REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

ANÁLISE TÉCNICA DOS RISCOS DA INTELIGÊNCIA ARTIFICIAL NOS CIBERATAQUES

TECHNICAL ANALYSIS OF THE RISKS OF ARTIFICIAL INTELLIGENCE IN CYBERATTACKS

ANÁLISIS TÉCNICO DE LOS RIESGOS DE LA INTELIGENCIA ARTIFICIAL EN CIBERATAQUES

Gabriel Laroche Borba[1], Luis Felipe Araujo Mota[2]

e44182
https://doi.org/10.63026/acertte.v4i4.182

PUBLICADO: 06/2024

RESUMO

A crescente integração da Inteligência Artificial, conhecida como IA, em várias áreas como medicina e indústria, traz avanços notáveis. Por outro lado, acarreta desafios, principalmente, na segurança cibernética. A utilização da IA em ciberataques e a evolução das ameaças são preocupações constantes e críticas. A análise técnica dos riscos da IA nos ciberataques é um campo crucial, visando entender as implicações dessa interseção. O estudo realizado teve como metodologia uma pesquisa bibliográfica, com base no tema proposto. A partir dessas fontes foi possível desenvolver essa investigação que analisa potenciais ameaças, táticas dos atacantes e estratégias de defesa. Ao abordar fontes e estudos relevantes, o trabalho busca fornecer uma compreensão abrangente dos desafios emergentes e das técnicas que cibercriminosos podem usar com o apoio da IA.

PALAVRAS-CHAVE: Inteligência artificial. Riscos. Segurança. Desafios. Ameaças cibernéticas.

ABSTRACT

The increasing integration of Artificial Intelligence, or AI, in various areas such as medicine and industry, brings remarkable advancements. On the other hand, it entails challenges, especially in cybersecurity. The use of AI in cyberattacks and the evolution of threats are constant and critical concerns. Technical analysis of the risks of AI in cyberattacks is a crucial field to understand the implications of this intersection. The study carried out had a bibliographic research methodology based on the proposed theme. From these sources, it was possible to develop this investigation that analyzes potential threats, attackers' tactics, and defense strategies. By addressing relevant sources and studies, the work seeks to provide a comprehensive understanding of emerging challenges and techniques that cybercriminals can use with the support of AI.

KEYWORDS: Artificial intelligence. Risks. Safety. Challenges. Cyber threats.

RESUMEN

La creciente integración de la Inteligencia Artificial, conocida como IA, en diversas áreas como la medicina y la industria, trae consigo avances notables. Por otro lado, conlleva retos, especialmente en ciberseguridad. El uso de la IA en los ciberataques y la evolución de las amenazas son preocupaciones constantes y críticas. El análisis técnico de los riesgos de la IA en los ciberataques es un campo crucial para entender las implicaciones de esta intersección. El estudio realizado tuvo como metodología una investigación bibliográfica, basada en el tema propuesto. A partir de estas fuentes, fue posible desarrollar esta investigación que analiza las amenazas potenciales, las tácticas de los atacantes y las estrategias de defensa. Al abordar fuentes y estudios relevantes, el trabajo busca proporcionar una comprensión integral de los desafíos y técnicas emergentes que los ciberdelincuentes pueden usar con el apoyo de la IA.

PALABRAS CLAVE: *Inteligencia artificial. Riesgos. Seguridad. Desafíos. Amenazas cibernéticas.*

[1] Graduando em Sistemas de Informação pela UFPE – Campus de Recife.
[2] Graduando em Sistemas de Informação pela UFPE – Campus de Recife.

REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

1 INTRODUÇÃO

Segundo Santos (2020, p.1), a inteligência artificial (IA) consiste em um sistema composto por software que trabalha em conjunto para simular uma inteligência semelhante à humana. Esse sistema é capaz de aprender diversos tipos de informações e tomar decisões com base em padrões extraídos de vastos bancos de dados. Assim como a capacidade adaptativa humana, a IA pode ajustar-se conforme os desafios que enfrenta. O objetivo das inteligências artificiais é aumentar a inteligência e a utilidade das máquinas. Portanto, o desenvolvimento desses sistemas não busca substituir completamente a capacidade humana de tomar decisões, mas complementá-la em situações específicas.

De acordo com De Almeida Souza & De Moraes (2021), a inteligência artificial (IA) tem avançado consideravelmente e está presente em diversas áreas, incluindo a cibersegurança. Contudo, é evidente que a IA também pode ser utilizada para ataques cibernéticos em larga escala, ampliando as possibilidades para os criminosos virtuais. Diante disso, percebe-se que a rápida evolução da IA traz consigo uma série de riscos e desafios que precisam ser compreendidos e enfrentados de maneira adequada.

O objetivo deste artigo é explorar os riscos associados à IA, com foco na segurança, futuro do trabalho e ameaças cibernéticas. Além disso, examinar os principais riscos e desafios da IA identificados na literatura científica, abordar aspectos relacionados à segurança, destacando as preocupações com sistemas autônomos, redes neurais e a potencial manipulação de algoritmos e dados. Ademais, serão discutidos os impactos da IA no mercado de trabalho e as ameaças cibernéticas associadas a essa tecnologia.

1.1 Contextualização

Antes de iniciar a apresentação dos resultados da pesquisa bibliográfica a respeito desse tema, é necessário apresentar o contexto no qual essa pesquisa está inserida. Assim, o leitor terá consciência da questão que está sendo explorada e o aporte teórico necessário para entender o tema que será abordado neste trabalho.

A evolução histórica da IA, conforme descrita por Samoili et al. (2020), é marcada por marcos significativos ao longo das décadas. Desde os anos 1940, quando Alan Turing estabeleceu os fundamentos da IA, até a última década, testemunhamos um progresso extraordinário. Durante as décadas de 1950 e 1960, foram desenvolvidos algoritmos e linguagens de programação, como o LISP, que impulsionaram a pesquisa em IA. Nos anos 1970 e 1980, a IA se baseou em sistemas especialistas, enquanto os anos 1990 viram a popularização do aprendizado de máquina com o advento de grandes bases de dados e o aumento do poder computacional.

A virada do século testemunhou avanços notáveis em mineração de dados, reconhecimento de padrões e redes neurais, culminando na explosão do aprendizado profundo na década de 2010, com tecnologias como o GPT-2, OpenAI, AphaGO, etc. Esse desenvolvimento histórico resultou na

REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

aplicação generalizada da IA em setores diversos, como indústria, finanças, saúde e transporte, transformando profundamente a forma como interagimos com a tecnologia e o mundo ao nosso redor.

1.2 Contextualização Teórica

Antes de relacionar a Inteligência artificial com a Cibersegurança, é necessário que sejam definidos alguns conceitos importantes para uma melhor compreensão desta pesquisa.

Segundo Nunes (2012), a cibersegurança envolve práticas adotadas por organizações e Estados para gerenciar riscos de segurança, visando proteger a confidencialidade, integridade e disponibilidade de dados e recursos no ciberespaço. Este conceito abrange diretrizes, políticas e a implementação de salvaguardas, tecnologias, ferramentas e treinamentos, garantindo um ambiente cibernético seguro tanto para o estado do ambiente quanto para seus usuários.

De acordo com De Almeida Souza & De Moraes (2021, p. 29), é reconhecido que o conceito de cibersegurança existia antes da implementação da inteligência artificial (IA) nesse contexto, porém é indiscutível que a aplicação da IA se tornou essencial para resolver a maioria dos desafios.

Segundo Teles (2015), enquanto a IA tem contribuído para melhorar a cibersegurança, também tem sido empregada em ataques. Os problemas de segurança são geralmente provocados por indivíduos mal-intencionados visando benefícios próprios. Assim, a IA se torna uma ferramenta ágil e poderosa para realizar ataques cibernéticos mais rápidos e sofisticados, explorando minuciosamente as vulnerabilidades de sistemas frágeis a um custo relativamente baixo. As consequências desses ataques são imprevisíveis e podem resultar em danos significativos e inesperados para empresas e indivíduos nos próximos anos.

Neste contexto, ciberataques referem-se à utilização de códigos maliciosos para manipular sistemas computacionais e suas redes, conforme definido por Graça (2014). Quando um hacker emprega inteligência artificial para propósitos maliciosos, isso é conhecido como IA adversária. Embora o aprendizado de máquina desempenhe um papel crucial no combate a ameaças cibernéticas, seus algoritmos operam com base em parâmetros específicos predefinidos, o que implica que ameaças não previamente identificadas podem passar despercebidas. Apesar da capacidade impressionante da IA para processar informações, ela é programada por humanos, o que implica a existência de vulnerabilidades (LAZIĆ, 2019).

Ainda como contextualização, pode-se citar um tipo de vírus que facilmente poderia ser automatizado com a IA, que é o ransomware, além também de mostrar o funcionamento de um vírus em uma aplicação de ciberataque. De acordo com Messias (2015), vírus são programas que penetram no computador que utilizamos sem a nossa autorização e executam ações que não solicitamos. Normalmente, essas ações prejudicam o equipamento ou seu desempenho.

Como exposto por Britto & Freitas (2017, p. 5),

REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

> [...] Ransomware é um vírus de resgate, que embaralha os arquivos do computador, impedindo seu funcionamento normal. Para restaurar os arquivos e recuperar o sistema, a vítima precisa fazer um pagamento. De acordo com Aliaksandr Trafimchuk, da empresa israelense de desenvolvimento de softwares Check Point, o vírus utilizado na prática criminosa, denominado Petya, é uma espécie do gênero Ransomware que apareceu pela primeira vez na cena do crime cibernético no início de 2016. Embora o Petya não tenha uma taxa de infecção impressionante como outros recursos de armazenamento, como CryptoWall ou TeslaCrypt, foi imediatamente marcado como o próximo passo na evolução do Ransomware.

Ainda, segundo esses autores, o Petya pode ser melhor descrito como um sistema de Ransomware de três estágios, onde cada etapa possui sua própria funcionalidade dedicada:

Figura 1: Etapas de infecção do Petya

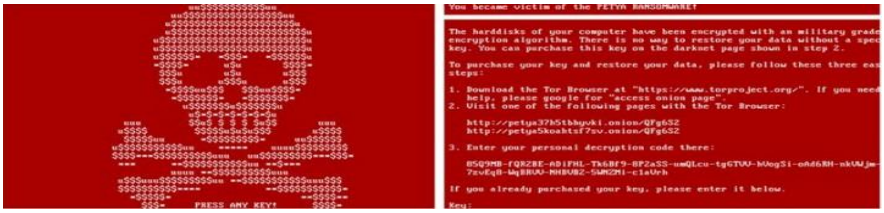
ETAPA	COMANDO	ROTINA
0	MBR Overwrite	Substituir o registro de inicialização mestre do disco rígido e implanta o boot-loader personalizado
1	MFT Encryption	Usar o carregador de inicialização personalizado introduzido no estágio 0 para criptografar todos os registros da Mestre-Tabela de arquivos (MFT), o que torna o sistema de arquivos completamente ilegível.
2	Ransom Demand	Exibe o logotipo da Petya e a nota de resgate detalhando o que deve ser feito para descriptografar o disco rígido.

Fonte: TRAFIMCHUK (2016) Decrypting the Petya Ransomware

Segundo Britto & Freitas (2017), no estágio inicial, o Petya gera um vetor de inicialização de 8 bytes e uma chave aleatória de 16 bytes, que é expandida para uma chave de criptografia de 32 bytes por meio de um algoritmo simples. Parar o processo de execução do Petya antes que o sistema reinicie completamente permite desfazer as operações realizadas pelo vírus e restaurar a unidade ao seu estado anterior.

Conforme descrito pelos autores, na fase inicial, o Petya carrega o código de inicialização implantado na fase 0. Se essa verificação for bem-sucedida, o Petya começa a enumerar os registros do Módulo de Arquivos Mestre (MFT) da unidade, ocultando-se sob a aparência de um aplicativo legítimo de reparo de arquivos. Além disso, a fase 2 da execução do vírus Petya ocorre após a codificação do MBR e do MFT, quando o computador é reiniciado e mostra o logo do crânio assustador, conforme detalhado pelos autores.

Figura 2: Estágio 2 "Demanda de resgate"



Fonte: TRAFIMCHUK (2016) Decrypting the Petya Ransomware

REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

Depois que a vítima pressiona qualquer tecla, a nota de resgate que contém as instruções de pagamento e descriptografia é exibida. E, com o advento da máquina de redes neurais e as IAs, ficaria bem mais fácil implementar esse tipo de vírus, o que traz uma preocupação ainda maior para os utilizadores de computador que se preocupam com a segurança deles.

1.3 Contextualização do Problema

Apesar da IA evoluir para a facilitação e o avanço tecnológico da humanidade, ao contextualizar historicamente a IA no escopo que está sendo explorado, deve-se levantar a questão da importância da conscientização sobre os possíveis riscos que essa tecnologia trás para uso tanto de bens civis como também de hackers, que utilizam os mesmos para bombardear sistemas e realizar ataques complexos em alvos aleatórios.

Segundo Wirkuttis (2017, p. 32)

> Embora a conscientização das empresas sobre as ameaças cibernéticas tenha aumentado, e muito dinheiro tenha sido investido para combater os crimes cibernéticos, a capacidade das organizações de proteger totalmente seus ativos virtuais ainda é desconhecida.

Sabendo disso, surge a dúvida do motivo pelo qual a IA é utilizada nos ataques cibernéticos. Ainda, de acordo com Wirkuttis (2017, p. 29)

> [...] a IA consegue reutilizar os padrões de ameaças para identificar novos Incidentes, o que reduz tempo e custos. As ameaças cibernéticas são basicamente causadas por ações maliciosas que podem ocorrer por motivos econômicos, políticos ou militares.

Conforme alertam alguns especialistas em segurança da informação, estamos enfrentando uma pandemia de ciberataques, prevendo-se que em 2021 esses ataques possam resultar em prejuízos financeiros globais de até 6 trilhões de dólares. Este aumento é impulsionado pelos consideráveis investimentos em inteligência artificial por parte de muitas empresas. Estima-se que o setor de segurança cibernética cresça aproximadamente 12,5% nos próximos cinco anos, projetando um mercado que poderá movimentar até US$403 bilhões até 2027, conforme De Almeida Souza & De Moraes (2021).

Por consequência, sabendo que esse assunto é alarmante para os próximos anos, artigo foca em trazer os principais ataques, ameaças, tendências e uma análise dos instrumentos tecnológicos, baseados em IA, que irão aumentar e provavelmente serão utilizados por cibercriminosos para realizar ataques mais complexos e que necessitam de maior segurança para serem interrompidos.

2 METODOLOGIA

Este estudo foca na "Análise Técnica dos Riscos da IA nos Ciberataques", fundamentando-se em uma revisão criteriosa da literatura científica e técnica relacionada à inteligência artificial (IA), seus avanços tecnológicos e os perigos específicos associados à sua aplicação em ataques

REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

cibernéticos. A metodologia adotada teve como objetivo assegurar a credibilidade e relevância dos dados, selecionando artigos que exploram profundamente esses temas. Os critérios de seleção incluíram a profundidade na exploração dos conceitos de IA aplicados à cibersegurança, a contribuição para o entendimento dos riscos emergentes e a clareza na identificação de tendências, desafios técnicos e estratégias defensivas potenciais. A partir dessa revisão crítica, foi possível obter uma compreensão abrangente dos riscos associados e das possíveis táticas que os adversários podem adotar com o suporte da IA proporcionando uma base sólida para a análise detalhada realizada neste estudo.

Para obter uma compreensão abrangente e atualizada do tema, iniciamos com a etapa fundamental do "Estado da Arte". Esta metodologia de pesquisa serve como um alicerce sólido onde consolidamos o conhecimento teórico existente e o integramos à nossa abordagem metodológica para aprofundar a compreensão e fornecer evidências robustas relacionadas ao tema em questão. O "Estado da Arte" envolve uma investigação detalhada das contribuições mais recentes na interseção entre IA e cibersegurança, incluindo tendências emergentes e debates pertinentes. É crucial realizar uma análise crítica e abrangente das fontes de pesquisa mais atuais e significativas para estabelecer uma base sólida para as investigações subsequentes.

2.1 Ataques Cibernéticos

Existem diversos tipos de ataques cibernéticos que são aplicados pelos criminosos que se utilizam do uso da IA, nesse tópico temos como objetivo descrever sobre alguns desses tipos, a fim de entender como esses criminosos se utilizam da IA para o malefício das pessoas. Nesta pesquisa vamos citar o ataque Dos/DDos e o chatbot, incluindo a sua definição, como o ataque funciona e algumas notícias relacionadas das consequências desses ataques. Além disso, irá ser citada a dinâmica dos ataques dentro de redes neurais e a diferença entre um ataque com IA e outro sem essa tecnologia.

2.1.1 Ataques DoS/DDoS

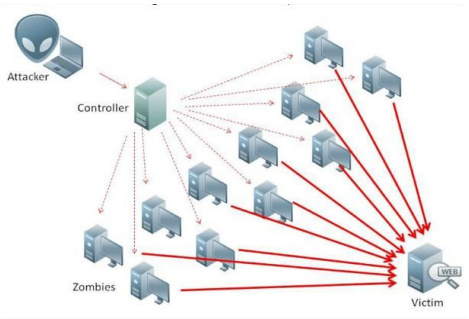
Segundo Jordão (2014, p.3), um ataque DDoS “[...] atinge sua meta excedendo os limites do servidor. Para tal façanha, os responsáveis pelo ataque criam programas maliciosos que são instalados em diversas máquinas, as quais realizarão múltiplos acessos simultâneos ao site em questão”.

Segundo, então, esta afirmação, é possível concluir que um ataque DDoS (Distributed Denial of Service) consiste em sobrecarregar um servidor ou rede com tráfego excessivo, tornando-o inacessível para usuários legítimos. A IA está sendo utilizada para tornar esses ataques mais sofisticados. Isso inclui a criação de botnets mais inteligentes que podem se adaptar ao ambiente de destino, a identificação de vulnerabilidades nos sistemas de defesa contra DDoS, o aprimoramento de táticas de engenharia social e até mesmo a tentativa de contornar sistemas de mitigação de defesas baseados em IA. A IA está transformando a paisagem dos ataques DDoS, tornando-os mais

REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

desafiadores de se defender e potencialmente mais prejudiciais, destacando a necessidade de constante inovação na segurança cibernética.

Figura 3: Como funciona um Ataque DoS



Fonte: Imagem da internet

Segundo Jordão (2014), para combater ataques como esse, profissionais geralmente precisam configurar filtros em dispositivos que levam ao site alvo, determinando quais IPs podem acessá-lo e quais são considerados perigosos. Uma alternativa eficaz é recorrer a empresas especializadas como a Akamai, que utilizam uma rede global de computadores para conter o ataque. Esses computadores distribuídos geograficamente dividem a tarefa de combater as máquinas zumbis, tornando a abordagem mais eficiente, já que cada computador de defesa lida com um número limitado de máquinas atacantes.

2.1.2 Chatbot

Os bots, derivados da palavra "robot", são programas criados para replicar ações humanas de forma repetitiva e simular interações entre humanos e computadores. Inicialmente mais simples e com interação limitada, os bots modernos realizam uma análise antecipada das necessidades dos usuários, proporcionando interações cada vez mais naturais e eficientes (DE CARVALHO JÚNIOR, et al., 2018).

Segundo Mowbray, Pearson & Shen (2012), os bots são programas de software que operam com base em um conjunto de instruções pré-definidas, em vez de serem controlados diretamente por um usuário humano. Eles são desenvolvidos para interagir de maneira transparente com seres humanos, podendo se adaptar ao contexto, responder perguntas, fornecer informações ou oferecer sugestões. Os chatbots, ou chatterbots (robôs de conversação), surgiram nesse contexto, com destaque para ELIZA, criada por Joseph Weizenbaum em 1976, sendo reconhecida como a primeira aplicação desse tipo.

Existem diversos algoritmos de chatbots utilizados por cibercriminosos para realizar ataques de injeção de IA, como A.L.I.C.E. e a API Pandorabots. A.L.I.C.E., ou Alicebot (Artificial Linguistic Internet Computer Entity), é um chatterbot de IA que utiliza linguagem natural. O software utilizado para criar A.L.I.C.E. é de código aberto, diferentemente de outros programas comerciais de chatbots

REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

que podem ser bastante caros. O motor Alicebot e a AIML (Artificial Intelligence Markup Language) estão disponíveis gratuitamente sob os termos da GNU General Public License, utilizada por projetos como GNU/Linux.

A API Pandorabots, também mencionada pelos mesmos autores, permite a integração de serviços de hospedagem de bots e processamento de linguagem natural em aplicações personalizadas. Nessa plataforma, os bots são baseados na AIML, a mesma linguagem utilizada pelo projeto A.L.I.C.E. APIs, ou Interfaces de Programação de Aplicativos, facilitam a criação de aplicativos a partir de um esquema pré-programado pelo desenvolvedor.A seguir, apresentamos um diálogo entre um usuário humano e um chatbot de uma empresa, especificamente desenvolvido para simular o funcionamento de uma aplicação inteligente. Este exemplo fictício ilustra a interação entre um cliente e um atendente virtual de uma clínica médica, onde o usuário solicita ao chatbot que agende uma consulta.

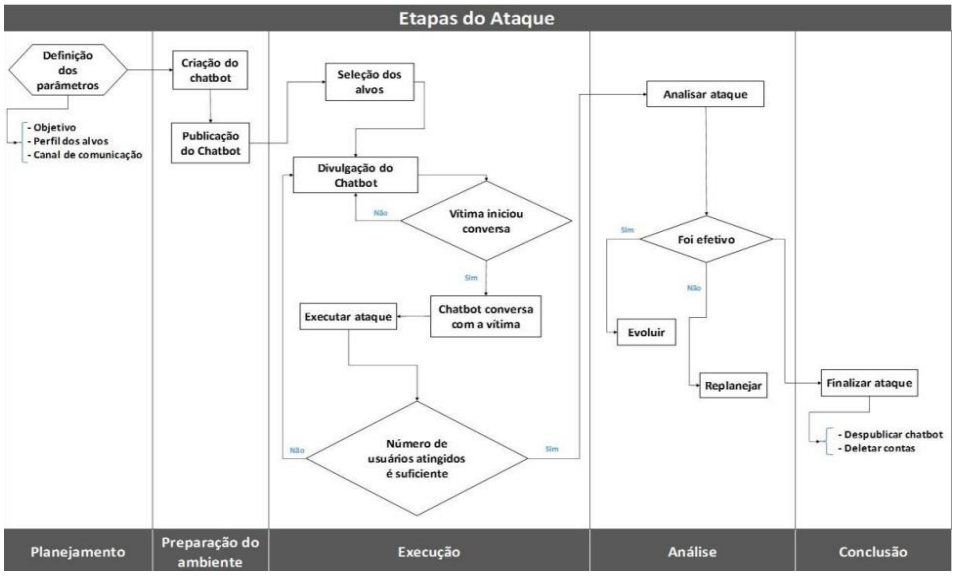
Figura 4: Conversa entre um humano e o chatbot com WhatsApp e Dialogflow.

Fonte: Mowbray, Pearson & Shen (2012)

De acordo com De Almeida Souza & De Moraes (2021), os ciberataques utilizando inteligência artificial podem amplificar a eficácia da engenharia social praticada por hackers. Nunes (2017) menciona um estudo de caso realizado pela empresa Take Blip, que avaliou o impacto bem-sucedido de um ataque de engenharia social utilizando chatbots.

REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

Figura 5: Etapa do estudo de caso pela empresa Take Blip

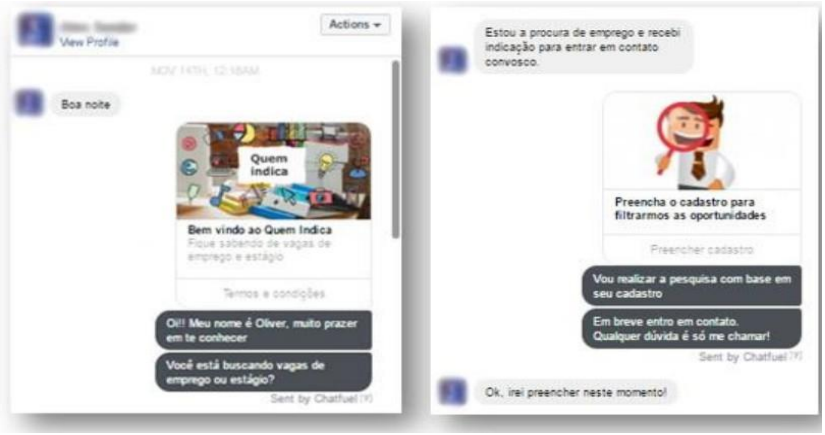


Fonte: NUNES, 2017.

Nessa representação, primeiro irão ser definidos os principais parâmetros (Objetivo, Perfil dos ativos, Canal de comunicação). Depois, irá ser criado o chatbot e o conseguinte deploy dele. Ele irá procurar vítimas para executar os ataques, até atingir uma quantidade ideal de vítimas. Caso o número de usuários atingidos não seja o suficiente, irá ocorrer mais divulgação. Após analisar ataques, a própria IA aplicada no ataque irá evoluir com os dados, ou senão se replanejar para realizar um ataque mais efetivo. Após isso, automaticamente, contas irão ser deletadas e o ataque irá ser finalizado.

Conforme De Almeida Souza & De Moraes (2021), desenvolveu-se um chatbot dedicado a oportunidades de emprego em uma página do Facebook. Ao enviar uma mensagem, os usuários iniciavam uma interação com o chatbot, conforme ilustrado na figura abaixo:

Figura 6: Estudo de caso interação do usuário

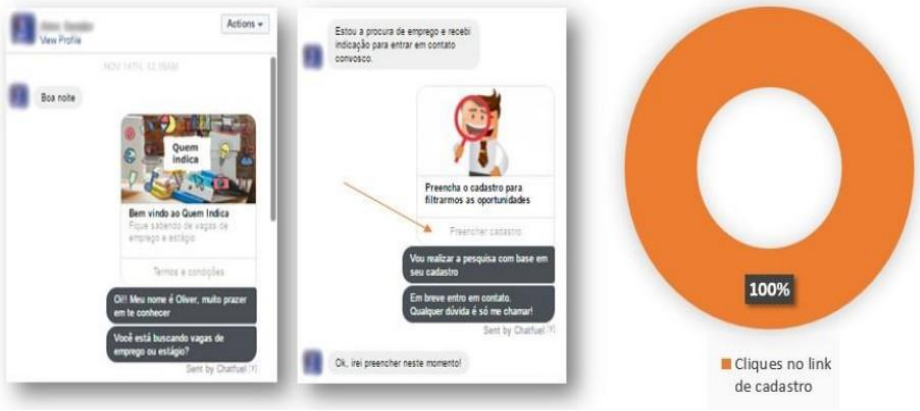


Fonte: NUNES, 2017.

REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

De acordo com os autores, o experimento ocorreu ao longo de 12 dias, durante os quais 18 pessoas interagiram com o chatbot. Todas as pessoas, ou seja, 100% delas, clicaram no primeiro link enviado, que solicitava o cadastro profissional, conforme ilustrado na figura abaixo.

Figura 7: Estudo de Caso envio de link

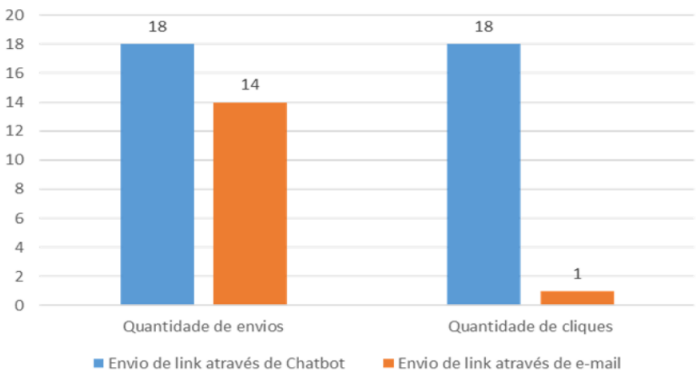


Fonte: NUNES, 2017.

Na pesquisa, verificou-se que 78% dos usuários que acessaram o link inicial preencheram o perfil. Posteriormente, foi enviado um segundo link com sugestões de vagas personalizadas com base nas respostas dos usuários, solicitando informações adicionais como nome, e-mail, telefone, última empresa onde trabalhou e cargo. Dos usuários que receberam este novo link, 71% completaram o preenchimento dessas informações.

Para avaliar a eficácia dessa abordagem, os participantes receberam um e-mail sobre a pesquisa de vagas, contendo um link para comparação, sendo que apenas um dos 14 e-mails enviados foi aberto. O gráfico 1 apresenta uma comparação visual desses resultados.

Gráfico 1: Comparativo entre chatbot x email



Fonte: NUNES, 2017.

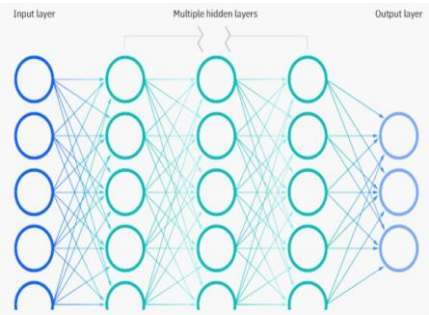
REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

Portanto, percebe-se que os códigos open-source, como A.L.I.C.E., e Pandorabots, que são baseados em API, podem ser um instrumento perigoso para o uso de hackers na execução de ataques, como foi efetivamente mostrado nos gráficos das figuras deste tópico. Por essa razão, destaca-se a importância de estudar sobre esse assunto para evitar esses tipos de ataques e investir na cibersegurança contra esses tipos de ataques.

2.2 Redes Neurais

Segundo De Almeida Souza & De Moraes (2021), a rede neural, uma técnica empregada em inteligência artificial, consiste em conexões direcionadas entre neurônios, sem pesos, que podem ser excitatórias (positivas) ou inibitórias (negativas), simulando as sinapses biológicas. Esse modelo simplificado é inspirado no funcionamento dos neurônios biológicos, como ilustrado na figura 8.

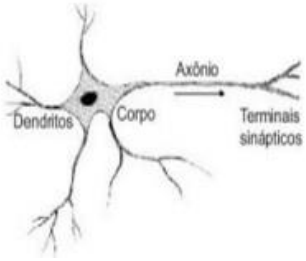
Figura 8: Esquema de uma rede neural



Fonte: Imagem da internet

Uma rede neural é um modelo computacional que simula o funcionamento de um cérebro, em que temos um esquema com vários círculos que se interligam entre si tendo entrada de dados e saída de dados que compartilhada de forma múltipla cada um fazendo cálculos e mandando para o outro, possibilitando a IA identificar padrões, fazer provisões, identificar linguagens, etc.

Figura 9: Representação simplificada de um neurônio humano

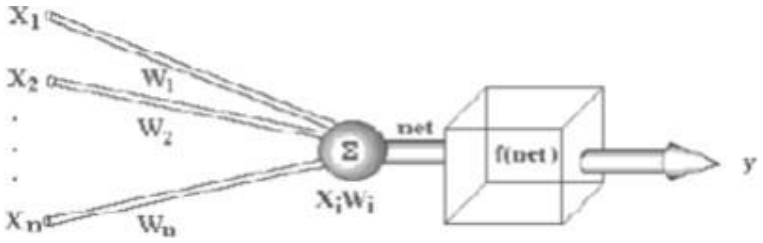


Fonte: FERNEDA, 2006.

REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

Conforme Rodrigues, Cândido & Silva (2018), os dendritos são responsáveis por captar os sinais recebidos e encaminhá-los ao corpo do neurônio para processamento. Em seguida, o corpo do neurônio gera um novo impulso que é transmitido pelo axônio aos neurônios adjacentes.

Figura 10: Neurônio artificial no modelo MCP



Fonte: MADSEN; ADAMATTI, 2011

Segundo Madsen e Adamatti (2011, p. 27), de acordo com a figura 10, cada neurônio artificial é constituído por um vetor de entradas X = [x1, x2,..., xn], que correspondem aos dendritos no modelo biológico. Cada entrada possui um peso Wi que determina a força da conexão entre os neurônios. Um peso positivo atua como um excitador, enquanto um peso negativo inibe a conexão. Caso o peso seja zero, a conexão entre os neurônios não existe.

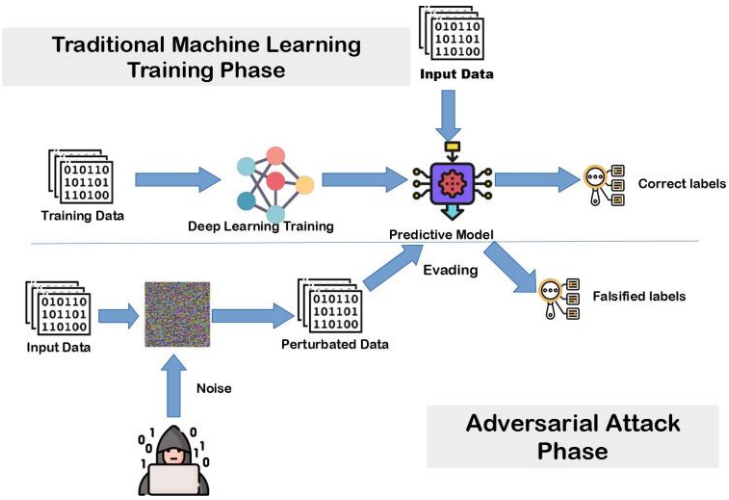
Segundo Delipetrev, Tsinaraki & Kostic (2020), uma rede neural é um componente da IA que, quando profunda e eficaz, é suscetível a ataques adversários. Isso ocorre porque, ao introduzir pequenas alterações nos dados de entrada, um invasor pode induzir a rede neural a tomar decisões incorretas. Para mitigar essa ameaça, estratégias como o treinamento contra adversários, considerando suposições específicas não aplicáveis aos testes de ataque com injeção falsa de dados, a hipótese de detecção de adversários em que o invasor utiliza dados diferentes da distribuição normal e o treinamento da rede neural para identificar tais exemplos antes de processá-los são necessárias.

Estudos recentes na área de Inteligência Artificial revelam que redes neurais profundas (DNNs) de alto desempenho são altamente suscetíveis a ataques adversários. Através de alterações sutis nas entradas das DNNs, um invasor pode manipulá-las para induzir decisões incorretas. Para mitigar essa ameaça, são necessários métodos como treinamento adversarial, onde o modelo é exposto a exemplos adversários que seguem distribuições diferentes das entradas normais. Isso permite que o DNN aprenda a distinguir exemplos legítimos de exemplos manipulados antes de processá-los (CHEN et al., 2021).

Dentre os ataques conhecidos, destaca-se o DeepAttacks, que utiliza conteúdo gerado por IA para atacar outras IA, através de redes neurais adversárias. Um exemplo mencionado pela Avast foi o uso desse algoritmo para enganar sistemas automáticos, fazendo-os interpretar erroneamente sinais de trânsito, como transformar um sinal de Pare em um limite de velocidade de 70 km/h (LIMA FILHO, 2019).

REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

Figura 11: Como funciona um Deep Attack

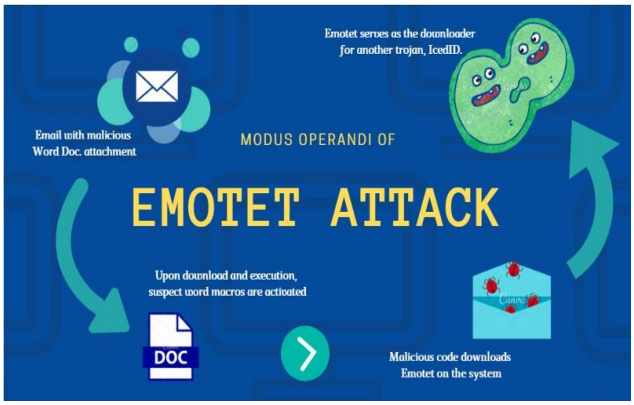


Fonte: Towards data science, 2020

Com a análise da figura acima, um Deep Attack funciona através do treinamento de uma Rede Neural Adversária Generativa (GAN) para gerar dados falsos que se assemelham a dados reais. Essa GAN é usada para criar exemplos enganosos que são então introduzidos no sistema de IA induzindo-o a tomar decisões incorretas ou prejudiciais. O objetivo é explorar a vulnerabilidade das redes neurais profundas, enganando-as com dados aparentemente autênticos, o que pode ter consequências graves em aplicações como reconhecimento de imagem, processamento de linguagem natural e outras tarefas de aprendizado de máquina.

Além disso, o Trojan Emotet exemplifica um avançado modelo de IA. Utilizando principalmente ataques de spam-phishing por e-mail, ele é capaz de se integrar automaticamente em conversas já existentes, aumentando assim sua autenticidade e persuadindo a vítima a clicar (SHOAIB, 2016).

Figura 12: Como funciona um Emotet Attack



Fonte: Itnow, 2022

REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

Com a análise da figura acima, o Trojan Emotet é um malware notório que se propaga por meio de campanhas de phishing. Em suas etapas de infecção, o Emotet geralmente começa com o envio de e-mails de phishing maliciosos para potenciais vítimas. Esses e-mails podem conter anexos maliciosos ou links para sites comprometidos. Quando uma vítima clica ou abre o anexo, o Emotet é instalado no sistema da vítima, permitindo que ele se espalhe pela rede da vítima e se auto-propague para outros computadores. Uma vez infectado, o Emotet é notável por sua capacidade de se auto-espalhar e baixar cargas úteis adicionais, tornando-o uma ameaça persistente e difícil de remover.

Em resumo, muitos ataques serão baseados em técnicas de Deep Learning e de redes neurais, já que possui rápida adaptabilidade e crescimento a partir de aprendizagem com dados e experimentação. A evolução das técnicas de treinamento contra adversários e a detecção de comportamento malicioso são essenciais para mitigar essas ameaças. À medida que a IA continua a avançar, é crucial que a comunidade de pesquisa e as empresas de tecnologia trabalhem em conjunto para desenvolver soluções robustas e proteger sistemas baseados em redes neurais contra possíveis ataques, garantindo assim um uso seguro e confiável dessa tecnologia inovadora.

2.3 Diferença entre o ataque com IA e sem IA

Emerge claramente a dimensão avassaladora dos ataques cibernéticos quando impulsionados pela IA superando, em muito, em devastação aqueles desprovidos dessa tecnologia. Esta realidade pode ser discernida ao analisarmos três componentes cruciais: Automatização, Letalidade e Rapidez.

Segundo Soprana (2017 apud DE ALMEIDA SOUZA & DE MORAES, 2021, p. 36):

> Qual é a principal diferença entre um ataque cibernético humano e usando inteligência artificial? para quem sofre o ataque não há muita diferença. Em sua forma mais básica, um ataque inteligente é uma versão automatizada de um ataque liderado por humanos. Um dos desafios que os atacantes enfrentam hoje, especialmente o phishing é que eles realmente necessitam entender o sistema que estão atacando ou seu alvo individual, que é uma atividade que leva muito tempo. Nesse caso, ataques automatizados com o uso da IA, personalizados e inteligentes podem ser mais eficazes e rápidos, porque cada objetivo leva muito menos tempo.

A IA capacita a automatização completa de ataques, possibilitando a orquestração de múltiplos ataques simultaneamente, sem a necessidade de intervenção humana, confiando exclusivamente em máquinas. A letalidade inerente a um ciberataque com IA frequentemente supera de maneira significativa a de ataques desprovidos dessa tecnologia, intensificando assim seu potencial destrutivo. De forma exemplar, o ciberataque desencadeado pelo vírus WannaCry em 2017 ilustra a impressionante velocidade intrínseca aos ataques cibernéticos habilitados por IA. Em questão de horas, esse ataque global disseminou-se de forma catastrófica, infectando sistemas e criptografando dados sensíveis, destacando o impacto devastador que a IA pode proporcionar.

Em maio de 2017, o vírus WannaCry afetou 200.000 computadores em 150 países, marcando um ataque sem precedentes na história da Internet. Esse incidente alarmou funcionários

REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

públicos e empresários quanto ao futuro dos ataques cibernéticos em larga escala (PASSARINHO, 2021 citado por DE ALMEIDA SOUZA E DE MORAES, 2021, p. 12). É incontestável que a utilização de inteligência artificial amplia consideravelmente a eficácia dos ataques cibernéticos, tornando-os mais poderosos e desafiadores de serem contidos por organizações e autoridades de segurança cibernética.

3 TENDÊNCIAS E DESAFIOS

Especialistas em cibersegurança indicam que os crimes cibernéticos estão em constante evolução, impulsionados pelo crescente conhecimento adquirido pelos hackers através da avaliação das máquinas infectadas. Além disso, os ataques de ransomware estão se tornando cada vez mais sofisticados, uma vez que algumas empresas optam por pagar os resgates (EMM, 2021 citado por DE ALMEIDA SOUZA E DE MORAES, 2021, p. 12). A capacidade essencial da inteligência artificial de aprender e se adaptar está inaugurando uma nova era, na qual ataques altamente personalizados, conduzidos por humanos, podem ser escalonados de forma escalável (DE MORAES et al., 2021, p. 13).

Os mecanismos ofensivos baseados em IA são capazes de mutação conforme aprendem sobre o ambiente, comprometendo sistemas de forma habilidosa com mínima chance de detecção. Consequentemente, os ataques futuros podem ser mais penetrantes, oferecendo uma maior garantia de alcançar seus objetivos. Por um lado, observa-se um avanço contínuo em sistemas baseados em IA, cada vez mais complexos e eficazes na detecção de anomalias. Por outro lado, a IA adversária representa um potencial de ataque extremamente eficaz.

De acordo com Zequim & Ribeiro (2022), houve um significativo aumento nos prejuízos relatados ao FBI devido aos ataques cibernéticos ao longo dos anos. Isso ressalta a importância crucial de investir em segurança da informação nas empresas para mitigar os impactos financeiros desses ataques. Em 2020, por exemplo, as empresas sofreram prejuízos de aproximadamente US$ 4 bilhões, sublinhando a importância vital dos sistemas de informação na proteção contra ataques cibernéticos.

Empresas que negligenciam a adoção de soluções baseadas em IA correm maior risco de enfrentar consequências severas. Portanto, é fundamental explorar o potencial da IA no contexto cibernético para uma compreensão mais profunda das ameaças digitais. Nesse sentido, enfrentamos dois desafios cruciais e interligados na área de IA: segurança e privacidade. A segurança abrange desde o acesso ilegal a informações até a violação da confidencialidade e dos dados pessoais (GOULART, conforme citado por DE ALMEIDA SOUZA E DE MORAES, 2021, p. 30).

Os desafios enfrentados no campo da IA e da cibersegurança são inegavelmente significativos e inter-relacionados. Como destacado por vários especialistas e evidenciado por eventos recentes, a proteção de dados e a privacidade são questões críticas. A ameaça de ciberataques continua a ser uma preocupação global com impactos financeiros substanciais.

REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

Um estudo realizado pela Mastercard em parceria com o Datafolha revelou que 57% das empresas brasileiras foram vítimas de ataques digitais. Dentro desse grupo, apenas 32% possuíam uma equipe dedicada à cibersegurança. Embora a maioria reconheça a importância da cibersegurança, 39% das empresas ainda não a consideram uma prioridade orçamentária. Embora muitas empresas afirmem ter planos de resposta a ataques cibernéticos, apenas um terço delas realizou testes de prevenção nos últimos três meses antes da pesquisa (VILELA, 2021 citado por DE ALMEIDA SOUZA E DE MORAES, 2021, p. 30).

No contexto brasileiro, a pesquisa realizada pela Mastercard/Datafolha destaca, ainda, a frequência de ataques cibernéticos nas empresas do país, com uma minoria tendo uma área dedicada à cibersegurança. Embora a maioria reconheça a importância da cibersegurança, a alocação de recursos e a implementação de medidas preventivas continuam sendo um desafio. A falta de testes regulares de prevenção e a escassez de profissionais qualificados na área ressaltam a necessidade de uma abordagem mais proativa na proteção de ativos digitais. (VILELA, 2021 apud DE ALMEIDA SOUZA E DE MORAES, 2021, p. 31)

Em suma, os desafios relacionados à segurança e à privacidade no contexto da IA e cibersegurança são complexos e dinâmicos, exigindo esforços constantes das organizações e da comunidade de pesquisa para enfrentá-los de maneira eficaz.

4 CONSIDERAÇÕES

Com base nas informações coletadas e nas discussões realizadas ao longo do trabalho, fica evidente que a IA apresenta riscos significativos que precisam ser abordados de forma adequada. A segurança, o futuro do trabalho e as ameaças cibernéticas são áreas críticas que requerem atenção e medidas de mitigação. É essencial estabelecer regulamentações e diretrizes claras para garantir a segurança dos sistemas de IA e proteger a privacidade e a integridade dos dados. Além disso, é necessário investir em pesquisas e desenvolvimento de estratégias de contramedidas para lidar com as ameaças emergentes. A IA tem o potencial de trazer benefícios significativos para a sociedade, mas também exige uma abordagem cautelosa e discernimento para evitar consequências indesejadas. Portanto, é fundamental promover uma abordagem equilibrada que leve em consideração tanto os benefícios quanto os riscos da IA.

Segundo Revoredo (2021, conforme citado por ZEQUIM & RIBEIRO, p. 29), várias organizações estão adotando a inteligência artificial em suas estratégias de conformidade para cumprir com os requisitos da Lei Geral de Proteção de Dados (LGPD). Além disso, a IA, utilizando Processamento de Linguagem Natural (NLP), está sendo empregada para auxiliar as organizações na interpretação do significado de seus contratos legais em contextos específicos, como no âmbito da LGPD, analisando cláusulas contratuais em relação a outros documentos corporativos.

Outro exemplo relevante da aplicação da IA é nos sistemas antivírus. De acordo com informações do site oficial da Avast, o sistema de IA utiliza aprendizado de máquina para extrair dados de toda a base de usuários e treinar continuamente cada módulo de segurança. Assim,

REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

quando uma nova amostra de malware é detectada, os produtos da Avast são atualizados automaticamente com os identificadores necessários para oferecer proteção atualizada.

Para mitigar os riscos associados à IA, é fundamental estabelecer políticas e regulamentações adequadas. Governos e organizações relevantes devem colaborar na criação de diretrizes claras que garantam a segurança e a ética na implementação da inteligência artificial. Essas regulamentações podem abordar questões como a responsabilidade legal dos sistemas autônomos de IA, a proteção dos dados pessoais e a transparência dos algoritmos.

Em conclusão, a IA apresenta riscos significativos que devem ser abordados de forma adequada. Os aspectos de segurança, futuro do trabalho e ameaças cibernéticas são de particular importância. Ao estabelecer regulamentações apropriadas, promover a conscientização e a educação, fomentar a colaboração e envolver a sociedade, podemos enfrentar os desafios da IA e aproveitar seus benefícios de maneira responsável. Com uma abordagem cautelosa, é possível construir um futuro em que a IA seja uma força impulsionadora para o progresso e o bem-estar da humanidade.

REFERÊNCIAS

BRITTO, Gladstone Avelino; FREITAS, Maristella Barros. Ciberataques em massa e os limites do poder punitivo na tipificação de crimes informáticos. **Revista de Direito Penal, Processo Penal e Constituição**, v. 3, n. 2, p. 1-16, 2017.

CHEN, Xing et al. DNNOff: offloading DNN-based intelligent IoT applications in mobile edge computing. **IEEE transactions on industrial informatics**, v. 18, n. 4, p. 2820-2829, 2021.

DE ALMEIDA SOUZA, Jaqueline Patrícia; DE MORAES, Maria José. Fortalezas e fragilidades no uso da Inteligência Artificial na Cibersegurança. **Revista Tecnológica da Fatec Americana**, v. 9, n. 02, p. 25-40, 2021.

DE CARVALHO JÚNIOR, Ciro Ferreira et al. Chatbot: uma visão geral sobre aplicações inteligentes. **Revista Sítio Novo**, v. 2, n. 2, p. 68-84, 2018.

DELIPETREV, Blagoj; TSINARAKI, Chrysi; KOSTIC, Uros. Historical evolution of artificial intelligence. 2020.

FERNEDA, Edberto. Redes neurais e sua aplicação em sistemas de recuperação de informação. **Ciência da Informação**, v. 35, p. 25-30, 2006.

GRAÇA, Pedro José Bentes. **O Ciberataque como Guerra de Guerrilha: o caso dos Ataques dos/DDos à Estónia, Geórgia e ao Google-China**. 2014. Tese de Doutorado. Universidade de Lisboa (Portugal).

JORDÃO, Fabio. DDoS: como funciona um ataque distribuído por negação de serviço. 2014.

LAZIĆ, LJUBOMIR. Benefit from Ai in cybersecurity. In: **The 11th International Conference on Business Information Security (BISEC-2019)**, 18th October. 2019.

REVISTA CIENTÍFICA ACERTTE
ISSN 2763-8928

LIMA FILHO, Francisco Sales de et al. Smart detection: an online approach for DoS/DDoS attack detection using machine learning. **Security and Communication Networks**, v. 2019, n. 1, p. 1574749, 2019.

MADSEN, Carlos Alberto; ADAMATTI, Diana Francisca. NeuroFURG: uma ferramenta de apoio ao ensino de Redes Neurais Artificiais. **Revista Brasileira de Informática na Educação**, v. 19, n. 02, p. 14, 2011.

MESSIAS, Lucas Fernando. Estudo sobre infecção de vírus de computador do tipo Ransomware. 2015.

MOWBRAY, Miranda; PEARSON, Siani; SHEN, Yun. Enhancing privacy in cloud computing via policy-based obfuscation. **The Journal of Supercomputing**, v. 61, n. 2, p. 267-291, 2012.

NUNES, Paulo Fernando Viegas. A definição de uma estratégia nacional de cibersegurança. **Nação e defesa**, 2012.

RODRIGUES, Lucas A.; CANDIDO, Renato; SILVA, Magno TM. Restauracao de imagens com redes neuronais. **methods**, v. 35, n. 1, p. 20-36, 2018.

SAMOILI, Sofia et al. AI watch. Defining artificial intelligence 2.0. Towards an operational definition and taxonomy of AI for the AI landscape. **JRC Research Reports**, n. JRC126426, 2021.

SANTOS, João Pedro Silva et al. Evolução da Inteligência Artificial. In: **Anais do Congresso Nacional Universidade**, EAD e Software Livre. 2020.

TELES, Tiago Miguel Fonseca Paiva de Sousa. **Cibersegurança**. 2015. Tese de Doutorado.

WIRKUTTIS, Nadine; KLEIN, Hadas. Artificial intelligence in cybersecurity. **Cyber, Intelligence, and Security**, v. 1, n. 1, p. 103-119, 2017.

ZEQUIM, Eduarda Pagim; RIBEIRO, Douglas Francisco. O PAPEL DA INTELIGÊNCIA ARTIFICIAL NA SEGURANÇA CIBERNÉTICA: o uso de sistemas inteligentes em benefício da segurança dos dados das empresas. **Revista Interface Tecnológica**, v. 19, n. 1, p. 21-33, 2022.