生成式AI新规亮点：有分寸、重实际、留空间、促发展

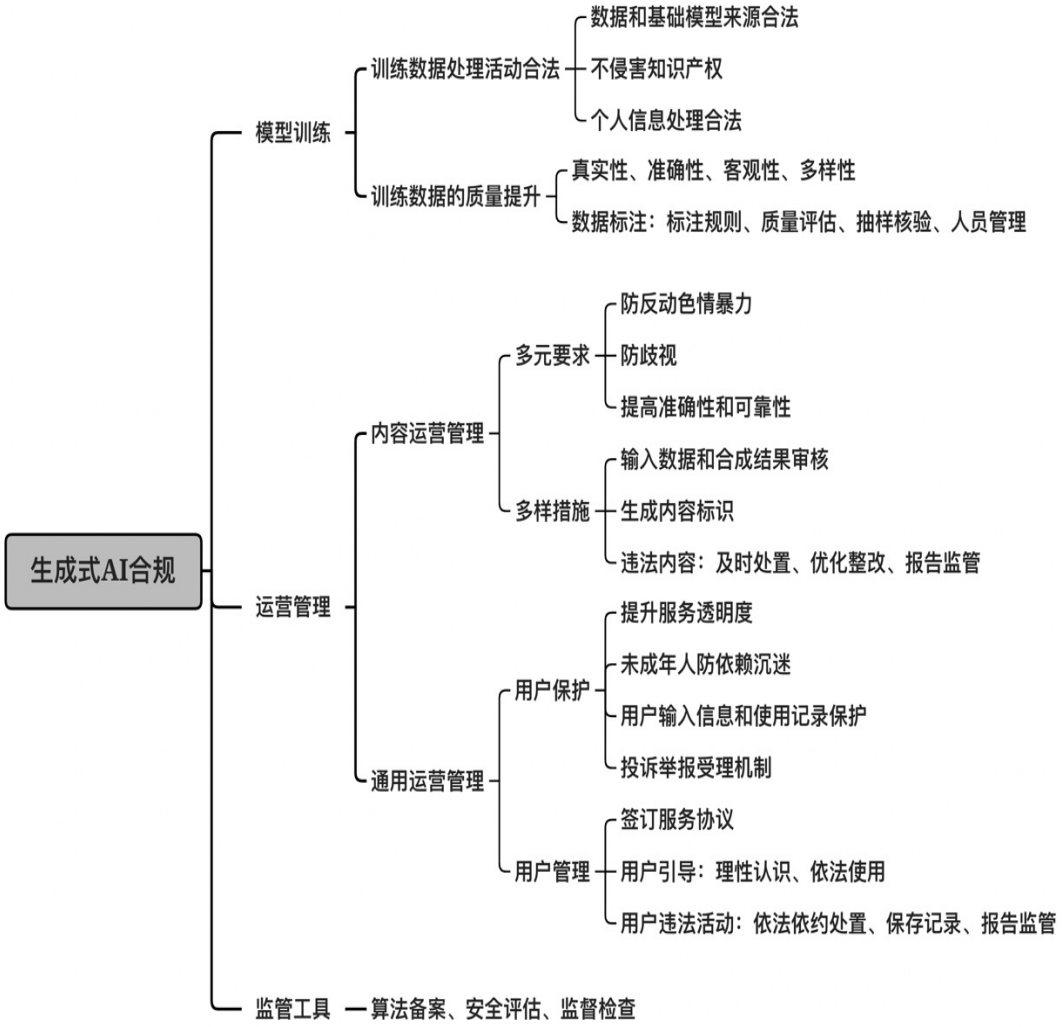
作者：杨建媛 李天烁

2023年7月13日，国家网信办、发改委、教育部等多部委联合发布了《生成式人工智能服务管理暂行办法》（“**《办法》**”）。相较于《生成式人工智能服务管理办法（征求意见稿）》（**“征求意见稿”**），《办法》的显著特征如下：

（1）将“坚持发展和安全并重、促进创新和依法治理相结合”作为基本原则，尊重技术发展的历史阶段，从实际出发合理减轻企业的合规负担，定下了助推生成式AI发展的总体基调。

（2）将不少结果性要求放宽为过程性要求，但明确强调合规措施的有效性，有效的合规措施将成为重要生产力助力生成式AI企业成为行业龙头。

本文拟基于海问自行总结的监管架构（见下图），就《办法》相较于征求意见稿的主要迭代亮点作出解读，帮助读者快速掌握生成式AI立法的最终基调。



一、适用范围：明确排除未向境内公众提供服务场景、客观开展外资监管

我们理解，**《办法》将适用范围明确为面向境内公众提供生成式AI服务，体现了保护个人发展、维护国家安全和公共利益的基本考量**。基于生成式AI技术的固有缺陷，现有技术条件尚无法避免其生成虚假和有害内容。一旦面向境内公众提供服务，则意味着潜在风险的大规模扩散。此处，问题关键即在于对“公众”的理解。

（一）内容和技术服务提供者均将受到监管

不同于《互联网信息服务深度合成管理规定》明确区分了“深度合成服务提供者”和“深度合成服务技术支持者”两类义务主体，《办法》仅规定了“生成式人工智能服务提供者”。鉴于《办法》明确将“通过提供可编程接口等方式”提供服务纳入监管范围，亦对模型训练作出了相应监管要求，因而我们理解，**除了直接提供服务的内容服务提供者，技术服务提**

供者也被涵盖在“生成式人工智能服务提供者”的范围项下。

至于《办法》未区分技术服务提供者和内容服务提供者的原因，或在于当前尚无法就最终生成内容实现明确的责任分割。当前实践中的业态较为复杂，最终违法内容的生成可能来源于多方因素，区分义务主体的同时即需要在规范层面对违法内容生成责任作出统一划分，但在业务实践快速发展的情况下预设规则难免与实践情况存在冲突之处。因此《办法》对“生成式人工智能服务提供者”作出宽泛定义，实践中的具体责任分配有待个案判定。

（二）未向境内公众提供服务的研发、自用明确不适用《办法》

相较于征求意见稿，《办法》新增第2条第三款，明确将企业、教育和科研机构、公共文化机构等研发、应用生成式AI技术，未向境内公众提供生成式AI服务的场景排除在适用范围之外。除了对技术创新具有关键作用的科研领域，生成式AI在企业内部的应用可以大幅提高企业生产效率，对我国数字经济发展具有重要意义。结合我们的项目经验，能够节省成本且提高专业化程度的行业大模型业已成为新的关注热点。

备受关注的一个问题是：to B服务是否可以豁免适用？解题核心在于如何理解“公众”。中国法语境下，“公众”是一个泛指，涵盖了个人和集体（例如组织、团体等），故解读法条难以直接得出to B服务可以豁免适用的结论。但就具体的场景该如何判断，则值得多推敲，例如：平台只给其上商户提供服务，商户也不会基于此再面向C提供服务，相当于只是平台内嵌的服务工具。此种情况和内部使用较为类似，而与一般的to B服务有所区别。

（三）将境外服务、外商投资客观纳入监管范畴

相较于征求意见稿，《办法》新增了外商投资条款，要求其符合外商投资相关法律、行政法规的规定，但目前针对生成式AI的外商投资规定尚不明确。同时，明确规定若境外向境内提供的生成式AI服务违反中国法律、法规、《办法》，国家网信部门应通知有关机构采取技术措施和其他必要措施予以处置，此条既为对监管部门的授权性规定，亦为对提供者的警示性规定。对于“封装”类的境内服务提供者，其境外技术支持存在因合规问题而中断的可能。《办法》明确将对境外服务及外商投资进行监管但视角较为客观，一方面注重针对境内公众、公共利益的全面保护，另一方面也有利于促进国内生成式AI市场的竞争。

二、模型训练：提出多层次规范体系、数据质量要求更加务实

《办法》针对模型训练提出多层次规范体系，与生成式AI算法备案中特别关注训练数据相关的合规制度建设一脉相承，原因在于训练数据为生成式AI所特别倚重。**《办法》的特有创新规定**（数据质量要求、使用具有合法来源的数据和基础模型等）、**重点法律的联动规定**（知识产权保护、个人信息处理合法等）、**相关规范的兜底规定**（《网络安全法》《数据安全法》等）**共同构成了针对训练数据的多层次规范体系。**

如上文所言，内容和技术服务提供者均为《办法》项下的义务主体，但**在模型训练阶段，相关义务责任则仅落于技术服务提供者，该等义务分配与现实情况相符。**《办法》第7条明确针对“训练数据处理活动”提出要求，意味着负责模型训练的技术服务提供者应当遵守相应义务并承担责任，而既不参与模型训练、亦不从事训练数据处理活动的内容服务提供者则被排除在外。

（一）仅就自身训练数据获取这一次流转的合法性负责

征求意见稿要求“提供者对训练数据来源的合法性负责”，而《办法》第7条将其修改为“使用具有合法来源的数据和基础模型”。对此我们理解，**《办法》一方面对“基础模型来源合法”增设要求，另一方面也明确了提供者对“训练数据来源合法”的合规义务边界。**

所谓“数据来源合法”，是指通过合法手段、而非窃取等非法方式获取数据。在征求意见稿中，数据来源合法要求指向训练数据本身，而该训练数据可能历经多次流转，如果其中某次流转存在非法获取的情况（即存在原罪），不排除提供者需为此承担责任的可能。而在《办法》中，**数据来源合法要求明确指向本次训练数据使用行为，关注重点聚焦于提供者为使用而获取训练数据这一次流转经历**，提供者的责任范围由此得到了限缩。

除此之外，**《办法》提出推动公共训练数据资源平台建设，亦将为“数据来源合法”提供基础保障。**《办法》第6条第二款规定，要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。与之相对，2023年5月23日出台的《北京市促进通用人工智能创新发展的若干措施》中亦提及将从“归集高质量基础训练数据集”、“谋划建设数据训练基地”、“搭建数据集精细化标注众包服务平台”三方面入手，提升高质量数据要素供给能力。随着数据垄断效应的加剧，企业若想合法获取高质量的训练数据集并非易事，而日后公共训练数据集的建立和开放或将有效缓解这一矛盾，降低企业获取具有合法来源的数据的成本。

不过，**深入探究“数据来源合法”的含义，会发现其中亦存在模糊之处。**例如，在服务提供者从第三方购买训练数据集的情况下，在数据交易所进行交易是否就必然可被认定为数据来源合法？如果是场外交易，仅靠要求数据提供方作出合规承诺，是否即满足要求？服务提供者事先是否需要进行尽职调查，是否需要数据提供方出示相关凭证以作证明？此类实操问题仍有待实践中进一步明确。

（二）不再要求对训练数据质量作出保证

相较于征求意见稿，**《办法》对训练数据的质量要求从“保证数据的真实性、准确性、客观性、多样性”软化为了“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”。**训练数据的质量对于最终生成内容的质量具有重要影响，如果训练数据中即存在大量虚假、有害信息，那最终生成结果也难以言及真实准确，因而有必要对训练数据作出要求。但是，由于模型训练需要海量数据，要求服务提供者对数据质量作出保证不仅会带来高昂的合规成本，亦缺乏可行性。不过还需注意的是，**《办法》对数据质量提升措施提出了“有效性”要求**，因此仍不可大意相关合规措施的具体落实。

除此之外，相较于征求意见稿提出的人工标注规则和抽样核验要求，**《办法》不仅关注到数据标注不仅限于人工标注的实践情况（实际上多数标注均非人工完成），扩张了数据标注要求的适用范围，还新增设了数据标注质量评估要求。**我们理解，日后此项亦将成为安全评估和算法备案的审查要点，建议企业事先作出合规部署。

三、内容运营管理：内容真实向善为目标、措施要求可落地性增强

在内容运营管理方面，**《办法》具有明确的监管目标，即防范虚假有害、内容向上向善**，但在同时，立法者也意识到在当前阶段要求一蹴而就不仅不现实，亦可能将新技术扼杀在摇篮之中，因而选择**尊重实际，旨在以循序渐进的方式达到监管目标**。例如，《办法》删除了征求意见稿中有关“歧视性内容生成”的禁止性规定，但仍保留了第4条中防歧视的原则性条款，并为此添加“措施有效性”的限定条件，亦即此意。

（一）放宽生成内容的真实准确要求

《办法》删除了征求意见稿中争议较大的“生成内容应当真实准确”要求，将其修改为了“提高生成内容的准确性和可靠性”。目前而言，“幻觉”源于大模型的固有技术缺陷，实践中尚且无法避免，在此背景下要求生成内容完全真实准确并不现实，因而《办法》对征求意见稿进行了修改，将结果性要求放宽为过程性要求。与之相对，企业应采取内容审核等合规措施证明其尽到了相应义务。

例如，OpenAI采用了两种技术方法以降低模型幻觉出现的频率。对于开放域的幻觉，OpenAI收集了ChatGPT被用户标为不真实的现实世界数据，以及用于训练奖励模型的对比标注数据；对于封闭域的幻觉，OpenAI设计了一系列步骤以使用GPT-4自行生成对比数据，并最终将生成结果准确性提升了30%左右。[1]

（二）删除模型优化训练的效果要求和时间限定

《办法》要求服务提供者在发现违法内容后采取模型优化训练等措施进行整改，删除了征求意见稿中“3个月内通过模型优化训练等方式防止再次生成”的要求。我们理解，无论是“3个月内”的时间限定，还是“防止再次生成”的效果目标，都对服务提供者提出了较为严苛的技术要求，实践中可能难以落地，所以《办法》对该条款作出了修改。不过另需注意的是，**《办法》新增了向主管部门报告的义务要求**，对此需要提供者建立相应机制以作回应。

四、通用运营管理：用户保护更加精细、用户管理契合实践

在通用运营管理问题上，体系化衔接和精细化处理是《办法》的突出特点。具体而言：

《办法》与《个人信息保护法》《互联网信息服务算法推荐管理规定》《网络安全法》等法律规范进行了体系化衔接，并据此作出了更精细、周全的规定。例如，《办法》第11条在依据《个人信息保护法》增设用户行权保障规定的同时，还将不得非法留存用户输入信息的范围从"推断用户身份"限缩为了"识别用户身份"。如下所列要点亦为此理。

（一）"防依赖沉迷义务"明确指向未成年人用户群体

征求意见稿第10条规定，提供者应"采取适当措施防范用户过分依赖或沉迷生成内容"，《办法》第10条将该表述修改为提供者应"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"。

具体而言，《办法》从保护对象和措施效果两方面入手，对上述防依赖沉迷义务进行了修订：

从保护对象来看，《办法》将保护对象具体界定为未成年人用户。正如我们在之前的文章《生成式AI（二）：体系化构建合规指南》中所分析，结合《互联网信息服务算法推荐管理规定》的未成年人保护条款进行体系解释，加之未成年人用户心智发育尚不成熟，更容易受到生成式AI影响，**建立未成年人保护机制是生成式AI服务防依赖沉迷义务的应然之义**。《办法》的规定明确印证了我们的观点。

从措施效果来看，一方面，《办法》将"适当措施"修改为"有效措施"，提高了合规措施的效果标准；另一方面，《办法》将"依赖或沉迷生成内容"修改为"依赖或沉迷生成式人工智能服务"，表述更加精准且合理。因为相较于具体的生成内容，未成年人更可能过度依赖或沉迷于服务功能本身，导致无法充分锻炼独立思考能力，阻碍未成年人的健康成长，当前世界范围内多所学校禁用ChatGPT亦为此理。

（二）新增服务协议签订义务

相较于征求意见稿，《办法》第9条要求服务提供者与使用者签订服务协议，明确双方权利义务。目前而言，**签订服务协议已成为提供生成式AI服务的实践惯例，该项规定继而将此举从实践惯例上升到了法定义务层面**。我们理解，服务协议不仅对划定双方权利义务具有重要意义，亦会落入安全评估、算法备案中用户权益保障情况的审查范畴。对此，我们将结合自身实务经验，在后续推送中专就生成式AI服务法律文本的合规要点作出解读，敬请期待。

（三）违法行为处置不再限于"暂停或终止服务"

相较于征求意见稿，《办法》在违法行为处置问题上给服务提供者提供的**措施选项不再限于"暂停或终止服务"，新增了"警示、限制功能"等措施选项**。我们理解，如果一旦发现用户利用服务从事违法或背德行为，无论程度轻重就施以暂停或终止服务的惩戒，未免过于严苛。《办法》给予了服务提供者

一定的自主研判空间，使其可以根据用户行为的严重程度施以相应的处置措施，更具合理性。

五、监管工具：基于风险的分类分级治理

《办法》采用了基于风险的分类分级监管方法，将服务的舆论属性或社会动员能力作为风险程度的判断标准，要求达到标准的服务提供者进行安全评估和算法备案。相较而言，征求意见稿仅规定在提供生成式AI服务前，应按照相应规定进行安全评估和算法备案——如此存在理解上的分歧：一种理解为凡是提供生成式AI服务即需进行安全评估和算法备案，另一种理解则是，符合相关规定中"具有舆论属性或社会动员能力"条件的生成式AI服务才需进行安全评估和算法备案。《办法》修改了征求意见稿的表述，明确采用了后一种理解，在规范层面上减轻了服务提供者的合规负担。

《生成式人工智能服务管理暂行办法》

与征求意见稿对比汇总

《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》
图例：删除 新增 调整	
	第一章 总则
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，维护国家安全和社会公共利益，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 （第一款）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。	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、视频等内容的服务（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）尊重他人合法利益，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 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 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，采取适当措施防范用户过分依赖或沉迷生成内容。 **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。~~	**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~~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，应当依法依约采取警示、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。
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 ~~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** 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

第二条 （第二款）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。	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。
第二十一条 本办法自2023年 月 日起实施。	第二十四条 本办法自2023年8月15日起施行。
《暂行办法》删除条款	
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。~~	

[1] OpenAI: GPT-4 Technical Report, https://arxiv.org/pdf/2303.08774.pdf

2023-07-14