

TC260-PG-20233A

---

# 网络安全标准实践指南

## —生成式人工智能服务内容标识方法

---

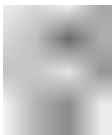
(v1.0-202308)

全国信息安全标准化技术委员会秘书处

2023年08月

本文档可从以下网址获得：

[www.tc260.org.cn/](http://www.tc260.org.cn/)



全国信息安全标准化技术委员会  
NATIONAL INFORMATION SECURITY STANDARDIZATION TECHNICAL COMMITTEE

# 前 言

《网络安全标准实践指南》（以下简称《实践指南》）是全国信息安全标准化技术委员会（以下简称“信安标委”）秘书处组织制定和发布的标准相关技术文件，旨在围绕网络安全法律法规政策、标准、网络安全热点和事件等主题，宣传网络安全相关标准及知识，提供标准化实践指引。



## 声 明

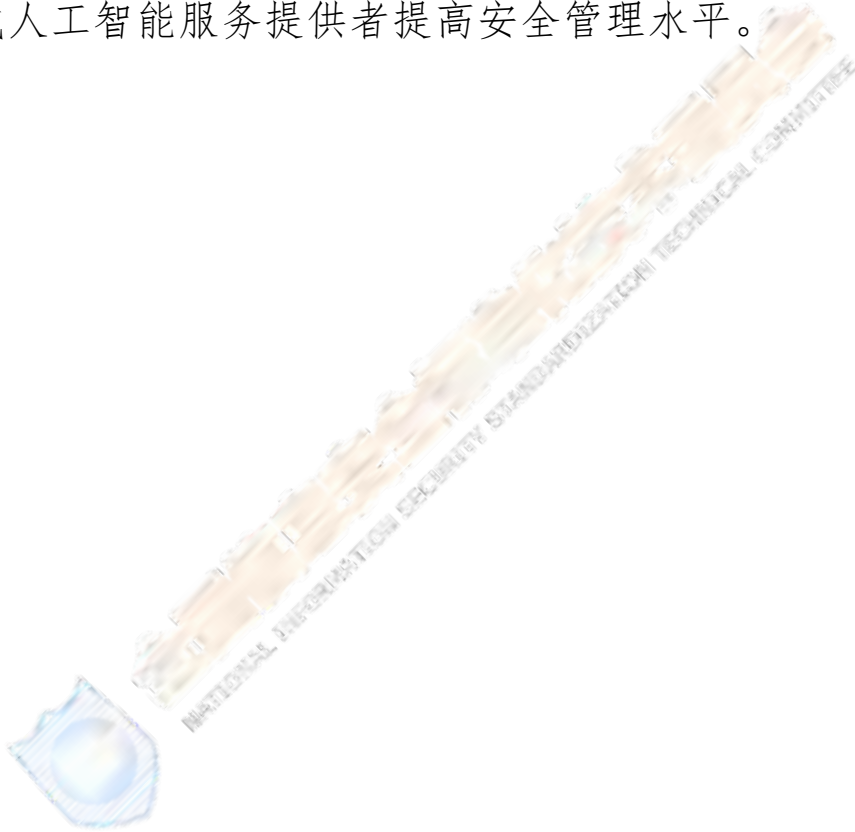
本《实践指南》版权属于信安标委秘书处，未经秘书处书面授权，不得以任何方式抄袭、翻译《实践指南》的任何部分。凡转载或引用本《实践指南》的观点、数据，请注明“来源：全国信息安全标准化技术委员会秘书处”。

## 技术支持单位

本《实践指南》得到中国电子技术标准化研究院、国家计算机网络应急技术处理协调中心、浙江大学、阿里云计算有限公司、科大讯飞股份有限公司、北京抖音信息服务有限公司、北京中关村实验室、北京邮电大学、北京百川智能科技有限公司、华为云计算技术有限公司、北京智谱华章科技有限公司、上海稀宇科技有限公司、上海人工智能创新中心、北京深言科技有限责任公司、北京百度网讯科技有限公司、北京面壁智能科技有限责任公司、中国科学技术大学等单位技术支持。

## 摘 要

为贯彻落实《生成式人工智能服务管理暂行办法》关于对生成内容进行标识的要求，指导生成式人工智能服务提供者等有关单位做好内容标识工作，本文件围绕文本、图片、音频、视频四类生成内容给出了内容标识方法，可用于指导生成式人工智能服务提供者提高安全管理水平。



# 目 录

1 范围 .....	1
2 术语和定义 .....	1
2.1 显式水印标识 .....	1
2.2 隐式水印标识 .....	1
3 标识方式和标识信息 .....	1



# 1 范围

本文件给出了生成式人工智能服务提供者对生成内容进行标识的方法。

本文件适用于生成式人工智能服务提供者利用生成式人工智能技术向公众提供生成文本、图片、音频、视频等内容时对生成内容进行标识。

## 2 术语和定义

### 2.1 显式水印标识

在交互界面内或背景中添加的半透明文字。

注：可通过调整文字图案分布密度、显示参数等使显式水印标识不影响正常使用，但仍可清晰分辨，例如将其透明度设为 90%等。

### 2.2 隐式水印标识

通过修改图片、音频、视频内容添加的，人类无法直接感知、但可通过技术手段从内容中提取的标识。

## 3 标识方式和标识信息

3.1 在人工智能生成内容的显示区域中，应在显示区域下方或使用者输入信息区域下方持续显示提示文字，或在显示区域的背景均匀添加包含提示文字的显式水印标识。提示文字应至少包含“由人工智能生成”或“由 AI 生成”等信息。

3.2 由人工智能生成图片、视频时，应采用在画面中添加提示文字的方式进行标识。提示文字宜处于画面的四角，所占面积应不低于画面的 0.3%或文字高度不低于 20 像素。提示文字内容应至少包含“人工

智能生成”或“AI生成”等信息。视频中由当前服务生成的画面应添加提示，其他画面可不添加提示。

3.3 由人工智能生成图片、音频、视频时，应按以下方式在生成内容中添加隐式水印标识。隐式水印标识中至少包含服务提供者名称，也可包含内容 ID 等其他内容。

注：内容 ID 是服务提供者对生成内容的唯一编号。

- a) 图片的隐式水印标识应通过空域水印或变换域水印的方式实现，含有隐式水印的原始生成图片应满足任意连续 50%以上面积且分辨率大于等于 384×384 的切片均包含完整标识信息；
- b) 视频的隐式水印标识应通过时空域水印或变换域水印的方式实现，含有隐式水印标识的原始生成视频应满足任意连续 5 秒以上视频内容中均包含完整标识信息；
- c) 音频的隐式水印标识应通过时域水印或变换域水印的方式实现，含有隐式水印标识的原始生成音频应满足任意连续 10 秒以上音频内容中均包含完整标识信息；
- d) 服务提供者应具有从该服务所生成内容中提取隐式水印标识的接口或工具。

3.4 由人工智能生成的图片、音频、视频以文件形式输出时，应在文件元数据中添加扩展字段进行标识。扩展字段内容应包含服务提供者名称、内容生成时间、内容 ID 等信息。扩展字段编码应采取以下键值对格式：

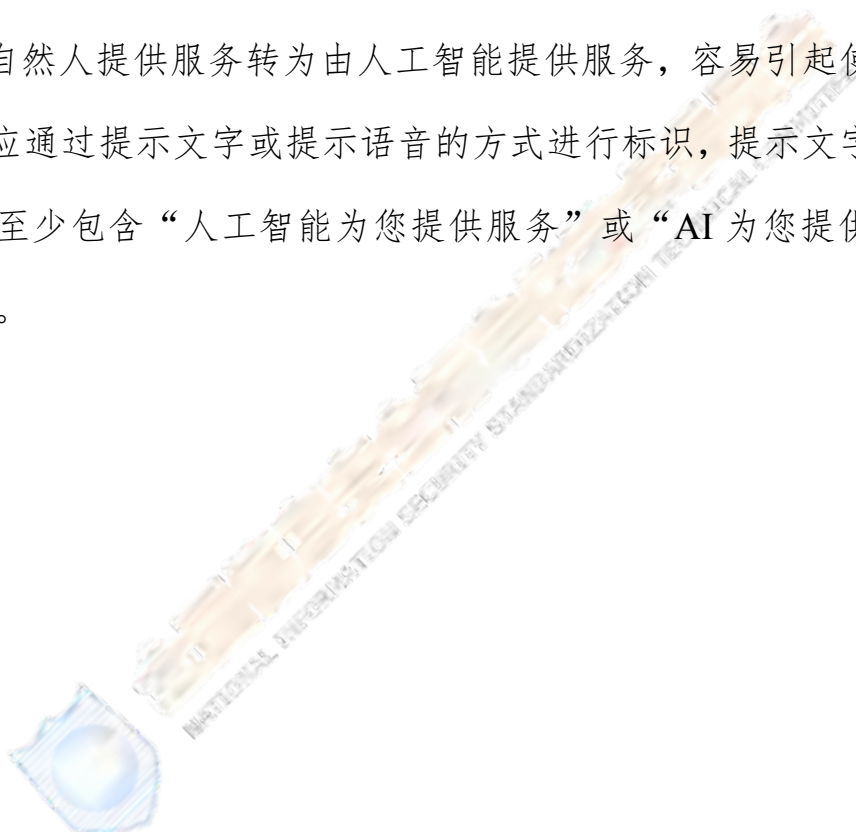
AIGC: {"ServiceProvider": value1, "Time": value2, "ContentID": value3}

注 1: 服务提供者名称 (ServiceProvider), 其值 value1 的类型为字符串, 长度不超过 32 字符;

注 2: 内容生成时间 (Time), 其值 value2 的类型为字符串, 精确到毫秒, 时间格式为 yyyy-MM-dd HH:mm:ss.SSS;

注 3: 内容 ID (ContentID), 其值 value3 的类型为字符串, 不超过 32 个字符。

3.5 由自然人提供服务转为由人工智能提供服务, 容易引起使用者混淆时, 应通过提示文字或提示语音的方式进行标识, 提示文字或提示语音应至少包含“人工智能为您提供服务”或“AI 为您提供服务”等信息。





## 参考文献

- [1] 王建民, 王朝坤, 余志伟. 软件保护技术[M]. 清华大学出版社, 2013.
- [2] 刘瑞祯, 谭铁牛. 数字图像水印研究综述 [J]. 通信学报, 2000(08):39-48.
- [3] 王翌妃, 周杨铭, 钱振兴等. 鲁棒视频水印研究进展[J]. 中国图象图形学报, 2022, 27(01):27-42.
- [4] Guang Hua, Jiwu Huang, Yun Q Shi, Jonathan Goh, and Vrizlynn LL Thing. Twenty Years of Digital Audio Watermarking - A Comprehensive Review[J]. Signal Processing, 2016, 128(Nov.): 222-242.

