情境的范围。

这些使用界线亦透过治理要求及使用者教育训练进一步强化。相关要求强调，使用者必须遵循适用的输入与输出处理规范、避免使用未经核准的插件或外部 AI 工具，且不得将 AI 用途扩展至核准目的之外，以防止非预期用途或任务范围持续扩张（Mission Creep）。

此外，USI 亦采用核准机制、监控及日志记录等控制措施，以确保 AI 的使用始终维持在所定义的边界内。所有使用者在依赖 AI 产出前，均须主动验证 AI 生成内容，并对其利用 AI 所产生的内容、决策或沟通结果负责。

与 AI 使用相关之核准纪录、已定义的使用界线、允许使用条件及重大例外事项，均应依公司治理程序妥善保存。

6. 禁止用途（Prohibited Uses）

禁止以 AI 的方式造成公司、客户、员工或其他受保护信息遭未经授权揭露、传输、保存或暴露。 除非已依适用要求完成特别审查、核准及治理，否则禁止于敏感、受限制或高影响领域中使用完全自动化的 AI 决策。

禁止未经授权使用 AI 执行跨系统操作、发起交易、进行大量处理，或绕过既有控制措施、核准程序及职责分离机制。

不得以 AI 从事非法、欺骗性、不道德、歧视性或其他受禁止的用途，包括不符合适用法律、合约义务或公司治理要求之使用。受禁止作法包括在适用法律或治理要求不允许的情况下，使用旨在实质扭曲人类行为的操控型 AI、非法社会评分，以及未经授权的实时生物辨识、监控或类似监视活动。

USI 实施方式

USI 透过治理限制、技术控制及针对不被允许之 AI 使用的管理升级机制，落实本要求。在适用情况下，未经核准之外部 AI 服务、插件或整合功能，将透过既有的访问控制、租户管理或网络管理措施加以限制；疑似违规或高风险例外情况，则须依公司治理程序进行管理审查与升级处理。

针对敏感或高影响 AI 使用案例，USI 实施强化之存取限制，包括角色型访问控制（RBAC）、最小权限原则，以及适用时的事前核准或多层级授权。此类使用案例亦可能接受额外监控、日志记录及定期审查。

USI 亦透过治理要求与使用者教育训练强调，员工不得绕过核准控制措施、不得将 AI 用于未授权之业务目的，亦不得以可能造成重大资安、合规、隐私或伦理风险的方式使用 AI。重大违规、例外情况及相关管理措施之纪录，均应依公司治理程序保存。

7. 风险管理（Risk Management）

公司应采用风险导向方法执行 AI 治理，并考虑使用案例之预期用途、涉及数据之性质、自主化程度、影响范围，以及对个人、营运或法规遵循义务的潜在影响。

当发生重大变更时，AI 使用案例应重新评估，包括用途、数据使用、整合范围、使用者族群、决策影响或外部法律与法规条件之变更。

较高风险的 AI 使用案例应接受强化审查、治理关注及控制要求，包括适当时采取更严格的监督与升级处理。

已核准之 AI 使用案例应接受定期监控或审查，以确认风险假设、治理条件及控制有效性随时间推移仍维持适当。

USI 实施方式

USI 透过治理程序，要求导入之 AI 解决方案及使用案例，依其性质、预期用途、数据敏感度、自主化程度，以及可能造成的业务或合规影响，进行审查、分类与追踪。AI 治理委员会负责审查重大 AI 使用案例；较高风险情境会依情况于导入前或导入过程中接受强化审查与额外治理关注。

如发生用途、数据使用、整合范围或决策影响等重大变更，相关使用案例应透过公司 AI 治理程序重新评估。USI 亦定期审查已核准 AI 使用案例，以持续评估演变中的风险、控制有效性及治理条件。使用案例审查、风险分类、重新评估及重大治理决策之纪录，均应依公司治理程序保存。

8. 数据保护（Data Protection）

使用于 AI 之数据应依公司信息分类及保护要求处理，并仅限于与相关目的相符、必要且经授权之范围。

敏感、机密、个人或其他受限制数据，不得于外部或未经核准之 AI 环境中使用，除非已依适用之治理、信息安全、隐私及法律要求明确允许。

AI 相关数据处理应于适当情况下，与公司的监控、日志记录、保存及事件响应能力整合，以支持问责性与可稽核性。

USI 实施方式

USI 透过数据治理措施，要求用于 AI 的信息遵循公司数据分类与保护要求，包括限制在未经核准或外部 AI 环境中使用机密、个人或其他受限制数据。企业数据源如需连接 AI 服务，应事先完成适当核准及治理审查；相关数据存取、传输或使用活动，则应于适用情况下接受日志记录与监控。

USI 亦将保存、删除及访问控制措施纳入整体信息安全与数据保护架构，并透过供货商或合约控制，处理第三方 AI 供货商对公司资料的机密性、安全性及二次利用限制。相关核准、监控活动及重大数据保护发现之纪录，均应依公司治理程序保存。

9. 培训和意识（Training and Awareness）

公司应维持教育训练与意识提升活动，以支持负责任、安全及合规的 AI 使用。相关训练应于新进人员报到时或适用情况下于取得权限前提供，后续定期更新，并于治理要求、技术或风险条件发生重大变更后更新。

训练内容可包括 AI 治理原则、信息安全、隐私保护、输出验证、适当使用限制，以及对影子 AI（Shadow AI）与误用风险的认知。

教育训练与认知提升措施可依公司内不同角色、责任及 AI 使用性质调整。

USI 实施方式

USI 透过员工到职期间及每年提供 AI 相关训练，并搭配针对相关使用者的角色别训练、可接受提示词与数据处理指引、常见 AI 失效模式认知，以及对使用较高风险或业务关键 AI 能力团队的重点强化，落实本要求。

USI 亦定期透过电子报及类似沟通管道，向员工发布额外的 AI 相关教育训练与认知内容，以强化重要治理、信息安全、隐私及负责任使用要求。训练对象可包括一般用户、开发人员、系统管理员、核准人员、审查人员，或其他负有较高 AI 相关责任的角色。

10. 通报机制（Reporting Mechanism）

AI 相关疑虑，包括疑似误用、控制弱点、重大错误、资安事件、隐私疑虑、伦理问题，或对具有重大影响之 AI 支持结果的申诉，均可透过公司既有的通报、伦理、资安、隐私或法规遵循管道提出。如适用程序设有申诉或救济受理机制，公司应提供合理可及之管道，并依相关治理程序进行审查与响应。

已通报事项应透过公司既有治理与调查程序进行审查及处理，并视情况采取矫正或预防措施。

USI 实施方式

USI 透过既有的资安、隐私、法规遵循、伦理或其他治理疑虑之通报与升级管道落实本要求，这些管道亦可用于通报 AI 相关问题。当 AI 服务提供给外部客户及第三方时，公司亦应提供 AI 专属之外部通报管道。

已通报之疑虑，包括疑似误用、异常输出、控制弱点或重大事件，应依其性质与严重程度进行审查及分流，并视情况由相关职能单位参与。必要时，重大 AI 相关疑虑可透过公司治理程序升级处理，包括提交 AI 治理委员会审查。USI 亦依公司治理程序保存重大通报、调查、后续行动及改善结果之纪录。

11. 治理指标（Governance Metrics）

公司得监控治理指标，以支持 AI 治理计划之监督与持续改善。指针范例可包括经核准之 AI 系统或使用案例的数量与类型、已通报事件或疑虑，以及改善活动之状态。

公司亦得追踪训练涵盖率、治理审查活动、政策例外、与重大 AI 使用相关的环境或资源效率指标，或其他有助于呈现风险可视性及计划成熟度的指标。

相关稽核观察事项、审查结果或重复发生的问题，可作为改善优先级及管理报告的依据。

USI 实施方式

USI 透过定期治理报告及指标审查，支持对 AI 相关管理事项之持续监督。AI 治理委员会每月审查活动之产出，可用于治理指标、趋势分析、后续行动及管理报告。

相关指标可包括已核准 AI 使用案例之状态、高风险案例审查、已通报事件或疑虑、训练完成情况、改善进度，以及其他与 AI 治理有效性及成熟度相关的指标。透过信息安全委员会（ISSC），相关治理报告亦可支持董事会层级定期掌握 AI 相关管理状况。相关指针、报告及重大后续行动之纪录，均应依公司治理程序保存。

12. 政策管理（Policy Management）

本政策应定期审查并视需要更新，以反映业务使用、技术、风险条件及适用法律或法规要求之变化。

AI 治理委员会负责维护本政策，并透过公司治理架构监督其实施。

13. 违规与例外（Violations and Exceptions）

违反本政策者，得依适用之公司规则、合约安排及当地法律要求，采取纪律处分或其他管理措施。

本政策之任何例外，均应依公司既有治理要求，完成适当之核准及风险接受程序。

附录（Appendix）

本附录以示例方式汇整禁止及高风险 AI 作法，包括操控、非法监控、歧视性使用、未经授权之自主行动，以及就业、金融、生产控制等敏感领域中可能需要强化治理审查的 AI 使用案例。

表 1：非法（禁止）AI 作法清单

下列 AI 作法违反适用法律与法规，或严重违反基本伦理原则，构成公司治理红线，任何情况下均严格禁止开发、采购或使用。

类别	治理红线（原则导向）	代表性范例（非完整列举）
操控与欺骗	在个人不知情或违反其真实意愿的情况下，使用 AI 操控其决策、行为或认知。	• 以潜意识或心理操控影响个人选择 • 使用深伪内容冒充他人或误导大众
利用弱势族群	针对弱势族群，例如儿童、高龄者、身心障碍者或经济弱势者，利用其特定弱点。	• 以 AI 诱导高龄者进行不必要的购买或财务决策 • 搜集或滥用未成年人的个人资料
非法监控与隐私侵害	在未取得法律授权或明确同意的情况下，部署 AI 进行大规模监控、追踪或生物辨识。	• 未经授权的人脸辨识或生物特征监控 • 使用 AI 推论员工情绪、心理状态或行为倾向
社会评分与歧视	使用 AI 以具有社会歧视性的方式，对个人或群体进行分类、评分或差别待遇。	• 以社会信用评分影响服务或机会之取得 • 依种族、性别、年龄等受保护特征给予歧视性待遇

公共安全与非法用途	将 AI 用于非法活动，或用于危害公共安全或国家安全之目的。	• 协助网络攻击、数据窃取或诈欺 • 产生伪造证据或伪造法律、商业文件

表 2：高风险 AI 作法清单

高风险领域	风险说明（原则导向）	代表性范例（非完整列举）
人力资源与人员管理	影响就业机会、职涯发展或劳资关系的 AI 系统。	• 招募筛选、履历评估、面试评分 • 绩效评估、升迁、薪酬或终止聘用建议
金融、信用与重大交易	对个人或组织产生重大经济影响的 AI 系统。	• 信用核准、保险定价或理赔决策 • 大额资金移转、投资或资产处分
关键基础设施与生产控制	影响安全、营运持续性或系统稳定性的 AI 系统。	• 生产线控制、设备排程或质量判定 • 能源管理、供应链或物流控制
具有法律效果的自动化决策	产生法律效果或类似重大影响的自动化决策。	• 完全自动化的服务拒绝、处罚或限制 • 合约生成、法律或合规判定
代理型 AI（自主型 / Agent-based AI）	具备自主决策及跨系统执行能力的 AI 系统。	• AI 代理使用员工凭证跨企业系统执行任务 • AI 系统自主规划及执行业务流程
敏感社会领域	应用于高度监管或公共利益领域的 AI 系统。	• 教育评量、考试评分或学术分发 • 医疗诊断、治疗建议或健康风险评估