This is a preprint of an article published by Taylor & Francis in International Information & Library Review, on Dec. 2019 available online:

<http://www.tandfonline.com/doi/full/10.1080/10572317.2019.1669937>

Column Title: Perspectives on Public Services

Column Editor: Andrea Falcone, Dean of W. Frank Steely Library, Northern Kentucky University, Highland Heights, KY, falcone1@nku.edu

This column examines advances in public services internal and external to libraries. The focus is on how public services, such as instruction and education, programming, research consulting, and circulation, evolve and impact users. The strength of the column is its broad, international focus and contributors are encouraged to explore issues and recent advances in public services relevant to their geographical region, as well as the larger, global audience. Interested authors are invited to submit proposals and articles to the column editor at falcone1@nku.edu.

Teaching Algorithmic Bias in a Credit-Bearing Course

CAROLYN CAFFREY GARDNER

Information Literacy Coordinator, California State University, Dominguez Hills, Carson, CA
USA

cgardner@csudh.edu

Abstract

Information literacy instruction has become increasingly nuanced with the widespread adoption of critical approaches to teaching and the ACRL Framework. Librarians are already teaching information evaluation strategies, however, more work can be done in the area of understanding algorithmic decision making and bias. This column describes how a public university integrated lessons on algorithmic bias into a credit-bearing information literacy course for a general undergraduate audience. The activities and readings can be adapted to a one-shot instruction environment and the collaborative process for designing the curriculum is also shared.

Keywords

Algorithmic bias; information literacy; evaluating information, academic libraries

Introduction

As a librarian in a college setting, I am frequently told humorous stories involving information systems ranging from accidental obscene typos to mistyped search queries. These information system stories have recently included anecdotes of algorithms gone awry. An algorithm is a set of instructions or rules used by a computer to perform a specific task such as organizing search results by relevance ("City Council Passes First Bill in Nation to Address Transparency, Bias in Government Use of Algorithms," 2017). In one such story, a student described getting in trouble from their parents based on the advertisements Google began to serve the family and the student. The student worked at a fast food location adjacent to a strip club and over time based on geolocation data the student began to receive nearly constant advertisements for this location.

Needless to say, the parents did not find this quite as humorous as the student. We've all likely experienced a mild algorithm gone wrong story and they are usually inconsequential. For example you may have personally experienced an odd advertisement or an inability to have a voice assistant understand what you're asking such as the viral video of a toddler asking Alexa for the song "Baby Shark" (Dawn, 2018). However, we encounter these opaque algorithmic forces in all aspects of our lives, many of which have more serious implications, such as job applications, college admissions, predictive policing, and financial credit instruments to name a few.

Research, public concerns, and activism calling out the power of major tech companies have even influenced legislation on the implications of algorithms in our society. In the United States, two related bills calling for governmental accountability within algorithmic decision-making have been introduced by the Senate and the House of Representatives. The Algorithmic Accountability Act (2019), introduced by senators Cory Booker and Ron Wyden, would require companies to perform impact assessments of their algorithms as they relate to bias, fairness, and discrimination. Another associated bill in the House of Representatives, the Deceptive Experiences to Online Users Reduction Act (S.1084) would prohibit large online platforms from designing interfaces that reduce autonomy in decision making. Both of these bills have been referred to committee and not discussed further at the time of this writing. These federal bills though are not isolated incidents, New York City is the first major U.S. city to enact legislation to develop a task force to review the city agencies' use of algorithms and potential discrimination ("City Council Passes First Bill in Nation to Address Transparency, Bias in Government Use of Algorithms," 2017) and Washington State is exploring a similar state-wide law ("Washington State Legislature," 2019). The American public is largely concerned about how algorithms can impact privacy and fairness. A recent study from Pew Research Center notes that 58% of Americans are concerned with bias in algorithms (A. Smith, 2018). The European Union is also enacting legislation in this area, specifically the General Data Protection Regulation (GDPR). While many know GDPR for protecting consumer privacy, components of the law also require the clear explanation of algorithmic decision making applied to the person seeking protection (Brkan, 2019). Finally, the Australian Competition and Consumer Commission just released an investigative report on the conduct of companies like Facebook and Google and among their recommendations is a closer look at how algorithms are used to match advertising with customers (Bryone & Westbrook, 2019). Countries not represented here may not be experiencing the same widespread universal adoption of Google and Facebook and subsequent backlash, for example in China where citizens deal with censorship of these platforms.

This public debate over information seeking and evaluation has likewise captivated librarians. Our world (online and off!) is increasingly mediated, filtered, personalized, and predicted by algorithms. It is imperative that libraries play a central role in teaching how to appraise, interrogate, and analyze the roles algorithms play in structuring our information seeking and use. Librarians are already beginning to tackle the role of algorithms in vendor discovery systems (Reidsma, 2019) and library search interfaces (C. L. Smith & Matteson, 2018). When it comes to educating library patrons, Safiya Noble's research on Google and algorithmic bias is foundational (Noble, 2018a). I myself became interested in this topic after attending a lecture by Safiya Noble where she described racialized gender bias experienced in search results. Specifically she described searching for "black girls" and finding sexualized information and

pornography. The implications for how girls see themselves misrepresented or not represented in information systems inspired me to focus more on this kind of information evaluation in library information literacy sessions.

As I began looking for inspiration in the LIS field on how to teach library users about algorithmic bias particularly when it comes to search results and recommendations, I found much of the available materials very theoretical or designed for advanced students (upper-division undergraduates or graduate students) in a particular major of study such as communication or computer science¹. I did find one mention of assigning Safiya Noble's work to undergraduates in an information literacy course (Fisher, 2017). The library where I coordinate the information literacy program serves a large percentage of Pell-grant eligible, first-generation, underrepresented racial minority, and/or undocumented students. Our information literacy program facilitates many one-time workshops within courses offered the first or second year of students' college careers, and they may not yet have the disciplinary or technical knowledge required for the example syllabi and course activities I found as examples. Furthermore, I found little guidance in the Association for College and Research Libraries (ACRL) *Framework for Information Literacy in Higher Education*. While this document describes how authority is constructed and contextual, the critical analysis of algorithms are not explicitly mentioned. At the same time as I was personally wrestling with these issues, I had submitted a proposal to teach a first-year seminar course which was accepted. This would be a topical course for incoming first-year students, which also provided college success skills. Library faculty in our Reference, Instruction, and Outreach (RIOS) department were simultaneously working on developing a proposal for a credit-bearing information literacy course, which would meet a general education designation. The department decided that the first-year seminar course, UNV 101, would be an excellent vehicle to pilot some ideas on how to teach algorithmic awareness for the information literacy course. This column will detail the activities from this pilot course and describe the creation of our permanent credit-bearing information literacy course, LIB 151: Fundamentals of Information Literacy, with a significant component on algorithmic bias. Many of the activities mentioned can be modified to fit a stand-alone information literacy session in the context of one-shot instruction programs as well. The process for how the RIOS department of eight faculty librarians collectively arrived at the curriculum will also be shared.

University 101: Hackers, Pirates, and Robots

To help understand what knowledge and interests students were bringing to the pilot course, UNV 101: Hackers, Pirates, and Robots - Finding and evaluating information in the digital age, I developed a short survey. Students were asked to identify topics they were familiar with and ones they were most excited about. Most students, 88% (n=25), said they were unfamiliar with algorithms as a concept either in search engines or other technologies, and 92% were eager to learn more.² I tried to use multi-modal materials that students would find engaging and were

---

¹ For examples of these advanced syllabi I recommend Professor Sandvig's Algorithmic Culture course at University of Michigan <http://www-personal.umich.edu/~csandvig/710F18/> and Mike Annany's Algorithm and Big Data course at the University of Southern California <http://mike.annany.org/ananny-asims-algorithmsBigData-syllabus-may29.pdf>.

² Approved human subject research *CSUDH IRB No. 19-024*

© C. Caffrey Gardner

Address correspondence to Carolyn Caffrey Gardner, Information Literacy Coordinator, California State University, Dominguez Hills, 1000 E. Victoria Street, Carson, CA 90747. E-mail: cgardner@csudh.edu

written for a non-scholarly audience, because these were first-year students with varied majors. The activities and readings are described in more depth below.

Homework: Analyzing Google Image Search Results

To introduce this assignment we watched a short video on how Google indexes the web designed for a popular audience (*How Search Works*, 2010) followed by a discussion on indexing and viewing cached websites. Next, students were introduced to algorithmic bias through a brief lecture and were assigned to read a *Time* magazine summary of Safiya Noble's research for a general audience (Noble, 2018b). This culminated in a homework assignment where students compared the narratives represented in two different search engines (Google, DuckDuckGo), identified advertisements in results, and analyzed bias within Google Image search results. I selected DuckDuckGo as a search engine to contrast with Google because of their marketing and focus on avoiding the filter bubble of personalized search.

Figure 1: Assignment Instructions

Part I - Analyzing & Comparing Search Engines:

- 1) Login to your personalized Google account (you can use your CSUDH google account if you do not have another one).
- 2) Run three separate searches in Google:
  - a. Climate Change
  - b. Immigration to the U.S
  - c. Your choice – pick something you have looked for recently!
- 3) For each search, take a screenshot image of your first page of results. To take a screenshot on a Windows computer use the “PrntScn” button. On a Mac use Shift-Command-4 and select the portion of your screen. You can also use free programs such as Jing or [Screenr](#) to take quick and customizable screenshots.
- 4) Next do the same searches in DuckDuckGo and take a screenshot image of your results.
- 5) In a document, paste your screenshots for each search and compare your results. Write approximately a paragraph for each of the three searches (3 paragraphs in total). Where do you see advertisements in your results? What services and websites are prioritized in each search system? How did each search system interpret what you were searching for? Which results are more relevant to you? What narrative is being told in the results?

Part II - Analyzing Google Images

- 1) Select a phrase to search for from the following list: God, working mom, Mexican, lead singer, doctor, handsome, beautiful women, cute children, happy family, professor, or expert.
- 2) Search Google Images for your phrase with Safe Search on.
- 3) Take a screenshot image of your results and in a short paragraph describe what you are seeing. What autosuggestions does Google recommend? What is the perspective represented by these searches? What narrative is being told in the results?

By pre-selecting particularly problematic Google image searches for concepts rather than commonly identifiable ones, like a video game or a place, students were able to more easily analyze harmful narratives in search results (see Figure 2). One Google image search result that generated a lot of engaging conversation was that of “God.” Search results often featured a white male representation of a Judeo-Christian God which many of the students felt was an accurate depiction of this concept. This resulted in a great chance to talk with students about how being part of the dominant culture or narrative represented may make it more difficult to spot biases and perspectives in an algorithm (Figure 1).

Figure 2: Sample Student Work

Part II – Analyzing Google Images
1. beautiful women

3. I am only seeing white women in the results for beautiful women. Google is suggesting that women of color are not beautiful which is not true. The perspective represented by these searches is one that seems to be racist. Google is not displaying any women of color, basically saying that only the white women are beautiful. These results can be offensive to many people especially women.

The assignment was graded with a rubric that assessed the depth of reflection, the clarity of images and text, and the linkages to course concepts and readings. Overall, students excelled at deciphering the Google image search results (see Figure 2) but had a difficult time identifying advertisements in text-based Google searches similar to what other researchers have found (Stanford History Education Group, 2016). The students commented verbally in class that they found the part asking for advertisements very easy and did not realize they had misidentified some initially. In future iterations, I would recommend requiring students to search Google on a desktop or laptop so that they are also able to analyze from where definition information on the sidebar with select search terms comes.

Activity: Creating an Algorithm

To help illustrate how algorithms make decisions using “if-then” coding syntax without doing any computer programming ourselves we wrote algorithmic directions for how to do a specific task. While there are other kinds of conditional logic used in computer programming I

© C. Caffrey Gardner

Address correspondence to Carolyn Caffrey Gardner, Information Literacy Coordinator, California State University, Dominguez Hills, 1000 E. Victoria Street, Carson, CA 90747. E-mail: cgardner@csudh.edu

anticipated that students would be able to more effectively write statements using this format, for example, “if the jar of peanut butter is closed then place your open hand on the jar and twist in a counter clockwise direction until the lid comes off the jar.” The purpose behind this activity was to illustrate the difficulty in developing rules that are comprehensive and clear without imparting assumptions and biases from the developer. I began by breaking the students into small groups and asked them to write down a list of “if-then” statements detailing how to make a peanut butter and jelly sandwich for someone who had never encountered a sandwich before and had no prior knowledge of how to make one. After 20 minutes of group work, we came back together as a class, and I attempted to make the sandwiches according to group directions. Some of the rules were clearer than others and resulted in something that looked like a sandwich, while others had me spreading peanut butter onto my hands to make a point about an unclear or missing step. While this activity was a lot of fun, students also discussed afterward that even what seems like a relatively simple task can be difficult to automate in this way. Some students who were initially skeptical of the idea of being aware or evaluating algorithms began to have a deeper appreciation for the potential unintended consequences of automated decision making. Depending on your space and comfort-levels with a sticky mess you could adapt this activity to feature any mechanically-inclined task such as putting together a jigsaw puzzle.

Video & Discussion

Over the course, we also had several in-class activities where students discussed a reading or a video in small groups using guided questions and then reported out to the rest of the class. Videos were especially well-received and described in mid-semester evaluation feedback as especially engaging by the students. Joy Buolamwini’s TED talk on her experience encountering racial bias in face recognition software resonated with students, and 22 out of 25 students wrote a response indicating they had faced as people of color with being recognized by Snapchat filters, Siri, and other voice or facial recognition software (Buolamwini, 2016). At the time of the class, Google’s Art and Culture app had just released a feature that matched your selfie to famous works of art and students read a short piece about how not having representation within the app matters (Shu, 2018). Not knowing how students would respond to racist imagery and ideas I tried to be sensitive to students’ experiences by selecting examples that seemed rather silly (like social media filters). As students gained an understanding of algorithmic bias we did dive into more substantial algorithmic harms by discussing a piece on off-line algorithms used to make decisions about parole and recidivism risk which is used by some court systems (Angwin, 2016). In free-write responses to the piece, students described how the real examples of this algorithm in a criminal justice setting was a significant turning point in their understanding of the ethics and implications of algorithm decision making. Finally, to balance the perspectives from the passionate readings and videos we also read research on public opinion on algorithms (Rainie, Janna, & Erson, 2017) and personalized music algorithms that enhance discovery of new music (Pasick, 2015).

Privacy Tools

At this point in the course, most students were persuaded that this is a topic people should be generally aware of and think critically about. Some students had stronger feelings that echoed some of the researchers we read or watched and felt a need to take action. Some of the students ideas included designing educational opportunities around understanding algorithms in high

school and changing web interfaces to be more transparent about what is happening behind the scenes. I had not initially considered that this would be of interest to students and in future iterations, I would like to spend more time collectively discussing how we can improve our algorithmic world including discussing the proposed U.S. legislation. I did introduce students to tools they can use while browsing the web to decrease their own personalized search results and break out of their own geolocation and past history filter bubbles. In exploring browser extensions the tools from the Electronic Frontier Foundation, such as Panoptickick (<https://panoptickick.eff.org/>), were especially popular with students.

Library 151: Fundamentals of Information Literacy

While UNV 101 was a one-time opportunity, the department was eager to capitalize on the success and enthusiasm students had for learning about algorithms, privacy, and information privilege as we were developing our own course. Because the *ACRL Framework* is a guiding document for our information literacy program we thought it would be best to align the learning outcomes for the module on algorithmic bias with concepts within the framework (see Table 1). Though algorithmic bias is not mentioned explicitly within the *ACRL Framework*, evaluating information and reflection on its creation is certainly covered (ACRL, 2016). Content on understanding and assessing algorithms can align with an information literacy program guided by these concepts.

Table 1: Outcomes Alignment Chart

LIB 151 Course Module Learning Outcome	ACRL Framework Concept
Discuss the ways that search systems privilege some perspectives and communities while presenting barriers to others	Scholarship as Conversation: “recognize that systems privilege authorities and that not having a fluency in the language and process of a discipline disempowers their ability to participation and engage” (Disposition)
Articulate that search results are not a natural reflection of the world	Information Creation is a Process: “Recognize the implications of information formats that contain static of dynamic information” (Knowledge Practice)
Question how algorithms shape your personal choices and perspectives on a given topic	Authority is Constructed and Contextual: “Develop awareness of the importance of assessing content with a skeptical stance with a self-awareness of their own biases and worldview” (Disposition)
Recognize that search engine revenue comes from advertising not information retrieval services	Information Has Value: “Understand how the commodification of their personal information and online interactions affects the information they receive” (Knowledge Practice)

Once we had agreed on the outcomes for the course and its modules we began to develop a shared curriculum using collaborative decision making strategies. All of the librarians are faculty peers and may have the opportunity to teach this course in the future. Our department does not have a department chair, and while we all have different responsibilities we do not have any formal supervision over each other. With that in mind, we wanted to design a course structure where we might all have the same core assignments and learning outcomes but still have the flexibility to adapt to our teaching styles and preferences as needed.

As part of the curriculum design process all librarians typed up ideas of activities, readings, assignments, and other course materials in a shared online document. I printed all of those ideas out on individual slips of paper and taped them to whiteboards in our meeting room. Each librarian was given some stickers to prioritize the activities on the board. We then aligned the prioritized activities with the learning outcomes and discussed gaps. In this way we were able to collaboratively design curriculum we all felt comfortable adapting in spite of the fact that course design is often an individual exercise on our campus.

The course's other main units focus on frameworks for evaluating other information sources, analyzing personal information seeking and use behaviors, privacy, and information privilege. Our course was proposed through our campus-wide curriculum approval process requiring approval from our entire library faculty, our Dean, and an academic senate curriculum committee. This process took nearly a year to complete and our course launches in January 2020, and we are excited to teach new students from all majors to appraise, interrogate, and analyze the roles algorithms play in structuring our information seeking and use.

Conclusion

Understanding and evaluating black box algorithms will be of continued importance in our information landscape. As librarians and educators we pride ourselves on teaching others to critically evaluate information, and algorithmic bias is a concept absent from our teaching resources and practices. Librarians should work to teach others about algorithmic bias and its harms. Char Booth notes in her work that it takes, "five times as much work to create something from scratch" than to modify an existing lesson (Booth, 2011, p.20). To that end some of these activities have been previously shared in an online repository for library instruction materials, Project CORA (Gardner, 2018) and I would love to see more lessons on algorithmic bias readily available to others. We hope that our attempts at teaching this content to undergraduate students in the United States can be adapted and modified for different educational experiences as well as inspire other librarians to develop curriculum in this area.

References

ACRL. (2016). Framework for information literacy for higher education. Retrieved from <http://www.ala.org/acrl/standards/ilframework>

Algorithmic Accountability Act of 2019, H.R. 2231, 116th Cong. (2019).

Angwin, J. L. (2016, May 23). Machine Bias. *ProPublica*. Retrieved from

© C. Caffrey Gardner

Address correspondence to Carolyn Caffrey Gardner, Information Literacy Coordinator, California State University, Dominguez Hills, 1000 E. Victoria Street, Carson, CA 90747. E-mail: cgardner@csudh.edu

- <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- Booth, C. (2011). *Reflective teaching, effective learning: Instructional literacy for library educators*. Chicago, IL: American Library Association.
- Brkan, M. (2019). Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond. *International Journal of Law and Information Technology*, 27(2), 91–121. <https://doi.org/10.1093/ijlit/eay017>
- Bryone, K., & Westbrook, T. (2019, July 26). Australia to “lift veil” on Facebook, Google algorithms to protect... *Reuters*. Retrieved from <https://www.reuters.com/article/us-australia-tech-regulator-idUSKCN1UL01P>
- Buolamwini, J. (2016). *How I’m fighting bias in algorithms*. Retrieved from https://www.ted.com/talks/joy_buolamwini_how_i_m_fighting_bias_in_algorithms
- City council passes first bill in nation to address transparency, bias in government use of algorithms. (2017, December 11). Retrieved from <https://www.nyclu.org/en/press-releases/city-council-passes-first-bill-nation-address-transparency-bias-government-use>
- Cutts, M. (2010). *How search works*. Google. Retrieved from <https://www.youtube.com/watch?v=BNHR6IQJGZs>
- Dawn, R. (2018, October 30). “Baby Shark” girl who stole our hearts with adorable Alexa request visits TODAY. Retrieved from <https://www.today.com/parents/baby-shark-girl-who-stole-our-hearts-adorable-alexa-request-t140886>
- Fisher, Z. (2017). Facing the frames: Using the Framework as a guide for a credit-bearing information literacy course. *College & Research Libraries News*, 78(7), 354–358. <https://doi.org/10.5860/crln.78.7.354>
- Gardner, C.C. (2018) Analyzing search engines: What narrative is told through the algorithm. Project CORA (*Community of Online Research Assignments*). Retrieved from <https://www.projectcora.org/assignment/analyzing-search-engines-what-narrative-told-through-algorithm>.
- Noble, S. (2018a). *Algorithms of oppression: How search engines reinforce racism*. New York: NYU Press.
- Noble, S. (2018b, March 26). Google has a striking history of bias against black girls. *Time*. Retrieved from <http://time.com/5209144/google-search-engine-algorithm-bias-racism/>
- Pasick, A. (2015, December 21). The magic that makes Spotify’s Discover Weekly playlists so damn good. Retrieved September 10, 2018, from: <https://qz.com/571007/the-magic-that-makes-spotifys-discover-weekly-playlists-so-damn-good/>

- Rainie, L., Janna, & erson. (2017, February 8). Code-Dependent: Pros and cons of the algorithm age. Retrieved from <http://www.pewinternet.org/2017/02/08/code-dependent-pros-and-cons-of-the-algorithm-age/>
- Reidsma, M. (2019). *Masked by Trust: Bias in Library Discovery*. Sacramento, CA: Library Juice Press.
- Shu, C. (2018). Why inclusion in the Google Arts & Culture selfie feature matters. *Tech Crunch*. Retrieved from <http://social.techcrunch.com/2018/01/21/why-inclusion-in-the-google-arts-culture-selfie-feature-matters/>
- Smith, A. (2018). *Public attitudes toward computer algorithms*. Retrieved from <https://www.pewinternet.org/2018/11/16/public-attitudes-toward-computer-algorithms/>
- Smith, C. L., & Matteson, M. L. (2018). Information literacy in the age of machines that learn: Desiderata for machines that teach. *Libri*, 68(2), 71–84. <https://doi.org/10.1515/libri-2017-0025>
- Stanford History Education Group. (2016). *Evaluating information: The cornerstone of civic online reasoning - Executive summary*. Retrieved from <https://stacks.stanford.edu/file/druid:fv751yt5934/SHEG%20Evaluating%20Information%20Online.pdf>
- Washington State Legislature. (2019). Senate Bill 5082. Retrieved July 8, 2019, from <https://app.leg.wa.gov/billsummary?BillNumber=5527&Year=2019&Initiative=false>