

## 互联网上, AI 的身份标识 ——《人工智能生成合成内容标识办法》解读 作者: 杨迅 | 张莹

1993年,《纽约客》杂志的一幅漫画描绘了一只狗坐在电脑前,对同伴说:“在互联网上,没人知道你是一条狗。”如今,随着人工智能技术的飞速发展,我们面临的问题是:“在互联网上,没人知道你是AI。”然而,AI并非人类,识别对方是人类还是AI,已成为人们迫切的需求,因此,人工智能标识的要求应运而生。

国家互联网信息办公室、工业和信息化部、公安部、国家广播电视总局近日联合发布了《人工智能生成合成内容标识办法》(以下简称《标识办法》),及其配套国标《网络安全技术 人工智能生成合成内容标识方法》将于2025年9月1日同步实施,共同指导相关主体规范开展标识活动。

### 一. 《标识办法》概述

#### 1. 出台背景

随着生成式AI技术的普及应用,机器创作与人类原创作品的界限日益模糊。如何确保用户辨别信息来源、预防深度伪造等技术滥用,已成为全球性挑战。我国GAI产业正处于快速发展阶段,亟需建立清晰的规则体系,为产业创新提供确定性预期。

本次出台的《标识办法》是在《互联网信息服务算法推荐管理规定》(以下简称《算法推荐管理规定》)、《互联网信息服务深度合成管理规定》(以下简称《深度合成管理规定》)和《生成式人工智能服务管理暂行办法》(以下简称《暂行办法》)的基础上,进一步细化了AI生成内容标识的具体实施规范。2022年发布的《深度合成管理规定》第十七条明确要求服务提供者在生成或编辑的信息内容适当位置进行显著标识,向公众提示深度合成情况。《标识办法》则是更系统全面地构建了从内容生成到传播全链条的标识管理框架。

.....  
如您需要了解我们的出版物,  
请联系:

Publication@llinkslaw.com

## 2. 意义和制度设计

人工智能生成内容的标识要求实际上是要求向公众公示该内容由人工智能生成，从而赋予公众是否信赖和欣赏该内容的选择权。而该些标识不可抹灭的特性则意味着无论该人工智能生成物如何流转，在他基因深处都打上了人工智能的烙印。

《标识办法》的推出在落实前述标识要求的基础上，重点解决“哪些是生成的”“谁生成的”“从哪里生成的”等关键问题，推动由生成到传播各环节的全流程管理，力争打造可信赖的人工智能技术。该办法采用了系统化的标识框架，通过显式与隐式标识相结合的方式，既确保用户直观感知内容属性，又保障技术追溯可能性。

在制度设计上，《标识办法》明确划分了生成环节与传播环节的责任边界，对不同主体提出差异化要求，形成了多层次、全覆盖的标识管理体系。这种设计既考虑了技术实现的可行性，又兼顾了用户体验和企业成本，体现了“统筹发展与安全”的治理理念。

## 二. 生成环节的标识要求

生成环节的标识要求是《标识办法》的核心部分，明确了谁应当进行标识、标识的分类以及具体标识要求，为人工智能生成内容的规范化管理提供了基础框架。

### 1. 标识主体

根据《标识办法》第二条，开展人工智能生成合成内容标识活动的主体为符合《算法推荐管理规定》、《深度合成管理规定》和《暂行办法》规定情形的网络信息服务提供者。这些主体在提供 AI 生成合成内容服务时，负有添加标识的法定义务。

行业组织、企业、教育和科研机构、公共文化机构、有关专业机构等研发、内部使用人工智能生成合成技术，但未向公众提供服务的，则不属于标识主体，不负标识义务。

### 2. 标识类别

《标识办法》将人工智能生成合成内容标识分为两大类：显式标识和隐式标识。

**显式标识：**是指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被用户明显感知到的标识。这种标识直接面向最终用户，目的是让普通用户能够直观地识别内容的生成来源，增强用户的知情权和选择权。

**隐式标识：**是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。这种标识通常嵌入在文件元数据或以数字水印等形式存在，虽然用户不易直接感知，但可以用于内容溯源、责任追究和技术监管。

### 3. 标识要求

#### 显式标识

根据《标识办法》第四条，服务提供者提供的生成合成服务属于《深度合成管理规定》第十七条第一款情形，即生成或者编辑的信息内容可能导致公众混淆或误认时，应按照下列要求对生成合成内容添加显示标识：

- (1) 文本内容：在文本的起始、末尾、中间适当位置添加文字提示或通用符号提示等标识，或在交互场景界面或文字周边添加显著的提示标识。
- (2) 音频内容：在音频的起始、末尾或中间适当位置添加语音提示或音频节奏提示等标识，或在交互场景界面中添加显著的提示标识。
- (3) 图片内容：在图片的适当位置添加显著的提示标识。
- (4) 视频内容：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，可在视频末尾和中间适当位置添加显著的提示标识。
- (5) 虚拟场景：在起始画面的适当位置添加显著的提示标识，可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。
- (6) 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。

服务提供者提供生成合成内容下载、复制、导出等功能时，应当确保文件中含有满足要求的显式标识。

#### 隐形标识

根据《标识办法》要求除了显式标识外，服务提供者应当按照《互联网信息服务深度合成管理规定》第十六条的要求，对于使用深度合成技术生成或者编辑的信息内容，该生成合成内容的文件元数据中添加隐式标识。隐式标识需要包含生成合成内容属性信息、服务提供者名称或编码、内容编号等制作要素信息。

《标识办法》对隐式标识采取了弹性处理方式，考虑到技术实现难度和企业成本因素，对于数字水印等更高级形式的隐式标识，《标识办法》并未作强制要求，而是采取鼓励态度，提出服务提供者可在生成合成内容中添加数字水印等形式的隐式标识。

### 三. 传播环节的标识要求

传播环节的标识要求是《人工智能生成合成内容标识办法》的重要组成部分，它规范了 AI 生成内容从生成到传播的全链条管理，确保标识信息能够有效跟随内容流转，并明确了各参与主体的责任与义务。

## 1. 网络信息服务提供者

作为 AI 生成合成内容的源头，网络信息服务提供者承担着最基础的标识责任。《标识办法》要求这类主体不仅要在生成内容时添加适当的显式、隐式标识，还需要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。以此确保用户在使用服务时对标识规则有充分了解，并为后续生成合成内容进入传播过程中标识合规奠定基础。

此外，《标识办法》第十三条还规定，服务提供者在履行算法备案、安全评估等手续时，应当按照本办法提供生成合成内容标识相关材料，并加强标识信息共享，为防范打击相关违法犯罪活动提供支持和帮助。

## 2. 网络信息内容传播平台

作为内容传播的关键环节，内容传播平台承担着重要的标识验证和补充责任。根据《标识办法》，平台应当采取措施，规范生成合成内容传播活动，具体包括：

- (1) 核验文件元数据中是否含有隐式标识，对于含有隐式标识的，应当采取适当方式在发布内容周边添加显著的提示标识，明确提醒用户该内容属于生成合成内容；
- (2) 对于文件元数据中未核验到隐式标识，但用户声明为生成合成内容的，应当采取适当方式添加显著的提示标识，提醒用户该内容可能为生成合成内容；
- (3) 对于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，但平台检测到显式标识或其他生成合成痕迹的，可识别为疑似生成合成内容，并采取适当方式在发布内容周边添加相应提示标识；
- (4) 对于以上三项确为、可能和疑似生成合成内容的，应当在文件元数据中添加生成合成内容属性信息、传播平台名称或编码、内容编号等传播要素信息；以及
- (5) 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。

这些要求体现了平台的“把关人”角色，通过多层次的标识验证和补充机制，确保 AI 生成内容在传播过程中的可识别性和透明度。

## 3. 互联网应用程序分发平台

《标识办法》明确规定，互联网应用程序分发平台在应用程序上架或上线审核时，应当要求互联网应用程序服务提供者说明是否提供人工智能生成合成服务。互联网应用程序服务提供者提供人工智能生成合成服务的，互联网应用程序分发平台应当核验其生成合成内容标识相关材料。

这一规定将监管前移至应用分发环节，通过应用商店的准入机制，从源头上确保提供 AI 生成内容服务的应用具备必要的标识功能，形成闭环管理。

#### 4. 用户

用户作为内容的最终接收者和潜在的二次传播者，也被纳入标识管理体系。《标识办法》规定，用户向提供网络信息内容传播平台服务的服务提供者发布生成合成内容时，应当主动声明并使用平台提供的标识功能进行标识。

这一规定体现了标识责任的延伸，不仅要求原始服务提供者添加标识，也要求内容的二次传播者保持标识的完整性，形成全链条的标识管理。同时，这也是对用户社会责任的强调，鼓励用户在享受 AI 技术便利的同时，也承担相应的信息透明义务。

此外，如果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，服务提供者可以在通过用户协议明确用户的标识义务和使用责任后，提供不含显式标识的生成合成内容，并依法留存提供对象信息等相关日志不少于六个月。

### 四. 建议

面对《标识办法》的生效和实施，人工智能企业必须迅速行动起来，做好合规的充分准备。

- (一) **构建完善的标识技术体系。** 迅速搭建一套符合《标识办法》的技术架构，涵盖显式与隐式标识的全方位解决方案。针对不同内容形态——文本、图片、音频、视频等，设计出既合规又直观的标识展示方式。同时，开发稳定而可靠的元数据嵌入和数字水印技术，确保标识在内容流转中坚不可摧，防止被轻易删除或篡改。这些技术要求必须与《网络安全技术 人工智能生成合成内容标识方法》国家标准同步，确保无缝对接。
- (二) **更新用户协议与服务条款。** 及时更新用户协议和服务条款，清晰阐述生成合成内容的标识方法和样式，建立醒目的提示机制，让用户对标识规则一目了然。对于用户在特殊情况下申请不包含显式标识的内容，协议中应明确其标识义务和使用责任，并建立完善的日志留存机制。
- (三) **优化产品交互设计。** 在产品设计上，重新审视并优化用户交互流程，将标识要求巧妙融入产品体验之中。例如，在内容生成界面加入标识预览功能，在内容导出时增设标识确认步骤，在内容分享界面提供清晰的标识说明等。通过精心设计，不仅满足合规要求，更将用户体验的影响降至最低，实现合规与体验的双赢。

如您希望就相关问题进一步交流, 请联系:



杨 迅  
+86 21 3135 8799  
xun.yang@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