Public Library Quarterly

ISSN: 0161-6846 (Print) 1541-1540 (Online) Journal homepage: http://www.tandfonline.com/loi/wplq20

Algorithms: avoiding the implementation of institutional biases

Lori Ayre & Jim Craner

To cite this article: Lori Ayre & Jim Craner (2018) Algorithms: avoiding the implementation of institutional biases, Public Library Quarterly, 37:3, 341-347

To link to this article: https://doi.org/10.1080/01616846.2018.1512811

Published online: 21 Sep 2018.

PUBLIC LIBRARY QUARTERLY
2018, VOL. 37, NO. 3, 341–347
https://doi.org/10.1080/01616846.2018.1512811

TECHNOLOGY COLUMN

Algorithms: avoiding the implementation of institutional biases

Lori Ayre and Jim Craner

The Galecia Group, Petaluma, CA, USA

ABSTRACT
Computer algorithms, the logic and code that power automated decision-making programs, increasingly dominate many aspects of modern society. There are already many examples of institutional biases – including ideological bias, racism, sexism, ableism – being solidified in algorithms, causing harm to already underprivileged populations. This article explores library-specific and society-wide examples as well as efforts to prevent the implementation of these biases in the future.

ARTICLE HISTORY
Received 5 July 2018
Accepted 26 July 2018

KEYWORDS
Algorithmic decision-making; algorithms; bias; filter bubble; inadvertent injustice; institutional bias; machine learning

You've probably heard of "algorithms" in the context of computer programs or Internet services, like Netflix's "Watch Next" feature. The Netflix algorithm analyzes the viewing habits of millions of people to suggest movies for people with similar tastes. Of course, most people intuitively know that a viewer who watched "Iron Man" and "The Avengers" is likely to watch the next "Captain America" movie, but the algorithm automatically compares the movies watched by millions of people to find those relationships among rarer, more esoteric films. Algorithms aren't just about movie recommendations or Amazon shopping suggestions. They affect our lives in every possible way, with serious and significant impacts. And as more of our services and infrastructure become digitized, algorithms influence parts of our lives that we might not have suspected.

At its most basic, an algorithm is just a sequence of steps. The Simple English Wikipedia cites a recipe as a good example of an algorithm: starting with the ingredients (input), certain steps are performed in a certain order (algorithm), resulting in a complete dish (output). Computers use algorithms in the form of software programs that define those steps, processing input data, resulting in output data.

Algorithms are implemented in many types of modern library software. Some library self-check software makes recommendations based on what a patron is checking out. One discovery layer product uses the circulation

CONTACT Lori Ayre lori.ayre@galecia.com The Galecia Group, 15 Old Creek Road, Petaluma, CA 94952, USA.

activities of all of their users in their "recommendations" algorithm. And librarians use algorithms every time they alphabetize a book or shelve Harry Potter DVDs in the proper sequel order.

For another example of library-specific algorithms and their use, consider the Library Journal's Index of Public Library Service. This survey index involves analyzing a number of service metrics, population figures, and budget expenditures to yield a numerical "score" so that libraries can compare those factors to other libraries in their budget peer group. This list is based on an algorithm: the various inputs – library visits, circulation and e-circulation, public computer usage, program attendance, and service population – are fed into a series of calculations yielding a single numeric score (LJ 2017a).

One would assume that a higher LJ score means a "better" library but according to the FAQ, these scores do **not** "measure the quality, excellence, effectiveness, value, or appropriateness of library services." To do so would involve assigning a value or weight to certain subjective qualities, ultimately favoring a certain type of library over others. In their words, they are trying **not** to endorse a certain strategic objective, such as "*'library as place' versus remote library use versus community outreach and engagement*. (LJ 2017b)" In fact, they are intentionally striving to ensure a particular neutrality and they are very clear and transparent about how they gather and synthesize their data, and the possible negative aspects of their decisions.

This is an example of a neutral implementation of an algorithm, with a very transparent explanation of the input data sources and the mathematical calculations themselves. Most importantly, considering some of our later examples, the results are intended strictly for informational purposes. If this algorithm was used to award grant funding, however, you can see that there might be considerable controversy about the details of the algorithm and which factors were weighted, or favored, over others. As we will see in other examples, such as algorithms that determine prison sentences and parole periods, both the source of the input data and the details of the algorithms are not transparent and need to be critically examined, especially when they have such a great impact on people's lives.

Internet content filters provide another example of algorithms. Depending on how they are implemented, they can have very negative effects. While filters can be effective at blocking out undesirable content, they have been proven to be invariably inaccurate to some degree (Ayre 2004; Houghton-Jan 2010). The worst example of an Internet content filter is one that blocks web pages based on keywords. These filters can end up blocking legitimate educational info, such as breast cancer treatment resources due to the appearance of the word "breast." (NCAC 2016).

Even the best Internet content filters rely upon questionable algorithms either because of the input data that is relied upon or how the program is written. For example, most content filters rely upon categories which can be blocked or not. Depending on how a web page is categorized, it might be appropriately blocked or it might be erroneously blocked. This depends on categories that have been applied to the web page. For example, one person's women's health page is another person's "pro-abortion" page. How the page is categorized depends on how, or who, assigns the category – that's part of the input. And then the program might block that page or it could block the entire website or possibly even everything at that IP address – it's all part of the recipe.

In the past, the ACLU has sued schools using Internet filters that blocked LGBT educational resources while allowing conversion therapy information to be accessed. In this case, the algorithm underlying the content filters enabled them to be used to support ideological choices. This is another way that filters often go terribly wrong (ACLU 2015).

The problem is that algorithms are not usually as transparent as the LJ's Index or content filters that block based on an editable list of keywords. The inputs and processing steps are often proprietary. For example, the list of websites and the filter categories to which they've been assigned is not published nor is the precise manner in which the categories are assigned.

It is often the case that all we know about an algorithm is based on the output or result of using it but because there are no human faces associated with algorithms, we tend to view them as unbiased, more trustworthy and more objective than an individual would ever be. In fact, algorithms are the result of human endeavor and human-generated data sets so they are just as biased as we are. We just can't see it.

Increasingly, algorithms are affecting every aspect of our lives in the modern world. Websites can take loan applications and immediately offer automated mortgage loan decisions – we assume those decisions are based on logic and fair data but what if some of that data, or the algorithms themselves, are rooted in false assumptions, biased data collection, or outright institutionalized racism? Are we re-implementing, digitally, the old real estate concept of "redlining," or discriminating against certain races of people in housing and financial transactions? According to an article in MIT Technology Review (Spielkamp 2017), this "automated decision-making" (ADM) has already been found to be problematic.

WIRED reports that US states are using algorithmic computer systems – developed, controlled, and kept secret by corporate developers – to determine criminal sentences and parole lengths but nobody knows how these computer systems determine someone to be "high risk (Tashea 2017)." Real estate transactions and the criminal justice system are just two of the many sectors being transformed with automated business logic based on algorithmic decision-making.

Just like with content filters, there are two ways that algorithmic decision-making can be problematic. One is that the data being input may be biased or inaccurate. Second is that the underlying programs (the recipe) can be biased. Some of these programmatic biases can be built into software purposefully while other times, biases find their way into the algorithms accidentally – a form of "inadvertent injustice," so to speak (Zhou 2018).

Machine Learning is the latest evolution of algorithmic decision-making. Machine learning is accomplished by "training" a computer program with thousands of "known" data points, which it then analyzes so it can then classify "unknown" data in the future (sometimes referred to as a simpler form of "Artificial Intelligence" or "AI"). A great example of machine learning is provided in radiology. Thousands of CAT scan or X-ray images are fed into a computer program along with data about whether or not the image represents a cancer diagnosis. The program then "learns" what features indicate cancer so that it can scan future images and present a diagnosis – providing a backup for human radiologists.

The radiology example has a laudable goal, with relatively straightforward input data: cancer or no cancer. But if the input data is based on questionable practices or collection methods (Daly and Olopade 2015), then future decisions by those algorithms will replicate the same errors.

Consider the "source data" that goes into an algorithm-based system to calculate prison and parole sentences. This data consists of a number of factors, some of which are based on the individual defendant's traits, some of which are based on historical data from society as a whole. As Nick Thieme writes in an article about "computational injustice": "*AI's unique talent for finding patterns has only perpetuated our legal system's history of discrimination… Since people of color are more likely to be stopped by police, more likely to be convicted by juries, and more likely to receive long sentences from human judges, the shared features identified are often race or proxies for race. Here, computational injustice codifies social injustice.*" (Thieme 2018) In other words, social bias and algorithmic bias can reinforce each other in a feedback loop – a vicious circle of injustice accelerated by our big data tools.

While on the topic of vicious circles and feedback loops, consider algorithm-driven portals like Google and Facebook. Both companies use an algorithm that presents new content based on previous choices. In other words, if a liberal person clicks on a news link from a presumed liberal source, shared by a liberal friend, then Facebook will be more likely to present more of that "liberal" content in the future, and vice versa. This leads to what is called "the filter bubble" effect, where people are put into a silo with little exposure to contrary points of view. As in the examples above, this can create an indefinite feedback loop. Given that over half of Millennials and almost half of Baby Boomers get their news from Facebook, that's one vicious and far-reaching feedback loop (Mitchell, Gottfried, and Matsa 2015).

Much has been written about the contribution of these political filter bubbles to the recent breakdown in civil discourse in the US, but these types of bubbles can also spread deliberate misinformation about other critical topics, such as health information. Renee DiResta writes in WIRED that anti-vax (vaccination) groups easily recruit new members via social media recommendation engines but also by buying ads in search engine results, ensuring that querying parents see deliberately-posted non-factual "medical" information, in direct opposition to accepted scientific consensus (DiResta 2018). She quotes Michael Golebiewski of Bing who describes this *as a "data void," or search void: a situation where searching for answers about a keyword returns content produced by a niche group with a particular agenda.*

Thankfully, there are developers, researchers, activists, and government officials aware of these issues and working to mitigate them, despite the daunting challenge posed. Just a few examples for further reading:

- A team of researchers at the Alan Turing Institute has developed a rigorous model for determining demographic and other types of bias in selection models (Kusner et al. 2017). From the paper: "[o]ur definition of counterfactual fairness captures the intuition that a decision is fair towards an individual if it is the same in (a) the actual world and (b) a counterfactual world where the individual belonged to a different demographic group." Humans aren't race-blind but we can strive to ensure that our algorithms are.
- A 2017 report from the Pew Research Center quoted industry analyst Susan Etlinger: "*So to ensure that we use algorithms successfully, whether for financial or human benefit or both, we need to have governance and accountability structures in place.*" (Rainie and Anderson 2017) This report is a comprehensive overview of the topic of algorithms and their role in society. The current state of the art is examined with plenty of examples, followed by speculative prediction about the future impact of algorithmic decision-making systems, including a call for a potential regulatory structure.
- The ACLU has begun working on cases related to systemic injustice due to hidden or faulty algorithm implementations, such as a first-of-its-kind measure in New York City that creates an advisory body to study algorithm-related topics (Richardson 2017): *The legislation will create a task force to review New York City agencies' use of algorithms and the policy issues they implicate. The task force will be made up of experts on transparency, fairness, and staff from non-profits that work with people most likely to be harmed by flawed algorithms. It will develop a set of recommendations addressing when and how algorithms should be made public, how to assess whether they are biased, and the impact of such bias.*
- The Obama administration issued a number of reports on the future of AI, including a detailed strategic plan calling for research "*to understand*

the ethical, legal, and social implications of AI, and to develop methods for designing AI systems that align with ethical, legal, and societal goals." (NSTC 2016) This plan is a high-level overview of recent and future government AI research efforts and funding, along with recommendations for promoting the growth and adoption of AI in society. (The current administration recently convened an industry/government meeting about the future of AI research in the United States, but reports from attendees appear to be mixed (Simonite 2018)).
- One of the most recent and helpful resources come from a team at Data & Society – *Algorithmic Accountability: A Primer* (Caplan et al. 2018). This report, prepared for Congress in 2018 by Data & Society researchers, provides a detailed overview of algorithmic bias with specific examples based on criminal justice policy. In addition, the report contains a discussion of various ethical facets involved in algorithm implementation and algorithmic decision-making. Finally, accountability structures are discussed, including by journalism and by regulation.

No matter how far removed computer algorithms seem from everyday life, they are not just a trustworthy series of 1's and 0's. The algorithms, the choice of data to use, how it is processed, the rules that are applied – these are all created by people, with their respective history and biases and values.

As humans, we all have implicit biases. And as we build these new systems – facial recognition, AI, analytical algorithms – we're creating them in our own image, with these biases baked in. It's critical that we examine our data, the logic, **and** the humans creating them rather than trusting that "the computer must be right."

Notes on contributor

Lori Bowen Ayre is the Principal Consultant at The Galecia Group, a library technology consulting firm. She is nationally recognized as an expert in automated materials handling, RFID, and library workflow optimization. Recognizing that technology is a catalyst of change, she works to help libraries harness the power of technology as they navigate the changing needs of their communities. She can be reached at lori.ayre@galecia.com.

Jim Craner is the Development & Operations Manager for The Galecia Group. A former Code for America Fellow and Americorps volunteer, he has extensive experience building websites and mapping applications for libraries and nonprofit organizations. He can be reached at jim@galecia.com.

ORCID

Lori Ayre http://orcid.org/0000-0002-3604-9530

References

American Civil Liberties Union (ACLU). 2015. Anti-LGBT web filtering. Accessed July 20, 2018. https://www.aclu.org/issues/lgbt-rights/lgbt-youth/anti-lgbt-web-filtering.

Ayre, L. 2004 March–April. Filtering and filter software. *Library Technology Reports* 20. doi:10.5860/ltr.40n2.

Caplan, R., J. Donovan, L. Hanson, and J. Matthews 2018. Algorithmic accountability: A primer. Accessed July 20, 2018. https://datasociety.net/output/algorithmic-accountability-a-primer/.

Daly, B., and O. Olopade. 2015. Race, ethnicity, and the diagnosis of breast cancer. *Journal of the American Medical Association* 313 (2):141–42. doi:10.1001/jama.2014.17323.

DiResta, R. 2018. The complexity of simply searching for medical advice. Accessed July 20, 2018. https://www.wired.com/story/the-complexity-of-simply-searching-for-medical-advice/.

Houghton-Jan, S. 2010. Internet filtering. *Library Technology Reports* 48 (1):25–33. doi:10.5860/ltr.46n8.

Kusner, M., J. Loftus, C. Russell, and R. Silva. 2017. Counterfactual fairness. *ArXiv*. https://arxiv.org/abs/1703.06856.

Library Journal (LJ). 2017a. The LJ index: Score calculation algorithm. Accessed June 6, 2018. https://lj.libraryjournal.com/americas-star-libraries-score-calculation-algorithm. Archive available: http://web.archive.org/web/20171018025809/https://lj.libraryjournal.com/americas-star-libraries-score-calculation-algorithm.

Library Journal (LJ). 2017b. The LJ index: Frequently asked questions. Accessed June 6, 2018. https://lj.libraryjournal.com/stars-faq/. Archive available: http://web.archive.org/web/20180102185241/http://lj.libraryjournal.com:80/stars-faq/.

Mitchell, A., J. Gottfried, and K. E. Matsa 2015. Facebook top source for political news among millennials. Accessed July 20, 2018. http://www.journalism.org/2015/06/01/facebook-top-source-for-political-news-among-millennials.

National Coalition Against Censorship (NCAC). 2016. Internet Filters. Accessed July 20, 2018. https://ncac.org/resource/internet-filters-2.

National Science and Technology Council (NSTC). 2016. The National artificial intelligence research and development strategic plan. Accessed July 20, 2018. https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/national_ai_rd_strategic_plan.pdf.

Rainie, L., and J. Anderson 2017. Code-dependent: Pros and cons of the algorithm age. Accessed July 20, 2018. http://www.pewinternet.org/2017/02/08/code-dependent-pros-and-cons-of-the-algorithm-age/.

Richardson, R. 2017. New York City takes on algorithmic discrimination. Accessed July 20, 2018. https://www.aclu.org/blog/privacy-technology/surveillance-technologies/new-york-city-takes-algorithmic-discrimination.

Simonite, T. 2018. The Trump administration plays catch-up on artificial intelligence. Accessed July 20, 2018. https://www.wired.com/story/trump-administration-plays-catch-up-artificial-intelligence/.

Spielkamp, M. 2017. Inspecting algorithms for bias. Accessed July 20, 2018. https://www.technologyreview.com/s/607955/inspecting-algorithms-for-bias/.

Tashea, J. 2017. Courts are using AI to sentence criminals. That must stop now. Accessed July 20, 2018. https://www.wired.com/2017/04/courts-using-ai-sentence-criminals-must-stop-now/.

Thieme, N. 2018. We are hard-coding injustices for generations to come. Accessed July 20, 2018. https://undark.org/article/ai-watchdog-computational-justice/.

Zhou, L. 2018. Is your software racist? Accessed July 20, 2018. https://www.politico.com/agenda/story/2018/02/07/algorithmic-bias-software-recommendations-000631.