A Cisco está adotando uma abordagem radical à segurança de IA em sua nova solução AI Defense.

Em uma entrevista exclusiva no domingo com Rowan Cheung da The Rundown AI, o vice-presidente executivo e CPO da Cisco, Jeetu Patel, disse que a AI Defense está “adotando uma abordagem radical para enfrentar os desafios que as soluções de segurança existentes não estão equipadas para enfrentar”.

A AI Defense, anunciada na semana passada, visa abordar os riscos no desenvolvimento e implantação de aplicações de IA, bem como identificar onde a IA é usada numa organização.

AI Defense pode proteger sistemas de IA contra ataques e salvaguardar o comportamento do modelo em plataformas com recursos como:

- Detecção de aplicativos de IA sombra e sancionados em nuvens públicas e privadas;
- Testes automatizados de modelos de IA para centenas de possíveis problemas de segurança; e
- A validação contínua protege contra possíveis ameaças à segurança, como injeção imediata, negação de serviço e vazamento de dados confidenciais.

A solução também permite que as equipes de segurança protejam melhor os dados de suas organizações, fornecendo uma visão abrangente dos aplicativos de IA usados pelos funcionários, criando políticas que restringem o acesso a ferramentas de IA não sancionadas e implementando proteções contra ameaças e perda de dados confidenciais, garantindo ao mesmo tempo a conformidade.

“A adoção da IA expõe as empresas a novos riscos que as soluções tradicionais de segurança cibernética não abordam”, disse Kent Noyes, chefe global de IA e inovação cibernética da empresa de serviços de tecnologia World Wide Technology em St. “O Cisco AI Defense representa um avanço significativo na segurança da IA, proporcionando visibilidade total dos ativos de IA de uma empresa e proteção contra ameaças em evolução.”

O que você vai ler:

- Passo positivo para segurança de IA
- Em relação às limitações
- Necessidade crítica no caminho para AGI
- Perdição Existencial?

Passo positivo para segurança de IA

MJ Kaufmann, autor e instrutor da O’Reilly Media, operadora de uma plataforma de

aprendizagem para profissionais de tecnologia, em Boston, afirmou a análise da Cisco sobre as soluções de segurança cibernética existentes. “A Cisco está certa”, disse ela ao TechNewsWorld. “As ferramentas existentes não conseguem resolver muitos ataques orientados operacionalmente contra sistemas de IA, como ataques de injeção imediata, vazamento de dados e ações não autorizadas de modelos.”

“Os implementadores devem agir e implementar soluções específicas para os resolver”, acrescentou ela.

A Cisco está em uma posição única para fornecer esse tipo de solução, observou Jack E. Gold, fundador e principal analista da J.Gold Associates, uma empresa de consultoria de TI em Northborough, Massachusetts. telemetria que pode ser usada para reforçar as capacidades de IA que desejam proteger”, disse ele ao TechNewsWorld.

A Cisco também deseja fornecer segurança em todas as plataformas – local, nuvem e multinuvem – e em todos os modelos, acrescentou.

“Será interessante ver quantas empresas adotarão isso”, disse ele. “A Cisco certamente está caminhando na direção certa com esse tipo de capacidade porque as empresas, em geral, não estão encarando isso de forma muito eficaz.”

Fornecer proteção multimodelo e multinuvem é importante para a segurança da IA.

“As soluções de IA multimodelo e multinuvem expandem a superfície de ataque de uma organização, introduzindo complexidade em ambientes distintos com protocolos de segurança inconsistentes, vários pontos de transferência de dados e desafios na coordenação do monitoramento e resposta a incidentes – fatores que os agentes de ameaças podem explorar mais facilmente”, Patricia Thaine, CEO e cofundadora da Private AI, uma empresa de segurança e privacidade de dados em Toronto, disse ao TechNewsWorld.

Em relação às limitações

Embora a abordagem da Cisco de incorporar controles de segurança na camada de rede por meio de sua malha de infraestrutura existente seja promissora, ela também revela limitações preocupantes, afirmou Dev Nag, CEO e fundador do QueryPal, um chatbot de suporte ao cliente com sede em São Francisco.

“Embora a visibilidade no nível da rede forneça telemetria valiosa, muitos ataques específicos de IA ocorrem nas camadas de aplicativos e modelos que o monitoramento de rede por si só não consegue detectar”, disse ele ao TechNewsWorld.

“A aquisição da Robust Intelligence no ano passado oferece à Cisco capacidades importantes em torno da validação de modelos e proteção de tempo de execução, mas seu foco na integração de rede pode levar a lacunas na segurança do ciclo de vida real de desenvolvimento de IA”, disse ele. “Áreas críticas como treinamento de segurança de

pipeline, verificação de modelo de cadeia de suprimentos e ajustes de proteção exigem integração profunda com ferramentas de MLOps que vão além do paradigma tradicional centrado em rede da Cisco.”

“Pense nas dores de cabeça que vimos com ataques à cadeia de suprimentos de código aberto, onde o código ofensivo é abertamente visível”, acrescentou. “Os ataques modelo à cadeia de suprimentos são quase impossíveis de detectar em comparação.”

Nag observou que, do ponto de vista da implementação, o Cisco AI Defense parece ser principalmente uma reembalagem de produtos de segurança existentes com alguns recursos de monitoramento específicos de IA em camadas.

“Embora sua extensa área de implantação ofereça vantagens para a visibilidade em toda a empresa, a solução parece mais reativa do que transformadora por enquanto”, afirmou ele. “Para algumas organizações que estão iniciando sua jornada de IA e que já trabalham com produtos de segurança da Cisco, o Cisco AI Defense pode fornecer controles úteis, mas aquelas que buscam recursos avançados de IA provavelmente precisarão de arquiteturas de segurança mais sofisticadas, desenvolvidas especificamente para sistemas de aprendizado de máquina.”

Para muitas organizações, a mitigação dos riscos de IA exige testadores de penetração humanos que saibam como fazer perguntas aos modelos que extraem informações confidenciais, acrescentou Karen Walsh, CEO da Allegro Solutions, uma empresa de consultoria em segurança cibernética em West Hartford, Connecticut.

“O lançamento da Cisco sugere que sua capacidade de criar proteções específicas para modelos mitigará esses riscos para evitar que a IA aprenda com dados ruins, responda a solicitações maliciosas e compartilhe informações não intencionais”, disse ela ao TechNewsWorld. “No mínimo, poderíamos esperar que isso identificasse e mitigasse problemas básicos para que os pen testers pudessem se concentrar em estratégias de comprometimento de IA mais sofisticadas.”

Necessidade crítica no caminho para AGI

Kevin Okemwa, escrevendo para o Windows Central, observa que o lançamento do AI Defense não poderia vir em melhor hora, já que os principais laboratórios de IA estão se aproximando da produção de uma verdadeira inteligência artificial geral (AGI), que supostamente replica a inteligência humana.

“À medida que a AGI se aproxima a cada ano que passa, os riscos não poderiam ser maiores”, disse James McQuiggan, defensor da conscientização de segurança na KnowBe4, um provedor de treinamento de conscientização de segurança em Clearwater, Flórida.

“A capacidade da AGI de pensar como um ser humano com intuição e orientação pode revolucionar as indústrias, mas também introduz riscos que podem ter consequências de longo alcance”, disse ele ao TechNewsWorld. “Uma solução robusta de segurança de IA garante que a AGI evolua de forma responsável, minimizando riscos como tomadas de decisão fraudulentas ou consequências não intencionais.”

“A segurança da IA não é apenas algo ‘bom de ter’ ou algo em que se pensar nos próximos anos”, acrescentou. “É fundamental à medida que avançamos em direção à AGI.”

Perdição Existencial?

Okemwa também escreveu: “Embora a Defesa de IA seja um passo na direção certa, sua adoção nas organizações e nos principais laboratórios de IA ainda está por ser vista. Curiosamente, o CEO da OpenAI (Sam Altman) reconhece a ameaça da tecnologia para a humanidade, mas acredita que a IA será inteligente o suficiente para evitar que a IA cause a destruição existencial.”

“Vejo algum otimismo sobre a capacidade da IA de se autorregular e prevenir resultados catastróficos, mas também noto na adoção que o alinhamento de sistemas avançados de IA com valores humanos ainda é uma reflexão tardia e não um imperativo”, Adam Ennamli, diretor de risco e segurança no General Bank of Canada disse ao TechNewsWorld.

“A noção de que a IA resolverá seus próprios riscos existenciais é perigosamente otimista, conforme demonstrado pelos atuais sistemas de IA que já podem ser manipulados para criar conteúdo prejudicial e contornar os controles de segurança”, acrescentou Stephen Kowski, CTO de campo da SlashNext, uma empresa de segurança de computadores e redes. empresa, em Pleasanton, Califórnia.

“As salvaguardas técnicas e a supervisão humana continuam a ser essenciais, uma vez que os sistemas de IA são fundamentalmente impulsionados pelos seus dados de formação e objectivos programados, e não por um desejo inerente de bem-estar humano”, disse ele ao TechNewsWorld.

“Os seres humanos são muito criativos”, acrescentou Gold. “Eu não acredito nessa bobagem

do Juízo Final. Descobriremos uma maneira de fazer a IA trabalhar para nós e fazê-lo com segurança. Isso não quer dizer que não haverá problemas ao longo do caminho, mas nem todos acabaremos em ‘Matrix’.”