Interrogating Human-centered Data Science: Taking Stock of Opportunities and Limitations

Anissa Tanweer
University of Washington
Seattle, WA, USA
tanweer@uw.edu

Cecilia Aragon
University of Washington
Seattle, WA, USA
aragon@uw.edu

Michael Muller
IBM Research
Cambridge, MA, USA
michael_muller@us.ibm.com

Shion Guha
University of Toronto
Toronto, Canada
shion.guha@utoronto.ca

Samir Passi
Microsoft
Redmond, WA, USA
passi.samir@gmail.com

Gina Neff
University of Cambridge
and University of Oxford
Cambridge, UK
gina.neff@oii.ox.ac.uk

Marina Kogan
University of Utah
Salt Lake City, UT, USA
kogan@cs.utah.edu

ABSTRACT

Data science has become an important topic for the CHI conference and community, as shown by many papers and a series of workshops. Previous workshops have taken a critical view of data science from an HCI perspective, working toward a more human–centered treatment of the work of data science and the people who perform the many activities of data science. However, those approaches have not thoroughly examined their own grounds of criticism. In this workshop, we deepen that critical view by turning a reflective lens on the HCI work itself that addresses data science. We invite new perspectives from the diverse research and practice traditions in the broader CHI community, and we hope to co-create a new research agenda that addresses both data science and human-centered approaches to data science.

CCS CONCEPTS

• **Human-centered computing → HCI design and evaluation methods**; *Collaborative and social computing theory, concepts and paradigms.*

KEYWORDS

Human centered data science, data science, critical computing, context

ACM Reference Format:
Anissa Tanweer, Cecilia Aragon, Michael Muller, Shion Guha, Samir Passi, Gina Neff, and Marina Kogan. 2022. Interrogating Human-centered Data Science: Taking Stock of Opportunities and Limitations. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '22 Extended Abstracts), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3491101.3503740

CHI '22 Extended Abstracts, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9156-6/22/04.
https://doi.org/10.1145/3491101.3503740

1 BACKGROUND

Data-intensive methods and technologies hold great promise, but also have the potential to produce misleading results [25], violate rights and norms [35], perpetuate structural disparities [5], and inhibit accountability [22]. Accordingly, in recent years we have seen a burgeoning response in the form of methodology [40], methods critique [28, 29], pedagogy [43], principles [11], and projects [21] aimed at rehabilitating data-intensive technologies and methods to make them more ethical, responsible, equitable, and beneficial. Much of this discourse either explicitly or implicitly advocates for making data-intensive methods and related technologies more "human-centered" [36] by designing for "human values" [8], considering "human contexts" [42], and prioritizing "human flourishing" [39].

Human-centered perspectives offer powerful antidotes to beliefs and practices that naively assume the beneficence and superiority of technology or pursue technological advancement regardless of harms inflicted upon people in the process—a phenomenon that Broussard [9] has called "technochauvinism." However, the HCI analyses (e.g., [3, 20, 26, 27]) have not thoroughly examined their own grounds of criticism. In this workshop, we deepen that critical view by turning a reflective lens on the HCI work itself that addresses data science.

Scholars have questioned the ability of anthropocentrism to adequately address the moral, political, environmental, and social challenges posed by sociotechnical systems. This includes perspectives of feminist and Indigenous epistemologists who focus on the web of relationships between humans and nonhuman entities rather than on the human as the preeminent location of attention. Such works break down simplistic ontologies of difference and explore how humans are constituted in relation to nonhumans (Haraway, 2016/1985), reminding us that "human life is embedded in a material world of great complexity, one on which we depend for our continued survival" [18, p. 5].

These observations seem particularly salient given recent revelations about the ecological impact of data-intensive methods and technologies [4, 10], as well as moral questions raised by the possibility of sentient artificial intelligence [41]. Lewis et al. [23] argue for a position on AI drawn from Indigenous ontologies that "take the world as the interconnected whole that it is" in order to "illuminate the full scale of relationships that sustain us, provide guidance

on recognizing non-human beings and building relationships with them founded on respect and reciprocity." And HCI researchers have begun exploring ways that decentering the human may open new avenues for equitable and sustainable technology design [6, 15, 38].

Such ideas point to multiple ways in which relational thinking may inform our approaches to interrogating data science. Sabina Leonelli [22] urges us to recognize data as "relational" and "fungible" objects that are "first and foremost, material artifacts" whose "physical characteristics, including their format and the medium through which they are conveyed, are as relevant to understanding their epistemic role as their social and conceptual functions" (for treatments of data as a medium of design, see [12, 13, 32–34]). Birhane and Cummins [7] call attention to the ways in which it is problematic to assume that technologies may ever be rendered beneficial to an undifferentiated set of all humans or achieve a common good, advocating instead for a "relational ethics" that prioritizes an algorithmic technology's impact on the most vulnerable, marginalized, or disproportionately affected.

In this workshop, we invite explorations of these and other provocations salient to "human-centeredness" as it relates to data science. Previous events convened by the organizers of this workshop and other colleagues [3, 20, 26, 27] under the guise of human-centered data science have critically interrogated data science and generated practical approaches to mitigating the social and ethical challenges posed by the field. These discussions led to the writing and publication of the book, *Human-Centered Data Science: An Introduction* [2], forthcoming from MIT Press. We leverage the publication of that textbook as an occasion to probe and perturb conversations and practices around human-centered data science—including the conversations that led to the book. Whereas the aforementioned workshops were tasked with "interrogating data science," here we turn our lens back on that interrogation itself, creating an opportunity for reflection that advances the field of human-centered data science in new directions.

We seek to take stock of the goal, broadly conceived, to humanize data science. We invite participation from individuals who have been involved in reading, writing, teaching, designing, and organizing with an aim to problematize data-intensive technologies and methods, and propose solutions to those problems. Workshop participants will submit short position papers along with their applications; we encourage authors to use these position papers as an opportunity to articulate their own experience and expertise as it relates to human-centered data science, and to consider the opportunities and limitations of efforts to intervene in data science. We hope to solicit a set of position papers and workshop discussions that collectively span theory and practice, provocation and pragmatism. As such, we invite participants to reflect on a broad set of questions ranging from general to specific:

- What are the strengths of a human-centered lens for interrogating data science, and what possibilities does it obscure or preclude?
- What theories, ideas, and methods can complement human-centered approaches or otherwise further the cause of making data science more ethical, responsible, equitable, and beneficial?
- What successes have been won in humanizing data science, and what challenges thus far have proved intractable?
- Can we map successes and obstacles in using data science in practice?
- How might a human-centered approach to data science improve the outputs of data science systems?
- What changes in data-science education are needed to expand this approach?
- The Zapatistas have stated, "In the world we want many worlds to fit" [24]. Data science prefers unitary and undisputed "facts"—preferably arranged in binary contrasts. How do we need to change our socio-technical practices so as to represent multiple, co-existent, simultaneous knowledges about the data of data science, and the standpoints that give contexts and meanings to those knowledges [17]?
- In *Ghostwork*, Gray and Suri [16] showed that there are many hidden humans "behind" the systems that end-users experience as "AI." Alkhatib and colleagues have shown in a series of papers that complex aggregate actions can be decomposed into micro-tasks that humans can perform without awareness of the other humans who perform other micro-tasks in the same aggregate action [1]. What is lost and what is gained in such configurations?
- Scholars from Fitts [14] to Shneiderman [37] have debated representations and dynamics of human-initiative and/vs. AI-initiative in systems. What is lost and what is gained when the human has the initiative versus when the data science system has the initiative?

We will launch our gathering with a panel discussion among authors and contributors of the forthcoming book, *Human-Centered Data Science: An Introduction* [2]. During that panel, we will open a space for unanswered questions and new challenges—as it were, a retrospective of the early-2020 thinking that is in the book. We will then continue to explore the opportunities and limitations of human-centered data science through small group discussions based on short position papers submitted by each participant. Finally, we will facilitate a discussion about future directions for research related to human-centered data science.

The goal of the workshop is to push against the current boundaries of human-centered data science and establish an agenda for further advancing this area of work. We intend to use the papers and discussions contributed by workshop participants to either develop a slate of journal articles for a special issue or to seed one or more funding proposals for innovative research.

2 ORGANIZERS

Anissa Tanweer (primary contact), *University of Washington, eScience Institute, US.* Anissa Tanweer is a research scientist at the University of Washington's eScience Institute focused on human-centered data science. She conducts ethnographic research on the practice and culture of data science, and brings a sociotechnical lens to bear on the design and implementation of data science training programs. She serves as Program Chair for the UW Data Science for Social Good internship and Associate Editor for the newly launched Data Analytics for Social Impact section of Frontiers in Big Data.

Cecilia Aragon, *University of Washington, Department of Human Centered Design and Engineering, US.* Cecilia Aragon is a Professor in the Department of Human Centered Design Engineering and Director of the Human-Centered Data Science Lab at the University of Washington. Her research focuses on enabling humans to explore and gain insight from vast data sets. In 2008, she received the Presidential Early Career Award for Scientists and Engineers (PECASE). She is co-author of the book, *Human-Centered Data Science: An Introduction*, forthcoming in 2022 from MIT Press [2].

Michael Muller, *IBM Research, Human Centered AI, US.* Michael Muller studies work-practices of data science workers at IBM Research (Cambridge MA USA). With colleagues, he has analyzed how humans intervene (individually and collaboratively) between "the data" and "the model" as aspects of responsible and accountable data science work. He is co-author of the book, *Human-Centered Data Science: An Introduction*, forthcoming in 2022 from MIT Press [2].

Shion Guha, *University of Toronto, Faculty of Information, Canada.* Shion Guha studies algorithms in the public sector with particular focus on child welfare and criminal justice systems. He is interested in the intersections of computing and critical methodologies. He is co-author of the book, *Human-Centered Data Science: An Introduction*, forthcoming in 2022 from MIT Press [2].

Samir Passi, *Microsoft, US.* Samir Passi studies the sociotechnical challenges with the development and use of Responsible AI systems at Microsoft, where he works as a Researcher with the AI, Ethics and Effects in Engineering and Research (Aether) initiative. He is particularly interested in mapping the relation between the human and organizational work involved in building AI systems and the social and normative implications emanating from the use of those systems.

Gina Neff, *University of Cambridge, Minderoo Centre for Technology & Democracy; and University of Oxford, Oxford Internet Institute & Department of Sociology, UK.* Gina Neff studies the effects of the rapid expansion of our digital information environment on workers, workplaces and in our everyday lives. Her books include *Venture Labor* [30], *Self-Tracking* [31] and *Human-Centered Data Science: An Introduction* [2]. She also led the team that won the 2021 Webby Award for the best educational website on the Internet, for the *A to Z of AI* [19], which has reached over 1 million people in 17 different languages.

Marina Kogan, *University of Utah, US.* Marina Kogan is an Assistant Professor in the School of Computing at the University of Utah. Her research interests are in Crisis Informatics, Social Computing, and Network Science. Her research centers around social group formation, dynamics, and cooperative work. More specifically, she focuses on the cooperative activities and the resulting large-scale social dynamics that emerge on social media in disruption events, such as disasters arising from natural hazards, collective action around political crises, and community disruptions. She is co-author of the book, *Human-Centered Data Science: An Introduction*, forthcoming in 2022 from MIT Press [2].

3 WEBSITE

https://sites.google.com/view/hcds-workshop-chi2022/home

The following content will be added to the website:

3.1 Prior to the Workshop

- Background and references
- Call for submissions
- Instructions for submissions
- Schedule
- Links to accepted position papers

3.2 Following the Workshop

- Links to recording of panel discussion
- Links to a transcript of panel discussion
- Links to a clean, public version of notes from the workshop

4 PRE-WORKSHOP PLANS

4.1 Recruitment

We hope to draw participation from an eclectic group of scholars and practitioners who approach human-centered data science from different disciplinary backgrounds and positionalities. Table 1 shows various academic communities we hope to participation from, and the recruitment venues we plan to use to reach them.

4.2 Reviewing position papers

We will form a committee to review submissions of position papers, drawing from a pool of individuals who participated in previous human-centered data science workshops held at the Computer Supported Cooperative Work conference and the GROUP conference (Aragon et al, 2016; Kogan et al., 2020; Muller et al, 2019, 2020). Reviewers will be asked to evaluate each paper using a very simple rubric assessing 1) Clarity 2) Relevance to the call for submissions 3) Uniqueness or originality of the articulated position. We intend to accept up to 25 submissions and run the workshop with 15-30 participants (including organizers). Depending on the number and nature of submissions, the review committee may also group papers according to similar or complementary themes and pre-assign breakout reading and discussion groups based on these affinities.

4.3 Circulating papers

We will circulate the papers well in advance of the workshop through multiple channels. If we end up grouping papers according to affinity as described above, then authors will be asked in particular to read a small number of papers by authors with whom they have been grouped. Papers will be posted on the workshop website, emailed to participants, and posted in a workshop Slack channel so as to provide an opportunity for asynchronous discussion.

4.4 Ensuring accessibility

In advance of the workshop we will survey participants to understand accessibility needs and will make any accommodations within our ability, including the possibility of providing closed captioning services for Zoom discussions and/or an ASL translator.

5 IN-PERSON, HYBRID, OR VIRTUAL-ONLY

This workshop will be a virtual-only event using Zoom and an optional asynchronous component on Slack. We will adjust our platforms depending on participants' accessibility needs based on a survey conducted in advance of the event.

Table 1: Recruitment

Academic community	Sample recruitment vehicles
HCI researchers and designers	HCI researchers and designers CHI-Announcements CSCW-SIG@jiscmail.ac.uk Eusset@listserv.uni.siegen.de bcs-hci@jiscmail.ac.uk chi-il@acm.org Facebook: CHI meta; CSCW meta Twitter lists: Algorithmic transparency; Design activists; Digital cultures; Dubbers
Science & Technology Studies (STS) and Critical Data Studies scholars	STS global mailing list Association of Internet Researchers mailing list
Reflexive data science practitioners	Academic Data Science Alliance newsletter, eScience Institute mailing list
Underrepresented scholars	Hispanics in Computing Women in Data Science AccessAdvance Fempower.tech AccessSIGCHI R.A.C.E. Twitter lists: Feminist design scholars; Women and technologies; Fewpower.tech; others TBD

6 ASYNCHRONOUS ENGAGEMENT

6.1 Asynchronous Plans for Part 1: Panel Discussion

The panel discussion that opens the workshop will be live-streamed on YouTube but also recorded and posted to the website following the workshop, along with a transcript.

6.2 Asynchronous Plans for Part 2: Position Paper Discussion

Position papers will be made available in advance of the workshop on the website, by email, and in a Slack workspace created for workshop participants. In the two weeks leading up to the event, organizers will post position papers and invite comments, questions, and reflections from other participants about each of the papers. Organizers will initiate a thread about each paper and seed it with their own comments or questions in order to encourage asynchronous interaction. This will give people the opportunity to asynchronously discuss the papers both before and after the workshop if they would like.

Participants who know in advance that they will not be able to join the discussion of position papers via video conference due to connectivity limitations or time zone incompatibility will also have the following options:

- Join the meeting by voice only.
- Submit a pre-recorded, 5-minute synopsis of their paper.
- Submit written reflections on one or more of the other position papers in advance. Written reflections will be read by moderators and recorded in notes on the session.
- Crowdsourced notes from the position paper discussion breakout rooms will be shared with participants following the workshop.

6.3 Asynchronous Plans for Part 3: Future Directions Discussion

Participants who know in advance that they will not be able to join the discussion of future directions via video conference due to connectivity limitations or time zone incompatibility will have the following options:

- Join the meeting by voice only.
- Submit written reflections on their ideas for future directions of research. Written reflections will be read by moderators and recorded in notes on the session.
- A transcript from this whole-group discussion will be shared with participants following the workshop. We will also post a streamlined version of notes from this discussion to the workshop Slack channel soon after the conclusion of the workshop, and will invite further discussion there.

7 WORKSHOP STRUCTURE

Depending on the geographic distribution of participants, we may decide to run two consecutive sessions of identical structure in order to accommodate widely distributed time zones. In this case, organizers will compile materials from both sessions following the

workshop and make them available to all participants so they can see what they missed in the other session.

We have five hours of workshop content planned, but intend to split this up with substantive breaks between segments to combat Zoom fatigue.

7.1 Introduction (1 hour)

We will use the first hour of the workshop to introduce our goals, review logistics, and get to know the workshop participants. During this time, we plan to conduct a fun and interactive icebreaker. For example, we may collect information about each participant's academic or professional biography in advance and create virtual Bingo cards with these details that individuals fill out as they interact with other participants. Or we may ask participants to create a research "trading card" in which they characterize their work according to several creative prompts such as: If academic researchers and scholars were a band of superheroes, which superhero would you be? If you could invite any academic scholar (living or deceased) to dinner, who would it be? If you could make up a new prize to nominate your own research for, what it be called?

7.2 Part I: Panel Discussion (30 min)

We will open the workshop with a discussion among panelists, composed of the authors and contributors who produced the new book, Human-Centered Data Science: An Introduction, which will be released by MIT Press in February 2022 (authors include the organizers of this workshop).

Break (30-60 min)

7.3 Part II: Position Paper Discussion (1.5 hours)

We will form small group virtual breakout rooms of 3-4 people in which each person provides a synopsis of their position paper that is no longer than 5 minutes, and the group discusses common themes and contrasting positions, resonance and dissonances between the papers in their group. After one hour, we will switch into new breakout groups to discuss a different set of papers. Some of these synopses may be pre-recorded to accommodate participants who require asynchronous participation (see section on "Asynchronous Engagement").

Break (30-60 min)

7.4 Part III: Future Directions (2 hours)

Will come back together for a large-group discussion in which we distill highlights from the breakout room conversations and chart areas for future work and research. We intend that this discussion will lead to either a cluster of papers for a special journal issue, or one or more collaborative research grant proposals.

8 POST-WORKSHOP PLANS

We anticipate writing a report of the workshop as an article or post for *Interactions*. If the workshop surfaces ideas that are close to being publication-ready, then we hope to propose a special issue and will use the discussions from the workshop to crystallize paper topics for that issue and form writing teams. The target journal for this special issue remains to be determined, depending on what themes and interests emerge among workshop participants. For example, if participants rally around pedagogical issues, we might submit our idea to the *Journal of Statistics and Data Science Education*; if participants express interest in critiques or provocations, then we might propose a special issue of *Big Data & Society*; if participants generate ideas for interventions to design or practice, we might propose a special issue of the *Journal of Collaborative Computing and Work Practices* (*JCSCW*).

If the workshop instead surfaces ideas for empirical research that has yet to be conducted, then we may use follow-up discussions to develop ideas for future funding opportunities and research projects, ultimately leading to one or more collaborative grant proposals.

We will facilitate conversations about these possible follow-on activities through a combination of email, Google docs, and Slack.

9 CALL FOR PARTICIPATION

In the workshop, *"Interrogating Human-Centered Data Science: Taking Stock of Opportunities and Limitations"*, we will consider efforts, broadly conceived, to humanize data science. We seek participation from individuals who have been involved in reading, writing, teaching, designing, and organizing with an aim to problematize data-intensive technologies and methods, and propose solutions to those problems. For background, details, and a link to the application, please see the workshop website, https://sites.google.com/view/hcds-workshop-chi2022/home.

Our intent is to push against the current boundaries of human-centered data science and establish an agenda for furthering this area of work. We plan to use this workshop as a launchpad to develop a slate of journal articles for a special issue or to seed one or more funding proposals for innovative research.

The workshop will be held remotely on [Date TBD]. To be considered for participation, submit a position paper of 2-3 pages in the ACM primary article template (https://www.acm.org/publications/taps/word-template-workflow) that articulates your own experience and expertise as it relates to human-centered data science, and examines the opportunities and limitations of efforts to intervene in data science. Applicants will be accepted based on a review of their position papers for clarity, relevance to the call for submissions, and uniqueness or originality. Position papers will be circulated and discussed among workshop participants and may be posted to the workshop website with author consent. At least one author of each accepted submission must attend the workshop. All workshop participants must register for both the workshop and at least one day of the conference.

REFERENCES

[1] Ali Alkhatib, Michael S. Bernstein, and Margaret Levi. 2017. Examining crowd work and gig work through the historical lens of piecework. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, Denver Colorado USA, 4599–4616. https://doi.org/10.1145/3025453.3025974

[2] Cecilia Aragon, Shion Guha, Marina Kogan, Michael Muller, and Gina Neff. 2022. *Human-Centered Data Science: An Introduction.* MIT Press, Cambridge, MA.

[3] Cecilia Aragon, Clayton Hutto, Andy Echenique, Brittany Fiore-Gartland, Yun Huang, Jinyoung Kim, Gina Neff, Wanli Xing, and Joseph Bayer. 2016. Developing a Research Agenda for Human-Centered Data Science. In *Conference Companion Publication of the 2016 Conference on Computer Supported Cooperative Work and*

Social Computing. ACM Press, San Francisco, California, USA, 529–535. https://doi.org/10.1145/2818052.2855518

[4] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Virtual Event Canada, 610–623. https://doi.org/10.1145/3442188.3445922

[5] Ruha Benjamin. 2019. *Race after Technology: Abolitionist Tools for the New Jim Code*. Polity, Medford, MA.

[6] Heidi R. Biggs, Jeffrey Bardzell, and Shaowen Bardzell. 2021. Watching myself watching birds: Abjection, ecological thinking, and posthuman design. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3411764.3445329

[7] Abeba Birhane and Fred Cummins. 2019. Algorithmic injustices: Towards a relational ethics. *arXiv:1912.07376* (Dec. 2019). http://arxiv.org/abs/1912.07376 arXiv: 1912.07376.

[8] Alan Borning and Michael Muller. 2012. Next steps for value sensitive design. In *Proceedings of the 2012 SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. Association for Computing Machinery, New York, NY, USA, 1125–1134. https://doi.org/10.1145/2207676.2208560

[9] Meredith Broussard. [n. d.]. *Artificial Unintelligence: How computers misunderstand the world*. MIT Press, Cambridge, MA.

[10] Kate Crawford. 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.

[11] Catherine D'Ignazio and Lauren F Klein. 2020. *Data Feminism*. MIT Press, Cambridge, MA.

[12] Melanie Feinberg. 2017. A design perspective on data. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 2952–2963. https://doi.org/10.1145/3025453.3025837

[13] Melanie Feinberg. 2017. Material vision. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 604–617. https://doi.org/10.1145/2998181.2998204

[14] Paul M. Fitts. 1951. *Human Engineering for an Effective Air-Navigation and Traffic-control System*. Technical Report. National Research Council. Pages: xxii, 84.

[15] Laura Forlano. 2016. Decentering the human in the design of collaborative cities. *Design Issues* 32, 3 (July 2016), 42–54. https://doi.org/10.1162/DESI_a_00398

[16] Mary L. Gray and Siddharth Suri. 2019. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Houghton Mifflin Harcourt, Boston, MA.

[17] Sandra G. Harding (Ed.). 2004. *The feminist standpoint theory reader: Intellectual and political controversies*. Routledge, New York, NY.

[18] N. Katherine Hayles. 1999. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. University of Chicago Press.

[19] Oxford Internet Institute. n.d.. A to Z of AI. Retrieved Oct 14, 2021 from https://atozofai.withgoogle.com/intl/en-US/

[20] Marina Kogan, Aaron Halfaker, Shion Guha, Cecilia Aragon, Michael Muller, and Stuart Geiger. 2020. Mapping out human-centered data science: Methods, approaches, and best practices. In *Companion of the 2020 ACM International Conference on Supporting Group Work*. ACM, Sanibel Island Florida USA, 151–156. https://doi.org/10.1145/3323994.3369898

[21] Ida B. Wells Just Data Lab. [n. d.]. About. https://www.thejustdatalab.com/about

[22] Sabina Leonelli. 2015. What counts as scientific data? A relational framework. *Philosophy of Science* 82, 5 (2015), 810–821. https://doi.org/10.1086/684083 Publisher: [The University of Chicago Press, Philosophy of Science Association].

[23] Jason Edward Lewis, Noelani Arista, Archer Pechawis, and Suzanne Kite. 2018. Making kin with the machines. *Journal of Design and Science* 3.5 (July 2018). https://doi.org/10.21428/bfafd97b

[24] Subcommandante Insurgente Marcos. 1996. *Fourth Declaration of the Lacandon Jungle*. Technical Report. https://schoolsforchiapas.org/wp-content/uploads/2014/03/Fourth-Declaration-of-the-Lacandona-Jungle-.pdf

[25] Deborah Mayo. 2020. P-Values on trial: Selective reporting of (best practice guides against) selective reporting. *Harvard Data Science Review* 2, 1 (2020). https://doi.org/10.1162/99608f92.e2473f6a

[26] Michael Muller, Cecilia Aragon, Shion Guha, Marina Kogan, Gina Neff, Cathrine Seidelin, Katie Shilton, and Anissa Tanweer. 2020. Interrogating data science. In *Conference Companion Publication of the 2020 Conference on Computer Supported Cooperative Work and Social Computing*. ACM, Virtual Event USA, 467–473. https://doi.org/10.1145/3406865.3418584

[27] Michael Muller, Melanie Feinberg, Timothy George, Steven J. Jackson, Bonnie E. John, Mary Beth Kery, and Samir Passi. 2019. Human-centered study of data science work practices. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–8. https://doi.org/10.1145/3290607.3299018

[28] Michael Muller, Ingrid Lange, Dakuo Wang, David Piorkowski, Jason Tsay, Q. Vera Liao, Casey Dugan, and Thomas Erickson. 2019. How data science workers work with data: Discovery, capture, curation, design, creation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3290605.3300356

[29] Michael Muller, Christine T. Wolf, Josh Andres, Michael Desmond, Narendra Nath Joshi, Zahra Ashktorab, Aabhas Sharma, Kristina Brimijoin, Qian Pan, Evelyn Duesterwald, and Casey Dugan. 2021. Designing ground truth and the social life of labels. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3411764.3445402

[30] Gina Neff. 2012. *Venture labor: Work and the burden of risk in innovative industries*. MIT press.

[31] Gina Neff and Dawn Nafus. 2016. *Self-tracking*. MIT Press.

[32] Samir Passi. 2021. *Making Data Work: The Human and Organizational Lifeworlds of Data Science*. Cornell University.

[33] Cathrine Seidelin. 2020. *Towards a co-design perspective on data: Foregrounding data in the design and innovation of data-based services*. Ph. D. Dissertation. IT University of Copenhagen. https://scienceopen.com/document?vid=50421150-5ba6-4ae7-ac17-ab0a5b3b78cc

[34] Cathrine Seidelin, Yvonne Dittrich, and Erik Grönvall. 2020. Co-designing data experiments: Domain experts' exploration and experimentation with self-selected data sources. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society (NordiCHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3419249.3420152

[35] Andrew D. Selbst, danah boyd, Sorelle Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2018. Fairness and abstraction in sociotechnical systems. In *Proceedings of the 2019 ACM Conference on Fairness, Accountability, and Transparency*. https://papers.ssrn.com/abstract=3265913

[36] Chirag Shah, Theresa Anderson, Loni Hagen, and Yin Zhang. 2021. An iSchool approach to data science: Human-centered, socially responsible, and context-driven. *Journal of the Association for Information Science and Technology* 72, 6 (2021), 793–796. https://doi.org/10.1002/asi.24444 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.24444.

[37] Ben Shneiderman. 2020. Human-centered artificial intelligence: Three fresh ideas. *AIS Transactions on Human-Computer Interaction* 12, 3 (2020), 109–124. https://doi.org/10.17705/1thci.00131

[38] Nancy Smith, Shaowen Bardzell, and Jeffrey Bardzell. 2017. Designing for cohabitation: Naturecultures, hybrids, and decentering the human in design. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 1714–1725. https://doi.org/10.1145/3025453.3025948

[39] Royal Society. n.d.. *Data Management and Use: Governance in the 21st Century*. Technical Report. British Academy and The Royal Society. https://royalsociety.org/topics-policy/projects/data-governance/

[40] Anissa Tanweer, Emily Kalah Gade, PM Krafft, and Sarah K Dreier. 2021. Why the Data Revolution Needs Qualitative Thinking. *Harvard Data Science Review* 3 (2021).

[41] Steve Torrance. 2011. Machine ethics and the idea of a more-than-human moral world. In *Machine ethics*, Michael Anderson and Susan Leigh Anderson (Eds.). Cambridge University Press, 115–137. https://doi.org/10.1017/CBO9780511978036

[42] UC Berkeley Division of Computing Data Science & Society. [n. d.]. Human Contexts and Ethics. https://data.berkeley.edu/hce

[43] Michelle Hoda Wilkerson and Joseph L. Polman. 2020. Situating data science: Exploring how relationships to data shape learning. *Journal of the Learning Sciences* 29, 1 (2020), 1–10. https://doi.org/10.1080/10508406.2019.1705664