生成式人工智能服务交易合规评估指引

Guidelines for Compliance Assessment of Transactions Involving Generative Artificial Intelligence Services

免责声明

重要提示：本文件仅为一般性信息指南，不构成任何法律或其他专业意见。

《生成式人工智能服务交易合规评估指引》（以下简称“本文件”）由深圳数据交易所牵头，与起草单位、起草人共同编写，旨在为生成式人工智能服务交易相关方提供一套具有可操作性的评估标准。

本文件的内容基于截至2026年2月9日所掌握的公开信息和对相关法律法规的理解。本文件的编写方和撰稿人已努力确保本文件所载信息的准确性、完整性和及时性。然而，数据隐私和合规领域的法律法规复杂多变，且可能存在解释上的差异。因此，本文件可能无法涵盖所有相关细节，也无法及时反映所有最新的法律发展、监管要求或实践变化等。

本文件不构成、亦不应被视为任何形式的法律、会计、投资或其他专业意见。本文件中的信息不能取代针对特定情况的专业税务、会计、法律或其他相关专业意见。在做出任何与数据合规相关的具体决策或采取任何相关行动之前，读者应当咨询具备相关资质的专业顾问，并向其提供与特定情况相关的所有事实。社会主体在具体业务中应以法律法规和国家监管部门的正式规定为准。

本文件的编写方和撰稿人对本文件所含信息的完整性、准确性或及时性不作任何明示或暗示的保证，包括但不限于对本文件的功能、时效性和特定用途适用性的保证。对于本文件中的任何错误和

疏漏，以及任何人因信赖本文件之全部或任一部分内容而采取或未采取的任何措施或因此而产生的任何后果，本文件的编写方和撰稿人不承担任何责任。

本文件的使用受以下条款约束:

1.本文件仅供信息参考，不得用于任何商业目的。

2.在引用或传播本文件的任何内容时，请务必注明出处，并明确其公益合作性质。

3.本文件的编写方保留随时修改、更新或撤回本文件的权利，恕不另行通知。

4.通过访问、阅读或使用本文件，即表示您已阅读、理解并同意本免责声明的所有条款。

编委会

指导单位

深圳市互联网信息办公室

深圳市政务服务和数据管理局

主编

王青兰 深圳数据交易所合规部总经理

副主编

陈一芊 深圳数据交易所数据合规经理

王　艺 北京市环球（深圳）律师事务所合伙人

起草单位

深圳数据交易所有限公司、深圳市福田区政务服务和数据管理局、深圳市北鹏前沿科技法律研究院、公安部第三研究所、深圳市华傲数据技术有限公司、深圳市标准技术研究院、深圳市优必选科技股份有限公司、中兴通讯股份有限公司、深圳 TCL 新技术有限公司、中国联合网络通信有限公司智能城市研究院

起草人

王青兰、陈一芊、王艺、李晓明、贾西贝、沈树雄、胡婧卓、胡敏喆、南红玉、余灏、刘晋名、姚迁、宁宣凤、吴涵、王永霞、杨宇鑫、罗康、许苗峰、钟晓瑜、谭丽、肖声高

目次

一、范围 6
二、规范性引用文件 6
三、术语和定义 7
四、评估原则 9
五、评估框架 10
六、主体合规评估 11
七、标的合规评估 15
八、流通合规评估 32
九、评估方法 36
附录 A 39
附录 B 41
附录 C 44
附录 D 47
附录 E 51

生成式人工智能服务交易合规评估指引

一、范围

本文件规定了生成式人工智能服务交易合规评估的评估原则、主体合规评估要求、标的合规评估要求、流通合规评估要求以及评估方法。

本文件适用于生成式人工智能服务的交易主体、服务使用者、专业第三方服务机构、数据交易机构等生成式人工智能服务交易相关方对生成式人工智能服务交易活动进行合规评估，也适用于主管部门和监督管理部门对生成式人工智能服务交易活动进行监督、管理和评估。

二、规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本适用于本文件。

GB/T 25069—2022　信息安全技术　术语

GB/T 35273—2020　信息安全技术　个人信息安全规范

GB/T 35295—2017　信息技术　大数据　术语

GB/T 41479—2022　信息安全技术　网络数据处理安全要求

GB 45438—2025　网络安全技术　人工智能生成合成内容标识方法

GB/T 45288—2025　人工智能　大模型　第 1 部分：通用要求

GB/T 45654—2025　网络安全技术　生成式人工智能服务安全基本要求

GB/T 45674—2025　网络安全技术　生成式人工智能数据标注安全规范

GB／T 37932—2025　数据安全技术 数据交易服务安全要求

DB4403/T 564—2024　数据交易合规评估规范

三、术语和定义

GB/T 25069—2022、GB/T 35295—2017、GB/T 45288—2025、GB 45438—2025、GB/T 45654—2025、GB/T 37932—2025、DB4403/T 564—2024界定的术语和定义，以及下列术语和定义适用于本文件。

（一）　生成式人工智能服务　generative artificial intelligence service

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，包括以交互界面、可编程接口等方式提供服务。

注：利用生成式人工智能技术包括自行开发的技术或利用第三方模型技术、调用第三方提供的可编程接口等技术服务。

（二）　训练数据　training data

用于训练机器学习模型的输入数据子集，包括预训练、优化训练过程中的输入数据。

（三）　大模型　large-scale model

基于大量数据训练得到，具有复杂计算架构，能处理复杂任务，且具备一定泛化性的深度学习模型。

（四） 生成式人工智能服务交易标的 object of generative artificial intelligence service transaction

即交易标的，指数据卖方或数据商与数据买方之间交易的训练数据、大模型、生成式人工智能服务。

（五） 交易主体 subject of transaction

生成式人工智能服务交易活动中的卖方、买方和数据商。

（六） 服务使用者 service user

使用生成式人工智能服务生成内容的组织、个人。

（七） 个人信息 personal information

以电子或者其他方式记录的与已识别或者可识别的自然人有关的各种信息。

注 1：个人信息包括姓名、出生日期、身份证件号码、个人生物识别信息、住址、通信通讯联系方式、通信记录和内容、账号密码、财产信息、征信信息、行踪轨迹、住宿信息、健康生理信息、交易信息等，不包括匿名化处理后的信息。

注 2：关于个人信息的判定方法、相关术语、子分类，参见 GB／T 35273—2020 附录 A。

注 3：个人信息处理者通过个人信息或其他信息加工处理后形成的信息，例如，用户画像或特征标签，能够单独或者与其他信息结合识别特定自然人身份或者反映自然人活动情况的，属于个人

信息。

（八） 敏感个人信息 sensitive personal information

一旦泄露或者非法使用，容易导致自然人的人格尊严受到侵害或者人身、财产安全受到危害的个人信息。

注：包括生物识别、宗教信仰、特定身份、医疗健康、金融账户、行踪轨迹等信息，以及不满十四周岁未成年人的个人信息。

（九） 重要数据 important data

特定领域、特定群体、特定区域或者达到一定精度和规模，一旦遭到篡改、破坏、泄露或者非法获取、非法利用，可能直接危害国家安全、经济运行、社会稳定、公共健康和安全的数据。

（十） 违法和不良信息 illegal and harmful information

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。

四、评估原则

（一）合法原则

生成式人工智能服务交易合规评估活动应依法合规开展，评估人员应了解并遵守所有适用的法律法规要求。

（二）客观公正原则

生成式人工智能服务交易合规评估活动应真实和准确地反映生成式人工智能服务交易活动现状，确保评估意见仅建立在评估证

据的基础上。

（三）保密原则

评估人员应遵守保密义务，审慎使用并保护在评估过程中获得的信息。

五、评估框架

根据相关法律法规的规定，结合生成式人工智能服务交易过程，首先针对交易主体的相关情况进行评估，再针对交易标的的相关情况进行评估，最后针对交易标的流通相关情况进行评估。每个环节均会从合法、安全、诚信、权益保障四个维度进行评估，评估过程包括评估准备、评估实施、评估报告三个阶段，评估框架见图1。

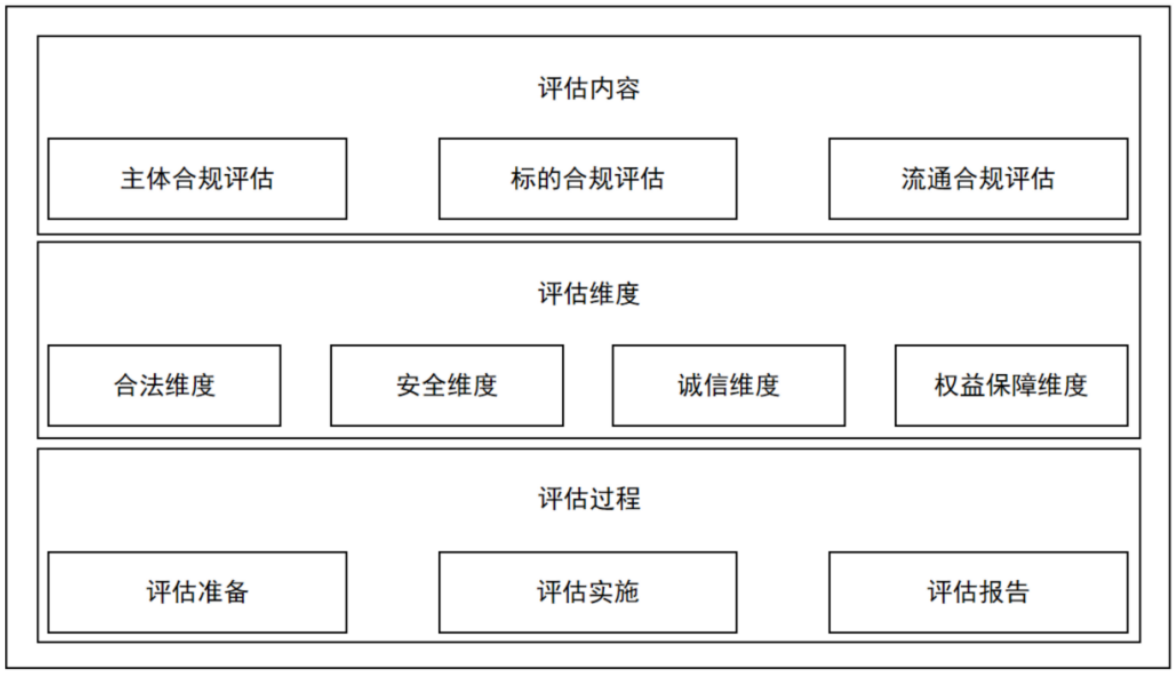


图1　生成式人工智能服务交易合规评估参考框架图

六、主体合规评估

（一）合法维度评估

1. 对交易主体的评估内容包括:

（1）营业执照合法有效且如期完成企业年度报告申报;

（2）交易主体具备开展生成式人工智能服务交易活动所涉及的资质文件、权属证明文件。

注：如互联网信息服务备案、增值电信业务许可、移动互联网应用程序备案，以及相关《计算机软件著作权证书》《电子版权证书》或《软件著作权认证证书》等。

（3）若交易标的涉及新闻出版、影视制作、文艺创作、药品信息、医疗器械、宗教信息、教育等特殊行业活动的，交易主体取得相关许可证或通过备案。

注：如网络文化经营许可证、互联网新闻信息许可证、互联网宗教信息服务许可证、教育移动互联网应用程序备案管理等。

（4）若交易主体涉及外商投资的，交易主体的资质还符合外商投资相关法律、行政法规的规定。

注：生成式人工智能服务可能涉及的互联网新闻信息服务、互联网文化经营（音乐除外）、互联网公众发布信息服务等均属于外商投资负面清单内容。

2.评估交易主体在合法维度所需的相关文件资料，可见附录B。

（二）安全维度评估

1. 对交易主体的评估内容包括:

（1）交易主体已建立生成式人工智能服务安全管理机构，对交易标的所涉及的大模型安全、数据安全、系统安全、使用安全等方面的工作承担总体责任，并明确职责分工。

（2）交易主体属于重要数据处理者的，已定期开展数据安全风险评估。

（3）交易主体属于重要数据处理者或处理个人信息达到国家网信部门规定数量的，已设立数据安全管理部门、个人信息合规管理部门，明确重要数据安全负责人、个人信息保护负责人及其职责。

（4）交易主体已建立算法安全自评估制度，阐述算法自评估的制度设计考虑，结合现状论证制度的合理性、完备性和可落地性，说明为夯实制度执行所采取的保障措施。

注：对于已完成互联网信息服务算法备案或生成式人工智能服务备案的，凭借公示的备案号、上线编号等信息，可以免予自评估制度等制度文件的提供。

（5）交易主体已制定生成式人工智能服务设计、开发、部署、应用全生命周期的风险监测与评估机制，并已根据风险评估的结果，定期更新风险监测方法和流程。

（6）交易主体已建立算法、网络安全和数据安全事件应急处置制度，涵盖即时监测、快速响应、事件隔离、协同处置与事后优化等环节，对安全事件的有效管理和最小化影响，同时定期进行合规性审查和优化，以适应不断变化的合规要求。

（7）交易主体属于个人信息处理者的，已依法开展个人信息

保护合规审计。

（8）交易主体已制定对涉及生成式人工智能服务交易的信息技术外包和数据处理供应商风险评估机制，明确监督管理要求。该机制可防范供应商任意终止关键技术或服务，以及未经授权处理数据等行为。

（9）交易主体已在管理及技术方面具有与当前业务及数据处理活动相适应的数据安全保障能力。

（10）交易主体已使用安全可靠的算力来源，具有足够的算力资源对大模型进行持续迭代和升级。

（11）交易主体已建立科技伦理审查机制，依法对生成式人工智能服务交易活动开展科技伦理风险评估。

（12）交易主体已建立健全针对生成式人工智能服务的辟谣机制，发现有利用该服务制作、复制、发布、传播虚假信息的，已及时采取辟谣措施，保存有关记录，确保记录的完整性，并向有关主管部门报告，确保与监管部门的有效协同。

2. 评估交易主体在安全维度所需的相关文件资料，可见附录B。

（三）诚信维度评估

1. 从诚信维度评估交易主体时，核查交易主体是否已提交真实、准确的材料，确保相关情况的真实可信，并如实披露其近三年网络安全、数据安全、个人信息保护、人工智能、知识产权相关的行政处罚、刑事处罚、被诉和被仲裁案件。

2. 评估交易主体在诚信维度所需的相关文件资料，可见附录B。

（四）权益保障维度评估

1.通则

（1）从权益保障维度评估交易主体时，核查交易主体是否已完善其权益保障机制，同时保障服务使用者、合作方、知识产权权利人、个人信息主体或其他经营者的合法权益。

（2）评估交易主体在权益维度所需的相关文件资料，可见附录B。

2.相关主体权益保障措施评估内容

（1）交易主体已对提供网络、数据、算力、大模型等信息技术支持的合作方进行全面核验。核验的内容包括其网络和数据安全合规能力、个人信息保护情况、知识产权保护情况及相关资质。交易主体已了解合作方的经营范围、资质证照、数据安全技术能力和管理能力、过往遵守相关法律法规的情况，以及合作方所在国家/地区对数据安全的相关规定等。

（2）交易主体已建立完善的第三方权益保护机制，包括建立有效的投诉、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。当收到有关不正当竞争、数据安全和个人信息保护等内容安全合规的投诉举报时，已及时受理并采取停止生成、停止传输、消除等处置措施。

（3）交易主体已遵守与合作方签订的合作协议约定的内容，

包括提供生成式人工智能服务中数据处理的授权范围以及其他约定。

3.服务使用者权益保障措施评估内容

（1）交易主体已建立服务使用者权益保障制度，包括权利告知、权利请求响应和处理流程等。

（2）交易主体已与信息技术合作方在合作协议中就服务使用者数据处理行为约定权益保障义务和责任。

（3）交易主体已制定关于生成式人工智能服务全生命周期的数据处理规则，确保符合网络和数据安全、个人信息保护、人工智能、知识产权保护等方面的法律法规。

七、标的合规评估

（一）合法维度评估

1.评估对象及评估路径

（1）生成式人工智能服务交易可分为训练数据交易、大模型交易及服务交易三类，相应的合规评估也分为三个层次，分别对应训练数据、大模型及服务本身的合规性评估。其中，大模型交易的合规评估除针对大模型自身外，还需涵盖其训练数据的合规性；服务交易的合规评估除针对服务本身外，还需向上追溯，涵盖其依托的大模型及该大模型的训练数据的合规性。

（2）评估交易主体在合法维度所需的相关文件资料，可见附录B。

2.训练数据合规

（1）数据来源

对训练数据来源的评估内容包括:

a）交易主体已确保用于训练生成式人工智能的训练数据来源合法合规，不使用非法的、可能侵犯他人知识产权、人格权或者其他合法权益的训练数据对生成式人工智能进行训练，数据收集满足DB4403／T 564—2024 第 8.1.2 条的相关要求。

b）面向特定训练数据来源采集前和采集后，已分别对每一来源训练数据的可靠性进行安全评估和核验，如某一来源训练数据内容中含违法不良信息超过 5％的，不宜采集或使用该来源的训练数据。

c）已确保训练数据来源可追溯，处理个人信息已具备合法性基础，或者虽然处理个人信息不具备合法性基础，但我国法律法规允许通过将训练数据中的个人信息匿名化处理后进行使用，以保证其生成式人工智能模型或产品没有处理个人信息。数据处理满足DB4403／T 564—2024 第 8.1.3 条的相关要求。

d）已核查、记录训练数据的来源及相关授权文件:

1）若使用开源训练数据时，具有该训练数据来源的开源许可协议或相关授权文件;

2）若使用自采训练数据时，具有采集记录，不采集他人已明确不可采集的数据、不可采集的网页数据，或个人已拒绝授权采集的个人信息等;

3）若使用商业训练数据时，有具备法律效力的交易合同、合作协议等，在来源方不能提供数据来源、质量、安全等方面的承诺以及相关证明材料时，不使用该训练数据，且已对交易方或合作方所提供训练数据、承诺、材料进行审核。

e）发现违法训练数据的，已立即停止使用相应训练数据。

（2）数据内容

对训练数据内容的评估范围包括:

a）交易主体在训练数据用于训练前，已采取如下措施，对训练数据内容进行审查:

1）采取关键词、分类模型、人工抽检等方式，充分过滤全部训练数据中的违法不良信息。

2）对训练数据内容的准确性进行核查，及时更正或删除不准确的内容。

3）对训练数据中的知识产权侵权风险进行识别，采取如下措施防范侵权他人知识产权风险:

①根据国家法律法规制定知识产权相关策略，并结合第三方投诉情况不断更新;

②发现存在知识产权侵权等问题的内容，及时从训练数据中删除。

4）对训练数据是否涉及个人信息进行识别，使用包含个人信息的训练数据时，评估是否已取得对应个人信息主体的同意，或满足其他合法使用该个人信息的条件。

b）训练数据内容满足 DB4403／T 564—2024 第 8.1.4 条的相关要求。

c）训练数据内容与生成式人工智能涉及的服务具有相关性和充分的代表性，且考虑到生成式人工智能服务预期使用的具体场景、功能或服务使用者行为的相关特征或要素。

（3）数据标注

如基于大模型训练，需要对训练数据进行标注的，评估交易主体是否已采取以下措施:

a）根据生成式人工智能服务的服务类型、服务功能和安全风险，按照 GB／T 45674—2025 第 6 条的要求，分别对功能性数据标注与安全性数据标注制定标注规则，标注规则至少包括标注目标、数据格式、标注方法、质量指标四个维度;

b）明确划分数据标注人员角色，在标注人员上岗前，对其开展安全培训与考核，考核内容包括相关法律法规知识、标注规则理解能力、标注平台或工具使用能力、安全风险判定能力、数据安全管理能力等，在标注过程中，对其标注能力进行持续评定;

c）涉及个人信息、重要数据、公共数据、商业秘密数据标注的，已与标注人员签署保密协议，明确其保密义务与责任;

d）对照标注规则，对标注后的内容进行人工、自动化或混合形式的抽检，对标注不合格的训练数据按照标注规则予以重新标注或删除;

e）如涉及使用基于生成式人工智能服务的自动化标注平台或

工具进行辅助标注，已确保自动化标注平台符合生成式人工智能服务相关法律法规要求。

3.大模型合规

（1）设计与研发

对于大模型设计和研发阶段，评估交易主体是否已采取以下措施:

a）在研发阶段已明确生成式人工智能服务的机制机理，确保生成式人工智能服务的应用场景合法正当，并明确提供生成式人工智能服务所需配备的网络、数据、算力、模型、技术、人员等资源配置;

b）已按照法律法规的要求，对服务涉及的科技伦理风险进行审查，若涉及具有舆论社会动员能力和社会意识引导能力的算法模型、应用程序及系统或面向存在安全、人身健康风险等场景的具有高度自主能力的自动化决策系统的研发，已按照法律法规的要求开展伦理审查复核;

c）在预训练环节，选择来源合法的基础模型，交易主体已对基础模型的来源进行审查，确保基础模型的安全性;

d）对涉及重要数据、个人信息、商业秘密、公共数据的训练数据访问和使用已进行认证和授权，并留存日志，日志留存时间至少为 6 个月;

e）在优化训练环节，交易主体使用已标注的训练数据进行优化训练;

f）若涉及接入第三方大模型服务的，第三方大模型具有合法来源，且使用用途、方式未违反与第三方之间的协议、模型使用说明、模型开源许可证的限制性要求，且满足以下要求：

1）若以 API 等方式接入第三方大模型服务的，使用已获得主管部门备案的大模型；

2）若以本地部署形式接入第三方开源大模型或其他第三方大模型的，使用境内主体开发的大模型。

（2）第三方算力服务的使用

若大模型涉及使用第三方算力服务的，评估交易主体是否满足以下要求：

a）原则上，算力基础设施部署于中华人民共和国境内，若因服务所必需，确有必要部署于中华人民共和国境外，流通过程已满足本文件第 8 条第 1 款第 2 项的（3）（4）关于数据出境、出口管制的规定；

b）算力基础设施采用绿色、安全可信的硬件和网络设施；

c）交易主体已与算力服务的提供者签署协议，明确算力服务计费规则，各方在网络安全、数据安全和个人信息保护、知识产权保护、保密责任等方面的权利和责任。

（3）验证与评估

在大模型验证与评估阶段，评估交易主体是否满足如下要求：

a）建立生成内容测试题库、拒答内容测试题库，对输出内容的合法性、准确性和可靠性进行评估，防范本文件附录 C 中列举的

安全风险;

b）在生成式人工智能服务正式上线前，已对大模型的机制机理、生成内容的安全风险、效果、质量等进行评估测试，确保符合法律法规、设计规范和预期目标。

（4）监测与更新

在大模型上线运行阶段，评估交易主体是否满足以下要求:

a）制定大模型更新、升级的安全管理策略，形成管理机制，对大模型进行定期更新、升级和维护;

b）定期对所使用的开发框架、开源代码等进行安全审计，关注开源框架安全及漏洞相关问题，识别和修复安全漏洞，并保留审计底稿;

c）定期通过抽检的手段，宜采取以下措施，对生成内容的安全情况、自动化审查的有效性进行评估，并对模型进行优化:

1）采取人工抽检、技术抽检等手段，按照 GB／T 45654—2025 中 B2.2.2.1 的测评方法对生成内容进行测评，内容抽样合格率（不含违法和不良信息即视为合格）不低于 90%;

2）根据抽检情况采取模型优化措施，降低生成违法或不良信息、侵权信息、错误信息的风险。

4.生成式人工智能服务合规

对生成式人工智能服务的评估内容包括:

（1）利用生成式人工智能服务提供生成文本、图片、音频、视频等内容时，交易主体已采取以下措施:

a）明确生成式人工智能服务应用的目的、用途以及部署应用生成式人工智能服务的预期环境，以满足不同应用场景下的合规要求；

b）部署事实核查技术，以验证人工智能生成内容的准确性和真实性。

（2）利用生成式人工智能服务提供生成文本、图片、音频、视频等内容，可能导致公众混淆或者误认的，交易主体已按照法律法规及 GB 45438—2025 的格式、内容要求，对生成合成内容添加以下显式标识:

a）在文本的起始、末尾或者中间适当位置添加文字提示或者通用符号提示等标识，或者在交互场景界面、文字周边添加显著的提示标识；

b）在音频的起始、末尾或者中间适当位置添加语音提示或者音频节奏提示等标识，或者在交互场景界面中添加显著的提示标识；

c）在图片的适当位置添加显著的提示标识；

d）在视频起始画面和视频播放周边的适当位置添加显著的提示标识，在视频末尾和中间适当位置添加显著的提示标识；

e）呈现虚拟场景时，在起始画面的适当位置添加显著的提示标识，在虚拟场景持续服务过程中的适当位置添加显著的提示标识；

f）提供生成合成内容下载、复制、导出等功能时，文件中含

有满足以上要求的显式标识。

（3）利用生成式人工智能服务提供生成文本、图片、音频、视频等内容，交易主体已按照法律法规及 GB 45438—2025 的格式、内容要求，在生成合成内容、文件元数据中添加隐式标识，并保存日志信息至少 6 个月。

（4）交易主体已基于移动电话号码、身份证件号码、统一社会信用代码或者国家网络身份认证公共服务等方式，依法对服务使用者进行真实身份信息认证。

（5）生成式人工智能服务按照法律法规要求不适用于未成年人的，交易主体已通过用户协议、产品界面等予以说明，涉及向未成年人提供生成式人工智能服务的，交易主体采取下述措施：

a）合理限制不同年龄阶段的未成年人在使用服务中的消费行为，对于超出未成年人行为能力认知范畴的充值消费，宜经由其监护人同意、确认，防止未成年人未经监护人许可的过度消费；

b）合理限制不同年龄阶段未成年人的使用时段、频率、时长等，防止未成年人沉迷、过度依赖；

c）向监护人提供未成年人充值购买记录、使用记录的查阅渠道，允许监护人对未成年人使用服务的时段、时长、频率、生成内容领域等进行合理限制；

d）采取更为严格的内容安全措施，展示有益于未成年人生活学习、身心健康的内容，促进未成年人开阔眼界、陶冶情操、提高素质，防范未成年人过度依赖服务，影响未成年人独立思考、培育

良好习惯。

（6）提供服务过程中，交易主体具备识别和过滤违法不良信息输入和输出能力，对服务使用者的输入数据和生成内容，自行或委托外部服务方，采用自动化方式（关键词、分类模型等方式）进行合规审核，避免传播违法和不良信息、侵权信息、错误信息，防范本文件附录C中列举的安全风险发生。

（7）交易主体发现违法或不良信息时，已采取停止生成、停止传输、消除等处置措施。

（8）提供服务过程中，已对服务使用者的个人信息、输入信息、使用记录予以保护，满足以下要求：

a）个人信息的收集满足最小必要原则，仅收集与提供服务目的直接相关的个人信息；

b）不得非法留存能够识别服务使用者身份的输入信息和使用记录，若基于网络日志留存、服务使用者同意等合法方式留存的，存储期限限于为实现处理目的所必要的最短时间；

c）未经服务使用者同意或具备其他合法性基础，不得非法向他人提供服务使用者的输入信息和使用记录；

d）利用服务使用者的个人信息、输入信息、使用记录进行模型训练、优化的，已明确告知服务使用者且取得其同意，并给予其选择退出的渠道和方式。

（二）安全维度评估

1. 通则

（1）从安全维度评估交易标的时，评估交易主体是否已根据交易标的所涉及的数据类型和重要程度采取不同级别的安全管理和技术措施，保障交易标的安全。

（2）评估标的安全维度所需的相关文件资料，可见附录B。

2. 数据安全

对交易标的数据安全的评估内容包括:

（1）交易主体已采取保障数据存储安全的措施:

a）采取必要的存储安全措施，防止数据泄露、篡改和丢失;

b）使用加密技术保护存储的数据，确保仅授权人员可访问和操作数据;

c）数据存储设施定期进行安全评估和风险监控;

d）定期对数据进行备份，防止数据损坏和丢失。

（2）交易主体已采取保障数据传输安全的要求:

a）在数据传输过程中，使用加密通道（如SSL／TLS）保障数据的完整性和机密性;

b）在数据传输过程中进行完整性校验，防止数据在传输过程中被篡改;

c）建立数据传输管理制度，明确数据传输的权限和流程，防止未经授权的数据传输行为。

（3）交易主体已采取保障数据访问安全的要求：

a）建立数据访问控制机制，确保仅经过授权的人员可访问和处理数据；

b）访问控制机制包括身份验证、多因素认证等手段，确保数据访问的安全性；

c）定期审查和更新访问权限，防止因人员变动或职责变更导致的数据泄露风险。

（4）交易主体已落实数据分类分级要求：

a）已按照 GB／T 43697—2024 和行业主管部门的要求，依据数据的来源、敏感程度以及数据泄露是否可能危害国家安全、社会公共利益等进行数据分类分级；

b）已按数据的不同类别和级别采取不同的保护措施，及时调整保护措施。

（5）交易主体已采取以下数据审计与监控措施：

a）建立数据使用和访问的审计制度，记录所有数据操作行为，确保数据处理过程及应用生成数据可追溯、可管控；

b）通过日志管理和行为监控，及时发现和应对异常数据操作，保障数据安全；

c）定期进行安全审计，评估数据安全管理体系的有效性，并根据审计结果进行改进。

（6）交易主体已采取以下数据销毁的管理和技术措施：

a）制定数据销毁安全管理制度，以确保不再需要的数据会被

及时销毁，以免造成数据泄露风险;

b）配置必要的数据销毁技术手段，如硬销毁和软销毁方法，确保以不可逆方式销毁数据。

3. 大模型安全

对交易标的使用的大模型安全的评估内容包括:

（1）交易主体已建立完善的模型鲁棒性和安全性评估机制，确保模型能够抵御对抗样本攻击、数据投毒攻击等，防止模型被恶意利用。

（2）交易主体已建立模型漏洞监察管理机制，定期进行安全审计、漏洞扫描和对抗性测试，及时发现和修复模型漏洞，并制定模型更新和维护策略，采取必要的安全措施，防止模型过时或带有安全漏洞的模型被恶意利用。

（3）交易主体已建立异议举报及受理机制，当服务使用者或举报者对生成内容合法合规性有异议，交易主体已按以下要求进行数据异议处理:

a）及时向服务使用者或举报者反馈，告知其已进入生成内容异议审核阶段;

b）及时判断被异议的生成内容是否违反法律法规、违反社会公德或伦理道德;

c）对已确认被异议的生成内容违反法律法规、违反社会公德或伦理道德的，已及时采取停止生成、停止传输、消除等处置措施，并采取模型优化训练等措施进行整改;

d）向服务使用者或举报者告知生成内容的异议处理情况，并视具体情况向有关主管部门报告。

（4）交易主体确保微调不会损害原有安全和保障机制。

4. 系统安全

对交易标的系统安全的评估内容包括:

（1）交易主体已采取以下系统漏洞管理措施:

a）建立系统漏洞管理机制，定期扫描和评估系统中的安全漏洞;

b）建立漏洞响应预案，快速响应和处理系统安全事件，降低安全风险;

c）对发现的安全漏洞及时修复。必要时可进行系统更新和补丁安装，防止被恶意利用。

（2）交易主体已满足以下安全开发管理要求:

a）生成式人工智能服务系统开发时遵循安全编码规范，防止因编码缺陷导致的安全问题;

b）开发阶段已进行安全评审和测试，发现并修复潜在的安全漏洞，确保系统上线前的安全性;

c）对于使用的第三方组件和开源软件，已按照 GB／T 43848—2024 的要求进行安全审查和测试，确保其不包含已知的安全漏洞;

d）已制定和实施详细的数据冗余管理机制，确保系统的可靠性和安全性，在系统出现故障时快速恢复，保证系统的持续运行。

（3）交易主体已采取以下入侵检测与防御措施:

a）生成式人工智能服务的数据存储、模型训练和服务部署在安全可信的网络环境中进行，已采用相关网络安全措施提供安全能力，包括防火墙、入侵防范、网络隔离等;

b）生成式人工智能服务设备已部署入侵检测系统（IDS）和入侵防御系统（IPS），实时监控系统的运行状态，及时发现和阻止恶意攻击;

c）通过日志分析和行为监控，识别潜在的安全威胁，采取相应的防御措施;

d）定期进行安全演练，测试和改进入侵检测与防御系统的有效性。

（4）交易主体已建立明确反映安全措施有效性的指标，例如未经授权的访问尝试次数、推理、绕过、提取、渗透或来源验证等。

5. 使用安全

对交易标的使用安全的评估内容包括:

（1）交易主体已建立服务使用者信息及业务数据的安全保护机制，对服务使用者的个人信息、输入信息和使用记录已依法履行如下保护义务:

a）根据必要性原则，仅收集与提供服务目的直接相关的个人信息;

b）不得非法留存能够识别服务使用者身份的输入信息和使用记录;

c）不得非法向他人提供服务使用者的输入信息和使用记录，除非获得服务使用者同意，或具有其他合法性基础；

d）未进行明确告知并取得服务使用者同意的，不得擅自将服务使用者的输入信息用于后续模型训练，除非具备其他合法性基础。

（2）交易主体已采取以下内容侵权维权保护措施：

a）已针对服务使用者不当使用生成式人工智能服务生成内容造成他人权益损害的场景，建立被侵权人维权支持机制；

b）在确认侵权事实属性的前提下，就被侵权人在法律框架内维护其合法权益提供合理配合，并采取必要措施防止侵害结果的扩大。

（三）诚信维度评估

1. 从诚信维度评估交易标的时，评估内容包括：

（1）若在数据交易机构内开展交易，交易主体提交的交易标的相关登记材料真实、准确、完整，已依据实际情况披露交易标的的描述说明、适用范围、计费方式等信息，并向数据交易机构提供服务样例。

（2）交易主体在交易标的推广和营销过程中未使用夸大、虚假或误导性的宣传语，确保所有宣传材料准确反映交易标的的实际功能和性能。

（3）交易主体已提供交易标的相关说明文档，包括但不限于训练数据、大模型的有关情况，以及生成式人工智能服务的工作原

理、逻辑概述等，材料内容完整、规范，并且与交易标的实际情况相一致。

（4）若在数据交易机构内交易，交易主体已向数据交易机构提交与交易相关的书面承诺，遵守相关法律法规、行业规范及本文件的各项要求。

（5）若在数据交易机构内交易，交易标的不存在违反与数据交易相关书面承诺或数据交易机构业务规则的情况。

（6）交易标的近三年不存在不符合网络安全、数据安全、个人信息保护、人工智能、知识产权等方面法律法规要求而被刑事处罚、行政处罚且未整改完毕的情形。

2. 评估交易主体在诚信维度所需的相关文件资料，可见附录 B。

（四）权益保障维度评估

1. 从权益保障维度评估交易标的时，评估内容包括:

（1）交易主体已制定产品协议或说明文档、页面，公开其服务的适用人群、场合、用途，明确各方主体的权利和义务，增强服务的可解释性和透明度，指导服务使用者科学理性认识和依法使用生成式人工智能技术，包括如下内容:

a）明确生成式人工智能服务的适用人群、场合、用途，是否可以用于科研、商用、特定行业使用等;

b）明确告知服务使用者使用生成式人工智能服务及其生成内容的规范、违反法律法规及协议约定的后果;

c）明确说明生成合成内容标识的方法、样式等规范内容，并提示用户仔细阅读并理解相关的标识管理要求；

d）明确生成式人工智能服务生成内容的知识产权归属；

e）对于涉及金融、医疗、法律、自动驾驶、生产安全方面的特定行业应用，向服务使用者着重提示使用风险。

（2）交易主体已在服务过程中，按照法律法规及其服务协议的约定，提供安全、稳定、持续的服务，保证生成式人工智能服务的应答率和正常使用。

2. 评估交易主体在权益维度所需的相关文件资料，可见附录B。

八、流通合规评估

（一）合法维度评估

1. 评估内容包括：

（1）交易主体已遵守法律、行政法规，尊重社会公德和伦理道德，在提供和使用生成式人工智能服务过程中未存在以下情形：

a）利用生成式人工智能服务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；

b）利用生成式人工智能服务生成民族、信仰、国别、地域、性别、年龄、职业、健康等歧视性内容；

c）恶意删除、篡改、伪造、隐匿生成合成内容的显式、隐式标识;

d）利用生成式人工智能服务实施市场垄断、不正当竞争、侵害知识产权或损害他人合法权益的行为。

（2）交易标的涉及具有舆论属性或者社会动员能力的生成式人工智能服务，交易主体在交易前已按照有关法律法规履行安全评估和备案或登记手续。

注：安全评估如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，备案或登记手续如互联网信息服务算法备案、生成式人工智能服务备案或登记。

（3）交易标的涉及数据出境的，交易主体已按照有关法律法规履行数据出境的相关要求。

注：例如因模型训练、内容生成等目的，将在中国境内采集的数据传输至境外，且数据中含有重要数据或达到网信部门规定数量的个人信息时，已按照有关法律法规履行数据出境的相关要求。

（4）交易标的涉及大模型源代码数据、目标代码、可执行代码等数据出境的，交易主体已遵守法律法规对出口管制的相关要求。

2. 评估交易主体在合法维度所需的相关文件资料，可见附录B。

（二）安全维度评估

1. 评估内容包括:

（1）交易主体已依据数据的来源、敏感程度以及数据泄露是否可能危害国家安全、社会公共利益等进行数据分类分级，并按照数据不同的类别和级别采取不同的保护措施。

（2）数据分类分级工作已贯穿于流通过程的各个环节，并在各个环节使用统一的识别规则并采取相应的技术措施。

（3）按照网络安全等级保护的要求，交易主体处理重要数据或处理一百万人以上个人信息的系统原则上已满足三级以上网络安全等级保护要求。

（4）交易主体全面加强数据处理系统、数据传输网络、数据存储环境等的安全保护，制定并落实数据安全管理制度。

2. 评估交易主体在安全维度所需的相关文件资料，可见附录B。

（三）诚信维度评估

1. 评估内容包括:

（1）交易主体已签署交易合同，交易合同约定内容公平合理，明确双方权利义务，且与本次交易的背景、目的等情况内容一致。

（2）交易主体已承诺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，已承诺依法承担个人信息处理者责任，履行个人信息保护义务。

2. 评估交易主体在诚信维度所需的相关文件资料，可见附录B。

（四）权益保障维度评估

1. 通则

（1）评估交易主体是否已公开接受关于知识产权侵权、不正当竞争等第三方权益、个人信息权益、违法和不良信息相关的投诉举报入口，公开行使权利的途径和方式、响应时限，并在法定时限内处理。

（2）评估交易主体在权益维度所需的相关文件资料，可见附录B。

2. 第三方权益保护

在第三方权益保护方面，评估交易主体是否已采取以下措施，保障第三方权利人的知识产权、竞争性利益或其他合法权益：

（1）如权利人提供的材料可初步证明侵权事实的，已及时对有关内容进行屏蔽，采取内容阻断、防止生成、优化训练等必要措施防止侵害结果的扩大；

（2）就权利人在法律框架内维护其合法权益提供合理配合，根据行政、司法等国家机关的要求配合提供证据。

3. 个人信息保护

在个人信息保护方面，评估交易主体是否已采取以下措施，保障个人信息主体权益：

（1）参考GB／T 41479—2022第5.11的要求，向个人信息主体告知权益范围、申诉投诉及举报途径和响应时限承诺；

（2）在公布时限内，响应个人关于查阅、复制、更正、补充、

删除其个人信息、撤回同意等个人信息权利请求。

4. 违法和不良信息防范

在违法和不良信息防范方面，评估交易主体在主动发现或经投诉举报发现服务使用者利用生成式人工智能服务制作、复制、发布、传播虚假信息，或利用生成式人工智能服务从事违法活动时，是否已依法依约采取辟谣、警示、限制功能、暂停或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。

九、评估方法

（一）评估准备

1. 评估团队组建

交易主体或其委托的数据专业第三方服务机构按照不同的生成式人工智能服务交易需求，组建包含评估组长、管理评估员、技术评估员、质量控制员等角色的评估团队。

2. 评估方案

（1）评估团队按照不同的生成式人工智能服务交易需求，制定评估方案和评估计划。

（2）评估团队按照本文件第六、七、八条的要求及参照本文件附录 B建议文档进行评估资料收集，并做好资料分类及内容标记。

（3）交易主体对拟进行生成式人工智能服务交易的申请资料进行确认，保证资料的完整性、正确性、一致性等符合相关法律要求。

（二）评估实施

1. 评估实施的工作流程如下:

（1）主体合规评估：评估团队依据本文件第六条的要求对主体合规进行评估，并记录评估结果和评估问题;

（2）标的合规评估：评估团队依据本文件第七条的要求对标的合规进行评估，并记录评估结果和评估问题;

（3）流通合规评估：评估团队依据本文件第八条的要求对流通合规进行评估，并记录评估结果和评估问题;

（4）评估问题反馈：评估团队对评估过程中发现的问题进行汇总分析并反馈给交易主体;

（5）问题整改确认：交易主体对问题进行整改后，评估团队对问题整改情况进行确认并将确认结果反馈给交易主体。

2. 安全合规评估可委托第三方技术评估机构进行评估，出具评估意见，供交易主体或其委托的第三方服务机构援引。

（三）评估报告

1. 评估报告编制

评估团队整理评估过程记录、评估底稿文件及评估结果，基于分析结果，撰写评估报告，报告中包含评估过程、主要发现、结论、改进建议等内容。

2. 结果沟通与反馈

评估团队签署评估报告并发送至交易主体，收集反馈意见。

3. 跟踪与改进

交易主体根据评估结果和反馈，制定改进措施或调整策略，设定跟踪机制以监督改进措施的实施情况和效果。

4. 存档与复审

评估团队对本次评估活动的全部材料进行归档封存。

附　录　A

生成式人工智能服务交易合规评估模板

A.1　评估概述

A.1.1　评估目的和依据

（说明本次评估的目的和依据）

A.1.2　评估对象和范围

（描述本次评估涉及的对象和范围）

A.1.3　评估结论概要

（生成式人工智能服务交易的概要情况，评估结果及处置建议概要）

A.2　交易主体基本情况

A.2.1　基本信息

（主体名称、统一社会信用代码等工商登记信息）

A.2.2　业务情况

（主营业务、发展状况等）

A.2.3　资质情况

（说明交易主体是否进行备案或取得行政许可、审批等）

A.2.4　相关领域信用情况

（交易主体在相关领域是否存在通报、约谈、负面舆情等）

A.2.5　有关的诉讼、仲裁、行政处罚、被执行情况

（交易主体在相关领域是否存在诉讼、仲裁、行政处罚、被执行情况）

A.3　交易标的基本情况

A.3.1　交易标的简介

（描述交易标的名称、行业、应用和服务场景、交付形态等）

A.3.2　数据处理及模型情况

（描述交易标的训练数据来源、数据标注、模型机制机理、模型开发、验证情况等）

A.3.3　服务情况

（描述交易标的内容标识、内容监测、未成年人保护、数据安全保护、其他权益保护情况等）

A.3.4　交易标的流通方式

（描述交易标的流通方式、安全保障措施等）

A.4　交易合规评估

A.4.1　主体合规评估

（对交易主体方面的合规情况予以评估）

A.4.2　标的合规评估

（对交易标的的合规情况予以评估）

A.4.3　流通合规评估

（对交易标的流通方面的合规情况予以评估）

A.5　总结

A.5.1　合规风险与对策建议（如有）

A.5.2　结论性意见

附　录　B

生成式人工智能服务交易合规评估建议文档

B.1　主体合规评估建议文档

主体合规评估的建议文档见下表 B.1。

表 B.1　主体合规评估建议文档

评估维度	建议文档
合法维度	a）主体营业信息，包括主体名称、性质、注册资金、营业范围及统一社会信用代码； b）主体经营相关的资质证书及行政许可证书、执照，包括合法有效的经营执照、ICP备案、算法备案、生成式人工智能服务备案或登记、APP备案、ICP许可证等； c）交易主体所获得的与生成式人工智能服务相关的行业认证、数据安全能力认证、备案证明、证书等材料，包括数据安全管理能力认证、数据安全能力成熟度认证、合规管理体系认证等。
安全维度	a）生成式人工智能服务安全专职机构组织建设情况； b）算法、数据安全管理体系文档，包含算法、数据安全总体方针、策略、程序文件、管理制度、相关记录表单等，如算法安全自评估制度、数据安全管理制度、数据安全关键岗位组织架构图、招聘流程、负责人背景资料、数据操作人员招聘流程、日常管理操作规程、数据安全事件应急响应制度及应急预案、数据销毁安全管理制度、个人信息合规审计制度、个人信息合规审计报告等； c）对提供生成式人工智能服务相关供应链上下游服务商的算法、数据安全风险评估报告及评估过程和评估结果材料； d）训练数据提供备份方案； e）训练数据的访问和使用日志。
诚信维度	a）承诺函，如不存在违法交易标的承诺、无违法违规承诺函、不存在舆情事件承诺函等； b）近三年相关的行政处罚、被诉和被仲裁案件的信息等（如有）； c）生成式人工智能服务工作原理； d）生成式人工智能服务宣传材料； e）同时参考其他维度提供的信息。

评估维度	建议文档
权益维度	a）个人权利的告知（例如个人信息保护政策／隐私政策、隐私通知的相关章节、使用协议等）； b）个人权利请求响应流程、管理政策或相关内部发文（如有）； c）个人权利请求处理和响应表单、记录（如有）； d）与合作方的合作协议、数据过程记录； e）第三方管理机制，包括对生成式人工智能服务领域内的第三方合作方管理的文件及机制说明，对第三方管理审查的记录等； f）生成式人工智能服务相关的知识产权权属证明、知识产权约定的协议（如有）； g）科技伦理审查风险评估报告（如有）； h）知识产权、内容合规监测措施和制度。

B.2　标的合规评估建议文档

标的合规评估建议文档见下表 B.2。

表 B.2　标的合规评估建议文档

评估维度	建议文档
合法维度	a）训练来源证明文件，包括数据来源清单、收集方式说明、数据来源合法性说明、爬虫协议（robots协议）、开源协议、采买协议、隐私政策等； b）训练数据内容安全审查方式、手段的说明、过滤关键词清单，如聘请第三方技术支持方的，提供服务协议； c）训练数据的预处理措施说明； d）训练数据标注规则、标注人员情况、培训及考核情况说明； e）模型机制机理评估及资源需求说明； f）科技伦理审查报告及复核结果（如有）； g）自研模型的知识产权证明，如《计算机软件著作权证书》《App 电子版权证书》或《软件著作权认证证书》等； h）第三方大模型的来源证明、备案证明、开源协议及合作协议等； i）与第三方算力服务方的合作协议； j）生成式人工智能服务自评估报告、测试题集等； k）对内容标识的方式、措施说明、截图； l）生成内容安全审查方式、手段的说明、过滤关键词清单，如聘请第三方技术支持方的，提供服务协议； m）对违法和不良内容进行处置的方式和措施、处置记录； n）模型优化措施的说明。

评估维度	建议文档
安全维度	a）存储、传输、访问、数据分类分级等安全措施规范； b）授权可访问和操作加密数据记录； c）数据存储设施安全评估和风险监控报告； d）数据传输完整性校验记录； e）定期审查和更新访问权限记录； f）数据操作行为日志记录； g）定期数据审计和生成内容审核记录； h）模型鲁棒性和安全性评估方案及相关评估记录； i）模型及系统漏洞监察管理方案及相关监察记录； j）系统开发安全编码规范； k）入侵检测系统（IDS）和入侵防御系统（IPS）日志记录。
诚信维度	a）交易标的的描述说明、适用范围、计费方式等信息、服务样例； b）数据交易承诺函、同类交易标的上市承诺函、无重大风险变化承诺函等。
权益维度	a）数据主体权利的告知情况，如隐私政策、授权书、界面截图等； b）数据主体权利请求响应方式、流程、管理政策的说明、文件； c）防范未成年人过度依赖或者沉迷的措施说明、执行证明、截图； d）投诉举报入口、公布处理流程和反馈时限的截图，受理、处理公众投诉举报的证明等； e）辟谣机制说明、保存利用深度合成服务制作、复制、发布、传播虚假信息有关记录的证明等。

B.3　流通合规评估建议文档

流通合规评估的建议文档见下表 B.3。

表 B.3　流通合规评估建议文档

评估维度	建议文档
合法维度	a）交易标的描述说明文档、应用场景说明等材料； b）通过安全评估、备案或登记等手续的证明文件。
安全维度	a）流通安全措施说明； b）网络安全等级保护测评结果通知书或其他数据安全能力证明文件。
诚信维度	a）交易合同； b）承诺函，如依法承担网络信息内容生产者责任、个人信息处理者责任、不存在违法标的承诺、无违法违规承诺函、不存在舆情事件承诺函等。
权益维度	a）接受关于知识产权侵权、不正当竞争等第三方权益、个人信息权益、违法和不良信息相关的权利请求或投诉举报的途径和方式说明（截图）、处理记录。

附　录　C

训练数据及生成内容的主要安全风险

C.1　包含违反社会主义核心价值观的内容

主要的安全风险包括：

a）煽动颠覆国家政权、推翻社会主义制度；

b）危害国家安全和利益、损害国家形象；

c）煽动分裂国家、破坏国家统一和社会稳定；

d）宣扬恐怖主义、极端主义；

e）宣扬民族仇恨；

f）宣扬暴力、淫秽色情；

g）传播虚假有害信息；

h）其他法律、行政法规禁止的内容。

C.2　包含歧视性内容

主要的安全风险包括：

a）民族歧视内容；

b）信仰歧视内容；

c）国别歧视内容；

d）地域歧视内容；

e）性别歧视内容；

f）年龄歧视内容；

g）职业歧视内容；

h）健康歧视内容；

i）其他方面歧视内容。

C.3　商业违法违规

主要的安全风险包括：

a）侵犯他人知识产权；

b）违反商业道德；

c）泄露他人商业秘密；

d）利用算法、数据、平台等优势，实施垄断和不正当竞争行为；

e）其他商业违法违规行为。

C.4　侵犯他人合法权益

主要的安全风险包括：

a）危害他人身心健康；

b）侵害他人肖像权；

c）侵害他人名誉权；

d）侵害他人荣誉权；

e）侵害他人隐私权；

f）侵害他人个人信息权益；

g）侵犯他人其他合法权益（如声音权益等）。

C.5 无法满足特定服务类型的安全需求

将生成式人工智能服务用于安全需求或合规需要较高的特定服务类型时，例如自动控制、医疗信息服务、金融信息服务、关键信息基础设施等，存在的安全风险包括：

a）内容不准确，严重不符合科学常识或主流认知；

b）内容不可靠，虽然不包含严重错误的内容，但无法对服务使用者形成帮助。

附　录　D

典型应用场景下的生成式人工智能服务常见违法违规行为

D.1　概述

本资料性附录结合法律法规和部分行业对生成式人工智能服务的应用，就其典型应用场景下的常见的违法违规行为进行列举说明。

D.2　移动互联网（智能设备）场景

包含但不限于以下行为：

a）提供生成式人工智能服务过程中收集和使用服务使用者的个人信息，如生活习惯、行踪轨迹、设备使用记录、家庭成员身份等，应取得但未获得服务使用者明确同意或单独同意的。

b）涉及家庭、单位等多人使用的智能设备，未采取权限管理和身份识别等技术手段，生成式人工智能服务内容涉及泄露企业商业秘密、个人信息及隐私的。

c）提供生成式人工智能服务过程中持续分析、监测服务使用者的个人信息及行为，如语音、视频、行踪轨迹、设备使用记录等，应取得但未获得服务使用者明确同意或单独同意的。

d）生成式人工智能服务未对服务使用者输入数据及指令中涉及本文件附录C的安全风险进行审核，即向服务使用者提供生成内容及开展操作指令联动的。

D.3　汽车应用场景

包含但不限于以下行为：

a）涉及汽车领域重要数据处理，未按照法律法规规定开展风险评估，并向省、自治区、直辖市网信部门和有关部门报送风险评估报告的。

b）收集和使用服务使用者的个人信息，如驾驶习惯、车辆性能、车辆定位、行驶路线提供生成式人工智能服务，应取得但未取得服务使用者同意或单独同意的。

c）生成式人工智能服务涉及车外处理驾驶习惯、车辆性能、车辆定位、行驶路线、交互语音等个人信息的，应取得但未取得服务使用者同意或单独同意的。

d）在智能座舱中的智能语音助手、导航系统、自动化驾驶等服务模块中利用生成式人工智能服务未进行安全评估，导致存在误判驾驶人意图、交通信号、行人或其他车辆位置等道路交通情况即生成内容或开展操作指令联动的。

e）生成的内容及联动的操作指令涉及违反相关道路和交通安全法规或存在交通安全风险的。

D.4　金融应用场景

包含但不限于以下行为：

a）收集和使用消费者金融信息、其他个人信息，通过生成式人工智能服务进行广告推送、营销，应取得但未取得消费者同意或单独同意的。

b）利用生成式人工智能服务提供金融产品服务，未履行适当性义务，未将合适的金融产品或者服务提供给适当的金融消费者的。

c）未对生成内容进行过滤，未向消费者提供准确的营销资料，致使向金融消费者营销宣传的内容存在虚假、欺诈、隐瞒或者引人误解的宣传。

d）利用生成式人工智能服务向金融消费者告知格式条款，未将与金融消费者有重大利害关系的内容提请金融消费者注意的。

e）利用生成式人工智能服务向金融消费者说明重要内容和风险等情况下，未依照法律法规和监管规定留存录音、录像资料或者系统日志及其他相关资料的。

f）征信机构利用生成式人工智能提供信用评价类产品和服务前，未履行必要的内部测试和评估验证程序，确保评价规则可解释、信息来源可追溯的。

D.5　医疗应用场景

包含但不限于以下行为：

a）利用病患医疗信息作为训练数据，但未进行匿名化处理且未取得病患本人单独同意的。

b）未对生成式人工智能服务医疗应用进行科技伦理审查的。

c）未聘请专业人员对医疗健康类训练数据的真实性、准确性进行审查，致使训练数据存在过时、失实或误导性风险的。

d）生成的内容涉及泄露可识别的病患身份信息的情形。

e）生成内容涉及受版权保护但未依法取得授权的医学资料、文件。

f）未对生成内容或界面进行标识，提示内容安全风险、内容不可直接作为医疗决策建议的。

D.6　人工智能体应用场景

包含但不限于以下行为：

a）非法利用人工智能体应用在服务过程中浏览、访问的网页、APP 等平台内的数据或服务使用者的数据作为训练数据。

b）人工智能体执行验证码自动填充、高频自动抢票、密码爆破、滥发邮件短信等违反公序良俗、违反法律法规或侵犯第三方合法权益的指令。

c）未经服务使用者同意，私自读取服务使用者设备中的其他数据。

d）恶意执行超出服务使用者指令意图的操作。

附　录　E

生成式人工智能服务涉及的常见备案、许可类手续

生成式人工智能服务涉及的常见备案、许可类手续的具体事项、涉及主体、完成时间、法律依据见表 E.1。

表 E.1　生成式人工智能服务常见备案、许可类手续表

序号	事项	主体	完成时间	法律依据
1	互联网信息服务备案	提供非经营性互联网信息服务的主体	从事互联网信息服务前	《互联网信息服务管理办法》第四条
2	增值电信业务许可	提供经营性互联网信息服务的主体	从事互联网信息服务前	《互联网信息服务管理办法》第四条
3	移动互联网应用程序备案	中华人民共和国境内从事互联网信息服务的APP主办者	通过移动互联网应用程序从事互联网信息服务前	《工业和信息化部关于开展移动互联网应用程序备案工作的通知》第二条第1款
4	公安备案	中华人民共和国境内网站/移动互联网应用程序的相应主办者	网络正式联通之日起三十日内	《计算机信息网络国际联网安全保护管理办法》第十二条
5	算法备案	具有舆论属性或者社会动员能力的算法推荐服务提供者	提供服务之日起十个工作日内	《互联网信息服务算法推荐管理规定》第二十四条
6	生成式人工智能服务备案	具有舆论属性或者社会动员能力的生成式人工智能服务提供者	提供服务之日起十个工作日内	《生成式人工智能服务管理暂行办法》第十七条
7	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	具有舆论属性或社会动员能力的互联网信息服务的主体	应用上线前	《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》第三条

参考文献

［1］ GB/T 25069—2022 信息安全技术 术语

［2］ GB/T 22239—2019 信息安全技术 网络安全等级保护基本要求

［3］ GB/T 35273—2020 信息安全技术 个人信息安全规范

［4］ GB/T 35295—2017 信息技术 大数据 术语

［5］ GB/T 37932—2019 信息安全技术 数据交易服务安全要求

［6］ GB/T 41479—2022 信息安全技术 网络数据处理安全要求

［7］ GB/T 41867—2022 信息技术 人工智能 术语

［8］ GB/T 42460—2023 信息安全技术 个人信息去标识化效果评估指南

［9］ GB/T 43697—2024 数据安全技术 数据分类分级规则

［10］ GB/T 45288—2025 人工智能 大模型 第1部分：通用要求

［11］ GB 45438—2025 网络安全技术 人工智能生成合成内容标识方法

［12］ GB/T 45654—2025 网络安全技术 生成式人工智能服务安全基本要求

［13］ GB/T 45674—2025 网络安全技术 生成式人工智能数据标注安全规范

［14］ GB/T 45652—2025 网络安全技术 生成式人工智能预训练和优化训练数据安全规范

［15］ DB4403/T 271—2022 公共数据安全要求

［16］ DB4403/T 439—2024 公共数据安全评估方法

［17］ DB4403/T 564—2024 数据交易合规评估规范

［18］ T/CECC 027—2024 生成式人工智能数据应用合规指南

［19］ TC260—003 生成式人工智能服务安全基本要求

［20］ 中华人民共和国民法典.第十三届全国人民代表大会第三次会议通过.2020 年 5 月 28 日

［21］ 中华人民共和国网络安全法.第十二届全国人民代表大会常务委员会第二十四次会议通过.2016 年 11 月 7 日

［22］ 中华人民共和国出口管制法.中华人民共和国第十三届全国人民代表大会常务委员会第二十二次会议通过.2020 年 10 月 17 日

［23］ 中华人民共和国数据安全法.第十三届全国人民代表大会常务委员会第二十九次会议通过.2021 年 6 月 10 日

［24］ 中华人民共和国个人信息保护法.第十三届全国人民代表大会常务委员会第三十次会议通过.2021 年 8 月 20 日

［25］ 中华人民共和国公司法.第十四届全国人民代表大会常

务委员会第七次会议第二次修订.2023 年 12 月 29 日

[26] 未成年人网络保护条例.国务院令第 766 号.2023 年 10 月 16 日

[27] 网络数据安全管理条例.国务院令第 790 号.2024 年 9 月 24 日

[28] 网络信息内容生态治理规定.国家互联网信息办公室令第 5 号.2019 年 12 月 15 日

[29] 互联网信息服务算法推荐管理规定.国家互联网信息办公室、工业和信息化部、公安部、国家市场监督管理总局令第 9 号.2021 年 12 月 31 日

[30] 数据出境安全评估办法.国家互联网信息办公室令第 11 号.2022 年 7 月 7 日

[31] 互联网信息服务深度合成管理规定.国家互联网信息办公室、工业和信息化部、公安部令第 12 号.2022 年 11 月 25 日

[32] 个人信息出境标准合同办法.国家互联网信息办公室令第 13 号.2023 年 2 月 22 日

[33] 生成式人工智能服务管理暂行办法.国家互联网信息办公室、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、公安部、国家广播电视总局令第 15 号. 2023 年 7 月 10 日

[34] 促进和规范数据跨境流动规定.国家互联网信息办公室令第 16 号.2024 年 3 月 22 日

[35] 国家互联网信息办公室、工业和信息化部、公安部、

国家广播电视总局.关于印发《人工智能生成合成内容标识办法》的通知：国信办通字〔2025〕2号.2025年3月7日

［36］ 深圳经济特区数据条例.深圳市第七届人民代表大会常务委员会第二次会议通过.2021年6月29日

［37］ 深圳市发展和改革委员会.关于印发《深圳市数据交易管理暂行办法》的通知：深发改规〔2023〕3号.2023年2月21日

［38］ 深圳市发展和改革委员会.关于印发《深圳市数据商和数据流通交易第三方服务机构管理暂行办法》的通知：深发改规〔2023〕4号.2023年3月2日