PONTIFÍCIA UNIVERSIDADE CATÓLICA DE CAMPINAS

CENTRO DE CIÊNCIAS EXATAS, AMBIENTAIS E DE TECNOLOGIA

CURSO DE CIÊNCIA DE DADOS E INTELIGÊNCIA ARTIFICIAL

JOÃO ROBERTO CRESPI JUNIOR

ATIVIDADE DE EXTENSÃO

ÉTICA NA INTELIGÊNCIA ARTIFICIAL: DESAFIOS MODERNOS

Relatório sobre impactos éticos na aplicação e uso de inteligências artificiais

CAMPINAS

2023

RESUMO

O principal objetivo deste relatório será apresentar e evidenciar os problemas do acesso livre à inteligências artificiais que levam a problemas que ferem a ética e moral do ser humano, discutindo e problematizando a utilização do ChatGPT e seu uso mal-intencionado. Apresentará uma introdução à inteligência artificial, caracterização do problema e proposta de intervenção.

SUMÁRIO

INTRODUÇÃO 4
CONTEXTUALIZAÇÃO 4
DEFINIÇÃO DO PROBLEMA 5
PROPOSTA DE INTERVENÇÃO 6
CONSIDERAÇÕES FINAIS 7
REFERÊNCIAS 7

INTRODUÇÃO

A inteligência artificial (IA) desempenha um papel cada vez mais relevante em nossa sociedade, trazendo consigo inúmeros benefícios e avanços. No entanto, o acesso livre e o uso indiscriminado de IA podem acarretar problemas éticos e morais que afetam o bem-estar humano. Este relatório visa apresentar e evidenciar os desafios associados ao acesso irrestrito à inteligências artificiais, destacando especificamente o ChatGPT e seu potencial uso mal-intencionado. Serão discutidos os problemas decorrentes desse uso indevido, como respostas imprecisas, disseminação de informações falsas e violação de privacidade, além de explorar os impactos no âmbito acadêmico. Em seguida, serão propostas intervenções que visam mitigar esses problemas, promovendo o uso ético e responsável da IA. Com essas medidas, busca-se preservar a integridade moral e ética, assegurando um ambiente seguro e confiável para a utilização dessa tecnologia.

CONTEXTUALIZAÇÃO

As inteligências artificiais não supervisionadas são projetadas para aprender e compreender informações sem depender de um conjunto de dados rotulados ou instruções específicas. Em vez disso, esses modelos são alimentados com grandes quantidades de dados não rotulados, como texto, e utilizam técnicas de aprendizado de máquina para descobrir padrões e estruturas nesses dados. Já as supervisionadas, são treinadas com conjuntos de dados rotulados, nos quais cada exemplo de entrada está associado a um rótulo ou uma resposta correta.

O ChatGPT é um exemplo de modelo de linguagem baseado em transformadores, treinado usando técnicas de aprendizado de máquina supervisionado e não supervisionado. Ele foi treinado em uma grande quantidade de texto proveniente de diversas fontes, como livros, artigos, sites e muito mais. No entanto, seu conhecimento se restringe a eventos ou informações ocorridas após a sua data de corte de conhecimento, setembro de 2021.

Ao receber uma pergunta ou uma entrada de texto, o ChatGPT processa essa informação e gera uma resposta baseada em seu acervo de conhecimento que é fruto de um treinamento

prévio. Ele é capaz de capturar informações relevantes e até mesmo fornecer explicações detalhadas.

Sendo assim o ChatGPT vem sendo usado por milhares de pessoas em inúmeras áreas, para sanar dúvidas, adquirir conhecimento, gerar códigos de programação ou até mesmo para conseguir informações confidencias.

DEFINIÇÃO DO PROBLEMA

Apesar de suas habilidades impressionantes, o ChatGPT é um modelo que utiliza de inteligência artificial não supervisionada, ou seja, grande parte dos dados são não rotulados, o que ocasionalmente pode gerar respostas que parecem plausíveis, mas não são factualmente corretas. Ele também não tem capacidade de verificar a veracidade das informações que fornece, podendo gerar respostas imprecisas, enganosas. Seu treinamento é baseado em grandes volumes de dados da internet, podendo levar ao modelo refletir preconceitos e desigualdades. Além disso, alguns usuários estão utilizando a ferramenta para facilitar a obtenção de informações confidenciais, como CPF, RPG e outras informações que já poderiam ser encontradas na internet com dificuldade, mediante linhas de comandos na tentativa de "desconfigurar" a inteligência, desse modo a ferramenta que era para ser inovadora e um avanço para a sociedade acaba sendo uma arma para pessoas mal-intencionadas.

O uso indevido do programa não é punido pela própria plataforma, que é de fácil acesso, permitindo que qualquer pessoa com um e-mail possa utilizá-lo para conseguir informações restritas, realizar atividades ou provas acadêmicas. Em um teste, a ferramenta foi capaz de resolver um exame de direito com 95 questões de múltipla escolha e 12 questões dissertativas, obtendo um C+ o que levaria em aprovação.

Com todo esse conhecimento centralizado em uma única ferramenta e de fácil acesso a população, as universidades enfrentam uma dificuldade para avaliar devidamente os alunos, já que as atividades podem ser facilmente burladas com essa ferramenta, assim, comprometendo o sistema de avaliação acadêmico e ferindo a ética e moral do ser humano.

O problema não se restringe apenas ao âmbito acadêmico e a gama de ferramentas não se limita apenas a texto, existem inteligências artificiais que produzem, além de texto, imagens, vídeos, *dashboards* e até mesmo voz artificial. Desse modo, indivíduos maliciosos possuem um arsenal de ferramentas para produzir desinformação, prejudicar pessoas e até mesmo conseguir um diploma.

PROPOSTA DE INTERVENÇÃO

Diante dos problemas associados ao acesso livre e ao uso mal-intencionado de inteligências artificiais, especialmente o ChatGPT, é necessário implementar medidas que visam mitigar esses impactos negativos e preservar a ética e a moral humana. A seguir, são apresentadas algumas propostas de intervenção:

1. Melhoria contínua do treinamento do ChatGPT: A empresa responsável pelo desenvolvimento deve investir em um treinamento mais rigoroso, abordando não apenas a quantidade de dados, mas também a qualidade e a diversidade desses dados, obtendo seus dados de fontes confiaveis e revisadas. Isso ajudará a reduzir a ocorrência de respostas imprecisas e a refletir menos preconceitos e desigualdades em suas respostas.
2. Implementação de filtros e compromisso com a veracidade: O ChatGPT deve ser aprimorado com mecanismos de filtragem e verificações de informações, apesar da plataforma, atualmente, não ter compromisso com a verdade, é de extrema importancia que ela crie esse compromisso, para evitar a disseminação de respostas falsas ou enganosas. Isso pode envolver a utilização de fontes confiáveis de informação e técnicas de validação de fatos durante o processamento das respostas.
3. Reforço das políticas de uso responsável: A empresa que fornece acesso ao ChatGPT deve estabelecer políticas claras e rigorosas de uso responsável, reforçando a proibição do uso da ferramenta para obter informações confidenciais, pessoais ou promover atividades ilegais. É importante que essas políticas sejam divulgadas amplamente e que haja consequências para o não cumprimento delas, como impedir a conta de utilizar o programa.
4. Monitoramento e relato de abusos: É necessário estabelecer um sistema eficaz de monitoramento para identificar o uso mal-intencionado. Em caso de suspeita de mal-uso deve ser criado um relato de abuso pelo próprio sistema, levando ao usuário ser investigado, podendo acarretar banimento da conta.
5. Educação sobre o uso ético da IA: É fundamental promover a conscientização e a educação sobre os princípios éticos junto ao uso da inteligência artificial. Isso pode ser realizado por meio de campanhas de divulgação feitas pela sociedade, treinamentos e materiais educativos direcionados aos usuários feitos pela empresa, destacando a importância de utilizar a tecnologia de maneira responsável e respeitosa.
6. Colaboração com instituições educacionais: os desenvolvedores devem buscar parcerias com instituições educacionais, especialmente universidades, para desenvolver soluções

conjuntas que ajudem a preservar a integridade dos métodos de avaliações das entidades educacionais, como uma IA que identifica se a resposta foi gerada pelo próprio ChatGPT ou derivada do mesmo.

CONSIDERAÇÕES FINAIS

As inteligências artificiais chegaram para ficar no nosso cotidiano e o impacto que elas trazem para a sociedade é muito benéfico, entretanto, o acesso livre para usuarios maliciosos traz consigo desafios éticos e morais que precisam ser abordados de forma proativa antes que se tornem problemas maiores. As propostas de intervenções apresentadas buscam enfrentar esses problemas por meio da melhoria do treinamento, implementações de filtros mais robustos, estabelecimento de políticas claras, monitoramento e relato de abusos, educação sobre o uso ético e da colaboração com instituições educacionais. Ao adotar essas medidas, espera-se mitigar o uso mal-intencionado da inteligência artificial, preservar a ética e moral dentro da sociedade e promover um ambiente mais seguro e responsável para o acesso a essas tecnologias

REFERÊNCIAS

ChatGPT bot passes law school exam. Disponível em: <https://www.cbsnews.com/news/chatgpt-bot-passes-law-school-exam/>.

Quais os impactos do ChatGPT e da Inteligência Artificial na Educação? Disponível em: <https://www.ifsc.edu.br/web/ifsc-verifica/w/quais-os-impactos-do-chatgpt-e-da-inteligencia-artificial-na-educacao->.

ChatGPT na educação: impactos, vantagens e desvantagens. Disponível em: <https://brasilescola.uol.com.br/noticias/chatgpt-na-educacao-especialista-comenta-sobre-a-inteligencia-artificial-no-campo-educacional/3129039.html#:~:text=ChatGPT%20na%20educa>. Acesso em: 30 mai. 2023.

The Impact of ChatGPT on Academic Integrity. Disponível em: <https://www.enago.com/thesis-editing/blog/the-impact-of-chatgpt-on-academic-integrity>. Acesso em: 30 mai. 2023.

TEAL, M. **The Ethics of College Students Using ChatGPT**. Disponível em: <https://ethicspolicy.unc.edu/news/2023/04/17/the-ethics-of-college-students-using-chatgpt/>.

EPSTEIN, D. **Ethical Implications of ChatGPT in the Educational Setting**. Disponível em: <https://insights.bu.edu/ethical-implications-of-chatgpt-in-the-educational-setting/>.