



Não-neutralidade da tecnologia: a ação social das mediações técnicas

Rafael Gonçalves

Pedro P. Ferreira

Resumo

Tendo em vista o problema da neutralidade ou não da tecnologia, a presente pesquisa explora como o uso do conceito de mediação técnica – proposto pela teoria ator-rede – pode contribuir para pensar a ação social da mesma, em especial de algoritmos de aprendizado de máquina e de sistemas inteligentes. Assim, ela visa analisar como a tecnologia atua na sociedade reproduzindo e produzindo vieses, mesmo que estas por vezes mantenham um discurso e uma aparência de neutralidade.

Palavras-chave: mediação técnica, neutralidade, aprendizado de máquina

1 Introdução

Esta pesquisa vem investigando a não-neutralidade da tecnologia, em particular das tecnologias de aprendizado de máquina, – i.e., a forma como valores são produzidos e reproduzidos pelos sistemas sociotécnicos nos quais ela funciona, mesmo quando eles são apresentados como “neutros”. Da perspectiva da sociologia, a concretização de valores e preconceitos em instituições e estruturas sociais não é nem surpreendente, e nem necessariamente problemática. Afinal, instituições e estruturas sociais são, por definição, concretizações dos valores da cultura que as criou (DURKHEIM, 1996), e se os sistemas inteligentes vêm cada vez mais atuando como instituições contemporâneas, é compreensível que concretizem os preconceitos e discriminações (viéses) dos grupos que os projetam, constroem e usam.

Existe, porém, uma crescente preocupação atualmente com as assimetrias repetidamente produzidas por sistemas inteligentes, com “a insustentável neutralidade da tecnologia” (CAMPOS; SANTOS DIAS, 2018), com a importância de verbalizar a sua não-neutralidade (ROSA; STRIEDER, 2020), e de reter os fios que articulam ciência, tecnologia e sociedade (DAGNINO, 2008). Neste texto, nos baseamos nos trabalhos de Latour (1994a,b, 2002, 2005) para apresentar uma perspectiva de agência social das mediações técnicas, que nos parece particularmente vantajosa para a investigação

do problema do viés (*bias*) e da produção-reprodução de opressões no aprendizado de máquina.

2 Viés no aprendizado de máquina

O viés (*bias*) no aprendizado de máquina (*machine learning*) pode ser entendido como “a preferência algorítmica, geralmente não intencional, por uma previsão em lugar de outras, resultando em consequências legalmente ou eticamente inapropriadas”, ou simplesmente como “programação que assume os preconceitos dos programadores ou dos dados” (SHADOWEN, 2017, p. 5). Esse problema não é novo (BOYD; CRAWFORD, 2012), e vem ganhando importância nos últimos anos, à medida em que sistemas “inteligentes” assumem mais tarefas e se inserem com maior profundidade na vida contemporânea (CRAWFORD; CALO, 2016; CRAWFORD; SCHULTZ, 2019; CRAWFORD; WHITTAKER et al., 2016; DYER-WITHEFORD; KJØSEN; STEINHOFF, 2019; JOLER; PASQUINELLI, 2020; SHADOWEN, 2017).

Apesar da imagem hegemônica de neutralidade técnica e automatismo dos sistemas inteligentes, muitas vezes apresentados como podendo nos ajudar a “detectar, remover ou reduzir viéses humanos” (AI100, 2016), já são bastante conhecidos na literatura os casos de *algoritmos racistas* que falham seletivamente no reconhecimento facial de negros e asiáticos, ou que os classifi-

cam negativamente, *algoritmos misóginos* que classificam mulheres inferiormente aos homens em diferentes contextos, *algoritmos capacitistas* que discriminam pessoas com deficiência, *algoritmos discriminatórios* que recusam serviços a pessoas com base em seus códigos postais e muitos outros (ANGWIN et al., 2016; CARRERA, 2020; CRAWFORD; CALO, 2016; CRAWFORD, 2019; CRAWFORD; PAGLEN, 2019; CRAWFORD, 2019; DEBRUSK, 2018; GALLOWAY, 2019; PRATES; AVELAR; LAMB, 2019; RICHARDSON; SCHULTZ; CRAWFORD, 2019; SHADOWEN, 2017; REESE, 2016).

Tipicamente, quando a não-neutralidade de sistemas inteligentes e algoritmos é reconhecida como um problema, as soluções tendem a ser técnicas, baseadas em alterações nos parâmetros ou procedimentos, na introdução de procedimentos corretivos no sistema ou até mesmo no desenvolvimento de sistemas voltados para a detecção e correção de vieses (SURESH; GUTTAG, 2020; PRATES; AVELAR; LAMB, 2019; SIMPSON, 2018). Como já o disseram Crawford e Paglen (2019), “em lugar de reconhecer a escala e a natureza sistêmica do problema, companhias de tecnologia têm respondido” com “desenviesamento técnico [*technical debiasing*]” por “correções nos algoritmos e diversificação dos conjuntos de dados [data sets]”, o que além de insuficiente, “levanta sérias preocupações sobre privacidade e consentimento”. Além disso, nesses casos nos deparamos com uma regressão infinita: como lidar com o viés do sistema usado para corrigir o viés de outro sistema? Crawford e Paglen (2019, p. 12) argumentam que o problema só é de fato encarado quando as “desigualdades estruturais subjacentes” (racismo, machismo, capacitismo, etc.) são abordadas, a começar pela “crescente assimetria de poder entre aqueles que produzem e se beneficiam da inteligência artificial, e aqueles sujeitados às suas aplicações”.

3 Teoria ator-rede e o conceito de mediação técnica

Neste contexto, a teoria ator-rede (TAR) é uma proposta teórico-metodológica que surge na França na década de 1980 nos nomes de Michel Callon, Bruno Latour, John Law, e outros. Ela se propõe a analisar a sociedade não como um objeto dado – *o social* –, mas sim como um emaranhado dinâmico de atores – humanos ou não-humanos – relacionados uns aos outros, ao mesmo

tempo produto e produtores dessa *rede de atores*. Por isso, ela é também chamada de *sociologia da associação*, ou ainda *sociologia da tradução* (LATOUR, 2005).

As redes de atores seriam então coletivos, conjuntos heterogêneos em que não haveria mais uma necessidade de distinção *a priori* entre natureza e cultura, ou ainda entre agentes humanos e não-humanos. Entende-se então que o que existem são híbridos: quase-sujeito, quase-objetos que essas redes heterogêneas não cessariam de multiplicar (LATOUR, 1994a).

Neste contexto, se torna importante a noção de mediação – em particular *mediação técnica* – na medida que o campo tecnológico passa a ser composto, não por artefatos intermediários e desprovidos de agência, mas por elementos ativos mediadores:

Um intermediário, em meu vocabulário, é o que transporta significado ou força sem transformação: definir sua entrada é suficiente para definir sua saída. Na prática, um intermediário pode ser tomado não só como uma caixa preta, mas também como uma caixa preta que conta como um, mesmo se é composto internamente por muitas partes. Mediadores, por outro lado, não podem ser contados como somente um; podem contar por um, por nenhum, por vários ou por infinito. Sua entrada nunca é um bom preditor de sua saída; sua especificidade tem que ser levada em conta toda vez. Mediadores transformam, traduzem, distorcem, e modificam o significado ou os elementos que eles devem carregar. (LATOUR, 2005, p. 29, tradução nossa).¹

Desta forma, toda ação mediada tecnicamente resultaria em desvio, mudança, tradução. Isso implica que, ao se utilizar uma forma de tecnologia, qualquer processo alteraria seu *programa de ação*, de forma que, ao se mudar os meios para concretização de uma tarefa, estaríamos também alterando sua finalidade (LATOUR, 2002, p. 252).

A hipótese latouriana vai além ao pensar que

¹An intermediary, in my vocabulary, is what transports meaning or force without transformation: defining its inputs is enough to define its outputs. For all practical purposes, an intermediary can be taken not only as a black box, but also as a black box counting for one, even if it is internally made of many parts. Mediators, on the other hand, cannot be counted as just one; they might count for one, for nothing, for several, or for infinity. Their input is never a good predictor of their output; their specificity has to be taken into account every time. Mediators transform, translate, distort, and modify the meaning or the elements they are supposed to carry.

ação mediada tecnicamente não é executada nem por um sujeito que *utiliza* uma tecnologia, nem unicamente pela tecnologia, mas por um terceiro agente (ou actante), um agenciamento de ambos. Nem o ser-humano, nem o avião voa, mas sim, o agenciamento humano-avião ou avião-humano. Ao se segurar uma arma de fogo, uma pessoa não se torna má, ao ser segurada por alguém uma arma não vira automaticamente causa de morte. Mas em uma situação concreta, um sujeito que participa de uma briga e uma arma que descansa em seu bolso podem, ao mesmo tempo e mutuamente, se transformar no humano-arma assassino (LATOUR, 1994b, p. 33).

No caso do aprendizado de máquina, não é nem o ser humano, e nem a máquina, que pensa ou aprende, mas sim o agenciamento humano-máquina. Ao ser programado, um computador não se torna um simples reflexo do programador humano e de seus preconceitos, e ao usar o computador, o humano não se liberta de seus preconceitos por meio de um dispositivo neutro. Mas em uma situação concreta, uma pessoa-com-computador pode aumentar a escala e o efeito de seus preconceitos, inclusive por apresentá-los como neutros e objetivos. Em nenhum dos casos (avião, arma, computador) podemos dizer que a mediação técnica determinou a sequência de fatos, mas também não podemos ignorar sua agência.

Outra forma de desvio, ao mesmo tempo diferente – por exibir mais atores e temporalidades – e similar – por também mostrar a expressão de valor inscrito na tecnologia –, é o exemplo das lombadas de trânsito como forma de garantir a segurança nos arredores de escolas ou hospitais:

O objetivo do motorista é traduzido, por meio de uma lombada, de “diminuir a velocidade para não colocar estudantes em risco” para “diminuir para proteger a suspensão de meu carro”. Os dois objetivos estão distantes um do outro, e nós reconhecemos aqui o mesmo deslocamento que na história da arma de fogo. A primeira versão apela para a moralidade, desinteresse esclarecido, e reflexão, enquanto que a segunda apela para puro egoísmo e ato reflexo. [...] O motorista modifica seu comportamento através da mediação da lombada: ele muda de moralidade para força. (LATOUR, 1994b, p. 38, tradução nossa).²

²The drivers goal is translated, by means of the speed bump, from “slow down so as not to endanger students” into “slow down and protect my cars suspension.” The two goals are far apart, and we recognize here the same displacement as in our gun story. The

Moralidade – e política, e cultura, e direito – está inscrita nas coisas. O exemplo da lombada mostra uma mudança clara de conotação moral, mas este não é o único tipo de desvio. Por vezes, antes de uma regra ser concretizada em leis e exigências, ela já está presente nos objetos tecnológicos (LATOUR, 2002, p. 258).

4 Agência social das mediações técnicas

Podemos perceber então que o conceito latouriano de mediação técnica incorpora imediatamente uma ruptura com a ideia de neutralidade da tecnologia. Com isso, artefatos, processos e serviços passam a uma posição ativa na sociedade, produzindo, reproduzindo ou alterando valor de diferentes formas.

Uma tecnologia criada, incorpora em si uma dobra de temporalidades, espacialidades e actantes, envolvidos desde sua criação, desenvolvimento e cristalização em uma forma específica (FERREIRA, 2010, p. 249). Além disso – e ao considerar que ação intencionada não é uma propriedade nem exclusiva de humanos, nem de não-humanos, mas sim de *coletivos* (LATOUR, 1994b, p. 46) –, se essas etapas se dão no interior de um *coletivo* (ou instituição) composto em sua maioria por agentes interessados numa determinada divisão de sociedade, a tecnologia produzida e reproduzida promoverá essa separação.

Isso nos oferece perspectivas potentes para pensar o problema do viés no aprendizado de máquina e nos sistemas inteligentes. Como vêm enfatizando diversas autoras e diversos autores (CAMPOLO; CRAWFORD, 2020; CRAWFORD, 2015; CRAWFORD; SCHULTZ, 2019; PASQUINELLI; JOLER, 2020), para além dos dados de entrada e dos diversos usos que um usuário poderia fazer de um determinado programa de computador, ele pode conter em si mesmo, instituído, formas de opressão. A agência social das mediações técnicas nos permite assim perceber que não é arbitrário que um sistema de tradução automática reproduza esteriótipos de gênero também presentes em discursos de “inteligências naturais” (PRATES; AVELAR; LAMB, 2019) ou que um sistema de busca associe palavras negativas à imagens

drivers first version appeals to morality, enlightened disinterest, and reflection, whereas the second appeals to pure selfishness and reflex action. [...] The driver modifies his behavior through the mediation of the speed bump: he falls back from morality to force.

de corpos negros e positivas à de corpos brancos (CARRERA, 2020).

5 Conclusão

O conceito de mediação técnica tem muito a contribuir para a análise da ação que os objetos técnicos desempenham na sociedade. Sobretudo ao pensar mediação como tradução ou desvio, podemos pensar que toda ação mediada tecnicamente incorpora valores intrínsecos à tecnologia utilizada para tal – ao mudar os meios, mudamos o fim.

Desta forma, é possível pensar que – para além do uso que se faz desta – tecnologia nascida no seio de uma sociedade que é hegemonicamente classista, racista, sexista e eurocêntrica incorpora, reproduz, transforma e produz valores e atos concretos que legitimam e amplificam essas opressões. Sim, existe agência própria à objetos técnicos, mas ela não é arbitrária, mas sim em favor da reprodução dos valores presentes no contexto de sua criação/reprodução, na instituição que os mobilizou.

Referências

- AI100. Artificial intelligence and life in 2030: one hundred year study on artificial intelligence. *Report of the 2015 Study Panel*, 2016. Stanford University.
- ANGWIN, J. et al. *Machine bias*. *ProPublica*, May 23, 2016. 2016. Disponível em: <<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>>. Acesso em: 14 out. 2020.
- BOYD, D.; CRAWFORD, K. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, v. 15, n. 5, p. 662–679, 2012. Publisher: Taylor & Francis.
- CAMPOLO, A.; CRAWFORD, K. Enchanted Determinism: Power without Responsibility in Artificial Intelligence. *Engaging Science, Technology, and Society*, v. 6, p. 1–19, 2020.
- CAMPOS, P. E. F. DE; SANTOS DIAS, H. J. DOS. A insustentável neutralidade da tecnologia: o dilema do Movimento Maker e dos Fab Labs. *Liinc em Revista*, v. 14, n. 1, 2018.
- CARRERA, F. Racismo e sexismo em bancos de imagens digitais: análise de resultados de busca e atribuição de relevância na dimensão financeira/profissional. *COMUNIDADES, ALGORITMOS E ATIVISMOS DIGITAIS*, p. 139, 2020.
- CRAWFORD, K.; WHITTAKER, M. et al. *The AI now report: The social and economic implications of artificial intelligence technologies in the near-term*. Report prepared for the AI Now public symposium, hosted by the White House and New York University's Information Law Institute. 2016.
- CRAWFORD, K. Can an algorithm be agonistic. *Ten Scenes from Life in Calculated Publics*. *ResearchGate*, v. 41, n. 1, 2015.
- Regulate facial-recognition technology. *Nature*, v. 572, n. 7771, p. 565–565, 2019. Publisher: Macmillan Publishers Ltd., London, England.
- CRAWFORD, K.; CALO, R. There is a blind spot in AI research. *Nature*, v. 538, n. 7625, p. 311–313, 2016. Publisher: Nature Publishing Group.
- CRAWFORD, K.; PAGLEN, T. Excavating AI: The politics of images in machine learning training sets. *Excavating AI*, 2019.
- CRAWFORD, K.; SCHULTZ, J. AI Systems as State Actors. *Columbia Law Review*, v. 119, n. 7, p. 1941–1972, 2019. Publisher: JSTOR.
- DAGNINO, R. *Neutralidade da ciência e determinismo tecnológico: um debate sobre a tecnociência*. Unicamp, 2008.
- DEBRUSK, C. The risk of machine-learning bias (and how to prevent it). *MIT Sloan Management Review*, 2018.
- DURKHEIM, É. *As formas elementares da vida religiosa*. (Trad. Paulo Neves). São Paulo: Martins Fontes, 1996.
- DYER-WITHEFORD, N.; KJØSEN, A. M.; STEINHOFF, J. *Inhuman Power: Artificial Intelligence and the Future of Capitalism*. Pluto Press, 2019. ISBN 978-0-7453-3860-6. Disponível em: <<http://www.jstor.org/stable/j.ctvj4sxc6>>.
- FERREIRA, P. P. Por uma definição dos processos tecnicamente mediados de associação. *Revista brasileira de ciência, tecnologia e sociedade*, v. 1, n. 2, p. 58–75, 2010.

- GALLOWAY, A. *Are Algorithms Biased?* Library Catalog: cultureandcommunication.org. 26 jan. 2019. Disponível em: <<http://cultureandcommunication.org/galloway/are-algorithms-biased>>. Acesso em: 30 mar. 2020.
- JOLER, V.; PASQUINELLI, M. Nooscope. *Künstliche Intelligenz und Medienphilosophie HfG Karlsruhe*, 2020. Disponível em: <<https://nooscope.ai>>.
- LATOURE, B. *Jamais fomos modernos*. Editora 34, 1994a.
- Morality and technology. Tradução: Couze Venn. *Theory, culture & society*, v. 19, n. 5, p. 247–260, 2002. Publisher: Sage Publications London, Thousand Oaks and New Delhi.
 - On technical mediation. *Common Knowledge*, v. 3, n. 2, p. 29–64, 1994b.
 - *Reassembling the social: an introduction to actor-network-theory*. Oxford ; New York: Oxford University Press, 2005. 301 p. (Clarendon lectures in management studies). OCLC: ocm58054359. ISBN 978-0-19-925604-4.
- PASQUINELLI, M.; JOLER, V. The Nooscope manifested: artificial intelligence as instrument of knowledge extractivism. *KIM research group (Karlsruhe University of Arts and Design) and Share Lab (Novi Sad)*, 2020. Disponível em: <<https://nooscope.ai>>.
- PRATES, M. O. R.; AVELAR, P. H.; LAMB, L. C. Assessing gender bias in machine translation: a case study with Google Translate. *Neural Computing and Applications*, 27 mar. 2019. ISSN 0941-0643, 1433-3058. DOI: [10.1007/s00521-019-04144-6](https://doi.org/10.1007/s00521-019-04144-6). Disponível em: <<http://link.springer.com/10.1007/s00521-019-04144-6>>. Acesso em: 16 abr. 2020.
- REESE, H. Bias in machine learning, and how to stop it. *TechRepublic*, 2016.
- RICHARDSON, R.; SCHULTZ, J. M.; CRAWFORD, K. Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. *NYUL Rev. Online*, v. 94, p. 15, 2019. Publisher: HeinOnline.
- ROSA, S. E. DA; STRIEDER, R. B. Não Neutralidade da Ciência-Tecnologia: verbalizações necessárias para potencializar a constituição de uma cultura de participação, 2020.
- SHADOWEN, A. *Ethics and Bias in Machine Learning: A Technical Study of What Makes Us Good*(Academic Works, CUNY). 2017.
- SIMPSON, M. IBM's new bias detection software, democratizing AI. *IT World Canada September 24th*, 2018. Disponível em: <<https://www.itworldcanada.com/article/how-ibms-new-bias-detection-software-has-already-helped-companies/409206>>.
- SURESH, H.; GUTTAG, J. V. *A Framework for Understanding Unintended Consequences of Machine Learning*. 2020. _eprint: 1901.10002.