Teaching Responsible Data Science

Julia Stoyanovich
New York University
New York, NY, USA
stoyanovich@nyu.edu

ABSTRACT

Responsible Data Science (RDS) and Responsible AI (RAI) have emerged as prominent areas of research and practice. Yet, educational materials and methodologies on this important subject still lack. In this paper, I will recount my experience in developing, teaching, and refining a technical course called "Responsible Data Science", which tackles the issues of ethics in AI, legal compliance, data quality, algorithmic fairness and diversity, transparency of data and algorithms, privacy, and data protection. I will also describe a public education course called "We are AI: Taking Control of Technology" that brings these topics of AI ethics to the general audience in a peer-learning setting. I made all course materials are publicly available online, hoping to inspire others in the community to come together to form a deeper understanding of the pedagogical needs of RDS and RAI, and to develop and share the much-needed concrete educational materials and methodologies.

CCS CONCEPTS

• **Social and professional topics** → **Socio-technical systems**; **Computing education**; *Adult education.*

KEYWORDS

Responsible Data Science, Responsible AI

ACM Reference Format:
Julia Stoyanovich. 2022. Teaching Responsible Data Science. In *1st International Workshop on Data Systems Education (DataEd'22), June 17, 2022, Philadelphia, PA, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3531072.3535318

1 INTRODUCTION

Automated Decision Systems (ADS) process data about people, some of which may be sensitive or proprietary, and help make decisions that are consequential to people's lives and livelihoods. These systems are used every more broadly, and with their wide-spread use comes the impetus to ensure that they are designed, developed and deployed *responsibly* — in accordance with ethical norms, and with legal and regulatory requirements. Current computer science and data science students will soon become practising data scientists, influencing how ADS are built, tested and deployed, and more generally, how technology impacts society. And while these students are increasingly aware of the potential societal risks of technology, few of them are equipped with responsible data science skills and competencies.

In this paper, I will recount my experience in developing, teaching, and refining a technical course on *Responsible Data Science (RDS)*, which tackles the issues of ethics in AI, legal compliance, data quality, algorithmic fairness and diversity, transparency of data and algorithms, privacy, and data protection. Although numerous ethics courses are available, with many focusing specifically on technology and computer ethics, pedagogical approaches used in these courses rely exclusively on texts rather than on software development or data analysis. For this reason, technical students often consider these courses unimportant and a distraction from the "real" material. How can we develop materials and instructional methodologies that are thoughtful and engaging, and that help students gain knowledge and skills useful in their future careers? I believe that, to do this, we must *strive for balance*: between texts and coding, between critique and solution, and between cutting-edge research and practical applicability. Finding such balance is both necessary and difficult in the nascent field of RDS, where we are only just starting to understand how to interface between the intrinsically different methodologies of engineering and social sciences. In Section 2, "Teaching Future Technologists," I will speak about my quest for such a balance.

An immediate realization is that RDS is not a purely technical discipline, rather, it is socio-legal-technical. When thinking about the responsible design of the central RDS artifact, an Automated Decision System, we must consider not only its technical components —the data and the model— but, first and foremost, its context of use: What goals does the ADS aim to achieve? Who are the stakeholders: individuals, groups or organizations whom the ADS impacts, directly or indirectly? What are the benefits when the ADS works well and who benefits? What are the risks, and the actual or potential harms, and who is harmed? Who decides on the appropriate *balance between the benefits and the risks*? And whose responsibility is it to mitigate the harms?

By pondering these questions, we come to another immediate realization, namely, that facilitating the responsible design, development and use of ADS is everyone's job. No single stakeholder group, no matter how well-resourced or well-intentioned, can do this alone. This realization has motivated me to developing educational materials and methodologies on Responsible Data Science (RDS) and Responsible AI (RAI) for a variety of audiences. In Section 3, "Teaching Members of the Public," I will describe a public education course *We are AI: Taking Control of Technology* that brings these topics to the general audience in a peer-learning setting.

While I write this article in the first person, the work discussed here would not have been possible without collaboration and input from many of my colleagues. Their names and specific contributions are discussed in the Acknowledgments section.

DataEd'22, June 17, 2022, Philadelphia, PA, USA

ACM ISBN 978-1-4503-9350-8/22/06.
https://doi.org/10.1145/3531072.3535318

2 TEACHING FUTURE TECHNOLOGISTS

I have been developing and teaching Responsible Data Science (RDS) courses at the Center for Data Science at New York University since Spring 2019. The graduate course is offered annually, with a substantial increase in enrollment and an addition of an undergraduate course in 2021. During the last 2 years, I have been co-teaching RDS with George Wood. We have been collaborating on making the course modular and scaling it up in response to increasing demand.

Slightly different versions of the course are offered to undergraduate and graduate students. Course structure, and most of the materials and methodologies, are in common between the two audiences, although the undergraduate course proceeds at a somewhat slower pace and its assignments are somewhat less demanding. Because the commonalities are more substantial than the differences, I will refer to different offerings collectively as "the course" in the remainder of this section.

2.1 Course Overview

The RDS course is structured as a sequence of lectures, with supplementary readings, labs, accompanying assignments, and a course project. The course relies on classroom-based instruction. It ran in hybrid mode during the COVID-19 pandemic but still followed the same general methodology, with lectures and labs offered synchronously to students over Zoom.

All course materials, including the syllabus, complete lecture slides, lab assignments, and reading materials, are publicly available on the course website.[1] Homework assignments, with solutions and grading rubrics, and a detailed description of the course project, will be made available to instructors upon request. Instructional materials and methodologies are discussed in Section 2.3.

Enrollment statistics. RDS is offered annually in the Spring semester. In Spring 2019 and 2020, the course enrolled 18 and 46 graduate students, respectively. In Spring 2021, an undergraduate course was added, and we enrolled around 60 students in each undergraduate and graduate cohort (120 students in total). In Spring 2022, there was, once again, substantial increase in demand, with around 90 students in the graduate course and 125 in the undergraduate. Enrollment was capped at these class sizes, and every available seat in both courses was filled. Notably, the undergraduate course is among the degree requirements of the new Bachelors in data science at NYU, while the graduate course serves as an elective for several Masters and PhD programs.

Prerequisites. The course has Introduction to Data Science or Introduction to Computer Science as its only prerequisite. A machine learning course is not a prerequisite for RDS. This is a deliberate choice that reflects our goals to (1) educate data science students on ethics and responsibility early in their program of study, and (2) to enroll a diverse group of students, including those who may not go on to take machine learning. Students are expected to have basic familiarity with the python programming language, which is used in labs and assignments.

[1] *Responsible Data Science* course website, https://dataresponsibly.github.io/courses/.

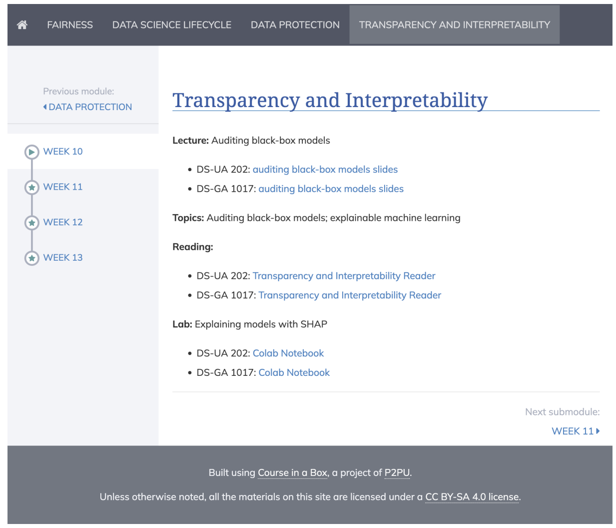

Figure 1: The transparency and interpretability module of the *Responsible Data Science* course.

2.2 Course Organization

During this semester-long course, students complete several thematic modules, in which content is delivered through a combination of case studies (often from the recent press), fundamental algorithmic techniques, and hands-on exercises with open-source datasets and software libraries. The order of topics, the extent to which they are covered, and their break-down into modules, has evolved over the years, and is likely to continue to evolve as the field of RDS and Responsible AI (RAI) develops and matures. Currently, the course consists of 4 modules:

- Module 1: Algorithmic fairness (4 weeks)
- Module 2: The data science lifecycle (2 weeks)
- Module 3: Data protection (3 weeks)
- Module 4: Transparency and interpretability (4 weeks)

Figure 1 shows the structure of the Transparency and Interpretability module on the course website, presenting the content of a typical week. This GitHub page contains links to lecture slides, assigned reading, and Google Colaboratory notebooks for that week's hands-on lab, for each undergraduate and graduate course.

In selecting the topics to cover and in structuring them as modules, I started with the technical topics that have been the focus of the Fairness, Accountability, Transparency, and Ethics (FATE) community, as represented by the ACM FAccT conference and by papers on relevant topics at major AI conferences like AAAI, IJCAI and NeuRIPS. The initial set of topics included algorithimic fairness, transparency, and interpretability. I then expanded this set to weave in material on ethics, law and regulation, and data engineering, as discussed below.

Module 1. This module begins with an introduction and overview of the course, centering the conversation on automated decision systems (ADS, see Introduction), stakeholders (those who are impacted by the operation of an ADS, directly or indirectly), benefits, and

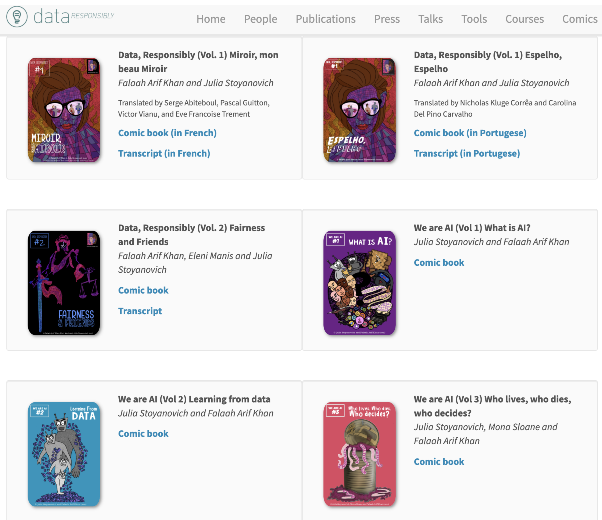

Figure 2: A sampling of educational comic books on responsible data science and AI, available in multiple languages.

harms. We then dive into algorithmic fairness, going beyond classification and risk assessment, as is currently typical in this line of research [4], to also include fairness in set selection and ranking [41], to consider intersectional discrimination [39], and to make connections between algorithmic fairness and diversity[8, 17, 38]. An important component of this module is the introduction of equality of opportunity doctrines from political philosophy [11, 12, 25, 27], and novel content on the mapping between algorithmic fairness and equality of opportunity [13–15, 41].

Module 2. This module takes a lifecycle view of responsible data science, as a step towards a more holistic (rather than reductionist) treatment of technology ethics [35]. Here, we make connections between RDS and responsible data management and data engineering, emphasizing the importance of both (1) the data lifecycle, and (2) the lifecycle of the design, development, deployment, and use of ADS [29, 35]. Data engineering topics are often overlooked in data science education in general, and have also received limited attention so far in RDS research. Yet, responsibility concerns, and important decision points, arise in data sharing, annotation, acquisition, curation, cleaning, and integration. Consequently, opportunities for improving data quality and representativeness, controlling for bias, and allowing humans to oversee the process, are missed if we do not consider these earlier lifecycle stages [16, 18, 34].

Module 3. This module tackles data protection and privacy, and makes a connection between these topics and applied ethics. We discuss reconstruction attacks (e.g., [23]), and then dive more deeply into the fundamental law of information recovery [7] and differential privacy [9, 10], discussing both the guarantees provided by the framework and the challenges that prevent its wide-spread adoption [22]. Students work with privacy-preserving synthetic data generators [21, 24, 42] to appreciate the trade-offs between privacy and utility. We then discuss data protection concerns that go beyond privacy, and move into a conversation about a principles-based approach to ethics in data-intensive research and practice [28].

Module 4. In the final course module, we take a broad view of transparency and interpretability. We return to the conversation about ADS stakeholders, and discuss methods for bring them "into the loop" of automated decision-making based on their unique needs, concerns, and responsibilities [32]. Technical topics include feature-based explanations of black-box models [1, 6, 20, 26], discrimination in online ad delivery [1, 5, 37], and nutritional labels for public disclosure [30, 32, 33, 40]. Legal and regulatory frameworks are also an important component of this module. Current international and local regulatory efforts are used to ground the discussion throughout the course, starting from the first lecture, and are discussed in detail at the end of the course. To help ground the course in current events in New York City, students are encouraged to attend public hearings of the New York City Committee on Technology, particularly those pertaining to regulation of ADS, and to reflect on these hearings during the discussion [36].

2.3 Instructional Materials and Methods

One of the challenges I faced when designing this course was the lack of a text book that offers comprehensive coverage of Responsible Data Science, balancing case studies, fundamental concepts and methodologies from the social sciences, and statistical and algorithmic techniques. The RDS course does not have a required textbook, but each module is accompanies by a "reader" that consists primarily of research papers but also sometimes includes articles from the popular press. Important concepts from the assigned papers are covered in class, and students are instructed on where to focus their attention while reading the papers and which parts to skim or skip. I have also been co-developing less traditional materials and methods for RDS instructions, discussed next.

Scientific comics. Falaah Arif Khan and I have been developing comic books on RDS and RAI.[2] We have produced a 5-volume general audience series called "We are AI" [31], and are working on a scientific comic series called "Data, Responsibly", of which 2 volumes have been published to date [2, 3]. We have been using the RDS course to guide the content of the scientific comic series, and use the published volumes as supplementary reading for the course. Figure 2 shows a sampling for the comics, some of which are available in French, Spanish, and Brazilian Portuguese in addition to English.

External resources. I have been participating in the development of a free online course called 'AI Ethics: Global Perspectives", where the goal is to create a living repository of instructional materials on RDS and RAI. This project is spearheaded by Stefaan Verhulst at the NYU Governance Lab (The GovLab). The course repository is currently comprised of 35 online modules — recorded lectures with supplementary readings — by instructors from 20 countries around the world, along with 4 webinars and 8 panels. Course materials are available online. [3]

I have been using materials from this course as part of RDS assignments, selecting a handful of lectures for each homework, and

[2] *Data, Responsibly* and *We are AI* comics: https://dataresponsibly.github.io/comics.
[3] *AI Ethics: Global Perspectives* course website, https://aiethicscourse.org.

instructing students to write a brief memo reflecting on issues raised in the lecture in response to specific prompts. In this way, students improved their written communication skills, and were exposed to a variety of important RDS and RAI themes that were not covered in my course, including the impact of ADS on individuals with disabilities, content moderation in social media, indigenous data sovereignty, the ethics of autonomous driving, and data activism.

Project-based learning. The course project pursues the broad learning goal of making ADS interpretable using the novel paradigm of an *object-to-interpret-with* [19]. Adhering to constructivist principles, students work in teams of 2 to audit an automated decision system (ADS) of their choice. Students are instructed to select a system developed by others in response to a Kaggle competition, but can also use other systems that are of interest to them.

As part of the project, students interrogate the assumptions behind the ADS of their choice, to understand the purpose of the ADS, its stated goals and any trade-offs that multiple goals may introduce. They describe the data collection process and the statistical properties of the data on which the ADS is trained and operates. They go on to interrogate the implementation of the ADS, auditing it for accuracy, fairness and robustness, and investigating the features that are most important to the predictions. Finally, the students come up with general recommendations regarding the ADS, expressing their position on its performance and fitness for use, and suggesting opportunities for improvement.

3 TEACHING MEMBERS OF THE PUBLIC

In collaboration with Peer-to-Peer University (P2PU), an open-education non-profit, and with the Queens Public Library (QPL) New York City, I have been developing and teaching a public education course called *We are AI*. The course was offered for the first time in Spring 2022, in person. A precursor to this offering was a series of online workshops with QPL's Jobs and Business Academy in the Queensbridge housing project in Fall 2020 [36], and an online pilot of We are AI with a group of QPL librarians in Summer 2021.

3.1 Course Overview

The course is structures as a *learning circle*, an facilitated study group where participants learn about topic of common interest together. A learning circle is a peer-learning modality: There are no teachers or students and everyone learns the material together. That being said, one of the participants is designated as a facilitator, who decides the meeting schedule, keeps the group on task during meetings, and supports individual learners' participation and goals.

In slight violation of learning circle protocol, I co-facilitated the course during its initial in-person offering in Spring 2022, together with Lucius Bynum and Lucas Rosenblatt, who also helped refine and simplify presentation. We plan to facilitate or co-facilitate several future iterations of the course to have an opportunity to reflect on the material, and further refine it.

The overarching idea behind this course is that technology and ethics are deeply intertwined, and that people must step up to control how technology is used. Therefore, the course interleaves technical and ethical concepts, and reinforces the importance of human agency, as summarized in the course description:

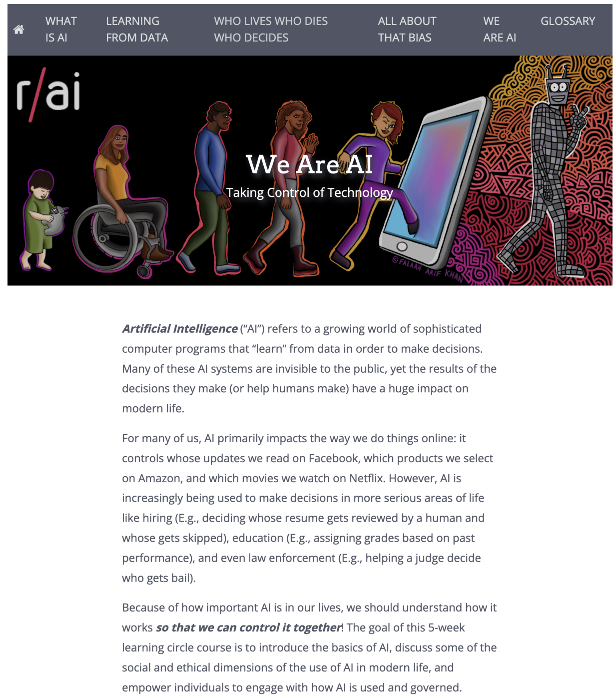

Figure 3: The landing page of *We are AI*, stating the goals of the course and giving an outline of the modules.

[...] Because of how important AI is in our lives, we should understand how it works so that we can control it together! The goal of this 5-week learning circle course is to introduce the basics of AI, discuss some of the social and ethical dimensions of the use of AI in modern life, and empower individuals to engage with how AI is used and governed."

The course is available for adoption by other public libraries in the US and internationally through the P2PU platform[4] and directly on the course website[5]. Figure 3 shows the landing page of the course website, with a verbal and visual statement of the goals of the course.

3.2 Course Organization and Materials

The course is organized as a collection of modules. Course participants meet in person to discuss each module in a focused 90-minute session. In our Spring 2022, these meetings were on a weekly basis, but they can also be condensed to take part over 1 or 2 weeks. The course website presents complete content for each module, with a simple and consistent navigation structure. Each module consists of a brief instructional video, individual and discussion-based exercises, and embeds notes and tips for the course facilitator. There is no homework (only optional supplemental readings) so all work takes place during the meeting. Modules are outlined below.

- Module 1: What is AI
- Module 2: Learning from data
- Module 3: Who lives, who dies, who decides

[4] *P2PU* website, https://learningcircles.p2pu.org/en/course/962/.
[5] *We are AI* course website, https://dataresponsibly.github.io/we-are-ai/.

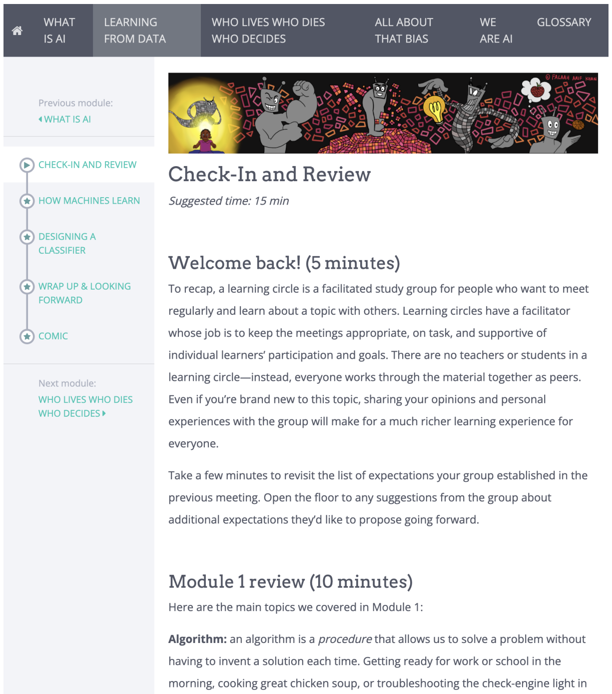

Figure 4: A module of the public education course. All course modules follow a consistent structure.

- Module 4: All about that bias
- Module 5: We are AI

All course materials, discussion prompts, and activities needed to run a group with minimal preparation on the part of the facilitator, are incorporated into the course and available on the website. The intention is to enable anyone to use this online material to facilitate a learning group or educate themselves on AI. No math, programming skills, or existing understanding of AI is required.

The course website also contains a glossary of terms and an extensive facilitator guide, describing the rationale behind course design and giving specific guidance on the content. We expect that the content of the course, the glossary, and the facilitator guide will continue to evolve as we iterate to improve presentation and make course material more accessible.

The *We are AI* comic book series was designed specifically as supplementary reading for this course, with one volume per module. We are preparing to release a Spanish-language version of this series, and are working on improving accessibility of the comic and of the course over-all for a range of abilities and levels of expertise.

4 NEXT STEPS

In this paper I described two ongoing education efforts, where the goal is to teach Responsible Data Science (RDS) and Responsible AI (RAI) to future data scientists (Section 2) and to members of the public (Section 3). I made all course materials are publicly available, hoping to inspire others in the community to come together to form a deeper understanding of the pedagogical needs of RDS and RAI, and to develop and share the much-needed concrete educational materials and methodologies.

Much work remains on addressing the educational and training needs of current and future data scientists, decision makers who use ADS, policy makers, auditors, and the public at large. And there is a need to rigorously evaluate the effectiveness of educational methodologies for these audiences. Both these directions are part of my ongoing and future work, and I hope that many others will join me in this quest for balance between research and practice.

ACKNOWLEDGMENTS

The work discussed here would not have been possible without collaboration and input from many of my colleagues.

Falaah Arif Khan, a talented artist and data scientist, is the driving force behind the comic books that have become a distinguishing feature of my teaching and, even more so, of how I think about and communicate complex socio-technical topics. Dr. George Wood, who holds a Ph.D. in Sociology and studies inequalities in public health and criminal justice, has been co-teaching RDS with me at NYU since Spring 2021, bringing the much-needed social science perspective to the course. Dr. Armanda Lewis, an education researcher, has been instrumental in connecting instructional methodologies for RDS with educational best-practices and developing assessment for components of the course. Dr. Debbie Yuster has been teaching RDS as "Ethics for Data Science" at Ramapo College since Spring 2021, and has helped me learn how to transfer the material across institutions and audiences.

Dr. Eric Corbett holds a Ph.D. in Digital Media and studies civic technology, community engagement and technical interaction design. He and I collaborated closely on the design of the We are AI public education course. Dr. Mona Sloane, Lucas Rosenblatt, Lucius Bynum, Meghan McDermott, Becky Margraf, Grif Peterson, Jeffrey Lambert, Sadie Coughlin-Prego, and Kaven Vohra all contributed to the development and refinement of the materials and methodologies used in this course.

This research was supported in part by NSF Awards No. 1934464, 1916505, and 1922658.

REFERENCES

[1] Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. 2019. Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Biased Outcomes. *Proc. ACM Hum. Comput. Interact.* 3, CSCW (2019), 199:1–199:30. https://doi.org/10.1145/3359301
[2] Falaah Arif Khan, Eleni Manis, and Julia Stoyanovich. 2021. Fairness and Friends. *Data, Responsibly Comic Series* 2 (2021). https://dataresponsibly.github.io/comics/
[3] Falaah Arif Khan and Julia Stoyanovich. 2020. Mirror, Mirror. *Data, Responsibly Comic Series* 1 (2020). https://dataresponsibly.github.io/comics/
[4] Alexandra Chouldechova and Aaron Roth. 2020. A snapshot of the frontiers of fairness in machine learning. *Commun. ACM* 63, 5 (2020), 82–89. https://doi.org/10.1145/3376898
[5] Amit Datta, Anupam Datta, Jael Makagon, Deirdre K. Mulligan, and Michael Carl Tschantz. 2018. Discrimination in Online Personalization: A Multidisciplinary Inquiry. In *Conference on Fairness, Accountability and Transparency, FAT 2018, 23-24 February 2018, New York, NY, USA*. 20–34. http://proceedings.mlr.press/v81/datta18a.html
[6] Anupam Datta, Shayak Sen, and Yair Zick. 2016. Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems. In *Proceedings of IEEE Symposium on Security and Privacy*. IEEE, 598–617.
[7] Irit Dinur and Kobbi Nissim. 2003. Revealing information while preserving privacy. In *Proceedings of the Twenty-Second ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, June 9-12, 2003, San Diego, CA, USA*, Frank Neven, Catriel Beeri, and Tova Milo (Eds.). ACM, 202–210. https://doi.org/10.1145/773153.773173

[8] Marina Drosou, H. V. Jagadish, Evaggelia Pitoura, and Julia Stoyanovich. 2017. Diversity in Big Data: A Review. *Big Data* 5, 2 (2017), 73–84. https://doi.org/10.1089/big.2016.0054

[9] Cynthia Dwork. 2011. A firm foundation for private data analysis. *Commun. ACM* 54, 1 (2011), 86–95. https://doi.org/10.1145/1866739.1866758

[10] Cynthia Dwork and Aaron Roth. 2014. The Algorithmic Foundations of Differential Privacy. *Found. Trends Theor. Comput. Sci.* 9, 3-4 (2014), 211–407. https://doi.org/10.1561/0400000042

[11] Ronald Dworkin. 1981. What is Equality? Part 1: Equality of Welfare. *Philosophy and Public Affairs* 10, 3 (1981), 185–246. http://www.jstor.org/stable/2264894

[12] Joseph Fishkin. 2014. *Bottlenecks: A New Theory of Equal Opportunity.* Oup Usa.

[13] Falaah Arif Khan, Eleni Manis, and Julia Stoyanovich. 2020. Fairness and Friends. https://dataresponsibly.github.io/comics/. *Data, Responsibly Comic Series* 2 (2020).

[14] Falaah Arif Khan, Eleni Manis, and Julia Stoyanovich. 2021. Fairness as Equality of Opportunity: Normative Guidance from Political Philosophy. *CoRR* abs/2106.08259 (2021). arXiv:2106.08259 https://arxiv.org/abs/2106.08259

[15] Falaah Arif Khan, Eleni Manis, and Julia Stoyanovich. 2021. Translational Tutorial: Fairness and Friends. In *ACM FAccT*.

[16] Keith Kirkpatrick. 2017. It's Not the Algorithm, It's the Data. *Commun. ACM* 60, 2 (Jan. 2017), 21–23. https://doi.org/10.1145/3022181

[17] Jon M. Kleinberg and Manish Raghavan. 2018. Selection Problems in the Presence of Implicit Bias. In *9th Innovations in Theoretical Computer Science Conference, ITCS 2018, January 11-14, 2018, Cambridge, MA, USA (LIPIcs, Vol. 94)*, Anna R. Karlin (Ed.). Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 33:1–33:17. https://doi.org/10.4230/LIPIcs.ITCS.2018.33

[18] David Lehr and Paul Ohm. 2017. Playing with the Data: What Legal Scholars Should Learn about Machine Learning. *UC Davis Law Review* 51, 2 (2017), 653–717.

[19] Armanda Lewis and Julia Stoyanovich. 2021. Teaching Responsible Data Science. *International Journal of Artificial Intelligence in Education (IJAIED)* (2021). https://doi.org/10.1007/s40593-021-00241-7

[20] Scott M. Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett (Eds.). 4765–4774. https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

[21] Ryan McKenna, Gerome Miklau, and Daniel Sheldon. 2021. Winning the NIST Contest: A scalable and general approach to differentially private synthetic data. *CoRR* abs/2108.04978 (2021). arXiv:2108.04978 https://arxiv.org/abs/2108.04978

[22] Jeffrey Mervis. 2019. Can a set of equations keep U.S. census data private? *Science Magazine* (2019). https://doi.org/doi:10.1126/science.aaw5470

[23] Arvind Narayanan and Vitaly Shmatikov. 2008. Robust De-anonymization of Large Sparse Datasets. In *2008 IEEE Symposium on Security and Privacy (S&P 2008), 18-21 May 2008, Oakland, California, USA*. IEEE Computer Society, 111–125. https://doi.org/10.1109/SP.2008.33

[24] Haoyue Ping, Julia Stoyanovich, and Bill Howe. 2017. DataSynthesizer: Privacy-Preserving Synthetic Datasets. In *Proceedings of the 29th International Conference on Scientific and Statistical Database Management, Chicago, IL, USA, June 27-29, 2017*. ACM, 42:1–42:5. https://doi.org/10.1145/3085504.3091117

[25] John Rawls. 1971. *A Theory of Justice*. Harvard University Press. http://www.jstor.org/stable/j.ctvjf9z6v

[26] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why Should I Trust You?": Explaining the predictions of any classifier. In *KDD '16 Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, San Francisco, CA, USA, 1135–1144. https://doi.org/10.1145/2939672.2939778

[27] John Roemer. 2002. Equality of opportunity: A progress report. *Social Choice and Welfare* 19, 2 (2002), 455–471. https://EconPapers.repec.org/RePEc:spr:sochwe:v:19:y:2002:i:2:p:455-471

[28] Matthew J. Salganik. 2017. Bit by Bit: Social Research in the Digital Age. Princeton University Press, Chapter Chapter 6: Ethics.

[29] Sebastian Schelter and Julia Stoyanovich. 2020. Taming Technical Bias in Machine Learning Pipelines. *IEEE Data Eng. Bull.* 43 (2020).

[30] Julia Stoyanovich. 2021. Hiring and AI: Let Job Candidates Know Why They Were Rejected. *The Wall Street Journal* (22 09 2021). https://www.wsj.com/articles/hiring-job-candidates-ai-11632244313

[31] Julia Stoyanovich and Falaah Arif Khan. 2021. What is AI? *We are AI Comic Series* (2021). https://dataresponsibly.github.io/we-are-ai/comics/

[32] Julia Stoyanovich, Jay J. Van Bavel, and Tessa West. 2020. The imperative of interpretable machines. *Nature Machine Intelligence* (2020). https://doi.org/10.1038/s42256-020-0171-8

[33] Julia Stoyanovich and Bill Howe. 2019. Nutritional Labels for Data and Models. *IEEE Data Eng. Bull.* 42, 3 (2019), 13–23. http://sites.computer.org/debull/A19sept/p13.pdf

[34] Julia Stoyanovich, Bill Howe, Serge Abiteboul, Gerome Miklau, Arnaud Sahuguet, and Gerhard Weikum. 2017. Fides: Towards a Platform for Responsible Data Science. In *Proceedings of the 29th International Conference on Scientific and Statistical Database Management, Chicago, IL, USA, June 27-29, 2017.* 26:1–26:6. https://doi.org/10.1145/3085504.3085530

[35] Julia Stoyanovich, Bill Howe, and H. V. Jagadish. 2020. Responsible Data Management. *Proc. VLDB Endow.* 13, 12 (2020), 3474–3488. http://www.vldb.org/pvldb/vol13/p3474-stoyanovich.pdf

[36] Julia Stoyanovich, Steven Kuyan, Meghan McDermott, Maria Grillo, and Mona Sloane. 2020. Public Engagement Showreel, Int 1894. *NYU Center for Responsible AI* (12 11 2020). https://dataresponsibly.github.io/documents/Bill1894Showreel.pdf

[37] Latanya Sweeney. 2013. Discrimination in Online Ad Delivery. *ACM Queue* 11, 3 (2013), 10. https://doi.org/10.1145/2460276.2460278

[38] Ke Yang, Vasilis Gkatzelis, and Julia Stoyanovich. 2019. Balanced Ranking with Diversity Constraints. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, Sarit Kraus (Ed.). ijcai.org, 6035–6042. https://doi.org/10.24963/ijcai.2019/836

[39] Ke Yang, Joshua R. Loftus, and Julia Stoyanovich. 2021. Causal Intersectionality and Fair Ranking. In *2nd Symposium on Foundations of Responsible Computing, FORC 2021, June 9-11, 2021, Virtual Conference (LIPIcs, Vol. 192)*, Katrina Ligett and Swati Gupta (Eds.). Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 7:1–7:20. https://doi.org/10.4230/LIPIcs.FORC.2021.7

[40] Ke Yang, Julia Stoyanovich, Abolfazl Asudeh, Bill Howe, H. V. Jagadish, and Gerome Miklau. 2018. A Nutritional Label for Rankings. In *Proceedings of the 2018 International Conference on Management of Data, SIGMOD Conference 2018, Houston, TX, USA, June 10-15, 2018.* 1773–1776. https://doi.org/10.1145/3183713.3193568

[41] Meike Zehlike, Ke Yang, and Julia Stoyanovich. 2021. Fairness in Ranking: A Survey. *CoRR* abs/2103.14000 (2021). arXiv:2103.14000 https://arxiv.org/abs/2103.14000

[42] Jun Zhang, Graham Cormode, Cecilia M. Procopiuc, Divesh Srivastava, and Xiaokui Xiao. 2017. PrivBayes: Private Data Release via Bayesian Networks. *ACM Trans. Database Syst.* 42, 4 (2017), 25:1–25:41. https://doi.org/10.1145/3134428