Número 113 | Abril 2026

NTT DATA

Radar

A revista de cibersegurança

IA e cibersegurança: a nova guerra entre algoritmos

Por Pedro Felipe del Jesús Canto Vela

Nunca tivemos tanta capacidade tecnológica para nos proteger — nem tanta para atacar. A inteligência artificial passou a ocupar um papel central na cibersegurança, redefinindo o equilíbrio entre a defesa e o risco. Os mesmos algoritmos que detectam ameaças em milissegundos também podem ser utilizados para lançar ataques mais rápidos, precisos e difíceis de rastrear. Nesse contexto, a automação deixa de ser apenas uma vantagem competitiva e passa a representar um novo campo de batalha.

Durante anos, a segurança esteve concentrada no fortalecimento das infraestruturas e na melhoria da resposta a incidentes. No entanto, a IA, especialmente a IA generativa e os modelos avançados de machine learning, transformou esse cenário. As organizações não enfrentam apenas ameaças mais sofisticadas, mas operam em um ambiente no qual a velocidade de adaptação se torna determinante. A IA acelera tanto o ataque quanto a defesa, exigindo que as organizações revisem suas estratégias para além da simples adoção de tecnologia.

No cenário ofensivo, a inteligência artificial reduziu significativamente a barreira de entrada para o cibercrime. As campanhas de *phishing* são personalizadas em tempo real, os *deepfakes* viabilizam falsificações de identidade altamente convincentes e o *malware* passa a se adaptar dinamicamente para contornar controles. A engenharia social passa a contar com modelos capazes de gerar mensagens contextualizadas e persuasivas. Ainda assim, o principal fator de transformação não é apenas a sofisticação, mas a velocidade. Os invasores conseguem testar variações e ajustar táticas em questão de horas, enquanto muitas organizações ainda operam com processos concebidos para um contexto muito mais lento.

A democratização dessas tecnologias amplia a superfície de ataque e dificulta a atribuição. Em um ambiente sem perímetro claramente definido, com infraestruturas em nuvem e cadeias de suprimentos digitais, surgem novos riscos, como a exposição de dados em modelos generativos, o *prompt injection* e o envenenamento de modelos. O risco deixa de estar restrito à infraestrutura e passa a envolver diretamente os algoritmos e os dados que alimentam esses sistemas.

Ao mesmo tempo, a IA se consolida como um pilar essencial da defesa moderna. A capacidade de analisar grandes volumes de dados em tempo real permite detectar anomalias, priorizar alertas e automatizar respostas, reduzindo significativamente os tempos de reação. A análise preditiva, por sua vez, contribui para antecipar ameaças antes que se tornem incidentes críticos, fortalecendo uma postura de segurança mais proativa.

A vantagem competitiva não está na tecnologia isoladamente, mas na forma como ela é integrada a um marco sólido de governança e gestão de riscos. A inteligência artificial amplia capacidades, mas não substitui a estratégia nem o julgamento humano. Sem a supervisão adequada e controles eficazes, o uso da IA pode resultar em decisões desalinhadas ou em uma falsa percepção de segurança.

Na era da automação, a verdadeira vantagem não está em adotar a IA mais rapidamente do que os invasores, mas em incorporá-la de forma consciente a um modelo de resiliência. Porque, mesmo quando os algoritmos aceleram o jogo, a responsabilidade de proteger o futuro continua sendo essencialmente humana.

Pedro Felipe del Jesús Canto Vela
Cybersecurity Expert Analyst

Quando a inteligência artificial aprende a enganar… e a operar

Cibercrônica por Juan Pablo Camperos

Entre 2024 e o início de 2026, a inteligência artificial deixou de ser apenas uma ferramenta de eficiência para se tornar um fator de risco operacional. Não se trata do surgimento de ameaças completamente novas, mas da capacidade da inteligência artificial de ampliar ameaças já existentes em uma velocidade que as organizações ainda não conseguem governar.

Um dos exemplos mais visíveis dessa evolução é o uso de *deepfakes* em fraudes corporativas. Na Ásia, um colaborador participou de uma videoconferência com pessoas que aparentavam ser executivos da própria organização. As vozes, os gestos e o contexto eram coerentes. A instrução foi direta: executar transferências urgentes. Não houve exploração técnica nem uso de *malware*. Apenas a exploração indevida da confiança. A identidade visual deixou de ser suficiente.

Um caso semelhante em Singapura reforça essa tendência. Um diretor financeiro foi induzido a transferir uma quantia significativa após interagir em uma reunião executiva inteiramente gerada por IA. O golpe só foi identificado quando os invasores tentaram escalar a operação. Em ambos os casos, a tecnologia não atacou o sistema — atacou a percepção.

Essa mudança redefine o ponto de entrada. A engenharia social deixa de depender exclusivamente da persuasão humana e passa a se apoiar na capacidade de replicar identidades com alta precisão. A questão deixa de ser se a mensagem é convincente e passa a ser se existe um processo estruturado para questioná-la.

Em paralelo, outro vetor passou a evidenciar riscos mais estruturais. A Microsoft iniciou ações legais contra um grupo que utilizava credenciais comprometidas para acessar serviços de IA generativa. O objetivo não era a exfiltração de dados, mas a exploração da infraestrutura para gerar conteúdo malicioso e comercializar seu uso. O problema não estava no modelo em si, mas no controle de acessos.

Ao mesmo tempo, a cadeia de suprimentos da IA começou a apresentar fragilidades semelhantes às do *software* tradicional. Pesquisas recentes identificaram modelos maliciosos publicados em repositórios abertos, projetados para executar código ao serem integrados a ambientes de desenvolvimento. Os modelos deixam de ser dados passivos e passam a atuar como componentes ativos dentro do sistema.

A isso se soma um elemento que muitas organizações ainda subestimam: a gestão da informação. Na Coreia do Sul, autoridades identificaram que uma plataforma de IA havia transferido dados de usuários e conteúdos de *prompts* sem consentimento. Esse incidente evidenciou um ponto crítico: os *prompts* contêm contexto de negócio, decisões estratégicas e informações sensíveis. Tratá-los como texto descartável, na prática, equivale a um vazamento de dados.

No entanto, a mudança mais relevante ocorre quando esses padrões atingem a operação. Na Amazon, uma série de incidentes ao longo de 2026 evidenciou como o uso de assistentes de código e de automação baseada em IA pode impactar diretamente sistemas produtivos. Alterações executadas em alta velocidade, sem validações adequadas e com controles insuficientes resultaram em interrupções e perdas operacionais significativas.

Esse tipo de situação evidencia a convergência entre TI e TO. Quando a inteligência artificial deixa de apenas apoiar e passa a influenciar a execução, o impacto deixa de ser exclusivamente digital. Um erro pode escalar de uma configuração isolada até comprometer a continuidade de um serviço completo.

E esse fenômeno não se limita a ambientes digitais. Em sistemas de abastecimento de água, já foram registrados acessos não autorizados nos quais parâmetros operacionais foram alterados. No setor de petróleo e gás, a manipulação de sensores gerou alarmes falsos e decisões equivocadas. Nesses casos, a IA não é a origem direta do ataque, mas atua como um acelerador ao reduzir o esforço necessário para sua execução.

Mais preocupante ainda é o fato de que o acesso nem sempre ocorre por meio da exploração de sistemas. Casos recentes mostram como agentes maliciosos utilizam IA para construir identidades plausíveis e atravessar processos de recrutamento. Uma vez dentro da organização, o invasor não depende mais de vulnerabilidades — passa a operar com acesso legítimo. Em ambientes nos quais sistemas industriais são conectados ou gerenciados remotamente, esse tipo de infiltração representa um risco estrutural.

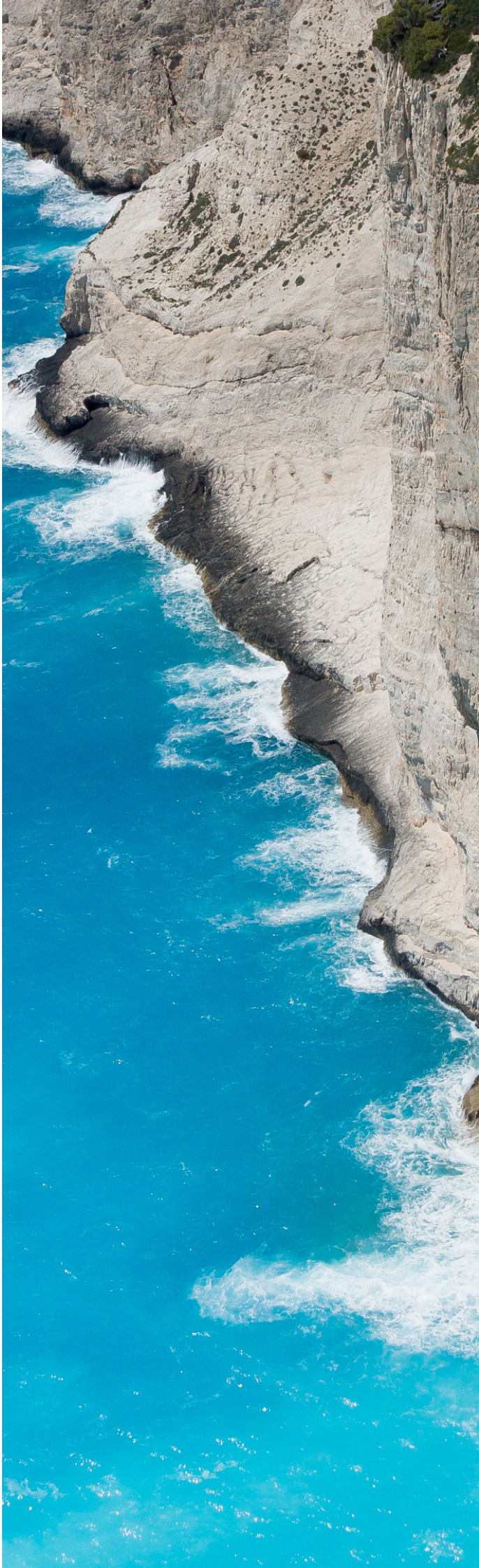
Em conjunto, esses incidentes revelam uma evolução clara. A identidade deixa de ser confiável pela aparência, os acessos continuam sendo o ponto mais vulnerável, a cadeia de suprimentos passa a incluir modelos e a automação amplifica qualquer erro. A IA não rompe sistemas — ela acelera suas falhas.

Quando esses padrões atingem ambientes operacionais, o impacto muda de natureza. Uma fraude pode interromper pagamentos a fornecedores críticos, uma credencial comprometida pode abrir acesso a sistemas industriais e uma alteração automatizada mal controlada pode comprometer processos produtivos inteiros. A cibersegurança deixa de ser apenas um problema técnico e passa a ser um fator crítico de continuidade operacional.

A resposta não está em limitar o uso da inteligência artificial, mas em estabelecer mecanismos eficazes de governança. As organizações que avançam não são necessariamente as que mais investem em IA, mas as que implementam controles claros sobre o seu uso — e, sobretudo, sobre o seu impacto.

Porque, nesse novo cenário, a pergunta já não é se a IA pode ser utilizada em um ataque, mas se as organizações estão preparadas para operar com ela sem perder o controle.

Juan Pablo Camperos
Cybersecurity Expert Architect



Além da ilusão: desinformação e deepfakes na era da IA

Artigo por Prakash Narayanamoorthy e Ben Colman*

**Este artigo foi publicado originalmente com o título "Beyond the illusion: Disinformation and deepfakes in the age of AI" na revista Cyber Frontiers, da NTT DATA, edição de janeiro de 2026. A reprodução deste conteúdo foi autorizada previamente.*

À medida que a IA torna mais tênue a linha entre o real e o falso, as organizações passam a enfrentar uma nova onda de manipulação digital que ameaça a confiança, a segurança e a própria percepção do que é verdadeiro.

A confiança das pessoas é a base de muitas interações digitais essenciais com organizações e órgãos governamentais. No entanto, essa confiança está sob ameaça, já que a IA facilita a criação de *deepfakes* capazes de reproduzir a aparência e a voz de pessoas reais com precisão impressionante. O que antes exigia estúdios especializados e recursos significativos agora está ao alcance de agentes maliciosos com recursos computacionais comuns e conhecimento técnico limitado.

O que estamos vendo é apenas a ponta do iceberg, já que a dimensão do problema cresce de forma exponencial.

Para órgãos governamentais, as implicações em termos de segurança nacional são profundas. O relatório de Riscos Globais de 2024, do Fórum Econômico Mundial, classifica a desinformação e a informação incorreta impulsionadas por IA como a principal ameaça que o mundo enfrentará nos próximos dois anos.

Sob a perspectiva da cibersegurança, os *deepfakes* representam um novo vetor de ameaça.
Os modelos tradicionais de segurança, historicamente centrados no controle de acesso e na proteção de dados, não foram concebidos para detectar manipulações baseadas em conteúdo. Esse cenário cria um ponto cego nos centros de operações de segurança, deixando as organizações vulneráveis a uma nova geração de ataques que tem como alvo a confiança humana, e não os sistemas técnicos.

A importância da cibersegurança

A detecção de deepfakes tornou-se um requisito fundamental para a cibersegurança. A convergência entre a falsificação de identidade viabilizada por IA e os vetores tradicionais de ataque resulta em ameaças compostas, capazes de contornar os mecanismos convencionais de proteção:

- Em videoconferências, a imagem e a identidade de executivos podem ser falsificadas para autorizar transferências fraudulentas, gerando riscos financeiros relevantes.
- A clonagem de voz vem sendo utilizada para contornar os sistemas de autenticação biométrica, comprometendo mecanismos que até pouco tempo eram considerados confiáveis.
- O uso de mídia sintética viabiliza campanhas de *phishing* altamente sofisticadas, praticamente indistinguíveis das comunicações legítimas.
- A manipulação de evidências passa a comprometer processos legais e regulatórios, afetando a integridade do sistema de justiça.
- Campanhas coordenadas de desinformação direcionadas a infraestruturas críticas levantam preocupações de segurança nacional ao enfraquecer a confiança pública e potencialmente interromper serviços essenciais.

Essas ameaças exigem capacidades especializadas de detecção, integradas diretamente aos fluxos de trabalho das operações de segurança. Sem proteção em tempo real contra a mídia sintética, até mesmo estruturas robustas de cibersegurança deixam lacunas relevantes de proteção.

Manipulação sintética: impacto em setores críticos

Os *deepfakes* representam uma ameaça significativa em setores de alto risco. No setor de serviços financeiros, invasores utilizam áudio e vídeo gerados por IA para se passar por clientes durante interações com centrais de atendimento, contornando processos de verificação de identidade e iniciando transações fraudulentas. Essas ações exploram vulnerabilidades tanto na validação digital quanto nos sistemas de autenticação baseados em voz.

No setor público, órgãos governamentais tornam-se alvos de falsificações envolvendo autoridades e informações manipuladas, o que amplia os riscos para a segurança nacional e compromete a confiança institucional.

Infraestruturas críticas, como energia, saúde e serviços de emergência, também são vulneráveis a campanhas de desinformação baseadas em *deepfakes*. Esses ataques podem simular comunicações de crise, comprometer a continuidade operacional e gerar confusão em situações de emergência. No setor de aviação, agentes maliciosos podem se passar por pilotos, controladores de tráfego aéreo ou executivos de companhias aéreas por meio de conteúdos sintéticos, o que pode provocar atrasos em voos e riscos à segurança.

À medida que essas ameaças evoluem, torna-se essencial que as organizações invistam em detecção de *deepfakes*, protocolos seguros de comunicação e inteligência de ameaças intersetorial, com o objetivo de preservar a confiança e fortalecer a resiliência.

Combater a ameaça dos *deepfakes*: uma estratégia de defesa multimodal

Diante de ataques que combinam texto, áudio, imagem e vídeo para criar fraudes altamente convincentes, as estratégias de defesa precisam adotar uma abordagem igualmente integrada. Considere o seguinte cenário, um executivo recebe uma ligação durante a madrugada. A voz é familiar e convincente — aparentemente, trata-se do CEO solicitando uma transferência bancária urgente. No entanto, a voz foi sinteticamente clonada com alto grau de precisão. Nesse contexto, entram em ação técnicas avançadas de análise forense de áudio. Algoritmos avançados de detecção analisam inconsistências sutis, como pausas artificiais, anomalias de frequência e padrões de respiração irregulares, para desmascarar a falsificação. Mesmo quando a voz parece autêntica, o sistema consegue distinguir a diferença.

Em outro cenário, uma videoconferência apresenta um interlocutor conhecido transmitindo instruções críticas. Ainda assim, trata-se de uma falsificação. Ferramentas de detecção de *deepfakes* em vídeo analisam microexpressões faciais, padrões de piscada e sinais comportamentais que indicam origem sintética. Essas ferramentas funcionam como detectores digitais de mentiras, protegendo os canais de comunicação visual contra a manipulação. No entanto, a detecção isolada não é suficiente.

A resposta em tempo real torna-se essencial. Sistemas modernos de segurança incorporam mecanismos contínuos de detecção de *deepfakes*, capazes de identificar conteúdos suspeitos no momento em que surgem.

Os alertas são classificados por nível de gravidade, permitindo priorizar ameaças de maior risco sem sobrecarregar as equipes de segurança. Cada incidente é registrado com metadados detalhados, incluindo marcação temporal, origem e indicadores de anomalia, criando um rastro forense que apoia investigações e processos de conformidade regulatória.

Além disso, esses sistemas aprendem por meio de seus recursos integrados de auditoria, analisando padrões entre incidentes e ajudando as organizações a fortalecer suas defesas ao longo do tempo.

Seja para mitigar ataques de engenharia social ou preservar a integridade das comunicações digitais, o objetivo é claro: restaurar a confiança no que vemos e ouvimos. Em um cenário marcado pelo avanço da manipulação sintética, uma defesa multimodal e inteligente deixa de ser opcional e passa a ser essencial.

Preparando as operações de segurança para o futuro: mantendo-se um passo à frente da curva sintética

A detecção de *deepfakes* deve ser tratada como uma transformação estratégica. O que hoje é considerado sofisticado pode rapidamente se tornar padrão. Nesse contexto, as organizações que atuam de forma proativa constroem as bases para uma resiliência sustentável.

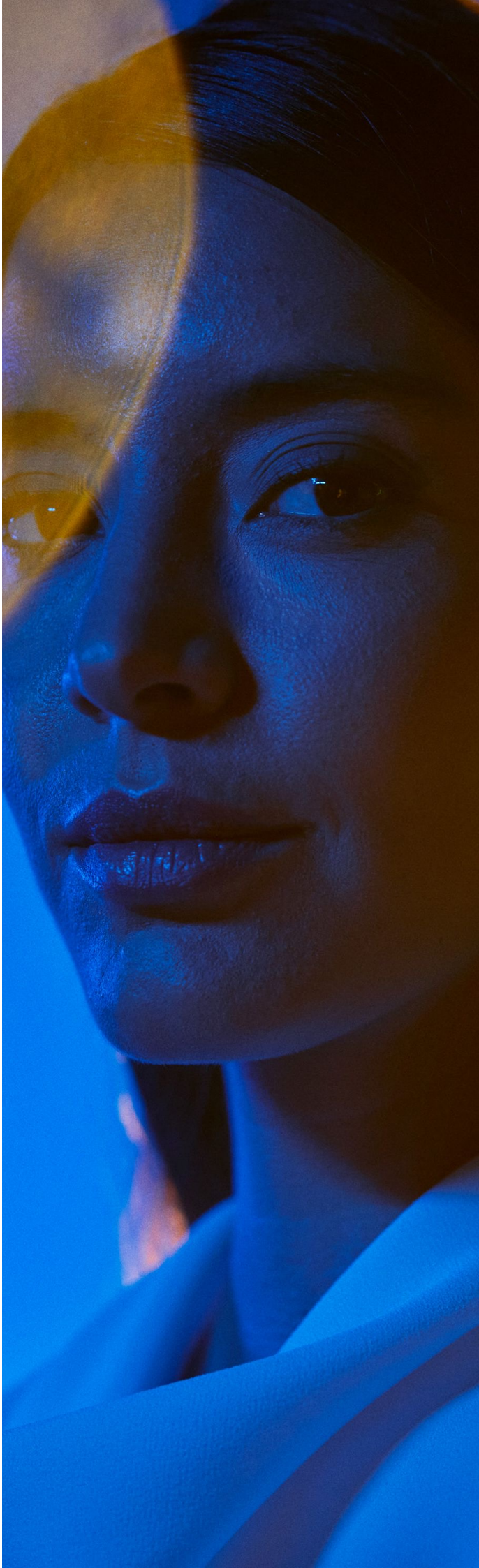
Vamos pensar na detecção de *deepfakes* como uma transformação estratégica. Ao integrar essas capacidades ao núcleo das operações de segurança, organizações e governos redefinem a forma como a veracidade é validada no ambiente digital. Já não basta confiar no que vemos ou ouvimos; precisamos verificar essas informações por meio de sistemas baseados em inteligência artificial que sejam tão ágeis quanto as ameaças que combatem.

Essa mudança vai além da tecnologia. Trata-se também de confiança. Proteger os canais de comunicação contra a falsificação de identidade impulsionada por IA garante que as decisões críticas sejam tomadas com base em informações autênticas. Isso preserva a integridade da liderança, a continuidade dos negócios e a confiança dos stakeholders.

Em um ambiente no qual conteúdos sintéticos se tornam cada vez mais convincentes, a defesa contra *deepfakes* se consolida como um pilar fundamental da confiança digital.

Prakash Narayanamoorthy
Global Capability Leader: Emerging Technology Security at NTT DATA

Ben Colman
Co-Founder and CEO at Reality Defender



Da regulamentação à prática: como as organizações devem governar a inteligência artificial

Artigo por Julissa Calderón Loayza e Melanie Brenis Valencia

O desenvolvimento e a adoção da Inteligência Artificial (IA) avançam em ritmo cada vez mais acelerado, gerando benefícios claros em eficiência, automação e tomada de decisão, mas também novos riscos relacionados à segurança, à privacidade e à transparência. À medida que esses sistemas passam a ser incorporados a processos cada vez mais críticos nas organizações, diferentes países vêm intensificando seus esforços para definir marcos regulatórios capazes de gerenciar esses riscos e proporcionar maior segurança jurídica. Nesse contexto, torna-se relevante analisar o estado atual da regulamentação da IA, apresentando, a seguir, uma visão geral de sua evolução em diferentes países da América e da Europa.

Europa: o primeiro marco vinculante de IA

Na Europa, a regulamentação da IA deu um passo decisivo com a aprovação do **Regulamento Europeu de Inteligência Artificial (EU AI Act)** em 2024, cuja aplicação ocorre de forma progressiva a partir de 2025. O regulamento é diretamente aplicável nos 27 Estados-membros da União Europeia e constitui o primeiro marco integral e vinculante sobre IA em nível regional. Suas disposições se aplicam a organizações públicas e privadas que desenvolvam, comercializem ou utilizem sistemas de IA no mercado europeu, inclusive quando estejam estabelecidas fora da União Europeia, desde que esses sistemas produzam efeitos em seu território.

De forma geral, o EU AI Act estabelece um modelo regulatório baseado no nível de risco dos sistemas de IA e define, entre outros pontos, as seguintes obrigações:

- Proíbe determinados usos de IA considerados inaceitáveis devido ao seu impacto sobre direitos fundamentais, como, por exemplo, a categorização biométrica sensível.
- Exige controles rigorosos para sistemas de IA de alto risco, incluindo a implementação de sistemas de gestão de riscos e de supervisão humana.
- Impõe obrigações de transparência, especialmente quando sistemas de IA interagem com pessoas ou geram conteúdo sintético, como no caso de deepfakes.
- Define responsabilidades claras para fornecedores, distribuidores e usuários de IA, estabelecendo mecanismos de controle por parte das autoridades competentes.

América: crescente interesse regulatório

Na América, diferentemente da Europa, a regulamentação da IA ainda se encontra em estágio inicial. A maioria dos países não possui leis específicas em vigor, embora se observe um interesse crescente, refletido em projetos de lei em países como Chile, Brasil, México, Colômbia, Argentina e Equador.

Essas iniciativas buscam estabelecer princípios éticos, proteger direitos fundamentais e definir obrigações básicas de transparência, mas, em sua maioria, ainda estão em fases iniciais e carecem de mecanismos claros de supervisão, o que gera incerteza regulatória.

Nesse panorama, os avanços mais relevantes concentram-se nos Estados Unidos e no Peru. Nos Estados Unidos, embora ainda não exista uma lei federal abrangente sobre IA, diversos Estados aprovaram regulamentações já em vigor, como o **Colorado AI Act**, o **Utah AI Policy Act** e o **Texas Responsible AI Governance Act**, que introduzem obrigações concretas em matéria de uso responsável, transparência e governança de sistemas de IA. Essa abordagem, ainda que fragmentada, é operacional e reflete um avanço regulatório em nível estadual. Por sua vez, o Peru foi um dos primeiros países da América Latina a desenvolver um marco normativo aplicado de forma efetiva, com a aprovção da **Lei nº 31814** em 2023, voltada ao uso da IA para o desenvolvimento econômico e social, e publicou posteriormente o **Decreto Supremo nº 115-2025-PCM**, estabelecendo disposições claras para sua implementação efetiva.

Em linhas gerais, o marco regulatório peruano — já em vigor — introduz obrigações que se destacam por:

- Classificar os sistemas de IA conforme seu nível de risco e finalidade, exigindo avaliações prévias de impacto e definindo o grau de supervisão aplicável antes de sua implantação.
- Exigir avaliações de impacto e análise de riscos para o uso de IA no setor público, considerando efeitos sobre direitos fundamentais, segurança e possíveis vieses.
- Estabelecer requisitos de rastreabilidade e documentação dos sistemas de IA, de modo a permitir a explicação de seu funcionamento, suas decisões e seus resultados perante as autoridades competentes.

- Referenciar expressamente padrões e boas práticas internacionais, como a ISO/IEC 42001, para orientar a gestão, o monitoramento contínuo e a melhoria dos sistemas de IA.

O panorama descrito evidencia que a regulamentação da IA avança em diferentes velocidades, mas com uma tendência comum: a necessidade de controlar os riscos associados ao seu uso. Diante disso, as organizações não podem se limitar a reagir a novas normas; precisam compreender em que patamar se encontram em relação ao uso da IA. Contar com um diagnóstico claro, identificar riscos e lacunas e implementar medidas adequadas é fundamental para estabelecer um modelo de governança de IA que complemente a regulamentação e permita antecipar exigências normativas futuras.

Governança de IA: da regulamentação à gestão efetiva do risco

O avanço regulatório descrito anteriormente confirma uma realidade: a Inteligência Artificial já não é apenas um tema tecnológico, mas uma questão estratégica, legal e de gestão de riscos. No entanto, a regulamentação por si só não garante um uso responsável. O verdadeiro desafio para as organizações é traduzir essas exigências em um modelo de governança interna que permita controlar riscos, assegurar a conformidade e sustentar a confiança.

Atualmente, uma parcela significativa das organizações já utiliza IA em processos críticos ou se encontra em fases avançadas de adoção. Diversos relatórios e estudos internacionais convergem em um ponto, embora o investimento em IA cresça de forma sustentada, a maturidade em governança, gestão de riscos e controle ainda é limitada. Em muitos casos, as iniciativas surgem de forma descentralizada, com modelos ou soluções implementados sem uma estrutura clara de supervisão, rastreabilidade ou responsabilidade formal. É nesse cenário que emerge a necessidade da Governança de IA.

O que significa governar a IA?

Governar a IA não significa frear a inovação, mas habilitá-la de forma segura. Implica estabelecer um framework estruturado que acompanhe todo o ciclo de vida do sistema — desde o design e o treinamento até o monitoramento e eventual desativação —, assegurando transparência, segurança, privacidade e conformidade normativa.

Um modelo sólido de governança de IA geralmente se apoia em cinco pilares fundamentais:

1) Definição clara de papéis e responsabilidades:

É indispensável delimitar quem é responsável pelos dados, pelo modelo, pelo controle de riscos e pela segurança. A ambiguidade organizacional é um dos principais fatores de exposição. Sem uma atribuição clara de responsabilidades, os riscos se diluem e as decisões críticas carecem de supervisão adequada.

A estrutura de governança dependerá do nível de maturidade digital e organizacional. Em empresas com maior desenvolvimento em análises e inovação, pode existir um Chief AI Officer (CAIO) com responsabilidade transversal sobre estratégia, riscos e supervisão da IA. Em outros casos, essa função é responsabilidade do Chief Data Officer (CDO), quando a IA está fortemente associada à governança de dados, ou sobre o CIO/CTO, quando o enfoque é mais tecnológico. Além disso, muitas organizações estão constituindo Comitês de IA interdisciplinares, com a participação de líderes de tecnologia, dados, risco, conformidade, jurídico e cibersegurança — e, em casos de maior exposição estratégica, membros do conselho ou da alta liderança. Essa abordagem permite alinhar a estratégia de inovação com a postura frente ao risco e a supervisão executiva.

2) Gestão de riscos desde a concepção:

A IA deve ser avaliada sob uma abordagem baseada em riscos. Isso inclui não apenas os riscos tecnológicos tradicionais, mas também riscos emergentes como vieses algorítmicos, falta de explicabilidade, vulnerabilidades a ataques adversariais, vazamento de informações, impacto reputacional e possíveis violações a direitos fundamentais. A gestão deve ser contínua, não pontual.

3) Integração com a governança de dados e a cibersegurança:

Não existe IA confiável sem dados governados nem sem controles de segurança robustos. A qualidade, a legitimidade e a classificação dos dados são determinantes. Além disso, a IA amplia a superfície de ataque, o que exige controles específicos frente a ameaças como manipulação de modelos, extração de informações ou uso indevido de ferramentas generativas por colaboradores ou terceiros.

4) Princípios corporativos de IA responsável:

Para além da conformidade normativa, as organizações devem definir seus próprios princípios de uso responsável — transparência, supervisão humana, não discriminação, segurança e prestação de contas — e integrá-los à sua cultura e aos seus processos de decisão.

5) Integração com a governança corporativa existente:

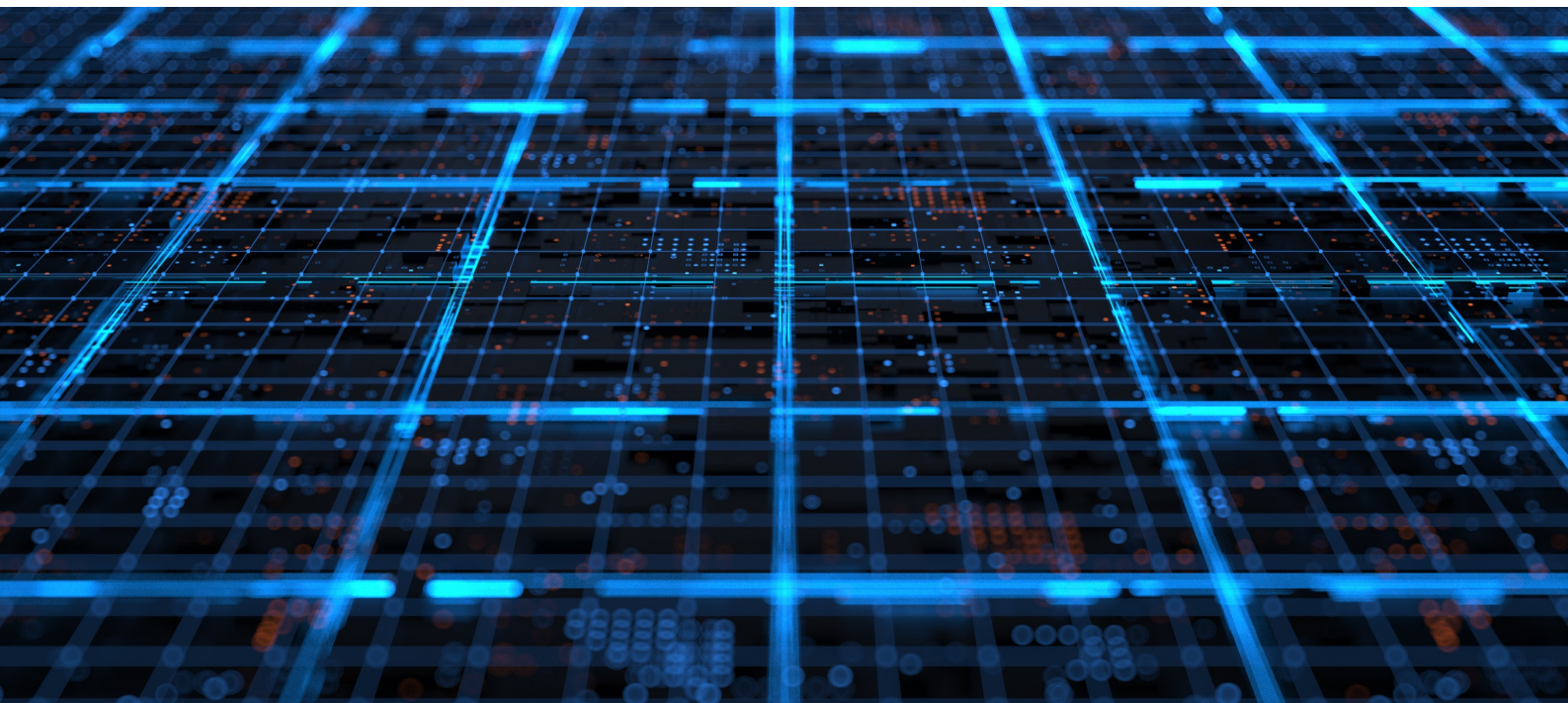
A governança de IA não deve operar como uma estrutura paralela, mas sim integrar-se à governança de dados, à gestão de riscos empresariais, à conformidade e à cibersegurança. Essa integração evita redundâncias e fortalece a coerência estratégica.

Confiança como vantagem competitiva

A evolução da IA não pode se limitar à dimensão tecnológica. As organizações que conseguirem estruturar um modelo de governança sólido estarão mais bem preparadas para responder a orgãos reguladores, mitigar incidentes e sustentar a confiança de clientes e stakeholders.
Em um ambiente em que a regulamentação avança e a exposição reputacional é imediata, o diferencial competitivo não estará apenas em quem inova mais rapidamente, mas em quem o faz de forma mais segura, transparente e responsável.
A governança de IA torna-se, assim, a ponte entre a inovação e a resiliência organizacional.

Julissa Calderón Loayza
Cybersecurity Expert Associate

Melanie Brenis Valencia
Cybersecurity Senior Consultant



Cibersegurança ofensiva e as tendências da Agentic AI: implicações para o Red Team moderno

Tendências por Marco Antonio Andazabal Rebaza

A evolução da Inteligência Artificial (IA) transformou a interação entre usuários e tecnologia, fazendo com que deixasse de ser uma ferramenta de apoio analítico para se tornar sistemas capazes de gerar conteúdo, automatizar processos e auxiliar em tarefas técnicas especializadas. Essa democratização reduziu a barreira de entrada a capacidades avançadas, permitindo que perfis não especializados acessem funcionalidades antes reservadas a especialistas. No entanto, essa acessibilidade também traz desafios significativos para a cibersegurança. No âmbito ofensivo, o surgimento de sistemas de Agentic AI introduz um novo paradigma: modelos capazes de planejar, executar e ajustar ações de forma autônoma para alcançar objetivos definidos.

Da IA generativa à Agentic AI

1) IA generativa e automação assistida

A IA generativa, baseada principalmente em modelos de linguagem de grande escala (LLMs), caracteriza-se por sua capacidade de produzir conteúdo, texto, código, configurações ou análises técnicas em resposta a instruções explícitas. Seu funcionamento é essencialmente reativo, respondendo a um prompt específico sem manter autonomia operacional além da interação imediata.

No contexto da cibersegurança ofensiva, sua aplicação se concentra em tarefas de apoio, como geração de scripts, criação de payloads e elaboração de e-mails para simulações de phishing, entre outras.

No entanto, esses sistemas exigem supervisão constante. Cada etapa do processo ofensivo — reconhecimento, exploração e pós-exploração — precisa ser realizada manualmente pelo operador. A IA não toma decisões estratégicas por conta própria; executa tarefas delimitadas por instruções humanas.

2) Agentic AI: sistemas orientados a objetivos e autonomia operacional

A Agentic AI representa uma evolução estrutural em relação aos modelos generativos tradicionais. Em vez de se limitar a responder a instruções pontuais, esses sistemas são projetados para alcançar objetivos definidos por meio de processos autônomos de planejamento e execução.

Do ponto de vista arquitetural, um sistema agentivo integra:

- Planejamento em múltiplas etapas: decompõe um objetivo complexo em subtarefas.
- Memória contextual persistente: mantém estado e histórico de ações.
- Execução autônoma: interage com ferramentas externas (APIs, sistemas operacionais, scanners).
- Avaliação e retroalimentação: analisa resultados e ajusta sua estratégia.

Arquiteturalmente, combina um modelo de linguagem com um ciclo de decisão iterativo (percepção → planejamento → ação → avaliação). Isso permite que opere como um "agente" dentro de um ambiente técnico, e não como um simples assistente passivo.

Transformação do Red Team na era da Agentic AI

1) Automatización avanzada del reconocimiento

Em operações de Red Team, a fase de reconhecimento (recon) é fundamental. Normalmente, envolve a coleta manual ou semiautomatizada de informações por meio de ferramentas OSINT, scanners de rede e análise de superfícies expostas.

Um sistema de Agentic AI poderia:

- Definir subobjetivos (identificar domínios, subdomínios e serviços expostos).
- Orquestrar múltiplas ferramentas de escaneamento.
- Correlacionar resultados.
- Priorizar vetores de ataque conforme a probabilidade de sucesso.
- Ajustar a estratégia com base em descobertas intermediárias.

Esse comportamento se assemelha mais ao de um operador humano assistido por automação inteligente do que a um simples script estático.

Geração dinâmica de exploits e provas de conceito

Embora a geração de exploits complexos ainda exija conhecimento especializado, a IA pode contribuir em:

- Adaptação de exploits públicos a contextos específicos.
- Modificação de payloads conforme as restrições do ambiente.
- Geração de scripts para provas de conceito (PoC).
- Análise preliminar de código vulnerável.

Em um cenário controlado de pentesting, isso pode reduzir significativamente o tempo entre a identificação de uma vulnerabilidade e sua validação técnica.

No entanto, a preocupação reside no fato de que agentes maliciosos poderiam utilizar capacidades semelhantes para escalar ataques com menor experiência técnica.

Engenharia social assistida por IA

Um dos impactos mais relevantes se observa na engenharia social. A Agentic AI pode:

- Analisar perfis públicos.
- Gerar e-mails altamente personalizados.
- Ajustar o tom conforme o perfil organizacional.
- Simular conversas coerentes em múltiplas iterações.

Isso aumenta a taxa potencial de sucesso em campanhas de phishing direcionado (*spear phishing*), reduzindo a necessidade de habilidades avançadas em manipulação psicológica.

A convergência entre cibersegurança ofensiva e Agentic AI redefine as dinâmicas tradicionais do Red Team e do pentesting técnico. A transição de modelos reativos para agentes autônomos orientados a objetivos introduz uma mudança estrutural na forma como as operações ofensivas são executadas e escaladas.

Mais do que uma simples ferramenta de automação, a Agentic AI representa um catalisador estratégico capaz de ampliar tanto a eficiência defensiva quanto o potencial ofensivo. Nesse contexto, o desafio não reside apenas em compreender a tecnologia, mas em antecipar suas implicações operacionais, éticas e regulatórias.

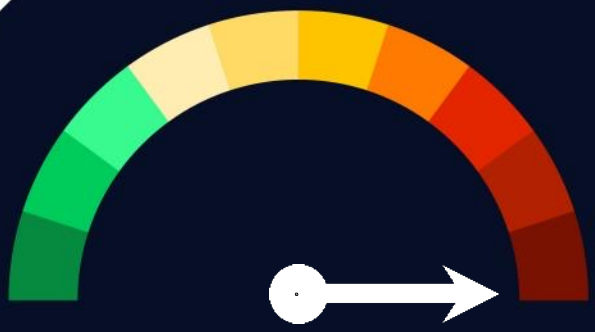
O futuro da cibersegurança não dependerá exclusivamente da capacidade técnica, mas da governança inteligente de sistemas cada vez mais autônomos.

Marco Antonio Andazabal Rebaza
Cybersecurity Analyst

Vulnerabilidades

Vulnerabilidades no Cisco Secure Firewall Management Center

Data: 4 de março de 2026
CVE: CVE-2026-20079 e CVE-2026-20131



CVSS: 10.0

CRÍTICA

Descrição

Foram identificadas duas vulnerabilidades críticas, catalogadas como CVE-2026-20079 e CVE-2026-20131, presentes no Cisco Secure Firewall Management Center.

A primeira vulnerabilidade (CVE-2026-20079) origina-se de um processo incorreto criado durante a inicialização do sistema. Um agente malicioso poderia explorar essa falha por meio do envio de uma requisição HTTP maliciosa. Isso poderia permitir que o invasor executasse scripts e comandos no sistema afetado e obtivesse privilégios de *root*.

A segunda vulnerabilidade (CVE-2026-20131) consiste na desserialização insegura de um fluxo de bytes Java fornecido pelo usuário. Um agente malicioso poderia enviar um objeto Java serializado para executar código no dispositivo e obter privilégios de *root*.

Solução

O fabricante recomenda atualizar os produtos afetados para a versão mais recente.

Para isso, a Cisco orienta seus clientes a utilizar o verificador de software da Cisco e seguir as instruções disponibilizadas.

Produtos afetados

Os produtos afetados são os seguintes:

- Cisco Secure FMC Software
- Cisco Security Cloud Control (SCC) Firewall Management

Referências

- sec.cloudapps.cisco.com
- sec.cloudapps.cisco.com
- incibe.es

TLP:WHITE

Vulnerabilidades

Vulnerabilidade crítica no Android (System)

Data: 2 de marzo de 2026
CVE: CVE-2026-0006

CVSS: 9.8

CRÍTICA

Descrição

Foi identificada uma vulnerabilidade crítica de execução remota de código em dispositivos Android em diferentes versões.

Essa vulnerabilidade (**CVE-2026-0006**) permite que um agente malicioso execute código malicioso sem necessidade de interação do usuário com o dispositivo e sem exigir privilégios adicionais, dado que a falha reside em um componente do sistema Android.

A exploração bem-sucedida poderia resultar em controle remoto total do dispositivo, exfiltração de dados, espionagem em tempo real, instalação de malware persistente e movimentação lateral na rede.

Solução

Recomedação:

- Recomenda-se instalar a atualização correspondente ao nível de patch 2026-03-01 ou posterior. Os fabricantes (Google, Samsung, Xiaomi, entre outros) estão disponibilizando essas atualizações via OTA (Over The Air).

Produtos afetados

São afetados por essa vulnerabilidade os dispositivos com as seguintes versões do Android: 12, 12L, 13, 14 e 15.

Referências

- source.android.com
- nvd.nist.gov

Android corrige 129 vulnerabilidades em seu patch de segurança de março

Data: 2 de marzo de 2026
CVE: CVE-2026-21385 e 128 outras

Crítica

Descrição

O boletim de segurança de março do Android corrige um total de 129 vulnerabilidades, sendo 10 de severidade crítica e 106 de severidade alta. A exploração dessas vulnerabilidades poderia permitir que um agente malicioso provocasse escalonamento de privilégios, execução remota de código ou estouro de buffer.

O Android indica que a vulnerabilidade identificada como CVE-2026-21385 pode estar sendo explorada de forma limitada. Essa vulnerabilidade está presente em um componente de tela da Qualcomm. Em comunicado oficial, a empresa informou que a falha se encontra no subcomponente Graphics e consiste em um estouro de inteiros que um agente malicioso pode explorar para provocar corrupção de memória no dispositivo.

Produtos afetados

Os produtos afetados por esta vulnerabilidade incluem:

- Versões do Android anteriores a 2026-03-01.

Solução

Atualizar os produtos afetados para a versão mais recente disponível.

Referências

- source.android.com
- docs.qualcomm.com

TLP:WHITE

Google corrige de forma emergencial 10 vulnerabilidades

Data: 3 de marzo de 2026
CVE: CVE-2026-3536 e 9 outras

Alta

Descrição

O Google lançou um patch de segurança emergencial para o Chrome com o objetivo de corrigir 10 vulnerabilidades, das quais 3 são de caráter crítico e 7 de severidade alta.

Entre os problemas identificados, destacam-se falhas de corrupção de memória, erros de implementação em mecanismos gráficos e JavaScript, que poderiam permitir a execução de código arbitrário e a exfiltração de informações, entre outros impactos.

As vulnerabilidades mais críticas (CVE-2026-3536, CVE-2026-3537 e CVE-2026-3538), relacionadas a lacunas na validação de codecs web e de navegação, poderiam dificultar a identificação de tentativas de phishing e facilitar downloads não autorizados.

Recomenda-se que usuários e organizações apliquem as atualizações disponíveis o mais brevemente possível.

Produtos afetados

Os produtos afetados por essa vulnerabilidade incluem todas as versões do Google Chrome anteriores a esta versão:

- 145.0.7632.159

Solução

Recomedação:

- Atualizar para as versões mais recentes do software.

Referências

- cyberpress.org
- nvd.nist.gov

Eventos

IAPP Global Summit 2026
30 de março – 2 de abril

O IAPP Global Privacy Summit 2026 será realizado de 30 de março a 2 de abril de 2026 em Washington, D.C., em formato presencial, consolidando-se como um dos principais encontros globais em privacidade e cibersegurança. O evento reunirá líderes dos setores público e privado para debater regulamentação internacional, governança de dados, ameaças cibernéticas emergentes e o impacto real da inteligência artificial na proteção da informação. Um espaço estratégico para antecipar tendências e desafios regulatórios.

Link

Gen AI Summit
17 e 18 de abril

O Gen AI Summit EU 2026 será realizado de 17 a 18 de abril de 2026 em Valência, Espanha, em formato presencial de dois dias, voltado para profissionais de IA, dados e machine learning. O encontro reunirá líderes de tecnologia, inovadores e equipes técnicas para explorar o potencial transformador da inteligência artificial generativa. A programação inclui palestras técnicas, painéis sobre ética, governança e segurança, além de espaços de networking e workshops práticos.

Link

CYBERUK 2026
21 a 23 de abril

O CYBERUK 2026 será realizado de 21 a 23 de abril de 2026 no Scottish Event Campus (SEC), em Glasgow, Reino Unido, em formato presencial. Organizado pelo National Cyber Security Centre (NCSC), é a conferência de referência do governo britânico para profissionais de cibersegurança, contando com mais de 100 palestrantes especializados. A programação aborda o tema "The Next Decade: Accelerating Our Cyber Defence", explorando estratégias, ameaças emergentes, defesa de infraestruturas críticas e colaboração público-privada. O evento contará ainda com áreas de exposição, networking e sessões técnicas especializadas.

Link

Cyber Intelligence Europa 2026
22 e 23 de abril

O Cyber Intelligence Europe 2026 será realizado de 22 a 23 de abril de 2026 em Bruxelas, Bélgica, em formato presencial de dois dias. Com foco na cooperação europeia, reunirá representantes de governos, forças armadas e agências de segurança para debater estratégias nacionais de cibersegurança e políticas públicas.
A programação inclui análise de tendências em cibercrimes, monitoramento de ameaças e preparação para ciberataques, com ênfase especial na proteção de infraestruturas críticas.
O evento explorará ainda abordagens coordenadas de compartilhamento de dados e respostas conjuntas a ameaças de grande escala.

Link

Recursos

➢ Integrating Cybersecurity and Enterprise Risk Management (ERM) - NIST

Publicação do NIST que orienta as organizações na integração da cibersegurança ao framework de Gestão de Riscos Empresariais (ERM), alinhando a identificação, avaliação e priorização de riscos cibernéticos aos objetivos estratégicos do negócio. O documento explica como utilizar registros de riscos para aprimorar a tomada de decisões em nível executivo e fortalecer a comunicação entre as áreas técnicas e a alta liderança, promovendo uma governança mais sólida e maior resiliência frente a ameaças digitais.

Link

➢ The ENISA Cybersecurity Exercise Methodology – ENISA

Publicação da ENISA que oferece um framework teórico completo para planejar, executar e avaliar exercícios de cibersegurança de forma eficaz, do início ao fim, assegurando a participação dos perfis e partes interessadas adequados no momento oportuno.
O documento tem como base as lições aprendidas, as práticas recomendadas do setor e a experiência em cibersegurança, além de ferramentas e modelos práticos para organizar e melhorar as simulações e os testes de resposta a incidentes.

Link

➢ Compêndio sobre proteção de dados pessoais: normativa e critérios interpretativos relevantes

Documento oficial publicado pelo Ministério da Justiça e Direitos Humanos do Peru que compila de forma atualizada a Lei de Proteção de Dados Pessoais nº 29733, seu regulamento e os critérios interpretativos relevantes emitidos pela Autoridade Nacional de Proteção de Dados Pessoais, com o objetivo de orientar a aplicação prática da normativa. Serve como guia de referência para entidades públicas e privadas na gestão e no cumprimento do marco legal de proteção de dados pessoais.

Link

Inscreva-se na RADAR
up.nttdata.com/suscribetearadar

Powered by the Cybersecurity NTT DATA Team

br.nttdata.com