

IA em Campo de Batalha

Ataques Automatizados, Defesa Inteligente e o Futuro da Cibersegurança

Ramicés dos Santos
Diretor de Inteligência Artificial | SCTI | Governo do Estado de Santa Catarina
ramices.silva@scti.sc.gov.br

linkedin.com/in/ramices

Seminário de Engenharia de Software

UDESC Alto Vale | 2026

Apresentação



Ramicés dos Santos

Diretor de Inteligência Artificial

SCTI | Governo do Estado de Santa Catarina

- Inteligência artificial aplicada ao setor público
- Segurança da informação e governança digital
- Diálogo entre academia, mercado e governo

Agenda:

1 2 3 4 5

- Contexto e fundamentos
- IA ofensiva e automação de ataques
- Caso TARGET com agentes de IA
- IA defensiva e SOC
- Desenvolvimento seguro e futuro
- Debate

Reflexão



“A segurança não é um produto, é um processo.”

Bruce Schneier

Secrets and Lies: Digital Security in a Networked World, 2000

Por que o tema é urgente

US$ 4,44 mi

custo médio global de uma violação de dados em 2025

241 dias

tempo médio para identificar e conter uma violação

≈ 60%

das violações envolvem o elemento humano

30%

envolvem terceiros, o dobro do ano anterior

Superfície de ataque ampla, escassez de profissionais e volume de dados superior à capacidade humana.

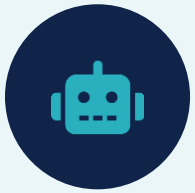
Fontes: IBM, Cost of a Data Breach Report 2025; Verizon, Data Breach Investigations Report 2025.

Conceitos fundamentais



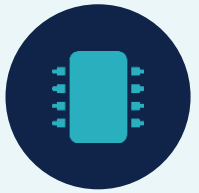
Inteligência artificial

Capacidade de sistemas computacionais de aprender, raciocinar e tomar decisões em tarefas que exigem inteligência humana.



Agente de IA

Sistema que percebe o ambiente, planeja e executa ações por meio de ferramentas, para cumprir um objetivo.



Autonomia computacional

Grau em que o sistema age sem intervenção humana a cada passo. Quanto maior, maior a necessidade de controles.

02

IA ofensiva: a automação do ataque

Como agentes inteligentes alteram a velocidade, a escala e o custo das ameaças

A natureza dual da inteligência artificial

Na defesa

- Detectar anomalias em grandes volumes de dados
- Priorizar alertas por risco
- Responder em segundos

No ataque

- Reconhecer alvos em larga escala
- Personalizar o engano para cada vítima
- Adaptar-se durante a ação

A mesma capacidade que acelera a defesa reduz o custo e o tempo do ataque.

Da estatística aos agentes: a evolução

1

Aprendizado de máquina

Filtros de spam e classificação de arquivos, em uso desde os anos 2000.

2

Aprendizado profundo

Reconhecimento de padrões em imagens, áudio e programas.

3

Modelos de linguagem

Geram texto e código a partir de instruções em linguagem natural.

4

Agentes de IA

Planejam, usam ferramentas e executam sequências de ações.

MITRE ATT&CK: o vocabulário comum dos ataques

Reconhecimento	**Desenvolvimento de recursos**	**Acesso inicial**	**Execução**	**Persistência**
Escalonamento de privilégios	**Evasão de defesa**	**Acesso a credenciais**	**Descoberta**	**Movimentação lateral**
Coleta	**Comando e controle**	**Exfiltração**	**Impacto**	

Em destaque: táticas observadas no caso TARGET. Cada etapa será analisada a seguir.

Reconhecimento automatizado

- Agentes navegam, leem e resumem fontes públicas continuamente
- Correlacionam organizações, fornecedores e pessoas
- Adaptam-se a mudanças nas páginas sem reprogramação
- Reduzem o tempo e a qualificação necessários



CONTRAPONTO DEFENSIVO

Reduzir a exposição de informações e monitorar a superfície pública da organização.

Engenharia social em escala: o phishing com modelos de linguagem

- Mensagens personalizadas, em bom português e no tom da vítima
- Milhares de abordagens com custo marginal baixo
- Resposta automática às reações das vítimas
- Autoridades já alertam para o uso de IA em ataques (Reuters, 2023)



CONTRAPONTO DEFENSIVO

Autenticação multifator resistente a phishing, treinamento contínuo e verificação por canal independente.

Deepfakes e fraudes de identidade

- Voz e vídeo sintéticos imitam pessoas conhecidas
- Casos documentados: pagamentos autorizados após falsa videoconferência com diretores
- A confiança em canais de comunicação passa a exigir verificação
- Impacto direto em processos de aprovação financeira



CONTRAPONTO DEFENSIVO

Dupla aprovação, confirmação por canal independente e procedimentos formais para pedidos urgentes.

Código assistido e análise de vulnerabilidades

- Assistentes explicam, adaptam e depuram código existente
- A barreira técnica para tarefas complexas diminui
- Falhas podem ser descobertas mais rápido, por ambos os lados
- Provedores adotam salvaguardas e monitoramento de uso



PONTO CENTRAL

O mesmo recurso serve a finalidades distintas: a autorização e o contexto definem o uso legítimo.

Ataques contra a própria inteligência artificial



Envenenamento de dados e modelos

Dados de treinamento adulterados alteram o comportamento do modelo de forma sutil.



Injeção de instruções

Conteúdo externo carrega comandos ocultos que o agente interpreta como legítimos.



Agência excessiva

O agente recebe permissões ou ferramentas além do necessário e causa danos amplos.

Fonte: OWASP, Top 10 for LLM Applications 2025 (LLM04, LLM01 e LLM06).

Model Context Protocol (MCP): conectando modelos a ferramentas



1

Modelo de linguagem

Interpreta o objetivo e planeja as ações.



2

Cliente MCP

Aplicação do agente que se conecta aos servidores.



3

Servidor MCP

Expõe funções descritas e suas permissões.



4

Sistemas e ferramentas

Executam a ação e devolvem o resultado.

Vantagem e risco: integração padronizada, porém o agente herda tudo o que o servidor permite.

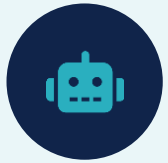
MCP e Kali Linux: testes de segurança autorizados com agentes



1

Analista responsável

Define escopo e autorização formal.



2

Agente de IA

Planeja e sequencia os testes.



3

Servidor MCP

Expõe categorias de ferramentas de teste.



4

Alvo autorizado

Laboratório ou sistema sob contrato.

Controles obrigatórios: escopo formal, aprovação humana, isolamento de rede e registro completo de auditoria.

Caso público de campanha conduzida majoritariamente por IA (2025)

≈ 30

alvos globais, entre empresas e órgãos governamentais

80–90%

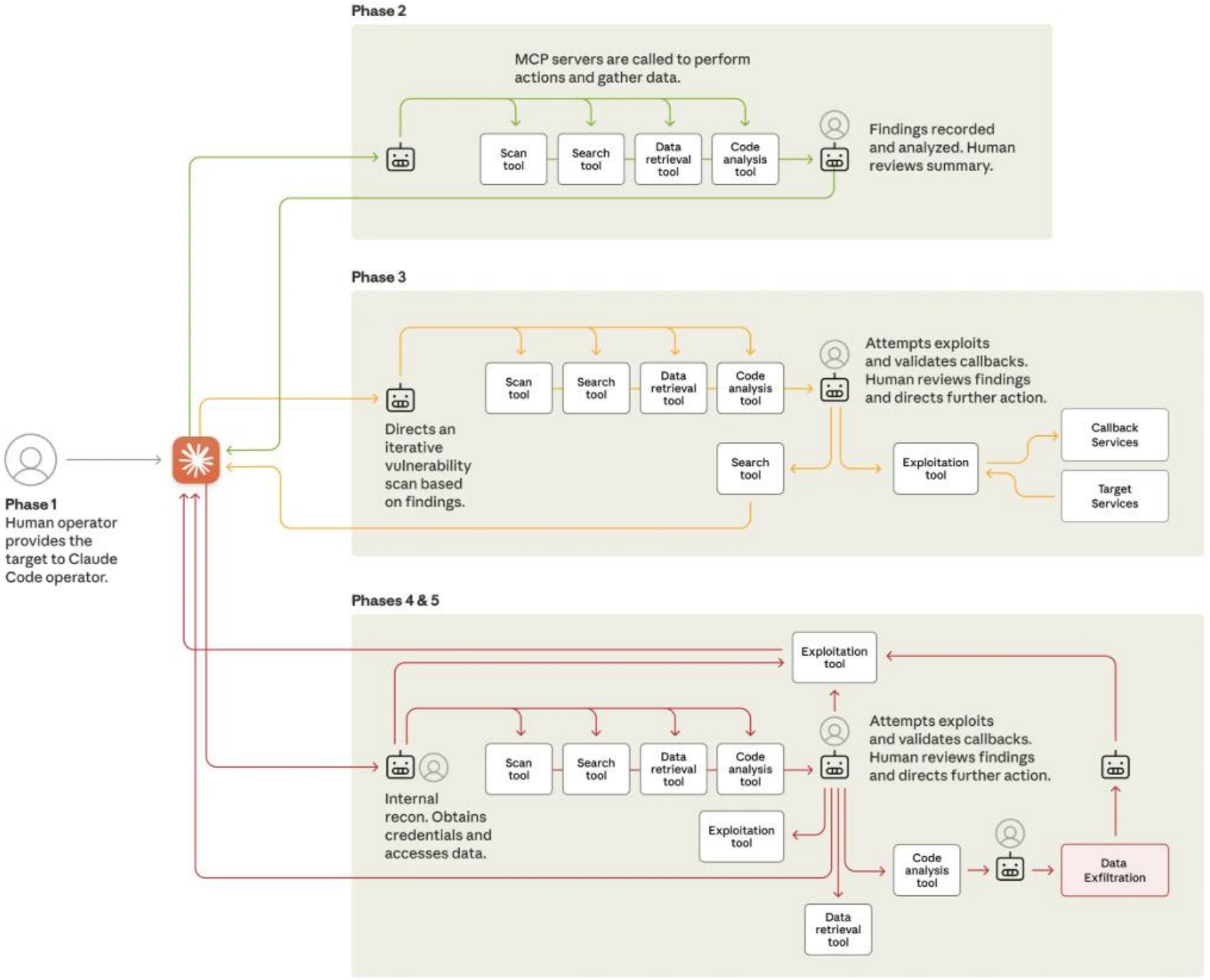
da campanha executada pelo agente de IA

4–6

pontos de decisão humana por campanha

Limitação observada: o agente por vezes inventou credenciais e exagerou achados.

Fonte: Anthropic, Disrupting the first reported AI-orchestrated cyber espionage campaign, 13/11/2025.



Fonte: Anthropic, Disrupting the first reported AI-orchestrated cyber espionage campaign, 13/11/2025.

03 Caso TARGET (2013) Como seria com agentes de IA

Por que estudar o caso TARGET



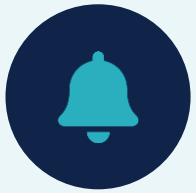
Documentação pública

Relatório do Senado dos EUA (2014) com análise da cadeia de ataque, etapa por etapa.



Risco de terceiros

A invasão começou em um fornecedor, e não na própria empresa.



Falha de resposta

A tecnologia detectou o ataque, mas a organização não agiu a tempo.

Dimensão do incidente

≈ 40 mi

cartões de pagamento comprometidos

até 110 mi

consumidores com dados expostos, incluindo registros pessoais

US$ 18,5 mi

acordo com 47 estados e o Distrito de Colúmbia (2017)

Fontes: Senado dos EUA, Comissão de Comércio (2014); NBC News (2017).

Linha do tempo do incidente



Fonte: Senado dos EUA, Comissão de Comércio, A Kill Chain Analysis of the 2013 Target Data Breach (2014).

Mapa do incidente: do fornecedor ao destino dos dados



Lacuna defensiva: os equipamentos de segurança geraram alertas, mas a Target os ignorou.

Etapa 1: reconhecimento e escolha do elo frágil

Etapa 1 de 8 | **MITRE ATT&CK: Reconhecimento**

O QUE OCORREU (2013)

- Escolha de um fornecedor de climatização com acesso remoto à rede
- Fornecedores costumam ter defesas menos maduras que a empresa principal
- Fontes públicas indicavam quais fornecedores se conectavam à empresa

AUTOMAÇÃO COM IA HOJE

- Agentes coletam e cruzam fontes públicas: contratos, portais, vagas e perfis
- Classificam fornecedores por exposição e provável baixa maturidade
- Repetem a coleta continuamente, sem esforço humano adicional

Contraponto defensivo: reduzir informações públicas sobre integrações e avaliar continuamente a segurança dos fornecedores.

Etapa 2: engenharia social contra o fornecedor

Etapa 2 de 8 | **MITRE ATT&CK: Acesso inicial**

O QUE OCORREU (2013)

- Provável phishing em setembro de 2013
- Credenciais de acesso à rede da Target foram roubadas
- O fornecedor usava antivírus gratuito, sem proteção em tempo real

AUTOMAÇÃO COM IA HOJE

- Modelos de linguagem redigem mensagens personalizadas e sem erros de idioma
- Agentes adaptam o texto ao cargo e ao perfil público de cada vítima
- A escala cresce sem aumento proporcional de custo

Contraponto defensivo: autenticação multifator resistente a phishing, filtros de mensagens e treinamento contínuo.

Etapa 3: uso de credenciais válidas no portal de fornecedores

Etapa 3 de 8

MITRE ATT&CK: Acesso a credenciais e contas válidas

O QUE OCORREU (2013)

- Com as credenciais, acessaram o sistema de cobrança e contratos
- Esse sistema não estava isolado da rede interna
- A atividade parecia legítima: usuário e senha corretos

AUTOMAÇÃO COM IA HOJE

- Agentes imitam o comportamento do usuário legítimo: horários e volume
- Regras com limites fixos perdem eficácia contra esse padrão
- A detecção passa a depender da análise de comportamento

Contraponto defensivo: privilégio mínimo, segmentação do acesso de terceiros e análise de comportamento de usuários.

Etapa 4: movimentação lateral e busca de servidores

Etapa 4 de 8

MITRE ATT&CK: Descoberta e movimentação lateral

O QUE OCORREU (2013)

- 12 de novembro: primeiro acesso à rede interna
- Deslocamento de áreas menos sensíveis para servidores com dados de clientes
- Indícios de segmentação insuficiente entre as redes

AUTOMAÇÃO COM IA HOJE

- O agente mapeia o ambiente e escolhe o próximo passo conforme o que observa
- Muda de rota diante de bloqueios, com pouca supervisão humana
- Ciclo contínuo de observação, decisão e ação

Contraponto defensivo: segmentação de redes, monitoramento do tráfego interno e detecção de varreduras.

Etapa 5: código malicioso nos caixas de pagamento

Etapa 5 de 8

MITRE ATT&CK: Execução e persistência

O QUE OCORREU (2013)

- De 15 a 30 de novembro, código malicioso foi instalado nos caixas
- Capturava dados dos cartões na memória, durante o processamento
- Variante de programa vendido em mercados clandestinos

AUTOMAÇÃO COM IA HOJE

- Assistentes de código reduzem a qualificação para adaptar programas existentes
- Provedores mantêm salvaguardas, mas o risco de uso indevido persiste
- A defesa observa o comportamento, e não a aparência do programa

Contraponto defensivo: lista de aplicações permitidas, monitoramento de integridade e criptografia ponta a ponta dos dados de pagamento.

Etapa 6: alertas gerados e ignorados

Etapa 6 de 8

MITRE ATT&CK: Detecção e resposta (lacuna defensiva)

O QUE OCORREU (2013)

- Em 30 de novembro, a ferramenta de segurança emitiu alertas urgentes
- A equipe não respondeu e a remoção automática estava desativada
- Novos alertas revelaram os destinos dos dados, também sem ação

AUTOMAÇÃO COM IA HOJE

- Ações automatizadas podem gerar volume de eventos capaz de saturar equipes
- Quanto maior o volume, maior a chance de perder o alerta crítico
- O desafio é de escala: a análise humana não acompanha a velocidade

Contraponto defensivo: triagem assistida por IA, priorização por risco e resposta automática sob supervisão humana.

Etapa 7: coleta e preparação dos dados

Etapa 7 de 8 | **MITRE ATT&CK: Coleta**

O QUE OCORREU (2013)

- Dados dos cartões reunidos em servidores internos intermediários
- Afetados os cartões usados nas lojas dos EUA de 27/11 a 18/12
- Resultado: cerca de 40 milhões de cartões comprometidos

AUTOMAÇÃO COM IA HOJE

- Agentes identificam e priorizam os dados de maior valor em grandes volumes
- Reconhecem dados pessoais e financeiros em formatos diversos
- Reduzem o tempo entre o acesso e a seleção do que será levado

Contraponto defensivo: minimização e classificação de dados, criptografia e alertas de acúmulo anômalo.

Etapa 8: exfiltração dos dados

Etapa 8 de 8 | **MITRE ATT&CK: Exfiltração**

O QUE OCORREU (2013)

- A partir de 2 de dezembro, dados foram enviados a servidores externos
- Cerca de 11 GB em aproximadamente duas semanas, em horário comercial
- As ferramentas identificaram os destinos, mas não houve ação

AUTOMAÇÃO COM IA HOJE

- Agentes ajustam volume, horário e rota para se confundir com o tráfego legítimo
- Alternam destinos e canais diante de bloqueios
- A detecção exige referência de comportamento normal e controle de saída

Contraponto defensivo: filtragem do tráfego de saída, detecção de volumes anômalos e bloqueio automático de destinos suspeitos.

Arquitetura agêntica do cenário



1

Operador

Define objetivo e limites.



2

Agente planejador

Decompõe o objetivo em etapas.



3

Servidores MCP

Conectam o agente às ferramentas.



4

Ferramentas

Executam as ações no ambiente.



5

Memória e avaliação

Registram achados e ajustam o plano.

Ciclo contínuo: observar, planejar, agir e avaliar, com poucos pontos de decisão humana.

O que muda e o que permanece com agentes de IA

O que muda

- Velocidade de execução
- Escala: vários alvos simultâneos
- Custo e qualificação exigidos
- Adaptação em tempo real

O que permanece

- Necessidade de um acesso inicial
- Dependência de falhas humanas e de configuração
- Erros e invenções do modelo
- Rastros passíveis de detecção

Falhas de governança evidenciadas pelo caso



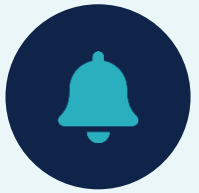
Gestão de terceiros

Fornecedor com acesso remoto e controles frágeis, sem isolamento adequado do portal.



Segmentação de rede

Deslocamento fácil de áreas menos sensíveis até os sistemas de pagamento.



Monitoramento sem resposta

Alertas urgentes sem tratamento e bloqueio automático desativado.

Controles que teriam interrompido o ataque

Etapa	Controle	Referência
Acesso do fornecedor	Autenticação multifator e privilégio mínimo	NIST CSF 2.0 (Proteger)
Rede	Segmentação entre portal de fornecedores e pagamentos	PCI DSS
Caixas de pagamento	Lista de aplicações permitidas e integridade	CIS Controls
Alertas	Triagem contínua, resposta documentada e bloqueio	NIST CSF 2.0 (Detectar, Responder)
Saída de dados	Controle de tráfego de saída e anomalias	MITRE ATT&CK (Exfiltração)

Lições do caso TARGET

- Terceiros fazem parte da superfície de ataque
- Detectar não basta: é preciso responder
- A segmentação limita a extensão do dano
- A IA reduz o tempo entre o acesso e o dano
- A defesa também precisa operar em velocidade de máquina



PARA DEBATE

Se os alertas do caso fossem tratados por um agente de IA, o desfecho seria diferente? Sob quais condições?

04

IA defensiva: automação da proteção

SOC automatizado, MCP com Wazuh e casos práticos em projetos de software

SEMESO 2026
Seminário de Engenharia de Software | UDESC Alto Vale

Evidências do benefício da automação defensiva

US$ 1,9 mi

economia média por violação com uso extensivo de IA e automação

80 dias

redução do ciclo de vida da violação nessas organizações

97%

das organizações com violação não tinham controles de acesso adequados

A automação reduz custo e tempo, mas exige governança e controles de acesso sobre a própria IA.

Fonte: IBM, Cost of a Data Breach Report 2025.

O Centro de Operações de Segurança e seus gargalos

- O SOC monitora eventos de segurança de forma contínua
- O volume de alertas supera a capacidade de análise
- Tarefas repetitivas de triagem causam fadiga e rotatividade
- A escassez de profissionais agrava o problema



CONSEQUÊNCIA

Ameaças reais ficam sem análise, como ocorreu no caso TARGET.

SOC automatizado: do evento à resposta



1

Coleta

Agentes e registros dos sistemas.



2

Correlação

SIEM relaciona eventos de várias fontes.



3

Triagem por IA

Prioriza e explica cada alerta.



4

Orquestração

SOAR executa fluxos de resposta.



5

Decisão humana

Aprova ações de maior impacto.

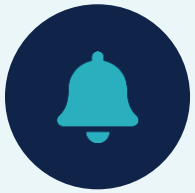
Siglas: SIEM, gestão de eventos e informações de segurança; SOAR, orquestração, automação e resposta.

Exemplo - Wazuh: plataforma aberta de monitoramento de segurança



Agentes

Coletam eventos, processos e portas em servidores e estações.



Regras e alertas

Detectam comportamentos suspeitos e verificam conformidade.



Vulnerabilidades

Inventariam falhas conhecidas e monitoram a integridade de arquivos.

MCP para Wazuh: o analista conversa com o SOC

1

Analista

Pergunta em linguagem natural.

2

Agente de IA

Interpreta e planeja as consultas.

3

Servidor MCP Wazuh

Expõe funções de consulta.

4

API do Wazuh

Agentes, alertas e vulnerabilidades.

Princípio: começar com funções de leitura; ações de resposta somente com aprovação humana.

IA em ação!!!

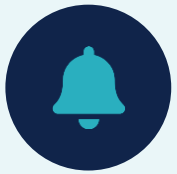

VULNERABILIDADES
40.669
+58% vs média 25.665
CRÍTICAS
3.169
-4% vs média 3.289
ALTAS
24.246
+73% vs média 14.019
ENDPOINTS
151
+10% 84 ativos · 63 offline
AGENTES OFFLINE
63
+80%
vs média 35 · domingo
CIS WIN 11
23%
0 pp meta ≥ 70%
EVOLUÇÃO & COMPARATIVO COM A MÉDIA DE 2026
Vulnerabilidades detectadas no ambiente
Total por snapshot mensal · fonte: índice wazuh-states-vulnerabilities
50k
40k
30k
20k
10k
0k
média 25.665
Fev
Mar
Abr
Mai
Jun
16/set
30/set
05/out
Atual vs. média de 2026
Barra = valor atual (05/out) · marca vertical = média 2026
Total
todas severidades
40.669
+58%
Críticas
3.169
-4%
Altas
24.246
+73%
Médias
12.337
+57%
Baixas
915
+80%
Offline
agentes
63
+80%
Atual (05/out)
Média 2026
Principais achados
Leitura executiva do cenário atual
CRÍTICO
Superfície estável em patamar alto
40.669 vulnerabilidades ativas (+58% vs média de 2026), praticamente igual a 30/set (+0,2%). A pressão está em Altas
(+73%) e Médias (+57%); as Críticas seguem estáveis (-4%) — o salto recente veio de novas divulgações de CVE, não de
Servidores
CVEs críticos (RCE) por servidor
srv-dc02
Controlador de Domínio · Server 2022
0
CRÍTICOS

Funções disponibilizadas na arquitetura de demonstração



Inventário de agentes

Situação dos agentes, processos e portas abertas.



Alertas

Resumos e estatísticas semanais de alertas.



Vulnerabilidades

Resumo geral e lista das vulnerabilidades críticas.



Saúde do cluster

Estado dos nós e do conjunto de serviços.



Regras e registros

Resumo de regras e busca nos registros do gerenciador.



Diagnóstico

Erros do gerenciador e estatísticas de coleta de logs.

Exemplo conceitual: triagem assistida de um alerta



1

Alerta recebido

Evento de gravidade média ou alta.



2

Contexto

Consulta ativo, processos e portas.



3

Classificação

Hipótese e gravidade, com justificativa.



4

Recomendação

Ação sugerida e evidências.



5

Decisão humana

Aprova, ajusta ou descarta, e registra.

Níveis de autonomia e supervisão humana

Nível 0: observação		Somente leitura e relatórios.
Nível 1: recomendação		O agente sugere; a pessoa decide e executa.
Nível 2: ação aprovada		O agente executa após aprovação humana.
Nível 3: ação automática limitada		Apenas ações reversíveis e de baixo risco.
Nível 4: autonomia ampla		Não recomendada para ações irreversíveis.

Caso prático 1: segurança contínua em um projeto de software



Papel da IA: explicar achados, sugerir correções e priorizar por risco, com revisão humana.

Caso prático 2: gestão de vulnerabilidades e dependências

- Inventário dos componentes do software (SBOM)
- Priorização por exploração real e criticidade do ativo
- A IA sugere a correção e abre solicitação de alteração
- Revisão humana antes da integração ao código



INDICADOR

Tempo médio de correção de vulnerabilidades críticas.

Caso prático 3: resposta a incidentes em um projeto

Sem automação

- Triagem manual de centenas de alertas
- Correlação feita em planilhas
- Relato escrito somente ao final

Com automação assistida

- Alertas enriquecidos automaticamente
- Hipóteses apresentadas com evidências
- Relatório preliminar gerado e revisado

Indicadores de acompanhamento: tempo de detecção, tempo de resposta e taxa de falsos positivos.

05

Desenvolvimento seguro na era da IA

Os desafios da Engenharia de Software e os novos conceitos

SEMESO 2026
Seminário de Engenharia de Software | UDESC Alto Vale

Desafios da Engenharia de Software com código gerado por IA

≈ 45%

do código gerado por IA apresentou falhas de segurança (mais de 100 modelos)

5,2–21,7%

de pacotes sugeridos não existem, em modelos comerciais e abertos

86%

de falha na proteção contra injeção de scripts em páginas web

Código plausível não é código seguro: revisão e testes continuam indispensáveis.

Fontes: Veracode, GenAI Code Security Report 2025; Spracklen et al., USENIX Security 2025.

Segurança desde a concepção



Secure by Design

A segurança é requisito do projeto, e não complemento posterior.



Ciclo seguro (NIST SSDF)

SP 800-218 e o perfil para IA generativa, SP 800-218A (2024).



Cadeia de suprimentos

Lista de componentes (SBOM), assinatura e verificação de dependências.

OWASP Top 10 para aplicações com modelos de linguagem (2025)

1. Injeção de instruções
2. Divulgação de informações sensíveis
3. Cadeia de suprimentos
4. Envenenamento de dados e modelos
5. Tratamento inadequado de saídas
6. Agência excessiva
7. Vazamento do prompt de sistema
8. Fragilidades em vetores e embeddings
9. Desinformação
10. Consumo ilimitado de recursos

Fonte: OWASP, Top 10 for LLM Applications 2025. Em destaque: itens mais críticos para agentes.

Segurança de agentes e de servidores MCP



Privilégio mínimo

Credenciais com escopo restrito e prazo curto.



Isolamento

Execução em ambientes controlados, com rede restrita.



Aprovação humana

Exigida para ações irreversíveis ou de alto impacto.



Auditoria

Registro integral das ações, entradas e decisões.

Todo conteúdo externo lido pelo agente deve ser tratado como não confiável.

Governança, normas e conformidade

Marco	Contribuição
ISO/IEC 27001	Sistema de gestão da segurança da informação
ISO/IEC 42001:2023	Sistema de gestão para inteligência artificial
NIST AI RMF 1.0 (2023)	Gestão de riscos de sistemas de IA
NIST CSF 2.0 (2024)	Funções: governar, identificar, proteger, detectar, responder e recuperar
LGPD (Lei 13.709/2018)	Proteção de dados pessoais, inclusive em sistemas de IA
PL 2.338/2023	Marco legal da IA, em tramitação na Câmara dos Deputados

Novas práticas para sistemas com agentes

- Modelagem de ameaças que inclua o agente e suas ferramentas
- Testes adversariais de modelos e agentes (red teaming)
- Avaliação contínua de segurança, com testes de regressão
- Observabilidade: rastrear decisões e chamadas de ferramentas
- Supervisão humana definida desde o projeto

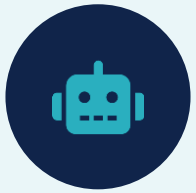


REFERÊNCIA

Catálogos públicos de ameaças à IA, como o MITRE ATLAS, apoiam a modelagem.

DFD - STRIDE

O futuro: segurança na era da autonomia computacional



Agentes contra agentes

Ataque e defesa operando em velocidade de máquina, com supervisão humana estratégica.



Segurança como propriedade

Engenharia responsável: requisitos, arquitetura, testes e operação seguros.



Formação profissional

Fundamentos sólidos, pensamento crítico, ética e aprendizado contínuo.

Referências

- SENADO DOS EUA, Comissão de Comércio. A “Kill Chain” Analysis of the 2013 Target Data Breach. 2014.
- NBC NEWS. Target settles 2013 hacked customer data breach for $18.5 million. 2017.
- ANTHROPIC. Disrupting the first reported AI-orchestrated cyber espionage campaign. 13 nov. 2025.
- IBM. Cost of a Data Breach Report 2025.
- VERIZON. 2025 Data Breach Investigations Report.
- VERACODE. 2025 GenAI Code Security Report.
- SPRACKLEN, J. et al. We Have a Package for You! USENIX Security, 2025 (arXiv:2406.10279).
- OWASP. Top 10 for LLM Applications 2025.
- NIST. SP 800-218A (2024); Cybersecurity Framework 2.0 (2024); AI RMF 1.0 (2023).
- MITRE. ATT&CK Enterprise Matrix; ATLAS.
- ISO/IEC 42001:2023; ISO/IEC 27001.
- BRASIL. Lei 13.709/2018 (LGPD); PL 2.338/2023.
- REUTERS. AI being used for hacking and misinfo, top Canadian cyber official says. 20 jul. 2023.
- SCHNEIER, B. Secrets and Lies: Digital Security in a Networked World. 2000.



Obrigado!

Perguntas e debate

Ramicés dos Santos
Diretor de Inteligência Artificial | SCTI | Governo do Estado de Santa Catarina

SEMESO 2026
Seminário de Engenharia de Software | UDESC Alto Vale