



“O QUE OS ROBÔS FARÃO COM A GENTE?”: UMA REVISÃO DE LITERATURA ACERCA DO CHATGPT

Nunah Souza Santos¹
Alessandro Novaes Pereira²
Sérgio Rodrigo da Silva Ferreira³
Ruth de Cássia dos Reis⁴

RESUMO

O avanço dos dispositivos de Inteligência Artificial (IA), especialmente aqueles baseados em processamentos de textos e imagens, tem despertado crescente interesse devido ao lançamento constante de novas ferramentas. A notoriedade atingiu patamares significativos com o lançamento do ChatGPT, um modelo capaz de criar diálogos extremamente similares aos humanos, proporcionando um novo paradigma na relação entre humanos e máquinas. A pesquisa aborda as crescentes preocupações sociais e éticas geradas por essa tecnologia, destacando questões como viés de dados, transparência, e o impacto potencial na prática de pesquisa e no trabalho criativo. Analisa-se a necessidade de compreensão dos aspectos evocados por pesquisadores no campo das humanidades digitais, a fim de estabelecer parâmetros para a investigação contínua desta tecnologia emergente.

Palavras-chave: ChatGPT, Inteligência Artificial, Viés de Dados, Vigilância, Tecnologias digitais.

INTRODUÇÃO

A recente ascensão de dispositivos de inteligência artificial (IA) baseados em processamentos de textos e imagens tem ganhado cada vez mais notoriedade através de lançamentos recorrentes de novas ferramentas. Esses programas, baseados em absorção de milhares de bancos de dados, possibilitam criações de conteúdos completos, diversos e interativos, promovendo um novo relacionamento entre o humano e a máquina.

Embora o surgimento de programas baseados em inteligência artificial já estivesse sendo observado, o assunto tomou proporções maiores a partir da apresentação pública do ChatGPT, em novembro de 2022. Esse modelo apresentou um diferencial até então ainda não

¹Graduanda do Curso de Comunicação Social - Jornalismo da Universidade Federal do Espírito Santo - UFES, nunah.santos@edu.ufes.br;

²Graduando do Curso de ABI Ciências Sociais da Universidade Federal do Espírito Santo - UFES, alessandro.n.pereira@edu.ufes.br;

³ Doutor em Comunicação e Cultura Contemporâneas pela UFBA, em estágio pós-doutoral no Programa de Pós-Graduação em Comunicação e Territorialidades da Universidade Federal do Espírito Santo (PÓSCOM/UFES), bolsista Capes, sergiorodrigosf@gmail.com;

⁴ Orientadora do trabalho, doutora em Comunicação e Cultura pela Universidade Federal do Rio de Janeiro e professora titular da Universidade Federal do Espírito Santo, ruth.reis@ufes.br.



difundido: a possibilidade de criação de diálogos com extrema semelhança aos de humanos.

Através de comandos simples, o ChatGPT é capaz de produzir conteúdos completos e diversificados, característica possível devido a sua estrutura pautada em redes neurais e *machine learning*. O ChatGPT se autodescreve como “uma versão específica do modelo de linguagem GPT (*Generative Pré-trained Transformer*), desenvolvido pela OpenAI,” e “é capaz de compreender o significado e a estrutura das frases, gerando respostas coerentes e naturais” (OPENAI, ChatGPT, 2021).

O entusiasmo recorrente da popularização desses dispositivos, no entanto, coloca sob disfarce o seu mecanismo de mediação social, reforçado pela enxurrada de dados arrebanhados pelas tecnologias ao longo dos anos. Por esse motivo, discussões acerca do termo “IA” vem se popularizando e criando um movimento de investigação em diversas esferas da sociedade, que buscam vislumbrar os possíveis impactos provocados em todos os setores.

O presente trabalho tem como objetivo principal elencar quais aspectos do ChatGPT têm sido evocados por pesquisadores do campo das humanidades digitais e quais vieses eles possuem a fim de estabelecer parâmetros de análise para pesquisarmos a respeito desta tecnologia emergente. A questão que queremos responder é de que modo esses autores têm especulado como o ChatGPT vai impactar a prática da pesquisa e dos usos no trabalho criativo e na mídia.

A importância deste trabalho se justifica a partir da discussão sobre as repercussões sociais da Inteligência Artificial e a influência das negociações ator-rede, difundidas pela delegação de comportamentos humanos para os robôs, principalmente ao considerar os artefatos deliberadamente projetados para moldar nossas ações e, por muitas vezes, substituir a ação humana, “é neste ponto que você tem uma escolha relativamente nova: ou disciplinar as pessoas ou substituir os humanos não confiáveis por um personagem não-humano delegado” (LATOUR, 1992, p. 238, tradução nossa).

Este artigo apresenta as primeiras inferências sobre o tema e faz parte de uma pesquisa maior que pretende identificar a questão do viés discursivo em textos e imagens gerados por dispositivos como o Chat GPT. Autores como Noble (2018), O’Neil (2016), Silva (2019) e Benjamin (2019) apontam para questões de impacto no que se refere a gênero, raça, religião, política entre outros fatores importantes, já extremamente afetados pelo colonialismo de dados. Dois aspectos serão elencados: a política-ideológica-eleitoral e as questões de gênero.



A pesquisa sobre as perspectivas de gênero nos discursos automatizados buscará identificar se os vieses nos sistemas algorítmicos podem corroborar os ideais hegemônicos, reforçando e renovando práticas machistas e misóginas. Já a pesquisa sobre política-ideológica-eleitoral investigará os possíveis vieses na construção discursiva automatizada, com ênfase em discursos sobre a polarização política que colocada em pontos extremos, esses vieses em sistemas algorítmicos, podendo reforçar desigualdades sociais e estereótipos, além do risco de incrementar ao fenômeno da desinformação na sociedade.

O presente artigo tem como estratégia metodológica a revisão narrativa da literatura, até então produzida sobre o ChatGPT no meio acadêmico, buscando compreender de que forma essa ferramenta vem sendo encarada e discutida cientificamente. A escolha dessa perspectiva metodológica, feita através de mapeamento de literatura recente e com uma temática mais aberta, nos possibilita encarar as construções atuais acerca de um assunto em plena fase de ebulição.

IMPLICAÇÕES SOCIAIS DO CHATGPT E DESAFIOS ÉTICOS

O ChatGPT trouxe, com toda sua notoriedade na mídia, preocupações constantes que norteiam pesquisas, produções científicas, narrativas controversas e mobilização de políticas públicas ao redor do mundo. Embora alguns aspectos já venham sendo evocados, os impactos da IA generativa a longo prazo ainda precisam ser mais esclarecidos. Para Sampaio Et Al (2023, p.17), “[a]s IAs revelam mais os absurdos das práticas humanas do que do fazer tecnológico”, já que seus decisores, aqueles responsáveis pelas escolhas de seus processos de criação, “são, em sua maioria, jovens, brancos, norte-americanos, que residem no Vale do Silício, com uma visão de mundo ocidental e neoliberal” (SAMPAIO ET AL 2023, p.5).

Essas preocupações são reforçadas principalmente por sua característica que Márcio Carneiro dos Santos (2023, p.24) identifica como “caixa opaca”, já que as questões de propriedade industrial permitem que seus processos internos sejam protegidos e apenas parcialmente compreendidos. Ou seja, os processos que levam à produção de seus algoritmos apenas são divulgados se assim quiserem seus detentores. Apesar dos resultados de sua pesquisa terem demonstrado que o modelo GPT não tenha sido treinado para mentir propositalmente, suas respostas não devem ser plenamente acatadas, já que questões sensíveis provindas de discurso neoliberal hegemônico precisam ser observadas e levadas em consideração.



Em estudo recente, Valência (2023) explora um outro campo da pesquisa em comunicação: os autoefeitos da mídia. A chegada das diversas tecnologias de inteligência artificial marca uma mudança na percepção dessas ferramentas nos processos de comunicação. Uma importante observação é que as tecnologias, que antes eram compreendidas como mediadoras no processo comunicacional, agora, com a IA, passam a ocupar um papel de sujeito comunicativo, assim como os próprios humanos. Dessa forma, esse relacionamento enfrenta modificações e uma mudança de perspectiva que, de acordo com Gusman e Lewis (2019, p. 73 apud. VALÊNCIA, 2023, p. 48) resultam na alteração de “sujeitos-máquina com os quais as pessoas criam significado, em vez de através dos quais as pessoas criam significado”.

Para o autor, a interação com essas tecnologias que são capazes de criar conteúdos comunicacionais faz com que o usuário enfrente efeitos únicos se comparados àqueles provocados pela comunicação de massa e que ainda precisam ser aprofundados. Valência defende, portanto, a necessidade de uma abordagem interdisciplinar a fim de mapear os autoefeitos resultantes da relação IA e usuário.

É importante elucidar também as diferenças existentes entre alguns termos, enquanto os *modelos gerativos* são conjuntos de regras e instruções que aprendem como as palavras são normalmente usadas de maneira conjunta, *os modelos de linguagem* são maneiras de prever as palavras ou símbolos que vêm a seguir em uma sentença. Nesse sentido, elas usam os modelos gerativos, para que posteriormente essas regras sejam utilizadas para antecipar como novas informações podem ser estruturadas, ilustrando quanto mais fiel possível a maneira como as palavras são usadas no cotidiano humano, “isso é feito mesmo quando não é possível compreender completamente todas as nuances e complexidades da linguagem” (SAMPAIO ET AL, 2023, p.4).

Esse modelo elabora cautelosamente uma palavra após a outra, baseado em que probabilidade a próxima seria usada por um humano. Exatamente por isso, tais modelos nunca repetem exatamente suas respostas e, com alguma frequência, eram o que tem sido chamado de alucinação, que acontecem quando o robô escolhe uma probabilidade que é sintaticamente possível, mas falsa em termos de fatos e dados (SAMPAIO ET AL, 2023).

Sampaio lembra que o modelo de linguagem gerativa pode eliminar o ato de descobrir coisas agradáveis por acaso – como referências inusitadas, *insights* oportunos e cruzamentos, ou seja, o termo que os ingleses chamam de *serendipity*, que em português conhecemos como *serendipidade*. A partir dessa ótica, surge mais uma problemática acerca do uso da



Inteligência Artificial no processo de sociabilidade humana: o relatório do Fórum Internacional sobre IA e o Futuro da Educação Desenvolvendo Competências for the AI Era (MIAO; HOLMES, 2020) aponta os riscos da IA, e aponta cautela no uso, difusão e implantação desses *robôs* na educação, a fim de atenuar o processo criativo e autônomo de aprendizagem.

Vale pontuar, inclusive, que o ChatGPT possui limitações, pois, por ser construído com base estatística, a qual se vale de padrões definidos por um grande conjunto de dados de texto, apresenta a possibilidade de que preconceitos e estereótipos presentes nos dados sejam replicados (IRIGARAY ET AL, 2023, p.1). O perigo da repetição de tais comportamentos problemáticos originados da captação de dados possivelmente influenciados por discursos neoliberais, torna-se ainda mais preocupante quando a ferramenta em questão apresenta um suposto caráter democratizante. Essa característica, mascarada através da objetividade e fluência de linguagem, pode criar uma ilusão de que o conhecimento esteja acessível a todos, reforçando a replicação desses padrões problemáticos (SANTAELLA, 2023, p. 66).

Para além dessas questões mencionadas, Santaella aponta para o dilema da inserção das ferramentas de Inteligência Artificial no meio acadêmico, e evoca problemáticas na produção científica. De acordo com a autora, embora esses programas apresentem um bom potencial, é imprescindível que se mantenha o papel da interferência humana no fazer científico, já que os sistemas nos colocam à frente de problemáticas relacionadas a “autoria e direitos autorais, custos e colonialismo de dados e os paradoxos da internacionalização da pesquisa, centro e periferia e produção de conhecimento” (2023, p.75).

Santaella (2023) reforça que a premissa é verdadeira e traça alguns paralelos que buscam mitigar ou até mesmo barrar a ferramenta, desde a proposta de diversas universidades do mundo de proibir o uso, fechando as suas portas e fluxos digitais a quaisquer intromissões nos seus tradicionais regramentos educativos (SANTAELLA, 2023, p. 9) ao relativo esforço da “União Europeia que, desde 2021, discute uma regulamentação para a Inteligência Artificial (IA)” (SANTAELLA, 2023, p. 9).

A autora lembra os abalos das projeções do ChatGPT na autoria, criatividade e autonomia humana, Santaella (2023) reforça a hipótese que, “enquanto a geração de imagens afeta o nicho dos produtores criativos no campo da visualidade, o ChatGPT afeta todos os seres humanos falantes e letrados” (SANTAELLA, 2023, p. 45). Nesse sentido, ela lembra os impactos da Inteligência Artificial Gerativa (IAG) nas “profissões e formações educacionais em todas as áreas em que a linguagem verbal” (SANTAELLA, 2023, p. 45). É importante



salientar também que, para além dos abalos cognitivos da formação da linguagem, a autora pontua que IAGs de imagem sensibiliza a autoria, a criatividade e autonomia, eixos centrais de questões pontuadas como o clímax do que é considerado unicamente humano (SANTAELLA, 2023, p. 45).

Outro tópico mensurado no livro é a velocidade difundida a partir da criação pelo uso das IAGs de imagem. Para o criador digital, Chad Nelson, “essa tecnologia leva você do lamejo em sua cabeça para um primeiro esboço em segundos [...] a velocidade com que você pode criar e explorar é revolucionária”. Concomitantemente Paul Trillo, artista digital, pontua a praticidade da ferramenta, que, embora para algumas pessoas o uso delas seja a morte de artistas de determinados nichos, ele acredita na possibilidade de alguns artistas não precisarem trabalhar à noite e nos fins de semana, a partir do uso desses mecanismos (SANTAELLA, 2023, p. 49). Nesse contexto, Santaella (2023) lembra que, “consideramos os humanos criativos e negamos tal caráter à IA não apenas porque entendemos relativamente como a IA funciona, mas porque não entendemos suficientemente como os humanos funcionam” (SANTAELLA, 2023, p. 55).

CONCLUSÃO

Levando em conta o caráter inicial da pesquisa, que buscará investigar os aspectos político-ideológico e de gênero no viés discursivo dos robôs, ChatGPT e outros, a contribuição deste trabalho reside na capacidade de possibilitar maior conhecimento às recentes discussões que permeiam o assunto no meio acadêmico e na mídia. Em conclusão, a ascensão dos dispositivos de inteligência artificial, exemplificada pelo ChatGPT, marca uma era de transformações significativas no relacionamento entre humanos e máquinas. A capacidade desses programas de gerar conteúdos completos e interativos, simulando diálogos humanos e imagens, levanta questões cruciais sobre seu impacto nas práticas de pesquisa, na criatividade e nos usos na comunicação e na mídia.

No entanto, a notoriedade do ChatGPT e de tecnologias semelhantes deve ser acompanhada por uma análise crítica dos desafios éticos e das implicações sociais que esses avanços introduzem. A opacidade das caixas de algoritmos e a concentração de poder nas mãos de decisores com determinadas características demográficas suscitam preocupações sobre vieses e representatividade na produção de IA. A complexidade desses sistemas, embora promissora, também traz à tona dilemas relacionados à autonomia humana, serendipidade na descoberta e a reprodução de preconceitos presentes nos dados.



Além disso, destacamos a necessidade de uma abordagem interdisciplinar para compreender os efeitos cognitivos, emocionais, atitudinais e comportamentais gerados pela interação entre humanos e IA. As implicações na educação, na produção científica e na criatividade são consideráveis, exigindo uma reflexão cuidadosa sobre como integrar essas tecnologias de maneira ética e eficaz.

REFERÊNCIAS

BENJAMIN, R. **Race After Technology: Abolitionist Tools for the New Jim Code**. Polity Press, 2019. IRIGARAY, H. A.R.; STOCKER F, **ChatGPT: um museu de grandes novidades**. In. cad. EBAPE, v. 21, nº 1, Rio de Janeiro, 2023.

IRIGARAY, H. A.R.; STOCKER F, **ChatGPT: um museu de grandes novidades**. In. cad. EBAPE, v. 21, nº 1, Rio de Janeiro, 2023.

LATOUR, B. **Where are the missing masses? The sociology of a few mundane artifacts**. In: Bijker, W. E. and Law, J. (orgs). *Shaping Technology/Building Society: Studies in Sociotechnical Change*, Cambridge, MA, MIT Press, p. 225-58, 1992.

NOBLE, S.. **Algorithms of Oppression: How Search Engines Reinforce Racism**. NYU Press, 2018

O'NEIL, C.. **Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy**. Crown Publishing Group, 2016

OPENAI. ChatGPT. Versão 1.0. 2021. Disponível em: <https://openai.com>. Acesso em: jun. 2023.

SAMPAIO, R. C. et al **ChatGPT and other AIs will change all scientific research: initial reflections on uses and consequences**. SciELO Preprints, 2023.

SANTAELLA, L. **Há como deter a invasão do ChatGPT?** Estação das letras e cores, 2023.

SANTOS, M. C. D. **What I learned interviewing a robot: Notes on the experimental application of the EIAF methodology using the ChatGPT artificial intelligence tool**. HIPERTEXT.NET: REVISTA ACADÉMICA SOBRE DOCUMENTACIÓN DIGITAL Y COMUNICACION INTERACTIVA, Barcelona, n 26, p. 23-29, 2023.

SILVA, T.. **Racismo Algorítmico em Plataformas Digitais: microagressões e discriminação em código**. In: Anais do Simpósio Internacional Lavits, 2019

SILVA, T.. **Racismo Algorítmico: inteligência artificial e discriminação nas redes digitais**. São Paulo: Edições Sesc, 2022. VALÊNCIA, V. A. Autoefectos en la comunicación mediada por IA. Hipertext.net, 2023, n.º 26, pp. 47-52.

VALÊNCIA, V. A. **Autoefectos en la comunicación mediada por IA**. Hipertext.net, 2023, n.º 26, pp. 47-52.