

Sabina Leonelli, *A pesquisa científica na era do big data: cinco maneiras como o big data prejudica a ciência, e como podemos salvá-la*. Tradução de Carla Cristina Munhoz Xavier. Rio de Janeiro, Editora Fiocruz, 2022.

Por Thiago Bicudo Castro  
Senac-SP  
<https://orcid.org/0000-0003-2825-956X>

*Perspectivas para a pesquisa sociológica na era do big data*

*A pesquisa científica na era do big data: cinco maneiras como o big data prejudica a ciência, e como podemos salvá-la*, de Sabina Leonelli, oferece um panorama conceitual sobre o *big data*, objeto de estudos do livro, e promove ao leitor um detalhamento crítico sobre suas implicações para a pesquisa científica. A obra

está organizada em quatro capítulos (além da introdução e conclusão), que viabilizam uma leitura clara, rápida e concisa. Não se trata de um livro destinado apenas aos especialistas nos assuntos referentes ao campo científico ou às tecnologias informacionais (TI). A utilização de uma linguagem direta, sem jargões técnicos ou acadêmicos, e mesmo sem excessos conceituais e categoriais, permite que o livro possa ser incorporado aos debates introdutórios sobre o tema em questão.

A autora, Sabina Leonelli, professora de Filosofia da Ciência na Universidade de Exeter (Reino Unido), possui larga experiência nos temas que relacionam a produção científica com as novas tecnologias informacionais. Sua preocupação em compreender a maneira como ocorre a produção social de dados digitais e as implicações disso no campo científico a conduz a uma discussão mais profunda sobre a epistemologia, que pode ser observada no conjunto de sua obra. No livro ora resenhado, Leonelli se propõe a esclarecer dois aspectos centrais quando se trata da relação “ciência x *big data*”:

A primeira é o elo entre a produção de dados e a produção de conhecimento e a forma como os dados devem ser gerenciados para confirmar ou negar uma asserção; a segunda é a enorme dificuldade e os imensos recursos necessários para processar e analisar os dados, a fim de usá-los para criar interpretações confiáveis sujeitas a avaliação crítica (p. 17).

Uma das contribuições mais valiosas do livro para aqueles que buscam não apenas saber em que consiste o *big data*, mas também aos cientistas que pretendem abordar o *big data* em suas pesquisas, seja como objeto de estudos ou ferramenta de pesquisa, é uma conceituação que consegue escapar dos limites conceituais propostos ou impostos pelas áreas da informá-

tica e TI. A autora problematiza, no primeiro capítulo, a ideia comum de traduzir o *big data* com base em suas características mais rígidas, os denominados “sete ‘V’s’ do big data”<sup>1</sup>. Sua crítica, portanto, toma esses dados a partir da forma como são usados. A maneira como são produzidos e como circulam entre os diferentes setores sociais, e a perspectiva de estabelecer conexões e diálogos entre setores que, em outros momentos, se mostravam inviáveis dadas as barreiras sociais e dificuldades técnicas (p. 24).

Portanto, a “mobilidade” aparece como uma das características do *big data*, e, do ponto de vista filosófico e epistemológico, proposto pela autora, seria o fator mais importante a ser considerado pelos cientistas ao se depararem com esses dados. A *mobilidade* decorre “de sua capacidade de viajar por diferentes situações de análise e reutilização e de se relacionar com o maior número possível de tipos de dados.” (p. 31).

O segundo capítulo, “Sinais de alerta: cinco maneiras como o big data prejudica a pesquisa”, conecta-se a uma parte da discussão proposta no capítulo anterior, a classificação, formatação, armazenamento e uso de bancos de dados e seus efeitos sobre a análise e a interpretação dos dados. Partindo disso, o *conservadorismo* e o *problema dos dados antigos* são o primeiro ponto apresentado por Leonelli para mostrar o problema que esses elementos trazem aos resultados de pesquisa. A pesquisa com *big data* não deve prescindir da teoria, como se a centralidade dos dados levasse à “morte da teoria”. Quando a autora fala em “conservadorismo”, existe uma crítica à utilização de dados antigos, “cujas características e métodos de gestão se tornam cada vez mais obscuros com o tempo, em vez de estimular a produção de novos dados cujas características respondam de maneira específica às necessidades e circunstân-

cias de quem os usa” (p. 53). Trata-se, portanto, de um aspecto prejudicial à ciência na medida em que leva ao desenvolvimento de pesquisas pouco criativas, que não exploram a evolução contínua e cada vez mais sofisticada do mundo. Embora possa ser constatada uma prevalência de exemplos das áreas das Ciências Naturais e médicas, eis que aparece um dos primeiros problemas epistemológicos para as Ciências Sociais, que vai decair em possibilidades de mitificação da realidade social.

O *big data* pode ainda trazer pouca segurança devido aos *dados não confiáveis*, o segundo problema apontado no capítulo. Aqui a discussão é centrada na maneira como as bases de dados e suas formas de gestão, classificação e organização são, quase sempre, responsáveis por descontextualizar os dados. Torna-se difícil, neste cenário, identificar a potencialidade dos dados como fonte de novos conhecimentos. Não existe uma solução universal para esse problema. De acordo com a autora, a responsabilidade de investigar a confiabilidade dos dados, bem como sua recontextualização, deve ser de cada pesquisador, orientado com base nos critérios de cada área de pesquisa. Para o campo das Ciências Sociais, podemos indagar sobre o problema dos analistas de dados sociais, que também é atravessado por essa questão abordada no tópico. O risco de analistas sem os conhecimentos teóricos e metodológicos específicos dessas ciências reunirem uma enorme quantidade de dados sem observarem suas nuances socioculturais e potenciais diferenças de outras ordens é acabar tomando-os como fonte de conhecimento. Uma ciência centrada em dados não pode existir na ausência de confiabilidade e do entendimento de como os dados são mobilizados (p. 53).

A reflexão sobre as maneiras como a pesquisa científica é afetada pelo *big data* proposta

por Leonelli segue uma ordem lógica, na qual cada elemento posterior que é levantado ajuda a complementar a discussão do aspecto precedente. O terceiro tópico da lista de problemas é a *mistificação*, que nos leva aos dados parciais, que representam uma parte muito seletiva da realidade. A autora problematiza essa questão a partir de um argumento que vai de encontro àqueles que são entusiastas da utilização do *big data* na pesquisa científica. Leonelli afirma que tais dados fornecem “informações sobre muito pouco e de uma forma que tende a impedir, ou pelo menos dificultar, qualquer tipo de oposição construtiva.” (p. 56). Há, inclusive, apontamentos conjunturais e um diagnóstico sociopolítico e científico-tecnológico:

Ao contrário da visão do Big Data e Dados Abertos como portadores da democracia e instigadores da participação social na pesquisa, a forma como a ciência é governada e financiada não parece ser desafiada pelo Big Data, mas, sim, pela desigualdade de poder e visibilidade entre as diferentes nações e comunidade científica que continua a crescer (p. 58).

A consequência direta à qual esse tópico leva é um desdobramento da situação de *divisão digital* para uma de *divisão de dados* (p. 58).

Se no tópico anterior a autora propõe uma análise mais ampla e diagnóstica do contexto de utilização de *big data* ao melhor estilo de uma sociologia compreensiva à Max Weber, no seguinte, ao abordar a *corrupção* e o *problema dos dados desonestos*, ela se volta à noção de mercadoria no sentido marxista para situar a produção e a posse de dados em amplos setores (p. 59). Aborda-se uma das questões que têm mobilizado discussões no campo da sociologia digital: a privatização do *big data* por grupos, denominados pela autora como *oligarquias* ou

*monopólios*. Seriam esses grupos os responsáveis por “participar da construção de ferramentas e estratégias de análise e interpretação” dos dados; e de produzir informação e conhecimento que ficam restritos a esses próprios grupos, sem grandes possibilidades de abertura desses dados privados.

Dois aspectos podem ser discutidos a partir dessas considerações de Leonelli. O primeiro é de caráter teórico-metodológico sobre a maneira como a autora mobiliza categorias marxianas (*produção, mercadoria e alienação*) como conceitos para adentrar no problema das classes sociais, que por sua vez não é citado na discussão. Portanto, aquelas categorias tomadas como conceitos, até mesmo apriorísticos, não desmistificam a realidade social, pois a mesma autora não assume uma posição teórica quanto a entender o sujeito da tal produção de informação. Fica a ideia de que os “grupos oligárquicos” ou o “monopólio”, ou seja, a classe social dominante é quem de fato produz toda a informação e conhecimento contido nos dados. A categoria *classe social*, portanto, fica de fora da análise de Leonelli. Existe ainda uma referência aos debates sociológicos sobre as “formas de exploração ou trabalho não remunerado” que ocorrem no setor das TI’s. Este é um debate complexo e que ainda tem repercutido entre sociólogos e economistas marxistas<sup>2</sup>.

O segundo aspecto é mais bem elaborado pela autora, pois mobiliza todo aquele conjunto de discussões já consagrados em sua obra. Refere-se aos impactos dessa privatização dos dados para o mundo da pesquisa científica. A rigor, os grupos econômicos, empresas etc. têm o poder de controlar a mobilidade dos dados, restringindo o acesso aberto e/ou quem poderá acessá-los.

Na sequência, ao levantar a problemática dos *dados sociais e confidenciais*, o livro abre

portas para uma discussão que encontra ressonância nas ciências políticas e na própria sociedade civil. Essa discussão diz respeito a todo o conjunto de dados que as pessoas produzem cotidianamente em redes sociais, compras *on-line*, trajetos na malha urbana e afins. Mas não apenas essa produção de dados a discussão abrange. Relacionando-se com o tópico anterior, aqui a autora comenta sobre as utilizações que as empresas fazem dessas bases de dados para fins políticos ou comerciais. E esse é um debate polêmico e relevante atualmente, que nos leva ao aspecto que fecha o capítulo 2, *a ética como parte integrante da ciência*.

Esclarecer sobre o uso das ferramentas, verificar a validade dos dados, a segurança e o potencial impacto do *big data* sobre as pessoas e as comunidades correspondem à compreensão de que as “*questões éticas e sociais sejam vistas como parte integrante das necessidades técnicas e científicas associadas à gestão e análise de dados*” (p. 71, itálico da autora).

O terceiro capítulo é, assim como o primeiro, o mais interessante para pesquisadores das ciências humanas e sociais. Nesse capítulo a autora promove uma discussão de caráter teórico que pode ser incorporada e aprofundada por cientistas de diversas áreas do saber. O cerne do debate é confrontar as perspectivas *representativas e relacionais* sobre como interpretar o *big data*.

Na perspectiva representativa, os dados são “objetos de conteúdo fixo e imutável, cujo significado como representações da realidade deve ser investigado e revelado progressivamente por meio do uso correto de métodos científicos.” (p. 75). Uma das críticas a essa visão é a de que se coaduna a uma retórica frequentemente associada ao *big data*, de acumulação de conhecimento de forma indutiva por meio de dados. Em síntese, seria a ideia

de que quanto mais dados, mais fatos temos, logo, maiores possibilidades de obter conhecimentos (p. 77).

A alternativa apresentada seria tomar os dados de forma *relacional*, ou seja, os dados não refletem imediatamente a sociedade, tampouco servem como representações da realidade; os dados devem ser percebidos como objetos que se relacionam com uma questão não resolvida. “Na visão relacional, *qualquer objeto pode desempenhar o papel de ‘dados’ desde que (1) seja tratado como uma fonte potencial de conhecimento empírico e (2) possa ser mobilizado para torná-lo acessível a mais pessoas*” (p. 83, itálico da autora). A abordagem relacional é uma solução para os problemas apresentados pela abordagem representativa. A maneira como a visão relacional interpreta os dados faz coro à proposta da socióloga Deborah Lupton (2015) de pensar o *big data* como artefatos socioculturais. No livro de Leonelli, no qual há uma defesa da visão relacional, os dados ganham contornos históricos e não estáticos, vinculados às experiências e reconstruídos a fim de verificar sua origem e garantir validade (p. 85).

Eu encerro esta resenha apontando para um fato importante. Algumas das críticas apresentadas por Leonelli podem ser consideradas mais relevantes em certos campos científicos do que em outros. O impacto do *big data* na pesquisa científica pode variar dependendo da disciplina, e nem todas as preocupações levantadas pela autora se aplicam universalmente. É importante notar que os leitores devem estar cientes de que as discussões trazidas pelo livro podem não se aplicar igualmente a todas as áreas de pesquisa.

As áreas das ciências humanas e sociais podem se beneficiar das discussões do livro, embora não tenham sido aprofundadas pela autora. Há poucas menções à maneira como

essas ciências estão adentrando nesse terreno do *big data*. O primeiro capítulo pode oferecer, além de uma conceituação preliminar e menos técnica do *big data*, também algumas possibilidades de temas e objetos de pesquisa. As problemáticas epistemológicas que surgem sobre a dimensão teórica das ciências humanas, como, por exemplo, o modelo de pesquisa tradicional – baseado em hipóteses amparadas nas teorias e em observações qualitativas para a coleta de dados amostrais, testes e generalizações etc. –, não são aplicáveis para o tipo de informação do *big data*.

E mesmo o terceiro capítulo pode fornecer um ponto de partida para reflexões mais densas sobre como evitar uma diminuição do papel da teoria a uma situação secundária, contornando análises fragmentadas e desunificadas da realidade social e exercendo um papel crítico diante das iniciativas que podem surgir.

Por fim, é plausível afirmar que o livro em questão traz um apelo por uma mudança cultural na prática científica. A autora enfatiza a necessidade de questionar as suposições e premissas por trás do uso do *big data*, bem como a importância de reconhecer as limitações e os vieses embutidos nessas abordagens. Ela encoraja os cientistas a se tornarem mais críticos e reflexivos, adotando uma abordagem mais holística em relação aos dados e às suas interpretações.

### Referências bibliográficas

- BOLAÑO, C. R. S. & VIEIRA, E. S. (2015a), “The political economy of the internet: Social networking sites and a reply to Fuchs”. *Television & New Media*, 16 (1): 52-61. Disponível em <https://doi.org/10.1177/1527476414527137>.
- BOLAÑO, C. R. S. & VIEIRA, E. S. (2015b), “Digitalisation and labour: A rejoinder to Christian Fuchs”. *Triple C*, 13 (1): 79-83. Disponível em <https://doi.org/10.31269/triplec.v13i1.666>
- DANTAS, Marcos. (2003), “Informação e trabalho no capitalismo contemporâneo”. *Lua Nova: Revista de Cultura e Política*, 60. Disponível em <https://doi.org/10.1590/S0102-64452003000300002>.
- FUCHS, Christian. (2015). “Against divisiveness: Digital workers of the world unite! A Rejoinder to César Bolaño and Eloy Vieira”. *Television & New Media*, 16 (1): 62-71. Disponível em <https://doi.org/10.1177/1527476414528053>.
- LEONELLI, Sabina. (2022), *A pesquisa científica na era do big data: cinco maneiras como o big data prejudica a ciência, e como podemos salvá-la*. Tradução de Carla Cristina Munhoz Xavier. Rio de Janeiro, Editora Fiocruz.
- LUPTON, Deborah. (2015), *Digital sociology*. Londres, Routledge.

### Notas

1. Refere-se às seguintes características: *volatilidade, velocidade, valor, veracidade, validade, volume e variedade* (Leonelli, p. 24, 2022)
2. Certamente, os debates não se esgotam, nem se limitam ao conjunto de autores aqui elencados, mas oferecem uma dimensão a seu respeito. Ao aprofundarmos essas questões no campo marxista, por exemplo, é possível fazer referência não apenas ao trabalho de Dantas (2003), como também ao debate travado entre Bolaño (2015a; 2015b) e Fuchs (2015) a respeito da problemática concernente às categorias trabalho produtivo, trabalho improdutivo e sujeito trabalhador, bem como à maneira como elas se inserem no âmbito do trabalho digital.

Texto recebido em 01/06/2023 e aprovado em 07/08/2023.

DOI: 10.11606/0103-2070.ts.2023.212702.

