北美洲

美国 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	《人工智能权利法案蓝图：让自动化系统为美国人民服务》（2022）	保障人民权利	旨在帮助指导自动化系统的设计、使用和部署，从而保护人工智能时代下美国公众的权利。这些原则是通过与美国公众广泛协商制定的，构建保护公民权利、公民自由和隐私的自动化系统的蓝图。	根据《人工智能权利法案蓝图》的内容，它不取代、修改或指导对任何现有法规、条例、政策或国际文书的解释。它不构成对公共或联邦机构具有约束力的指导，因此不要求遵守本文所述的原则。它也不能决定美国政府在任何国际谈判中的立场。这些原则并非旨在禁止或限制政府机构的任何合法活动，包括执法、国家安全或情报活动。本白皮书中阐述的原则的适当应用在很大程度上取决于使用自动化系统的环境。蓝图中认为，执法活动需要进行利益平衡。
		技术指南	•安全有效的系统：它应该采取积极并持续的保障措施防止公众受到伤害；避免使用不适合或任务无关的数据，包括可能造成二次伤害的重复使用；并证明系统的安全性和有效性。 •算法歧视保护：任何自动化系统都应该经过测试，以确保在出售或使用之前不包含算法歧视。 •数据隐私：应通过内置的隐私保护、数据最小化、限制并且公开透明地使用和收集数据等方式来保护公众，此外，公众还应当有权拥有明确的机制，以主动、知情和持续的方式来控制对数据（包括元数据）的访问和使用。任何收集、使用、共享或存储个人数据的自动化系统都应满足这些期望。	

2	国会众议院《深度伪造问责法》2023.9	数字溯源	•要求记录内容来源技术，以表明音频或视觉元素被修改或完全由人工智能生成 •制定处罚措施，包含刑事处罚和民事处罚	与其他会创建全新机构、繁琐的许可制度或开辟广泛新责任类别的方法不同，《算法问责法案》是针对人工智能和自动化系统已经造成的问题而提出的有针对性的应对措施。消费者和监管机构缺乏对这些“自动化关键决策流程”应用场景的洞察，这使得追究企业责任和消费者做出明智选择变得困难。美国公众和政府需要更多信息来了解人工智能的应用场景和原因，而企业则需要清晰的架构，以确保影响评估流程的有效性。
		深度伪造检测工作	•国土安全部部长应建立深度伪造特别工作组； •国土安全部部长每年提交关于特别工作组活动、技术进展、国家安全威胁新发展及应对措施的报告； •国土安全部应建立信息共享计划，并及时向新闻机构通报； •政府可将可靠的检测技术提供给美国私营部门互联网平台。	
3	《关于推进美国在人工智能领域的领导地位的备忘录》(2024.10)	高风险评估	•在本备忘录发布之日起180 天内，在私营部门合作的情况下，AISI 应在公开部发布之前对至少两个前沿人工智能模型进行自愿初步测试，以评估可能对国家安 全构成威胁的能力。•保护美国人工智能免受外国情报威胁：一加强对美国 AI 生态系统及其关键支撑行业（如半导体设计和生产）的外国情报威胁的识别和评估。二识别 AI 供应链中的关键节点，并制定最有可能被外国行为主体破坏或危及的节点和途径清单。三防范 AI 知识产权竞争风险，以确保技术优势不被侵蚀。	此次国家安全备忘录的出台标志着美国在全球人工智能领域战略部署已进入一个新阶段，在政策上明确加强运用 AI 实现国家安全目标。值得注意的是，当前美国正指导各机构获取前沿 AI 系统并投入使用，这通常涉及针对特定要求的技术和产品大量的采购工作，近期将会对全球人工智能产业链产生较大影响。

4	国防部 “媒体取证”“语义取证”项目 (2015 至今)	深度伪造检测工作	旨在通过开发能够自动检测、归因和描述伪造媒体资产特征的技术，让分析人员在检测者和操纵者之间的斗争中占据上风。	• 2015 启动媒体取证（MediFor）开发自动评估图像或视频完整性的技术，集成到 取证平台；• 2021 年启动语义取证（SemaFor）开发能自动检测、归因和描述各种类型深度伪 算法；• 2024 年发起 AI FORCE 开放社区研究工作，开发机器学习或深度学习模型检测 AI 图像。