FACULDADE DE TECNOLOGIA E CIÊNCIAS SOCIAIS APLICADAS – FATECS
CURSO

Gabriel Verleun Boechat
21650430

ANÁLISE DA APLICAÇÃO DA INTELIGÊNCIA ARTIFICIAL NO ATUAL CENÁRIO DE CIBERSEGURANÇA.

BRASÍLIA
2022

Gabriel Verleun Boechat

ANÁLISE DA APLICAÇÃO DA INTELIGÊNCIA ARTIFICIAL NO ATUAL CENÁRIO DE CIBERSEGURANÇA.

Trabalho de Conclusão de Curso (TCC) apresentado como um dos requisitos para a conclusão do curso de Engenharia de Computação do CEUB– Centro Universitário de Brasília

Orientador (a)**: Me. Francisco Javier De Obaldía Díaz**

BRASÍLIA
2022

Gabriel Verleun Boechat

ANÁLISE DA APLICAÇÃO DA INTELIGÊNCIA ARTIFICIAL NO ATUAL CENÁRIO DE CIBERSEGURANÇA

Trabalho de Conclusão de Curso (TCC) apresentado como um dos requisitos para a conclusão do curso de Engenharia de Computação do CEUB – Centro Universitário de Brasília

Orientador (a)**: Me. Francisco Javier De Obaldía Díaz**

Brasília, 2022.

BANCA EXAMINADORA

_____
Francisco Javier De Obaldía Díaz -Mestre
Orientador (a)

_____
Fábio Oliveira Guimarães - Mestre
Examinador (a)

_____
Ingrid Maria Dittert - Doutora
Examinador (a)

ANÁLISE DA APLICAÇÃO DA INTELIGÊNCIA ARTIFICIAL NO ATUAL CENÁRIO DE CIBERSEGURANÇA

Gabriel Verleun Boechat[1], Francisco Javier De Obaldía Díaz[2], Fábio Oliveira Guimarães[3], Ingrid Maria Dittert[4]

RESUMO

Este trabalho tem como principal objetivo analisar o atual cenário de inteligência artificial aplicada dentro de sistemas de informação para áreas de segurança da informação. A análise tem como base relatórios anuais de cibersegurança, disponibilizados online ou cedidos através de terceiros, e também serão utilizados artigos científicos publicados e artigos advindos da internet para citar as formas de atuação de inteligências artificiais no meio de cibersegurança. Foram traçados comparativos referentes ao uso de IA na área de cibersegurança, assim como o custo associado com invasões. Ademais, uma análise do atual cenário de cibersegurança é feita, bem como, o que esperar de um futuro próximo..

Palavras-chave: “Inteligência Artificial”, "Cibersegurança", “Malware”

Abstract: The main objective of this work is to analyze the current scenario of artificial intelligence applied within information systems for information security areas. The analysis is based on annual cybersecurity reports, available online or transferred through third parties, published scientific articles and articles from the internet will also be used to cite the ways in which artificial intelligences act in the context of cybersecurity. Comparative plots were drawn regarding the use of AI in the area of cybersecurity, as well as the cost associated with intrusions. An analysis is made of the current cybersecurity scenario and what to expect in the near future.

keywords: “Artificial Intelligence”, “Cybersecurity”, “Malware”

[1] UniCEUB, Gabriel Verleun Boechat.
[2] UniCEUB, Francisco Javier De Obaldía Díaz.
3 UniCEUB, Fábio Oliveira Guimarães.
[4] UniCEUB, Ingrid Maria Dittert.

1 INTRODUÇÃO

> “Nas últimas décadas, a utilização da rede de computadores mundiais aumentou significativamente. O uso de dispositivos virou parte do cotidiano do ser humano moderno. Na mesma proporção que aumentou a funcionalidade da rede no dia a dia das pessoas, aumentou também o número de ataques a dispositivos conectados. Isso resultou em uma redução da funcionalidade, taxa de transferência e desempenho da rede. Dessa forma, para prevenir tais ataques, avanços tecnológicos foram realizados para que consigamos cada vez mais ser dependentes das redes globais ao nos envolvermos em atividades sociais, empresariais e educacionais de forma segura.” (Anwar *et al.* (2017)).

De acordo com a linha de pensamento apresentada por Anwar *et al.* (2017), devido ao aumento significativo do uso de computadores e redes, novos questionamentos e problemas são levantados. O que tornou a segurança de dispositivos foco de estudo para maior desenvolvimento de tecnologias capazes de garantir integridade e disponibilidade do sistema contra ameaças. Nessa perspectiva, o trabalho de Comerlato *et al.* (2022) nos introduz ao conceito de Sistemas Críticos quanto à Segurança SCS, que são responsáveis por identificar e mitigar falhas de segurança.

A inteligência artificial continua sendo implementada na cibersegurança e para desenvolver tecnologias de ponta. Para manter um certo nível de desempenho aceitável, é necessário localizar e adaptar-se contra a evolução de todas as ameaças dentro da rede. Isso evidencia que os recursos avançados de defesa estão sendo utilizados em maior número e com maior grau de relevância no cenário atual.

Com o avanço do uso de Inteligência Artificial (IA) nas tecnologias mundiais, a utilização IA dentro de empresas em áreas de segurança vem aumentando ao longo dos anos, visto que a demanda por prevenção, identificação e correção de ataques a sistemas de informação tem crescido juntamente ao aumento e popularização do cibercrime.

As formas como os ataques são coordenados vem evoluindo juntamente às tecnologias e novas formas de reagir aos mesmos vem surgindo e se aprimorando cada vez mais. A maneira como IA vem sendo utilizada para providenciar a melhor segurança dentro de sistemas de informação facilita, em parte, que grandes ameaças conhecidas sejam prevenidas de forma quase que automática, porém, conforme a tecnologia avança, os métodos de invasão também tendem a se tornar mais sofisticados.

Um ataque cibernético é geralmente uma tentativa maliciosa e coordenada por um indivíduo ou organização, com intenções maliciosas, para violar outro indivíduo ou sistema de informação da organização com o intuito de obter acesso e/ou dados do mesmo ou de terceiros.

Nessa perspectiva, há o questionamento do porquê os ciberataques estão aumentando. Uma das teorias que se irá debater é que os ciberataques são mais baratos, mais convenientes e possuem um risco menor quando comparados a ataques físicos. Cibercriminosos precisam de pouco menos que um computador e acesso à internet para conduzir ataques, não possuem limitações geográficas que os impeça de acessar um sistema e, devido a natureza da internet, são capazes de esconder sua identidade e localização de forma muito simples e eficaz.

A cibersegurança tem como objetivo entender a forma como ataques funcionam e gerar medidas de resposta para os mesmos, garantindo a confidencialidade, integridade e disponibilidade de informações.

A utilização da Inteligência Artificial está, nos dias de hoje, vulgarizada entre as mais variadas áreas de computação, onde existe uma promessa de grandes avanços através da utilização da Inteligência Artificial e da cibersegurança.

O artigo tem o intuito de analisar a forma como a inteligência artificial é utilizada atualmente na área de cibersegurança, através da análise de dados referentes a ataques, estudos de caso envolvendo o uso de IA e artigos que abordam o tema.

2 REVISÃO BIBLIOGRÁFICA

> "Nas primeiras décadas de sua existência, as redes de computadores foram principalmente usadas por pesquisadores universitários para enviar e-mails e por funcionários corporativos, para compartilhamento de impressoras. Nessas condições, a segurança não recebeu muita atenção. Agora, como milhões de cidadãos comuns estão usando redes para serviços bancários, compras e preenchimento de suas declarações de impostos, e fraqueza após fraqueza é encontrada, segurança de rede tornou-se um problema de grandes proporções." (Tanenbaum, 2011, p. 763).

Seguindo a linha de pensamento de Tanenbaum, a utilização ampla de cidadãos comuns nas redes faz que a segurança da rede possa vir a ser um dos principais desafios de empresas, governos e Estados no futuro. E para entendermos melhor a magnitude da segurança cibernética, precisamos compreender a diferença entre a mesma e a segurança da informação.

Dessa forma, pode-se entender utilizando linha de pesquisa e entendimento de Von Solms e Van Niekerk, que evidencia que a segurança da informação é a proteção do dado, da informação em si, resultantes de várias ameaças e vulnerabilidades. Entretanto, a segurança cibernética não é necessariamente apenas a proteção do ciberespaço, como também a proteção daqueles que operam no ciberespaço e de qualquer um que pode ser alcançado via ciberespaço. (Von Solms & Van Niekerk, 2013).

Outrossim, o objetivo da segurança da informação é garantir a continuidade dos negócios e minimizar os danos aos negócios, limitando o impacto de incidentes de segurança. Segurança da tecnologia da informação e comunicação (TIC) lida com a proteção da real tecnologia baseada em sistemas nos quais as informações são comumente armazenadas e/ou transmitidas.

Se segurança cibernética é sinônimo de segurança da informação, seria razoável supormos que os incidentes de segurança cibernética também podem ser descritos em termos das características usadas para definir a segurança da informação. Assim, um incidente de segurança cibernética, por exemplo, também levaria a uma violação da confidencialidade, integridade ou disponibilidade de formação, que também são pilares que compõem a segurança da informação. Porém, existem ameaças de segurança cibernética que não fazem parte do âmbito da segurança da informação, como casos de cyberbullying e invasões de equipamentos IoT.

Podemos entender por Tanenbaum (2011), que as informações guardadas dentro de empresas e organizações variam entre informações técnicas, comerciais, financeiras ou legais. A maioria das informações estão guardadas em computadores. Computadores domésticos também contém dados que podem ser valiosos para atacantes, como informações pessoais, financeiras ou de caráter sigiloso. (Tanenbaum, 2007).

A cibersegurança teve seu início em uma época em que não havia necessidade em desenvolver técnicas com foco em melhorar a segurança do meio digital e de seus ativos. Até o início dos anos 1990, o número de

computadores pessoais presentes em casas de cidadãos comuns era mínimo e boa parte dos computadores se encontravam em empresas, universidades e outras organizações. A maior parte destas máquinas se encontrava isolada dentro de seu próprio ambiente, desconectadas de qualquer rede.

Em comparativo, no cenário atual, é possível ver uma grande mudança na popularidade do termo "cibersegurança" como evidenciado no trabalho "*AI-Driven Cybersecurity: An Overview, Security Intelligence Modeling and Research Directions*", já que, como constatado anteriormente, há uma necessidade crescente de desenvolver novos métodos de segurança para garantir um sistema que consiga se manter dentro dos padrões de segurança básica (também conhecidos como a tríade da CIA).

> "O crescimento do trabalho remoto e do ensino a distância desencadeado pela pandemia obrigou empresas e instituições de vários tamanhos a ampliar ou até superar o conceito do perímetro de rede. O ambiente tecnológico não mais se resume às salas de aula ou ao escritório, ele está em qualquer lugar do mundo onde haja conexão. Soluções de segurança que enxergam a rede da empresa como um mundo interno e sempre confiável, dificilmente terão efetividade nesse cenário." (Relatório anual de segurança RNP, 2021).

Ainda enfrentando e descobrindo novas mudanças que vieram advindas do período de pandemia, o mundo digital está avançando cada vez mais rápido devido à uma necessidade de aperfeiçoamento constante. Novos avanços são feitos no campo de TI quase que diariamente, com isso, observamos também que, devido ao alto grau de utilização de tecnologias conectadas, setores não focados em segurança de dados também sofrem com vazamentos e intrusões, gerando custos. Iremos analisar tais custos por meio do *IBM security: Cost of a Data Breach Report* 2022.

> "Ao analisar o comparativo histórico das últimas décadas, antes que houvesse qualquer conversa sobre o uso de inteligência artificial (IA), *machine learning* (ML) e *deep learning* (DL) para paradigmas de segurança cibernética na tecnologia (TI), a implementação de padrões básicos de segurança foi predicado em respostas manuais diagnósticas e reativas a informações de interações conhecidas anteriores, ou de assinaturas baseadas em *hash* criadas a partir de infecções anteriores, ataques ou anomalias. Esta foi a base do *host*, rede, aplicativo, servidor, dispositivos e requisitos de segurança de perímetro." (Badhwar, R. (2021)).

A segurança cibernética antes da aplicação de IA não tinha a capacidade de reagir em tempo real, os ataques teriam que ser descobertos dentro do sistema e em seguida seriam contornados e/ou resolvidos. Como o atual cenário exibe ataques mais frequentes, sofisticados e distribuídos, vê-se necessário a implementação de uma IA para auxiliar o sistema de segurança, visto que apenas a segurança baseada em IA pode detectar padrões de ataque anteriormente desconhecidos, ataques de dia zero (*zero-day-attack*) e outros desafios apresentados por *ransomware*, *malware* polimórfico e ameaça persistente avançada (APT).

Com os ataques cibernéticos em rápida evolução e a rápida multiplicação de dispositivos acontecendo hoje, a IA e o aprendizado de máquina podem ajudar a acompanhar os cibercriminosos, automatizar a detecção de ameaças e responder com mais eficácia do que as técnicas convencionais orientadas por *software* ou manuais.

O termo inteligência artificial foi criado por John McCarthy em um congresso sobre o tema realizado na Universidade de

Dartmouth em 1956. Os primeiros anos da IA foram repletos de sucessos. Considerando-se as ferramentas e computadores primitivos da época, o fato de os computadores serem capazes de efetuar atividades consideradas inteligentes e não somente operações aritméticas era surpreendente.

Atualmente, a IA procurou se utilizar das teorias existentes como base, ao invés de procurar por soluções completamente novas. O surgimento da internet, dos grandes volumes de dados e dos mecanismos de pesquisa deu um novo fôlego ao campo da IA. O conceito mais aceito nos dias de hoje é o de um agente inteligente, em que as abordagens simbólicas e conexionistas podem trabalhar de forma colaborativa para a resolução de problemas através de um sistema computacional. O termo agente refere-se a algo que pode perceber seu ambiente por meio de sensores e de agir sobre este ambiente por intermédio de atuadores.

Muitas tecnologias modernas acabam por serem confundidas com técnicas de IA, como em técnicas de processamento de imagens ou aparelhos relacionados à IoT, normalmente chamados de *smart devices*. Apesar do termo inteligência artificial ainda não estar inteiramente definido, é importante focar no conceito de um agente inteligente. De forma geral, um agente deve ser capaz de representar conhecimento e incerteza; de raciocinar; de tomar decisões; de aprender com experiências e instruções; de se comunicar e interagir com pares e com o mundo.

Em relação ao futuro da cibersegurança, podemos citar o livro do pesquisador Badhwar, R. (2021). *The CISO's Next Frontier*. De acordo com o autor, as tecnologias atuais possuem limitações e que a grande próxima revolução tecnológica seria baseada na computação quântica. Tais modelos e exemplos serão discorridos em etapa futura do desenvolvimento do trabalho.

Apesar de possuir múltiplas áreas de atuação, a Inteligência Artificial aplicada à cibersegurança vem tomando um grande espaço no mercado e tem se tornado uma das principais ferramentas para auxiliar no combate contra os cibercrimes, já que a mesma não possui as mesmas necessidades que um trabalhador humano e também possui uma capacidade maior de desempenho quando diante de situações já conhecidas ou facilmente contornáveis. Porém, com base no artigo: *The cold war online*, vemos que o atual cenário está dividido e ambos os lados possuem interesses que os levam a cometer ou prevenir crimes. Apesar dos grandes avanços feitos com o uso de IA, ainda existem problemas relacionados a mesma que podem ser melhor observados a partir da interpretação do artigo: *Deceiving AI*.

A seguir, serão expostas algumas técnicas de IA que permitem que uma máquina aparente ter inteligência e como estas funcionam, ao longo do trabalho, serão realizadas comparações entre as aplicações das técnicas utilizadas em cibersegurança, porém será primeiro necessário contextualizar as mesmas.

> "*Machine Learning*: O *Machine Learning* consiste na aquisição de conhecimento automático por parte das máquinas, sem a necessidade explícita de ser programado, a aprendizagem centra-se em fatos conhecidos. Além disso, explora o estudo e construção de algoritmos que podem aprender e efetuar previsões sobre os dados" (Ongsulee, 2017).

Ainda tendo a óptica de Ongsulee como base, podemos definir os tipos de aprendizagem do machine learning em:

– Aprendizagem Supervisionada: O algoritmo recebe um conjunto de dados de entrada junto com saídas corretas correspondentes. O algoritmo aprende comparando suas saídas reais com saídas

corretas para encontrar erros.

— Aprendizagem Não Supervisionada: A máquina recebe um conjunto de entradas, mas nenhum conjunto de saídas correspondente. Nesse caso, cabe à máquina encontrar padrões de semelhanças e diferenças entre os dados e com esses padrões gerar novas saídas corretas.

— Aprendizagem Semi-supervisionada: É usada para as mesmas aplicações como aprendizado supervisionado. Mas ele usa dados rotulados e não rotulados para treinamento. Esse tipo de aprendizagem pode ser usada com métodos como classificação, regressão e previsão.

— Aprendizagem por Reforço: é frequentemente usado para robótica, jogos e navegação. Com o aprendizado por reforço, o algoritmo descobre por tentativa e erro quais ações produzem as maiores recompensas. Este tipo de aprendizagem tem três componentes: o agente (o aprendiz ou tomador de decisão), o ambiente (tudo com que o agente interage) e ações (o que o agente pode fazer).

Entretanto, com base no estudo: *The Burden of Artificial Intelligence on Internal Security Detection*, entenderemos que existem problemas relacionados ao uso de IA em diversos setores, assim como a cibersegurança também enfrenta dificuldades ao fazer o uso da ferramenta.

3 METODOLOGIA/DESENVOLVIMENTO DO TRABALHO

O artigo possui natureza aplicada, já que o objetivo do trabalho é analisar as formas como a IA e seus métodos são utilizados dentro do ambiente de segurança cibernética, assim, gerando uma visão mais generalizada sobre o tema e suas aplicações.

É uma pesquisa de revisão bibliográfica qualitativa referente à aplicação atual de técnicas de IA para auxílio em cibersegurança e que faz uso de artigos e outras fontes para descrever as características da aplicação propriamente dita. A forma como as informações são utilizadas, confere uma pesquisa explicativa, pois há uma interpretação dos fatos referentes ao atual cenário computacional envolvendo segurança cibernética e o uso de técnicas de IA.

Quanto aos procedimentos técnicos, é uma pesquisa bibliográfica centrada em artigos científicos de fontes como google acadêmico; IEEE; repositórios de faculdades nacionais e internacionais; simpósios de tecnologia com foco em segurança e IA e relatórios de empresas com foco em segurança cibernética e pesquisa e aplicação de IA.

O trabalho foi desenvolvido em algumas etapas, como a seguir:

Etapa 1: Pesquisa na literatura e fontes de informação sobre incidentes de cibersegurança, impactos e formas de mitigar ataques à segurança, sendo utilizadas. Na seção anterior foram expostos vários casos os quais serão analisados, diante um cenário com uso de IA. O levantamento de dados referente a pesquisa foi realizado utilizando meios de pesquisa como o google acadêmico, IEEE, Sci HUB, acervos universitários, periódicos e livros científicos.

Etapa 2: Conceitualização de Agentes de Inteligência Artificial e os métodos algorítmicos de IA. Outrossim, estudo inicial da operacionalidade de cibercriminosos - em relação a praticidade com base no método de exfiltração de dados - DNS.

Podemos concluir que para obter vantagem e velocidade na criação de soluções na proteção das redes, o *Machine Learning* é um dos métodos mais utilizados na proteção do ciberespaço. Entretanto, ele é amplo e também engloba agentes e métodos, além dos já citados anteriormente, que são fundamentais para o futuro da proteção e utilização segura do ciberespaço.

Para compreender o conceito de IA nos dias de hoje, necessitamos ter em mente o conceito de um agente inteligente, capaz de obter informações a partir do meio em que

este se encontra e, também, ser capaz de tomar decisões baseadas em informações advindas do próprio ambiente, assim como informações prévias, sendo essas geradas pelo próprio agente ou trazidas de um banco de dados de apoio do mesmo. O agente também necessita de aprendizado contínuo, para que possa continuar se aprimorando e desenvolvendo sua inteligência.

Existem cinco tipos de agentes:

- Agentes reativos: São máquinas simples que se limitam a reagir aos estímulos que recebem do ambiente, no entanto, poderão agir mesmo sem a recepção de estímulos, apenas como resposta ao ambiente na qual se encontram.
- Agentes de procura: Estes agentes devem ser capazes de entender os estados existentes nas ações e construir com base nisso uma representação interna dos mesmos, para além disso, devem possuir a capacidade de agir sobre eles, tendo em consideração as regras de funcionamento dos sistemas em questão.
- Agentes baseados em conhecimento: Estes agentes necessitam de ter conhecimento e raciocínio para aumentarem o seu desempenho, assim sendo, um agente tem de construir a sua imagem do mundo, como tal, é necessário saber representar o conhecimento e interagir para conseguir desenvolver o raciocínio, tendo assim a capacidade de decidir.
- Agentes aprendizes: Um agente aprendiz baseia-se na percepção e na ação (agentes reativos), bem como na capacidade de decidir e aprender.
- Agentes adaptativos: Utilizam algoritmos genéticos, estes são técnicas que permitem efetuar otimização, sendo úteis para a resolução de problemas, podem ser consideradas técnicas inteligentes, pois permitem trabalhar simultaneamente em soluções alternativas, sendo ferramentas poderosas quando aplicadas para resolver problemas.

Em relação aos métodos que podem ser utilizados, temos os seguintes:

- *Artificial Immune Systems* (AIS): Inspirada nos princípios e processos de um sistema imune, os algoritmos são modelados considerando características de aprendizagem e memória para ajudar na resolução de problemas.
- Redes Neurais: Composto por unidades interligadas, como os neurônios, que processam informações, sendo inspiradas no funcionamento do cérebro humano, organizam-se em camadas com diversos nós capazes de trocar estímulos por estarem conectados em rede.
- *Deep Learning*: Utiliza redes neurais com várias camadas de unidades de processamento, utilizando o avanço tecnológico que tem existido ao longo dos anos. Inclui reconhecimento de imagens e de fala.
- Árvores de Decisão: Representa a função que recebe um vetor de parâmetro de entrada e retorna uma decisão. Os valores de entrada e de saída podem ser discretos e contínuos.
- *Natural Language Processing* (NLP): O Processamento de Linguagem Natural (PLN) é uma área de pesquisa e aplicação que explora como os computadores podem ser usados para entender e manipular texto ou fala em linguagem natural para fazer coisas úteis.
- *Machine Vision*: Pesquisadores em visão computacional vêm desenvolvendo, paralelamente, técnicas matemáticas para recuperar a forma tridimensional e a aparência de objetos em imagens.
- *Fuzzy Logic*: Os sistemas *Fuzzy Logic*, são métodos de raciocínio, que propõem cálculos matemáticos para traduzir

o conhecimento humano relativamente aos processos reais. O mecanismo de inferência dos sistemas *Fuzzy Logic* consiste em 3 etapas, a primeira consiste no mapeamento, utilizando uma função com os valores numéricos da entrada. Na segunda etapa, o sistema *Fuzzy Logic* processa as regras de acordo com a robustez da entrada. Na terceira etapa, os valores resultantes são novamente transformados em valores numéricos.

Podemos entender também de Pariwat Ongsulee sobre o estudo do *deep learning*. Entendendo que o *deep learning* ingere grandes quantidades de dados para treinar uma rede neural profunda que aprende por conta própria ao longo do tempo, como identificar imagens ou executar outras tarefas. Modelos de *deep learning* podem atingir altas taxas de precisão mesmo para atividades de ataque que são apenas vagamente definidas. Eles são usados para identificar imagens não seguras para o trabalho e outras (como logotipos) ou para detectar melhor e-mails de *spam* e tentativas de *phishing*.

Na perspectiva do estudo inicial da operacionalidade de cibercriminosos, os mesmos estão determinados a contornar as defesas cibernéticas existentes, como *firewalls* e sistemas de detecção e prevenção de intrusão. Aqueles empenhados em roubar informações valiosas de clientes ou negócios estão usando cada vez mais o sistema de nomes de domínio (DNS), o diretório de endereços da Internet, que pode ser "um elo fraco nas práticas de segurança cibernética".

Os dados DNS geralmente podem passar por *firewalls*, e os invasores os sequestram para transportar seu *malware*, assumir o controle de dispositivos e roubar registros de clientes, e-mails e outros dados confidenciais. O aprendizado de máquina pode detectar e impedir o chamado "túnel DNS" para extração de dados, com modelos treinando continuamente em trilhões de consultas DNS geradas e coletadas diariamente em todo o mundo.

Podemos entender melhor sobre a utilização do DNS no artigo online *DNS AS A PATHWAY FOR INFILTRATION AND EXFILTRATION* em que fala:

> "Os *hackers* podem usar vários caminhos para roubar dados, mas o que geralmente é deixado aberto sem saber é o DNS, ou Sistema de Nomes de Domínio. O DNS está sendo cada vez mais usado para extração de dados por dispositivos infectados por *malware* ou por funcionários desonestos. De acordo com uma recente pesquisa de segurança de DNS de empresas sediadas na América do Norte e Europa, 46% dos entrevistados experimentaram exfiltração de DNS e 45% experimentaram tunelamento de DNS. O DNS não é usado apenas para vazamento de dados, mas também para mover código malicioso para uma rede. Essa infiltração é mais fácil do que você pensa. Os *hackers* podem preparar um comando binário, codificá-lo e transportá-lo através de firewalls e filtros de conteúdo via DNS para a rede de uma organização. Os *hackers* enviam e recebem dados via DNS, convertendo-os efetivamente em um protocolo de transporte secreto. As soluções de prevenção de perda de dados (DLP) normalmente analisam o vazamento de dados por e-mail, web, FTP e outros vetores, mas não têm visibilidade da exfiltração baseada em DNS." (Blackhat. **2018).**

Etapa 3: Realizar análise dos custos relativos ao gasto gerado por invasões e gastos mitigados com uso de IA, assim como descrição e análise dos tipos de cibercrimes mais comuns no mercado. Outrossim, estudo sobre a base da cibersegurança - Tríade da CIA - além de análise do cenário atual político e empresarial da cibersegurança e da

Inteligência Artificial.

Em relação aos investimentos e custos de proteção fazendo um comparativo entre regiões (Figura 1), tendo como base o Relatório internacional de segurança da IBM, aqui já anteriormente citado.

> "O custo total médio global de uma violação de dados aumentou em US $0,11 milhões a US $4,35 milhões em 2022, o maior valor da história deste relatório. O aumento de USD 4,24 milhões no relatório de 2021 para US $4,35 milhões no relatório de 2022 representa um aumento de 2,6%. Nos últimos dois anos, a média do custo total aumentou 12,7%, sendo de US $3,86 milhões no relatório 2020." *(IBM security: Cost of a Data Breach Report* 2022,página 9).

Figura 1.Custo médio de uma violação de dados por um país ou região.

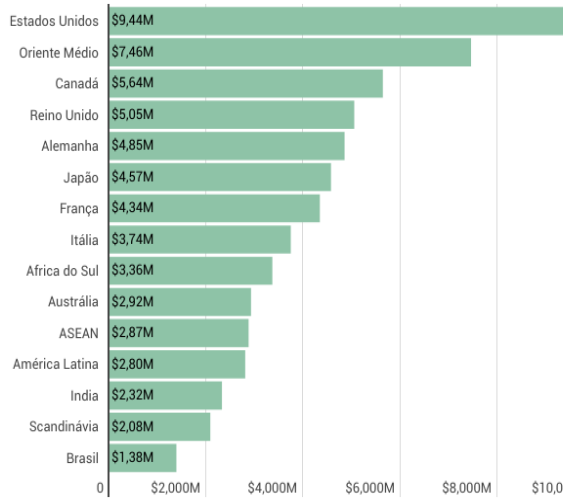


Gráfico adaptado do relatório: IBM security: Cost of a Data Breach Report 2022

Fonte: *IBM SECURITY: COST OF A DATA BREACH REPORT* 2022 (2022).

Outrossim, o relatório também afirma que:

> "Os Estados Unidos lideram a lista por 12 anos consecutivos. Enquanto isso, o país com a taxa de crescimento mais rápida nos últimos anos foi o Brasil, um aumento de 27,8% de USD 1,08 milhão para US $1,38 milhão." *(IBM security: Cost of a Data Breach Report* 2022, página 7)

Assim, a análise do relatório permite inferir que inovações estão se tornando cada vez mais frequentes e os custos referentes ao processo de reparação dos danos causados são maiores entre aqueles que não possuíam ferramentas de segurança para remediar as invasões. Dessa forma, podemos entender que se torna mais caro reparar os danos do que preveni-los.

Ainda abordando o relatório da IBM, é constatado no levantamento da pesquisa que não só os custos relacionados à invasões diminuíram com o uso de uma tecnologia de inteligência artificial, mas o tempo necessário para encontrar e contornar a invasão também diminui. Na verdade, o principal fator relacionado à diminuição do custo de identificação, manutenção e prevenção de ataques é o uso de IA para auxiliar os técnicos de segurança, como evidenciado nos gráficos das Figuras 1 e 2, adaptados do mesmo relatório.

Figura 2.Custo médio de uma violação de dados por IA de segurança em nível de implantação de automação.

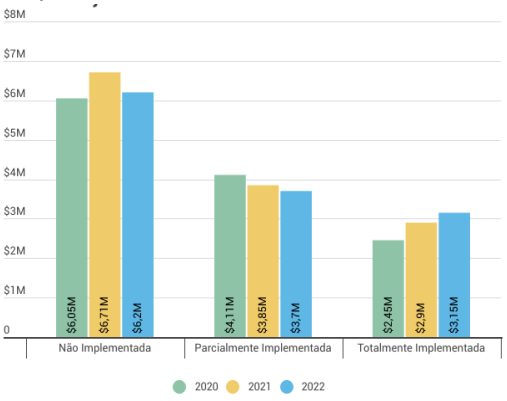


Gráfico adaptado do relatório: IBM security: Cost of a Data Breach Report 2022

Fonte: *IBM SECURITY: COST OF A DATA BREACH REPORT* 2022 (2022).

O gráfico da Figura 2 mostra a diferença entre os custos relativos à uma invasão ao longo de um ano (com o comparativo de 2020, 2021 e 2022). Podemos notar que o uso de uma tecnologia de IA aplicada a cibersegurança tem influência direta nos custos gerados por uma invasão, organizações que fazem uso de IA totalmente implementada tem um custo duas vezes menor do que organizações que não a utilizam, também podemos inferir, através do gráfico da Figura 3, sobre o aumento do uso de uma ferramenta de IA em organizações.

Figura 3.Estado de segurança IA e automação comparando três níveis de implementação.

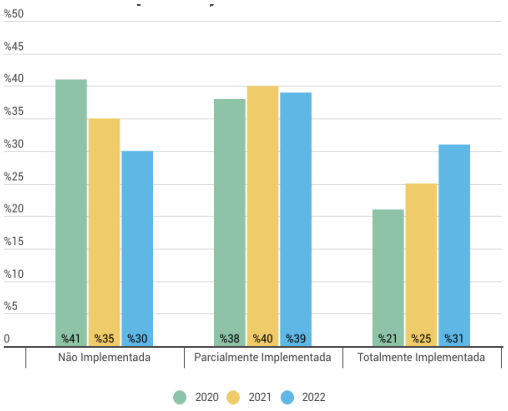


Gráfico adaptado do relatório: IBM security: Cost of a Data Breach Report 2022

Fonte: *IBM SECURITY: COST OF A DATA BREACH REPORT* 2022 (2022).

Conforme ataques se tornam mais frequentes e mais complexos, é observado no gráfico um aumento no uso de IA em apoio à cibersegurança, sendo que a implementação total da ferramenta se encontra crescendo com o passar do tempo, tendo em vista o aumento na popularização de seu uso assim como da necessidade do mesmo para conseguir acompanhar o atual cenário da cibersegurança.

Com o passar dos anos, nos tornamos mais e mais dependentes da tecnologia, até o ponto em que boa parte de nossas informações, sejam essas sigilosas ou não, se encontram online ou em máquinas conectadas à alguma rede. O custo da perda ou violação destes dados vem crescendo ao longo dos anos e, com a influência da pandemia, aumentou de forma significativa. O relatório da IBM, citado anteriormente, nos dá uma visão do ponto de vista financeiro sobre a situação gerada após os ataques e o custo total de gerado após uma invasão ou ataque, assim como nos permite analisar avanços e resultados gerados a partir de um investimento feito em segurança.

O relatório também nos mostra os setores mais afetados durante o ano de 2022 com uma comparação com o ano de 2021.

Figura 4.Custo médio de uma violação de dados por setor.

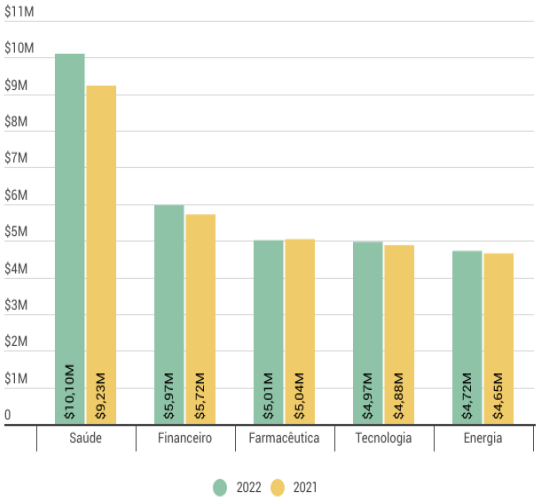


Gráfico adaptado do relatório: IBM security: Cost of a Data Breach Report 2022

Fonte: *IBM SECURITY: COST OF A DATA BREACH REPORT* 2022 (2022).

> "As cinco principais indústrias por custo permaneceram inalteradas na ordem de classificação do relatório de 2021"(*IBM security: Cost of a Data Breach Report* 2022, página 11).

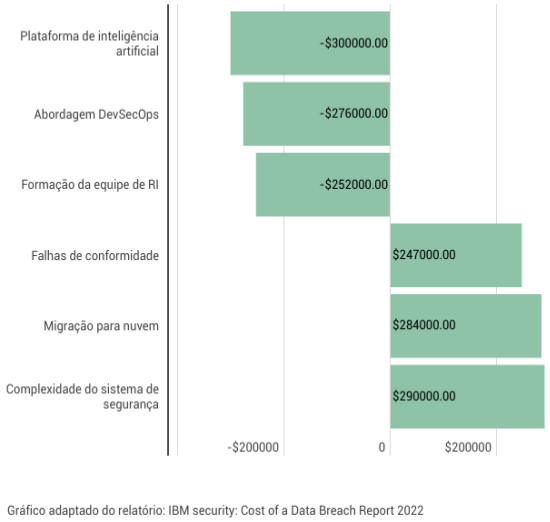
Os setores citados no gráfico são grandes alvos de ciberataques pois são setores, em sua maioria, que possuem grandes quantidades de dados, estes que podem

representar condições socioeconômicas de indivíduos (identificar pessoas vulneráveis para coordenar ataques de engenharia social) ou roubo de informações sigilosas. Dentro do contexto moderno, todo o dado pode ser explorado de alguma forma.

Os riscos envolvendo a segurança cibernética se espalham por diversos setores. O custo, apesar de flutuar a depender do setor analisado, continua sendo de valores altos, criando uma preocupação já reforçada antes por motivos não relativos aos mesmos, e reforçando cada vez mais uma desconfiança relativa aos métodos de segurança mais comuns atualmente empregados.

O relatório também reconhece que o uso de uma IA para cibersegurança é um dos fatores chave para diminuir gastos gerados a partir de invasões, como evidenciado no gráfico da Figura 5.:

Figura 5.Principais fatores no custo total médio de uma violação de dados.



Fonte: *IBM SECURITY: COST OF A DATA BREACH REPORT* 2022 (2022).

A demanda por novas técnicas, capazes de suprir as necessidades digitais atuais de segurança, é impulsionada pelos números de casos de crimes cibernéticos atualmente, pela facilidade com os quais os mesmos são cometidos e, principalmente, para garantir que documentos sigilosos sejam protegidos de terceiros mal intencionados (como em casos que envolvem grandes roubos de dados por governos).

Podemos observar que está se tornando mais comum o uso de IA aplicadas como apoio aos SCS no meio corporativo e, tendo em vista o cenário geopolítico atual no ambiente cibernético, vemos que há uma necessidade, cada vez maior, de proteger dados sensíveis assim como outros ativos, sendo esses de pessoas comuns, empresas ou governos. A IA é vista como a chave para a quarta revolução tecnológica e sua aplicação na área de segurança se torna mais marcante conforme o número de ataques continua crescendo.

> "As informações sobre explorações futuras geralmente aparecem em fóruns de discussão enquanto a exploração está sendo projetada e testada. Uma vez que a exploração seja aperfeiçoada, provavelmente será colocada à venda em um mercado da *darknet* antes de aparecer na natureza. A equipe da *Arizona State* aproveitou isso e construiu um sistema de aprendizado de máquina que monitora sites *darknet* e *deepnet* para tráfego sobre explorações de segurança. Com efeito, seu sistema de aprendizado de máquina transforma as atividades de comunicação e *marketing* dos *hackers* em um sistema de alerta antecipado que protege os desenvolvedores de *software* contra explorações de dia zero. Eles hackearam os *hackers*." (Murnane 2016).

Existem vários tipos de ciberataques que ocorrem nos dias de hoje e com os quais os utilizadores de Sistemas de Informação devem ter atenção e cuidado de forma a não serem alvos dos mesmos. Os ataques mais

conhecidos são os *malwares*, *scarewares*, *botnets* e ataques de DoS.

— *Malware*: É um *software* malicioso que engloba todo e qualquer *software* que tenha sido alterado com o objetivo de danificar dispositivos, roubar informação e assumir controle, seja a nível individual ou a nível organizacional. Existem vários tipos de *Malware*, como *backdoors*, *spyware*, cavalo de tróia, vírus, entre outros.

— *Ataque Denial of Service* (DoS): É um ataque que tem o objetivo principal de inativar uma máquina ou uma rede, para que esta se torne inacessível aos utilizadores, este é executado através do bloqueio do tráfego, nomeadamente, enchendo-o de pedidos ou então acionando uma falha no sistema.

— *Phishing*: é um tipo de ciberataque que usa o e-mail como arma, o receptor da mensagem acredita na credibilidade do remetente, abre o e-mail, seleciona a hiperligação que normalmente é disponibilizada e a informação sobre o utilizador é obtida sem a sua permissão. Consegue obter informações como senhas ou dados do cartão de crédito.

— *Ransomware*: Este tipo de ataque cibernético restringe o acesso ao sistema do computador ou aparelho que se pretende atacar e tentam pedir um resgate ao dono, de forma a libertar o acesso. O *Ransomware* consegue atingir uma máquina através da recepção de um anexo enviado por e-mail ou através do *browser*, normalmente quando se visita uma página que já tenha sido infectada com o vírus.

— SQL *Injection*: É um tipo de ciberataque na qual se aproveita falhas em sistemas ligados ou que interagem com bases de dados. O ataque é executado através de comandos SQL, onde o atacante insere uma instrução SQL personalizada dentro que uma query.

— *Cross site scripting*: É uma vulnerabilidade definida no sistema de um computador, presente normalmente em aplicações web que ativam ataques maliciosos ao inserirem scripts dentro de páginas web que são acedidas por outros utilizadores. Estes *scripts* permitem que os atacantes consigam escapar do controle efetuado durante o acesso.

— *Credential Stuffing*: Este tipo de ciberataque consiste no roubo de credenciais de acesso, normalmente de utilizadores e endereços de e-mail, bem como as senhas correspondentes, sendo que estas são utilizadas posteriormente para obter acesso não autorizado em aplicações web.

> "A esmagadora maioria dos incidentes de *ransomware* foram motivados financeiramente, com um conjunto limitado de ataques susceptíveis que foram politicamente motivados e intencionalmente destrutivos."(Relatório anual de segurança RNP, 2021).

Ransomware continuou sendo a ameaça cibernética mais significativa enfrentada pela maioria das organizações em 2021. A esmagadora maioria dos incidentes de *ransomware* foram motivados financeiramente, com um conjunto limitado de ataques que provavelmente foram politicamente motivados e intencionalmente destrutivos. Em 2020, aproximadamente 1.300 vítimas de *ransomware* tiveram seus dados expostos em sites de vazamento. Isso quase dobrou em 2021, com 2.435 vítimas expostas. O período de pandemia possibilitou que um número maior de vítimas fosse prejudicado no meio digital, já que havia um grande número de pessoas conectadas e trocando dados e mensagens ao mesmo tempo.

> "Em relação a tantos ataques, podemos utilizar a Tríade da CIA como base para estudo e comparação. "Confidencialidade, integridade e disponibilidade, também conhecido como a tríade da CIA, é um modelo geralmente desenhado para guiar informações políticas de segurança dentro de uma organização."(Sarker

et al. (2021).)

- Confidencialidade: É uma propriedade da política de segurança que normalmente se refere à proteção de informações e sistemas de partes não autorizadas. A ameaça de confidencialidade pode normalmente destinar-se a bancos de dados, servidores de aplicativos e administradores de sistema e podem ser considerados como “roubo de dados”.
- Integridade: Evitar qualquer tipo de destruição ou modificação de informações por terceiros não autorizados. A ameaça de integridade normalmente inclui ameaças relacionadas a finanças, como alterar dados financeiros, roubar dinheiro, redirecionar depósitos, e danos à confiabilidade da organização, podendo ser considerado como “alteração de dados”.
- Disponibilidade: Garantir o acesso de sistemas de informação ou ativos para uma parte autorizada ou entidade de forma confiável e tempestiva. A ameaça de disponibilidade normalmente inclui negação de serviço ou destruição, e pode ser considerado como “negação de acesso de dados".

Atualmente, a internet está sob constante ataque, não somente por *hackers*, ladrões de credenciais e espiões de empresas, governos que insistem em sua própria soberania digital, estão cada vez mais atacando a ideia desse ambiente digitalizado como um bem comum transnacional. O ciberespaço está se tornando uma zona de guerra em uma nova era de combate ideológico.

> “Os combatentes são primariamente pertencentes a dois grupos, as forças do ideal 'Internet livre', que favorece o fluxo irrestrito de informação, independente de fronteiras ou barreiras culturais, e o campo da “cibersoberania”, liderado pela Rússia e pela China, que exige maior controle do governo sobre a Internet e da informação. Para sustentar sua operação de censura massiva, o Grande *Firewall*” chinês (*China 's great Firewall*) emprega mais pessoas do que as forças armadas do país.” (Aftergood. (2017)).

Com a crescente onda de ataques envolvendo grandes governos e o crescente interesse em estudo e aplicação de IA, vemos como soluções atuais podem ser aplicadas para contornar o atual cenário instável e preocupante.

O relatório citado acima traz dados que mostram comparativos entre o custo dos danos advindos de invasões e roubo de dados de três tipos de empresas, as que não possuem uma IA de apoio, as que possuem uma IA parcialmente implementada no SCS e as que possuíam um SCS com uma IA integralmente implementada e funcional.

Etapa 4: Estabelecer as atuais limitações envolvendo o resultado e entradas de dados referente a modelos de Inteligência Artificial, assim como comparativos para novos estudos futuros com base em tecnologias emergentes na área da computação, não limitando-se somente à Inteligência artificial, mas complementando-a.

Conforme a tecnologia continua evoluindo, versões mais antigas de métodos, até então, revolucionários, se transformam rapidamente em tecnologias ultrapassadas. A Inteligência Artificial também se encaixa nos padrões da obsolescência, mas continua avançando a passos largos devido ao alto foco que a tecnologia vem recebendo desde o início dos anos 2000.

Métodos de IA são baseados em identificação de padrões, uma vez que os dados que formam o *baseline* (treinamento) são comprometidos, tornando mais fácil para os criminosos desenvolver uma IA com o objetivo de contornar as medidas de segurança estabelecidas por uma ferramenta de segurança.

> “IA em segurança e automação referem-se a habilitar as

> tecnologias segurança que aumentam ou substituem a intervenção humana na identificação e contenção de incidentes e tentativas de intrusão. Essas tecnologias dependem de IA, aprendizado de máquina, análise e orquestração de segurança automatizada." (*IBM security: Cost of a Data Breach Report* 2022).

Quer envolva sensores do mundo real ou manipulação posterior dos dados digitais resultantes, a alteração da entrada para um classificador existente é chamada "evasão". Outro tipo de vulnerabilidade ocorre se um invasor puder inserir dados adulterados no conjunto de treinamento, o que é conhecido como "envenenamento".

> "Na última década, os sistemas de aprendizagem mostraram uma surpreendente capacidade de classificar imagens, traduzir idiomas e realizar outras tarefas que antes pareciam exclusivamente humanas. No entanto, esses sistemas funcionam de forma opaca e às vezes cometem erros elementares, e essa fragilidade pode ser explorada intencionalmente para ameaçar a segurança". (Monroe. 2021).

Mesmo que os detalhes do sistema estejam ocultos, no entanto, os pesquisadores descobriram que ataques que funcionam contra um sistema frequentemente funcionam contra outros que têm uma estrutura interna diferente ou desconhecida. Esta observação inicialmente surpreendente reflete o poder dos sistemas de aprendizado profundo para encontrar padrões em dados.

> "Essa "transferibilidade" de ataques destaca o risco de um conjunto de treinamento comum como o *ImageNet*, que contém um enorme conjunto de imagens anotadas que é amplamente utilizado para o treinamento de sistemas de visão. Embora esse corpus de treinamento comum torna fácil comparar o desempenho de diferentes classificadores, torna todos eles potencialmente vulneráveis aos mesmos problemas, sejam maliciosos ou não." (Monroe. (2021)).

O problema principal envolvendo a IA é como a mesma é treinada, o algoritmo criado para identificar padrões deve ser testado previamente e a segurança ou credibilidade dos dados devem ser atestadas para maior desempenho na segurança. Como os métodos de inteligência artificial, muitas vezes, utilizam bases de dados públicas, ou até mesmo *inputs* de usuários comuns da internet, as mesmas podem sofrer com problemas envolvendo desinformação na base de dados, o que resulta em uma identificação de padrões com baixa acurácia.

Existem, também, problemas relacionados a limitações ainda encontradas na ferramenta. Problemas como os de alta acurácia, que, ao demonstrar um nível de acerto de 0,99 ainda deixam escapar uma quantidade que, para o mercado de segurança, ainda significam milhares de falhas de segurança. O alto número de acertos ainda se mostra incapaz de atender a todos os chamados e incidentes que uma empresa possa ter.

> "No cenário do mundo real, geralmente há um milhão de consultas DNS de comportamento normal em um único ambiente em um mês. A precisão de 0,99% do modelo AI significa que ainda há dez mil consultas que podem ser enganosas, como classificar padrões benignos como maliciosos." (Ho *et al.* (2020)).

A falta de explicação relativa aos resultados gerados a partir de modelos de aprendizagem de máquina é um dos principais problemas. Devido ao grau de acurácia ser obtido através de padrões encontrados por um algoritmo de aprendizado de máquina, a forma como alguns incidentes ou dados são vistos como

maliciosos ou não, pode ser considerada incorreta. Já que a maior parte dos dados gerados na saída do algoritmo de Inteligência artificial não possui uma explicação exata por trás, a não ser a detecção de um padrão. É difícil, ainda, ajustar algoritmos de IA ou ML para que os mesmos possam gerar resultados ainda mais precisos e específicos, de forma que sejam explicáveis e reaplicados para casos específicos ou parecidos.

> "Principalmente a saída da IA não pode ser explicada, o resultado da IA geralmente é inexplicável porque esse resultado é mais semelhante ao malicioso, não verdadeiro ou falso." (Ho *et al.* (2020)).

O problema do aprendizado online está centrado na questão da confiabilidade e segurança de uma base de dados. Em casos em que existe uma aferição manual relativa à saída de dados, esse problema pode ser contornado a depender do caso, visto que um ser humano seria capaz, em um ambiente de segurança, de atestar a integridade dos dados na saída. Quando são utilizadas grandes bases de dados conhecidas, dois problemas podem ser identificados. O conhecimento prévio dos dados dá uma vantagem aos cibercriminosos para melhor aperfeiçoarem suas técnicas e ferramentas. Outro caso está no conteúdo das bases de dados conhecidas, caso as mesmas sejam *open source*, se torna mais difícil garantir que os dados lá presentes sejam realmente dados válidos, seguros e reais.

> "No entanto, no mundo real da segurança, o padrão de *malware* geralmente muda dia a dia. Por exemplo, a implementação da análise DGA pode aprender 30 famílias diferentes hoje em dia, mas pode haver 100 famílias adicionais no futuro. Assim, a qualidade dos dados de treinamento desempenha um papel essencial na precisão da previsão. No entanto, a coleta de dados rotulados geralmente custam muito em segurança." (Ho *et al.* (2020)).

Por mais que a área de inteligência artificial tenha se tornado foco dentro do mercado, novas tecnologias vêm sendo desenvolvidas, essas com capacidade de mudar como a inteligência artificial será aplicada na área de segurança.

As idéias levantadas por Badhwar (2021) acerca dos avanços tecnológicos em pesquisas referentes à computação quântica preocupa tecnólogos de segurança, devido a suposta capacidade de quebrar diversos algoritmos de criptografia, atualmente vistos como extremamente difícil, trazendo incerteza quanto à confidencialidade e integridade dos dados confidenciais em repouso ou em trânsito.

Na computação clássica, o estado de um computador (convencional) pode ser descrito pela sequência de *bits* (de dois estados) (0 ou 1) ou configurações binárias do transistores dentro da CPU ou um dispositivo de armazenamento, por exemplo, um registro de dois *bits* em qualquer dado tempo pode armazenar qualquer um dos (2^2) quatro estados binários (00, 01, 10 ou 11), portanto, com N transistores, pode haver 2^N estados (binários) possíveis a qualquer momento.

Atualmente, a indústria depende da criptografia para proteger dados confidenciais. Dados criptografados com algoritmos de criptografia padrão da indústria são mais seguros, pois uma descriptografia maliciosa e não autorizada demoraria muito, devido ao grande período de tempo que levaria um ator malicioso para realizar um ataque de força bruta.

Precisamos de criptografia pós-quântica agora porque haverá um dia no futuro, quando os profissionais de segurança cibernética enfrentarão as consequências previstas pelo algoritmo de Shor, no qual os computadores quânticos terão a capacidade teórica de quebrar ou forçar toda a criptografia e algoritmos de *hash*, a depender do grau de dificuldade computacional de

fatoração ou de logaritmos discretos em nosso mundo pré-quântico. O algoritmo de Shor é um algoritmo quântico que encontra com alta probabilidade a ordem de um elemento.

> "Há uma necessidade muito válida para o uso de inteligência artificial (IA) ou aprendizado de máquina (ML) em cibersegurança. Essa necessidade será suprida com o uso de alguns algoritmos de aprendizado de máquina supervisionados e/ou não supervisionados em plataformas de computação que podem fornecer recursos reativos preditivos, cognitivos e automatizados com recursos de resposta e orquestração, sem exigir nenhuma propagação de dados ou estática assinaturas, entrada humana, inferência, análise e/ou programação." (Badhwar. (2021)).

4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

Com base nos estudos apresentados, concluímos que existem cinco aplicações chave da IA dentro da área de cibersegurança atualmente.

A primeira sendo a detecção de engenharia social e *spam*, englobando crimes como os de *phishing* e *credential stuffing*, uma vez que boa parte destes crimes são realizados via email, com indivíduos mal intencionados se passando por uma outra pessoa ou organização. Basicamente, ele usa análise de dados e aprendizado de máquina para examinar metadados, conteúdo, contexto e comportamento típico do usuário, criando um padrão comportamental do usuário. O uso de técnicas como *machine vision* para identificar imagens e ou logos falsos, também é uma forma de combater crimes que envolvam um alto grau de engenharia social.

Tendo o Gmail como referência, pois o mesmo tem experiência utilizando IA e filtros baseados em regras, tomamos que os filtros - baseados em regra - possuem a capacidade de bloquear o tipo mais óbvio de spam. Contudo, o aprendizado de máquina busca novos padrões que podem apontar o email como suspeito. Isso evidencia que com o uso dos algoritmos podemos ter mais parâmetros para categorizar um email como não seguro.

A segunda aplicação é a detecção de anomalias. A detecção de padrões sofisticados é um dos melhores usos do aprendizado de máquina para segurança cibernética. Os invasores cibernéticos geralmente se escondem nas redes e evitam a detecção criptografando suas comunicações, usando credenciais roubadas e excluindo ou modificando *logs*, mas um algoritmo de aprendizado de máquina projetado para sinalizar comportamentos incomuns ainda pode pegá-los antes que possam realizar uma de suas atividades contra o sistema. Isso atesta que os modelos também podem ser programados para observar ameaças internas, analisando a forma como os próprios colaboradores se comportam dentro da organização, podendo, assim, definir o mesmo como sendo uma pessoa autorizada dentro do ambiente, ou um atacante utilizando credenciais falsas, tudo a partir de uma análise comportamental do usuário dentro do ambiente. Além disso, o aprendizado de máquina pode se ajustar às mudanças ingerindo novos dados e adaptando-se a ambientes dinâmicos.

Já a terceira aplicação é a detecção avançada de *malware.* Tradicionalmente, a detecção de *malware* envolve o monitoramento e a pesquisa de tráfego de rede em busca de correspondências de assinatura, ou seja, semelhanças com indicadores conhecidos de comprometimento. O *deep learning*, no entanto, oferece uma oportunidade de analisar grandes quantidades de dados para fazer inferências sobre *malware* antes mesmo de ser aberto. À medida que o

malware evolui rapidamente, os modelos de aprendizado profundo têm a capacidade de acompanhar sua evolução. Muitas das tarefas que são resolvidas por redes neurais são apresentadas como problemas de classificação e uma das mais importantes tarefas de classificação no domínio da segurança estão a ser capazes de diferenciar entre *malware* e *goodware*.

A quarta aplicação é combater a fadiga de alerta. A IA na segurança cibernética pode ajudar a evitar que a equipe no centro de operações de segurança (SOC) fique sobrecarregada por alertas de incidentes ininterruptos. Dessa forma, o aprendizado de máquina pode intervir para fazer a triagem de alertas de baixo risco, assumir tarefas repetitivas e aumentar os níveis básicos de inteligência de ameaças que exigem intervenção humana. Profissionais e analistas de segurança permanecem no comando, mas seus recursos tecnológicos baseados em IA podem liberá-los para se concentrarem em tarefas de nível superior e na tomada de decisões. Os modelos de aprendizado de máquina sinalizam possíveis ataques, analistas humanos os revisam e esse *feedback* é incorporado de volta ao modelo. Os analistas de segurança podem ser mais produtivos e o algoritmo pode otimizar seu desempenho ao longo do tempo.

Destarte, a quinta aplicação se baseia em encontrar *Exploits* de Dia Zero (*Zero day exploit*). *Zero-days* estão entre as maiores preocupações que as equipes de segurança enfrentam na era da tecnologia moderna e das redes. Defender sistemas críticos de comprometimentos de dia zero é uma tarefa que a maioria das soluções de segurança herdadas geralmente não consegue realizar. Devido à complexidade de descobrir novas falhas de segurança e desenvolver códigos elaborados que possam explorá-las, esses ataques geralmente são realizados por grupos financiados ou experientes, como atores do estado-nação e APTs. Os métodos tradicionais de segurança de *endpoint*, como *software* antivírus ou soluções de gerenciamento de *patches*, não podem detectar ou impedir uma exploração de dia zero — é muito novo para ferramentas baseadas em assinaturas.

As arquiteturas de aprendizado profundo podem ser usadas para descobrir padrões ocultos ou latentes e tornar-se mais conscientes do contexto ao longo do tempo - ambos são úteis na identificação de vulnerabilidades ou atividades de dia zero. O processamento de linguagem natural pode vasculhar o código-fonte para sinalizar arquivos maliciosos. As “redes adversárias generativas”, que podem aprender a imitar qualquer distribuição de dados, também podem ser úteis para identificar vulnerabilidades complexas.

A IA também pode ser usada por profissionais de segurança de TI para aplicar boas práticas de segurança cibernética e reduzir a superfície de ataque em vez de perseguir constantemente atividades maliciosas. Ao mesmo tempo, invasores patrocinados pelo Estado, gangues cibernéticas criminosas e *hackers* ideológicos podem empregar essas mesmas técnicas de IA para derrotar as defesas e evitar a sua detecção.

Dessa forma, o estudo conclui que com o avanço da tecnologia, analisando os métodos estudados e a utilização de IA no cenário atual da cibersegurança, cada tipificação de crime precisa de um método específico para solucionar a mácula utilizando IA, como mostrado na tabela 1:

Tabela 1.Tabela de resultados comparativos.

Crimes	Métodos	Aplicação	Custo anual médio
Malware	Machine Learning AIS	Detecção avançada de Malware Combate a fadiga de alerta Detecção de anomalias	$2.4M
Ataques DoS	Machine Learning Fuzzy Logic AIS	Detecção de anomalias Combate a fadiga de alerta	$1.89M
Phishing	Deep Learning Fuzzy logic NLP Machine Vision	Detecção de engenharia social	$4.91M
Ransomware	Deep Learning NLP AIS	Detecção avançada de Malware Combate a fadiga de alerta Detecção de anomalias	$4.54M
SQL Injection	Machine Learning	Combate a fadiga de alerta Detecção de anomalias	$3.86M
Cros site scripting	Machine Learning	Combate a fadiga de alerta Detecção de anomalias	$1.03M
Zero day exploits	Deep Learning NLP	Detecção avançada de Malware Combate a fadiga de alerta Detecção de anomalias	$2.0M

Com base na tabela apresentada, o primeiro método é do *Artificial Immune Systems* (AIS) que atua dentro da rede da mesma forma que o sistema imunológico de um ser humano, células especializadas são criadas para combater patógenos. Com isso, é constatado que o AIS pode ser utilizado para detectar tipos de *ransomware* e *malware* que possam estar presentes no sistema. Usado juntamente de um IDS, podem ser utilizados para detecção de intrusões dentro da rede. Também, pode ser utilizado para checagem periódicas em relação a integridade do sistema como porta de comunicação de rede, aferição quanto à confiabilidade de transferência de arquivos na rede. Isso evidenciou um alto nível de adaptabilidade. Como foi citado anteriormente, como se assemelha a um sistema imunológico, esse método é vantajoso e bem aplicável por ter manter um alto grau de adaptabilidade em frente ao atual cenário, mantendo-se informado quanto a vulnerabilidades desconhecidas.

Outro método que obteve resultado no estudo e é comparado ao cérebro humano foi o de Redes Neurais que são bem utilizadas em técnicas de detecção de intrusão e de anomalias dentro da rede, por possuírem uma alta capacidade de abstração, conseguem delimitar o que é um comportamento normal dentro da rede através de uma análise dos dados que trafegam pela mesma. Ao serem introduzidos em um sistema de detecção e prevenção de intrusão, podem classificar juntamente do mesmo atividades suspeitas na rede. O método é muito utilizado também para monitorar as atividades do usuário o classificando como malicioso ou não.

O *Deep Learning*, algoritmo que apresenta um nível de complexidade maior comparado ao *Machine Learning*, mostrou na pesquisa que não necessita que seja efetuada uma extração de dados, pois consegue detectar correlações não lineares escondidas nos dados. Outrossim, suporta qualquer tipo de novos ficheiros e permite detectar ataques desconhecidos, sendo utilizado se torna uma vantagem ao nível de segurança. O mesmo ainda tem sido usado na detecção de *Ransomware*, pois a sua técnica permite que se aprenda a representação abstrata de dados. Outrossim, é possível criar um classificador que permite detectar possíveis ataques de *Ransomware*. No entanto, também pode ser efetuada uma correlação à informação, permitindo extrair uma melhor representação dos dados. Por fim, o *Deep Learning* possui a capacidade de aprender sobre recursos com múltiplos níveis de abstração, o que permite que o sistema aprenda funções complexas de mapeamento. É utilizado para classificar os URLs enquanto URLs de *Phishing* ou legítimos.

Outro método que foi estudado é o *Natural Language Processing* (NLP) que

analisa a forma como as palavras foram utilizadas e tem a capacidade de derivar os conjuntos de recursos dos dados de texto não estruturados. Foi constatado que são bem utilizadas para analisar artefatos no desenvolvimento de *software*, focando no compromisso das mensagens e no defeito dos relatórios. Com o uso das informações geradas, é possível criar medidas relativas ao processo do *software*, sua qualidade e sua segurança. Também utilizado para combate de *ransomware*, já que pode analisar componentes textuais do programa, assim como verificar comportamentos diferentes no sistema, o que ajuda no seu combate. Por trabalhar com análise contextual de palavras e seu uso semântico, é possível aplicar métodos de NLP para aferir quanto à legitimidade de emails que possam ser tentativas de *phishing*. Com isso, o uso do método permite analisar os diferentes contextos que se aplicam para um golpe, classificando a ameaça antes que se torne um problema.

Machine Vision tem como seu maior uso o reconhecimento facial, já que a mesma opera como com a análise e classificação de imagens digitais. O reconhecimento facial pode ser utilizado tanto para autenticação em momentos de acesso a sistemas ou dados em uma organização, assim como para monitoramento de locais físicos dentro de uma empresa ou organização. Outro uso é para combater crimes de *phishing*. Como boa parte destes crimes são cometidos via email ou mensagens de telefone, os criminosos muitas vezes fazem uso de logotipos de bancos, empresas ou até governos e com o uso de *machine vision*, é possível aferir quanto a confiabilidade de um email a partir de imagens não oficiais identificadas no mesmo ou previamente classificadas como maliciosas.

Em ataques de *Denial of Service*, temos uma grande quantidade de requerimentos chegando ao servidor com o objetivo de indisponibilizar recursos da rede para requisições legítimas, com o intuito de explorar vulnerabilidades no sistema. Com o uso de lógica *fuzzy*, é possível classificar possíveis ataques de rede, pois a lógica do algoritmo se baseia em produzir um resultado verdadeiro entre zero e um, criando alertas de possíveis ataques ao classificar comportamentos relacionados à requisição dentro da rede. São gerados alertas que são então classificados por profissionais de rede, que possuem o veredito final. É bem utilizado para dar apoio aos profissionais de rede para prevenir ataques DoS. A mesma pode ser utilizada para gerar relatórios quanto ao comportamento de usuários de rede, contudo ainda necessita de uma validação final de um profissional de rede, uma vez que as ações de usuários podem variar muito dentro da rede.

Ainda no contexto de invasões e métodos e analisando a tabela 1, ataques de *zero-day-exploits* se mantêm como os ataques mais perigosos. Graças a ausência de informações relativas ao método de invasão descoberto/criado. Como explicado no artigo *Machine Learning Goes Dark And Deep To Find Zero-Day Exploits Before Day Zero Day*, técnicas de machine learning são empregadas para vasculhar a web, em locais como fóruns de atividade maliciosa ou outros endereços suspeitos (principalmente os localizados na *dark web*), atrás de menções à novas técnicas de invasão. Como visto anteriormente, o mercado de *Ransomware as a service* (RaaS) vem crescendo muito nos últimos anos, e com o aumento de usuários maliciosos utilizando a rede para fins ilícitos, também crescem e evoluem os meios por onde os mesmos se conectam. Vasculhar a web atrás de ataques nunca antes vistos é uma forma de prevenir que consequências desastrosas venham a ocorrer caso as medidas necessárias não sejam tomadas antecipadamente.

5 CONSIDERAÇÕES FINAIS

A IA tem o potencial de melhorar com o

tempo utilizando *Machine Learning* e *Deep Learning*. Foi observado que o *Machine Learning* tem crescido rapidamente para avançar a tecnologia muito além da nossa capacidade de pensamento. Podemos notar que as tecnologias de aprendizagem são usadas para análise de dados e reconhecimento de padrões com um mínimo de interação humana. Tivemos a possibilidade de constatar que combater os crescentes ataques cibernéticos e ameaças de *malware* ocasionou que muitas organizações de TI agora estão implantando inteligência artificial em seus negócios.

Os dados coletados diariamente de IA interna e de terceiros mostram que há o aumento exponencial de sua geração. Conforme evidenciado por violações de dados mais frequentes, esses grandes *pools* de dados são alvos muito procurados por cibercriminosos - em *firewalls* e produtos *antimalware*. Isso evidenciou que está se tornando cada vez mais difícil para os analistas cibernéticos monitorar a sofisticação das ameaças.

Foi evidenciado que os *hackers* estão usando métodos cada vez mais sofisticados para violar a segurança de TI, coletar informações e lançar ataques. A utilidade do aprendizado de máquina e da IA também beneficia os cibercriminosos. Algumas ameaças estão evoluindo mais rapidamente que outras ao se alinharem ao uso de *machine learning.*

Ao analisar os artigos, concluímos que os *hackers* podem usar o aprendizado de máquina para alterar de forma criativa os e-mails de *phishing* para que eles não apareçam em listas de e-mail em massa e sejam otimizados para incentivar o engajamento e os cliques. Eles vão além do texto do e-mail. Os *hackers* usam a IA para produzir imagens realistas, falsas identidades em mídias sociais e outros conteúdos para dar à interação a melhor legitimidade possível. Outrossim, foi identificado no estudo que criminosos usam IA e aprendizado de máquina para melhorar suas habilidades de adivinhação de senha. É evidente que os mecanismos de adivinhação de senhas agora possuem técnicas mais sofisticadas com base na frequência e nas taxas de sucesso das tentativas de *hackers* criminosos. Entretanto, a capacidade de hackear *hashes* roubados também está melhorando, pois os criminosos estão criando dicionários melhores.

Os cibercriminosos usam a tática da engenharia social para enganar e convencer as vítimas a divulgar detalhes confidenciais ou realizar uma ação específica. Ao tornar mais simples e rápido coletar dados sobre empresas, funcionários e parceiros, a IA e o aprendizado de máquina fazem uso das ações dos criminosos. Em outras palavras, ataques baseados em engenharia social são fortalecidos por inteligência artificial e aprendizado de máquina.

Mesmo em seu estado atual a inteligência artificial, ainda que muito utilizada por grandes organizações para gerenciamento de segurança, enfrenta diversos problemas acerca de suas limitações e de seu uso com fins maliciosos. É uma ferramenta poderosa que marcou o início da revolução 4.0 (apesar de ainda ser muito rudimentar no início da mesma) e continua se desenvolvendo conforme a tecnologia avança.

Este trabalho teve como foco explicar o conceito de cibersegurança e sua diferença de “segurança da informação”, inteligência artificial e seus métodos (aplicados dentro do mercado da cibersegurança), assim como as limitações da mesma dentro da área de cibersegurança, sendo aplicada para garantir ou não a segurança de um sistema, foram categorizados os principais tipos de ataques a sistemas de informação juntamente de perspectivas relacionadas ao atual estado geopolítico.

Foi evidenciado que determinados métodos de IA aplicados ao cenário de cibersegurança possuem maior efetividade do que outros métodos quando aplicados

para prevenir determinados tipos de crimes.

Ao introduzir o conceito de computação quântica, é abordada novamente a questão e limitações referentes à inteligência artificial e como a mesma estará sujeita a novas melhorias ao incorporar conceitos e ferramentas advindas da computação quântica.

Com isso, chegamos à conclusão de que o mundo da cibersegurança mudou muito nas últimas décadas e há formas diferentes de analisar e identificar qual a melhor forma de IA para resolver problemas. Obtivemos resultados nas pesquisas que nos levaram a entender as diferentes formas e estratégias utilizadas para cada tipo de crime e como preveni-lo. Outrossim, a conclusão que foi atingida é que a tecnologia de segurança de dados avança de forma desproporcional em relação aos criminosos e mecanismos futuro do ciberespaço e como ele impactará na vida da sociedade está no início, mas pode atingir novos níveis e avançar a forma como lidamos com a rede conectada e sua cibersegurança.

AGRADECIMENTOS

Este trabalho é dedicado aos meus pais, Jefferson, o vento em minha vela e as estrelas que me guiam, Julieta. Minha família e professores, as partes me ajudaram a montar meu navio. Dedicado ao Vitor, a âncora que me mantém firme e à Vitória, o farol que me guiou em meio a escuridão. Também dedico esse trabalho aos vários amigos e colegas que me acompanharam ao longo dessa jornada em águas desconhecidas, tanto os que foram quanto os que ficaram. Um novo porto aguarda, para uma nova jornada.

REFERÊNCIAS

- AFTERGOOD, S. *Cybersecurity: The cold war online.* Nature. **2017**.
- ANWAR, S.; MOHAMAD ZAIN, J.; ZOLKIPLI, M.F.; INAYAT, Z.; KHAN, S.; ANTHONY, B.; CHANG, V. *From Intrusion Detection to an Intrusion Response System: Fundamentals, Requirements, and Future Directions*. Algorithms **2017**.
- BADHWAR, R. *The CISO's Next Frontier.* 1a edição. Estados Unidos. CISCO **2021**.
- BLACKHAT. *DNS as a pathway for infiltration and exfiltration.* **2018.** https://www.blackhat.com/sponsor-posts/03192018.html 7/10/2022.
- COMERLATO, C.; CARVALHO, L.; SCHIRRU, R.; CARLOS, J.; MEDEIROS, J. *Sistemas críticos quanto à segurança utilizando redes neurais artificiais.* COPPE / UFRJ. **2022**.
- (HO, T; CHEN, W.; HUANG, C.; *The Burden of Artificial Intelligence on Internal Security Detection.* IEEE 17th International Conference on Smart Communities: Improving Quality of Life Using ICT, IoT and AI (HONET). **2020.**
- IBM security: Cost of a Data Breach Report 2022. **2022.**
- MONROE, D. *Deceiving AI.* Communications of the ACM. **2021**
- MURNANE, K. *Machine learning goes dark and deep to find zero-day exploits before day zero.* Forbes. **2016** https://www.forbes.com/sites/kevinmurnane/2016/08/08/machine-learning-goes-dark-and-deep-to-find-zero-day-exploits-before-day-zero/?sh=7955afe2417b 13/09/2022.
- ONGSULEE, P. *Artificial Intelligence, Machine Learning and Deep Learning.* Fifteenth International Conference on ICT and Knowledge Engineering. **2017.**
- REN, S.Q.; TAN, B.H.M.; SUNDARAM, S.; WANG, T.; NG, Y.; CHANG, V.; AUNG, K.M.M. *Secure searching on cloud storage enhanced by homomorphic indexing.* Future Gener. Comput. Syst. **2016**.
- 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**).
- SOLMS, R.; NIEKERK J. *From information security to cyber security.* Elsevier. **2013.**
- TANENBAUM, A.S.; BOS, H. *Modern Operating Systems*. 4a edição. Estados Unidos. Pearson Education, Inc. **2007.**
- TANENBAUM, A.S.; WETHERALL, D.J. *Computer Networks.* 5a edição. Estados Unidos. Pearson Education, Inc. **2011.**