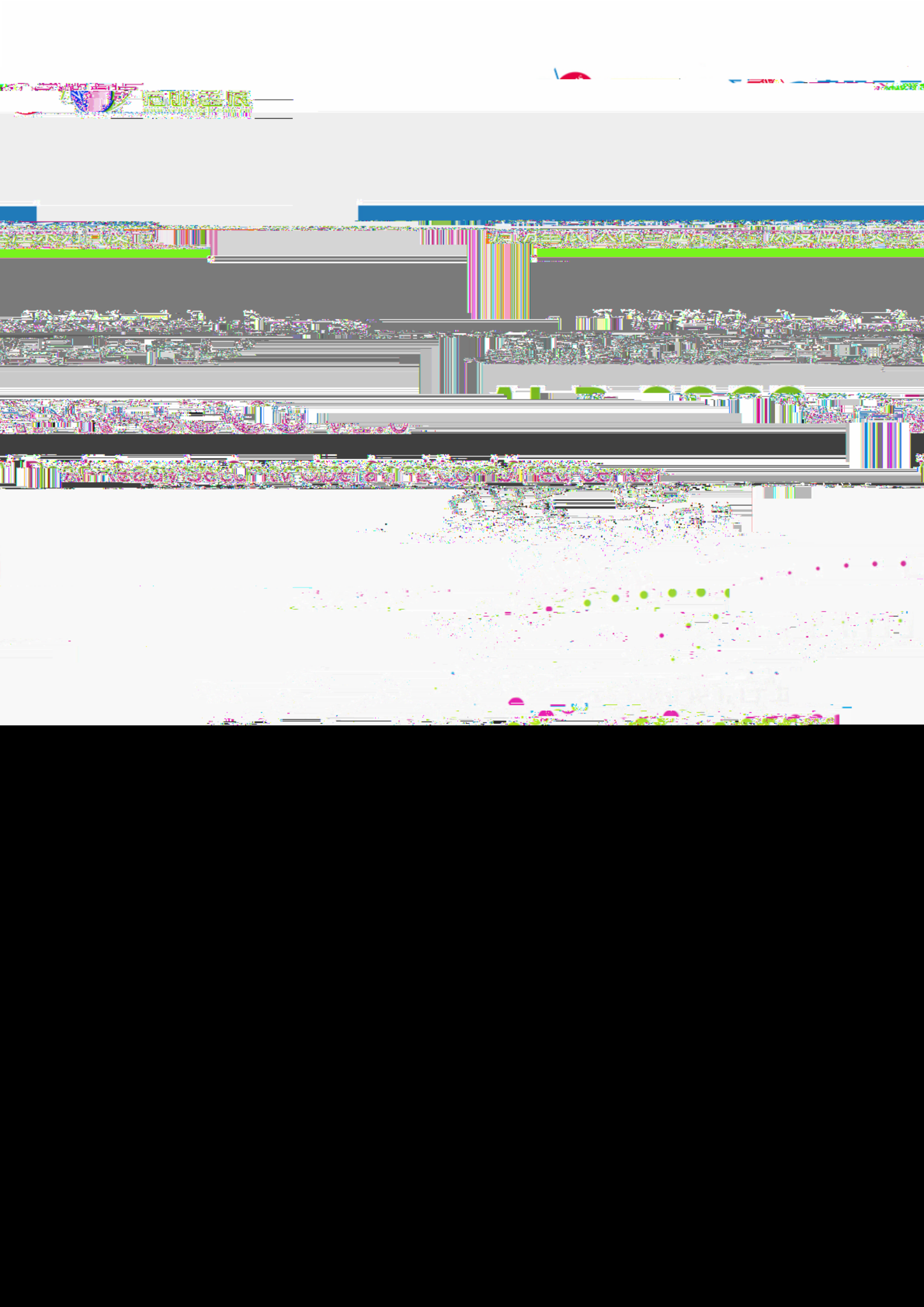




北京启明星辰信息安全技术有限公司版权所有，并保留对本文档及本声明的最终解释权和修

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

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安全技术有限公司。未经北京

启明星辰信息安全技术有限公司事先书面许可，任何人不得复制或向任何第三方提供或传播本文档的任何部分，否则将承担法律责任。

或部公用于商业用途。

部公进行复制、摘录、修改、传播、翻译成其它语言、将其全部

制作，其内容如有更改，恕不另行通知。

本文档依据现有信息制

免责声明

启明星辰信息安全技术有限公司在编写该文档的时候已尽力保证其准确性，但对其内容准

确性，但北京启明星辰信息安全技术有限公司

的准确性和及时性。

的准确性和及时性。

信息反馈

如有任何宝贵意见，请反馈：

8 号中关村软件园 21 号楼启明星辰大厦 邮编：100193

信箱：北京市海淀区东北旺西路

9088

100193 电话：010-82779000

传真：010-82779000

www.yanustech.com.cn 获得最新技术和产品信息

你可以访问启明星辰网站。

目录

1	公司简介.....	6
2	背景与挑战.....	8
2.1	背景分析.....	8
2.2	面临挑战.....	9

2.2.1	缺少大模型资产使用合规性持续监测评估.....	10
2.2.2	缺少大模型资产使用的合规性持续监测评估.....	10

3	定义和建设思路.....	13
---	--------------	----

3	定义和建设思路.....	13
---	--------------	----

3.1	AI-R-SOCC定义.....	13
-----	------------------	----

3.2	建设思路.....	14
-----	-----------	----

4	核心能力.....	16
---	-----------	----

4.1	运营能力.....	16
-----	-----------	----

4.1.1	能力市场.....	16
-------	-----------	----

大模型资产实体定义.....	19	4.2.1
----------------	----	-------

企业/单位白建士模型.....	20	4.2.2
-----------------	----	-------

4.2.3	内部私搭大模型.....	20
4.3	大模型安全分析.....	21
4.3.1	大模型风险关联分析.....	21
4.3.2	基于用户自认的行为分析.....	22
4.3.3	大模型安全图谱分析.....	23
4.4	智能告警降噪.....	24
4.4.1	安全事件.....	24
4.5	大模型自身风险监测.....	25
4.5.1	大模型自身风险监测.....	25
4.5.2	风险人员监测.....	26
4.5.3	风险行为/事件监测.....	27
4.5.4	智能治理建议生成.....	27
4.6	行为审计溯源.....	28
4.7	大模型安全态势呈现.....	28
5	部署和典型应用场景.....	30
5.1	场景一：.....	30
5.2	场景二：.....	30
5.3	场景三：.....	30
5.4	场景四：.....	30
5.5	场景五：.....	30
5.6	场景六：.....	30
5.7	场景七：.....	30
5.8	场景八：.....	30
5.9	场景九：.....	30
5.10	场景十：.....	30
5.11	场景十一：.....	30
5.12	场景十二：.....	30
5.13	场景十三：.....	30
5.14	场景十四：.....	30
5.15	场景十五：.....	30
5.16	场景十六：.....	30
5.17	场景十七：.....	30
5.18	场景十八：.....	30
5.19	场景十九：.....	30
5.20	场景二十：.....	30
5.21	场景二十一：.....	30
5.22	场景二十二：.....	30
5.23	场景二十三：.....	30
5.24	场景二十四：.....	30
5.25	场景二十五：.....	30
5.26	场景二十六：.....	30
5.27	场景二十七：.....	30
5.28	场景二十八：.....	30
5.29	场景二十九：.....	30
5.30	场景三十：.....	30
5.31	场景三十一：.....	30
5.32	场景三十二：.....	30
5.33	场景三十三：.....	30
5.34	场景三十四：.....	30
5.35	场景三十五：.....	30
5.36	场景三十六：.....	30
5.37	场景三十七：.....	30
5.38	场景三十八：.....	30
5.39	场景三十九：.....	30
5.40	场景四十：.....	30
5.41	场景四十一：.....	30
5.42	场景四十二：.....	30
5.43	场景四十三：.....	30
5.44	场景四十四：.....	30
5.45	场景四十五：.....	30
5.46	场景四十六：.....	30
5.47	场景四十七：.....	30
5.48	场景四十八：.....	30
5.49	场景四十九：.....	30
5.50	场景五十：.....	30
5.51	场景五十一：.....	30
5.52	场景五十二：.....	30
5.53	场景五十三：.....	30
5.54	场景五十四：.....	30
5.55	场景五十五：.....	30
5.56	场景五十六：.....	30
5.57	场景五十七：.....	30
5.58	场景五十八：.....	30
5.59	场景五十九：.....	30
5.60	场景六十：.....	30
5.61	场景六十一：.....	30
5.62	场景六十二：.....	30
5.63	场景六十三：.....	30
5.64	场景六十四：.....	30
5.65	场景六十五：.....	30
5.66	场景六十六：.....	30
5.67	场景六十七：.....	30
5.68	场景六十八：.....	30
5.69	场景六十九：.....	30
5.70	场景七十：.....	30
5.71	场景七十一：.....	30
5.72	场景七十二：.....	30
5.73	场景七十三：.....	30
5.74	场景七十四：.....	30
5.75	场景七十五：.....	30
5.76	场景七十六：.....	30
5.77	场景七十七：.....	30
5.78	场景七十八：.....	30
5.79	场景七十九：.....	30
5.80	场景八十：.....	30
5.81	场景八十一：.....	30
5.82	场景八十二：.....	30
5.83	场景八十三：.....	30
5.84	场景八十四：.....	30
5.85	场景八十五：.....	30
5.86	场景八十六：.....	30
5.87	场景八十七：.....	30
5.88	场景八十八：.....	30
5.89	场景八十九：.....	30
5.90	场景九十：.....	30
5.91	场景九十一：.....	30
5.92	场景九十二：.....	30
5.93	场景九十三：.....	30
5.94	场景九十四：.....	30
5.95	场景九十五：.....	30
5.96	场景九十六：.....	30
5.97	场景九十七：.....	30
5.98	场景九十八：.....	30
5.99	场景九十九：.....	30
6	附录.....	31
6.1	附录一：.....	31
6.2	附录二：.....	31
6.3	附录三：.....	31
6.4	附录四：.....	31
6.5	附录五：.....	31
6.6	附录六：.....	31
6.7	附录七：.....	31
6.8	附录八：.....	31
6.9	附录九：.....	31
6.10	附录十：.....	31
6.11	附录十一：.....	31
6.12	附录十二：.....	31
6.13	附录十三：.....	31
6.14	附录十四：.....	31
6.15	附录十五：.....	31
6.16	附录十六：.....	31
6.17	附录十七：.....	31
6.18	附录十八：.....	31
6.19	附录十九：.....	31
6.20	附录二十：.....	31
6.21	附录二十一：.....	31
6.22	附录二十二：.....	31
6.23	附录二十三：.....	31
6.24	附录二十四：.....	31
6.25	附录二十五：.....	31
6.26	附录二十六：.....	31
6.27	附录二十七：.....	31
6.28	附录二十八：.....	31
6.29	附录二十九：.....	31
6.30	附录三十：.....	31
6.31	附录三十一：.....	31
6.32	附录三十二：.....	31
6.33	附录三十三：.....	31
6.34	附录三十四：.....	31
6.35	附录三十五：.....	31
6.36	附录三十六：.....	31
6.37	附录三十七：.....	31
6.38	附录三十八：.....	31
6.39	附录三十九：.....	31
6.40	附录四十：.....	31
6.41	附录四十一：.....	31
6.42	附录四十二：.....	31
6.43	附录四十三：.....	31
6.44	附录四十四：.....	31
6.45	附录四十五：.....	31
6.46	附录四十六：.....	31
6.47	附录四十七：.....	31
6.48	附录四十八：.....	31
6.49	附录四十九：.....	31
6.50	附录五十：.....	31
6.51	附录五十一：.....	31
6.52	附录五十二：.....	31
6.53	附录五十三：.....	31
6.54	附录五十四：.....	31
6.55	附录五十五：.....	31
6.56	附录五十六：.....	31
6.57	附录五十七：.....	31
6.58	附录五十八：.....	31
6.59	附录五十九：.....	31
6.60	附录六十：.....	31
6.61	附录六十一：.....	31
6.62	附录六十二：.....	31
6.63	附录六十三：.....	31
6.64	附录六十四：.....	31
6.65	附录六十五：.....	31
6.66	附录六十六：.....	31
6.67	附录六十七：.....	31
6.68	附录六十八：.....	31
6.69	附录六十九：.....	31
6.70	附录七十：.....	31
6.71	附录七十一：.....	31
6.72	附录七十二：.....	31
6.73	附录七十三：.....	31
6.74	附录七十四：.....	31
6.75	附录七十五：.....	31
6.76	附录七十六：.....	31
6.77	附录七十七：.....	31
6.78	附录七十八：.....	31
6.79	附录七十九：.....	31
6.80	附录八十：.....	31
6.81	附录八十一：.....	31
6.82	附录八十二：.....	31
6.83	附录八十三：.....	31
6.84	附录八十四：.....	31
6.85	附录八十五：.....	31
6.86	附录八十六：.....	31
6.87	附录八十七：.....	31
6.88	附录八十八：.....	31
6.89	附录八十九：.....	31
6.90	附录九十：.....	31
6.91	附录九十一：.....	31
6.92	附录九十二：.....	31
6.93	附录九十三：.....	31
6.94	附录九十四：.....	31
6.95	附录九十五：.....	31
6.96	附录九十六：.....	31
6.97	附录九十七：.....	31
6.98	附录九十八：.....	31
6.99	附录九十九：.....	31
7	索引.....	32

5.3	场景三：运行期间的大模型自身安全性.....	33
5.4	场景四：影子大模型监测与治理.....	35

6 运营安全

6.1	智能运营	36
-----	------------	----

6.1

6.2 应急处置

6.2.1 应急处置流程

37

6.2.2 应急处置预案

6.4 安全专家

37

6.5 全天候值守

1 公司简介

最具实力的、拥有

启明星辰公司成立于 1996 年，由留美博士严望佳女士创建，是国内

度最高的网络安全

安全事件调查取证、恶意病毒木马查杀、网络钓鱼攻击、垃圾邮件

2010 年 6 月 23 日，启明星辰在深交所中小板正式挂牌上市。

易除病毒、网络审计、

启明星辰拥有自主研发的工业安全产品、防火墙、防病毒、入侵检测、

客户需求不断增加，启明星

终端管理、加密认证等技术领域，共有百余个产品型号，并根据客

需，将客户的安全保障体系

辰解决方案为客户的安全需求与信息安全产品、服务之间建立起紧密

的安全保障体系。

与信息安全核心技术紧密相连，帮助其建立完

、建设、运营和维护市场占有率第一，运行安

自 2002 年起，启明星辰持续保持国内

发展成为国内终端、威胁管理、安全管理综合国内市场第一位、安全性能、安全专业服务能力

场领导者。目前，公司在全国省市自治区设立三十多家分支机构，拥有覆盖全国的销售

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 11 月江泽民

启辰公司；2003 年 1 月，胡锦涛总书记接

李锐委员、曾庆红常委等中央领导人亲切视察启辰

切接见了启明星辰公司 CEO 严望佳女士。

启明星辰入选中国软件企业、上市公司社会责任

任等多项权威荣誉，启明星辰连续多年

计算

件产业优秀企业，中国电子政务 100 强第 26 名，及拥有产品涉密等级涉及国家秘密所

级信息安全等级保护系统

化。

启明星辰目前是国内规模最大的国家级网络安全研究中心，也是国内首家为公安

部 351 工程、公安部科技部和 311 工程、上海科技之都提供过国家级科研项目的企业，创造首工

类攻击，给受害者带来财产损失和隐私泄露等严重后果，填补了该领域安全防御领域的多项空白。

启明星辰安全产品与服务，在保障国家“两网”安全、保障国家重要基础设施安全、保障金融安全、保障能源安全、保障公共安全、保障国防安全等方面，发挥着重要作用。启明星辰安全产品与服务，在保障国家“两网”安全、保障国家重要基础设施安全、保障金融安全、保障能源安全、保障公共安全、保障国防安全等方面，发挥着重要作用。

启明星辰安全产品与服务，在保障国家“两网”安全、保障国家重要基础设施安全、保障金融安全、保障能源安全、保障公共安全、保障国防安全等方面，发挥着重要作用。启明星辰安全产品与服务，在保障国家“两网”安全、保障国家重要基础设施安全、保障金融安全、保障能源安全、保障公共安全、保障国防安全等方面，发挥着重要作用。

启明星辰安全产品与服务，在保障国家“两网”安全、保障国家重要基础设施安全、保障金融安全、保障能源安全、保障公共安全、保障国防安全等方面，发挥着重要作用。启明星辰安全产品与服务，在保障国家“两网”安全、保障国家重要基础设施安全、保障金融安全、保障能源安全、保障公共安全、保障国防安全等方面，发挥着重要作用。

启明星辰安全产品与服务，在保障国家“两网”安全、保障国家重要基础设施安全、保障金融安全、保障能源安全、保障公共安全、保障国防安全等方面，发挥着重要作用。启明星辰安全产品与服务，在保障国家“两网”安全、保障国家重要基础设施安全、保障金融安全、保障能源安全、保障公共安全、保障国防安全等方面，发挥着重要作用。

启明星辰安全产品与服务，在保障国家“两网”安全、保障国家重要基础设施安全、保障金融安全、保障能源安全、保障公共安全、保障国防安全等方面，发挥着重要作用。

启明星辰安全产品与服务，在保障国家“两网”安全、保障国家重要基础设施安全、保障金融安全、保障能源安全、保障公共安全、保障国防安全等方面，发挥着重要作用。启明星辰安全产品与服务，在保障国家“两网”安全、保障国家重要基础设施安全、保障金融安全、保障能源安全、保障公共安全、保障国防安全等方面，发挥着重要作用。

启明星辰安全产品与服务，在保障国家“两网”安全、保障国家重要基础设施安全、保障金融安全、保障能源安全、保障公共安全、保障国防安全等方面，发挥着重要作用。启明星辰安全产品与服务，在保障国家“两网”安全、保障国家重要基础设施安全、保障金融安全、保障能源安全、保障公共安全、保障国防安全等方面，发挥着重要作用。

启明星辰安全产品与服务，在保障国家“两网”安全、保障国家重要基础设施安全、保障金融安全、保障能源安全、保障公共安全、保障国防安全等方面，发挥着重要作用。启明星辰安全产品与服务，在保障国家“两网”安全、保障国家重要基础设施安全、保障金融安全、保障能源安全、保障公共安全、保障国防安全等方面，发挥着重要作用。

挑战：

斤

2 背景与概述

2.1 背景分析

随着人工智能技术的飞速发展，大模型（Large Language Models, LLMs）已成为当前科技领域的热点。LLMs在自然语言处理、机器翻译、代码生成等领域展现出卓越的能力，广泛应用于企业级应用、个人助理、内容创作等场景。然而，大模型在提供强大功能的同时，也面临着数据安全、隐私保护、模型偏见、幻觉生成等挑战。因此，深入分析大模型的安全风险，制定有效的防护措施，对于保障企业和个人利益至关重要。

本报告旨在系统性地分析大模型应用中的安全风险，包括数据泄露、模型攻击、内容生成风险、模型偏见与歧视、以及模型滥用等。报告将详细阐述这些风险的表现形式、潜在危害，并提出相应的安全策略和最佳实践，以帮助企业和个人有效识别、评估和缓解大模型带来的安全风险。

本报告的研究范围主要聚焦于大模型在生成式人工智能（AIGC）领域的应用，特别是文本生成、图像生成、视频生成等方面。报告将结合最新的行业趋势和研究成果，提供具有前瞻性和实用性的安全建议。

本报告的研究范围主要聚焦于大模型在生成式人工智能（AIGC）领域的应用，特别是文本生成、图像生成、视频生成等方面。报告将结合最新的行业趋势和研究成果，提供具有前瞻性和实用性的安全建议。

本报告的研究范围主要聚焦于大模型在生成式人工智能（AIGC）领域的应用，特别是文本生成、图像生成、视频生成等方面。报告将结合最新的行业趋势和研究成果，提供具有前瞻性和实用性的安全建议。

本报告的研究范围主要聚焦于大模型在生成式人工智能（AIGC）领域的应用，特别是文本生成、图像生成、视频生成等方面。报告将结合最新的行业趋势和研究成果，提供具有前瞻性和实用性的安全建议。

本报告的研究范围主要聚焦于大模型在生成式人工智能（AIGC）领域的应用，特别是文本生成、图像生成、视频生成等方面。报告将结合最新的行业趋势和研究成果，提供具有前瞻性和实用性的安全建议。

- **数据泄露：**训练和应用中，数据含大量敏感信息，一旦泄露，会侵犯个人隐私、损害企业竞争力和声誉，引发法律风险。
- **数据投毒：**攻击者向训练数据注入恶意样本，干扰正常训练，使模型性能下降、准确性降低。
- **模型窃取：**通过对输入数据微小扰动，让模型产生错误输出，在图像识别领域会导致分类错误。

本报告的研究范围主要聚焦于大模型在生成式人工智能（AIGC）领域的应用，特别是文本生成、图像生成、视频生成等方面。报告将结合最新的行业趋势和研究成果，提供具有前瞻性和实用性的安全建议。

本报告的研究范围主要聚焦于大模型在生成式人工智能（AIGC）领域的应用，特别是文本生成、图像生成、视频生成等方面。报告将结合最新的行业趋势和研究成果，提供具有前瞻性和实用性的安全建议。

本报告的研究范围主要聚焦于大模型在生成式人工智能（AIGC）领域的应用，特别是文本生成、图像生成、视频生成等方面。报告将结合最新的行业趋势和研究成果，提供具有前瞻性和实用性的安全建议。

本报告的研究范围主要聚焦于大模型在生成式人工智能（AIGC）领域的应用，特别是文本生成、图像生成、视频生成等方面。报告将结合最新的行业趋势和研究成果，提供具有前瞻性和实用性的安全建议。

本报告的研究范围主要聚焦于大模型在生成式人工智能（AIGC）领域的应用，特别是文本生成、图像生成、视频生成等方面。报告将结合最新的行业趋势和研究成果，提供具有前瞻性和实用性的安全建议。

放在企业全景视角的大模型安全治理面临三大核心矛盾：

○ **数据孤岛化**：各子系统独立运行，缺乏对大模型全生命周期（输入-推理-输出）的

端到端检测能力。

具体挑战：

○ **数据孤岛化**：安全日志分散存储，无法实现跨系统攻击链溯源。

○ **数据孤岛化**：AI模型数据存储在私有云，难以与外部安全平台联动。

环，实现大模型应用的全局可视、风险可管、

子系统，构建“监测-分析-处置-优化”智能闭

处置可溯。

2.2 面临挑战：

其次，随着大模型应用渗透率持续提升，安全风险随之增加。

型的应用给企业带来便捷的同时也带来了新的安全边界问题，例如大模型输入/输出

大模型

信息泄露、对大模型的注入攻击等。应对这些新问题企业会新增新型的大模型安全

出的敏感

多大模型的安全

防护手段，但这些手段基本针对某一单一的风险点，对于企业管理者面对众

上，企业面临三大核心挑战：

——数据孤岛化：安全日志分散存储，无法实现跨系统攻击链溯源。

1. 数据孤岛效应：

目前数据存储在多个AI模型集的私有云，难以与外部安全平台联动。

据报告，AI模型的训练数据往往存储在私有云中，难以与外部安全平台联动。

攻击链。

覆盖的私有API通道注入恶意指令时，MASB无法独立识别跨系统

攻击链，无法及时发现攻击行为。

攻击链，无法及时发现攻击行为。

“勿以一时便利，而致永久信任危机”的隐患也。）

2. 分析能力局限性：

模型输出结果往往呈现碎片化特征，难以关联分析跨系统、跨设备的异常行为。例如，当模型检测到某设备存在异常行为时，往往只能识别出该设备的异常行为，而无法追溯到该设备所连接的其他设备或系统，导致安全团队难以全面评估潜在威胁。

链)；

理：当模型输出泄露客户隐私时，现有工具无法追溯至具体训练数据

问行为。

无法保障：

不下：各子系统独立告警导致重复通知（如MAF和MAVAS同时报生同

均处理告警量超千条，有效告警识别率不足20%。

题通过“策略一致性与模型自断策略”解决。

分散的子系统监控难以捕捉此类长期潜伏威胁。

4. 性能与安全的平衡困境：

监控影响业务，MAVAS深度扫描导致模型推理延迟增加20%，企业被迫在安全性

与业务连续性间取舍；

不足40%，运维成本飙升。

● 资源浪费严重：多套子系统独立运行，硬件资源利用率不

模型资产使用的合规性持续监测评估

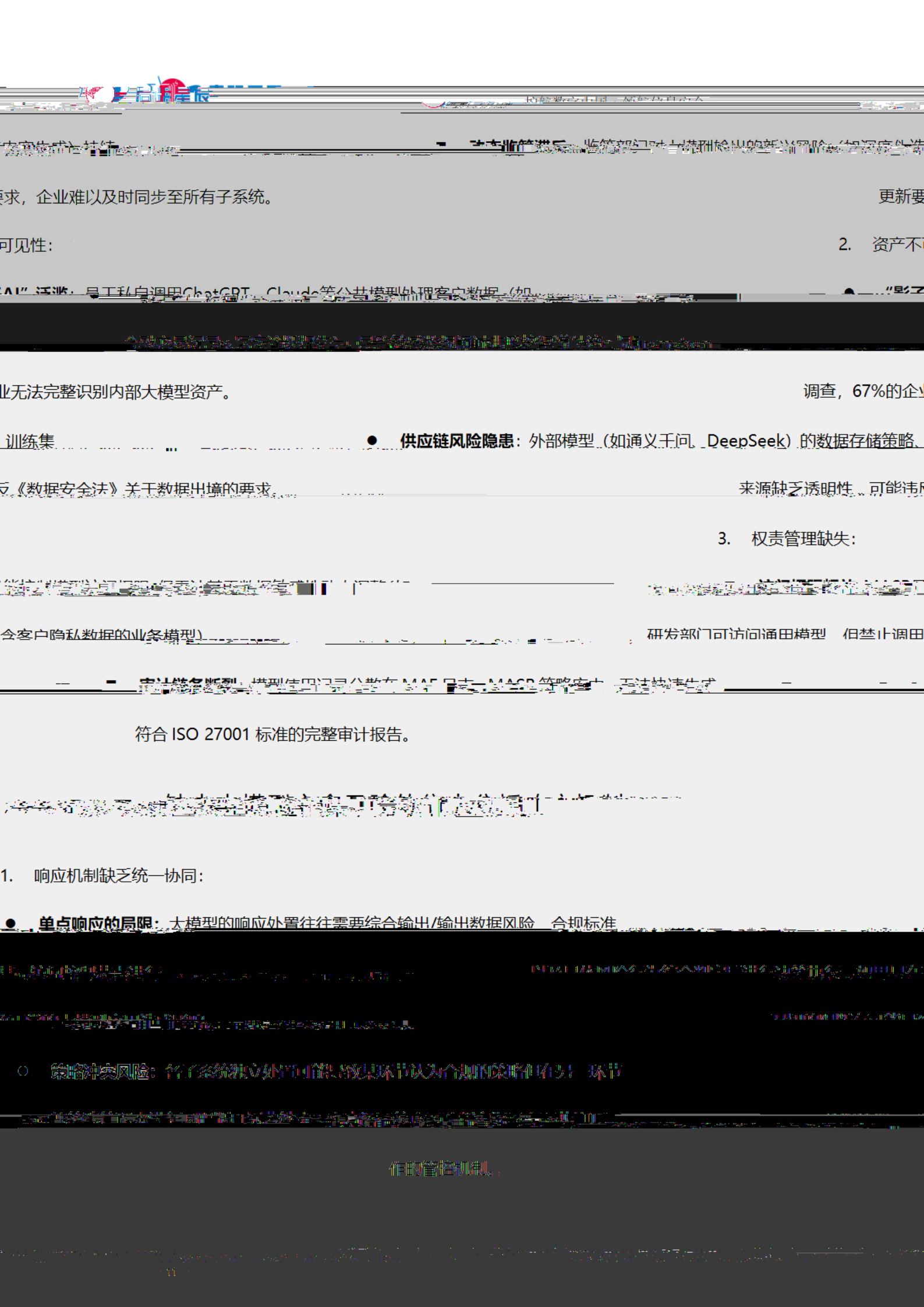
2.2.2 缺少大模

界造成不，而安全合规管理的结构性难题。

1. 大模型合规评估监管不足：

一、缺乏系统化的监管机制：企业主管部门要求AI模型安全评估，但因缺乏合

的各个环节，导致“带病运行”风险，加剧系统脆弱性，威胁系统安全。



内容生成) 持续

— 技术监管滞后：监管部门尚未适配由AI模型输出的新型网络攻击深度伪造

需求，企业难以及时同步至所有子系统。

更新要

可见性：

2. 资产不

“AI”泛滥：员工私自调用ChatGPT、Claude等公共模型处理客户数据（如

“显

业无法完整识别内部大模型资产。

调查，67%的企业

训练集

● 供应链风险隐患：外部模型（如通义千问、DeepSeek）的数据存储策略、

反《数据安全法》关于数据出境的要求。

来源缺乏透明性，可能违

3. 权责管理缺失：

数据“模型访问权限”与“数据使用策略”

数据“模型访问权限”与“数据使用策略”

合客户隐私数据的业务模型)

研发部门可访问通田模型，但禁止调田

符合 ISO 27001 标准的完整审计报告。

1. 响应机制缺乏统一协同：

● 单点响应的局限：大模型的响应外置往往需要综合输出/输出数据风险、合规标准

○ 策略冲突风险：各子系统独立处置逻辑可能导致各环节认为合规的策略存在“环节

作的管控机制。

2. 闭环治理缺失：

“当前网络安全态势感知能力有限，且态势感知数据未能有效利用”

“态势感知数据未能有效利用”

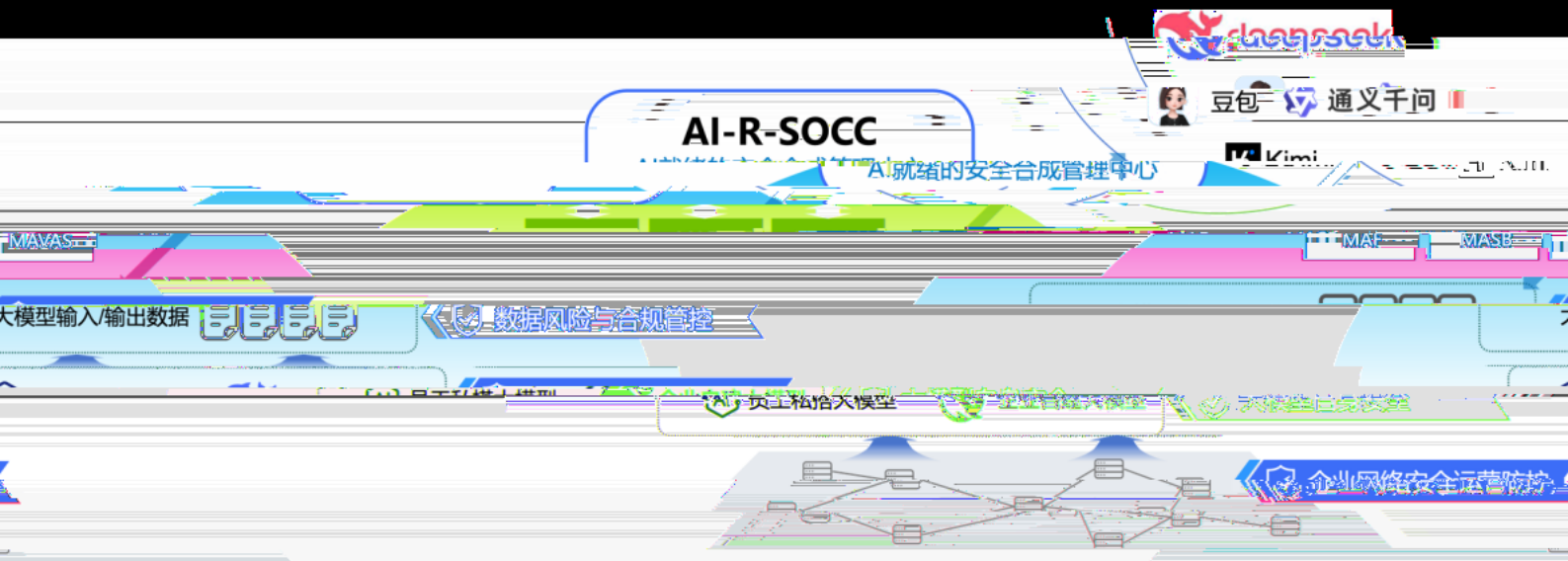
- 知识沉淀不足：处置经验沉淀在人员头脑中，缺乏标准化剧本库供复用。

3 定义和建设思路

3.1 AI-R-SOCC 定义

面对大模型安全挑战，亟需构建一体化的AI安全管理与融合大模型应用安全、数据安全

全融合管理中心 (AI-R-SOCC) 应运而生。



的趋势下,大模型也成为了企业中的信息交互中心,

在企业员工大量使用大模型辅助办公

边界问题,给企业安全运营工作带来了新的挑战。AI-R-SOCC 依托

这势必带来了新的安全边

一种最直接的体现是,在数据治理方面,企业需要建立统一的数据治理平台,实现数据

MAVAS、MASR、MAVAS 等大模型安全的管控手段,可形成对大模型应用白盒的安全性评估

MA

说,通过这些过

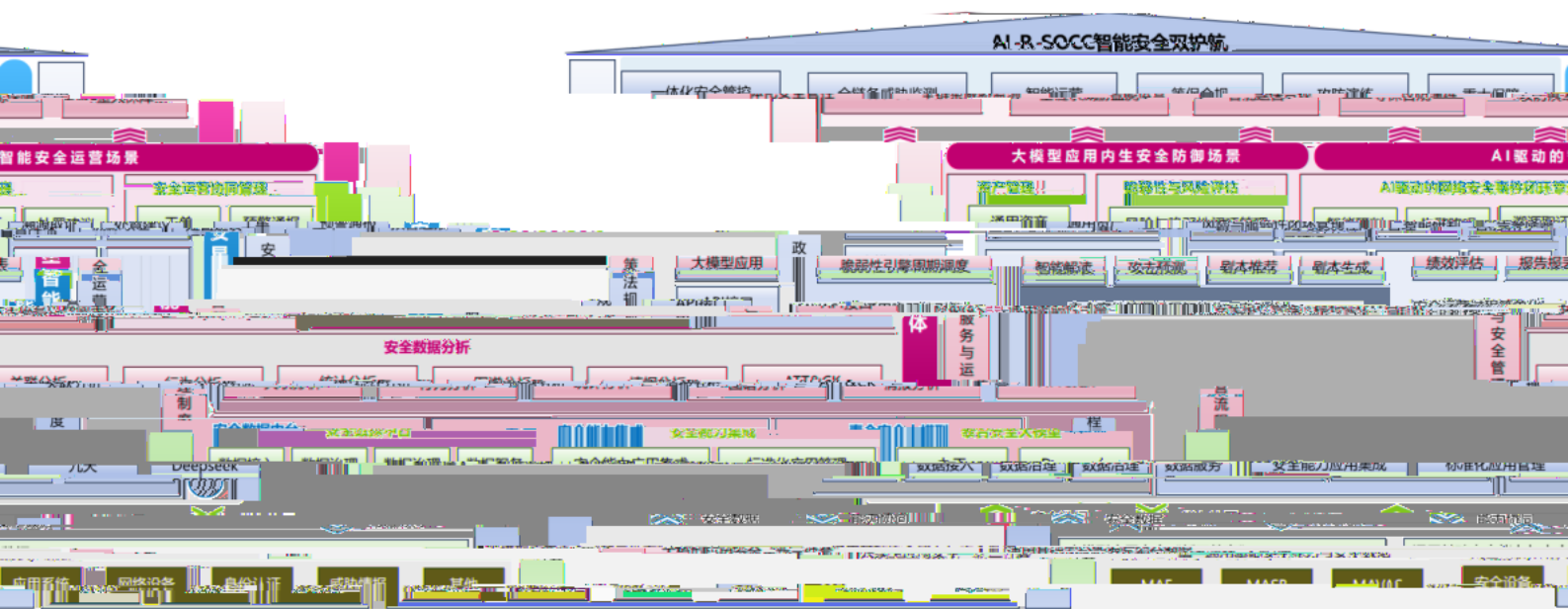
什么?只点然评估时大模型结果中所左给,给出数据的风险性和合规性的

放之于企业网络安全的具体视角由驱动统一运营驱动,形成大模型应用

将会被 AI-R-SOCC

网络安全的一体化安全运营

安全,同时安全



台提供多源异构的海量数据进行接入和治理，可接入融合包括大模型应用安全、“新三

平

（MAF 大模型应用防火墙、MASB 大模型访问安全代理、MAVAS 大模型安全评估

件套”

化分析引擎来支撑上层能力构建，包括处置海量多源异构信息的数据中台能力、提供安星智

能辅助研判及智能决策的融合安全大模型能力，以及多种关联分析引擎、行为分析引擎

图谱分析引擎。

（提供了面向大模型应用生命周期内的安全管理能力以及 AI 驱动

在此基础上 AI-R-SOCC

的大模型的安全管理能力深度融合 MAF、MASB 和 MAVAS 的

的智能化安全管理能力。面

的综合评估能力、基于综合分析的大模型应用风险监测及安全事件

大模型资产安全性及合规性

策略联动响应能力。智能化安全管理能力深度融合 AI 驱动资产

件管理能力以及风险行为底

4 核心能力

4.1 人工智能能力

人工智能能力是智眼目目平台的核心能力，也是平台区别于传统视频监控平台的关键能力。平台通过引入人工智能技术，实现了对视频数据的智能分析和处理，提高了安全事件的检测效率和准确性。

人工智能能力主要体现在以下几个方面：1. 智能识别：平台能够自动识别视频中的目标物体，如车辆、行人、动物等，并对其进行分类和跟踪。2. 行为分析：平台能够对目标物体的行为进行分析，如异常行为、聚集行为等，及时发现潜在的安全隐患。3. 智能检索：平台支持基于关键词、时间、地点等多种条件的智能检索，方便用户快速查找所需信息。

平台使安全人员能够在统一的界面上高效完成指令传递和执行，同时调用预设安全剧本，大

大提升安全事件处理效率。平台还支持与第三方系统对接，实现数据共享和协同作战。

平台还支持多种部署方式，如本地部署、云部署等，满足不同场景的需求。平台具有良好的兼容性和扩展性，能够适应未来技术的发展。

平台还支持多种部署方式，如本地部署、云部署等，满足不同场景的需求。平台具有良好的兼容性和扩展性，能够适应未来技术的发展。

平台还支持多种部署方式，如本地部署、云部署等，满足不同场景的需求。平台具有良好的兼容性和扩展性，能够适应未来技术的发展。

平台还支持多种部署方式，如本地部署、云部署等，满足不同场景的需求。平台具有良好的兼容性和扩展性，能够适应未来技术的发展。

平台还支持多种部署方式，如本地部署、云部署等，满足不同场景的需求。平台具有良好的兼容性和扩展性，能够适应未来技术的发展。

平台还支持多种部署方式，如本地部署、云部署等，满足不同场景的需求。平台具有良好的兼容性和扩展性，能够适应未来技术的发展。

平台还支持多种部署方式，如本地部署、云部署等，满足不同场景的需求。平台具有良好的兼容性和扩展性，能够适应未来技术的发展。

4.1.1 能力市场

能力市场为安全运营提供了一个聚合能力资源平台，将各种安全能力封装成能力包，

方便用户按需订阅和使用。能力市场支持多种计费模式，如按次计费、按月订阅等，满足不同用户的需求。

平台还支持多种部署方式，如本地部署、云部署等，满足不同场景的需求。平台具有良好的兼容性和扩展性，能够适应未来技术的发展。

平台还支持多种部署方式，如本地部署、云部署等，满足不同场景的需求。平台具有良好的兼容性和扩展性，能够适应未来技术的发展。

平台还支持多种部署方式，如本地部署、云部署等，满足不同场景的需求。平台具有良好的兼容性和扩展性，能够适应未来技术的发展。

复杂场景下的自动化响应流程

提升设备运维的准确性、性能

同时，系统提供对脚本执行情况进行对设备状态的实时监控，全程

发现问题并采取防范措施，提升整体安全运营的效率。

状态及潜在安全隐患，帮助用户提前发

4.1.2 任务管理

系统涵盖人工任务、周期任务和脆弱性任务三

任务管理提供了全面的任务调度与管理功能

● 人工任务管理

通过直观的界面展示待办任务详情，用户可以快速查看、分配和处理各项任务，确保任

理

● 周期任务管理

务调度功能，支持用户根据需求设置任务执行周期，并配置具体的执行

提供自动化的任

任务，减少人工干预，提升任务执行的自动化

动作，帮助用户高效管理日常定期检查与维护

● 脆弱性任务管理

户快速识别和修复系统中的安全漏洞，全面保障资产安全。

4.1.3 大模型安全风险智能响应

智能响应围绕脆弱性监测这一核心任务，包括智能监测、智能预警和研判任务三大功能。

构建了高效的安全监控体系。系统通过实时分析和动态跟踪，提前实现对脆弱性、告警事件

的感知能力，有效提升脆弱性监测的效率和准确性，及时发现潜在的安全风险，提升系统的

潜在风险，为安全态势感知与持续优化提供

● 智能监测

对脆弱性进行持续监控，支持用户创建自定义监

通过大模型强大的实时分析能力，对系

智能监测安全风险。

通过大模型强大的实时分析能力，对系

及时发现和处理。

提供实时风险监控和告警能力，通过图表和列表展示脆弱性监测任务

实时监控任务

帮助用户实时监控任务进展，了解任务状态。系统结合大模型智能分析能力，将

任务执行情况，帮助

对脆弱性任务进行可视化呈现，为用户提供实时监控和综合分析的数据支撑。

任务执行情况，帮助

系统支持实时监控和告警能力。

系统支持实时监控和告警能力。

○ 研判任务

系统支持实时监控和告警能力，通过图表和列表展示脆弱性监测任务

潜在风险，帮助用户

智能有效判断，结合智能分析以及图像分析，全面评估任务的脆弱性

及时发现和处理响应效

快速识别和评估处理安全风险事件，降低安全风险，减少不必要的干扰，

率。

○ 大模型风险处置

系统支持实时监控和告警能力，通过图表和列表展示脆弱性监测任务

实时监控任务



- **提示词攻击处置：**MAF 检测到恶意指令注入 → AI-R-SOCC 触发 MASB 冻结账号

实时脱敏输出内容 → MAVAS

- **数据泄露阻断：**MASB 识别敏感文件上传 → MAF

并告警数据治理部门。

扫描关联模型训练集残留风险 → 隔离高风险模型并

封合法用户）。

误封

司资产管理

4.2 全

提供全面的资产管理能力，帮助企业“摸清家底”，通过多种适配器的有机融

平台提

景下，新增了针对大模型的资产管控，使组织能够从全局掌控大模型应用相关的

模型应用场景

产状况，为风险管控、全局监测提供基础支撑。

资产

大模型应用资产管控能力，帮助企业摸清家底，实现大模型应用资产的统一

，这有助于组织更好地管理大模型应用资产，提升大模型应用的安全性和合规性。

资源和组件。这类实体资产主要包括：

▲ **基础设施：**大模型应用部署的服务器、云主机、算力服务器（CDI D），以及关联

的网络资源，提供完善的管理能力。

库、数据湖、

- **应用软件：**对大模型应用相关的应用、组件进行统一管理，包括数据

应用组件管理

应用组件管理

4.3 大模型安全分析

同 通計公在式共野公析引敬 已党行為公析引敬

那么甘工夕酒已粉的全量完全数也

ATT&CK 分析 知识图谱分析 并结合告警智能降噪 事件自动溯源等

UFRA 画像分析

4.3.1 大模型风险关联分析

1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023, 2024, 2025, 2026, 2027, 2028, 2029, 2030, 2031, 2032, 2033, 2034, 2035, 2036, 2037, 2038, 2039, 2040, 2041, 2042, 2043, 2044, 2045, 2046, 2047, 2048, 2049, 2050, 2051, 2052, 2053, 2054, 2055, 2056, 2057, 2058, 2059, 2060, 2061, 2062, 2063, 2064, 2065, 2066, 2067, 2068, 2069, 2070, 2071, 2072, 2073, 2074, 2075, 2076, 2077, 2078, 2079, 2080, 2081, 2082, 2083, 2084, 2085, 2086, 2087, 2088, 2089, 2090, 2091, 2092, 2093, 2094, 2095, 2096, 2097, 2098, 2099, 2100, 2101, 2102, 2103, 2104, 2105, 2106, 2107, 2108, 2109, 2110, 2111, 2112, 2113, 2114, 2115, 2116, 2117, 2118, 2119, 2120, 2121, 2122, 2123, 2124, 2125, 2126, 2127, 2128, 2129, 2130, 2131, 2132, 2133, 2134, 2135, 2136, 2137, 2138, 2139, 2140, 2141, 2142, 2143, 2144, 2145, 2146, 2147, 2148, 2149, 2150, 2151, 2152, 2153, 2154, 2155, 2156, 2157, 2158, 2159, 2160, 2161, 2162, 2163, 2164, 2165, 2166, 2167, 2168, 2169, 2170, 2171, 2172, 2173, 2174, 2175, 2176, 2177, 2178, 2179, 2180, 2181, 2182, 2183, 2184, 2185, 2186, 2187, 2188, 2189, 2190, 2191, 2192, 2193, 2194, 2195, 2196, 2197, 2198, 2199, 2200, 2201, 2202, 2203, 2204, 2205, 2206, 2207, 2208, 2209, 2210, 2211, 2212, 2213, 2214, 2215, 2216, 2217, 2218, 2219, 2220, 2221, 2222, 2223, 2224, 2225, 2226, 2227, 2228, 2229, 2230, 2231, 2232, 2233, 2234, 2235, 2236, 2237, 2238, 2239, 2240, 2241, 2242, 2243, 2244, 2245, 2246, 2247, 2248, 2249, 2250, 2251, 2252, 2253, 2254, 2255, 2256, 2257, 2258, 2259, 2260, 2261, 2262, 2263, 2264, 2265, 2266, 2267, 2268, 2269, 2270, 2271, 2272, 2273, 2274, 2275, 2276, 2277, 2278, 2279, 2280, 2281, 2282, 2283, 2284, 2285, 2286, 2287, 2288, 2289, 2290, 2291, 2292, 2293, 2294, 2295, 2296, 2297, 2298, 2299, 2300, 2301, 2302, 2303, 2304, 2305, 2306, 2307, 2308, 2309, 2310, 2311, 2312, 2313, 2314, 2315, 2316, 2317, 2318, 2319, 2320, 2321, 2322, 2323, 2324, 2325, 2326, 2327, 2328, 2329, 2330, 2331, 2332, 2333, 2334, 2335, 2336, 2337, 2338, 2339, 2340, 2341, 2342, 2343, 2344, 2345, 2346, 2347, 2348, 2349, 2350, 2351, 2352, 2353, 2354, 2355, 2356, 2357, 2358, 2359, 2360, 2361, 2362, 2363, 2364, 2365, 2366, 2367, 2368, 2369, 2370, 2371, 2372, 2373, 2374, 2375, 2376, 2377, 2378, 2379, 2380, 2381, 2382, 2383, 2384, 2385, 2386, 2387, 2388, 2389, 2390, 2391, 2392, 2393, 2394, 2395, 2396, 2397, 2398, 2399, 2400, 2401, 2402, 2403, 2404, 2405, 2406, 2407, 2408, 2409, 2410, 2411, 2412, 2413, 2414, 2415, 2416, 2417, 2418, 2419, 2420, 2421, 2422, 2423, 2424, 2425, 2426, 2427, 2428, 2429, 2430, 2431, 2432, 2433, 2434, 2435, 2436, 2437, 2438, 2439, 2440, 2441, 2442, 2443, 2444, 2445, 2446, 2447, 2448, 2449, 2450, 2451, 2452, 2453, 2454, 2455, 2456, 2457, 2458, 2459, 2460, 2461, 2462, 2463, 2464, 2465, 2466, 2467, 2468, 2469, 2470, 2471, 2472, 2473, 2474, 2475, 2476, 2477, 2478, 2479, 2480, 2481, 2482, 2483, 2484, 2485, 2486, 2487, 2488, 2489, 2490, 2491, 2492, 2493, 2494, 2495, 2496, 2497, 2498, 2499, 2500, 2501, 2502, 2503, 2504, 2505, 2506, 2507, 2508, 2509, 2510, 2511, 2512, 2513, 2514, 2515, 2516, 2517, 2518, 2519, 2520, 2521, 2522, 2523, 2524, 2525, 2526, 2527, 2528, 2529, 2530, 2531, 2532, 2533, 2534, 2535, 2536, 2537, 2538, 2539, 2540, 2541, 2542, 2543, 2544, 2545, 2546, 2547, 2548, 2549, 2550, 2551, 2552, 2553, 2554, 2555, 2556, 2557, 2558, 2559, 2560, 2561, 2562, 2563, 2564, 2565, 2566, 2567, 2568, 2569, 2570, 2571, 2572, 2573, 2574, 2575, 2576, 2577, 2578, 2579, 2580, 2581, 2582, 2583, 2584, 2585, 2586, 2587, 2588, 2589, 2590, 2591, 2592, 2593, 2594, 2595, 2596, 2597, 2598, 2599, 2600, 2601, 2602, 2603, 2604, 2605, 2606, 2607, 2608, 2609, 2610, 2611, 2612, 2613, 2614, 2615, 2616, 2617, 2618, 2619, 2620, 2621, 2622, 2623, 2624, 2625, 2626, 2627, 2628, 2629, 2630, 2631, 2632, 2633, 2634, 2635, 2636, 2637, 2638, 2639, 2640, 2641, 2642, 2643, 2644, 2645, 2646, 2647, 2648, 2649, 2650, 2651, 2652, 2653, 2654, 2655, 2656, 2657, 2658, 2659, 2660, 2661, 2662, 2663, 2664, 2665, 2666, 2667, 2668, 2669, 2670, 2671, 2672, 2673, 2674, 2675, 2676, 2677, 2678, 2679, 26

其他类型的安全数据进行实时分析, 不仅覆盖安全运营场

关联能力，将 MAF、MSBA 以及其

目的数据关联八城需求, 建立 1 城型六四六八七月提供了三河、古站。

最早的右效荔枝

这棵树下面就是253号井

通心脉、聚谷、缩肝、关腹等自设，发抽除藏。

生成器函数的安全使用

术。它可以辅助识别网络故障原因,并修复故障。

MAVA MAE MASIRI

的不安全漏洞进行“实时检测”发现,实时多维漏洞风险评估,动态关联模型,SI(如 MAVAS

检测的(CVE编号)、异常使用行为(如MAF捕获的漏洞注入攻击)、构造性优化



變異：高：低頻變異重組

【关键词】MAE; 网络模型; 控制; 误差; 结合; MASB

介绍和风险分析到生产现场管理。

4.3.2. 基于用户身份的行为分析

针对主机、终端、应用、网络流量和数据库等，基于历史

数据构建用户行为特征模型，对异常行为进行识别和告警。

通过构建用户行为特征模型，对异常行为进行识别和告警。

件。

对用户和实体的行为进行收集、处理和解析，主要关注网络中

行为管理负责

检测。通过对网络中的行为深度解析和识别，将行为数据

的行为模型和异常行为

管理，构建行为分析模型。

和特征数据提供给模型

在大数据应用安全场景中，平台基于 MASB 平台的细粒度身份信息（包括设备角色、

部门归属、权限等级），AI-R-SOCC 构建动态用户画像，实现“身份-行为-数据”三位一

体分析。通过关联 MAF 的输入输出审计日志、MASB 的访问控制记录及 MAVAS 的合规检

测结果，系统可精准识别高风险行为模式，分析能力例如：

● **风险行为评估**：整合各大模型安全子系统的数据，对用户的大模型使用行为进行风

险和合规分析。

通过构建用户行为特征模型，对异常行为进行识别和告警。

通过构建用户行为特征模型，对异常行为进行识别和告警。

通过构建用户行为特征模型，对异常行为进行识别和告警。

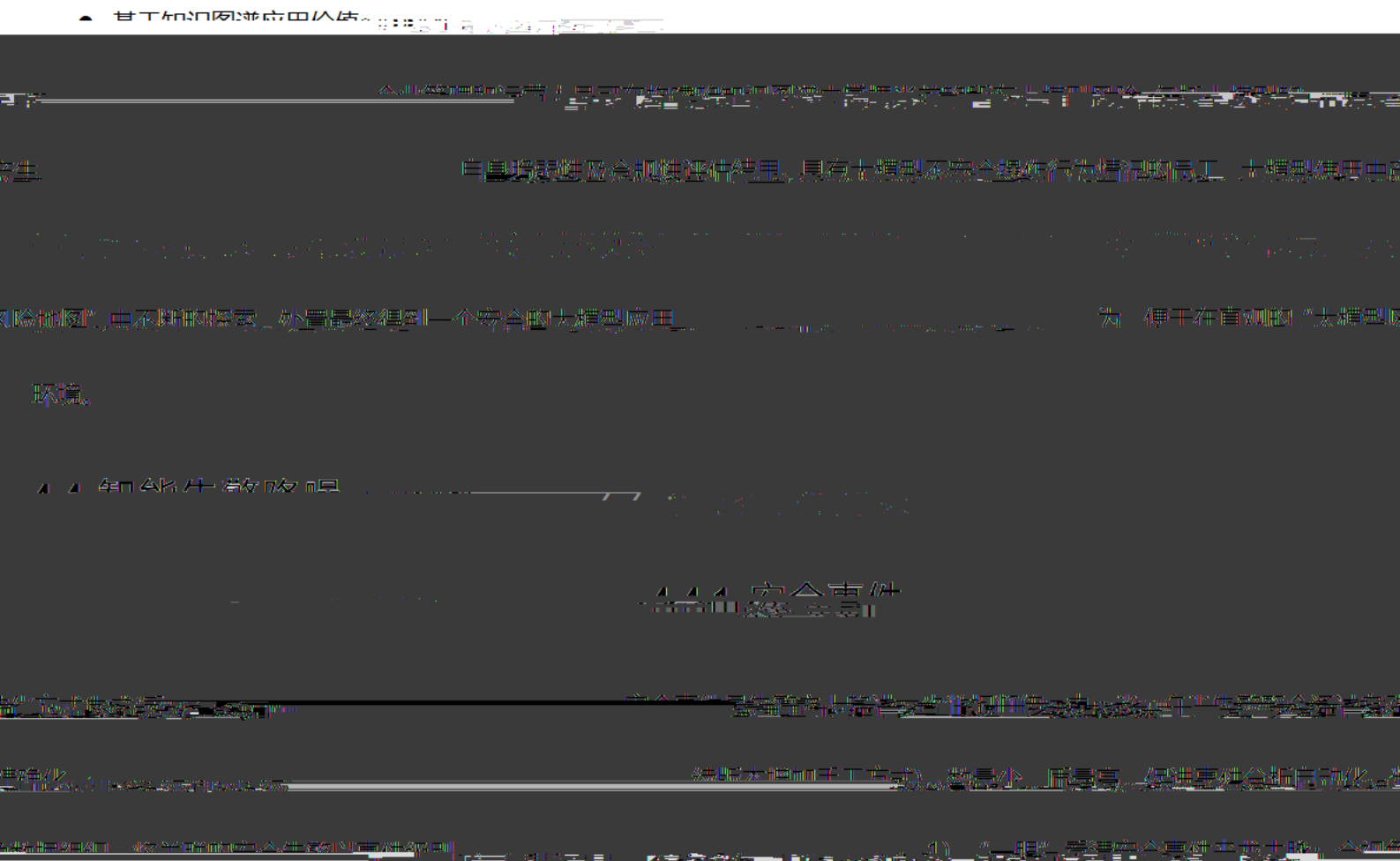
通过构建用户行为特征模型，对异常行为进行识别和告警。

通过构建用户行为特征模型，对异常行为进行识别和告警。

通过构建用户行为特征模型，对异常行为进行识别和告警。

- 图谱实体涵盖：
 - **风险实体**：用户（身份/权限）、大模型、业务资产、敏感数据信息元素、安全告警/事件、合规策略等。
 - **风险关系**：使用关系、会话关联、输入关系、输出关系、事件归属、策略违规映射

等



关键路径、ATT&CK 等方式进行呈现，一眼看出安全告警的重要信息，同时在安全事件的概览页面将 IP、资产等信息进行展示，清晰展示安全事件的影响范围。

2) “一图”展示安全事件攻击故事线：以拓扑形式展示安全事件的攻击关系，方便用户的安全事件进行研判。拓扑图以 IP/资产为节点，以关联的告警为边，并支持数据的下钻，

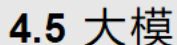
可以表示为如下形式：

3) 全分岐型道岔—安全

4.4.2

告警疲劳中解脱出来，投入到安全运营

全息降噪引擎将安全专家的知识与



多维度监测体系,实现从风险发现的全局可见,并可指导后续的联动处置形成大模型

大模型应用实战 | 深度应用安全其座系列白皮书



2) 由安全抓守时检测：动态维护国家到行业再到企业单位的十模形使用政策及规范

4) **服务能力监控**: 监测大模型的响应延迟、资源利用率, 可与企业内多大模型的服务接口联动进行服务自动触发熔断或负载均衡策略。

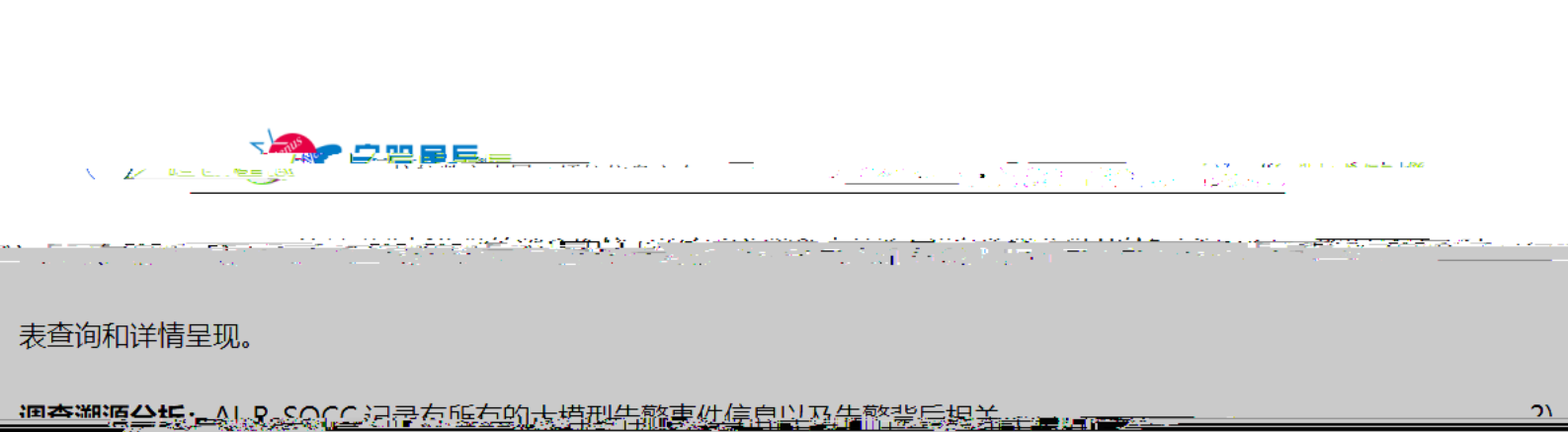
5) **系统层风险预警：**扫描模型依赖环境风险（如部署服务器的系统漏洞或相关组件的

4.5.2 风险人员监测

2) 高危用户画像: 其人员为高风险非企业用户

4.5.3 风险行为/事件监测

1) 多维度信息聚合: 将 MAE、MASE 所监测的人模型攻击行为



表查询和详情呈现。

图 4-1-2 AI R SOCC 记录有所有的大模型告警事件信息以及告警处理后相关

的原始日志、行为分析甚至流量数据，可根据安全运营的诉求对告警事件进行深入

溯源和溯源分析，对事件来源、传播路径、影响范围进行溯源和溯源分析，及时发

入排查并减少大模型的应用风险。

4.1.4 智能治理建议生成

处理建议，并综合企业的大模型应用环境给出治理建议，帮助用户推进大模型的安全合规化

使用。

4.1.5 大模型风险管控处置

具体参见“4.1 安全风险治理”章节。

行为审计能力建设

针对针对大模型应用的安全行为进行全程监测记录，行为审计能力（审计内容、审计范围、审计

策略、审计数据存储、审计报告生成、审计数据共享等）。

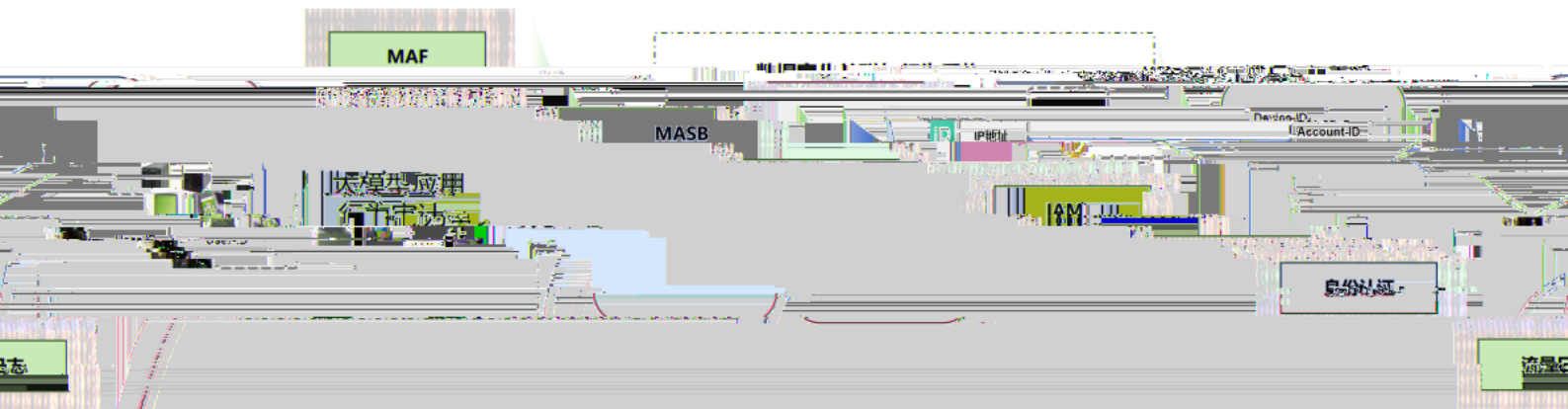
行为审计能力是安全运营的重要组成部分，通过行为审计能力，可以实现对大模型应用的安全行为进行全程监测记录，及时发现和处置安全风险。

行为审计能力包括“行为审计”、“行为审计策略”、“行为审计策略配置”、“行为审计策略管理”等。

行为审计能力是安全运营的重要组成部分，通过行为审计能力，可以实现对大模型应用的安全行为进行全程监测记录，及时发现和处置安全风险。

行为的访问行为进行全程记录。

用户、业务协同对大模型应用



大模型安全态势呈现

4.7 大模型安全态势呈现

在大模型安全运营过程中，各种安全风险评估、监测、管控的过程中将产生大量的监测数据，AI-R-SOCC 在大模型安全运营过程中，通过各项数据融合引擎与 2D+3D 可视化技术的应用，将海量的合规探测、行为监测、风险

通过各项数据融合引擎与 2D+3D 可视化技术的应用，将海量的合规探测、行为监测、风险

通过各项数据融合引擎与 2D+3D 可视化技术的应用，将海量的合规探测、行为监测、风险

通过各项数据融合引擎与 2D+3D 可视化技术的应用，将海量的合规探测、行为监测、风险

险。

AI-R-SOCC 提供了大模型综合风险态势、大模型资产风险态势、大模型使用风险监测

态势、大模型合规运营态势等综合风险态势，

大模型综合风险态势，还有大模型合规运营态势的呈现，综合所有的风险因素呈现

险、使用中的安全威胁、用户行为安全的高维

总体安全健康指数以及大模型自身风险

态势。

指标，便于全局掌控安全风险和安全态

通过各项数据融合引擎与 2D+3D 可视化技术的应用，将海量的合规探测、行为监测、风险

通过各项数据融合引擎与 2D+3D 可视化技术的应用，将海量的合规探测、行为监测、风险

通过各项数据融合引擎与 2D+3D 可视化技术的应用，将海量的合规探测、行为监测、风险

为广量早流动风险的实时监控

● 上提型住田品哈吹本势 万会上提型住田品和卡上只仁

佳田白权阻泄田 敏感信自泄屋 已常操作等威胁维度 通过多源行为口主

控 取

融合分析，便于快速识别高危人员并阻断违规行为。

信息，智能运营任务完成情况

● 大模型智能运营态势：智能运营态势大屏集成了总览

状，展示工具即生成数TOP10和十只即写写TOP10

成助当型分大TOP10 成助当型分大

有效优化资源分配与业务流程，为快速响应精准防控提供了坚实基础。



5 部署和典型应用场景

AI-R-SOCC 作为大模型应用安全方案的核心基座，对接大模型应用安全“新三件套”

(MAF 大模型应用防火墙、MASB 大模型访问安全代理、MAVAS 大模型安全评估系统)，

和高效 全运营，针对大模型在训练、推理、部署等全生命周期的复杂安全需求，提供针对性

暂行办法》）、伦理准则（如歧视性言论），引发法律处罚与品牌危机等。

管理

● 解决方案

数据泄漏防控

1. 敏感数

MASD 的 AI C 通过 MASD 引擎，实时监控用户上传与指令中的敏感内容，敏感内容（如客户住址、商业机密）...

动态拦截高风险内容；

去违抑内 ✓ 实时监控输出内容，识别敏感信息泄露（如客户住址、商业秘密）...

敏感信息泄露

敏感信息泄露，如客户住址、商业机密等

及时抑制风险的产生。

2. 生成内容滥用阻断

✓ 语义深度解析：基于多轮对话上下文理解，识别隐晦攻击指令（如“生成绕过中

术”）；

核的贷款申请话术

生成内容滥用，如生成绕过安全策略的指令、恶意攻击脚本等

生成内容滥用，如生成绕过安全策略的指令、恶意攻击脚本等

实时拦截违规内容：

求合规策略（如禁止生成“保本保收益”承诺），

生成内容滥用，如生成绕过安全策略的指令、恶意攻击脚本等

生成内容滥用，如生成绕过安全策略的指令、恶意攻击脚本等

大模型投毒攻击防御

3.

输入数据验毒：通过 MAE 检测输入的异常数据（如噪声注入），联动 MASD 阻

✓

源 IP；

断污染源

训练或投喂数据批量认证：对模型训练集进行 MAVAS 安全扫描，拦截未

✓ 大模型训

认证数据源。

生成内容滥用，如生成绕过安全策略的指令、恶意攻击脚本等

型对话交互

✓ 大模型对话敏感数据防泄漏：通过对输出数据的全面解析，识别大模型输出过程中的敏感信息输出实时识别、封堵或脱敏，保障大模型服务合规。

各管理环节

✓ 动态合规监测，输出内容实时动态监管方式（如《生成式人工智能管理办法》1.1.1 对

不合规内容生成告警。

● 价值成果

大幅降低

低；

3. 减少业务损失：避免因模型滥用导致的品牌声誉损害与用户流失。

5.2 场景一：模型上线准入管控

● 痛点

数据污染、隐私泄露漏洞）；

“带病上线”风险（如训练数

● 解决方案

1. 自动化安全审查：

资源过载 (CPU利

✓ 服务能力退化风险: 模型API响应延迟飙升 (P95>1000ms)

超时等待。

利用率>95%), 引发业务中断, 且缺乏实时预警与自动

性能退化。

1.500055

1. 漏洞实时监测与修复:

境 (如CUDA版本、容器镜像)

✓ 漏洞扫描: 联动MAVAS每小时扫描模型依赖环境

漏洞告警。

识别高危漏洞 (如CVE-2023-1234)。

自动推荐修复建议 (如升级TensorFlow至2.12)。

✓ 自动化补丁: 对低风险漏洞

高风险漏洞触发模型隔离并告警。

2. 合规性动态适配:

动态合规引擎 (支持法规更新), 实时扫描输出内容。

✓ 敏感词过滤: 集成最新敏感词库, 动态更新。

(如禁止生成敏感话题相关内容)。

违规内容实时拦截。

引擎实时扫描模型输出与合规知识库, 违规内容实时脱敏或

✓ 输出内容校验: 通过NLP

阻断。

3. 服务能力保障:

✓ 性能基线监控: 设定API响应延迟 (<500ms)、错误率 (<1%)、资源利用率

(CPU<85%)、动态扩容。异常时触发熔断与告警。

至轻量化模型版本。

● 价值成果

“AI+政务”深度融合, 显著提升政务服务效率与用户体验, 助力数字政府建设。

2. 政务问答模型输出内容合规率提升至 100%。

5.4 场景四：影子大模型监测与治理

痛点

- ✓ 企业大模型资产清单分散在多个部门，存在未登记的“影子模型”；

企业员工私搭模型（如 Llama 2 代码生成工具）成为数据泄露与攻击跳板

解决方案

包括容器化部署的私搭模型），构建动态资产库；

- ✓ 自动发现企业内部模型实例（包

从特征、与代码构建（检测私有模型仓库）、自动构建模型指纹库；通过流量检测（模型特征、API

搭模型指纹库。

建私

对发现的影子大模型进行风险探测和输入/输出监测，对风险进行提示并在必

- ✓ 通过

实时阻断。

要时

价值成果

模型资产黑洞消除：识别并治理若干个未登记的私搭模型，收缩企业内大模型攻

大模型

击面风险。

6 能力优势

图 6-1 大模型安全能力优势

大模型安全能力优势主要体现在以下几个方面：一是模型安全，通过模型训练数据溯源监测、推理过程行为分析等任务，实现大模型安全运营的日常工作的自动化；二是数据安全，通过数据流向、风险检测点、任务执行进度等关键信息清晰直观；三是模型安全，通过模型训练数据溯源监测、推理过程行为分析等任务，实现大模型安全运营的日常工作的自动化；四是数据安全，通过数据流向、风险检测点、任务执行进度等关键信息清晰直观。

大模型安全能力优势主要体现在以下几个方面：一是模型安全，通过模型训练数据溯源监测、推理过程行为分析等任务，实现大模型安全运营的日常工作的自动化；二是数据安全，通过数据流向、风险检测点、任务执行进度等关键信息清晰直观；三是模型安全，通过模型训练数据溯源监测、推理过程行为分析等任务，实现大模型安全运营的日常工作的自动化；四是数据安全，通过数据流向、风险检测点、任务执行进度等关键信息清晰直观。

的环境下持续运行。

6.2 集中调度

大模型安全能力优势主要体现在以下几个方面：一是模型安全，通过模型训练数据溯源监测、推理过程行为分析等任务，实现大模型安全运营的日常工作的自动化；二是数据安全，通过数据流向、风险检测点、任务执行进度等关键信息清晰直观；三是模型安全，通过模型训练数据溯源监测、推理过程行为分析等任务，实现大模型安全运营的日常工作的自动化；四是数据安全，通过数据流向、风险检测点、任务执行进度等关键信息清晰直观。

大模型安全能力优势主要体现在以下几个方面：一是模型安全，通过模型训练数据溯源监测、推理过程行为分析等任务，实现大模型安全运营的日常工作的自动化；二是数据安全，通过数据流向、风险检测点、任务执行进度等关键信息清晰直观；三是模型安全，通过模型训练数据溯源监测、推理过程行为分析等任务，实现大模型安全运营的日常工作的自动化；四是数据安全，通过数据流向、风险检测点、任务执行进度等关键信息清晰直观。

安全运营效率，为大模型构建全方位的安全防护网。

6.3 快速响应

大模型安全能力优势主要体现在以下几个方面：一是模型安全，通过模型训练数据溯源监测、推理过程行为分析等任务，实现大模型安全运营的日常工作的自动化；二是数据安全，通过数据流向、风险检测点、任务执行进度等关键信息清晰直观；三是模型安全，通过模型训练数据溯源监测、推理过程行为分析等任务，实现大模型安全运营的日常工作的自动化；四是数据安全，通过数据流向、风险检测点、任务执行进度等关键信息清晰直观。

大模型安全能力优势主要体现在以下几个方面：一是模型安全，通过模型训练数据溯源监测、推理过程行为分析等任务，实现大模型安全运营的日常工作的自动化；二是数据安全，通过数据流向、风险检测点、任务执行进度等关键信息清晰直观；三是模型安全，通过模型训练数据溯源监测、推理过程行为分析等任务，实现大模型安全运营的日常工作的自动化；四是数据安全，通过数据流向、风险检测点、任务执行进度等关键信息清晰直观。

6.4 安全专家

数据分析能力，提供针对大模型安全的实战安全知识智能

依托强大的大模型知识图谱与类

我们的大模型安全专家，能够针对大模型安全威胁，提供精准的安全防护策略，帮助企业有效应对大模型安全威胁。

通过大模型安全专家，企业可以实时监测大模型安全威胁，并及时采取相应的安全防护措施，确保大模型的安全运行。

大模型安全专家，是企业大模型安全的重要保障，能够帮助企业有效应对大模型安全威胁，确保大模型的安全运行。

体作战实力，从容应对各类大模型安全威胁。

大模型安全专家，是企业大模型安全的重要保障，能够帮助企业有效应对大模型安全威胁，确保大模型的安全运行。

警技术，实现 7×24 小时全年无间断的大模型安全运

采用先进的自动化监控与智能预

大模型安全专家，是企业大模型安全的重要保障，能够帮助企业有效应对大模型安全威胁，确保大模型的安全运行。

大模型安全专家，是企业大模型安全的重要保障，能够帮助企业有效应对大模型安全威胁，确保大模型的安全运行。

大模型安全专家，是企业大模型安全的重要保障，能够帮助企业有效应对大模型安全威胁，确保大模型的安全运行。

