汉坤法律评述
2023 年 7 月 14 日

HANKUN
汉 坤 律 师 事 务 所
Han Kun Law Offices

平衡与发展：《生成式人工智能服务管理暂行办法》正式发布

作者：段志超 | 蔡克蒙 | 金今

2023 年 7 月 13 日，国家互联网信息办公室会（“**网信办**”）同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了《生成式人工智能服务管理暂行办法》（“**《暂行办法》**”），将于 2023 年 8 月 15 日正式生效实施。《暂行办法》在网信办此前 4 月 11 日公布征求意见的《生成式人工智能服务管理办法》（征求意见稿）（“**《征求意见稿》**”）的基础上广泛吸收了公众反馈意见，做出了大幅度修改。与《征求意见稿》相比，《暂行办法》体现出了对尚不成熟的生成式人工智能服务更多的包容，更侧重鼓励技术发展创新，在原理和制度层面更好地统筹兼顾了发展与安全。本文将从适用范围、监管路径、与现有制度对接、数据训练、服务应用、外资准入等多方面简析《暂行办法》，着重关注《暂行办法》在《征求意见稿》基础上的优化创新及其潜在影响。

一、适用范围强调向“境内公众”提供服务，排除研发和内部应用

《暂行办法》第 2 条规定“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、视频等内容的服务（以下称“**生成式人工智能服务**”），适用本办法”，《暂行办法》强调其规制对象为“向境内公众”提供生成式人工智能服务。相较于《征求意见稿》，《暂行办法》吸纳了公众意见，明确将未向境内公众提供服务的生成式人工智能技术的研发、应用排除出适用范围。前者大大减轻了模型研发阶段的合规负担，后者则缓解了许多企业接入生成式人工智能服务用于改善工作效率等内部应用目的时的合规顾虑，体现了《暂行办法》审慎包容、鼓励创新的监管思路。

二、包容审慎、分类分级监管的路径，强调统筹协调多部门监管

《暂行办法》提出了对生成式人工智能服务试行“包容审慎和分级分类”的监管思路。《暂行办法》新增了《中华人民共和国科学技术进步法》作为立法依据，更加突出了鼓励科技创新的政策导向。此外，《暂行办法》还增加了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、国家广播电视总局等多部门作为相关的监管机关，规定各部门依职权加强对生成式人工智能服务的管理。

分级分类的监管思路可能借鉴了欧盟《人工智能法案》（草案）将人工智能系统分为不可接受风险、高风险、有限风险的规定。由于生成式人工智能具有通用性，“包容审慎和分级分类”的监管思路有助于《暂行办法》作为生成式人工智能领域的“基本法”保留一定灵活性，各监管部门、行业主管部门、标准化组织亦可以在此基础上制定更加细化的生成式人工智能分级分类规则，并针对特定行业、特定应用或某些高风险的生成式人工智能服务制定更为严格的规范。此外，《暂行办法》针对生成式人工智能服务一些主要的应用

场景，规定利用生成式人工智能服务从事新闻出版、影视制作、文艺创作等活动需遵守相关领域的监管规定，与现有制度对接（第 2 条）。

三、针对业界发展生成式人工智能的实际问题、利好创新的政策措施

在美国对华技术“脱钩”、“卡脖子”，中国企业获取先进芯片、算力存在诸多障碍的大背景下，《暂行办法》针对生成式人工智能研发、应用提出了一系列政策鼓励措施，包括：

- 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。
- 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。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。促进算力资源协同共享，提升算力资源利用效能（第 5 条、第 6 条）。

目前，一些地方已在算力、数据等基础设施建设、统筹方面走在前列。例如 2023 年 5 月发布的《北京市加快建设具有全球影响力的人工智能创新策源地实施方案（2023-2025 年）》提出了“提升算力资源统筹供给能力”、“加强公共数据开放共享”、“构建高效协同的大模型技术产业生态”等方面的具体鼓励措施。

四、数据训练、模型应用和优化方面具体制度上的适度“松绑”

《暂行办法》吸取了业界的反馈意见，更充分考虑了当前训练数据质量、生成内容可靠性、准确性等方面的技术障碍与局限，适度放宽了生成式人工智能数据训练、模型应用和优化的具体合规要求，主要体现在以下方面：

- 《暂行办法》适当放宽了对于训练数据质量的要求。第 7 条将原《征求意见稿》中“保证”数据“真实性、准确性、客观性、多样性”改为“采取有效措施增强”，减轻了服务提供者在训练数据质量方面的责任。
- 删除《征求意见稿》第 4 条要求“生成内容真实准确，采取措施防止生成虚假信息”，不得含有“可能扰乱经济秩序和社会秩序的内容”，改为要求提供者“基于服务类型特点采取有效措施，提升生成内容的准确性和可靠性”，一定程度上减轻了服务提供者在生成内容上的责任。
- 删除《征求意见稿》第 9 条真实身份验证义务的要求。这可能是考虑到当前生成式人工智能服务主要是“对话式”而非“发布式”的特点，且许多通过提供可编程接口（API）提供生成式人工智能服务情况下无需也难以落实真实身份验证义务的情况。但另一方面，如果生成式人工智能被用于提供互联网信息服务，仍可能需要依据相关监管规定履行实名身份认证义务。
- 《暂行办法》第 11 条删除了不得进行用户画像的要求，并将禁止向他人提供使用者输入信息改为不得收集非必要信息和“非法”向他人提供使用者输入信息，即，排除了用户同意或者法律法规另有规定的情形。这一修订更加符合《个人信息保护法》的知情同意和必要性原则，也为服务提供者

使用用户输入信息优化模型提升服务质量留下了更多空间。

- 《暂行办法》减轻了违法违规内容的监测及处置义务，删除了《征求意见稿》第 13 条对提供者“发现、知悉”违法违规内容时即应采取措施停止生成的规定，放宽为要求建立投诉举报机制并及时处理违法违规信息，“采取模型优化训练等措施进行整改”；《暂行办法》还删除了受到广泛争议的《征求意见稿》第 13 条设定的 3 个月内通过模型优化训练等防止再次生成违法内容的严格时限，留下了一定的灵活空间。

- 《暂行办法》减轻了服务提供者在算法透明度的义务，删除了《征求意见稿》第 17 条对提供者“提供可以影响用户信任、选择的必要信息”的详细规定，仅通过第 4 条第 5 项要求提供者“采取有效措施提高算法透明度”，使得服务提供者可以通过更加灵活的方式探索提高算法透明度。

- 《征求意见稿》要求利用生成式人工智能产品向公众提供服务前进行具有舆论属性或者社会动员能力互联网信息服务安全评估。《暂行办法》第 17 条则澄清需履行安全评估义务的主体为“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提供者，适当限缩了需进行安全评估的范围，与已有规范保持一致。

五、外资准入和境外服务

《暂行办法》第 20 条规定，“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。”结合《暂行办法》第 2 条关于适用范围的规定，我们倾向于认为本条主要旨在限制不符合《暂行办法》要求的境外生成式人工智能向境内公众提供服务。在此情况下，主管部门可能采取屏蔽等技术措施阻断对相关境外服务网站、应用的访问。此外，如果境内服务提供者将境外生成式人工智能服务嵌入自己的产品向境内公众提供服务，则需要遵守《暂行办法》的相关规定，否则主管部门可能依据《暂行办法》第 21 条对境内服务提供者进行处罚。

《暂行办法》第 23 条新增规定“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。”目前，法律法规并未对提供生成式人工智能服务本身设定行政许可或外资准入限制，但如果生成式人工智能服务被应用于存在许可或外资准入的领域，如提供增值电信服务、网络视听节目服务、互联网文化经营等，则需遵守相关许可或市场准入规定。

六、影响与展望

综合来看，监管部门从善如流，《暂行办法》成稿体现了产业界和公众对《征求意见稿》提出的许多建议，更充分的考虑当前生成式人工智能的技术局限，在明确安全规范“红线”的基础上，基于“审慎包容”的原则适度放松了生成式人工智能从研发、模型训练到应用、优化各阶段的合规要求，体现出了鼓励新技术发展、应用的政策导向。但从具体规则设置来看，《暂行办法》在训练数据合规性、生成内容安全准确、透明度等方面的要求需要企业结合技术与法律力量提出创造性的解决方案，以缓解监管机构的安全顾虑，为产业发展赢得更多的制度空间。

特别声明

汉坤律师事务所编写《汉坤法律评述》的目的仅为帮助客户及时了解中国或其他相关司法管辖区法律及实务的最新动态和发展，仅供参考，不应被视为任何意义上的法律意见或法律依据。

如您对本期《汉坤法律评述》内容有任何问题或建议，请与汉坤律师事务所以下人员联系：

段志超

电话： +86 10 8516 4123
Email： kevin.duan@hankunlaw.com

蔡克蒙

电话： +86 10 8516 4289
Email： kemeng.cai@hankunlaw.com