

启明星辰 | 大模型应用安全 | 深度应用安全 | 基础安全 | 安全运营 | 安全服务

AI就绪的大模型身份与访问管理

AI-R-IAM v1.0

AI-Ready Identity and Access Management

启明星辰信息技术集团股份有限公司

2025年2月



版权声明

解释权

北京启明星辰信息技术有限公司版权所有，并保留对本文档及本声明的最终

又。

和修改权

文档中出现的任何文字叙述、文档格式、插图、照片、方法、过程等内容，除另有特

本

别注明外，其著作权或其他相关权利均属于北京启明星辰信息技术有限公司。未经北京启明星辰信息技术有限公司书面同意，任何人不得以任何方式或形式对本手册内的任何部分进行复制、摘录、备份、修改、传播、翻译成其它语言、将其全部或部分用于商业用途。

免责声明

2008年10月

如不另行通知

本文档依据现有信息制作，其中内容如有更改，恕

北京启明星辰信息技术有限公司在编写本手册的时候已尽量保证其准确性，但并不能完全保证本手册中的内容完全正确，北京启明星辰信息技术有限公司

确可靠，但北京启明星辰信息技术有限公司不对本文档中的遗漏、不准确、或错误导致的损失和损害承担责任。

信息反馈

如有任何宝贵意见，请反馈：

信箱：北京启明星辰信息技术有限公司 010中关村软件园 21 号楼启明星辰大厦 10 楼

100193 电话：010-82779088

传真：010-82779000

获得最新技术和产品信息

您可以访问启明星辰网站

启明星辰

2000年12月，启明星辰在深交所挂牌上市。

2001年，启明星辰成立。

2002年，启明星辰在深交所挂牌上市。

2003年，启明星辰成立。

2004年，启明星辰在深交所挂牌上市。

2010年6月23日，启明星辰在深交所中小板正式挂牌上市。

2011年，启明星辰在深交所挂牌上市。

终端管理、加密认证等技术领域，共有百余个产品型号，并获国家级产品认证。

2012年，启明星辰在深交所挂牌上市。

2013年，启明星辰在深交所挂牌上市。

2014年，启明星辰在深交所挂牌上市。

2015年，启明星辰在深交所挂牌上市。

发展成为国内统一威胁管理、安全管理平台国内市场第一位，安全审计、安全专业服务市

多家分支机构，拥有覆盖全国的渠道和

场领导者。目前，公司在全国各省市自治区设立三十

售后服务体系。

的关怀与鼓励。2000年1月，江泽民、

长期以来，启明星辰公司得到了党和国家领导人

公司；2003年1月，胡锦涛总书记亲

李岚清、曾庆红等党和国家领导人亲切视察启明星辰

切接见了启明星辰公司 CEO 严望佳博士。

布局内重点软件企业，国家火炬计划软

凭借多年来的潜心研发，启明星辰获得国家规划

及拥有最高级别的涉及国家秘密的计算

件产业优秀企业，中国电子政务 IT100 强等荣誉，及

机信息系统集成资质证书。

研究基地，完成包括国家发改委产业

启明星辰是目前我国规模最大的国家级网络安全

示范工程，国家科技部 863 计划、国家科技支撑计划等国家级科研项目近百项。创造了百

余项专利和软件著作权 参与制订国家及行业网络安全标准,填补了我国信息安全科研领域

牛乳と母乳

家庭暴力 - 家庭暴力

[illegible]

为世界五百强中 80% 的中国企业/客户提供安全产品及服务，在金融领域，自明星后对政策

性指标, 同时控制除碳效率, 全水型除碳效率达到 90% 以上, 在 100℃ 以上。

[illegible]

人 群 中 的 金 子 王 子 家 庭

小学

全产品和最佳实践服务，帮助客户全面提升其 IT 基础设

推进国际化的品牌佳自宁企率川第_只曲画不除奴十

2 大模型的发展与风险

2.1 新质生产力“大模型”

领域的核心驱动力。从 ChatGPT, GPT-4, Gemini 等, 大模型在自然语言处理、

近年来, 大模型技术取得了突破性进展, 成为人工智能领域的重要突破。从 OpenAI 的 GPT-4, 再到百度文心一言、阿里通义千问、九天大模型、DeepSeek

能力, 其可实现文本生成、智能问答、图像

够学习到丰富的语言知识, 语义理解和逻辑推理

多样化的任务。大模型发展的历程分为四个阶段:

生成、代码编写等

2022 年 11 月)

● 准备期 (

, OpenAI 发布 ChatGPT, 其强大的自然语言处理能力震撼全球, 开启

2022 年 11 月

促使各方加大在该领域的投入与研究

大模型发展浪潮

长期 (2023 年)

● 成

OpenAI 发布 GPT-4, 谷歌发布 Gemini Pro, 微软发布 Copilot, 百度发布文心一言

(1)

合自身技术优势; 阿里发布通义千问, 立足丰富业

(2) 国内: 百度推出文心一言, 整合

产学研, 高校和科研机构纷纷开展研究, 为

友达昇, 到工业级推出工业大模型, 关注

模型发展基础技术储备。

模型训练技术储备

模型训练技术储备

模型训练技术储备

模型训练技术储备

4月, 在首届移动人工智能生态大会上, 中国移动发布

(2) 国内: 2024年5月12

模型训练技术储备

了“九天”人工智能基座, 包含千亿参数的灵犀、千亿多模态大

模型不断迭代，众多初创企业也凭借特色大模型在细分领域崭露头角。

● 大规模应用期（2025 年 1 月-）

(1) 国外：大模型深入医疗、金融、教育等多行业，助力疾病诊断、风险评估、个性化学习等。OpenAI 还在探索新应用领域。

(2) 国内：DeepSeek 凭借创新算法和架构，展现出低成本、高效能以及开源优势，

用中安全挑战

2.2 大模型应用

选择模型的安全挑战

安全上

《原景》调查：78%的高管同意在未来必须为 AI Agent 构

根据埃森哲的《2025 年技术

Identity, NHI)，都暴露出诸多安全问题。

类身份代表自然人身份实体，在十模型应用在巨场景中的些

1. 类身份安全问题：人

2. AI Identity 安全问题：越来越多的 AI Agent 在 I 模型中扮演 AI Agent

3. 类身份安全问题：人

4. 类身份安全问题：人

AI Agent 应视为具有身份的独立新兴技术，应参照参考 AI Agent 作为员工，在大

模型中扮演 AI Agent 角色，AI Identity 安全问题：越来越多的 AI Agent 在 I 模型中扮演 AI Agent

种。可能会出现如下风险：

1. 攻击者可能利用 AI Agent 的漏洞对模型进行攻击，如通过提示注入、

攻击者可能利用 AI Agent 的漏洞对模型进行攻击，如通过提示注入、

2. 攻击者可能利用 AI Agent 的漏洞对模型进行攻击，如通过提示注入、

攻击者可能利用 AI Agent 的漏洞对模型进行攻击，如通过提示注入、

非法访问权限。

(2) 访问权限放大：与人类身份不同，AI Agent 不会以可预测的方式请求访问权限，

如果权限管理机制不完善，可能导致用户或应用程序获得超出其应有的权限，从而能够访问

数据或系统破坏。

和操作敏感资源，造成数据泄

3. 类身份安全问题：人

(3) 类身份安全问题：人

隐私泄露，进而可能被用于精准诈骗、

系统遭受黑客攻击，这些数据可能被窃取，导致用户

定向广告骚扰等。

定义为企业技术内的应用程序、服务或

3、非人类身份（NHI）安全问题：非人类身份

密码、账号、OAuth 令牌、API

机器等实体绑定的数字身份。这些包括机器人、API

2025 年 2 月发布《十模型应用安全白皮书》

2025 年 2 月发布《十模型应用安全白皮书》

2025 年 2 月发布《十模型应用安全白皮书》

2025 年 2 月发布《十模型应用安全白皮书》



NHI 在大模型应用中，大多数存在于大模型基础设施交互过程中，NHI 的身份生命周期管

理薄弱，存在诸多安全风险与问题：

(1) 影子账号：大模型应用迭代快，加之生命周期流程弱、账号使用信息不可见，许多 NHI 账号不活跃，增加攻击面。

(2) 影子账号：大模型应用迭代快，加之生命周期流程弱、账号使用信息不可见，许多 NHI 账号不活跃，增加攻击面。

(3) 缺乏账号所有权：多数参与大模型应用的组织中，NHI 无明确所有权信息，而明确

管理者对安全维护和问题补救很关键。

(4) 缺乏环境隔离：在许多场景中，相同 NHI 会在生产环境非生产环境

在每个环境中都有相同的密码，从而增加了横向移动的风险。

的逻辑 NHI 在

杂，难以清晰理清依赖关系。

2.2.2 数据安全挑战

大模型应用在数据安全方面面临如下挑战：

(1) 提示词输入/输出风险：提示词可能包含敏感信息，若大模型系统对输入

不完善，敏感信息可能在输入环节直接暴露。在输出结果时，若对输

示词的验证和过滤机制不

完善，敏感信息可能在输入环节直接暴露。在输出结果时，若对输

示词的验证和过滤机制不

的提示词用于工业场景中模型进行诊断推理时，输出结果若敏感处理，可能导致企业生产数据泄露。

存在漏洞，不法分

(2) 大模型 API 调用风险：API 调用过程中，若身份认证和授权机制

API 接口若缺乏有效

子可能冒用合法身份调用 API，获取敏感数据或进行恶意操作。同时，A

安全防护，易遭受攻击，导致数据泄露或篡改。比如攻击者通过漏洞获取 API 调用权限，

非法获取金融大模型中的用户资产数据。

(3) RAG 知识库查询风险：RAG 知识库整合大量数据，若查询权限控制不当，用户可能

数据泄露或篡改风险。

数据泄露或篡改风险。攻击者可能通过漏洞获取 API 调用权限，非法获取金融大模型中的用户资产数据。

数据泄露或篡改风险。攻击者可能通过漏洞获取 API 调用权限，非法获取金融大模型中的用户资产数据。

(4) 数据泄露风险：在大模型训练阶段，企业

数据泄露或篡改风险。攻击者可能通过漏洞获取 API 调用权限，非法获取金融大模型中的用户资产数据。

数据泄露或篡改风险。攻击者可能通过漏洞获取 API 调用权限，非法获取金融大模型中的用户资产数据。

数据泄露或篡改风险。攻击者可能通过漏洞获取 API 调用权限，非法获取金融大模型中的用户资产数据。

数据泄露或篡改风险。攻击者可能通过漏洞获取 API 调用权限，非法获取金融大模型中的用户资产数据。

数据泄露或篡改风险。攻击者可能通过漏洞获取 API 调用权限，非法获取金融大模型中的用户资产数据。

(5) 合规

数据泄露或篡改风险。攻击者可能通过漏洞获取 API 调用权限，非法获取金融大模型中的用户资产数据。

数据泄露或篡改风险。攻击者可能通过漏洞获取 API 调用权限，非法获取金融大模型中的用户资产数据。

化原则、未获授权使用公开数据等，可能面临法律诉讼和罚款。

、行为风险、数据风险、业务风险、设备

可信环境基于大数据安全能力，实现网络风险

风险的全链路安全风险审计。

围扩展到安全设备、系统资源、应用资源、

访问可信与数据可信作为一体两面，将防控范围

、终端、全网设备、员工操作、可、AI资源等，通过“ID、自公”的维度形成可信身份

组合性和可扩展性，使企业能够以更少的资源实现更好的安全性，搭配一体化安全机制的融

全协同，实现全程全网、端到端的安全保护

身份、

随着大模型作为新质生产力，在政府、企业、个人中应用越来越深入，需要针对其

化全程

访问管理、数据安全等方面的需求，建立一套基于大模型技术特点和业务场景的一体化

可信的数字身份基座，即 AI 就绪的大模型身份与访问管理。



大模型身份与访问管理

4 AI 就绪的

身份管理(以下简称: AI-R-IAM)是为大规模人工智能模型(如

AI 就绪的人模型身

身份管理(以下简称: AI-R-IAM)是为大规模人工智能模型(如

AI 就绪的人模型身

从身份、访问、行为、数据等领域为人模型在训练、部署和运行阶段多个场景中

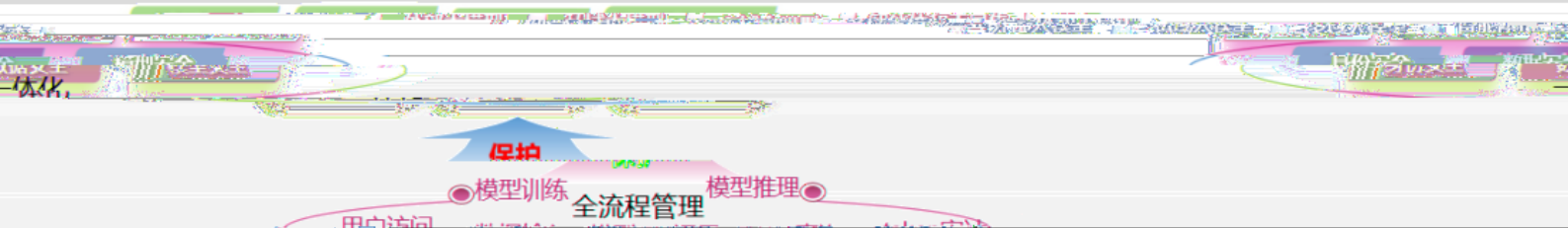
提供统一

一体化全程可信能力, 构建大模型安全从“单点可控”迈向“一体化全程可信”保护体

构建一体

为其它安全能力提供身份管理属性支撑。

系, 同时



机制，确保只有经过授权的用户和系统能够访问大模型。系统不仅支持传统验证，多因素认证，还引入基于中国移动号卡特性的实名认证能力，有效防止未经授权的操作和潜在的安全威胁。

先进的身份认证
的用户名和密码
防止未经授权的操作

● 可信访问权限管理

化的权限

AI-R-IAM 针对多个大模型访问、AI 接口访问、RAG 访问提供了集中的精细

限管理不仅提高了系统的安全性，还增强了用户的操作体验。

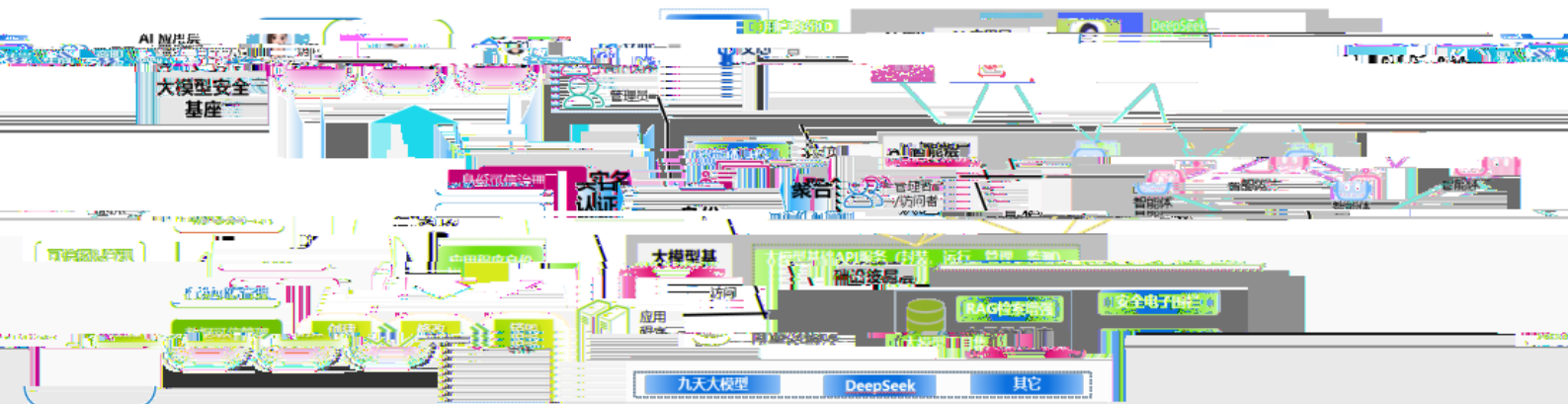
● 可信审计管理

AI-R-IAM 对大模型使用过程和大模型内部组件业务调用过程进行安全审计和实时监控，记录用户的所有操作行为，包括用户登录时间、IP 地址、访问的大模型资源、执行的操作等信息，实时监测用户和组件的访问行为，及时发现异常风险。

AI-R-IAM 通过构建一体化的全程可信能力，能够有效应对当前大模型面临的安全挑战，

人工智能技术的安全发展奠定了坚实的基础。

4.1 可信身份治理



AI-R-IAM 支撑基于大模型特性的多维度身份和属性的治理和管理，AI-R-IAM 的“身份 ID”超越了传统身份概念的范围，聚合人类身份、AI Identity 以及非人类身份（NLI）

集中一体的身份标识，对身份和属性进行全生命周期管理，针对用户、智能体、API 应用程序建立统一的管理，是数字世界秩序的底层支撑。

机制，同时引入中国移动基于号卡特性的实名认证能力，AI-R-IAM 采用了先进的身份认证技术，有效防止未经授权访问和潜在的安全威胁。

不仅涵盖传统人类身份从入职到离职的全生命周期管理，还创新性地将 NLI 纳入统一治理范畴，助力企业实现数智化转型，有效降低安全风险，提升合规性。在具体流程如下：

身份在入职、转岗、离职时，进行信息收集验证、权限调整及账号注销等操作：

1、智能体身份在创建时生成唯一 ID 并设定权限，修改时变更信息，销毁时注销身份；

2、回收资源并处理数据；

3、应用程序身份在创建时注册认证，授予权限，修改时更新信息，变更权限，销毁时注销身份；

删除身份、注销身份并清理数据。

即可提供友好的身份识别和访问控制。随着数据治理的深入，信息治理和档案管理中用户身份识别的引入，则可以打通数据、信息、知识、档案的治理流程，实现跨系统、跨数据、跨知识、跨档案的治理。图 4-2-1 展示了数据治理、信息治理、知识治理和档案治理的治理流程。

权限管理

4.2 可信数据空间

数据治理、信息治理、知识治理和档案治理的治理流程，如图 4-2-1 所示。

风险。这种数据治理和权限管理不仅提高了系统的安全性，还提升了用户的操作体验。

（一）用户权限管理

在数据治理、信息治理、知识治理和档案治理的治理流程中，权限管理是一个关键环节。权限管理是指对系统资源的使用权限进行分配和管理，以确保系统的安全性和数据的完整性。

在数据治理、信息治理、知识治理和档案治理的治理流程中，权限管理是一个关键环节。权限管理是指对系统资源的使用权限进行分配和管理，以确保系统的安全性和数据的完整性。

（二）用户角色权限管理

在数据治理、信息治理、知识治理和档案治理的治理流程中，权限管理是一个关键环节。权限管理是指对系统资源的使用权限进行分配和管理，以确保系统的安全性和数据的完整性。

在数据治理、信息治理、知识治理和档案治理的治理流程中，权限管理是一个关键环节。权限管理是指对系统资源的使用权限进行分配和管理，以确保系统的安全性和数据的完整性。

在数据治理、信息治理、知识治理和档案治理的治理流程中，权限管理是一个关键环节。权限管理是指对系统资源的使用权限进行分配和管理，以确保系统的安全性和数据的完整性。

算法工程师	✓	✓	✓	×	×	×
-------	---	---	---	---	---	---

在数据治理、信息治理、知识治理和档案治理的治理流程中，权限管理是一个关键环节。权限管理是指对系统资源的使用权限进行分配和管理，以确保系统的安全性和数据的完整性。

业务用户	×	受限调用	×	×	×	×
------	---	------	---	---	---	---

（三）AI 接口权限管理

包括：进口接口策略、接口鉴权、接口防垃圾、接口鉴权、接口信息数据保护、接口鉴权

设备管理：设备管理、设备管理、设备管理、设备管理、设备管理、设备管理、设备管理、设备管理

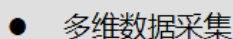
AC（检测增强生成）：检测增强生成、检测增强生成、检测增强生成、检测增强生成、检测增强生成、检测增强生成、检测增强生成、检测增强生成

权限控制：不同部门只能访问其业务相关的知识库；文档

内容管理：内容管理、内容管理、内容管理、内容管理、内容管理、内容管理、内容管理、内容管理

结果（如涉及竞争对手的信息）实时内容过滤：结合语义分析引擎，自动屏蔽无权限的检索结果





● 安全元数据管理

分發、檢閱、修繕、保管

◎ 社會文化批判

○ 表紙のデザイン

期而後 新期而後 并期而後 期而後 知此早見而後

- ## 通过数据多源融合

关系构建, 提供从 “关

- ## 十、缓刑、假释、减刑、赦免

景分析等, 对大模型

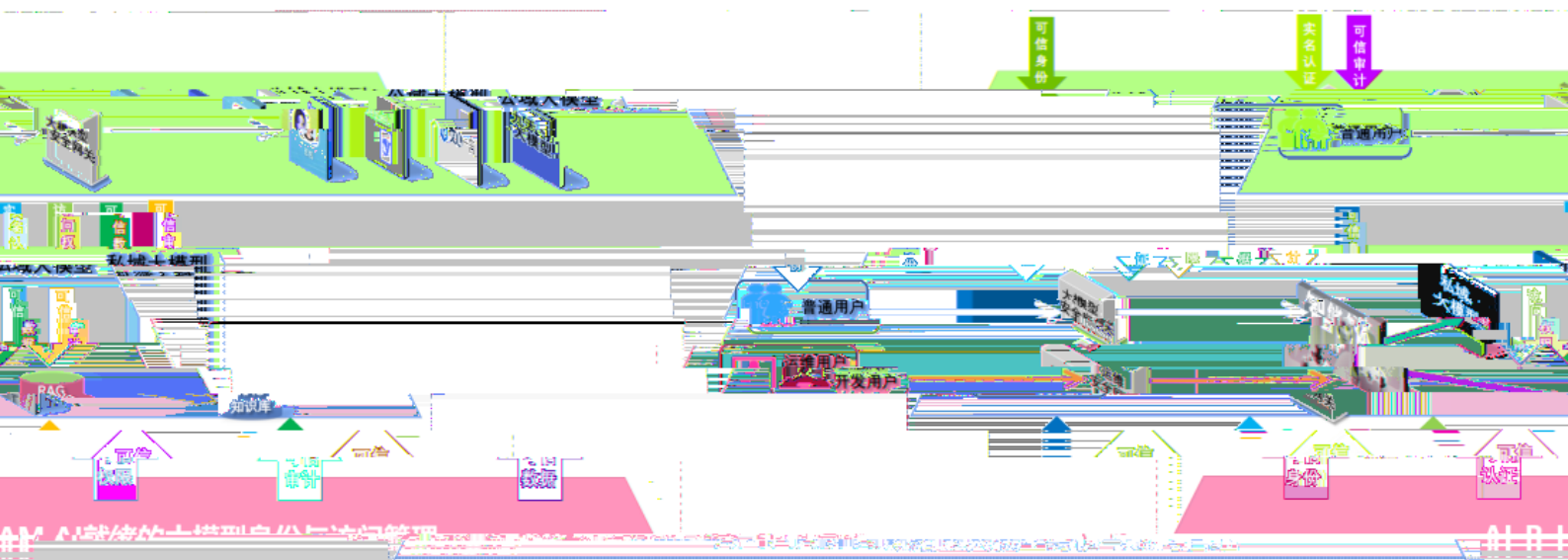
- 智能追踪

- 世界全史网提供全球知识资源, 采用国际统一标准, 保证信息准确。

回聲山 > 山脈分類 >



户访问公域大模型、用户访问私域大模型、



1. 普通用户访问公域大模型

普通用户：企业员工还是个人用户，都可以借助自己的终端设备，像 PC、手机

覆盖多种类型的服

或者平板等，去访问那些部署在公有云上的大模型应用。这些大模型应用涵

盖多种类型的服

的物体，提取出人物特征等信息，语音转文字服

内容：图像识别服务可以精准地识别图片中

务则可以利用占目制的语音文件准确识别语音转成对应的文字内容

的标签服务用户通过标签，新开账户而无需注册账号

大模型服务，从而消除服务使用

要利用文本生成服务撰写一份产品推广文案时，只需打

非常简便。比如，当一个企业员工想要

企业宣传语的生成服务，只需输入一些关键词，即可生成

工模型服务，从而消除服务使用

务提供。又如个人用户使用图像识别服务识别一张拍

成识别，很快就能识别出一份企业宣传文案

的操作，就能从 Web 界面上获取到详细的识别分

物识别服务时，也能通过简单的操作

类结果。

用户数据安全和服务质量至关重要。AI 与 IAM

然而，在这种机制的辅助下，保障

提供可信身份认证及安全审计的能力。

在普通用户和公域大模型场景中，主要

同大模型应用时，

可信身份认证方面，构建了一套严谨的身份验证机制。当用户首次访问

在身份验证方面，

系统会要求用户进行身份验证，这可能包括用户名和密码组合、生物特征、

身份验证的方式，只有经过严格的验证流程，系统才会授予访问权限。此外，系统还支持可验证凭证（Verifiable Credential, VC）的

使用，用户可以在需要时出示自己的数字身份凭证，系统会自动验证其有效性，从而简化了身份验证流程，提高了用户体验。

保护用户的数据不被未授权访问。

在安全审计方面，AI-R-IAM 持续监控整个系统的运行状况。它会记录下所有用户的操

作日志，包括登录时间、访问的资源、操作类型等。通过大数据分析，系统能够识别出异常行为模式，如频繁的登录失败、异常的访问时间等，并及时发出警报。

此外，系统还支持定期的安全审计，管理员可以查看审计日志，了解系统的安全状况，并根据需要进行相应的配置调整。

当系统检测到异常行为时，会立即触发警报提示管理员进行进一步调查，从而

了大量不符合常规模式的数据访问请求，就可以及时

有效地维护系统的稳定性和安全性。

2. 普通用户访问私域大模型

在企业数字化转型过程中，大模型应用已成为提升效率、优化决策的重要工具。然而，大模型应用的安全性和可控性是企业在部署时必须考虑的关键因素。私域大模型应用是指部署在企业内部、仅面向特定用户群体提供服务的模型应用。普通用户访问私域大模型应用时，需要遵循严格的安全规范，以确保数据的安全性和模型的可靠性。

普通用户通过终端设备访问部署在企业内部的私域大模型应用（例如知识问答、文档生成等

普通

场景），与公共大模型相比，私域大模型更注重数据的安全性和可控性，确保企业核心数据不被泄露，同时保证模型的输出符合企业的合规要求。

在部署私域大模型应用时，企业需要建立完善的安全体系，包括身份认证、访问控制、数据加密、审计日志等，以保障模型应用的安全运行。

在用户身份可信认证基础上，私域场景下需要对不同角色进行更加细致的权限划分。企

业内部不同角色的用户，其访问大模型应用的需求和权限各不相同。例如，高层管理人员可能需要访问所有数据，而普通员工则只能访问与其工作相关的数据。

部门主管则只需要关注本部门的数据，

了解企业的运营状况；包括各个部门的数据汇总；部门

普通员工的权限最为有限，只能访问与

并且能够对本部门员工的权限进行一定程度的调整；普通

员工只能访问与自己工作直接相关的数据，无法访问其他部门的数据。通过精细化的权限管理，企业可以确保数据的安全性和模型的可靠性。

性和准确性，以满足不同角色的需求同时保障数据安全。

3. 运营维护人员访问私域大模型

运营维护人员访问私域大模型应用服务器，负责日常维护和故障排查。由于运维人员权

限与业务人员不同，在受到边界管控管理人员授权时，运维人员只能访问特定资源，

上，还想要从数据权限和权限管理、日志操作、全局管理等方面能力。

在安全运营平台中，日志操作和权限管理是核心功能。

在安全运营平台中，日志操作和权限管理是核心功能。

在安全运营平台中，日志操作和权限管理是核心功能。

在安全运营平台中，日志操作和权限管理是核心功能。

应该只关注核心组件和关键组件的日志，而不能做到所有组件的日志。

数据权限管理是核心功能，日志操作和权限管理是核心功能。

在安全运营平台中，日志操作和权限管理是核心功能。

在安全运营平台中，日志操作和权限管理是核心功能。

日志操作和权限管理是核心功能。

日志操作和权限管理是核心功能。

在安全运营平台中，日志操作和权限管理是核心功能。

日志操作和权限管理是核心功能。

日志操作和权限管理是核心功能。

日志操作和权限管理是核心功能。

在安全运营平台中，日志操作和权限管理是核心功能。

在安全运营平台中，日志操作和权限管理是核心功能。

在安全运营平台中，日志操作和权限管理是核心功能。

在安全运营平台中，日志操作和权限管理是核心功能。

在安全运营平台中，日志操作和权限管理是核心功能。

在安全运营平台中，日志操作和权限管理是核心功能。

在安全运营平台中，日志操作和权限管理是核心功能。

在安全运营平台中，日志操作和权限管理是核心功能。

在安全运营平台中，日志操作和权限管理是核心功能。

在安全运营平台中，日志操作和权限管理是核心功能。

下段, 单击 **新建设备** 按钮, 新建设备。

该设备人员信息来自阿里云账号, 设备名称为设备名称, 设备 ID 为设备 ID。

同时该设备人员信息来自阿里云账号, 设备名称为设备名称, 设备 ID 为设备 ID。

单击 **新建设备** 按钮, 新建设备。

单击 **新建设备** 按钮, 新建设备。

单击 **新建设备** 按钮, 新建设备。

单击 **新建设备** 按钮, 新建设备。

单击 **新建设备** 按钮, 新建设备。

单击 **新建设备** 按钮, 新建设备。

单击 **新建设备** 按钮, 新建设备。

单击 **新建设备** 按钮, 新建设备。

单击 **新建设备** 按钮, 新建设备。

单击 **新建设备** 按钮, 新建设备。

单击 **新建设备** 按钮, 新建设备。

单击 **新建设备** 按钮, 新建设备。

单击 **新建设备** 按钮, 新建设备。

单击 **新建设备** 按钮, 新建设备。

单击 **新建设备** 按钮, 新建设备。

单击 **新建设备** 按钮, 新建设备。

单击 **新建设备** 按钮, 新建设备。

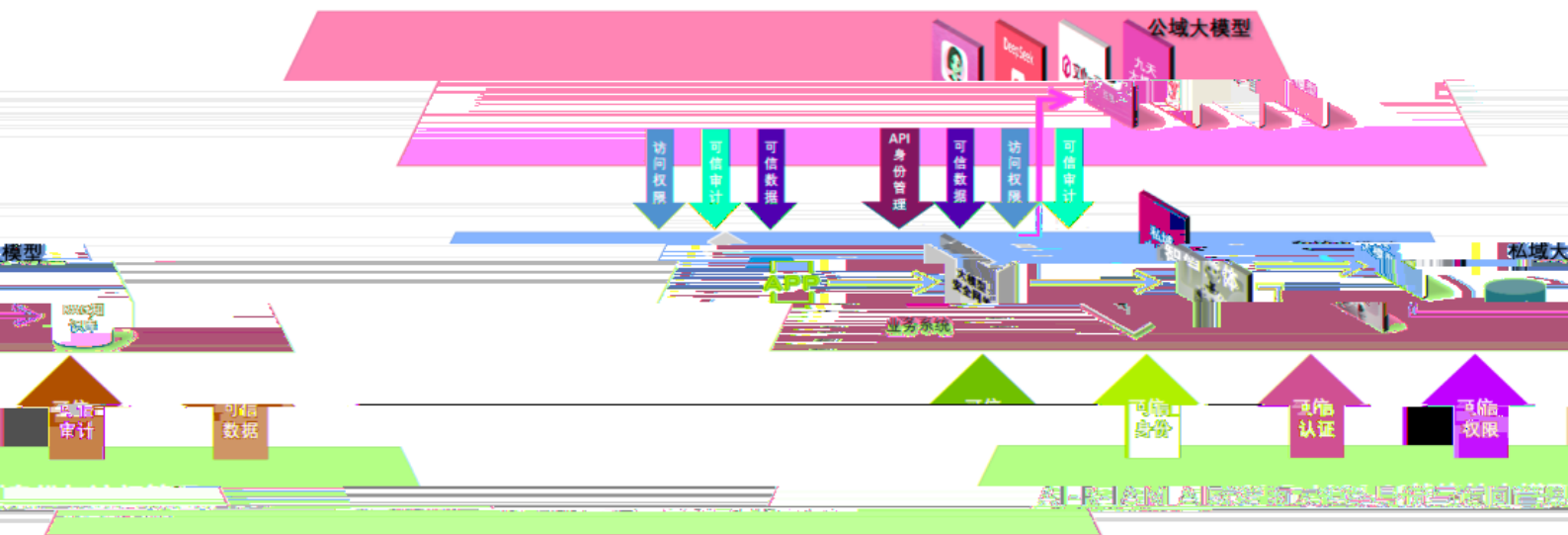
单击 **新建设备** 按钮, 新建设备。

单击 **新建设备** 按钮, 新建设备。

单击 **新建设备** 按钮, 新建设备。

单击 **新建设备** 按钮, 新建设备。

5.3 场景二：应用访问大模型应用场景



应用访问大模型场景主要包含业务系统访问公域大模型、业务系统访问私域大模型两种。

。AI-R-IAM 应用场景，业务系统能够通过安全智能体访问公域、私域大模型、RAG 知识库。对业务访问提供以下安全措施：

1 业务身份管控

业务系统身份管控是指对业务系统身份信息进行统一管理，确保身份信息的准确性和安全性。

硬编码在代码中，或者存储在不够安全的地方，攻击者可能通过伪造身份来访。如密钥可能被窃取或篡改，导致业务系统被非法访问。此外，身份信息可能被滥用，造成资源浪费或进行恶意操作。

AI-R-IAM 从业务身份创建、修改和销毁的全流程管控，实现业务身份可信。

身份创建：业务系统身份创建与认证。通过身份认证或身份识别进行身份创建，确保身份信息的准确性和安全性。

最小访问权限。

身份修改：更新业务系统身份信息和权限，确保合规并满足业务需求。身份修改：需根据功能和任务变更身份。

遵循最小权限原则。

身份销毁：删除业务系统身份信息并销毁凭证，同时进行关联资源清理和安全检查，确

保系统的安全稳定运行。

2. 业务权限管控

业务系统在访问公域、私域大模型时，如果没有基于最小权限原则来分配访问权限，如某些功能可能只需要读取权限，却被赋予了写入或管理权限。由于权限划分不清晰，可能导致内部人员误操作或恶意操作，进而影响系统整体安全。

业务系统应基于最小权限原则，对大模型访问权限进行精细化管控，确保权限分配的合理分配与安全性，降低安全风险，实现大模型业务权限的精准控制。

业务系统应基于最小权限原则，对大模型访问权限进行精细化管控，确保权限分配的合理分配与安全性，降低安全风险，实现大模型业务权限的精准控制。

业务系统应基于最小权限原则，对大模型访问权限进行精细化管控，确保权限分配的合理分配与安全性，降低安全风险，实现大模型业务权限的精准控制。

业务系统应基于最小权限原则，对大模型访问权限进行精细化管控，确保权限分配的合理分配与安全性，降低安全风险，实现大模型业务权限的精准控制。

业务系统应基于最小权限原则，对大模型访问权限进行精细化管控，确保权限分配的合理分配与安全性，降低安全风险，实现大模型业务权限的精准控制。

业务系统应基于最小权限原则，对大模型访问权限进行精细化管控，确保权限分配的合理分配与安全性，降低安全风险，实现大模型业务权限的精准控制。

业务系统应基于最小权限原则，对大模型访问权限进行精细化管控，确保权限分配的合理分配与安全性，降低安全风险，实现大模型业务权限的精准控制。

业务系统应基于最小权限原则，对大模型访问权限进行精细化管控，确保权限分配的合理分配与安全性，降低安全风险，实现大模型业务权限的精准控制。

业务系统应基于最小权限原则，对大模型访问权限进行精细化管控，确保权限分配的合理分配与安全性，降低安全风险，实现大模型业务权限的精准控制。

3. 业务行为审计

业务系统应基于最小权限原则，对大模型访问权限进行精细化管控，确保权限分配的合理分配与安全性，降低安全风险，实现大模型业务权限的精准控制。

业务系统应基于最小权限原则，对大模型访问权限进行精细化管控，确保权限分配的合理分配与安全性，降低安全风险，实现大模型业务权限的精准控制。

业务系统应基于最小权限原则，对大模型访问权限进行精细化管控，确保权限分配的合理分配与安全性，降低安全风险，实现大模型业务权限的精准控制。

业务系统应基于最小权限原则，对大模型访问权限进行精细化管控，确保权限分配的合理分配与安全性，降低安全风险，实现大模型业务权限的精准控制。

业务系统应基于最小权限原则，对大模型访问权限进行精细化管控，确保权限分配的合理分配与安全性，降低安全风险，实现大模型业务权限的精准控制。

业务系统应基于最小权限原则，对大模型访问权限进行精细化管控，确保权限分配的合理分配与安全性，降低安全风险，实现大模型业务权限的精准控制。

业务系统应基于最小权限原则，对大模型访问权限进行精细化管控，确保权限分配的合理分配与安全性，降低安全风险，实现大模型业务权限的精准控制。

业务系统应基于最小权限原则，对大模型访问权限进行精细化管控，确保权限分配的合理分配与安全性，降低安全风险，实现大模型业务权限的精准控制。

业务系统应基于最小权限原则，对大模型访问权限进行精细化管控，确保权限分配的合理分配与安全性，降低安全风险，实现大模型业务权限的精准控制。

业务系统应基于最小权限原则，对大模型访问权限进行精细化管控，确保权限分配的合理分配与安全性，降低安全风险，实现大模型业务权限的精准控制。

- 28 -

(user) 与智能体身份 (AI

通过 AI-R-IAM 提供可信身份管理，赋予用户可信身份

识别。如果智能体身份做了封

Identity)，由统一身份 ID 进行绑定管理，并生成审计日志

谁让智能体进行操作，方便界定相应责任。

项操作，系统可追溯到是

被恶意篡改或仿冒，智能体会与特定的代码及模型紧密绑定。运用

为了有效防止智能体

签名或证书技术，明确证明某个智能体是由谁指定的模型和代码构建，并成功部署

数字签

信任下，当智能体接收到该智能体所发出的请求时，系统会对其进行身份验证，并生成审计

高度，从而确保系统运行的安

格验证，以此来精准确认该智能体的真实身份，同时评估其可信

全性和可靠性。

在访问智能体场景下，普通用户通过大模型安全网关与 AI-R-IAM 实现认证能力建立实

身份，系统会记录该用户身份，并由 AI-R-IAM 负责管理该身份，并生成审计日志

名

同时，认证流程不再依赖于普通用户，每一个智能体都将有独立身份进行认证控制，并

在认证流程中签署信任认证令牌，通过短期令牌 (Short Access Token) 或一次性密钥

(One-Time Key) 等方式，让授权更加安全性和可追溯性。当智能体调用 RAG 时，通过

(如运行环境指纹，地理位置等) 进行增强认证。

3. 关联审计与监管

通过身份绑定机制，对普通用户使用智能体行为，以及智能体自身的行为进行审计和监

管。当智能体发生越权操作或异常行为时，AI-R-IAM 系统有能力快速定位并冻结该代理

权限。监管部门可记录并追溯智能体的操作细节，例如记录了哪些数据，进行了哪些个

依据何种规则。

造成训练数据外泄

在企业内部，智能体易成为攻击目标，被攻击后可能执行恶意操作，

数据到外部。

新风险，如自动提交敏感数

提供用户和智能体，智能体与数据存储服务器，RAG 知识库，

在此场景下，AI-R-IAM

限边界，实行最小化策略，制定智能体对 RAG 的数据访问权限

内外部大模型之间的数据权

文检测等安全控制，可及时阻断和审计智能体异常行为等。避免

范围，辅以模型护栏、上下

模安全事故。

因智能体自主决策引发大规

6 发展趋势展望

网络安全态势感知与威胁情报

网络安全态势感知与威胁情报

网络安全态势感知与威胁情报

网络安全态势感知与威胁情报

网络安全态势感知与威胁情报

网络安全态势感知与威胁情报

网络安全态势感知与威胁情报

网络安全态势感知与威胁情报

网络安全态势感知与威胁情报

网络安全态势感知与威胁情报

异常行为。

网络安全态势感知与威胁情报

或管理系统进行结合，实现对网络攻击的主动防御。基于零信任架构，AI-R-IAM 将对所有

网络安全态势感知与威胁情报

网络安全态势感知与威胁情报

网络安全态势感知与威胁情报

网络安全态势感知与威胁情报

5G/6G 网

术，保障大模型在物联网场景下与海量设备交互时的数据安全和身份可信；借助 5G

网络安全态势感知与威胁情报