IRON MOUNTAIN®

White paper

Como a IA generativa está redefinindo a segurança

Resumo

Quando se trata de segurança, a GenAI é uma ferramenta para o bem ou para o mal? As equipes de segurança corporativa estão se debatendo com o que essa tecnologia emergente representa para suas organizações. Elas percebem cada vez mais que a GenAI traz tanto oportunidades quanto riscos.

Conteúdo

GenAI: uma nova ferramenta para tudo

Certamente, uma das mais importantes inovações tecnológicas recentes é a inteligência artificial generativa (GenAI). Quando o ChatGPT foi lançado no final de novembro de 2022, ele marcou um divisor de águas em uma nova era da informática.

Você pode usar a GenAI para escrever um artigo ou para gerar o código de sua última atualização de software. Ela é capaz de gerar imagens que se parecem com fotos realistas ou criar vídeos e arquivos de áudio quase idênticos aos reais. Além disso, ela permite que você pesquise na Internet respostas para praticamente qualquer pergunta. E nós apenas arranhamos a superfície do que essa nova tecnologia pode fazer.

É compreensível que as organizações estejam empolgadas com as oportunidades que a GenAI representa. Uma enquete da AWS e o MIT constatou que 80% dos gestores de dados acreditam que a IA generativa transformará suas organizações, e 45% disseram que suas empresas já a adotaram amplamente.

"A IA generativa tem o potencial de ser a tecnologia mais disruptiva que já vimos até hoje", diz Steve Chase, líder de consultoria da KPMG nos EUA. "Ela mudará fundamentalmente os modelos de negócios, oferecendo novas oportunidades de crescimento, eficiência e inovação, ao mesmo tempo em que surgem riscos e desafios significativos. Para os líderes aproveitarem o enorme potencial da IA generativa, eles devem estabelecer uma estratégia clara que leve sua organização rapidamente da fase de experimentação para a industrialização".

Diante desse cenário, os executivos estão sob pressão para começar a implementar a GenAI o mais rápido possível. No entanto, é natural que alguns ainda estejam receosos. Apesar de todas as suas possíveis vantagens, a IA generativa também apresenta alguns riscos notáveis.

Muitos líderes de segurança ainda estão lutando para elaborar um plano sobre como lidar com a IA generativa. Na enquete da AWS e o MIT, 16% dos líderes de dados informaram que suas empresas proibiram completamente o uso de IA generativa, e apenas 6% disseram estarem usando IA generativa na produção.

O restante estava experimentando ao nível individual, de equipe ou organizacional, tentando entender melhor os recursos dessa ferramenta. Em essência, essa é a GenAI – uma ferramenta. Como qualquer ferramenta, ela pode ser usada para o bem ou para o mal.

Do ponto de vista da segurança cibernética, está claro que a GenAI tem implicações importantes para três grupos diferentes de pessoas: equipes de segurança, funcionários em geral e agentes mal-intencionados dentro ou fora da organização.

Equipes de segurança + GenAI = Melhor proteção

SOs fornecedores de segurança vêm integrando recursos de IA e aprendizado automático em seus produtos há anos, para o grande benefício de seus clientes. O surgimento da GenAI parece estar acelerando essa tendência ao aprimorar os recursos do software de segurança.

As ferramentas de IA generativa são particularmente boas na detecção de ameaças e na identificação de ataques. Segundo os analistas da Bain, "a identificação de ameaças tem o maior potencial para a IA generativa melhorar a segurança cibernética..." Eles também acrescentam que "a IA generativa já está ajudando os analistas a descobrir mais rapidamente um ataque e seu possível impacto. Por exemplo, ela pode ajudá-los a filtrar com mais eficiência alerta de incidentes, rejeitando falsos positivos. A capacidade da IA generativa de detectar e caçar ameaças se tornará cada vez mais dinâmica e automatizada".

Um estudo independente da IBM constatou que "as organizações com uso extensivo de IA e automação tiveram um ciclo de vida de vazamento de dados 108 dias mais curto em comparação com as organizações estudadas que não as implementaram (214 dias em relação a 322 dias)". Entretanto, o mesmo estudo da IBM constatou que 40% das organizações ainda não haviam implementado a IA e a automação da segurança.

Algumas equipes de segurança também estão usando a IA generativa para melhorar as estratégias de confiança zero. A GenAI pode ajudar a criar perfis de risco para diferentes endpoints. E seus recursos de correspondência de padrões podem ajudar a detectar qualquer evento anômalo.

A GenAI também pode aumentar as funções da equipe de outras maneiras. Por exemplo, ela permite investigar ameaças emergentes ou fornecedores potenciais. Ela também analisa históricos para encontrar padrões, auxilia na elaboração de relatórios e desenvolve políticas projetadas para evitar ou atenuar futuros incidentes de segurança.

Em resumo, a IA generativa pode aprimorar as habilidades da equipe existente, ajudando-a a ser mais eficiente e eficaz e, por fim, melhorar a segurança da organização.

Infelizmente, nem todos os possíveis efeitos da IA generativa são totalmente positivos.

Funcionários + GenAI = maior risco

Grande parte dos testes da GenAI ocorre fora do departamento de segurança. É provável que os colaboradores de quase todas as equipes de uma organização façam testes com a tecnologia. Entretanto, ninguém sabe realmente até que ponto os funcionários usam a GenAI, ou mesmo quais ferramentas eles usam. Assim como as organizações há muito tempo têm a "TI invisível", agora elas têm a "IA invisível".

Isso é um problema porque a IA generativa não só oferece grandes vantagens, mas também apresenta riscos significativos. E um dos potenciais problemas é o vazamento de dados.

As ferramentas de IA generativa são baseadas em grandes modelos de linguagem (LLMs – pela sigla em inglês). Esses LLMs absorvem um grande volume de texto, usado para prever a próxima palavra ao gerar um novo texto. Um de seus casos de uso mais úteis é a incorporação de texto existente e sua otimização.

No entanto, se os funcionários inserirem informações de identificação pessoal, códigos privados ou outros segredos da empresa no LLM, a ferramenta poderá armazenar esses dados e enviá-los para outra pessoa.

Os analistas da PwC explicam que "os aplicativos de GenAI podem exacerbar os riscos de privacidade e proteção de dados. Afinal, a promessa dos grandes modelos de linguagem é que eles usam uma grande quantidade de dados e criam ainda mais dados novos, que são vulneráveis a distorções, baixa qualidade, acesso não autorizado e perda".

Outro possível problema é que os desenvolvedores novatos podem, às vezes, usar chatbots para escrever códigos inseguros. "No contexto da segurança cibernética, devemos esperar que programadores inexperientes recorram a ferramentas de modelagem de linguagem preditiva para ajudá-los em seus projetos quando se depararem com um problema de codificação complexo", relata a InfoWorld.

Acrescenta que "embora não sejam inerentemente negativos, os problemas podem surgir quando as organizações não têm processos de revisão de código adequadamente estabelecidos e o código é implantado sem análise".

Os seres humanos poderiam facilmente vazar segredos da empresa ou escrever códigos inseguros sem usar a GenAI, é claro. Porém, essa nova ferramenta representa um novo vetor de vazamento de dados – especialmente um que é difícil de proteger.

Isso significa que a equipe precisará de mais treinamento sobre como usá-la de forma segura. As equipes de segurança e gerenciamento de riscos precisarão encontrar maneiras de monitorar o uso das ferramentas pelos funcionários para garantir que eles não exponham desnecessariamente a organização a um risco maior.

Agentes mal-intencionados + GenAI = Possíveis desastre

A implicação mais significativa da GenAI para a segurança empresarial é a possibilidade de agentes mal-intencionados usarem a IA generativa para cometer crimes. Ao mesmo tempo que torna mais fácil para as equipes de segurança encontrarem ataques, também facilita para os criminosos da web encontrarem novas maneiras de executar esses ataques.

Na verdade, já estão surgindo evidências de criminosos usando IA generativa. De acordo com a Bain, "as menções à IA generativa na dark web proliferaram em 2023. É comum ver hackers se gabando do uso do ChatGPT".

Para alguns aspirantes a hackers, a IA generativa facilita o início de suas atividades como criminosos cibernéticos. Em vez de aprender a escrever seu próprio código, eles podem fazer com que uma das ferramentas de codificação de IA faça isso por eles.

As ferramentas têm algumas ressalvas – você não pode pedir ao ChatGPT para escrever malware para você, por exemplo. Contudo, não é preciso ser muito inteligente para encontrar soluções para algumas dessas barreiras. Até mesmo os criminosos cibernéticos compartilham seus métodos com outras pessoas.

De qualquer forma, esses hackers novatos não são a maior ameaça às redes corporativas. As ferramentas de segurança corporativa devem ser capazes de impedir a maioria dos ataques de amadores. Uma ameaça muito maior é o possível ataque de phishing.

A maioria sabe que não deve confiar em um e-mail com muitas palavras escritas incorretamente. Inclusive, as iniciativas educacionais das empresas têm feito um bom trabalho ao treinar as pessoas para verificar novamente e-mails incomuns.

Agora, o que acontece quando esses e-mails soam exatamente iguais aos de todos os outros? E se for anexado um vídeo ou uma nota de voz que se pareça e soe exatamente como seu chefe?

Ferramentas de GenAI

As empresas estão incorporando funcionalidades da IA generativa a uma ampla variedade de produtos, e a cada dia surgem novos mais. Alguns dos aplicativos de GenAI mais populares são os seguintes:

ChatGPT – o revolucionário modelo de linguagem grande da OpenAI, que responde a perguntas e mantém conversas.

GitHub Copilot – assistente de codificação da IA que se autodenomina "a ferramenta de desenvolvedor de IA mais adotada do mundo".

Copy.ai – ferramenta de redação projetada para tarefas como blog e conteúdo de marketing.

Scribe – assistente de redação especializado na criação de documentação e guias.

Bing – motor de busca da Microsoft que agora incorpora as respostas do GPT-4.

Gemini – alternativa do Google ao ChatGPT e ao Bing.

Dall-E2 – ferramenta da OpenAI para criar imagens fotorrealistas a partir de texto.

Synthesia – plataforma de IA que transforma texto em vídeos realistas.

Rephrase.ai – plataforma de teste para vídeo com avatares de estoque e personalizados.

Bardeen – ferramenta de automação de fluxo de trabalho que cuida de tarefas tediosas.

Murf.ai – gerador de áudio para criar locuções com base em vozes humanas reais.

Designs.ai – ferramenta de design gráfico para criar logotipos, vídeos, anúncios, e muito mais.

A PwC aponta: "O risco mais imediato com o qual devemos nos preocupar? O phishing mais sofisticado. Iscas mais convincentes e personalizadas em bate-papos, vídeos, ou áudios ou vídeos falsos gerados em lives, passando-se por alguém conhecido ou em uma posição de autoridade".

Essas mesmas ferramentas também podem ser usadas para prejudicar a reputação de sua empresa na Internet. Agentes mal-intencionados podem criar imagens, áudio ou vídeo falsos e publicá-los nas mídias sociais. Eles também podem ameaçar publicá-los on-line com o intuito de exigir um resgate.

Além dos *deep fakes*, a GenAI também pode ser usada para gerar códigos maliciosos. Por exemplo, os pesquisadores de segurança da HYAS usaram a GenAI para criar um novo tipo de malware polimórfico que eles apelidaram de *"Black Mamba"*. Embora pareça um código benigno, o *Black Mamba* na verdade se reescreve em tempo de execução para se tornar um *keylogger* malicioso que extrai dados por meio do Microsoft Teams. Como o malware se reescreve continuamente, ele escapa até mesmo das mais fortes defesas de segurança cibernética.

"Usando essas novas técnicas, um agente de ameaças pode combinar vários comportamentos altamente detectáveis de uma forma incomum e evitar a detecção explorando a incapacidade do modelo de reconhecê-lo como um padrão malicioso", explicou HYAS. Além disso, ele acrescenta, *"esse problema é agravado quando é a IA que conduz os ataques cibernéticos, pois os métodos que ela escolhe podem ser altamente atípicos em comparação com os usados pelos humanos. Também, a velocidade com que esses ataques podem ser executados complica exponencialmente a ameaça"*.

Por mais assustador que seja o Black Mamba, os pesquisadores de segurança dizem que ele provavelmente não é a pior ameaça representada pela IA generativa. Esse título talvez pertença à incorporação indireta de prompts, um tipo de ataque que subverte ferramentas populares de GenAI, alimentando-as com dados maliciosos por meio de sites aparentemente normais.

A Wired informou: *"Em um experimento realizado em fevereiro, pesquisadores de segurança forçaram o chatbot do Bing da Microsoft a se comportar como um golpista"*. E acrescenta: *"Nas instruções ocultas em um site, os pesquisadores pediram ao chatbot que solicitasse à pessoa que o utilizasse os detalhes de sua conta bancária. Esse tipo de ataque, em que informações ocultas podem fazer com que a IA se comporte de maneira indesejável, é apenas o começo"*.

É quase certo que os criminosos cibernéticos estejam – neste exato momento – pensando em outras maneiras de usar a GenAI como método de ataque. Os responsáveis pela segurança e TI das empresas estão cientes da ameaça, mas até agora fizeram pouco para combatê-la.

Um estudo da McKinsey descobriu que 53% dos entrevistados acreditavam que a GenAI representava um risco de segurança, mas apenas 38% estão trabalhando para mitigar esse risco. Isso levanta a questão: o que eles deveriam estar fazendo?

Alternativas de ação

Se a GenAI fosse apenas uma ameaça, a resposta seria óbvia: as organizações bloqueariam seus sistemas para evitar que os colaboradores tivessem acesso a ela. Elas empregariam os métodos mais rigorosos possíveis para evitar e atenuar os ataques que incorporam as funções da GenAI. Agora, a GenAI não é apenas uma ameaça. Ela também é uma oportunidade.

As equipes de segurança inteligentes estão procurando maneiras de integrar essa ferramenta em suas organizações para ajudá-las a atingir suas metas e, ao mesmo tempo, combater a ameaça. Com isso em mente, considere as seguintes alternativas de ação.

1. Siga as medidas de segurança existentes. A boa notícia é que as ferramentas de segurança cibernética atuais oferecem essas medidas contra as ameaças da GenAI. Se você já conta com medidas de segurança robustas, está preparado para o novo mundo da IA generativa.

2. Melhore a proteção de seus modelos de IA. À medida que sua organização expande o uso da IA, seus modelos representam um alvo muito atraente. A PwC observa: "A GenAI acrescenta um ativo com valor para os agentes de ameaças e para o controle de sua organização. Eles poderiam manipular os sistemas de IA para fazer previsões incorretas ou negar serviços aos clientes". Portanto, a empresa recomenda: "Sua linguagem proprietária e modelos fundamentais, dados e novos conteúdos precisarão de proteções de defesa cibernética mais fortes".

3. Adicione a GenAI e a automação ao seu arsenal de segurança. As organizações que implementam a IA e a automação como parte de suas defesas encontram malware muito mais rapidamente do que aquelas que não fazem isso. Além disso, a GenAI também pode tornar sua equipe de segurança mais produtiva de muitas outras maneiras. Ao incorporar essas novas ferramentas em seus esforços contínuos, você estará mais bem preparado para impedir ataques que tentem usar a IA generativa contra você.

4. Eduque sua equipe. Como a IA generativa é muito nova, o conteúdo está em constante mudança. Incentive suas equipes a se manterem atualizadas com as informações mais recentes. A InfoWorld aconselha: "Agora, seria sensato repensar o treinamento de seus funcionários para incorporar diretrizes sobre o uso responsável de ferramentas de IA no local de trabalho". A empresa ainda menciona: "O treinamento dos profissionais também deve levar em conta a sofisticação da IA em novas técnicas de engenharia social".

5. Monitore a regulamentação. Além de monitorar as pesquisas ou informações — você também precisa ficar de olho nas respostas do governo às novas ferramentas. O Gartner observa: "A Lei de Inteligência Artificial da UE e outras bases regulatórias na América do Norte, na China e na Índia já estão estabelecendo normas para gerenciar os riscos dos aplicativos de IA". Ele acrescenta: "Esteja preparado para a conformidade, além do que já é exigido pelos regulamentos, como a proteção da privacidade".

6. Elabore e aplique políticas. Seus funcionários já usam IA generativa. Agora, se você for como a maioria das organizações, provavelmente ainda não implementou nenhuma regra sobre o que considera apropriado para o seu local de trabalho. Segundo a McKinsey, "apenas 21% dos entrevistados que relataram a adoção de IA afirmam que suas organizações estabeleceram políticas que regulam o uso dessas tecnologias pelos funcionários em seu trabalho". Isso é um problema porque, como declara a PwC, "sem a governança e a supervisão adequadas, o uso da GenAI pode criar ou ampliar os riscos legais".

7. Formalize o gerenciamento de riscos. O risco é inerente a todos os negócios. No entanto, as empresas inteligentes escolhem cuidadosamente os riscos que estão dispostas a correr e como mitigar os possíveis perigos. Esse processo é muito difícil, a menos que você disponha de um processo formal de gerenciamento de riscos. O Gartner prevê que "até 2026, os modelos de IA das organizações que implementarem a transparência, a confiança e a segurança da IA alcançarão uma melhoria de 50% em termos de adoção de metas de negócios e aceitação do usuário". Se sua organização não tiver um processo formal de gerenciamento de riscos ou se você não acreditar que seus processos são adequados para o desafio da IA generativa, procure ajuda de um parceiro externo, como os consultores de gerenciamento de riscos da Iron Mountain.

8. Aprimore o gerenciamento de informações. Você também pode proteger sua organização contra os perigos da GenAI governando e dando suporte eficaz aos seus dados. Ao seguir boas práticas de gerenciamento de informações, você terá menos chances de sofrer um vazamento de dados. E se seus sistemas forem violados, ter backups e recuperação de desastres em vigor pode proteger você contra a perda de dados. Novamente, é melhor procurar ajuda de um provedor de gerenciamento de informações como a Iron Mountain, para garantir que você esteja preparado para as ameaças da GenAI.

9. Avalie cuidadosamente os fornecedores. Ao procurar ferramentas GenAI ou ajuda relacionada à segurança cibernética e à proteção de dados, certifique-se de analisar todas as empresas externas com as quais você trabalha. Com as novas tecnologias, é tentador se apressar para implantar novas tecnologias. Embora você não deva se atrasar, é preciso dedicar tempo suficiente para ter certeza de que pode confiar nos parceiros escolhidos para ajudar você a integrar a GenAI aos seus fluxos de trabalho e proteger sua empresa dos riscos associados à GenAI.

Sobre a Iron Mountain

A Iron Mountain Incorporated (NYSE: IRM), fundada em 1951, é líder global em serviços de armazenamento e gerenciamento de informações. Com a confiança de mais de 225.000 organizações no mundo todo e uma rede imobiliária de mais de 85 milhões de pés quadrados em mais de 1.400 instalações localizadas em mais de 50 países, a Iron Mountain armazena e protege bilhões de ativos de informação, incluindo informações críticas de negócios, dados altamente confidenciais e artefatos culturais e históricos. Fornecendo soluções que incluem armazenamento seguro, gerenciamento de informações, transformação digital, destruição segura, bem como centros de dados, armazenamento e logística de arte e serviços em nuvem, a Iron Mountain ajuda as organizações a reduzir custos e riscos, cumprir as regulamentações, recuperar-se de desastres e possibilitar uma maneira mais digital de trabalhar.

IRON MOUNTAIN®

ironmountain.com/pt-br