

亚洲

新加坡 AI 虚假信息治理政策表

序号	名称	主要 条目	主要内容	实施过程
1	《生成式人工智能治理 模型框架》(2024.1)	数字溯源	政府、行业和社会认识到需采用数字水印、密码来源证明等技术解决方案，以跟上人	该框架旨在为人工智能在新加坡的发展和应用提供指导，其中包括了人工智能伦理、风险管理、隐私保护、安全性和透明度等方面的指导原则。此新加坡在道德标准方面比较注重社会责任和透明度。新加坡政府推出了一系列的伦理准则，包括机器人准则、数据伦理准则等，以确保人工智能的开发和使用符合道德和社会责任要求。此外，新加坡还鼓励人工智能的开发和应用具有透明度，以便公众了解人工智能系统的运作和决策过程
		开放核证 / 多方合作	•与内容生命周期中的关键方合作，支持嵌入和显示数字水印和来源细节。例如，由于出版商对最终用户验证内容真实性至关重要，和出版商合作十分关键	
		类型标准化	标准化要标记的编辑类型，以帮助最终用户区分非人工智能和人工智能生成的内容。	
		用户要求	最终用户需更好地理解内容来源，学会使用工具验证真实性，关键利益相关者应与政策制定者合作提高用户认识，简化显示的来源细节以方便用户理解	

日本 AI 虚假信息治理政策表

序号	名称	主要 条目	主要内容	实施过程
1	《人工智能运营指南（草案）》 (2024.1)	开放核证 / 弱势群体关怀	要求使社会弱势群体更易使用人工智能，避免造成“信息弱势群体”或“技术弱势群体”	<p>政府过去一直避免用硬法来监管人工智能，2023 年 5 月成立了人工智能战略委员会。</p> <p>到目前为止，人工智能战略委员会已经讨论了以下几点：</p> <ul style="list-style-type: none">•应对人工智能风险：作为风险应对的基本方针，人工智能开发者、提供者、用户等应自行评估风险，履行治理职能。•透明度和可靠性：人工智能开发人员和提供商必须根据现行法律和准则主动披露信息。政府还应考虑直接鼓励主要人工智能开发商和提供商的透明度和可靠性。
		数字溯源 / 信息可验证性	<p>在合理范围内记录和保存人工智能训练过程、推理过程、判断依据等日志，确保可验证性。</p> <p>向利益相关者提供关于人工智能系统和服务的多方面信息</p>	
		开放核证 / 多元主体要求	<ul style="list-style-type: none">•人工智能开发者：依法依规处理训练数据，保障数据安全；管理数据质量，记录开发信息，保障系统可追溯性；合理设置安全使用范围，正确选择预训练模型；开发后向利益相关者提供多方面系统信息，向人工智能提供者解释合规情况•人工智能提供者：保障系统性能，采用风险控制方法；确认数据公平性，评估模型偏差，促使开发者改进；引入隐私和安全机制，降低数据滥用风险；记录系统架构和数据处理过程，便于监督排查；定期验证系统使用目的，检查处理安全漏洞•人工智能企业用户：安全使用人工智能系统和服务；注意隐私保护；妥善存储利用文件，遵守服务条款，	

			规范使用行为	
--	--	--	--------	--

韩国 AI 虚假信息治理政策表

序号	名称	主要 条目	主要内容	实施过程
1	《人工智能时代个人信息安全使用指南》 2023 年 8 月	政府和私营部门共同设计监管制度	<ul style="list-style-type: none">•建立以原则为基础的纪律，以解决不确定性•为人工智能开发和服务阶段制定个人信息的处理标准•通过公共机构和私营机构的合作，为每个部门制定指导方针•加强国际合作体系	个人信息保护委员会特别指出，由于人工智能的使用引发了对侵犯个人信息可能性的担忧，考虑到以信息主体意想不到的方式处理数据的案例增多。该指南，以指导相关方处理个人信息，从而最大限度地降低侵犯隐私的风险，同时确保安全使用数据，促进人工智能的发展。

印度 AI 虚假信息治理政策表

序号	名称	主要 条目	主要内容	实施过程
1	《印度拟要求模型上线前需得到政府批准》	发布许可制度	印度政府表示，所有目前正在测试中、处于开发的测试阶段或任何形式上不可靠的人工智能（AI）模型、大型语言模型（LLMs）、使用生成式 AI 的软件或任何算法，在印度互联网上为用户部署前，必须获得“印度政府的明确许可”。 所有平台确保“它们的计算资源不会通过使用 AI、生成式 AI、LLMs 或任何其他算法而产生任何偏见、歧视或威胁选举过程的完整性”。	印度电子和信息技术部部长拉吉夫·钱德拉谢卡尔表示：“这表明了未来的监管方向。我们今天发布这个意见是要求您（AI 平台）遵守它。”

沙特 AI 虚假信息治理政策表

序号	名称	主要 条目	主要内容	实施过程
1	《生成式人工智能的指南》	提高人们对于人工智能技术重要性的认识水平	《政府生成式人工智能指南》与《公众生成式人工智能指南》，为公务人员及公众提供了生成式人工智能相关指引。两份指引均介绍了生成式人工智能的优势以及可能遇到的风险，提出了合理使用生成式人工智能的基本原则以及推荐实践。	/

中国（大陆）AI 虚假信息治理政策表

序号	名称	主要 条目	主要内容	实施过程
1	《互联网信息服务算法推荐管理规定》2021	算法 推荐 技术	要求提供算法推荐服务、尊重社会公德和伦理，遵守商业道德和职业道德，遵循公正公平、公开透明、科学合理和诚实信用原则	在监管的方针上，中国对于生成式人工智能采取分类分级监管的原则，未来有关主管部门（网信办）将会发布具体的分类分级规则和指引。 在目前负责算法备案、评估与监管的机构主要是网信办（中央网信办网络管理技术局）。与此同时，电信主管部门、公安部门也会依据职责负责监督管理工作。
2	《互联网信息服务深度合成管理规定》2022	深度 合成 技术	从技术角度对人工智能进行了一定的约束	
3	《生成式人工智能服务管理暂行办法》2023	生成 式人 工智 能技 术	《办法》仅适用于向中国境内公众提供服务的人工智能产品。但需要注意的是，“公众”一词目前尚未有明确的定义和解释，因此不能排除to B的服务将完全豁免于该办法的适用。并且来源于中国境外向境内提供的生成式人工智能服务也会受到该办法的管辖。	
4	《人工智能安全治理框架》(2024.9)	安全 风险 分类	<ul style="list-style-type: none">•人工智能内生安全风险：模型算法安全风险、数据安全风险、系统安全风险•人工智能应用安全风险：网络域安全风险、现实域安全风险、认知域安全风险、伦理域安全风险	

		技术应对措施	<ul style="list-style-type: none">•针对人工智能内生安全风险：提高可解释性和可预测性；遵循数据安全规则，加强知识产权保护；提高系统透明性，加强平台风险防控。•针对人工智能应用安全风险：建立安全防护和数据护栏机制；设置服务边界；加强检测技术研发；建立应急管控措施	
		综合治理措施	<ul style="list-style-type: none">•对人工智能系统分类分级，加强最终用途管理和登记备案•建立可追溯管理制度，对公众服务系统标识管理•完善数据安全和个人信息保护规范•构建负责任研发应用体系，完善伦理审查制度•强化供应链安全保障，共享成果，开源技术•推进可解释性研究，提高可解释性和可预测性•建立信息共享和应急处置机制•推动安全教育与学科同步发展，培养多领域安全人才•加强教育培训和宣传，建立行业自律和社会监督机制。•促进国际交流合作，制定国际标准	

中国香港地区 AI 虚假信息治理政策表

序号	名称	主要 条目	主要内容	实施过程
1	《香港生成式人工智能技术及应用指引》 2025.4.15	人工智能和大数据分析技术实务操作指引	核心一：问责。非常类似于数据保护中对于数据处理者的要求！ 核心二：风险为本的人为监督 核心三：全生命周期合规 核心四：与持股者的沟通及交流	涵盖生成式人工智能技术的应用范围、应用局限、潜在风险与治理原则，包括资料泄露、模型偏见和错误等技术偏见。
2	《中国香港的道德问题框架》	道德数据管理	<div>道德数据影响评估模型。这评估了数据收集、使用和披露以及数据驱动活动对所有利益相关者利益的影响。</div> <div>流程监督模型。这着眼于组织如何将组织道德价值观转化为原则和政策以及“设计道德”计划。它还考虑如何实施内部审查流程，例如进行 EDIA 和建立有效的个人问责制度。</div>	2019 年香港金管局（简称 HKMA）发布通告，鼓励各认证机构在金融科技产品和服务开发过程中积极采用上述道德问责框架机制，合理规范地收集用户个人资料。

3	《开发及使用人工智能道德标准指引》(《人工智能指引》)	道德原则	目的是协助机构在开发及使用人工智能时，明白并遵从《个人资料(私隐)条例》(第486章)《私隐条例》就保障个人资料私隐的相关规定。《人工智能指引》的内容包括数据管理价值及人工智能道德原则，并提供人工智能策略管治的实务指引，以协助机构制订合适人工智能策略及管理模式，并作出风险评估及制定相关监察安排等。	香港特区政府将保持开放的态度，紧密关注技术发展，正在与业界商量合作，在生成式 AI 大模型、开源功能方面对香港做出独特贡献。 此外，香港也在 2 月底发布的财政预算案上提出，就建立人工智能超算中心进行可行性研究，以推动本地科研和产业发展提供足够算力基础。
---	-----------------------------	------	---	--