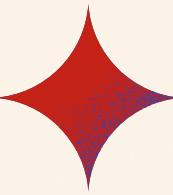
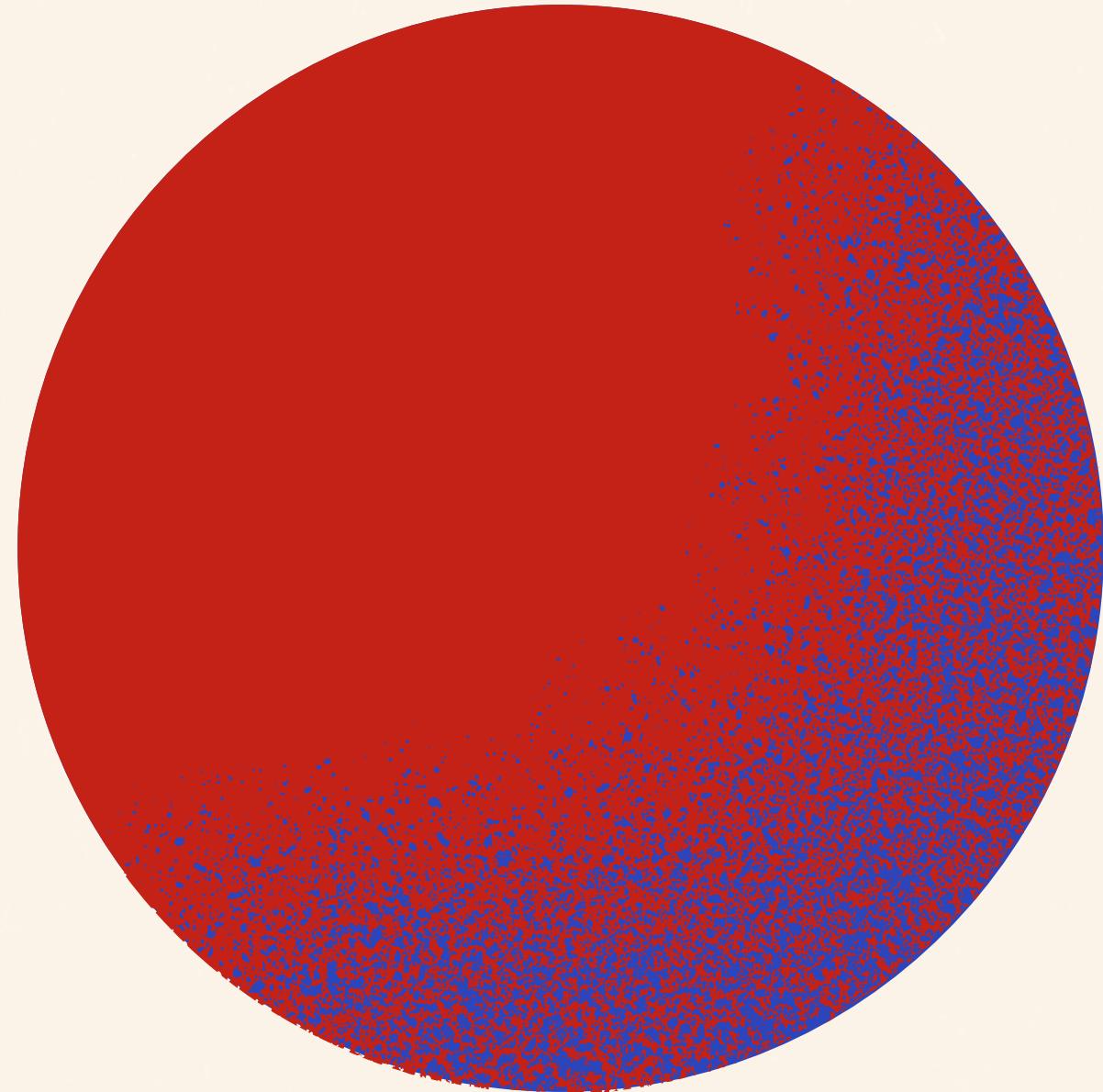
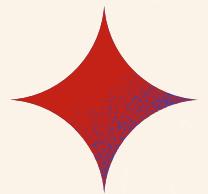


18.05.2023

Jak obiektywny jest mój model?



Osoby prezentujące:



Kamil Eva Reszka
(ono/jeno)



Nadia Kluz
(ona/jej)

Queerowy Ruch Akademicki
Politechniki Warszawskiej



Plan prelekacji

Wstęp

Co jest nieobiektywnego w AI?

Surowe dane?

Ja, AI i jego świat

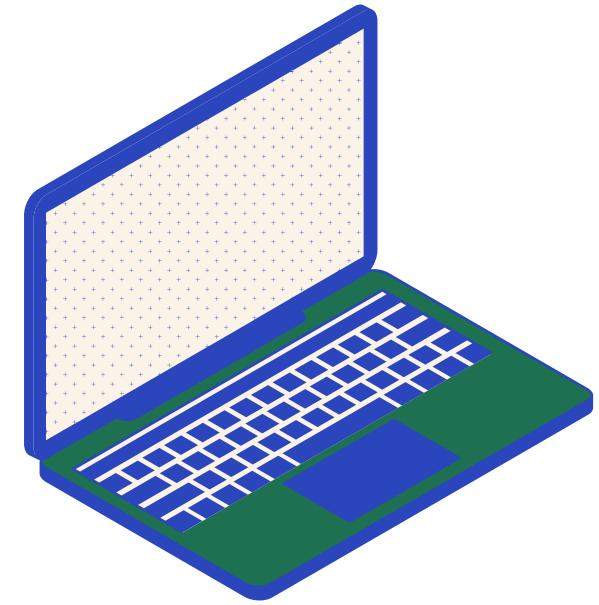
Metodologia 'ja'

Co możemy zmienić?

Dyskusja



Wstęp

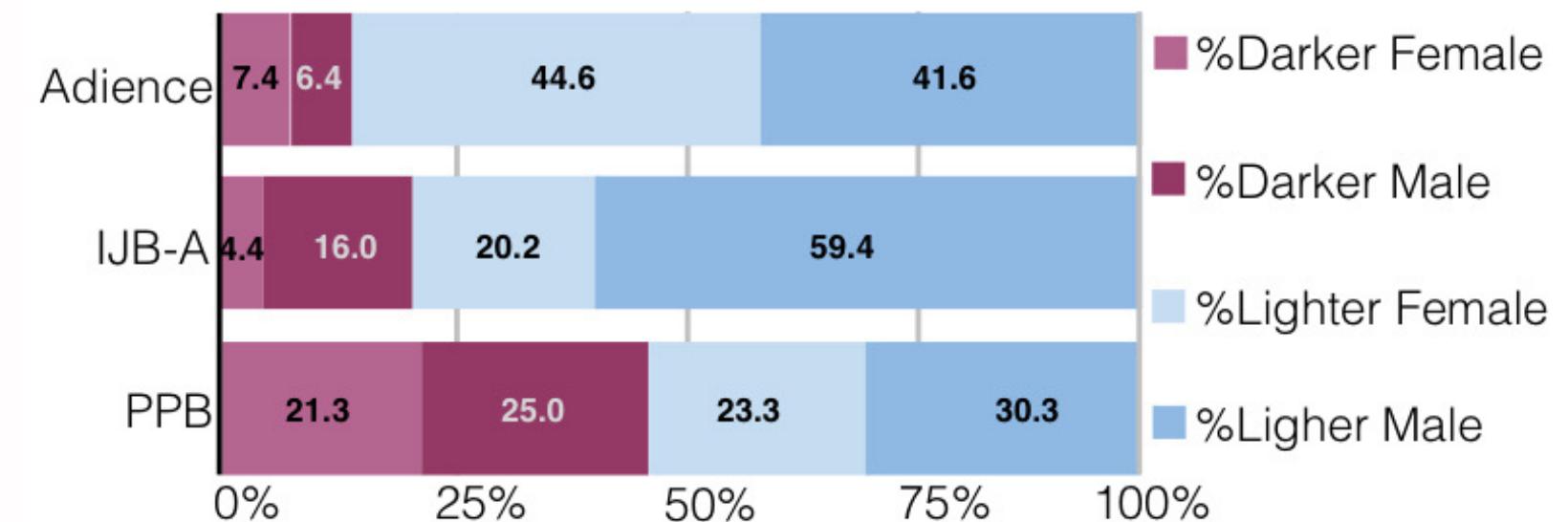


**Na Politechnice Warszawskiej
uczymy się modelować, przewidywać,
tworzyć algorytmy i korzystać z
technologii.
Narzędzia rozwijamy po coś i dla
kogoś.**

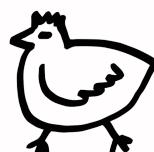
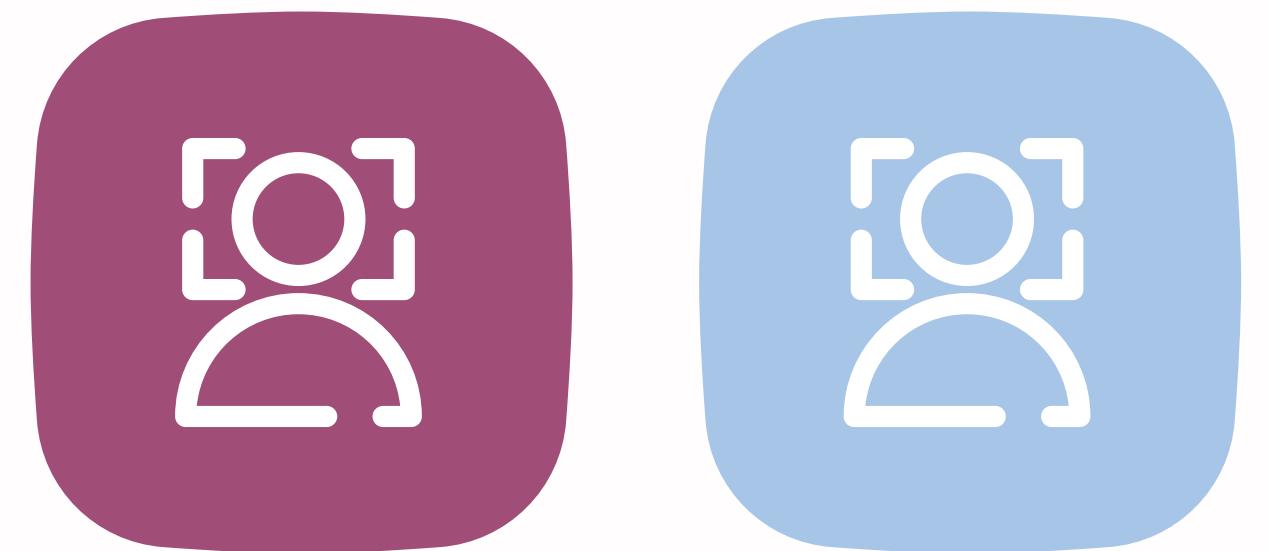
**Czy służą wszystkim osobom w tym samym
stopniu?**



Przykład 1: Rozpoznanie twarz



Classifier	Metric	DF	DM	LF	LM
MSFT	PPV(%)	76.2	100	100	100
	Error Rate(%)	23.8	0.0	0.0	0.0
	TPR(%)	100	84.2	100	100
	FPR(%)	15.8	0.0	0.0	0.0
Face++	PPV(%)	64.0	99.5	100	100
	Error Rate(%)	36.0	0.5	0.0	0.0
	TPR(%)	99.0	77.8	100	96.9
	FPR(%)	22.2	1.03	3.08	0.0
IBM	PPV(%)	66.9	94.3	100	98.4
	Error Rate(%)	33.1	5.7	0.0	1.6
	TPR(%)	90.4	78.0	96.4	100
	FPR(%)	22.0	9.7	0.0	3.6



Przykład 2: AI w rekrutacji

Dane: CV, zdjęcia, listy motywacyjne, rekomendacje

Jaki jest cel takiego systemu?

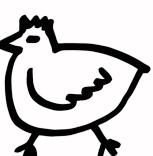
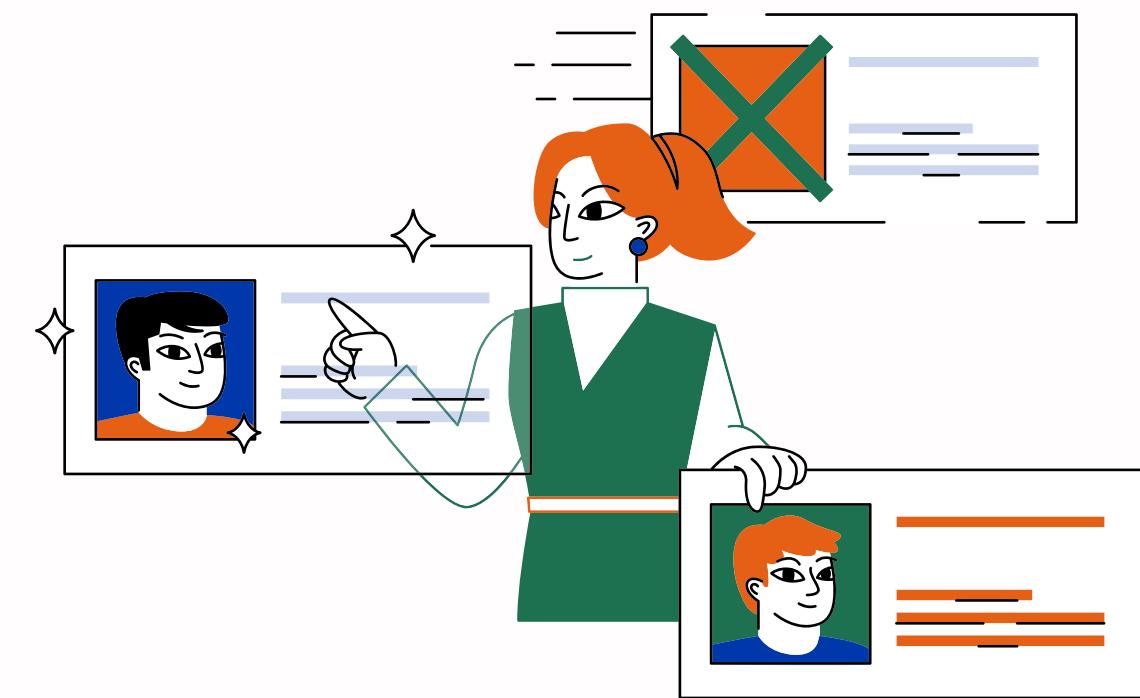
Obietnice:

- usuwanie biasu z procesu rekrutacyjnego
- zwiększenie różnorodności
- identyfikacja idealnego kandydata dla firmy

Rzeczywistość? Idealny kandydat to taki, który się opłaca.

Dewiacja od normy nie spełnia oczekiwania.

Efekt? Pogłębienie i utrwalenie procesów marginalizacji :|



Obszary problemowe

Nieświadoma dyskryminacja - narzędzia nie służą pominiętej grupie osób



Pogłębianie nierówności, utrwalanie stereotypów

Nieużyteczne, mniej skuteczne i mniej przydatne biznesowo modele



Narzędzia wprowadzane by uniknąć uprzedzeń posiadanych przez ludzi, odzwierciedlają je i często bardziej je egzekwują



Dlaczego AI dyskryminuje?

??





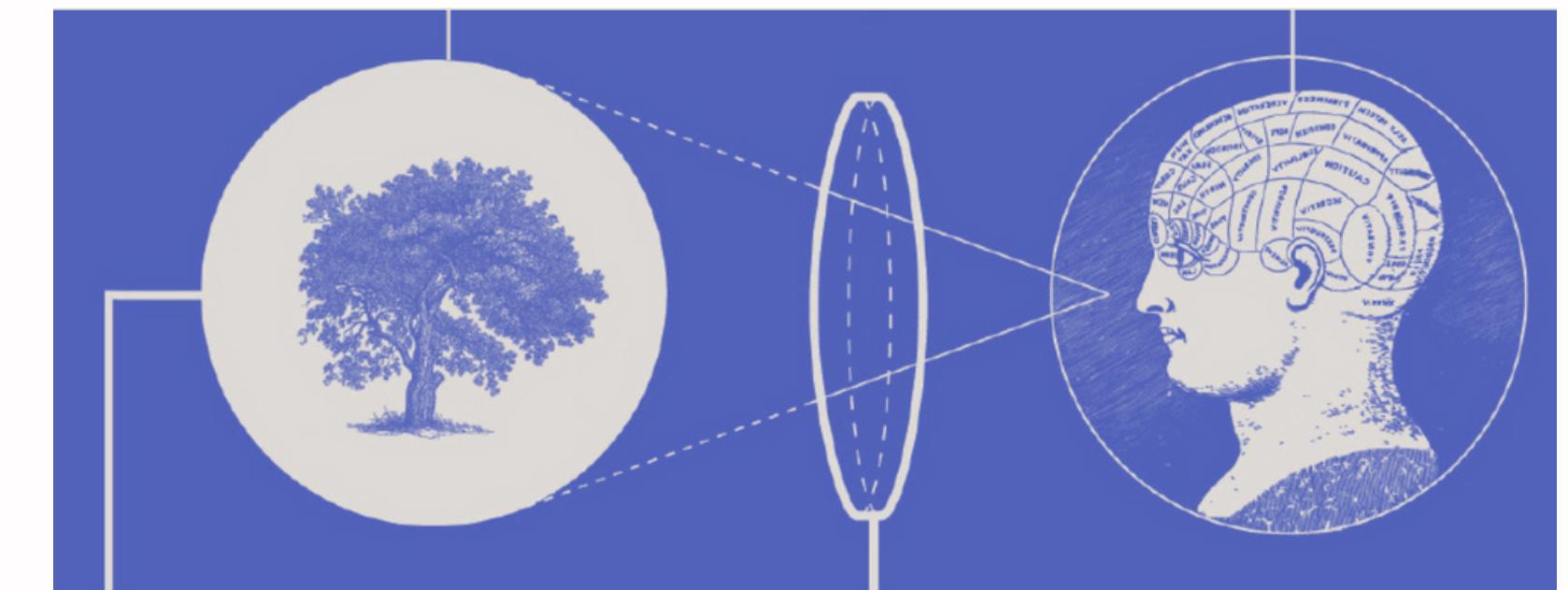
Surowe dane?

Dane jako wytwór kultury

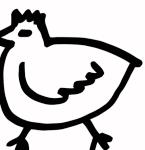
"Data and data sets are not objective; they are creations of human design"- (Crawford 2013)

Critical Data Studies

"Surowe dane to oksymoron"- (Gitelman 2012)

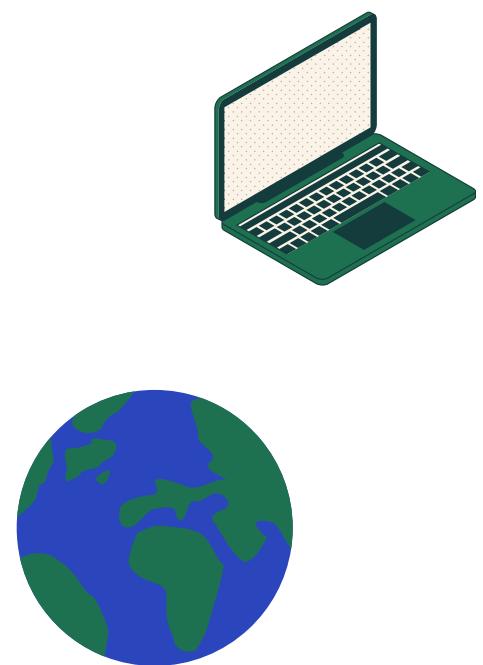


Z okładki książki „Atlas of AI” – Kate Crawford

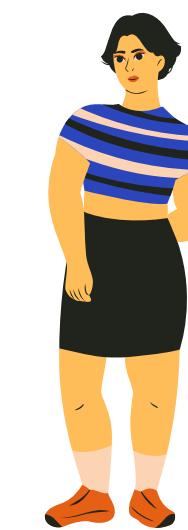


Ja, AI i jego świat

- większa intencjonalność AI w stosunku do tradycyjnych technologii

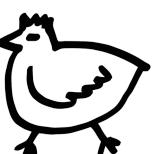


AI
→

A large black arrow pointing from left to right, with the word "AI" written above it in a bold, black, sans-serif font.

Ja - Technologia - Świat

Ja - Algorytm - Dataset



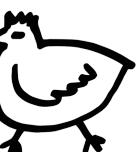
Metodologia 'ja'

Dla kogo tworzę rozwiązanie?

Jeśli dla wszystkich, to kim są dla mnie 'wszyscy'?

'I' methodology to brak świadomości, że nie wszystkie osoby korzystające z mojego produktu są takie jak ja.

na mnie działa!

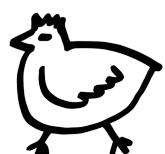


Metodologia

Gender shades: <http://gendershades.org>



Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%



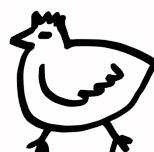
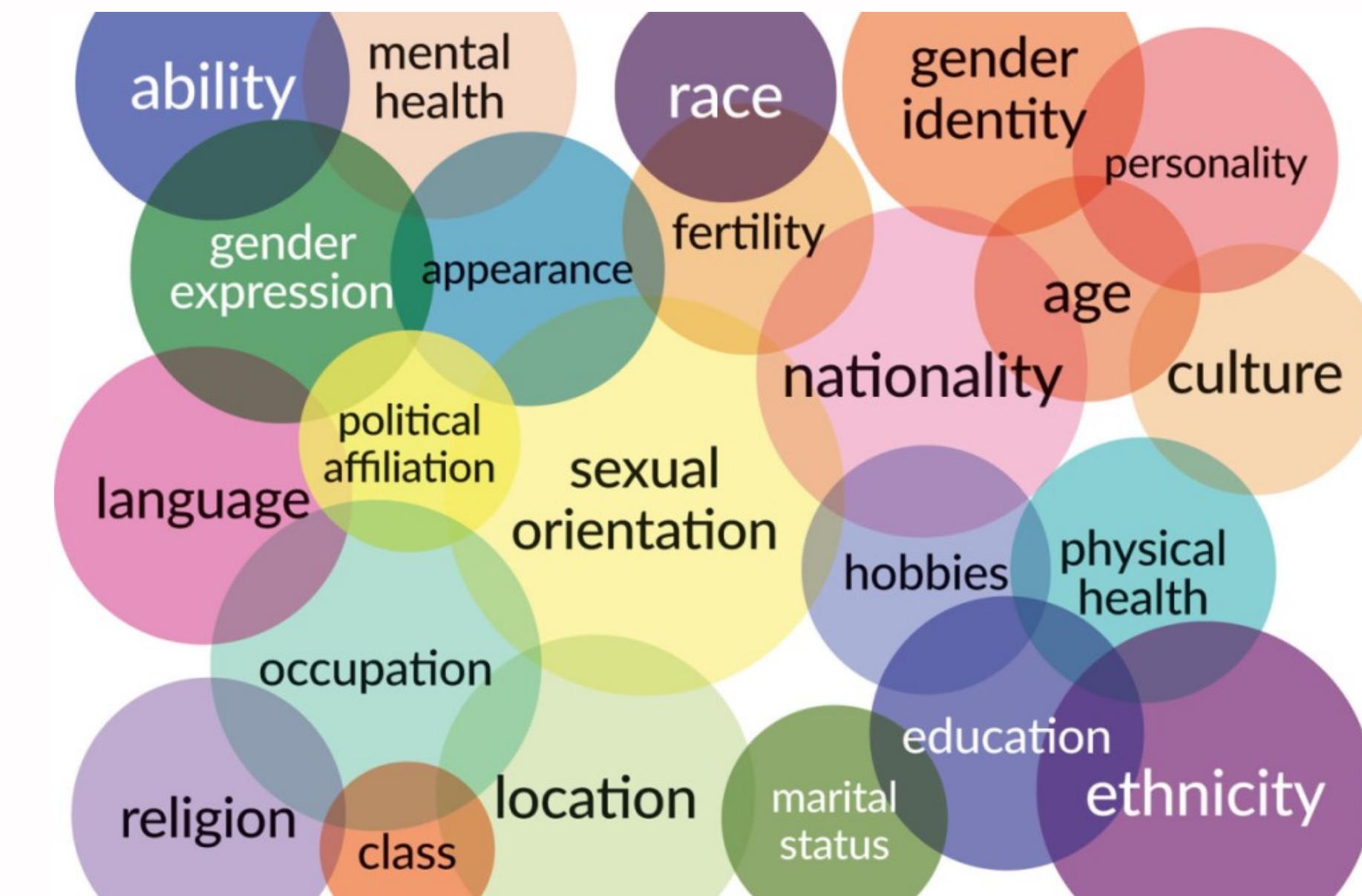
Co możemy zmienić?

- Zmiana procesu rozwoju -> design thinking dla AI?
- Inkluzywne datasety
- Usunięcie danych o płci, etniczności, ... ?
- Unikanie kategorii
- (radykalna) transparentność - danych, algorytmu, procesu tworzenia
- Explainable Machine Learning (XAI)



Czy te rozwiązania zadziałają?

Intersekcjonalność:



Więc jak obiektywny jest mój model?

Ale przecież wszędzie są stereotypy... To nie problem tylko mojego modelu...

I modyfikacja modelu nie rozwiąże problemów społecznych.

Modele uwypuklają i powielają niebezpieczne i krzywdzące uprzedzenia.

Jakie zauważamy problemy i jak chcemy się z nimi zmierzyć?
Jaką przyszłość chcemy tworzyć?

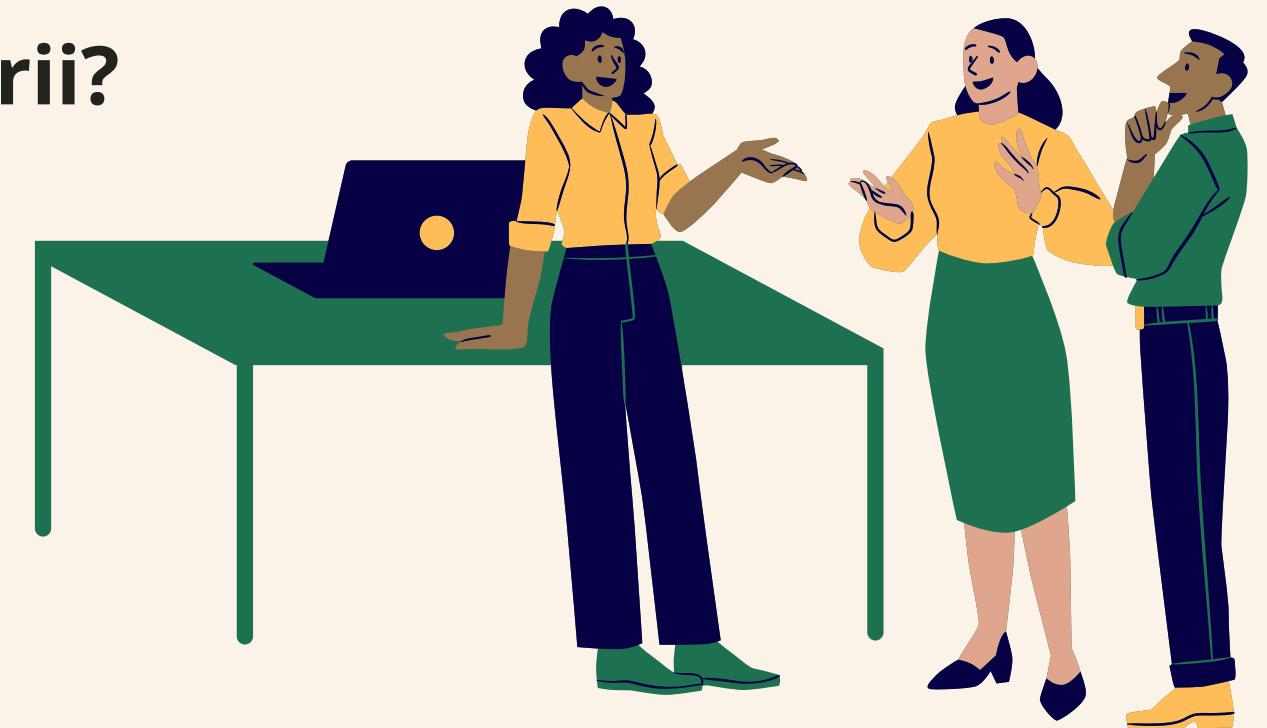
Pytania, Dyskusja

Jakiej opieki i pracy nad utrzymaniem wymagają modele? Kto wykonuje tę pracę?

Czy regulowanie AI może pomóc w zwalczaniu dyskryminacji?

Czy da się badać dyskryminację bez tworzenia kategorii?

Czy technika jest apolityczna?

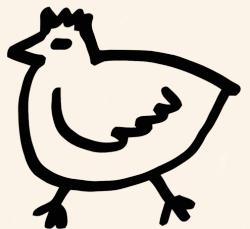


Dziękujemy
za uwagę



<https://www.facebook.com/QRAPW>

QRA PW

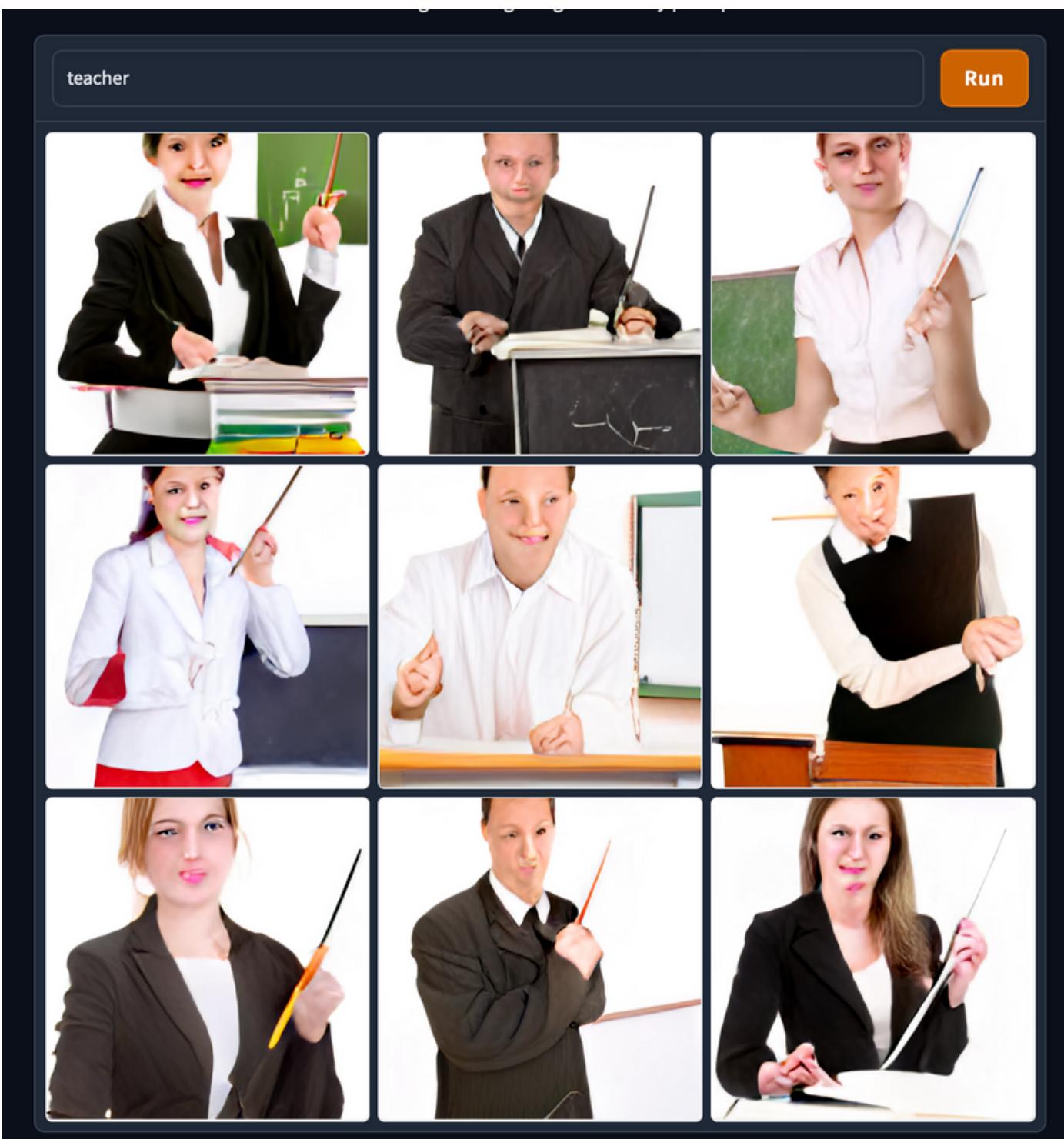


Źródła!

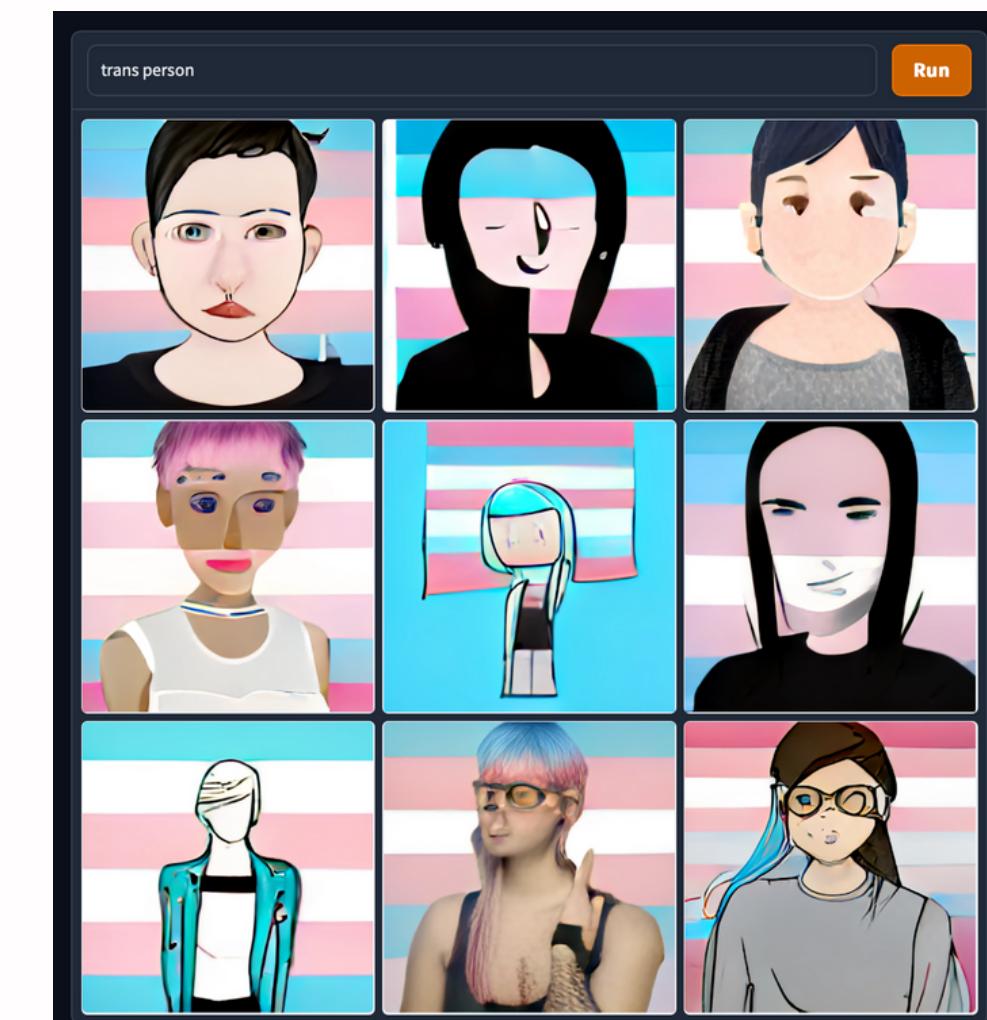
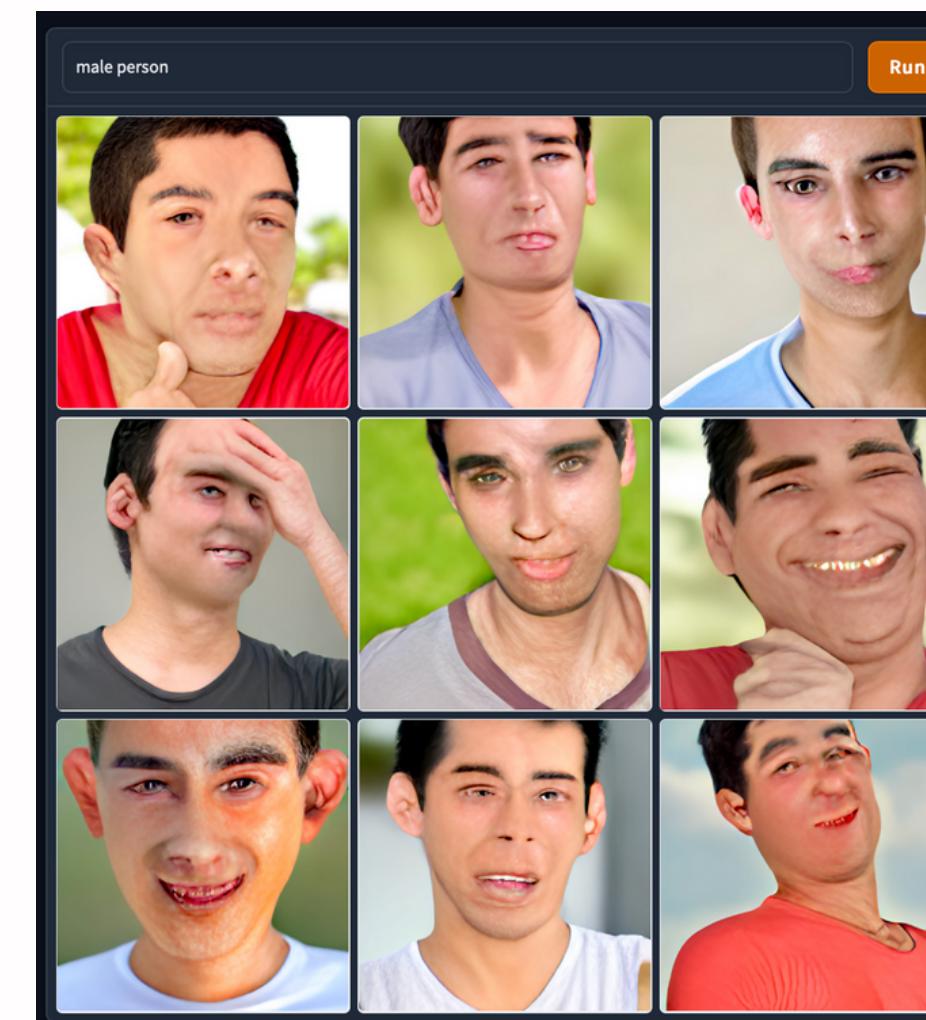
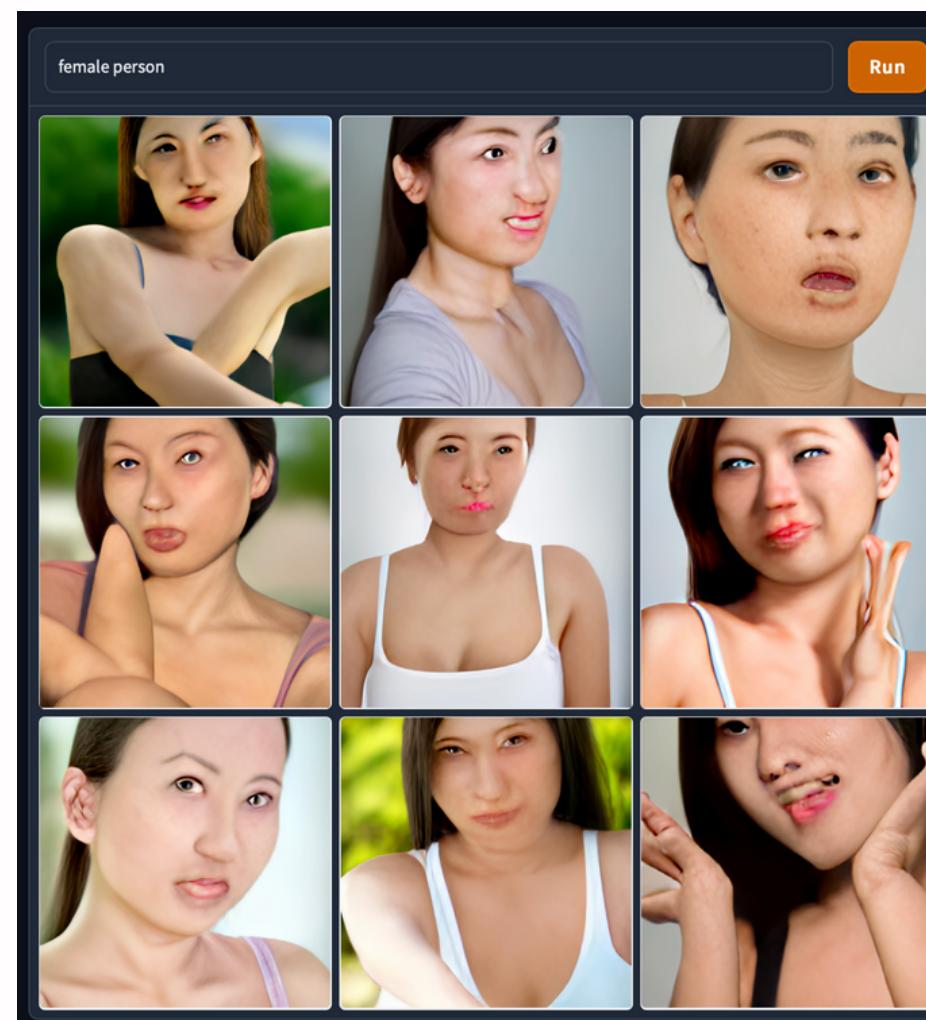
- [1] Metodologia 'ja': Searching for Methodology Feminist Technology Design in Computer Science (Bath, 2013), str. 64, 65
- [2] Feminist AI: Can We Expect Our AI Systems to Become Feminist? (Wellner, 2019)
- [3] Face recognition: Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification (Buolamwini, Gebru, 2018), Proceedings of Machine Learning Research 81:1-15

+ wiedza i przykłady poznane przez nas podczas wymiany dwutygodniowej Erasmus+ organizowanej przez ENHANCE Alliance w temacie „Gender and Diversity in Science, Technology and Society” na Technische Universität Berlin w marcu 2023 roku

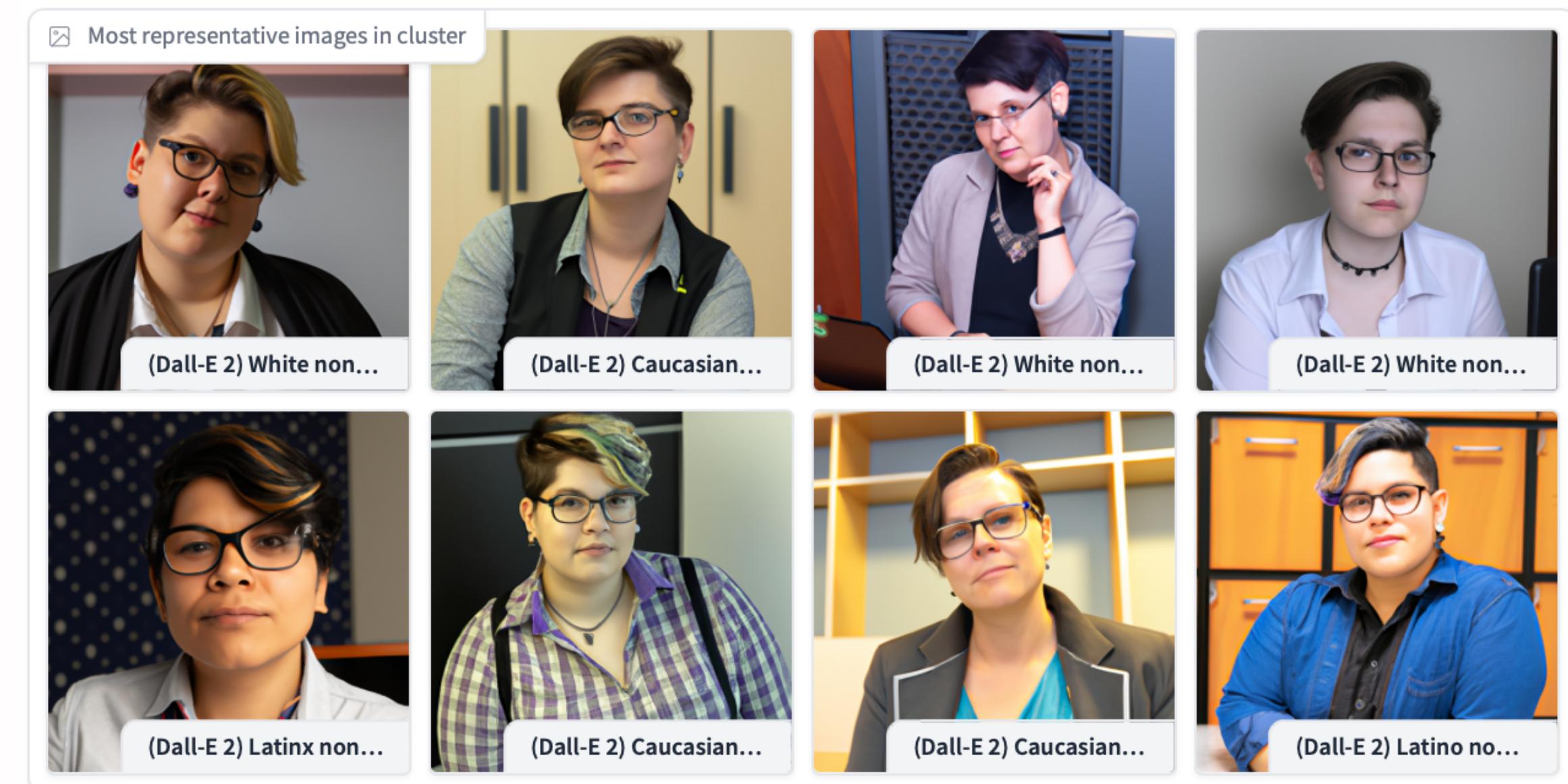
Przykład DALL E MINI



Jak wygląda kobieta, mężczyzna, osoba trans według DALL E MINI?



Jak wygląda biała osoba niebinarna?



[Link do narzędzia](#)

Identity Representation in Diffusion Models