



**AALBORG UNIVERSITY**  
DENMARK

**Aalborg Universitet**

## **Designing for Inclusive AI**

Jochum, Elizabeth; Schoux Casey, Christina

*Publication date:*  
2019

[Link to publication from Aalborg University](#)

*Citation for published version (APA):*

Jochum, E., & Schoux Casey, C. (2019). *Designing for Inclusive AI*. Poster presented at Kick-off: AI for the people, Aalborg, Denmark.

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

### **Take down policy**

If you believe that this document breaches copyright please contact us at [vbn@aub.aau.dk](mailto:vbn@aub.aau.dk) providing details, and we will remove access to the work immediately and investigate your claim.

# Designing for Inclusive AI

Elizabeth Jochum / Christina Schoux Casey  
Communication and Psychology / English, German, and Cultural Studies  
Aalborg University  
jochum@hum.aau.dk / casey@hum.aau.dk

*The resulting gap between the design and operation of algorithms and our understanding of their ethical implications can have severe consequences affecting individuals, groups and whole segments of society.*

-Mittelstadt, et al.<sup>1</sup>

## The Problem

AI has the potential to address economic and political injustice and contribute to more fair, equitable and sustainable societies. However, studies repeatedly demonstrate that AI tools are biased and actively discriminate against people of color, people with disabilities, poor people, and women. The reason for this is not necessarily intentional, but rather the result of uncritical assumptions and implicit bias in tools and processes used in the development, training and implementation of AI.

AI bias has real-world consequences. Race, gender, ability, and class biases are manifest in:

- Facial recognition and classification
- Motion recognition and classification
- Gender recognition and classification
- Emotion recognition and classification
- Natural Language Processing
- Video captioning
- Health Care assessment
- Predictive policing
- Prison sentencing
- CV evaluation
- Bank loan risk evaluation
- Admissions evaluation
- Hate speech detection +

*Currently there is no standard way to identify how a dataset was created, and what characteristics, motivations, and potential skews it represents.*

-Gebru et al.<sup>3</sup>

## The Opportunities

Interdisciplinary teams can work towards implementing inclusivity and fairness in AI at every level, including

- Planning and design

Are a range of sociodemographic categories included in the design and dataset (e.g., ability, age, race/ethnicity, gender, class, geographic region)?

How could the project potentially hurt, control, or profile protected (e.g., ability, race, gender, sexuality) or vulnerable (e.g., poor, imprisoned) groups?

- Dataset coding

How are fairness and inclusivity being addressed in the process of labeling, cleaning, sampling, scaling, and aggregation?

- Participatory & Inclusive Design Procedures

Involve test subjects and study participants earlier in the development  
Are usability testers and study participants drawn from diverse sociodemographic categories?

- Implementation

How does the project work with different user populations?

- Improvement and maintenance

How does unanticipated bias emerge after implementation?

What measures are in place for addressing bias after systems are implemented?

*We must hold our designs accountable to a reasonable degree of freedom from bias.*

-Friedman & Nissenbaum<sup>4</sup>

## Developing Best Practices

How do we transform machine learning and AI into ethical practices?

- *For researchers:* Inclusive design, coding, testing and practices
- *For students:* Interdisciplinary certificate in Ethical Artificial Intelligence, BA and MA level.

Incorporate ethical considerations throughout the data science curriculum. (National Academies of Science 2018)<sup>6</sup>

- *For projects:* Establish a protocol for data transparency and fairness (e.g., using openethics.ai).

Adopt the Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS) from IEEE.

*Who codes matters, how we code matters, why we code matters.*

-Boulamwini<sup>5</sup>

## Next Steps

Design BA/MA courses in Ethics and AI/ML

“Humans In the Loop: Robots, Automation and Humanities Research” (SP2020)

“Humans, Ethics, and AI” course; crosslisted

EU Centers of Excellence Network for Trustworthy Robotics and Intelligent Systems. [www.centris-ai.eu](http://www.centris-ai.eu)



[www.ajlunited.org](http://www.ajlunited.org)



[aiethicslab.com](http://aiethicslab.com)



[d4bl.org](http://d4bl.org)



[ainowinstitute.org](http://ainowinstitute.org)

## Potential Partners

The Just Data Lab \* Design Justice Network \* Deep Learning Indaba \* The Partnership on AI  
Leverhulme Centre for the Future of Intelligence \* The Institute for Ethical AI & Machine Learning  
\* MIT-Watson AI Lab \* The Berkman Klein Center for Internet and Society \* Digital Ethics Lab

## References

1. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*. <https://doi.org/10.1177/2053951716679679>
2. Jochum, E., & Derks, J. (2019). Tonight We Improvise! Real-time tracking for human-robot improvisational dance. Manuscript submitted for publication. In *International Conference on Movement and Computing (MOCO) 2019*.
3. Gebru, Timnit, & et al. (2018). Datasheets for Datasets. Retrieved from arXiv:1803.09010v4
4. Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems*, 14(3), 330–347. <https://doi.org/10.1145/230538.230561>
5. Boulamwini, J. (2016). “InCoding: In the Beginning was the Coded Gaze.” <https://medium.com/mit-media-lab/incoding-in-the-beginning-4e2a5c51a45d>
6. National Academies of Sciences. (2018). *Envisioning the Data Science Discipline: The Undergraduate Perspective*. NA Press. <https://www.nap.edu/catalog/24886/>



Figure 1. *The Dynamic Still*. Jochum et al. Pictured: Sandro Masai



Figure 2. Joy Boulamwini. [time.com](http://time.com)