Ciberataques potenciados por IA: a nova realidade que os CISO enfrentam

itsecurity.pt/news/analysis/ciberataques-potenciados-por-ia-a-nova-realidade-que-os-ciso-enfrentam

A inteligência artificial transformou-se na mais recente fronteira da cibersegurança, criando uma realidade paradoxal onde a mesma tecnologia promete revolucionar a produtividade empresarial e se tornou, também, na arma de eleição dos cibercriminosos. Esta nova geração de ataques exige uma reavaliação fundamental das estratégias de cibersegurança onde a velocidade de resposta e a sofisticação das defesas se tornaram questões de sobrevivência organizacional

Por Rui Damião . 01/10/2025

O panorama de cibersegurança enfrentou uma transformação radical em 2025, com a Inteligência Artificial (IA) a emergir simultaneamente como ferramenta de defesa e arma ofensiva. O relatório “Cost of Data Breach 2025” da IBM revelou que 13% das organizações reportaram violações de modelos ou aplicações de IA, enquanto 16% das violações de dados envolveram atacantes que utilizaram inteligência artificial, mais frequentemente para phishing gerado por IA (37%) e ataques de deepfake (35%).

Estas estatísticas ganham maior relevância quando analisadas no contexto empresarial global. O “Cybersecurity Readiness Index 2025” da Cisco descobriu que quase nove em cada dez (86%) dos líderes empresariais com responsabilidade de cibersegurança reportaram pelo menos um incidente relacionado com IA nos últimos 12 meses, demonstrando que esta já não é uma ameaça futura, mas uma realidade presente.

A sofisticação dos ataques

A Microsoft, por seu lado, observa uma evolução preocupante nas técnicas dos atacantes, tendo identificado a utilização de IA por atores de ameaça para reconhecimento, investigação de vulnerabilidades, tradução, técnicas de comando operacional refinadas por LLM, desenvolvimento de recursos, técnicas de scripting, evasão de deteção, engenharia social e ataques de força bruta.

A sofisticação destes ataques é particularmente evidente na personalização em escala. A IBM reportou que a IA generativa reduziu o tempo necessário para escrever um email de phishing convincente de 16 horas para apenas cinco minutos, demonstrando como a tecnologia democratiza capacidade que antes eram exclusivas de grupos altamente especializados.

O custo de não agir

O estudo “Global Digital Trust Insights 2025” da PwC revelou que apenas 2% das organizações implementaram ciber-resiliência em toda a organização, apesar de 77% das organizações esperarem que o seu orçamento de cibersegurança aumente no próximo ano.

Esta lacuna entre consciencialização e implementação é especialmente crítica num contexto que dois terços dos líderes de segurança notam que a IA generativa expandiu a superfície de ciberataque no último ano, à frente de outras tecnologias como cloud (66%) e produtos conectados (67%).

Os custos financeiros destas ameaças são substanciais: embora os custos globais médios das violações de dados tenham diminuído para 4,44 milhões de dólares em 2025, segundo o “Cost of Data Breach 2025” da IBM, 20% das organizações experienciaram violações ligadas ao uso não autorizado de inteligência artificial, adicionando uma média de 670 mil dólares aos custos de violação.

No entanto, de acordo com a IBM, 97% das organizações que sofreram violações relacionadas com inteligência artificial não tinham controlos de acesso à IA em vigor. Segundo a Cisco, apenas 48% das empresas acredita que os seus funcionários compreendem totalmente como é que os atores maliciosos estão a usar a IA para melhorar os seus ataques.

Para os CISO portugueses, estes dados representam mais do que apenas estatísticas: são um alerta de que a era dos ataques previsíveis acabou. No seu lugar emerge uma nova geração de ameaças que aprendem, adaptam-se e evoluem a uma

velocidade sem precedentes, exigindo uma reavaliação fundamental das estratégias de cibersegurança organizacional.

A escala e velocidade dos novos ataques

A transformação mais significativa no panorama de cibersegurança não reside apenas na sofisticação dos ataques, mas na sua velocidade e escala sem precedentes. A IBM registou um aumento de 71% ano após ano em ataques que utilizam credenciais comprometidas, enquanto a IBM X-Force observou um aumento de 84% em emails de phishing que entregam infostealers numa base semanal.

"Na maioria dos casos, simplesmente não sabemos se um email de phishing foi escrito por uma IA, nem podemos dizer se um script shell ofuscado foi criado manualmente ou gerado por IA"

Chester Wisniewski, Director, Global Field CISO da Sophos

Chester Wisniewski, Director, Global Field CISO da Sophos, contextualiza esta evolução e afirma que é na velocidade e escala que se "espera ver o maior avanço por parte dos agentes de ameaças que utilizam IA", especificando que "os ataques, em si mesmos, não deverão tornar-se muito mais sofisticados, mas sim mais frequentes e de maior qualidade", onde os "atacantes menos qualificados vão aumentar as suas capacidades enquanto os atacantes de nível médio vão encontrar mais vítimas e explorá- -las mais rapidamente".

Bruno Castro, Fundador e CEO da VisionWare, afirma que "quantificar exatamente a diferença em velocidade ou escala entre ataques tradicionais e ataques potenciados por IA é difícil porque os operadores misturam várias técnicas". No entanto, diz, "o que antes estava ao alcance de grupos altamente especializados passa agora a ser acessível a atores com menos recursos, pois a IA reduz o conhecimento técnico necessário para executar ataques sofisticados".

Esta democratização é corroborada por dados da Cisco, que apontam que 52% das organizações reportam que os funcionários não compreendem totalmente como os atores maliciosos estão a usar IA, enquanto apenas 49% dos inquiridos acreditam que os funcionários compreendem totalmente as ameaças de cibersegurança relacionadas com IA.

Automação e personalização em escala

A diferença fundamental reside na capacidade de automação inteligente. Bruno Castro explica que se observa “um aumento expressivo da personalização de phishing, crescimento rápido de casos de voice cloning usados em fraudes e automação de tarefas que antes exigiam intervenção humana”.

O “Digital Defense Report 2024” da Microsoft observou um aumento de 2,75 vezes ano após ano em ataques de ransomware, demonstrando como os atacantes estão a combinar inteligência artificial com táticas tradicionais para aumentar a eficácia.

Wisniewski refere que, “na maioria dos casos, simplesmente não sabemos se um email de phishing foi escrito por uma IA, nem podemos dizer se um script Shell ofuscado foi criado manualmente ou gerado por IA”. No entanto, destaca um padrão revelador: “o aumento do volume de ataques com pequenas variações no tema, algo que antes poderia ser difícil para os humanos sem contarem com automação sofisticada”.

A Cisco revelou que 71% dos líderes acreditam que um incidente de cibersegurança terá lugar e vai afetar a organização dentro dos próximos 12 a 24 meses, o que reflete a urgência das ameaças. Bruno Castro resume a situação atual: “a IA reduz o custo por alvo, permitindo atingir milhares de vítimas com conteúdo adaptado, ou seja, mais velocidade e escala operacional”. Esta transformação leva a que os CISO enfrentam não apenas ameaças mais sofisticadas, mas também mais frequentes e economicamente viáveis para os atacantes, alterando completamento o cálculo de risco-benefício no panorama de cibersegurança.

O espetro atual das ameaças

A diversidade de ataques potenciados por inteligência artificial expandiu dramaticamente, abrangendo desde técnicas tradicionais aprimoradas até vetores de ataque completamente novos. Chester Wisniewski refere que “os dois tipos mais prolíficos, e confirmados como gerados por IA, são os emails de phishing e as campanhas de mensagens fraudulentas (SMS, WhatsApp, etc.). A principal razão pela qual podemos confirmar que são gerados por IA é que atacantes ficaram sem dinheiro para fazer melhor, e então mensagens cheias de erros geradas por LLM passaram a aparecer nas tentativas de fraude”.

Bruno Castro oferece uma catalogação mais abrangente e indica que os ataques “que se têm observado com maior frequência são principalmente phishing altamente personalizado”, como spear phishing e vishing com scripts criados por IA, “clonagem de voz para fraudes telefónicas, deepfake vídeo/áudio para chantagem ou manipulação pública, malware com componentes adaptativos (malware que ajusta a sua assinatura ou comportamento segundo o host), automação de BEC e negociação de resgates com chatbots”.

Esta evolução representa uma mudança fundamental no panorama de ameaças onde os atacantes já não dependem exclusivamente de uma técnica, mas combinam múltiplos vetores potenciados por IA para maximizar o sucesso. Desde emails de phishing gerados em minutos até malware que se adapta autonomamente aos

sistemas que infeta, a IA democratizou capacidades avançados e criou um ecossistema de ameaças onde a velocidade, personalização e automação convergem para desafiar as defesas tradicionais.

A dificuldade técnica de identificação

Um dos aspetos mais preocupantes dos ataques potenciados por IA é a sua capacidade de mimificar comportamentos legítimos, tornando a deteção extremamente complexa. Chester Wisniewski, da Sophos, diz, diretamente, que “é muito difícil, se não impossível, na maioria dos casos” distinguir tecnicamente entre um ataque tradicional automatizado e um verdadeiramente potenciado por inteligência artificial. “A menos que existam artefactos específicos deixados pelo LLM do qual estão a abusar, não há nenhuma forma óbvia de determinar isto”, explica.

“Um dos pontos destacados no mais recente relatório de Riscos e Conflitos do CNCS é que a IA generativa está a emergia como veículo facilitador de ataques: simplifica campanhas de phishing altamente direcionadas, com a criação de deepfakes utilizados em fraudes empresariais e a automação da identificação de vulnerabilidades”

Bruno Castro, Fundador e CEO da VisionWare

Bruno Castro, da VisionWare, oferece uma abordagem mais estruturada para a identificação. “Tecnicamente, um CISO pode distinguir (ou, pelo menos, suspeitar) entre um ataque tradicional automatizado e um verdadeiramente “AI-powered” observando características como: (1) grau de personalização sem precedentes nas mensagens (contexto recente, referências pessoais que não vêm de leaks públicos comuns), (2) texto/linguagem que se adapta o tom ao interlocutor em tempo real, (3) voz ou vídeo sintetizados de alta qualidade e sincronizados e (4) interação conversacional com variações muito ligeiras e sistemáticas entre mensagens que revelam alguns padrões”, afirma.

O Fundador & CEO da VisionWare detalha, também, os indicadores que diz serem mais fiáveis para deteção: “anomalias de conteúdo, por exemplo, em linguagem que não corresponde ao histórico do remetente, e metadados inconsistentes. Anomalias pronunciadas, e anomalias em vídeo detetáveis através de artefactos subtis, inconsistência de iluminação, incongruência de movimentos”, entre outros.

Bruno Castro enfatiza que a deteção eficaz requer uma combinação de técnicas. “A um nível mais técnico, é essencial combinar análise forense de conteúdo, ML-detectors de deepfake e enriquecimento de indicators of compromise, que aumentam a fiabilidade de deteção”.

A realidade, contudo, permanece desafiante: enquanto as ferramentas de deteção evoluem, os atacantes também refinam as suas técnicas. Como diz Wisniewski, da Sophos, a distinção entre ataques tradicionais e potenciados por IA muitas vezes só se torna aparente após análise forense detalhada, quando os em áudio como pausas não naturais e palavras mal danos já podem estar consumados. Isto coloca os CISO numa posição reativa, enfatizando a importância de estratégias de defesa em profundidade que assumam a presença de conteúdo gerado por IA no panorama de ameaças.

O panorama português

Em Portugal, os padrões de ataque seguem tendências específica que refletem tanto a estrutura económica nacional, como as vulnerabilidades setoriais, Bruno Castro identifica os alvos preferenciais: “financeiro, saúde e administração pública e/ou infraestruturas críticas”, em linha com o panorama europeu.

A preocupação, reforça Bruno Castro, é reconhecida ao mais alto nível e “um dos pontos destacados no mais recente relatório de Riscos e Conflitos do CNCS é que a IA generativa está a emergir como veículo facilitador de ataques: simplifica campanhas de phishing altamente direcionadas, com a criação de deepfakes utilizados em fraudes empresariais e a automação da identificação de vulnerabilidades”.

Em Portugal, os gaps mais críticos “abrangem tanto a deteção de conteúdos sintéticos, como deepfakes de voz e vídeo, como a resposta a campanhas de phishing avançadas e em larga escala”, explica Bruno Castro, acrescentando que “muitas empresas não dispõem de ferramentas capazes de identificar manipulações multimédia credíveis nem de soluções de filtragem preparadas para mensagens altamente personalizadas, criadas em vários idiomas e adaptadas ao contexto da vítima”. Ao mesmo tempo, “os processos de verificação de identidade também continuam frágeis, sobretudo em fluxos financeiros ou de comunicação executiva, onde pedidos falsificados podem ser aceites sem autenticação robusta ou validações alternativas”.

O Fundador & CEO da VisionWare observa que são poucas as organizações portuguesas que têm políticas específicas para enfrentar conteúdo gerado por inteligência artificial. “Os dados públicos são limitados e não existe um levantamento nacional público e atualizado que diga quantas organizações têm políticas específicas para deepfakes/IA. Ainda assim, e uma vez que as ameaças deepfake têm ganho cada vez uma maior expressão em Portugal, diria que a grande maioria ainda não tem protocolos formais para deepfakes, principalmente, claro, as empresas de menor dimensão”.

Bruno Castro refere, ainda, a questão da preparação humana. “Em muitos casos, a formação de colaboradores ainda está orientada para sinais tradicionais de fraude e não para conteúdos criados por IA e poucas organizações têm políticas o exercícios

de red teaming específicos para estas ameaças", afirma.

Abordagem pragmática para os CISO

Face às limitações orçamentais que caracterizam muitas organizações portuguesas, a questão central não é implementar todas as soluções disponíveis, mas priorizar investimentos que ofereçam o maior retorno em termos de segurança. Chester Wisniewski partilha que, hoje, "a IA é utilizada principalmente para aumentar a eficácia da engenharia social. O passo mais importante que um CISO pode dar é rever os seus processos de verificação humana quando realmente importa e implementar uma autenticação multifator eficaz e resistente a phishing".

Já Bruno Castro defende que, "independente de serem ameaças potenciadas por IA", um CISO com um orçamento limitado deve "apostar numa fórmula que funciona à data de hoje para a maioria das organizações: implementar um modelo formal de governação de segurança com o envolvimento obrigatório do top management que permita avaliar, detetar e corrigir continuamente as principais vulnerabilidades dentro de casa – técnicas ou procedimentais".

Para Wisniewski, da Sophos, há três áreas críticas que requerem atenção imediata: "os processos de suporte técnico para verificar a identidade do utilizador antes de este poder redefinir credenciais de acesso; os processos de verificação para que as transferências bancárias e outras transações financeiras tenham autenticação mútua; e a atualização das formações de sensibilização de segurança, para que vão além da procura por erros gramaticais e ortográficos ou links suspeitos". Estas técnicas básicas, acrescenta, "já não são eficazes num mundo de conteúdos gerados por LLM que se tornam cada vez mais inteligentes".

A nova realidade da cibersegurança

Os ciberataques potenciados por inteligência artificial deixaram de ser uma ameaça futura para se tornarem uma realidade presente e premente para as organizações de todo o mundo, onde se inclui, naturalmente, Portugal. Como demonstrado pelos dados da IBM, da Microsoft, da Cisco e da PwC, estamos perante uma transformação fundamental no panorama de ameaças: 13% das organizações já reportaram violações de modelos ou aplicações de IA, enquanto 86% dos líderes empresariais experienciaram pelo menos um incidente relacionado com IA nos últimos 12 meses.

Chester Wisniewski resume a evolução: "os ataques, em si mesmos, não deverão tornar-se muito mais sofisticados, mas sim mais frequentes e de maior qualidade". Bruno Castro complementa e observa que "o que antes estava ao alcance de grupos altamente especializados passa, agora, a ser acessível a atores com menos recursos, pois a IA reduz o conhecimento técnico necessário para executar ataques sofisticados".

Os dados revelam uma janela crítica de oportunidade. Apenas 2% das organizações implementaram resiliência cibernética em toda a organização, apesar de 77% esperarem aumentar os orçamentos de cibersegurança. Esta discrepância entre consciencialização e ação efetiva representa simultaneamente o problema e a solução.

A magnitude desta transformação exige colaboração entre setores, partilha de inteligência de ameaças e uma abordagem coordenada a nível nacional. Com 67% dos líderes a observarem que a IA generativa expandiu a superfície de ataque, a resposta não pode ser individual.

A era dos ataques previsíveis terminou e, no seu lugar, emerge uma nova realidade onde a inteligência artificial democratizou capacidades avançadas de ataque, mas também oferece ferramentas poderosas para a defesa. O sucesso dos CISO portugueses vai depender da sua capacidade de navegar esta dualidade, transformando uma ameaça existencial numa oportunidade de fortalecimento organizacional através de estratégias inteligentes, investimentos direcionados e uma governação de segurança verdadeiramente integrada no ADN empresarial.