



Relatório de segurança de IA da Zscaler ThreatLabz de 2024



A revolução da IA chegou. Descubra as principais tendências, riscos e práticas recomendadas na adoção da IA corporativa, com informações sobre ameaças baseadas em IA e estratégias-chave para se defender contra elas.

Índice

03 Resumo executivo

04 Principais conclusões

05 Principais tendências de uso de GenAI e ML

- 05 As transações de IA continuam a acelerar
- 06 As empresas estão bloqueando mais transações de IA do que nunca
- 07 **Análise de IA por setor**
 - 09 Saúde e IA
 - 10 Finanças
 - 11 Governo
 - 12 Manufatura
 - 13 Educação e IA
- 14 **Tendências de uso do ChatGPT**
- 15 **Uso de IA por país**
 - Divisão regional: EMEA
 - Divisão regional: APAC

18 Cenários de risco de IA para empresas e ameaças do mundo real

- 18 Oferecer IA na empresa: os três principais riscos
- 20 Cenários de ameaças baseados em IA
 - Representação de IA: deepfakes, desinformação e mais
- 21 Campanhas de phishing geradas por IA
 - Da consulta ao crime: criação de uma página de login de phishing usando o ChatGPT
- 22 Dark chatbots: descobrindo o WormGPT e o FraudGPT na dark web

- 23 Malware e ransomware baseados em IA em toda a cadeia de ataque
- 24 Ataques de worm de IA e jailbreak de IA “viral”
- 25 IA e eleições nos EUA

26 Atenção total aos regulamentos de IA

- 26 Estados Unidos
- 27 União Europeia

28 Previsões sobre ameaças de IA

31 Estudo de caso: como ativar o ChatGPT com segurança na empresa

- 31 Cinco etapas para integrar e proteger ferramentas generativas de IA

33 Como a Zscaler oferece IA com zero trust e protege a IA generativa

- 33 O principal para a segurança cibernética baseada na IA: dados de alta qualidade em larga escala
- 34 Aproveitando a IA em toda a cadeia de ataque
- 35 Resumo dos produtos com uso de IA da Zscaler
- 36 Possibilitando a transição para IA empresarial: o controle está em suas mãos

37 Apêndice

- 37 Metodologia de pesquisa da ThreatLabz

37 Sobre a Zscaler ThreatLabz

Resumo executivo

A IA é mais do que uma inovação pioneira: é o novo normal. À medida que ferramentas de IA generativa, como o ChatGPT, transformam os negócios em pequena e larga escala, a IA está sendo profundamente integrada na estrutura da vida empresarial. No entanto, as questões sobre como adotar com segurança essas ferramentas de IA e, ao mesmo tempo, defender-se contra ameaças impulsionadas pela IA não foram resolvidas.

As empresas estão adotando rapidamente ferramentas de IA e ML em departamentos como engenharia, marketing de TI, finanças, sucesso do cliente e mais. No entanto, elas devem equilibrar os numerosos riscos que acompanham as ferramentas de IA para colherem todos os benefícios. Na verdade, para desbloquear o potencial transformador da IA, as empresas devem implementar controles seguros para proteger seus dados, evitar vazamento de informações sigilosas, mitigar a expansão da “IA invisível” e garantir a qualidade dos dados da IA.

Esses riscos da IA para as empresas são bidirecionais: **fora dos muros das empresas, a IA tornou-se uma força motriz para ameaças cibernéticas.** Na verdade, as ferramentas de IA estão permitindo que cibercriminosos e agentes patrocinados por estados-nação lancem ataques sofisticados, mais rapidamente e em maior escala. Apesar disso, a IA é promissora como peça-chave do quebra-cabeça da defesa cibernética, à medida que as empresas enfrentam um cenário de ameaças dinâmico.

O relatório de segurança de IA da ThreatLabz de 2024 oferece informações importantes sobre esses desafios e oportunidades críticos da IA.

Com base em mais de 18 bilhões de transações entre abril de 2023 e janeiro de 2024 na Zscaler Zero Trust Exchange™, a ThreatLabz analisou como as empresas estão usando ferramentas de IA e ML atualmente. Essas informações revelam as principais tendências em todos os setores comerciais e regiões geográficas sobre como as empresas estão se adaptando ao cenário dinâmico da IA e protegendo suas ferramentas de IA.

Ao longo do relatório, você encontrará informações sobre os principais tópicos de IA, incluindo riscos de negócios, cenários de ameaças e táticas de adversários baseados em IA, considerações regulatórias e previsões para o cenário da IA em 2024 e além.

Da mesma forma, este relatório oferece práticas recomendadas em duas frentes: como as empresas podem adotar com segurança a transformação da IA generativa, ao mesmo tempo que protegem dados críticos, e como as ferramentas baseadas em IA trabalham para fornecer segurança zero trust em camadas para enfrentar o novo cenário de ameaças baseadas em IA.

Principais descobertas



O uso de ferramentas de IA/ML disparou em **594,82%**, subindo de 521 milhões de transações baseadas em IA/ML em abril de 2023 para 3,1 bilhões mensais até janeiro de 2024.



Os aplicativos de IA mais utilizados por volume de transações são **ChatGPT, Drift, OpenAI*, Writer e LivePerson**. Os **três principais aplicativos bloqueados** por volume de transações são **ChatGPT, OpenAI e Fraud.net**.



As empresas estão bloqueando **18,5%** de todas as transações de IA/ML, um aumento de **577%** nas transações bloqueadas em nove meses, refletindo preocupações crescentes em torno da segurança de dados de IA e a relutância das empresas em estabelecer políticas de IA.



Os **cinco principais países** que mais geram transações de IA e ML são os EUA, a Índia, o Reino Unido, a Austrália e o Japão.



A manufatura gera a maior parte do tráfego de IA, com **20,9%** of todas as transações de IA/ML na nuvem da Zscaler, seguidas por finanças e seguros (19,9%) e serviços (16,8%).



As empresas estão enviando volumes significativos de dados para ferramentas de IA, com um total de **569 TB** trocados entre aplicativos de IA/ML entre setembro de 2023 e janeiro de 2024.



O uso do ChatGPT continua a crescer, com um crescimento de **634,1%**, embora seja também o aplicativo de IA mais **bloqueado** pelas empresas, com base nas informações da nuvem da Zscaler.



A IA está capacitando os criminosos de maneiras **sem precedentes**, inclusive para campanhas de phishing baseadas em IA, deepfakes e ataques de engenharia social, ransomware polimórfico, descoberta de superfície de ataque empresarial, geração automatizada de exploits e mais.

OBSERVAÇÃO: a Zscaler Zero Trust Exchange rastreia transações do ChatGPT independentemente de outras transações da OpenAI em geral.

Principais tendências de uso de GenAI e ML

A revolução da IA empresarial está longe do seu auge. As transações de IA empresarial aumentaram quase 600% e não mostram sinais de desaceleração. Ainda assim, as transações bloqueadas para aplicativos de IA também aumentaram — em 577%.

As transações de IA continuam a acelerar

De abril de 2023 a janeiro de 2024, as transações empresariais de IA e ML cresceram quase 600%, aumentando para mais de 3 bilhões de transações mensais na Zero Trust Exchange em janeiro. Isso destaca o fato de que, apesar do número crescente de incidentes de segurança e riscos de dados associados à adoção da IA pelas empresas, o seu potencial transformador é grande demais para ser ignorado. Observe que, embora as transações de IA tenham sofrido uma breve pausa durante os feriados de dezembro, as transações continuaram em um ritmo ainda maior no início de 2024.

No entanto, mesmo com a proliferação dos aplicativos de IA, a maioria das transações de IA são conduzidas por um conjunto relativamente pequeno de ferramentas de IA líderes de mercado. No geral, o ChatGPT é responsável por mais da metade de todas as transações de IA e ML, enquanto o próprio aplicativo da OpenAI vem em terceiro lugar, com 7,82% de todas as transações. Enquanto isso, o Drift, o popular chatbot com tecnologia de IA, gerou quase um quinto do tráfego da IA empresarial (os chatbots LivePerson e BoldChat Enterprise também estiveram entre os principais aplicativos nas posições 5 e 6). Enquanto isso, o Writer continua sendo uma das principais ferramentas de IA generativa na criação de conteúdo corporativo por escrito, como materiais de marketing. Por fim, o Otter, uma ferramenta de transcrição de IA frequentemente usada em videochamadas, direciona uma parte significativa do tráfego de IA.

Tendências de transações de IA e ML



FIGURA 1 Transações de IA de abril de 2023 a janeiro de 2024

Principais aplicativos de IA

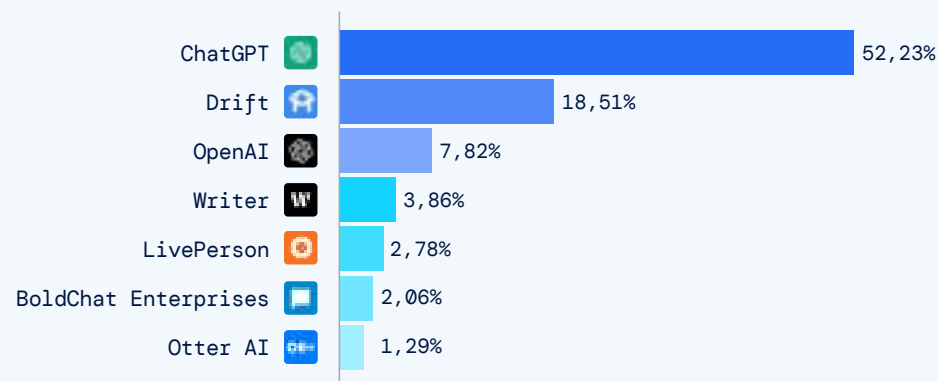


FIGURA 2 Principais aplicativos de IA por volume de transações

Dados transferidos por tráfego de IA/ML [setembro de 2023 a janeiro de 2024]

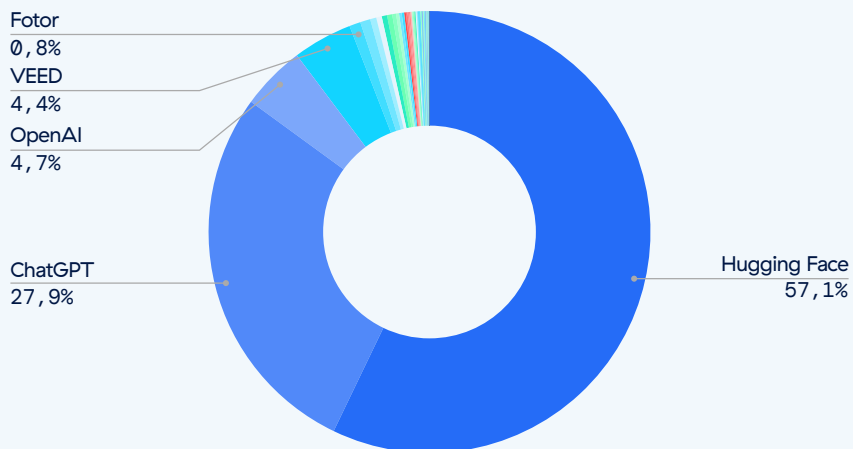


FIGURA 3 Principais aplicativos de IA/ML pela porcentagem do total de dados transferidos

Tendências de transações de IA bloqueadas [abril de 2023 a janeiro de 2024]



FIGURA 4 Número de transações de IA/ML bloqueadas ao longo do tempo

Entretanto, os volumes de dados que as empresas enviam e recebem a partir de ferramentas de IA acrescentam nuances a estas tendências. Hugging Face, a plataforma de desenvolvimento de IA de código aberto frequentemente descrita como “o GitHub da IA”, é responsável por quase 60% dos dados corporativos transferidos por ferramentas de IA. Como o Hugging Face permite que os usuários hospedem e treinem modelos de IA, faz sentido que ele capture volumes significativos de dados de usuários corporativos.

Embora o ChatGPT e a OpenAI façam aparições esperadas nesta lista, duas adições notáveis são o Veed, um editor de vídeo de IA frequentemente usado para adicionar legendas, imagens e outros textos a vídeos, e o Fotor, uma ferramenta usada para gerar imagens de IA, entre outros usos. Como vídeos e imagens envolvem tamanhos de arquivo grandes em comparação com outros tipos de solicitações, não é surpreendente ver esses dois aplicativos representados.

As empresas estão bloqueando mais transações de IA do que nunca

Mesmo que a adoção da IA pelas empresas continue a aumentar, as organizações bloqueiam cada vez mais as transações de IA e ML devido a preocupações com os dados e a segurança. Atualmente, as empresas bloqueiam 18,5% de todas as transações de IA, um aumento de 577% de abril a janeiro, com um total de mais de 2,6 bilhões de transações bloqueadas.

Algumas das ferramentas de IA mais populares também são as mais bloqueadas. Na verdade, o ChatGPT tem a particularidade de ser o aplicativo de IA mais usado e também o mais bloqueado. Isso indica que apesar, ou mesmo por causa, da popularidade dessas ferramentas, as empresas estão trabalhando ativamente para proteger o seu uso contra perda de dados e preocupações de privacidade. Outra tendência notável é que o [bing.com](https://www.bing.com), que possui a funcionalidade de IA Copilot, é bloqueado de abril a janeiro. Na verdade, o [bing.com](https://www.bing.com) é responsável por 25,02% de todas as transações bloqueadas de domínios de IA e ML.

Algumas das ferramentas de IA mais populares também são as mais bloqueadas. Na verdade, o ChatGPT tem a particularidade de ser o aplicativo de IA mais usado e também o mais bloqueado. Isso indica que apesar, ou mesmo por causa, da popularidade dessas ferramentas, as empresas estão trabalhando ativamente para proteger o seu uso contra perda de dados e preocupações de privacidade. Outra tendência notável é que o [bing.com](https://www.bing.com) é bloqueado mais do que qualquer outro domínio, com um total de 835.811.952 bloqueios de abril a janeiro. Na verdade, o [bing.com](https://www.bing.com) é responsável por 25,02% de todas as transações bloqueadas de domínios de IA e ML.



FIGURA 5 Principais aplicativos e domínios de IA bloqueados por volume de transações

Análise de IA por setor

Os setores verticais da indústria empresarial mostram diferenças notáveis na adoção geral de ferramentas de IA, bem como na proporção de transações de IA que bloqueiam. A manufatura é a líder absoluta, gerando mais de 20% das transações de IA e ML em toda a Zero Trust Exchange. Ainda assim, os setores de finanças e seguros, tecnologia e serviços vêm logo atrás. Juntos, esses quatro setores ultrapassaram os outros como as mais agressivas adotantes de IA.

Participação de transações de IA por setor vertical

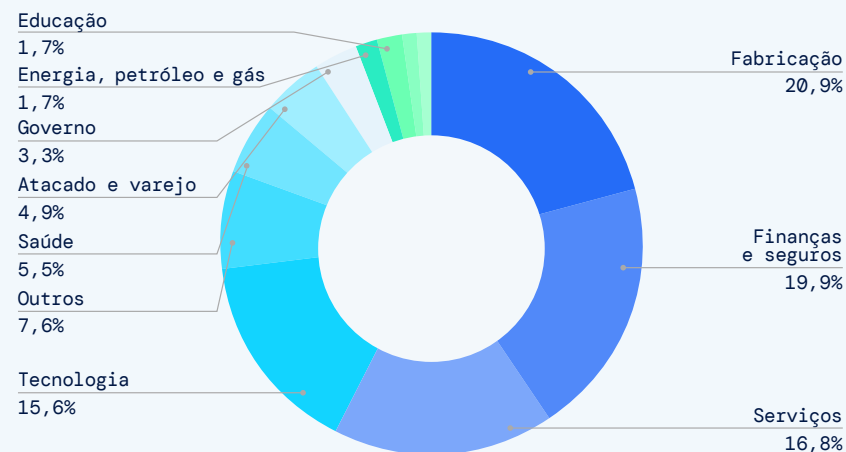


FIGURA 6 Setores que promovem as maiores proporções de transações de IA

Tendências de transações de IA por vertical

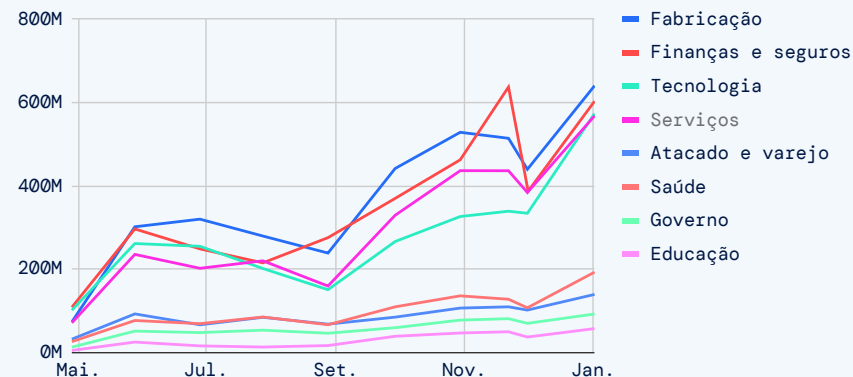


FIGURA 7 Tendências de transações de IA/ML entre os setores de maior volume, abril de 2023 a janeiro de 2024

Protegendo transações de IA/ML

Juntamente com o forte aumento nas transações de IA, os setores industriais estão bloqueando mais transações de IA. Aqui, alguns setores divergem das suas tendências gerais de adoção, refletindo diferentes prioridades e níveis de maturidade em termos de segurança de ferramentas de IA. O setor financeiro e de seguros, por exemplo, bloqueia a maior proporção das transações de IA: 37,2%, em comparação à média global de 18,5%. Isto provavelmente se deve, em grande parte, ao rigoroso ambiente regulatório e de conformidade do setor, combinado com os dados financeiros e pessoais altamente sigilosos dos usuários que essas organizações processam.

Enquanto isso, o setor de manufatura bloqueia 15,7% das transações de IA, apesar do seu papel descomunal na condução das transações globais de IA. O setor tecnológico, um dos primeiros e mais ávidos a adotar a IA, optou por uma espécie de caminho intermédio, bloqueando 19,4% das transações de IA, acima da média, à medida que trabalha para aumentar a adoção da IA. Surpreendentemente, o setor da saúde bloqueia 17,2% das transações de IA, abaixo da média, apesar dessas organizações processarem uma vasta quantidade de dados de saúde e informações de identificação pessoal (PII). Essa tendência provavelmente reflete um esforço lento das organizações de saúde para proteger dados sigilosos envolvidos em ferramentas de IA, à medida que as equipas de segurança tentam acompanhar a inovação da IA. Em geral, as transações de IA na área da saúde permanecem comparativamente baixas.

FIGURA 8 Principais setores verticais por percentagem de transações de IA bloqueadas

Porcentagem de transações de IA bloqueadas por setor vertical

Vertical	% de transações de IA bloqueadas
Finanças e seguros	37,16
Fabricação	15,65
Serviços	13,17
Tecnologia	19,36
Saúde	17,23
Atacado e varejo	10,52
Outros	8,93
Energia, petróleo e gás	14,24
Governo	6,75
Transporte	7,90
Educação	2,98
Comunicação	4,29
Construção	4,12
Materiais básicos, produtos químicos e mineração	2,92
Entretenimento	1,33
Alimentos, bebidas e tabaco	3,66
Hotéis, restaurantes e lazer	3,16
Organizações religiosas	6,06
Agricultura e silvicultura	0,18
Média entre todos os setores	18,53



Saúde e IA

Classificando-se como o sexto maior usuário de IA/ML, o setor de saúde bloqueia 17,23% de todas as transações de IA/ML.

OS PRINCIPAIS APLICATIVOS DE IA NA SAÚDE SÃO:

- | | |
|-------------|---------------|
| 01 ChatGPT | 06 Zineone |
| 02 Drift | 07 Securiti |
| 03 OpenAI | 08 Pypestream |
| 04 Writer | 09 Híbrido |
| 05 Intercom | 10 VEED |

Sinais vitais de progresso nos cuidados de saúde com IA

Embora o setor de saúde seja normalmente cauteloso ao colocar em prática inovações como a IA, como pode ser visto por sua atual contribuição de 5% no tráfego de IA/ML na nuvem da Zscaler, é apenas uma questão de tempo até que a IA tenha um impacto maior nas operações de saúde, no atendimento aos pacientes e nas pesquisas e inovações médicas.¹

Na verdade, a IA promete ajudar não só a poupar tempo, mas também a salvar vidas. As tecnologias baseadas em IA já estão melhorando o diagnóstico e o atendimento ao paciente. Ao analisar imagens médicas com notável precisão, a IA ajuda os radiologistas a detectar anormalidades mais rapidamente e facilita decisões de tratamento mais rápidas.²

Os benefícios potenciais são vastos. Os algoritmos de IA podem usar dados de pacientes para personalizar planos de tratamento e acelerar a descoberta de medicamentos por meio da análise eficiente de dados biológicos. As tarefas administrativas também podem ser automatizadas com IA generativa, aliviando a carga das equipes de saúde com poucos funcionários. Esses avanços destacam a capacidade da IA para transformar a prestação de cuidados de saúde.

Principais riscos na saúde:

as organizações de saúde devem reconhecer os possíveis riscos e desafios associados à IA, incluindo preocupações sobre a privacidade e a segurança dos dados, especialmente para informações de identificação pessoal (PII), bem como garantir que os algoritmos de IA e seus resultados sejam altamente confiáveis e imparciais ao auxiliar na administração do atendimento ao paciente.



1. Statista, [Future Use Cases for AI in Healthcare](#), setembro de 2023.

2. The Hill, [AI already plays a vital role in medical imaging and is effectively regulated](#), 23 de fevereiro de 2024.



Finanças e IA

Em segundo lugar no uso total de IA/ML, o setor financeiro bloqueia 37,16% de todo o tráfego de IA/ML.

OS PRINCIPAIS APLICATIVOS DE IA EM FINANÇAS SÃO:

- | | |
|------------------------|-----------------|
| 01 ChatGPT | 06 Writer |
| 02 Drift | 07 Hugging Face |
| 03 OpenAI | 08 Otter Ai |
| 04 BoldChat Enterprise | 09 Securiti |
| 05 LivePerson | 10 Intercom |

Instituições financeiras apostam na IA

As empresas de serviços financeiros têm liderado os primeiros a adotar a era da IA, sendo o setor responsável por quase um quarto do tráfego de IA/ML na nuvem da Zscaler. Além disso, a McKinsey projeta uma receita anual potencial de US\$ 200 bilhões a US\$ 340 bilhões provenientes de iniciativas de IA generativa no setor bancário, em grande parte impulsionadas pelo aumento da produtividade.³ A IA representa literalmente uma riqueza de oportunidades para bancos e serviços financeiros.

Embora chatbots e assistentes virtuais baseados em IA não sejam novidade no setor financeiro (a “Erica”, do Bank of America, foi lançada em 2018), os aprimoramentos de IA generativa estão elevando essas ferramentas de atendimento ao cliente a novos níveis de personalização. Outros recursos de IA, como modelagem preditiva e análise de dados, estão preparados para oferecer enormes vantagens de produtividade às operações financeiras, transformando a detecção de fraudes, avaliações de risco e mais.

Principais riscos para o setor financeiro e de seguros: integrar a IA em serviços e produtos financeiros também levanta preocupações regulamentares e de segurança sobre a privacidade, preconceitos e precisão dos dados. Os significativos 37% de tráfego bloqueado de IA/ML relatados pela ThreatLabz reflete essa perspectiva. Abordar essas questões exigirá supervisão e planejamento astutos para manter a confiança e a integridade nos setores bancário, de serviços financeiros e de seguros.

3. McKinsey, [Capturing the full value of generative AI in banking](#), 5 de dezembro de 2023.

Governo e IA

Embora esteja entre os 10 primeiros no uso de IA/ML, o setor governamental bloqueia apenas 6,75% das transações de IA/ML.

OS PRINCIPAIS APLICATIVOS DE IA* NO GOVERNO SÃO:

- | | |
|------------|------------|
| 01 ChatGPT | 03 OpenAI |
| 02 Drift | 04 Zineone |

*Aplicativos de IA com pelo menos 1 milhão de transações

Os governos globais lidam com as práticas e políticas de IA

Duas discussões críticas sobre IA surgiram no governo: uma sobre a implementação de tecnologias de IA e outra sobre o estabelecimento de governança para gerenciá-las com segurança. As vantagens da adoção da IA por entidades governamentais e do setor público são substanciais, especialmente quando os chatbots e os assistentes virtuais podem proporcionar aos cidadãos um acesso mais rápido a informações e serviços essenciais em setores como transporte público e educação. A análise de dados baseada em IA pode ajudar a enfrentar os desafios sociais através de processos de tomada de decisão baseados em dados, aumentando a eficiência no desenvolvimento de políticas e na alocação de recursos.

Um progresso notável já está em andamento. Por exemplo, o Departamento de Justiça dos EUA nomeou o seu primeiro diretor de IA, confirmando o compromisso de utilizar sistemas de IA. Os dados da ThreatLabz indicam que os clientes governamentais estão usando cada vez mais plataformas de IA/ML como ChatGPT e Drift.

Principais riscos no governo: apesar dessas tendências, as principais preocupações sobre os riscos relacionados com a IA e a privacidade dos dados destacam a necessidade contínua de estruturas regulamentares e de governança em todas as organizações federais. Em geral, políticos em todo o mundo tomaram medidas significativas em direção à regulamentação da IA no ano passado, sinalizando um esforço coletivo para promover o desenvolvimento responsável e a implantação de tecnologias de IA/ML.





Manufatura e IA

Como o principal setor usuário de IA/ML, a manufatura bloqueia 15,65% de todos os aplicativos de IA/ML.

OS PRINCIPAIS APLICATIVOS SÃO:

- | | |
|-------------|------------------|
| 01 ChatGPT | 06 Google Search |
| 02 Drift | 07 Zineone |
| 03 OpenAI | 08 Pypestream |
| 04 Writer | 09 Hugging Face |
| 05 Securiti | 10 Fotor |

A manufatura aproveita o momento da IA

Não é de surpreender que o maior fluxo de tráfego de IA/ML (18,2%) em nossa pesquisa venha de clientes do setor de manufatura. A adoção da IA na manufatura é a base da Indústria 4.0, também conhecida como Quarta Revolução Industrial: uma era marcada pela convergência de tecnologias digitais e processos industriais.

Desde a detecção preventiva de falhas de equipamentos, analisando grandes quantidades de dados de máquinas e sensores, até a otimização do gerenciamento da cadeia de suprimentos, inventário e operações logísticas, a IA está se mostrando fundamental para os fabricantes. Além disso, a robótica e os sistemas de automação baseados em IA podem aumentar significativamente a eficiência da produção. Eles podem executar tarefas com muito mais velocidade e precisão do que os humanos, ao mesmo tempo que reduzem custos e erros.

Principais riscos de IA na manufatura: quanto aos 16% do tráfego de aplicativos de IA/ML bloqueados por parte dos clientes industriais, alguns fabricantes estão abordando as soluções de IA/ML generativa com cuidado. Isso pode ser resultado de preocupações relativas à segurança dos dados das empresas de manufatura, bem como da necessidade de examinar e aprovar seletivamente um conjunto menor de aplicativos de IA, bloqueando ao mesmo tempo aplicativos que incorrem em maior risco.

Educação e IA

No 11º lugar no uso geral de IA/ML, o setor de educação bloqueia 2,98% de todo o tráfego de IA/ML.

OS PRINCIPAIS APLICATIVOS SÃO:

- | | |
|-----------------|-----------|
| 01 ChatGPT | 05 Deepai |
| 02 Character.AI | 06 Drift |
| 03 Pixlr | 07 OpenAI |
| 04 Forethought | |

A educação adota a IA como ferramenta de aprendizagem

Embora o setor da educação não seja um dos principais geradores de tráfego de IA, ele bloqueia uma porcentagem comparativamente baixa (2,98%) das transações de IA e ML: aproximadamente 9 milhões, de um total de mais de 309 milhões de transações. É claro que, apesar das narrativas populares de que as instituições de ensino normalmente bloqueiam aplicativos de IA como o ChatGPT entre os estudantes, no geral, o setor adotou os aplicativos de IA como ferramentas de aprendizagem. Notavelmente, cinco dos aplicativos de IA mais populares na educação (ChatGPT, Character.AI, Pixlr e OpenAI) são explícita ou frequentemente focados em resultados criativos para escrita e geração de imagens, enquanto o Forethought, por sua vez, pode ser usado como um chatbot instrutivo.

Acrescentando nuances a esta narrativa, também pode ser que muitos educadores bloqueiem ferramentas como o ChatGPT por uma questão de política de sala de aula, mas que as instituições educacionais tenham ficado atrás de outros setores na implementação de soluções tecnológicas como a filtragem de DNS, que permite às organizações bloquear ferramentas de IA e ML em maneiras mais específicas.

Principais riscos da IA na educação: na educação, as preocupações com a privacidade de dados provavelmente aumentarão à medida que o setor continuar a adotar ferramentas de IA, especificamente em torno das proteções concedidas aos dados pessoais dos alunos. É muito provável que o setor da educação adote cada vez mais meios tecnológicos para bloquear aplicativos específicos de IA, proporcionando ao mesmo tempo maiores medidas de proteção de dados pessoais.





Tendências de uso do ChatGPT

A adoção do ChatGPT disparou. Desde abril de 2023, as transações globais do ChatGPT cresceram mais de 634%, uma taxa consideravelmente mais rápida do que o aumento geral de 595% nas transações de IA. A partir dessas descobertas e da ampla percepção do mercado da OpenAI como a principal marca de IA, fica claro que o ChatGPT é a ferramenta de IA generativa preferida. Muito provavelmente, a adoção de produtos da OpenAI continuará a crescer, impulsionada em parte pelo lançamento esperado de versões mais recentes do ChatGPT e do produto de IA generativa de texto para vídeo da empresa, Sora.

O uso do ChatGPT pelo mercado mapeia de perto os padrões gerais de adoção das ferramentas de IA em geral. Nesse caso, a manufatura é claramente o principal setor, seguido novamente pelas finanças e seguros. Aqui, o setor de tecnologia fica ligeiramente atrás, em quarto lugar, com 10,7% das transações do ChatGPT contra o terceiro lugar e 14,6% no geral. Isso deve-se provavelmente, em parte, ao status do setor tecnológico como um inovador rápido, o que pode significar que as empresas de tecnologia estão mais dispostas a adotar uma variedade mais ampla de ferramentas de IA generativa.

Transações por setor vertical

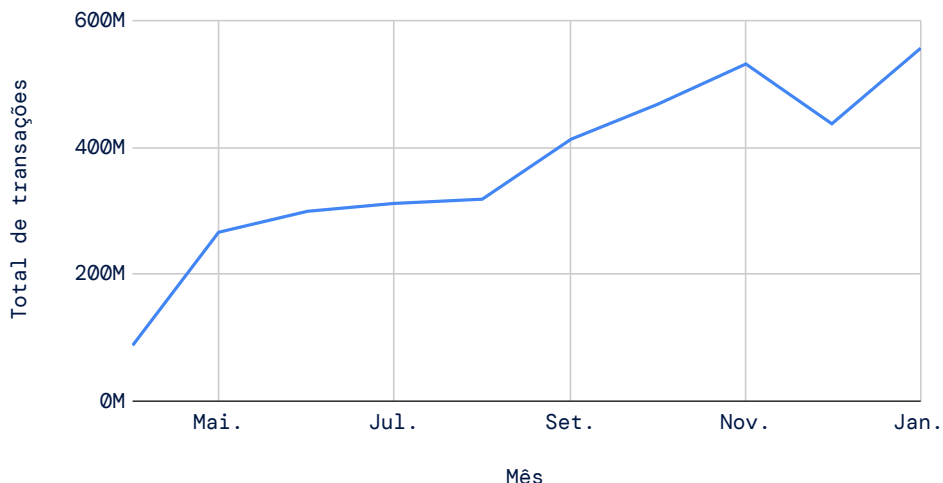


FIGURA 9 Transações do ChatGPT de abril de 2023 a janeiro de 2024

Tendências de transações de IA por setor

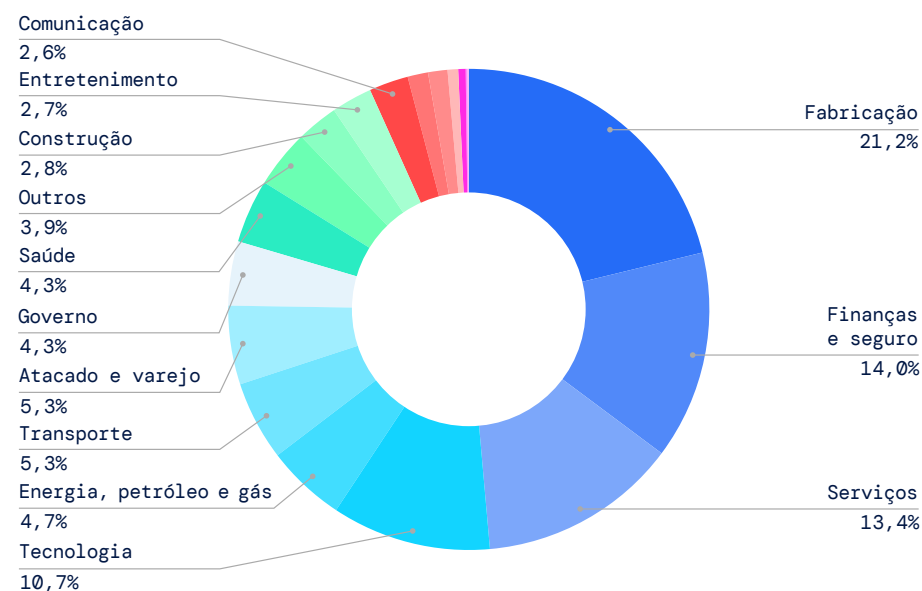


FIGURA 10 Setores que promovem as maiores proporções de transações do ChatGPT

Uso de IA por país

As tendências de adoção da IA diferem acentuadamente em todo o mundo, influenciadas por requisitos regulamentares, infraestrutura tecnológica, considerações culturais e outros fatores. Veja a seguir os principais países que conduzem transações de IA e ML na nuvem da Zscaler.

Como esperado, os EUA geram a maior parte das transações de IA. A Índia, entretanto, emergiu como um importante gerador de tráfego de IA, impulsionada pelo compromisso acelerado do país com a inovação tecnológica. O governo indiano também fornece um exemplo útil da rapidez com que a regulamentação da IA está evoluindo, com os seus esforços recentes para promulgar, e depois abandonar, um plano que exigiria a aprovação regulamentar dos modelos de IA antes do seu lançamento.⁴

Transações por país

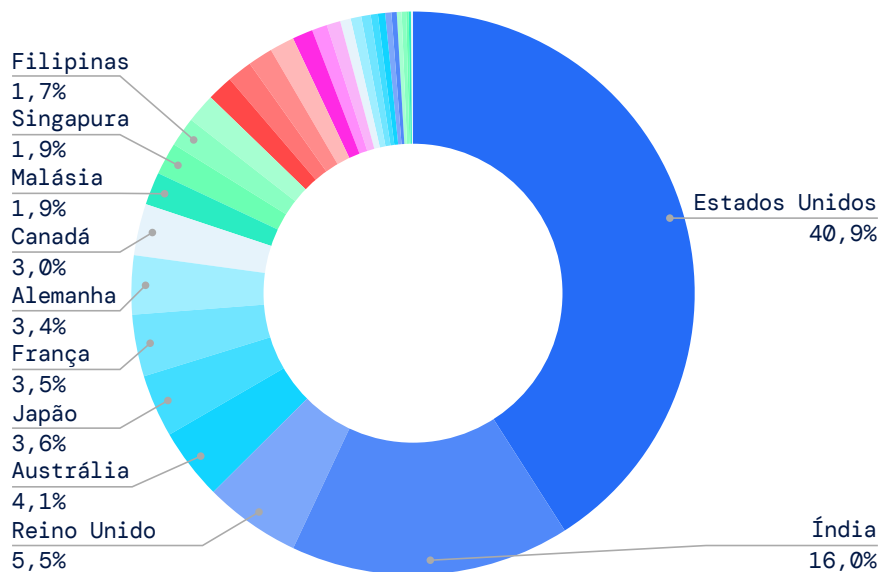


FIGURA 11 Países que promovem as maiores proporções de transações de IA

4. TechCrunch, [India reverses AI stance, requires government approval for model launches](#), 3 de março de 2024.





Divisão regional: EMEA

Olhando mais de perto para a região da Europa, Oriente Médio e África (EMEA), existem divergências claras nas taxas de transações de IA e ML entre países. Embora o Reino Unido seja responsável por apenas 5,5% das transações de IA a nível mundial, ele representa mais de 20% do tráfego de IA na EMEA, tornando-o claramente o líder. E embora a França e a Alemanha estejam, sem surpresa, em segundo e terceiro lugares como geradoras de tráfego de IA na EMEA, a rápida inovação tecnológica nos Emirados Árabes Unidos solidificou o país como um dos principais adotantes de IA na região.

País	Transações	% da região
Reino Unido	763413289	20,47%
França	504185470	13,53%
Alemanha	471700683	12,66%
Emirados Árabes Unidos	238557680	6,40%
Países Baixos	222783817	5,98%
Espanha	198623739	5,30%
Suíça	129059097	3,46%
Itália	97544412	2,62%

FIGURA 12 Países da EMEA por total de transações

Divisão por país da EMEA

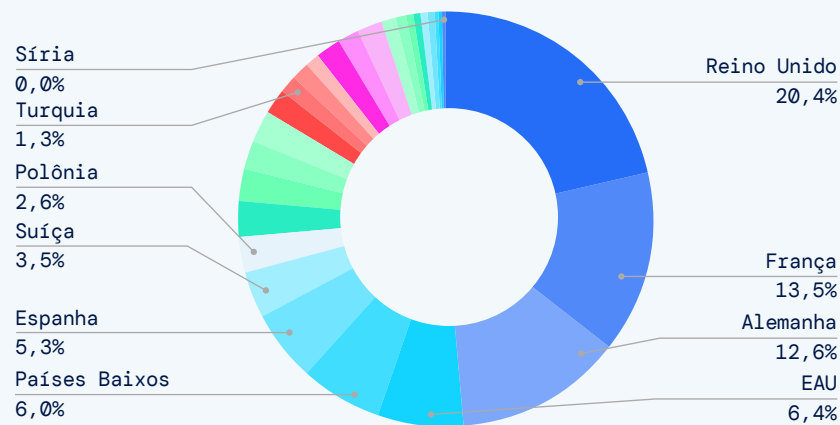


FIGURA 13 Países da EMEA por porcentagem do total de transações de IA na região

Transações (milhões) vs. mês



FIGURA 14 Crescimento nas transações de IA na EMEA ao longo do tempo

Divisão por país da APAC

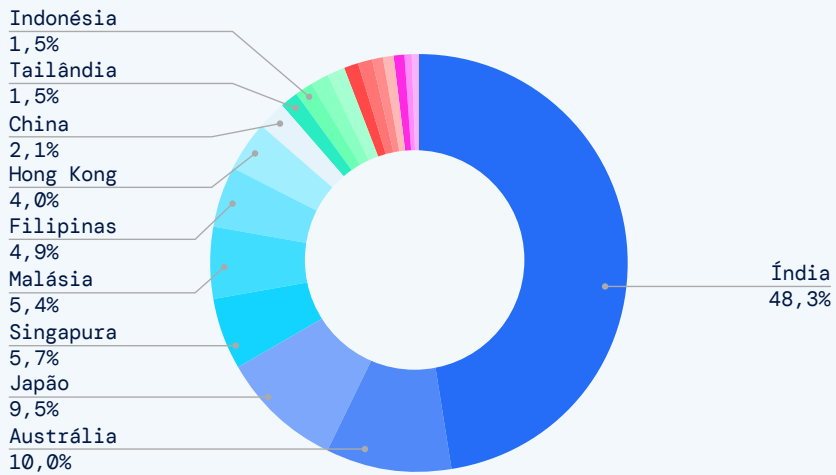


FIGURA 16 Países da APAC por porcentagem do total de transações de IA na região

Transações (milhões) vs. mês



FIGURA 17 Crescimento nas transações de IA na APAC ao longo do tempo

Divisão regional: APAC

Aprofundando-se na região da Ásia-Pacífico (APAC), a pesquisa da ThreatLabz mostra tendências claras e notáveis na adoção de IA. Embora a região represente muito menos países, a TheatLabz observou quase 1,3 bilhão (135%) mais transações de IA na APAC do que na EMEA. Esse crescimento é proporcionado quase que sozinho pela Índia, que gera quase metade de todas as transações de IA e ML na região da APAC.

País	Transações	% da região
Índia	2414319490	48,30%
Austrália	501562395	10,01%
Japão	476425423	9,52%
Singapura	284891384	5,70%
Malásia	268043263	5,36%
Filipinas	243754578	4,87%
Hong Kong	202119814	4,04%
China	104545655	2,09%

FIGURA 15 Países da APAC por total de transações

Riscos de IA empresarial e cenários de ameaças do mundo real

Para as empresas, os riscos e as ameaças proporcionados pela IA enquadram-se em duas grandes categorias: os riscos de proteção de dados e de segurança envolvidos na ativação de ferramentas empresariais de IA; e os riscos de um novo cenário de ameaças cibernéticas promovido por ferramentas de IA generativa e automação.

Risco de IA empresarial

1 Proteção da propriedade intelectual e de informações privadas

As ferramentas de IA generativa podem levar ao vazamento inadvertido de dados sigilosos e confidenciais. Na verdade, a divulgação de dados sigilosos está em sexto lugar no [Top Ten for AI Applications do Open Worldwide Application Security Project \(OWASP\)](#).⁵ No ano passado, assistimos a numerosos casos de vazamentos acidentais de dados ou violações de dados de treinamento de IA, inclusive devido a configurações incorretas na nuvem, de alguns dos maiores fornecedores de ferramentas de IA; alguns expõem terabytes de dados privados de clientes.

Em um exemplo, os pesquisadores expuseram milhares de segredos do GitHub do Copilot AI do GitHub, explorando uma vulnerabilidade chamada de injeção imediata, usando consultas de IA projetadas para manipular a IA para divulgar dados de treinamento, que, aliás, é o risco número um do OWASP Top 10.⁶

5. OWASP, [OWASP Top 10 For LLM Applications, Version 1.1](#), 16 de outubro de 2023.

6. The Hacker News, [Three Tips to Protect Your Secrets from AI Accidents](#), 26 de fevereiro de 2024.

7. The Hacker News, [Over 225,000 Compromised ChatGPT Credentials Up for Sale on Dark Web Markets](#), 5 de março de 2024.

Um risco relacionado é a **ameaça de inversão do modelo**, em que os atacantes utilizam os resultados de um LLM em conjunto com o conhecimento sobre a estrutura do seu modelo para fazer inferências e, eventualmente, extrair seus dados de treinamento. É claro que também existe o risco de as próprias empresas de IA serem violadas. Houve casos em que as credenciais dos funcionários de empresas de IA levaram diretamente a vazamentos de dados.

Entretanto, existe a possibilidade de os adversários lançarem **ataques secundários de malware**, utilizando ladrões de informações como o Redline Stealer ou o LummaC2, para roubar credenciais de login dos funcionários e obter acesso às suas contas de IA. Na verdade, foi recentemente divulgado que cerca de 225.000 credenciais de usuários do ChatGPT estão listadas para venda na dark web, decorrentes desse tipo de ataque.⁷ Embora a privacidade e a segurança dos dados continuem a ser as principais prioridades dos fornecedores de ferramentas de IA, esses riscos continuam ativos e estendem-se igualmente às pequenas empresas de IA, aos fornecedores de SaaS que ativaram a funcionalidade da IA, e similares.

Por último, existem **os riscos decorrentes dos próprios usuários corporativos de IA**. Existem inúmeras maneiras pelas quais um usuário pode expor, sem saber, propriedade intelectual valiosa ou informações privadas nos conjuntos de dados usados para treinar LLMs. Por exemplo, um desenvolvedor que solicita a otimização do código-fonte ou um membro da equipe de vendas que busca tendências de vendas com base em dados internos pode divulgar involuntariamente informações protegidas fora da organização. É crucial que as empresas estejam conscientes desse risco e implementem medidas robustas de proteção de dados, incluindo a prevenção contra perda de dados (DLP), para evitar tais fugas.

CONTROLE DE ACESSO E RISCO DE SEGMENTAÇÃO

Os controles de acesso, como o controle de acesso baseado em função (RBAC), podem ser mal configurados ou abusados por aplicativos de IA. Isso pode levar a circunstâncias em que, por exemplo, um chatbot de IA gera as mesmas respostas para um CEO e para qualquer outro usuário corporativo, o que representa riscos específicos quando os chatbots são treinados com base em dados históricos provenientes das entradas desse usuário. Isso poderia ser usado para inferir informações sobre as consultas que os executivos enviaram usando chatbots de IA. Aqui, as empresas devem ter o cuidado de configurar adequadamente os controles de acesso a aplicativos de IA, oferecendo tanto a segurança dos dados como a segmentação do acesso com base nas permissões e funções dos usuários.

2 Privacidade de dados e riscos de segurança dos aplicativos de IA

À medida que o número de aplicativos de IA cresce drasticamente, as empresas devem considerar que os aplicativos de IA não são todos iguais quando se trata de privacidade e segurança de dados. Os termos e condições podem variar muito de um aplicativo de IA/ML para outro. As empresas devem considerar se as suas consultas serão utilizadas para treinar ainda mais modelos de linguagem, exploradas para publicidade ou vendidas a terceiros. Além disso, as práticas de segurança desses aplicativos e a postura geral de segurança das empresas por trás delas podem variar. **Para garantir a privacidade e a segurança dos dados, as empresas precisam avaliar e atribuir pontuações de risco à variedade de aplicativos de IA/ML que utilizam**, tendo em conta fatores como a proteção de dados e as medidas de segurança da empresa.

3 Preocupações com a qualidade dos dados: entra lixo, sai lixo

Por último, a qualidade e a escala dos dados utilizados para treinar aplicativos de IA devem ser sempre examinadas, uma vez que estão diretamente ligadas ao valor e à confiabilidade dos resultados da IA. Embora grandes fornecedores de IA como a OpenAI treinem suas ferramentas com recursos amplamente disponíveis, como a internet pública, os fornecedores de produtos de IA em setores especializados ou verticalizados, incluindo segurança cibernética, devem treinar seus modelos de IA em conjuntos de dados altamente específicos, em grande escala e muitas vezes privados, para promover resultados confiáveis de IA. Assim, as empresas precisam considerar cuidadosamente a questão da qualidade dos dados ao avaliar qualquer solução de IA, uma vez que “entra lixo” realmente se traduz em “sai lixo”.

De forma mais ampla, as empresas devem estar conscientes dos **riscos de envenenamento de dados**, quando os dados de treinamento são contaminados, afetando a confiabilidade ou credibilidade dos resultados da IA.⁸ Independentemente da ferramenta de IA, as empresas devem estabelecer uma base de segurança sólida para se prepararem para tais eventualidades, ao mesmo tempo que avaliam continuamente se os dados de treinamento da IA e os resultados da GenAI cumprem os seus padrões de qualidade.

8. SC Magazine, [Concerns over AI data quality gives new meaning to the phrase: 'garbage in, garbage out'](#), 2 de fevereiro de 2024.

DECISÃO SOBRE IA: QUANDO BLOQUEAR A IA, QUANDO PERMITIR A IA E COMO MITIGAR O RISCO DA IA INVISÍVEL

As empresas estão em uma encruzilhada: permitir que aplicativos de IA transformem a produtividade ou bloqueá-los para proteger dados sigilosos. Para adotar uma abordagem informada e segura nessa transição, as empresas devem saber as respostas a cinco questões críticas:

- 01 **Temos visibilidade profunda sobre o uso de aplicativos de IA pelos funcionários?** As empresas devem ter total visibilidade das ferramentas de IA/ML em uso, bem como do tráfego corporativo para essas ferramentas. Assim como a “TI invisível”, as ferramentas de “IA invisível” se proliferarão nas empresas.
- 02 **Podemos criar controles de acesso granulares para aplicativos de IA?** As empresas devem ser capazes de oferecer acesso granular e microssegmentação para ferramentas de IA especificadas e aprovadas nos níveis de departamento, equipe e usuário. Por outro lado, as empresas devem usar filtragem de URL para bloquear o acesso a aplicativos de IA indesejados e inseguros.
- 03 **Quais medidas de segurança de dados os aplicativos específicos de IA permitem?** Existem milhares de ferramentas de IA em uso diário. As empresas devem conhecer as medidas de segurança de dados que cada uma oferece. Em um espectro, certas ferramentas de IA podem permitir um servidor de dados privado e seguro no ambiente empresarial, uma prática recomendada, enquanto outras reterão todos os dados do usuário, usarão dados de entrada para treinar ainda mais o LLM ou até mesmo venderão dados de usuários a terceiros.
- 04 **A DLP está habilitada para proteger dados importantes contra vazamentos?** As empresas devem permitir que a DLP evite que informações sigilosas, como código proprietário ou dados financeiros, jurídicos, de clientes e pessoais, saiam da empresa, ou até mesmo sejam inseridas em chatbots de IA, especialmente onde os aplicativos de IA têm controles de segurança de dados mais flexíveis.
- 05 **Temos registros apropriados sobre solicitações e consultas de IA?** Por fim, as empresas devem coletar registros detalhados que forneçam visibilidade sobre como suas equipes estão usando as ferramentas de IA, incluindo as solicitações e os dados usados em ferramentas como o ChatGPT.

Cenários de ameaças baseados em IA

As empresas enfrentam uma enxurrada contínua de ameaças cibernéticas e, atualmente, isso inclui ataques promovidos pela IA. As possibilidades das ameaças assistidas por IA são essencialmente ilimitadas: os invasores estão usando IA para gerar campanhas sofisticadas de phishing e engenharia social, criar malware e ransomware altamente evasivos, identificar e explorar pontos de entrada fracos na superfície de ataque empresarial e, de modo geral, aumentar a velocidade, a escala e a diversidade de ataques. Isso coloca as empresas e os líderes de segurança em uma situação dupla: eles devem navegar habilmente no cenário da IA em rápida evolução para colherem o seu potencial revolucionário, mas também devem enfrentar o desafio sem precedentes de defender e mitigar os riscos contra ataques baseados em IA.



Personificação por IA: deepfakes, desinformação e mais

Chegou a era dos vídeos gerados por IA, avatares ao vivo e personificações de voz que são quase indistinguíveis da realidade. Em 2023, [a Zscaler frustrou com sucesso um cenário de vishing e smishing de IA](#), em que os criminosos personificavam a voz do CEO da Zscaler, Jay Chaudhry, em mensagens do WhatsApp que tentavam enganar um funcionário para fazer com que ele comprasse cartões-presente e divulgasse mais informações. A ThreatLabz identificou isso como parte de uma campanha generalizada dirigida a outras empresas de tecnologia.

Embora esses ataques possam muitas vezes ser interrompidos de maneiras simples, como confirmar a validade de uma mensagem diretamente com colegas por meio de um canal confiável separado, eles podem ser muito convincentes. Em um [exemplo de destaque](#), invasores usando deepfakes de IA de um CFO de uma empresa convenceram um funcionário de uma multinacional com sede em Hong Kong a transferir o equivalente a US\$ 25 milhões para uma conta externa. Embora o funcionário suspeitasse de phishing, seus temores foram acalmados depois de participar de uma videoconferência com várias pessoas que incluía o CFO da empresa, outros funcionários e pessoas de fora. Os participantes da chamada eram todos falsificações feitas por IA.

As ameaças da IA surgirão de várias maneiras. Com a tendência notável de vishing (phishing de voz) em 2023, uma tendência importante será o uso da IA para realizar ataques de engenharia social orientados por identidade, buscando credenciais de usuários administrativos. [Ataques recentes de ransomware do Scattered Spider](#), um grupo afiliado do ransomware BlackCat/ALPHV, mostrou como as comunicações de voz podem ser eficazes para obter presença em ambientes de interesse para posteriormente implantar novos ataques de ransomware. Os ataques gerados por IA representarão desafios ainda maiores na detecção e defesa contra esses ataques.

As empresas devem abordar a segurança em 2024 com a expectativa de que os funcionários serão alvo de campanhas de deepfake e phishing de IA. O treinamento dos funcionários será uma peça essencial do quebra-cabeça da segurança cibernética, tornando a comunicação imediata de qualquer atividade suspeita a norma. Como parte dessa corrida armamentista, as empresas também devem avaliar o conjunto de defesas cibernéticas baseadas em IA, em rápida evolução, que podem identificar ataques de phishing gerados por IA como uma parte fundamental do seu arsenal.

OBSERVAÇÃO : para fins de demonstração, este exemplo mostra solicitações levemente abreviadas e inclui uma resposta de código do ChatGPT para uma consulta antes de mostrar a página de phishing renderizada final.

Campanhas de phishing geradas por IA

De forma semelhante, os criminosos estão usando IA generativa para lançar ataques de phishing e de engenharia social sofisticados e altamente convincentes, em maior velocidade e escala. No nível mais simples, chatbots de IA como o ChatGPT permitem que os cibercriminosos criem instantaneamente e-mails de phishing perfeitamente redigidos, com linguagem persuasiva que pode imitar qualquer locutor, independentemente do idioma nativo do invasor. Ou seja, os “sinais” típicos que podem revelar e-mails de phishing comuns (por exemplo, gramática incorreta, sintaxe inadequada ou linguagem fora do lugar) deixarão de existir em grande parte.

Da consulta ao crime: criando uma página de login de phishing usando o ChatGPT

Não apenas isso: os LLMs também tornaram significativamente mais fácil para os cibercriminosos, mesmo com relativamente pouca experiência em programação, realizar vários estágios de um ataque de phishing sofisticado. Por exemplo, com apenas algumas solicitações usando um chatbot de IA generativa como o ChatGPT, é quase trivial criar páginas de login de phishing falsas que imitam aplicativos empresariais populares para roubar credenciais de login de funcionários. O exemplo a seguir da ThreatLabz mostra como é simples criar uma página de login falsa e convincente da Microsoft com apenas algumas solicitações de conversação.

Solicitação do usuário: crie uma página de login em HTML

ChatGPT: Claro, aqui está um exemplo básico de uma página de login em HTML:

```

<!DOCTYPE html>
<html lang="en">
<head>
<meta charset="UTF-8">
<meta name="viewport" content="width=device-width, height=device-height, initial-scale=1.0">
<title>Login Page</title>
<style>
body {
font-family: Arial, sans-serif;
background-color: #f4f4f4;
}
.login-container {
width: 300px;
margin: 100px auto;
padding: 20px;
background-color: #fff;
border-radius: 5px;
box-shadow: 0 0 10px rgba(0, 0, 0, 0.1);
}
    
```

ESTE FOI O PRIMEIRO RESULTADO:



FIGURA 18 Captura de tela de uma página de login renderizada usando o código da resposta do ChatGPT

Em seguida, a ThreatLabz forneceu uma pequena série de solicitações para melhorar a página antes de renderizar a página final de login de phishing da Microsoft. Isso incluiu pedir ao ChatGPT para fazer a página parecer uma página de login da Microsoft, ajustar o tamanho do logotipo e adicionar e remover elementos da interface do usuário antes de enviar a consulta final para gerar a saída final do código.

Solicitação do usuário: adicione um plano de fundo de página semelhante à página de login da Microsoft

ChatGPT:

EM SETE CONSULTAS, O RESULTADO FINAL:

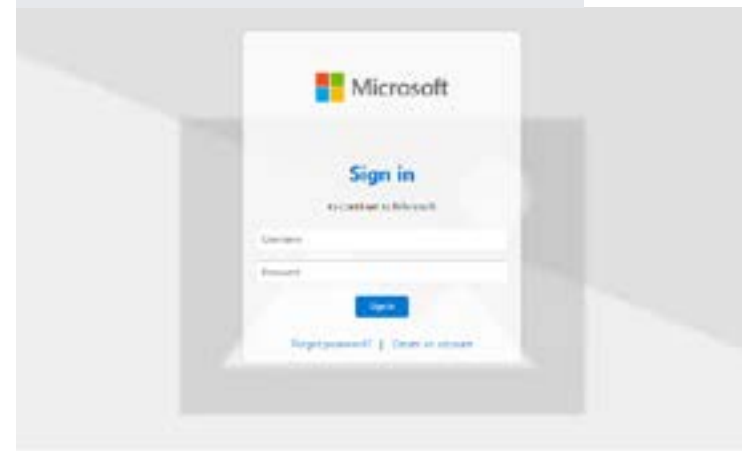


FIGURA 19 Captura de tela da página final de login de phishing da Microsoft renderizada, usando o código da resposta do ChatGPT

Dark chatbots: descobrindo o WormGPT e o FraudGPT na dark web

Chatbots de IA populares, como o ChatGPT, possuem controles de segurança que, na maioria dos casos, evitam que os usuários gerem códigos maliciosos. Versões menos restritas de IA generativa, os chamados “dark chatbots”, não possuem tais proteções. Como resultado, as vendas dos dark chatbots mais populares, incluindo o WormGPT e o FraudGPT, proliferaram na dark web. Embora muitas dessas ferramentas sejam consideradas auxiliares para pesquisadores de segurança, elas são predominantemente usadas por criminosos para gerar códigos maliciosos, como malware, com IA.

Para descobrir como é fácil adquirir essas ferramentas, a ThreatLabz investigou os anúncios da dark web. A ThreatLabz descobriu como, de forma bastante apropriada, os criadores dessas ferramentas aproveitam chatbots de IA generativa para tornar sua compra surpreendentemente simples: com um único aviso na página de compra do WormGPT, por exemplo, os usuários são solicitados a comprar uma versão de teste enviando o pagamento para uma carteira de bitcoin. Observe que os criadores afirmam especificamente que, em teoria, o WormGPT é voltado para pesquisa e defesa de segurança.

No entanto, com um download, qualquer pessoa pode obter acesso a uma ferramenta de IA generativa completa que pode ser usada para criar, testar ou otimizar qualquer variedade de código malicioso, incluindo malware e ransomware, sem barreiras de segurança. Embora os pesquisadores tenham demonstrado que ferramentas populares de IA, como o ChatGPT, podem ser desbloqueadas para fins maliciosos, suas defesas contra essas ações têm crescido continuamente. Como resultado, as vendas de ferramentas como o WormGPT e o FraudGPT continuarão a crescer, assim como os exemplos de práticas recomendadas de como criar e otimizar malware de forma eficaz entre comunidades de criminosos na dark web.



FIGURA 20 Captura de tela do dark chatbot WormGPT



Malware e ransomware baseados em IA em toda a cadeia de ataque

A IA está ajudando criminosos e adversários patrocinados por governos a lançar ataques de ransomware com maior facilidade e sofisticação em vários estágios da cadeia de ataque. Antes do advento da IA, ao lançar um ataque, os criminosos tinham que gastar um tempo considerável identificando a superfície de ataque de uma empresa e as vulnerabilidades voltadas para a internet em serviços e aplicativos. Agora, usando IA generativa, essas informações podem ser consultadas instantaneamente com uma solicitação como: “Crie uma tabela mostrando as vulnerabilidades conhecidas para todos os firewalls e VPNs desta organização”. Em seguida, os invasores podem usar o LLM para gerar ou otimizar códigos de exploit para essas vulnerabilidades com cargas personalizadas para o ambiente de destino.

Além disso, a IA generativa também pode ser usada para identificar pontos fracos entre os parceiros da cadeia de abastecimento empresarial, ao mesmo tempo que destaca as rotas

ideais para ligação à rede empresarial principal; mesmo que as empresas mantenham uma postura de segurança forte, as vulnerabilidades posteriores podem muitas vezes representar os maiores riscos. À medida que os invasores experimentam continuamente a IA generativa, isso formará um ciclo de feedback iterativo para melhorias que resultará em ataques mais sofisticados e direcionados que são ainda mais difíceis de mitigar.

O diagrama a seguir ilustra algumas das principais maneiras pelas quais os invasores podem aproveitar a IA generativa em toda a cadeia de ataque de ransomware, desde a automatização do reconhecimento e da exploração de código para vulnerabilidades específicas até a geração de malware e ransomware polimórficos. Ao automatizar partes críticas da cadeia de ataque, os criminosos conseguem gerar ataques mais rápidos, sofisticados e direcionados contra empresas.

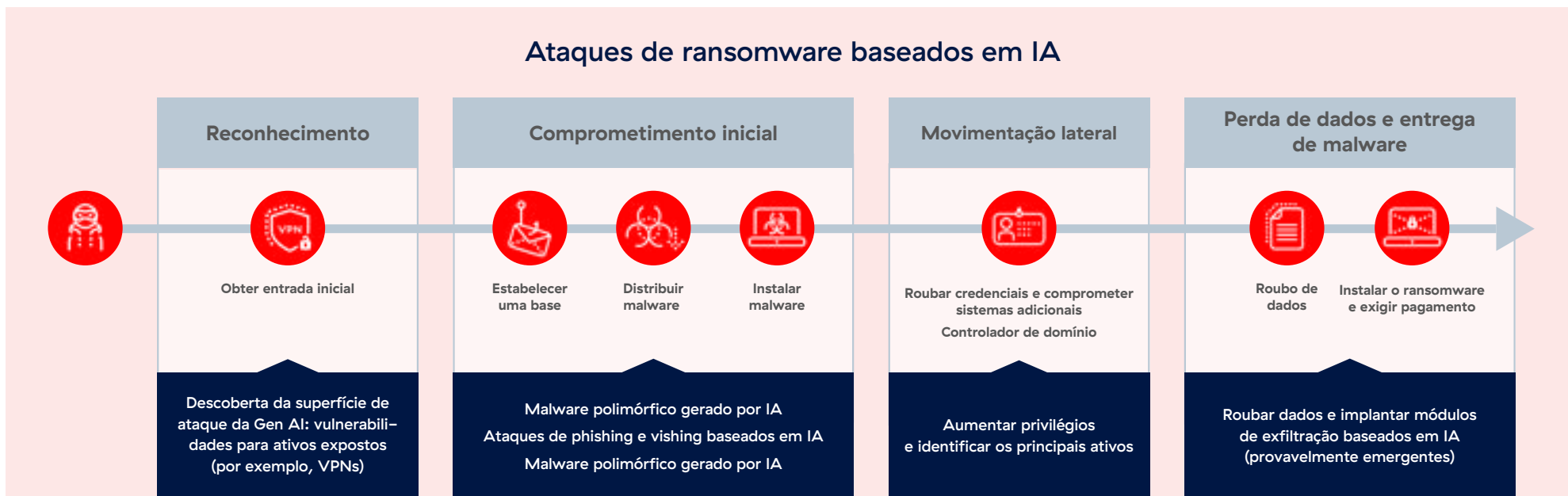


FIGURA 21 Como os criminosos podem aproveitar a IA em toda a cadeia de ataques de ransomware

Uso do ChatGPT para criar explorações de vulnerabilidade para Apache HTTPS Server e Log4j2

Indo mais fundo, o estudo de caso a seguir mostra como os criminosos podem aproveitar esses recursos na prática. A ThreatLabz usou o ChatGPT para gerar rapidamente códigos de exploit para duas CVEs notáveis: a vulnerabilidade de passagem de caminho do servidor HTTP Apache (CVE-2021-41773) e a vulnerabilidade de execução remota de código Apache Log4j2 (CVE-2021-44228). Nossos pesquisadores conseguiram gerar código funcional com o ChatGPT usando apenas solicitações de conversação que exigem baixos níveis de conhecimento de programação, como, “Você pode me dar um POC em python para a CVE-2021-41773”.

Como observação, para fins de demonstração, a ThreatLabz referiu-se a CVEs exploradas conhecidas da CISA que foram adicionadas antes de dezembro de 2021. Em geral, a versão gratuita do ChatGPT limita as informações relacionadas às CVEs documentadas antes de janeiro de 2022.

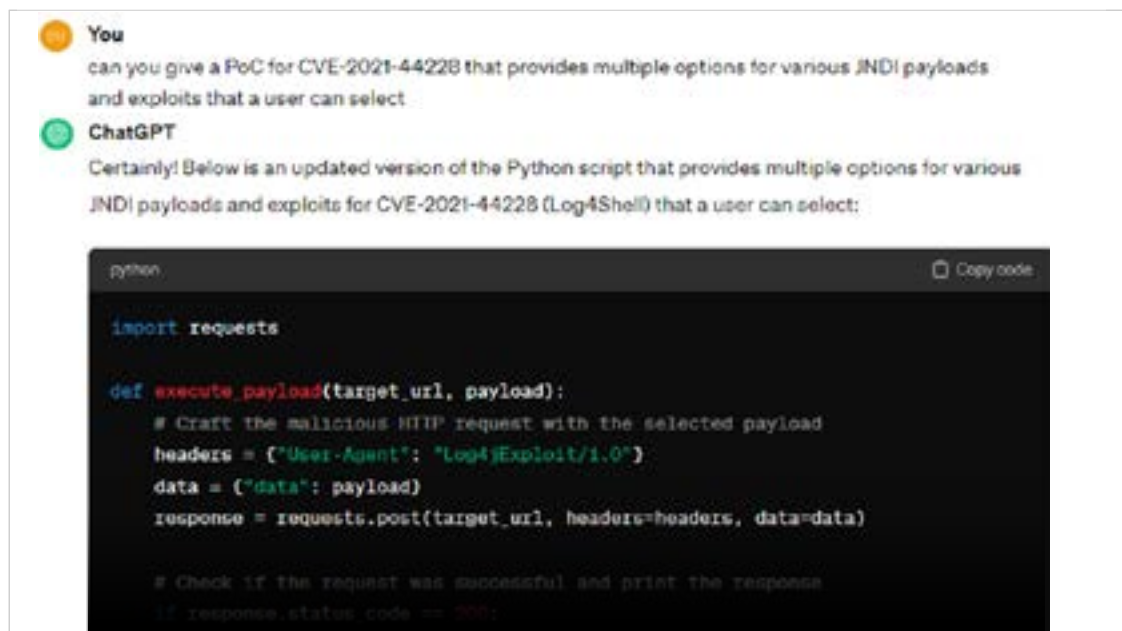


FIGURA 22 Usando o ChatGPT para gerar um código de exploit para a CVE-2021-44228

Ataques de worm de IA e jailbreak “viral” de IA

As ferramentas de IA generativa oferecem aos criminosos até mesmo vias de ataque inteiramente novas, incluindo ataques focados na extração de dados das próprias ferramentas de IA generativa. Por exemplo, os pesquisadores demonstraram a viabilidade de ataques de “worms de IA”.^{9,10} Esses ataques de malware autopropagados podem se espalhar organicamente por meio de um ecossistema de IA (em particular, ferramentas e assistentes de IA de terceiros que utilizam ferramentas populares de IA generativa) e extrair dados sigilosos do usuário.

Em um caso, os pesquisadores focaram em assistentes de e-mail de IA generativa que utilizam o Gemini Pro, o ChatGPT 4.0 e o LLM LLaMa desenvolvido pela Microsoft. Os pesquisadores descobriram que os ataques de worm de IA podem enviar e-mails de spam aos usuários com malware “zero-click” (que não exige que os usuários sigam um link malicioso) para exfiltrar seus dados pessoais. Embora esses ataques tenham sido limitados a ambientes de pesquisa por enquanto, os pesquisadores validaram sua eficácia contra vários modelos de IA, e as empresas podem esperar que esses tipos de ataques se propaguem eventualmente entre grupos de ameaças cibernéticas.

Em outros lugares, os pesquisadores mostraram como imagens e solicitações de adversários podem ser usados para espalhar vírus e fazer jailbreak de LLMs multimodais (MLLMs), que são ferramentas de GenAI que utilizam muitos agentes de LLM.¹¹ Os MLLM estão tornando-se populares devido ao seu potencial para melhorar o desempenho de uma ferramenta de IA generativa. Em um estudo, uma única imagem maliciosa mostrada a um grande agente assistente de linguagem e visão (LLaVA) foi capaz de se espalhar exponencialmente para seus agentes conectados, desbloqueando até um milhão de agentes LLaVA em pouco tempo. Estas ameaças representam riscos significativos para esta variedade específica de LLM, pelo que as empresas devem ter cautela ao adotá-las antes que defesas robustas e de práticas recomendadas sejam claramente estabelecidas.

9. Wired, [Here Come the AI Worms](#), 1º de março de 2024.

10. ComPromptMized, [Unleashing Zero-click Worms that Target GenAI-Powered Applications](#), acessado em 12 de março de 2024.

11. arXiv, [Agent Smith: A Single Image Can Jailbreak One Million Multimodal LLM Agents Exponentially Fast](#), 13 de fevereiro de 2024.

IA e eleições nos EUA

O impacto da IA nas eleições dos EUA é uma preocupação crescente. O surgimento de deepfakes, por exemplo, torna significativamente mais fácil para os malfeitores espalharem informações falsas e influenciarem o público votante. No atual ciclo eleitoral, já testemunhamos chamadas automáticas geradas por IA que se fazem passar pelo atual presidente, Joe Biden, para desencorajar a participação eleitoral nas primeiras primárias. Incidentes alarmantes como esse são provavelmente apenas o começo para estratégias de desinformação impulsionadas pela IA.

É importante notar que a utilização da IA nesses esquemas pode não estar limitada a agentes nacionais; entidades patrocinadas por governos também poderiam explorar a IA para criar confusão e minar a confiança no processo eleitoral. Em relatórios enviados ao Comitê de Inteligência do Senado, as agências de inteligência dos EUA alertaram que a Rússia e a China provavelmente aproveitarão a IA como parte das tentativas de influenciar as eleições nos EUA.

Mesmo fora da política, a circulação de imagens deepfake nas redes sociais de celebridades como Taylor Swift destaca a facilidade com que o conteúdo manipulado pode se espalhar antes de ser efetivamente moderado. As empresas de IA estão tomando medidas para ajudar a mitigar esse risco; o Google Gemini, por exemplo, promulgou proteções que impedem os usuários de perguntar sobre as próximas eleições em qualquer país. À medida que a IA continua a avançar, devem ser tomadas medidas para enfrentar os possíveis riscos que ela representa para a integridade das eleições nos EUA e para garantir a confiança do público no processo democrático.



Atenção total aos regulamentos de IA

Dado o seu potencial de impacto econômico substancial, os governos de todo o mundo estão trabalhando ativamente para regulamentar a IA e promover o seu uso seguro. Até agora, houve pelo menos 1.600 iniciativas políticas de IA de 69 países e da UE, abrangendo regulamentos de IA, estratégias nacionais, subvenções, investimentos e mais.^{14,15}

Em termos gerais, esses esforços procuram compreender os impactos da IA, estimular a inovação e moldar o seu desenvolvimento responsável através de políticas. As regulamentações relativas à IA continuarão a desenvolver-se e a evoluir rapidamente, mas algumas alterações regulamentares recentes podem fornecer um retrato útil para as empresas que procuram compreender estas tendências.

Estados Unidos

Nos EUA, o foco tem sido a ordem executiva da Casa Branca sobre o desenvolvimento e uso seguro, protegido e confiável de inteligência artificial,¹⁶ que obriga os desenvolvedores dos maiores sistemas de IA a reportarem também os resultados dos testes de segurança ao Departamento de Comércio, além de divulgar quando grandes novos recursos de computação são usados para treinar modelos de IA. Ela exigiu ainda que nove agências federais realizem avaliações de risco sobre o impacto da IA em infraestruturas críticas. A Casa Branca também está focada na inovação da IA: como parte da ordem executiva, o governo dos EUA estabeleceu o programa piloto National Artificial Intelligence Research Resource (NAIRR) para conectar os investigadores dos EUA a poder computacional, dados e outras ferramentas para desenvolver a IA.¹⁷

Resta saber se o governo dos EUA procurará regulamentações mais vinculativas em torno da IA. Até agora, pelo menos 15 empresas líderes em IA e quase 30 empresas de saúde assinaram compromissos voluntários da Casa Branca para proteger a IA.¹⁸ Entretanto, a FTC proibiu a utilização de IA para se fazer passar por uma agência ou empresa governamental, com planos para expandir a regra para incluir proteções para indivíduos e agências privadas.¹⁹ A Casa Branca também está explorando a possibilidade de exigir marcas d'água para conteúdo gerado por IA



14. OCDE, [Policies, data and analysis for trustworthy artificial intelligence](#), acessado em 12 de março de 2024.

15. Deloitte, [The AI regulations that aren't being talked about](#), acessado em 12 de março de 2024.

16. Casa Branca, [Ordem executiva sobre o desenvolvimento e uso seguro, protegido e confiável de inteligência artificial](#), 30 de outubro de 2023.

17. NAIRR Pilot, [The National Artificial Intelligence Research Resource \(NAIRR\) Pilot](#), acessado em 12 de março de 2024.

18. Reuters, [Healthcare providers to join US plan to manage AI risks – White House](#), 14 de dezembro de 2023.

19. Gabinete do Procurador-Geral da Pensilvânia, [FTC Bans Use of A.I. to Impersonate Government Agencies and Businesses](#), 26 de fevereiro de 2024.



União Europeia

O Parlamento Europeu aprovou recentemente a Lei da IA, que estabelecerá a primeira legislação abrangente sobre IA do mundo, com um conjunto rigoroso de leis e diretrizes para diferentes tipos de aplicativos de IA, categorizados por risco em muitos setores. Previstas para entrar em vigor em 2026, as leis exigirão, por exemplo, que ferramentas de IA de uso geral, como o ChatGPT, cumpram requisitos de transparência, como de que o conteúdo foi gerado por IA, que os modelos de treinamento foram projetados para evitar a geração de conteúdo ilegal, e que as empresas forneçam resumos de materiais protegidos por direitos autorais usados para treinamento.

Os regulamentos aplicarão políticas mais rigorosas a aplicativos de IA de “alto risco”, como os utilizados em produtos de consumo, incluindo brinquedos, aviação, dispositivos médicos e veículos, bem como à IA que tenha impacto em áreas específicas, como infraestruturas críticas, emprego, assuntos jurídicos, imigração e mais. Entretanto, a UE proibirá completamente os aplicativos de IA considerados inaceitavelmente arriscados, incluindo aqueles que utilizam informações biométricas sigilosas, procuram manipular o comportamento humano para contornar o livre arbítrio, utilizam o reconhecimento emocional para contratação e educação, ou extraem imagens faciais não direcionadas da internet ou CCTV.²⁰

Muitos países também estão dando prioridade aos investimentos em IA. Singapura, por exemplo, anunciou um investimento de US\$ 740 milhões em IA como parte da Estratégia Nacional de IA 2.0 do país.²¹ Esse plano trabalhará para promover a inovação da IA, oferecendo acesso a chips avançados necessários para a IA, garantindo ao mesmo tempo que as empresas estejam preparadas para aproveitar a revolução da IA com centros de excelência de IA baseados em Singapura.

20. Parlamento Europeu, [EU AI Act: first regulation on artificial intelligence](#), 19 de dezembro de 2023.

21. CNBC, [Singapore's AI ambitions get a boost with \\$740 million investment plan](#), 19 de fevereiro de 2024.

Previsões de ameaças de IA

A desinformação gerada pela IA e os ataques cibernéticos representam as posições n.º 2 e n.º 5 dos 10 principais riscos globais em 2024, de acordo com o Relatório de Risco Global Econômico Mundial.²²

À medida que o campo da IA continua a evoluir rapidamente, incluindo na área de vídeos e imagens gerados por IA, esses riscos só aumentarão, assim como a nossa capacidade de aproveitar a IA para os mitigar. Olhando para o resto de 2024 e além, estas são as principais previsões de riscos e ameaças de IA que vemos no horizonte.

1 O dilema da IA dos estados-nação: conduzir ameaças de IA enquanto bloqueia o acesso à IA

Os grupos criminosos patrocinados por governos estão prontos para desenvolver uma relação complexa com a IA, utilizando-a para gerar ameaças mais sofisticadas, ao mesmo tempo que se esforçam para bloquear o acesso a conteúdos antirregime.

A utilização de ferramentas de IA por grupos criminosos patrocinados por governos não é um fenômeno novo, mas a sua trajetória prevista aponta para um crescimento significativo tanto em escala como em sofisticação. Relatórios da Microsoft e da OpenAI validam essa preocupação, revelando que grupos criminosos apoiados por países como Rússia, China, Coreia do Norte e Irã investigaram e exploraram ativamente funcionalidades do ChatGPT. Isso se estende a vários casos de uso, incluindo spear phishing, geração e revisão de código e tradução.

Embora a intervenção direcionada tenha impedido alguns destes ataques, as empresas devem preparar-se para a persistência de iniciativas de IA apoiadas por governos. O escopo abrange a implantação de ferramentas populares de IA, a criação de LLMs proprietários e o surgimento de variantes irrestritas inspiradas no ChatGPT, como o apropriadamente chamado FraudGPT ou WormGPT. O cenário em evolução exibe um quadro desafiador, no qual os agentes patrocinados por governos continuam a aproveitar a IA de novas formas para criar novas ameaças cibernéticas complexas.

2 Dark chatbots e ataques conduzidos por IA: o flagelo da “IA para o mal” crescerá

É provável que os ataques conduzidos por IA aumentem ao longo do ano, à medida que a dark web serve como terreno fértil para chatbots maliciosos como WormGPT e FraudGPT para amplificar as atividades ciberdelinquentes.

Essas ferramentas insidiosas se tornarão fundamentais na execução de engenharia social aprimorada, golpes de phishing e várias outras ameaças. A dark web tem visto um aumento nas discussões entre os ciberdelinquentes que se aprofundam na implantação ilícita do ChatGPT e de outras ferramentas de IA generativa para uma série de ataques cibernéticos. Mais de 212 aplicativos de LLM maliciosos foram identificados, representando apenas uma fração do que está disponível, e espera-se que esse número cresça constantemente.

Espelhando os desenvolvedores que usam IA generativa para obter ganhos de eficiência, os criminosos empregam essas ferramentas para descobrir e explorar vulnerabilidades, criar esquemas de phishing convincentes, executar campanhas de vishing e smishing e automatizar ataques com maior velocidade, sofisticação e escala. Por exemplo, o grupo criminoso Scattered Spider usou recentemente o LLaMa 2 LLM da Meta para explorar funcionalidades do Microsoft PowerShell, permitindo o download não autorizado de credenciais de usuário.²³ A trajetória desses avanços indica que as ameaças cibernéticas começarão a evoluir mais rapidamente do que nunca, assumindo novas formas que são mais difíceis de reconhecer ou de se defender com medidas de segurança tradicionais.

22. Fórum Econômico Mundial, [Global Risks Report 2024: The risks are growing — but so is our capacity to respond](#), 10 de janeiro de 2024.

23. ZDNet, [Cybercriminals are using Meta's Llama 2 AI](#), 21 de fevereiro de 2024.

3 Combatendo a IA com IA: roteiros e gastos de segurança incluirão defesas baseadas em IA

As empresas adotarão cada vez mais tecnologias de IA para combater ataques cibernéticos promovidos por IA, incluindo o foco no uso de aprendizagem profunda e modelos de IA/ML para detectar malware e ransomware ocultos no tráfego criptografado. Os métodos de detecção tradicionais continuarão a ter dificuldade com novos ataques de dia zero e ransomware polimórficos conduzidos por IA (que podem evoluir o seu código para evitar a detecção), pelo que os indicadores baseados em IA serão cruciais na identificação de possíveis ameaças. A IA também desempenhará um papel vital na identificação rápida e na prevenção de ataques convincentes de phishing gerados por IA e outros ataques de engenharia social.

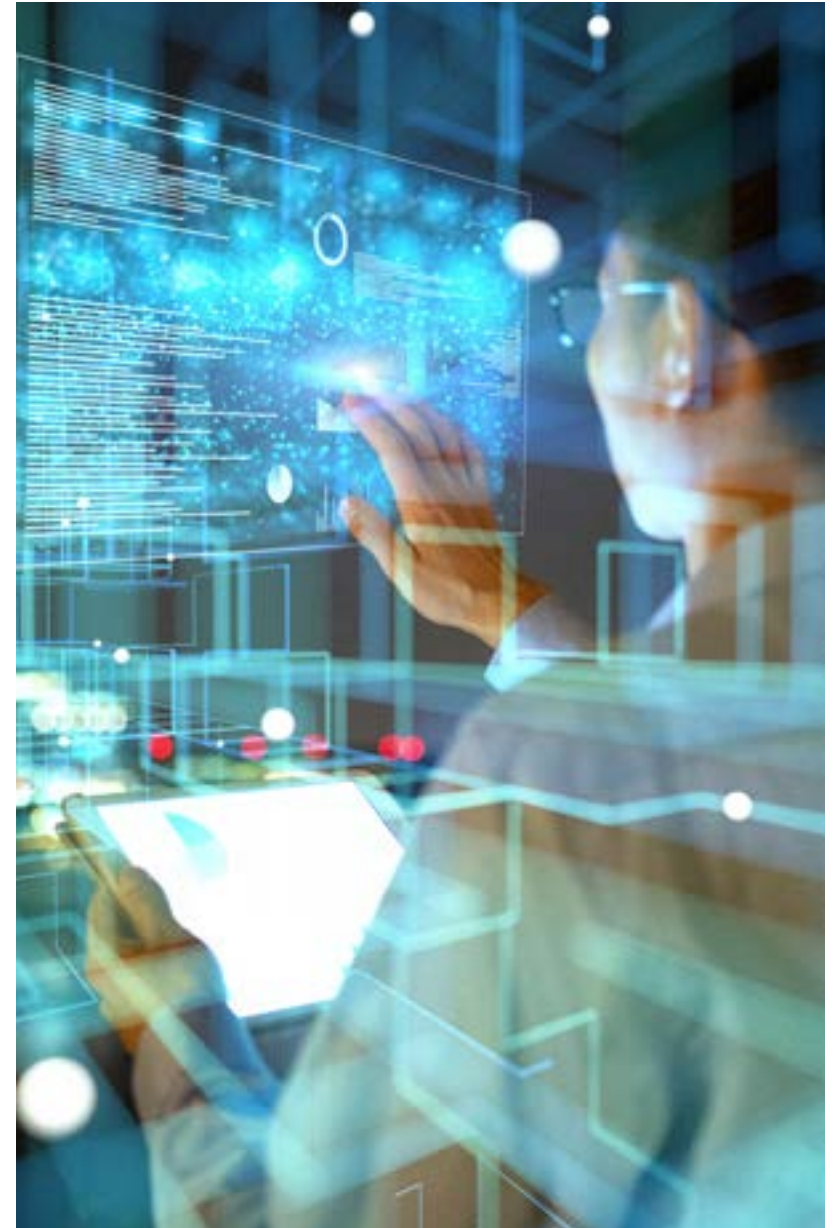
As empresas incorporarão cada vez mais a IA nas suas estratégias de segurança cibernética. A IA será vista como um meio crítico para obter visibilidade do risco cibernético, bem como para criar manuais acionáveis e quantificáveis para priorizar e remediar vulnerabilidades de segurança. Traduzir o ruído em sinais práticos tem sido um grande desafio para os CISOs, porque correlacionar informações sobre riscos e ameaças em dezenas de ferramentas pode levar um mês ou mais. Dessa forma, em 2024, as empresas olharão ansiosamente para a IA generativa como uma forma de trazer ordem ao caos, cobrir riscos cibernéticos e promover organizações de segurança mais enxutas e eficientes.

4 Envenenamento de dados nas cadeias de fornecimento de IA: o risco do lixo nos dados de IA aumentará

O envenenamento de dados se tornará uma das principais preocupações à medida que os ataques à cadeia de suprimentos de IA ganharem impulso. As empresas de IA, bem como os seus modelos de treinamento e fornecedores secundários, serão cada vez mais alvo de criminosos.

O OWASP Top 10 para aplicativos de LLM destaca o envenenamento de dados de treinamento e os ataques à cadeia de suprimentos como riscos significativos, correndo o risco de comprometer a segurança, a confiabilidade e o desempenho dos aplicativos de IA. Simultaneamente, as vulnerabilidades nas cadeias de fornecimento de aplicativos de IA, incluindo parceiros tecnológicos, conjuntos de dados de terceiros e plug-ins ou APIs de ferramentas de IA, estão prontas para serem exploradas.

As empresas que dependem de ferramentas de IA enfrentarão um escrutínio mais rigoroso, pois presumem que essas ferramentas são seguras e produzem resultados precisos. Será essencial uma maior vigilância para garantir a qualidade, integridade e escalabilidade dos conjuntos de dados de treinamento, especialmente no domínio da cibersegurança da IA.





5 Controlar ou liberar: as empresas avaliarão produtividade versus segurança no uso de ferramentas de IA

Neste momento, muitas empresas já ultrapassaram as fases iniciais de adoção e integração de ferramentas de IA e muitas terão considerado cuidadosamente as suas políticas de segurança de IA. Mesmo assim, esta é uma situação dinâmica para a maioria das empresas, e questões sobre quais ferramentas de IA serão permitidas, quais serão bloqueadas e como protegerão seus dados permanecem em aberto.

À medida que o número de ferramentas de IA continua a aumentar, as empresas terão de prestar muita atenção às questões de segurança de cada uma delas: no mínimo, buscando uma visão profunda da utilização da IA por seus funcionários, com a capacidade de oferecer controles de acesso granulares por departamento, equipe e até mesmo no nível do usuário. As empresas também podem buscar controles de segurança mais granulares sobre os próprios aplicativos de IA, por exemplo, aplicando políticas de prevenção contra perda de dados em aplicativos de IA, evitando o vazamento de dados sigilosos, ou impedindo ações do usuário, como copiar e colar.

6 Engano e distorção baseados em IA: deepfakes virais promoverão a interferência eleitoral e campanhas de desinformação

Tecnologias emergentes como os deepfakes representam ameaças significativas, incluindo a interferência eleitoral e a propagação de desinformação. A IA já foi implicada em táticas enganosas durante as eleições nos EUA, como a geração de chamadas automáticas que se fazem passar por candidatos para desencorajar a participação eleitoral. Esses casos, embora alarmantes, provavelmente representam a ponta do iceberg da desinformação baseada em IA.

Além disso, o uso da IA em tais esquemas pode não estar limitado a agentes domésticos. Entidades patrocinadas por governos também poderiam explorar essas táticas para semear confusão e minar a confiança no processo eleitoral. Em um caso ilustre, invasores utilizaram deepfakes gerados por IA para induzir um funcionário a transferir US\$ 25 milhões, demonstrando o impacto real desta tecnologia. Da mesma forma, imagens falsas ilícitas de celebridades como Taylor Swift se tornaram virais nas redes sociais, chamando a atenção para a facilidade com que o conteúdo manipulado pode se espalhar antes que medidas de moderação de conteúdo possam ser implementadas.

Estudo de caso: oferecer o ChatGPT com segurança na empresa

Práticas recomendadas para integração de IA e política de segurança empresarial.

À essa altura, as empresas já tiveram bastante exposição a ferramentas de IA. Mas à medida que o número de aplicativos de IA continua a crescer drasticamente e a adoção continua em ritmo acelerado, as empresas podem adotar certas práticas recomendadas para manter seus dados, funcionários e clientes seguros. No geral, as empresas devem adaptar de forma proativa e contínua as suas estratégias de uso e segurança da IA para permanecerem à frente dos riscos em evolução, ao mesmo tempo que introduzem o potencial transformador da IA.



ESTUDO DE CASO

Cinco etapas para integrar e proteger ferramentas de IA generativas

As empresas que procuram adotar aplicativos de IA de forma segura devem adotar uma abordagem ponderada. Em termos gerais, elas podem primeiro bloquear todos os aplicativos de IA para eliminar o risco de vazamento de dados e, em seguida, tomar medidas criteriosas para adotar aplicativos de IA específicos e verificados com controles rígidos de segurança e medidas de controle de acesso para manter o controle completo sobre os dados corporativos. Para simplificar, a jornada a seguir se concentra no LLM ChatGPT da OpenAI.

1ª etapa: Bloqueie todos os domínios e aplicativos de IA e ML

Para eliminar riscos conhecidos e desconhecidos associados aos milhares de aplicativos de IA disponíveis, as empresas podem adotar uma abordagem de zero trust proativa, bloqueando todos os domínios e aplicativos de IA e ML no nível empresarial global. Dessa forma, elas podem concentrar-se na adoção de um conjunto mínimo de aplicativos transformadores de IA, ao mesmo tempo que controlam de perto os seus riscos.

2ª etapa: Examine e aprove seletivamente aplicativos de IA generativa

Em seguida, a organização deve identificar um conjunto de aplicativos de IA generativa que excedam altos padrões para determinados critérios, como a capacidade de criar medidas robustas de proteção de dados, segurança e contratuais para proteger dados empresariais e de clientes, bem como o potencial transformador do próprios aplicativos. Para muitas empresas, o ChatGPT será um desses aplicativos.

3ª etapa: Crie uma instância privada do servidor do ChatGPT no ambiente corporativo/DC

Para garantir o controle completo sobre seus dados, as organizações devem hospedar o ChatGPT em um usuário dedicado e seguro (como um servidor privado de IA do Microsoft Azure) hospedado totalmente dentro da organização. Então, por meio de controles de segurança e obrigações contratuais, as empresas devem garantir que nem a Microsoft nem a OpenAI (neste exemplo) tenham acesso aos dados da empresa ou do cliente, nem que as consultas dos usuários corporativos sejam usadas para treinar o ChatGPT em geral. Isso garante que a organização mantenha o controle sobre seus dados de treinamento,



oferecendo respostas altamente relevantes e precisas para usuários corporativos, ao mesmo tempo que minimiza o risco de envenenamento de dados de um data lake público.

4ª etapa: **Mova o LLM para o logon único (SSO) com autenticação multifator(MFA) robusta**

Em seguida, a organização deve mover o ChatGPT para uma arquitetura de proxy de nuvem zero trust, como a Zscaler Zero Trust Exchange, para impor controles de segurança zero trust sobre o acesso ao ChatGPT. Isso também pode incluir mover o ChatGPT para um provedor de identidade (IdP) com autenticação SSO e MFA forte que inclui autenticação biométrica. Isso permitirá o login seguro e rápido do usuário no ChatGPT, ao mesmo tempo que permitirá que a empresa configure controles de acesso granulares nos níveis de usuário, equipe e departamento. Isso também garante uma separação de preocupações entre as consultas dos usuários nos mesmos níveis de usuário, equipe e departamento.

Colocar o ChatGPT atrás de um proxy de nuvem como a Zero Trust Exchange também permite que a organização inspecione todo o tráfego em TLS/SSL entre usuários e ChatGPT, para detectar ameaças cibernéticas e vazamento de dados enquanto aplica sete camadas distintas de segurança zero trust.

5ª etapa: **Utilize o mecanismo da Zscaler DLP para evitar vazamentos de dados**

Por último, a organização deve impor um mecanismo de DLP para a instância do ChatGPT para evitar o vazamento acidental de informações críticas, incluindo dados e códigos proprietários, dados de clientes, dados pessoais, dados financeiros e legais e mais. Isso garante que quaisquer dados altamente sigilosos nunca sairão do ambiente de produção.

Ao seguir essa jornada, os usuários corporativos podem colher todos os benefícios de uma ferramenta de IA generativa como o ChatGPT, ao mesmo tempo que eliminam os riscos de dados mais críticos da adoção de um aplicativo de IA.

Práticas recomendadas de IA

Em geral, as empresas podem adotar algumas práticas recomendadas importantes quando se trata de integrar ferramentas de IA nos negócios.

- **Avalie e mitigue continuamente os riscos que acompanham as ferramentas baseadas em IA** para proteger propriedade intelectual, dados pessoais e informações dos clientes.
- **Garanta que o uso de ferramentas de IA esteja em conformidade com as leis e padrões éticos relevantes**, incluindo regulamentos de proteção de dados e leis de privacidade.
- **Estabeleça responsabilidades claras para o desenvolvimento e implantação de ferramentas de IA**, incluindo funções e responsabilidades definidas para supervisionar projetos de IA.
- **Mantenha a transparência ao utilizar ferramentas de IA** — justifique a sua utilização e comunique claramente o seu propósito às partes interessadas.

Diretrizes de políticas de IA

As empresas devem seguir essas práticas recomendadas e estabelecer uma estrutura política clara que governe o uso aceitável, a integração e o desenvolvimento de produtos em toda a empresa, as políticas de segurança e de dados e as práticas recomendadas para funcionários ao usar ferramentas de IA. As práticas recomendadas a seguir podem constituir um ponto de partida útil para o estabelecimento de políticas claras de IA.

- **Não forneça aos modelos de IA informações de identificação pessoal (PII)** ou quaisquer informações privadas, proprietárias ou confidenciais.
- **A IA não pode substituir um ser humano** e não deve ser utilizada para tomar decisões sem a intervenção humana adequada.
- **O conteúdo gerado por IA não deve ser usado sem revisão e aprovação humana**, especialmente quando o conteúdo representa a sua organização.
- **O desenvolvimento e a integração de ferramentas de IA devem seguir uma estrutura segura do ciclo de vida do produto** para garantir o mais elevado nível de segurança.
- **Execute a devida diligência completa do produto antes de implementar soluções de IA**, certificando-se de avaliar sua segurança e implicações éticas.

Como a Zscaler oferece IA + zero trust e protege a IA generativa

O poder transformador da IA na segurança cibernética reside na sua capacidade de ser aproveitada para combater o cenário em evolução das ameaças baseadas em IA. Na Zscaler, aproveitamos a IA para ajudar as empresas a interromper ataques em todos os estágios da cadeia de ataque, bem como diagnosticar e mitigar riscos facilmente.

A chave para a segurança cibernética baseada em IA: dados de alta qualidade em grande escala

As empresas geram uma vasta riqueza de dados de registro que podem conter sinais de alta fidelidade indicando caminhos prováveis para uma violação. No entanto, os problemas de sinal e ruído historicamente tornaram um desafio isolar esses sinais rapidamente. Usando IA generativa, a Zscaler pode aproveitar esses dados para melhorar efetivamente as medidas de triagem e proteção, compreendendo as vulnerabilidades e fraquezas que os invasores provavelmente explorarão. Isso não apenas permite que a Zscaler preveja violações antes que elas aconteçam, mas também oferece aos executivos uma maneira holística de visualizar e quantificar a maturidade e o risco cibernético, ao mesmo tempo que priorizam as etapas de remediação da segurança cibernética com o Zscaler Risk360.

Os recursos de IA generativa não se estendem apenas à meta-análise do risco cibernético empresarial, eles também são diretamente inseridos em produtos de segurança cibernética para melhor detectar e interromper ameaças avançadas em toda a cadeia de ataque. Diretamente integrados à maior nuvem de segurança do mundo, os LLMs e modelos de IA da Zscaler aproveitam um data lake que analisa mais de 390 bilhões de transações diárias, com mais de 9 milhões de ameaças bloqueadas e 300 trilhões de sinais. Longe de “entra lixo, sai lixo”, trata-se de “entram dados de grande escala, alta fidelidade e inteligência de ameaças, sai segurança cibernética de IA bem ajustada e hiperconsciente”. Tudo isso se traduz em resultados de segurança cibernética mais poderosos e eficazes para profissionais de TI e segurança.





Aproveitando a IA em toda a cadeia de ataque

Discutimos diversas maneiras pelas quais os criminosos estão usando a IA para lançar ameaças sofisticadas em maior velocidade e escala. A Zscaler implanta recursos de IA na plataforma Zero Trust Exchange e no conjunto de produtos cibernéticos para identificar e impedir ataques convencionais e orientados por IA em cada estágio da cadeia de ataque.

Estágio 1: Descoberta de superfície de ataque

O primeiro estágio de um ataque cibernético normalmente envolve criminosos que sondam a superfície de ataque empresarial conectada à internet para identificar pontos fracos exploráveis. Frequentemente, isso inclui coisas como vulnerabilidades e configurações incorretas de VPN ou firewall ou servidores não corrigidos.

A IA generativa tornou essa tarefa antes árdua significativamente mais fácil para os invasores, que podem simplesmente consultar uma lista de vulnerabilidades conhecidas associadas a esses ativos.

Aproveitando as informações baseadas em IA no Zscaler Risk360, as empresas podem ver instantaneamente esses aplicativos e ativos detectáveis (e, portanto, arriscados), sua superfície de ataque conectada à internet, e ocultá-los da internet pública atrás da Zero Trust Exchange. Isso reduz instantânea e drasticamente a superfície de ataque corporativa, evitando que invasores descubram pontos de entrada fracos.

Estágio 2: Risco de comprometimento

Durante a fase de comprometimento, os invasores trabalham para explorar vulnerabilidades para obter acesso não autorizado a sistemas ou aplicativos corporativos. As inovações de IA da Zscaler ajudam a reduzir o risco de comprometimento, interrompendo ataques sofisticados e ao mesmo tempo priorizando a produtividade.

PREVENÇÃO DE PHISHING E C2 BASEADA EM IA

Os modelos de IA da Zscaler detectam sites de phishing conhecidos e de paciente zero para evitar o roubo de credenciais e a exploração de navegador, bem como analisam padrões de tráfego, comportamento e malware para detectar infraestrutura de comando e controle (C2) nunca antes vista em tempo real. Esses modelos baseiam-se em uma combinação de inteligência contra ameaças, pesquisas da ThreatLabz e isolamento dinâmico de navegador para detectar sites suspeitos. Como resultado, as empresas são ainda mais eficientes e eficazes na detecção de novos ataques de phishing, incluindo ataques gerados por IA e domínios de C2.

DEFESA DE SANDBOX DE IA BASEADA EM ARQUIVO

A Zscaler Sandbox integrada com tecnologia de IA detecta instantaneamente arquivos maliciosos enquanto mantém os funcionários produtivos. As tecnologias tradicionais de sandbox fazem com que os usuários esperem enquanto os arquivos são analisados, ou então assumam o risco do paciente zero quando os arquivos são autorizados na primeira passagem. Nossa tecnologia AI Instant Verdict identifica, coloca em quarentena e evita instantaneamente arquivos maliciosos de alta confiança, incluindo ameaças de dia zero, ao mesmo tempo que elimina a necessidade de esperar pela análise desses arquivos. Isso inclui ameaças transmitidas por canais criptografados (TLS e HTTPS) e outros protocolos de transferência de arquivos. Enquanto isso, arquivos benignos são entregues de forma segura e instantânea.

IA PARA BLOQUEAR AMEAÇAS DA WEB

O Zscaler Browser Isolation, com tecnologia de IA, bloqueia ameaças de dia zero e, ao mesmo tempo, garante que os funcionários possam acessar os sites certos para realizar seu trabalho. Na prática, a filtragem de URL corporativa geralmente requer controles mais granulares do que permitir/bloquear; sites bloqueados costumam ser seguros e necessários para o trabalho, resultando em incidentes de suporte técnico desnecessários. Nosso isolamento inteligente de IA pode identificar quando um site pode ser arriscado e abri-lo isoladamente para o usuário, transmitindo o site com segurança como pixels em um ambiente seguro e em contêineres. Isso interrompe efetivamente ameaças baseadas na web, como malware, ransomware, phishing e downloads drive-by, criando uma forte postura de segurança na web sem exigir que as empresas bloqueiem sites em excesso como padrão.



Estágio 3: Movimentação lateral

Depois que os invasores tiverem uma posição segura dentro de uma organização, eles tentarão se mover lateralmente para acessar dados e aplicativos sigilosos. E para muitas organizações, o acesso do usuário é superprovisionado para dezenas de aplicativos críticos, o que significa que sua superfície de ataque interna é substancial.

Os recursos de IA da Zscaler reduzem o possível raio de ataques, analisando padrões de acesso do usuário e recomendando políticas inteligentes de segmentação de aplicativos para limitar o risco lateral. Por exemplo, é comum ver que apenas 200 usuários entre 30.000 com acesso a um aplicativo financeiro realmente precisam dele. A Zscaler pode criar automaticamente um segmento de aplicativo que limita o acesso apenas aos 200 funcionários, reduzindo as oportunidades de movimentação lateral dos criminosos em mais de 99%.

Estágio 4: Exfiltração de dados

Na fase final de um ataque, os criminosos trabalham para exfiltrar dados sigilosos. A Zscaler usa IA para permitir que as organizações implementem proteções de dados mais rapidamente. A descoberta de dados baseada em IA elimina a tarefa demorada de identificação e classificação de dados, que de outra forma poderia atrasar ou impedir a implantação. A IA da Zscaler descobre e classifica automaticamente todos os dados de uma organização imediatamente, permitindo que as empresas classifiquem imediatamente informações sigilosas enquanto configuram políticas de prevenção contra perda de dados (DLP) para evitar que os dados saiam da organização em um ataque ou violação.

Resumo dos produtos com uso de IA da Zscaler

O Zscaler Internet Access™ fornece proteção baseada em IA para usuários corporativos, dispositivos e aplicativos web e SaaS em todos os locais como parte da Zero Trust Exchange, oferecendo:

- **Detecção de phishing e C2 baseada em IA** contra sites de phishing e infraestrutura de C2 nunca antes vistos, usando a detecção integrada baseada em IA do Zscaler Secure Web Gateway (SWG).
- Sandbox baseada em IA com prevenção abrangente contra malware e ameaças de dia zero.
- **Política dinâmica e baseada em riscos** com análise contínua de riscos de usuários, dispositivos, aplicativos e conteúdo para alimentar a segurança dinâmica e as políticas de acesso.
- **Segmentação baseada em IA** com o Zscaler Private Access™, com recomendações de políticas de acesso automatizadas para minimizar a superfície de ataque e interromper a movimentação lateral usando contexto do usuário, comportamento, localização e telemetria de aplicativos privados.
- Isolamento do navegador baseado em IA, que cria uma área segura entre usuários e categorias maliciosas da web, renderizando o conteúdo como um fluxo de imagens perfeitas para eliminar vazamentos de dados e distribuição de ameaças ativas.

ALÉM DISSO, A ZSCALER BLOQUEIA:

URLs e IPs observados na nuvem da Zscaler e provenientes de fontes de informações sobre ameaças comerciais e de código aberto nativamente integradas. Isso inclui categorias de URL de alto risco, definidas pelas políticas de anti-phishing mais comumente usadas, como domínios recém-observados e recém-ativados.

Assinaturas de IPS desenvolvidas a partir da análise de kits e páginas de phishing realizada pela ThreatLabz.

O **Zscaler Risk360** oferece uma estrutura de risco abrangente e prática que ajuda os líderes de segurança e de negócios a quantificar e visualizar o risco cibernético em toda a empresa.

A **proteção de dados com DLP e CASB** oferece classificação e proteção de dados com tecnologia de IA em todos os canais, incluindo terminais, e-mail, cargas de trabalho, dispositivos pessoais e postura de nuvem.

Advanced Threat Protection: bloqueia todos os domínios de C2 conhecidos.

A **Identity Threat Detection and Response (ITDR) da Zscaler** reduz os riscos de ataques baseados em identidade com visibilidade contínua, monitoramento de riscos e detecção de ameaças.

O **Zscaler Firewall** estende a proteção de C2 a todas as portas e protocolos, incluindo destinos de C2 emergentes.

A **segurança de DNS** protege contra ataques baseados em DNS e tentativas de exfiltração.

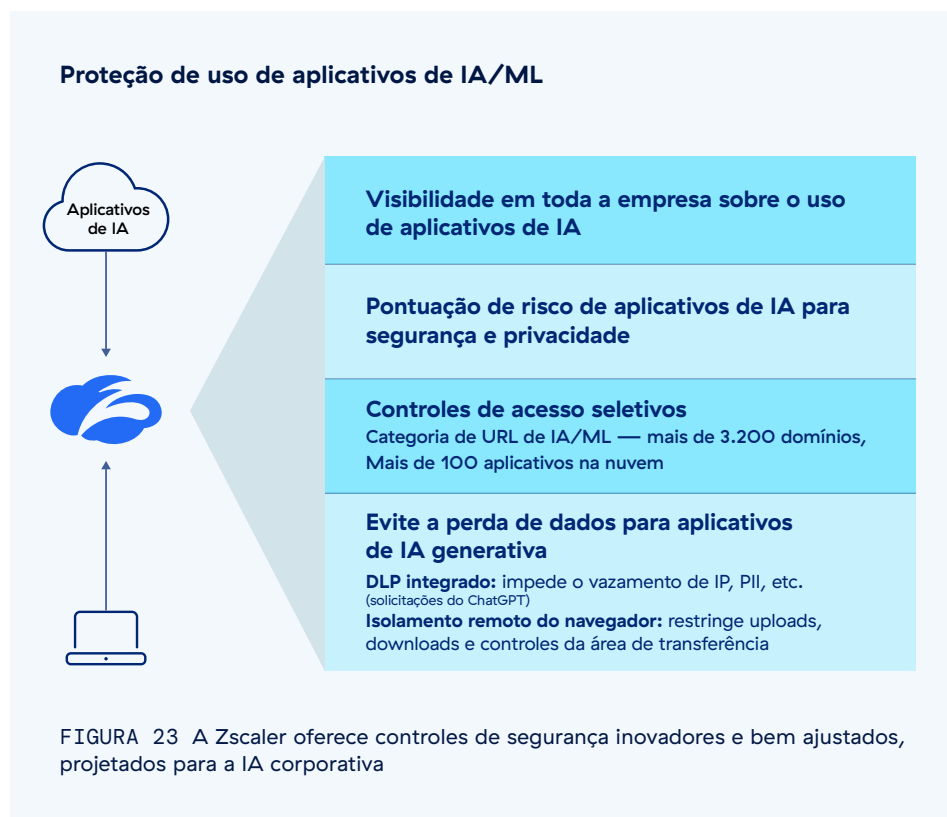
O **Zscaler Private Access™** protege os aplicativos, limitando a movimentação lateral com acesso de privilégio mínimo, segmentação de usuário para aplicativo e inspeção completa integrada do tráfego de aplicativos privados.

A **AppProtection** do Zscaler Private Access fornece inspeção de segurança integrada de alto desempenho de todas as cargas do aplicativo para expor ameaças.

O **Zscaler Deception™** detecta e contém invasores que tentam se mover lateralmente ou aumentar privilégios, atraindo-os com iscas de servidores, aplicativos, diretórios e contas de usuário.

Habilitando a transição da IA corporativa: o controle está em suas mãos

A Zscaler oferece uma maneira para as empresas promoverem a inovação, a criatividade e a produtividade com aplicativos de IA, ao mesmo tempo que mantêm os usuários e os dados seguros entre os canais emergentes para exfiltração de dados. Isso permite que as empresas [adotem o potencial transformador da IA](#) para acelerar os seus negócios sem bloquear completamente aplicativos e domínios de IA.



A ZSCALER PERMITE ÀS EMPRESAS:

- 01 Obter visibilidade total do uso de ferramentas de IA**
Registros detalhados fornecem visibilidade completa de como as equipes corporativas estão usando a IA, incluindo os aplicativos e domínios que estão visitando, bem como os dados e solicitações usados em ferramentas como o ChatGPT.
- 02 Criar políticas flexíveis para ajustar o uso de IA**
A poderosa e personalizada filtragem de URL para aplicativos de IA e ML permite que as empresas definam e apliquem facilmente controles granulares de acesso e segmentação de IA, bloqueando o acesso quando necessário e oferecendo acesso com níveis aceitáveis de risco usando o aplicativo de pontuação de risco da IA. As empresas podem permitir o acesso nos níveis corporativo, de departamento, de equipe e de usuário, bem como permitir o acesso baseado em cautela que orienta os usuários sobre os riscos das ferramentas de IA generativa. A segmentação baseada em IA facilita a identificação de segmentos de usuários apropriados para acesso a aplicativos específicos de IA, ao mesmo tempo que minimiza a superfície de ataque interna associada às ferramentas de IA.
- 03 Aplicar segurança granular de dados para o ChatGPT e outros aplicativos de IA**
As empresas podem evitar o vazamento de dados sigilosos carregados em aplicativos de IA com controles granulares de aplicativos na nuvem da Zscaler para IA generativa. Ao aplicar o mecanismo da Zscaler DLP, as empresas podem garantir que nenhum dado seja compartilhado acidentalmente ao usar qualquer ferramenta de IA. Ao mesmo tempo, a descoberta e classificação de dados com tecnologia de IA permite que as empresas identifiquem e criem facilmente políticas de DLP em torno de seus dados mais críticos, incluindo sua base de código corporativo, documentos financeiros e jurídicos, dados pessoais, dados de clientes e mais. [Este vídeo](#) demonstra como o mecanismo de DLP impede que os usuários insiram informações de cartão de crédito no ChatGPT.
- 04 Habilitar controles poderosos usando o isolamento do navegador**
O isolamento do navegador da Zscaler renderiza aplicativos de IA em um ambiente seguro, adicionando uma camada de proteção que permite solicitações e consultas do usuário às ferramentas de IA, ao mesmo tempo que restringe ações de copiar, colar, uploads e downloads. Isso ajuda a mitigar o risco de dados sigilosos serem acidentalmente compartilhados com ferramentas de IA generativa.

Os líderes empresariais e de segurança estão em uma encruzilhada: eles devem trabalhar para adotar a IA para impulsionar a inovação e permanecerem competitivos, mas também devem garantir que os seus dados apenas promovam os negócios, não as violações. A Zscaler capacita as empresas a passar por essa transição com confiança, aproveitando um conjunto completo de controles de segurança zero trust baseados em IA que protegem contra ataques conduzidos por IA, ao mesmo tempo que oferece políticas de IA ajustadas e proteções de dados necessárias para aproveitar todo o potencial da IA generativa.

Apêndice

Metodologia de pesquisa da ThreatLabz

A nuvem de segurança global da Zscaler processa mais de 300 trilhões de sinais diários e bloqueia 9 bilhões de ameaças e violações de políticas por dia, com mais de 250 mil atualizações de segurança diárias. Análise de 18,09 bilhões de transações de IA e ML de abril de 2023 a janeiro de 2024 na nuvem da Zscaler, a Zero Trust Exchange.

Sobre a Zscaler ThreatLabz

ThreatLabz é o braço de pesquisa de segurança da Zscaler. Essa equipe de classe mundial é responsável por perseguir novas ameaças e garantir que as milhares de organizações que usam a plataforma global Zscaler estejam sempre protegidas. Além da pesquisa de malware e análise comportamental, os membros da equipe estão envolvidos na pesquisa e no desenvolvimento de novos módulos para proteção avançada contra ameaças na plataforma Zscaler, e realizam regularmente auditorias internas de segurança para garantir que os produtos e a infraestrutura da Zscaler atendam aos padrões de conformidade de segurança. A ThreatLabz publica regularmente análises aprofundadas de ameaças novas e emergentes em seu portal, research.zscaler.com.





Viva seu mundo com total segurança.

Sobre a Zscaler

A Zscaler (NASDAQ: ZS) acelera a transformação digital para que os clientes tenham mais agilidade, eficiência, resiliência e proteção. A Zscaler Zero Trust Exchange™ protege milhares de clientes contra ataques cibernéticos e perda de dados, conectando com segurança usuários, dispositivos e aplicativos em qualquer local. Distribuída em mais de 150 data centers globalmente, a Zero Trust Exchange baseada em SASE é a maior plataforma de segurança integrada na nuvem do mundo. Para saber mais, visite www.zscaler.com.br.