Keynote ASSETS '19, October 18-20, 2019, Philadelphia PA

My Algorithms Have Determined You're Not Human: AI-ML, Reverse Turing-Tests, and the Disability Experience

Karen Nakamura
Robert and Colleen Haas Distinguished Chair in Disability Studies
University of California, Berkeley
knak@berkeley.edu

Abstract

The past decade has seen an exponential growth in the capabilities and deployment of artificial intelligence systems based on deep neural networks. These are visible through the speech recognition and natural language processing of Alexa/Siri/Google that structure many of our everyday interactions, and the promise of SAE Level 5 autonomous driving provided by Tesla and Waze. Aside from these shiny and visible applications of AI-ML are many other uses that are more subtle: AI-ML is now being used to screen job applicants as well as determine which web ads we are shown. And while many vendors of AI-ML technologies have promised that these tools provide for greater access and freedom from human prejudice, disabled users have found that these tools can embed and deploy newer, subtler forms of discrimination against disabled people. At their worst, AI-ML systems can deny disabled people their humanity.

The explosion of AI-ML technologies in the last decade has been driven by at least three factors. First, the deep neural networks algorithms that currently drive much machine learning have been improved dramatically through the use of backpropagation [1], generative adversarial nets [2], and convolution [3], allowing for their deployment across a broad variety of datasets. Second, the cost of computing hardware (especially GPUs) has dropped dramatically while large scale cloud computing facilities and widespread fiber/broadband/4G has provided for universal availability. Finally, large datasets have come online to aid in the training of the neural nets – for example, the image datasets provided through Google and Facebook, the large natural language datasets

driving Amazon Alexa, and so forth.

Deep neural networks themselves have two key features or flaws, depending on the perspective. First, they are highly dependent on the diversity of the training dataset used. Second, their internal operations when deployed are entirely opaque not only to the end-user but also to the designers of the system itself.

Lack of Diversity in Training Datasets

Discussion of the lack of diversity in training datasets has for the most part centered on the lack of racial and gender diversity in facial recognition datasets, what has been called artificial intelligence's "white guy problem" [4]. To be sure, this is a serious issue, especially the egregious case of Google Photos in 2018 tagging African American individuals as "gorillas." With the deployment of facial recognition in law enforcement and immigration and border patrol situations, there are serious concerns about the higher error rate for black and brown individuals compared against white individuals. When gender intersects race, error rates also increase [5].

Much less is being said, though, about the misrecognition of disabled individuals, whether through vision, voice, or user-interface interactions. Part of this is because these limitations are seen as "natural" – for example, "of course" Alexa might not recognize the 'accent' of a hard of hearing user or a user whose cerebral palsy affects their vocal cords. That's only "natural," because "normal" people also have trouble understanding them. Thus, little effort has been placed into improving systems in these directions (although to be fair, many companies such as Amazon have recently been hiring disabled engineers to work in these areas).

Coding of the Implicit Biases and Limitations of the Trainers

Most of us are familiar with the reverse Turing-tests that are being used to prove that we're not robots when filling out web forms. These have taken many shapes, from the early OCR-type tests ("type in the letters or numbers you see") to the more complex vision recognition systems ("click on all of the tile squares that have busses"). These reverse Turing-tests exclude disabled people two ways. The obvious one is that some people who are blind or who are deaf cannot access the reCAPTCHA images or sound samples. Second, more insipidly, is that the reverse Turing-tests are being used as part of the

training system for AI-ML systems and thus code trainer bias.

When we click on "busses" for Google reCAPTCHA, we not only proving our humanity, we're providing Google with considerable data on what busses are and what busses aren't and through this we are helping train their next iterations of image recognition and autonomous vehicle technologies.

However, we (the end-users and system designers) do not actually know what the systems are learning. There is a (perhaps apocryphal) story in AI circles of a US Army machine learning system that was trained to differentiate between Russian and US army tanks and that passed tests 100%, but failed in the field. It was revealed that what it was really learning to recognize was that photographs taken on cloudy days were Russian and those taken on sunny days were American.

The consequences for disabled people being misrecognized can be serious. Wheelchair users are constantly being run over by human car drivers that do not recognize them as humans (the "I didn't even see you" excuse that bicyclists are also familiar with), yet the datasets being used to train automobile vision systems also embed similar limitations not only due to the lack of wheelchairs and scooters in training datasets, but the trainers themselves may be misrecognizing them.

The issue is that this lacuna is opaque to the system designers. There is no way to detect determine that AI-ML systems will fail to recognize disabled people as humans without testing them, and no way to explicitly program them without providing them with a rich dataset of disabled individuals in a wide variety of contexts – which doesn't exist. In all of these cases, the lack of input from disabled designers, programmers, trainers, and users leads to serious problems.

One final example is the use of AI-ML in screening job applicant resumes [6]. We've all seen the ads from recruiting companies who promise to find the right candidate resumes out of hundreds or thousands of applicants. These are often based on deep neural networks that have learned what a "successful" resume is for any particular position, but in doing so, it can also code the implicit bias of the training dataset. If recruiters have had a past implicit bias (say for people who engage in collegiate team sports) that is in the training set then the system will learn this – and worse, once trained, it is impossible to detect this bias by looking at the neural net itself (i.e., no forensic examination of the internals of the system will detect the bias, it will only be detectable by testing it on other

datasets). HR departments are attracted to this as a feature rather than a bug – there is absolute deniability of any hiring bias against protected categories.

How do we move forward?

While AI-ML has been advertised as progressive, the way that these systems have been trained can embed deeply conservative social values. Many facial recognition training datasets pre-code values of trinary 'race' and binary 'gender' which reify shifting social categories, and they omit faces that are not 'normal,' such as those of people who are disabled, ill, or disfigured. And once trained, those systems carry these biases forward. This short piece has not touched upon some of the other key ethical issues involved with working with datasets involving disabled individuals [7]

For us to move forward, we need to push to ensure that the designers, programmers, training sets, trainers, and users of emerging AI-ML systems represent the full diversity of our user population. Diversity and inclusion need to be programmed in from the start, it cannot be an afterthought. In response to the "gorillas" incident, Google Photos reacted by removing categories of "chimpanzees," "gorillas," and "apes" from the system. The work of actually training AI-ML systems to not only be better in recognizing African Americans and black faces let alone the diversity within the disability community will take considerably more work and the full inclusion of minoritized individuals in the planning, implementation, and training process.

Author Keywords

Artificial intelligence; deep neural networks; disabilities; race; bias

BIOGRAPHY

Karen Nakamura is a cultural and visual anthropologist at UC Berkeley and the Robert and Colleen Haas Chair in Disability Studies. Nakamura's research focuses on disability, sexuality, and minority social movements in contemporary Japan and the role technology and social policy in the inclusion and exclusion of minoritized individuals. She is currently creating the Berkeley Disability Lab — a nexus for thinking about technology, accessibility, and disability as well as design studio intentionally designed by and for disabled bodies

and minds.

References

[1] David E. Rumelhart, Geoffrey E. Hinton and Ronald J. Williams. 1986. Learning representations by back-propagating errors. Nature, 323, 533–536. https://doi.org/10.1038/323533a0

[2] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio (2014). Generative adversarial nets. Advances in neural information processing systems, 2672-2680.

[3] Pierre Sermanet, Soumith Chintala, and Yann LeCun. Convolutional neural networks applied to house numbers digit classification. arXiv preprint arXiv:1204.3968 (2012).

[4] Kate Crawford. 2016. Artificial Intelligence's White Guy Problem. NY Times, June 25. http://nyti.ms/28YaKg7

[5] Joy Buolamwini and Timnit Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of Machine Learning Research 81:1–15, 2018 Conference on Fairness, Accountability, and Transparency.

[6] Jutta Treviranus (2017). AI's problem with disability and diversity. https://www.cbc.ca/radio/spark/362-machine-learning-outliers-smart-device-ownership-and-more-1.4279433/ai-s-problem-with-disability-and-diversity-1.4279444

[7] Shari Trewi. 2018. AI Fairness for People with Disabilities: Point of View. arXiv e-prints arXiv:1811.10670.