生成式 AI 管理办法 合规白皮书

Argus Gateway · 企业 AI 隐私网关 · 示意版 v0.1 · 更新于 2026-05-23 · 爱科拓 (agilist.cn)

本白皮书示意《生成式人工智能服务管理暂行办法》（2023-08-15 国家网信办等七部门发布）相关条款与 Argus Gateway 能力的对应方向。 Argus Gateway 是企业与上游 LLM 服务之间的代理网关，不是算法提供方、不是模型训练方，亦不是生成式 AI 服务的备案主体。 本文档面向企业法务 / 合规团队评估 Argus Gateway 选型时的方向性参考。

> 阅读说明： 本文件仅就 Argus Gateway 作为代理网关组件时，可为客户在生成式 AI 服务相关合规要求中提供的技术支撑进行说明， 不构成法律意见，不替代企业自有的合规 / 法务团队判断，也不替客户履行算法备案 / 安全评估等法定义务。 详细映射 / evidence 字段 / 配套技术说明，请在申请 demo 后由 Argus Gateway 团队随部署方案一起提供。

1. Argus Gateway 在生成式 AI 管理办法合规链路中的定位

《生成式人工智能服务管理暂行办法》（下称"办法"）规范的对象是"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"的服务提供者。 办法下，合规备案主体是生成式 AI 服务提供者（即模型 / 算法提供方），备案与安全评估义务不可转移给基础设施层供应商。

Argus Gateway 处于企业内部用户与上游生成式 AI 服务（DeepSeek / Kimi / 通义 / 豆包 / Anthropic 等）之间，承担三类与办法相关的技术职责：

- 输入侧脱敏：在请求进入上游生成式 AI 服务前，对常见 PII 做脱敏处理，降低敏感数据在生成请求中的暴露面。
- 审计留痕：记录每次生成请求的身份、时间、脱敏类型、路由决策、上游后端等信息，形成可审计的证据链。
- 路由可控：企业可选择已完成网信办备案的国产生成式 AI 后端，避免接入未备案模型带来的合规风险。

Argus Gateway 不是算法提供方、不进行模型训练、不存储推理内容用于训练，不属于办法第 17 条算法备案义务的承担主体。 企业在选型上游生成式 AI 服务时，应自行核实上游服务的备案状态。

2. 办法条款 × Argus Gateway 能力方向

条款	要求	Argus Gateway 能力方向
第 4 条	生成内容应坚持社会主义核心价值观，不得生成法律、行政法规禁止的内容	Argus Gateway 不参与内容生成，本条主体责任由上游生成式 AI 服务提供方承担。企业层面可结合 Argus Gateway 的审计日志做事后回溯。
第 7 条	训练数据来源合法，不含侵犯知识产权的内容；涉及个人信息的应取得同意或符合法律规定	Argus Gateway 为代理网关，不存储推理内容用于训练，亦不参与上游模型的训练数据收集。本条主体责任由上游算法提供方承担。
第 14 条	服务提供者发现违法内容应及时采取处置措施、保存记录并报告	Argus Gateway 提供按部门 / 用户 / 上游 / 时间维度的审计日志，可作为企业层面"使用记录留存"的支撑材料，配合企业内部内容安全流程。
第 17 条	具有舆论属性或者社会动员能力的生成式 AI 服务应履行算法备案、安全评估	Argus Gateway 不是算法提供方，本条不适用于 Argus Gateway 自身。可向客户提供算法原理 / 路由策略 / 安全机制等技术说明，支持客户在选用上游服务时核验对方备案信息。
第 19 条	提供者应明确并公开使用人群、场合、用途等，处理个人信息应符合 PIPL 等法律规定	Argus Gateway 的多层 PII 脱敏 + 审计留痕 + 国产 LLM 路由组合，是客户在 PIPL 框架下处理"涉个人信息的 AI 调用"场景的可选技术措施。详见《PIPL 合规白皮书》。

3. 脱敏与生成式 AI 输入安全

办法第 19 条要求服务提供者"依法承担个人信息处理者责任，履行个人信息保护义务"。当企业内部用户向上游生成式 AI 服务发起请求时，请求内容中可能携带客户信息、员工数据、内部业务数据等敏感字段。

Argus Gateway 在请求进入上游服务前进行多层 PII 识别与脱敏：

- L1 规则层：覆盖中国常见结构化 PII（身份证 / 银行卡 / 手机 / 邮箱 / 地址 / 港澳台身份证 / 企业统一社会信用代码等）。
- L1b 证据评分层：结合上下文判断识别结果的置信度，降低误判率。
- L2 开放词汇层：面向中文姓名 / 组织名 / 地名等规则难以穷尽的开放词汇 PII 类型。

脱敏后的请求转发到上游生成式 AI 服务，原始 PII 仅在 Argus Gateway 本地保留映射，响应阶段做还原。核心算法（argus-redact）已开源，企业可独立验证识别效果。

4. 审计与"使用记录"留存

办法第 14 条、第 19 条要求服务提供者保存使用记录，对违法内容采取处置措施。Argus Gateway 为每次生成请求生成结构化审计日志，包含：

- 请求方身份（API Key / 部门 / 员工）
- 脱敏层命中情况与 PII 类型分布
- 路由决策与上游后端（含备案状态可登记字段）
- 响应时长与成本归因

审计数据可按部门 / 用户 / 上游 / 时间维度聚合导出，对接企业已有的 SIEM 或合规审计系统。日志保留期可配置，与办法第 19 条"个人信息处理者保护义务" + PIPL 第 19 条"合理期限"协调。

5. 国产 LLM 路由与备案状态登记

办法第 17 条对"具有舆论属性或者社会动员能力"的生成式 AI 服务设定了算法备案要求。企业选型上游服务时，建议优先接入已完成网信办备案的国产生成式 AI 服务。

Argus Gateway 提供国产 LLM 后端的统一接入（DeepSeek / Kimi / 通义 / 豆包等），并在上游配置中保留可登记"备案号 / 服务名称 / 服务提供者"等元数据字段， 作为企业内部审计材料的引用基线。Argus Gateway 不替客户判断上游是否已完成备案，客户应自行核验。

6. 使用建议

- 企业接入生成式 AI 服务前，应由法务 / 合规官完成上游服务的备案状态核验，并将核验结果登记到 Argus Gateway 上游配置。

- 对涉敏感业务的生成请求，建议启用 Argus Gateway 多层 PII 脱敏与本地 / 私有云部署组合，降低数据外泄面。
- 审计日志保留期需与办法、PIPL 第 19 条、网安法第 21 条（6 个月）协调，并预留事后回溯所需时长。
- Argus Gateway 提供的技术说明材料、配置模板均为支撑性参考，最终合规判定以企业法务审定版本为准。

免责声明： 本文件为 Argus Gateway 在生成式 AI 管理办法相关场景下的产品能力方向性介绍，不构成法律意见，亦不替代企业自有的合规 / 法务团队判断。 生成式 AI 服务的算法备案、安全评估、内容合规等法定义务，依法由生成式 AI 服务提供者承担，Argus Gateway 仅作为代理网关提供技术支撑，不替客户履行上述义务。 本文件所述结论以编制日产品状态为准。
