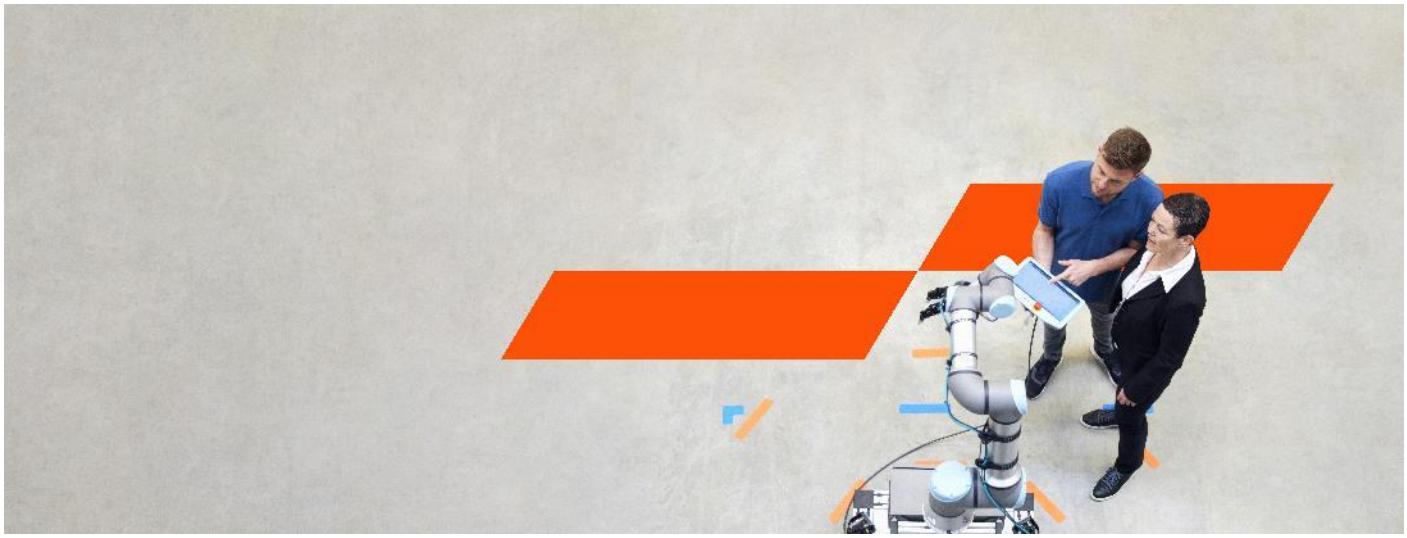


《人工智能生成合成内容标识办法》合规解码：服务者、平台及用户的三维责任

2025年9月1日，国家互联网信息办公室、工业和信息化部、公安部、国家广播电视总局联合发布的《人工智能生成合成内容标识办法》（以下简称**《标识办法》**）正式实施，标志着我国正式就人工智能生成合成内容标识落实全链条、全主体的治理体系。同步实施的还有《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**《标识方法》**），对标识方式、应用场景、标识格式等细节作出统一指导，为产业提供了明确可行的“操作手册”。

同日，国内六大主流社交媒体平台同步公告，上线“AI 生成”显式角标和隐式元数据标识功能，创作者发布时必须主动标注，平台将自动检测并补充标注，且严禁任何人删除、篡改或隐匿 AI 标识，违者将按社区规则和国家法规予以处置。

本文以《标识办法》划定的三大责任主体，即服务提供者、传播平台与用户为切入点，对照境外最新立法，简要梳理我国标识合规的要求。

一. 标识

《标识办法》将标识分为显式标识（Explicit Labeling）和隐式标识（Implicit Labeling）两类：

1. 显式标识

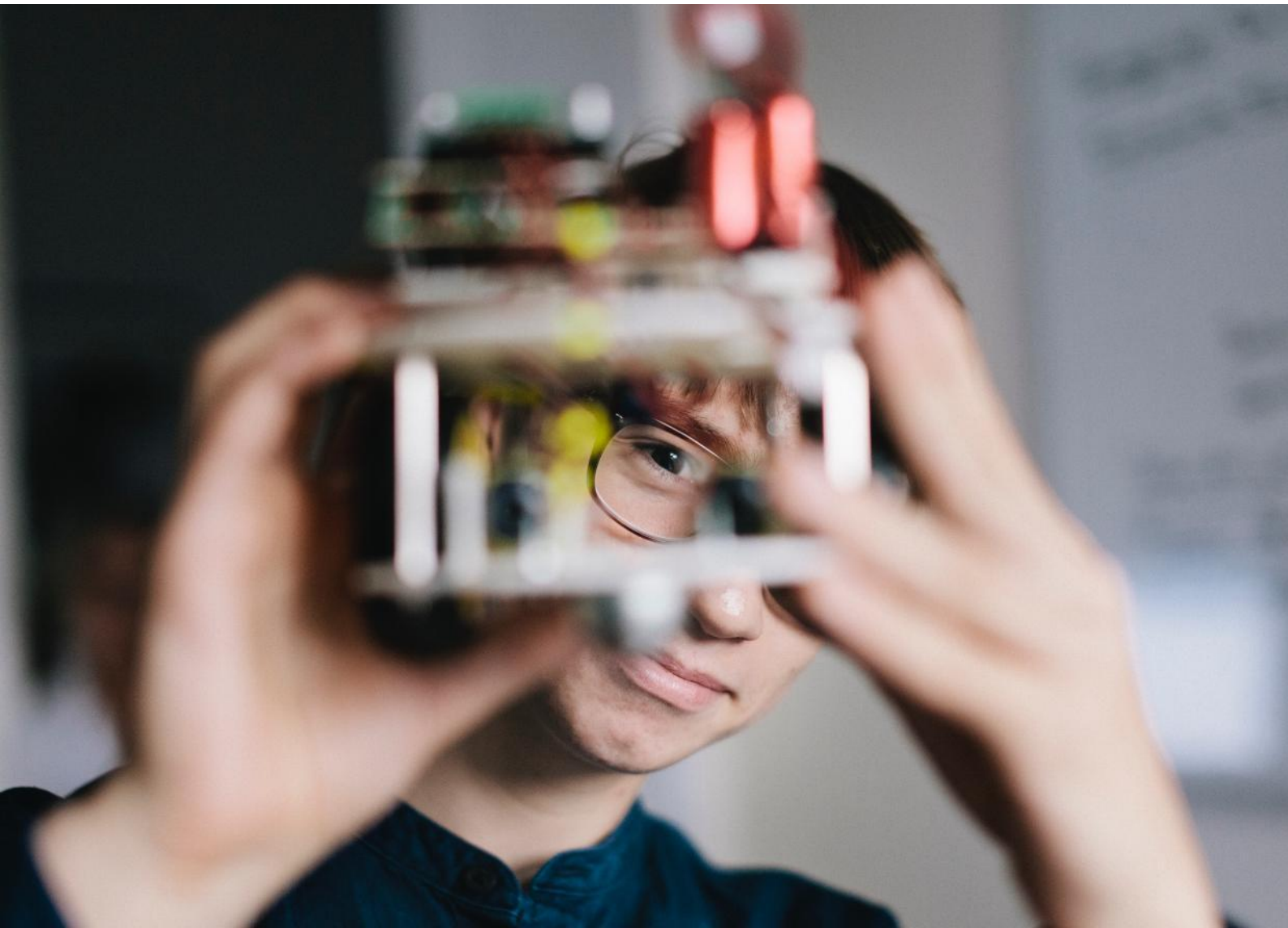
显式标识通过文字、声音、图形等形式呈现在人工智能生成的合成内容或交互场景界面中。其可被用户明显感知，用以向公众提示内容由人工智能生成合成。如文本中的“AI 生成”提示、图片或视频上的水印图标、音频文件开头的语音提醒或特定提示音。这类标识的目的是在第一时间告知受众“该内容来自人工智能”，防止用户将其误认成人工创作或真实事件。各类生成合成内容的显式标识要求如下表所示：

内容	标识方式	核心要点
文本	在开头、结尾或正文中适当位置添加文字/符号提示，或在界面周边提示	必须包含“AI”+“生成/合成”等要素，清晰可辨
音频	在开头、结尾或中间插入语音提示/提示音，或在界面显示文字提示	语音标识：正常语速，音频节奏“短长短短”清晰可辨
图片	在边缘或角落添加文字提示	包含“AI”+“生成/合成”，字体清晰，大小适度（文字高度不应低于画面最短边长度的5%）
视频	在起始画面和播放界面显著标注，可在中段或结尾补充	文字清晰可辨，大小适度（文字高度不应低于画面最短边长度的5%），至少显示两秒
虚拟场景	在起始画面显著标注，可在运行过程中适当位置显示	与视频要求类似
其他	根据自身应用特点添加显著的提示标识	

2. 隐式标识

隐式标识是采取技术措施在人工智能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。其核心功能是可追溯与防篡改。

文件元数据隐式标识方法包括五个要素：生成合成标签要素（内容的人工智能生成合成属性信息）、生成合成服务提供者要素（生成合成服务提供者的名称或编码）、内容制作编号要素（生成合成服务提供者对该内容的唯一编号）、内容传播服务提供者要素（内容传播服务提供者的名称或编码）以及内容传播编号要素（内容传播服务提供者对该内容的唯一编号）。文件元数据隐式标识格式具体参考《标识方法》附录E的规定，且在人工智能生成合成的内容文件中应参考《标识方法》附录F的范例仅保留一份文件元数据隐式标识。



二．服务提供者：标识与留痕

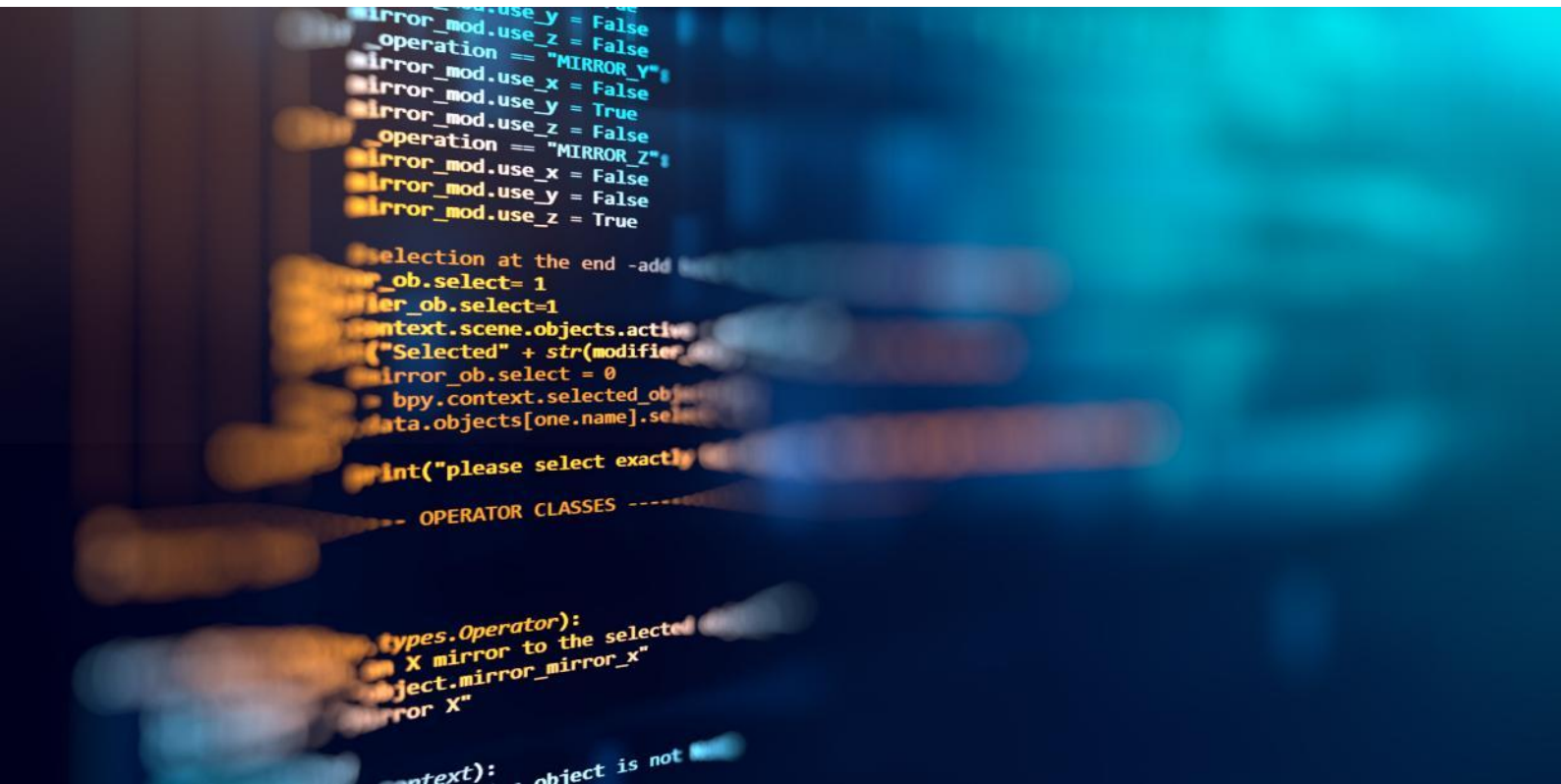
1. 在“生成点”同步完成显式标识

根据《互联网信息服务深度合成管理规定》第十七条及《标识办法》第四条的要求，深度合成服务提供者在提供**可能导致公众混淆或误认的服务时**，应当在生成或编辑的信息内容的合理位置加注**显著的提示标识**。这一显式标识可以是文字、图形或语音提示，并且必须在**下载、复制、导出**环节随内容保留，防止在跨平台传播过程中“脱落”。

值得注意的是，《标识办法》第九条在兼顾用户实际需求的同时，提供了有限豁免：在用户明确提出申请，且服务提供者通过协议界定其标识义务和使用责任的前提下，可以提供不含显式标识的合成内容。**但这种“去标”仅限于源头生成环节；一旦用户使用网络信息平台将合成内容公开传播，仍需依第十条规定主动声明并补充显式标识**。

2. 在文件元数据中添加隐式标识

与“显式标识”相辅相成，《标识办法》第五条**强制**要求服务提供者在生成合成内容的文件元数据中**添加隐式标识**。 同时，鼓励引入**数字水印**等技术手段，提高在跨平台复制、截图、转码场景下的持久性和防篡改能力。



3. 留痕：把标识义务写进协议、留存日志

实现双标识仅是合规起点，服务提供者还需将标识义务写入合同、留存日志。《标识办法》第八条明确要求服务提供者应当在**用户服务协议**中明确说明生成合成内容标识的方法、样式等规范内容，并提示用户仔细阅读并理解相关的标识管理要求。《标识办法》第九条进一步规定，用户申请服务提供者提供没有添加显式标识的生成合成内容的，服务提供者应依法留存提供对象信息等相关日志**不少于六个月。**

对比来看，欧盟的《人工智能法案》在透明度义务上与中国的“双重标签”有一定的相似之处，但在适用范围和义务主体存在差异。在隐式标识方面，欧盟要求 AI 系统提供者为其输出内容加上“机器可读标记”，这一点与中国的隐式标识相对应，均强调通过元数据或水印的形式实现跨平台检测与可追溯。

而在显式标识方面，中国《标识办法》规定，若合成内容落入《互联网信息服务深度合成管理规定》第十七条所指的 **“可能导致公众混淆或误认”**的情形，服务提供者就必须在生成点添加显著提示。与此不同，欧盟并未强制性地要求AI系统提供者添加显著标识，而将责任主体定位于部署者（deployer）（即使用或发布内容的人）。当AI生成的图像、音频或视频**“会被合理的人视为具有真实性或可信性”**时，部署者必须以“**清晰且可区分**”（clearly and distinguishably）的方式标明该内容为人工生成。此外，如果 AI 生成或操纵的文本“**以公共利益为目的向公众发布**”，部署者同样负有披露义务。总体而言，虽然中欧两种制度在显示标识的责任主体和适用范围存在差异，但他们都以防止公众被误导为核心，围绕透明性和真实性来设定显示标识的触发标准。

美国加利福尼亚州对标识的规定与中国和欧盟的规定存在显著不同。 根据SB 942 《AI 透明法案》，被界定为 **“Covered Provider”** 的 AI 提供者（即提供公开可访问的生成式人工智能系统，并在**加州拥有每月超过100 万用户或访问者**）必须对其生成或修改的音频、图像、视频内容嵌入隐式披露（latent disclosure）。该隐式披露要求在内容的元数据中加入可检测且难以移除的标识信息，用以标识内容来源等属性。与中国和欧盟强调显示标识不同，加州SB 942法案**对显式披露（manifest disclosure）采取“用户选项性”（option to include）**，即提供者需为用户提供启用显性标签的功能，但并未要求在所有情形下强制展示显性提示。同时，加州法律还要求 Covered Provider 提供一个对公众开放的检测工具，用于识别并验证隐式标识，并确保这些标识不易被删除或篡改。

三．网络传播服务提供者及互联网应用程序分发平台：核验与监管

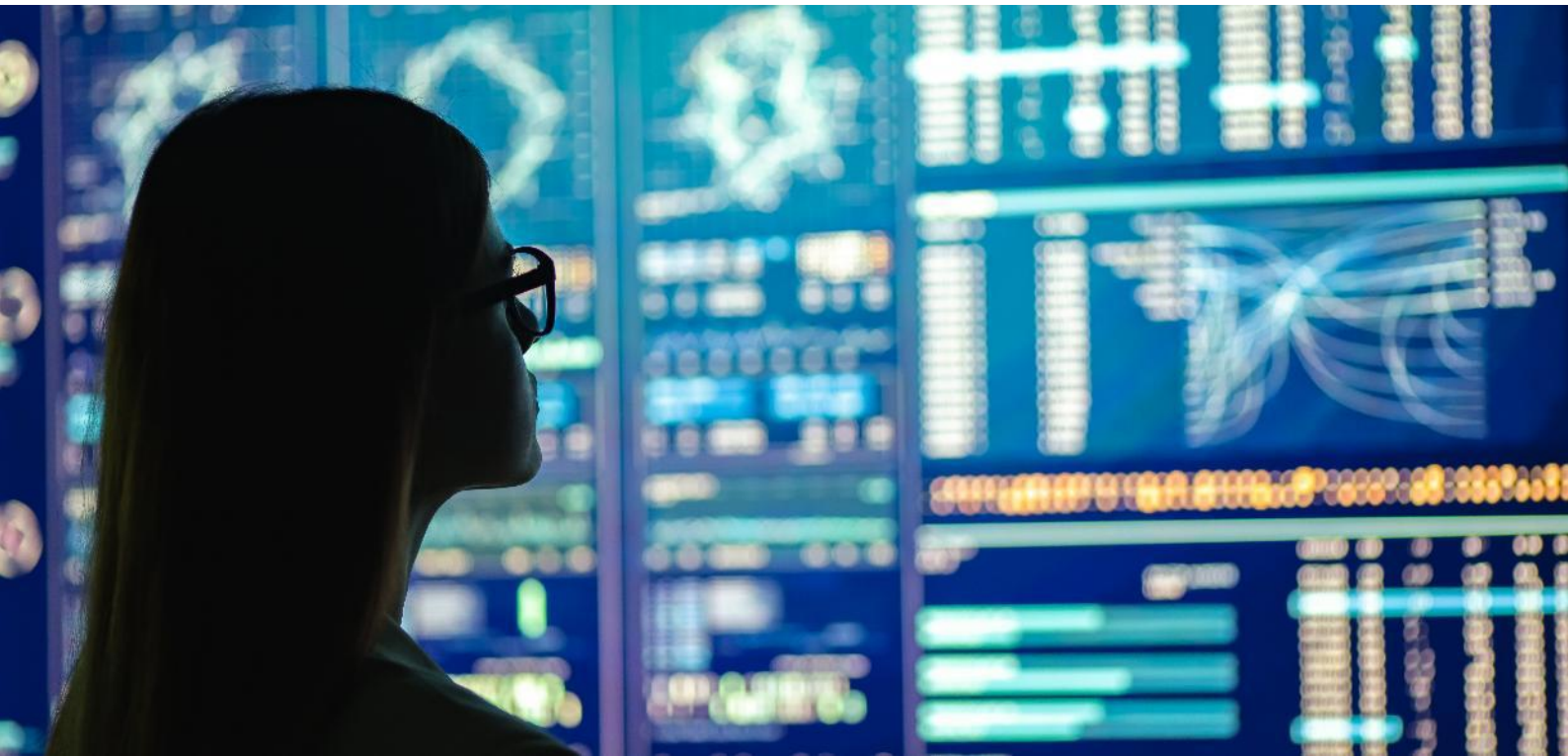
《标识办法》第六条、第七条将“标识”责任从内容生成端延伸至**传播端（社交平台、内容分发平台）和应用分发端（App Store）**，形成了一条贯通**生成 - 传播 - 分发**的全链条治理路径。

1. 网络传播服务提供者的“四步筛查”机制

根据《标识办法》第六条，平台在用户上传内容时，必须依次完成以下筛查：

- **核验元数据**：如文件元数据明确标注为 AI 合成内容，平台需在内容展示时添加显著提示标识，并告知公众其合成属性；
- **用户声明**：若元数据未标识，但用户主动声明为 AI 内容，平台同样需加注显著提示；
- **迹象识别**：若既无元数据标识、用户也未声明，但平台检测到显示提示或其他合成痕迹，应将其识别为“疑似 AI 内容”，并提示公众；
- **申报功能**：平台需提供必要的标识功能，提醒用户主动声明其上传内容是否包含 AI 合成成分。

在前三种情形中，平台不仅要做显式提示，还必须将“**合成属性 + 平台编码 + 内容编号**”写入文件元数据，以确保溯源和责任链条的完整。



2. 应用分发平台的把关义务

根据《标识办法》第七条，应用商店在程序上架或上线审核时，必须要求开发者“自报家门”。凡含有 AI 生成或合成功能的应用，必须提交其标识方案及相关材料，以便平台审查，防止“未合规的生成服务”进入市场。

与中国的“先验式全覆盖”不同，欧盟《数字服务法案》（DSA）的逻辑更偏向于分层治理：

- **普通平台：**主要适用“通知-删除”机制（notice & action）。即当用户举报或监管机构通知某一内容违法或具有风险时，平台需在 24 小时内处理并向用户说明理由。

- **超大型在线平台（VLOP/VLOSE）：**要求其在年度风险评估中主动评估和减轻“被操纵内容 / AI 生成内容”带来的系统性风险，并提升检测与标注能力（DSA 第 34、35 条）。

换言之，中国的模式要求所有平台在用户上传内容的环节都必须建立起“四步筛查”机制，从源头进行强制性的先验核验；而欧盟的模式则更具分层性，只有超大型平台才被要求主动提升检测与披露能力，普通平台的义务以“通知 - 删除”程序为主。

四．终端用户

在中国的监管框架下，《标识办法》第十条首次将标识义务明确延伸至终端用户，形成了“最后一厘米”的合规责任闭环。根据该条规定，**凡是用户借助网络信息内容传播服务发布合成内容的，均必须主动声明，并启用平台所提供的标识功能完成打标**。这一要求使得用户不再只是被动接受服务或依赖平台把关，而是被纳入整个合规链条的积极义务主体。

同时，《标识办法》还进一步规定，任何组织或个人不得删除、篡改、伪造或隐匿依法添加的标识，更不得为他人提供去标工具或服务，亦不得通过不正当的标识方式侵害他人合法权益。这些条款意味着，在中国的监管语境中，用户恶意“去标”或规避标识的行为被视为破坏透明机制的红线，可能面临行政处罚或承担相应的责任。需要说明的是，用户的责任更多是触发和配合：只要用户如实声明，平台会自动在前端展示显性标识（如“AI生成”字样或水印）。换句话说，用户的打标义务是通过使用平台内置功能来完成的。用户不需要自行设计或插入标签，只要不去规避或篡改，就算履行了法律义务。

与之相比，欧盟《人工智能法案》对用户的责任在形式上更严格，因为当AI生成内容可能被误认为真实时， deployer必须**亲自确保**披露“该内容为人工生成或被操纵”，而不是单纯依赖平台的自动工具。**在范围上，欧盟的显性标识义务明确豁免“私人非职业用途”用户，这意味着如果个人在非商业场景中使用生成式 AI，通常无需履行显性标识义务。**同时，欧盟并未在法律条文本身规定禁止用户删除或篡改标识。

就美国而言，在联邦层面，尚无统一的用户端打标义务。联邦贸易委员会（FTC）主要通过消费者保护与反欺诈框架来规制 AI 内容，重点在于防止误导性广告或虚假评论；一旦 AI 内容被用于欺诈，责任主体会被追责，但并不是以“未打标”为法律切入点。在州法层面，SB 942 法案要求大型模型提供者在输出端嵌入隐式水印，并提供显性标签功能，但其约束对象依旧是服务提供者与平台，而非普通用户。因此，美国的监管重心在企业端和市场秩序，而不是用户端。

综上，中国的模式体现出“形式宽松、范围最广”的特点；欧盟则是“形式严格、范围有限”，强调由用户自己打标，但只在特定高风险场景适用；美国则没有统一的用户标识义务，更依赖于消费者保护法和州法的企业端要求。这种差异不仅反映了三大法域在监管思路上的不同，也意味着在跨境合规时，企业和用户可能面临完全不同的责任分配逻辑。

基于上文分析，下表列出了中国、欧盟以及美国加州在服务提供者、平台以及用户三类主体责任上的主要差异：

主体	中国	欧盟	美国加州
服务提供者	隐性标识：强制嵌入元数据（生成方编码、编号等）。 显性标识：在生成点添加（可能导致公众混淆/误认的内容）	隐性标识：必须确保输出具备“机器可读标记” 显性标识：不直接由提供者承担，而是提供必要功能支持。	隐性标识：Covered Provider 必须嵌入元数据，内容来源、系统版本、唯一标识符等。 显性标识：非强制，提供者仅需向用户提供“选项性”工具。 必须提供公开可用的免费检测工具。
平台	四步筛查、元数据补录； App 分发平台须要求含 AI 功能的应用提交标识方案材料。	普通平台：主要适用“通知-删除”机制。 VLOP/VLOSE：需在年度风险评估中主动提升对 AI/操纵内容的检测和披露能力。	对平台并无单独提出类似中国“四步筛查”的强制要求
用户	显性标识义务：发布 AI 内容时必须主动声明并使用平台提供的标识功能。 禁止去标：任何人不得删除、篡改、隐匿或提供“去标工具”。	显性标识义务：部署者必须在内容“会被合理人视为真实或可信”时，清晰披露内容为 AI 生成或操纵。 文本义务：若以公共利益为目的发布，需披露。	无统一用户义务：联邦层面无强制打标要求。

五．法律责任

根据《标识办法》第十三条，违反标识义务的，由网信、工信、公安、广电等主管部门依据各自职责，依照相关法律、行政法规和部门规章予以处理。这意味着标识义务的违法风险并不限于单一部门执法，而是可能触发多部门联合监管。从实践角度看，违反标识义务的法律责任体系呈现多层次、多维度的特征，不仅涉及行政监管层面的处罚，还可能延伸至民事赔偿与刑事追责，形成全方位的法律约束机制。具体表现为：

1. 行政责任

在行政层面，主管部门可依据《生成式人工智能服务管理暂行办法》采取警告、通报批评、责令整改、暂停服务等措施；如同时违反《中华人民共和国网络安全法》《中华人民共和国数据安全法》或《中华人民共和国个人信息保护法》，或将面临更高强度的处罚。例如，《中华人民共和国网络安全法》《中华人民共和国个人信息保护法》最高罚款可达一百万元，《中华人民共和国数据安全法》则最高可达一千万元。此外，直接负责的主管人员和相关责任人员也可能被处以罚款，严重者甚至面临停业整顿、吊销许可证、关闭网站或应用下架等业务限制。

2. 民事责任：侵权行为引发的赔偿风险

违反标识义务不仅是合规风险，也可能引发民事侵权责任。如果 AI 生成内容未依法标注，造成他人权益受损，侵权方可能需要承担停止侵害、消除影响、赔礼道歉及赔偿损失的责任。典型情形包括：

- **人格权侵害**：如未标注的 AI 合成视频伪造他人肖像、声音或贬损名誉，可能侵害肖像权、名誉权、隐私权等。

- **知识产权侵害**：如 AI 内容未加标识而被冒充为原创，或抄袭他人作品、商标，可能构成著作权或商标权侵权。

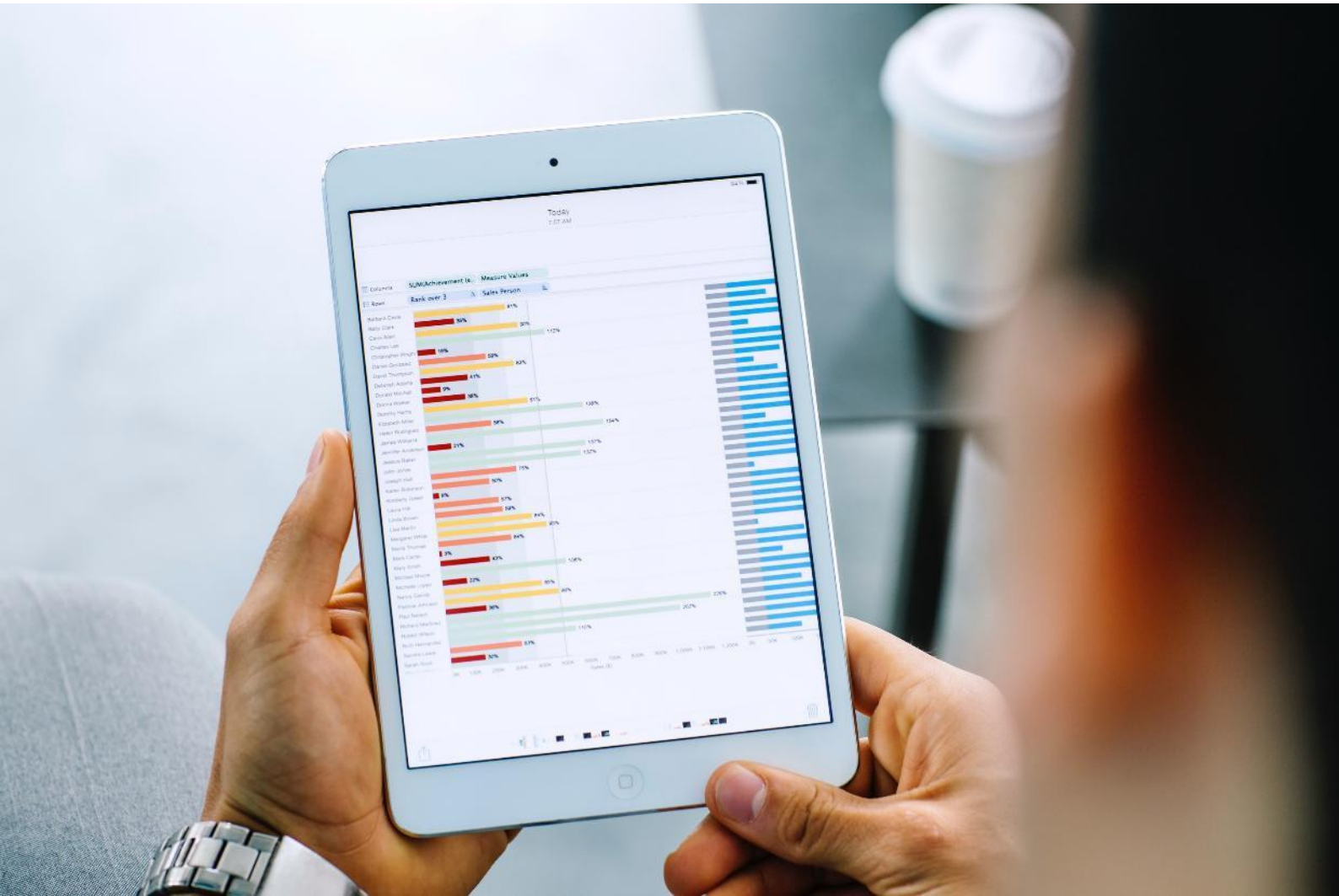
3. 刑事责任：情节严重时的刑事追责

若违法行为满足一定要件，社会危害性大，将突破行政与民事边界，上升为刑事犯罪。例如：

- **编造、故意传播虚假信息罪：** 利用未标识的 AI 内容制造虚假险情、疫情、灾情或警情，严重扰乱社会秩序的，可处三年以下有期徒刑，后果严重的则处三年以上七年以下有期徒刑；

- **帮助信息网络犯罪活动罪：** 为他人提供未标识的 AI 内容作为犯罪工具，情节严重的，可处三年以下有期徒刑或拘役，并处罚金；

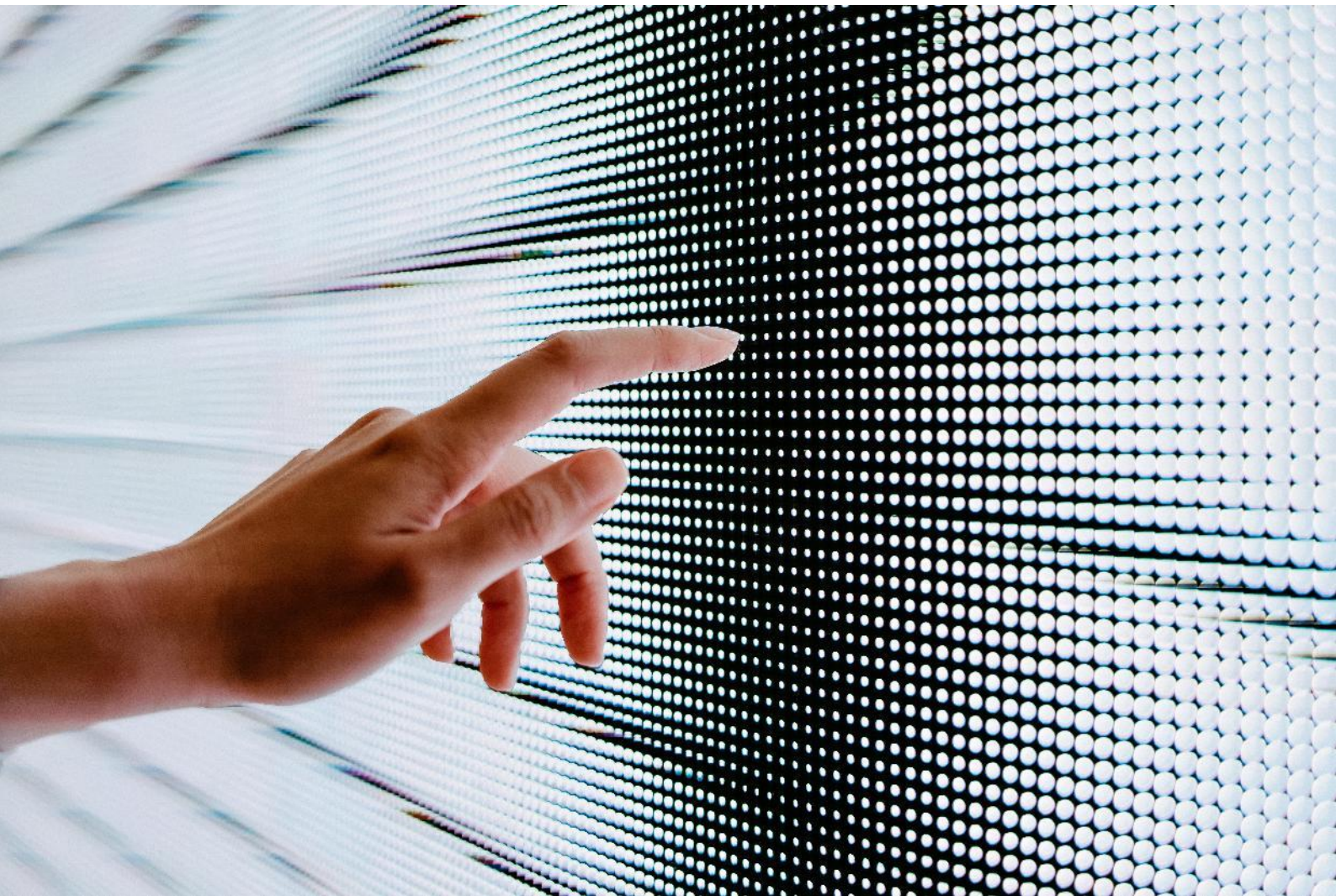
- **诈骗罪：** 利用未标识的 AI 内容伪造身份、合同骗取财物，数额较大的，刑期三年以下；数额巨大或有严重情节的，三年以上十年以下；特别巨大或特别严重的，十年以上直至无期徒刑，并处罚金或没收财产。



六. 合规建议

在该新规实施的背景下，企业既要满足本土监管要求，也要兼顾全球透明化趋势。为此，我们建议从以下几个方面着手：

主体	核心措施	风险防范效果
服务提供者（AI 模型/工具开发商）	在生成环节嵌入隐性标识（元数据/水印）；输出界面加显性提示；必要时引入第三方合规技术	实现“源头打标”，降低合规缺口，保障内容可追溯性
平台（传播服务商/应用分发方）	提供用户声明入口（如上传勾选 AI 生成）；前端显性提示；后台扫描检测与补标；在用户协议中明确标识义务与违规后果	把控传播环节风险，建立制度约束，形成“分发合规”
用户（内容生产者/发布者）	在发布时主动声明并启用平台标识功能；不得删除、篡改或规避已加标识	避免行政处罚、民事侵权或刑事追责，确保“终端合规”



结语

随着人工智能技术的飞速发展，科技与伦理的平衡愈发关键。这种动态的平衡并非简单的技术完善或伦理约束就能实现，而需要从多维度的治理框架出发：技术开发者需要承担更大的社会责任；政策制定者需要在规则中兼顾技术的灵活性和公平性；而公众则需要提高信息素养，增强对虚假内容的辨别能力。多方协作构建一个既高效又可信，既创新又负责任的数字内容生态。这不仅是人工智能时代的挑战，也是人类文明在科技浪潮中自我审视的一个重要契机。

联系我们

周伟然

普华永道中国科技、媒体及通信行业主管合伙人

电话： +86 (755) 8261 8886

电邮： wilson.wy.chow@cn.pwc.com

庄树清

普华永道中国内地税务数字化与转型主管合伙人

电话： +86 (21) 2323 3219

电邮： j.chong@cn.pwc.com

蒋亮

普华永道中国内地公司及监管服务税务及商务咨询合伙人

电话： +86 (21) 2323 8873

电邮： liang.l.jiang@cn.pwc.com

Michelle Taylor

程伟宾律师事务所合伙人

电话：+852 2833 4994

邮箱：michelle.a.taylor@tiangandpartners.com

*程伟宾律师事务所是一家独立的香港地区律师事务所且为普华永道网络成员。

感谢普华永道中国内地公司及监管服务税务及商务咨询高级顾问李雨婷对本文的贡献与支持。