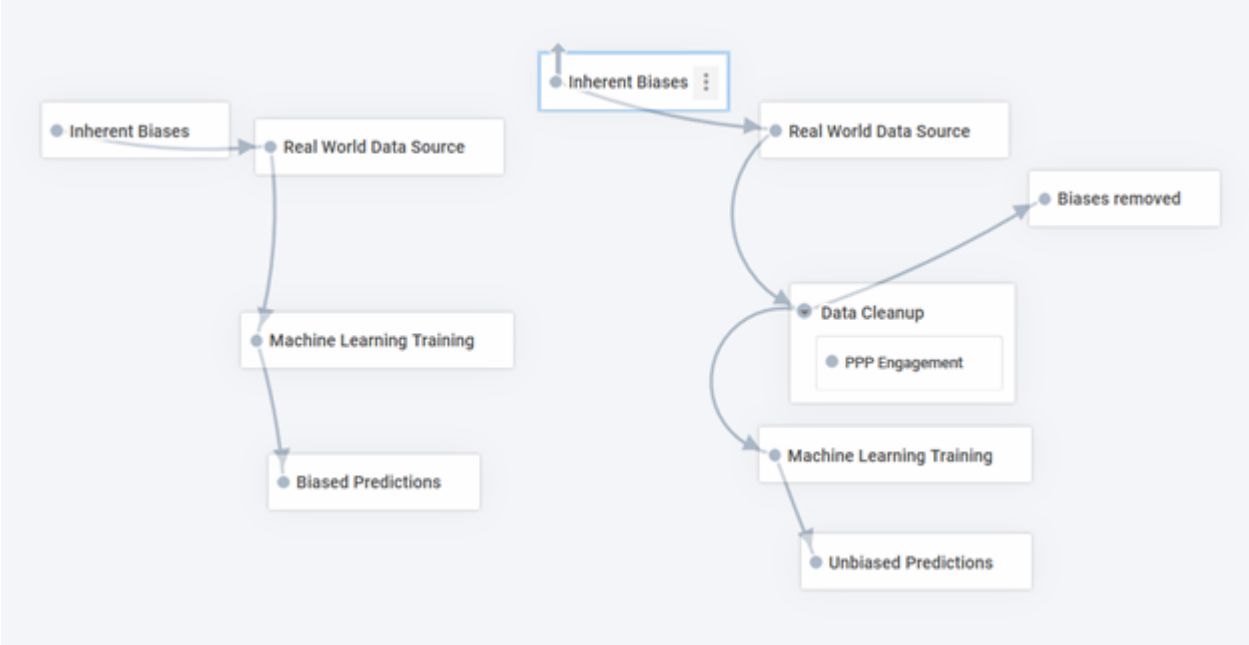
By Matt Chessen

Edited by Joe VerValin

Policy-makers must engage with technologists who are experts in the development of machine-based artificial intelligence. This is especially the case with machine learning systems that allow computers to learn without being explicitly programmed. If policy-makers do not engage in this area, then they are implicitly adopting a utilitarian approach to the results these systems produce, focusing only on the fairness of the outcomes and not the means that achieved those ends. There is a major problem with this hands-off approach. Technologists often, consciously or unconsciously, encode laws, policies, and virtues in decision-making machine learning systems.Public policy development faces certain challenges as policy-makers do not have the technical knowledge to balance these variables while technologists do not hone the art of making trade-offs. In fact, technologists focusing on system optimization may have no idea they are encoding laws, policies, and virtues at all.

The media has frequently reported on bias in artificial intelligence tools and companies pioneering machine-learning tools are a frequent target (e.g. racial bias,[1] gender bias,[2] Facebook,[3] Google,[4] and HP,[5] among others). These issues are especially apparent in machine learning systems that can automatically and unintentionally encode biases. Machine learning systems require large data-sets for training. These data-sets are frequently drawn from real-world data and, as we know, the real world has a significant number of biases. Biases that are not corrected before training will pass through to the output of the machine learning system.

Figure 1: Map showing how biases impact AI Predictions.



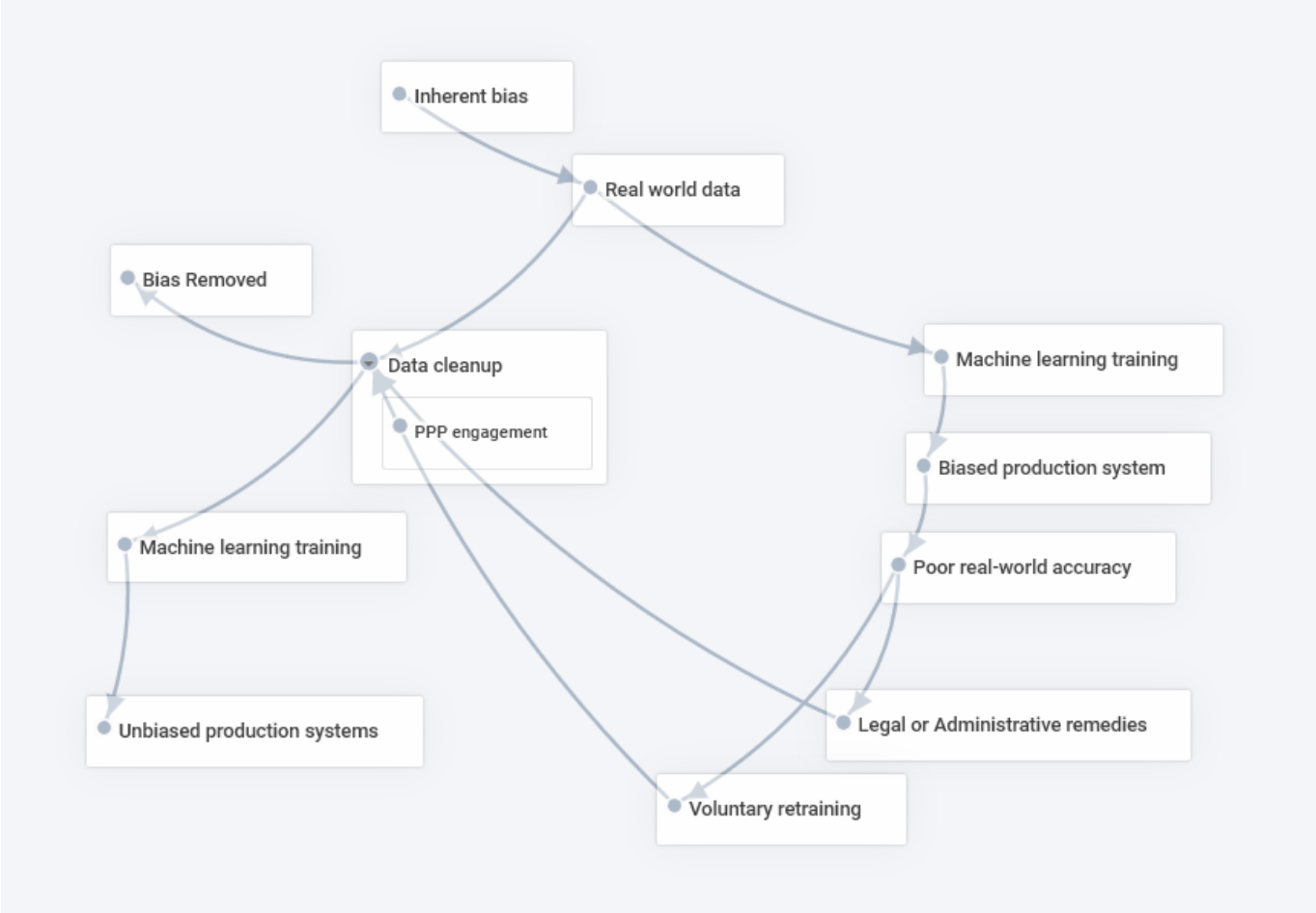
A Google machine learning system delivered ads for job openings to women that paid less than the job openings that were targeted towards men.[6] Why might this be? Currently in the United States,[7] women are paid approximately 0.75-0.80 cents for every dollar men earn.[8] One theory is that the Google system was trained with real world data where this gender bias was baked in, so the results reflected this bias. It wasn't the machine's fault – it was just sending men and women jobs it thought were the best matches based on the training data it was given. And you could blame the technologists who created it, but their skill is in creating effective software, not correcting gender bias, which is the specialty of a public-policy professional.

Compounding the problem, most artificial intelligence (AI) researchers are white men.[9] This causes an issue known colloquially as the "sea of dudes" or "white guy" problem,[10] where machine intelligence winds up mimicking the biases and privileges of its creators due to said group of creators being demographically monolithic.[11] Furthermore, AI job ads tend to target male applicants, perpetuating the problem.[12] This issue of bias is potentially lurking in many machine learning systems and we need new cross-disciplinary collaboration to solve it.

A New Public Policy Specialization

The public policy profession needs a new specialization in machine learning and data science ethics. If policy professionals are not engaged with technologists in the creation of machine learning systems, society will be forced to take a utilitarian approach towards these systems – focusing on the fairness of outcomes. This is a costly and inefficient approach, because unfair outcomes will need to be challenged in court or through administrative procedures. Machine learning complicates blame, legal liability and other post hoc remedies because these systems don't make decisions like traditional software. They are more statistical than algorithmic; more 'black-box' than auditable. There is no linear decision-tree that can be followed to determine why a machine learning tool came to a particular conclusion. (And systems that do have auditable decision trees simply aren't as effective as machine learning tools). While researchers are hard at work on this issue, machine learning systems can't yet explain to you why they made a decision. So it's entirely possible these systems could be making the right decision most of the time, but for the wrong reasons. And the incorrect decisions could be inexplicable without a thorough audit of the data used to train them.

Figure 2: Map showing factors that can aid in data cleanup.



If an unjust machine learning tool is challenged in court, the costs of a lawsuit may just be the beginning. If a machine learning system is found to be unfair, its data will need to be audited and cleaned, and the system will need to be retrained, adding considerable expense. A preferable approach is for policy and machine learning experts to work together on the creation of these systems to ensure they encode the laws, policies, and virtues we want to be reflected in their outputs. Fairness is a problem that should be fixed on the front-end, not policed on the back-end through legal liability.

AI researchers are examining the tradeoffs involved in creating fair machine learning systems.[13] Scientists at the Department of Computer and Information Science at the University of Pennsylvania conducted research on how to encode classical legal and philosophical definitions of fairness into machine learning.[14] They found that it is extremely difficult to create systems that fully satisfy multiple definitions of fairness at once.[15] There's

a trade-off between measures of fairness, and the relative importance of each measure must be balanced.

Replace the word ‘fairness’ with ‘justice,’ ‘liberty,’ or ‘non-partisanship,’ and we start to see the challenge. Technologists may be unconsciously codifying existing biases by using data-sets that demonstrate those biases or could be creating new biases through simple ignorance of their potential existence. Technologists should consciously remove these biases and encode laws, policies, and virtues, (together, shortened for our purposes to ‘values’) into machine learning systems. These values can be mathematized, but there will always be tradeoffs among different values that actually impact on people and society in the real world.

The same University of Pennsylvania scientists found that there can be a cost to fairness in machine learning.[16] The penalty for a machine learning training rate from encoding fairness can be mild to significant depending on the complexity of the model. The cost of encoding values (and the benefits) must be balanced against the potentially lower costs of systems that are optimized without regard to values or bias. For example, it may not be worth the cost to encode values in a machine learning system that delivers ads for clothing. But for a system that determines eligibility for food stamps or advises on criminal sentences, the cost of bias is severe and likely worth the expense of encoding values. Technologists can make these trade offs, but this is not their area of expertise. It falls squarely in the realm of public policy. Therefore, public policy professionals must engage with technologists if we want machine learning systems to make decisions that reflect society’s laws, policies, and virtues.

A real world example illustrates why it is so important for public-policy professionals to engage before machine learning systems are created, rather than relying on post hoc remedies. Imagine that a U.S. state wants to eliminate gerrymandering by using a machine learning system that can draw legislative districts automatically. There are a number of different policy goals that could be encoded for drawing fair districts, like: seeking representation proportional to the number of votes for each party statewide; creating districts that keep population centers and communities intact; generating districts with equal population; ensuring racial and partisan fairness; and maintaining compact, contiguous districts. The manner in which these criteria are encoded in the machine

learning system will make significant differences in how the congressional districts are drawn and in the resulting balance of political power in the state.

As you can imagine, everyone from politicians to lawmakers to lobbyists would have a strong interest in working with technologists to ensure their equities are represented in the construction of a redistricting system. People wouldn't just leave it up to the technologists to build the system, and then rely on legal challenges if they didn't like the outcome. Interested parties would get involved as the system was being built. The stakes are simply too high for the politicians, lawyers, and policy experts to leave these value trade-offs to technologists that don't specialize in value tradeoffs. In this gerrymandering example, policy experts' equities are obvious because the impacts on businesses, organizations, and people are so profound.

Governments, businesses, and academia have interests

Beyond gerrymandering, the full range of machine learning applications will be used to make choices that impact nearly every decision in our lives. These choices range from credit eligibility for businesses and individuals;[17] to the ability to secure government services and public assistance; to determining whether you are a security risk;[18] to eligibility for insurance.[19] The machines will start out as advisors to humans, but as they become more trustworthy, people will outsource some complex decision making completely to machines. The EU was concerned about this possibility and it drafted rules that would ban significant autonomous decision making about EU citizens unless appropriate protections,[20] laws, or consent are in place.[21] This type of precautionary principle is laudable, but we should work to make these types of laws unnecessary. An industry code of conduct is one solution, but such an agreement is moot without people who can actualize it. What we really need are experts who can build the right kinds of systems, rather than laws that could stifle innovation and prevent the development of machine learning decision-making tools.

Governments need to be aware of these concerns and public-policy professionals should be assigned to work closely with technologists who are developing machine learning systems where bias could have significant impacts on public services. Conversely, technology companies need to be aware of the implicit value judgments they are building into machine learning systems. The smart companies will hire their own public-policy experts to help data-scientists and technologists uncover hidden biases and ensure the optimal balance

between relevant laws, policies, and virtues. Woe be to the company or government that neglects this critical element of machine learning system design for a tool with significant public impact. They are likely to find themselves in a court of law, explaining their negligent discrimination. And if U.S. corporate machine learning systems can not demonstrate fairness in their creation, they may be unable to access the E.U. market due to its cautious autonomous decision making rules.

Academic institutions also need to adapt. University machine learning curricula need more policy-based instruction so technologists are aware of these concerns. And the public-policy profession needs a new sub-specialty: the data and machine learning ethics analyst. These analysts will have a solid grounding in both public-policy and technical fields. They will speak the language of machine learning developers, data scientists, and policy professionals. They will work with technologists to help them understand the value judgments inherent in their systems, assist them in removing bias from data-sets, and advise them how to encode values to remove bias. Universities such as Harvard,[22] Cornell,[23] and the University of Edinburgh have introduced courses on AI ethics.[24] These should be required for any institution providing AI instruction and mandatory for policy and technology students studying AI.

If public-policy and AI can spawn this offspring, it will give us a chance to encode values in our machine learning decision-making tools. Machine learning tools have the potential to take inevitable human biases out of decisions and significantly reduce the bias in critical decisions. These decisions require public-policy professionals and technologists to work together and ensure that values are built into machine learning systems during development.[25] The values we encode in our AI systems should be chosen affirmatively and consciously so that they reflect the laws, policies, and virtues of our entire society.

Values are the most important choices we make as human beings. Our intelligent machines should reflect the importance of those values.

—

Opinions expressed are the author's own and do not represent the views of the U.S

government or any other organization.

Works Cited

"AI Now Summary Report." AI NOW. Accessed December 15, 2017. https://artificialintelligencenow.com/media/documents/AINowSummaryReport_3_RpmwKHu.pdf.

"Art. 22 GDPR – Automated individual decision-making, including profiling." General Data Protection Regulation (GDPR). https://gdpr-info.eu/art-22-gdpr/.

Barr, Alistair. "Google Mistakenly Tags Black People as 'Gorillas,' Showing Limits of Algorithms." The Wall Street Journal. July 02, 2015. http://blogs.wsj.com/digits/2015/07/01/google-mistakenly-tags-black-people-as-gorillas-showing-limits-of-algorithms/.

BIG DATA: A Tool for Fighting Discrimination and Empowering Groups . 2015. Https://fpf.org/wp-content/uploads/Big-Data-A-Tool-for-Fighting-Discrimination-and-Empowering-Groups-FINAL.pdf

Chen, Brian X. "HP Investigates Claims of 'Racist' Computers." Wired. December 22, 2009. https://www.wired.com/2009/12/hp-notebooks-racist/.

"CIS Research Areas." CIS – Research Areas. http://www.cis.upenn.edu/about-research/index.php.

Clark, Jack. "Artificial Intelligence Has a 'Sea of Dudes' Problem." Bloomberg.com. June 23, 2016. https://www.bloomberg.com/news/articles/2016-06-23/artificial-intelligence-has-a-sea-of-dudes-problem.

"CS 4732: Ethical and Social Issues in AI (Spring, 2017)." CS 4732 (Spring, 2017) Ethical and Social Issues in AI. Accessed November 19, 2017. http://www.cs.cornell.edu/courses/cs4732/2017sp/

Crawford, Kate. "Opinion | Artificial Intelligence's White Guy Problem." The New York Times. June 25, 2016. https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-probl

em.html.

Crosman, Penny. “Can AI Be Programmed to Make Fair Lending Decisions?” American Banker. September 27, 2016. https://www.americanbanker.com/news/can-ai-be-programmed-to-make-fair-lending-decisions.

Datta, Amit, Michael Carl Tschantz, and Anupam Datta. “Automated Experiments on Ad Privacy Settings.” *Proceedings on Privacy Enhancing Technologies* 2015, no. 1 (2015). doi:10.1515/popets-2015-0007.

“DEGREE REGULATIONS & PROGRAMMES OF STUDY 2017/2018.” Course Catalogue – Ethics of Artificial Intelligence (PHIL10167). Accessed November 28, 2017. http://www.drps.ed.ac.uk/17-18/dpt/cxphil10167.htm.

Dickler, Jessica. “Men still earn more than women for the same jobs.” CNBC. December 06, 2016. http://www.cnbc.com/2016/12/05/men-still-earn-more-than-women-with-the-same-jobs.html.

Dickson, Ben. “How data science fights modern insider threats.” TechCrunch. August 25, 2016. https://techcrunch.com/2016/08/25/how-data-science-fights-modern-insider-threats/.

“DPI-687: The Ethics and Governance of Artificial Intelligence.” Harvard Kennedy School. Accessed November 19, 2017. https://www.hks.harvard.edu/courses/ethics-and-governance-artificial-intelligence

Hayasaki, Erika. “Is AI Sexist?” Foreign Policy. January 19, 2017. http://foreignpolicy.com/2017/01/16/women-vs-the-machine/.

“Home Fairness, Accountability, and Transparency in Machine Learning.” FAT ML 2016. Accessed November 15, 2017. https://www.fatml.org/.

Jabbari, Shahin, Matthew Joseph, Michael Kearns, Jamie Morgenstern, and Aaron Roth. “Fairness in Reinforcement Learning.” [1611.03071] Fairness in Reinforcement Learning. August 06, 2017. https://arxiv.org/abs/1611.03071.

Joseph, Matthew, Michael Kearns, Jamie Morgenstern, and Aaron Roth. “Fairness in Learning: Classic and Contextual Bandits.” [1605.07139] Fairness in Learning: Classic and Contextual Bandits. November 07, 2016. https://arxiv.org/abs/1605.07139.

Joseph, Matthew, Michael Kearns, Jamie Morgenstern, Seth Neel, and Aaron Roth. “Fair Algorithms for Infinite and Contextual Bandits.” [1610.09559] Fair Algorithms for Infinite and Contextual Bandits. June 29, 2017. https://arxiv.org/abs/1610.09559.

10152513950536061. “Language in your job post predicts the gender of your hire.” Textio Word Nerd. June 21, 2016. https://textio.ai/gendered-language-in-your-job-post-predicts-the-gender-of-the-person-youll-hire-cd150452407d#.6ap21s9jb.

Loukides, Mike. “The ethics of artificial intelligence.” O’Reilly Media. November 14, 2016. Accessed December 15, 2017. https://www.oreilly.com/ideas/the-ethics-of-artificial-intelligence.

Manjoo, Farhad. “Facebook’s Bias Is Built-In, and Bears Watching.” The New York Times. May 11, 2016. https://www.nytimes.com/2016/05/12/technology/facebooks-bias-is-built-in-and-bears-watching.html.

Mattu, Julia Angwin Jeff Larson Lauren Kirchner Surya. “Machine Bias.” ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Metz, Cade. “Artificial Intelligence Is Setting Up the Internet for a Huge Clash With Europe.” Wired. June 03, 2017. https://www.wired.com/2016/07/artificial-intelligence-setting-internet-huge-clash-europe.

Openness and Oversight of Artificial Intelligence | Berkman Klein Center. Accessed November 15, 2017. https://cyber.harvard.edu/node/99783.

“Pay Equity & Discrimination.” Institute for Women’s Policy Research. https://iwpr.org/issue/employment-education-economic-change/pay-equity-discrimination/.

“Principles for Accountable Algorithms and a Social Impact Statement for Algorithms.” FAT ML 2016. Accessed December 15, 2018. https://www.fatml.org/resources/principles-for-accountable-algorithms.

Rao, Anand, Jaime Yoder, and Scott Buisse. *AI in Insurance: Hype or Reality?* March 2016. Https://www.pwc.com/us/en/insurance/publications/assets/pwc-top-issues-artificial-intelligence.pdf

Roth , Aaron. “Fairness in Learning.” Adventures in Computation. Accessed November 15,

2017. http://aaronsadventures.blogspot.com/2016/05/fairness-in-learning.html

“Scholarship.” FAT ML 2016. Accessed December 15, 2017. http://www.fatml.org/resources/relevant-scholarship.

“Video Series.” Video Series | Berkman Klein Center. Accessed December 15, 2017. https://cyber.harvard.edu/node/99765.

1. Mattu, Julia Angwin Jeff Larson Lauren Kirchner Surya. “Machine Bias.” ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing Manjoo, Farhad. Chen, Brian X. “HP Investigates Claims of ‘Racist’ Computers.” Wired. December 22, 2009. https://www.wired.com/2009/12/hp-notebooks-racist/. ↑
2. Hayasaki, Erika. “Is AI Sexist?” Foreign Policy. January 19, 2017. http://foreignpolicy.com/2017/01/16/women-vs-the-machine/ ↑
3. “Facebook’s Bias Is Built-In, and Bears Watching.” The New York Times. May 11, 2016. https://www.nytimes.com/2016/05/12/technology/facebooks-bias-is-built-in-and-bears-watching.html ↑
4. Barr, Alistair. “Google Mistakenly Tags Black People as ‘Gorillas,’ Showing Limits of Algorithms.” The Wall Street Journal. July 02, 2015. http://blogs.wsj.com/digits/2015/07/01/google-mistakenly-tags-black-people-as-gorillas-showing-limits-of-algorithms/\ ↑
5. Chen, Brian X. “HP Investigates Claims of ‘Racist’ Computers.” Wired. December 22, 2009. https://www.wired.com/2009/12/hp-notebooks-racist/. ↑
6. Datta, Amit, Michael Carl Tschantz, and Anupam Datta. “Automated Experiments on Ad Privacy Settings.” *Proceedings on Privacy Enhancing Technologies* 2015, no. 1 (2015). doi:10.1515/popets-2015-0007. ↑
7. Dickler, Jessica. “Men still earn more than women for the same jobs.” CNBC. December 06, 2016. http://www.cnbc.com/2016/12/05/men-still-earn-more-than-women-with-the-same-jobs.html ↑
8. “Pay Equity & Discrimination.” Institute for Women’s Policy Research. https://iwpr.org/issue/employment-education-economic-change/pay-equity-discrimination/. ↑
9. Clark, Jack. “Artificial Intelligence Has a ‘Sea of Dudes’ Problem.” Bloomberg.com. June 23, 2016. https://www.bloomberg.com/news/articles/2016-06-23/artificial-intelligence-has-a-sea-of-dudes-problem ↑
10. Crawford, Kate. “Opinion | Artificial Intelligence’s White Guy Problem.” The New York

Times. June 25, 2016. https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html 10152513950536061. ↑
11. Furness, Dyllan. "Will AI built by a 'sea of dudes' understand women? AI's inclusivity problem." Digital Trends. December 05, 2016. Accessed January 1, 2018. https://www.digitaltrends.com/cool-tech/women-in-artificial-intelligence/. ↑
12. "Language in your job post predicts the gender of your hire." Textio Word Nerd. June 21, 2016. https://textio.ai/gendered-language-in-your-job-post-predicts-the-gender-of-the-person-youll-hire-cd150452407d#.6ap21s9jb. ↑
13. "CIS Research Areas." CIS – Research Areas. http://www.cis.upenn.edu/about-research/index.php ↑
14. Jabbari, Shahin, Matthew Joseph, Michael Kearns, Jamie Morgenstern, and Aaron Roth. "Fairness in Reinforcement Learning." [1611.03071] Fairness in Reinforcement Learning. August 06, 2017. https://arxiv.org/abs/1611.03071 ↑
15. "Fair Algorithms for Infinite and Contextual Bandits." [1610.09559] Fair Algorithms for Infinite and Contextual Bandits. June 29, 2017. https://arxiv.org/abs/1610.09559 ↑
16. Joseph, Matthew, Michael Kearns, Jamie Morgenstern, and Aaron Roth. "Fairness in Learning: Classic and Contextual Bandits." [1605.07139] Fairness in Learning: Classic and Contextual Bandits. November 07, 2016. https://arxiv.org/abs/1605.07139. ↑
17. Crosman, Penny. "Can AI Be Programmed to Make Fair Lending Decisions?" American Banker. September 27, 2016. https://www.americanbanker.com/news/can-ai-be-programmed-to-make-fair-lending-decisions ↑
18. Dickson, Ben. "How data science fights modern insider threats." TechCrunch. August 25, 2016. https://techcrunch.com/2016/08/25/how-data-science-fights-modern-insider-threats/ ↑
19. Rao, Anand, Jaime Yoder, and Scott Buisse. *AI in Insurance: Hype or Reality?* March 2016.Https://www.pwc.com/us/en/insurance/publications/assets/pwc-top-issues-artificial-intelligence.pdf ↑
20. Metz, Cade. "Artificial Intelligence Is Setting Up the Internet for a Huge Clash With Europe." Wired. June 03, 2017. https://www.wired.com/2016/07/artificial-intelligence-setting-internet-huge-clash-europe/ ↑
21. "Art. 22 GDPR – Automated individual decision-making, including profiling." General Data Protection Regulation (GDPR). https://gdpr-info.eu/art-22-gdpr/. ↑
22. "DPI-687: The Ethics and Governance of Artificial Intelligence." Harvard Kennedy School. Accessed November 19, 2017.

https://www.hks.harvard.edu/courses/ethics-and-governance-artificial-intelligence "CS 4732: Ethical and Social Issues in AI (Spring, 2017). ↑
23. "CS 4732: Ethical and Social Issues in AI (Spring, 2017)." CS 4732 (Spring, 2017) Ethical and Social Issues in AI. Accessed November 19, 2017. http://www.cs.cornell.edu/courses/cs4732/2017sp/; ↑
24. "DEGREE REGULATIONS & PROGRAMMES OF STUDY 2017/2018." Course Catalogue – Ethics of Artificial Intelligence (PHIL10167). Accessed November 28, 2017. http://www.drps.ed.ac.uk/17-18/dpt/cxphil10167.htm. ↑
25. *BIG DATA: A Tool for Fighting Discrimination and Empowering Groups* . 2015. Https://fpf.org/wp-content/uploads/Big-Data-A-Tool-for-Fighting-Discrimination-and-Empowering-Groups-FINAL.pdf ↑

Matt Chessen

Matt Chessen is a career U.S. diplomat, technologist, and author who is currently serving as a Senior Technology Policy Adviser in the Office of the Science and Technology Adviser to the Secretary of State. From 2016-2017, Matt was the State Department Science and Technology Policy Fellow at the George Washington University, where he researched the international implications of artificial intelligence, computational propaganda, cognitive security, and machine-driven communications. From 2014-2016, Matt was the Coordinator for International Cyber Policy for the Bureau of East Asian and Pacific Affairs where he led the regional implementation of the US International Strategy for Cyberspace. Matt holds a J.D. from Georgetown University, and an M.B.A. and B.A. from the University of Arizona. He has earned eight honor awards for his service at the Department of State, including Superior Honor Awards for his work on the Afghan Peace Process and his efforts advancing US international cyber policy.
View all posts