

## 新规生效!

# 简析《人工智能生成合成内容标识办法》及其配套标准

作者：潘永建 | 朱晓阳 | 邓梓珊 | 左嘉玮

2025 年 3 月 7 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《人工智能生成合成内容标识办法》(以下简称“《标识办法》”), 并同步推出强制性国家标准《网络安全技术 人工智能生成合成内容标识方法 GB45438—2025》(以下简称“《标识方法》”), 《标识办法》及其配套标准《标识方法》将于 2025 年 9 月 1 日正式实施。对于其中的人工智能生成合成内容标识要求、相关主体(服务提供者、相关生态企业和用户)的义务及平台责任、人工智能生成内容不规范带来新的安全风险以及企业合规要点等问题, 本文将进行重点解析。

## 1. 适用范围: 网络信息服务提供者开展人工智能生成合成内容标识活动

《标识办法》第二条规定, 其主要适用对象为符合《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管理暂行办法》规定情形的 **网络信息服务提供者**(以下简称“服务提供者”)开展人工智能生成合成内容标识活动。包括:

- **直接面向用户的服务提供者:** 如运营 AI 聊天小程序、生成式写作工具或图像生成平台的企业。
- **提供技术支持的服务提供者:** 如通过 API 接口为其他企业提供生成合成服务的企业。

.....  
如您需要了解我们的出版物,  
请联系:

Publication@llinkslaw.com

因此, 相关技术支持者即使不直接面向用户, 也必须要在技术层面落实标识措施。

## 2. 人工智能生成合成内容标识要求：显式标识 vs. 隐式标识

《标识办法》主要规定了网络信息服务提供者如何进行标识(详见下表列示), 那首先应准确理解何为“标识”。《标识办法》规定了两种标识方式: 显式标识和隐式标识:

- **显式标识**主要用于向公众提示内容由人工智能生成合成(例如, 在文章开头标注“AI 生成”);
- **隐式标识**则类似于后台的日志记录或水印信息, 用于保存和标记生成合成内容的相关信息。

|             | 定义                                                | 适用场景                                                                                                                                                                                                                                                                                                                                                | 标识方式                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|-------------|---------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>显式标识</b> | 在生成合成内容或者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 | <p>按照《标识办法》并参照《标识方法》, 提供以下深度合成服务, <b>可能导致公众混淆或者误认的</b>:</p> <p>(一)智能对话、智能写作等模拟自然人进行文本的生成或者编辑服务;</p> <p>(二)合成人声、仿声等语音生成或者显著改变个人身份特征的编辑服务;</p> <p>(三)人脸生成、人脸替换、人脸操控、姿态操控等人物图像、视频生成或者显著改变个人身份特征的编辑服务;</p> <p>(四)沉浸式拟真场景等生成或者编辑服务;</p> <p>(五)文生图片等图片内容生成服务;</p> <p>(六)音乐创作等音频内容生成服务;</p> <p>(七)文生视频、图生视频等视频内容生成服务;</p> <p>(八)其他具有生成或者显著改变信息内容功能的服务。</p> | <p>(一)在文本的起始、末尾或者中间适当位置添加文字提示或者通用符号提示等标识, 或者在交互场景界面、文字周边添加显著的提示标识;</p> <p>参照《标识方法》, 文本标识应包含人工智能要素(包含“人工智能”或“AI”, 表明使用人工智能技术)和生成合成要素(包含“生成”和/或“合成”, 表明内容制作方式为生成和/或合成), 且使用的字体和颜色应清晰可辨。</p> <p>(二)在音频的起始、末尾或者中间适当位置添加语音提示或者音频节奏提示等标识, 或者在交互场景界面中添加显著的提示标识;</p> <p>参照《标识方法》, 语音标识和交互场景界面标识均应包含人工智能要素和生成合成要素。其中, 语音标识应使用正常语速; 音频节奏标识应为“短长 短短”的节奏, 应清晰可辨。</p> <p>同时, 交互场景界面标识可采用在内容附近持续显示提示文字, 或在交互场景界面顶部、底部、背景等适当位置持续显示提示文字等一种或多种方式, 使用的字体和颜色应清晰可辨。</p> |

|  |  |  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|--|--|--|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  |  |  | <p>(三)在图片的适当位置添加显著的提示标识;</p> <p>参照《标识方法》, 图片标识应采用文字提示, 且包含人工智能要素和生成合成要素。特别地, 图片标识应位于图片的边或角, 且文字高度不应低于画面最短边长度的 5%; 图片标识使用的字型应清晰可辨。</p> <p>(四)在视频起始画面和视频播放周边的适当位置添加显著的提示标识, 可以在视频末尾和中间适当位置添加显著的提示标识;</p> <p>参照《标识方法》, 视频标识应采用文字提示, 且包含人工智能要素和生成合成要素。特别地, 视频标识应位于视频画面的边或角, 使用的字型应清晰可辨, 文字高度不应低于画面最短边长度的 5%; 在视频正常播放速度下, 视频标识持续时间不应少两秒。</p> <p>(五)呈现虚拟场景时, 在起始画面的适当位置添加显著的提示标识, 可以在虚拟场景持续服务过程中的适当位置添加显著的提示标识;</p> <p>参照《标识方法》, 虚拟场景标识与视频标识类似, 应采用文字提示, 且应包含人工智能要素和生成合成要素。虚拟场景标识应位于画面的边或角, 使用的字型应清晰可辨, 文字高度不应低于画面最短边长度的 5%。</p> <p>(六)其他生成合成服务场景根据自身应用特点添加显著的提示标识。</p> |
|--|--|--|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|      |                               |                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                                                                       |
|------|-------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 隐式标识 | 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 | <p>使用相关服务生成或者编辑的信息内容。</p> <p>笔者注：理论上所有生成合成内容均应附带隐式标识，但实践中可能更多适用于向用户输出相关文件的场景。对于纯文本内容，由于其仅由字符组成，缺乏嵌入元数据或数字水印所需的复杂结构，技术上难以实现隐式标识。</p> | <p>在文件元数据中添加隐式标识，隐式标识包含生成合成内容属性信息、服务提供者名称或者编码、内容编号等制作要素信息。</p> <p>参照《标识方法》，具体而言，隐式标识应包括以下要素：</p> <p>(1) 生成合成标签要素：内容的人工智能生成合成属性信息；</p> <p>(2) 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；</p> <p>(3) 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；</p> <p>(4) 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；</p> <p>(5) 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。</p> <p>关于文件元数据隐式标识的具体格式，可参照《标识方法》的附录E。</p> <p>鼓励在生成合成内容中添加数字水印等形式的隐式标识，但不作强制要求。</p> |
|------|-------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

3. 其他相关主体的义务(平台责任)

需要注意的是，虽然标识的主要义务在于服务提供者，但传播服务提供者(如社交媒体平台、视频平台)、应用程序分发平台(如手机商城)和用户等其他相关主体也必须履行相应义务。为构建全链条责任主体体系、强化合成内容传播的规范治理，《标识办法》对相关主体的义务作出了明确规定。整理如下表：

| 主体类型             | 义务                                                                                                      | 对应条款                  |
|------------------|---------------------------------------------------------------------------------------------------------|-----------------------|
| 提供网络信息内容传播服务的提供者 | 核验文件元数据中的隐式标识                                                                                           | 《标识办法》第六条第一款第(一)至(三)项 |
|                  | 文件元数据明确标明为生成合成内容的, 采取适当方式在发布内容周边添加显著的提示标识                                                               |                       |
|                  | 未核验到隐式标识, 但用户声明为生成合成内容的, 采取适当方式在发布内容周边添加显著的提示标识                                                         |                       |
|                  | 文件元数据中未核验到隐式标识, 用户也未声明为生成合成内容, 但提供网络信息内容传播服务的提供者检测到显式标识或者其他生成合成痕迹的, 识别为疑似生成合成内容, 采取适当方式在发布内容周边添加显著的提示标识 |                       |
|                  | 提供必要的标识功能, 并提醒用户主动声明发布内容中是否包含生成合成内容                                                                     | 《标识办法》第六条第一款第(四)项     |
| 互联网应用程序分发平台      | 存在第六条第一款第(一)至(三)项情形的, 在文件元数据中添加传播要素信息(如生成合成内容属性信息、传播平台名称或者编码、内容编号等)                                     | 《标识办法》第六条第二款          |
|                  | 在应用程序上架或上线审核时, 要求应用程序服务提供者说明是否提供生成合成服务                                                                  | 《标识办法》第七条             |
|                  | 应用程序服务提供者提供生成合成服务的, 核验其生成合成内容标识相关材料                                                                     |                       |
| 用户               | 使用网络信息内容传播服务发布生成合成内容的, 应当主动声明并使用服务提供者提供的标识功能进行标识                                                        | 《标识办法》第十条             |
| 任何组织/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恶意行为提供工具或者服务, 不得通过不正当标识手段损害他人合法权益                                     | 《标识办法》第十条             |

4. 可以不添加显式标识的场景

考虑到用户在使用人工智能服务时的实际需求,《标识办法》第九条规定,在同时满足以下条件时,服务提供者可以不添加显式标识:

- 用户主动申请获取未附加显式标识的生成合成内容;
- 用户协议中已明确规定用户的标识义务与使用责任;
- 依法留存提供对象信息等相关日志不少于六个月。

若用户在后续使用中拟将生成合成内容公开发布或传播,则仍应遵守《标识办法》第十条等相关要求,主动声明生成合成情况并添加显式标识。

值得讨论的是，用户进行一定修改后的生成合成内容后是否仍需声明和标识？笔者认为，应依据修改幅度区分：若仅进行简单调整(如图片色调调整)，则仍应当声明、标识；若为重大、实质性修改(例如图像大幅重绘)，则可豁免相关要求。不过我们注意到，具体的判定标准目前仍不明确，有待监管部门进一步明确。

## 5. 安全风险及企业合规要点

随着生成式人工智能技术的广泛应用，内容生产门槛显著降低，信息传播效率得到极大提升，但如果生成合成内容未依法标识或标识不规范，将可能引发一系列新型安全风险，包括传播虚假信息、加剧舆论操控风险，身份冒用和欺诈风险上升，以及平台信任机制受损等等。

因此，为有效应对《标识办法》及其配套标准的实施要求，我们建议，企业宜从技术、流程、制度、培训等多角度出发，构建覆盖内容生成、传播、分发、使用全生命周期的合规治理体系：

- 除了按照本文前述要求嵌入合规标识的技术处理工作，企业应格外留心流程和制度规范，建立和健全内容发布前的审核流程及用户请求记录、标识嵌入记录等日志留存及可追溯机制。
- 企业应制定《生成式人工智能内容标识操作手册》等内部合规制度，细化各类内容标识的标准及相应负责部门/人员以落实责任。
- AI 相关的内容合规培训亦有助于企业员工提高内容标识的合规意识，对于产品、技术、法务、运营等关键岗位，企业更应适当提高培训频次并建立相应考核机制，优化培训效果。
- 平台或服务的《用户协议》也有需要更新的部分，应明确用户在获取未附加显式标识内容时的使用限制与再传播义务。

如您希望就相关问题进一步交流，请联系：



潘永建  
+86 21 3135 8701  
david.pan@llinkslaw.com



朱晓阳  
+86 21 3135 8683  
nigel.zhu@llinkslaw.com



邓梓珊  
+86 21 6043 3793  
susan.deng@llinkslaw.com

如您希望就其他问题进一步交流或有其他业务咨询需求，请随时与我们联系: [master@llinkslaw.com](mailto:master@llinkslaw.com)

#### 上海

上海市银城中路 68 号  
时代金融中心 19 楼  
T: +86 21 3135 8666  
F: +86 21 3135 8600

#### 北京

北京市朝阳区光华东里 8 号  
中海广场中楼 30 层  
T: +86 10 5081 3888  
F: +86 10 5081 3866

#### 深圳

深圳市南山区科苑南路 2666 号  
中国华润大厦 18 楼  
T: +86 755 3391 7666  
F: +86 755 3391 7668

#### 香港

香港中环遮打道 18 号  
历山大厦 32 楼 3201 室  
T: +852 2592 1978  
F: +852 2868 0883

#### 伦敦

1/F, 3 More London Riverside  
London SE1 2RE  
T: +44 (0)20 3283 4337  
D: +44 (0)20 3283 4323



[www.llinkslaw.com](http://www.llinkslaw.com)



Wechat: Llinkslaw

#### 本土化资源 国际化视野

##### 免责声明：

本出版物仅供一般性参考，并无意提供任何法律或其他建议。我们明示不对任何依赖本出版物的任何内容而采取或不采取行动所导致的后果承担责任。我们保留所有对本出版物的权利。

© 通力律师事务所 2025