



教职工政治学习参考资料

(2025 年第 8 期)

苏州大学党委宣传部编

2025 年 11 月 13 日

教职工政治学习参考资料

(2025 年第 8 期)

苏州大学党委宣传部编

2025 年 11 月 13 日

● 学习内容

网络安全专题学习

● 参考资料

1	新一代人工智能伦理规范	1
1.1	伦理规范全文	1
1.2	伦理规范解读	5
2	生成式人工智能服务管理暂行办法	8
2.1	暂行办法全文	8
2.2	暂行办法解读	13
3	人工智能生成合成内容标识办法	16
3.1	标识办法全文	16
3.2	标识办法解读	19

1 新一代人工智能伦理规范

1.1 伦理规范全文

第一章 总则

第一条 本规范旨在将伦理道德融入人工智能全生命周期，促进公平、公正、和谐、安全，避免偏见、歧视、隐私和信息泄露等问题。

第二条 本规范适用于从事人工智能管理、研发、供应、使用等相关活动的自然人、法人和其他相关机构等。

（一）管理活动主要指人工智能相关的战略规划、政策法规和技术标准制定实施，资源配置以及监督审查等。

（二）研发活动主要指人工智能相关的科学研究、技术开发、产品研制等。

（三）供应活动主要指人工智能产品与服务相关的生产、运营、销售等。

（四）使用活动主要指人工智能产品与服务相关的采购、消费、操作等。

第三条 人工智能各类活动应遵循以下基本伦理规范。

（一）增进人类福祉。坚持以人为本，遵循人类共同价值观，尊重人权和人类根本利益诉求，遵守国家或地区伦理道德。坚持公共利益优先，促进人机和谐友好，改善民生，增强获得感幸福感，推动经济、社会及生态可持续发展，共建人类命运共同体。

（二）促进公平公正。坚持普惠性和包容性，切实保护各相关主体合法权益，推动全社会公平共享人工智能带来的益处，促进社会公平正义和机会均等。在提供人工智能产品和服务时，应充分尊重和帮助弱势群体、特殊群体，并根据需要提供相应替代方案。

（三）保护隐私安全。充分尊重个人信息知情、同意等权利，依照合法、正当、必要和诚信原则处理个人信息，保障个人隐私与数据安全，不得损害个

人合法数据权益，不得以窃取、篡改、泄露等方式非法收集利用个人信息，不得侵害个人隐私权。

（四）确保可控可信。保障人类拥有充分自主决策权，有权选择是否接受人工智能提供的服务，有权随时退出与人工智能的交互，有权随时中止人工智能系统的运行，确保人工智能始终处于人类控制之下。

（五）强化责任担当。坚持人类是最终责任主体，明确利益相关者的责任，全面增强责任意识，在人工智能全生命周期各环节自省自律，建立人工智能问责机制，不回避责任审查，不逃避应负责任。

（六）提升伦理素养。积极学习和普及人工智能伦理知识，客观认识伦理问题，不低估不夸大伦理风险。主动开展或参与人工智能伦理问题讨论，深入推动人工智能伦理治理实践，提升应对能力。

第四条 人工智能特定活动应遵守的伦理规范包括管理规范、研发规范、供应规范和使用规范。

第二章 管理规范

第五条 推动敏捷治理。尊重人工智能发展规律，充分认识人工智能的潜力与局限，持续优化治理机制和方式，在战略决策、制度建设、资源配置过程中，不脱离实际、不急功近利，有序推动人工智能健康和可持续发展。

第六条 积极实践示范。遵守人工智能相关法规、政策和标准，主动将人工智能伦理道德融入管理全过程，率先成为人工智能伦理治理的实践者和推动者，及时总结推广人工智能治理经验，积极回应社会对人工智能的伦理关切。

第七条 正确行权用权。明确人工智能相关管理活动的职责和权力边界，规范权力运行条件和程序。充分尊重并保障相关主体的隐私、自由、尊严、安全

等权利及其他合法权益，禁止权力不当行使对自然人、法人和其他组织合法权益造成侵害。

第八条 加强风险防范。增强底线思维和风险意识，加强人工智能发展的潜在风险研判，及时开展系统的风险监测和评估，建立有效的风险预警机制，提升人工智能伦理风险管控和处置能力。

第九条 促进包容开放。充分重视人工智能各利益相关主体的权益与诉求，鼓励应用多样化的人工智能技术解决经济社会发展实际问题，鼓励跨学科、跨领域、跨地区、跨国界的交流与合作，推动形成具有广泛共识的人工智能治理框架和标准规范。

第三章 研发规范

第十条 强化自律意识。加强人工智能研发相关活动的自我约束，主动将人工智能伦理道德融入技术研发各环节，自觉开展自我审查，加强自我管理，不从事违背伦理道德的人工智能研发。

第十一条 提升数据质量。在数据收集、存储、使用、加工、传输、提供、公开等环节，严格遵守数据相关法律、标准与规范，提升数据的完整性、及时性、一致性、规范性和准确性等。

第十二条 增强安全透明。在算法设计、实现、应用等环节，提升透明性、可解释性、可理解性、可靠性、可控性，增强人工智能系统的韧性、自适应性和抗干扰能力，逐步实现可验证、可审核、可监督、可追溯、可预测、可信赖。

第十三条 避免偏见歧视。在数据采集和算法开发中，加强伦理审查，充分考虑差异化诉求，避免可能存在的数据与算法偏见，努力实现人工智能系统的普惠性、公平性和非歧视性。

第四章 供应规范

第十四条 尊重市场规则。严格遵守市场准入、竞争、交易等活动的各种规章制度，积极维护市场秩序，营造有利于人工智能发展的市场环境，不得以数据垄断、平台垄断等破坏市场有序竞争，禁止以任何手段侵犯其他主体的知识产权。

第十五条 加强质量管控。强化人工智能产品与服务的质量监测和使用评估，避免因设计和产品缺陷等问题导致的人身安全、财产安全、用户隐私等侵害，不得经营、销售或提供不符合质量标准的产品与服务。

第十六条 保障用户权益。在产品与服务中使用人工智能技术应明确告知用户，应标识人工智能产品与服务的功能与局限，保障用户知情、同意等权利。为用户选择使用或退出人工智能模式提供简便易懂的解决方案，不得为用户平等使用人工智能设置障碍。

第十七条 强化应急保障。研究制定应急机制和损失补偿方案或措施，及时监测人工智能系统，及时响应和处理用户的反馈信息，及时防范系统性故障，随时准备协助相关主体依法依规对人工智能系统进行干预，减少损失，规避风险。

第五章 使用规范

第十八条 提倡善意使用。加强人工智能产品与服务使用前的论证和评估，充分了解人工智能产品与服务带来的益处，充分考虑各利益相关主体的合法权益，更好促进经济繁荣、社会进步和可持续发展。

第十九条 避免误用滥用。充分了解人工智能产品与服务的适用范围和负面影响，切实尊重相关主体不使用人工智能产品或服务的权利，避免不当使用和滥用人工智能产品与服务，避免非故意造成对他人合法权益的损害。

第二十条 禁止违规恶用。禁止使用不符合法律法规、伦理道德和标准规范的人工智能产品与服务，禁止使用人工智能产品与服务从事不法活动，严禁危害国家安全、公共安全和生产安全，严禁损害社会公共利益等。

第二十一条 及时主动反馈。积极参与人工智能伦理治理实践，对使用人工智能产品与服务过程中发现的技术安全漏洞、政策法规真空、监管滞后等问题，应及时向相关主体反馈，并协助解决。

第二十二条 提高使用能力。积极学习人工智能相关知识，主动掌握人工智能产品与服务的运营、维护、应急处置等各使用环节所需技能，确保人工智能产品与服务安全使用和高效利用。

第六章 组织实施

第二十三条 本规范由国家新一代人工智能治理专业委员会发布，并负责解释和指导实施。

第二十四条 各级管理部门、企业、高校、科研院所、协会学会和其他相关机构可依据本规范，结合实际需求，制订更为具体的伦理规范和相关措施。

第二十五条 本规范自公布之日起施行，并根据经济社会发展需求和人工智能发展情况适时修订。

1.2 伦理规范解读

2021年9月25日，国家新一代人工智能治理专业委员会发布了《新一代人工智能伦理规范》（以下简称《伦理规范》），旨在将伦理道德融入人工智能全生命周期，为从事人工智能相关活动的自然人、法人和其他相关机构等提供伦理指引。《伦理规范》经过专题调研、集中起草、意见征询等环节，充分考虑当前社会各界有关隐私、偏见、歧视、公平等伦理关切，包括总则、特定活动

伦理规范和组织实施等内容。

（1）明确 6 项基本伦理规范

《伦理规范》明确提出，人工智能各类活动应遵循增进人类福祉、促进公平公正、保护隐私安全、确保可控可信、强化责任担当、提升伦理素养等六项基本伦理规范。

（2）提出 18 项具体伦理要求

人工智能特定活动应遵守的伦理规范包括管理规范、研发规范、供应规范和使用规范共 18 项具体伦理要求。

管理规范包含推动敏捷治理、积极实践示范、正确行权用权、加强风险防范、促进包容开放；

研发规范包含强化自律意识、提升数据质量、增强安全透明、避免偏见歧视；

供应规范包括尊重市场规则、加强质量管控、保障用户权益、强化应急保障；

使用规范包含提倡善意使用、避免误用滥用、禁止违规恶用、及时主动反馈、提高使用能力。

（3）加强风险防范

加强风险防范方面，增强底线思维和风险意识，加强人工智能发展的潜在风险研判，及时开展系统的风险监测和评估，建立有效的风险预警机制，提升人工智能伦理风险管控和处置能力。

（4）保障用户权益

保障用户权益方面，在产品与服务中使用人工智能技术应明确告知用户，应标识人工智能产品与服务的功能与局限，保障用户知情、同意等权利。为用

户选择使用或退出人工智能模式提供简便易懂的解决方案，不得为用户平等使用人工智能设置障碍。

2 生成式人工智能服务管理暂行办法

2.1 暂行办法全文

第一章 总 则

第一条 为了促进生成式人工智能健康发展和规范应用,维护国家安全和公共利益,保护公民、法人和其他组织的合法权益,根据《中华人民共和国网络安全法》、《中华人民共和国数据安全法》、《中华人民共和国个人信息保护法》、《中华人民共和国科学技术进步法》等法律、行政法规,制定本办法。

第二条 利用生成式人工智能技术向中华人民共和国境内公众提供生成文本、图片、音频、视频等服务(以下称生成式人工智能服务),适用本办法。

国家对利用生成式人工智能服务从事新闻出版、影视制作、文艺创作等活动另有规定的,从其规定。

行业组织、企业、教育和科研机构、公共文化机构、有关专业机构等研发、应用生成式人工智能技术,未向境内公众提供生成式人工智能服务的,不适用本办法的规定。

第三条 国家坚持发展和安全并重、促进创新和依法治理相结合的原则,采取有效措施鼓励生成式人工智能创新发展,对生成式人工智能服务实行包容审慎和分类分级监管。

第四条 提供和使用生成式人工智能服务,应当遵守法律、行政法规,尊重社会公德和伦理道德,遵守以下规定:

(一) 坚持社会主义核心价值观,不得生成煽动颠覆国家政权、推翻社会主义制度,危害国家安全和利益、损害国家形象,煽动分裂国家、破坏国家统一和社会稳定,宣扬恐怖主义、极端主义,宣扬民族仇恨、民族歧视,暴力、淫秽色情,以及虚假有害信息等法律、行政法规禁止的内容;

（二）在算法设计、训练数据选择、模型生成和优化、提供服务等过程中，采取有效措施防止产生民族、信仰、国别、地域、性别、年龄、职业、健康等歧视；

（三）尊重知识产权、商业道德，保守商业秘密，不得利用算法、数据、平台等优势，实施垄断和不正当竞争行为；

（四）尊重他人合法权益，不得危害他人身心健康，不得侵害他人肖像权、名誉权、荣誉权、隐私权和个人信息权益；

（五）基于服务类型特点，采取有效措施，提升生成式人工智能服务的透明度，提高生成内容的准确性和可靠性。

第二章 技术发展与治理

第五条 鼓励生成式人工智能技术在各行业、各领域的创新应用，生成积极健康、向上向善的优质内容，探索优化应用场景，构建应用生态体系。

支持行业组织、企业、教育和科研机构、公共文化机构、有关专业机构等在生成式人工智能技术创新、数据资源建设、转化应用、风险防范等方面开展协作。

第六条 鼓励生成式人工智能算法、框架、芯片及配套软件平台等基础技术的自主创新，平等互利开展国际交流与合作，参与生成式人工智能相关国际规则制定。

推动生成式人工智能基础设施和公共训练数据资源平台建设。促进算力资源协同共享，提升算力资源利用效能。推动公共数据分类分级有序开放，扩展高质量的公共训练数据资源。鼓励采用安全可信的芯片、软件、工具、算力和数据资源。

第七条 生成式人工智能服务提供者（以下称提供者）应当依法开展预训练、优化训练等训练数据处理活动，遵守以下规定：

（一）使用具有合法来源的数据和基础模型；

（二）涉及知识产权的，不得侵害他人依法享有的知识产权；

（三）涉及个人信息的，应当取得个人同意或者符合法律、行政法规规定的其他情形；

（四）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；

（五）《中华人民共和国网络安全法》、《中华人民共和国数据安全法》、《中华人民共和国个人信息保护法》等法律、行政法规的其他有关规定和有关主管部门的相关监管要求。

第八条 在生成式人工智能技术研发过程中进行数据标注的，提供者应当制定符合本办法要求的清晰、具体、可操作的标注规则；开展数据标注质量评估，抽样核验标注内容的准确性；对标注人员进行必要培训，提升尊法守法意识，监督指导标注人员规范开展标注工作。

第三章 服务规范

第九条 提供者应当依法承担网络信息内容生产者责任，履行网络信息安全义务。涉及个人信息的，依法承担个人信息处理者责任，履行个人信息保护义务。

提供者应当与注册其服务的生成式人工智能服务使用者（以下称使用者）签订服务协议，明确双方权利义务。

第十条 提供者应当明确并公开其服务的适用人群、场合、用途，指导使用者科学理性认识和依法使用生成式人工智能技术，采取有效措施防范未成年人用户过度依赖或者沉迷生成式人工智能服务。

第十一条 提供者对使用者的输入信息和使用记录应当依法履行保护义务，不得收集非必要个人信息，不得非法留存能够识别使用者身份的输入信息和使用记录，不得非法向他人提供使用者的输入信息和使用记录。

提供者应当依法及时受理和处理个人关于查阅、复制、更正、补充、删除其个人信息等的请求。

第十二条 提供者应当按照《互联网信息服务深度合成管理规定》对图片、视频等生成内容进行标识。

第十三条 提供者应当在其服务过程中，提供安全、稳定、持续的服务，保障用户正常使用。

第十四条 提供者发现违法内容的，应当及时采取停止生成、停止传输、消除等处置措施，采取模型优化训练等措施进行整改，并向有关主管部门报告。

提供者发现使用者利用生成式人工智能服务从事违法活动的，应当依法依规采取警示、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。

第十五条 提供者应当建立健全投诉、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、处理公众投诉举报并反馈处理结果。

第四章 监督检查和法律责任

第十六条 网信、发展改革、教育、科技、工业和信息化、公安、广播电视、新闻出版等部门，依据各自职责依法加强对生成式人工智能服务的管理。

国家有关主管部门针对生成式人工智能技术特点及其在有关行业和领域的服务应用，完善与创新发展的科学监管方式，制定相应的分类分级监管规则或者指引。

第十七条 提供具有舆论属性或者社会动员能力的生成式人工智能服务的，应当按照国家有关规定开展安全评估，并按照《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。

第十八条 使用者发现生成式人工智能服务不符合法律、行政法规和本办法规定的，有权向有关主管部门投诉、举报。

第十九条 有关主管部门依据职责对生成式人工智能服务开展监督检查，提供者应当依法予以配合，按要求对训练数据来源、规模、类型、标注规则、算法机制机理等予以说明，并提供必要的技术、数据等支持和协助。

参与生成式人工智能服务安全评估和监督检查的相关机构和人员对在履行职责中知悉的国家秘密、商业秘密、个人隐私和个人信息应当依法予以保密，不得泄露或者非法向他人提供。

第二十条 对来源于中华人民共和国境外向境内提供生成式人工智能服务不符合法律、行政法规和本办法规定的，国家网信部门应当通知有关机构采取技术措施和其他必要措施予以处置。

第二十一条 提供者违反本办法规定的，由有关主管部门依照《中华人民共和国网络安全法》、《中华人民共和国数据安全法》、《中华人民共和国个人信息保护法》、《中华人民共和国科学技术进步法》等法律、行政法规的规定予以处罚；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。

构成违反治安管理行为的，依法给予治安管理处罚；构成犯罪的，依法追究刑事责任。

第五章 附 则

第二十二条 本办法下列用语的含义是：

（一）生成式人工智能技术，是指具有文本、图片、音频、视频等内容生成能力的模型及相关技术。

（二）生成式人工智能服务提供者，是指利用生成式人工智能技术提供生成式人工智能服务（包括通过提供可编程接口等方式提供生成式人工智能服务）的组织、个人。

（三）生成式人工智能服务使用者，是指使用生成式人工智能服务生成内容的组织、个人。

第二十三条 法律、行政法规规定提供生成式人工智能服务应当取得相关行政许可的，提供者应当依法取得许可。

外商投资生成式人工智能服务，应当符合外商投资相关法律、行政法规的规定。

第二十四条 本办法自 2023 年 8 月 15 日起施行。

2.2 暂行办法解读

《生成式人工智能服务管理暂行办法》（以下称《暂行办法》），是国家网信办联合国家发展改革委、教育部、科技部、工业和信息化部、公安部、广电总局七部门于 2023 年 7 月 10 日联合公布，自 2023 年 8 月 15 日起施行的办法。

《暂行办法》旨在促进生成式人工智能健康发展和规范应用，维护国家安全和公共利益，保护公民、法人和其他组织的合法权益。出台《暂行办法》，既是促进生成式人工智能健康发展的重要要求，也是防范生成式人工智能服务风险的现实需要。

（1）《暂行办法》的适用范围

《暂行办法》规定，利用生成式人工智能技术向中华人民共和国境内公众提供生成文本、图片、音频、视频等服务的内容，适用本办法。国家对利用生成式人工智能服务从事新闻出版、影视制作、文艺创作等活动另有规定的，从其规定。行业组织、企业、教育和科研机构、公共文化机构、有关专业机构等研发、应用生成式人工智能技术，未向境内公众提供生成式人工智能服务的，不适用本办法的规定。

（2）《暂行办法》坚持的主要原则

《暂行办法》提出国家坚持发展和安全并重、促进创新和依法治理相结合的原则，采取有效措施鼓励生成式人工智能创新发展，对生成式人工智能服务实行包容审慎和分类分级监管。

（3）定义了生成式人工智能技术和生成式人工智能服务提供者

生成式人工智能技术，是指具有文本、图片、音频、视频等内容生成能力的模型及相关技术；生成式人工智能服务提供者，是指利用生成式人工智能技术提供生成式人工智能服务（包括通过提供可编程接口等方式提供生成式人工智能服务）的组织、个人。

（4）明确了提供和使用生成式人工智能服务应当遵守的规定

明确提供和使用生成式人工智能服务应当坚持社会主义核心价值观，不得生成煽动颠覆国家政权、推翻社会主义制度，危害国家安全和利益、损害国家形象，煽动分裂国家、破坏国家统一和社会稳定，宣扬恐怖主义、极端主义，宣扬民族仇恨、民族歧视，暴力、淫秽色情，以及虚假有害信息等法律、行政法规禁止的内容；在算法设计、训练数据选择、模型生成和优化、提供服务等过程中，采取有效措施防止产生民族、信仰、国别、地域、性别、年龄、职业、健康等歧视；尊重知识产权、商业道德，保守商业秘密，不得利用算法、数据、

平台等优势，实施垄断和不正当竞争行为；尊重他人合法权益，不得危害他人身心健康，不得侵害他人肖像权、名誉权、荣誉权、隐私权和个人信息权益；基于服务类型特点，采取有效措施，提升生成式人工智能服务的透明度，提高生成内容的准确性和可靠性。

（5）规定了相关治理制度

明确生成式人工智能服务提供者应当依法开展预训练、优化训练等训练数据处理活动，使用具有合法来源的数据和基础模型；涉及知识产权的，不得侵害他人依法享有的知识产权；涉及个人信息的，应当取得个人同意或者符合法律、行政法规规定的其他情形；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。此外，明确了数据标注的相关要求。

（6）明确了生成式人工智能的服务规范

要求采取有效措施防范未成年人用户过度依赖或者沉迷生成式人工智能服务。规定提供者应当按照《互联网信息服务深度合成管理规定》对图片、视频等生成内容进行标识。规定提供者发现违法内容的，应当及时采取停止生成、停止传输、消除等处置措施并采取模型优化训练等措施进行整改。明确提供者发现使用者利用生成式人工智能服务从事违法活动的，应当依法依约采取有关处置措施，保存有关记录，并向有关主管部门报告。

（7）对投诉、举报作了相关规定

规定提供者应当建立健全投诉、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、处理公众投诉举报并反馈处理结果。明确使用者发现生成式人工智能服务不符合法律、行政法规和本办法规定的，有权向有关主管部门投诉、举报。

3 人工智能生成合成内容标识办法

3.1 标识办法全文

第一条 为了促进人工智能健康发展，规范人工智能生成合成内容标识，保护公民、法人和其他组织合法权益，维护社会公共利益，根据《中华人民共和国网络安全法》、《互联网信息服务算法推荐管理规定》、《互联网信息服务深度合成管理规定》、《生成式人工智能服务管理暂行办法》等法律、行政法规和部门规章，制定本办法。

第二条 符合《互联网信息服务算法推荐管理规定》、《互联网信息服务深度合成管理规定》、《生成式人工智能服务管理暂行办法》规定情形的网络信息服务提供者（以下简称“服务提供者”）开展人工智能生成合成内容标识活动，适用本办法。

第三条 人工智能生成合成内容是指利用人工智能技术生成、合成的文本、图片、音频、视频、虚拟场景等信息。

人工智能生成合成内容标识包括显式标识和隐式标识。

显式标识是指在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以被用户明显感知到的标识。

隐式标识是指采取技术措施在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。

第四条 服务提供者提供的生成合成服务属于《互联网信息服务深度合成管理规定》第十七条第一款情形的，应当按照下列要求对生成合成内容添加显式标识：

（一）在文本的起始、末尾或者中间适当位置添加文字提示或者通用符号提示等标识，或者在交互场景界面、文字周边添加显著的提示标识；

（二）在音频的起始、末尾或者中间适当位置添加语音提示或者音频节奏提示等标识，或者在交互场景界面中添加显著的提示标识；

（三）在图片的适当位置添加显著的提示标识；

（四）在视频起始画面和视频播放周边的适当位置添加显著的提示标识，可以在视频末尾和中间适当位置添加显著的提示标识；

（五）呈现虚拟场景时，在起始画面的适当位置添加显著的提示标识，可以在虚拟场景持续服务过程中的适当位置添加显著的提示标识；

（六）其他生成合成服务场景根据自身应用特点添加显著的提示标识。

服务提供者提供生成合成内容下载、复制、导出等功能时，应当确保文件中含有满足要求的显式标识。

第五条 服务提供者应当按照《互联网信息服务深度合成管理规定》第十六条的规定，在生成合成内容的文件元数据中添加隐式标识，隐式标识包含生成合成内容属性信息、服务提供者名称或者编码、内容编号等制作要素信息。

鼓励服务提供者在生成合成内容中添加数字水印等形式的隐式标识。

文件元数据是指按照特定编码格式嵌入到文件头部的描述性信息，用于记录文件来源、属性、用途等信息内容。

第六条 提供网络信息内容传播服务的提供者应当采取下列措施，规范生成合成内容传播活动：

（一）核验文件元数据中是否含有隐式标识，文件元数据明确标明为生成合成内容的，采取适当方式在发布内容周边添加显著的提示标识，明确提醒公众该内容属于生成合成内容；

（二）文件元数据中未核验到隐式标识，但用户声明为生成合成内容的，采取适当方式在发布内容周边添加显著的提示标识，提醒公众该内容可能为生成合成内容；

（三）文件元数据中未核验到隐式标识，用户也未声明为生成合成内容，但提供网络信息内容传播服务的服务提供者检测到显式标识或者其他生成合成痕迹的，识别为疑似生成合成内容，采取适当方式在发布内容周边添加显著的提示标识，提醒公众该内容疑似生成合成内容；

（四）提供必要的标识功能，并提醒用户主动声明发布内容中是否包含生成合成内容。

有前款第一项至第三项情形的，应当在文件元数据中添加生成合成内容属性信息、传播平台名称或者编码、内容编号等传播要素信息。

第七条 互联网应用程序分发平台在应用程序上架或者上线审核时，应当要求互联网应用程序服务提供者说明是否提供人工智能生成合成服务。互联网应用程序服务提供者提供人工智能生成合成服务的，互联网应用程序分发平台应当核验其生成合成内容标识相关材料。

第八条 服务提供者应当在用户服务协议中明确说明生成合成内容标识的方法、样式等规范内容，并提示用户仔细阅读并理解相关的标识管理要求。

第九条 用户申请服务提供者提供没有添加显式标识的生成合成内容的，服务提供者可以在通过用户协议明确用户的标识义务和使用责任后，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。

第十条 用户使用网络信息内容传播服务发布生成合成内容的，应当主动声明并使用服务提供者提供的标识功能进行标识。

任何组织和个人不得恶意删除、篡改、伪造、隐匿本办法规定的生成合成内容标识，不得为他人实施上述恶意行为提供工具或者服务，不得通过不正当标识手段损害他人合法权益。

第十一条 服务提供者开展标识活动的，还应当符合相关法律、行政法规、部门规章和强制性国家标准的要求。

第十二条 服务提供者在履行算法备案、安全评估等手续时，应当按照本办法提供生成合成内容标识相关材料，并加强标识信息共享，为防范打击相关违法犯罪活动提供支持和帮助。

第十三条 违反本办法规定的，由网信、电信、公安和广播电视等有关主管部门依据职责，按照有关法律、行政法规、部门规章的规定予以处理。

第十四条 本办法自 2025 年 9 月 1 日起施行。

3.2 标识办法解读

《人工智能生成合成内容标识办法》（国信办通字〔2025〕2 号，以下简称《标识办法》）是国家互联网信息办公室、工业和信息化部、公安部、国家广播电视总局于 2025 年 3 月 7 日印发，自 2025 年 9 月 1 日起施行的政策文件。

《标识办法》旨在促进人工智能健康发展，规范人工智能生成合成内容标识，保护公民、法人和其他组织合法权益，维护社会公共利益。

（1）《标识办法》重点解决的问题

《标识办法》的推出重点解决“哪些是生成的”“谁生成的”“从哪里生成的”等关键问题，推动由生成到传播各环节的全流程管理，力争打造可信赖的人工智能技术。该办法采用了系统化的标识框架，通过显式与隐式标识相结合的方式，既确保用户直观感知内容属性，又保障技术追溯可能性。

（2）《标识办法》的核心目标

保障公众知情权：通过显式标识（如文字提示、角标）和隐式标识（如元数据嵌入），确保用户能够清晰区分 AI 生成内容与真实内容，减少误导风险。

压实主体责任：明确服务提供者、传播平台、用户三方责任，形成“生成-传播-使用”全链条监管。

推动技术合规发展：以“低成本标识”为原则，避免过度增加企业负担。

（3）明确了对服务提供者的要求

明确服务提供者应当对文本、音频、图片、视频、虚拟场景等生成合成内容添加显式标识，在提供生成合成内容下载、复制、导出等功能时，应当确保文件中含有满足要求的显式标识；应当在生成合成内容的文件元数据中添加隐式标识，隐式标识包含生成合成内容属性信息、服务提供者名称或者编码、内容编号等制作要素信息；应当在用户服务协议中明确说明生成合成内容标识的方法、样式等规范内容，并提示用户仔细阅读并理解相关的标识管理要求。

（4）明确了对互联网应用程序分发平台的要求

在应用程序上架或者上线审核时，应当要求互联网应用程序服务提供者说明是否提供人工智能生成合成服务，并核验其生成合成内容标识相关材料。若检测到其中明确标明该内容为 AI 生成合成，则应当在发布界面或内容周边自动加注醒目的标识，明确提示公众该内容为 AI 生成。

（5）明确了对个人用户的要求

用户作为内容的最终接收者和潜在的二次传播者，也被纳入标识管理体系。《标识办法》规定，用户向提供网络信息内容传播平台服务的提供者发布生成合成内容时，应当主动声明并使用平台提供的标识功能进行标识。

这一规定体现了标识责任的延伸，不仅要求原始服务提供者添加标识，也要求内容的二次传播者保持标识的完整性，形成全链条的标识管理。同时，这也是对用户社会责任的强调，鼓励用户在享受 AI 技术便利的同时，也承担相应的信息透明义务。

（6）明确了合规获得没有添加显式标识的生成合成内容的要求

充分考虑生成合成内容在实际场景中的落地应用，积极回应社会关切和产业需要，《标识办法》第九条提出，在用户主动要求提供未添加显式标识内容时，网站平台在不违反相关法律法规要求前提下，可通过在用户协议中明确责任义务并依法留存相关日志信息后，面向用户予以提供。同时，用户在后续使用过程中，需遵守《标识办法》第十条等相关要求，主动声明生成合成情况并添加显式标识后，方可面向公众发布和传播。