

## 新法速递 | 《生成式人工智能服务管理暂行办法》对比稿

CAI PENG

7月13日，网信办会同六部委发布《生成式人工智能服务管理暂行办法》（“《办法》”）。这是我国首次对生成式人工智能服务进行专门的规范，旨在促进生成式人工智能健康发展和规范应用。

《办法》共分为五章二十四条，涵盖了生成式人工智能服务的技术发展与治理、服务规范、监督检查和法律责任等方面的内容。

《办法》明确了提供和使用生成式人工智能服务应当遵守的基本原则和义务规定，并强调我国对生成式人工智能服务实行包容审慎和分类分级的监管原则。此前，我们曾对征求意见稿进行过全景式的逐条解读和分析（<https://www.zhonglun.com/Content/2023/04-12/1708534432.html>）。总体而言，我们认为正式稿反映了国家对人工智能发展的鼓励与支持态度，希望通过立法为相关服务提供者提供安全、宽松的发展环境。落实在具体法条中，我们认为《办法》的义务条款更具谦抑性，更加强调监管对生成式人工智能的鼓励态度。

《办法》将于2023年8月15日起施行。

为方便读者查阅此次立法变化，我们对征求意见稿与正式版《办法》制作了新旧对比表如下：

《生成式人工智能服务管理办法（征求意见稿）》	《生成式人工智能服务管理暂行办法》	修改方式
<p>第一条 为促进生成式人工智能健康发展和规范应用，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》等法律、行政法规，制定本办法。</p>	<p>第一条 为了促进生成式人工智能健康发展和规范应用，<b>维护国家和社会公共利益，保护公民、法人和其他组织的合法权益</b>，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》、<b>《中华人民共和国科学技术进步法》</b>等法律、行政法规，制定本办法。</p>	调整
<p>第二条 <del>研发、利用生成式人工智能产品，面向中华人民共和国境内公众提供服务的</del>，适用本办法。 <del>本办法所称生成式人工智能，是指基于算法、模型、规则生成文本、图片、声音、视频、代码等技术。</del></p>	<p>第二条 利用生成式人工智能<b>技术</b>向中华人民共和国境内公众提供<b>生成文本、图片、音频、视频</b>等服务（以下称生成式人工智能服务），适用本办法。</p> <p><b>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。</b></p> <p><b>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。</b></p>	调整+新增
/	<p>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，<b>对生成式人工智能服务实行包容审慎和分类分级监管。</b></p>	新增
<p>第四条 提供生成式人工智能产品或服务应当遵守法律法规的要求，尊重社会公德、<del>公序良俗</del>，<del>符合以下要求：</del></p> <p>（一）<del>利用生成式人工智能生成的内容应当体现社会主义核心价值观，不得含有颠覆国家政权、推翻社会主义制度，煽动分裂国家、破坏国家统一，宣</del></p>	<p>第四条 提供<b>和使用</b>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<b>伦理道德</b>，<b>遵守以下规定：</b></p> <p>（一）<b>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</b></p>	新增+调整

<p>扬恐怖主义、极端主义，宣扬民族仇恨、民族歧视，暴力、淫秽色情信息，虚假信息，<del>以及可能扰乱经济秩序和社会秩序的内容。</del></p> <p>(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，采取措施防止出现种族、民族、信仰、国别、地域、性别、年龄、职业等歧视。</p> <p>(三) 尊重知识产权、商业道德，不得利用算法、数据、平台等优势实施不公平竞争。</p> <p>(四) 利用生成式人工智能生成的内容应当真实准确，<del>采取措施防止生成虚假信息。</del></p> <p><del>(五) 尊重他人合法权益，防止伤害他人身心健康，损害肖像权、名誉权和个人隐私，侵犯知识产权。禁止非法获取、披露、利用个人信息和隐私、商业秘密。</del></p>	<p>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；</p> <p>(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；</p> <p>(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；</p> <p>(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；</p> <p>(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。</p>	
<p>/</p>	<p>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。</p>	<p>新增</p>
<p>第三条 国家支持人工智能算法、框架等基础技术的自主创新，<del>推广应用</del>国际合作，鼓励优先采用安全可信的软件、工具、<del>计算</del>和数据资源。</p>	<p>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。</p> <p>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</p>	<p>调整+新增</p>

	<p>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。</p>	
<p>第七条 提供者应当对生成式人工智能产品的预训练数据、优化训练数据来源的合法性负责。</p> <p>用于生成式人工智能产品的预训练、优化训练数据，应满足以下要求：</p> <p>（一）符合《中华人民共和国网络安全法》等法律法规的要求；</p> <p>（二）不含有侵犯知识产权的内容；</p> <p>（三）数据包含个人信息的，应当征得个人信息主体同意或者符合法律、行政法规规定的其他情形；</p> <p>（四）能够保证数据的真实性、准确性、客观性、多样性；</p> <p>（五）国家网信部门关于生成式人工智能服务的其他监管要求。</p>	<p>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：</p> <p>（一）使用具有合法来源的数据和基础模型；</p> <p>（二）涉及知识产权的，不得侵害他人依法享有的知识产权；</p> <p>（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；</p> <p>（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；</p> <p>（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。</p>	<p>调整</p>
<p>第八条 生成式人工智能产品研制中采用人工标注时，提供者应当制定符合本办法要求，清晰、具体、可操作的标注规则，对标注人员进行必要培训，抽样核验标注内容的正确性。</p>	<p>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。</p>	<p>调整+新增</p>
<p>第五条 利用生成式人工智能产品提供聊天和文本、图像、声音生成等服务的组织和个人（以下称“提供者”），包括通过提供可编程接口等方式支持他人自行生成文本、图像、声音等，承担该产品生成内容生产者的责任；涉及个人信息的，承担个人信息处理者的法定责任，履行个人信息保护义</p>	<p>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。</p> <p>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</p>	<p>调整+新增+删除</p>

<p>务。</p> <p><del>第九条 提供生成式人工智能服务应当按照《中华人民共和国网络安全法》规定，要求用户提供真实身份信息。</del></p>	<p>方权利义务。</p>	
<p>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，采取适当措施防范用户过分依赖或沉迷生成内容。</p>	<p>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，<b>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</b>，采取有效措施防范<b>未成年人用户过度依赖或者沉迷生成式人工智能服务</b>。</p>	调整
<p>第十一条 <del>提供者</del>在提供服务过程中，<del>对用户的</del>输入信息和使用记录<b>承担</b>保护义务。<del>不得非法留存能够推断出用户身份的输入信息，不得根据用户输入信息和使用情况进行画像，不得向他人提供用户输入信息。法律法规另有规定的，从其规定。</del></p>	<p>第十一条 提供者对<b>使用者</b>的输入信息和使用记录应当<b>依法履行</b>保护义务，<b>不得收集非必要个人信息</b>，不得非法留存能够<b>识别使用者身份</b>的输入信息<b>和使用记录</b>，不得<b>非法</b>向他人提供<b>使用者</b>的输入信息<b>和使用记录</b>。</p> <p>提供者应当依法及时受理和处理个人关于<b>查阅、复制、更正、补充、删除其个人信息</b>等的请求。</p>	调整+新增
<p><del>第十六条 提供者应当按照《互联网信息服务深度合成管理规定》对生成的图片、视频等内容进行标识。</del></p> <p><del>第十二条 提供者不得根据用户的种族、国别、性别等进行带有歧视性的内容生成。</del></p>	<p><b>第十二条</b> 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<b>生成内容</b>进行标识。</p>	调整+删除
<p><b>第十四条</b> 提供者应当在<b>生命周期内</b>，提供安全、稳健、持续的服务，保障用户正常使用。</p>	<p><b>第十三条</b> 提供者应当在<b>其服务过程中</b>，提供安全、稳定、持续的服务，保障用户正常使用。</p>	调整
<p><del>第十五条 对于运行中发现、用户举报的不符合本办法要求的生成内容，除采取内容过滤等措施外，应在3个月内通过模型优化训练等方式防止再次生成。</del></p> <p><del>第十九条 提供者发现用户利用生成式人工智能产品过程中违反法律法规，违背商业道德、社会公德</del></p>	<p><b>第十四条</b> 提供者发现<b>违法内容</b>的，应当<b>及时</b>采取<b>停止生成、停止传输、消除等</b>处置措施，采取<b>模型优化训练等措施进行整改</b>，并向有关主管部门<b>报告</b>。</p> <p>提供者发现<b>使用者</b>利用生成式人工智能<b>服务从事违法活动</b>的，应当<b>依法依约</b>采取<b>警示、限制功能、</b></p>	调整+合并

<p>行为时，包括从事网络炒作、恶意发帖跟评、制造垃圾邮件、编写恶意软件，实施不正当的商业营销等，应当暂停或者终止服务。</p>	<p>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。</p>	
<p><del>第十三条</del>提供者应当建立用户投诉接收处理机制，及时处置个人关于更正、删除、屏蔽其个人信息的请求；发现、知悉生成的文本、图片、声音、视频等侵害他人肖像权、名誉权、个人隐私、商业秘密，或者不符合本办法要求时，应当采取措施，停止生成，防止危害持续。</p>	<p><b>第十五条</b> 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。</p>	<p>调整+删除+新增</p>
<p>/</p>	<p><b>第十六条</b> 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。</p> <p>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，完善与创新相适应的科学监管方式，制定相应的分类分级监管规则或者指引。</p>	<p>新增</p>
<p><del>第六条</del>利用生成式人工智能产品向公众提供服务前，应当按照《具有舆论属性或社会动员能力的互联网信息服务安全评估规定》向国家网信部门申报安全评估，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。</p>	<p><b>第十七条</b> 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。</p>	<p>调整</p>
<p><del>第十八条</del> 提供者应当指导用户科学认识和理性使用生成式人工智能生成的内容，不利用生成内容损害他人形象、名誉以及其他合法权益，不进行商业炒作、不正当营销。</p> <p>用户发现生成内容不符合本办法要求时，有权向网信部门或者有关主管部门举报。</p>	<p><b>第十八条</b> 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。</p>	<p>删除+调整</p>
<p><del>第十七条</del> 提供者应当根据国家网信部门和有关主</p>	<p><b>第十九条</b> 有关主管部门依据职责对生成式人工</p>	<p>调整+新增</p>

<p><del>管部门的要求，提供可以影响用户信任、选择的必要信息，包括预训练和优化训练数据的来源、规模、类型、质量等描述，人工标注规则，人工标注数据的规模和类型，基础算法和技术体系等。</del></p>	<p>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。</p> <p>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。</p>	
<p>/</p>	<p>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。</p>	<p>新增</p>
<p><del>第三十条</del>提供者违反本办法规定的，由网信部门和有关主管部门按照《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》等法律、行政法规的规定予以处罚。</p> <p>法律、行政法规没有规定的，由网信部门和有关主管部门依据职责给予警告、通报批评，责令限期改正；拒不改正或者情节严重的，责令暂停或者终止其利用生成式人工智能提供服务，并处一万元以上十万元以下罚款。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。</p>	<p>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。</p> <p>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。</p>	<p>调整</p>
<p><del>第二条</del>研发、利用生成式人工智能产品，面向中华人民共和国境内公众提供服务的，适用本办法。</p> <p>本办法所称生成式人工智能，是指基于算法、模型、规则生成文本、图片、声音、视频、代码等内容的技术。</p>	<p>第二十二条 本办法下列用语的含义是：</p> <p>（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。</p> <p>（二）生成式人工智能服务提供者，是指利用生</p>	<p>新增+调整</p>

	<p>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。</p> <p>（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。</p>	
/	<p>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。</p> <p>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。</p>	新增
<del>第三十一条</del> 本办法自 2023 年 月 日起实施。	<del>第二十四条</del> 本办法自 2023 年 8 月 15 日起施行。	调整