PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS

Bacharelado em Ciência da Computação

Guilherme Otávio de Oliveira
Daniel Lucas Soares Madureira

Além da Programação: IA na Fronteira da Segurança do Trabalho

Belo Horizonte

2021

Guilherme Otávio de Oliveira
Daniel Lucas Soares Madureira

Além da Programação: IA na Fronteira da Segurança do Trabalho

Projeto de Pesquisa apresentado na disciplina Trabalho Interdisciplinar III - Pesquisa Aplicada do curso de Ciência da Computação da Pontifícia Universidade Católica de Minas Gerais.

Belo Horizonte

2021

RESUMO

A IA tem o potencial de reduzir a necessidade da presença humana em atividades perigosas, monótonas e cansativas, permitindo que as pessoas se dediquem a tarefas menos arriscadas e mais estimulantes. No entanto, a IA também pode aumentar riscos existentes e introduzir novos desafios. Para mitigar esses riscos, é essencial o desenvolvimento de novos algoritmos de IA ou a aplicação de métodos inovadores, sempre considerando aspectos éticos, sociais e legais. O artigo discute como as tecnologias baseadas em Inteligência Artificial (IA) estão sendo cada vez mais utilizadas em diversos sistemas e ferramentas. Ele ainda enfatiza a importância de não transformar a IA em armas e de garantir que as aplicações de IA sejam claras quanto à sua natureza não humana.

Palavras-chave: Inteligência Artificial; Aprendizado de Máquina; IA Responsável .

SUMÁRIO

1 INTRODUÇÃO

A Inteligência Artificial (IA) tem se tornado cada vez mais presente em nossa sociedade, desempenhando um papel crucial em diversos setores e substituindo o trabalho humano em muitas áreas. Este avanço tecnológico traz consigo uma série de implicações éticas, sociais e econômicas que precisam ser cuidadosamente consideradas.

A IA tem o potencial de trazer benefícios significativos, como a redução de custos, aumento da produtividade, melhoria na tomada de decisões, maior competitividade e melhoria na experiência do trabalhador como mostrado por Carvalho [1]. No entanto, também existem riscos associados ao seu uso, incluindo a possibilidade de ataques cibernéticos aos sistemas de IA e a reprodução de preconceitos por parte dos algoritmos de IA a partir dos conjuntos de dados utilizados para treiná-los.

Este artigo busca explorar essas questões, com o objetivo de propor soluções para diminuir os riscos de ataques cibernéticos direcionados aos sistemas de IA e mitigar o aprendizado e reprodução de preconceitos por parte dos algoritmos de IA. Através de uma análise cuidadosa e consideração das implicações da IA, esperamos contribuir para o uso responsável e ético desta tecnologia em nossa sociedade.

2 OBJETIVOS

Este artigo visa explorar as implicações éticas, sociais e econômicas da Inteligência Artificial (IA) na sociedade contemporânea com foco específico em dois desafios críticos, a segurança contra ataques cibernéticos direcionados aos sistemas de IA e a mitigação do viés algorítmico nos algoritmos de IA, o objetivo é propor soluções eficazes e sustentáveis para reduzir esses riscos e promover um uso mais responsável e ético da IA em diversos setores, diante da urgente necessidade de proteger contra vazamentos de dados, que comprometem a privacidade e a segurança dos usuários finais. Além disso, a disseminação de dados falsos por meio de sistemas de IA pode distorcer informações, influenciando decisões sociais e econômicas de maneiras potencialmente prejudiciais. Portanto, análises detalhadas dessas questões são essenciais para contribuir para um debate informado e colaborativo sobre o futuro da IA, visando seu impacto positivo na sociedade e na economia global

3 REVISÃO BIBLIOGRÁFICA

Nesta revisão bibliográfica, serão abordados dois artigos relevantes sobre a Inteligência Artificial (IA) e seus impactos na sociedade. O primeiro parágrafo discutirá o conteúdo do primeiro artigo, enquanto o segundo parágrafo se concentrará no segundo artigo.

3.1 Inteligência Artificial: riscos, benefícios e uso responsável

Estamos usando tecnologias baseadas em Inteligência Artificial em um número crescente de sistemas e ferramentas. A Inteligência Artificial pode tornar reduzir a necessidade da presença humana em muitas atividades perigosas, monótonas e cansativas, nos liberando para atividades menos perigosas e mais desafiadoras e estimulantes. Ao mesmo tempo, a Inteligência Artificial pode aumentar riscos existentes e trazer novos riscos. Para evitar ou reduzir esses riscos, é necessário o desenvolvimento de novos algoritmos de Inteligência Artificial, ou seu uso de maneiras novas e inovadoras, levando em consideração questões éticas, sociais e legais.

3.2 IMPACTOS DA INTELIGÊCIA ARTIFICIAL NA SOCIEDADE

Esse artigo aborda os efeitos da Inteligência Artificial (IA) na sociedade, destacando tanto seus impactos positivos quanto os riscos associados. O principal objetivo deste estudo é analisar como a disseminação da IA está transformando diversos setores da sociedade e quais são as implicações para o futuro. A metodologia utilizada envolveu uma revisão da literatura existente sobre IA e seus impactos. Os resultados da pesquisa indicam que a IA está impulsionando avanços significativos em áreas como medicina, automação industrial e mobilidade, melhorando a eficiência e qualidade de vida. No entanto, também são destacados os riscos, incluindo preocupações éticas, perda de empregos devido à automação e ameaças à privacidade. As conclusões do artigo enfatizam a importância de abordar esses desafios de maneira proativa, desenvolvendo políticas e regulamentações adequadas para orientar o uso responsável da IA na sociedade. Além disso, destaca-se a necessidade de educação e conscientização sobre os impactos da IA, a fim de promover

um debate informado e moldar um futuro mais equilibrado e ético para a tecnologia.

4 METODOLOGIA

Este capítulo descreve a metodologia adotada para realizar a pesquisa proposta. Inicialmente, serão apresentadas as etapas e procedimentos utilizados para coleta e análise dos dados. Em seguida, será discutida a classificação da pesquisa conforme os critérios estabelecidos na literatura especializada.

4.1 Desenvolver e Implementar Medidas de Segurança Cibernética Robustas

4.1.1 Atividade 1: Identificação de Ameaças

Os sistemas de Inteligência Artificial (IA) estão expostos a diversas ameaças cibernéticas que podem comprometer sua segurança, confidencialidade e disponibilidade. Para mitigar esses riscos, é crucial compreender as principais formas de ataque. Primeiramente, destacam-se força bruta, que consiste em tentativas repetitivas e sistemáticas de adivinhar senhas ou chaves de criptografia, visando o acesso não autorizado ao sistema, utilizando ferramentas como Hydra ou Hashcat. Ademais, há a exploração de vulnerabilidades, na qual os invasores aproveitam falhas de segurança presentes no software ou hardware para obter acesso indevido ou causar danos ao sistema, frequentemente utilizando ferramentas como Metasploit. Assim como também existem os ataques de interferência adversária, onde os dados de entrada são manipulados com o intuito de enganar ou comprometer o modelo de IA, resultando em respostas incorretas ou maliciosas, sendo comuns técnicas como ataques de adversários. Por fim, a injeção de dados maliciosos envolve a introdução de dados falsos ou prejudiciais durante o treinamento do modelo, com o objetivo de corromper os resultados ou desviar o comportamento do sistema, utilizando scripts personalizados ou ferramentas de injeção de SQL, como SQLmap ou Havij.

4.1.2 Atividade 2: Análise Detalhada dos Dados de Treinamento

A análise detalhada dos dados de treinamento é uma etapa que envolve a identificação, coleta e análise crítica de conjuntos de dados relevantes que serão utilizados para treinar e validar os algoritmos de IA dos quais os principais passos incluem a identificação das fontes de dados pertinentes ao problema, como bancos de dados públicos utilizando

APIs como a do Kaggle ou repositórios como o UCI Machine Learning Repository, dados internos da organização ou dados adquiridos de terceiros.

Após isso, há o processo de aquisição ética e legal dos dados, utilizando ferramentas como Python e bibliotecas como requests para obtenção de permissões ou pandas para manipulação de dados estruturados. Em seguida, ocorre o processo de limpeza e pré-processamento dos dados, que inclui normalização com scikit-learn, remoção de ruídos com técnicas estatísticas ou de processamento de sinais, imputação de valores ausentes usando métodos como média ou mediana, e transformação dos dados para o formato adequado ao treinamento de modelos de IA, utilizando pipelines de pré-processamento em scikit-learn ou tensorFlow. Por fim os dados processados são então armazenados de forma segura com bancos de dados como MySQL, MongoDB ou serviços de armazenamento em nuvem como AWS S3 ou Google Cloud Storage, garantindo sua integridade e disponibilidade futura.

4.1.3 Atividade 3: Estabelecer Protocolos para Monitoramento Contínuo da Segurança

Estabelecer protocolos para o monitoramento contínuo da segurança envolve definir métricas apropriadas, implementar alertas e monitorar constantemente os sistemas e para realizar tal tarefa é necessário implementar vários métodos, os quais incluem escolher métricas adequadas, como detecção de intrusões e análise de padrões de acesso, para uma avaliação objetiva da segurança, configurar alertas baseados nessas métricas ajuda na detecção imediata de atividades suspeitas, de forma a garantir uma resposta rápida. O monitoramento contínuo dos logs, tráfego de rede e políticas de segurança permite identificar e mitigar ameaças, assegurando a proteção dos sistemas de IA contra tentativas de comprometimento. Ferramentas como SIEM (Security Information and Event Management), IDS/IPS (Intrusion Detection/Prevention Systems) e plataformas de monitoramento de rede serão utilizadas para implementar eficazmente esses métodos de segurança.

4.1.4 Atividade 4: Avaliação e Correção de Viés Algorítmico

Para implementar avaliação e correção de viés algorítmico faz necessário garantir o monitoramento contínuo dos resultados dos algoritmos em produção e a aplicação de técnicas para corrigir quaisquer vieses para identificar padrões que possam indicar discriminação ou tratamento desigual de grupos específicos. Uma vez identificados, os vieses são corrigidos utilizando técnicas como reamostragem de dados, ajuste de pesos ou modificação dos dados de treinamento utilizando ferramentas como Apache Spark para processamento de dados distribuídos, juntamente com bibliotecas de análise estatística como Apache Commons Math, promovendo confiança e aceitação dos usuários nos resul-

tados produzidos.

4.2 Cronograma

O cronograma apresentado na Tabela 1 detalha as atividades planejadas para o desenvolvimento seguro de sistemas de Inteligência Artificial (IA).

Tabela 1 – Cronograma

	Meses 1-3	**Meses 4-6**	**Meses 7-9**	**Meses 10-11**
Identificação de Ameaças				X
Análise Detalhada dos Dados de Treinamento			X	
Estabelecer Protocolos de Monitoramento		X		
Avaliação e Correção de Viés Algorítmico	X			

Referências Bibliográficas

1. CARVALHO, André Carlos Ponce de Leon Ferreira. Inteligência Artificial: riscos, benefícios e uso responsável. Estudos Avançados, São Paulo, Brasil, v. 35, n. 101, p. 21–36, 2021. DOI: 10.1590/s0103-4014.2021.35101.003. Disponível em: https://www.revistas.usp.br/eav/article/view/185020.. Acesso em: 16 maio. 2024.

2. FILHO, L. C. A. .; CONCEIÇÃO, G. C. da. IMPACTOS DA INTELIGÊCIA ARTIFICIAL NA SOCIEDADE. Revista Interface Tecnológica, [S. l.], v. 20, n. 2, p. 134–145, 2023. DOI: 10.31510/infa.v20i2.1777. Disponível em: https://revista.fatectq.edu.br/interfacetecnologica/article/view/1777. Acesso em: 16 maio. 2024.