O mercado negro da inteligência artificial: cibercriminosos já vendem serviços de IA para elaborar malware, campanhas de phishing e ataques DDoS

Autores: Marcelo Branquinho e Plataforma Safer

29/08/2025

Resumo: O presente artigo explora o surgimento de um mercado clandestino de **inteligência artificial (IA)** voltado a atividades ilícitas. Nele, cibercriminosos comercializam serviços de IA para **desenvolvimento de malware sob medida, automatização de campanhas de phishing altamente convincentes e condução de ataques de negação de serviço distribuídos (DDoS)** em larga escala. Abordam-se as principais modalidades de uso malicioso de IA – da geração de códigos maliciosos polimórficos à criação de *deepfakes* para fraudes –, bem como as implicações para a segurança de infraestruturas críticas e empresas globais. Por fim, discutem-se estratégias de defesa, destacando a necessidade de contramedidas que empreguem a própria IA para combater ameaças cada vez mais sofisticadas.

Palavras-chave: Inteligência artificial, mercado negro, cibercrime, malware, phishing, DDoS, deepfake, segurança cibernética.

1. Introdução

Nos últimos anos, a inteligência artificial (IA) deixou de ser exclusividade de laboratórios e big techs para figurar também no arsenal de cibercriminosos. Surge assim um **mercado negro da IA** no qual criminosos oferecem ferramentas e serviços baseados em IA para potencializar ataques digitais. Modelos de linguagem generativa sem restrições éticas estão sendo empregados para **criar malware sob medida, redigir e-mails de phishing altamente convincentes e otimizar ataques de negação de serviço distribuídos (DDoS)**. O que antes poderia soar como ameaça distante hoje já é uma realidade concreta.

Essa convergência entre IA e cibercrime está reconfigurando o cenário de ameaças, **reduzindo as barreiras de habilidade e custo para invasores**. Até criminosos com pouca experiência técnica agora podem lançar campanhas maliciosas sofisticadas ao empregar assistentes de IA treinados para fins ilícitos. Para gestores de cibersegurança e líderes de empresas – sobretudo em infraestruturas críticas –, compreender essa tendência é crucial. Este artigo explora como a IA está sendo explorada no submundo do crime digital e os impactos disso na segurança corporativa, além de discutir medidas de mitigação para enfrentar essas **ameaças emergentes impulsionadas por IA**.

2. IA no Ecossistema do Cibercrime

O cibercrime contemporâneo funciona como um negócio estruturado, com criminosos vendendo kits de ataque e acessos ilícitos na dark web em um modelo de "Crime como Serviço". Nesse contexto, a IA emergiu como mais um produto valioso do submundo. Desenvolvedores clandestinos criam **ferramentas de IA sob medida para o crime** e as comercializam para outros agentes maliciosos – seja para automatizar golpes de phishing, escrever malware, burlar captchas ou aprimorar ataques de força bruta.

As primeiras evidências dessa tendência vieram no fim de 2022: pesquisadores flagraram discussões em fóruns hackers sobre o uso do ChatGPT para escrever códigos maliciosos e scripts de fraude. Em janeiro de 2023, especialistas já alertavam que, embora os casos iniciais fossem relativamente básicos, era "apenas uma questão de

tempo” até cibercriminosos mais avançados aperfeiçoarem o uso da IA para fins malignos. De lá para cá, a profecia se concretizou. **Ferramentas de IA generativa sem filtros, treinadas para auxiliar em crimes, explodiram em popularidade** no último ano. Em 2024, menções a plataformas como **WormGPT**, **DarkGPT** e **FraudGPT** dispararam cerca de 200% nos fóruns clandestinos em comparação a 2023, evidenciando o interesse crescente nesses recursos. A IA tornou-se, definitivamente, parte integrante do arsenal dos atacantes, baixando o nível de habilidade necessário para se lançar no cibercrime e abrindo espaço para ataques mais numerosos e sofisticados.

3. Malware Sob Medida com IA

A criação de malware é um dos campos onde a IA maliciosa mais vem mostrando suas capacidades. Mesmo criminosos sem habilidades avançadas de programação já conseguem gerar código nocivo funcional com auxílio de assistentes de IA. Por exemplo, em fóruns hackers um usuário divulgou um **script “infostealer” (ladrão de dados)** escrito inteiramente pelo ChatGPT, capaz de vasculhar e exfiltrar documentos do computador da vítima. Outro caso envolveu um novato que usou o ChatGPT para desenvolver uma **ferramenta de criptografia de arquivos em Python**, admitindo ser seu primeiro código escrito graças à “boa ajuda” da IA. Esses casos ilustram como modelos generativos permitem que atores de baixo nível técnico criem malwares básicos.

Já cibercriminosos mais experientes procuram elevar o patamar: **ferramentas clandestinas como o FraudGPT anunciam a capacidade de produzir malwares “indetectáveis”** e encontrar vulnerabilidades em sistemas automaticamente. O uso de IA possibilita também a geração de **malware polimórfico em escala** – ou seja, códigos maliciosos que se reescrevem constantemente para escapar de antivírus. Tarefas complexas como criar variações mutantes de vírus, antes restritas a programadores especializados, agora podem ser realizadas em segundos por meio de motores de IA como DarkGPT e WormGPT. Com isso, malwares alimentados por IA tendem a se tornar mais **adaptáveis e difíceis de detectar** pelas defesas tradicionais, pois a cada execução podem surgir em uma forma inédita. Em suma, a IA passou a funcionar como um “consultor

técnico” do cibercriminoso, acelerando o desenvolvimento de códigos maliciosos e ampliando seu potencial destrutivo.

4. Phishing e Engenharia Social com IA

As campanhas de **phishing** – envios de comunicações fraudulentas para enganar usuários – tornaram-se muito mais perigosas com a ajuda da IA. Tradicionalmente, muitos e-mails de phishing denunciavam-se por erros de ortografia ou traduções incorretas. Agora, porém, **modelos de linguagem conseguem redigir mensagens impecáveis em qualquer idioma**, ajustando tom e contexto para imitar comunicações legítimas. Ataques de **Business Email Compromise (BEC)**, nos quais golpistas se fazem passar por executivos de alto escalão em e-mails urgentes, ganharam eficiência com esses modelos. O WormGPT, por exemplo, foi divulgado em fóruns como um assistente capaz de escrever e-mails persuasivos para fraudes do tipo BEC e phishing em massa. De forma geral, criminosos já utilizam IA para automatizar a criação de e-mails falsos extremamente convincentes e personalizar mensagens de texto maliciosas em larga escala, eliminando sinais típicos de golpe. Um estudo da *Harvard Business Review* demonstrou que ataques de phishing automatizados por IA podem **reduzir em 95% os custos** das campanhas maliciosas, obtendo taxas de sucesso iguais ou até superiores às dos esquemas tradicionais.

Além do texto, a IA também ampliou as fronteiras da **engenharia social visual e sonora**. Hoje é trivial para um fraudador gerar **deepfakes de voz ou vídeo** imitando pessoas reais. Já houve caso de criminosos usando uma voz sintética de um diretor financeiro para convencer um funcionário a transferir US$ 25 milhões, num golpe de alto nível de sofisticação. Em fóruns, circulam tutoriais sobre como aproveitar ferramentas de áudio *deepfake* para imitar a voz de CEOs em chamadas telefônicas de fraude. O resultado são engodos cada vez mais realistas: um funcionário desavisado pode receber um recado de voz que **soa exatamente como seu gerente**, pedindo uma ação urgente, e acreditar genuinamente que se trata de uma instrução legítima. Autoridades já soam o alarme – no fim de 2024, o FBI alertou para o uso crescente de IA generativa em esquemas de fraude online, que os torna mais convincentes e amplos em alcancekelacyber.com. Diante dessa evolução, confiar em julgamentos baseados apenas na

aparência ou voz não é mais seguro; empresas precisam adotar verificações adicionais de autenticidade em comunicações sensíveis.

5. Ataques Automatizados e DDoS com IA

A inteligência artificial também vem sendo aplicada para tornar ataques automatizados mais eficientes e devastadores. Um exemplo é na **quebra de senhas e credenciais**: algoritmos de *machine learning* podem ser treinados em bases de dados de senhas vazadas para **prever novas senhas** prováveis, acelerando ataques de força bruta ou tentativas de login em massa (*credential stuffing*). Da mesma forma, ferramentas de IA ajudam a automatizar a varredura de sistemas em busca de vulnerabilidades, agilizando o processo de identificar brechas exploráveis antes que equipes de defesa consigam corrigi-las. Em suma, tarefas que exigiam tempo e esforço humano – como testar milhares de combinações de senha ou mapear pontos fracos de rede – agora podem ser executadas em alta velocidade por bots inteligentes.

No contexto de ataques de **negação de serviço distribuído (DDoS)**, a IA permite otimizar estratégias e ampliar a escala de forma impressionante. Redes de bots (*botnets*) tradicionalmente são alugadas no mercado negro para sobrecarregar alvos com tráfego, mas com algoritmos inteligentes coordenando essas botnets, os ataques se tornam mais **adaptativos e difíceis de mitigar**. Relatos indicam que bots alimentados por IA conseguem gerar até **200 vezes mais requisições contra um alvo** do que ataques anteriores, atingindo uma intensidade sem precedentes. Além disso, esses bots avançados conseguem **driblar mecanismos de segurança automatizados**, como CAPTCHAs e filtros anti-bot, adaptando seu comportamento para se passar por tráfego legítimo. Isso significa que defesas tradicionais contra DDoS e automações maliciosas – baseadas em identificar padrões repetitivos de bots – podem falhar diante de ataques conduzidos por IA. Organizações precisam, portanto, reconsiderar suas estratégias de proteção contra uma nova geração de ameaças automatizadas e inteligentes.

6. Serviços Clandestinos de IA: WormGPT, FraudGPT e Outros

Os exemplos mais emblemáticos do “mercado negro da IA” são as recentes plataformas WormGPT e FraudGPT. Em junho de 2023, um desenvolvedor anônimo lançou o **WormGPT**, apresentando-o como

uma “alternativa do ChatGPT para hackers *black hat*”. Construído sobre um modelo de linguagem open source (GPT-J 6B), o WormGPT prometia oferecer respostas de IA **sem nenhum filtro moral ou restrição de segurança**, aptas a auxiliar na programação de malwares e na elaboração de golpes. Seu acesso foi comercializado via fóruns da dark web e Telegram por cerca de US$ 100 ao mês, ou até US$ 5.000 por uma instância dedicada privada. Dias depois, outro vendedor sob o pseudônimo “CanadianKingpin12” anunciou o **FraudGPT**, promovido de forma similar em mercados clandestinos e canais fechados. O mesmo indivíduo chegou a divulgar variantes batizadas de **DarkBERT**, **DarkBARD** e **DarkGPT**, tentando diversificar a oferta de IAs voltadas ao crime.

De modo geral, essas ferramentas de IA ilegais compartilham funcionalidades: **geração de código malicioso sob demanda, criação de páginas de phishing personalizadas, descoberta de vulnerabilidades exploráveis** e até sugestões para fraudes financeiras. Em suma, oferecem uma espécie de “assistente do hacker”, suprindo etapas técnicas de um ataque. O FraudGPT, por exemplo, foi anunciado como capaz de produzir *phishing* kits e encontrar números de cartões de crédito que escapassem de verificações de segurança, sendo ofertado por assinatura em torno de **US$ 200 mensais** (ou US$ 1.700 ao ano). Vale notar que, como em qualquer mercado ilícito em ascensão, também surgiram golpistas vendendo **IA falsas**: canais no Telegram e sites prometendo ferramentas milagrosas que jamais são entregues, apenas para enganar interessados desavisados.

Outro desafio desse tipo de “negócio” é a volatilidade diante de escrutínio. Em agosto de 2023, poucas semanas após o lançamento, o criador do WormGPT anunciou abruptamente o fim do projeto e apagou todo o conteúdo relacionado, possivelmente reagindo à exposição de sua identidade real por pesquisadores de segurança. De forma similar, o operador do FraudGPT removeu suas postagens e desapareceu dos fóruns, indicando temor de ser rastreado. Contudo, a demanda subjacente permanece: rapidamente **novos bots como “EvilGPT” e “WolfGPT”** surgiram reivindicando o posto de “próximo WormGPT” (ainda que em versões mais simples), sinalizando que o comércio de IA maliciosa continua evoluindo no subterrâneo digital.

7. Estratégias de Defesa e Mitigação

Os avanços da IA exigem uma evolução correspondente nas defesas cibernéticas. Especialistas apontam que **ferramentas de segurança tradicionais tornam-se insuficientes contra ataques turbinados por IA**, sendo necessária uma postura mais proativa e inovadora. Em vez de apenas reagir a incidentes conhecidos, as organizações devem antecipar-se e **empregar a própria IA para reforçar a segurança** – o conceito de "fight fire with fire" ou "combater IA com IA". Abaixo, listamos algumas estratégias recomendadas para enfrentar essa nova realidade:

- **IA na Ciberdefesa**: Adotar soluções de segurança que incorporem **inteligência artificial defensiva**, capazes de analisar grandes volumes de dados em busca de padrões anômalos que indiquem ataques sutis. Ferramentas de detecção de intrusão e sistemas de monitoramento comportamental baseados em IA podem reagir a táticas evasivas que passariam despercebidas por filtros tradicionais.
- **Treinamento e Procedimentos**: Reforçar a **conscientização dos colaboradores** quanto a ameaças potencializadas por IA. Isso inclui alertá-los de que e-mails bem escritos podem ser fraudulentos e que **vozes ao telefone podem ser imitações digitais**. Estabelecer protocolos de verificação adicional para solicitações sensíveis – por exemplo, confirmar ordens de pagamento acima de certo valor por um segundo canal – para reduzir riscos de fraude via phishing ou deepfake.
- **Verificação de Conteúdo e Identidade**: Investir em tecnologias emergentes de **detecção de deepfakes** em áudio e vídeo, bem como em sistemas de verificação de mensagens geradas por IA. Ao receber um áudio suspeito de um executivo, por exemplo, é recomendável submetê-lo a análise para verificar autenticidade. Da mesma forma, e-mails podem ser avaliados por filtros que buscam traços de geração artificial ou contexto incompatível.
- **Inteligência de Ameaças**: Acompanhar ativamente fóruns clandestinos e relatórios de inteligência para **identificar novas ferramentas de IA maliciosa** assim que elas surjam. Estar ciente das táticas que os criminosos estão comercializando (como WormGPT, FraudGPT e sucessores) permite atualizar controles de segurança e bloquear indicadores relacionados a essas

ferramentas (ex: padrões de linguagem, domínios de phishing conhecidos etc.).

- **Resiliência e Redundância**: Preparar planos de resposta que considerem cenários de ataques aprimorados por IA. Por exemplo, em caso de um ataque DDoS inteligente, ter acordos com provedores de mitigação com capacidade elástica e filtros adaptativos. Assegurar que políticas de bloqueio de contas após múltiplas tentativas de login estejam calibradas para resistir a ataques de força bruta acelerados por IA, sem prejudicar usuários legítimos.

Em essência, a defesa cibernética precisa tornar-se **mais ágil, inteligente e multicamadas** na era da IA maliciosa. Isso inclui combinar medidas tecnológicas avançadas com processos humanos robustos de verificação. Ao mesmo tempo, é fundamental manter-se atualizado sobre as tendências de ataques baseados em IA e promover uma cultura de cibersegurança onde todos estejam alertas para sinais de atividades suspeitas, por mais autênticas que pareçam.

8. Conclusão

O surgimento de uma "indústria sombra" de inteligência artificial voltada ao crime cibernético evidencia como a inovação tecnológica pode ser subvertida para fins maliciosos em larga escala. Ao longo deste artigo, vimos que **modelos de IA generativa estão sendo usados para criar malware mais sofisticado, tornar ataques de phishing quase indistinguíveis de comunicações legítimas e automatizar ofensivas digitais com eficiência inédita**. Essa realidade impõe um novo paradigma de riscos para empresas e governos: agora enfrentamos **adversários potencializados por IA**, capazes de amplificar sua taxa de sucesso e reduzir custos de ataque de forma dramática.

Por outro lado, também ficou claro que a mesma agilidade e potência proporcionadas pela IA aos criminosos podem – e devem – ser contrapostas com IA defensiva e estratégias de segurança aprimoradas. Ignorar essa tendência não é opção: as evidências indicam que a popularidade e a lucratividade das ferramentas de IA no submundo do crime só aumentarão, fomentando uma corrida armamentista digital. Portanto, **o momento de agir é agora**. Executivos

de alto escalão, gestores de cibersegurança e profissionais da área precisam incorporar essa perspectiva em seus planos de proteção, investindo em inovação responsável e mantendo-se informados sobre as ameaças emergentes.

Em suma, o mercado negro da IA representa um divisor de águas na cibersegurança. Enfrentá-lo exigirá uma combinação de tecnologia avançada, políticas robustas e conscientização humana. Aqueles que estiverem preparados para **evoluir suas defesas na mesma velocidade em que os atacantes evoluem suas ferramentas** estarão mais aptos a preservar a segurança de suas informações e operações. A batalha contra a criminalidade digital entra numa nova fase – e a capacidade de **adaptar-se e inovar** será o fator decisivo para prevalecer neste embate entre ataques e defesas movidos por inteligência artificial.

Referências

1. KELA Cyber. *AI Jailbreaking Interest Surged 50% in 2024 – How Defenders Can Keep Up*. Relatório de Inteligência, 31 de mar. 2025. Disponível em: **https://www.kelacyber.com/blog/ai-jailbreaking-interest-surge/**.
2. Cloudflare. *Combating AI-powered threats – Building resilience in the new AI era of cybersecurity*. *The NET* (Cloudflare), 2023. Disponível em: **https://www.cloudflare.com/the-net/security-signals/ai-powered-threats/**.
3. **Inforchannel** – Redação. *Cibercriminosos adotam o ChatGPT para desenvolver ferramentas maliciosas*. 10 jan. 2023. Disponível em: **https://inforchannel.com.br/2023/01/10/cibercriminosos-adotam-o-chatgpt-para-desenvolver-ferramentas-maliciosas/**.
4. Burgess, Matt. *Criminals Have Created Their Own ChatGPT Clones*. *Wired*, 7 ago. 2023. Disponível em: **https://www.wired.com/story/chatgpt-scams-fraudgpt-wormgpt-crime/**.
5. Moraes, Lucas C. *Malware Polimórfico: A Nova Era das Ameaças com IA*. Blog Dolutech, 27 jul. 2025. Disponível em: **https://dolutech.com/malware-polimorfico-com-ia-darkgpt-e-novas-ameacas/**.

6. Fan, Ricardo. *O mercado negro do cibercrime: preços e serviços oferecidos na dark web*. *DefesaNet*, 18 jan. 2019. Disponível em: **https://www.defesanet.com.br/sof/o-mercado-negro-do-cibercrime-precos-e-servicos-oferecidos-na-dark-web/**.
7. Del Val, Margarita. *Dark AI tools: How profitable are they on the dark web?* Outpost24 (KrakenLabs Research), 04 jul. 2025. Disponível em: **https://outpost24.com/blog/dark-ai-tools/**.