Ethnography of Data Science and Algorithmic Systems

Bibliography by Ian Lowrie, lowrie.ian@gmail.com

The social study of algorithms is a growth industry. There is a wealth of resources out there on specific topics, such as data cleaning, metadata communication, establishing training sets, or developing hardware. Here, though, I focus on high-level food for thought. It should also be pointed out that much of this work has been focused on academic and public-sector projects. The case studies that I have included as the last section of this bibliography are split between public and private contexts, but I think that the insights are mutually applicable.

For still further reading, you can check out the Critical Algorithm Studies reading list, or this literature review I put together with collaborators from Intel Labs.

Foundational statements

boyd, d., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication and Society*, 15(5), 662–679.

Extremely influential early examination of big data epistemologies and practices.

Seaver, N. (2015). The nice thing about context is that everyone has it. *Media, Culture & Society 37*(7):1101-1109.

Important response to boyd & Crawford.

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1)

Big, high-level summary. It's impossible to tell a story about the social study of algorithms without Kitchin's work.

Lowrie, I. (2017). Algorithmic rationality: Epistemology and efficiency in the data sciences. *Big Data & Society,* 4(1).

My own modest attempt to understand the epistemological procedures of algorithmic computation, with a particular focus on the consequences of "efficiency" replacing "truth" as an evaluative standard in big data science.

On machine learning, in particular

Burrell, J. (2015). How the machine "thinks:" Understanding opacity in machine learning algorithms. *Big Data & Society* 3(1).

This piece has a phenomenal analysis of how organizational and technological factors conspire to prevent algorithmic transparency in machine learning applications.

Stevens, H. (2017). A feeling for the algorithm: Working knowledge and big data in biology. *Osiris*, 32(1), 151–174.

Though based in a case study of academic biology, this has a widely applicable and sophisticated demonstration that algorithmic computation is not a scaled-up version of "pen-and-paper methods," but a new form of expert practice with its own social and epistemological dynamics.

How do users think about algorithmic systems?

These are all fairly straightforward case studies of specific engagements between users and algorithms. I know this literature less well than I would like, but these researchers seem to be at the forefront of this conversation. Highly recommend following the citation trail forward and backward here. "Algorithmic awareness" seems to be the term of art, here.

Berg, M. (2014). Participatory trouble: Towards an understanding of algorithmic structures on Facebook.
Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 8(3).

Bucher, T. (2017). The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms. *Information, Communication and Society, 20*(1), 30–44.

Lustig, C., & Nardi, B. (2015). Algorithmic authority: The case of bitcoin. In 2015 *48th Hawaii International Conference on System Sciences* (pp. 743–752).

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144*(1), 114-126.

Hamilton, K., Karahalios, K., Sandvig, C., and Eslami, M. (2014). A path to understanding the effects of algorithm awareness. In *CHI 2014*, 631-642. ACM Press.

Ethics

The literature on algorithms and ethics is *huge*. Here, I have limited myself to articles that directly touch on issues that surfaced in our tutorial.

Ananny, M., & Crawford, K. (2016). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society, 20*(3), 973–989.

On the hazards of importing moral regimes from one domain into another.

Gurses, S., & van Hoboken, J. (2017). Privacy after the Agile Turn.
https://doi.org/10.17605/OSF.IO/9GY73

Sophisticated and practical analysis of privacy and software development in an age of Agile and

IaaS/SaaS orientations.

Neyland, D. (2015). Bearing account-able witness to the ethical algorithmic system. *Science, Technology & Human Values*, 41(1), 50–76.

Excellent case study of a collaboration between an anthropologist and people implementing facial recognition. Does double duty in highlighting the sociotechnical aspects of data ethics and demonstrating how ethnography might productive engage with them.

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). Auditing algorithms: Research methods for detecting discrimination on internet platforms. Presentation at *Data and Discrimination: Converting Critical Concerns into Productive Inquiry*.

On algorithm audit as a method of inquiry. There are a few examples of this, including in Burrell's article on machine opacity, but I have yet to see a particularly convincing case study.

Case Studies

Bruun, H., & Sierla, S. (2008). Distributed problem solving in software development: The case of an automation project. *Social Studies of Science*, *38*(1), 133–158.

More focused on development than on data science as such, this piece nevertheless introduces a crucial distinction between "modular," "integral," and "translational" problem solving in software design that I have found incredibly illuminating when imported to other contexts of data work.

Fiore-Silfvast, B., & Neff, G. (2013). What we talk about when we talk data: Valences and the social performance of multiple metrics in digital health. *Conference Proceedings. Ethnographic Praxis in Industry Conference, 2013*(1), 74–87.

Early and influential study of data practices in eHealth, which presents a range of ongoing collaborations and provides a clear framework for studying data practices in other contexts.

Garnett, E. (2016). Developing a feeling for error: Practices of monitoring and modelling air pollution data. *Big Data & Society*, *3*(2).

Nice look at the daily work practices and professional identities of the various folks involved in large-scale remote sensing.

Jaton, F. (2017). We get the algorithms of our ground truths: Designing referential databases in digital image processing. *Social Studies of Science, 47*(6), 811–840.

Case study of an image recognition algorithm development project. The project is academic, but the dynamics surveyed -- of the relationship between constructing algorithms, curating data sets that train those algorithms, and convincing peers of the superiority of a new algorithmic approach-- absolutely parallel those in industrial contexts.

Nadim, T. (2016). Data Labours: How the Sequence Databases GenBank and EMBL-Bank Make Data. *Science as Culture*, 25(4), 496–519.

Cool, A. (2016). Detaching data from the state: Biobanking and building Big Data in Sweden. *BioSocieties 11*(3):277-295

Both are excellent studies of databasing, data sharing, and collaboration between domain experts, government watchdogs, users, and maintainers.

Stilgoe, J. (2018). Machine learning, social learning and the governance of self-driving cars. *Social Studies of Scienc*e, 48(1), 25–56.

Like most studies of large-scale industrial algorithm development, Stilgoe is definitely looking from "outside," but nevertheless makes a number of important points about the relationships between business models, technical decisions, and the public sphere.