SEGURANÇA DA INFORMAÇÃO EM RISCO: O IMPACTO DA FORMAÇÃO DE IAS MALICIOSAS

INFORMATION SECURITY AT RISK: THE IMPACT OF MALICIOUS AI TRAINING

Fernanda de Souza Silva
Faculdade de Tecnologia de Americana
fernanda.silva220@fatec.sp.gov.br

Katelyn dos Santos Alves
Faculdade de Tecnologia de Americana
katelyn.alves@fatec.sp.gov.br

Wagner José da Silva
Faculdade de Tecnologia de Americana
wagner.silva@fatec.sp.gov.br

Resumo

Com o acelerado avanço das tecnologias de inteligência artificial (IA), diversos desafios surgem acompanhados dos benefícios, sobretudo relacionados a como informações publicadas na internet são utilizadas para prover essas tecnologias com dados sensíveis, como imagens e informações pessoais expostas de forma inadequada. O impacto que a falta de proteção de dados pode ter na formação de IAs mal-intencionadas já tem suas consequências observadas em fraudes de identidade, invasões de sistemas privados e espionagem cibernética. Ao analisar as formas de vulnerabilidade em estruturas de segurança digital utilizadas de forma mal-intencionada para treinar essas tecnologias, é possível reconhecer as repercussões na privacidade de indivíduos e até mesmo em sua integridade. Examinando ataques reais alimentados por dados comprometidos, é possível propor políticas de proteção cibernética mais eficazes e, até mesmo, regulamentação no uso de inteligências artificiais.

Palavras-chave: Inteligência artificial, proteção de dados, segurança digital.

Abstract

With the rapid advancement of artificial intelligence (AI) technologies, various challenges arise alongside the benefits, especially related to how information published on the internet is used to feed these technologies with sensitive data such as images and personal information exposed inadequately. The impact that the lack of data protection can have on the development of malicious AIs has already been observed in identity fraud, private system breaches, and cyber espionage. By analyzing the vulnerabilities in digital security structures used maliciously to train these technologies, it is possible to recognize the repercussions on individuals' privacy and even their integrity. Examining real attacks fueled by compromised data allows for the proposal of more effective cybersecurity policies and even the regulation of artificial intelligence usage.

Keywords: Artificial intelligence, data protection, digital security.

1. Introdução

A segurança da informação contempla diversos pilares; dentre eles, a confidencialidade busca garantir que os dados expostos por alguém sejam mantidos em sigilo e protegidos por seu receptor, de forma a não ficarem vulneráveis. Entretanto, com o avanço da tecnologia e a implementação de inteligências artificiais cada vez mais próximas do uso rotineiro da internet, nota-se um crescimento mútuo em fraudes de identidade, invasões de sistemas privados e espionagem cibernética, impulsionadas pela forma como informações mal protegidas são utilizadas para incorporar o aprendizado de algoritmos maliciosos.

O potencial desta tecnologia para ser utilizada com fins maléficos é um risco vagamente discutido entre a sociedade, causando com que esta não se atente ao tipo de informação sensível que compartilha nos meios sociais e, consequentemente, resultando em brechas de segurança originárias do próprio emissor da informação.

Este artigo tem como objetivo explorar a forma em que as vulnerabilidades do sistema e as falhas na proteção de dados compartilhados em meios sociais contribuem para a formação destas tecnologias que podem lícita, mas implicitamente, disseminar informações de forma mal-intencionada. Serão abordadas também as maneiras de identificar uma informação formada de tal forma e como impedir a disseminação de informações sensíveis. A abordagem será adotada a partir de uma perspectiva de segurança da informação.

2. Metodologia

Este artigo foi desenvolvido com base em uma abordagem mista, utilizando métodos quantitativos e qualitativos para coletar e analisar dados relacionados à segurança da informação e ao impacto da falta de proteção para alimentar inteligências artificiais maliciosas. A pesquisa começou com uma ampla revisão bibliográfica, consultando artigos científicos, estudos acadêmicos e fontes online confiáveis, abordando temas como segurança da informação, proteção de dados e uso de inteligência artificial para fins maliciosos. Para isso, foram utilizadas bases de dados como Google Scholar, SciELO e IEEE Xplore, além de artigos de especialistas e notícias sobre violações de dados e segurança cibernética.

Em seguida, os dados primários foram coletados por meio de um questionário online distribuído a um grupo diversificado de indivíduos. O questionário abrange temas como uso de mídias sociais e exposição de informações pessoais, percepções dos usuários sobre publicidade personalizada e coleta de dados, interações com inteligência artificial e preocupações com privacidade, além de compartilhar informações acadêmicas e profissionais com assistentes virtuais e sistemas de IA. Os dados coletados foram analisados quantitativamente, auxiliando na identificação de padrões e tendências de comportamento dos usuários em relação à segurança de suas informações.

Ao mesmo tempo, a análise qualitativa das respostas aos questionários abertos e a revisão da literatura aprofundaram a compreensão das práticas de proteção de dados. Além disso, foram examinados estudos de caso documentados em outros artigos e notícias, destacando como a falta de segurança das informações pessoais contribuiu para o desenvolvimento de inteligências artificiais maliciosas, ilustrando assim que as violações da privacidade podem ser exploradas.

Com esta abordagem, o estudo buscou fornecer um panorama das vulnerabilidades na segurança das informações pessoais e como essas vulnerabilidades podem ser exploradas pelos sistemas de IA de forma antiética, colocando o usuário em uma posição de exposição social e fragilidade emocional e destacando os riscos crescentes em um cenário digital cada vez mais exposto.

3. Estrutura de uma inteligência artificial

O termo “Inteligência Artificial”, em seu princípio mencionado pelo cientista americano John McCarthy e anteriormente referido pelo pai da computação Alan Turing como “Máquina que pensa”, possui diversas vertentes em seu significado. Turing definiu o conceito ao realizar o famoso “Teste de Turing”, um exame desenvolvido em 1950 que pretendia descobrir se um computador conseguiria demonstrar a mesma inteligência de uma pessoa. Por outro lado, os pesquisadores Allen Newell, J.C. Shaw e Herbert Simon buscaram o desenvolvimento de IAs como formas de solucionar problemas matemáticos complexos com uma linha de raciocínio mais precisa do que a de seres humanos, a fim de provar teoremas.

É inegável que, independentemente da definição mais aceita pela sociedade do que é uma inteligência artificial, ela se aproxima cada vez mais do cotidiano, de forma a ter passado em um curto espaço de tempo de uma tecnologia inovadora e exclusiva para uma ferramenta quase indispensável no estilo de vida moderno. Implementada em assistentes virtuais como a Alexa da Amazon, o Google Assistente e até mesmo a Siri da Apple, o contato entre cidadão e máquina se tornou paulatinamente algo tão natural que se torna inevitável o desaparecimento da linha tênue que separa os dois ao longo do tempo.

Por conta dessa aproximação e constante troca de conhecimentos, não apenas o ser humano utiliza ferramentas de IA para aprender, mas também as ferramentas de IA utilizam o ser humano para observar e desenvolver características mais próximas às dos seres humanos. Zhi-Hua Zhou (2001) afirma que, assim como as pessoas, máquinas também aprendem com as experiências.

Zhi-Hua Zhou caracteriza em seu livro Machine Learning a forma como o aprendizado da máquina se baseia simplesmente em um algoritmo, citando Mitchell (1977) ao dizer que um computador é ensinado a aprender de experiências E para uma classe de tarefas T e cálculo de performance P, se a performance em tarefas T, medidas por P, progride com a experiência E.

A entrada de dados — experiência — pode ser originária de diversas fontes, sejam elas legais ou não; com esses dados, a performance será aperfeiçoada, gerando a realização de tarefas mais precisas. Um exemplo corrente do quanto máquinas tiveram seu conhecimento complementado nos últimos anos pode ser observado nas imagens abaixo, todas formadas pela conhecida plataforma de IA generativa de imagens, o Midjourney.

A primeira imagem demonstra a geração de uma ilustração de uma mulher, formada ainda no início da plataforma em 2022; seguindo com as imagens a seguir, pode-se observar que ao longo dos meses seguintes a ilustração formada com os mesmos elementos de composição foi ficando mais apurada, em um espaço de tempo tão curto, para entender, agregar detalhes, ser mais coerente e deixar de lado elementos óbvios de artificialidade, ao ponto de se tornar quase indistinguível de uma imagem real.

Figura 1 – Evolução de imagens geradas por inteligência artificial

Fonte: https://www.hardware.com.br/artigos/evolucao-das-imagens-geradas-por-ia/

É necessário que, para que a performance aumente com o intuito de gerar ilustrações cada vez mais parecidas com o mundo real, sejam absorvidas imagens reais para terem seus elementos de composição analisados, trazendo à tona a tese que discute o fato dessas imagens reais serem tiradas muitas vezes de plataformas sociais, onde constantemente os usuários não têm conhecimento de que suas fotos e vídeos são utilizados para tal finalidade.

4. Aspectos da segurança da informação

A proteção dos dados digitais tornou-se uma prioridade vital num mundo cada vez mais interligado. A segurança da informação inclui uma variedade de práticas, políticas e tecnologias destinadas a garantir a integridade, confidencialidade e disponibilidade das informações.

À medida que os sistemas computacionais evoluíram, especialmente desde a década de 1970, surgiram novos desafios, que vão desde ataques cibernéticos sofisticados a falhas internas, exigindo soluções eficazes para proteger os dados sensíveis das empresas, dos governos e dos indivíduos. Uma das abordagens mais básicas para a segurança da

informação é a implementação da criptografia, uma técnica que permite que os dados sejam codificados para que apenas destinatários autorizados possam acessá-los.

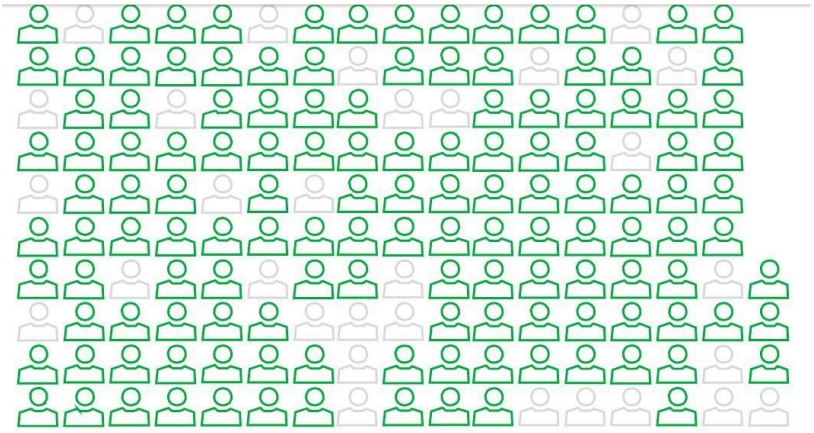
Carla Oliveira (2001) descreve em seu estudo a explicação feita pelo cientista Whitfield Diffie durante o desenvolvimento da criptografia de chave pública na década de 1970; a criptografia garante que, mesmo que os dados sejam interceptados, eles não serão inteligíveis para quem não possui a chave correta. Esta técnica evoluiu ao longo dos anos e é amplamente utilizada em sistemas de comunicação, comércio eletrônico e transferência de dados confidenciais, como em instituições bancárias.

Além da criptografia, outro elemento essencial da segurança da informação é o controle de acesso. Esse conceito vai desde senhas e autenticação de dois fatores até sistemas biométricos mais sofisticados, como reconhecimento facial ou de impressão digital. Howard e Longstaff (1988) afirmam que a segurança baseada em controle de acesso é fundamental para garantir que apenas indivíduos autorizados possam ter acesso a informações e sistemas sensíveis.

A evolução das ameaças cibernéticas também estimulou o desenvolvimento de firewalls, antivírus e sistemas de detecção de intrusões (IDS), ferramentas que monitoram e analisam o tráfego de rede em busca de atividades suspeitas. A Organização Internacional de Padronização (ISO) estabeleceu normas como a ISO/IEC 27001, que fornece uma estrutura para a implementação de sistemas de gestão de segurança da informação (SGSI) para proteger informações críticas de forma contínua e estruturada. Além disso, no Brasil existem legislações com foco na garantia da segurança de informações pessoais, como a Lei Geral de Proteção de Dados (Lei Nº 13.709/2018).

No entanto, apesar destes níveis de proteção, os maiores riscos para a segurança da informação advêm frequentemente do fator humano. A falta de treinamento adequado, o uso de senhas fracas e a vulnerabilidade a ataques de phishing são algumas das principais causas de violações de segurança. De acordo com Verizon (2022) em seu relatório de Investigações de Violações de Dados (DBIR), grande parte das violações de segurança cibernética envolvem erro humano de alguma forma, seja por negligência ou falta de conhecimento das melhores práticas de segurança.

Figura 2 - Quebras de segurança causadas por falhas de segurança



Fonte: https://www.verizon.com/business/dam/img/resources/reports/2022/dbir/figure-9.svg

A segurança da informação, portanto, é um campo dinâmico que exige atualização constante para lidar com novas ameaças e tecnologias. A combinação de ferramentas avançadas, políticas bem definidas e usuários conscientes é fundamental para manter a segurança em um cenário digital em rápida evolução.

5. Proteção de informações em mídias sociais

À medida que entramos numa era digital cada vez mais dependente de tecnologias emergentes, a Inteligência Artificial (IA) emerge como uma ferramenta indiscutível de progresso. Contudo, os riscos associados ao seu uso inadequado são cada vez mais evidentes. A falta de controle adequado das informações utilizadas para treinar essas IAs pode levar a graves violações de privacidade, além de ameaçar a segurança dos dados pessoais. O avanço dessas tecnologias, quando alimentadas por dados comprometidos, tem permitido o desenvolvimento de algoritmos capazes de realizar fraudes de identidade e outras atividades cibernéticas ilegais de forma assustadoramente eficaz.

De acordo com Nakamura e Geus (2002, p. 9): “A informação é um bem que, como qualquer outro ativo importante, é essencial para os negócios de uma organização e, consequentemente, necessita ser adequadamente protegida."

A informação sensível, seja governamental, comercial ou pessoal, torna-se vulnerável a ataques, especialmente quando o controle sobre esses dados é fraco. Confiar na IA para gerir essas informações pode intensificar o risco, uma vez que sistemas mal treinados ou destinados a fins ilegítimos podem comprometer profundamente a integridade e a confidencialidade dos dados.

Com base em uma pesquisa realizada durante o mês de agosto de 2024 de elaboração autoral, sobre as práticas de segurança digital, percebe-se que a conscientização dos usuários é insuficiente. Quase metade das pessoas compartilha dados pessoais sem verificar as políticas de privacidade, e uma parcela significativa não lê os termos de serviço antes de usar plataformas online. Esses comportamentos refletem o despreparo da sociedade diante de um cenário digital em que dados são constantemente explorados e manipulados, muitas vezes sem que o usuário tenha pleno conhecimento disso.

O aprendizado automático, que é a base para o funcionamento da IA, precisa de grandes volumes de dados para se aprimorar, e a fonte desses dados nem sempre é segura. Plataformas como redes sociais, bancos de dados públicos e até dispositivos de segurança doméstica fornecem inadvertidamente materiais para a evolução de inteligências maliciosas. Isso é especialmente preocupante considerando que 75% dos usuários percebem anúncios relacionados às suas buscas (Figura 3) e 68,8% relatam que esses anúncios aparecem com base em conversas presenciais, sugerindo uma coleta ativa e não transparente de dados (Figura 4). Esse tipo de rastreamento alimenta IA generativa, capaz de violar a privacidade e criar perfis falsos ou deepfakes, sem que os usuários estejam plenamente cientes.

Figuras 3 e 4 – Análise de gráficos com resultados de pesquisa autoral

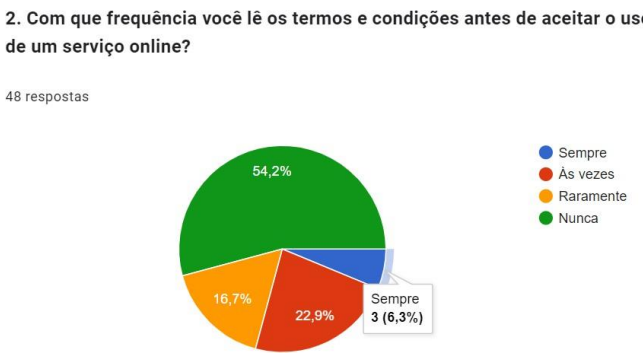
Fonte: Figuras elaboradas pelos autores.

Além da invasão dos sistemas, a manipulação dos próprios dados é um problema crescente. A IA maliciosa pode explorar fraquezas nos sistemas de segurança, gerando desinformação ou perpetuando fraudes de identidade. Deepfakes, por exemplo, tornaram-se uma ferramenta poderosa e perigosa nas mãos de criminosos digitais. Essas falsificações de vídeos e imagens são cada vez mais difíceis de serem detectadas por seres humanos ou pelos sistemas de verificação tradicionais. Os impactos podem ser devastadores, atingindo desde indivíduos — cujas reputações podem ser manchadas — até grandes corporações e governos, que podem ser vítimas de ataques de desinformação massiva.

A pesquisa revela que o comportamento negligente dos usuários contribui significativamente para o treinamento de IAs maliciosas. A ausência de uma cultura de segurança cibernética sólida, aliada à falta de conhecimento técnico, facilita a exploração de vulnerabilidades por algoritmos maliciosos. Isso é evidenciado pelo fato de que 54,2% das pessoas nunca leem os termos e condições, e 16,7% já foram vítimas de vazamento de dados.

Esses vazamentos servem como combustível para IAs treinadas para roubar informações sensíveis e realizar atividades criminosas.

Figuras 5 e 6 – Análise de gráficos com resultados de pesquisa autoral



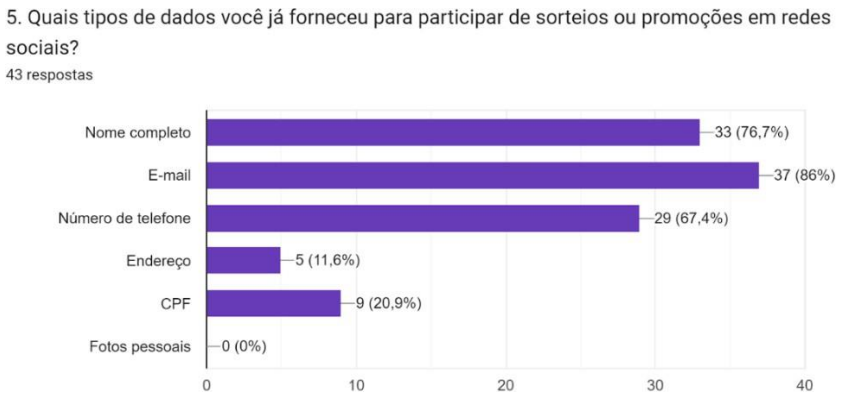
Fonte: Figuras elaboradas pelos autores.

A coleta de dados em massa, muitas vezes sem o conhecimento ou consentimento dos indivíduos, expõe usuários e organizações a riscos que vão além da invasão de privacidade. Algoritmos maliciosos de IA podem usar essas informações para espionagem cibernética, manipulação de dados e criação de perfis detalhados que auxiliam no desenvolvimento de ciberataques altamente direcionados. As falhas nos sistemas de segurança tradicionais, como a criptografia e a autenticação multifatorial, são superadas por tecnologias de IA que se adaptam rapidamente a novos métodos de defesa, revelando brechas que antes eram consideradas seguras.

Estudos também apontam o impacto crescente da IA generativa na privacidade e segurança cibernética: quanto maior a quantidade de dados disponíveis, maior a possibilidade de exploração de vulnerabilidades. As plataformas de redes sociais, por exemplo,

proporcionam um terreno fértil para treinar IA capaz de violar a privacidade ou criar perfis falsos, como nos sorteios e promoções onde dados como nome completo (76,7%) e e-mail (86%) são compartilhados. Isso facilita a coleta de dados e o uso subsequente para fins maliciosos, como a criação de identidades falsas ou ataques personalizados.

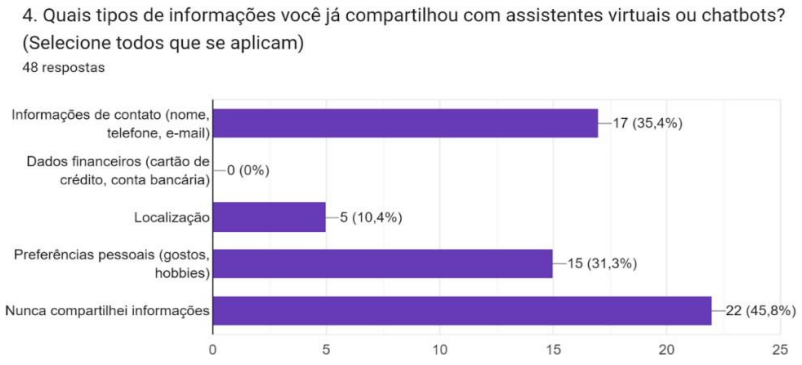
Figura 7 – Análise de gráficos com resultados de pesquisa autoral



Fonte: Figura elaborada pelos autores.

O surgimento de deepfakes, vídeos e imagens gerados por IA que imitam quase perfeitamente pessoas reais, levantou preocupações sobre a integridade das informações digitais. Estas falsificações tornaram-se cada vez mais difíceis de detectar, ameaçando não só a reputação dos indivíduos, mas também a confiança do público na autenticidade dos conteúdos digitais. A pesquisa revelou que 45,8% dos usuários nunca compartilharam dados sensíveis com assistentes virtuais, mas uma porcentagem significativa o fez, com informações de contato (35,4%) e preferências pessoais (31,3%), o que pode ser facilmente explorado por IAs maliciosas para atividades ilícitas.

Figura 8 – Análise de gráficos com resultados de pesquisa autoral



Fonte: Figura elaborada pelos autores.

É notável que, informações expostas sem consentimento resultam também em impactos poucos discutidos no meio atual, como a repercussão social e emocional sobre as vítimas destes. Um estudo realizado pela Fia Business School (2022) afirma que as vítimas de vazamentos de dados de cunho íntimo estão diretamente expostas a danos psicológicos incalculáveis, além de estarem sujeitas a chantagem de seus transgressores e constantes ameaças à sua imagem e reputação, situação que desencoraja a denúncia de tais atos.

Portanto, a implementação de medidas de segurança mais robustas e a conscientização sobre o uso responsável das informações são imperativas para mitigar os riscos associados ao desenvolvimento de IAs maliciosas. Apenas através da educação digital e da vigilância contínua das práticas de segurança cibernética será possível conter os impactos negativos dessas tecnologias emergentes.

6. Resultados

Para mitigar os riscos de segurança da informação decorrentes da formação de IAs maliciosas, é essencial adotar uma abordagem estratégica e multidimensional. A solução deve incluir camadas de proteção que atuem na prevenção, detecção e mitigação dos impactos cibernéticos, com foco na integridade dos dados e no controle das informações usadas no treinamento das inteligências artificiais.

6.1 Educação e Conscientização Digital

A segurança da informação começa com o conhecimento e conscientização de indivíduos e empresas. Medidas de educação digital são cruciais para evitar a exposição inadvertida de dados, como treinamentos em:

- Segurança de dados para profissionais de IA e usuários, garantindo que entendam a importância de proteger informações compartilhadas.
- Privacidade digital para práticas seguras de navegação e identificação de ameaças cibernéticas, como phishing e coleta ilegal de dados em redes sociais.

6.2 Políticas de Privacidade e Regulações Rígidas

As regulamentações como a GDPR na Europa e a LGPD no Brasil são vitais para estabelecer limites claros sobre o uso e tratamento de dados. Essas leis 11 demandam que empresas tratem as informações com transparência e responsabilidade, aplicando penalidades severas às violações:

- Aplicação rigorosa dessas leis em nível internacional, mesmo em um contexto de fronteiras digitais difusas.
- Monitoramento e auditoria contínua de empresas que utilizam IA em suas operações para garantir que respeitem essas regulamentações.

Figura 9 – Fundamentos da Lei Geral de Proteção de Dados

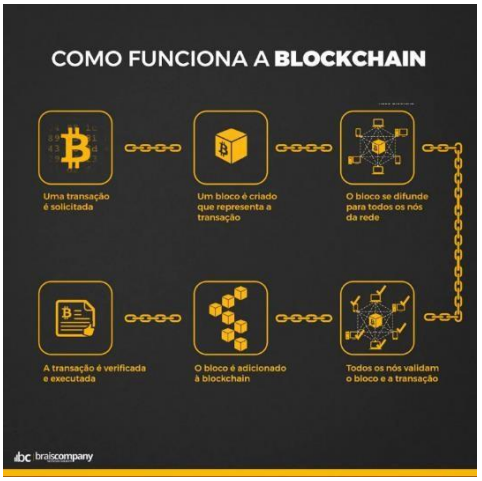
Fonte: https://www.serpro.gov.br/lgpd/menu/tratamento-dos-dados/objetivo-e-abrangencia-da-lgpd

6.3 Uso de Criptografia e Medidas de Segurança Cibernética Avançadas

Dado o cenário de risco representado pelo uso inadequado da IA, é essencial a implementação cumulativa de técnicas e políticas robustas de proteção de dados como: *criptografia avançada*, como o *protocolo Diffie-Hellman*, que permite troca segura de chaves criptográficas, protegendo as informações durante a transmissão e a*utenticação multifatorial (MFA)* e *controle de acesso baseado em funções (RBAC)* para restringir o acesso a dados sensíveis, garantindo que apenas indivíduos autorizados possam manipulá-los.

Além disso, *sistemas baseados em blockchain* podem ser utilizados para garantir a i*mutabilidade e rastreabilidade dos dados*, tornando mais difícil para invasores alterarem informações de maneira ilegal. O blockchain oferece uma camada extra de segurança, garantindo que qualquer alteração nos dados seja rastreada de forma transparente.

Figura 10 – Esquematização do funcionamento do blockchain



Fonte: Renan Lima/Braiscompany

Essa camada extra de proteção é comumente observada em operações como transações monetárias de criptomoedas, que são realizadas de forma a serem submetidas a diversas etapas, diminuindo assim, as chances de uma quebra de segurança ou ataque cibernético durante a operação e aumentando a rastreabilidade dos dados de forma segura.

6.4 Desenvolvimento de IA Ética e Segura

Para garantir que as inteligências artificiais sejam treinadas com integridade, é necessário: auditoria técnica contínua das IAs, avaliando suas fontes de dados e algoritmos para evitar o uso de informações comprometidas ou maliciosas; utilização de dados sintéticos no treinamento dos modelos, reduzindo a dependência de dados reais, muitas vezes extraídos sem consentimento adequado; códigos abertos e transparentes para que especialistas e órgãos reguladores possam auditar e verificar as práticas seguras no desenvolvimento de IA.

6.5 Fortalecimento das Medidas de Segurança Cibernética

A proteção dos dados que alimentam as IA requer soluções mais avançadas: c*riptografia de ponta a ponta e MFA* para proteger o acesso a sistemas críticos e b*lockchain* para garantir a imutabilidade dos dados e *sistemas de detecção de intrusões* para monitorar atividades suspeitas em tempo real. Além disso, *estratégias de monitoramento proativas e inteligência de ameaças* são necessárias para identificar, isolar e neutralizar atividades maliciosas, como tentativas de vazamento de dados ou manipulação de informações por IAs maliciosas.

6.6 Colaboração Global e Compartilhamento de Informações

A proteção contra IA maliciosas exige uma colaboração internacional entre governos, empresas de tecnologia e organizações de segurança cibernética: criação de redes seguras para compartilhar informações sobre novas ameaças e vulnerabilidades, facilitando uma resposta coordenada em grande escala; estabelecimento de padrões globais de segurança, garantindo que, independentemente da localização, todos sigam diretrizes claras para o uso ético e seguro de dados e IA.

6.7 Desenvolvimento de Ferramentas de Detecção de Deepfakes e Algoritmos Maliciosos

Com o surgimento de deepfakes e outras falsificações digitais, é essencial investir em: ferramentas avançadas de detecção de deepfakes, que utilizem IA para identificar padrões anômalos em vídeos e imagens e monitoramento de comportamento de IA, com sistemas capazes de detectar e bloquear ações suspeitas antes que causem danos

7. Conclusão

Os avanços na inteligência artificial (IA) trouxeram oportunidades e desafios significativos para o campo da segurança da informação. O uso inadequado de dados para treinar essas IA, aliado à crescente sofisticação dos ataques cibernéticos, tem revelado um cenário alarmante de vulnerabilidades em sistemas antes considerados seguros. Algoritmos e técnicas maliciosos, como deepfakes, tornaram-se cada vez mais difíceis de detectar e combater, ameaçando a privacidade e a integridade das informações numa escala sem precedentes.

Portanto, a necessidade de implementar uma abordagem de segurança multidimensional torna-se essencial. Tecnologias avançadas como a criptografia Diffie-Hellman - um algoritmo de chave assimétrica que permite a dois ou mais usuários criar e compartilhar uma chave de criptografia secreta - a autenticação multifatorial – autenticação realizada em duas ou mais etapas de formas diferentes - e a blockchain – armazenamento de dados em blocos interligados em uma cadeia - revelaram-se eficazes na mitigação destes riscos, mas devem ser continuamente melhoradas e complementadas por políticas rigorosas de governança de dados. Leis como o regulamento Geral de Proteção de Dados (GDPR), lei da União Europeia que entrou em vigor em 25 de maio de 2018 GDPR e a lei Geral de Proteção de Dados Pessoais (LGPD), n° 13.709/2018, são essenciais para a proteção da privacidade individual e devem ser aplicadas com rigor, com atenção especial ao cenário global, onde a aplicação dessas leis ainda enfrenta desafios.

Além disso, é fundamental que o treinamento e o desenvolvimento da IA sejam feitos de forma ética, com transparência sobre as fontes de dados e a intenção de utilização dessas tecnologias. Somente combinando inovação tecnológica, regulamentação forte e consciência global será possível reduzir os impactos negativos da inteligência artificial maliciosa e proteger a sociedade de futuras ameaças cibernéticas.

Em suma, a segurança da informação na era da IA exige esforços contínuos e colaborativos entre governos, empresas e desenvolvedores para criar um ambiente digital seguro e confiável que valorize a privacidade e os direitos dos indivíduos num cenário em constante evolução.

Referências

BRAISCOMPANY. Blockchain: quais os segredos da tecnologia mais segura do mundo? **Portal de notícias G1.** 2021. Disponível em: https://g1.globo.com/pb/paraiba/especial-publicitario/braiscompany/braiscompany/noticia/2021/04/07/blockchain-quais-os-segredos-da-tecnologia-mais-segura-do-mundo.ghtml. Último acesso em: 20 out. 2024.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. **Lei Geral de Proteção de Dados Pessoais (LGPD).** Brasília, DF: Presidência da República, 2020. Último acesso em: 12 de nov. 2024.

BRASIL. Projeto de Lei n° 2338 de 2023. **Dispõe sobre o uso da Inteligência Artificial**. Brasília, DF: Senado Federal. Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/157233 . Último acesso em: 12 de nov. 2024.

BRASIL. Projeto de Lei n° 872 de 2021. **Dispõe sobre o uso da Inteligência Artificial.**, DF: Senado Federal. Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/147434#:~:text=Disp%C3%B5e%20sobre%20o%20uso%20da,da%20Intelig%C3%AAncia%20Artificial%20no%20Brasil. Último acesso em: 12 de nov. 2024.

DINIZ, ANDRÉ K**.** O impacto da inteligência artificial no direito**. Revista da Faculdade de Direito da Universidade de São Paulo**, v. 114, p. 173-199, 2019. Disponível em: https://www.revistas.usp.br/rfdusp/article/view/156553/152042. Último acesso em: 20 out. 2024.

FERNANDES, F. **O que é inteligência artificial: veja como surgiu, exemplos e polêmicas**. Canal de notícias TechTudo, 2023. Disponível em: https://www.techtudo.com.br/guia/2023/03/o-que-e-inteligencia-artificial-veja-como-surgiu-exemplos-e-polemicas-edsoftwares.ghtml. Último acesso em: 20 out. 2024.

FIA. **Vazamento de dados: o que fazer e principais casos do Brasil**. FIA Business School, 2022. Disponível em: https://fia.com.br/blog/vazamento-de-dados/. Último acesso em: 12 nov. 2024.

GRASSI, Paul A. et al. NIST Special Publication 800-63b: Digital Identity Guidelines. **National Institute of Standards and Technology (NIST)**, v. 27, 2017. Disponível em: https://pages.nist.gov/800-63-3/sp800-63b.html. Último acesso em: 20 out. 2024.

HAND, D. J.; Mannila, H.; Smyth, P. **Principles of data mining**. Cambridge: MIT Press, 2001.

HOWARD, J. D.; Longstaff, T. A. **A common language for computer security incidents**. Disponível em: https://nsarchive.gwu.edu/sites/default/files/documents/4530309/John-D-Howard-Thomas-A-Longstaff-Sandia-National.pdf. Último acesso em: 20 out. 2024.

ISO/IEC. International Standarts Organization - ISO/IEC 27001, 27001:2022. Disponível em: https://www.iso.org/standard/27001. Último acesso em: 13 nov. 2024.

KAUFMAN, D. Desmistificando a Inteligência artificial**. Ed. Autêntica**. Belo Horizonte, 2022. Disponível em: https://books.google.com.br/books?hl=pt-BR&lr=&id=3LxtEAAAQBAJ&oi=fnd&pg=PT5&dq=intelig%C3%AAncia+artificial+artigo&ots=9Adkf6LCyO&sig=fsCX7DDjsMaP_uHeL2j66qTTUNc#v=onepage&q=intelig%C3%AAncia%20artificial%20artigo&f=false. Último acesso em: 20 out. 2024.

MITCHELL, T. M. Machine learning**.** New York: **McGraw Hill**, 1997.

MITCHELL, T. M. Version spaces: a candidate elimination approach to rule learning. In*:* **Proceedings of the 5th International Joint Conference on Artificial Intelligence (IJCAI),** Cambridge, MA, p. 305-310, 1977.

PAGIM ZEQUIM, E.; FRANCISCO RIBEIRO, D. . O PAPEL DA INTELIGENCIA ARTIFICIAL NA SEGURANÇA CIBERNETICA : o uso de sistemas inteligentes em benefício da segurança dos dados das empresas. **Revista Interface Tecnológica**, *[S. l.]*, v. 19, n. 1, p. 21–33, 2022. DOI: 10.31510/infa.v19i1.1358. Disponível em: https://revista.fatectq.edu.br/interfacetecnologica/article/view/1358. Acesso em: 20 out. 2024.

PEREIRA, L. M. **Inteligência artificial: mito e ciência**. Universidade NOVA de Lisboa. Lisboa, 1988. Disponível em: https://www.researchgate.net/profile/Luis-Pereira-25/publication/237130636_Inteligencia_Artificial_Mito_e_Ciencia/links/00463527ca46b52079000000/Inteligencia-Artificial-Mito-e-Ciencia.pdf. Último acesso em: 20 out. 2024.

PLAZA, R.W. A evolução das imagens geradas por IA**. Portal de notícias de tecnologia Hardware.com.br.** Disponível em: https://www.hardware.com.br/artigos/evolucao-das-imagens-geradas-por-ia/. Último acesso em: 20 out. 2024.

SERPRO. **Objetivo e abrangência da LGPD**. Disponível em: https://www.serpro.gov.br/lgpd/menu/tratamento-dos-dados/objetivo-e-abrangencia-da-lgpd. Último acesso em: 20 out. 2024.

VERIZON. **Data Breach Investigations Report (DBIR)**. Portal Verizon Business, 2022. Disponível em: https://www.verizon.com/business/resources/reports/dbir/2022/summary-of-findings/. Último acesso em: 20 out. 2024.

ZHOU Z. Machine learning**. Ed. Springer Nature - Nanjing University**. Jiangsu – China, 2021. Disponível em: https://books.google.com.br/books?hl=pt-BR&lr=&id=ctM-EAAAQBAJ&oi=fnd&pg=PR6&dq=machine+learning&ots=o_Kn_7XzZq&sig=KQBH0hyjOqL32OH7O8mn58Ct0J8#v=onepage&q=machine%20learning&f=false. Último acesso em: 20 out. 2024.