

Alex Hansen

DSCI-403-OL

11 September 2023

### Data Ethics Case Study: Biases in AI Art

AI Art (more appropriately Machine Learning Art) has been an exploding field over the last several years, and with it have come serious legal and ethical questions related to copyright law/infringement and artistic licensing, compensation, and ownership, however these are not the only concerns regarding AI Art; with its growing use, the societal and cultural biases and bigotries are being noticed. From major journalistic names and researchers to professional bloggers and TikTok'ers there have been a growing number of publications calling out the inherent racism, ableism, agism, and sexism found in AI Art generation algorithms. This case study would perfectly fit as an example in Cathy O'Neil's TedTalk as an example of blind faith in big data and the need for algorithm auditing and correction.

Cathy begins her talk by describing how algorithms are used to sort people (data) into categories, one being the successful (target) output. She goes on to discuss examples of how an inappropriate definition of success and blind faith in big data led to the loss of jobs for many New York teachers; how no one in the state of New York even knew what the actual model was to generate these outputs, yet it was used to take away the livelihood of many teachers. This is similar to the lack of understanding of input data for many AI models. Time Magazine (2019) discusses how early facial recognition models had much higher errors (35% vs 1%) for identifying black women vs white men; this is easy to understand when the government dataset utilized consisted of 70% men, 80% light-skinned individuals, and less than 5% women of color [1]. Stephanie Dinkins in a New York Times interview discussed similar issues in the early stages of AI Art and its inability to depict black women, attributed to the lack of relevant, input data. Linda Dounia Rebeiz demonstrated during an interview how Western cultural biases on the views of Africa lead DALLE-E2, an AI art generator, to depict the Senegalese capital of Dakar as an "arid desert landscapes [with] ruined buildings" [2] (for the unaware of how wrong this depiction is, google Dakar). In VentureBeat's 2020 article on Ramya Srinivasan's and Kanji Uchino's, of Fujitsu Laboratories of America, study "Biases in Generative Art – A Causal Look from the Lens of Art History" they highlight the author's discussion of how annotators' preferences, cultures, and beliefs might lead to inconsistencies in the labeling process for input data, thus generating biased models [3, 4]. These highlight Cathy's argument about how input data can already be skewed before it even reaches an algorithm, and that blind faith in the data is dangerous.

Cathy goes on to discuss how an algorithm's measure of success can be biased, discussing an AI used to filter "good" potential Fox News hires by looking at previous applicants for the last 20 years and of those of which were hired and stayed for multiple years; this measure of success excluded women as potentially good hires. TikTok influencer, writer, director, producer, and disability consultant, Jeremy Andrew Davis posted a video highlighting the bias of AI art generator MidJourney when it comes to the depiction of autistic individuals; he demonstrated

that in 148 queries only 2 were female, 5 were over 30 years old, 0 were not white, and 0 were smiling [5]. Not only is this an example of removal of “outliers” from the data enhancing biases, but it is also an example of AI reinforcing negative stereotypes; in every image the individual was depicted as melancholy with a dark and gloomy aesthetic. This form of stereotype reinforcement is light compared to other findings; Cathy O’Neil calls out the codification of sexism and bigotry in AI through the analysis of arrest data.

If the input arrest data is skewed towards men and minorities, due to cultural biases and bigotry leading to the over policing of targeted groups, then a model asked to ‘predict the location of criminal activities’ or ‘the appearance of a suspect’ it is going to skew to producing minority neighborhoods and appearances. In fact, Alexandra Luccioni, et al.’s study “Stable Bias: Analyzing Societal Representations in Diffusion Models” featured in Business Insider (2023) observes how Text-to-Image (TTI) systems could be used to build “virtual sketch artist” software for use by police departments to generate photo-realistic images of suspects based on witnesses’ verbal testimonies. They demonstrated how typical adjectives used by witnesses are gender coded and ethnically biased in TTI systems [6, 7]. These types of biases are harmful to individuals, communities, and society. AI companies have a duty to rectify their data, and forms of “data laundering”, as Cathy puts it, should be considered unethical and in some cases illegal. The questions of regulation and legality are outside the scope of this paper not discussed here, but without question the author of this paper, Cathy O’Neil, and the individual’s authoring the referenced research are proponents of instituting changes to how AI is managed from a legal and ethical perspective [4, 6].

Checking for fairness, auditing algorithms, and correcting them are critical to prevent long term effects; Cathy mentions feedback loops, but the erasure of culture is an even more detrimental prospect. Dinkins and Auriea Harvey both experienced cases where AI art generators refused to display results, and threatened to ban the users, for including the term “slave”. This, and other terms, were banned from the AI for being racially biased, however Harvey and Dinkins point out that this also erases a part of history [2]. A most interesting case of culture erasure is discussed in the ‘Fujitsu Art History’ study where DeepArt was asked to repaint images and failed to account for the contributions of important/key minorities in artistic movements, and recolored black figures in important works, such as Clementine Hunter’s Black Matriarch, while not recoloring white faces in works of art [4].

Compelling quotes from the study describe art as “much more than an aesthetic entity... art imparts “moral knowledge”, i.e., knowledge about what [we] ought to do and not to do... Art also enables “empathic knowledge”, something through which one can compare different views of the world through direct experience... A generated style translation should thus preserve such important characteristics of art movements, or else they will be contributing to a bias in understanding history... People have a propensity to favor suggestions from automated systems and to ignore contradictory information, even if it is correct. [Known] as “automation bias”. [4]” I believe Cathy would agree with those quotes, as she ended her talk on the importance of data scientist to be ethical translators of the discussions happening in society, and for the masses to demand greater accountability of companies and to end the blind faith in big data.

## References

- [1] Buolamwini, Joy. 2019. "Artificial Intelligence Has a Problem with Gender and Racial Bias. Here's How to Solve It." Time, February 7, 2019. <https://time.com>
- [2] Small, Zachary. 2023. "Black Artists Say A.I. Shows Bias, with Algorithms Erasing Their History." The New York Times, July 4, 2023. <https://www.nytimes.com>
- [3] Wiggers, Kyle. 2020. "Researchers Find Race, Gender, and Style Biases in Art-Generating AI Systems." VentureBeat, December 28, 2020. <https://venturebeat.com>
- [4] Srinivasan, Ramya, and Kanji Uchino. 2020. "Biases in Generative Art -- A Causal Look from the Lens of Art History." arXiv. <https://doi.org/10.48550/ARXIV.2010.13266>.
- [5] Jeremy Andrew Davis, "Is ai Ableist", TikTok, 7th August 2023
- [6] Luccioni, Alexandra Sasha, Christopher Akiki, Margaret Mitchell, and Yacine Jernite. 2023. "Stable Bias: Analyzing Societal Representations in Diffusion Models." arXiv. <https://doi.org/10.48550/ARXIV.2303.11408>.
- [7] Mok, Aaron. 2023. "What Do AI Art Generators Think a CEO Looks like? Most of the Time a White Guy." Business Insider, March 28, 2023. <https://www.businessinsider.com>