ICS 33.030
CCS M.21

团　　体　　标　　准

T/CCSA 792–2026　　T/CAAAD 043–2026

互联网广告 生成式人工智能创意素材生产平台能力要求

Internet advertisement—— Generative artificial intelligence creative materials production platform capacity requirements

2026–03–02 发布　　2026–06–01 实施

中国通信标准化协会　　中国广告协会　　发布

版权声明

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》的规则起草。

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。

本文件由中国通信标准化协会和中国广告协会共同提出，并分别归口。

本文件起草单位：北京巨量引擎网络技术有限公司、中国信息通信研究院、北京快手科技有限公司、中国传媒大学、北京沃东天骏信息技术有限公司、利欧集团数字科技有限公司、秒针信息技术有限公司。

本文件主要起草人：张贝贝、侯佳敏、杨正军、朱岩、杜蕾、马月、朱小荔、赵乃萱、陈琳、马澈、周楠、沈雅姣、任露、张清芳、刘帅、谷晨、张泽华、吕晶晶、梁悦然、周崧骏、胡春磊、孙方超、何敏。

引　言

随着算法和模型的飞速进步，人工智能领域也进入了新的发展阶段，在广告营销领域的应用也在不断探索拓展。其中，基于大模型的广告营销创意素材平台，已经成为广告行业AI应用最为重要的方向之一。广告营销领域是合规重地，素材生成须考虑广告领域的多重合规要求。本文件针对生成式人工智能营销创意素材平台业务场景，对于此类平台的模型训练数据、模型能力要求、素材审核（含广告合规、内容安全等多重标准）、生成合成内容标识、生成合成内容的风险识别与处置等多环节明确技术能力、合规等多个维度要求，推进AI技术在互联网广告营销创意素材生成场景的应用。

互联网广告 生成式人工智能创意素材生产平台能力要求

1 范围

本文件规定了互联网广告生成式人工智能营销创意素材平台的模型训练数据、模型选取、素材审核（含广告合规、内容安全等多重标准）、生成合成内容标识、生成合成内容的风险识别与处置等环节的技术及能力要求。

本文件适用于指导生成式人工智能营销创意素材生产平台运营者参考使用。

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 45438-2025 网络安全技术 人工智能生成合成内容标识方法

3 术语和定义

下列术语和定义适用于本文件。

3.1

生成式人工智能服务 generative artificial intelligence service

基于生成式人工智能技术，能够根据使用者的提示生成文本、图片、音频、视频等内容的服务。

3.2

互联网广告生成式人工智能营销创意素材生产平台 generative AI marketing creative production platform for internet advertising

基于生成式人工智能技术，生产文本、图片、音频、视频等内容，应用于互联网广告场景的技术服务提供者，又称为“互联网广告生成式人工智能营销创意素材服务提供者”，以下简称“服务提供者”或“平台”。

3.3

互联网广告生成式人工智能营销创意素材服务使用者 service user of generative AI marketing creative production platform for internet advertising

使用互联网广告生成式人工智能营销创意素材生产平台的技术服务，在平台注册账户，发起创意素材生成指令、完成创意素材制作的组织或个人，以下简称“服务使用者”或“用户”。

3.4

训练数据 training data

所有用于生成式人工智能模型预训练、优化训练等模型训练过程中的各类数据。

3.5

基础模型 foundation model

使用大量数据进行训练，用于普适性目标、可优化适配多种下游任务的深度神经网络模型。

3.6

第三方模型 third party model

互联网广告生成式人工智能营销创意素材服务提供者可以直接选取使用第三方模型服务生成素材，或者在第三方基础模型上精调生成自己的模型用于素材生成。

3.7

显式标识 explicit label

显式标识是指在人工智能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。

3.8

隐式标识 implicit label

隐式标识是指采取技术措施在人工智能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。

4　基于生成式人工智能生成广告营销创意素材平台能力的基本原则

4.1　概述

互联网广告生成式人工智能营销创意素材生产平台服务用户生产和创作广告营销创意素材时，应当基于产业发展需求与现行法律法规所组成的合规体系，遵循一系列基本原则，包括技术与安全平衡发展原则、模型算法合规原则、数据安全原则、内容安全原则、广告法合规原则及知识产权保护原则，以此实现平台技术进步与承担社会责任相统一。

4.2　技术与安全平衡发展

平台利用生成式人工智能技术提升广告营销创意素材的生产效率和质量，应当关注模型系统安全、数据安全、内容安全，建立安全防护措施，防范模型发生被外部攻击、数据泄露、输出违法有害信息等问题，实现广告营销创意素材生成服务的稳定性、鲁棒性、准确性和可靠性。

4.3　模型算法合规

模型与算法合规是生成符合法律规定和商业伦理标准的广告营销创意素材的前提，生成式人工智能所依赖的模型和算法应当严格遵循我国现行法律法规和行业标准，实现模型和算法的公平公正性，防范产生歧视或误导性的结果。

4.4　数据安全

平台在广告营销创意素材生成过程中，需对大量数据进行收集、传输、存储和使用，应当实施相应的数据保护措施，保护用户隐私和企业商业秘密的安全；通过建立健全的平台数据安全管理体系，防止数据泄露、篡改或滥用，确保数据的完整性和安全性。

4.5　内容安全

平台应当从模型设计、安全防御机制、训练数据筛选过滤等方面，提高模型质量，降低生成违法、不良信息的风险；建立机审与人审相结合的审核能力，对输入输出内容进行内容安全审核；对用户进行积极引导，并建立有效监督机制、投诉应对处理机制，发现不良信息及时处理，提高生成内容的合规性和质量。

4.6　广告合规

在平台提供广告投放服务且服务使用者将生成素材用于广告投放时，平台应当按照广告法的法律法规要求，审核生成内容的真实性、准确性和合法性，避免发生虚假宣传、欺骗和误导消费者的行为。

4.7　权利保护

平台应积极引导用户在尊重知识产权的前提下使用相关服务，并建立便捷有效的知识产权及其他权益侵权的投诉途径、投诉机制和处理流程，积极响应权利人的侵害知识产权及其他权益的投诉，及时采取处理措施。

5　模型训练数据要求

5.1　训练数据收集要求

5.1.1　数据来源安全

平台开展训练数据收集前，应当对数据的来源进行审核、安全评估。

5.1.2　训练数据来源记录

平台应当对不同来源的训练数据进行核验、记录：

a) 使用他人开源的数据集的，应当开展开源合规评估；
b) 使用商业交易或合作获取的外部数据，应要求交易方或合作方对数据的来源、质量、合法性、安全性等进行承诺；
c) 使用其他来源数据的，应当保留数据来源记录。

5.1.3　训练数据质量

平台应当采取有效措施提高训练数据质量，增强训练数据的真实性、准确性、客观性、多样性；通过搜集不同来源、不同领域的数据，实现训练数据集的有效广泛覆盖，避免模型的过拟合或欠拟合问题；建立对违法有害数据内容的过滤机制，防范模型错误决策或生成不公平结果。

5.2　训练数据处理要求

5.2.1　内容安全风险识别

平台应当通过关键词、分类模型、人工抽检等方式，对训练数据中的内容安全风险进行充分识别、过滤。内容安全风险审核识别要求可参照本文件第6.2条。

5.2.2 个人信息处理

平台个人信息的处理应满足以下要求：

a) 训练数据在标注、训练使用过程中，不得以识别特定用户个人为目的进行使用，不得损害个人信息主体的利益。如须以识别特定用户个人为目的进行使用，须取得用户的授权同意；
b) 平台应根据“最小够用原则”进行训练数据过滤处理，宜采用数据识别脱敏能力、数据合成技术、差分隐私能力等技术方法，在保证数据可用性的前提下，对于敏感个人信息进行处理。

5.2.3 数据清洗

平台宜通过数据去重、数据聚类、排列组合、数据杂质去除、除错和一致化等方式，提升数据的真实性、准确性。

5.2.4 数据标注

平台数据标注应满足以下要求：

a) 平台应当制定清晰的标注规则，标注规则应包括标注目标、数据格式、标注方法、质量控制等内容；
b) 对标注人员开展安全培训、考核，并为标注人员执行标注任务预留充足、合理的标注时间，形成标准化、规范化的操作规程；
c) 平台应当制定及执行标注质检方案，通过多人验证、埋题验证和机器验证等方式进行质量控制，并根据模型训练、测试及应用结果不断改进数据标注技术及流程；
d) 平台应当提供安全的标注环境，对标注数据的访问控制、标注操作等进行权限管理；
e) 平台应当根据模型训练、测试及生成结果，持续改进、完善数据标注技术与流程。

5.3 训练数据安全要求

平台对训练数据的存储、访问、流转等管理应当符合以下要求，防范数据泄露、数据滥用、数据丢失等风险：

a) 应当根据数据的敏感性及机密性对训练数据进行分类分级，以匹配适当的数据安全管控措施；
b) 应当对训练数据中包含的敏感个人信息或重要业务数据采取加密存储、访问权限控制等安全管控措施；
c) 应当对训练数据所涉应用系统从客户端至服务端之间的数据传输进行加密处理，采用HTTPS 防止链路被嗅探、流量镜像及传输过程中被篡改，其中敏感个人信息数据的传输应采用字段级加密手段；
d) 应当采取访问控制措施，确保经过审批授权的人员才可以访问不同等级的训练数据；
e) 应当通过系统平台等方式记录训练数据的用途、分类、权限等信息，以便对数据进行跟踪和监控；
f) 应当具有训练数据收集及准备阶段关键活动日志，基于日志可对关键操作进行审计与追溯；
g) 应定期对训练数据进行复制和备份，并建立数据备份恢复的定期检查机制，以保证训练数据的可用性。

6 生成式人工智能广告营销创意素材平台的模型能力要求

6.1 模型安全

平台模型安全应符合以下要求:

a) 平台应当对模型系统所采用的芯片、软件、组件、 算力等方面的供应能力安全开展评估，增强供应的持续性、稳定性与可靠性；

b) 平台应通过构建入侵防控体系、加强系统管理与监控等技术方法，具备能力对恶意攻击、非法访问等进行识别，包括但不限于识别提示词攻击、对抗攻击等，并增强模型防护能力，降低数据泄漏及违法不良信息输出、传播等风险；

c) 平台应采取技术措施和防护能力，防范外部攻击者通过成员推理攻击、梯度攻击等方式还原模型参数和训练数据，从而破坏模型拥有者的知识产权或潜在商业优势；

d) 平台应通过线上异常流量检测、行为检测等措施，识别外部攻击者投毒反馈、垃圾提问、接口攻击等侵占模型服务资源、进行负向和恶意反馈投毒等，并采取流量反爬、接口防刷、人机识别、频控限制等方式降低风险；

e) 平台应当对服务提供过程中发现的模型安全问题，及时通过增强学习、优化训练等措施优化模型，提升模型与服务的持续运营能力；

f) 如涉及接入第三方模型为服务使用者提供生成式人工智能广告营销创意素材生成服务，或基于第三方基础模型进行二次训练、精调，平台应当采取如下措施保障模型安全：

   1) 从内容安全、提示词攻击、隐私保护、商业敏感信息保护、模型幻觉防护等维度对所接入的第三方模型进行安全测评或对专业机构出具的安全测评报告进行审核；

   2) 与第三方模型提供方签署相关协议，约定模型合规、输入输出数据处理、内容安全管控策略和要求、精调模型知识产权归属等内容。如第三方模型为开源模型，应当遵守模型开源许可证及开源协议的约定；

   3) 将敏感个人信息、重要业务数据输入第三方模型前，应通过数据识别脱敏、数据合成等技术措施，在保证数据可用性的前提下，降低数据泄露风险；

   4) 输入三方模型训练的数据应参考 5.2 训练数据处理要求进行处理后使用，避免有害数据对于模型应用结果的影响。

6.2 模型生成内容的准确性、可靠性

平台应基于服务使用者的意图与需求，采取多种措施，如调整训练数据量、调整标注规则、加大数据清洗、引入更多广告营销创意领域专业知识、优化模型结构、开展事实性检测、引入人工审核和反馈机制进行监督与纠偏等，以提高模型生成内容的准确性及可靠性，减少误导性内容、错误内容、超出用户输入意图的内容。对于明显偏激以及明显诱导生成违法不良内容的问题，应拒绝回答或返回兜底答案。

6.3 模型算法管理

平台模型算法管理应符合以下要求:

a) 平台在算法设计与模型训练、推理过程中，应对算法机制机理、模型、数据和模型应用结果开展审核、评估和验证；

b) 平台不得利用算法、数据、模型等优势，实施垄断和不正当竞争行为；

c) 平台应当采取有效措施，防范模型算法生成营销素材产生民族、信仰、国别、地域、性别、年龄、职业、健康等歧视。

7　生成式人工智能广告营销创意素材内容审核要求

7.1　审核机制

平台应建立如下审核机制保障内容安全：

a) 平台应当组建内容安全审核团队，针对模型生成内容开展审核能力建设，防范生成违法违规或有害内容；
b) 平台应当通过广泛的拦截词建设、过滤技术或其他方式，对输入内容开展安全审核，对具有严重攻击性或其他风险的问题进行拦截和干预，以避免产生不良影响；
c) 平台应当根据相关法律法规规定适时调整内容安全及广告内容相关的审核规则。

7.2　通用内容审核要求

平台在提供内容生成过程中应当采取措施防范、制止违法内容，及时识别、筛除以下内容：

a) 反对宪法所确定的基本原则，煽动抗拒或者破坏宪法、法律、行政法规实施；
b) 危害国家安全，颠覆国家政权，破坏国家统一；
c) 宣扬恐怖主义、极端主义或者煽动实施恐怖活动、极端主义活动；
d) 煽动民族仇恨、民族歧视，有损民族团结；
e) 破坏国家宗教政策，宣扬邪教和封建迷信的；
f) 散布淫秽、色情、赌博、暴力、凶杀、恐怖或者教唆犯罪的；
g) 侵害他人权益，侮辱或者诽谤他人，侵害他人名誉、隐私和其他合法权益；
h) 法律、法规和国家规定禁止的其他内容。

7.3　广告内容审核要求

平台应对广告内容进行审核，并满足以下要求：

a) 平台仅为平台用户提供营销创意素材生产技术服务，属于互联网信息服务提供者，应及时识别、筛除 7.2 条相关内容；
b) 平台宜采取合理措施防范违法广告，明知应知利用平台技术生产内容构成违法广告时及时制止；
c) 平台根据相关法律法规规定构成广告发布者时，应遵循以下广告内容准则审核生产内容。
    1) 广告内容不得涉及以下情形：使用或者变相使用中华人民共和国的国旗、国歌、国徽，军旗、军歌、军徽；使用或者变相使用国家机关、国家机关工作人员的名义或者形象；使用“国家级”、“最高级”、“最佳”等用语；危害人身、财产安全，泄露个人隐私；妨碍社会公共秩序或者违背社会良好风尚；妨碍环境、自然资源或者文化遗产保护；法律、行政法规规定禁止生产、销售、发布广告的产品或者提供的服务；
    2) 广告中涉及专利产品或者专利方法的应当标明专利号和专利种类，不得涉及虚假或无效专利；
    3) 广告不得含有虚假或者引人误解的内容，不得欺骗、误导消费者；
    4) 医疗、药品、医疗器械、农药、兽药、饲料和饲料添加剂、保健食品、酒类、教育培训、招商投资类、房地产、农作物种子、林木种子、草种子、种畜禽、水产苗种和种养殖等特殊行业广告在广告内容或广告审查资质上有特殊要求，应按照广告法要求完成审核。

8　生成合成内容标识要求

8.1 标识方法

平台作为生成式人工智能服务提供者，遵守《生成式人工智能服务管理暂行办法》、《互联网信息服务深度合成管理规定》、《人工智能生成合成内容标识办法》等法律、行政法规和部门规章，对可能导致公众混淆或者误认的文本、图片、音频、视频、虚拟场景等生成合成内容进行标识。生成合成内容标识包含显式标识方法及隐式标识方法两类，具体标识方式应遵守《人工智能生成合成内容标识办法》、《网络安全技术 人工智能生成合成内容标识方法》相关规定予以执行。

8.2 标识管理

平台应在平台服务协议、平台规则中明确说明生成合成内容标识的方法、样式等内容，提示用户仔细阅读并理解平台相关的标识管理要求。

若用户主动申请平台提供不添加显式标识的生成合成内容的，平台可以通过平台服务协议、产品页面提示用户的标识义务及使用责任后，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。

9 生成合成内容的风险识别与处置机制

9.1 用户身份认证

平台应当对申请注册的用户进行基于移动电话号码、身份证件号码或者统一社会信用代码等方式的真实身份信息认证。用户不提供真实身份信息，或者冒用组织机构、他人身份信息进行虚假注册的，平台不得为其提供相关服务。

9.2 平台服务规则

平台在向用户提供服务前，应当与用户签署平台服务协议，明确约定平台提供服务过程中双方的权利、义务、责任；平台应当制定和公开平台管理规则等，依法依约履行管理责任，一旦发现用户违反平台协议、规则的情形，平台应当及时采取处理措施。

9.3 主动识别与违规处置

9.3.1 主动识别

平台应对生成内容主动进行识别，具体要求如下：

a) 用户输入行为检测：平台应当采用关键词、分类模型等方式，对于用户输入信息进行检测，对于识别到的用户输入违法不良信息或明显诱导生成违法不良信息，采取暂停提供服务等处置措施；
b) 建立内容审核机制：平台应当通过第 7.1 条的内容审核机制，主动识别模型风险，防范模型生成违法或有害信息并向用户提供；
c) 建立有效的监督机制：平台应当审查模型输出内容，并根据反馈和发现进行调整和改进，以不断提高模型的质量和符合性；
d) 建立业务风控机制：平台应当通过分析用户行为模式，识别出与正常行为模式显著不同的行为，识别出风险账号或行为，及时采取干预机制。

9.3.2 违规处置

平台应对识别出的违规内容或违规人员进行处置：

a) 平台发现违法和不良信息或用户利用生成式人工智能营销创意素材生成技术服务从事违法活动的，应当依法采取处置措施；
b) 平台违规处置措施包括但不限于警示、公示、限制功能、暂停或终止提供服务等处置措施。

9.4 投诉举报处理机制

平台应设置便捷的用户申诉和公众投诉、举报入口，公布处理流程，及时受理、处理和反馈处理结果。

9.5 风险提示

因生成式人工智能的产品功能仍处于发展阶段，平台宜通过用户协议、平台规则、产品页面提示等方式，明确告知用户平台技术服务的局限性及使用生成内容的知识产权风险，提示用户生成内容仅供参考，用户应自行对其加以审慎识别和判断。

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] 中华人民共和国网络安全法（2016 年 11 月 7 日第十二届全国人民代表大会常务委员会第二十四次会议通过）

[3] 生成式人工智能服务管理暂行办法（2023 年 5 月 23 日国家互联网信息办公室国家发展和改革委员会、教育部、科学技术部、工业和信息化部、公安部、国家广播电视总局发布）

[4] 人工智能生成合成内容标识办法（2025 年 3 月 14 日国家互联网信息办公室、工业和信息化部、公安部、国家广播电视总局发布）

[5] 人工智能安全治理框架（2024 年 9 月 9 日，全国网络安全标准化技术委员会发布）