Cibersegurança e a Revolução Tecnológica nas PME

Inteligência Artificial na cibersegurança: problema ou solução?

IAPMEI

academia PME

Esta apresentação não foi gerada por IA

(Espero não alucinar)

6%

"IA de bolso"

As empresas estão a adoptar IA, mas têm preocupações

Sob Pressão

64%

Das empresas sofrem pressão para acelerar iniciativas de IA

Preocupadas com os riscos

84%

Consideram a cibersegurança como o bloqueador

#1 à adopção de IA generativa

Investindo em novas defesas

64%

Identificaram a segurança como prioridade na criação de use cases de IA

Outros Temas:

- Propriedade Intelectual
- Transparência
- Residência
- Em suma, soberania

Fonte: “Enterprise generative AI: State of the market” 2023 IBM Institute of Business Value Study

Cibersegurança e a Revolução Tecnológica nas PME

Ataques à IA	IA deve ser tratada como nova superfície de ataque. Necessidade de novas estratégias de deteção e resposta para evasão de modelo, extração, inferência e poisoning	Prompt injection pode evadir as defesas contra a geração de conteúdos indesejáveis, para além do acesso a integrações exploráveis e a um volume de dados de treino potencialmente sensíveis	Modelos maliciosos podem ser carregados para repositórios abertos, com comportamentos escondidos, desencadeados muito depois de terem sido instalados
Ataques utilizando *IA*	Generative AI irá dar escala ao ciber crime, e reduzir barreiras à entrada na cibersegurança – baixando o nível de competências necessários para atacar	O Phishing será cada vez mais "targetado" e o as técnicas de IA generativa para vídeo e áudio irão requerer novas abordagens de defesa	Os atacantes irão adaptar-se mais depressa a estratégias defensivas e melhorar a evasividade, descoberta de vulnerabilidades e adaptação de malware

Segurança para a IA

Proteger os modelos fundacionais, a IA generativa e os data sets é essencial

Proteger os dados de treino da IA contra roubo, manipulação ou violações de PI

Modelos seguros contra vulnerabilidades no pipeline e nas integrações e com políticas restritas de Gestão de acessos

Segurar a utilização dos modelos de IA detetando exfiltrações de dados ou de prompts, e estabelecer monitorização

ver. IBM Adversarial Robustness Toolkit

IA para Segurança

Ganhos de produtividade irão reduzir os "gargalos" humanos na Segurança

IA irá gerir tarefas repetitivas como a explicação e documentação de casos,ou a resposta a problemas conhecidos

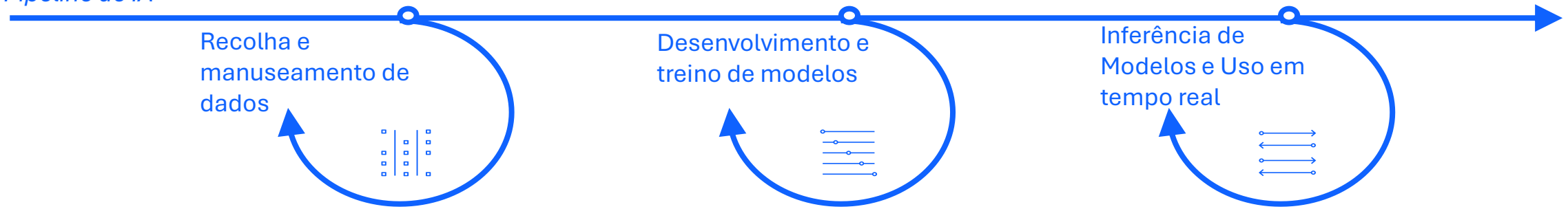
IA poderá gerar conteúdos mais rapidamente, levando em consideração elementos multidisciplinares – novos workflows, políticas de deteção

IA irá aprender e criar respostas a novos desafios – novas analogias para novos ataques, patching e Gestão de vulnerabilidades

IA pode trabalhar 24x7, e poderá ter menos bias que os humanos!

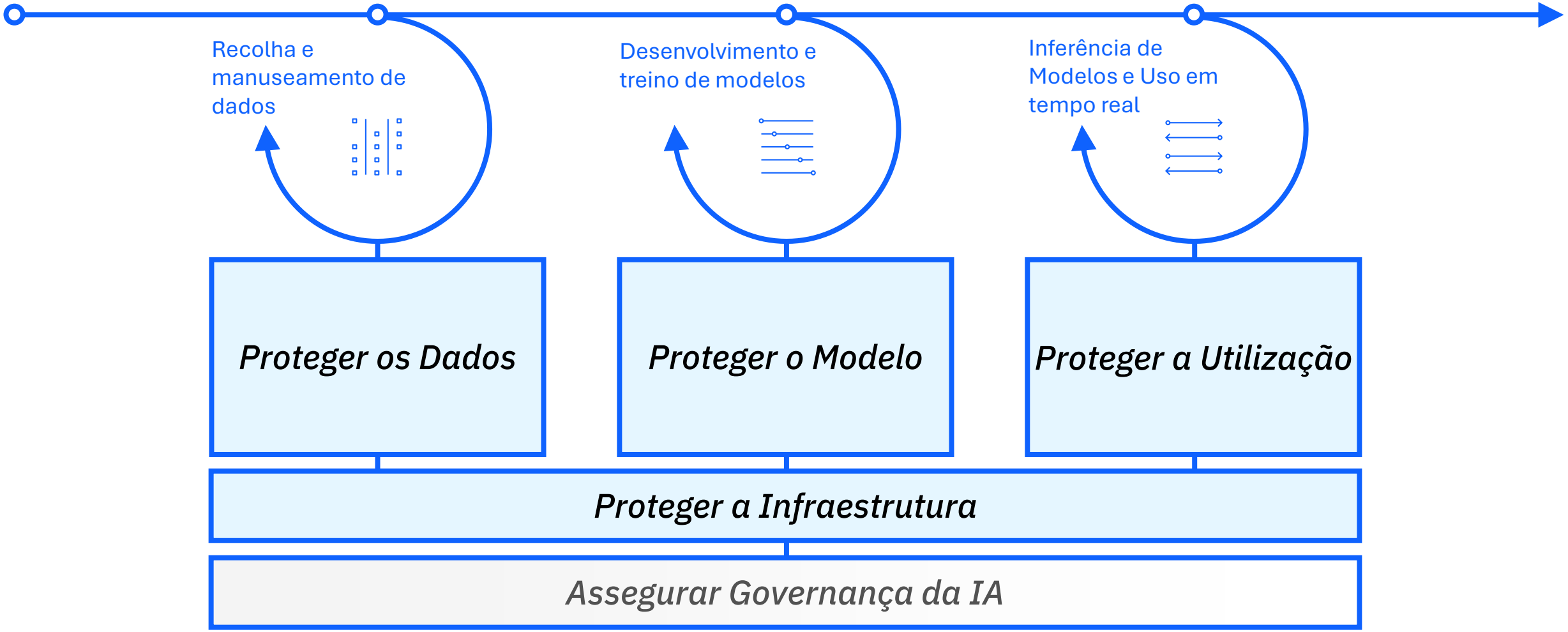
Segurança para IA
(IA como um problema)

Riscos adversariais no pipeline de IA



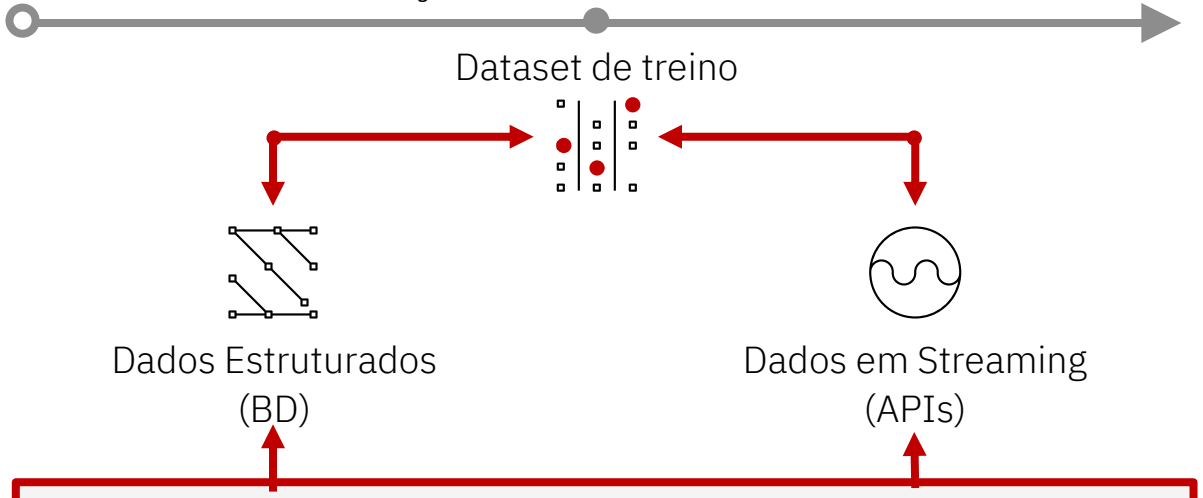
Ataques sobre...	Dados sensíveis que são centralizados e acedidos para treino	Novas vulnerabilidades dos modelos	Inferência de Modelo para fazer “hijack” ou manipulação do comportamento

Abordagem à Segurança da IA



Cibersegurança e a Revolução Tecnológica nas PME

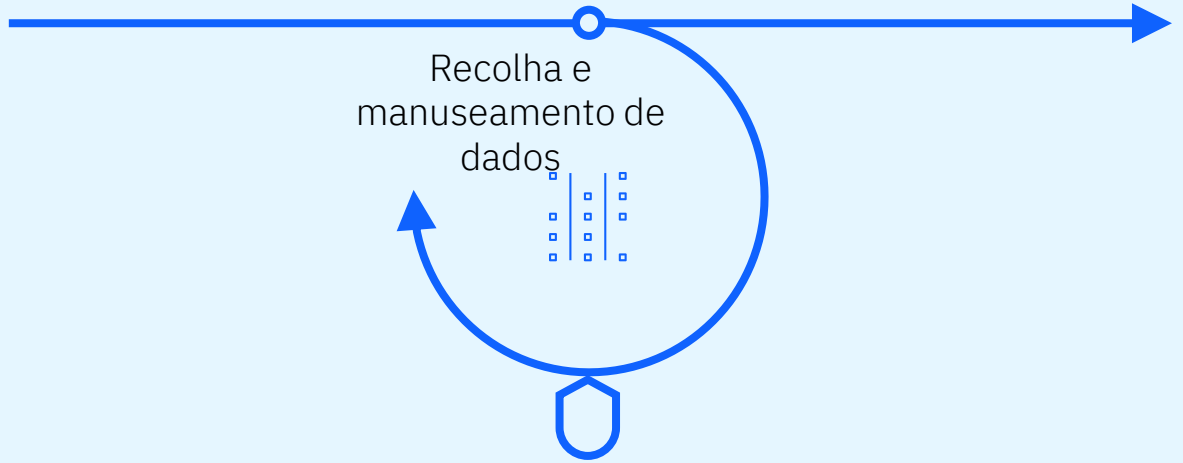
Riscos de coleção e manuseamento de dados



Atacantes irão atacar os datasets

- Os modelos de ML são data intensive e consomem volumes massivos de dados, incluíndo dados sensíveis
- Exfiltração pode resultar de uma vulnerabilidade técnica ou de controlos de acesso insuficientes
- Os atacantes podem explorar vulnrabilidades ou usar esquemas de phishing para ter acesso e roubar dados sensíveis usados no treino e afinação dos modelos de ML

Melhores práticas de coleção e manuseamento de dados

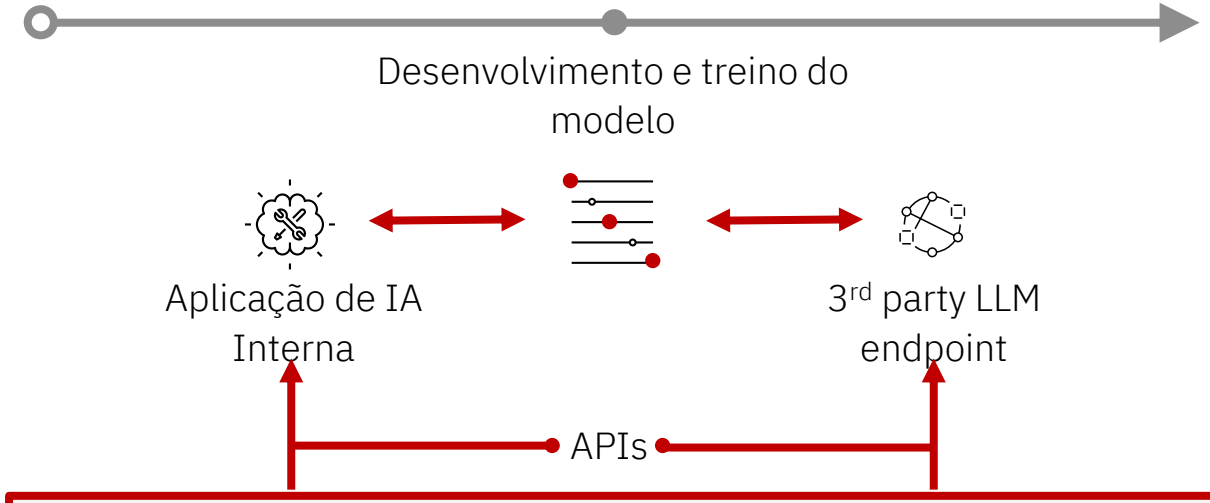


Proteger os dados

- Aplicar descoberta e classificação de dados para detetar dados sensíveis usados no treino ou afinação do modelo
- Implementar controlos de segurança de dados: cifra, controlo de acessos e monitorização continua com capacidade de resposta. Atenção aos utilizadores privilegiados
- Consciencializar os riscos de segurança em todas as etapas do pipeline de IA, e garantir que há elementos de segurança a trabalhar em proximidade dos cientistas de dados

Cibersegurança e a Revolução Tecnológica nas PME

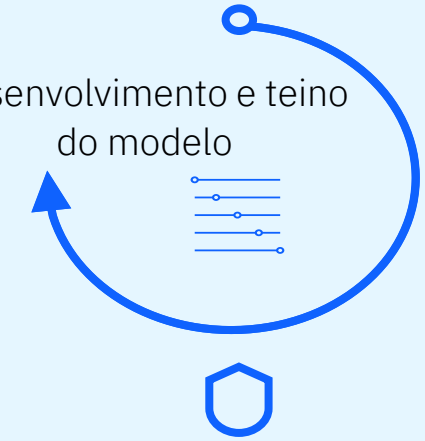
Riscos do desenvolvimento e treino do modelo



Atacantes explorarão vulnerabilidades e dependências

- **Ataques ao Supply chain:** Explorar vulnerabilidades em modelos open toolchains, libraries de terceiros, pacotes de software e outras dependências
- **Ataques às API** vulneráveis que transportem dados sensíveis e integrem ferramentas e aplicações
- **Privilege escalation:** explorar agentes de LLM ou plug-ins com privilégios excessivos no acesso a funções ou sistemas

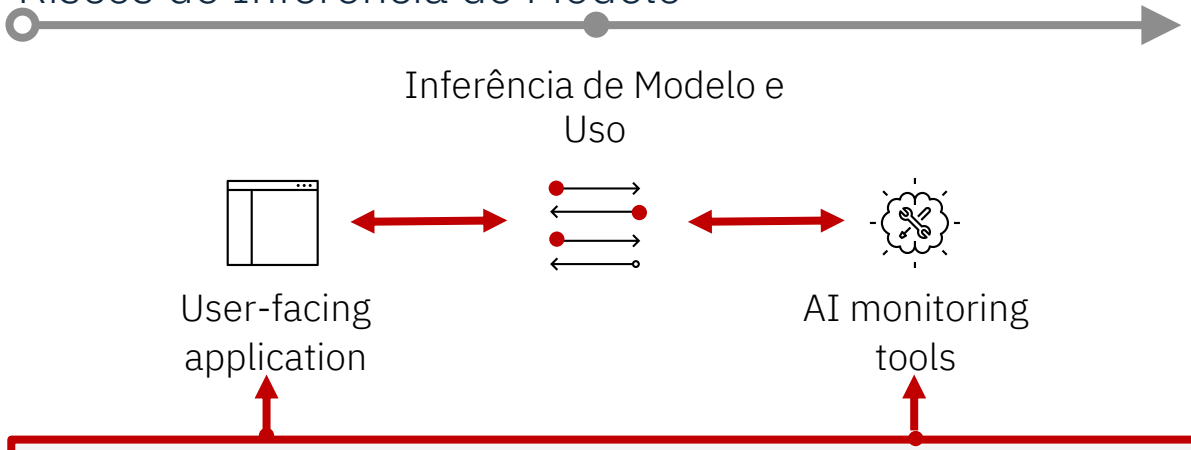
Melhores Práticas de desenvolvimento e treino do modelo



Proteger o Modelo

- Monitorização continua de vulnerabilidades, malware ou corrupção no pipeline de IA/ML
- Discobrir e fazer “harden” às API e integrações com modelos terceiros
- Configurar e fazer cumprir políticas, controlos de acesso e outros em modelos, artefactos e datasets

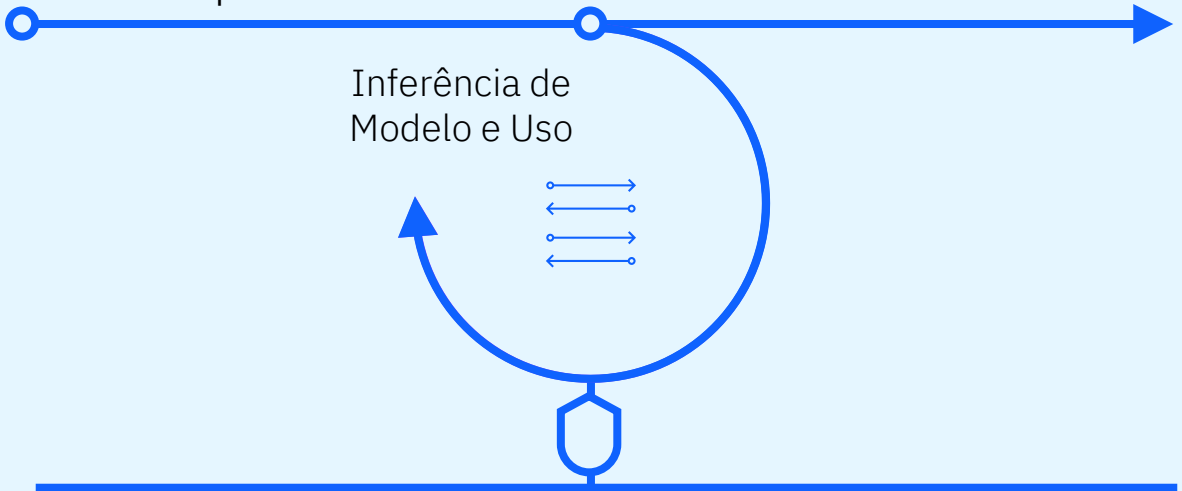
Riscos de Inferência de Modelo



Atacantes explorarão vulnerabilidades e dependências

- **Prompt injection:** Prompts maliciosas podem fazer jailbreak aos LLMs, permitindo acessos indevidos, roubar dados sensíveis ou gerar outputs adulterados / com bias
- **Model denial of service:** os atacantes podem sobrecarregar os LLM com inputs que degradem a qualidade do serviço e/ou aumentem substancialmente os custos
- **Model theft:** Os atacantes podem desenvolver inputs para obter outputs do modelo, acumulatndo um dataset de pares de input-output, de forma a treinar um outro modelo mimetizando o primeiro, e desta forma "roubando" as suas capacidades

Melhores práticas de Inferência de Modelo e Uso



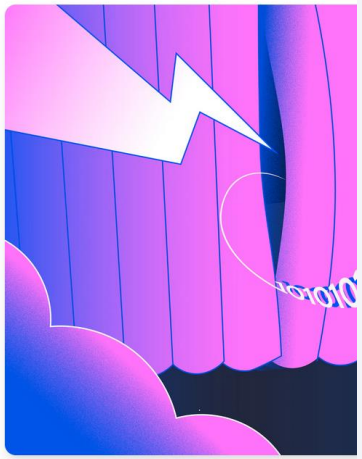
Proteger a Utilização

- Monitorizar inputs maliciosos como prompt injections, e outputs que contenham dados sensíveis ou conteúdo inapropriado
- Implementar soluções de Segurança de IA que possam detetar e responder a ataques específicos de IA (e.g., data poisoning, model evasion, model extraction)
- Desenvolver playbooks de resposta que neguem acesso, quarantena, e/ou desconectar modelos comprometidos

38TB of da
exposed b
rese

Wiz Research found a data exposure incid
30,000 internal Microsoft Teams mes

Hillai Ben-Sasson, Ronny Greenberg
September 18, 2023

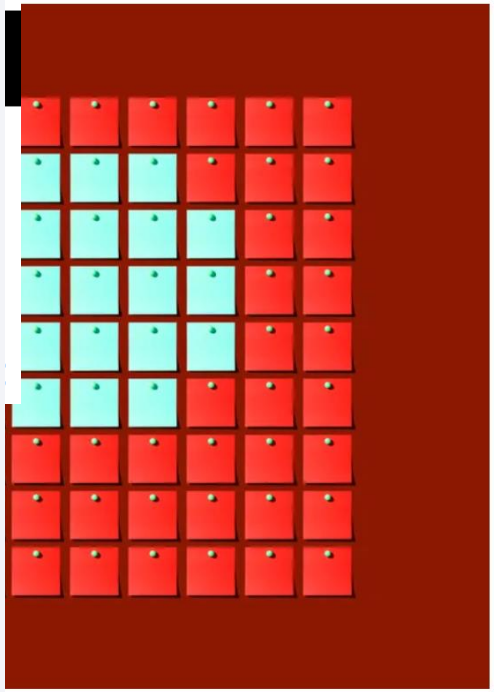


10 AI dangers and risks and how to manage them

https://www.ibm.com/blog/10-ai-dangers-and-risks-and-how-to-manage-them/

Major AI Chatbots—
v to Stop It

ChatGPT, Bard, and other
ard to tame.



IA para a Segurança (IA como Solução)

As operações de Segurança precisam de IA para melhorar a eficiência

49%

membros das equipas de SOC estão realmente a ver apenas ***metade dos alertas*** que deveriam ver num dia típico[1]

80%

das organizações usa pelo menos ***10 soluções diferentes*** para gerir a sua segurança[2]

2 em cada 3

organizações ***expandiram*** a sua superfície de ataque no último ano[2]

52%

dos ambientes de Segurança ficaram ***mais difíceis de gerir*** nos últimos 2 anos[3]

51%

ads organizações têm dificuldade em ***detetar e responder*** a ameaças avançadas[3]

29%

dos processos de operações de Segurança são ***imaturos*** e precisariam de reengenharia antes de ser automatizados[2]

Source: 1. Global Security Operations Center Study Results, IBM, March 2023.
Source: 2. The State of Attack Surface Management 2022, Randori.
Source: 3. ESG: SOC Modernization and the Role of XDR, 2022.

Cibersegurança e a Revolução Tecnológica nas PME

Àreas em que a IA é usada actualmente

Threat Detection and Response

Proceder a data mining de contexto, re-avaliar risco, e depois gerar um attack timeline mapeado com MITRE, com ações de resposta recomendadas

55%

Aumento de rapidez em triagem de alertas

Data Security

Monotorização do comportamento de utilizadores privilegiados no acesso a dados críticos

40%

Redução de violações de Segurança nos ambientes monitorizados2

Identity and Access Management

Risk-based authentication, pelo cálculo de um score de risco baseado em biometria, análise comportamental e histórico de acesso

15x

Redução na fricção de utilizadores3

[1] IBM Press Release, "IBM Launches New QRadar Security Suite to Speed Threat Detection and Response,"; [2] *The Total Economic Impact of IBM Security Guardium*, Forrester Report; [3] IBM Security Verify customer, leading U.S. insurance company

Interpretar Machine Data

Ajudar os analistas a ter uma menor dependência do entendimento sobre os Sistemas / aplicações protegidas, baixando os seus requisitos técnicos e acelerando a investigação

- Interpretação de Logs
- Diagnóstico
- Threat Intelligence

Geral Linguagem natural

Usar a IA para gerar texto baseado em eventos entendível por não técnicos

- Automatizar Reporting
- Acelerar Threat Hunting
- Aconselhar

Prever e remediar ataques com base em histórico de IOC

Notas Finais:

- O tamanho importa?
- Quantum
- Cidadania

Cibersegurança e a Revolução Tecnológica nas PME

Obrigado!

IAPMEI

academia PME