

ARTIGO

TECNOLOGIAS DA LINGUAGEM, ÉTICA EM IA E REGULAMENTAÇÃO

POR

Marlo Souza
msouza1@ufba.br

Tecnologias da Inteligência e Linguagem Humana (TIL) ganharam bastante proeminência na última década com o surgimento de aplicações como sistemas de tradução automatizada e assistentes virtuais. Uma das grandes tendências da área é a utilização de métodos de aprendizagem de máquina e grandes conjuntos de dados. Entre eles, há um recente interesse na

aplicação de grandes Modelos de Linguagem (ML) como os modelos BERT e GPT-3.

Uma crença implícita na literatura é que tais modelos codificam conhecimento “da língua” que pode ser usado durante o “raciocínio”. Tais trabalhos, entretanto, ignoram que a linguagem não pode ser isolada das estruturas sociais que a produzem. Assim, um ML representa necessariamente a lingua-

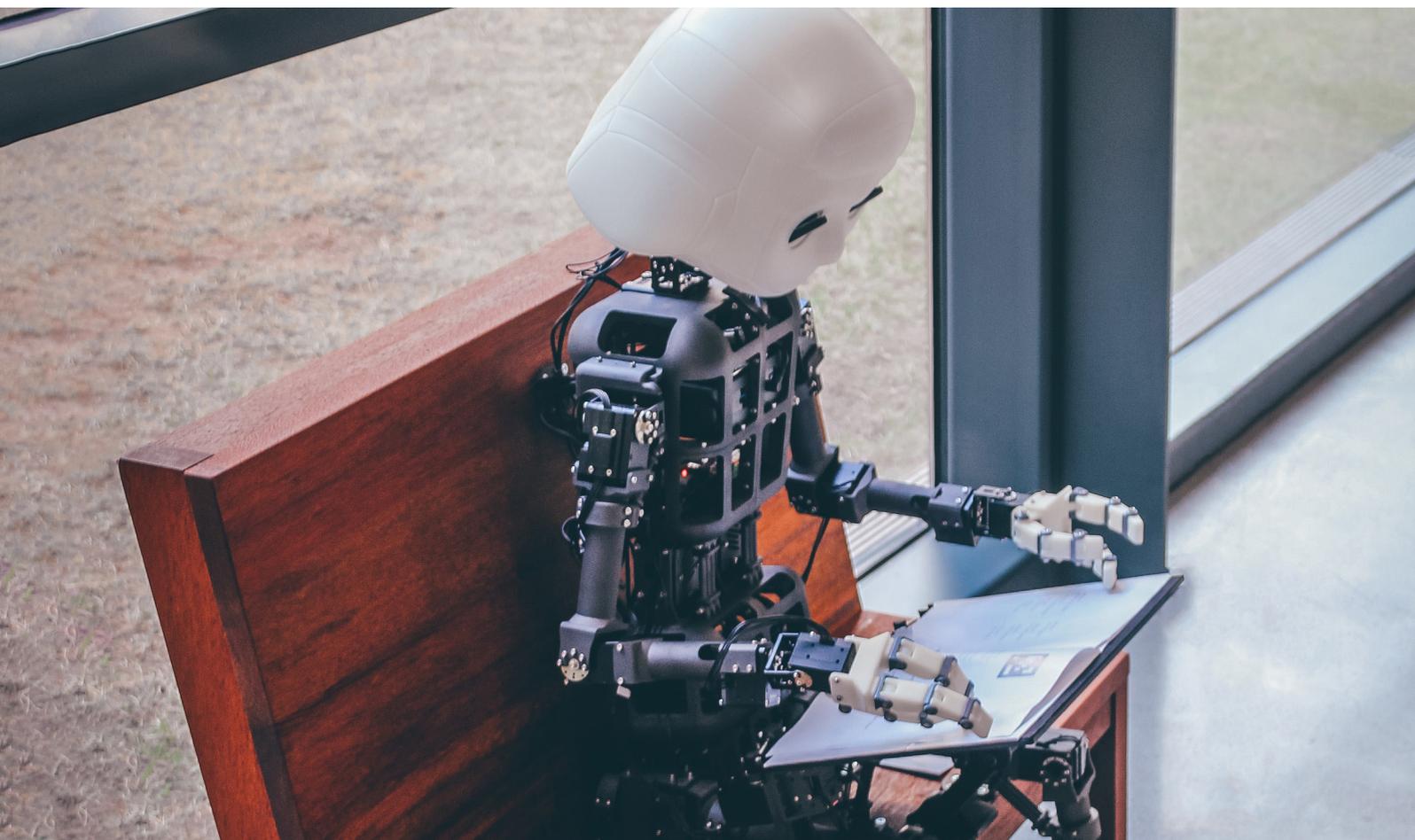
gem de alguém, em algum lugar, de uma classe social.

Uma vez que a linguagem é indissociável de sua função social, qualquer modelo está imbuído de uma ideologia de linguagem. Assumir que uma grande coleção de textos obtidos da Web como universalmente representante da variabilidade de uma dada língua é, como argumentam Bender et al. [2], desconsiderar a existência das vozes menos representadas na sociedade - aqueles com menor acesso à educação e cuja variante é desprestigiada [1].

Por fim, a atual pesquisa baseada em grandes ML está atrelada a um elevado custo econômico e ambiental. Enquanto os custos monetários implicam na dupla exclusão de acesso de comunidades em desvantagem econômica, tanto para sua utilização, quanto para a construção de tecnologias, o impacto socioambiental de seu desenvolvimento será mais sentido justamente pelas comunidades mais marginalizadas [2].

Qual Política Está Embutida em um TIL?

Existe uma crescente preocupação em como as TIL podem perpetuar ou agudizar iniquidades sociais, e.g., [6]. É preciso enfatizar que os danos causados por tais tecnologias variam em escala e natureza. Por um lado, ao naturalizar uma variante privilegiada enquanto norma padrão [1], tais tecnologias cometem violência simbólica ao classificarem variantes como “erradas” ou que “precisam passar por normalização” [4]. Por outro lado, podem também produzir violência material ao mascarar processos de tomada de decisão inerentemente injustos sob uma ilusão de objetividade técnica. Por exemplo, a ineficácia ao processar variantes em um sistema de triagem automatizado na Saúde [8] pode impor barreiras ao seu acesso. Similarmente, a automatização de decisões em Tribunais [9] levanta questões sobre como relações de poder codificadas na língua são reproduzidas nas decisões desses sistemas.



Ademais, sistemas sociotécnicos são construídos a partir de escolhas de projeto que evidenciam posições políticas - não existe artefato neutro ou puramente objetivo [12]. Assim, a própria escolha de quais problemas podem ser automatizados é uma escolha política que impregna de viés o processo de desenvolvimento das TIL. De fato, Winner [12, p. 122, tradução nossa] já enfatizava a necessidade de “olhar por trás das coisas técnicas e notar as circunstâncias sociais do desenvolvimento e uso [dos sistemas computacionais]”. É preciso, então, examinar TILs não somente sob a perspectiva técnica, mas também quem os desenvolve, quem os usará, sobre quem eles serão usados e como se dá a sua utilização.

Regulamentação e Educação em Inteligência Artificial

Primeiramente, uma vez que as TIL se tornam mais ubíquas, é essencial que as próximas gerações sejam educadas para entendê-las e seus impactos. Para tanto, é necessária a inclusão de tópicos de Inteligência Artificial, Letramento Digital e Ética da IA nos currículos da Educação Básica e Superior.

Do ponto de vista de regulamentação do desenvolvimento e uso das TIL, é preciso entender quais os riscos da implantação de tais sistemas na sociedade. Para tanto, propostas de regulamentação como a da União Europeia¹ propõem a classificação de tipos de risco, que limitam o uso e estabelecem as responsabilidades daqueles que desenvolvem tais tecnologias.

¹ <https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence>

Por um lado, toda tomada de decisão por uma TIL precisa ser transparente, i.e., as decisões de projeto têm de estar claras e disponíveis às entidades controladoras. Isso significa também que qualquer pessoa sujeita a uma decisão algorítmica precisa estar ciente de tal fato, assim como quais de seus dados são utilizados em tais sistemas. Para tanto, há propostas na literatura para documentação de modelos [11] e dados usados para sua criação [7].

Além da transparência, é preciso garantir que os danos causados por TIL sejam conhecidos e atacados. A literatura em Avaliação de Impacto Algorítmico mostra que sua capacidade de atacar danos se dá através das relações de poder estabelecidas entre as partes interessadas. Mais ainda, é preciso entender que os impactos identificados são nada mais que construtos avaliativos e só podem ser fidedignos se capazes de representar os prejuízos reais [10].

Nesse sentido, recentemente foram propostas iniciativas de inclusão de diferentes partes interessadas, especialmente aquelas sujeitas às decisões de tais sistemas, no processo de projeto e avaliação das TIL [3]. Dada a natureza dos sistemas de decisão baseado em dados, entretanto, é preciso ainda uma forte regulação de como os dados dos participantes serão utilizados e quem terá acesso a eles - questões que tocam aspectos sobre privacidade de dados e ética em pesquisa.

Como Dignum [5] enfatiza, o objetivo da pesquisa em IA ética e responsável não é produzir tecnologias que sejam ‘completamente justas, pois esse é um objetivo inal-

cançável. Antes, procuramos estabelecer métricas, metodologias e processos com os quais podemos garantir a confiança nesses sistemas. Nesse artigo, discutimos de forma não exaustiva algumas das dificuldades em tal construção e quais ações são necessárias para limitar os possíveis danos por meio de políticas e regulamenta-

ções necessárias para a área.

Há um grande caminho a percorrer na construção de um futuro em que as TIL sejam totalmente integradas à sociedade de forma construtiva, mas sem dúvida é um desafio que, enquanto estudiosos da Computação e Tecnologia, não podemos nos abster de atacar.

Referências

1. BAGNO, Marcos. A norma oculta: língua & poder na sociedade brasileira. Parábola, 2003.
2. BENDER, Emily M. et al. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? . In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. 2021. p. 610-623.
3. BERDITCHEVSKAIA, Aleks; MALLIARAKI, Eirini; PEACH, Kathy. Participatory AI for humanitarian innovation. 2021.
4. BERTAGLIA, Thales Felipe Costa. Normalização textual de conteúdo gerado por usuário. 2017. Tese de Doutorado. Universidade de São Paulo.
5. DIGNUM, Virginia. The myth of complete ai-fairness. In: International Conference on Artificial Intelligence in Medicine. Springer, Cham, 2021. p. 3-8.
6. EUBANKS, Virginia. Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press, 2018.
7. GEBRU, Timnit et al. Datasheets for datasets. Communications of the ACM, v. 64, n. 12, p. 86-92, 2021.
8. INOUE, Alexandre; PRADO, Marcus; COZMAN, Fábio. Automated Emergency Room Triage: Helping Patients Get the Best Treatment. In: Anais do XVII Encontro Nacional de Inteligência Artificial e Computacional. SBC, 2020. p. 591-602.
9. LAGE-FREITAS, André et al. Predicting Brazilian court decisions. PeerJ Computer Science, v. 8, p. e904, 2022.
10. METCALF, Jacob et al. Algorithmic impact assessments and accountability: The co-construction of impacts. In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. 2021. p. 735-746.
11. MITCHELL, Margaret et al. Model cards for model reporting. In: Proceedings of the conference on fairness, accountability, and transparency. 2019. p. 220-229.
12. WINNER, Langdon. Do artifacts have politics?. In: Computer Ethics. Routledge, 2017. p. 177-192.



MARLO SOUZA é Professor Adjunto do Departamento de Computação Interdisciplinar na Universidade Federal da Bahia e exerce atualmente a presidência da Comissão Especial de Processamento de Linguagem Natural da Sociedade Brasileira de Computação além de atuar nos Grupos de Pesquisa e Extensão em Informática Educação e Sociedade – Onda Digital e no Núcleo de IA Ética e Confiável do Rio de Janeiro. Atua nas áreas de Lógica, Representação de Conhecimento e Processamento de Linguagens Naturais.