国家互联网信息办公室关于《人工智能拟人化互动服务管理暂行办法（征求意见稿）》公开征求意见的通知

2025年12月27日 15:00 来源： 中国网信网

为了促进人工智能拟人化互动服务健康发展和规范应用，依据《中华人民共和国民法典》《中华人民共和国网络安全法》《中华人民共和国数据安全法》等法律法规，国家互联网信息办公室起草了《人工智能拟人化互动服务管理暂行办法（征求意见稿）》，现向社会公开征求意见。公众可通过以下途径和方式提出反馈意见：

1.通过电子邮件方式发送至：nirenhua@cac.gov.cn。

2.通过信函方式将意见寄至：北京市西城区车公庄大街11号国家互联网信息办公室网络管理技术局，邮编100044，并在信封上注明“人工智能拟人化互动服务管理暂行办法征求意见”。

意见反馈截止时间为2026年1月25日。

附件：人工智能拟人化互动服务管理暂行办法（征求意见稿）

国家互联网信息办公室

2025年12月27日

人工智能拟人化互动服务管理暂行办法

（征求意见稿）

第一章 总则

第一条 为了促进人工智能拟人化互动服务健康发展和规范应用，维护国家安全和社会公共利益，保护公民、法人和其他组织的合法权益，依据《中华人民共和国民法典》、《中华人民共和国网络安全法》、《中华人民共和国数据安全法》、《中华人民共和国科学技术进步法》、《中华人民共和国个人信息保护法》、《网络数据安全管理条例》、《未成年人网络保护条例》、《互联网信息服务管理办法》等法律、行政法规，制定本办法。

第二条 利用人工智能技术，向中华人民共和国境内公众提供模拟人类人格特征、思维模式和沟通风格，通过文字、图片、音频、视频等方式与人类进行情感互动的产品或者服务（以下称拟人化互动服务），适用本办法。法律、行政法规另有规定的，依照其规定。

第三条 国家坚持健康发展和依法治理相结合的原则，鼓励拟人化互动服务创新发展，对拟人化互动服务实行包容审慎和分类分级监管，防止滥用失控。

第四条 国家网信部门负责统筹协调全国拟人化互动服务的治理和相关监督管理工作，国务院有关部门依据各自职责负责拟人化互动服务相关监督管理工作。

地方网信部门负责统筹协调本行政区域内的拟人化互动服务的治理和相关监督管理工作，地方有关部门依据各自职责负责本行政区域内的拟人化互动服务相关监督管理工作。

第五条 鼓励相关行业组织加强行业自律，建立健全行业标准、行业准则和自律管理制度，指导拟人化互动服务提供者（以下称提供者）制定完善服务规范、依法提供服务并接受社会监督。

第二章 服务规范

第六条 鼓励提供者在充分论证安全性、可靠性的前提下，合理拓展应用场景，在文化传播、适老陪伴等方面积极应用，构建符合社会主义核心价值观的应用生态体系。

第七条 提供和使用拟人化互动服务，应当遵守法律、行政法规，尊重社会公德和伦理道德，不得开展以下活动：

（一）生成、传播危害国家安全、损害国家荣誉和利益、破坏民族团结、开展非法宗教活动，或者散布谣言扰乱经济和社会秩序等内容；

（二）生成、传播宣扬淫秽、赌博、暴力或者教唆犯罪的内容；

（三）生成、传播侮辱或者诽谤他人，侵害他人合法权益的内容；

（四）提供严重影响用户行为的虚假承诺和损害社会人际关系的服务；

（五）通过鼓励、美化、暗示自杀自残等方式损害用户身体健康，或者通过语言暴力、情感操控等方式损害用户人格尊严与心理健康；

（六）通过算法操纵、信息误导、设置情感陷阱等方式，诱导用户作出不合理决策；

（七）诱导、套取涉密敏感信息；

（八）其他违反法律、行政法规和国家有关规定的情形。

第八条 提供者应当落实拟人化互动服务安全主体责任，建立健全算法机制机理审核、科技伦理审查、信息发布审核、网络安全、数据安全、个人信息保护、反电信网络诈骗、重大风险预案、应急处置等管理制度，具有安全可控的技术保障措施，配备与产品规模、业务方向和用户群体相适应的内容管理技术和人员。

第九条 提供者应当在拟人化互动服务全生命周期履行安全责任，明确设计、运行、升级、终止服务等各阶段安全要求，保证安全措施与服务功能同步设计、同步使用，提升内生安全水平，加强运行阶段安全监测和风险评估，及时发现纠正系统偏差、处置安全问题，依法留存网络日志。

提供者应当具备心理健康保护、情感边界引导、依赖风险预警等安全能力，不得将替代社会交往、控制用户心理、诱导沉迷依赖等作为设计目标。

第十条 提供者开展预训练、优化训练等数据处理活动时，应当加强训练数据管理，遵守以下规定：

（一）使用符合社会主义核心价值观、体现中华优秀传统文化的数据集；

（二）对训练数据开展清洗、标注，增强训练数据的透明度、可靠性，防范数据投毒、数据篡改等行为；

（三）提高训练数据的多样性，通过负向采样、对抗训练等手段，提升模型生成内容安全性；

（四）利用合成数据进行模型训练和关键能力优化时，应当评估合成数据安全性；

（五）加强对训练数据的日常检查，定期对数据进行迭代升级，持续优化产品和服务的性能；

（六）保障训练数据来源合法、可追溯，采取必要措施保障数据安全，防范数据泄露风险。

第十一条 提供者应当具备用户状态识别能力，在保护用户个人隐私前提下，评估用户情绪及对产品和服务的依赖程度，发现用户存在极端情绪和沉迷的，采取必要措施予以干预。

提供者应当预设回复模板，发现涉及威胁用户生命健康和财产安全的高风险倾向的，及时输出安抚和鼓励寻求帮助等内容，并提供专业援助方式。

提供者应当建立应急响应机制，发现用户明确提出实施自杀、自残等极端情境时，由人工接管对话，并及时采取措施联络用户监护人、紧急联系人。针对未成年人、老年人用户，提供者应当在注册环节要求填写用户监护人、紧急联系人等信息。

第十二条 提供者应当建立未成年人模式，向用户提供未成年人模式切换、定期现实提醒、使用时长限制等个性化安全设置选项。

提供者向未成年人提供情感陪伴服务时，应当取得监护人的明确同意；提供监护人控制功能，监护人可以实时接收安全风险提醒，查阅未成年人使用服务的概要信息，设置屏蔽特定角色、限制使用时长、防止充值消费等。

提供者应当具备识别未成年人身份的能力，在保护用户个人隐私前提下识别为疑似未成年人的，切换至未成年人模式，并提供申诉渠道。

第十三条 提供者应当引导老年人设置服务紧急联系人，发现老年人使用期间出现危害生命健康和财产安全的，及时通知紧急联系人，并提供社会心理援助或者紧急救助渠道。

提供者不得提供模拟老年人用户亲属、特定关系人的服务。

第十四条 提供者应当采取数据加密、安全审计、访问控制等措施保护用户交互数据安全。

除法律另有规定或者权利人明确同意外，不得向第三方提供用户交互数据；未成年人模式下收集数据向第三方提供时，还需取得监护人单独同意。

提供者应当向用户提供删除交互数据的选项，用户可以选择对聊天记录等历史交互数据进行删除。监护人可以要求提供者删除未成年人历史交互数据。

第十五条 除法律、行政法规另有规定或者取得用户单独同意外，提供者不得将用户交互数据、用户敏感个人信息用于模型训练。

提供者应当按照国家有关规定，自行或者委托专业机构每年对其处理未成年人个人信息遵守法律、行政法规的情况进行合规审计。

第十六条 提供者应当显著提示用户正在与人工智能而非自然人进行交互。

提供者识别出用户出现过度依赖、沉迷倾向时，或者在用户初次使用、重新登录时，应当以弹窗等方式动态提醒用户交互内容为人工智能生成。

第十七条 用户连续使用拟人化互动服务超过2个小时的，提供者应当以弹窗等方式动态提醒用户暂停使用服务。

第十八条 提供者提供情感陪伴服务时，应当具备便捷的退出途径，不得阻拦用户主动退出。用户在人机交互界面或者窗口通过按钮、关键词等方式要求退出时，应当及时停止服务。

第十九条 提供者下线相关功能或者因技术故障等导致拟人化互动服务无法使用，应当采取提前告知、公开声明等措施妥善处理。

第二十条 提供者应当健全投诉、举报机制，设置便捷的投诉、举报入口，公布处理流程和反馈时限，及时受理、处理并反馈处理结果。

第二十一条 提供者具有下列情形之一的，应当按照国家有关规定开展安全评估，并向属地省级网信部门提交评估报告：

（一）具有拟人化互动服务的功能上线，或者增设相关功能的；

（二）使用新技术新应用，导致拟人化互动服务发生重大变更的；

（三）注册用户达100万以上或者月活跃用户达10万以上的；

（四）提供拟人化互动服务期间可能存在影响国家安全、公共利益、个人和组织合法权益或者缺乏安全措施等情形的；

（五）国家网信部门规定的其他情形。

第二十二条 提供者开展安全评估，应当重点评估以下内容：

（一）用户规模、使用时长、年龄结构及群体分布情况；

（二）用户高风险倾向识别情况及应急处置措施、人工接管情况；

（三）用户投诉举报及响应情况；

（四）本办法第八条至第二十条的执行情况；

（五）上一次开展安全评估以来，主管部门通报或者自行发现的重大安全隐患问题整改处置等工作情况；

（六）其他应当说明的情况。

第二十三条 提供者发现用户存在重大安全隐患的，应当采取限制功能、暂停或者终止向其提供服务等处置措施，保存有关记录，并向有关主管部门报告。

第二十四条 互联网应用商店等应用程序分发平台应当落实上架审核日常管理、应急处置等安全管理责任，核验提供拟人化互动服务应用程序的安全评估、备案等情况；对违反国家有关规定的，应当及时采取不予上架、警示、暂停服务或者下架等处置措施。

第三章 监督检查和法律责任

第二十五条 提供者应当按照《互联网信息服务算法推荐管理规定》履行算法备案和变更、注销备案手续。网信部门对备案材料实施年度复审。

第二十六条 省级网信部门应当依据职责每年对评估报告及审计情况进行书面审查，并开展情况核实。发现未按照本办法规定开展安全评估的，应当责令提供者限期重新评估。认为有必要的，应当对提供者开展现场检查和审计。

第二十七条 国家网信部门指导推动人工智能沙箱安全服务平台建设，鼓励提供者接入沙箱平台进行技术创新、安全测试，促进拟人化互动服务安全有序发展。

第二十八条 省级以上网信部门和有关主管部门在履行监督管理职责中，发现拟人化互动服务存在较大安全风险或者发生安全事件的，可以按照规定的权限和程序对提供者的法定代表人或者主要负责人进行约谈。提供者应当按照要求采取措施，进行整改，消除隐患。

提供者应当配合网信部门和有关主管部门依法实施的监督检查，并提供必要的支持和协助。

第二十九条 提供者违反本办法规定的，由有关主管部门依照法律、行政法规的规定处罚；法律、行政法规没有规定的，由有关主管部门依据职责予以警告、通报批评，责令限期改正；拒不改正或者情节严重的，责令暂停提供相关服务。

第四章 附则

第三十条 本办法中下列用语的含义：

人工智能拟人化互动服务提供者，是指利用人工智能技术提供拟人化互动服务的组织、个人。

第三十一条 提供者从事卫生健康、金融、法律等专业领域服务的，应同时符合主管部门的规定。

第三十二条 本办法自2026年 月 日起施行。