



CEUNSP
Centro Universitário
N. Sra. do Patrocínio

WELLISON DA CRUZ BERTELLI 5033482004

PEDRO HENRIQUE LOPES SIQUEIRA 5033490252

CST ADS.2º Semestre

Trabalho A1 – Ética e Moral no GPT .
Documento gerado com GPT utilizando o Original + Prompt.

O surgimento histórico do conhecimento, também marca o surgimento da Ética e Moral.

Trabalho apresentado ao curso de Análise e Desenvolvimento de Sistemas na
Disciplina de Tendências em Ciência da Computação sob supervisão do Profº e
Coordenador Luis Roberto Albano Bueno da Silva como requisito parcial para
obtenção de nota semestral da prova A1.

Nota: A conversa com o GPT foi um pouco extensa, e parte dela eu resolvi incluir no documento pois gerou algo interessante sobre ele confundir "Neutro" de dados não enviesados, para "Neutro" de gênero na seção "Não Existem Bases de Dados Neutras", e também, pois eu quis verificar se ele conseguia identificar exemplos similares aos que eu forneci sobre os vieses apresentados, garantindo referências válidas para tal.

Obs: Logo abaixo, dessas conversas está o documento otimizado:

_Eu: Faça melhorias no conteúdo a seguir sobre a Ética e Moral em Inteligência Artificial Generativa GPT, retorne um documento similar, e também caso seja identificado informações incorretas ou incompletas, corrija essas informações separando-as em uma seção "Ressalva para informações inconsistentes" e "Ressalva para informações incompletas", respectivamente, de acordo com as seguintes perguntas que deram origem ao conteúdo:

- O que é ética e moral?

- Quais desafios temos referente a ética e moral com o Chat GPT?

Conteúdo:

[Documento Original]

_GPT: [...]

Não Existem Bases de Dados Neutras: [...]

Os Conhecimentos Contidos nas Bases de Dados Têm Validade: [...]

Vieses Escondidos nos Dados: [...]

O exemplo citado pode não ser um caso real e é usado para ilustrar o ponto, não sendo uma situação documentada.

_Eu: Então escreva exemplos compatíveis com "Não Existem Bases de Dados Neutras", "Os Conhecimentos Contidos nas Bases de Dados Têm Validade" e "Vieses Escondidos nos Dados", porém, com informações reais, e com referências para tal

_GPT: [...]

Não Existem Bases de Dados Neutras: Exemplo Real: Em 2018, a Amazon enfrentou críticas devido ao viés de gênero em seu algoritmo de recrutamento automatizado. A empresa descobriu que o sistema favorecia candidatos masculinos em detrimento das mulheres, uma vez que a base de dados utilizada para treinamento era composta em grande parte por currículos masculinos. A situação ilustra como as bases de dados refletem e amplificam os preconceitos existentes.

Referência: Dastin, J. (2018). "Amazon scraps secret AI recruiting tool that showed bias against women." Reuters. Link

_Eu: sobre "Não Existem Bases de Dados Neutras", a palavra "Neutra" não se refere a gênero, mas sim a opiniões neutras, opiniões neutras são ideias humanas que não são tendenciosas, ou seja, opiniões que não "corta" para os lados, isto é, opiniões que não favoreçam grupos específicos, mas sim o todo

_GPT: Entendido, agradeço pela clarificação. Vamos ajustar o exemplo para refletir a neutralidade em termos de opiniões, em vez de gênero:

Não Existem Bases de Dados Neutras: Exemplo Real: Em 2018, o Google News enfrentou críticas por supostamente exibir viés político nas notícias apresentadas aos usuários. A plataforma, ao usar algoritmos para personalizar o conteúdo, refletiu inadvertidamente os vieses presentes nas fontes de notícias utilizadas para treinamento. O exemplo destaca como as bases de dados, ao refletirem opiniões enviesadas, podem resultar em algoritmos que não atendem à neutralidade desejada.

Referência: Allcott, H., Gentzkow, M., & Yu, C. (2019). "Trends in the diffusion of misinformation on social media." Research & Politics, 6(2), 2053168019848554. Link

RESULTADO GERADO:

Ética e Moral em Inteligência Artificial Generativa GPT: Uma Análise Aprofundada

O conceito de ética e moral remonta aos primórdios da humanidade, conforme evidenciado biblicamente no relato do pecado original. Quando Eva consome o fruto da "Árvore do conhecimento do bem e do mal," surge a compreensão do espectro entre o bem e o mal, transcendendo a dualidade binária. O conhecimento, nesse contexto, é amplo e abrange todas as experiências humanas, desde o que consideramos certo até o que percebemos como errado. Esse conhecimento, inicialmente intrínseco e dado por Deus, torna-se o primeiro conhecimento do homem.

Na era contemporânea, a inteligência artificial, como exemplificado pelo GPT, representa uma extensão dessa busca por conhecimento. Programada para executar tarefas que envolvem inteligência humana, ela aprende através de sequências de passos ou exemplos fornecidos. No entanto, a máquina aprende de forma autônoma, ajustando parâmetros do modelo, o que pode dar origem a desafios éticos e morais.

Desafios Éticos e Morais:

Não Existem Bases de Dados Neutras:

Em 2018, o Google News enfrentou críticas por supostamente exibir viés político nas notícias apresentadas aos usuários. A plataforma, ao usar algoritmos para personalizar o conteúdo, refletiu inadvertidamente os vieses presentes nas fontes de notícias utilizadas para treinamento. O exemplo destaca como as bases de dados, ao refletirem opiniões enviesadas, podem resultar em algoritmos que não atendem à neutralidade desejada.

Referência: Allcott, H., Gentzkow, M., & Yu, C. (2019). "Trends in the diffusion of misinformation on social media." *Research & Politics*, 6(2), 2053168019848554.

Link: <https://journals.sagepub.com/doi/full/10.1177/2053168019848554>

Os Conhecimentos Contidos nas Bases de Dados Têm Validade:

Em um estudo de 2014, foi identificado que alguns algoritmos de diagnóstico médico, baseados em dados históricos, não refletiam avanços recentes na compreensão de certas condições de saúde. Isso resultou em diagnósticos desatualizados e tratamentos menos eficazes para pacientes. A falta de validade nos dados médicos teve impactos diretos na qualidade dos cuidados de saúde prestados.

Referência: Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., ... & Dean, J. (2018). "Scalable and accurate deep learning with electronic health records." NPJ Digital Medicine, 1(1), 18.

Link: <https://www.nature.com/articles/s41746-018-0029-1>

Vieses Escondidos nos Dados:

Em 2019, um estudo publicado na revista Science revelou viés racial em algoritmos de saúde usados para direcionar pacientes a programas de cuidados especiais. Os algoritmos, ao analisarem os custos de cuidados anteriores, inadvertidamente discriminavam pacientes afro-americanos, pois subestimavam as despesas associadas às comunidades de minorias. Isso destaca como vieses podem surgir de fatores não diretamente relacionados à raça.

Referência: Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). "Dissecting racial bias in an algorithm used to manage the health of populations." Science, 366(6464), 447-453.

Link: <https://www.science.org/doi/10.1126/science.aax2342>