人工智能生成合成内容标识的管理与挑战

张继红

- 《人工智能生成合成内容标识办法》采用了标识双轨制设计，通过明确服务提供者、传播平台、互联网应用程序分发平台以及用户四类主体的责任和行为规范，构建了一整套标识管理责任体系。
- 美国加州立法更侧重于对大型生成式人工智能系统的提供者的监管，但适用范围较窄；欧盟区分了风险等级，但未细化标识方法，可操作性较弱。虽然不同国家和地区的规定各有不同，但均秉承了事后救济转向事前风险内嵌控制的治理理念。
- 我国人工智能标识制度落地过程中，技术可靠、执行成本与国际协同仍是待解难题，唯有通过技术创新、法律协同、全球合作三管齐下，才能构建既防范风险又促进发展的治理体系。

随着人工智能技术的快速发展，生成合成内容（AIGC）的应用呈现爆炸式增长，但也带来了传播虚假信息引发社会舆情、利用深度伪造实施诈骗、侵害知识产权等一系列问题，严重侵蚀了公众对网络信息的信任基础。在此背景下，国家互联网信息办公室等四部门于2025年3月联合发布《人工智能生成合成内容标识办法》（以下简称《标识办法》）和配套强制性国家标准《网络安全技术 人工智能生成合成内容标识方法》（GB 45438-2025）（以下简称“标识国标”），并将于2025年9月1日正式施行。同时，为了指导人工智能生成合成内容服务提供者和网络信息内容传播服务提供者开展人工智能生成合成内容的文件元数据隐式标识工作，《网络安全标准实践指南——人工智能生成合成内容标识 服务提供者编码规则》（TC260-PG-20252A）亦同步出台，构建了一套“规章+标准”的协同治理框架，让人工智能标识管理制度得以落地施行。

我国人工智能标识制度的主要框架

首先，《标识办法》采用了标识双轨制设计，创造性地构建了“显式+隐式”标识的双重体系。显式标识是指在生成合成内容或交互场景界面中添加的，以文字、声音、图形等方式呈现并可以被用户明显感知到的标识。例如，文本起始/末尾标注“AI生成”，音频插入特定节奏提示，视频起始画面或播放周边添加角标等。隐式标识是指采取技术措施在生成合成内容文件数据中添加的，不易被用户明显感知到的标识。隐式标识采用文件元数据嵌入技术，记录生成合成内容属性、服务提供者名称或编码、内容编号等制作要素信息，便于溯源和核验，为监管部门提供取证依据。标识双轨制设计兼顾了公众知情和技术溯源的双重需求，显式标识降低了用户误认风险，而隐式标识则可强化全链条监管。

其次，《标识办法》明确了服务提供者、传播平台、互联网应用程序分发平台以及用户四类主体的义务。服务提供者应履行显式及隐式标识义务，说明并提示用户仔细阅读、理解相关标识管理要求。提供网络信息内容传播服务的服务提供者，应当核验文件元数据中的隐式标识，根据核验结果做二次标注。互联网应用程序分发平台，应当核验互联网应用程序服务提供者生成合成内容标识的相关材料。用户应履行告知义务，主动声明并使用服务提供者提供的标识功能进行标识。此外，《标识办法》还做了三项禁止性规定，即任何组织和个人不得恶意删除、篡改、伪造、隐匿生成合成内容标识，不得为他人实施上述行为提供工具或服务，不能通过不正当标识手段损害他人合法权益。

通过明确各方责任和行为规范，构建了一整套标识管理责任体系，涵盖了从内容生成、传播到分发等全生命周期，由此形成闭环管理，确保标识信息的完整性和可追溯性。

第三，作为强制性国家标准，“标识国标”从技术层面提供了操作指南，对文本、图片、音频、视频、虚拟场景等不同内容类型提出差异化显式标识要求，不仅督促企业提高标识的准确性和有效性，切实保障了用户的知情权。例如，文本应采用文字或角标形式；图片标识需位于图片的边或角，文字高度不低于画面最短边长度的5%；音频标识应采用语音或音频节奏标识，使用正常语速；在正常播放速度下，视频标识持续时间不少于2s。

此外，考虑到中小企业的成本负担，《标识办法》允许低成本方案，仅提出“鼓励服务提供者添加数字水印等形式的隐式标识”，为人工智能产业营造一个公平的竞争环境。同时，还设置了6个月的过渡期，为企业预留技术适配时间。

国际比较：标识管理的制度差异

在适用范围上，我国《标识办法》适用于利用生成式人工智能技术向中国境内公众提供生成文本、图片、音频、视频等内容服务的网络信息服务提供者，覆盖面较广，旨在全面规范生成式AI技术的使用和传播。相较而言，美国《加州人工智能透明度法案》（2026年1月1日起生效），更集中在具有较大社会影响力的生成式人工智能系统上。主要针对拥有每月超过100万用户的大型生成式人工智能系统的提供者，且仅限于生成“图像、视频或音频”的系统。欧盟《人工智能法案》则根据人工智能应用的风险级别进行分类监管，适用范围涵盖了各类人工智能应用。对生成或操纵构成深度伪造的图像、音频或视频内容的人工智能系统的部署者，必须披露该内容是AI生成。

在标识要求方面，《标识办法》规定的显式和隐式标识属于提供者的法定义务，但同时还保留了一定灵活性，允许服务提供者在通过用户协议明确用户的标识义务和使用责任后，可提供不含显式标识的内容，并依法留存相关日志。这种灵活性为企业用户在特定场景下提供了操作空间。相比之下，美国《加州AI透明法案》对标识要求更为严格，所有生成内容均须强制性地进行“显性披露”（manifest disclosure）和“潜在披露”（latent disclosure），不允许豁免或通过协议免除上述披露义务。且在技术可行的情况下，应是永久性或者极难移除的。欧盟《人工智能法案》仅提出“应在首次互动或接触时以清晰可辨的方式提供给自然人”，并未区分显式和隐式标识，亦未对标识形式、方法和标准做进一步具体规定，而是“鼓励和促进在欧盟一级起草行为守则”以细化该义务。此外，还对透明度义务做了多场景下的特别豁免，彰显了欧盟平衡人工智能产业发展与安全的审慎考量。

在监管重点问题上，我国《标识办法》延伸了标识管理的监管链条，将传播平台也纳入了监管范围。而美国《加州AI透明法案》则采用了更为灵活的监管模式，将第三方标识义务的设定主要纳入合同范畴，具体由授权方进行监督。欧盟《人工智能法案》则聚焦生成合成内容的提供者和部署者，较少涉及信息传播平台的合规义务。但其区分了风险等级，对深度伪造等人工智能系统施以更严格的透明度义务要求。

此外，在技术手段与工具要求方面，我国《标识办法》未对检测工具的使用和可访问性作出具体要求。美国《加州人工智能透明法案》则明确规定生成式AI系统的提供者应当免费向用户公开提供具有AI检测功能的工具，大大加强了公众对生成合成内容进行验证的能力，有利于用户监督。欧盟《人工智能法案》也要求系统提供者嵌入技术解决方案，便于技术检测。

就责任机制而言，《标识办法》采用指示性条款，即由网信、电信、公安和广播电视等主管部门按照“有关法律、行政法规、部门规章的规定”予以处理。再结合《互联网信息服务算法推荐管理规定》《互联网信息服务深度合成管理规定》《生成式人工智能服务暂行办法》相关规定，仅对算法生成合成信息提出了标识要求并做了明确的法律责任规定，但其他违法行为罚则尚处于模糊状态。欧盟《人工智能法案》规定了违反透明度义务的高额罚款；美国《加州人工智能透明度法案》则规定，如果发现提供者违规，可每天处以5000美元的民事罚款。

综上，美国加州立法更侧重于对大型生成式人工智能系统的提供者的监管，并确保技术检测工具的可访问性，但适用范围较窄；欧盟则区分了风险等级，注重平衡人工智能产业与安全，但未细化标识方法，可操作性较弱；我国实现了“法律+技术”双轮驱动，覆盖从内容生产到传播的全链条治理，但缺乏明确的责任机制，后续落地实施的效果可能会被大打折扣。虽然不同国家和地区对标识方式、标识技术的规制内容各有不同，但都积极构建了人工智能生成合成的标识管理制度，且均秉承了事后救济转向事前风险内嵌控制的治理理念。

我国人工智能标识制度落地还需解决的问题

第一，标识技术仍有风险隐患。例如，显式标识可能会被恶意截图、转码、清洗等手段移除，而隐式标识的嵌入和识别亦可能会受到文件格式、压缩算法等因素的影响，导致标识信息的丢失或损坏。尽管《标识办法》对此做了禁止性规定，但缺乏有效的技术对抗手段和处罚细则，法律威慑力严重不足。未来需要进一步完善标识技术，提高其安全性和可靠性。

第二，法律衔接存在断层。《标识办法》作为我国网络信息内容和人工智能治理体系的重要一环，与《网络安全法》一脉相承，与《互联网信息服务算法推荐管理规定》《互联网信息服务深度合成管理规定》《生成式人工智能服务管理暂行办法》做了衔接，但仍存在协同不足的问题。例如，元数据隐式标识的存储与跨境传输也需符合数据出境安全评估要求，但现有法规未明确相应的操作细则。而且，算法备案、安全评估等程序与标识义务如何衔接亦缺乏指引，后续还应制定相应的操作性细则，指引企业实现合规。

第三，跨境传播引发法律冲突。全球化背景下人工智能生成合成内容的传播具有跨国界特点。如前所述，不同国家和地区的标识规范存在明显差异，可能会导致生成合成内容在跨境流动过程中出现标识不一致、监管空白或重复监管等问题。例如，美国加州规定水印不可移除，而我国允许用户协议声明后传播未标识内容。这种立法规范上的差异性可能产生法律适用冲突，标识内容在跨境场景下或被重复标注或被选择性遗漏，从而引发争议。因此，需要加强国际间的沟通、协调与合作，推动建立统一国际标识标准和规范，共同应对人工智能生成合成内容带来的挑战。

《标识办法》及相关标准的出台标志着我国在AIGC治理领域从“被动应对”转向“主动规制”，不仅从管理层面明确了标识规范，而且在技术层面提供了操作指南，避免了规范空洞化问题。然而，技术可靠、执行成本与国际协同仍是待解难题，唯有通过技术创新、法律协同、全球合作三管齐下，才能构建既防范风险又促进发展的治理体系。

（作者系上海政法学院教授、博导）