欧洲

欧盟 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	《人工智能法》(2024.7) 包含 180 个序言、113 条条款和 13 个附件，为欧盟境内人工智能系统的开发、部署和使用建立了全面的框架。	多元主体要求	•人工智能技术提供者：要提供能以机器可读格式标记内容的技术解决方案 •人工智能内容部署者（用户）：需适当保障第三方权利和自由	（1）2024 年 11 月 2 日之前，各成员国指定人工智能监管机构； （2）2025 年 2 月 2 日开始，法案的一般规定和禁令规定将会生效，针对社会信用排名系统等具有“不可接受的风险”的人工智能的禁令清单开始适用，并将于法律颁布实施后的六个月内开始执行； （3）2025 年 5 月 2 日之前，欧盟人工智能办公室将会发布行为准则； （4）2025 年 8 月 2 日开始，通用人工智能模型（GPAI）相关规定开始生效，提供者将须履行透明度等义务； （5）2026 年 2 月 2 日之前，欧盟委员会需要提出建立后市场监测计划模板，并列出计划所列要素清单；
		数字溯源	•开发者和使用者必须明确披露内容是由人工智能生成的 •对深度伪造内容进行标记是强制性的，通过进行分类和添加水印来实现	
		风险评估	•不可接受风险。定义为对安全、生计和基本权利构成明显威胁的人工智能系统，根据法案第 5 条禁止。 •高风险。定义为可能危害安全、基本权利或导致重大不利影响的人工智能系统，受法案第 6-51 条规定的严格合规和透明度义务约束。 •有限风险(Limited Risk)。定义为需要特定透明度措施的人工智能系统。例如，聊天机器人，情绪识别系统等。法案第 52-54 条概述的透明度义务。 •最小风险(Minimal Risk)。定义为对公民权利或安全构成最小或无风险的人工智能应用。	
		问责机制	•从平台主体义务、监管机构职责、处罚措施等方面，构建了人工智能算法的外部问责机制。	

			•该法案要求平台主体遵守规定和标准，并接受监管机构的监督和管理，定期提交审查和风险评估报告。	
2	《反虚假信息行为准则》(2018年) 全球首个行业内通过自我监管来打击虚假信息的倡议	自愿签署原则	要求互联公司采取措施应对在线虚假信息: •自我监管:欧盟鼓励互联网平台通过自我监管的方式，来应对虚假信息。 •透明度:准则要求平台提高透明度，让用户了解平台如何处理虚假信息。 •合作:准则鼓励平台与事实核查机构合作，共同打击虚假信息。 •广告限制:准则也涉及对虚假信息网站的广告投放进行限制。	（1）2021年5月26日欧盟委员会发布关于强化《反虚假信息行为准则》的政策指引。 （2）2025年2月13日，欧盟委员会和欧洲数字服务委员会正式批准将《打击虚假信息自律行为准则》纳入《数字服务法》(DSA）框架。
3	《人工智能与虚假信息：国家支持的信息行动与公共领域的扭曲》 (2022)	操纵公众舆论规制	•重点讨论了生成式人工智能对舆论传播的影响，列举了国家层面传播虚假信息新的表现形式，即通过国家与媒体的“结盟”，煽动、放大、并传递虚假信息，为政客提供“客观报道”，为其网络雇佣军的更广泛协调和参与提供掩护，实现全面动员。	•降低各类行为者通信平台武器化能力； •开发识别和跟踪深度伪造的方法；充分训练信息完整性模型； •对技术平台实施共同监管；研究”协同化虚假信息“和网络骚扰运动对各类语言、训练数据和对抗性生成网络的影响等

4	《生成式人工智能与数据安全指南》(2024 年)	数据准确性原则	数据准确性原则要求数据应当准确、最新，控制者需更新或删除不准确的数据。欧盟机构在所有阶段都必要措施确保数据准确性，并对输出数据进行定期监控；在使用第三方提供的数据集或系统时，应获得确保数据准确性的合同保证和文档。	（1）指南为欧盟机构负责任地使用生成性人工智能系统提供了重要参考，指南重点在于数据最小化、数据准确性和数据安全，强调了透明和问责和在人工智能全生命周期内保护个人权利的重要性。 （2）该指南不影响《人工智能法》的规定，不具有法律效力。
		完善数据使用许可体系	生成式人工智能系统的个人数据处理涵盖系统生命周期内的所有处理活动，包括数据收集、训练、系统交互和内容生成。 欧盟机构处理个人数据必须有法律基础，有关情况下需征求数据被处理方同意。	

德国 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	《针对深度伪造人格权进行刑事保护的法案（草案)》 (2024.7)	核心规定	对侵犯他人人格权，通过计算机技术制作或修改并传播虚假媒体内容的行为 进行处罚，刑期最高两年或处以罚款	•技术解决方案：联邦参议院建议联邦政府发起和支持深度伪造识别与标记技术的研发和 实施项目，使企业和政府机构受益。 •评估与调整：期望定期评估涉及人工智能的法律实施情况，相关机构应提出建议，以优 化人格权保护措施，使其与技术发展和简化官僚程序相
		豁免条款	在符合特定正当利益（如艺术、科学、研究、教学、新闻报道等）的情况下， 相关行为可豁免刑事责任，涉及个人隐私领域通常很难适用豁免。	

		没收措施	可没收犯罪者或参与者使用的图像、音频载体及其他技术手段。	协调。 •支持经济发展：建议设立中央联络点，为中小企业提供深度伪造识别、标记及法律要求 方面的咨询，以考虑经济利益并减少官僚障碍。
		加重处罚情形	若将特定媒体内容向公众传播或传播涉及个人隐私领域的内容，刑期最 高可达五年或罚款。	
2	《人工智能软件物料清单的共同愿景》 （该文件由德国和意大利联合发布）	AI 治理与网络安全的“透明底账”	•所谓供应链攻击的目标通常是收集/窃取敏感信息、预先定位以及更广泛地造成损害，无论是利益相关者之间的关系还是经济方面。 •通过增加透明度，特别是关于最终 AI 系统的创建过程以及其各个组件和依赖关系的信息，可以实现通过 AI 供应链来提升网络安全。AI 软件物料清单（SBOM for AI）可以增加供应链的透明度，从而有助于网络安全。 •能够捕捉 AI 系统的静态和动态方面（例如用于训练、测试和验证的生命周期中的数据集或学习成果） •能够以机器可读格式轻松自动处理和工具生成； •尽可能利用结构化数据格式，以确保相关信息披露透明，按需向所有利益相关者提供。	（1）AI 软件物料清单（SBOM for AI）应包含区分 AI 系统的独特特征与传统软件组件的差异。在引入 AI 软件物料清单（SBOM for AI）之前，系统卡和模型卡等工具已被私营公司和 AI 监管机构提出，作为提供 AI 模型透明度的工具。尽管在某些情境中这些工具具有有效性，但目前这些工具存在缺乏协调一致且机器可读的格式、自动化和与其他工具的互操作性等问题。数据管理也是一个重要考虑因素。 （2）同时，AI 软件物料清单（SBOM for AI）需要能够提供训练管道和数据集的可追溯性，特别是在涉及专有闭源模型、合成数据和预训练信息的情况下，这些情况下使用了大量多样化的数据语料库和复杂的数据处理管道来创建基础模型。 此外，鉴于 AI 技术发展的速度，保持 AI 软件物料清单（SBOM for AI）的更新至关重要，当需要时添加新信息和相关字段，例如模型蒸馏这一新兴
		最小元素原则	•文件旨在突出一组核心数据字段，这些字段是可由机器生成和处理的。重要的是要强调，这些最小元素仅是合理建议，并应根据具体使用情境相应决定。 •为了增强度可信并避免给人一种虚假的安全感，AI 软件物料清单（SBOM for AI）应作为一个整体进行验	

			证。这意味着不仅要验证其各个组件（例如通过相应制造商的加密哈希或数字签名）的真实性，还要验证整个 AI 软件物料清单（SBOM for AI）。为了实现这一目标，可行的 AI 软件物料清单（SBOM for AI）至少应由其制造商进行数字签名。虽然 AI 软件物料清单（SBOM for AI）中的各个组件已签名，但整个 AI 软件物料清单（SBOM for AI）的签名应可从外部进行验证。	且有意义的技术，就应被纳入 AI 软件物料清单（SBOM for AI）。

英国 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	《生成式人工智能框架》2024 年 1 月	伦理和法律问题	•多样化和包容性的参与需要构建到项目生命周期中，以应对生成式人工智能可能带来的伦理问题。 •生成式人工智能模型可以处理个人数据，所以需要保护个人数据以避免个人隐私泄露。 •生成式人工智能模型在大型数据集上进行训练，这些数据集可能包含有偏见或有害的材料以及个人数据，数据集所带来的偏差将会贯穿生成式人工智能模型的整个生命周期，所以需要在所有阶段测试中努力降低偏差。 •生成式人工智能不应该用来取代战略决策。 人工智能系统不应损害个	英国政府生成式人工智能框架提供的实用建议： （1）确定使用生成式人工智能的明确目标，并确保这些目标与组织的人工智能发展路线图相一致； （2）选择满足需求并符合生成式人工智能功能的使用案例； （3）了解生成式人工智能的局限性，避免高风险的使用案例； （4）了解其他政府组织正在考虑哪些使用案例，思考是否可以共享信息或重复利用其工作成果；

			人和组织的合法权利，他们也不应该歧视个人或造成不公平的市场结果。	
		安全性问题	•政府数据可能包含敏感信息和个人信息，必须在任何时候都保证合法、安全和公平地处理这些信息。 •相关组织需要建立保障措施，并进行适当的技术控制，包括检测恶意活动的内容过滤和验证检查，以确保正确响应且不泄漏数据。	
		人工控制	•使用生成人工智能时确保有质量保证控制流程，其中包括由适当培训的合格工作人员来审查生成人工智能工具的输出，并验证生成人工智能输出的所有决策。 •纳入最终用户的反馈至关重要，即引入允许最终用户报告内容并触发人工审查流程的机制。	
2	《教育领域中的生成式人工智能》	数据保护	保护个人隐私，个人信息和敏感数据不应输入生成式人工智能工具。英国信息专员办公室也提示开发或使用 AIGC 的企业或个人需承担数据保护的法定义务。	/

意大利 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	利用《欧盟通用数据保护条例》作出限制禁令	跨境数据传输	GDPR 对数据跨境传输提出了严格的要求，确保数据传输到欧盟外的国家或地区时，接收方必须提供等同于 GDPR 的保护水平。这可以通过多种机制实现，包括标准合同条款、绑定性企业规则（BCRs）或获得适当的国家等级。	意大利数据保护监管机构利用该条例打响的国别监管第一枪，通过临时限制、全面禁用、违规调查、限期整改、解禁等系列措施，督促 OpenAI 在数据隐私方面达到合规要求，值得借鉴。
2	《2024-2026 年意大利人工智能战略》 2024 年 8 月	战略框架	分为四个宏观领域：研究、公共管理、企业和培训。该战略还提出了一个实施监测系统，并对监管环境进行了分析，概述了该战略必须部署的框架。	一方面，人工智能技术的运用将能够提高行政效率并优化公共资源管理；另一方面，人工智能技术将能够促进公民和企业对服务的使用，并促进这些主体与中央和地方政府之间的互动。战略强调“在公共行政部门中运用人工智能是打造全新、更高效、更便捷的公民服务形象的主要创新方向之一”，同时也强调了这些举措将在竞争力方面产生的影响。

芬兰 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	《人工智能系统的开发和使用中的数据保护指南》 2025 年 05 月	个人数据的处理合法合规	• 在处理个人数据之前，从数据主体的角度评估 AI 系统的数据保护风险； • 根据风险水平，决定所需的安全保障措施； •在开发、训练或使用 AI 系统时，为个人数据处理选择合适的法律依据； • 遵守《通用数据保护条例》(GDPR)中规定的数据保护原则； •明确定义哪些个人数据是必要的，以及这些数据在 AI 系统中的用途； •在 AI 系统的设计和开发阶段，应确保数据主体能够行使其权利。	指南还解释了 AI 系统的定义、何为自动化决策、何时需要开展数据保护影响评估(DPIA)，以及何种情况下个人数据处理构成高风险。

法国 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	国家人工智能战略（SNIA）	/	该战略由四个部分组成：宣布一项国家人工智能计划；一项开放数据政策；一个监管框架；以及制定道德规则，以确保人工智能的使用和发展是透明、可解释和非歧视性的。	自 2018 年以来，该战略一直由一名国家协调员负责协调，其任务是通过九个部委和其他公共机构实施该战略。 截至 2024 年，法国正处于该战略的第二阶段，重点是在整个经济领域推广人工智能技术，支持优先领域的发展和创新。国家人工智能战略的新阶段主要侧重于培训和吸引人才。该战略由国家人工智能协调员指导，并将以各利益相关方的工作为基础。