汉坤法律评述
2023 年 8 月 22 日

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

生成式人工智能服务算法备案和安全评估要求初步分析

作者：解石坡 | 赵攀

2023 年 8 月 15 日，《生成式人工智能服务管理暂行办法》（下称“**《人工智能暂行办法》**”）正式生效。《人工智能暂行办法》由国家互联网信息办公室（“**网信办**”）、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（“**工信部**”）、公安部、国家广播电视总局等七部委于 2023 年 7 月 10 日联合发布。在此之前，4 月 10 日，网信办发布《生成式人工智能服务管理办法（征求意见稿）》并公开征求意见。从征求意见到正式出台仅历时三个月。

2023 年 7 月 31 日，某头部应用商店运营商宣布下架近 100 款提供 ChatGPT 相关服务的应用程序。对此，该公司提供的说明表示“正如您所知，政府一直在加强与深度合成技术（DST）和生成式人工智能服务（包括 ChatGPT）相关的监管措施。DST 必须满足在中国运营的许可要求，包括从工业和信息化部获得许可证。根据我们的审核，您的应用与 ChatGPT 相关，但没有获得在中国运营所需的许可证。如果您想在中国提供您的应用，请咨询专业人士，以确保符合《互联网信息服务深度合成管理规定》的规定。”

根据《人工智能暂行办法》第 17 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》（下称“**《算法推荐管理规定》**”）履行算法备案和变更、注销备案手续。因此，AIGC 的算法备案与安全评估，是对相关产品和服务的核心监管要求，以下将基于相关规定和有限实践经验，对算法备案和安全评估制度进行详细解析。

一、算法备案和安全评估的适用范围

根据《人工智能暂行办法》，对于 AIGC 产品进行算法备案与安全评估，有两个关键的适用要素，即（1）向中华人民共和国境内公众提供；和（2）产品具有舆论属性或者社会动员能力。

（一）向中华人民共和国境内公众提供

《人工智能暂行办法》第 2 条同时从正面和反面规定了办法的适用范围，即“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……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。”因此，简而言之，向境内公众提供 AIGC 服务的，适用办法的规定，包括履行算法备案和安全评估要求；未向境内公众提供 AIGC 服务的，则不适用办法的规定。对此，根据《人工智能暂行办法》第 20 条，境外 AIGC 产品如用于向中国境内公众提供 AIGC 服务，同样适用《人工智能暂行办法》。

《人工智能暂行办法》并未明确“境内公众”的范围，从而引发了较多的关于适用范围的疑问。

- B2C 和 B2B 业务模式：B2C 的业务模式下，AIGC 产品将向最终用户和消费者提供，因而较为明确地落入向“境内公众”提供服务的范围。而对于 B2B 的业务模式，则并非当然排除，而是需要根据结合 AIGC 产品的使用方式和场景进行判断。例如，如果 AIGC 产品虽然是向企业客户提供，但企业客户将其用于向不特定公众提供服务（如客服类产品），则其被认定为向境内公众提供服务的可能性仍然较高。

- 内测或定向公测版本：测试阶段的产品或服务并非当然排除办法适用，尤其是当测试用户数量已经达到了一定规模，仍可能被认为已经开始向公众提供服务。

- “封装”、“嵌套”AIGC 产品：企业利用第三方 AIGC 产品向境内客户提供服务，也构成向境内公众提供 AIGC 服务。例如企业通过调用 API 接口提供 AIGC 服务[1]，或对已有的 AIGC 产品进行“封装”、“嵌套”后提供 AIGC 服务，都可能被认定为 AIGC 服务提供者。

（二）具有舆论属性或者社会动员能力

具有舆论属性或者社会动员能力是判断相关主体是否需要进行算法备案和安全评估的另一个关键性的适用要素。2021 年 12 月由网信办、工信部、公安部和国家市场监督管理总局联合发布的《算法推荐管理规定》和 2022 年 11 月由网信办、工信部和公安部联合发布的《互联网信息服务深度合成管理规定》（“**《深度合成管理规定》**”）均提出对具有舆论属性或社会动员能力的算法推荐服务/深度合成技术要求履行算法备案。

《人工智能暂行办法》、《算法推荐管理规定》和《深度合成管理规定》均未对“具有舆论属性或社会动员能力”进行解释。对此，可参照《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》（下称“**《安全评估规定》**”）第 2 条列举的“具有舆论属性或社会动员能力的互联网信息服务”情形，包括：（1）开办论坛、博客、微博客、聊天室、通讯群组、公众账号、短视频、网络直播、信息分享、小程序等信息服务或者附设相应功能；（2）开办提供公众舆论表达渠道或者具有发动社会公众从事特定活动能力的其他互联网信息服务。

根据目前的经验，对于“具有舆论属性或社会动员能力”应当做广义理解。只要相关服务的直接或间接受众较为公开且面对不特定用户，或者具备评论、留言、发布等信息交互功能，就有可能被认定为具有舆论属性或社会动员能力。

二、算法备案

《算法推荐管理规定》和《深度合成管理规定》对具有舆论属性或社会动员能力的算法推荐服务/深度合成服务相关主体提出了算法备案的要求。

- 《算法推荐管理规定》第 24 条规定，具有舆论属性或者社会动员能力的**算法推荐服务提供者**应当在提供服务之日起十个工作日内通过互联网信息服务算法备案系统填报服务提供者的名称、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，履行备案手续。

[1] 参见，国家互联网信息办公室有关负责人就《生成式人工智能服务管理暂行办法》答记者问。其中，在对人工智能服务提供者的释义中提到“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。”访问地址：https://www.gov.cn/zhengce/202307/content_6892001.htm。

- 《深度合成管理规定》第 19 条规定，具有舆论属性或者社会动员能力的**深度合成服务提供者**，应当按照《算法推荐管理规定》履行备案和变更、注销备案手续。**深度合成服务技术支持者**应当参照前款规定履行备案和变更、注销备案手续。

（一）备案主体和备案对象

《算法推荐管理规定》和《深度合成管理规定》均对备案主体和备案对象作出规定。

- 《算法推荐管理规定》下，备案主体为算法推荐服务提供者，备案的对象是算法推荐服务。根据《算法推荐管理规定》第 2 条，具体的算法类别包括生成合成类、个性化推送类、排序精选类、检索过滤类、调度决策类等算法技术。

- 《深度合成管理规定》下，备案主体不仅包括深度合成服务提供者，还包括深度合成服务技术支持者，备案的对象是深度合成服务，即应用深度合成技术提供的互联网信息服务。

在“互联网信息服务算法备案系统”算法类型选项中，深度合成被并入了生成合成类算法，一并显示为“生成合成类（深度合成）”，事实上打消了《算法推荐管理规定》中“生成合成类”与《深度合成管理规定》中“深度合成”是否存在区别的争议。

《深度合成管理规定》明确，深度合成服务提供者，是指提供深度合成服务的组织、个人；深度合成服务技术支持者，是指为深度合成服务提供技术支持的组织、个人。因此，应用层的服务提供商，以及基础层的模型或算法技术支持者，均需要履行备案义务。

（二）算法备案申请的办理流程

企业通过网信办的互联网信息服务算法备案系统（https://beian.cac.gov.cn/#/index）提交算法备案申请。算法备案申请需要填写并提交的信息主要包含主体信息、算法信息和产品及功能（或技术服务）信息三个板块。实践中，需要先完成主体信息备案，才能推进算法信息和产品功能信息的备案。备案中比较重要的文件包括主体信息中的《落实算法安全主体责任基本情况》以及算法信息中需提交的《算法安全自评估报告》，包含企业机构设置、制度建设和安全措施方面较为实质的内容。

自 2022 年 8 月 12 日起截至目前，网信办发布了四批境内互联网信息服务算法备案清单[2]和首批境内深度合成服务算法备案清单[3]。互联网信息服务算法备案清单中涉及生成合成、个性化推送、排序精选、检索过滤、调度决策等算法类别。其中，个性化推送类和检索过滤类是主要的类别。通过深度合成服务算法备案共 41 个深度合成服务算法中，生成结果包含文本、语音内容、图片的算法类型居多。

（三）审查时长

根据《算法推荐管理规定》第 25 条，算法备案的法定时限为三十个工作日。但在实践中往往存在需要企业补正材料的情形，所以完成算法备案通常需要 2－3 个月。备案结果分批次公示，一般 2－3 个月发布一个批次。近来由于累积申请较多，截至目前，最新一批备案清单仍未发布。

[2] 第一批次为 2022 年 8 月，第二批次为 2022 年 10 月，第三批次为 2023 年 1 月，第四批次为 2023 年 4 月，参见：http://www.cac.gov.cn/2022-08/12/c_1661927474338504.htm。

[3] 参见：http://www.cac.gov.cn/2023-06/20/c_1688910683316256.htm。

三、安全评估

《人工智能暂行办法》第 17 条重申了对 AIGC 的安全评估的要求。事实上，安全评估并非《人工智能暂行办法》新提出的监管要求，而是可以追溯到 2016 年和 2017 年不同监管机构的政策制定和执法活动。除了行业普遍关注的网信办安全评估，实践中也存在工信、公安等部门主要负责的安全评估工作，容易引起混淆。

（一）网信办安全评估

网信办的安全评估要求最早见于其 2017 年 10 月 30 日发布的《互联网新闻信息服务新技术新应用安全评估管理规定》，其第 7 条规定："有下列情形之一的，互联网新闻信息服务提供者应当自行组织开展新技术新应用安全评估，编制书面安全评估报告，并对评估结果负责：（一）应用新技术、调整增设具有新闻舆论属性或社会动员能力的应用功能的；（二）新技术、新应用功能在用户规模、功能属性、技术实现方式、基础资源配置等方面的改变导致新闻舆论属性或社会动员能力发生重大变化的。"第 9 条进一步要求在自安全评估完成之日起 10 个工作日内报请国家或者省、自治区、直辖市网信办组织开展安全评估。

2018 年 11 月 15 日，网信办和公安部联合发布了《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》（《安全评估规定》），其第 3 条规定："互联网信息服务提供者具有下列情形之一的，应当依照本规定自行开展安全评估，并对评估结果负责：（一）具有舆论属性或社会动员能力的信息服务上线，或者信息服务增设相关功能的；（二）使用新技术新应用，使信息服务的功能属性、技术实现方式、基础资源配置等发生重大变更，导致舆论属性或者社会动员能力发生重大变化的；（三）用户规模显著增加，导致信息服务的舆论属性或者社会动员能力发生重大变化的；（四）发生违法有害信息传播扩散，表明已有安全措施难以有效防控网络安全风险的；（五）地市级以上网信部门或者公安机关书面通知需要进行安全评估的其他情形。"虽然《安全评估规定》由网信办和公安部联合发布、共同主管，但如下文所述，实践中其主要对接机构是公安机关。

此后，这一要求在《算法推荐管理规定》《深度合成管理规定》[4]和《人工智能暂行办法》中均得到进一步重申，要求具有舆论属性或者社会动员能力的服务提供者，按照国家有关规定开展安全评估。"国家有关规定"的措辞，显示安全评估的范围可能超出《安全评估规定》的要求。

目前，尚未出台针对 AIGC 产品和服务安全评估的进一步细则和指导，但根据目前行业实践来看，安全评估的内容十分实质，包括但不限于企业机构设置、人员配置、制度建设要求，数据安全、运营维护等多项基础安全评估内容，以及针对账号管理、各类信息审核、各类功能的专项安全评估工作。

（二）公安机关安全评估

如上文所述，2018 年网信办和公安部联合发布了《安全评估规定》。实践中，其主要对接部门是公安机关，安全评估报告的提交流程也在公安机关的全国互联网安全管理服务平台（http://www.beian.gov.cn）上完成。

[4] 《深度合成管理规定》第 15 条还规定了需要自行或者委托专业机构开展安全评估的情形："深度合成服务提供者和技术支持者提供具有以下功能的模型、模板等工具的，应当依法自行或者委托专业机构开展安全评估：（一）生成或者编辑人脸、人声等生物识别信息的；（二）生成或者编辑可能涉及国家安全、国家形象、国家利益和社会公共利益的特殊物体、场景等非生物识别信息的。"

（三）工信部门安全评估

2016 年 7 月，工信部发布了《互联网新技术新业务信息安全评估指南》，此后，又陆续出台了《互联网新技术新业务安全评估指南》（替代了 2016 年《互联网新技术新业务信息安全评估指南》）、《互联网新技术新业务安全评估实施要求》以及即时通信、内容分发、信息搜索查询、大数据技术和应用等领域的《互联网新技术新业务安全评估要求》等通信行业标准，确立了工信部对基础电信企业和增值电信企业运营中涉及的互联网新技术新业务进行安全评估的管理框架。

2017 年 6 月 8 日，工信部曾发布《互联网新业务安全评估管理办法（征求意见稿）》，其第 10 条的规定，电信业务经营者应当在以下两种情形进行安全评估，包括：（1）拟将互联网新业务面向社会公众上线的（含合作推广、试点、商用试验）；或（2）电信管理机构书面要求电信业务经营者进行安全评估的。虽然公开信息显示该征求意见稿尚未正式生效，但相关行业标准确立的安全评估体系已经在实际运行，且涉及较为实质的安全评估内容。

因此，实践中，除了履行网信办的安全评估流程外，AIGC 产品和服务的提供商，还可能涉及公安机关和工信部门（通管或信管）的安全评估要求。但其具体的要求、未来是否将发生整合目前尚未完全明确。

综上所述，算法备案和安全评估是在中国境内合法开展面向公众的 AIGC 业务的基础合规要求。其适用范围较为广泛，对于直接或间接向公众提供的产品和服务均可能适用，对于应用层的服务提供和基础层的技术支持也均可能适用。未来，很可能成为某些重点行业客户（例如金融、电信、医疗等强监管行业）对相关产品和服务提供商的准入必然要求。对此，AIGC 产品和服务的提供者应当密切关注监管要求，采取积极措施，履行相关程序，避免产生合规风险，影响业务开展。

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：

解石坡

电话： +86 10 8524 5866
Email： angus.xie@hankunlaw.com