

March 14, 2025

韩国广播通信委员会发布《生成式人工智能服务用户保护指南》

I. 出台背景及核心内容

2025年2月28日，韩国广播通信委员会（下称“**广电委**”）正式发布《生成式人工智能服务用户保护指南》（下称“**《指南》**”），旨在防范生成式人工智能（generative AI，下称“**生成式AI**”）服务可能引发的用户权益风险，推动构建更可信赖的AI环境。《指南》明确规定了AI开发企业及服务提供商应遵循的原则，并明确制定了责任化运营的具体实施细则。

前述“开发企业”指研发生成式AI基础模型并将其投入市场，供服务提供商使用或自行开展服务的主体；“服务提供商”指借助生成式AI系统提供数字工具及相关服务的主体（《指南》第13页）。

《指南》将于2025年3月28日起正式实施，其要求开发企业及服务提供商在全服务周期中遵循4项基本原则，并配套提出落实这些原则的6项具体实施细则。

II. 《指南》核心原则

《指南》针对生成式AI服务全流程提出以下4项基本原则（《指南》第14页）：

- （1）生成式AI服务应保障人类尊严，维护个人自由与权利，确保服务处于人类有效控制与监督之下；
- （2）生成式AI服务应以通俗易懂的方式向用户说明AI系统运行原理、输出结果及其可能对用户造成的影响；
- （3）生成式AI服务应安全运行，最大限度降低意外损害风险，防止被恶意利用或篡改；
- （4）生成式AI服务须确保避免对用户造成歧视性或非公正性后果。

III. 实施细则与具体要求

《指南》为达成4项基本原则，向开发企业及服务提供商提出了6项实施细则，并为每项细则提供了具体操作范例（《指南》第15至26页）。

1. 保护生成式AI服务用户的人格权

开发企业应着力构建算法系统，识别并控制生成式AI基础系统中可能侵犯用户人格权的内容要素，并通过有效手段予以管控；服务提供商则需承担起事前防范的管理责任，防止用户利用生成式AI生成侵犯他人人格权的内容。为此，服务提供商应建立内部监测体系，并设立用户举报程序。此外，服务提供商还应审慎考虑提供服务的范围与方式，确保生成式AI的输出内容不会侵犯用户权利。

具体而言，前述主体应在用户使用服务前，对可生成内容的尺度加以引导，避免出现仇恨性言论等不当内容；当用户输入侵犯他人权利的信息时，系统及时给出警告提示，情节严重者可在一定时间内封禁相关用户账号；同时，应采取专门措施保护特定年龄段用户的权益。

2. 保证用户对生成式AI服务生成过程的知情权

服务提供商有义务明确告知用户产出物是由生成式AI生成，并且，在不损害企业正当利益的前提下，应向用户提供有助于理解生成式AI生成过程的基本信息。

具体而言，应为生成物添加出处标注或提供出处链接；提供生成式AI模型的相关信息；当客户要求对生成式AI的生成过程进行说明时，应在不超出企业正当利益范围的前提下，提供相应的说明方案等。

3. 维护生成式AI服务的多样性

开发企业和服务提供商应在技术可行的范围内，致力于减少生成式AI的偏见与歧视现象，并制定相应的内部原则或标准。此外，服务提供商应建立相关机制，防止用户歧视性地使用生成式AI服务，并筛除相关行为；当用户认为生成物存在偏见时，应提供便于用户直接举报的直观方式，并制定处理举报内容的规范程序。

具体而言，开发企业和服务提供商应在服务正式运营前，在多个场景下开展歧视性内容测试；同时，应将用户举报渠道设置于显著位置，并提供侵权举报的格式文本等切实可行的方案。

4. 对生成式AI服务时收集、利用数据的过程进行合规管理

当服务提供商计划利用用户数据训练AI时，必须事先明确告知用户，并建立同意或拒绝机制。同时，应在企业内部设立监督组织或指定专门负责人，以确保用户数据能够得到安全、合理的使用。

《指南》中提及的具体操作范例包括：制定相关政策，明确生成式AI服务是否涉及利用训练数据，并通过服务条款、常见问题解答等形式告知用户；在服务注册阶段或用户使用阶

段设置相关环节，让用户事先选择是否同意将数据用以训练；组建负责风险管理的内部组织或指定专门负责人，以应对收集及使用训练数据过程中出现的各类问题。

5. 使用生成内容时，明确责任与协同参与

服务提供商应当针对可能出现的问题，清晰界定自身及用户的责任范围，并引导用户明确认知自身责任。此外，为最大程度降低用户难以察觉的损害风险，服务提供商还须构建监测系统及其他风险管理体系。

《指南》中提及的具体操作范例包括：在服务条款或协议中，明确用户因不当使用而引发问题时需承担的责任；采用通俗易懂的方式与术语，向用户阐释其应承担的责任，确保用户充分理解；在服务界面的显著位置设置举报链接，方便用户在问题发生时进行反馈与求助。

6. 致力于生成内容的合规流通与传播

服务提供商应引导用户避免利用生成式AI服务生成或分享不当内容，同时审查、管理用户的输入内容及生成式AI的生成物，确保其符合道德和伦理标准，尤其应着力防范用户向青少年传播有害内容。

在具体操作层面，《指南》提出的可行方案包括：明确告知用户生成式AI生成物可能存在于不准确或错误之处；采取输入输出双重过滤机制，防止生成有害内容；当用户在与生成式AI对话输入对青少年有害的内容时，及时给出警告提示。

IV. 未来展望及实务建议

随着生成式AI技术蓬勃发展，AI的应用领域不断拓展，然而，人格权侵犯、用户数据使用问题、偏见性问题等困境也随之而来。此次《指南》为AI企业在开展负责任的技术研发与服务运营时，明确了必须遵循的重要准则，同时也为强化用户权益保护提供了清晰的政策导向。

广电委计划每两年对该《指南》的合理性进行审查，以确保其能在AI生态系统中有效落地；同时，也将紧密契合相关法律及技术环境的动态变迁，持续对《指南》加以完善。AI开发企业与服务提供商亦需未雨绸缪，时刻把握未来AI监管环境变化，并以此为依据，协同制定兼具法律智慧与技术创新的应对策略。

韩国太平洋律师事务所的AI团队将秉持专业精神，持续追踪韩国国内及全球AI相关监管动态，积极助力企业深入研读、妥善应对《指南》等相关规则，为AI服务的可持续发展注入强劲动力。

姜 泰 旭

律师

T 82.2.3404.0485

E taeuk.kang@bkl.co.kr

池 涌 泉

资深外国律师 (澳洲)

T 82.2.3404.0245

E yongquan.chi@bkl.co.kr

金 玉

资深外国律师 (中国)

T 86.10.6461.3657

E yu.jin@bkl.co.kr