豆包手机助手隐私安全白皮书

1. 概述

AI 技术的创新为移动智能终端的发展带来了更多的可能，以 AI 助手为代表的应用，充分发挥手机的交互优势，为用户提供更智能、更便捷的助手服务。AI 助手能够识别用户的意图并快速执行任务，还能根据用户习惯提供个性化回复。在带来便利性的同时，我们也深知用户隐私安全的重要性，因此豆包手机助手在所有产品和服务的设计过程中，始终将用户的数据与隐私安全性作为首要考虑因素之一。

本安全白皮书详细介绍了豆包手机助手为提升产品和服务的信息安全能力而实施的具体管理和技术措施，主要分为如下四个章节进行介绍：AI 隐私安全基础、AI 应用隐私安全、数据安全及隐私保护、安全管理体系。

在互联互通的世界中，不断发展的技术给我们带来了新的信息安全风险。豆包手机助手力求为用户提供安全、放心、可靠的产品。我们希望本白皮书让用户能够更好的理解我们对豆包手机助手产品和服务的信息安全及隐私保护的理念和实施。如果您对本白皮书有任何疑虑或疑问，请随时发送邮件至 o-feedback@mail.doubao.com 告知我们。

2. AI 隐私安全基础

豆包手机助手坚信，保障用户对数据的控制权是 AI 服务价值的前提。我们不将隐私安全视为“附加功能”，而是将其嵌入 AI 业务的全生命周期中——从模型能力的早期设计，到用户交互的信息传输，再到服务结束的敏感数据销毁，每一环均以 “主动防御”替代 “被动应对”，同时保障用户知情权，确保 AI 技术的创新与应用，始终在尊重用户隐私、守护数据安全的轨道上前进。让每一位用户在享受 AI 便捷的同时，无需为隐私安全担忧。

2.1 AI 业务架构

豆包手机助手以端云 AI 安全防护体系为架构基础，结合字节跳动的自研大模型，构建了一系列的具有创新体验的智能 AI 业务。

2.2 AI 平台隐私安全

2.2.1 端侧 AI 隐私安全

运行环境安全

豆包手机助手依托安全硬件平台运行，硬件平台配备有 NPU（Neural-network Processing Units）推理引擎，同时具备 TEE（Trusted Execution Environment）硬件安全能力。豆包手机助手借助硬件级的隔离与加密保护机制，实现对用户数据的全方位守护。

本地模型安全

豆包手机助手基于可信 TEE 与硬件级文件加密技术，对模型数据的预置、分发、更新、存储，以及模型的推理应用进行了全链路的防护，以保护模型算法的安全性，以及防止端侧推理过程的用户数据泄露。

1. **模型数据保护：**模型数据由预置于可信 TEE 内的平台密钥进行加密，分发更新模型之前，基于 PKI 的可信设备认证技术确保模型分发到安全的硬件设备上。加密模型在设备上由 TEE 安全服务进行解密，并经硬件级加密技术，再次加密存储于专用文件加密目录下。

2. **模型推理安全：**在系统完整性度量检测服务的监控下，模型从本地加密目录解密并加载到推理框架服务中，当系统检测到特权攻击时，将拒绝提供模型解密与加载服务，从而保障模型与推理过程的用户数据安全。

本地模型安全推理

豆包手机助手在手机本地部署了一部分的 AI 模型，以实现数据仅在设备本地处理的高敏感隐私数据保护功能。目前主要的本地大模型包括以下几类，分别应用在 AI 业务的各种高敏感数据处理场景，并且后续会持续优化并应用到更多的场景。

OCR 模型	用于对用户本地图片中的文本提取，识别。
文本向量化模型	用于对用户数据进行向量化处理与快速检索。
命名实体识别模型	用于提取业务数据中的实体命名信息。
内容识别模型	用于本地模型推理输入输出的内容安全与合规审查。
多模态大模型	用于对本地图片、视频理解，文本总结等。

2.2.2 云端 AI 隐私安全

敏感数据不参与训练

豆包手机助手具备完整的安全架构设计，从技术架构上确保不会使用用户敏感数据训练模型，原始用户数据仅用户于处理业务请求，相关生成的业务数据受严格保护且与训练数据隔离。

数据严格权限管控

云端平台通过细致的权限管控，确保在用户未明确授权情况下，避免用户敏感数据在非必要的场景下以任何形式被任何第三方存储、访问、使用。

全链路数据安全保护

技术层面安全保护措施覆盖，不依赖人为约定的保障措施，数据在传输和存储过程中均为加密状态，涉及到模型推理会通过脱敏、去标识化等手段剥离与自然人关系，无法定向查看可关联用户的明文敏感数据。

1. **传输安全：**端云链路数据传输加密，服务间调用具备身份管控和鉴权，保障数据的机密性和完整性
2. **存储安全：**云端数据通过 TKMS（可信密钥管理服务）加密存储，加密算法使用 AES-256-GCM
3. **日志安全：**所有日志打印均遵循豆包手机助手日志打印规范，不允许打印用户个人信息等相关敏感数据，且有专门的安全团队执行验收

可信计算

为保障云端安全能力的可信执行，云端的关键服务部署于机密计算容器中，所有处理逻辑，通过机密计算安全隔离特性，实现明文数据与外部环境隔离。后续通过长期构建的可信计算推理、远程证明等技术，保障从技术理论角度证明无人可访问用户的敏感数据。

模型安全防护服务

为了保障云本地大模型能够安全可靠的服务于 AI 业务，我们构建了模型防护服务，对大语言模型与多模态模型的输入与输出进行分析检测与过滤，防范提示词攻击、意图执行偏离、敏感信息泄露等安全风险。

2.2.3 云端服务安全

云端基础设施安全

豆包手机助手遵从纵深防御的安全理念，建立了完善、先进的云端安全防护架构，在网络到主机与容器的各个层面均采用了先进的安全防护技术。

包括但不限于，在网络边界部署高级网络威胁检测系统（Network Traffic Analysis，NTA），对恶意网络入侵行为进行检测和告警。所有流量都经过应用防火墙（Web Application Firewall，WAF）进行攻击检测和验证，确保其安全性与合法性，对恶意的请求予以实时阻断。物理服务器全面部署了服务器入侵检测系统（Host-based Intrusion Detection System，HIDS），可以实时监控发现异常进程、捕捉异常外连、木马后门等异常行为，并及时作出响应。为了应对分布式拒绝服务攻击（Distributed Denial-of-Service Attack，DDoS），我们还部署了 DDoS 防护系统，自动对攻击流量进行拦截，再将正常流量转发回服务器，从而保证业务不会受 DDoS 攻击的影响。

此外，安全团队会密切跟踪安全态势和最新的攻击手法，不断迭代和升级安全防御措施，持续保障业务安全。

云端业务系统安全

为提升豆包手机助手云端服务相关系统的安全性及可用性，防止系统在上线后遭受潜在的安全攻击，我们会在其上线前进行严格的安全测试，测试通过后才可上线。在测试时，我们会参考行业最佳实践（如 OWASP Top 10 Web Application Security Risks），结合实际业务情况，量身定制技术测试方案，并由专业的安全测试人员进行测试。此外，豆包手机助手接入层均采用了高可用的方式接入，采用多实例方式运行，能够保证服务的可靠性和稳定性。

3. AI 应用隐私安全

豆包手机助手深知用户隐私安全的重要性，并将其视为产品设计的核心基础，功能设计均紧紧围绕这一核心考量展开，遵循授权同意、最小必要、用户可控原则。

在授权同意方面，无论是豆包助手开启时，还是启用其他应用时，我们都会以向用户阐释数据处理规则。只有在获得用户同意后，相关功能才会正式启用，确保用户对自身数据的使用拥有知情权。

用户可控原则是指用户对数据和功能的高度控制权。以 AI 操作手机为例，当涉及可能影响用户重要权益的操作时，助手会主动征求用户的确认，确保操作符合用户的真实意图。

3.1 豆包助手

在唤醒豆包助手时，如果用户手机处于锁屏状态，为避免未经授权执行高敏感操作，豆包助手会启用安全访问鉴权，只有鉴别通过后才能执行后续高级操作或询问敏感数据，保障用户的隐私安全。

当用户通过豆包助手对本地数据进行访问和检索时，用户可依据自己的需求和隐私偏好，按应用进行控制，自行选择支持检索的数据范围，把控个人数据的访问范围，保护用户隐私。

豆包助手在提供意图执行的能力时，为了确保意图执行时安全性，在意图注册、意图执行以及数据访问与使用等核心操作中提供意图访问控制能力，所有操作都需通过安全访问鉴权后才能执行，以此保证操作发起方的合法性，防止越权操作。

为了更灵活地完成用户指令，豆包手机助手增加了 AI 自动帮用户操作手机的功能。在操作时，为避免 AI 过度授权带来潜在风险，我们对涉及用户敏感数据的操作制定了严格管控机制，必须得到用户明确确认和授权才能进行。特别是高敏操作，需要用户二次确认后才能接入并执行。具体包括以下六大类操作：权限与隐私管理、金融与交易操作、系统与设备控制、法律与合规要求相关操作、高风险不可逆操作、健康与医疗数据处理。

3.2 全局记忆

全局记忆在数据获取和处理过程中均采取了严格的安全防护措施，同时严格遵循用户可控原则。具体的保护措施如下：

1. **合理必要原则：**数据获取遵循合理必要原则，只收集完成全局记忆功能目的所需的数据，避免过度采集用户数据，从源头降低隐私泄露风险。

2. **优先端侧处理原则：**数据处理尽可能在端侧完成，处理完成后加密存储在端侧。如果因为设备计算能力或其他技术限制，必须把数据上传到云端处理时，传输和处理过程均采取严格的保护措施保障用户数据安全，数据使用完后会立即删除。

3. **数据用途明确：**全局记忆记录的数据仅用于为用户提供记忆查看与智能回复，不会用于非功能所需的场景，不会用于训练，确保数据用途的单一性和安全性。

用户可控原则方面：

1. **授权可控制：**开启全局记忆的各项功能时，都会给出清晰的隐私声明，告诉用户涉及的数据范围，只有在用户授权同意后才会开启功能。

2. **功能开关可控制：**开启记忆功能后可以随时关闭，控制中心有快速关闭按钮，方便用户随时停止记忆功能，及时保护个人隐私。

3. **数据管理可控制：**支持删除部分数据或清空全部记录，用户可以根据自己的需求灵活管理自己的记忆数据，决定数据的保留或删除。

4. **记录感知可控制：**记忆过程实时状态栏提示，用户能知道记忆功能什么时候在记录信息，对信息记录有清晰的了解和掌控。

5. **查看权限可控制：**查看记忆时需要验证锁屏密码，确保只有用户自己能查看记忆内容，防止他人未经允许访问个人记忆数据。

6. **记录范围可控制：**支持手动设置允许被记忆的应用、自定义记忆的时间段、自定义地理围栏，也可以用自然语言要求不记录某些信息，精确控制记忆的记录范围。

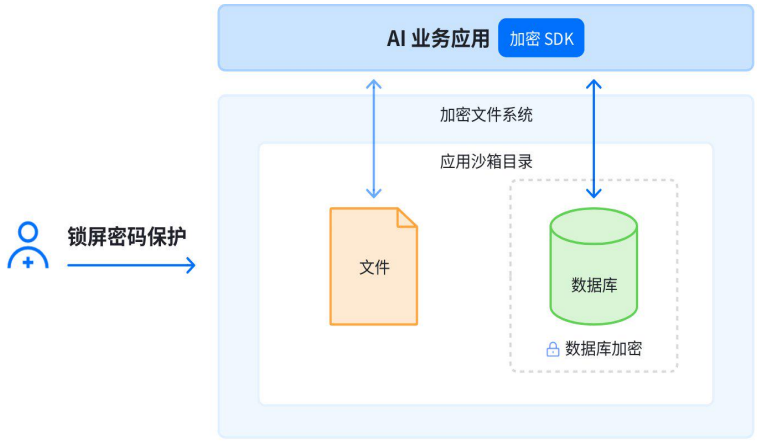
4. 数据安全及隐私保护

豆包手机助手为用户提供了全生命周期的数据安全防护解决方案，确保用户数据在采集、存储、传输、访问、销毁等每一个阶段，都能获得安全保护。 我们同时参考业内的最佳实践定义了数据分类分级标准，以及对应的安全保护措施要求，确保安全防护措施与个人数据敏感程度、数据资产价值相匹配，最大限度地减少未经授权的数据使用、数据暴露和泄露的风险。

4.1 数据加密保护

4.1.1 端侧数据存储加密

基于系统文件加密机制，AI 业务采集的用户数据与文件存储于受设备锁屏密码保护的应用沙箱加密目录内。对于高敏感的个人记忆数据库，还采用了数据库加密技术，仅在 AI 业务运行时解密使用，进一步增强数据安全。



4.1.2 端云全链路加密

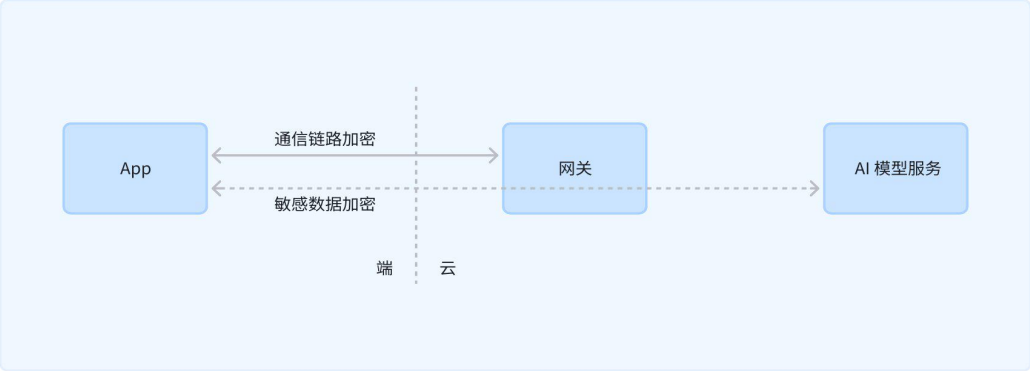
通信链路加密

豆包手机助手预置了字节签发的 TLS 安全证书，端云通信采用双向认证加密通信技术。端侧 AI 业务在进行上云推理请求时，由端、云安全服务，对设备，云服务进行

双向的身份认证，确保设备、云服务均为可信通信实体，之后建立可信加密通道，保障通信数据安全。

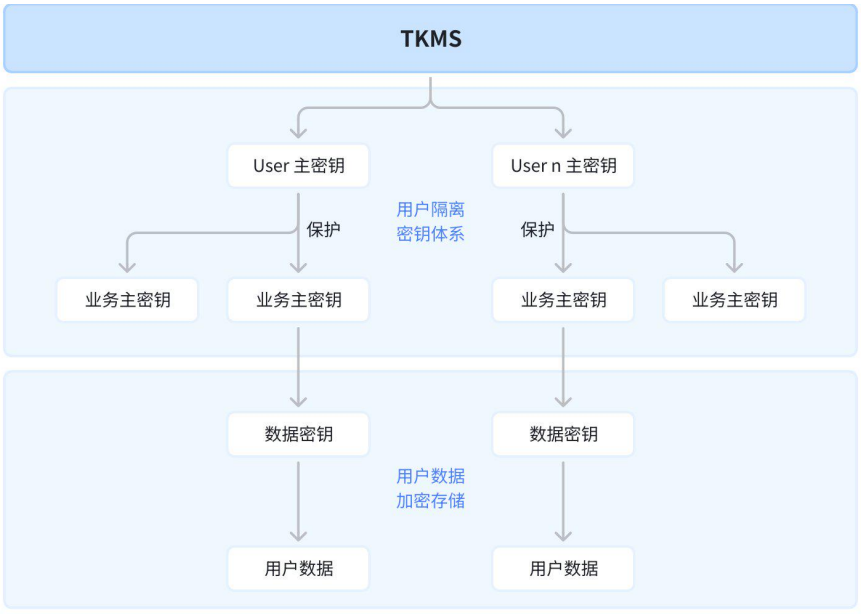
敏感数据加密

为了进一步防范用户数据泄露隐患，我们对涉及用户隐私的上云推理请求数据，进行了端到云的二次加密，从而保障推理数据在端 APP 与云服务之间的安全加密传输。在会话建立之初，端侧隐私安全服务通过与云端安全保护服务协商一次性用的加密密钥，并以此加密推理请求数据，保障数据的安全。推理完成之后，加密密钥与数据立即销毁，从而防止非法的数据分析以及与本次请求不相关的数据使用。



4.1.3 云端数据存储加密

豆包手机助手为用户构建了用户隔离的数据加密存储方案。每个用户拥有独立的数据加密密钥体系，对云端存储的数据与文件实现一文一密的加密保护，以防止跨越用户的非法数据访问。用户数据在上云之前在端侧执行加密，密文数据上云后，业务在未授权情况下无法直接使用明文数据。



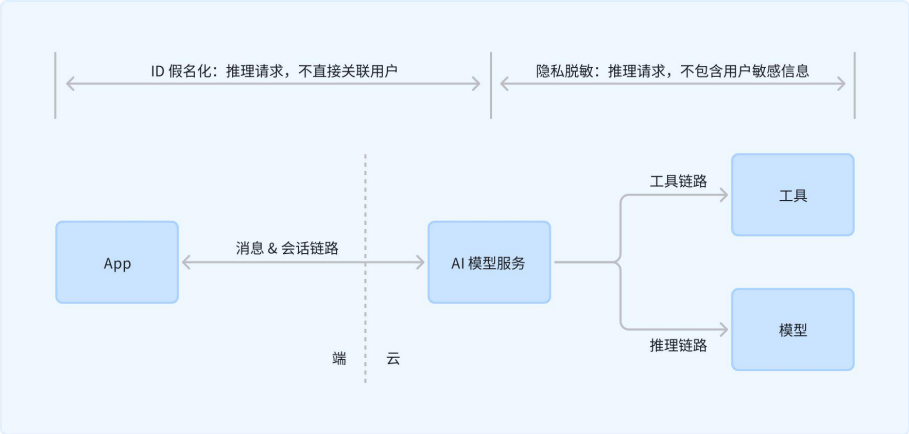
4.2 数据脱敏保护

4.2.1 ID 假名化

豆包手机助手为用户构建了保护隐私的推理链路方案。部分端侧涉及用户隐私数据的上云的推理请求，会通过端云协同的安全服务会对用户账号 UID，设备 DID 进行假名化处理，即采用一个伪 ID（pseudonym id）替代真实 ID（identity）来标识数据主体，使得 AI 业务上云请求不包含真实的用户 UID 与设备 ID，业务端数据分析和数据处理时无法直接跟踪分析用户。

4.2.2 隐私脱敏

推理请求在进入到模型推理前还会对用户信息做进一步的匿名化与脱敏处理，对可直接关联用户个人的敏感信息进行识别并脱敏，以减少用户信息的暴露可能。



4.3 隐私保护

4.3.1 隐私保护制度

豆包手机助手积极响应中国法律法规要求，以《个人信息保护法》、《数据安全法》、《网络安全法》、GB/T 35273-2020《信息安全技术个人信息安全规范》、GB/T 41391-2022《信息安全技术 移动互联网应用程序（App）收集个人信息基本要求》、YD/T 2407-2021《移动智能终端安全能力技术要求》等重点法律标准为核心，进行细化解读，并结合豆包手机助手业务特点制定《数据安全&隐私合规&安全认证要求指引》、《App 违规收集、处理个人信息认定标准与自查指引》、《业务安全设计流程规范》、《端云安全设计基线》等内部隐私合规指引规范与隐私合规自查清单。

4.3.2 隐私保护实践

隐私保护影响评估

为了加强对用户个人数据的管理，通过开展个人数据隐私影响评估（PIA，"Privacy Impact Assessment"），降低隐私安全风险发生概率，豆包手机助手建立了隐私影响评估管理流程。针对个人数据处理活动，检验其合法合规程度，判断其对个人数据主体合法权益造成损害的各种风险，评估用于保护个人数据主体的各项措施有效性，确保业务收集、处理、使用个人数据的合理合法性和正当必要性，保障个人数据主体的合法权益。

为了保障 PIA 流程的高效执行，我们设立了隐私合规评审组，受理隐私合规相关评估需求，输出合规评估意见和合规方案，跟进合规方案落地等工作。PIA 流程贯彻在豆包手机助手开发生命周期的完整流程中，基于 PIA 范围开展 PIA 评估，识别隐私合规相关风险并采取相应的风险处置措施，确保豆包手机助手满足集团安全隐私合规要求。豆包手机助手会在以下场景发起 PIA 流程：

1. 涉及收集新的个人数据
2. 涉及强制用户提供个人数据
3. 涉及向之前没有访问权限的组织或个人披露个人数据
4. 涉及将已收集的个人数据用于新的使用目的
5. 涉及使用可能被视为侵犯隐私的新技术
6. 涉及使用敏感个人信息
7. 法律法规/国家标准或者监管政策发生变化
8. 对个人数据的处理内容与方式发生变化可能带来较高风险的其他场景

隐私合规检测能力

为了有效落地隐私合规要求，我们在业务开发和迭代流程中采用人工检测与自动化工具检测结合的方式开展隐私合规检测。隐私合规检测项覆盖了数据全生命周期涉及的合规要求，并同时结合了个人信息保护法中对个人信息处理的各项要求。隐私合规检测对象包含了所有自研预置应用软件和第三方预置应用软件等。通过动静态结合的隐私合规检测技术，识别预置应用软件中存在的隐私合规的风险，静态检测技术包含敏感 API 扫描分析、敏感权限扫描分析等，动态检测技术包含敏感函数调用分析等。豆包手机助手针对数据“端侧产生、端云流转” 的特点，围绕数据生命周期各个阶段对用户个人数据进行保护，兼顾“安全防护” 与 “用户体验”，通过底层技术阻断隐私泄露风险，同时让用户能直观感知、自主管控隐私权限和个人数据。

4.3.3 隐私保护表现

豆包手机助手对于用户个人数据的隐私保护贯穿在数据全生命周期中：

1. **授权：**用户对个人数据拥有自主控制权，可在客户端自主管理各 AI 应用的数据使用权限，并从数据类型维度进一步细化控制开关，确保用户实时掌控各类数据使用，

以“可知、可控、可退”的核心原则，保障用户权益。

2. **传输**：应用间的数据传输、用户发送的语音、提示词、图片等信息均为加密传输，防止中途被截取，确保数据从用户端到服务端全程安全。

3. **存储**：采用加密技术将数据转化为密文存储，仅保留必要交互数据，且关联的账号标识也加密处理，超期数据自动清理，降低存储风险。

4. **使用**：语音数据仅用于转文字和理解需求，提示词敏感信息自动脱敏，图片个人信息 AI 识别后加强保护，所有数据仅服务于响应需求，敏感信息均会做匿名化处理，并且默认不用于模型训练。

5. **管理**：实施“最小权限”访问控制，内部操作全程留痕，用户可随时查看、定向删除或一键清除数据，注销账号则同步删除所有关联信息，保障用户对数据的控制权。

5. 安全管理体系

豆包手机助手为 AI 安全搭建的完整管理体系——从组建专业安全团队，到在项目设计开发时就融入安全考量，再到严格管理合作的供应商，以及制定突发安全事件的应对方案。我们希望通过制度规范、技术保障和专业团队的配合，为用户数据的全生命周期筑起安全屏障。

5.1 专职的安全与隐私团队

豆包手机助手业务设立了专职的安全团队与隐私团队，形成技术与合规的双重保障：安全团队聚焦技术防护，贯穿 AI 系统全生命周期，深度参与软件开发生命周期（SDLC），在需求分析阶段明确安全指标，设计阶段评审架构安全性，编码阶段提供安全培训与指导，测试阶段开展安全测试，上线后持续监控并快速处置异常；隐私团队专注合规管控，审核数据采集范围与授权规范、监督数据加密脱敏落地，还跟踪法规更新并开展合规审计，守护用户数据权益。

5.2 安全设计与安全开发流程

为了向用户提供更好的安全防护，豆包手机助手深入践行安全设计（Security By Design）和隐私设计（Privacy By Design）理念。产品设计阶段中，我们会将信息安全和隐私保护作为重要的考虑因素之一，从而为您的 AI 体验的安全与隐私提供充分保护。

此外，我们围绕 AI 业务特点，建立了从产品需求、设计、研发、测试到上线整个生命周期的安全开发流程，分几大阶段落实安全管控：

1. **风险需求识别**：由安全团队牵头，联合业务、开发团队梳理 AI 系统的业务场景与功能需求，同步识别潜在安全风险（如数据泄露、模型被攻击等），明确风险防控目标

与安全需求，形成“风险清单”，作为后续开发的安全基准。

2. **安全评审**：在系统设计方案定稿后，组织跨部门安全评审会，安全团队从方案安全性、数据流转合规性等维度进行评审，重点核查设计方案是否覆盖风险需求识别阶段明确的防控目标，未通过评审的方案需优化后重新提交，确保设计环节无安全遗漏。

3. **安全测试**：开发团队完成功能模块开发后，向安全团队提交安全测试工单，工单需明确模块功能、涉及数据类型及测试范围；安全团队根据工单内容制定专项测试方案，明确测试方法（如渗透测试、漏洞验证等）与验收标准，确保测试针对性与有效性。

4. **增量扫描**：针对开发过程中的代码迭代与功能更新，启用自动化扫描工具，聚焦新增代码、修改模块开展实时安全扫描，及时发现代码漏洞、依赖组件风险等问题；扫描结果同步至开发团队，明确整改要求与时限，安全团队跟踪整改进度，确保风险修复。

5. **上线前安全验收**：产品发布前，安全团队对照“风险清单”开展全面验收，核查风险防控措施落地情况、增量扫描问题整改效果，同时进行全量渗透测试；验收通过后出具验收报告作为产品发布的必要依据，未通过验收的项目需整改后重新验收。

5.3 安全漏洞管理与响应

豆包手机助手团队建立了完善的安全漏洞管理流程，能够及时发现、响应和分析新的安全漏洞，并在最短时间内提供风险补救和修复措施。我们的安全专业团队会定期对产品进行安全测试和攻防演练，尽可能地发现产品中存在的安全漏洞和风险。流程各阶段明确了角色分工与时限要求。例如，在风险收集阶段，需在 9 小时内完成情报捕获与研判；在应急响应阶段，要在 96 小时内实现重点业务修复。

豆包手机助手业务拥有专业协作团队，团队成员涵盖漏洞情报、安全攻防、PMO 等角色，各角色各司其职。其中，PMO 全程跟踪流程进度，以确保流程得以有效落地。在技术支撑方面，公司通过多维度情报监测（覆盖 CVE、社交媒体等渠道）、全场景

影响面排查（包括服务端、客户端、办公网等场景）以及 7x24 小时攻击监控，并结合 WAF、HIDS 等工具实现安全防护。

豆包手机助手团队还建立了完善的漏洞修复机制，涵盖对外通报（通报对象包括内部全员、云上客户、监管部门）到存量修复推进、增量管控，再到复盘优化等环节，从而保障漏洞修复率。