君合专题研究报告 JUNHE

2023 年 4 月 17 日

生成式 AI 监管征求意见稿对行业影响的几点思考

引言

2023 年 4 月 11 日，国家互联网信息办公室（“网信办”）发布《生成式人工智能服务管理办法（征求意见稿）》（“**《征求意见稿》**”），向社会公开征求意见。

随着 ChatGPT 横空出世引爆生成式人工智能（或称“**生成式 AI**”或“**AIGC**”）产品浪潮，与之相伴的数据滥用、侵犯隐私、虚假信息和道德伦理等问题也开始呈现。在欧美已经开始酝酿或着手监管 AIGC 产品之时，《征求意见稿》的发布意味着我国的 AIGC 监管框架将在短期内形成并落地执行。

《征求意见稿》短短 21 条，但已经有针对性地覆盖 AIGC 产品和服务各个环节的监管原则甚至具体要求，预期正式文件出台后，国内日趋白热化的生成式 AI 产品和服务上线周期将被拉长，有些产品甚至可能因无法满足监管要求而被搁置。

《征求意见稿》出台后迅速引发广泛关注和热议，本文仅抓取《征求意见稿》部分重点内容提出我们的行业观察和评论，供各方参考。

一、 办法适用范围

《征求意见稿》第二条规定，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。

✧ **笔者简评：根据《征求意见稿》，面向中华人民共和国境内公众提供生成式AI服务的，均应适用本办法。那么，境外产品提供境内用户注册入口的，是否也应适用本办法？根据我国现行法律监管体系，除非法律法规明确规定并经有关主管部门批准，在我国境内提供信息服务的，应当在境内成立法律实体，以境内实体经营业务。因此，若境外产品提供方主动提供境内用户注册入口，并形成面向我国境内公众提供服务的效果，监管机关可要求境外生成式AI产品提供方在中国境内成立实体，以境内实体提供服务，并适用本办法。**

二、 服务提供者

《征求意见稿》第五条将生成式AI服务“提供者”划定为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，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、声音等的组织和个人。提供者应受制于《征求意见稿》中关于服务提供者的监管要求，承担产品生成内容生产者的责任和个人信息保护义务。

✧ **笔者简评：目前看，生成式AI产品有三大类，一是基础大语言模型，二是大模型在垂直行业的精调，即垂直大模型，三是利用大模型的API打造的应用。按照《征求意见稿》第五条，似乎以上三类产品的参与方均应被认定为“提供者”而受制于《征求意见稿》中关于服务提供者的监管要求。然而，根据《征求意见稿》第二条，该办法的适用范围应是面向我国境内公众提供服务的服务提供方。因此，若基础大语言模型提供方仅**

向垂直大模型提供方提供其产品、而非面向公众提供服务，似乎不应适用《征求意见稿》。我们理解随着正式文件的出台和落地，这点疑问将得到澄清。

三、 服务准入条件

《征求意见稿》第六条规定，利用生成式人工智能产品向公众提供服务前应满足两个条件：一是按照《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；二是按照《互联网信息服务算法推荐管理规定》履行算法备案手续。

❶ *互联网信息服务安全评估*

根据《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》，服务提供者可以自行或委托第三方实施安全评估，并将安全评估报告提交所在地地市级以上网信部门和公安机关。

安全评估的重点内容包括如下：

(1) 确定与所提供服务相适应的安全管理负责人、信息审核人员或者建立安全管理机构的情况；

(2) 用户真实身份核验以及注册信息留存措施；

(3) 对用户的账号、操作时间、操作类型、网络源地址和目标地址、网络源端口、客户端硬件特征等日志信息，以及用户发布信息记录的留存措施；

(4) 对用户账号和通讯群组名称、昵称、简介、备注、标识，信息发布、转发、评论和通讯群组等服务功能中违法有害信息的防范处置和有关记录保存措施；

(5) 个人信息保护以及防范违法有害信息传播扩散、社会动员功能失控风险的技术措施；

(6) 建立投诉、举报制度，公布投诉、举报方式等信息，及时受理并处理有关投诉和举报的情况；

(7) 建立为网信部门依法履行互联网信息服务监督管理职责提供技术、数据支持和协助的工作机制的情况；

(8) 建立为公安机关、国家安全机关依法维护国家安全和查处违法犯罪提供技术、数据支持和协助的工作机制的情况。

❶ *算法备案*

根据《互联网信息服务算法推荐管理规定》，算法备案的内容包括服务提供者的名称、服务形式、应用领域、算法类型、算法自评估报告、拟公示内容等信息。而《互联网信息服务算法推荐管理规定》中规定的算法备案时限为提供服务之日起十个工作日内，《征求意见稿》将生成式人工智能产品的算法备案时限提前到产品上线前。

✧ **笔者简评：网信办将生成式AI定性为“具有舆论属性或者社会动员能力的互联网信息服务提供者”符合监管预期，但较比现行规定，《征求意见稿》更为严格地将安全评估和算法备案设定为生成式AI产品上线的前置条件。《征求意见稿》此举将拉长国内生成式AI的上线周期，给市场争先恐后推出自家生成式AI产品的局面踩上一脚刹车。**

四、 训练数据要求

《征求意见稿》第七条规定，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：

(1) 符合《网络安全法》等法律法规的要求；

(2) 不含有侵犯知识产权的内容；

(3) 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；

(4) 能够保证数据的真实性、准确性、客观性、多样性；

(5) 国家网信部门关于生成式人工智能服务的其他监管要求。

✧ **笔者简评：生成式AI研发和服务过程中，需要使用大量数据进行预训练和优化训练。按照《征求意见稿》第七条要求，生成式AI服务提供者在使用这些数据过程中要确保数据的来源和使用合法合规并不侵犯他人知识产权。鉴于我国公开途径可以获取的可用数据源有限，目前我国的大模型训练过程可能会同时采用国外的公开途径可以获取的数据源。预计错综复杂的数据源获取途径将导致提供者满足《征求意见稿》第七条要求有实际难度；相应地，为了满足数据源的合规要求所需的时间、人力甚至金钱成本也会增加。**

五、 人工标注要求

《征求意见稿》第八条要求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。

✧ **笔者简评：众所周知，生成式AI产品研制过程中需要对大量训练数据集进行标注，标记训练对象的特征，以作为机器学习的基础素材。高质量的数据标注是模型训练的关键。虽然目前市场上已有智能化数据标注的产品，但为实现对人类指令的精准理解，像ChatGPT模型的训练过程也是运用了海量的人工标注。目前，国内大模型训练过程所需的人工标注采用自行雇佣或人力外包或服务外包的方式，无论采用哪种方式，相信《征求意见稿》提出的标注规则及对标注人员进行必要培训的要求，应该符合目前国内的人工标注实践要求。但是，在服务外包的场景下，提供者需注意其自身方是人工标注的相关法律要求承担者，因此应在服务外包过程中要求人工标注服务公司协助其满足法律要求。**

六、 防止生成虚假信息

《征求意见稿》第四条第（四）款要求，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。

《征求意见稿》第十五条进一步规定，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。

✧ **笔者简评：在AI语言模型训练中，最大的问题之一就是如何阻止模型胡编乱造。但是，测试使用国内大模型产品成语绘图功能的用户会发现其生成图片多文不对题。即使使用表现力更强的GPT-4,亦会不时出现所获内容和正确答案南辕北辙，AI对你一本正经“胡说八道”的现象。那么，这些AI “编造”的生成内容是否属于虚假信息？如是，按照《征求意见稿》第十五条，服务提供者应在3个月内通过模型优化训练等方式防止再次生成。但是，对于实则属于大语言模型研发过程中不成熟产品表现的“虚假信息”，三个月的优化训练时间是否足够？如果不够，网信办是否会要求服务提供商停止提供服务并下架产品，而使得产品失去通过实际使用而继续优化训练的机会？**

七、 国际合作

《征求意见稿》开篇第三条即提出，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。“国际合作”作为国家支持的生成式 AI 发展方向之一被特别提及。

✧ **笔者简评：除了开篇提及国家支持国际合作，整篇《征求意见稿》并未再提及国际合作的可行模式或者探索方向。**

2022年9月前后，美国商务部向英伟达和AMD发出通知，限制英伟达的A100、H100和AMD的MI 250系列及未来的高端GPU产品向中国出口。而该类高端GPU芯片对中国的出口限制将极大拖慢我国生成式AI大模型训练的速度。因此，为尽快将生成式AI转化成我国各行业垂直领域生产力，避免我国在AI时代掉队，

我们确有必要探索引入GPT-4或以上的生成式AI产品的国际合作之路。

考虑《征求意见稿》的整体监管要求，尤其是准入条件和对服务提供者的监管，国际合作似乎应有一家境内公司与境外生成式AI产品提供方合作，由境内公司作为提供服务的接入口并承担服务提供者的角色，方能满足《征求意见稿》的监管规定。

值得一提的是，《征求意见稿》第十七条规定，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。此信息报送义务并未就国际合作和境内生成式AI产品进行区分。可以想象，此类信息报送义务将对国际合作产生一定的障碍。因此，国际合作的场景下，如境内提供者能够满足《征求意见稿》其他监管要求，此条信息报送义务是否能够就国际合作的生成式AI有一定豁免？

小结

《征求意见稿》的发布一方面体现了监管部门支持生成式AI发展的积极态度，另一方面也为其健康发展划定了底线，避免生成式AI野蛮生长。生成式AI产品和服务提供者应当承担产品生成内容生产者责任，落实网络安全责任，对生成内容从意识形态、隐私保护、知识产权保护等多个角度把好关，在监管框架内推动人工智能技术的健康合规发展。

虽然《征求意见稿》已经覆盖生成式AI大多重要环节，其中部分原则性规定或有待更多细节规定出台。我们将持续跟进网信办未来正式发布的《生成式人工智能服务管理办法》及人工智能方面的其他监管规定，并适时分享我们的评论意见。

陈　伟　合伙人　电话：86 10 8553 7988　邮箱地址：chenwei@junhe.com
颜炳琳　律　师　电话：86 10 8553 7940　邮箱地址：yanbl@junhe.com