O COLAPSO DAS DEFESAS CONVENCIONAIS

Por que NGFW, EDR e SIEM não são mais suficientes diante de ameaças cibernéticas impulsionadas por Inteligência Artificial

AUTORES: Marcelo Branquinho e Plataforma Safer

20/02/2026

RESUMO: A ascensão da inteligência artificial como instrumento ofensivo transformou profundamente o cenário da segurança cibernética, especialmente em ambientes que integram tecnologia da informação (TI), tecnologia operacional (TO), IoT e computação em nuvem, caracterizando os chamados sistemas ciberfísicos (CPS).

Ferramentas tradicionais como firewalls de próxima geração (NGFW), sistemas de detecção e resposta a endpoints (EDR) e plataformas de gerenciamento de eventos e informações de segurança (SIEM), historicamente eficazes contra ameaças baseadas em assinaturas e padrões previsíveis, demonstram limitações estruturais diante de ataques dinâmicos, polimórficos e adaptativos construídos com apoio de modelos de linguagem de grande escala (LLMs) e algoritmos de aprendizado de máquina.

Este artigo analisa tecnicamente o esgotamento do paradigma de defesa em profundidade quando operado sob lógica estática e reativa, demonstra como LLMs podem subverter controles baseados em regras e discute a necessidade de um novo modelo de detecção fundamentado em comportamento e contexto operacional. Argumenta-se que a sobrevivência cibernética de infraestruturas críticas exige a incorporação de inteligência artificial defensiva como núcleo arquitetural da defesa, superando a fragmentação entre camadas técnicas e promovendo correlação contextual contínua entre eventos, ativos e impactos operacionais.

PALAVRAS-CHAVE: Inteligência Artificial; LLMs; Defesa em Profundidade; NGFW; EDR; SIEM; Sistemas Ciberfísicos; CPS; Segurança Cibernética Industrial; Detecção Comportamental; Contexto Operacional; IA Ofensiva.

1. INTRODUÇÃO

A segurança cibernética atravessa uma inflexão histórica comparável à transição da segurança perimetral tradicional para a defesa em profundidade ocorrida no início dos anos 2000. Entretanto, a transformação atual é ainda mais profunda, pois não decorre apenas do aumento do número de ameaças, mas da mudança qualitativa na natureza dos ataques. A incorporação sistemática de inteligência artificial aos arsenais ofensivos alterou a assimetria entre atacantes e defensores, introduzindo velocidade, adaptabilidade e capacidade de personalização em escala industrial.

Historicamente, os modelos de defesa foram concebidos sob a premissa de que ataques poderiam ser identificados por padrões recorrentes. Assinaturas de malware, indicadores de comprometimento (IoCs), reputação de IP e regras heurísticas sempre constituíram o núcleo dos mecanismos de detecção. Essa lógica foi eficaz em um cenário em que o desenvolvimento de ataques exigia tempo, habilidade técnica significativa e infraestrutura especializada.

Contudo, com a popularização de modelos de linguagem de grande escala (LLMs), frameworks de machine learning open source e ferramentas automatizadas de geração de código, o ciclo de criação de ataques foi drasticamente reduzido. Um invasor, munido de ferramentas baseadas em IA, pode hoje gerar variações praticamente ilimitadas de um mesmo artefato malicioso, testar múltiplos vetores simultaneamente e adaptar sua abordagem conforme as respostas do ambiente alvo.

Essa mudança não é apenas tecnológica, mas estrutural. A defesa baseada em previsibilidade enfrenta uma ameaça baseada em adaptabilidade. A lógica determinística encontra uma lógica probabilística. E, nesse choque de paradigmas, as defesas tradicionais começam a demonstrar limitações críticas.

Em sistemas ciberfísicos (CPS), onde a convergência entre TI, TO, IoT e nuvem amplia a superfície de ataque, essa vulnerabilidade estrutural torna-se ainda mais preocupante. A falha não implica apenas

vazamento de dados, mas potencial interrupção de processos industriais, danos ambientais ou riscos à vida humana.

É nesse contexto que se insere a análise do colapso das defesas convencionais.

2. O ESGOTAMENTO DA DEFESA EM PROFUNDIDADE BASEADA EM CAMADAS ESTÁTICAS

A defesa em profundidade consolidou-se como pilar arquitetural da segurança moderna, recomendada por normas como IEC 62443 e NIST SP 800-82. Seu princípio é simples e elegante: múltiplas camadas independentes aumentam o custo e a complexidade do ataque, reduzindo a probabilidade de comprometimento total.

Contudo, essa arquitetura foi concebida sob uma hipótese implícita: cada camada opera com eficácia previsível diante de vetores de ataque conhecidos ou modeláveis.

O problema emerge quando as camadas são projetadas para reagir a padrões estáticos, enquanto o atacante opera com padrões dinâmicos.

A defesa em profundidade tradicional é composta por:

- Controles perimetrais;
- Segmentação de rede;
- Proteção de endpoints;
- Monitoramento centralizado;
- Políticas e governança.

Embora cada elemento possua mérito individual, a integração entre eles raramente é cognitiva. Em muitos ambientes, as camadas funcionam como silos tecnológicos, trocando eventos, mas não necessariamente compartilhando contexto operacional.

Essa fragmentação cria um fenômeno crítico: excesso de dados e escassez de inteligência.

Quando um atacante utiliza IA para alternar vetores, modular intensidade e operar dentro de limites aceitáveis de comportamento, cada camada isoladamente pode considerar o evento como tolerável. O ataque, contudo, emerge da correlação longitudinal de múltiplos pequenos desvios.

Assim, o modelo não colapsa por ausência de tecnologia, mas por ausência de contextualização integrada.

3. LIMITAÇÕES DOS NGFWS DIANTE DE AMEAÇAS ADAPTATIVAS

Os firewalls de próxima geração representaram evolução significativa em relação aos firewalls tradicionais, incorporando inspeção profunda de pacotes, controle granular por aplicação e integração com feeds de inteligência de ameaças. No entanto, sua lógica central permanece baseada em regras declarativas e padrões identificáveis.

Em ataques impulsionados por IA, essa previsibilidade é explorada sistematicamente.

Modelos ofensivos podem:

- Analisar regras expostas por tentativa e erro;
- Ajustar o tamanho e a fragmentação de pacotes;
- Modular a frequência de requisições para evitar limiares de detecção;
- Utilizar protocolos legítimos como canal de comando e controle.

Em ambientes industriais, a situação torna-se ainda mais complexa. Protocolos como Modbus, Profinet ou DNP3 foram concebidos para funcionalidade e disponibilidade, não para segurança robusta. Muitos carecem de autenticação ou criptografia nativa.

Um atacante que utiliza IA para modelar tráfego compatível com o comportamento esperado do processo industrial consegue operar sob o radar do firewall, pois não viola regras formais — apenas manipula semântica operacional.

O firewall enxerga conformidade protocolar.O processo industrial sofre degradação funcional.

Esse desalinhamento entre visão técnica e impacto operacional é uma das maiores fragilidades da arquitetura tradicional.

4 EDRS E O PROBLEMA DOS FALSOS POSITIVOS EM ESCALA

EDRs foram desenvolvidos para superar limitações de antivírus tradicionais, incorporando análise comportamental e visibilidade em tempo real sobre endpoints. Contudo, enfrentam uma tensão permanente entre sensibilidade e precisão.

Ambientes industriais apresentam peculiaridades que desafiam modelos comportamentais genéricos:

- Atualizações esporádicas;
- Softwares legados;
- Processos de manutenção não padronizados;
- Acessos remotos emergenciais.

Essa variabilidade gera ruído operacional significativo.

Ataques baseados em IA exploram exatamente essa ambiguidade. Malwares podem operar com comportamento “low and slow”, ajustando-se para permanecer abaixo dos limiares configurados para evitar sobrecarga de alertas.

Além disso, técnicas como injeção de código em processos legítimos e execução em memória dificultam detecção por assinatura.

O resultado é paradoxal: quanto mais complexo o ambiente, maior o volume de alertas; quanto maior o volume, menor a capacidade humana de priorização eficaz.

O EDR detecta eventos.Mas não necessariamente entende o impacto sistêmico desses eventos.

5 SIEMS E A ILUSÃO DA CORRELAÇÃO SUFICIENTE

SIEMs consolidam logs de múltiplas fontes e aplicam regras de correlação para identificar padrões suspeitos. Entretanto, sua eficácia depende da antecipação de cenários.

Ataques impulsionados por IA são caracterizados por:

- Sequências inéditas de ações;
- Alternância dinâmica de vetores;
- Uso legítimo de credenciais comprometidas;
- Exploração de configurações válidas.

Se não há violação explícita de regra, o SIEM pode não gerar alerta crítico.

Além disso, a explosão de dados provenientes de dispositivos IoT e sensores industriais amplia exponencialmente o volume de eventos, dificultando modelagem precisa.

Sem contexto operacional, a correlação permanece técnica, não estratégica.

Um login fora de horário pode ser irrelevante em um ambiente administrativo e crítico em uma planta industrial sensível.

A ausência de contextualização transforma o SIEM em um repositório volumoso de dados, mas não necessariamente em uma ferramenta decisória eficaz.

6 COMO LLMS SUBVERTEM CONTROLES BASEADOS EM REGRAS

Modelos de linguagem de grande escala democratizaram capacidades antes restritas a especialistas em desenvolvimento de exploits e engenharia social.

Hoje é possível:

- Gerar código ofuscado sob demanda;

- Reescrever payloads para evitar detecção por hash;
- Criar campanhas de phishing hiperpersonalizadas;
- Automatizar análise de superfície de ataque.

A adaptabilidade é o diferencial crítico. Enquanto regras são fixas, LLMs produzem variações infinitas.

Além disso, a IA pode aprender com respostas defensivas. Caso um payload seja bloqueado, o modelo pode sugerir variações alternativas quase instantaneamente.

Esse ciclo iterativo reduz o tempo entre tentativa e sucesso, pressionando defesas reativas.

7 A NECESSIDADE DE DETECÇÃO BASEADA EM COMPORTAMENTO E CONTEXTO

Superar as limitações descritas exige mudança estrutural na lógica de detecção.

A análise deve transcender eventos isolados e considerar:

- Histórico comportamental de ativos;
- Interdependência entre sistemas;
- Criticidade do processo afetado;
- Consequências físicas potenciais.

Detecção comportamental eficaz requer aprendizado contínuo e modelagem específica por ambiente.

Em CPS, não basta saber que um comando foi executado; é necessário compreender se aquele comando é compatível com o estado operacional esperado do sistema naquele momento.

A defesa deixa de ser puramente técnica e passa a ser contextual-operacional.

8 DO MODELO REATIVO AO MODELO COGNITIVO DE DEFESA

O modelo reativo baseia-se em detectar e responder. O modelo cognitivo baseia-se em compreender, antecipar e orquestrar.

A introdução de IA defensiva como camada central permite:

- Correlação multivariada de eventos;
- Priorização por impacto;
- Simulação de cenários futuros;
- Automatização rastreável de resposta.

Nesse paradigma, ferramentas tradicionais tornam-se sensores de um sistema maior.

A inteligência reside na integração, não na ferramenta isolada.

9 IMPLICAÇÕES PARA SISTEMAS CIBERFÍSICOS (CPS)

Sistemas ciberfísicos combinam mundo digital e físico. Isso altera radicalmente a equação de risco.

Um incidente pode gerar:

- Interrupção de fornecimento de energia;
- Contaminação de água;
- Parada de produção industrial;
- Instabilidade em cadeias logísticas.

A segurança deixa de ser apenas proteção de dados e passa a ser proteção de processos críticos.

Ataques impulsionados por IA aumentam a probabilidade de manipulação sutil e progressiva de parâmetros industriais.

Sem inteligência contextual e capacidade preditiva, a detecção pode ocorrer apenas após o impacto físico.

Em CPS, tempo de detecção equivale a tempo de mitigação de danos reais.

10. CONCLUSÃO

O chamado “colapso das defesas convencionais” não deve ser interpretado como a falência tecnológica de firewalls, EDRs ou SIEMs. Essas soluções continuam sendo componentes fundamentais de qualquer arquitetura de segurança. O que colapsa, na realidade, é a lógica estrutural que sustenta sua aplicação isolada e predominantemente reativa diante de um adversário que opera sob paradigma adaptativo e probabilístico impulsionado por inteligência artificial.

O cenário atual caracteriza-se por uma assimetria inédita. Enquanto as organizações ainda operam majoritariamente com arquiteturas desenhadas para detectar padrões conhecidos, atacantes utilizam IA para criar padrões inéditos em tempo real. A defesa tradicional depende da previsibilidade; a IA ofensiva prospera na variabilidade. Essa tensão estrutural expõe um limite conceitual da segurança baseada exclusivamente em regras, assinaturas e correlação estática de eventos.

Em ambientes corporativos convencionais, esse descompasso já representa riscos significativos de vazamento de dados, fraude e interrupção operacional. Em sistemas ciberfísicos (CPS), contudo, as consequências extrapolam o domínio digital. A convergência entre TI, TO, IoT e nuvem transforma incidentes cibernéticos em potenciais eventos físicos, capazes de impactar produção industrial, distribuição de energia, abastecimento de água, transporte e cadeias logísticas críticas. Nesse contexto, a insuficiência de modelos reativos deixa de ser apenas problema técnico e passa a ser questão de resiliência sistêmica.

A análise desenvolvida ao longo deste artigo demonstra que a principal fragilidade das defesas convencionais não está na ausência de ferramentas, mas na ausência de integração cognitiva entre elas. A

fragmentação entre camadas técnicas produz grande volume de dados, mas não necessariamente inteligência contextual. Alertas isolados, mesmo tecnicamente corretos, tornam-se irrelevantes se não forem correlacionados ao impacto real sobre processos operacionais. O excesso de sinais sem priorização estratégica gera fadiga analítica, reduz capacidade de resposta e cria zonas de invisibilidade exploráveis por atacantes adaptativos.

A emergência de LLMs e algoritmos ofensivos baseados em aprendizado de máquina adiciona nova dimensão ao problema. A capacidade de gerar código dinâmico, modular ataques conforme a reação do ambiente e executar campanhas de engenharia social hiperpersonalizadas reduz drasticamente o tempo de iteração entre tentativa e sucesso. O atacante aprende com a defesa. Se a defesa não aprende com o ambiente na mesma velocidade, a assimetria amplia-se progressivamente.

Diante desse cenário, torna-se imperativo abandonar a ilusão de que a simples multiplicação de camadas técnicas elevará indefinidamente o nível de proteção. Mais camadas estáticas não compensam ausência de inteligência contextual. O que se impõe é a transição para um modelo cognitivo de defesa, no qual a inteligência artificial defensiva ocupa posição central, não periférica.

Esse modelo implica transformação estrutural:

Primeiro, a segurança precisa ser orientada por comportamento e contexto operacional, não apenas por severidade técnica. Um evento deve ser avaliado segundo sua relevância para a continuidade do processo, criticidade do ativo envolvido e potencial de impacto físico ou econômico.

Segundo, a arquitetura deve permitir aprendizado contínuo do ambiente. O histórico operacional passa a ser elemento essencial de modelagem de normalidade, permitindo que desvios significativos sejam identificados mesmo quando não correspondem a assinaturas conhecidas.

Terceiro, a resposta deve ser automatizada de forma rastreável e governável. A velocidade dos ataques impulsionados por IA não é compatível com modelos exclusivamente manuais de decisão. Entretanto, automação sem governança gera novos riscos. O equilíbrio reside em respostas inteligentes, parametrizadas e auditáveis.

Quarto, a segurança deve ser integrada ao negócio e à operação industrial. Em CPS, proteção não pode ser tratada como camada externa. Ela precisa compreender processos físicos, fluxos de energia, ciclos de produção e interdependências sistêmicas. A defesa passa a ser parte da engenharia operacional.

A transição para esse paradigma não é trivial. Envolve revisão arquitetural, maturidade organizacional, capacitação técnica e mudança cultural. Contudo, a alternativa — manter-se dependente de modelos concebidos para uma realidade anterior — representa risco estrutural crescente.

Em última análise, o futuro da segurança cibernética será definido pela capacidade de as organizações internalizarem que a inteligência é o novo perímetro. Não se trata de substituir tecnologias existentes, mas de integrá-las sob lógica superior de compreensão contextual e adaptação contínua. Ferramentas deixam de ser barreiras isoladas e passam a ser sensores de um sistema cognitivo maior.

A batalha contra a IA ofensiva não será vencida por mais regras, mais assinaturas ou mais alertas. Será vencida por arquiteturas capazes de aprender, antecipar e orquestrar defesa com a mesma velocidade e sofisticação do adversário.

Em sistemas ciberfísicos, onde o digital controla o físico, essa evolução não é apenas desejável. É condição para a sustentabilidade operacional, a resiliência econômica e a segurança da sociedade.

A defesa do futuro não será apenas tecnológica. Será metodológica, contextual e inteligente.