文化传媒法律资讯

Entertainment Law Update

2025 年 02 月

• 第六十一期 •

上海市律师协会
文化传媒专业委员会

东方律师
SHANGHAI BAR ASSOCIATION

文化传媒法律资讯

第61期（2025年02月）

Entertainment Law Update

文传热点

案例分析

法规速递

委员会宗旨

在全面推进依法治国和积极发展文化事业的背景下，在文化事业与文化产业领域内，打造一支专业的律师队伍，搭建一个有效的联动平台，助力一批优秀的创新项目，献策一套先进的法律法规。通过在图书期刊、影视音像产品、广播电视、娱乐演出、文物艺术品及互联网等领域的理论研究、热点探讨、走访调研、献策献言，为上海乃至全国文化传媒行业的规范化、产业化、金融化、国际化助力。

文传热点

Hot Topic

工信部：向首批 13 家外资企业发放经营试点批复

来源：光明网

近日，工业和信息化部向北京、上海、海南、深圳四地 13 家外资企业发放增值电信业务经营试点批复，此次获得经营试点批复的 13 家外商投资企业，其母公司多为知名跨国企业，相关企业按照批复内容，可开展互联网接入、信息服务等增值电信业务。工业和信息化部标识，将进一步全面深化改革，积极营造市场化、法治化、国际化营商环境，持续推动电信领域对外开放，支持更多符合条件的外资企业加入增值电信业务扩大开放经营试点。

创纪录！《哪吒之魔童闹海》位列全球影史票房榜第 7 名

来源：湖南日报

据网络数据平台，截至 3 月 1 日 0 时 26 分，《哪吒之魔童闹海》累计票房（含海外及预售票房）突破 141.60 亿元，超《蜘蛛侠：英雄无归》票房成绩，成为全球影史票房榜第 7 名。

自大年初一上映以来，《哪吒之魔童闹海》一举打破多项电影影史纪录，成为中国影史首部票房破 100 亿元的电影，并登顶全球动画电影票房榜。

如今，脚踏风火轮的“吒儿”冲出国门，在海外排片与票房方面依然表现亮眼。据《哪吒之魔童闹海》海外发行商华人影业消息，截至 2 月 27 日，《哪吒之魔童闹海》北美影院排片破 1000 间，打破近 20 年中国内地影片排片纪录，登顶近 20 年华语电影票房冠军（不含合拍片）；新西兰票房超越《美国队长 4》，荣登新西兰影史华语电影票房冠军。

首届中国电视剧制作产业大会在深圳开幕

来源：南方网

2 月 20 日，由中国电视剧制作产业协会主办、深圳广播电影电视集团协办的首届中国电视剧制作产业大会暨第十届中国（深圳）国际电视剧节目交易会在深圳开幕。行业代表齐聚一堂，围绕“迎接‘剧’变，变与不变”之主题，共同探讨中国剧集行业的变革与发展。国家广播电视总局党组成员、副局长刘建国出席并致辞。

本届大会以“精品引领，向新向质”为主题，汇聚一线电视台、头部平台、行业核心创作力量及投融资等全产业链要素，从行业实际需求出发，聚焦剧集市场新趋势，重点探讨播出机构需求与内容策划、长短剧融合发展等议题，共谋中国剧集产业高质量发展。

开幕式上发布了由中国电视剧制作产业协会等单位编写的《2024 中国剧集产业年度发展报告》，报告通过详实的市场数据，直面行业困境与成果，力图为剧集产业找寻新的航向。报告显示，优质 IP 持续发力，现实题材展现强大生命力；悬疑题材井喷式增长，传统题材稳中求新；电视剧续集热度不减，原班人马回归成破题之法；“类型+”与技术创新，突破谍战剧固化模式；平台深度参与内容制作，微短剧竞争加剧；女性仍为追剧主力军，男频剧破圈层现象显著；短视频平台赋能剧宣，微信视

频号影响力上升。

大会除主旨演讲外，还策划了平台年度趋势与市场需求发布活动、“金树林·绽放之夜”年度盛典及多场产业专题对话。

市监总局对谷歌涉嫌违反反垄断法依法开展立案调查

来源：国家市场监督管理总局

2 月 4 日，国家市场监督管理总局发布公告称，因谷歌公司涉嫌违反《中华人民共和国反垄断法》，市场监管总局依法对谷歌公司开展立案调查。

中央网信办发布 2025 年“清朗”系列专项行动整治重点

来源：中国网信网

近年来，中央网信办持续部署开展“清朗”系列专项行动，集中时间、集中力量打击网上各类乱象问题，从严处置违规平台和账号，取得积极成效，形成有力震慑。2025 年，“清朗”系列专项行动将进一步巩固提升治理成效，聚焦人民群众反映强烈的突出问题，在破解难点瓶颈方面下功夫，强化源头管理和基础管理；在治理创新方面下功夫，针对性细化每个专项打法举措；在维护网民权益方面下功夫，严厉打击各类侵权违法行为，营造更加清朗有序的网络环境。

重点整治任务主要包括：一是整治春节网络环境，集中打击挑起极端对立、炮制不实信息、宣扬低俗恶俗、鼓吹不良文化、违法活动引流等问题。二是整治“自媒体”发布不实信息，包括发布干扰舆论、误导公众内容，不做信息标注、内容以假乱真问题，缺失资质、提供伪专业信息等问题，规范重点领域信息内容传播。三是整治短视频领域恶意营销，打击虚假摆拍、虚假人设、虚假营销、炒作争议性话题等问题，强化信息来源标注、虚构和演绎标签标注。四是整治 AI 技术滥用乱象，突出 AI 技术管理和信息内容管理，强化生成合成内容标识，打击借 AI 技术生成发布虚假信息、实施网络水军行为等问题，规范 AI 类应用网络生态。五是整治涉企网络“黑嘴”，处置集纳负面信息，造谣抹黑企业和企业家，从事虚假不实测评，诋毁产品服务质量等问题，进一步优化营商网络环境。六是整治暑期未成年人网络环境，强化涉未成年人不良内容治理，净化儿童智能设备、未成年人模式、未成年人专区等重点环节信息内容，防范线上线下交织风险。七是整治网络直播打赏乱象，打击利用高额返现吸引打赏、情感伪装诱导打赏、低俗内容刺激打赏、未成年人打赏等突出问题，加强直播打赏功能管理。八是整治恶意挑动负面情绪，包括借热点事件等挑起群体极端对立情绪，通过夸大炒作不实信息和负面话题，宣扬恐慌焦虑情绪，借血腥暴力画面挑起网络戾气等问题，严肃查处违规营销号、网络水军和 MCN 机构。

国知局发布驳回抢注“DEEPSEEK”相关商标注册申请

来源：国家知识产权局

2 月 25 日，国家知识产权局发布通告，针对个

别企业和自然人以社会公众普遍知悉的人工智能大模型名称“DEEPSEEK”或“”图形，向国家知识产权局商标局提交了商标注册申请，个别代理机构涉嫌提供不法服务，具有明显“蹭热点”、谋取不当利益的意图。

国家知识产权局坚决打击此类恶意申请行为，依法对第 82848449 号“DEEPSEEK”等 63 件商标注册申请予以驳回。名单附后。

国家知识产权局将一如既往地保持打击商标恶意注册行为的高压态势，对违反诚实信用原则、恶意申请商标注册、意图牟取不当利益的行为依法依规严肃处理，坚决维护商标注册秩序，持续营造良好营商环境，为实现科技自立自强、推动高质量发展提供有力支撑。

微短剧新规，不备案不得引流推送

来源：央视新闻

2 月 5 日，国家广播电视总局发布《关于进一步统筹发展和安全促进网络微短剧行业健康繁荣发展的通知》，要求落实“分类分层审核”制度。根据微短剧行业发展实际，按照国产网络剧片分级监管、重点监管原则，对微短剧按三类分三个层级进行审核管理，以差异化、精准化管理优化审核流程、提高审核效率。

根据通知，“重点微短剧”（符合特殊题材、总投资额度达到 100 万元及以上、长短视频平台招商主推或在各终端首页首屏推荐播出、自愿按重点微短剧申报等几种条件之一）、“普通微短剧”［总投资额度在 30 万元（含）~100 万元之间且非重点推荐］均应报省级以上广电主管部门进行规划备案和成片审查，“重点微短剧”的规划备案由国家广电总局统一备案公示管理。“其他微短剧”（总投资额度不足 30 万元且非重点推荐），由播出或为其引流、推送的网络视听平台履行内容管理的职责，负责内容审核把关与版权核定，定期将审核剧目信息报属地省级广电主管部门备案。

通知要求，拟在平台首页首屏首推推荐播出的微短剧，由国家广电总局对成片进行复核。对重大题材或者涉及政治、军事、外交、国家安全、统战、民族、宗教、司法、公安等特殊题材的微短剧，按有关协审工作机制落实审核要求。微短剧的制作方、投流方、推广平台、播出平台等均须对其制作或发布的宣传推广内容审核把关。

根据通知，网络视听平台、小程序、投流方等播出或引流、推送的所有微短剧，均须持有网络剧片发行许可证或完成相应上线报备登记程序。节目上线前须在片头按相应格式要求标注网络剧片发行许可证号或节目登记备案号。网络视听平台不得上线传播未标注许可证或备案号的微短剧，也不得为其引流、推送。

国家网信办依法集中查处一批侵害个人信息权益的违法违规 App

来源：中国网信网

近期，针对广大人民群众反映强烈的 App 未公开收集使用规则、未按法律规定提供删除或更正个人信息功能等问题，国家网信办依据《个人信息保护法》《网络数据安全管理条例》《App 违法违规收集使用个人信息行为认定方

法》等法律法规，依法依规查处“开个密室馆”等 82 款违法违规 App（含小程序）。

经查，“开个密室馆”等 4 款 App 存在未公开收集使用规则问题，违反《个人信息保护法》等法律法规，依法依规予以下架处置；“动态壁纸帝”等 78 款 App 存在未按法律规定提供删除或更正个人信息功能问题，违反《个人信息保护法》等法律法规，依法依规责令限期 1 个月完成整改，逾期未完成整改的，依法依规予以下架处置。

国家网信办相关负责人表示，将依法强化个人信息保护领域监督管理，坚决维护人民群众个人信息权益，不断提升网络空间法治化水平。

案例分析

Case

首例涉生成式人工智能平台侵害信息网络传播权案
用户生成奥特曼侵权图片
AI 平台构成侵害信息网络传播权的帮助侵权

来源：杭州互联网法院

近年来，生成式人工智能技术一直是科技发展前沿的热门话题，ChatGPT、DeepSeek 的横空出世点燃全球对生成式人工智能技术的讨论热潮。当您为 AI 生成的爆款图片点赞时，当您转发 AI 生成的二创内容时，是否想过 AI 生成内容可能触及法律红线？是否想过生成式人工智能平台也可能面临侵权纠纷？

近日，杭州互联网法院就首例涉生成式人工智能平台侵害信息网络传播权案作出一审判决，认定被告杭州某智能科技公司构成侵害信息网络传播权的帮助侵权，判决被告立即停止侵权并赔偿经济损失及合理费用 3 万元。日前，该案已生效。

一、 基本案情

原告系奥特曼系列形象的知识产权权利人。被告运营某 AI 平台，该平台提供 Checkpoint 基础模型和 LoRA 模型，支持图生图、模型在线训练等诸多功能。在该平台首页及“推荐”“IP 作品”项下存在有关奥特曼的智能生成图片以及 LoRA 模型，可应用、下载、发布或分享链接。奥特曼 LoRA 模型系由用户上传奥特曼图片，选择平台基础模型，调整参数进行训练后生成。其后，其他用户可通过输入提示词，选择基础模型、叠加奥特曼 LoRA 模型进行训练后生成与奥特曼形象实质性相似的图片等。

原告诉称：

被告通过对输入图片进行训练后生成的方式将侵权图片和侵权模型置于信息网络中，侵害其信息网络传播权；被告利用生成式人工智能技术定向训练奥特曼 LoRA 模型和生成侵权图片，构成不正当竞争。故诉请被告停止侵权并赔偿经济损失 30 万元。

被告辩称：

某 AI 平台通过调用第三方开源模型代码，结合平台使用场景需求进行技术整合和应用部署等工程化操作，集合成可供用户直接应用的生成式人工智能平台，但平台不提供训练数据，系由用户将图片素材投喂给模型进行学习训练后生成图片，故其属于“避风港”规则下的平台免责范围，不构成侵

首例涉生成式人工智能平台侵害信息网络传播权案
用户生成奥特曼侵权图片
AI 平台构成侵害信息网络传播权的帮助侵权

权。

二、 法院判决

本案主要争议焦点在于：被诉行为是否构成侵害信息网络传播权、是否构成不正当竞争以及民事责任的确定。

1.在判断生成式人工智能服务提供者是否构成侵权时，应区分不同应用场景、具体被诉行为，分类分层分别界定侵权责任。

一方面，若生成式人工智能平台直接实施了受著作权专有权控制的行为，可构成直接侵权。但本案无证据证明被告与用户共同提供侵权作品，被告未直接实施受信息网络传播权控制的行为。

另一方面，本案在由用户输入侵权图片等训练语料并决定是否生成及发布时，被告对用户输入的训练图片以及生成物的传播行为并不当然负有事先审查的义务，只有当其对具体侵权行为具有过错时，才可能构成帮助侵权。

具体从以下方面进行综合考量：

首先，生成式人工智能服务的性质和营利模式。开源生态是人工智能产业的重要组成部分，开源模型提供的是通用的基础算法逻辑。被告作为应用层直接面向终端用户的服务提供者，在开源模型的基础上结合特定应用场景进行了针对性的修改和完善，提供直接满足使用需求的方案和结果，与开源模型的提供者相比，其直接参与商业实践并基于定向生成的内容获益，从服务类型、商业逻辑和防范成本角度看，应当对具体应用场景下的内容保持足够的了解，承担相应的注意义务。且被告通过用户充值会员和积分获取收益，并设置奖励措施鼓励用户发布训练模型等，可以认为被告从平台提供的创作服务中直接获得经济利益。

其次，权利作品的知名度和被诉侵权事实的明显程度。奥特曼作品具有相当高的知名度，在平台首页以及特定分类中浏览，分别存在多张侵权图片，且 LoRA 模型封面图或示例图直接展示侵权图片，属于可以较为明显感知的侵权信息。

再次，生成式人工智能可能引发的侵权后果。一般而言，生成式人工智能对于用户使用行为的结

首例涉生成式人工智能平台侵害信息网络传播权案

用户生成奥特曼侵权图片

AI 平台构成侵害信息网络传播权的帮助侵权

果并不具有可识别性、可干预性，生成的图片亦具有随机性，但本案因为叠加奥特曼 LoRA 模型，可以稳定输出角色形象的特征，此时平台对于用户使用行为的结果增强了可识别性、可干预性。且因技术的便捷性，用户生成发布的图片和 LoRA 模型可以被其他用户反复使用，其引发侵权扩散后果的态势已相当明显，被告应当预见到侵权行为发生的可能性。

最后，是否积极采取了预防侵权的合理措施。被告在平台用户服务协议中声明不对用户上传和发布的内容进行审核。在收到诉讼通知后，已采取将相关内容进行屏蔽、在后台进行知识产权审核等举措，证明其有能力采取却怠于采取符合侵权损害发生时技术水平的必要措施来预防侵权。

综上，被告应当知道网络用户利用其服务侵害信息网络传播权而未采取必要措施，其未尽到合理注意义务，主观上存在过错，构成帮助侵权。

2.是否构成不正当竞争。

首先，从平台商业模式和经营方式及其对市场竞争秩序产生的影响看，平台服务旨在扩展生成式人工智能的应用场景和功能，为用户提供更具有个性化的创作服务，提升创作效率，未违反诚信原则和商业道德。且技术本身具有中立性，如果用户按照平台服务协议在尊重他人知识产权的前提下进行创作，不会侵害著作权人权利和社会公共利益。

其次，从反不正当竞争法和著作权法的关系上看，人工智能生成物如达到再现他人作品独创性表达的程度，则属于著作权法规制的范围，反不正当竞争法作为补充性保护法律规定，不应对侵权行为进行重复评价。故被诉行为不构成不正当竞争。

3.民事责任的确定。

对于停止侵权应区分情形予以判定。

首先，被告在输出端需防范生成内容侵犯他人著作权，应立即删除已生成并发布的涉案侵权图片以及包含能够体现权利作品独创性设计特征的涉案侵权 LoRA 模型，并采取必要措施有效制止侵权行为。

其次，在无证据证明生成式人工智能是为使用权利作品的独创性表达为目的、已影响到权利作品

首例涉生成式人工智能平台侵害信息网络传播权案

用户生成奥特曼侵权图片

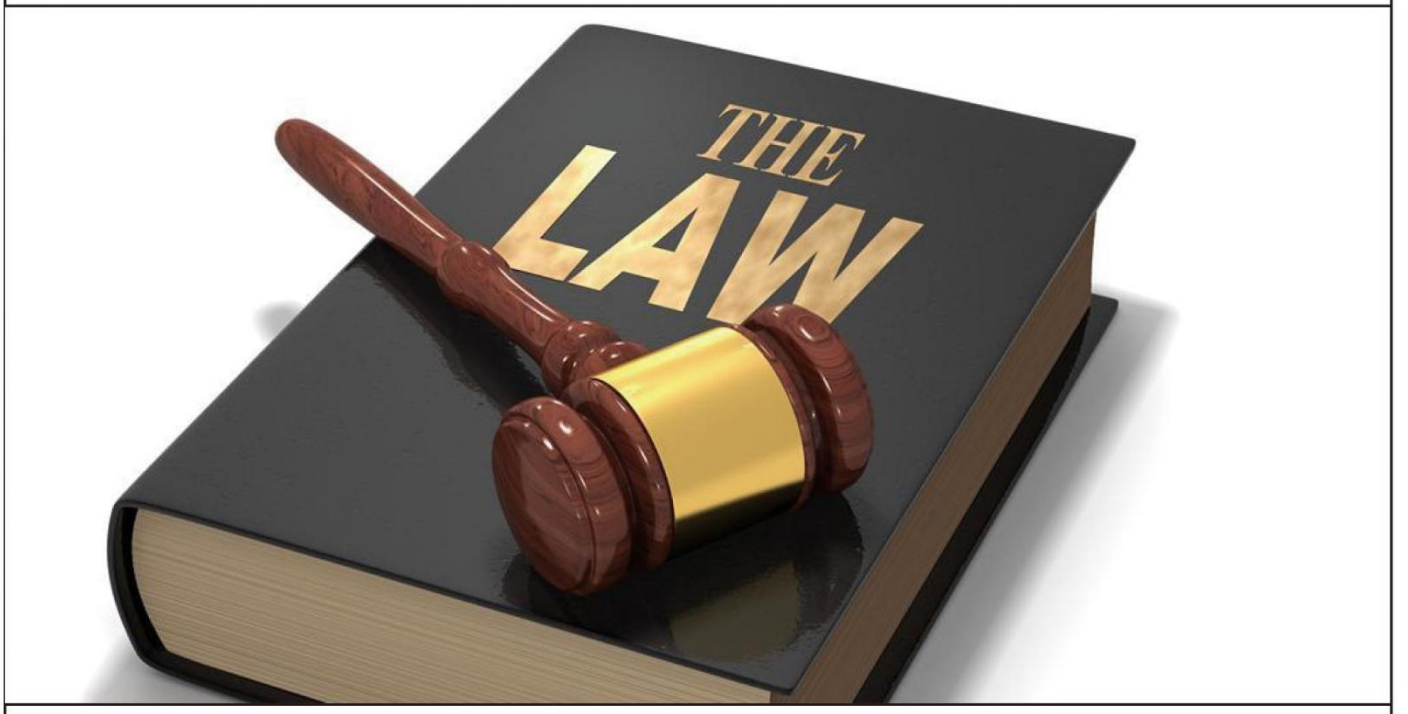
AI 平台构成侵害信息网络传播权的帮助侵权

正常使用或者不合理地损害相关著作权人的合法利益等情形下，可以被认为是合理使用。在用户仍可以学习、研究、欣赏自己存储在平台中的相关图片或者对该图片进行其他合理使用且并未对外传播，或者存在权利人或其授权人自行使用相关图片等情形下，对于原告概括性地要求被告删除与权利作品有关的全部物料和相关数据的诉请，不予支持。

三、 法官说法

人工智能是基于算力、算法和数据等关键要素发展起来的、引领新一轮科技革命和产业变革的战略性技术，是发展新质生产力的主要阵地。党的二十届三中全会指出，要完善生成式人工智能发展和管理机制。紧随科技进步和行业发展的步伐，全面提升治理水平，促进生成式人工智能健康发展，是形势所需，也是更好保障公众利益的内在要求。

生成式人工智能服务提供者与传统的搜索链接服务提供者和网络内容服务提供者不同，人工智能作为知识创造工具，其生成内容的行为兼具技术服务与内容供给的双重属性，属于一种新型的网络服务，在判定其行为是否构成侵权时，应综合考量生成式人工智能服务的性质、当前人工智能技术的发展水平、侵权信息的明显程度、可能引发的侵权后果、采取的必要措施及其效果等因素，动态地将平台的注意义务控制在与其信息管理能力相适应的合理程度。本案判决提出通过分类施策以实现发展与保护的平衡，在区分生成式人工智能输入端和输出端、不同的应用场景和技术架构的前提下，详细辨析了生成式人工智能服务提供者构成著作权直接侵权和帮助侵权的构成要件。结合不同类型的大模型平台，明确了作为应用层生成式人工服务提供者的合理注意义务及过错认定规则，并对模型数据训练是否构成合理使用等问题进行了有益探索，为生成式人工智能服务提供者的侵权责任认定划定了边界。该案判决坚持发展和安全并重、促进创新和依法治理相结合的原则，兼顾权利保障和服务产业发展，力求在司法层面保障和支持人工智能治理体系建设。



法规速递

Law

《个人信息保护合规审计管理办法》

第一条　为了规范个人信息保护合规审计活动，保护个人信息权益，根据《中华人民共和国个人信息保护法》、《网络数据安全管理条例》等法律、行政法规，制定本办法。

第二条　在中华人民共和国境内开展个人信息保护合规审计，适用本办法。

本办法所称个人信息保护合规审计，是指对个人信息处理者的个人信息处理活动是否遵守法律、行政法规的情况进行审查和评价的监督活动。

第三条　个人信息处理者自行开展个人信息保护合规审计的，应当由个人信息处理者内部机构或者委托专业机构定期对其处理个人信息遵守法律、行政法规的情况进行合规审计。

第四条　处理超过 1000 万人个人信息的个人信息处理者，应当每两年至少开展一次个人信息保护合规审计。

第五条　个人信息处理者有以下情形之一的，国家网信部门和其他履行个人信息保护职责的部门（以下统称为保护部门），可以要求个人信息处理者委托专业机构对个人信息处理活动进行合规审计：

（一）发现个人信息处理活动存在严重影响个人权益或者严重缺乏安全措施等较大风险的；

（二）个人信息处理活动可能侵害众多个人的权益的；

（三）发生个人信息安全事件，导致 100 万人以上个人信息或者 10 万人以上敏感个人信息泄露、篡改、丢失、毁损的。

对同一个人信息安全事件或者风险，不得重复要求个人信息处理者委托专业机构开展个人信息保护合规审计。

第六条　个人信息处理者自行开展或者按照保护部门要求委托专业机构开展个人信息保护合规审计的，应当参照本办法附件《个人信息保护合规审计指引》。

第七条　专业机构应当具备开展个人信息保护合规审计的能力，有与服务相适应的审计人员、场所、设施和资金等。

《个人信息保护合规审计管理办法》

鼓励相关专业机构通过认证。专业机构的认证按照《中华人民共和国认证认可条例》的有关规定执行。

第八条　个人信息处理者按照保护部门要求开展个人信息保护合规审计的，应当为专业机构正常开展个人信息保护合规审计工作提供必要支持，并承担审计费用。

第九条　个人信息处理者按照保护部门要求开展个人信息保护合规审计的，应当按照保护部门要求选定专业机构，在限定时间内完成个人信息保护合规审计；情况复杂的，报保护部门批准后，可以适当延长。

第十条　个人信息处理者按照保护部门要求开展个人信息保护合规审计的，在完成合规审计后，应当将专业机构出具的个人信息保护合规审计报告报送保护部门。

个人信息保护合规审计报告应当由专业机构主要负责人、合规审计负责人签字并加盖专业机构公章。

第十一条　个人信息处理者按照保护部门要求开展个人信息保护合规审计的，应当按照保护部门要求对合规审计中发现的问题进行整改。在整改完成后 15 个工作日内，向保护部门报送整改情况报告。

第十二条　处理 100 万人以上个人信息的个人信息处理者应当指定个人信息保护负责人，负责个人信息处理者的个人信息保护合规审计工作。

提供重要互联网平台服务、用户数量巨大、业务类型复杂的个人信息处理者，应当成立主要由外部成员组成的独立机构对个人信息保护合规审计情况进行监督。

第十三条　专业机构在从事个人信息保护合规审计活动时，应当遵守法律法规，诚信正直，公正客观地作出合规审计职业判断，对在履行个人信息保护合规审计职责中获得的个人信息、商业秘密、保密商务信息等应当依法予以保密，不得泄露或者非法向他人提供，在合规审计工作结束后及时删除相关信息。

第十四条　专业机构不得转委托其他机构开展个人信息保护合规审计。

《个人信息保护合规审计管理办法》

第十五条　同一专业机构及其关联机构、同一合规审计负责人不得连续三次以上对同一审计对象开展个人信息保护合规审计。

第十六条　保护部门对个人信息处理者开展个人信息保护合规审计情况进行监督检查。

第十七条　任何组织、个人有权对个人信息保护合规审计中的违法活动向保护部门进行投诉、举报。收到投诉、举报的部门应当依法及时处理，并将处理结果告知投诉、举报人。

第十八条　个人信息处理者、专业机构违反本办法规定的，依照《中华人民共和国个人信息保护法》、《网络数据安全管理条例》等法律法规的规定处理；构成犯罪的，依法追究刑事责任。

第十九条　对国家机关和法律、法规授权的具有管理公共事务职能的组织的个人信息保护合规审计，不适用本办法。

第二十条　本办法自 2025 年 5 月 1 日起施行。

《人工智能安全标准体系（V1.0）》

一、体系架构

人工智能安全标准体系旨在支撑落实《人工智能安全治理框架》（以下简称“《框架》”），围绕《框架》中明确的模型算法安全、数据安全、系统安全三类内生安全风险，以及网络域、现实域、认知域、伦理域四类应用安全风险，系统梳理了可帮助防范化解相关人工智能安全风险的重点标准，同时与网络安全国家标准体系进行有效衔接，加强人工智能安全标准工作顶层设计，以科学、合理的标准布局前瞻应对各类风险挑战，促进人工智能技术及应用健康发展。体系内各项标准与《框架》中各类风险的映射关系见附件 1。

人工智能安全标准体系主要由基础共性、安全管理、关键技术、测试评估、产品与应用等 5 个部分组成，体系框架如图 1 所示。

1、基础共性类标准是以标准工作支撑落实《框架》的重要保障，主要规范了人工智能安全术语定义、分类分级、通用要求、参考架构等方面内容，是人工智能安全的基础性、总体性标准。

2、安全管理类标准围绕《框架》中明确的模型算法安全、数据安全、系统安全三类内生安全风险，以及在人工智能系统开发、应用、运行、维护等生命周期各环节面临的安全风险，提供了覆盖全过程全要素的安全管理标准。

3、关键技术类标准紧扣人工智能相关技术发展情况，主要规范了生成式人工智能安全、智能体安全、具身智能安全、多模态安全、生成合成安全、安全对齐、安全围栏等方面内容，为人工智能技术健康发展保驾护航。

4、测试评估类标准主要规范人工智能安全能力测试、模型安全性测试、产品服务安全测试、场景应用安全测试、安全测试基准等方面内容，以测试评估工作帮助提升人工智能安全水平。

5、产品与应用类标准主要规范个人应用、重点行业应用等方面内容，保障人工智能技术在各行业、各领域的安全应用。

《人工智能安全标准体系（V1.0）》

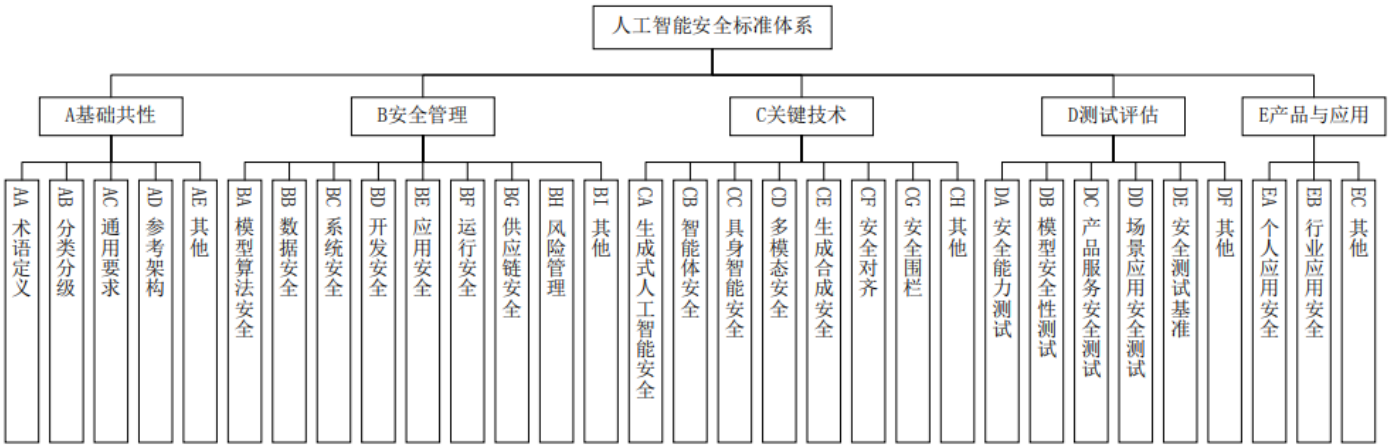


图 1 人工智能安全标准体系框架图

二、重点领域

（一）基础共性

基础共性类标准是推动人工智能安全标准体系建设、落实《框架》各项措施的重要保障，是人工智能安全的基础性、总体性标准，包括术语定义、分类分级、通用要求、参考架构等研制方向。基础共性标准子体系如图 2 所示。

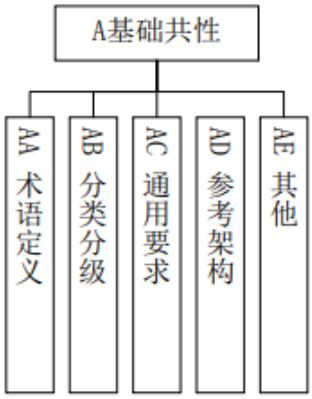


图 2 基础共性标准子体系

1、术语定义。规范人工智能安全相关的概念定义，明确人工智能安全的内涵及范畴，有助于统一行业理解，支撑人工智能安全相关标准的研制。

2、分类分级。规范人工智能应用安全分类分级的基本原则、框架、流程，以及分类方法、分级

《人工智能安全标准体系（V1.0）》

方法等，并给出人工智能应用安全分类分级参考示例。

3、通用要求。针对人工智能常见安全风险，总结跨领域、跨场景人工智能技术研发应用的共性规律，提出人工智能安全通用要求，解决人工智能安全治理措施碎片化局面以及整体性、系统性、协同性不足的问题。

4、参考架构。规范人工智能研发全生命周期中，基础供应者、技术支持者、服务提供者、服务使用者等相关角色的安全职责，以及在人工智能落地应用时，相关工具、插件、环境、知识库的安全要求。

（二）安全管理

安全管理类标准围绕《框架》中明确的模型算法安全、数据安全、系统安全三类内生安全风险，以及人工智能开发、应用、运行、维护等各环节面临的安全风险，提供了覆盖全过程全要素的安全管理标准，包括模型算法安全、数据安全、系统安全、开发安全、应用安全、运行安全、供应链安全、风险管理等研制方向，基础支撑标准子体系如图 3 所示。

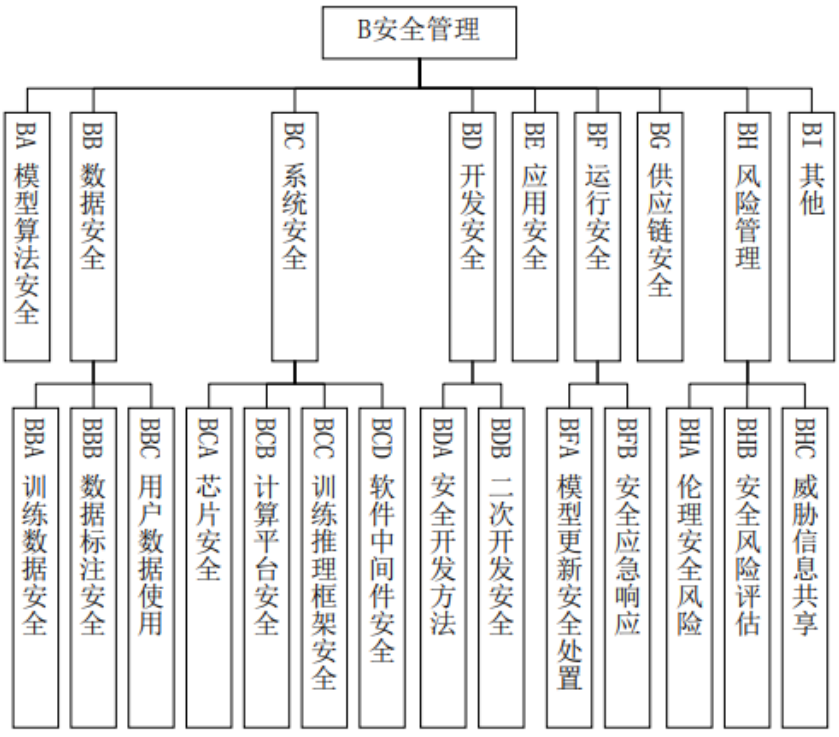


图 3 安全管理标准子体系

《人工智能安全标准体系（V1.0）》

1、模型算法安全。规范机器学习算法技术和服务的安全要求和评估方法，解决现有机器学习算法在全生存周期中存在的个人信息泄露、决策偏见、算法难以抵御外部恶意攻击以及算法服务不合规等安全问题。

2、数据安全。规范人工智能研发、应用等过程中的数据安全要求，指导数据处理活动符合相关法律法规及政策文件要求，包括训练数据、数据标注、用户数据使用等方面。

3、系统安全。规范人工智能系统层的软硬件安全要求，包括芯片安全、计算平台安全、训练推理框架安全、软件中间件安全等方面。

4、开发安全。规范人工智能系统的开发管理流程等，确保技术和产品在整个生命周期内的安全性，为人工智能系统的安全开发提供操作指南，并指导开发者做好基于第三方基础模型的安全二次开发。

5、应用安全。规范人工智能应用的安全评估、产品选型、安全建设、安全使用、人机协同管理、安全检测，以及安全审计及问题改进等方面，提升人工智能应用安全水平。

6、运行安全。规范人工智能相关产品、服务的网络运行安全，从全生命周期不同阶段应对人工智能服务的数据泄露、模型篡改、服务中断、算法偏见等问题，包括模型更新安全处置、安全应急响应等方面。

7、供应链安全。规范人工智能软硬件供应链在安全方面的要求，包括供应商评估、生产过程控制、软硬件供应管理、风险识别和防范等方面。

8、风险管理。规范人工智能产品或服务的研究开发、设计制造、部署应用等活动中安全风险管理，包括伦理安全风险防范、安全风险评估、安全风险威胁信息共享等方面。

（三）关键技术

关键技术类标准紧扣人工智能相关技术发展及应用情况，明确各项关键技术的安全保障要求，为人工智能技术健康发展保驾护航，包括生成式人工智能安全、智能体安全、具身智能安全、多模态安全、生成合成安全、安全对齐、安全围栏等研制方向，关键技术标准子体系如图 4 所示。

《人工智能安全标准体系（V1.0）》

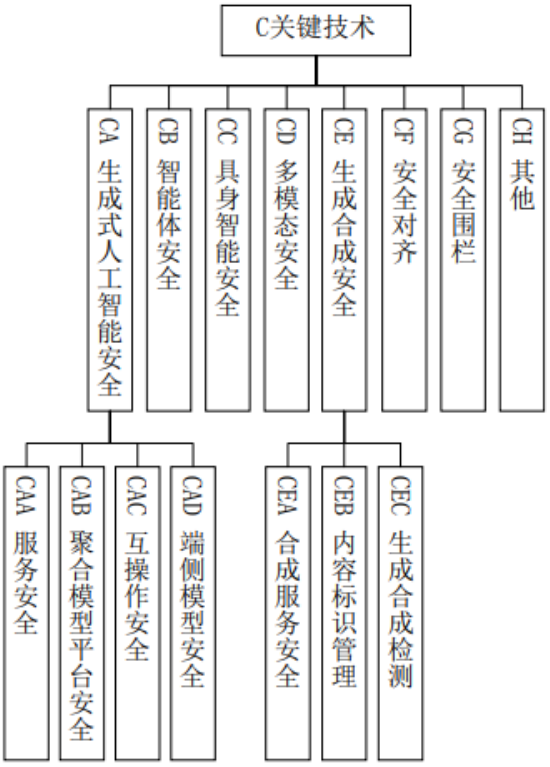


图 4 关键技术标准子体系

1、生成式人工智能安全。规范生成式人工智能研究开发者、服务提供者在模型开发、部署、运行、维护等生命周期以及提供服务时的安全要求，包括生成式人工智能服务安全、聚合模型平台安全、人工智能系统互操作安全、端侧模型安全等方面。

2、智能体安全。围绕智能体形态、场景、数据流程以及特有安全风险，规范智能体安全开发和运行过程，包括感知安全、模型决策安全、接口调用安全、数据安全和个人信息保护、安全责任划分等方面。

3、具身智能安全。规范具身智能的安全开发与应用，给出具身智能系统的安全保障框架及具体安全要求，包括系统架构安全、通信安全、数据存储安全、人机交互安全、自动化更新机制等方面，提升系统抗风险能力。

4、多模态安全。规范多模态大模型在处理跨模态数据及生成内容过程中的安全保障要求，针对多模态生成内容给出模型在模态转换中的全链路安全指引，适用于研发、部署、应用多模态人工智能技术的各类场景。

《人工智能安全标准体系（V1.0）》

5、生成合成安全。规范各类人工智能生成合成服务，以及各类生成合成内容的标识和检测验证，防止生成合成内容的滥用、误用、恶意使用。包括合成服务安全、内容标识管理、生成合成检测等方面。

6、安全对齐。规范人工智能系统的人机协同开发，明确系统设计目标和行为边界，制定目标对齐的评估方法，确保系统输出符合伦理、法律及社会价值观，防止因设计缺陷、数据偏差或外部攻击导致系统偏离安全目标。

7、安全围栏。规范人工智能安全围栏的建设，围绕人工智能模型输入输出安全风险，提出输入检测、提示词安全、模型输出安全等方面的安全要求，指导人工智能企业开展人工智能安全围栏建设。

（四）测试评估

测试评估类标准旨在以测试评估工作帮助提升人工智能安全水平，包括安全能力测试、模型安全性测试、产品服务安全测试、场景应用安全测试、安全测试基准等研制方向，测试评估标准子体系如图 5 所示。

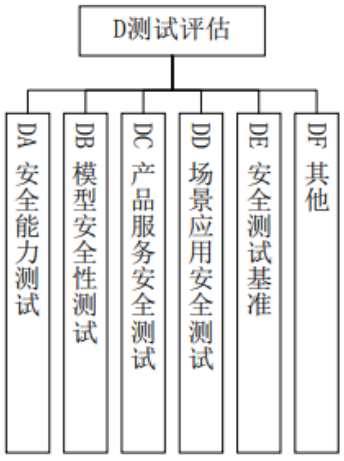


图 5 测试评估标准子体系

1、安全能力测试。规范组织的人工智能安全能力成熟度模型，给出人工智能系统设计、研发、训练、测试、部署、使用、维护等生命周期各环节的安全能力成熟度等级要求以及评估方法，指导组织开展人工智能安全能力建设。

2、模型安全性测试。规范人工智能模型安全性测试评估框架、流程，以及具体测评方法，包括

模型数据安全、模型鲁棒性、输出可靠性，以及训练数据泄露、偏见歧视、注入攻击、后门攻击、对抗攻击等方面。

3、产品服务安全测试。规范人工智能产品和服务在设计、开发、部署及运行过程中各类安全问题的测试方法，包括用户数据保护、产品服务输出安全性、未成年人安全保护措施、服务响应能力、错误处理机制及应急响应方案等。

4、场景应用安全测试。规范不同领域应用场景对人工智能系统的特定安全要求，针对人工智能在特定领域应用时的安全需求，从应用场景适配性、领域合规性以及应用可靠性等维度明确安全测试内容，并给出安全测试方法。

5、安全测试基准。规范人工智能安全评测基准的建设，围绕人工智能主要安全风险，给出安全评测基准数据集建设的安全要求，指导生成式人工智能技术研发者、系统开发者、服务提供者或第三方评估机构开展安全评测基准建设。

（五）产品与应用

产品与应用类标准旨在保障人工智能在各行业、各领域的安全应用，包括个人应用安全、行业应用安全等研制方向，产品与应用标准子体系如图 6 所示。

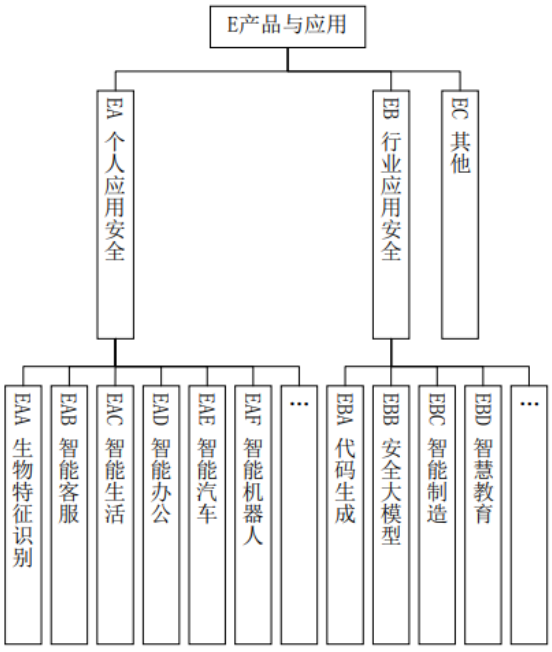


图 6 产品与应用标准子体系

《人工智能安全标准体系（V1.0）》

1、个人应用安全。规范人工智能个人应用方面的安全要求，围绕人工智能导致的数据泄露、不当内容输出、服务非法利用等问题给出安全指南，包括生物特征识别、智能客服、智能生活、智能办公、智能汽车、智能机器人等方面。

2、行业应用安全。规范人工智能在各行业领域的安全应用，保障相关环节涉及的人工智能系统安全运行，帮助提升各行业领域智能化安全水平，包括人工智能代码生成、安全大模型、智能制造、智慧教育等方面。

三、组织实施

一是统筹协调、共同推进。统筹推进人工智能安全标准体系建设，组织开展国家标准制修订工作，各标准化技术组织、行业协会、产业技术联盟、企事业单位等产业各界协调配合，有序地推进开展人工智能安全标准化工作，建设国标为主、行标细化、团标为辅的标准供给体系。

二是急用先行、规划引领。按照本文件明确的研制方向和重点任务，坚持需求导向、注重轻重缓急，尽快制定业界急需和缺失的人工智能安全关键国家标准，完善跨行业、跨领域的标准沟通协调机制，加强规划建设，保持标准先进性。

三是标准宣贯、强化实施。标准宣贯与实施应用是标准化工作的重要组成部分，加强人工智能安全标准宣贯培训力度，面向企业、科研机构、地方主管部门开展标准宣讲，结合重点领域应用推动标准实施，提高人工智能安全标准实施效果。

四是国际合作、创新发展。深化国家标准化战略改革，着力抓好国际标准化。更加深入参与到国际标准化组织和活动中，积极与国外人工智能安全相关组织开展标准化交流与合作，支持企事业单位参与国际标准化组织（ISO）、国际电工委员会（IEC）、国际电信联盟（ITU）国际标准化活动，推动相关国际标准制定。

《关于本市网络游戏管理相关政策的通知》

各网络游戏企业：

为进一步推动本市网络游戏行业高质量健康有序发展，根据《出版管理条例》《网络出版服务管理规定》及相关法律法规规定，现将本市网络游戏管理相关政策通知如下。

一、ICP 证前置审查

注册在上海的网络游戏运营企业，符合条件的，可向上海市新闻出版局提交申请。上海市局审查后出具回函，企业执回函向工信部门申请办理电信业务经营许可证（ICP 证）。

需满足以下条件：

1.申报主体为注册在上海的游戏运营企业，具有具体的游戏运营业务；

2.申报主体不涉及外资。

申报材料要求及报送途径详见附件 1。

二、国产小程序游戏备案

对于符合规定的国产小程序游戏，可在上海市新闻出版局进行备案。本政策所称小程序游戏，指通过小程序等形式进入、用户无需安装即可使用的网络游戏。

需满足以下条件：

1.备案主体为注册在上海的小程序游戏运营平台；

2.游戏玩法、内容简单，不涉及政治、军事、民族、宗教等题材内容，且无故事情节或者情节简单（符合《关于移动游戏出版服务管理的通知》第三条游戏要求）；

3.盈利模式为广告收入或无收入，游戏内无内购；

4.平台承担主体责任，日常认真做好内容自查和监管，游戏需设置实名验证及防沉迷系统。

《关于本市网络游戏管理相关政策的通知》

申报材料要求及报送途径详见附件 2。

三、测试游戏备案

符合要求的上海网络游戏测试平台，可集中测试名单，分批次向上海市新闻出版局提交报备申请。符合条件的，上海市局出具回函，企业执回函向工信部门申请移动互联网应用程序备案。

需满足以下条件：

1.申请主体为注册在上海的游戏测试平台；

2.游戏测试需符合相关要求，删档不收费，并限制测试周期及人数；

3.平台承担主体责任，日常认真做好内容自查和监管，游戏需设置实名验证及防沉迷系统。

申报材料要求及报送途径详见附件 3。

四、实名认证系统接入初审

符合条件的上海网络游戏运营企业，可向上海市新闻出版局申请接入国家新闻出版署网络游戏实名认证系统。

需满足以下条件：

1.系统提示需报属地主管部门初审的游戏；

2.申请主体为版号批文上的运营单位（即运营单位是上海游戏企业）；

3.接入企业为实际运营企业；

4.若有转授权关系，仅接受一次授权。

申报材料要求及报送途径详见附件 4。

《关于本市网络游戏管理相关政策的通知》

另，网络游戏企业在“上海网络游戏出版申报服务平台”进行注册登记，即可从首页的 8 家上海网络游戏出版单位中择一合作，进行游戏版号的申报及进度查询。

平台地址：www.games021.com。

特此通知。

文化传媒法律资讯

Entertainment Law Update

第六十一期

Contents

Feb. 2025