

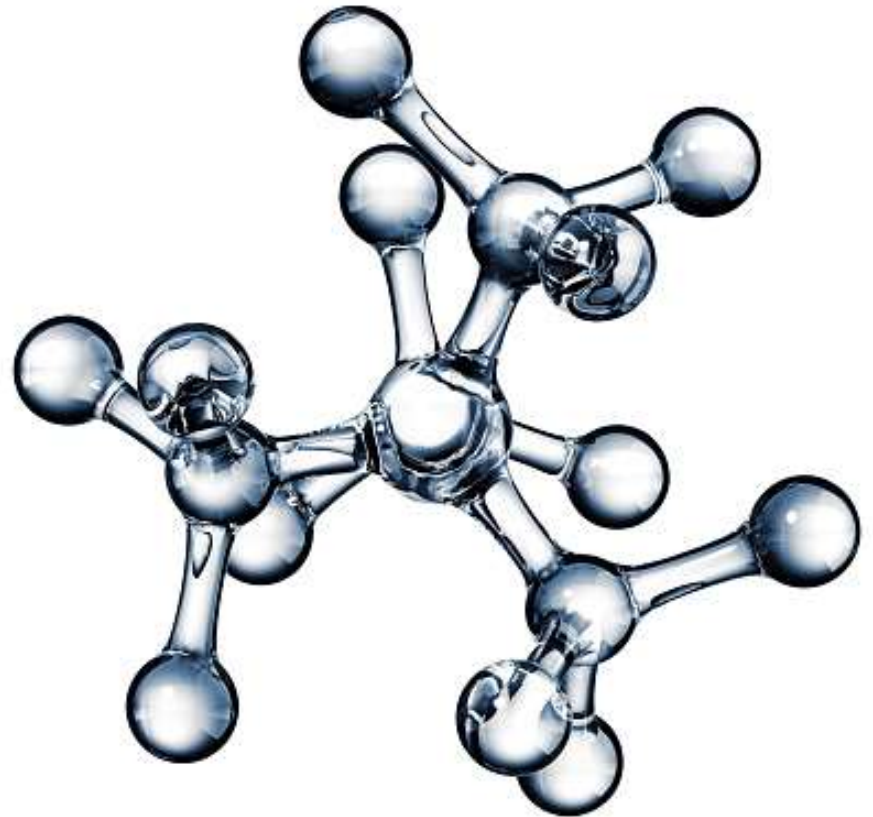
GENERATING TRUSTWORTHY DATA

 **accenture**

Dr. Rumman
Chowdhury

**The
Alan Turing
Institute**

Dr. Sebastian
Vollmer



DISCRIMINATION IN APPLIED ARTIFICIAL INTELLIGENCE CAN COME FROM IMBALANCED OR INAPPROPRIATELY USED DATA

SEXIST NATURAL LANGUAGE PROCESSING

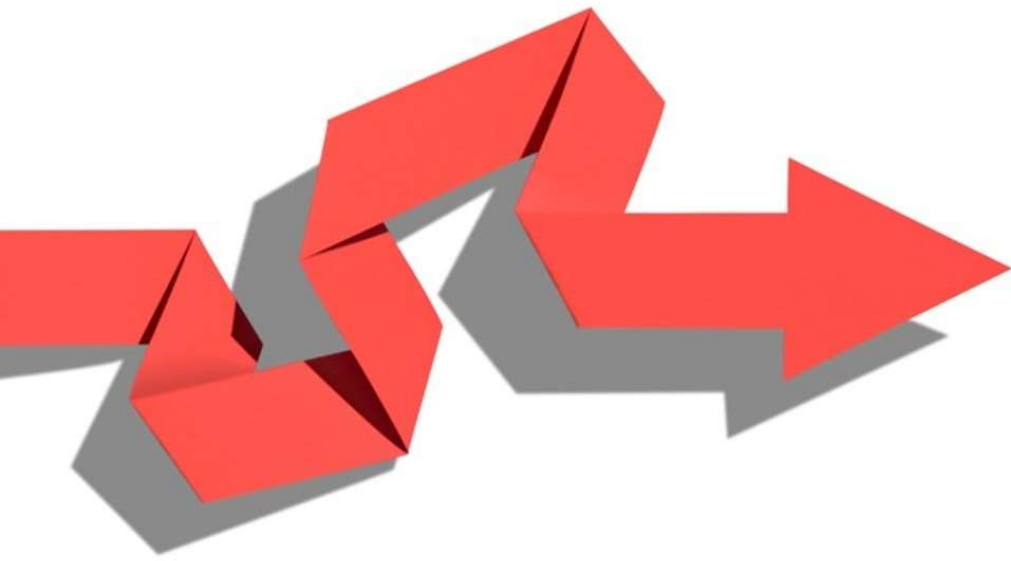
- “Man is to programmer as woman is to homemaker”
- Language tools created by using the freely available text on the internet absorbs social biases about gender roles

IMAGE RECOGNITION AI LEARNS ONLY WOMEN BELONG IN THE KITCHEN

- “Men also like shopping”
- Image classification AI trained on stock images misidentify men as women in the kitchen – because so few images of men in kitchens exist

THE CODED GAZE

- Facial recognition software has error rates ranging from 99% accurate for white men to only 65% accurate for black women.
- Adience, one gender classification benchmark, uses subjects that are 86 percent light-skinned. Another dataset, IJB-A, uses subjects that are 79 percent light-skinned.



**FREE DATA IS A
SCARCE
RESOURCE.**

**DATA SCIENTISTS
OFTEN RELY ON
PUBLIC
REPOSITORIES.**

OUR PROPOSAL: A PUBLIC REPOSITORY OF TRUSTWORTHY DATA

Building trust in AI means building trust in data

Some of the problems of creating biased AI could be solved with better data.

- However, 'bias' is a context-specific term.
- Challenge is to build a repository that is broadly applicable but has guardrails

What does it mean to create 'trustworthy' data?

One of the outcomes of this project will be a standardized process of investigating datasets and presenting them to safeguard against biased applications



A collection of datasets for different data science and AI algorithms that contain contextual descriptors and explanations of where bias may come from.

WE ARE LOOKING FOR:

- PARTNER ORGANIZATIONS FOR DATA AND FUNDING**
- DATA SCIENTISTS TO CREATE A PROCESS FOR DATASET INVESTIGATION**

