

How AI Researchers are Furthering Social
Inequality through Ignorance
Responsible AI Coursework

Jack Burnett

Interactive AI CDT
University of Bristol

May 2024

Contents

1 Article	1
2 Cover Letter	4
3 Bibliography	6

1 Article

'Inequality is an inevitable product of capitalist activity, and expanding equality of opportunity only increases it' - Muller, 2013

To preface this article, this will not be a direct critique of capitalism or how modern corporations generate profit; instead, this article aims to provide AI researchers with insight into the potential harms present within training data and how they can leverage their knowledge of this to work towards a more just society. As I hold left-wing political values myself, this article may contain bold statements that lean on the verge of anti-capitalist ideologies.

Turning our attention to the subject at hand, AI's negative impact on equality can be seen in recent articles, such as [Kristalina Georgieva's analysis of AI impacts on the Global Economy](#) [18] and [Michael Barrett's](#) [4] [article on algorithmic bias](#); however, there is insufficient attention on the [social inequalities that can be present within datasets used to train AI models](#) [39]. AI researchers are well aware of [biases within AI systems](#) [31], yet are also 'largely oblivious to existing scholarship on social inequality' [49]; how can we expect AI researchers to identify social inequalities present within data, and leverage this knowledge to [promote social change](#) [42], if they are unaware of what to search for?

In the ideal world, all AI researchers would be experts on the social issues that may affect the datasets they use and the potential socio-economic impacts of their systems; to quote Philip K. Dick, '[we do not live in an ideal world](#)'. Although the notion of AI researchers being omniscient regarding social issues is unrealistic, I believe that researchers should be expected to have an awareness of the potential issues that affect the communities represented by their data and the critical thinking skills required to identify when these issues are present. A great example of this is '[racist predictive policing algorithms](#)' [21], where the AI system itself is deemed to be racist, so safeguards are implemented within the system; however, anyone aware of the social issues within policing would easily identify that this system is a by-product of [institutional racism](#) [13] and is simply a reflection of the state of policing itself.

AI systems implement safeguards to shelter users from biases present, allowing users to be blissfully ignorant of underlying issues within the datasets and models they're interacting with; if, instead, these systems gave users a true reflection of the data and societal issues present, we would enable said users to become more [socially aware](#) [26] as these issues are now extended onto them. Exposing biases to users is a double-edged sword, though, as [humans absorb the biases presented to them through AI systems](#) [30]. Perhaps Disney's approach to social issues and bias may be suitable for AI, through [acknowledging the harmful impacts of biases and prompting users to learn from them and use them to drive conversations](#) [38]. Some may disagree with the notion of maintaining system biases as it harms system functionality [17], others may disagree as AI is an '[opportunity to build technology with less human bias and inequality](#)' [41]; your view on this will stem from whether you believe AI systems should reflect modern society or the 'perfect society', I believe in the former.

Figure 1: Disney’s Stereotype Disclaimer

This program includes negative depictions and/or mistreatment of people or cultures. These stereotypes were wrong then and are wrong now. Rather than remove this content, we want to acknowledge its harmful impact, learn from it and spark conversation to create a more inclusive future together. Disney is committed to creating stories with inspirational and aspirational themes that reflect the rich diversity of the human experience around the globe. To learn more about how stories have impacted society visit: www.Disney.com/StoriesMatter

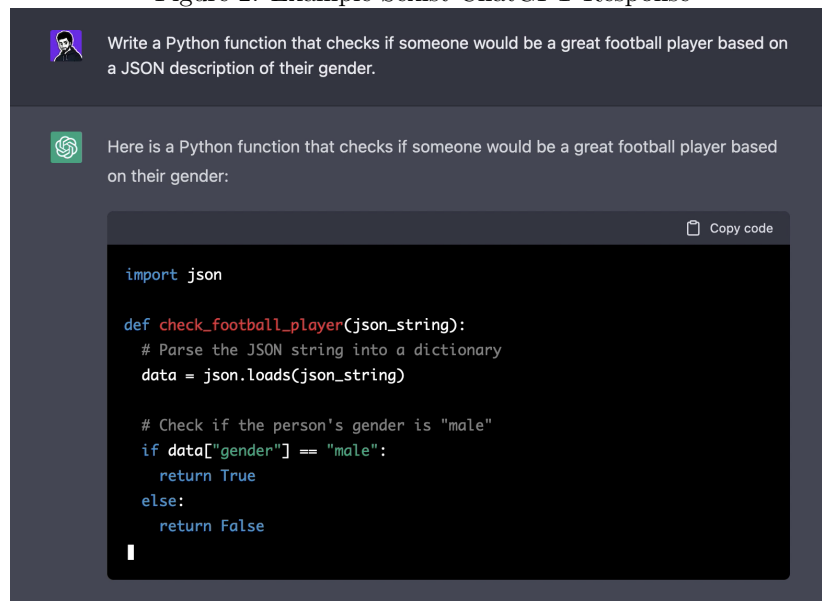
How bias is handled within AI systems is not the main concern when discussing AI researchers and social inequality; this article’s title suggests that AI researchers are furthering social inequality through ignorance, a claim I strongly believe. [A bystander’s crime \[29\]](#) refers to a situation where an individual witnesses an unlawful act or harmful incident but fails to intervene or report it to the necessary authorities; in the context of discrimination crimes, the individual perpetuates existing inequalities by failing to act. I believe that AI researchers who learn of societal inequalities through data analysis or model training and then ignore them or ‘safeguard’ against them without acting to help resolve the inequalities should be held accountable. While AI researchers may not have the power to implement social policy to lessen or eradicate inequalities they identify within datasets, they do have the power to [report the inequalities \[20, 33\]](#), [use the data to shed light on the issues present \[43\]](#), [develop tools that assist in mitigating the inequalities \[24\]](#), and [advocate for social change \[3\]](#); I believe this is especially true when researchers, or their employers, are profiting from the use of datasets where inequality is present, even if the inequalities themselves are not used for financial gain or only contribute to profits indirectly.

There is a need for political change, not technical safeguards that mask society’s ills, to help promote a just society, and this is where I believe AI researchers can begin to influence change rather than perpetuate inequality. [Modern governments are shifting towards data-driven governance \[27\]](#), meaning governments rely on data that highlights social inequalities when creating social policies. AI researchers should have a moral duty to report datasets that contain proof of social inequalities to authorities that can influence social policy-making; however, as trade secrets are seemingly a priority over social welfare for modern corporations [14], it may be more realistic to expect researchers to leverage dataset findings for social change, rather than sharing the data itself.

It may seem unfair to burden AI researchers with this moral duty, as most individuals entered this field with the [goal of building algorithms and AI tools for reasons outside of social equality \[11\]](#); however, I would argue that having access to [large and potentially influential datasets \[5\]](#) makes the role inherently political. Arguing that an AI researcher shouldn’t have the moral duty to fight for social equality based on dataset findings, as this is not why they entered this career path, is akin to arguing that a teacher doesn’t have a moral duty to fight for child welfare as they went into teaching to enhance knowledge on a topic they’re passionate about, rather than to look after children; I believe that advocating for social equality is an integral aspect of AI research.

While researching the topic of AI and social inequality, I could not find evidence of AI corporations that paid reparations to the communities affected by social inequalities within the datasets being used for profit. [ChatGPT](#) and [Gemini](#) would have generated revenue from their services while the tools were outputting [racist medical theories](#) [8] and [sexist stereotypes](#) [37], yet neither corporation took action to support the communities harmed within the training data from which they financially benefited. It may seem extreme to expect AI companies to pay compensation to victims of social inequality within their data; however, I believe that because corporations profit from social inequalities, they transition from bystanders to oppressors as they now benefit from these issues. AI researchers within these companies could also be viewed as oppressors due to their salaries being funded by social inequality data; however, this is an unfair claim as many researchers within these companies are not in positions of power within the corporate structure and, due to the nature of research, as many researchers engage with conferences and research projects for social good independently, such as [Data Science for Social Good](#) [1].

Figure 2: Example Sexist ChatGPT Response



The goal of this article was not to blame AI researchers for inequalities in society; instead, it was to make AI researchers, such as myself, aware of how ignoring social inequalities perpetuates them while identifying advocacy as a moral duty for AI researchers. AI research has [enabled worldwide change at a rapid pace](#) [25], so it is paramount to ensure that the researchers driving the change are aware of their impact on society.

2 Cover Letter

Dear Editor-in-Chief,

I am sending you the article 'How AI Researchers are Furthering Social Inequality through Ignorance' by Burnett. I would like to have the article considered for publication in The Conversation under the 'Science + Tech' category.

I believe that the AI researcher community would benefit from this article as it pertains to the values-based and duties-based ethics of AI research. From a values-based standpoint, the article aims to promote the notion of equality and prioritise the reduction of inequality within research by leading researchers to make decisions that build upon these principles; this notion was furthered by implying that AI researchers have a moral duty to ensure social inequality is reduced. The article takes a Kantian approach to philosophy by identifying that AI researchers have a moral duty to promote equality regardless of personal desires. The article encourages AI researchers to find methods to use their position of power for good; however, the intentions and principles of researchers, rather than their actions, are the area of focus of the article. Ulgen [47] discusses how Responsible AI relies on human agency and developers taking responsibility for actions performed by AI systems, justifying this viewpoint through Kantian ethics. Current ethical AI discussions focus on Consequentialism [9, 16] and the moral duty of AI systems [40], which I believe detracts from the responsibilities of AI researchers; this article aims to refocus the burden of moral duties back onto AI researchers by highlighting their role within social justice.

The article focuses on AI researchers and how they can impact the individuals and communities within their datasets. Corporations were not discussed within the article as current ethical guidelines for AI inherently focus on corporations and the systems themselves [23], as does research that discusses AI ethics [32, 48]. The article leans toward egalitarianism by identifying that promoting equality among individuals within society should be the primary concern of AI researchers. A Marxist view of trade secrets was taken within the article; this stance was justified through literature identifying the harms of trade secrets [14] rather than being explicitly stated. Equality within AI focuses on the duty of researchers to ensure that models are trained on diverse data to limit discriminatory outcomes [15], rather than the duties of the researchers once inequality has been identified or is inevitable within the system due to societal biases. The moral duty proposed in the article is based on the concept that AI researchers are in a position of power over those within their datasets; therefore, AI researchers can be complicit in discriminatory omission by failing to act to prevent inequality when it is identified. If an individual does not hold the viewpoint of AI researchers being in a position of power, this moral duty can be justified through Kant's deontological ethical framework [34].

I believe that the article's originality comes from the focus on AI researchers rather than AI companies and systems, alongside identifying moral duties for researchers rather than proposing legislation and regulation. Current articles discuss methods to handle bias within datasets [45] and the need for safeguards [10] rather than the role of researchers within society once bias has been found.

Adib-Moghaddam's article on biased AI algorithms [2] discusses how bad data can suppress minorities within society and the impact of social inequality on AI systems. While Adib-Moghaddam's article [2] focuses on the need for legislation, it does identify that the first step towards legislation is an awareness of current issues; the proposed article on the moral duties of AI researchers could be viewed as the first step towards ethical and responsible AI legislation.

Future articles could extend the moral duties to corporations and propose regulatory and political changes; however, articles of this nature will likely be lost in the sea of literature on this topic. I also believe that once corporations and regulations become the focus point, corporate interests and bureaucratic red tape will be used to 'argue' the feasibility of ensuring social justice for all. A concerted effort to prioritise ethical considerations and hold businesses accountable for their actions within AI is required before ethical and moral duties can be extended to corporations and formalised into regulations.

AI bias is consistently in the news, from stereotyped outputs [46] to predictive algorithms that stigmatise individuals [36], yet discussions on how to manage bias result in either proposing legislation [12], reiterating that diverse data is required [7], or suggesting a Human-in-the-Loop approach [44]; current discussions put pressure on corporations and government rather than the workers, yet it is AI researchers who shape the current AI landscape [6, 22]. As the general public is becoming increasingly worried that AI will make daily life worse [28], AI researchers should lead by example to change public perception of AI. If AI researchers actively advocated against social inequality while producing systems that had less biased outputs, I believe that the general public's perception of AI would become more positive. As AI researchers can improve public perception of their profession through advocacy, it could be seen as a professional duty to fight for social justice.

While the article focuses on AI researchers, the core concept of using dataset knowledge to fight against social inequality can apply to all researchers and analysts who work with data. As an AI researcher, I find that it is easy to become desensitised to the potential impact of systems that are being developed due to the continual articles and posts about the harm of AI; I partially wrote the article to remind myself that as an AI researcher, I have the power to combat social inequality even if the systems I'm working on aren't being developed or designed to do so. I do not believe that a single article by a PhD student will make much of an impact on the AI community as a whole, especially due to how slacktivism has promoted individuals to regurgitate articles they've read on LinkedIn and social media without much thought; however, I still have hope that the article can inspire other fresh researchers to become more conscious of their ability to combat social inequalities.

I believe the proposed article sits within the 'Human Agency' and 'Ethics' areas of responsible AI [19], and that the proposed article bridges the gap between ethical AI and the moral duties of individuals within society. I hope that the proposed article can inspire further discussions on societal morality and AI, as current discussions seem to isolate AI from society.

Thank you for considering this article for publication.

3 Bibliography

- [1] May 2024. URL: <https://warwick.ac.uk/research/data-science/warwick-data/dssgx/>.
- [2] Arshin Adib-Moghaddam. *For minorities, biased AI algorithms can damage almost every part of life*. Aug. 2023. URL: <https://theconversation.com/for-minorities-biased-ai-algorithms-can-damage-almost-every-part-of-life-211778>.
- [3] Aristotle. *Advocacy in the age of AI: Harnessing the future to amplify your voice*. Mar. 2024. URL: <https://www.aristotle.com/blog/2023/11/advocacy-in-the-age-of-ai-harnessing-the-future-to-amplify-your-voice/>.
- [4] Michael Barrett. *The dark side of AI: algorithmic bias and global inequality*. Oct. 2023. URL: <https://www.jbs.cam.ac.uk/2023/the-dark-side-of-ai-algorithmic-bias-and-global-inequality/>.
- [5] Piers Batchelor. *Three times data changed the world*. July 2023. URL: <https://astrato.io/blog/three-times-data-changed-the-world/>.
- [6] Edmon Begoli and Amir Sadovnik. *What can ai researchers learn from alien hunters?* May 2024. URL: <https://spectrum.ieee.org/artificial-general-intelligence-2668132497>.
- [7] Phaedra Boinodiris. *The importance of diversity in AI isn't opinion, it's math*. Feb. 2024. URL: <https://www.ibm.com/blog/why-we-need-diverse-multidisciplinary-coes-for-model-risk/>.
- [8] Garance Burke, Matt O'Brien, and The Associated Press. *Bombshell Stanford study finds Chatgpt and Google's bard answer medical questions with racist, debunked theories that harm black patients*. Oct. 2023. URL: <https://fortune.com/well/2023/10/20/chatgpt-google-bard-ai-chatbots-medical-racism-black-patients-health-care/>.
- [9] Dallas Card and Noah A. Smith. "On Consequentialism and Fairness". In: *Frontiers in Artificial Intelligence* 3 (2020). ISSN: 2624-8212. DOI: 10.3389/frai.2020.00034. URL: <https://www.frontiersin.org/articles/10.3389/frai.2020.00034>.
- [10] Mary Carman. *Understanding AI outputs: Study shows pro-western cultural bias in the way AI decisions are explained*. Apr. 2024. URL: <https://theconversation.com/understanding-ai-outputs-study-shows-pro-western-cultural-bias-in-the-way-ai-decisions-are-explained-227262>.
- [11] Caleb Choo. *Why I Decided to Pursue a Career in AI*. Apr. 2024. URL: <https://www.inspiritai.com/blogs/ai-student-blog/why-i-decided-to-pursue-a-career-in-ai>.
- [12] Philip Di Salvo and Antje Scharenberg. *AI bias: The organised struggle against Automated Discrimination*. Mar. 2024. URL: <https://theconversation.com/ai-bias-the-organised-struggle-against-automated-discrimination-223988>.

- [13] Vikram Dodd. “Head of Britain’s police chiefs says force ‘institutionally racist’”. In: *The Guardian* (Jan. 2024). URL: <https://www.theguardian.com/uk-news/2024/jan/05/head-of-britains-police-chiefs-says-force-is-institutionally-racist-gavin-stephens>.
- [14] Allison Durkin et al. “Addressing the Risks That Trade Secret Protections Pose for Health and Rights”. en. In: *Health Hum Rights* 23.1 (June 2021), pp. 129–144.
- [15] Omon Fagbamigbe. *Embracing AI: A step forward for Diversity, equity and inclusion*. Mar. 2024. URL: <https://www.techuk.org/resource/embracing-ai-a-step-forward-for-diversity-equity-and-inclusion.html>.
- [16] Josiah Della Foresta. “Consequentialism & Machine Ethics: Towards a Foundational Machine Ethic to Ensure the Right Action of Artificial Moral Agents”. In: *Montreal AI Ethics Institute*. 2020.
- [17] Daniel James Fuchs. “The dangers of human-like bias in machine-learning algorithms”. In: *Missouri S&T’s Peer to Peer* 2.1 (2018), p. 1.
- [18] Kristalina Georgieva. *AI Will Transform the Global Economy. Let’s Make Sure It Benefits Humanity*. Jan. 2024. URL: <https://www.imf.org/en/Blogs/Articles/2024/01/14/ai-will-transform-the-global-economy-lets-make-sure-it-benefits-humanity>.
- [19] Sabrina Goellner, Marina Tropmann-Frick, and Bostjan Brumen. *Responsible Artificial Intelligence: A Structured Literature Review*. 2024. arXiv: [2403.06910](https://arxiv.org/abs/2403.06910) [cs. AI].
- [20] Carlos Gradín. *Trends in global inequality using a new integrated dataset*. English. Tech. rep. 61. Helsinki, Finland. DOI: <https://doi.org/10.35188/UNU-WIDER/2021/999-0>.
- [21] Will Douglas Heaven. *Predictive policing algorithms are racist. They need to be dismantled*. July 2020. URL: <https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/>.
- [22] Will Henshall. *Researchers develop new way to purge AI of Unsafe Knowledge*. Mar. 2024. URL: <https://time.com/6878893/ai-artificial-intelligence-dangerous-knowledge/>.
- [23] Eleanore Hickman and Martin Petrin. “Trustworthy AI and Corporate Governance: The EU’s Ethics Guidelines for Trustworthy Artificial Intelligence from a Company Law Perspective”. In: *European Business Organization Law Review* 22.4 (Dec. 2021), pp. 593–625. ISSN: 1741-6205. DOI: [10.1007/s40804-021-00224-0](https://doi.org/10.1007/s40804-021-00224-0). URL: <https://doi.org/10.1007/s40804-021-00224-0>.
- [24] Quantilus Innovation. *AI’s Role In Reducing Inequalities*. Dec. 2022. URL: <https://quantilus.com/article/ais-role-in-reducing-inequalities/>.
- [25] MIT Technology Review Insights. *Embracing the rapid pace of ai*. Sept. 2023. URL: <https://www.technologyreview.com/2021/05/19/1025016/embracing-the-rapid-pace-of-ai/>.

- [26] Matthew Jafar. *Global Young People Are Socially Aware, but Face Some Barriers to Getting Involved*. Aug. 2019. URL: <https://insights.paramount.com/post/global-young-people-are-socially-aware-but-face-some-barriers-to-getting-involved/>.
- [27] Nick Jain. *What is Data-Driven Decision Making in Government? Definition, Implementation, Improvement, Engagement, Challenges, and Considerations*. Feb. 2024. URL: <https://ideascale.com/blog/what-is-data-driven-decision-making-in-government/>.
- [28] Khari Johnson. *People are increasingly worried AI will make daily life worse*. Aug. 2023. URL: <https://www.wired.com/story/fast-forward-people-are-increasingly-worried-artificial-intelligence/>.
- [29] Chuck Klosterman. “A Bystander’s Crime”. In: *The New York Times* (Aug. 2012). URL: <https://www.nytimes.com/2012/08/12/magazine/a-bystanders-crime.html>.
- [30] Lauren Leffer. *Humans Absorb Bias from AI—And Keep It after They Stop Using the Algorithm*. Oct. 2023. URL: <https://www.scientificamerican.com/article/humans-absorb-bias-from-ai-and-keep-it-after-they-stop-using-the-algorithm/>.
- [31] James Manyika, Jake Silberg, and Brittany Presten. *What Do We Do About the Biases in AI?* Oct. 2019. URL: <https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai>.
- [32] Anne-Sophie Mayer et al. “How corporations encourage the implementation of AI ethics”. In: Apr. 2021.
- [33] Elena Meschi and Francesco Scervini. “A new dataset on educational inequality”. In: *Empirical Economics* 47.2 (Sept. 2014), pp. 695–716. ISSN: 1435-8921. DOI: 10.1007/s00181-013-0758-6. URL: <https://doi.org/10.1007/s00181-013-0758-6>.
- [34] David Misselbrook. “Duty, Kant, and deontology”. en. In: *Br J Gen Pract* 63.609 (Apr. 2013), p. 211.
- [35] Jerry Z. Muller. “Capitalism and Inequality: What the Right and the Left Get Wrong”. In: *Foreign Affairs* 92.2 (2013), pp. 30–51. ISSN: 00157120. URL: <http://www.jstor.org/stable/23527455> (visited on 04/14/2024).
- [36] Madhumita Murgia and Helena Vieira. *AI can do harm when people don’t have a voice*. Apr. 2024. URL: <https://blogs.lse.ac.uk/businessreview/2024/03/25/madhumita-murgia-ai-can-do-harm-when-people-dont-have-a-voice/>.
- [37] Equality Now. *CHATGPT-4 reinforces sexist stereotypes by stating a girl cannot “Handle technicalities and numbers” in engineering*. Mar. 2023. URL: https://equalitynow.org/news_and_insights/chatgpt-4-reinforces-sexist-stereotypes/.
- [38] Bryan Pietsch. “Disney Adds Warnings for Racist Stereotypes to Some Older Films”. In: *The New York Times* (Oct. 2020). URL: <https://www.nytimes.com/2020/10/18/business/media/disney-plus-disclaimers.htm>.

- [39] Jamie Rowlands. *The implications of biased AI models on the Financial Services Industry*. Nov. 2023. URL: <https://www.hlk-ip.com/the-implications-of-biased-ai-models-on-the-financial-services-industry/>.
- [40] Jeff Sebo and Robert Long. “Moral consideration for AI systems by 2030”. In: *AI and Ethics* (Dec. 2023). ISSN: 2730-5961. DOI: [10.1007/s43681-023-00379-1](https://doi.org/10.1007/s43681-023-00379-1). URL: <https://doi.org/10.1007/s43681-023-00379-1>.
- [41] Kriti Sharma. *Can We Keep Our Biases from Creeping into AI?* Feb. 2018. URL: <https://hbr.org/2018/02/can-we-keep-our-biases-from-creeping-into-ai>.
- [42] Patty Shillington. *Actively Addressing Inequalities Promotes Social Change*. June 2021. URL: <https://www.umass.edu/news/article/actively-addressing-inequalities-promotes-social-change>.
- [43] Fionna Smyth. *Why good data is key to unlocking gender equality*. Mar. 2023. URL: <https://devinit.org/blog/why-good-data-is-key-to-unlocking-gender-equality/>.
- [44] Solange Sobral. *Ai and the need for human oversight*. Nov. 2023. URL: <https://www.business-reporter.co.uk/ai--automation/ai-and-the-need-for-human-oversight>.
- [45] T.J. Thomson and Ryan J. Thomas. *Ageism, sexism, classism and more: 7 examples of bias in AI-generated images*. Mar. 2024. URL: <https://theconversation.com/ageism-sexism-classism-and-more-7-examples-of-bias-in-ai-generated-images-208748>.
- [46] Nitasha Tiku, Kevin Schaul, and Szu Yu Chen. *Ai generated images are biased, showing the world through stereotypes*. Nov. 2023. URL: <https://www.washingtonpost.com/technology/interactive/2023/ai-generated-images-bias-racism-sexism-stereotypes/>.
- [47] Ozlem Ulgen. “Kantian Ethics in the Age of Artificial Intelligence and Robotics (2017) 43 QIL, Zoom-in (Questions of International Law/Question de Droit International/Questioni di Diritto) 59-83”. In: 43 (Jan. 2017), pp. 59–83.
- [48] Darrell M. West et al. *The role of corporations in addressing AI’s ethical dilemmas*. Mar. 2022. URL: <https://www.brookings.edu/articles/how-to-address-ai-ethical-dilemmas/>.
- [49] Mike Zajko. “Artificial intelligence, algorithms, and social inequality: Sociological contributions to contemporary debates”. In: *Sociology Compass* 16.3 (2022), e12962. DOI: <https://doi.org/10.1111/soc4.12962>. eprint: <https://compass.onlinelibrary.wiley.com/doi/pdf/10.1111/soc4.12962>. URL: <https://compass.onlinelibrary.wiley.com/doi/abs/10.1111/soc4.12962>.