



UNIVERSIDADE FEDERAL DE SANTA CATARINA

CIÊNCIA, TECNOLOGIA E SOCIEDADE

Trabalho Individual - Comentário crítico do artigo **Pós-realidade e Teoria da Desinformação: inquietações sobre o uso massivo de IA Generativa**

Pamela Cristina dos Santos Monteiro

O texto aborda tanto aspectos científicos como tecnológicos, pois o mesmo relata o impacto de uma tecnologia recente que foi desenvolvida na academia, que são as os Large Language Models (LLMs), mais especificamente, o produto ChatGPT. No entanto, os autores do artigo não aprofundam essas questões. O funcionamento dos LLMs, ou como eles são criados, não é detalhado. Com isso, o leitor que não tenha um conhecimento prévio sobre o assunto fica a mercê das descrições dos autores do artigo ou de fazer pesquisas em outras fontes cujo linguajar pode ser muito técnico e formal, dificultando o entendimento do assunto. Considerando o tema do artigo, que é a preocupação do uso de IA generativas (LLMs são um exemplo disso) na proliferação de desinformações, o fato dos autores não explicarem em detalhes como funciona ou o que exatamente são LLMs pode-se dizer que eles apresentam um viés em exagerar os perigos dessa ferramenta. Por exemplo, no trecho "*Para que o aprendizado de máquina possa ser saudável por um longo período de tempo é preciso garantir que o acesso às fontes de dados originais humanos seja preservado e que os dados adicionais, não gerados pelos LLMs, permaneçam disponíveis.*", os autores falham em dizer que as LLMs já são treinadas em dados originais humanos. O grande problema de qualquer modelo de aprendizado de máquina, seja ele pequeno ou gigante como uma LLM, é garantir que os dados sejam de qualidade e que apresentem um mínimo de viés possíveis dos seus criadores. Por exemplo, os dados utilizados para treinar modelos de reconhecimento de feições humanas precisam conter faces de todas as etnias e gêneros. E essa característica, aparentemente óbvia, é de responsabilidade do criador do modelo. E essa mesma característica deve ser observada nos LLMs.

O texto aborda aspectos sociais na questão do impacto que as LLMs podem ter na proliferação de desinformação na nossa sociedade. Um exemplo disso são as recentes *fake news*. Isso impacta diretamente em questões éticas e morais. A desinformação sempre existiu em nossa sociedade. Por exemplo, antes da invenção do Photoshop, já era possível manipular fotos. No entanto, tecnologias recentes como as LLMs facilitam a criação dessa desinformação. Uma pessoa com pouco conhecimento literário pode criar um texto gramaticalmente correto e formal em uma LLM (por exemplo, ChatGPT) simplesmente passando um *prompt* pedindo algo semelhante ao prompt a seguir: Crie uma notícia sobre o político X que fale que ele foi flagrado desviando dinheiro destinado a população. Uma LLM vai criar uma notícia

nesse estilo baseando-se em notícias utilizadas em seu treinamento. Destaca-se, no entanto, que uma LLM nunca terá garantia de relatar se um texto gerado é verídico ou não. Pois, uma LLM, de forma resumida, simplesmente prevê a próxima palavra mais provável de ocorrer ao gerar um texto. Esse funcionamento não é detalhado pelos autores, dando a entender, que há algo de errado com a LLM caso ela produza algum texto que não está atrelado a realidade. A regulação das LLM deve ocorrer, pois apesar de produzirem textos fora da realidade, o que não é errado, facilita muito o trabalho de pessoas com intenções erradas. É relativamente simples criar imagens ou textos convincentes de qualquer pessoa atrelada a qualquer assunto. Isso gera o problema que as pessoas atingidas pelo material malicioso produzidos pelas LLM devem provar sua inocência ou não realização de algum ato. Isso afeta de forma mais severa a população de baixa renda, que não tem recursos para se proteger ou buscar seus direitos. A regulação também vai auxiliar na proteção da saúde das pessoas, pois, como observado durante a pandemia da Covid-19, a produção de *fake news* relacionadas a área da saúde tem um impacto massivo na vida das pessoas. Diversas mortes poderiam ter sido evitadas caso parte da população tivesse seguido os protocolos corretos ao invés daqueles detalhados em *fake news*. As LLMs podem ser reguladas através de Projeto de Lei que definam certos limites éticos como sociais para os conteúdos produzidos pelas LLMs. Por exemplo, pode-se impedir que uma LLM como o ChatGPT produza conteúdos *fake news* relacionados a saúde. Um ponto interessante trazido pelos autores é a preservação do *input* e *output* dos usuários das LLM. Com isso, facilitaria o rastreamento do usuário que produzir informações maliciosas com ferramentas estilo ChatGPT. Apesar do impacto das desinformações geradas pela LLM, os autores do artigo destacam poucos desse impacto, focando mais na questão que as LLMs perdem a qualidade com o passar do tempo. Apesar desse fato ser possível (toda modelo de aprendizado de máquina degrada com o tempo), esse fato não deveria ser o ponto focal da discussão do texto.

O texto não foi relevante para minha formação em qualquer aspecto pois o ChatGPT já vem sendo amplamente discutido e a maioria das pessoas ligadas a área de tecnologia já o testou alguma vez, inclusive o autor deste texto. O ChatGPT é uma ferramenta que pessoas da área de tecnologia deverão aprender a utilizar. Com a regulamentação correta e amadurecimento dos modelos, os LLMs serão ferramentas indispensáveis para os profissionais que queiram ter uma produtividade competitiva.