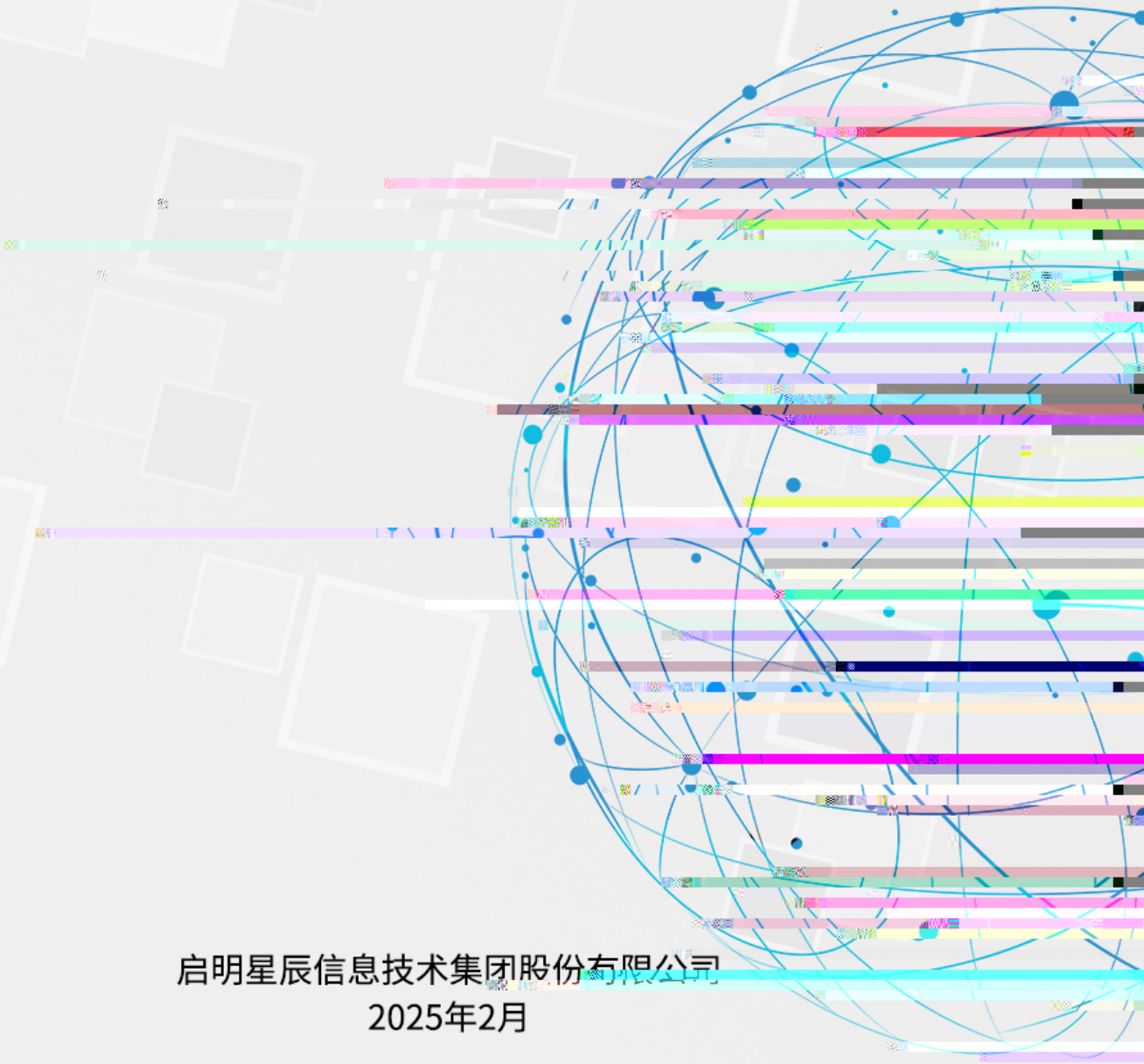
启明星辰

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

版权申明

北京启明星辰信息安全技术有限公司版权所有，并保留对本文档及本声明的最终解释权和修改权。

本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[illegible]

[illegible]，且著作权及其他相关权利归北京启明星辰信息安全技术有限公司所有。未经北京启明星辰信息安全技术有限公司事先书面许可，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、翻译成其他语言、将其全部或部分用于商业用途。

免责申明

本文档依据现有信息制作，其中资如有更改，恕不另行通知。

[illegible]

确可靠，但北京启明星辰信息安全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：[illegible]

100193 电话：010-82779088

传真：010-82779000

[illegible]

目录

1 公司简介 4

2.1 新质生产力与“大模型” 6

2.2 大模型应用中安全挑战 7

2.2.1 身份与访问控制的安全挑战 7

2.2.2 数据安全挑战 9

3 身份与访问管理系统（IAM） 11

4 AI 就绪的大模型身份与访问管理 13

4.2 可信权限管理 16

4.4 可信行为审计 19

5 应用场景 21

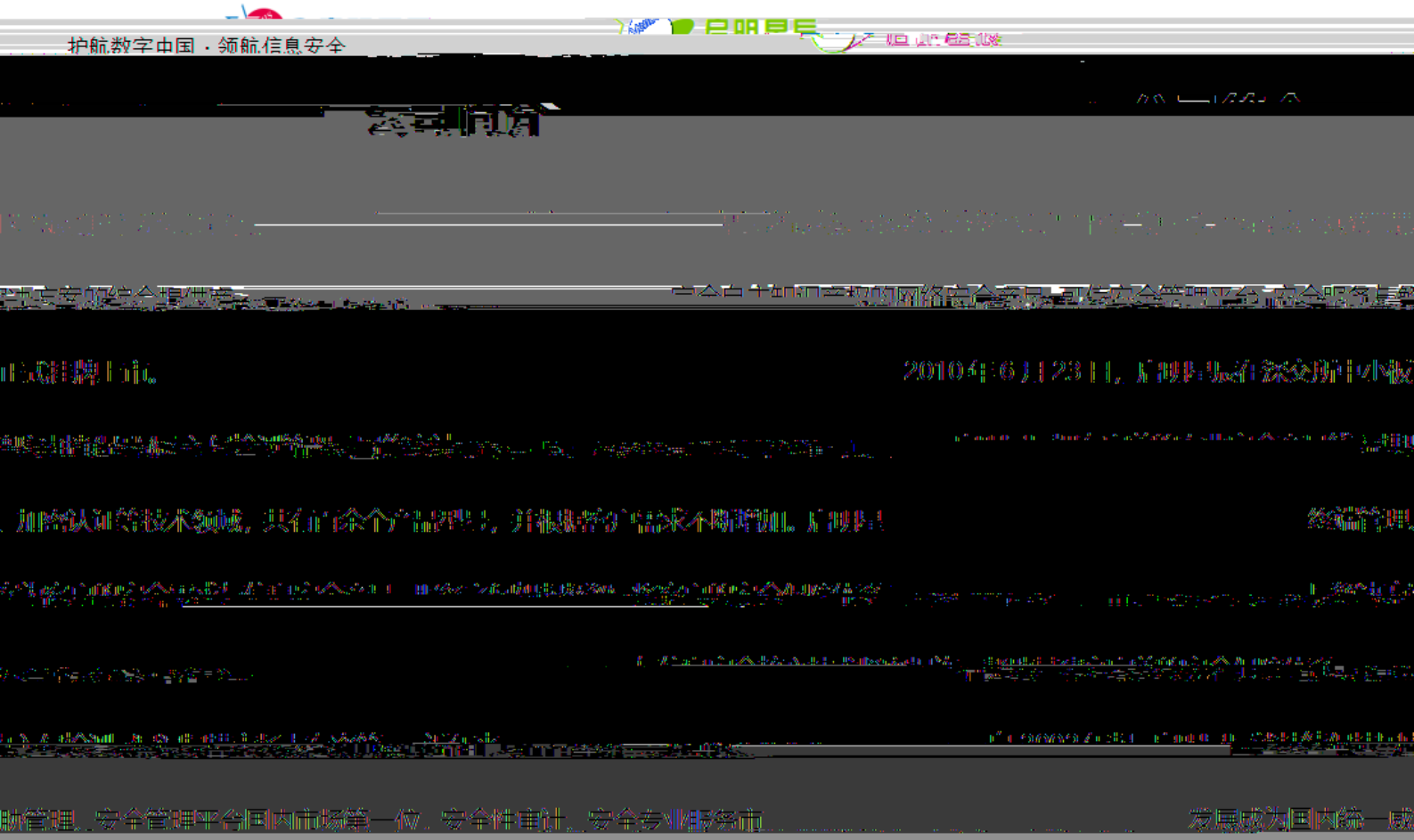
5.1 大模型应用场景数据流 21

5.2 场景一：用户访问大模型应用场景 21

5.3 场景二：应用访问大模型应用场景 26

5.4 场景三：用户-AI智能体的安全应用管理场景 31

6 发展趋势展望 33



场领导者。目前，公司在全国各省市自治区设立三十多家分支机构，拥有覆盖全国的渠道和售后服务体系。

长期以来，启明星辰公司得到了党和国家领导人的关怀与鼓励。2000 年 1 月，江泽民、李岚清、曾庆红等党和国家领导人亲切视察启明星辰公司；2003 年 1 月，胡锦涛总书记亲切接见了启明星辰公司 CEO 严望佳博士。

凭借多年来的潜心研发，启明星辰获得国家规划布局内重点软件企业，国家火炬计划软件产业优秀企业，中国电子政务 IT100 强等荣誉，及拥有最高级别的涉及国家秘密的计算机信息系统集成资质证书。

启明星辰目前是我国拥有最大的国家级网络安全研究基地，完成包括国家发改委产业化

、国家科技支撑计划等国家级科研项目近百项。创造了百

示范工程，国家科技部 863 计划

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

为世界五百强中 80%的中国企业客户提供安全产品及服

全产品和最佳实践服务，帮助客户全面提升其 IT 基础设施的安全性和生产效能，为打造和

2 大模型的发展与风险

2.1 新质生产力“大模型”

动力。从 ChatGPT，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核心驱动

型在自然语言处理、再到百度文心一言、阿里通义千问、九天大模型、DeepSeek 等，大模型

[illegible]

[illegible]

[illegible]

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，

2.2 大模型应用中安全挑战

Identity，NHI），指数据指纹令牌或

的背

1、人类身份安全问题：人类身份代表自然人身份实体，在大模型应用交互场景中

管理进一步出现滥用和泄露问题，大模型对用户数据使用不透明也可能泄露隐私

Agent 具

2、AI Identity 安全问题：越来越多的 AI Agent 在大模型场景下应用，AI

身份》中指出 AI Agent 具有复杂的身份关系，需要身份和专用身份管理，《人工智能代理的

AI Agent 应视为具有身份的不同员工的新兴范式，应重新思考 AI Agent 作为员工，在大

模型应用场景下建立身份层，即 AI Identity，它既有管理者又有访问者，自身具备身份属

非法访问权限。

则的方式请求访问权限，

(2) 访问权限放大：与人类身份不同，AI Agent 不会以可

有的权限，从而能够访问

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

或系统破坏。

和操作敏感资源，造成数据泄露

身份系统通常存储大量用户的个人信息、行为数据等，如果

(3) 自身安全：AI Agent

导致用户隐私泄露，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

定向广告骚扰等。

类身份定义为企业技术内的应用程序、服务或

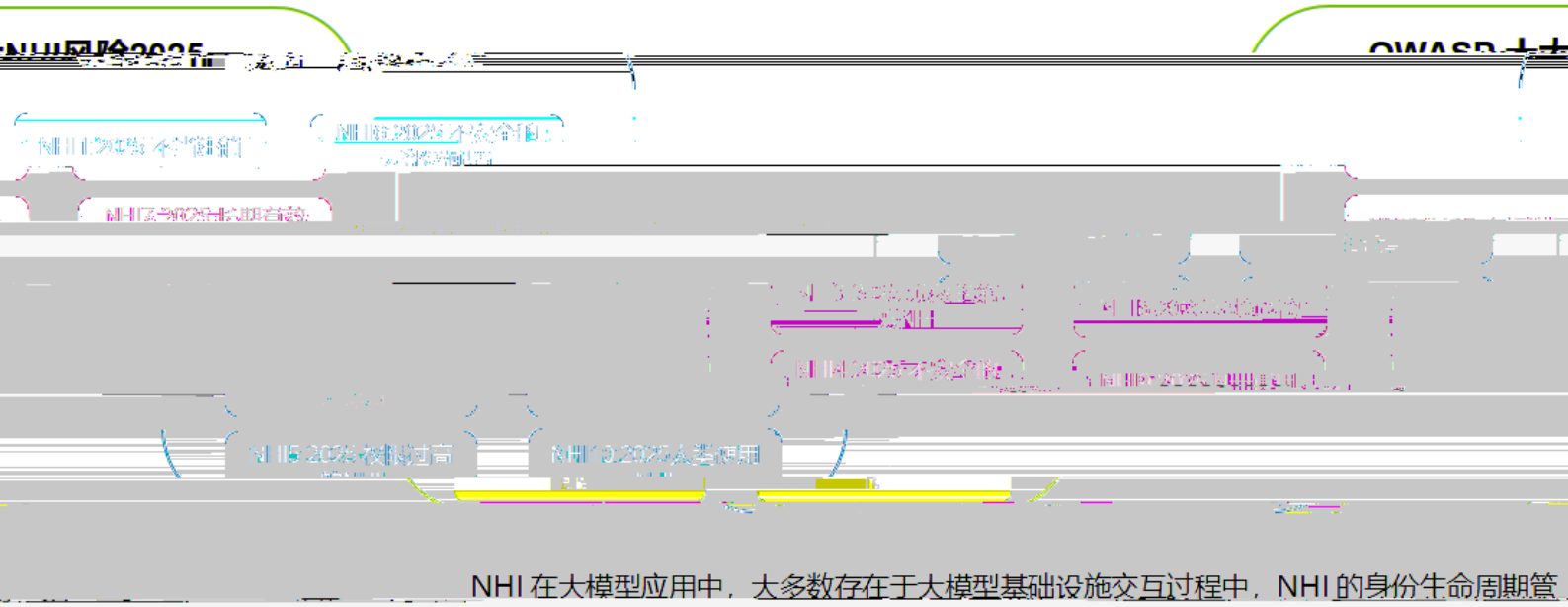
3、非人类身份（NHI）安全问题：非人类

、API 密钥、服务帐号、OAuth 令牌、云服

机器等实体绑定的数字身份，这些包括机器人

F》，如下图所示：

2025 年 2 月发布《十大 NHI 风险 202



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：

被内外部威胁者发现。

(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

NHI 账号不活跃，增加攻击面。

管理者对安全维护和问题补救很关键。

(4) 缺乏环境隔离：在许多情况下，相同 N

的逻辑 NHI 在每个环境中都有相同的密码，从

杂，难以掌握所有依赖关系。

2.2.2 数据安全挑战

大模型应用在数据安全方面呈现如下挑战：

(1) 用户提示词输入输出风险：用户提示词可能包含敏感信息，若大模型系统对输入提

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

的提示词用于医疗大模型进行诊断辅助，输出结果未做敏感处理，可能

调用风险：API 调用过程中，若身份认证和授权机制存在漏洞，不法分 (2) 大模型 API 调

调用 API，获取敏感数据或进行恶意操作。同时，API 接口若缺乏有效 子可能冒用合法身份

PI 调用权限， 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 A

非法获取金融大模型中的用户资产数据。

不当，用户可能 (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

(4) 数

(5) 合规风险：随着数字金融法规的不断完善，大模型的应用必须严格符合

地区、行业的数据处理、使用、存储和安全保护法律法规，如金融机构

化应用，未获得用户的授权，可能面临法律诉讼和监管处罚。

可信环境基于大数据安全能力，实现网络风险、行为风险、数据风险、业务风险、设备风险的全链路安全风险审计。

访问可信与数据可信作为一体两面，将防控范围扩展到安全设备、系统资源、应用资源、

更好的安全性，搭配一体化安全机制的融

组合性和可扩展性，使企业能够以更少的资源实现

产力，在政府、企业、个人中应用越来越深入，需要针对其身份、

随着大模型作为新质生

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全程

访问管理、数据安全等方面的

就绪的大模型身份与访问管理。

可信的数字身份基座，即 AI

4 AI 就绪的大模型身份与访问管理

AI 就绪的大模型身份与访问管理 (以下简称: AI-R-IAM) 是为大规模人工智能模型 (如

构建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”保护体系，同时为其它安全能力提供身份管理属性支撑。



先进的身份认证机制，确保只有经过授权的用户和系统能够访问大模型。系统不仅支持传统的用户名和密码验证，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访问和潜在的安全威胁。

- 可信访问权限管理

  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化的权限

理不仅提高了系统的安全性，还增强了用户的操作体验。

限管理

管理

- 可信审计

M 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

AI-R-IAM

中的所有操作行为，包括用户登录时间、IP 地址、访问的大模型资源、执行的

控，记录用户

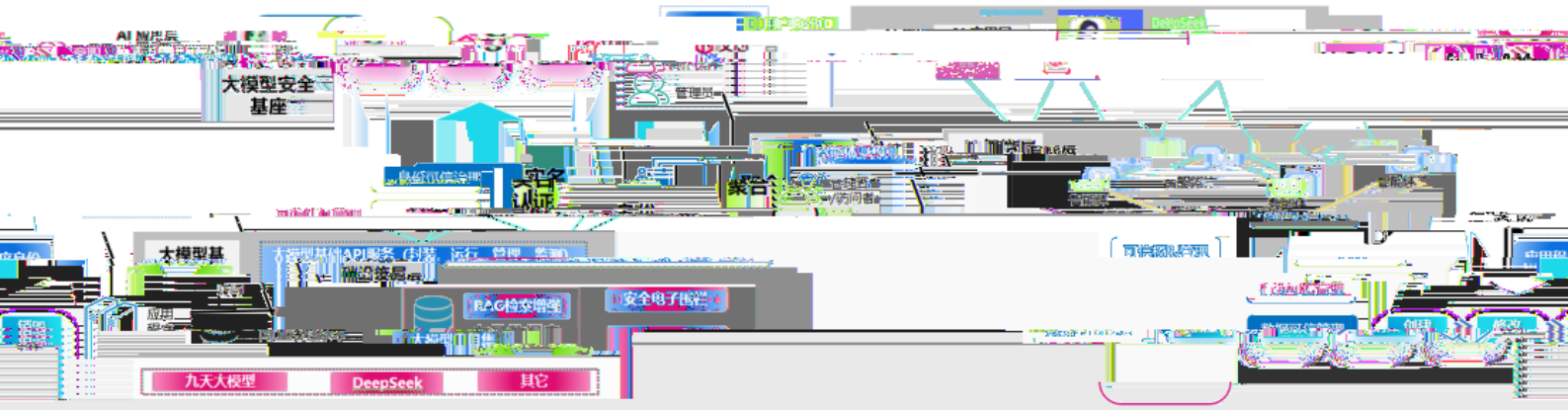
实时监测用户和组件的访问行为，及时发现异常风险。

操作等信息，

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，

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



多维度身份和属性的治理和管理，AI-R-IAM 的“身

AI-R-IAM 支撑基于大模型特性的多

全生命周期

针对用户、智能体、API 应用程序建立集中一体的身份标识，对身份和属性进行

的管理，是数字世界秩序的底层支撑。

实名认证能

AI-R-IAM 采用了先进的身份认证机制，同时引入中国移动基于号卡特性的实

力，有效防止未经授权的访问和潜在的安全威胁

提升合

和 AI Identity 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，

规性。在具体流程如下：

等操作；

(1) 用户身份在入职、转岗、离职时，进行信息收集验证、权限调整及账号注销等

据；

销毁时注销身份、回收资源并处理数

删除身份、销毁凭证并清理资源。

4.2 可信权限管理

型用户的角色和职责，动态调整其对

供了多场景、精细化权限管理功能，系统可以根据大模

风险。这种精细化的权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。

智能体对 RAG 的访问、用户对 RAG 的访问关联关系，建立大模型访问控制列表。

- 用户角色权限管理

角色级权限管理依托 RBAC 模型角色的访问控制技术，根据大模型不同的工作岗位

✓	✓	✓	×	×	×	算法工程师
用户	×	受限调用	×	×	×	业务用

口权限管理

- AI 接

包括上供型接口管理、接口密钥、接口冻结黑白、接口违离、接口传输数据内容、接口[illegible]

黑白名单等进行权限参数管理。

- RAG（检索增强生成）权限隔离

识库片段；

文档级权限控制：不同部门只能访问其业务相关的知识

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（如涉及竞对信息的段落）

4.4 可信行为审计



- 多维数据采集

构建全面的用户行为基线分析提供支撑

- 安全元数据管理

分发，实时数据管道

- 可信行为基线

- 数据画像分析

构建多维分析模型，反映大模型用户的特征、行为习惯、个人便好，生成用户画像、模型画像、数据画像等，并根据风险画像评分，智能量化风险水平。

知识图谱

- 数据知

多源融合、关系排序、深度去重等深度治理及分析，实现不同实体之间的关联

通过数据多

供从“关系”的角度分析大模型风险的能力。

关系构建，提供

深度分析

- 数据深

提全生命周期各个环节进行合规风险分析，重点数据审计、敏感数据和异常场

大模型数据

对大模型安全风险进行评估，识别已知风险和未知风险。

景分析等，

智能追踪溯源

- 智

分析，层层递进分析，实现风险智能追

可疑 IP->访问数据分布->可疑路径->可疑日志链路分

踪。

5 应用场景

5.1 大模型应用场景数据流

、使用、调用、训练等多 AI-R-IAM AI 就绪的大模型身份与访问管理，为大模型的访问

能力。相关场景包括：用 种场景提供可信身份、可信认证、可信权限、可信审计、可信数据

域大模型场景；业务系统调用智能体、RAG 知识库、私域大模型场 户访问私有大模型、公域

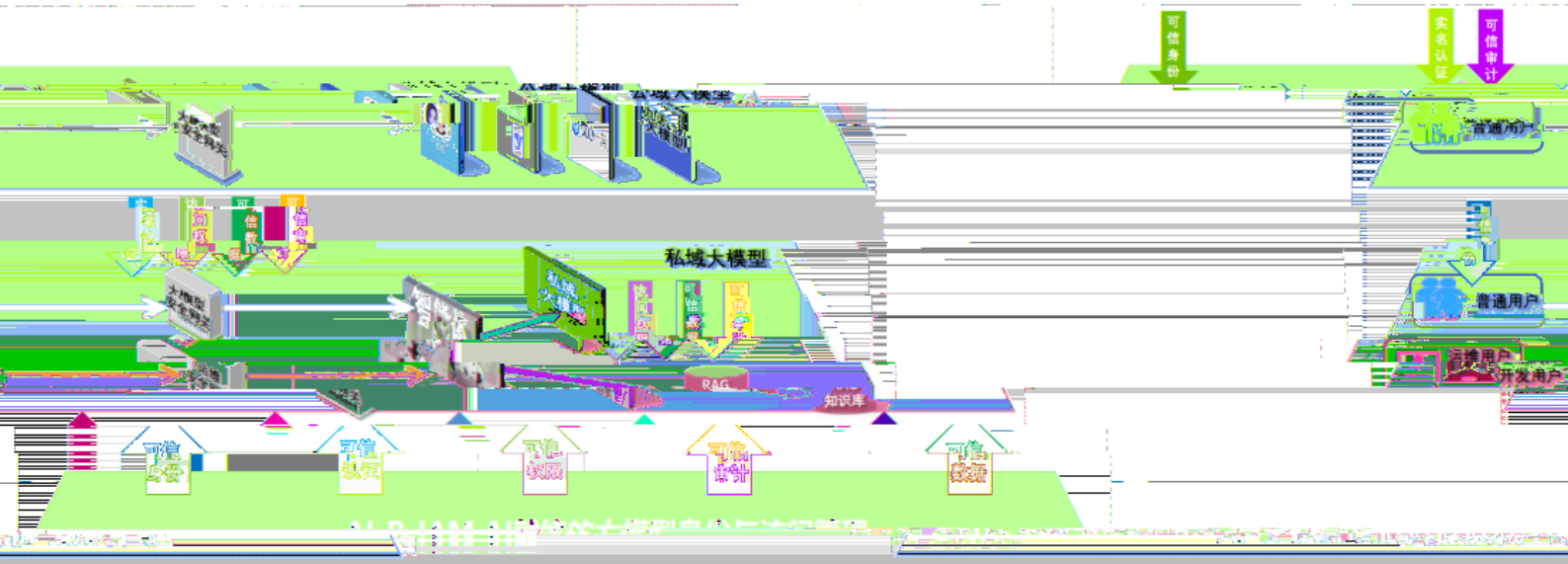
景、大模型数据运维场景等。 景；大模型数据训练场景



5.2 场景一：用户访问大模型应用场景

AI 就绪的大模型身份与访问管理能够为用户访问大模型应用场景提供可信 AI-R-IAM

私域大模型等。 户访问公域大模型、用户访问私域大模型、运营维护人员访问



1. 普通用户访问公域大模型

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

可信身份认证方面，构建了一套严谨的身份验证机制。当用户首次访问大模型应用时，

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

了大量不符合常规模式的数据访问请求，就可以

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

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

上，还需要重点加强精细化权限管理、高危操作、全库管理等方面能力。

查 网络连接状况

大模型应用服务器。然而，如果是在非工作

发更加严格的验证机制。

权限配置等。一旦识别出高危操作，就需要立即启动阻

层次。在第一层次是警告提示，当运维人员尝试执行高

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括业务应用访问公域大模型、业务应用访问私域大模型两种

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。AI-R-IAM 对业务访问提供以下安全措施：

1 业务身份管控

业务系统访问公域、私域大模型身份安全极为重要，如果业务系统的身份信息被窃

者可能通过伪造身份来访

如密钥可能被硬编码在代码中，或者存储在不够安全的地方，攻击者

问大模型服务，进而滥用资源或进行恶意操作

实现业务身份可信，

AI-R-IAM 从业务身份创建、修改和删除的全流程管控，

身份创建，对业务系统身份进行必要的认证，通过数字证书

以下权限。

系统身份下的权限，确保合规运营

身份修改：当业务系统身份发生变更时，比如业务

则。

能以下权限。

删除业务系统身份下的所有权限，同时将该业务系统

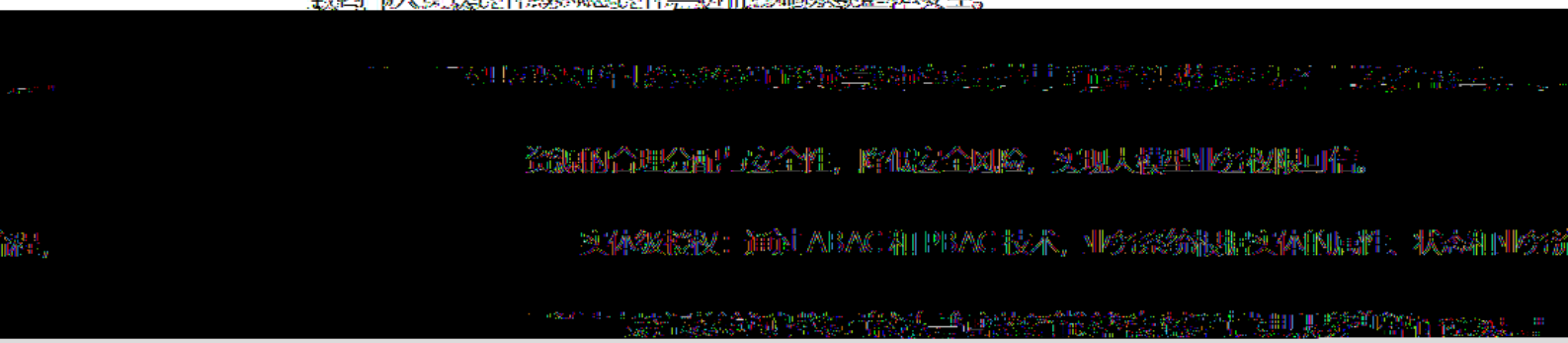
统身份下。

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。



数据级授权：结合 RBAC 和 ABAC 技术，系统依据业务系统的角色、属性及数据敏感

整业务系统的 Token 分

Token 动态授权：通过 RBAC、ABAC 和 PBAC 技术，动态调

使用。

配，实现对资源访问的弹性管控，确保资源得到高效合理的分配和

3. 业务行为审计

整、合规性检查不足、数据隐私泄漏等安全问题。

AI-R-IAM 在业务系统访问公域、私域大模型时，通过全链路日志采集、智能化分析



获有限权限，同时通过身份绑定对双方行为审计监管，保障访问数据和训练数据的安全。通过 AI-R-IAM 能力解决如下问题：

1. **用户到智能体全链路身份可信**

身份（AI

通过 AI-R-IAM 提供可信身份管理，赋予用户可信身份（user）与智能体

操作，方便界定相应责任。

项操作，系统可追溯到是谁让智能体进行操

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被恶意掉包或仿冒

数字签名或证书技术，明确证明某个智能体

统运行的安

格验证，以此来精准确认该智能体的真实身份，同时评估其可信度，从而确保系统

全性和可靠性。

(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 系统应有能力快速定位并冻结该代理

易成为攻击目标，被攻击后可能执行恶意操作，造成训练数据外泄 在企业内部，智能体

新风险，如自动提交敏感数据到外部。

在此场景下，AI-R-IAM 提供用户和智能体，智能体与数据存储服务器，RAG 知识库，内外部大模型之间的数据权限边界，实行最小化策略，制定智能体对 RAG 的数据访问权限范围，辅以模型护栏、上下文检测等安全控制，可及时阻断和审计智能体异常行为等。避免因智能体自主决策引发大规模安全事故。

6 发展趋势展望

异常行为。

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/6G 网