ICS 33.030
CCS M 21

团　体　标　准

T/TAF 237—2024

深度合成测评规范　人脸信息保护

Evaluation specification for deep synthesis—
Face information protection

2024-09-02 发布　　2024-09-02 实施

电信终端产业协会　发布

目　　次

前　言

本文件按照GB/T 1.1—2020《标准化工作导则　第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由电信终端产业协会提出并归口。

本文件起草单位：中国信息通信研究院、北京快手科技有限公司、华为终端有限公司、维沃移动通信有限公司、百度在线网络技术（北京）有限公司、荣耀终端有限公司、北京三星通信技术研究有限公司、北京微梦创科网络技术有限公司、OPPO广东移动通信有限公司、郑州信大捷安信息技术股份有限公司。

本文件主要起草人：冯金金、傅山、魏凡星、王嘉义、杨萌科、刘陶、陈鑫爱、王艳红、落红卫、谷晨、李实、赵盈洁、姚一楠、郭建领、王颖华、李辰淑、吴越、任资政、王天、李根、李腾、刘献伦。

引　　言

近年来，随着深度合成技术不断发展，越来越多的组织、个人参与到基于人脸信息的深度合成应用领域。但人脸信息属于敏感个人信息，任其随意使用会带来两方面的危害：一是人脸信息滥采滥用，造成用户权益损害；二是个人难以辨别合成的虚假信息，造成社会舆论负面影响。

因此，为落实相关法律法规要求，指导行业规范使用深度合成服务涉及的人脸信息，提升用户对深度合成人脸信息的感知度，保障用户在个人信息处理活动中的权益，需制定深度合成测评规范 人脸信息保护。

深度合成测评规范　人脸信息保护

1　范围

本文件规定了电信和互联网行业深度合成服务提供者在提供涉及人脸信息的深度合成服务时应满足的明示同意要求、显示标识要求、隐式标识要求和测评流程等内容。

本文件适用于深度合成服务提供者进行自评估，同时也适用于第三方评估机构开展评估工作。

2　规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 35273-2022　信息安全技术　个人信息安全规范

GB/T 42574-2023　信息安全技术　个人信息处理中告知和同意的实施指南

3　术语和定义

3.1

深度合成技术　deep synthesis technology

利用深度学习、虚拟现实等生成合成类算法制作文本、图像、音频、视频、虚拟场景等信息的技术。

3.2

深度合成服务　deep synthesis service

利用深度学习技术，合成出类似于真实世界的图片、视频、音频等数字化内容的服务。

3.3

深度合成服务提供者　deep synthesis service provider

提供深度合成服务的组织、个人。

3.4

深度合成服务使用者　deep synthesis service user

利用深度合成服务生成图像、视频等数字化内容的组织、个人。

3.5

人脸信息　face information

对特定自然人的人脸特征进行技术处理得到的、能够单独或者与其他信息结合识别该特定自然人身份的个人信息。本文件指处于任何处理阶段的人脸样本、人脸参考、人脸特征项或人脸特性的通称，不包括匿名化处理后的信息。

3.6

敏感个人信息 sensitive personal information

一旦泄露或者非法使用，容易导致自然人的人格尊严受到侵害或者人身、财产安全受到危害的个人信息，包括生物识别、宗教信仰、特定身份、医疗健康、金融账户、行踪轨迹等信息，以及不满十四周岁未成年人的个人信息。

[来源：GB/T 35273-2022，有修改]

3.7

明示同意 explicit consent

个人通过书面、口头等方式主动做出声明，或者自主作出肯定性动作，对其个人信息进行处理作出明确授权的行为。

注：肯定性动作包括个人主动勾选、主动点击“同意”“注册”“发送”“拨打”，主动填写或提供等。

[来源：GB/T 42574-2023，3.6]

3.8

显式标识 explicit identifier

提示生成内容深度合成情况的标识，易于用户（包括深度合成服务使用者、观看者等）感知。

3.9

隐式标识 implicit identifier

记录深度合成情况，嵌入生成内容中且用户（包括深度合成服务使用者、观看者等）不可直接感知的标识。

4 缩略语

下列缩略语适用于本文件。

APP：移动互联网应用程序（Mobile Internet Application）

SDK：软件开发包（Software Development Kit）

5 深度合成人脸信息涉及应用场景

5.1 概述

深度合成人脸信息，是指利用深度合成技术处理合成人脸信息过程中涉及到的特定自然人的人脸信息。按照合成的程度，深度合成人脸信息涉及的应用场景主要包括三类：人脸生成、人脸替换、人脸操控。

5.2 人脸生成

利用深度合成技术，基于特定自然人的人脸信息，创建全新的人脸图像、视频。生成的人脸具有非特定自然人的特征，逼真度高，多用作广告制作、虚拟头像生成等场景。

5.3 人脸替换

人脸替换通常有两种模式，一种是将特定自然人的人脸信息直接替换到特定的图像或视频模板中，实现抠图的效果，例如利用已有图片制作证件照的功能；另一种是将特定自然人的脸部特征数据与特定目标模板深度融合，生成兼具两者特征的人脸信息，例如仿古特效功能。

5.4 人脸操控

人脸操控主要是指利用深度合成技术改变特定自然人的人脸特征，包括嘴部、眉毛、眼睛和头部的倾斜角度等，从而呈现“说没说过的话”的效果。

5.5 典型工作场景

经深度合成服务生成的人脸信息，可能导致公众混淆或者误认的，应在人脸信息内容合理位置、区域进行显式标识和隐式标识。

深度合成人脸信息典型场景涉及深度合成服务提供者、深度合成服务使用者、观看者及监管方等角色，四者之间与深度合成服务、深度合成人脸信息图像视频、显式标识、隐式标识之间的关系具体如图1所示。

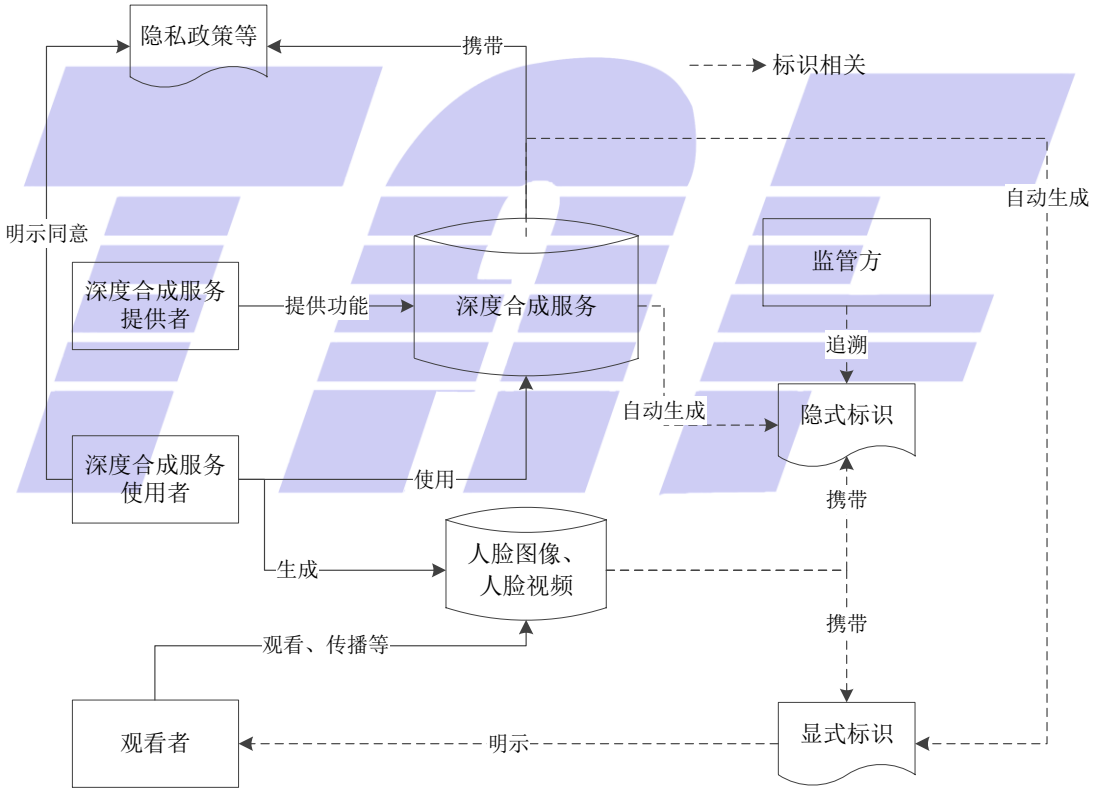

图1 典型工作场景

6 明示同意

6.1 明示的基本原则

深度合成服务提供者在实施明示时需考虑以下原则：

a) 合理性原则：明示内容应与当前提供的服务一致；

b) 正当性原则：应当在个人同意的范围内合理收集；

c) 充分性原则：履行充分告知的义务，明示内容应清晰准确，保障深度合成服务使用者对个人信息收集使用的知情权；

d) 友好性原则：明示形式应具有友好的界面，便于深度合成服务使用者理解和选择。

6.2 明示的方式

深度合成服务提供者应在以下场景实施明示：

a) 首次明示：在使用深度合成人脸信息服务前，应采用公开且便于查阅的方式，向深度合成服务使用者全面阐述深度合成服务使用情况，例如可采用制定、展示个人信息保护政策（或被称为“隐私政策”、“隐私协议”、“隐私权政策”等）或服务协议的形式进行明示；

b) 提示明示：在使用深度合成人脸信息服务过程中必要时（如需向深度合成服务使用者明示是否取得人脸信息个人主体单独授权同意时），需采用深度合成服务使用者不可绕过的方式向其明示相关信息，包括采用弹窗、专门页面、单独步骤或显著位置等方式，即时向深度合成服务使用者直接展示关键内容。

6.3 明示的内容

深度合成服务提供者应明示的内容包括：

a) 首次明示：

1) 应向深度合成服务使用者明示处理规则、方式、范围等，如，应对如何使用来自服务使用者上传的人脸信息（是否存储人脸信息、人脸信息是否会用于其他服务等）进行明示。

2) 应向深度合成服务使用者明示涉及敏感个人信息；

3) 应向深度合成服务使用者明示对个人权益的影响；

4) 应向深度合成服务使用者明示不应采用技术手段删除、篡改、隐匿标识。

b) 提示明示：在使用深度合成人脸信息服务过程中必要时，应向深度合成服务使用者明示需依法告知被编辑人脸信息的个人主体，并提示深度合成服务使用者需取得个人主体的单独同意。

注1：单独同意是指使用深度合成服务处理人脸信息时，深度合成服务使用者应确认已经单独征得人脸信息个人主体的授权同意。深度合成服务提供者应提供按钮等方式，以供深度合成服务使用者确认已取得授权。

注2：对编辑前的人脸信息使用应仅限当次服务范围内，不宜利用深度合成服务存储编辑前的人脸信息。

7 显式标识

显式标识应符合以下要求：

a) 已确认深度合成服务使用者知悉需取得个人主体单独同意后，经深度合成服务生成的可能导致公众混淆或者误认的人脸图片或视频时，应及时在生成的内容上添加显式标识，用于向观看者明示该图片或视频经深度合成服务处理；

注：若生成人脸图片或视频的存储和发布阶段分开，则应在存储时添加显示标识；否则，可在发布阶段再添加标识。

b) 显式标识应清晰，具有可读性，易于观看者理解和观看。

注：显式标识使用的文字颜色应与图像或视频内容有一定的对比度，文字字号不应过小，不得采用连体字等不便于观看的字体。显式标识位置应合理。标识内容应完整。显式标识不得采用过度简写，如，仅标注“AI”。

8 隐式标识

深度合成服务提供者应提供隐式标识能力，包括：

a) 应对利用深度合成技术生成涉及人脸信息的图片、视频添加隐式标识；

b） 隐式标识内容应至少包括深度合成服务提供者唯一标识、内容唯一标识等
c） 应具备基于隐式标识溯源的能力，
d） 隐式标识应对利用深度合成技术生成涉及人脸信息的图片、视频的展示、编辑等功能无影响；
e） 对图像或视频画面进行少量修改时，隐式标识应具备一定的鲁棒性，包括但不限于以下情景：
1） 通过裁剪、缩放、旋转、拼接等方式修改画面的尺寸或形状；
2） 调整亮度、对比度和色彩平衡等属性参数；
3） 修改编码格式。

9 测评流程

9.1 总体要求

深度合成人脸信息测评流程应符合图2的规定。包括确定测评目标、选择测评指标、制定测评计划、实施测评任务及得出测评结论等活动。

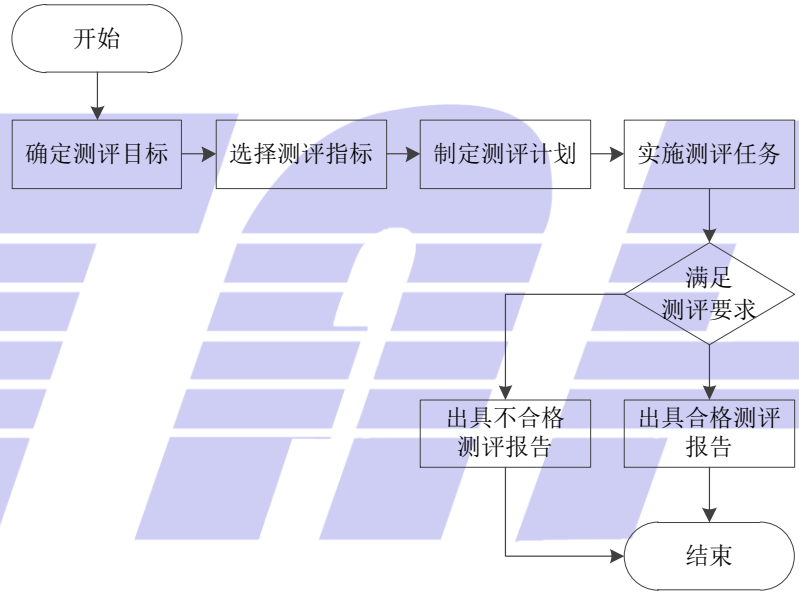

图2 评估流程图

9.2 测评方与被测评方

应考虑以下方面，确定测评方和被测评方：
a） 测评方可为深度合成服务提供者、开发者和运营者，也可为第三方实验室；
b） 被测评方可为APP、SDK、终端，或者产品中的深度合成服务。

9.3 选择测评指标

测评方应考虑以下方面，确定测评指标：
a） 测评方根据被测评方提供的技术说明文档、被测评样品等，确定涉及的深度合成服务类型；
b） 选择合适的测评规范标准；
c） 根据测评目标选择合适的测评指标。

9.4 制定测评计划

测评方应考虑以下方面，制定测评计划：

a） 测评计划应至少包括测评对象和范围、测评依据、测评环境、测评工具、时间进度安排等；

b） 测评计划中应明确测评通过/不通过的判断标准。

9.5 实施测评任务

测评方应考虑以下方面，实施测评任务：

a） 测评方应根据测评目标，本着公平、公正、公开原则开展测评工作；

b） 根据对应的测评规范标准开展实施测评活动；

c） 测评任务可顺序开展也可并行开展，无完整的顺序关系。

9.6 测评结论

测评方应考虑以下方面，给出测评结论：

a） 针对深度合成服务进行测评，明示同意、内容标识等均未见不符合项方可通过测评，否则未通过测评；

b） 在测评报告中，应包含被测评方基本信息、测评环境、测评指标、测评步骤和每一项测评的结果；

c） 如测评结论包含未通过项，则测评报告中应包含未通过原因的具体描述。

参 考 文 献

[1] 中华人民共和国个人信息保护法（2021 年 8 月 20 日第十三届全国人民代表大会常务委员会第三十次会议通过）

[2] 互联网信息服务算法推荐管理规定（2021 年 11 月 16 日国家互联网信息办公室 2021 年第 20 次室务会议审议通过）

[3] 互联网信息服务深度合成管理规定

[4] 工业和信息化部关于进一步提升移动互联网应用服务能力的通知

电信终端产业协会团体标准

深度合成测评规范　人脸信息保护

T/TAF 237—2024

*

电信终端产业协会印发
地址：北京市西城区新街口外大街 28 号
电话：010-82052809
电子版发行网址：www.taf.org.cn