IJoL International Journal of Librarianship

ISSN: 2474-3542 Journal homepage: http://journal.calaijol.org

Managing Bias When Library Collections Become Data

Catherine Nicole Coleman

Abstract:

Developments in AI research have dramatically changed what we can do with data and how we can learn from data. At the same time, implementations of AI amplify the prejudices in data often framed as ‘data bias’ and ‘algorithmic bias.’ Libraries, tasked with deciding what is worth keeping, are inherently discriminatory and yet remain trusted sources of information. As libraries begin to systematically approach their collections as data, will they be able to adopt and adapt the AI-driven tools to traditional practices?

Drawing on the work of the AI initiative within Stanford Libraries, the Fantastic Futures conference on AI for libraries, archives, and museums, and recent scholarship on data bias and algorithmic bias, this article encourages libraries to engage critically with AI and help shape applications of the technology to reflect the ethos of libraries for the benefit of libraries themselves and the patrons they serve. A brief examination of two core concepts in machine learning, generalization and unstructured data, provides points of comparison to library practices in order to uncover the theoretical assumptions driving the different domains. The comparison also offers a point of entry for libraries to adopt machine learning methods on their own terms.

To cite this article:

Coleman, C.N. (2020). Managing bias when library collections become data. *International Journal of Librarianship*, *5*(1), 8-19. https://doi.org/10.23974/ijol.2020.vol5.1.162

INTERNATIONAL JOURNAL OF LIBRARIANSHIP, 5(1), 8-19.

ISSN:2474-3542

Managing Bias When Library Collections Become Data

Catherine Nicole Coleman

Stanford University, CA, USA

ABSTRACT

Developments in AI research have dramatically changed what we can do with data and how we can learn from data. At the same time, implementations of AI amplify the prejudices in data often framed as ‘data bias’ and ‘algorithmic bias.’ Libraries, tasked with deciding what is worth keeping, are inherently discriminatory and yet remain trusted sources of information. As libraries begin to systematically approach their collections as data, will they be able to adopt and adapt the AI-driven tools to traditional practices?

Drawing on the work of the AI initiative within Stanford Libraries, the Fantastic Futures conference on AI for libraries, archives, and museums, and recent scholarship on data bias and algorithmic bias, this article encourages libraries to engage critically with AI and help shape applications of the technology to reflect the ethos of libraries for the benefit of libraries themselves and the patrons they serve. A brief examination of two core concepts in machine learning, generalization and unstructured data, provides points of comparison to library practices in order to uncover the theoretical assumptions driving the different domains. The comparison also offers a point of entry for libraries to adopt machine learning methods on their own terms.

Keywords: Data Bias, Algorithmic Bias, Collections as Data, Accountability, Critical Data Practice, Artificial Intelligence

Bryan Catanzaro, VP of Applied Deep Learning at Nvidia, told the crowd attending a conference for libraries, archives, and museums (Fantastic Futures, 2019) that computer scientists need the collected and curated data that libraries have. Catanzaro, who leads a research team at Nvidia, explained that developments in deep learning, where once massive amounts of data were necessary, are moving in the direction of more clearly defined and domain specific data sets. He was bringing this message to a library conference because data is now recognized to be a fundamental part of the algorithm. “That means,” said Catanzaro, “that the collection and curation of data sets, the skills that you guys practice—the resources that you have access to—can enable the creation of new algorithms and new applications.” (Catanzaro, 2019, 5:30)

For most machine learning researchers, acquiring the training data with which to build models is challenging for a number of reasons including copyright restrictions, privacy concerns, and pay walls. As a result, computer scientists have tended to use whatever data are easily accessible. The model that Catanzaro’s team and many others use for building text generators was trained on Wikipedia and unpublished books. (Devlin et al, 2018; Zhu et al, 2015) The limitations and potential problems of using data available 'in the wild' without attention to curation presents problems. A cautionary tale that Catanzaro shared was IBM's facial recognition program. The 2018 paper, *Gender shades: Intersectional accuracy disparities in commercial gender classification* by Joy Buolamwini and Timnit Gebru, uncovered significant demographic bias in benchmark datasets used for face recognition by IBM, Google, Microsoft and Face++. In an attempt to make their model "more fair and accurate," (Merler et al, 2019) IBM created the Diversity in Faces dataset from a subset of the YFCC100M dataset: 99.2 million photos and 0.8 million videos from Flickr that account holders had tagged with Creative Commons licenses. (Thomee et al, 2016) The owners of the Flickr images sued, insisting that giving license for re-use of the image did not include re-use for the purpose of building facial recognition technology. (Rizzi, 2020) The problems that IBM encountered and created were, in large part, failures of critical data practice that could have been avoided by following *Ten simple rules for responsible big data research* or similar guidelines. (Zook et al, 2019)

What if researchers had come to the library asking for images to train a face recognition model? Catanzaro spoke of 'collecting data' as scraping text or images from the web. Libraries instead think in terms of Collections as Data. (Padilla et al, 2019) Starting a search for images at a library or archive begins with collections. Image collections sourced from libraries, archives, and museums are already described and bounded by the circumstances and terms of collection. Provenance would have been integral to the selection of images. 'Which data sets?' and 'Why?' are questions that would have come first. The license for the data selected would also have been an essential criterion for selection. Though in this case the terms of the license were subject to dispute, at least the lesson learned after the fact would redound to the benefit of future data selection by informing library and archival practice. As Christine Borgman explains in *Big Data, Little Data, No Data*, data is difficult to define but always has context. Libraries provide knowledge infrastructure that is adaptable to the questions being asked, and the collection is a fundamental component of organizing information. (Borgman, 2015, p 173)

Would library staff have identified demographic bias in the data sets before the *Gender Shades* study was published? Likely not. Developing that awareness requires more than critical library skills and subject specialization. Diversity in staffing as well as the skills and tools to analyze the content of digitized collections rather than relying on metadata alone would be necessary. The recently published report, *Responsible Operations: Data Science, Machine Learning, and AI in Libraries* is an example of the self-reflexive work that goes on within libraries to address social and organizational challenges hand-in-hand with technical ones. In the report Thomas Padilla writes, referring to the individuals in libraries, digital humanities, and affiliated fields he consulted: "All agreed that the challenge of doing this work responsibly requires fostering organizational capacities for critical engagement, managing bias, and mitigating potential harm." (2019, p. 6)

This paper expands on the paradigm of managing bias in libraries that was introduced in the *Responsible Operations* report*:* "Managing bias rather than working to eliminate bias is a distinction born of the sense that elimination is not possible because elimination would be a kind of bias itself—essentially a well-meaning, if ultimately futile, ouroboros." (Padilla, 2019) To manage bias is to emphasize the active engagement and vigilance required to balance the inherently discriminatory ordering of information for library retrieval systems with the responsibility libraries have to serve people and help them find relevant information. The library is not neutral and does not have the right models. Rather, through the quotidian activities of catalogers, bibliographers, archivists, and by other library staff at the reference desk, in the acquisitions department and in circulation, the quiet struggle to adapt and respond to the changing information landscape, to decide what should be collected and preserved, and to help people find what they are seeking, plays out. The day-to-day work of libraries is local and human scale.

Libraries, archives, and museums are not just about storing data, organizing it, and providing access. They are vital institutions full of committed individuals whose work lies in the tension between the inherently discriminatory mediating practice of organizing and categorizing and the desire to make information freely available and discoverable. That tension, or friction, provides stability and drives change. Cultural heritage institutions are perpetually confronting the questions: ‘Are we preserving the right things? Are we making the right choices?’ There is no right answer. The bases of decision-making change over time and are distributed. Each institution has its own character; the forces acting on decision-making are many. Most importantly, there are human beings behind the decisions and the institutional norms who are accountable. Attempts to de-bias algorithms or de-bias data have been introduced recently in response to a crisis in machine learning. But seeking to avoid accountability, disguised as objectivity or worse, neutrality, is a technocratic fallacy. Bias is an unavoidable consequence of situated decision-making that we have to reckon with. Who decides how to classify things? Who decides which things are in the system and which are not? These questions are not new and they are integral to the work of libraries.

FROM GENERALIZABILITY TO ACCOUNTABILITY

A machine learning model is not useful if it is not generalizable. The example that Andrew Ing uses often to explain this concept involves building a model to predict the potential sale price of a home. (Ing, n.d.) The exercise begins with data about houses that have already been sold. The data must include the sale price of the home and the values for some number of corresponding features of each house, for example, the number of bedrooms, the number of bathrooms, and the square footage. If too many features or features that are too specific to a home are included, the model cannot be applied broadly. In other words, the variety of features used to train the model is limited by design.

This limitation is known as the ‘bias-variance tradeoff.’ When your model includes fewer features and those features are common to all homes (i.e. all homes have some measure of bedrooms, all can be measured in square feet), the model is more broadly applicable. Conversely, the more unique the qualities, the greater ‘variance’ in the model, making it less useful for prediction. Mitigating prediction error is also described in terms of overfitting or underfitting. Overfitting results when the model is very good at identifying distinguishing features in your training data, but when you try to apply the model to new data, it fails because the features are too specific to the training data. Underfitting describes the situation when your model is so general that it does not make useful distinctions in the data.

The implications of the bias-variance dilemma become clearer when it is framed in terms of equity vs. efficiency. Fast-growing internet business models tend toward efficiency, which is why we see so much bias in applications of AI emerging from the private sector. As long as there is no accountability for businesses to be equitable (even knowing what it means to be equitable is considered too expensive a question to answer), efficiency and overall increased performance will win out. In the abstract, the bias-variance balance is simply a matter of optimizing a model to produce the best result. But that aseptic presentation belies all of the consequential choices left unexamined in the process.

When Andrew Ing explains how machine learning works, he is very careful to point out that machine learning is limited to discrete tasks. And yet, he perpetuates the mythology that AI is the new electricity, a new currency; the basis of a new economy. The choice of which features to use and the consequences of those choices are not discussed in his training. Are the size of the house, the number of rooms and number of bathrooms selected because they are the best measures of value or simply because they are readily available and do not require additional research? Or are those features defining the model because the data for those features are the most complete? The age of the roof or the strength of the home's foundation might be better indicators of the value of the house to the homebuyer. But if there is spotty or inconsistent data gathered about those features, they may be ignored. It is easy to imagine a construction market being driven by the model: homes are built to maximize value according to those measures of size rather than that focusing on energy efficiency, affordability, or any number of other qualities that would benefit society.

The discriminatory effects of applying mathematical models either uncritically or without accountability has been documented incisively by Cathy O'Neil in her book *Weapons of Math Destruction* (2016), Meredith Broussard in *Artificial Unintelligence: How Computers Misunderstand the World* (2018), Safiya Noble in *Algorithms of Oppression* (2018), and Ruha Benjamin in *Race After Technology* (2019). Noble brings home to libraries the social implications of bias embedded in algorithms, what she terms *technological redlining*. Noble's focus is on not only Google Search but how search engines and knowledge discovery systems in the library reinforce racism. The discriminatory effects of the systems, Noble points out, is in not only how they limit our access to information but also how the results returned from queries shape attitudes. The promise of 'automation' and the focus on the 'most popular' are implementations of models that are generalizable for convenience while disguising the underlying decision-making and accountability.

AI is offered as a technocratic solution that perpetuates the recursively false notion that machine-generated decisions are better, or less biased, than human decisions. When that idea is put into practice, human discretion is given over to generalized prediction. Generalized prediction minimizes variation, privileging what is deemed most common or statistically significant. And, once again, what is registered as statistically significant is under question. (Armheim et al, 2019) Statistically significant measures are used to uphold an artificial 'center' to justify whatever exclusion is convenient for the sake of efficiency. Benjamin has captured this in the context of AI as the "New Jim Code." She refers to this as "the allure of objectivity without public accountability." (Benjamin, 2019, p. 53) This inclination of AI contradicts the strengths of a research library as a holder of unique objects and a place where collections are developed to serve the changing interests of a research community.

Libraries do need to find efficiencies to keep up with the deluge of increasing information and to continue the work of digitizing the past. Machine learning, computer vision, natural language processing, and related technologies can be extremely helpful in this effort. Can libraries employ the powerful predictive models without perpetuating and further entrenching bias? Can they do more? Might they also use the tools critically to examine bias in existing collections? This entails shifting thinking away from the ‘technology as solution’ mindset that is implicit in the way AI is taught and promoted, to thinking first of the usefulness of AI to the work of libraries, to the benefit of society and the production of new knowledge. The algorithm only sees what it has been programmed to see, which is to say, the power is not in the technology, but rather it is in the people who employ it. By instrumentalizing the technology and putting it in the hands of librarians, the technology can be truly useful in libraries.

The collaboration between the Frank-Ratchye STUDIO for Creative Inquiry and the Carnegie Museum of Art is an instructive example of how cultural heritage image collections can benefit from machine learning and computer vision to go beyond metadata and into the images themselves. The team classified 60,000 images from the Teenie Harris Photography Archive using a convolutional neural network trained on the ImageNet benchmark data set, resulting in multiple labels for each image, each with a percentage confidence measure. (Howard, 2017) They then clustered the images based on a calculation of similarity in the proposed labels. The results were groupings of images within the 60,000 that the archivist would not have otherwise discovered like women in fur coats, brides, and car crashes. (Dominique Luster, archivist; Golan Levin, Frank-Ratchye STUDIO for Creative Inquiry; and Caroline Record, Innovation Studio; personal communication, August 2019.) This opens exciting new possibilities for metadata creation and object discovery. (Wevers & Smits, 2019)

Let us approach this another way. A data-centered, rather than technology-centered look at those experiments could position the Teenie Harris Photography Archive as a benchmark against which to test the biases in the ImageNet training data set. One of Harris’s photographs, of basketball players posing after a game, was classified with the confidence of 38% volleyball, 36% bikini, and 4.6% basketball based on labeled data from ImageNet. It may not be possible to know with certainty why the confidence score of the label “volleyball” was so much higher than “basketball” but knowing that the archive is a collection of the work of the photographer for the Pittsburgh Courier working from 1935 to 1975 described as “one of the most detailed and intimate records of the black urban experience known today” gives context for a critical reading of the algorithm’s results that is not available from ImageNet. (In the ‘70s basketball players still wore knee pads. Knee pads are uncommon in basketball today but are still common in volleyball.)

Datasets that are currently considered gold standards, like ImageNet and BooksCorpus (which Catanzaro’s team used), are in need of their own benchmarks to challenge the context and provenance of both the images and the labels. In “Excavating AI” Kate Crawford and Trevor Paglen give particular attention to ImageNet in their dive deep into the many-layered problematic assumptions and politics behind the taxonomies, classes, and individual labels that make up image training sets. (2019) They draw parallels to the politically fraught Library of Congress Subject Headings— the controlled vocabulary used for indexing, cataloging, and searching for bibliographic records in library catalogs and electronic databases. But classification is both necessary and inherently problematic. (Bowker & Star 2000) This paradox is not only known to libraries, it is managed within their structure. Cataloging work, which is standards based, is balanced with collection development and the selection of materials, both of which play a more decentralized role in overall information management. And it is those who operate within libraries and information science who both call out and advocate for change. (Berman, 1973; Gross, 2017, Noble, 2019)

UNSTRUCTURED DATA AND ITS CONTEXT

How do we go from collections to data? Unstructured data, for the machine learning community, are data from sources without a defined structure or ‘schema.’ Digitized texts, images, audio, and video, for example, are taken to be unstructured at the level of the pixels of the image, the signal in the audio, or the sentences in the text. The research performed with these sources is oriented to extracting the linguistic, auditory, or visual structure inherent in the contained information in the form of patterns. Here it is helpful to consider Borgman’s distinction between source and resource in defining data. (Borgman, 2007, pp. 121-122) In machine learning research into texts, the focus is not in the individual sources but in a general understanding of the semantic and syntactic context of words. But is that distillation understood as a resource —a derivative of the source—or is the distinction blurred? The argument for building training corpora on undifferentiated texts is explained in 2010 by Google researchers in the article “The Unreasonable Effectiveness of Data.” The authors refer to the Brown Corpus, a corpus of 500 samples of English-language text, totaling more than one million words, compiled from works published in 1961 (Francis & Kucera, 1979):

> “In some ways this corpus is a step backwards from the Brown Corpus: it’s taken from unfiltered Web pages and thus contains incomplete sentences, spelling errors, grammatical errors, and all sorts of other errors. It’s not annotated with carefully hand-corrected part-of-speech tags. But the fact that it’s a million times larger than the Brown Corpus outweighs these drawbacks.”

The needs for machine learning research have changed from 2010 to today. Researchers like Catanzaro are seeking domain specific data sets to fine-tune models and achieve better results. And yet the underlying methods of data collection and corpus collection still exhibit the assumptions of Big Data that size trumps quality and any context other than linguistic continuity is irrelevant.

The distinction between structured and unstructured has as much to do with how researchers choose to use the data and their willingness to uncover the structure as it does with any intrinsic properties. From a library perspective, books, newspaper articles, and photographs are structured by their form, the intended audience and a number of other qualities that provide context. They are created, situated in place and time, and collected or maintained for some reason. And yet all of that situatedness, if it is not readily available to computation, is ignored when patterns and structure within the data set is sought through machine learning methods. Context is difficult to trace when digitized sources are aggregated, processed, and recombined into massive datasets.

What would it mean for libraries to support computational use of their collections? Librarians speaking at the Fantastic Futures conference who have already embraced the possibilities of AI-related technologies expressed frustration that, despite the considerable effort that has gone into digitizing collections, libraries still deliver books, periodicals, photographs, and precious manuscripts as individual objects when new interfaces oriented to AI-generated metadata afford opportunities to explore them at scale. And, moreover, researchers increasingly want to access them this way. Medieval manuscripts, 20th century photography collections, audio recordings, coins, web archives, books, and government documents are all data available to machine learning once they are digitized. Anything that can take the form of digital text, image, or audio signal can be represented numerically and computed. This reality is a paradigm shift that libraries have not yet come to terms with. The digitized photograph, for example, is no longer *only* a digital object that can be duplicated, distributed, layered, annotated, and compared one to another. Each individual image is *also* a data set at the pixel level: a source of information that can be machine read. And when an entire collection is analyzed at the pixel level, each gains new context in relationship to other images. This complicates our understanding of the digital object from a library perspective, but it also opens up new possibilities for managing those objects and making them discoverable. (Wevers & Smits, 2020; Arnold & Tilton, 2020)

As the Collections as Data project attests, the libraries that support computational access to their collections are still few. (Padilla et al, 2019) The shift in thinking of digital materials as data sets rather than as discrete objects is not only about changing how we support users. It is not enough to provide APIs to researchers who want access to collections amenable to computation. Ben Schmidt, who works with the HathiTrust digital library, a massive aggregation of texts, has argued that simply providing access to OCR'd text puts too great a burden on the researcher to manage that data, navigate it, and find what is relevant. He suggests, instead, dimensionality reduction; a means to algorithmically cluster texts based on their similarity. (Schmidt, 2018) Meaningful clusters help researchers filter, navigate and limit downloads to only the content they need. Schmidt writes,

> "Treating dimensionality reduction as infrastructure means thinking of digital representations of books not just as 'machine-readable' texts, but as 'machine-read' texts: data that has already been partially digested by an algorithm. The choices we make for what this machine reading looks like shape the universe of possible research."

The possibilities of Schmidt's recommendation go beyond pre-processing texts for access. His proposal for libraries to machine read collections as data opens the opportunity for librarians to run analysis for their own purposes and to better understand collections at the level of content and therein provide support to researchers. A more local, collection-level approach to modeling data is consistent with how traditional libraries provide services and just makes sense.

Barbara McGillivray, who spoke at the Fantastic Futures 2018 conference in Oslo, Norway, like Schmidt, is a digital humanist. The digital humanities have a significant overlap with both research in machine learning that uses sources oriented to human communication and data analysis in the library. McGillivray's work, which emphasizes the importance of applying algorithms at a human scale, is particularly appropriate to the work of subject experts who can evaluate the results of statistical analysis based on their familiarity with the data. (McGillivray, 2018) McGillivray found, for example, that ancient languages are not amenable to the standard word embedding models. She pointed out that the type of text determines the meaning of a word, so it is necessary to incorporate expert knowledge of the corpus iteratively into the model as you are tuning it. She shared examples of false positives where what appears, in the abstract, as semantic change, is just a shift in meaning either over time or given the context. While this degree of close reading is the work of the researcher, the curator also needs to be able to read the features in a collection.

Rebecca Wingfield, curator for American and British Literature at Stanford Libraries, brought a similar problem to the Stanford Libraries AI Studio. Wingfield was interested in analyzing a collection of texts, the Single Volume Novels Collection of 1,674 19th century titles. She was not in pursuit of a particular research question but sought to understand the anatomy of the collection–to see its shape and make it navigable–so she could share it with researchers. The collection was acquired in response to the Stanford English Department's interest in the history and theory of the novel. What makes the collection unique is that the titles are under-collected, scarcely held novels–'the great unread' in English literature. That also means that there is minimal cataloguing information about the content. In Wingfield's words, "they are being served to patrons as a mass of undifferentiated texts." (personal communication, August 2018) To make these novels more easily discoverable to scholars, Wingfield wanted richer facet data: dates that help determine the period in which the novel is set, place names to determine where it is set, and indicators of genre or topics that might provoke new avenues of research. Even if the extracted data does not make it into facets, this is a case where a range of techniques, from text similarity, to named entity extraction and topic modeling would help the librarian become more familiar with the collection and the relationship between the texts without having to read all of them. Librarians have methods for developing and managing collections; the challenge is to adapt those methods to new tools for reading 'machine read' collections. AI-assisted collection analysis empowers subject specialists whose domain knowledge makes them critical partners to researchers.

Technology-first methods to address data bias and algorithmic bias seem to be completely unaware of the work that goes on in libraries, archives, and museums to make cultural heritage available for research. Well-meaning projects like Datasheets for Datasets and Nutrition Labels for Data do not acknowledge the long tradition of information management and access that has addressed data bias and algorithmic bias long before machine learning emerged as a field of study. (Gebru et al, 2018; Holland et al, 2018) The suggestion that, because data collection remains overlooked within machine learning, the field of machine learning needs to create its own brute force methods for data management, including a sub-field within machine learning, is misguided. (Jo & Gebru, 2020; Jordan, 2018) Collaboration with libraries would be more fruitful. Critical data studies and the reflective practice of the digital humanities provides practical guidelines for libraries to begin adopting and adapting AI to collections. Yanni Alexander Loukissas's *local reading*, which gives attention to the provenance of data and how the local conditions of their creation shape research and practice, is a helpful point of reference. (Loukissas, 2017) Local reading provides a theoretical framework for a new set of tools to help curators adapt their methods to 21st century digital data curation.

CONCLUSION

Peter Norvig, co-author of the "The Unreasonable Effectiveness of Data" article quoted above, was famously misquoted by Wired author Chris Anderson in the 2008 article "The End of Theory:

The Data Deluge Makes the Scientific Method Obsolete" as supporting the notion that with enough data, theory is irrelevant. In his rebuttal, Norvig explained that "theory has not ended, it is expanding into new forms." (Norvig, n.d.) Figuring out those new forms, according to Norvig, requires the observational and experimental approach of a natural scientist. "Having more data, and more ways to process it," writes Norvig, "means that we can develop different kinds of theories and models." During his visit to Stanford Libraries, Norvig pointed out that libraries are in a privileged position to develop theories and models because we have content experts and subject specialists. (Peter Norvig, personal communication, November 2018.)

If libraries simply hand over collections as data to researchers, it would be a disservice to both libraries and the patrons. Librarians need to master the instruments of AI and employ them both to learn more about their own resources—to see and analyze them in new ways—and to help shape applications of AI with the expertise and ethos of libraries. At this moment when there are as many papers about the successes of AI research as there are papers calling out algorithmic bias, data bias, and setting forth principles of AI practice, libraries need to do much more than provide curated data to AI researchers. Libraries need to apply the principles of the profession to managing bias in AI-based systems.

AI practices are not ends in themselves. They are problem-solving techniques that need to be applied within a societal context. This might seem obvious, but there is a strong tendency to look to technology, particularly new technology, for solutions and easy answers. When we reify the technology, we lose the opportunity to use it for our own ends and effectively give over our responsibility to it. Libraries need what AI has to offer, but AI needs what librarians have to offer even more.

References

Amrhein, V., Greenland, S., & McShane, B. (2019, March 20). Scientists rise up against statistical significance. Retrieved March 30, 2019, from https://www.nature.com/articles/d41586-019-00857-9

Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired magazine*, *16*(7), 16-07.

Arnold, T., & Tilton, L. (2020). Distant Viewing Toolkit: A Python Package for the Analysis of Visual Culture. *Journal of Open Source Software*, *5*(45), 1800.

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *Calif. L. Rev.*, *104*, 671.

Benjamin, R. (2019). *Race after Technology: Abolitionist Tools for the New Jim Code*. Polity.

Berman, S. (1993). *Prejudices and antipathies: A tract on the LC subject heads concerning people*. McFarland & Company Incorporated Pub.

Berman, S., & Gross, T. (2017). Expand, Humanize, Simplify: An Interview with Sandy Berman. *Cataloging & Classification Quarterly*, *55*(6), 347-360. DOI: 10.1080/01639374.2017.1327468

Bermès, E. (2019, December 4). "The Corpus project at the French National Library." [Video file]. Retrieved from https://library.stanford.edu/projects/fantastic-futures

Bolukbasi, T., Chang, K. W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? debiasing word embeddings. In *Advances in neural information processing systems* (pp. 4349-4357).

Borgman, C. (2015). Big data, little data, no data: Scholarship in the networked world. MIT Press.

Borgman, C. (2007). Scholarship in the digital age: information, infrastructure, and the Internet. MIT Press.

Bowker, G. C., & Star, S. L. (2000). Sorting things out: Classification and its consequences. MIT Press.

Broussard, M. (2018). *Artificial unintelligence: How computers misunderstand the world*. MIT Press.

Buolamwini, J. & Gebru, T.. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency, in PMLR* 81:77-91

Catanzaro, B. (2019, December 4). "Datasets make algorithms: how creating, curating, and distributing data creates modern AI." [Video file]. Retrieved from https://library.stanford.edu/projects/fantastic-futures

Devlin, J., et al. (2018) "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805*.

Francis, W. N., & Kucera, H. (1979, July). BROWN CORPUS MAUNAL. Retrieved July 11, 2020, from http://korpus.uib.no/icame/manuals/BROWN/INDEX.HTM

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2018). Datasheets for datasets. *arXiv preprint arXiv:1803.09010*.

Gross, T. (2017). Examining the Subject Heading" Illegal aliens". Paper at CaMMS Forum: Working Within and Going Beyond: Approaches to Problematic Terminology or Gaps in Established Vocabularies. American Library Association Midwinter, Atlanta, GA.

Halevy, A., Norvig, P., & Pereira, F. (2009). The unreasonable effectiveness of data. *IEEE Intelligent Systems*, *24*(2), 8-12.

Hickerson, T., & Brosz, J. (2017). Remaining Relevant: Critical Roles for Libraries in the Research Enterprise.

Holland, S., Hosny, A., Newman, S., Joseph, J., & Chmielinski, K. (2018). The dataset nutrition label: A framework to drive higher data quality standards. *arXiv preprint arXiv:1805.03677*.

Howard, Z. (2017, April 8). Finding Patterns in the Content of Teenie Harris's Photos (with Convolutional Neural Networks and Agglomerative Clustering). Retrieved July 11, 2020, from https://zariahoward.github.io/TeenieHarris/ObjectDetection.html

Ing, A. (n.d.). AI For Everyone. Retrieved from https://www.deeplearning.ai/ai-for-everyone/

Jo, E. S., & Gebru, T. (2020, January). Lessons from archives: strategies for collecting sociocultural data in machine learning. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 306-316).

Jordan, M. I. (2018, April 30). Artificial Intelligence - The Revolution Hasn't Happened Yet. Retrieved from https://medium.com/@mijordan3/artificial-intelligence-the-revolution-hasnt-happened-yet-5e1d5812e1e7

Larson, E. (2020). Big Questions: Digital Preservation of Big Data in Government. *The American Archivist*, *83*(1), 5-20.

Leonard, P. (2019, December 4). "Yale DH Lab’s Pix Plot." [Video file]. Retrieved from https://library.stanford.edu/projects/fantastic-futures

Loukissas, Y. (2017) Taking Big Data apart: local readings of composite media collections. *Information, Communication & Society* 20, no. 5. pp 651-664. https://doi-org.stanford.idm.oclc.org/10.1080/1369118X.2016.1211722

McGillivray, B. (2018, December 05). Fantastic Futures. AI-conference. Retrieved July 11, 2020, from https://www.nb.no/hva-skjer/ai-conference/

Merler, M., Ratha, N., Feris, R. S., & Smith, J. R. (2019). Diversity in faces. *arXiv preprint arXiv:1901.10436*.

Mittelstadt, Brent, Chris Russell, and Sandra Wachter. "Explaining explanations in AI." In *Proceedings of the conference on fairness, accountability, and transparency*, pp. 279-288. 2019. DOI:https://doi-org.stanford.idm.oclc.org/10.1145/3287560.3287574

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.

Norvig, P. (n.d.). All we want are the facts, ma'am. Retrieved from https://norvig.com/fact-check.html

O'Donovan, M., Richardson, Z., Powell, S., & Moriarty, A. (2018). Open by default?: Images of Maori and Moana pacific subjects at Auckland war memorial museum Tamaki Paenga Hira, New Zealand. *Journal of the Australasian Registrars Committee*, (74), 44.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books.

Padilla, T., Allen, L., Frost, H., Potvin, S., Russey Roke, E., & Varner, S. (2019, May 22). Final Report --- Always Already Computational: Collections as Data (Version 1). Zenodo. http://doi.org/10.5281/zenodo.3152935

Padilla, T. (2019). *Responsible Operations: Data Science, Machine Learning, and AI in Libraries*. OCLC Research Position Paper. https://doi.org/10.25333/xk7z-9g97.

Rizzi, C. (2020, January 30). Class Action Accuses IBM of 'Flagrant Violations' of Illinois Biometric Privacy Law to Develop Facial Recognition Tech. Retrieved July 10, 2020, from https://www.classaction.org/news/class-action-accuses-ibm-of-flagrant-violations-of-illinois-biometric-privacy-law-to-develop-facial-recognition-tech

Schmidt, B. (2018) Stable random projection: lightweight, general-purpose dimensionality reduction for digitized libraries. *Journal of Cultural Analytics*. DOI:10.22148/16.025

Selbst, A. D.; Barocas, S. (2018). The intuitive appeal of explainable machines. Fordham Law Review, 87(3), 1085-1140.

Thomas, P. S., da Silva, B. C., Barto, A. G., Giguere, S., Brun, Y., & Brunskill, E. (2019). Preventing undesirable behavior of intelligent machines. *Science, 366*(6468), 999-1004.

Thomee, B., Shamma, D. A., Friedland, G., Elizalde, B., Ni, K., Poland, D., ... & Li, L. J. (2016). YFCC100M: The new data in multimedia research. *Communications of the ACM*, *59*(2), 64-73. https://doi.org/10.1145/2812802

Wevers, M. & Smits, T. (2020) The visual digital turn: Using neural networks to study historical images. *Digital Scholarship in the Humanities*, Volume 35, Issue 1, pp. 194–207. https://doi.org/10.1093/llc/fqy085

Zhu, Y., Kiros, R., Zemel, R., Salakhutdinov, R., Urtasun, R., Torralba, A., & Fidler, S. (2015). Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. In *Proceedings of the IEEE international conference on computer vision* (pp. 19-27).

Zook, M., Barocas, S., Boyd, D., Crawford, K., Keller, E., Gangadharan, S. P., ... & Narayanan, A. (2017). Ten simple rules for responsible big data research. https://doi.org/10.1371/journal.pcbi.1005399

About the Author

Catherine Nicole Coleman is Digital Research Architect for the Stanford University Libraries and Research Director for Humanities+Design, a research lab at the Center for Spatial and Textual Analysis at Stanford University.