

Katrina Ingram

Constructing AI: Examining how AI is shaped by data, models and people

Abstract:

Artificial Intelligence (AI) is a technology that is quickly becoming part of our digital infrastructure and woven into aspects of daily life. AI has the potential to impact society in many positive ways. However, there are numerous examples of AI systems that are operating in ways that are harmful, unjust and discriminatory. AI systems are constructs of the choices made in their design. They exist within a socio-cultural context that reflects the data used in their training, the design of their mathematical models and the values of their creators. If we want to build AI systems that benefit society, we need to change how we construct AI.

Keywords: Artificial Intelligence, Consent, Collection, Diversity, Privacy, Surveillance Capitalism,

Outline:

1. Introduction	2
2. The Devil is in the Data	2
2.1. Incomplete, Siloed and Biased	2
2.2. Surveillance Capitalism	3
2.3. Consent	3
2.4. Privacy	3
2.5. Collection and Storage	3
3. Making Mathematical Models	4
3.1. Values Inform Models	4
3.2. The Unexplainable Black Box	4
3.3. Designing Affordances	4
4. Who Makes AI?	5
4.1. Diversity	5
4.2. Funding	5
4.3. Purpose	6
5. Conclusion	6
6. References	7

Author(s):

Katrina Ingram, MA Communications and Technology, reganing@ualberta.ca

1. Introduction

Our world is shaped by the people who build our infrastructure. We understand this relationship clearly when we think about physical infrastructure. Urban planners, engineers and architects construct our transportation corridors and skylines, guiding the daily interactions of people within a place. Our digital infrastructure shapes our societal interaction in similar ways. Increasingly, the digital realm itself is informed by artificial intelligence. AI can be defined as machines that think and learn in ways that perform on par with, or better than, humans. As research advances and more AI technology is deployed, from autonomous vehicles, to the Internet of Things, to a wide range of AI enabled algorithms and robots, it will have an almost ubiquitous reach. Ample evidence suggests that “values are instantiated in sociotechnical systems” (Shilton, 2015, p. 2). AI systems are not an exception. They exist within a socio-cultural context that reflects the data used in their training, the design of their mathematical models and the values of their creators. AI systems are comprised of data, algorithmic models and computing power. Each of these elements are constructed by people and exist within a greater socio-cultural sphere. By examining these elements in more detail, we can start to see how ethical considerations intersect with the construction of AI systems.

2. The Devil is in the Data

Big datasets are necessary inputs for training AI systems. They are also one of the key ingredients that can lead to bias, discrimination and unequal outcomes. To understand data, we need to consider it within a bigger context. Data are abstracted elements or representations that seek to categorize and measure phenomena (Kitchin, 2017). While data is often seen as conveying the facts, data is not purely objective, for “Counting is political”. How data is collected, stored and constructed, for whom and for what purposes, reflects the values of those compiling the data (Lohr, 2015, p. 91). Data are part of a “complex socio-technical system” that forms a “data assemblage” which includes “ideas, techniques, technologies, systems, people and contexts” that “evolve and mutate over time” (Kitchin, 2017, p. 22). All of this exists independent of AI; however, AI can amplify and codify these data agendas in ways that have profound societal implications.

2.1. Incomplete, Siloed and Biased

Datasets are often incomplete, stuck in silos and contain historical biases. Examples of AI systems that discriminate against a particular group often have their roots in the use of data that contain a bias. For example, Amazon’s hiring algorithm, that discriminates against women for certain roles, was trained on datasets that privileged men (Dastin, 2018). Convenience of access to data creates a more subtle bias. Researchers gravitate towards datasets that they can more easily access, especially if the data is in the public domain and requires little if any clearance or permission to use. For example, Twitter data tends to get used frequently (Ahmed, Bath & Demartini, 2017). This also raises questions about information shared for one purpose but used in another context (Ahmed et al, 2017). In addition, machine readable data is preferred as it is in a workable format and easily aggregated as opposed to non-digital data or data that is stuck in siloed systems (Kitchin, 2017). Finally, incomplete data sets can result in unequal outcomes in either being used, resulting in error, or not being applied, resulting in lack of inclusivity (Kitchin, 2017). It is important to look at data and where it comes from in a holistic way to understand different types of data bias.

2.2. Surveillance Capitalism

The relative ease of gathering data, the low cost of storage, and AI's ability to process vast datasets have enabled a new business model, surveillance capitalism, that predicts behaviour by capturing and claiming human experience (Zuboff, 2019). According to Zuboff, "There was a time when you searched Google, but now Google searches you" (2019, p. 262). Products and services are designed in ways that normalize and reward data collection and sharing (Zuboff, 2019). We are living through the results of this experiment and its impacts on how we work, shop, vote, stay informed, relate to each other and to ourselves – essentially, how we live our lives. The deployment of more AI systems, which are data hungry, will further intensify and drive the demand for extractive practices while continuing to concentrate power in the hands of a few large players (Zuboff, 2019). What choices do we want to make as we enable an AI empowered future? Some might say it is too late, that we live in a post-privacy era and that we have freely consented to these conditions of surveillance in exchange for convenience.

2.3. Consent

The concepts of privacy and consent underpin how our society functions. Privacy is a value that has evolved over time and has both legal and moral grounds. Generally speaking, people have an expectation of privacy that allows them to "restrict access" to certain information and to have that decision respected (Moor, 1990). Even if there is "nothing to hide", privacy is an important value for both individuals and societal cohesion (Solove, 2013). Consent is the mechanism by which we grant access to information and the gold standard is informed consent, a voluntary agreement between parties with equal bargaining power (Richards & Hartzog, 2019). Our laws and norms are based on this foundational understanding of privacy and consent, but the digital sphere is challenging this understanding. Online terms of use policies "often serve more as liability disclaimers" than real protections of informed consent (Kitchin, 2017, p. 8). Thus, while we may legally "consent", we actually have no real choice (Richards & Hartzog, 2019). Online consent raises particularly challenging ethical concerns as big data sets gathered from platforms like Facebook are increasingly informing research in sensitive areas such as mental health (Gomes, Pawson, Muriello et al., 2018).

2.4. Privacy

Our data, including what we share publicly or within a selected online group, as well as the clicks, swipes, geo-location information and other "meta data" becomes the cost to participate in modern society. There is a blurring of public and private space that makes navigating the consent relationship more challenging (Ahmed et al., 2017). The increased use of biometric data, fueled by facial recognition, has led to calls to reconsider current privacy standards and stronger legal oversight (Kugler, 2018). Facial recognition technology used in law enforcement has recently become a hotly contested issue, leading some companies to bow out of this sector or limit its use for this purpose (Bajarin, 2020), while other sectors such as aviation are advancing new uses of facial recognition technology as a screening mechanism (Burt, 2020). Even if data is anonymized, many studies show how datasets can be aggregated to reidentify people (Heffetz & Ligett, 2013; Zang, Dummit, Lisker, & Sweeney, 2015; Henle, Matthews & Harel, 2019). Some researchers are working on technical solutions to solve this problem such as "k-anonymity", "differential privacy" and the use of synthetic data (Henle et al., 2019) or blockchain security protocols (Smith, 2019). There is increasing pressure to address these issues as the ethical choices of companies and governments involving invasive data collection practices have come under increasing public scrutiny in a post-Snowden era.

2.5. Collection and Storage

Data collection and storage practices raise ethical concerns that are part of the overall impact of AI. Online gig economy workers are often paid well below minimum wage to label data sets used to train AI (Semeuls, 2019). Research has also revealed the high carbon footprint created by running computations on massive sets of data. One study found that training a single AI model emitted the carbon equivalent of five cars over their lifetime (Strubell, Ganesh, McCallum, 2019). Anatomyof.ai is a visualization that maps an AI-ecosystem's full impact from the minerals mined as inputs for technology to the disposal of the e-waste at the end of its lifecycle (Anatomyof.ai). These externalities raise important ethical issues to consider as part of the total cost of producing and using AI. These are some of the reasons why data is a crucial element to consider in developing ethically aligned AI. However, data is not the only area that can present ethical challenges for AI.

3. Making Mathematical Models

It might seem surprising that mathematical models are not purely objective. In *Weapons of Math Destruction*, Cathy O'Neil makes the case that mathematical models are far from neutral, but rather "opinions embedded in mathematics" (O'Neil, 2017, p. 21). The choices made in designing AI models are another way that AI is socially constructed.

3.1. Values Inform Models

Lohr states that "In computing, a model is the equivalent of a metaphor, an explanatory simplification" (2015, p. 160). There are different schools of thought as to which types of mathematical models to apply in AI as Pedros Domingos highlights in the *Master Algorithm*. His exploration of the five tribes of machine learning demonstrate how values and design choices inform mathematical models. Symbolists value logic and inverse deduction while connectionists reverse engineer the brain with backpropagation and neural networks (Domingos, 2015). Decisions are made about how much "noise" to eliminate, what is considered an outlier in the model, and to determine acceptable levels of trade-offs between accuracy, generalizability and explainability (Domingos, 2015). Popular algorithms are also reused across applications, further amplifying and encoding their reach. Underlying all of this, is tacit agreement that efficiency, prediction and optimization, the rationale to build AI, are worthy goals. Those higher order values both drive and are driven by the ability to harness big data. It enables a "data-religion" which sees data as scientific, objective proof by which to explain the world (Harari, 2016). Like religion, the tribes within machine learning carry ideological perspectives that are encoded within the mathematical toolsets they choose to apply.

3.2. The Unexplainable Black Box

AI cannot always explain its decisions. Part of the power of deep neural networks, which is the focus of much current AI research, is that they learn on their own, in a "black box" that even their creators can't fully grasp (Marcus, 2019). That does not instill a lot of confidence and trust in using AI, especially in high stakes environments like healthcare. It also poses problems from a legal standpoint in assigning responsibility and accountability. Recently, even deep learning pioneers like Yoshua Bengio, have been candid about the need to work on understanding the "why" behind deep learning (Knight, 2019). However, some researchers wonder if there is a double standard being applied to machines when in many cases, humans cannot fully explain their decision-making processes (Zerilli, Knott, Maclaurin & Gavaghan, 2018). Finding a reasonable balance for how much explainability is required from an AI system might be context dependent and require discussions about how and where it will be used and

whether it's intended to support human decision making or replace it. Choices made in the model can lead to more or less explainability making this an important area for ethical discussion.

3.3. Designing Affordances

There is a need to design affordances into AI systems. Affordances speak to how we interact with objects based on their capabilities. We encounter affordances in the digital realm when we're faced with online forms that force answers to questions, categories in a drop-down menu or stereotypically gendered characters in a video game (Wittkower, 2017). These decisions may result in "disaffordances" that "fail to recognize differential embodied experiences" and include individual attributes like "race, gender, disability, and religion" (Wittkower, 2017, p. 2). IDEO, a leading design firm, thinks that human-centered design principles might be a way to address this problem in AI. They believe the design community can help "figure out how to apply the skills and abilities that data scientists typically have in service of people's needs" (Budds, 2017). Ensuring that AI is designed ethically will require data scientists to think about and design the affordances of their systems. The literature challenges the notion that math is neutral, impartial and objective. Instead, it illustrates how choices made inform mathematical models, which leads us to consider, who makes those choices?

4. Who Makes AI?

Beyond data and models, the people who build AI warrant consideration as they are the designers who's choices shape technology. Some ethical issues in AI stem from the very conception of the problem. How problems are framed, and the corresponding solutions encoded within AI systems links directly to the people involved in the design process. A 2019 report by AI Now, which looks at the issue of inclusion and diversity, notes that "as the focus on AI bias and ethics grows the scope of inquiry should expand to consider not only how AI tools can be biased technically, but how they are shaped by the environments in which they are built and the people that build them" (West, Whittaker & Crawford, 2019, p. 6). According to a global AI talent survey, there are 22,400 people working in AI based on those who publish research, however 36,524 people self-report as AI specialists on Linked In (Mantha & Kiser, 2019). It is an elite community comprised primarily of highly educated white men.

4.1. Diversity

The demographic backgrounds of those who develop technology play a role in how and what technology is developed. There are numerous studies and reports about the lack of gender diversity in the field of computing science. AI is even less representative with women making up only 16% of AI researchers (Mantha & Kiser, 2019). Silicon Valley likes to present itself as a meritocracy and has framed the gender diversity problem as a "pipeline issue" of not having enough qualified female candidates (West, Whittaker & Crawford, 2019). Yet, the masculinization of computing work was a long process that edged out women, as programming shifted from a "low-status, feminized task to work that was seen as central to control of corporate and state resources" (Miltner, 2019). Decades of research also shows that who makes a technology impacts what is constructed, and that gender plays a role in creating and reinforcing stereotypes through technology design (Cassell, 2001).

In *Algorithms of Oppression*, Safiya Noble challenges the notion that commercial search engines, fuelled by powerful AI algorithms, provide an unbiased and value-free service. She presents numerous examples that highlight Google's racial and gender biased search results, particularly as they impact Black women. Search results seek to further entrench and protect the interests of those already in power, misrepresent marginalized groups and keep women and minorities locked out of participating in the creation of technology (Noble, 2018). Only "2.5% of Google's workforce is Black" (West et al., 2019, p. 3). "The diversity problem is...most fundamentally about power. It affects how AI companies

work, what products get built, who they are designed for and who benefits from their development (West et al., 2019, p. 5).

4.2. Funding

In addition to individual traits, the socio-cultural environments in which people operate is influential. One example is the relationship between funding and the direction of research. Industry generally focuses on research with commercial applications which can mean that areas important to the public good, but not commercially viable, such as public health, may receive less attention (Fabbri, Lai, Grundy & Bero, 2018, p. e9). This has consequences, as we are seeing how chronic, systemic underfunding of public health infrastructure has contributed to a lack of readiness to address the COVID-19 pandemic at the expense of many lives being lost (Maani & Galea, 2020). Universities are not only looking to corporate sources for funding, but they are also becoming increasingly interested in holding patents to commercialize research, leaving scientists with an escalating imperative to engage with market or economic considerations (Slaughter, Archerd & Campbell, 2004). This creates a host of ethical "quandaries" to be navigated such as conflicts of incentives between publishing vs patenting, financial conflicts of interest and ethical considerations around student labour and the use of public funds to advance private intellectual property (Slaughter et al., 2004). AI is attracting significant investment from both government and industry making questions about funding and the associated ethical quandaries timely. The social construction of technology is informed by personal demographics, mixed with factors like funding, that can impact not only how AI is built, but what gets built and who owns it.

4.3. Purpose

Finally, it is also important to reflect on the overall project and ask is the purpose ethical? A controversial study of an AI system that uses facial recognition to determine sexual orientation raises questions about its purpose (Burdick, 2017). Regardless of whether this system works or not (which is debatable), do we want to live in a world that uses of this type of technology? The authors of the study say they conducted this research in order to expose discrimination in existing facial recognition algorithms (Wang & Kosinski, 2018). However, their research received considerable backlash, particularly from LGBTQ communities who might be harmed by this technology (Burdick, 2017). Questions of purpose and stakeholders impacted should be central considerations for researchers. Working proactively with AI researchers to have frank discussions that surface ethical issues and "blind-spots" while acknowledging the socio-cultural context in which they operate can enable necessary adjustments to the development workflow or at least acknowledge where potential gaps exist.

5. Conclusion

AI is a technology that is already having profound impacts on society. Yet, AI does not exist separate or outside of society, but within it. AI is constructed by people using data and models, all of which exists within and are informed by a socio-cultural context. There are numerous examples of how AI systems have served to encode harmful power structures by further amplifying inequity, discrimination, and bias. These systemic issues can be linked to the data, models and ultimately, the people involved, as well as the business models and social systems which currently govern how AI is constructed.

If we want to drive different, better, outcomes, then we need to change the inputs. Ethical practices can bring a more thoughtful approach to AI development. This starts with the purpose of examining if the use of AI is the appropriate in the first place rather than defaulting to a technological determinism, whereby technology is seen as the inevitable solution. More thoughtful decisions can be

made about the data and models used to construct AI that consider the impact of the systems being constructed. However, perhaps the most important input that needs to change is people. By having a more diverse, inclusive group of people involved in constructing AI, we can establish better goals, objectives consider a holistic perspective. We can design our AI-enabled digital infrastructure with the same care as we design our physical infrastructure.

6. References

- Ahmed, W., Bath, P. & Demartini, G. Chapter 4 Using Twitter as a Data Source: An Overview of Ethical, Legal, and Methodological Challenges. In: Woodfield, K., (ed.) *The Ethics of Online Research. Advances in Research Ethics and Integrity* (2). Emerald, pp. 79-107. ISBN 978-1-78714-486-6, 2017.
- Bajarin, T. Why it matters that IBM has abandoned its facial recognition technology. *Forbes*, 2020. Retrieved from - <https://www.forbes.com/sites/timbajarin/2020/06/18/why-it-matters-that-ibm-has-abandoned-its-facial-recognition-technology/#f66ec3fafaf3>
- Budds, D. Exclusive: Ideo's Plan To Stage An AI Revolution, 2017. Retrieved December 9, 2019, from *Fast Company* website: <https://www.fastcompany.com/90147010/exclusive-ideos-plan-to-stage-an-ai-revolution>
- Burdick, A. The A.I. "Gaydar" Study and the Real Dangers of Big Data. *The New Yorker*, 2017. Retrieved from <https://www.newyorker.com/news/daily-comment/the-ai-gaydar-study-and-the-real-dangers-of-big-data>
- Cassell, J. Genderizing HCI, 2001. Retrieved from <https://pdfs.semanticscholar.org/4810/c28fe3523b52bd39b1eeb3b6225ab2145fa7.pdf>
- Dastin, J. Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*, 2018. Retrieved from - <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>
- Domingos, P. *The Master Algorithm: How the quest for the ultimate learning machine will remake our world*. New York, NY: Basic Books, 2015.
- Fabbri, A., Lai, A., Grundy, Q., Bero, L. A. The influence of industry sponsorship on the research agenda: A scoping review. *American Journal of Public Health*. 108 (11), e9-e16, 2018.
- Gomes de Andrade, N., Pawson, D., Muriello, D. et al. Ethics and Artificial Intelligence: Suicide Prevention on Facebook. *Philos. Technol.* 31, 669–684, 2018. doi:10.1007/s13347-018-0336-0
- Harari, Y.N. *The Data Religion*. In *Homo Deus A Brief History of Tomorrow* (pp 428-462). Toronto, ON. Penguin Random House Canada, 2015.
- Heffetz, O., & Ligett, K. Privacy and Data-Based Research. *The Journal of Economic Perspectives: A Journal of the American Economic Association*, 28(2), 75–98, 2014. <https://doi.org/10.1257/jep.28.2.75>
- Henle, T., Matthews, G. J., & Harel, O. Data Confidentiality. In A. Levy, S. Goring, C. Gatsonis, B. Sobolev, E. van Ginneken, & R. Busse (Eds.), *Health Services Evaluation* (pp. 717–731), 2019. https://doi.org/10.1007/978-1-4939-8715-3_28
- Kitchin, Rob. *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. Sage, UK, 2017.
- Knight, W. An AI pioneer wants his algorithms to understand the why. *Wired*, 2019. Retrieved from - <https://www.wired.com/story/ai-pioneer-algorithms-understand-why/>
- Kugler, Matthew B., *From Identification to Identity Theft: Public Perceptions of Biometric Privacy Harms*. U.C. Irvine Law Review. (Forthcoming), 2018. Available at SSRN: <https://ssrn.com/abstract=3289850> or <http://dx.doi.org/10.2139/ssrn.3289850>
- Lohr, S. *Data-ism: The Revolution Transforming Decision Making, Consumer Behaviour And Almost Everything Else*. New York, NY: Harper Collins, 2015.
- Maani, N., & Galea, S. COVID-19 and Underinvestment in the Public Health Infrastructure of the United States. *The Milbank Quarterly*, 98(2), 250–259, 2020. <https://doi.org/10.1111/1468-0009.12463>
- Mantha, Y. and Kiser, G. *Global AI Talent Report*, 2019. Retrieved from <https://jfgagne.ai/talent-2019/>
- Marcus, G. and Davis, E. *Rebooting AI Building Artificial Intelligence We Can Trust*. New York, NY: Penguin Random House, 2019.

- Miltner, K. *Girls who coded gender in twentieth century U.K. and U.S. computing. [Review of the books Programmed Inequality: How Britain Discarded Women Technologists and Lost its Edge by Marie Hicks, Recoding Gender: Women's changing participation in computing by Janet Abbatte and The Computer Boys Take Over: Programmers and the Politics of Technical Expertise by Nathan Ensmenger]. Science, Technology & Human Values. 44(1), 161-176, 2019.*
- Noble, S. U. *Algorithms of Oppression. New York, NY: New York University Press, 2018.*
- O'Neil, C. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. New York: Crown Publishers, 2016.*
- Richards, N. M. and Hartzog, W. *The Pathologies of Digital Consent. Washington University Law Review, 2019. Retrieved from <https://ssrn.com/abstract=3370433>*
- Samuels, A. *The Internet Is Enabling a New Kind of Poorly Paid Hell. The Atlantic, 2018. Retrieved from - <https://www.theatlantic.com/business/archive/2018/01/amazon-mechanical-turk/551192/>*
- Shilton, K. *Anticipatory ethics for a future Internet: analyzing values during the design of an Internet infrastructure. Science and Engineering Ethics, 21(1), 1-18, 2015. <https://doi.org/10.1007/s11948-013-9510-z>*
- Slaughter, S., Archerd, C. J., Campbell, T. I. D. *Boundaries and Quandaries: How Professors Negotiate Market Relations. The Review of Higher Education. 28(1), 129-165, 2004.*
- Smith, C. S. *Building a World Where Data Privacy Exists Online. The New York Times, 2019. Retrieved from <https://www.nytimes.com/2019/11/19/technology/artificial-intelligence-dawn-song.html>*
- Solove, D. J. *Nothing to Hide: The False Trade-off between Privacy and Security. New Haven, CT: Yale University Press, 2013.*
- Strubell, E., Ganesh, A., & McCallum, A. *Energy and Policy Considerations for Deep Learning in NLP, 2019. Retrieved from <http://arxiv.org/abs/1906.02243>*
- Wang, Y., & Kosinski, M. *Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. Journal of personality and social psychology, 114(2), 246, 2018.*
- West, S.M. Whittaker, M and Crawford, K. *Discriminating systems: Gender, race and power in AI. AI Now Institute, 2019. Retrieved from <https://ainowinstitute.org/discriminatingystems.html>*
- Wittkower, D.E. *Disaffordances and dysaffordances in code. Paper presented at AoIR 2017: The 18th Annual Conference of the Association of Internet Researchers. Tartu, Estonia: AoIR, 2017. Retrieved from - <http://spir.aoir.org>*
- Zang, J., Dummit, K., Graves, J., Lisker, P., & Sweeney, A. L. *Who Knows What About Me? A Survey of Behind the Scenes Personal Data Sharing to Third Parties by Mobile Apps. Technology Science, 2015. Retrieved from <https://techscience.org/a/2015103001.pdf>*
- Zerilli, J., Knott, A., Maclaurin, J., & Gavaghan, C. *Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard? Philosophy & Technology, 2018. <https://doi.org/10.1007/s13347-018-0330-6>*
- Zuboff, S. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. New York: PublicAffairs, 2019.*