"AI Ethics Then & Now: A Look Back on the Last Five Years"

Selected Bibliography

Willie Costello, August 27, 2020

The ethics of the outputs

Fairness of the algorithm

Barocas, Solon, and Andrew D. Selbst. "Big Data's Disparate Impact." *California Law Review* 104, no. 3 (2016): 671–732.

Mitchell, Margaret, Dylan Baker, Nyalleng Moorosi, Emily Denton, Ben Hutchinson, Alex Hanna, Timnit Gebru, and Jamie Morgenstern. "Diversity and Inclusion Metrics in Subset Selection." *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, February 7, 2020, 117–23. https://doi.org/10.1145/3375627.3375832.

Verma, Sahil, and Julia Rubin. "Fairness Definitions Explained." In *2018 IEEE/ACM International Workshop on Software Fairness (FairWare)*, 1–7, 2018. https://doi.org/10.23919/FAIRWARE.2018.8452913.

Fairness of the use

Agre, Philip E., "Your Face Is Not a Bar Code: Arguments Against Automatic Face Recognition in Public Places" (2001) https://pages.gseis.ucla.edu/faculty/agre/bar-code.html

Buolamwini, Joy, and Timnit Gebru. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." In *Conference on Fairness, Accountability and Transparency*, 77–91, 2018. http://proceedings.mlr.press/v81/buolamwini18a.html.

Eubanks, Virginia. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Publishing Group, 2018.

Prabhu, Vinay Uday, and Abeba Birhane. "Large Image Datasets: A Pyrrhic Win for Computer Vision?" *ArXiv:2006.16923 [Cs, Stat]*, July 23, 2020. http://arxiv.org/abs/2006.16923.

Raji, Inioluwa Deborah, Timnit Gebru, Margaret Mitchell, Joy Buolamwini, Joonseok Lee, and Emily Denton. "Saving Face: Investigating the Ethical Concerns of Facial Recognition Auditing." *ArXiv:2001.00964 [Cs]*, January 3, 2020. http://arxiv.org/abs/2001.00964.

Selbst, Andrew D., Danah Boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi. "Fairness and Abstraction in Sociotechnical Systems." In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 59–68. FAT* '19. Atlanta, GA, USA: Association for Computing Machinery, 2019. https://doi.org/10.1145/3287560.3287598.

The ethics of the inputs

Documenting datasets

Bender, Emily M., and Batya Friedman. "Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science." *Transactions of the Association for Computational Linguistics* 6 (December 1, 2018): 587–604. https://doi.org/10.1162/tacl_a_00041.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. "Datasheets for Datasets." *ArXiv:1803.09010 [Cs]*, March 19, 2020. http://arxiv.org/abs/1803.09010.

Mitchell, Margaret, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. "Model Cards for Model Reporting." *Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT* '19*, 2019, 220–29. https://doi.org/10.1145/3287560.3287596.

Raji, Inioluwa Deborah, Andrew Smart, Rebecca N. White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. "Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing." *ArXiv:2001.00973 [Cs]*, January 3, 2020. http://arxiv.org/abs/2001.00973.

Data collection

Denton, Emily, Alex Hanna, Razvan Amironesei, Andrew Smart, Hilary Nicole, and Morgan Klaus Scheuerman. "Bringing the People Back In: Contesting Benchmark Machine Learning Datasets." *ArXiv:2007.07399 [Cs]*, July 14, 2020. http://arxiv.org/abs/2007.07399.

Jo, Eun Seo, and Timnit Gebru. "Lessons from Archives: Strategies for Collecting Sociocultural Data in Machine Learning." *ArXiv:1912.10389 [Cs]*, December 22, 2019. https://doi.org/10.1145/3351095.3372829.

Race critical code studies

Benjamin, Ruha. *Race After Technology: Abolitionist Tools for the New Jim Code.* John Wiley & Sons, 2019.

Browne, Simone. *Dark Matters: On the Surveillance of Blackness*. Duke University Press, 2015.

Chun, Wendy Hui Kyong. "Introduction: Race and/as Technology; or, How to Do Things to Race." *Camera Obscura: Feminism, Culture, and Media Studies* 24, no. 1 (70) (May 1, 2009): 7–35. https://doi.org/10.1215/02705346-2008-013.

Coleman, Beth. "Race as Technology." *Camera Obscura: Feminism, Culture, and Media Studies* 24, no. 1 (70) (May 1, 2009): 177–207. https://doi.org/10.1215/02705346-2008-018.

D'Ignazio, Catherine, and Lauren F. Klein. *Data Feminism*. MIT Press, 2020.

Noble, Safiya Umoja. *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press, 2018.

Hanna, Alex, Emily Denton, Andrew Smart, and Jamila Smith-Loud. "Towards a Critical Race Methodology in Algorithmic Fairness." *ArXiv:1912.03593 [Cs]*, December 7, 2019. https://doi.org/10.1145/3351095.3372826.

The ethics of the creators

"Abolish the #TechToPrisonPipeline" https://bit.ly/TechToPrisonPipeline

Kalluri, Ria. "Don't ask if artificial intelligence is good or fair, ask how it shifts power" *Nature*, July 7, 2020. https://www.nature.com/articles/d41586-020-02003-2

Winner, Langdon. "Do Artifacts Have Politics?" *Daedalus* 109, no. 1 (1980): 121–36.

News articles referenced

https://www.nytimes.com/2015/07/10/upshot/when-algorithms-discriminate.html

https://www.propublica.org/article/asians-nearly-twice-as-likely-to-get-higher-price-from-princeton-review

https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

https://www.reuters.com/investigates/special-report/usa-riteaid-software/

https://www.theatlantic.com/business/archive/2015/09/discrimination-algorithms-disparate-impact/403969/

https://www.theguardian.com/technology/2015/may/20/flickr-complaints-offensive-auto-tagging-photos

https://www.washingtonpost.com/news/the-intersect/wp/2015/05/20/google-maps-white-house-glitch-flickr-auto-tag-and-the-case-of-the-racist-algorithm/