FACULDADE DE TECNOLOGIA DE SÃO PAULO – FATEC-SP
CURSO DE ANÁLISE E DESENVOLVIMENTO DE SISTEMAS – ADS

BEATRIZ SILVA PINHEIRO DE GODOY

TÉCNICAS E DESAFIOS NA APLICAÇÃO DA INTELIGÊNCIA ARTIFICIAL NA SEGURANÇA CIBERNÉTICA

São Paulo
2024

BEATRIZ SILVA PINHEIRO DE GODOY

TÉCNICAS E DESAFIOS NA APLICAÇÃO DA INTELIGÊNCIA ARTIFICIAL NA SEGURANÇA CIBERNÉTICA

Trabalho de Conclusão de Curso orientado pelo professor Carlos Hideo Arima, apresentado como exigência parcial na obtenção do título de Tecnólogo no Curso Superior de Análise e Desenvolvimento de Sistemas pelo Centro Paula Souza na FATEC São Paulo.

São Paulo
2024

Este trabalho é dedicado à memória da minha querida avó, Maria Carvalho de Godoy, que deixou esse mundo no Dia das Mães deste ano. Sua presença em nossas vidas foi um exemplo de amor, força, humildade, determinação e dedicação, e seu legado continuará a inspirar-me em todas as minhas jornadas. Que sua alma possa descansar em paz.

AGRADECIMENTOS

Em primeiro lugar, agradeço a Deus, que me dá forças para enfrentar os desafios da vida, sem Sua graça e amparo, não teria chegado até aqui.

À minha mãe, minha eterna conselheira, que, mesmo diante das dificuldades, nunca deixou de me apoiar e orientar. Seu amor e dedicação foram fundamentais para superar os momentos difíceis.

Ao meu namorado, que sempre está ao meu lado, celebrando minhas conquistas e me encorajando, onde apoio constante foi essencial para que eu pudesse perseverar e alcançar meus objetivos com este trabalho.

Às minhas irmãs, que dispuseram a me ajudar em minhas tarefas para que eu pudesse concluir esta pesquisa, oferecendo suporte e palavras de incentivo.

Agradeço também ao Professor Carlos Hideo Arima, pela orientação e dedicação durante a realização deste trabalho. Sua paciência e conhecimento foram essenciais para a conclusão desta etapa.

Um agradecimento especial aos colegas da empresa JCAdvisor, do ramo de consultoria em segurança da informação, que me acolheram na equipe e foram a favor da realização desta pesquisa.

Meu agradecimento a todos que, de alguma forma, contribuíram para a realização deste trabalho. Cada gesto de apoio, cada palavra de incentivo e cada ajuda oferecida foram importantes para que eu pudesse concluir esta jornada.

Muito obrigada a todos.

“Se você conhece o inimigo e conhece a si mesmo, não precisa temer o resultado de cem batalhas. Se você se conhece mas não conhece o inimigo, para cada vitória ganha, sofrerá também uma derrota. Se você não conhece nem o inimigo nem a si mesmo, perderá todas as batalhas.”

Sun Tzu

RESUMO

Esta pesquisa analisa as técnicas de inteligência artificial (IA) aplicadas na segurança cibernética, destacando os desafios inerentes a essa abordagem. O estudo começa citando a evolução contínua das ameaças cibernéticas e as limitações dos sistemas de segurança tradicionais baseados em regras. Em seguida, são levantadas algumas das aplicações da IA na segurança da informação, incluindo sua capacidade de processar grandes volumes de dados, para detectar anomalias e atividades incomuns, automatizar respostas a ameaças e fornecer insights em tempo real sobre eventos de segurança. A pesquisa destaca os desafios técnicos, operacionais, éticos e de privacidade que surgem com a aplicação da IA na segurança cibernética. Ao contextualizar a crescente demanda por soluções inovadoras em segurança da informação, o estudo destaca a IA como uma ferramenta promissora, porém complexa, de implementar, evidenciando seus desafios. Este trabalho enfatiza o potencial da IA para aprimorar os resultados de segurança da informação, enquanto também ressalta os desafios significativos e propõe direções para pesquisas futuras. Essas pesquisas visam enfrentar os obstáculos da IA e garantir seu uso transparente e ético no campo da segurança da informação.

Palavras-chave: Inteligência Artificial, Segurança Cibernética, Técnicas, Desafios, Ética.

ABSTRACT

This research analyzes artificial intelligence (AI) techniques applied in cybersecurity, highlighting the challenges inherent to this approach. The study begins by citing the continued evolution of cyber threats and the limitations of traditional rules-based security systems. Next, some of the applications of AI in information security are raised, including its ability to process large volumes of data, to detect anomalies and extraordinary activities, automate responses to threats, and provide real-time insights into security events. The research highlights the technical, operational, ethical and privacy challenges that arise with the application of AI in cybersecurity. By contextualizing the growing demand for innovative information security solutions, the study highlights AI as a promising, but complex, implementation tool, highlighting its challenges. This work emphasizes the potential of AI to improve information security outcomes, while also highlighting significant challenges and proposed proposals for future research. These researches aim to overcome AI obstacles and ensure its transparent and ethical use in the field of information security.

Keywords: Artificial Intelligence, Cybersecurity, Techniques, Challenges, Ethics.

INDICE DE FIGURAS

TABELAS

LISTA DE SIGLAS

NIST	Instituto Nacional de Padrões e Tecnologia dos Estados Unidos
GDPR	Regulamento Geral de Proteção de Dados
LGPD	Lei Geral de Proteção de Dados Pessoais
CID	Confidencialidade, Integridade e Disponibilidade
SVMs	Máquinas de Vetores de Suporte
FGSM	Fast Gradient Sign Method
ML	Aprendizado de Máquina (*Machine Learning)*
DL	Aprendizado Profundo (*Deep Learning)*
PLN	Processamento de Linguagem Natural (*Natural Language Processing*)
RNAs	Redes Neurais Artificiais
DNNs	Redes Neurais Profundas *(Deep Neural Networks)*
CNNs	Redes Neurais Convolucionais
NER	Reconhecimento de Entidades Mencionadas
FGSM	*Fast Gradient Sign Method*
URLs	*Uniform Resource Locator*

SUMÁRIO

1. INTRODUÇÃO

A evolução tecnológica e a crescente digitalização das atividades humanas têm trazido avanços significativos, mas também têm exposto sistemas e dados a um número crescente de ameaças cibernéticas. Nesse cenário, a Inteligência Artificial (IA) surge como uma aliada para reforçar a segurança cibernética, oferecendo soluções para a detecção e mitigação de ameaças. Contudo, a aplicação da IA nesse campo apresenta desafios complexos e multifacetados, que vão desde aspectos técnicos e operacionais até questões éticas e de privacidade.

Segurança da informação é a prática de proteger sistemas e redes de computadores contra acesso não autorizado, roubo, dano ou interrupção. À medida que as organizações se tornam mais dependentes da tecnologia para gerenciar suas operações, a proteção de dados confidenciais se torna cada vez mais importante. Na era moderna, as violações de dados podem causar danos significativos à reputação, à estabilidade financeira e à confiança do cliente de uma organização. Os cibercriminosos estão continuamente evoluindo suas técnicas e estratégias, tornando desafiador para os profissionais de segurança da informação acompanhar.

Estudos indicam que o volume de ataques cibernéticos tem crescido exponencialmente, demandando abordagens inovadoras para sua mitigação (Buczak; Guven, 2016). A pesquisa se justifica pela necessidade em entender e superar os desafios associados a essa implementação. A compreensão desses desafios é relevante para desenvolver estratégias que maximizem os benefícios da IA, minimizando riscos potenciais.

Segundo Song et al. (2018), a IA tem o potencial de contribuir significativamente com a segurança cibernética, fornecendo mecanismos mais sofisticados de detecção de anomalias e respostas a incidentes. A utilização da IA para aprimorar essa segurança representa uma fronteira promissora, mas também desafiadora.

O foco desta pesquisa é compreender e abordar as dificuldades inerentes à aplicação da IA na segurança cibernética. Isso inclui desafios técnicos, como a precisão e a robustez dos algoritmos de IA, a necessidade de grandes volumes de dados de alta qualidade e a capacidade de adaptação a novas ameaças. Também são abordados desafios operacionais, incluindo a integração da IA com sistemas de segurança existentes, a manutenção e atualização contínua dos modelos de IA e a capacitação de profissionais. Questões éticas são igualmente importantes, como a transparência dos algoritmos, o risco de viés e a responsabilidade em caso de falhas. Desafios de privacidade também são considerados, especialmente a proteção dos dados utilizados para treinar modelos de IA e a conformidade com regulamentos de privacidade, como a LGPD e GDPR.

1.1. Questão problema da pesquisa

Levanta-se, conforme o contexto apresentado, a seguinte questão problema: "Quais são as técnicas e os principais desafios enfrentados na aplicação da Inteligência Artificial na segurança cibernética, considerando os aspectos técnicos, operacionais, éticos e de privacidade?"

1.2. Objetivos

1.2.1. Geral

Para responder à questão problema da pesquisa, o objetivo deste trabalho é analisar as técnicas e desafios envolvidos na aplicação da Inteligência Artificial na segurança cibernética, proporcionando uma visão abrangente dos obstáculos na implementação.

1.2.2. Específicos

Para atingir esse objetivo geral, são estabelecidos alguns objetivos específicos. Primeiramente, a realização da análise bibliométrica aplicando métricas de impacto e tendências com o modelo PRISMA, elaborar uma abordagem sistemática para analisar os desafios técnicos, investigando as limitações tecnológicas, em seguida, explorar os desafios operacionais, identificando barreiras na implementação e manutenção de soluções de IA. Também se examinar os desafios éticos e de privacidade, avaliando as questões relacionadas ao uso da IA, incluindo transparência e responsabilidade, analisando as implicações para a privacidade dos dados e as medidas necessárias para assegurar a conformidade com as regulamentações. Por fim, comunicar os resultados obtidos.

1.3. Contribuições

No campo acadêmico, esta pesquisa oferece uma análise das técnicas de Inteligência Artificial (IA) aplicadas na segurança cibernética, contribuindo para o corpo de conhecimento existente sobre o tema. A revisão sistemática da literatura, combinada com a análise bibliométrica, permite identificar as abordagens mais eficazes e inovadoras, bem como as tendências emergentes na aplicação da IA para segurança cibernética. Isso serve como uma base para futuros estudos, proporcionando aos pesquisadores uma compreensão dos métodos que têm sido mais bem-sucedidos e das áreas que ainda requerem investigação.

No âmbito prático, a pesquisa proporciona insights valiosos para profissionais de segurança da informação, desenvolvedores de IA e gestores de tecnologia. A identificação e avaliação das técnicas de IA mais eficazes permitirão que os profissionais de segurança

cibernética escolham as ferramentas e métodos mais adequados para suas necessidades específicas, aumentando a eficiência na detecção e mitigação de ameaças.

A pesquisa tem também um impacto social ao contribuir para uma sociedade digital mais segura. Com a crescente dependência da tecnologia em todos os aspectos da vida moderna, desde serviços financeiros até cuidados de saúde, a segurança cibernética se torna fundamental para a proteção de informações pessoais e sensíveis. Ao melhorar a eficácia das medidas de segurança através da IA e ao abordar os desafios associados, este trabalho contribui para a criação de um ambiente digital mais confiável e resiliente, beneficiando tanto indivíduos quanto organizações.

1.4. Estrutura

A estrutura utilizada na dissertação segue um formato tradicional e lógico, composto pelas seguintes seções principais:

Este capítulo de introdução, apresenta o contexto geral da pesquisa, a relevância do tema, a questão problema, os objetivos da pesquisa e a justificativa, para situar sobre a importância do estudo e o que se pretende alcançar.

No capítulo 2, é apresentado o embasamento teórico necessário para a compreensão do tema. É dividida em sub-seções que abordam conceitos essenciais para a pesquisa como a Interseção entre IA e Segurança Cibernética bem como seus conceitos individuais e as Técnicas de IA na Cibersegurança.

A metodologia situada no capítulo 3, descreve os métodos e procedimentos adotados para conduzir a pesquisa. Isso inclui a abordagem de pesquisa, técnicas de coleta e análise de dados, bem como os critérios de seleção dos estudos revisados.

O capítulo 4 é dedicado à apresentação e discussão dos resultados obtidos na pesquisa. Está organizada também em sub-seções para uma análise detalhada dos diferentes tipos de desafios: técnicos, operacionais, éticos e de privacidade. Também resume os principais achados e discute suas implicações para o campo da segurança cibernética na conclusão da análise.

O capítulo 5 conclui a dissertação com as considerações finais, recapitulando os principais resultados, discutindo as contribuições do estudo e propondo recomendações para futuras pesquisas.

2. FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, apresenta-se uma visão geral da IA, incluindo seu histórico, áreas de aplicação e subdisciplinas relevantes. Em seguida, discute-se os princípios e práticas da Segurança da Informação, destacando a importância da proteção de dados e sistemas contra ameaças cibernéticas. Explora-se também a convergência entre IA e Segurança Cibernética, analisando como técnicas de IA podem auxiliar a defesa contra ataques, além de detalhar as diversas técnicas de IA empregadas na cibersegurança, como algoritmos de aprendizado de máquina, redes neurais, sistemas especialistas e análise preditiva.

2.1. Inteligência artificial

A Inteligência Artificial (IA) é um dos campos mais atuais em ciências e engenharia, fascinando um variado grupo de pessoas e empresas (Russel e Norvig 2013; Polson e Scott 2020). Após a Segunda Guerra Mundial, muitos pesquisadores, incluindo o matemático inglês Alan Turing, dedicaram-se ao desenvolvimento de máquinas inteligentes. Turing foi pioneiro ao sugerir que a IA deveria ser pesquisada em programas de computador ao invés de criar máquinas físicas (McCarthy 2007). Seu Teste de Turing, apresentado no artigo de 1950 "*Computing Machinery and Intelligence*", desempenhou um papel crucial tanto na ciência da computação quanto na IA (Bostrom 2018; Isaacson 2014; Russel e Norvig 2013; Tegmark 2019).

A evolução da IA foi impulsionada pela questão de saber se algum dia um computador poderia agir de maneira inteligente como um ser humano, gerando interesse entre cientistas, engenheiros e filósofos (Oliveira, 2019). Os fundamentos da IA são estruturados em várias disciplinas, incluindo Filosofia, Matemática, Economia, Neurociência, Psicologia, Engenharia de Computadores, Teoria de Controle & Cibernética, e Linguística. Cada uma dessas disciplinas contribuiu com ideias, perspectivas e técnicas essenciais para o desenvolvimento da IA (Russel e Norvig 2013).

Na Filosofia, Aristóteles foi o primeiro a elaborar um conjunto de leis que regulam a parte racional da mente, enquanto Ramon Lull propôs a ideia de que o raciocínio útil poderia ser gerido por um dispositivo mecânico. Thomas Hobbes sugeriu que o raciocínio era similar à computação numérica (Russel e Norvig 2013). A Matemática contribuiu com a Lógica, Computação e Probabilidade, enquanto a Economia forneceu a teoria da decisão, que combina a teoria da probabilidade com a teoria da utilidade (Russel e Norvig 2013). A Neurociência investiga como o cérebro processa informações, e a Psicologia busca entender como os seres humanos e os animais pensam e se comportam. A Engenharia de

Computadores, a Teoria de Controle e Cibernética, e a Linguística também desempenham papéis cruciais para essa tecnologia (Russel e Norvig 2013).

O termo "Inteligência Artificial" foi cunhado por John McCarthy em 1956, durante uma reunião em Dartmouth, onde cientistas interessados em redes neurais, teoria dos autômatos e estudo da inteligência se reuniram para explorar a possibilidade de criar máquinas que pudessem realizar qualquer tipo de atributo da inteligência (Bostrom 2018; Russel e Norvig 2013). No entanto, o excesso de confiança tecnológica dos fundadores da área não levou aos resultados esperados (Tegmark 2019).

Russel e Norvig (2013) definem a IA em duas dimensões inter-relacionadas: processos de pensamento e raciocínio (pensando como um humano e pensando racionalmente) e comportamento (agindo como seres humanos e agindo racionalmente). O objetivo da IA é usar computadores para realizar tarefas que atualmente os humanos fazem melhor, especialmente aprender, que é a mais importante dessas tarefas (Domingos 2017). A IA é a ciência e a engenharia de criar máquinas inteligentes, principalmente programas de computador inteligentes (McCarthy 2007).

A história da IA passou por ciclos de entusiasmo e desapontamento, conhecidos como "invernos da IA". O primeiro ocorreu na década de 1970, quando se percebeu que a IA não conseguiria atingir as expectativas iniciais, resultando em escassez de financiamentos e ceticismo. O segundo "inverno" ocorreu no final da década de 1980, quando os computadores de quinta geração do Japão não alcançaram os resultados esperados, levando à retirada dos investimentos na área (Bostrom 2018).

A IA já ultrapassou a inteligência humana em diversas áreas, como em jogos, mas ainda enfrenta desafios significativos, como realizar atividades que os humanos fazem sem pensar e entender a linguagem natural (Bostrom 2018; Tegmark 2019). Os ramos da IA incluem reconhecimento de padrões, aprendizado com experiência e ontologia, aplicando-se em áreas como jogos, reconhecimento de fala e compreensão da linguagem natural (McCarthy 2007).

Uma das subáreas mais proeminentes da IA é o *machine learning*, que visa criar algoritmos que permitem que os computadores aprendam e melhorem seu desempenho com a experiência (Domingos 2017). O *machine learning* é crucial para atividades que requerem inteligência e melhora os resultados de pesquisas no Google e recomendações em plataformas como Netflix (Oliveira, 2019).

A colaboração entre humanos e máquinas é vista como essencial para o futuro da IA. Embora os algoritmos possam realizar muitas tarefas detalhadas, a intervenção humana é frequentemente necessária na tomada de decisões finais (Domingos 2017). Trabalhando

juntos, humanos e máquinas podem potencializar suas habilidades, criando uma sinergia onde cada um faz o seu melhor (Isaacson 2014; Domingos 2017).

Em suma, a IA é uma ciência e engenharia que busca criar máquinas inteligentes capazes de realizar tarefas humanas. Com uma história marcada por altos e baixos, a IA continua a evoluir, beneficiando-se de contribuições de várias disciplinas e da colaboração entre humanos e máquinas. Como destaca Domingos (2017), "Não é homem contra máquina; é homem com máquina versus homem sem ela".

2.2. Segurança da Informação

A segurança da informação é uma prática de grande importância para proteger os ativos digitais contra ameaças e ataques. De acordo com o Instituto Nacional de Padrões e Tecnologia dos Estados Unidos (NIST), a segurança cibernética envolve a prevenção de danos, proteção e restauração de computadores, sistemas de comunicações eletrônicas e as informações neles contidas, para garantir sua disponibilidade, integridade, autenticação, confidencialidade e não repúdio. É um conjunto de práticas, tecnologias e políticas usadas para proteger ativos digitais contra acesso, uso, divulgação, interrupção, modificação ou destruição não autorizados (NIST, 2017).

A origem do termo "segurança cibernética" remonta aos primórdios da computação, surgindo da necessidade de proteger sistemas de informação e redes de computadores contra ameaças e ataques virtuais. Trabalhos acadêmicos indicam que o termo começou a ser utilizado no final da década de 1980 (DENNING, 1999; ANDREASSON, 2011; BORAH, 2015). Anderson (2001) menciona que especialistas em segurança da informação do governo dos Estados Unidos utilizaram o termo na década de 1980, sendo que a primeira referência oficial apareceu em um relatório da Comissão Presidencial de Segurança Cibernética em 1997. Denning (1999) destaca que o termo foi inicialmente empregado por pesquisadores da computação para descrever a proteção de sistemas e redes contra ameaças virtuais, surgindo como uma extensão da segurança da informação.

Com o avanço da tecnologia e a crescente dependência da sociedade em relação à infraestrutura digital, a segurança cibernética evoluiu e se expandiu, abrangendo não apenas a segurança de computadores, mas também segurança de rede, segurança da informação e outros campos relacionados. Hoje, é uma área crítica de pesquisa e prática, com aplicações em finanças, saúde e governo (VIGANÒ; LOI; YAGHMAEI, 2020). A segurança cibernética evoluiu como ciência, inicialmente focada em aspectos técnicos como criptografia, firewalls e detecção de invasões. Com o tempo, o escopo se ampliou para

incluir aspectos organizacionais e sociais, como gerenciamento de riscos, governança e comportamento humano (SURYOTRISONGKO; MUSASHI, 2019).

Um dos primeiros marcos na legislação de segurança cibernética foi a Convenção de Budapeste, adotada pelo Conselho da Europa em 2001, que estabeleceu normas internacionais para combater o cibercrime, definindo crimes como invasão de sistemas, fraude e terrorismo, além de medidas para investigação e cooperação internacional (SOUZA; PEREIRA, 2009). Outro marco significativo é o Regulamento Geral de Proteção de Dados (GDPR) da União Europeia, em vigor desde 2018, que impõe regras rigorosas para a proteção de dados pessoais e privacidade dos cidadãos europeus (ZERLANG, 2017). No Brasil, a Lei Geral de Proteção de Dados Pessoais (LGPD), em vigor desde 2020, também busca proteger os direitos de privacidade e dados pessoais dos cidadãos (CASTRO; SILVA; CANEDO, 2022).

A segurança da informação, caracterizada pela aplicação de dispositivos de proteção sobre ativos visando preservar seu valor para as organizações, baseia-se nos princípios de Confidencialidade, Integridade e Disponibilidade (CID). A confidencialidade garante que o acesso à informação seja restrito a usuários legítimos (BEAL, 2008), a integridade assegura que a informação seja mantida sem alterações indevidas (SÊMOLA, 2003), e a disponibilidade garante que a informação esteja acessível para os usuários legítimos de forma oportuna (BEAL, 2008).

O maior desafio na área de segurança da informação vai além de software ou hardware, sendo a vulnerabilidade humana a mais crítica. Treinamentos para funcionários, implementação de políticas de segurança e conscientização sobre a importância da proteção de informações são essenciais para mitigar riscos. Qualquer ato irresponsável pode causar grandes prejuízos e perda de reputação para a organização (STALLINGS, 2014). Assim, a segurança da informação não só envolve métodos técnicos, mas também uma forte componente educacional e organizacional, necessária para enfrentar as complexas ameaças atuais.

2.3. Interseção entre IA e Segurança Cibernética

A combinação de Inteligência Artificial (IA) e segurança cibernética está se tornando cada vez mais presente na automação de tarefas que protegem sistemas e dados contra ameaças digitais, uma vez que hackers também avançam em tentativas mais sofisticadas de invasão. Em 2023 no Brasil aconteceram mais de 365 tentativas de ataque por minuto em pequenas e médias empresas (Kaspersky, 2024). Embora a IA apresente algumas falhas,

como a geração de falsos positivos na automação de tarefas, sua aplicação na segurança cibernética tem mostrado ser uma ferramenta poderosa na prevenção de crimes cibernéticos (CHAN; SIMON; MIN et al., 2019).

Os sistemas de IA, quando aplicados corretamente, são capazes de monitorar redes, identificar padrões de atividade suspeitos e neutralizar ameaças ainda em estágios iniciais. Esse processo é comparável ao sistema imunológico do corpo humano, onde a IA pode detectar e reagir a ameaças de maneira eficaz, aprendendo e se fortalecendo com cada interação (PINTO, 2020). A Inteligência Artificial (IA), utiliza redes neurais e sistemas especialistas para localizar e consultar domínios específicos, o que permite uma análise mais precisa e uma resposta rápida a ameaças (CHAN; SIMON; MIN et al., 2019).

As redes neurais profundas (*Deep Neural Networks*, DNN) não são apenas usadas para proteger organizações, mas também para prever ataques. A IA em ferramentas ou sistemas pode ser dividida em dois tipos: baseada em casos de raciocínio e em regras. Os raciocínios baseados em casos permitem a resolução de problemas relembrando situações semelhantes anteriormente resolvidas pela máquina. Já os sistemas baseados em regras resolvem problemas com base em regras definidas por especialistas, onde cada regra possui uma condição e uma ação correspondente (CHAN; SIMON; MIN et al., 2019).

A inteligência artificial (IA) é amplamente utilizada em negócios e organizações. Os líderes dessas organizações buscam entender os riscos e usam a IA para combater vulnerabilidades. Há muitas aplicações de IA atualmente, como o reconhecimento facial, que é usado para melhorar a eficiência e a segurança dos softwares, e câmeras de segurança para proteger vários ambientes. Com os avanços tecnológicos, cada vez mais pessoas investem nessas técnicas (PINTO, 2020). O objetivo da IA é proteger o negócio e manter a segurança das aplicações (TIMOCHENCO, 2020).

A crescente adoção da IA em diversas aplicações, desde assistentes virtuais até carros autônomos, destaca sua importância. A IA melhora a eficiência e a capacidade de resposta dessas aplicações, mas também aumenta as preocupações com a segurança, especialmente quanto a vulnerabilidades e ameaças cibernéticas que podem comprometer dados e sistemas (Sezer et al., 2018; Sharma e Bali, 2020). Também enfrenta desafios como o envenenamento de dados, onde atacantes manipulam os dados de treinamento para produzir resultados incorretos, e ataques adversariais, onde dados de entrada são manipulados para enganar os sistemas de IA (Sharma e Bali, 2020). É fundamental implementar medidas de segurança robustas para proteger esses sistemas e garantir a privacidade dos usuários (Sharma e Bali, 2020).

Assistentes pessoais, como a Siri da Apple, são amplamente usadas em diversos dispositivos. Esses softwares podem identificar comportamentos pessoais e ajudar os usuários com informações graças ao aprendizado de máquina. Bancos também usam IA para realizar tarefas com mais eficiência e rapidez. Na segurança, a IA pode proteger redes contra ataques, que são cada vez mais frequentes no mundo digital. Por exemplo, ataques a sistemas de Internet Banking podem ser identificados e neutralizados com a ajuda da IA, dificultando a ação de hackers.

Empresas como a IBM estão na vanguarda da integração da IA em suas operações, priorizando a segurança através de criptografia e parcerias que ajudam a prevenir ações não autorizadas (IBM, 2017). A automação é uma parte essencial do planejamento da IBM, permitindo maior eficiência e segurança nos negócios (TIMOCHENCO, 2020). Aplicações como chatbots, assistentes pessoais e sistemas de gestão beneficiam-se da IA para melhorar a experiência do usuário e a eficiência operacional (Stefanini, 2020; TIMOCHENCO, 2020).

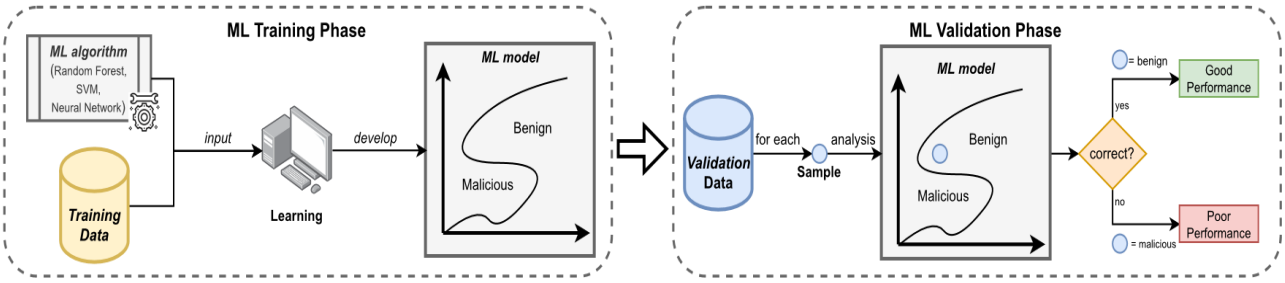
2.4. Técnicas de IA na Cibersegurança

A implementação da Inteligência Artificial (IA) na cibersegurança tem gerado diversas técnicas para prever, prevenir e responder a ameaças cibernéticas. Estudos recentes têm oferecido insights valiosos sobre esses métodos, destacando a eficácia da IA em diferentes áreas da cibersegurança.

2.4.1. Machine Learning (ML) e Deep Learning (DL)

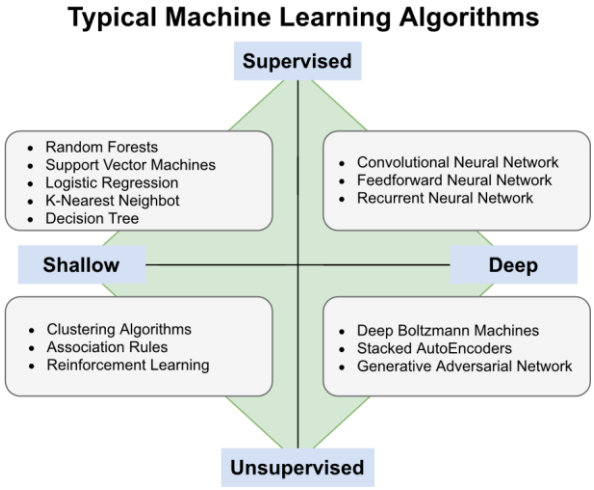
Machine Learning (ML), ou Aprendizado de Máquina, é um subcampo da Inteligência Artificial (IA) que se concentra no desenvolvimento de algoritmos que permitem aos computadores aprenderem a partir de dados e fazer previsões ou tomar decisões sem serem explicitamente programados para tal. A premissa básica do ML é que sistemas podem melhorar seu desempenho em uma tarefa específica através da experiência, ou seja, da exposição a grandes volumes de dados relevantes (BUCZAK, 2016). Na **Figura 1**, é exemplificado o Desenvolvimento de *Machine Learning*. Depois de coletar alguns dados de treinamento e analisar esses dados por meio de um algoritmo de ML, um modelo de ML é obtido. Esse modelo de ML deve ser testado por meio de alguns dados de validação. Se o desempenho de tal avaliação for apreciável, o modelo de ML poderá ser implantado na produção.

Figura 1: Desenvolvimento de Aprendizado de Máquina.



Fonte: Apruzzese et. al, 2023.

Figura 2: Algoritmos típicos de aprendizado de máquina.



Fonte: Apruzzese et. al, 2023.

Na **Figura 2**, há os algoritmos típicos de aprendizado de máquina. Um algoritmo pode ser "profundo" se depender de redes neurais, caso contrário, é "raso". Algoritmos que exigem dados rotulados são usados para tarefas "supervisionadas", caso contrário, eles podem ser usados também em tarefas "não supervisionadas". No aprendizado supervisionado, o modelo é treinado com um conjunto de dados etiquetados, onde a resposta correta é fornecida para cada exemplo. Técnicas comuns incluem regressão linear, máquinas de vetores de suporte (SVMs) e redes neurais. No aprendizado não supervisionado, o modelo deve encontrar padrões e relações nos dados sem referências externas, utilizando métodos como *clustering* e análise de componentes principais (PCA). O aprendizado por reforço envolve a aprendizagem por meio da interação com um ambiente e a obtenção de recompensas ou penalidades com base nas ações tomadas (Sutton & Barto, 2018).

O DL, a sigla para *Deep Learning*, significa “Aprendizado Profundo”. É uma subárea do ML que utiliza redes neurais artificiais com muitas camadas (daí o termo "profundo") para modelar padrões complexos em grandes volumes de dados. A principal diferença entre DL e outras técnicas de ML está na profundidade das redes utilizadas.

Enquanto modelos de ML tradicionais podem utilizar redes com uma ou duas camadas, os modelos de DL podem ter dezenas ou até centenas de camadas (LeCun, Bengio & Hinton, 2015).

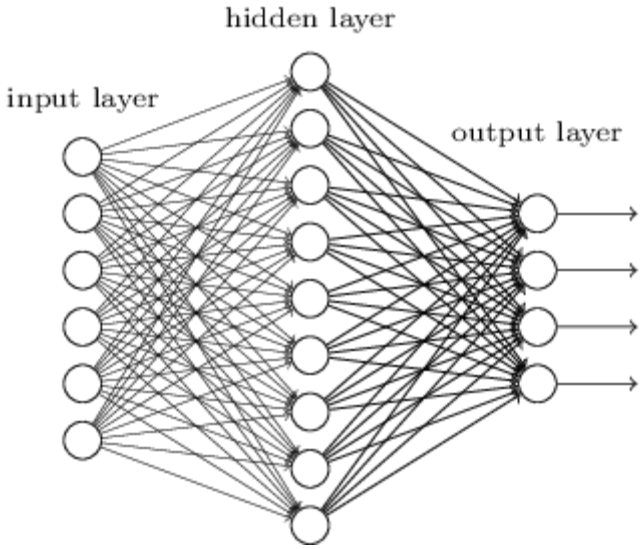
As redes neurais profundas, ou *Deep Neural Networks* (DNNs), são especialmente eficazes em tarefas como reconhecimento de imagens, processamento de linguagem natural e detecção de fraudes. Elas são capazes de aprender representações hierárquicas dos dados, onde níveis superiores da rede capturam características mais abstratas e complexas (Goodfellow, Bengio & Courville, 2016). Este poder de representação é uma das razões pelas quais o DL tem revolucionado muitos campos da IA, incluindo a segurança cibernética.

2.4.2. Redes Neurais Artificiais (RNA) e Redes Neurais Profundas (DNN)

As redes neurais, inspiradas na estrutura do cérebro humano, são um dos pilares fundamentais da Inteligência Artificial (IA) e têm se mostrado altamente eficazes em diversas aplicações, incluindo a segurança cibernética. Uma rede neural artificial (RNA) é composta por unidades de processamento interconectadas chamadas neurônios, que trabalham em conjunto para processar dados e aprender padrões a partir deles (Goodfellow, Bengio e Courville, 2016).

Essas redes funcionam através de camadas, como mostra a **Figura 3**: a camada de entrada recebe os dados brutos, uma ou mais camadas ocultas processam esses dados, e a camada de saída gera a resposta ou classificação. Cada neurônio em uma camada está conectado a todos os neurônios da camada subsequente, com cada conexão possuindo um peso que é ajustado durante o processo de treinamento para minimizar o erro na saída da rede (LeCun, Bengio e Hinton, 2015).

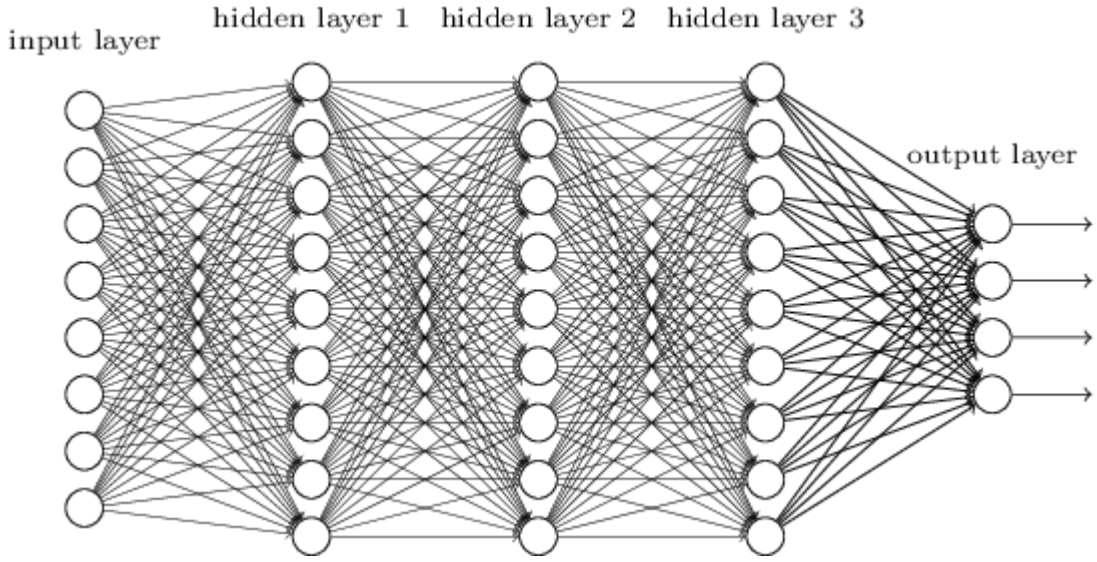
Figura 3: Estrutura de Redes Neurais Artificiais



Fonte: Deeplearningbook, 2024.

As redes neurais profundas (*Deep Neural Networks*, DNNs) são uma evolução das RNAs. Na **Figura 4**, é possível observar que possuem um número significativamente maior de camadas ocultas, o que lhes permite capturar e aprender representações de dados muito mais complexas. Essa profundidade adicional capacita as DNNs a resolver problemas que são intratáveis para redes neurais mais rasas, especialmente em áreas como reconhecimento de imagem, processamento de linguagem natural e, crucialmente, na segurança cibernética (Schmidhuber, 2015).

Figura 4: Estrutura de Redes Neurais Profundas



Fonte: Deeplearningbook, 2024.

2.4.3. Processamento de Linguagem Natural (PLN)

Processamento de Linguagem Natural (PLN), ou *Natural Language Processing* (NLP), é um subcampo da Inteligência Artificial (IA) que se concentra na interação entre computadores e linguagem humana. O objetivo principal do PLN é permitir que os computadores compreendam, interpretem e gerem linguagem humana de maneira significativa e útil. Este campo combina técnicas de linguística, ciência da computação e aprendizado de máquina para processar e analisar grandes quantidades de dados de linguagem natural (Manning & Schütze, 1999).

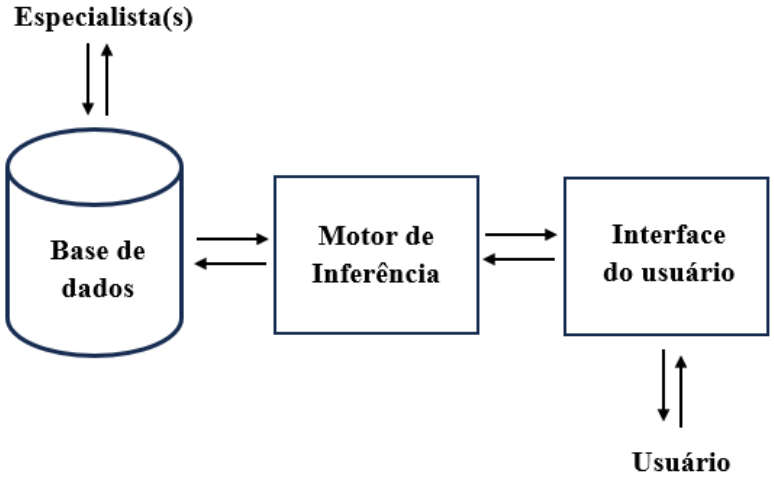
O PLN envolve várias tarefas, incluindo análise sintática, análise semântica, reconhecimento de entidades mencionadas (NER), tradução automática, análise de sentimentos, e geração de texto. Estas tarefas são fundamentais para a construção de sistemas que podem entender e responder a entradas em linguagem natural, como chatbots, assistentes virtuais e sistemas de tradução (Jurafsky & Martin, 2021).

2.4.4. Sistemas especialistas

Em inteligência artificial, um sistema especialista é geralmente um sistema de computador que emula a capacidade de tomada de decisão de um especialista humano. Um sistema especialista em segurança cibernética é uma instância de um sistema baseado em conhecimento ou em regras no qual as decisões podem ser tomadas com base em diretrizes de segurança (SARKER et. al, 2021).

Os sistemas especialistas são compostos por diversos componentes principais, como a base de conhecimento, o motor de inferência, a interface do usuário e, em alguns casos, módulos de aprendizagem e explicação. A **Figura 5** mostra a interação entre esses componentes. A base de conhecimento é um repositório de fatos e heurísticas que representam o conhecimento de especialistas humanos na área de interesse. Este conhecimento pode ser derivado de diversas fontes, incluindo entrevistas com especialistas, literatura especializada e registros históricos. O motor de inferência é responsável por processar a base de conhecimento e aplicar as regras para gerar conclusões ou recomendações. A interface do usuário permite a interação entre o sistema e os usuários finais, facilitando a inserção de dados e a visualização dos resultados.

Figura 5: Interação entre componentes de Sistemas Especialistas



Fonte: Elaborado pelo autor

2.4.5. Análise preditiva

Análise preditiva é uma técnica que utiliza métodos estatísticos, algoritmos de machine learning e inteligência artificial para analisar dados atuais e históricos, a fim de fazer previsões sobre eventos futuros. O objetivo principal é identificar padrões e relações nos dados que possam ser utilizados para prever tendências, comportamentos e resultados futuros com uma certa probabilidade de acerto (Kolias et al.,2017). O processo de análise preditiva começa com a coleta de dados, reunindo informações históricas relevantes e atuais de diversas fontes. Em seguida, passa-se pelo pré-processamento de dados, que envolve a limpeza e transformação dos dados para remover inconsistências e prepará-los para a análise. A etapa de modelagem é onde se selecionam e aplicam algoritmos estatísticos e de machine learning para construir modelos preditivos. Após a modelagem, vem a avaliação do modelo, que consiste em testar e validar os modelos preditivos utilizando técnicas como validação cruzada para garantir sua precisão e robustez. Uma vez validados, os modelos são implementados, aplicando-os em dados novos ou em tempo real para fazer previsões (Kim e Koo, 2021). É necessário monitorar continuamente o desempenho do modelo e ajustá-lo conforme necessário para manter a precisão das previsões (GARCIA, GUTIERREZ, DA SILVA, 2022).

3. METODOLOGIA DA PESQUISA

Para investigar as técnicas e desafios na aplicação da Inteligência Artificial (IA) na segurança cibernética, foi adotada a análise bibliométrica e de conteúdo como metodologias principais. A análise de conteúdo é uma técnica sistemática que permite extrair conclusões válidas a partir dos dados coletados de maneira objetiva, proporcionando uma visão abrangente sobre os aspectos significativos de estudos anteriores. Esta abordagem possibilita tanto ajustes qualitativos quanto quantitativos, assegurando que as conclusões deste estudo sejam aceitáveis, dada a abrangência das aplicações de IA nos domínios da segurança e da privacidade.

As amostras para esta investigação foram obtidas por meio de busca e seleção de artigos previamente revisados por pares, provenientes de fontes acadêmicas confiáveis. As bases de dados utilizadas incluíram Web of Science, Scopus, Science Direct, ASCE Library, IEEE, Wiley Online Library, Sage e Emerald. A ferramenta Google Acadêmico foi utilizada para pesquisa avançada e coleta de dados. Termos de busca como "inteligência artificial", "inteligência artificial em segurança da informação" e "desafios da segurança cibernética" foram empregados para identificar artigos relevantes que discutem os desafios e aplicações da IA na segurança cibernética. No total, 122 artigos potenciais foram identificados no período de 2016 a 2023.

A **Figura 6** apresenta uma nuvem de palavras gerada a partir das palavras-chave mais frequentes nos artigos analisados. Para gerar esta nuvem de palavras, a plataforma WordArt foi utilizada, onde foram inseridas as palavras previamente relacionadas ao objetivo da pesquisa. As palavras "inteligência artificial", "cibersegurança" e "desafios" destacam-se, refletindo a centralidade destes termos na literatura revisada. Esta visualização facilita a identificação dos temas mais abordados e das tendências de pesquisa no campo da IA aplicada à segurança cibernética.

Figura 6: Nuvem de palavras

Fonte: Resultado de pesquisa

Após a busca inicial, os artigos identificados foram submetidos a um processo de triagem para eliminar trabalhos irrelevantes. Este processo envolveu a aplicação de critérios de inclusão e exclusão para garantir uma seleção significativa e gerenciável de estudos pertinentes. Os critérios de inclusão foram:

- Artigos escritos em inglês ou português;
- Artigos completos de pesquisa (excluindo apresentações ou suplementos de pôsteres);
- Artigos que apresentam o uso da IA na cibersegurança ou discutem os desafios da IA na segurança da informação como parte central da metodologia;
- Artigos que respondem diretamente a uma ou mais das questões de pesquisa propostas;
- Para artigos publicados em múltiplos periódicos ou conferências, a versão mais recente foi considerada.

A seleção dos artigos seguiu um método de duas etapas com várias iterações. Na primeira etapa, títulos, resumos e palavras-chave das publicações foram examinados para verificar sua conformidade com os critérios estabelecidos. Na segunda etapa, uma leitura completa dos artigos foi realizada para confirmar sua pertinência ao objetivo geral da revisão. Este processo resultou na seleção de 76 artigos.

Os critérios de exclusão foram:

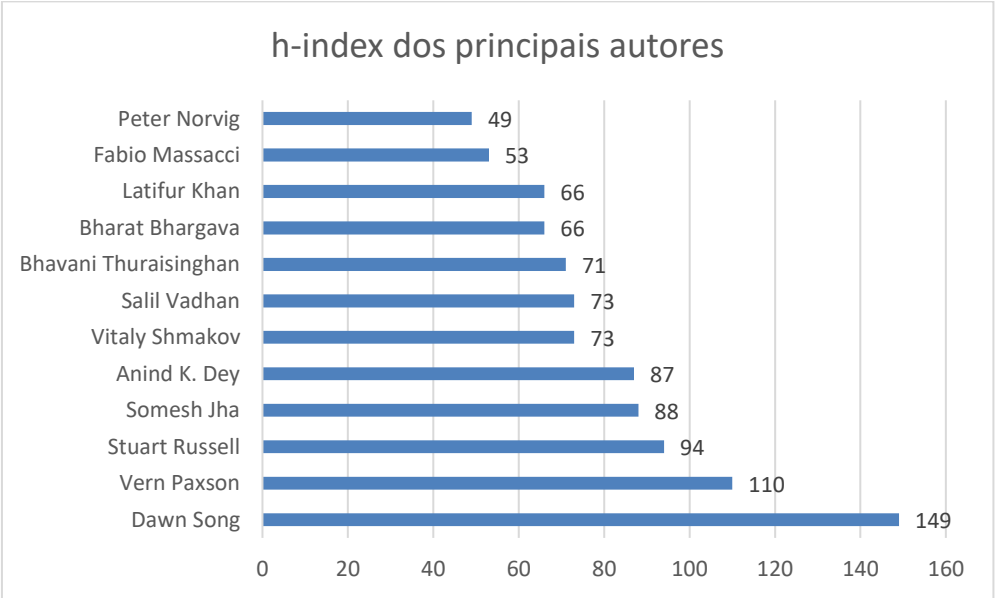
- Artigos em idiomas diferentes de inglês ou português;

- Artigos duplicados;
- Artigos bloqueados ou inacessíveis;
- Artigos incompletos e anteriores ao ano de 2016.

Após o processo de exclusão, a quantidade de artigos resultou na seleção de 38 estudos. A revisão empregou abordagens qualitativas e quantitativas para identificar as novas aplicações da IA em segurança e privacidade, os algoritmos de IA utilizados nessas aplicações e a análise da aplicabilidade desses algoritmos. Esta metodologia permitiu identificar as técnicas de IA mais promissoras em cibersegurança e os desafios frequentes na adoção e uso da IA na segurança da informação. A combinação destas abordagens garantiu uma compreensão detalhada e abrangente das potencialidades e limitações da IA no contexto da segurança cibernética.

Para avaliar a relevância e o impacto dos autores cujos alguns trabalhos foram incluídos na revisão, foi calculado usando o Google Acadêmico como ferramenta de busca, o h-index dos principais autores que publicam sobre assuntos relacionados ao problema de pesquisa, conforme mostrado na **Figura 7**. Este índice mede tanto a produtividade quanto o impacto das publicações de um autor, sendo uma métrica importante na identificação dos líderes de pesquisa no campo da IA e segurança cibernética.

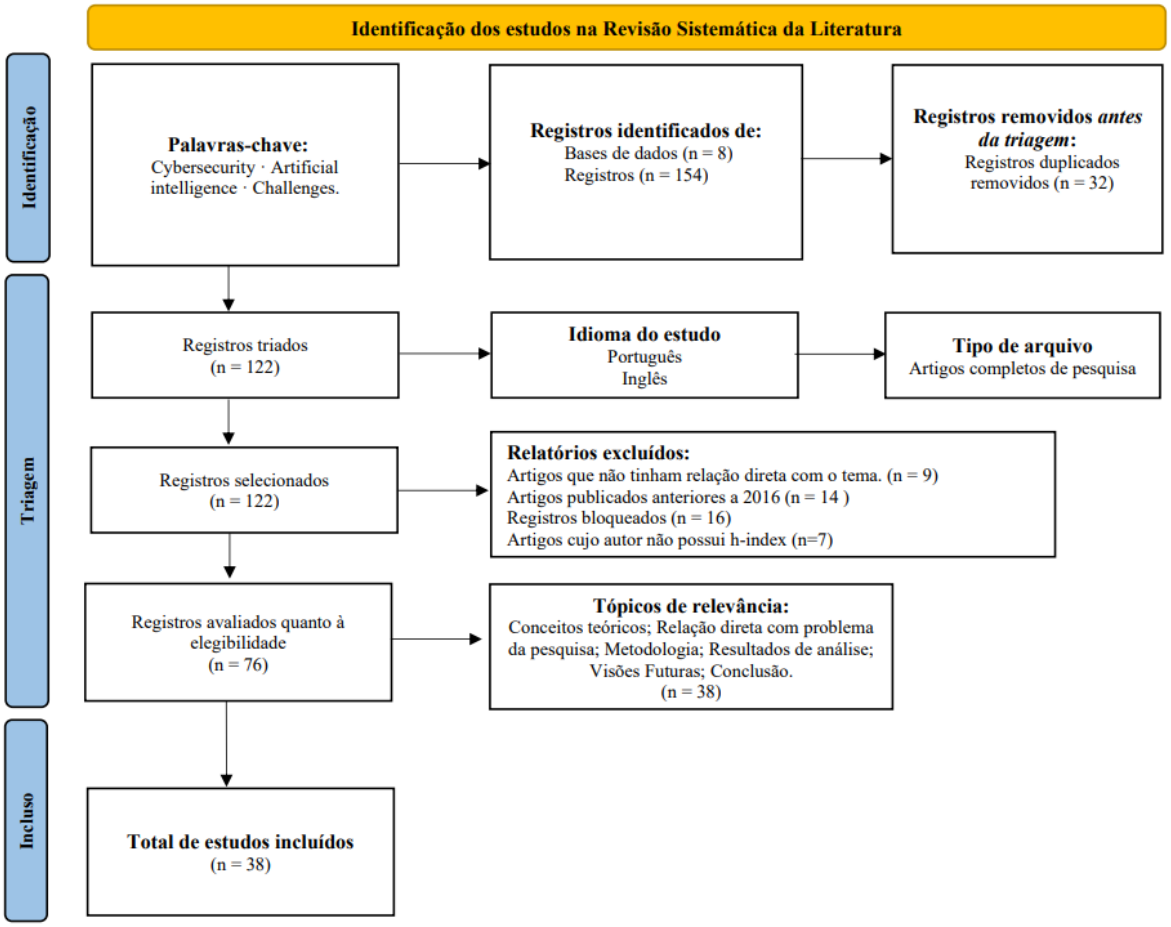
Figura 7: h-index dos principais autores



Fonte: Elaborado pelo autor.

O processo de seleção e triagem dos artigos seguiu o modelo PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), ilustrado na **Figura 8**. Esta figura demonstra as etapas de identificação, triagem, elegibilidade e inclusão dos artigos, evidenciando a transparência e a sistematicidade da revisão realizada.

Figura 8: Análise sistemática da literatura usando PRISMA



Fonte: Elaborado pelo autor.

A fim de analisar os resultados, o histograma ilustrado na **Figura 9** que mostra a quantidade de artigos selecionados a cada ano de 2016 até 2023 foi construído. Este gráfico facilita a visualização das tendências de publicação ao longo do tempo, evidenciando possíveis picos de interesse e áreas de foco na pesquisa sobre IA e segurança cibernética. O ano de 2024 não foi considerado, pois a pesquisa foi realizada no primeiro semestre de 2024 e não haveria resultados completos para o ano ainda.

Figura 9: Histograma da quantidade de artigos selecionados por ano (2016-2023)

Fonte: Elaborado pelo autor.

Para assegurar a replicabilidade deste estudo, é documentado a seguir, detalhadamente os processos e decisões tomadas durante a pesquisa:

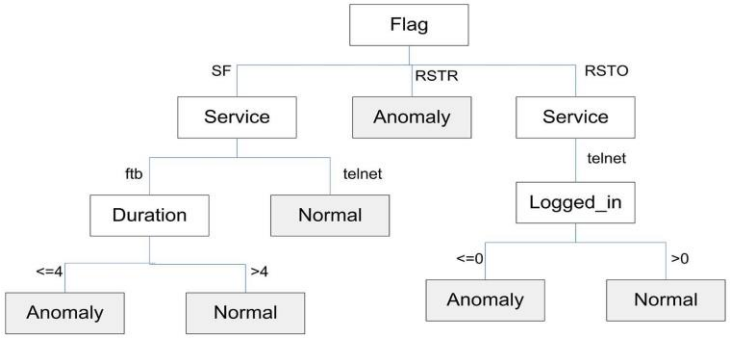
BUSCA DE ARTIGOS: Utilizou-se as bases de dados Web of Science, Scopus, Science Direct, ASCE Library, IEEE, Wiley Online Library, Sage, Emerald e Google Acadêmico. Termos específicos de busca foram aplicados para garantir a relevância dos artigos.
FERRAMENTAS DE ANÁLISE: A análise bibliométrica foi realizada utilizando ferramentas padrão de análise de citações e Google Acadêmico. A nuvem de palavras foi gerada na plataforma WordArt com base nas palavras-chave mais frequentes relacionadas a pesquisa.
CRITÉRIOS DE INCLUSÃO E EXCLUSÃO: Aplicação dos critérios estabelecidos para garantir a pertinência e qualidade dos artigos selecionados.
MÉTRICAS DE IMPACTO: Calculado o h-index dos principais autores para avaliar a relevância dos trabalhos incluídos.
TENDÊNCIAS: Histograma evidenciando anos com mais publicações e consequentemente possíveis picos de interesse e áreas de foco na pesquisa sobre IA e segurança cibernética.
MODELO PRISMA: Seguido o modelo PRISMA para documentar cada etapa do processo de revisão sistemática, desde a identificação até a seleção final dos artigos.

4. ANÁLISE DOS RESULTADOS

Baseando-se na revisão da literatura, foram identificados os principais desafios na aplicação de técnicas de Inteligência Artificial (IA) na segurança cibernética, questões como a complexidade e a variabilidade dos ataques cibernéticos, a necessidade de grandes volumes de dados de qualidade para treinar modelos eficazes, e as preocupações éticas e de privacidade inerentes ao uso de IA. Visões futuras sobre o desenvolvimento e a mitigação desses desafios, foram exploradas, incluindo avanços tecnológicos emergentes, e políticas regulatórias que possam suportar um ecossistema mais seguro e resiliente.

Buczak e Guven (2016) destacam que o ML é essencial para identificar padrões e prever comportamentos maliciosos. Algoritmos de aprendizado supervisionado são frequentemente usados para detectar intrusões na rede ao aprenderem a partir de dados históricos de ataques e comportamentos maliciosos conhecidos. Por exemplo, árvores de decisão, como mostra a **Figura 10**, funcionam de forma semelhante a um fluxograma, onde cada nó interno representa uma "pergunta" sobre um atributo, cada ramo representa o resultado da pergunta, e cada nó folha representa uma classe ou valor final. Juntamente com o uso de SVMs, podem ajudar a identificar padrões que indicam atividades anômalas em tráfego de rede (Buczak & Guven, 2016).

Figura 10: Detecção de anomalias cibernéticas usando ML baseado em árvore de decisão.



Fonte: Sarker et al, 2021.

O aprendizado não supervisionado é útil na detecção de anomalias, onde o sistema é treinado para identificar desvios de comportamento normal que podem indicar possíveis intrusões ou atividades suspeitas. Métodos de *clustering*, como *k-means*, podem ser aplicados para agrupar comportamentos de rede similares e destacar aqueles que são atípicos (BRAEI, 2020).

O DL em paralelo, também tem mostrado um potencial significativo na segurança cibernética, principalmente devido à sua capacidade de processar grandes volumes de dados e extrair características complexas automaticamente. São usadas para analisar logs de

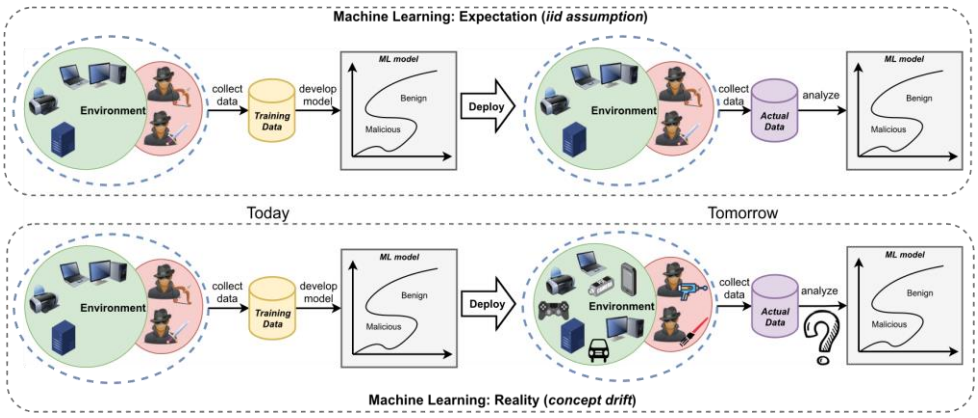
sistema, tráfego de rede e até mesmo comportamento de usuários para detectar atividades maliciosas que poderiam passar despercebidas por técnicas tradicionais (Kim, 2017).

Uma aplicação notável de DL é na detecção de malware. Modelos de DL, como redes neurais convolucionais (CNNs), podem ser treinados para reconhecer padrões associados a malware em arquivos executáveis e detectar novas variantes de malware que não foram previamente identificadas (Shone et al., 2018). Além disso, técnicas de DL são eficazes na prevenção de *phishing*, onde modelos são treinados para analisar e-mails e URLs em busca de características que indicam tentativas maliciosas (Abbasi et al., 2019).

Mas para modelos de aprendizado de máquina (ML) e *deep learning* (DL), há diversos desafios na segurança cibernética. Um dos principais obstáculos é a necessidade de grandes volumes de dados etiquetados para treinar os modelos de forma eficaz. A obtenção e etiquetagem desses dados pode ser um processo caro e demorado, além da dificuldade em *overfitting* e *underfitting*, ou seja, de fazer com que os modelos generalizem bem sem memorizar ou simplificar excessivamente os dados de treinamento (APRUZZESE, 2023).

A robustez dos modelos de ML e DL frente a ataques adversariais, é outro desafio significativo. Atacantes podem manipular dados de entrada de forma sutil para enganar os modelos e evitar a detecção. Isso é particularmente preocupante em cenários de segurança cibernética, onde a precisão na detecção de ameaças é crítica (Ren et al., 2020). Como mostra a **Figura 11**, o modelo de ML espera que os dados não se desviem dos vistos durante seu treinamento. Na cibersegurança, no entanto, o ambiente evolui e os adversários também se tornam mais poderosos (Apruzzese et. al, 2023).

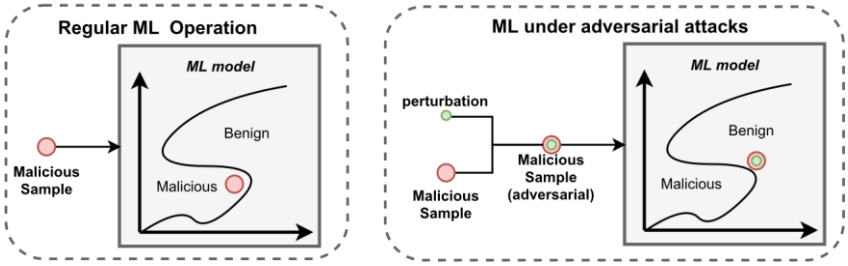
Figura 11: Aprendizado de máquina na presença de desvio de conceito.



Fonte: Apruzzese, 2023.

Na **Figura 12**, é exemplificado um ataque adversário típico contra um modelo de ML implantado. Ao inserir pequenas perturbações nos dados de entrada, é possível enganar um modelo de ML e induzir uma previsão incorreta (Apruzzese et. al, 2023).

Figura 12: Ataque adversário típico contra um modelo ML.

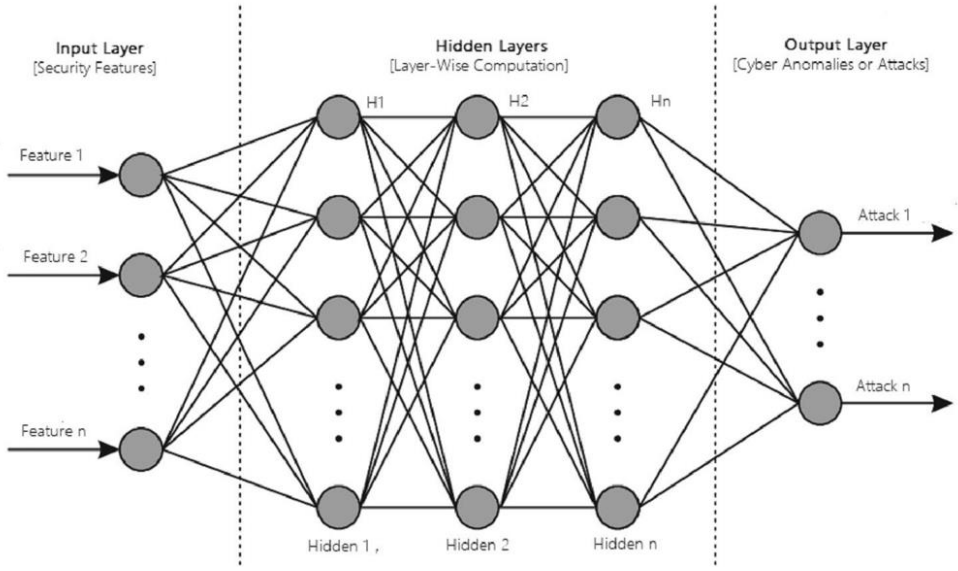


Fonte: Apruzzese et. al, 2023.

Lipton (2018) comenta que a complexidade dos modelos de DL pode dificultar a interpretação dos resultados, tornando mais difícil entender como o modelo chegou a uma determinada conclusão. Isso pode ser problemático para auditorias e conformidade regulatória, onde a explicabilidade é uma exigência.

Já com redes neurais artificiais (RNAs) e as redes neurais profundas (DNNs), são aplicadas na segurança da informação em várias frentes, como detecção de intrusões, análise de *malware* e proteção contra os ataques de *phishing*. As RNAs podem ser treinadas para identificar padrões de tráfego de rede normais e, consequentemente, detectar anomalias que possam indicar possíveis intrusões (Buczak & Guven, 2016). Por exemplo, como mostra a **Figura 13**, uma RNA treinada pode detectar um aumento incomum no tráfego de saída de um servidor, potencialmente indicando um ataque de exfiltração de dados, usando múltiplas camadas de processamento.

Figura 13: RNA para detectar anomalias com múltiplas camadas de processamento.



Fonte: Sarker et al, 2021.

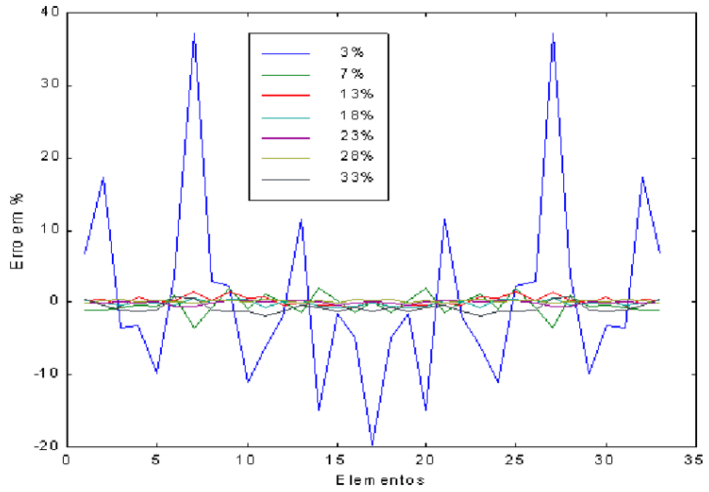
As DNNs, podem ser treinadas em vastos conjuntos de dados de amostras de *malware* e comportamento de programas para distinguir entre software benigno e malicioso com alta precisão (Shone et al., 2018). DNNs têm também mostrado eficácia na detecção de ataques de *phishing* ao analisar características sutis em e-mails que seriam difíceis de detectar por métodos tradicionais (Abbasi et al., 2019).

A aplicação, entretanto, têm demonstrado que as DNNs são vulneráveis a perturbações adversárias em trabalhos recentes (SONG et. al, 2018). Atacantes podem explorar as vulnerabilidades das redes neurais manipulando os dados de entrada de forma imperceptível para os humanos, que causam erros significativos nos modelos de IA (SONG et. al, 2018). Isso é particularmente problemático em ambientes de segurança cibernética, onde a precisão é crucial.

Um dos principais desafios, é a em relação ao alto consumo de recursos computacionais para treinamento e inferência e também a dificuldade em explicar e interpretar as decisões dos modelos complexos. Há a necessidade de grandes volumes de dados e a coleta e etiquetagem desses dados podem ser dispendiosas e demoradas, além de levantar preocupações de privacidade (Lipton, 2018). A **Figura 14**, mostra o resultado de erro pontual cometido na quantificação do treinamento de uma DNN sem métodos de regularização.

A interpretação dos resultados produzidos por DNNs pode ser difícil devido à sua complexidade. A "caixa preta" das DNNs torna desafiador entender exatamente como a rede chegou a uma determinada conclusão, o que é problemático para auditorias e conformidade regulatória (Lipton, 2018).

Figura 14: Erro pontual de uma RNA quando submetida ao conjunto de treinamento



Fonte: Researchgate, 2024.

A revisão de literatura revelou que o Processamento de Linguagem Natural (PLN) tem se mostrado uma ferramenta poderosa para detectar e mitigar ameaças. Uma das principais aplicações do PLN é a análise de grandes volumes de dados textuais provenientes de fontes como *logs* de sistema, e-mails, redes sociais e fóruns da *dark web*. Estes dados podem conter informações valiosas sobre potenciais ameaças, comportamentos anômalos e vulnerabilidades (Zhou et al., 2020).

Sahoo et al. (2017) destacam a aplicação da PNL na detecção de ações maliciosas em links, onde sistemas de IA são treinados para identificar URLs suspeitas, aumentando a eficácia na prevenção de ataques de engenharia social. Estudos demonstram que a análise de linguagem pode melhorar significativamente a precisão na detecção de *phishing* em comparação com métodos tradicionais baseados em heurísticas (Sahoo et al., 2017)

Uma outra aplicação crítica do PLN é na análise de ameaças internas. Os algoritmos de PLN podem monitorar e analisar comunicações internas para identificar sinais de comportamento malicioso ou negligente que possam indicar uma ameaça interna. Isso inclui a detecção de comunicações que mencionam explicitamente planos ou intenções maliciosas, bem como a análise de padrões linguísticos que podem sugerir insatisfação ou deslealdade entre os funcionários (Salem, Hossain & Kamhoua, 2018).

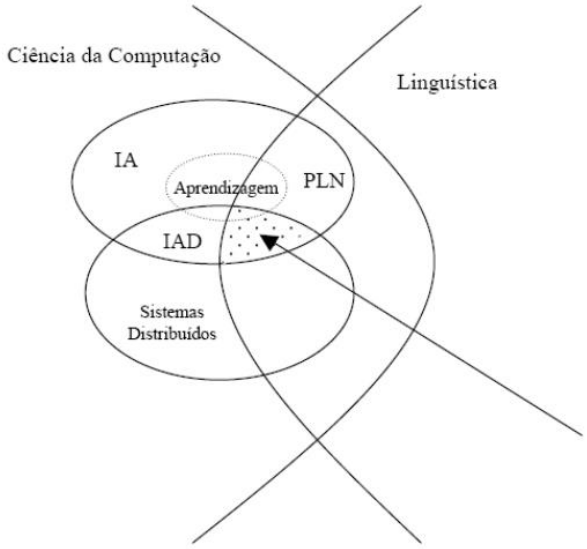
Embora o PLN ofereça muitas oportunidades na segurança cibernética, sua implementação eficaz enfrenta vários desafios. Como por exemplo a diversidade e complexidade da linguagem humana. A ambiguidade lexical, onde uma palavra pode ter múltiplos significados, e a ambiguidade sintática, onde uma sentença pode ser interpretada de diferentes maneiras, tornam a análise precisa de linguagem natural uma tarefa difícil (GEORGESCU, 2020).

Modelos de PLN, especialmente aqueles baseados em aprendizado profundo, como em outras técnicas de inteligência artificial que foram abordadas, requerem grandes quantidades de dados etiquetados para alcançar alta precisão para uma captura e interpretação adequada do contexto e da semântica em textos complexos. A coleta e etiquetagem desses dados podem ser caras e demoradas, além de levantar questões de privacidade e segurança (Devlin et al., 2019).

A adaptação a novos domínios e linguagens também é um desafio. Modelos de PLN treinados em um determinado conjunto de dados ou contexto podem não se generalizar bem para outros domínios ou linguagens e com múltiplos idiomas e dialetos. Isso é particularmente problemático em segurança cibernética, onde as ameaças evoluem rapidamente e novas formas de comunicação, como gírias ou jargões técnicos, são constantemente introduzidas (Ruder, Peters & Swayamdipta, 2019).

As ferramentas atuais desenvolvidas para o Processamento de Linguagem Natural buscam superar os fatores que prejudicam seu desempenho (GEORGESCU, 2020). Elas visam alcançar robustez (por exemplo, é essencial que um sistema de PLN consiga processar frases com erros ortográficos simples) e aprendizagem. Dessa forma, essas ferramentas são cada vez mais criadas utilizando o paralelismo. Como ilustrado na **Figura 15**, os sistemas de PLN mais avançados situam-se na interseção entre a Linguística e a Ciência da Computação, combinando ideias da Inteligência Artificial e dos Sistemas Distribuídos.

Figura 15: Sistemas atuais de PLN

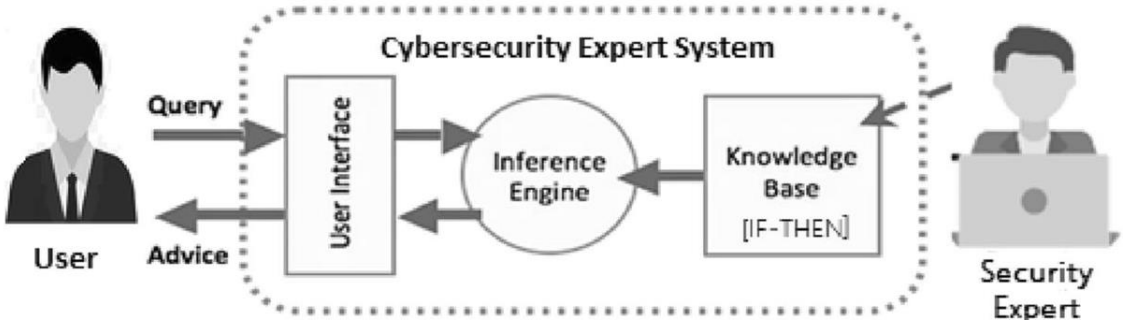


Fonte: Arruda e Carvalho, 2007.

Quanto aos sistemas especialistas, eles têm se mostrado extremamente valiosos na segurança digital. Sua aplicação abrange diversas áreas críticas, incluindo a detecção de intrusões, a análise de vulnerabilidades, e a resposta a incidentes. Estes sistemas são projetados para emular o conhecimento e a expertise de analistas humanos, oferecendo uma automação avançada e decisões rápidas baseadas em vastas quantidades de dados (Sarker et al, 2021).

A base dessa estrutura de especialistas em segurança cibernética é a base de conhecimento mostrada na **Figura 16**, pois consiste no conhecimento do domínio da aplicação de cibersegurança alvo, bem como conhecimento operacional das regras de decisões de segurança. O mecanismo de inferência mostrado, por outro lado, aplica as regras a fatos conhecidos do ponto de vista da segurança para deduzir fatos novos. A interface do usuário, reconhece os fatos de segurança originais e invoca o mecanismo de inferência para acionar as regras de decisão da base de conhecimento.

Figura 16: Estrutura de modelagem de um sistema especialista em cibersegurança.



Fonte: Sarker et al, 2021.

Apesar das vantagens, a implementação de sistemas especialistas em segurança cibernética tem a necessidade de manter os sistemas atualizados com as últimas informações sobre ameaças e técnicas de ataque, o que é um grande desafio. O cenário de ameaças cibernéticas está em constante evolução, e os sistemas especialistas precisam ser continuamente alimentados com novas regras e conhecimento para permanecerem eficazes, para diminuir as limitações na adaptação rápida a novas ameaças e técnicas de ataque emergentes. (Sarker et al, 2021).

Além de desafios na integração com outras tecnologias de segurança e sistemas de TI, há também a gestão de falsos positivos e falsos negativos. Um falso positivo ocorre quando o sistema identifica uma atividade legítima como uma ameaça, enquanto um falso negativo ocorre quando uma ameaça real não é detectada. Ambos os tipos de erro podem ter consequências graves, seja interrompendo operações normais ou deixando a rede vulnerável a ataques (Bhuyan et al., 2016). Em muitos casos, os dados analisados pelos sistemas especialistas podem ser incompletos ou pouco claros, o que pode comprometer a precisão das decisões tomadas pelo sistema (Tuptuk et al., 2018).

Para a técnica de análise preditiva, ela é aplicada para antecipar e mitigar ameaças antes que elas causem danos significativos. Mas a qualidade e quantidade de dados são cruciais para a eficácia dessa técnica. Dados incompletos, imprecisos ou enviesados podem comprometer a precisão dos modelos preditivos e a coleta de uma quantidade suficiente de dados relevantes pode ser um desafio devido à natureza heterogênea e distribuída dos sistemas de TI (Buczak & Guven, 2016).

Os modelos preditivos utilizados em segurança da informação têm desafios significativos na análise de dados em tempo real para detecção rápida de ameaças e problemas relacionados ao balanceamento de dados desbalanceados (muitas mais amostras de comportamento benigno do que malicioso). Esses modelos, tem em paralelo, obstáculos

com sinal-ruído, que é a dificuldade em distinguir sinais válidos de ruído em grandes volumes de dados. (Kim e Koo, 2021).

O cenário de ameaças cibernéticas está em constante evolução, com novos tipos de ataques e vulnerabilidades surgindo regularmente. Isso requer que os modelos preditivos sejam atualizados continuamente para manter sua eficácia. A adaptação rápida a novas ameaças é um desafio significativo, pois os modelos precisam ser treinados com novos dados e reavaliados regularmente (Kim et al., 2019).

O uso de dados pessoais e sensíveis em análise preditiva levanta questões de privacidade e conformidade regulatória. Garantir que os dados sejam utilizados de maneira ética e em conformidade com regulamentações como o GDPR é um desafio adicional (Veale et al., 2018).

4.1. Desafios técnicos gerais

A aplicação de IA na segurança cibernética enfrenta diversos desafios técnicos. Um dos principais é a necessidade de grandes volumes de dados para treinar modelos de *machine learning* (ML) e *deep learning* (DL), como foi explorado. Conforme apontado por Apruzzese et. al, (2023), a eficácia dos sistemas de IA depende diretamente da qualidade e quantidade dos dados disponíveis. A obtenção de dados relevantes e representativos pode ser difícil, especialmente em ambientes onde a privacidade dos usuários deve ser protegida. Além disso, a coleta de dados relevantes deve ser realizada de forma contínua para garantir que os modelos permaneçam atualizados diante de novas ameaças, o que aumenta a complexidade do gerenciamento de dados.

Há também outro desafio técnico significativo, que é a capacidade dos modelos de IA em detectar ameaças novas e desconhecidas. Muitos sistemas atuais ainda dependem de padrões pré-estabelecidos para identificar atividades maliciosas, o que limita sua eficácia contra ataques zero-day (Buczak & Guven, 2016). Esses modelos baseados em assinatura falham ao identificar comportamentos anômalos que não foram previamente catalogados. Além disso, a robustez dos modelos de IA contra ataques adversários, onde os atacantes manipulam dados de entrada para enganar os algoritmos, é uma área crítica que necessita de mais pesquisa e desenvolvimento (Ren et al., 2020). Métodos adversariais, como o FGSM (*Fast Gradient Sign Method*), têm mostrado que é possível gerar exemplos adversários que podem enganar facilmente sistemas de detecção baseados em IA (Goodfellow, Bengio e Courville, 2016).

4.2. Desafios Operacionais

Operacionalmente, a integração de sistemas de IA em infraestruturas de segurança cibernética existentes é complexa. Muitos sistemas legados não foram projetados para suportar a integração com tecnologias de IA, exigindo atualizações significativas ou substituições completas (Moustafa et al., 2019). Essa complexidade pode levar a interrupções significativas nas operações e a custos elevados de implementação. A implementação eficaz de soluções de IA requer profissionais altamente qualificados, o que pode ser um obstáculo significativo para muitas organizações devido à escassez de especialistas em IA e segurança cibernética (Joseph et al., 2019). A falta de conhecimento especializado pode resultar em uma subutilização das capacidades da IA ou em configurações inadequadas que não conseguem detectar ameaças de forma eficiente.

4.3. Desafios Éticos e de Privacidade

Os desafios éticos e de privacidade são igualmente significativos. A utilização de IA em segurança cibernética frequentemente requer o processamento de grandes quantidades de dados pessoais, o que levanta preocupações sobre a privacidade dos usuários (Crawford & Calo, 2016). A coleta massiva de dados pode levar a uma vigilância excessiva e à erosão da privacidade individual. Os algoritmos de IA podem inadvertidamente perpetuar ou exacerbar vieses existentes nos dados de treinamento, resultando em decisões injustas ou discriminatórias (Barocas & Selbst, 2016). Por exemplo, um sistema de detecção de intrusões treinado com dados enviesados pode detectar falsos positivos com maior frequência em determinados grupos de usuários.

As implicações de privacidade são particularmente preocupantes em contextos onde os dados coletados para fins de segurança podem ser usados de maneiras que os usuários não previram ou consentiram. Por exemplo, a análise de tráfego de rede para detectar intrusões pode revelar informações sensíveis sobre os hábitos e comportamentos dos usuários (Buczak & Guven, 2016). Isso pode levar a violações inadvertidas de privacidade, especialmente se os dados forem mal geridos ou compartilhados sem o devido controle.

4.4. Conclusão da Análise

Para mitigar esses desafios, pesquisas futuras devem focar no desenvolvimento de técnicas mais eficientes para a coleta e processamento de dados, bem como na criação de modelos de IA que sejam robustos contra ataques adversários. Por exemplo, técnicas de aprendizado federado, que permitem o treinamento de modelos de IA sem a necessidade de centralizar os dados dos usuários, podem oferecer uma solução promissora para os problemas de privacidade (Kairouz et al., 2019). É crucial desenvolver *frameworks* éticos e regulatórios que garantam a privacidade dos usuários e minimizem os vieses algorítmicos (Crawford & Calo, 2016).

A integração de IA em sistemas de segurança cibernética pode ser facilitada por meio do desenvolvimento de padrões de interoperabilidade e melhores práticas para a implementação e operação dessas tecnologias. A adoção de normas e protocolos padronizados pode ajudar a harmonizar a integração de IA em diferentes sistemas e infraestruturas. Investimentos em educação e treinamento também são essenciais para preparar a próxima geração de profissionais capacitados a lidar com esses desafios (Joseph et al., 2019). Programas de formação especializados em IA aplicada à segurança cibernética podem ajudar a mitigar a escassez de habilidades e garantir que as tecnologias sejam utilizadas de forma eficaz e segura.

A análise dos desafios técnicos, éticos e operacionais revela que, embora a IA ofereça significativas vantagens para a segurança cibernética, sua implementação eficaz requer abordagens multifacetadas que considerem tanto os aspectos tecnológicos quanto os sociais. Estudos futuros devem continuar a explorar essas áreas, visando desenvolver soluções que sejam ao mesmo tempo eficientes e responsáveis.

A **Tabela 1** resume os desafios da aplicação das técnicas de Inteligência Artificial (IA) em cibersegurança conforme identificado em diversos estudos revisados. Estes desafios refletem as complexidades e limitações enfrentadas ao implementar diferentes abordagens de IA para proteger sistemas contra ameaças cibernéticas.

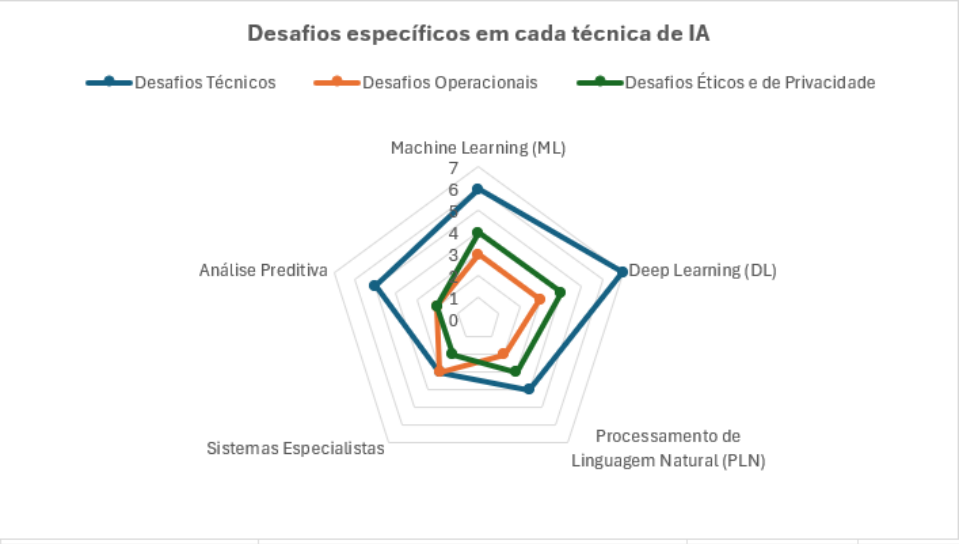
Tabela 1: Desafios associados as técnicas de IA abordadas

Autor	Técnica	Principais Desafios
Apruzzese et. al, 2023.	Machine Learning e Deep Learning (ML e DL)	Grandes volumes de dados para treinamento; Overfitting e Underfitting; Falta de robustez dos modelos frente a ataques adversariais;
Song et. al, 2018	Redes Neurais (RNAs e DNNs)	Vulnerabilidade a perturbações adversárias; Alto consumo de recursos computacionais para treinamento e inferência; Complexidade dos modelos pode levar a problemas de "caixa preta"; Limitações na eficácia contra ataques zero-day por falta de padrões pré-estabelecidos.
Georgescu, 2020	Processamento de Linguagem Natural (PLN)	Grandes volumes de dados para treinamento; Adaptação a novos domínios e linguagens. Ambiguidade linguística; Multilinguismo; Contexto e semântica.
Sarker et al, 2021	Sistemas Especialistas	Gestão de falsos positivos e falsos negativos; Dificuldade em manter os sistemas atualizados com as últimas informações sobre ameaças e técnicas de ataque; Manutenção de regras; Adaptabilidade; Integração com outras tecnologias.
Buczak & Guven, 2016	Análise preditiva	Escassez de qualidade alta e precisa para dados de treinamento; Exigência que sejam reavaliados dados de treinamento regularmente; Dificuldade em garantir a ética e conformidade com regulamentações; Análise de dados em tempo real; Balanceamento de dados; Sinal-ruído

Fonte: Elaborado pelo Autor.

Foi desenvolvido o gráfico de radar ilustrado na **Figura 17** com o objetivo de comparar qual tipo de desafio (técnicos, operacionais, éticos e de privacidade) é mais presente em cada uma das técnicas de Inteligência Artificial exploradas durante este artigo. Evidenciou-se então, através da análise desses dados, que os desafios de maior número, são os técnicos, seguido dos desafios éticos e de privacidade e por fim os operacionais.

Figura 17: Gráfico de radar com a análise dos tipos de desafios mais identificados.



Fonte: Elaborado pelo Autor.

5. CONSIDERAÇÕES FINAIS

O presente trabalho analisou as técnicas e desafios na aplicação da Inteligência Artificial (IA) na segurança cibernética, abrangendo aspectos técnicos, operacionais, éticos e de privacidade. Através de uma revisão da literatura, buscou-se fornecer uma visão dos obstáculos enfrentados.

O objetivo geral da pesquisa foi alcançado ao identificar e analisar os desafios técnicos, como a precisão e robustez dos algoritmos de IA, a necessidade de grandes volumes de dados de alta qualidade e a capacidade de adaptação a novas ameaças. As dificuldades operacionais foram também investigadas, incluindo a integração da IA com sistemas de segurança existentes, a manutenção e atualização contínua dos modelos de IA e a capacitação de profissionais. Questões éticas, como a transparência dos algoritmos, o risco de viés e a responsabilidade em caso de falhas, foram avaliadas, assim como os desafios de privacidade relacionados à proteção dos dados utilizados para treinar modelos de IA e a conformidade com regulamentos de privacidade.

Como proposta de pesquisa futura, sugere-se um aprofundamento nas técnicas de aprendizado contínuo e adaptativo para IA em segurança cibernética, permitindo que os modelos se ajustem de maneira mais eficiente às novas ameaças. A investigação de métodos para garantir a transparência e a ética dos algoritmos de IA, assim como o desenvolvimento de estratégias para assegurar a privacidade dos dados, são áreas promissoras para futuras pesquisas.

A continuidade desta pesquisa pode ser derivada da análise dos resultados obtidos, com foco em estudos de caso específicos que demonstrem a aplicação prática das soluções propostas. Este artigo visa contribuir de alguma forma para uma melhor compreensão das complexidades envolvidas na aplicação da IA na segurança cibernética. Espera-se que as conclusões aqui apresentadas possam servir de base para futuras investigações e desenvolvimentos nesta área crucial.

REFERÊNCIAS

ABBASI, S. A.; SALEEM, Y.; KIM, D. H. Enhanced Phishing Email Detection Using Deep Learning and NLP. In: Proceedings of the International Conference on Information and Communication Technology Convergence (ICTC), 2019. p. 114-119.

ANDERSON, Ross. E. Why information security is hard - an economic perspective. ***Research Advances in Database and Information Systems Security***, 2001. p. 1-16.

ANDREASSON, Kim J. Cybersecurity: Public sector threats and responses. **Taylor & Francis**, 2011.

Apruzzese, G.; Laskov, P.; Montes de Oca, E.; Mallouli, W.; Rapa, L. B.; Grammatopoulos, A. V.; Di Franco, F. The Role of Machine Learning in Cybersecurity. *Digital Threats: Research and Practice*, v. 4, n. 1, Artigo 8, mar. 2023, 38 p.

ARRUDA, Egio; CARVALHO, Cedric. Processamento de Linguagens Naturais e o Arcabouço GATE. Inf.ufg 2007.

BAROCAS, S.; SELBST, A. D. Big Data's Disparate Impact. California Law Review, v. 104, n. 3, p. 671-732, 2016.

BEAL, Adriana. Segurança da informação: Princípios e melhores práticas para a proteção dos ativos de informação nas organizações. São Paulo: Atlas, 2008.

BORAH, Chandra Kamal. Cyber war: the next threat to national security and what to do about it? by Richard A. Clarke and Robert K. Knake. 2015.

BOSTROM, Nick. 2018. **Superinteligência.** Rio de Janeiro: DarkSide Books.

BUCZAK, A. L.; GUVEN, E. A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection. IEEE Communications Surveys & Tutorials, v. 18, n. 2, p. 1153-1176, 2016.

Bhuyan, M. H., Bhattacharyya, D. K., & Kalita, J. K. (2016). "Network anomaly detection: Methods, systems and tools." *IEEE Communications Surveys & Tutorials, 16*(1), 303-336.

CHAN; SIMON; MIN; at al. Survey of AI in Cybersecurity for Information Technology ***Management*** (TEMSCON). Europa: IEEE Technology & Engineering Management Conference, 2019.

CRAWFORD, K.; CALO, R. There is a Blind Spot in AI Research. Nature, v. 538, n. 7625, p. 311-313, 2016.

CASTRO, Evandro Thalles Vale de; SILVA, Geovana RS; CANEDO, Edna Dias. Ensuring privacy in the application of the brazilian general data protection law (lgpd). In: ***Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing***. 2022. p. 1228-1235.

DENNING, Dorothy Elizabeth Robling. ***Information warfare and security***. AddisonWesley Professional, 1999.

DEVLIN, J.; CHANG, M.-W.; LEE, K.; TOUTANOVA, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Proceedings of the 2019 Conference of

the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), 2019. p. 4171-4186.

GARCIA, GUTIERREZ, DA SILVA, Caio Cruz Alfonso, Carolina de Carvalho, Nathan Brito. Inteligência Artificial Aplicada a Reconhecimento de detecção de Ataque Cibernético. Escola de Engenharia Mauá do Centro Universitário do Instituto Mauá de Tecnologia, São Caetano do Sul, p. 15-30, 2 jan. 2022.

GEORGESCU, Tiberiu Marian. Natural Language Processing Model for Automatic Analysis of Cybersecurity-Related Documents. *Symmetry*, v. 12, n. 3, artigo 354, 2020. Disponível em: https://doi.org/10.3390/sym12030354.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Deep Learning. Cambridge: MIT Press, 2016.

ISAACSON, Walter. 2014. "**Os inovadores"**. São Paulo: Companhia das Letras.

JOSEPH, A. D.; NELSON, B.; RUBINSTEIN, B. I.; TYGAR, J. D. Adversarial Machine Learning. Cambridge: Cambridge University Press, 2019.

JURAFSKY, D.; MARTIN, J. H. Speech and Language Processing. 3rd ed. Upper Saddle River: Pearson, 2021.

KAIROUZ, P.; McMAHAN, H. B.; AVELAR, P.; et al. Advances and Open Problems in Federated Learning. Foundations and Trends® in Machine Learning, v. 14, n. 1/2, p. 1-210, 2021.

KASPERSKY, Portal. **Os 10 hackers mais famosos de todos os tempos.** Disponível em: https://www.kaspersky.com.br/resource-center/threats/top-ten-greatest-hackers. Acesso em: 20/04/2024.

KIM, D. Deep Learning-Based Approach to Detect Zero-day Malware. In: Proceedings of the IEEE Conference on Computer and Communications Workshops (INFOCOM WKSHPS), 2017. p. 1-6.

KIM, Sung-Hyun; KOO, Jung-Hoon. ***Enhancing Cybersecurity with Artificial Intelligence: Current Techniques and Future Challenges.*** Journal of Cybersecurity, v. 7, n. 1, p. 45-60, 2021. DOI: 10.1093/cybsec/tyab007.

Kim, J., Kim, J., & Kim, H. (2019). "A novel hybrid intrusion detection method integrating anomaly detection with misuse detection." *Expert Systems with Applications, 41*(4), 1690-1700.

KOLIAS, Constantinos; KOUTSOULIDIS, Andreas; GIANNAKOPOULOS, Stamatios; MAGLARAS, Leandros; KAPRAVELI, Vivian; VAMVAKAS, Grigorios. **Automatic Detection of Online Recruitment Frauds: Characteristics, Methods, and a Public Dataset**. *Future Internet*, v. 9, n. 3, p. 1-13, 2017. DOI: 10.3390/fi9030046. 2017.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep Learning. Nature, v. 521, n. 7553, p. 436-444, 2015.

LIPTON, Z. C. The Mythos of Model Interpretability. Communications of the ACM, v. 61, n. 10, p. 36-43, 2018.

MANNING, C. D.; SCHÜTZE, H. Foundations of Statistical Natural Language Processing. Cambridge: MIT Press, 1999.

McCarthy, John. 2007. ***"What is artificial intelligence?".*** *John McCarthy's Home Page*, 12 de novembro de 2007. http://www-formal.stanford.edu/jmc/whatisai.pdf.

MOUSTAFA, N.; TURNBULL, B.; CHOO, K. K. R. An Ensemble Intrusion Detection Technique Based on Proposed Statistical Flow Features for Protecting Network Traffic of Internet of Things. IEEE Internet of Things Journal, v. 6, n. 3, p. 4815-4830, 2019.

NIST - National Institute of Standards And Technology. Security and Privacy Controls for Information Systems and Organizations. **Draft NIST Special Publication 800–53 Revision 5,** 2017.

OLIVEIRA, Arlindo. 2019. "**Inteligência Artificial"**. Lisboa: Fundação Francisco Manuel dos Santos.

Oliveira, Marlene e Zayr Claudio Gomes da Silva. 2020. **"Caminhos da ciência da informação: da library and information science às i-Schools."** Perspectivas em Ciência da Informação, 25: 8–27. doi:10.1590/1981-5344/4297.

PINTO, H. A. A utilização da inteligência artificial no processo de tomada de decisões. Brasília: Revista da informação legislativa, 2020.

POLSON, Nick e James SCOTT. 2020. "**Inteligência Artificial"**. Amadora: Vogais.

Portal Deeplearningbook. Redes Neurais. Disponível em: https://www.deeplearningbook.com.br/. Acesso em: 22/05/2024.

Portal Researchgate. Erro pontual em uma RNN. Disponível em: https://www.researchgate.net/. Acesso em: 22/05/2024.

REN, Kui; ZHENG, Tianwei; QU, Zhan; LIU, Xue. Adversarial attacks and defenses in deep learning. *Engineering*, v. 6, n. 3, p. 346-360, 2020.

RUDER, S.; PETERS, M. E.; SWAYAMDIPTA, S. Transfer Learning in Natural Language Processing. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2019. p. 27-38.

RUSSEL, Stuart e Peter NORVIG. 2013**. "Parte I: Inteligência Artificial".** Em Inteligência Artificial, 22-60. Rio de Janeiro: Elsevier.

SAHOO, Doyen; LU, Chenghao; HOI, Steven C. H.; ZHANG, Peilin; SUN, Yuting. Online Deep Learning: Learning Deep Neural Networks on the Fly. 2017. Disponível em: https://doi.org/10.48550/arXiv.1711.03705. DOI: 10.48550/arXiv.1711.03705. Acesso em: 23 jun. 2024.

SALEM, M. B.; HOSSAIN, M. E.; KAMHOUA, C. A. Analyzing Insider Threats Using Language Models. In: IEEE Symposium on Privacy-Aware Computing (PAC), 2018. p. 1-8.

SARKER, I. H.; Furhad, M. H.; Nowrozy, R. AI-Driven Cybersecurity: An Overview, Security Intelligence Modeling and Research Directions. Springer Nature Singapore Pte Ltd 2021.

SÊMOLA, Marcos**. Gestão da Segurança da Informação: uma visão executiva**. Rio de Janeiro: Elsevier Campus, 2003.

SHARMA, Sandeep; BALI, Ramesh. Artificial Intelligence for Cybersecurity: Threats, Attacks and Mitigation. New York: Springer, 2020.

Sharma, B. K., & Chen, K. (2019). Emerging Cyber Threats and Security Measures in the Era of New Technologies: Big Data, Cloud Computing, Internet of Things, and Social Media. In r1.

SCHMIDHUBER, J. Deep Learning in Neural Networks: An Overview. Neural Networks, v. 61, p. 85-117, 2015.

SHONE, N.; NGOC, T. N.; PHAI, V. D.; SHENG, Q. Intrusion Detection with Deep Learning: DoHNet Approach. In: Proceedings of the International Conference on Communication Systems and Network Technologies (CSNT), 2018. p. 1-6.

SOMMER, R.; PAXSON, V. Outside the Closed World: On Using Machine Learning for Network Intrusion Detection. In: IEEE SYMPOSIUM ON SECURITY AND PRIVACY (SP), 31., 2010, Oakland. Proceedings... Oakland: IEEE, 2010. p. 305-316.

SONG, Dawn, EYKHOLT, Kevin; EVTIMOV, Ivan; FERNANDES, Earlence; LI, Bo; RAHMATI, Amir; XIAO, Chaowei; PRAKASH, Atul; KOHNO, Tadayoshi;. Robust Physical-World Attacks on Deep Learning Visual Classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 1625-1634

SOUZA, Gills Lopes Macêdo; PEREIRA, Dalliana Vilar. A Convenção de Budapeste e as leis brasileiras. **Seminário Cibercrime e Cooperação Penal Internacional**, n. 1, 2009.

STALLINGS, W. **Criptografia e segurança de redes: princípios e práticas**. São Paulo: Pearson Education do Brasil, 2014.

SURYOTRISONGKO, David; VAN ECK, Nees Jan; WALTMAN, Ludo. Review of cybersecurity research topics, taxonomy and challenges: interdisciplinary perspective. **Journal of Cybersecurity**, v. 6. 2019.

TEGMARK, Max. “Life 3.0: ser-se humano na era da Inteligência Artificial”. Alfragide: D.Quixote. 2019.

TIMOCHENCO, L. **Inteligência Artificial na Segurança da Informação.** São Paulo: Revista Digital Online, 2020. Disponível em: https://infranewstelecom.com.br/inteligenciaartificial-na-seguranca-da-informacao. Acesso em: 25 out. 2020.

TUPTUK, N., & Hailes, S. "Security of smart manufacturing systems." *Journal of Manufacturing Systems*, *47*, 93-106. 2018.

VEALE, M., Binns, R., & Edwards, L. (2018). "Algorithms that remember: Model inversion attacks and data protection law." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2133), 20180083.

VIGANÒ, Eleonora; LOI, Michele; YAGHMAEI, Emad. Cybersecurity of critical infrastructure. ***The Ethics of Cybersecurity***, p. 157-177, 2020.

ZERLANG, Jesper. ***GDPR: a milestone in convergence for cyber-security and compliance***. Network Security, v. 2017, n. 6, p. 8-11, 2017.

ZHOU, Y.; QI, L.; HONG, Z.; CHEN, X. A Survey on Cyber Security Threats and Detection Methods in Textual Data. IEEE Access, v. 8, p. 132755-132772, 2020.