Trabalho 2 - Ética em IA

Gabriel Luciano (RA265673)
Victor Juliani (RA225162)

3 de janeiro de 2021

1 O que é ética em Inteligência Artificial?

Existem dois significados que engloba "Ética em Inteligência Artificial". O primeiro diz respeito às atitudes humanas na construção de um futuro tecnológico ético, com "IAs responsáveis-– também conhecido como roboética — que respeitem princípios básicos; por exemplo, privacidade de dados, transparência, apreço pela vida humana e liberdade, entendimento de vieses. Já na segunda definição, condiz a uma outra vertente estudada: ética na IA, a ética das próprias máquinas. Nesta é explorada um conceito mais futurístico, onde preocupa-se com as questões morais das máquinas; com o avanço da tecnologia, robôs e agentes de inteligência artificial devem ter comportamentos e valores morais [1]. Ou seja, como prover para as máquinas um guia de princípios éticos gerais que podem ser utilizados nas situações por elas enfrentadas. [2]

Seguindo esta definição, notamos que a primeira vertente é a mais prática e utilitarista no sentido de proteger e guardar o que consideramos justo e ético hoje. Por outro lado, a segunda explora um contexto onde diversas empresas almejam chegar, e consequentemente, não pode ser ignorado como campo de pesquisa. Muito pelo contrário, como exercício de pensamento, ela é proponente de diversas vertentes que podem ser ampliadas para várias áreas do conhecimento humano e que vão muito além da ética pela ética.

Sendo assim, é possível explorar diversos temas, como por exemplo, o direito civil das máquinas, indagações similares a apresentada em [3]: "A declaração universal dos direitos humanos se aplicaria para uma IA senciente?". Por conseguinte, perguntas que, embora não existam respostas claras, podem propor pensamentos que ajudem inclusive na primeira vertente do pensamento de ética em IA. Tendo um exemplo de: Como construir um futuro ético, sem pensar em um presente ético em IA? Deste modo, percebe-se que os dois significados iniciais são complementares, no sentido que buscam moldar um mundo mais ético e justo para coexistência de máquinas e humanos — sejam essas máquinas robôs com o propósito de vender mais com bases em anúncios direcionados ou um robô que é ciente de si e do universo à sua volta.

Por fim, existe também uma nova vertente emergente, que propõe construir uma IA para definir o que é ético [4] e o que não é. Porém, se estamos em dúvida sobre como construir uma IA ética, será que faz sentido deixar uma IA decidir tais princípios? Por não termos como responder isso, deixamos essa terceira vertente de lado.

2 Notícia recente de um problema de ética em IA (2020)

- Título da notícia: *What a machine learning tool that turns Obama white can (and can't) tell us about AI bias.*
- Veículo de publicação: The Verge
- Data: Junho 23, 2020

- Autor: James Vincent
- Acesso: The Verge - Face Depixelizer

Nesta notícia, é abordado uma aplicação que tem como objetivo realizar o *upscaling* de imagens pixeladas (de baixa qualidade). Entretanto, o resultado do processo revelou características enviesadas de forma racista. Além de que, ao utilizar a solução proposta, um outro problema pode se fazer presente, ao exemplo da remoção de anonimato de rostos embaçados em vídeos.

A notícia, por tanto, aborda o problema de que ao "melhorar"as imagens, houve um branqueamento dos indivíduos das fotos, além de inclusão de características enviesadas. O artigo cita exemplos do presidente Barack Obama, que ao ser *despixelizado*, se transforma em um homem branco. Além disso, a notícia é complementada com *tweets* de várias pessoas que ao utilizarem a ferramenta de IA, também sofreram com esse embranquecimento étnico, apagando também traços étnicos de mulheres asiáticas.

Sendo assim, a notícia se enquadra na primeira vertente abordada na questão 1, no sentido de que não houve preocupação dos autores do artigo em tratar o viés racial do algoritmo. Ou seja, não houve o cuidado de treinar a inteligência artificial com uma base suficientemente diversa para o problema que se propuseram a resolver. Outro indício de que o problema foi de fato uma inaptidão dos autores, de ter esse cuidado com a base de dados, é que na própria notícia é apresentada uma outra solução, do cientista de dados Mario Klingemann, com uma solução similar para o problema de melhoramento de imagens paxalizadas sem esse viés.

3 Artigo para solução de um problema de ética em 2020.

- Artigo: *Liu, Yang and Y. Wu. "FNED: A Deep Network for Fake News Early Detection on Social Media." ACM Trans. Inf. Syst. 38 (2020): 25:1-25:33.*
- Citações: 03 (Google Scholar)
- Acesso: https://dl.acm.org/doi/10.1145/3386253

O trabalho selecionado explicita o problema da rápida divulgação de *Fake News* em redes sociais. Este cenário é alarmante pois, devido a quantidade de tempo que os usuários passam na plataforma, acabam por substituir outros meios de acesso às notícias por aquelas ali divulgadas. Junto a isso, existem outros motivos pelo qual a aplicação difunde facilmente a informação, pondo em risco sua veracidade, são elas: Baixo custo, tanto em tempo quanto dinheiro, para consumir a informação; Facilidade de disseminar a informação na rede social; Consumidores da notícia viram divulgadores por meio de compartilhamentos e afins; Requer menos monitoramento de censura para ser difundida. Consequentemente a isso, possibilita divulgação de informações falsas, podendo alienar os usuários que as consomem [5].

Por este motivo, o trabalho analisado propõe o FNED, um aplicação para detecção de *Fake News* antecipadamente em redes sociais. Para tanto, a aplicação é composta por três partes: i) extrator de características baseado em grande volume de repostas sensível a status; ii) classificador de notícias baseado em *CNN*; iii) um *PU-Learning framework*.

A primeira delas é responsável por associar cada uma as respostas de um grupo a um perfil de usuário que envia as informações, formando um vetor de características conhecidas. Seguindo, explora na segunda etapa a saída da primeira, uma vez que esta apresenta um mapa de características que consiste em uma sequência de k concatenações de características de usuários e de texto, que será utilizado em uma CNN para classificar os textos analisados. Por fim, em sua última etapa, após o treinamento da CNN, indicado na etapa anterior [apenas com *fake news* (P) e amostras não rotuladas (U)], levando em consideração que $|P| < |U|$, para criar um novo dataset balanceado para criação de um classificador de notícias binário. Esse processo será repetido k vezes para a produção de um classificador forte o suficiente para classificar as amostras não rotuladas, obtendo acurácia superior a 92%.

Referências

[1] Keng Siau and Weiyu Wang. Artificial intelligence (ai) ethics: Ethics of ai and ethical ai. *Journal of Database Management*, 31:74–87, 03 2020.

[2] M. Anderson and S.L. Anderson. *Machine Ethics*. Cambridge University Press, 2011.

[3] Erik Sandewall. Ethics, human rights, the intelligent robot, and its subsystem for moral beliefs. *International Journal of Social Robotics*, Mar 2019.

[4] Bruce M. McLaren. Extensionally defining principles and cases in ethics: An ai model. *Artificial Intelligence*, 150(1):145 – 181, 2003. AI and Law.

[5] Yang Liu and Yi-Fang Brook Wu. Fned: A deep network for fake news early detection on social media. *ACM Trans. Inf. Syst.*, 38(3), May 2020.