君合研究简讯

JUNHE

2024 年 9 月 30 日

# 人工智能法律热点问题

## AIGC 内容如何标识及溯源？——网信办发布办法及标准征求意见稿

随着 AI 技术发展，生成合成技术不仅在逼真程度上日臻成熟，技术工具的可及性及易用性极大提高。当人人均可低成本地制作、传播深度伪造的内容时，现实与虚拟的界限不再显而易见，互联网信息传播的基本逻辑遭到挑战。例如，2024 年 8 月曝光的韩国“N 号房 2.0”事件中，不法分子将熟人女性的照片换脸到不雅视频中，并在 Telegram 上进行传播。类似地，9 月央视报道多地不法分子通过“换脸”技术，利用受害者在社交媒体发布的自拍伪造不雅照进行敲诈勒索。

互联网信息传播的信任基础需要被重塑，且迫在眉睫，而生成合成内容的标识将是重建信任边界的重要工具。网信办分别于 2022 年及 2023 年发布《互联网信息服务深度合成管理规定》及《生成式人工智能服务管理暂行办法》，对于生成合成服务提供者的标识义务进行了一般性规定。2024 年 9 月 14 日，网信办发布《人工智能生成合成内容标识办法（征求意见稿）》（以下简称“《办法征求意见稿》”）及其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（征求意见稿）》（以下简称“《标准征求意见稿》”），以进一步明确添加标识的的具体要求，《办法征求意见稿》意见反馈截止时间为 2024 年 10 月 14 日，《标准征求意见稿》意见反馈截止时间为 2024 年 11 月 13 日。

### 一、适用范围

在中国境内应用算法推荐技术、深度合成技术或生成式人工智能技术提供互联网信息服务的服务提供者（以下简称“提供者”）是标识办法及标准征求意见稿的主要义务主体。（《办法征求意见稿》第二条）此外，提供网络信息内容传播平台服务的服务提供者、互联网应用程序分发平台、用户均具有义务履行或者配合履行标识义务。

从事人工智能生成合成技术的研发或应用，但没有向中国境内公众提供相关服务的行业组织、企业、教育和科研机构、公共文化机构、专业机构等不适用相关规定。（《办法征求意见稿》第二条）

### 二、提供者的义务

根据服务性质的不同，提供者有义务对相关生成合成内容添加显式或/和隐式标识，《办法征求意见稿》相关规定介绍如下。

#### 显式标识义务

服务提供者提供的生成合成服务属于如下情形的（即《互联网信息服务深度合成管理规定》第十七条第一款），应当按照下列要求对生成合成内容添加显式标识。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。

**文本内容**：提供智能对话、智能写作等模拟自

然人进行文本的生成或者编辑服务的，应在文本的起始、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；
[1]

**人声、仿声内容**：提供合成人声、仿声等语音生成或者显著改变个人身份特征的编辑服务的，应在音频的起始、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；

**人脸内容**：提供人脸生成、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的，应在图片的适当位置添加显著的提示标识；涉及视频的，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；

**虚拟场景**：提供沉浸式拟真场景等生成或者编辑服务的，应当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；

**其他场景**：其他可能导致公众混淆或者误认的生成合成服务场景，应当根据自身应用特点添加具有显著提示效果的显式标识。**（第四条）**

根据《互联网信息服务深度合成管理规定》第十七条，深度合成服务提供者提供上述规定之外的深度合成服务的，应当提供显著标识功能，并提示深度合成服务使用者可以进行显著标识。我们理解这些场景可能包括非人声生成合成、非人脸图像视频生成合成等。

《标准征求意见稿》未额外创设新的义务，而是进一步提供了显式/隐式标识的方法（位置、字型和表达方式等）及示例，可供公司在具体场景中进行参考。例如，文本内容显式标识应采用文字或角标形式，并应同时包含以下要素:1)人工智能要素:包含“人工智能”或“AI”；2)生成合成要素:包含“生成”和/或“合成”。文本内容显式标识应位于以下一个或多个位置:1) 文本的起始位置；2)文本的末尾位置；3)文本的中间适当位置。文本内容显式标识使用的字型和颜色应清晰可辨。（第 5.1 条）如下为文本内容显式标识的一种示例。


图 C.1 位于文本开头的文字内容显式标识

## 隐式标识

根据《互联网信息服务深度合成管理规定》第十六条，深度合成服务提供者对使用其服务生成或者编辑的信息内容，应当采取技术措施添加不影响用户使用的标识，并依照法律、行政法规和国家有关规定保存日志信息。

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。隐式标识存在不同技术实现方式。具体而言，《办法征求意见稿》第五条要求在生成合成内容的文件元数据[2]中添加隐式标识，这些隐式标识可能包括但不限于内容的属性信息、服务提供者名称或编码、内容编号等。鼓励服务提供者在生成合成内容中添加数字水印等形式的隐式标识。

《标准征求意见稿》对于隐式标识应包括的要素规定如下：1）生成合成标签要素：内容的人工智能生成合成属性信息，包括确定、可能、疑似; 2）生成合成服务提供者的名称或编码; 3）生成合成服务提供者对该内容的唯一编号；4）内容传播服务提供者的名称或编码；5）内容传播服务提供者对

[1]《标准征求意见稿》附件包含的图片示例。

[2] 文件元数据是指按照特定编码格式嵌入到文件头部的描述性信息，用于记录文件来源、属性、用途、版权等信息内容。

该内容的唯一编号。附录 E、F 给出了元数据隐式标识应符合的具体格式及示例。

### 用户管理义务

服务提供者应在用户服务协议中详细说明生成合成内容标识的方法、样式和其他相关规范内容，并提示用户仔细阅读并理解相关的标识管理要求。（第八条）如果用户需要服务提供者提供没有添加显式标识的生成合成内容，服务提供者应在用户协议中明确用户的标识义务和使用责任，并留存相关日志不少于六个月。（第九条）

## 三、内容传播平台服务提供者的义务

网络信息内容传播平台服务提供者具有相关义务核验平台传播内容的隐式标识、对内容进行审核并对于确为、可能和疑似生成合成内容进行显著标识及隐式标识。具体义务如下：

**核验隐式标识**：应当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；

**显著提示可能、疑似的生成合成内容**：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，应当采取适当方式在发布内容周边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；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平台服务的服务提供者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，应当采取适当方式在发布内容周边添加显著的提示标识，提醒用户该内容疑似为生成合成内容；

**添加隐式标识**：对于确为、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；

**用户提示**：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。（第六条）

## 四、互联网应用程序分发平台及用户的义务

互联网应用程序分发平台在应用程序上架或上线审核时，应当核验服务提供者是否按要求提供生成合成内容标识功能。（第七条）对于应用分发平台应核验的具体要求及未成功核验相关应用程序未按要求提供标识功能的法律后果，《办法征求意见稿》未进行明确规定。我们理解应用分发平台应至少具有协助监管部门对违规应用下架的义务。

用户向提供网络信息内容传播平台服务的服务提供者上传生成合成内容时，应当主动声明并使用平台提供的标识功能进行标识。（第十条）

## 五、世界范围立法趋势

中国、欧盟、美国在内的各国立法机构均已具有相当程度的共识，需对人工智能生成合成内容的溯源及真实性核验加强监管。欧盟《人工智能法》第 50 条具有与《办法征求意见稿》类似的显式标识及隐式标识规定：对于与自然人直接交互的人工智能系统，人工智能系统提供者应确保人明确告知用户他们正在与人工智能系统进行互动。人工智能系统提供者应确保人工智能系统的输出以机器可读格式标记，并可被检测。部署生成合成图像、音频或视频内容的人工智能系统的部署者应披露该内容已被人工生成或操纵。

美国加利福尼亚州议会于 2024 年 2 月提出《加州数字内容溯源标识法案》尚在立法进程中。此法案要求生成式人工智能提供商在合成内容中嵌入包含来源数据的水印，并开发易于使用的可下载水印解析器，以便用户快速确定内容是否由提供商的系统创建。大型在线平台必须使用标签明确披露平台内容中的来源数据，说明是否为完全合成、部分合成、真实、真实但有轻微修改或不含水印，平台还需使用最新技术检测和标记被移除水印或无水印功能的合成内容。

在产业领域，相关公司也正在开发应用人工智能内容标记和检测技术。例如，Meta 已经在其 AI 工具生成的内容上应用了可见的水印，并计划对其

他公司如 Google、OpenAI、Microsoft、Adobe、Midjourney、Shutterstock 等创建的图像采取同样的处理方式。Google 宣布计划在其产品中实施 C2PA 内容认证技术，以帮助用户区分 AI 生成图像。

## 六、我们的观察

在算法备案、安全评估等流程中，生成合成内容标识义务是否履行已经是监管部门重点审核对象。生成合成内容标识义务的具体规定及国标的出台，为服务提供者提供更为详细的合规指引。此外，内容创作者、网络信息内容传播平台服务提供者以及应用分发平台在信息制作、传播中也均对内容标识负有相关义务。建议相关服务提供者、内容创作者、网络信息内容传播平台服务提供者以及应用分发平台根据现有法规及相关征求意见稿对合规义务的履行进行自查。可以预见，生成合成内容的标识将成为未来人工智能领域执法的重点。

尽管包括中国在内的立法者均认可应加强人工智能生成内容的透明度，但上述义务如何通过元数据、水印等技术方式进行落实仍存在争议，特别是考虑到目前的标签、水印技术仍正发展过程中，相关技术解决方案是否足够有效、具有可互操作性且不易被篡改，用户、内容传播平台及监管部门是否有有效的生成合成内容识别工具等对于上述制度的有效性及各方责任分配将产生较大影响。这些问题有待进一步观察。


董潇　合伙人　电话：86 10 8519 1718　邮箱地址：dongx@junhe.com
郭静荷　律　师　电话：86 10 8553 7947　邮箱地址：guojh@junhe.com
栗晔　电话：86 10 8519 1324　邮箱地址：liye@junhe.com