

	DIGICHINA		

Home » Work »

Translation: Measures for the Management of Generative Artificial Intelligence Services (Draft for Comment) – April 2023

Novel rules about training data and accuracy of generated media circulated for comment



by:
Seaton Huang

Published
April 12, 2023

Helen Toner

Zac Haluza

Rogier Creemers

Graham Webster

This translation is by (in randomized order), Seaton Huang, Helen Toner, Zac Haluza, and Rogier Creemers, and was edited by Graham Webster. During editing, an alternative translation from China Law Translate was consulted.

For analysis of this draft, please see our DigiChina Forum, compiling analysis from a group of invited specialists.

Translation

April 11, 2023

Measures for the Management of Generative Artificial Intelligence Services (Draft for Comment)

Article 1: In order to stimulate the healthy development and standardized application of generative artificial intelligence (AI), on the basis of the Cybersecurity Law of the People's Republic of China, the Data Security Law of the People's Republic of China, the Personal Information Protection Law of the People's Republic of China, and other such laws and administrative regulations, these Measures are formulated.

Article 2: These Measures apply to the research, development, and use of products with generative AI functions, and to the provision of services to the public within the [mainland] territory of the People's Republic of China.

Generative AI, as mentioned in these Measures, refers to technologies generating text, image, audio, video, code, or other such content based on algorithms, models, or rules.

Article 3: The State supports indigenous innovation, broad application, and international cooperation in foundational technologies such as AI algorithms and frameworks, and encourages the prioritized use of secure and reliable software, tools, computing, and data resources.

Article 4: The provision of generative AI products or services shall abide by the requirements of laws and regulations, respect social virtue and good public customs, and conform to the following requirements:

1. Content generated through the use of generative AI shall reflect the Socialist Core Values, and may not contain: subversion of state power; overturning of the socialist system; incitement of separatism; harm to national unity; propagation of terrorism or extremism; propagation of ethnic hatred or ethnic discrimination; violent, obscene, or sexual information; false information; as well as content that may upset economic order or social order.
2. In processes such as algorithm design, selecting training data, model generation and model optimization, service provision, etc., adopt measures to prevent the emergence of discrimination on the basis of race, ethnicity, religious belief, nationality, region, sex, age, or profession.
3. Respect intellectual property rights and commercial ethics; advantages in algorithms, data, platforms, etc., may not be used to engage in unfair competition.
4. Content generated through the use of generative AI shall be true and accurate, and measures are to be adopted to prevent the generation of false information.
5. Respect the lawful rights and interests of others; prevent harm to the physical and mental health of others, infringement of their likeness rights, reputation rights and personal privacy, as well as infringement of intellectual property rights. It is prohibited to illegally obtain, divulge or use personal information and private [information], as well as commercial secrets.

Article 5: Organizations or individuals that use generative AI to provide services such as chat, text, image, or audio generation (hereinafter referred to as “providers”); including providing programmable interfaces [i.e., APIs] and other means which support others to themselves generate text, images, audio, etc.; bear responsibility as the producer of the content generated by the product. Where personal information is involved, they bear legal responsibility as personal information handlers and are to fulfill personal information protection obligations.

Article 6: Before using generative AI products to provide services to the public, a security assessment must be submitted to the state cyberspace and information department [i.e., the Cyberspace Administration of China] in accordance with the Provisions on the Security Assessment of Internet Information Services With Public Opinion Properties or Social Mobilization Capacity, and the procedures of algorithm filing, modification, and cancellation

of filing must be carried out in accordance with the Internet Information Service Algorithmic Recommendation Management Provisions.

Article 7: Providers shall bear responsibility for the legality of the sources of generative AI product pre-training data and optimization training data.

Data used for generative AI product pre-training and optimization training shall satisfy the following requirements:

1. Conforming to the requirements of the Cybersecurity Law of the People's Republic of China and other such laws and regulations;
2. Not containing content infringing intellectual property rights;
3. Where data includes personal information, the consent of the personal information subject shall be obtained, or other procedures conforming with the provisions of laws and administrative regulations followed;
4. Be able to ensure the data's veracity, accuracy, objectivity, and diversity;
5. Other supervision requirements of the state cybersecurity and informatization department concerning generative AI functions and services.

Article 8: When human annotation is used in the development of generative AI products, providers shall formulate clear, specific, and practicable annotation rules conforming to the requirements of these Measures; necessary training of annotation personnel shall be conducted; and the validity of annotation content shall be spot checked.

Article 9: When providing generative AI services, users shall be required to provide real identity information in accordance with the provisions of the Cybersecurity Law of the People's Republic of China.

Article 10: Providers shall explicitly disclose the user groups, occasions, and uses for their services, and adopt appropriate measures to prevent users from excessive reliance on or addiction to generated content.

Article 11: In the process of providing services, providers have the duty to protect information input by users and usage records. They may not illegally preserve input information from which it is possible to deduce the identity of users, they may not conduct profiling on the basis of information input by users and their usage details, and they may not provide information input by users to others. Where laws or regulations provide otherwise, those provisions are to be followed.

Article 12: Providers may not engage in content generation that is discriminatory based on a user's race, nationality, sex, etc.

Article 13: Providers shall establish mechanisms for receiving and handling user complaints and promptly handle individual requests concerning revision, deletion, or masking of their personal information; and when they discover or learn that generated text, images, audio, video, etc., infringe other persons' likeness rights, reputation rights, personal privacy, or commercial secrets, or do not conform to the demands of these Measures, they shall adopt measures to cease generation and prevent the expansion of the harm.

Article 14: Providers shall, throughout the lifecycle, provide secure, stable and sustained services, and ensure users' normal usage.

Article 15: When generated content that does not conform to the requirements of these Measures is discovered during operations or reported by users, aside from adopting content filtering and other such measures, repeat generation is to be prevented through such methods as optimization training within three months.

Article 16: Providers shall mark generated images, videos, and other content in accordance with the Internet Information Service Deep Synthesis Management Provisions.

Article 17: Providers shall, in accordance with the requirements of the state cybersecurity and informatization department and relevant responsible departments, provide necessary information that could influence users trust or choices, including descriptions of the source, scale, type, quality, etc., of pre-training and optimization training data; rules for human annotation; the scale and type of human-annotated data; and foundational algorithms and technological systems.

Article 18: Providers shall guide users to scientifically understand and rationally use content generated by generative AI; not to use generated content to damage others' image, reputation, or other lawful rights and interests; and not to engage in commercial hype or improper marketing.

When users discover generated content that does not meet the requirements of these measures, they have the right to report this to cybersecurity and informatization departments or relevant responsible departments.

Article 19: If a provider finds that a user has used generative AI products to violate laws or regulations; violate business ethics or social virtue, including engaging in online hype,

malicious posting and commenting, creating spam, or writing malicious software; or engage in improper business marketing; etc.; service shall be suspended or terminated.

Article 20: If a provider violates the provisions of the Measures, the cybersecurity and informatization department and relevant responsible departments are to impose penalties in accordance with the provisions of Cybersecurity Law of the People's Republic of China, the Data Security Law of the People's Republic of China, the Personal Information Protection Law of the People's Republic of China, and other such laws and administrative regulations.

Where there are no provisions of law or administrative regulation, the cybersecurity and informatization department and relevant responsible departments are to, in accordance with their duties, issue warnings, circulate criticisms, and order corrections within a set period of time. Where corrections are refused or circumstances are grave, they are to order suspension or termination of their use of generative AI provider services, and a penalty more than 10,000 yuan and less than 100,000 yuan is to be imposed. Where behavior constitutes a violation of public security management, public security management penalties are to be imposed in accordance with the law. Where a crime is constituted, criminal responsibility shall be pursued in accordance with the law.

Article 21: These measures are effective beginning [day] [month], 2023.

Chinese-language Original

[[Source](#)] [[Archived copy](#)]

生成式人工智能服务管理办法 (征求意见稿)

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。

第二条 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。

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。

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、推广应用、国际合作，鼓励优先采用安全可信的软件、工具、计算和数据资源。

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、公序良俗，符合以下要求：

（一）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，以及可能扰乱经济秩序和社会秩序的内容。

（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。

（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。

（四）利用生成式人工智能生成的内容应当真实准确，采取措施防止生成虚假信息。

（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。

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务。

第六条 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。

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。

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：

（一）符合《中华人民共和国网络安全法》等法律法规的要求；

（二）不含有侵犯知识产权的内容；

（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；

(四) 能够保证数据的真实性、准确性、客观性、多样性;

(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。

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。

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。

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。

第十一条 提供者在提供服务过程中, 对用户的输入信息和使用记录承担保护义务。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。

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。

第十三条 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。

第十四条 提供者应当在生命周期内, 提供安全、稳健、持续的服务, 保障用户正常使用。

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。

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。

第十七条 提供者应当根据国家网信部门和有关主管部门的要求, 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。

第十八条 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。

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。

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。

第二十条 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。

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。

第二十一条 本办法自2023年 月 日起实施。