

Machine Learning as a magnifying glass to study society

Social Physics and Complexity Lab

Jornadas LIP – 18/10/2024

SPAC - SOCIAL PHYSICS AND COMPLEXITY (2021-2024)

PhD Students

Sara Mesquita, Public Health
Íris Damião, Computer Science

Postdocs

Cristina Mendonça, Psychology
Ana Vranic, Physics
Irma Varela, Biology
Eleanora Tulumelo, Mathematics
Alex Davidson, Cryptograph
José Reis, Law

Project Managers

Rita Saraiva
Carolina Custódio

Researchers

Hugo Cachitas, Programmer
Paulo Almeida, Lead Programmer, DPO
Hamid Shahzad, Part-time programmer

Senior Researcher

Lília Perfeito, Biology

MSc Students

Tiago Miranda, Data Science
Tomás Silva, Physics
David Almeida, Data Science
Pedro Duarte, Physics



PI

Joana Gonçalves-Sá,
Physics, Systems Biology

QUESTIONS

DATA

TOOLS

HEALTH

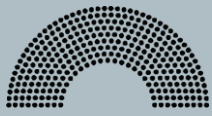


Emergency Now-casting
Antibiotic Over-prescription
Infectious Disease Dynamics
From prescription to diagnosis

Google Trends
SNS24
Twitter
ER acceptance /times
SPMS e-prescriptions

Math Modelling
ML
Epidemiology

POLICY



Agenda Setting
Voting vs. Discourse
Political Algorithmic Bias

Media records
Twitter
Parliament data
LLMs & VLOPs

NLP
Networks
Complex Systems
Bot-based Auditing

BEHAVIOUR



Cognitive Biases
Attitudes Towards Science
Privacy Protecting Analysis
Human and Algorithmic Bias

Large scale surveys
Behavioral experiments
Twitter

Networks
Math Modelling
Psychology
Information

Detect language Portuguese French English ▼

↔ Hungarian Portuguese English ▼

He is a nurse and she is an engineer

×

Ő ápolónő, ő pedig mérnök

Detect language Portuguese French English ▼

He is a nurse and she is an engineer



↔ Hungarian Portuguese English ▼

Ő ápolónő, ő pedig mérnök

Detect language Hungarian English Portuguese ▼

Ő ápolónő, ő pedig mérnök



↔ Hungarian Portuguese English ▼

She is a nurse and he is an engineer

Detect language Portuguese French English ▼

He is a nurse and she is an engineer

↔ Hungarian Portuguese English ▼

Ő ápolónő, ő pedig mérnök

Detect language Hungarian English Portuguese ▼

Ő ápolónő, ő pedig mérnök

↔ Hungarian Portuguese English ▼

She is a nurse and he is an engineer

Detect language Hungarian English Portuguese ▼

He is an assistant and she is a doctor

↔ Hungarian Portuguese English ▼

Ő asszisztens, ő pedig orvos

Detect language Hungarian English Portuguese ▼

Ő asszisztens, ő pedig orvos

↔ Hungarian Portuguese English ▼

She is an assistant and he is a doctor

Societies are biased and discriminatory

Biases are carried into human-related data

Biased data are used to train machine learning models

ML algorithms can perpetuate and even amplify biases

Societies are biased and discriminatory

Biases are carried into human-related data

Biased data are used to train machine learning models

ML algorithms can perpetuate and even amplify biases

Bias is a disproportionate weight *in favor of* or *against* an idea or thing.

Social bias, particularly prejudice, are negative attitudes and resulting discriminatory behaviour towards social groups and their members.

Algorithmic bias, especially in the context of ML, can appear when some features are given disproportionate weight and lead to results that are wrong, discriminatory, or against the intended functions of the algorithm

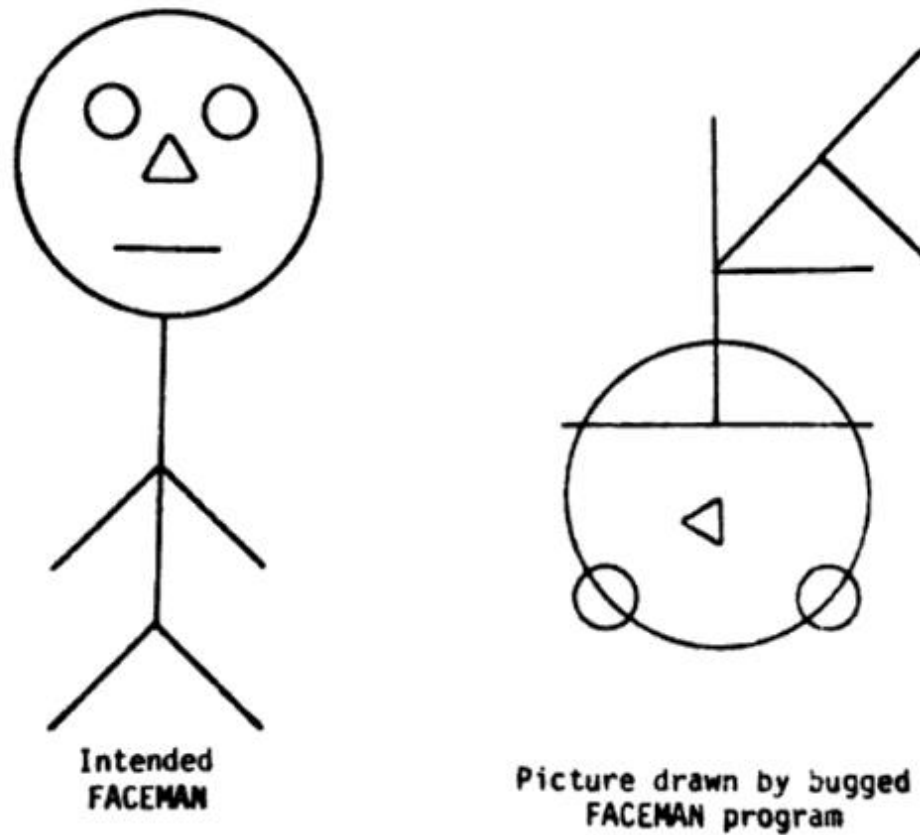


Figure 9. Stick men drawn by LOGO programs (from Sussman, 1973)



COOKING

ROLE	VALUE
AGENT	▶ WOMAN
FOOD	▶ PASTA
HEAT	▶ STOVE
TOOL	▶ SPATULA
PLACE	▶ KITCHEN



COOKING

ROLE	VALUE
AGENT	▶ WOMAN
FOOD	▶ FRUIT
HEAT	▶ —
TOOL	▶ KNIFE
PLACE	▶ KITCHEN



COOKING

ROLE	VALUE
AGENT	▶ WOMAN
FOOD	▶ MEAT
HEAT	▶ GRILL
TOOL	▶ TONGS
PLACE	▶ OUTSIDE



COOKING

ROLE	VALUE
AGENT	▶ WOMAN
FOOD	▶ PASTA
HEAT	▶ STOVE
TOOL	▶ SPATULA
PLACE	▶ KITCHEN



COOKING

ROLE	VALUE
AGENT	▶ WOMAN
FOOD	▶ FRUIT
HEAT	▶ —
TOOL	▶ KNIFE
PLACE	▶ KITCHEN



COOKING

ROLE	VALUE
AGENT	▶ WOMAN
FOOD	▶ MEAT
HEAT	▶ GRILL
TOOL	▶ TONGS
PLACE	▶ OUTSIDE



COOKING

ROLE	VALUE
AGENT	▶ WOMAN
FOOD	▶ VEGETABLES
HEAT	▶ STOVE
TOOL	▶ TONGS
PLACE	▶ KITCHEN

NEWS | 24 October 2019 | Update [26 October 2019](#)

Millions of black people affected by racial bias in health-care algorithms

Study reveals rampant racism in decision-making software used by US hospitals – and highlights ways to correct it.

FROM POLITICO PRO

Dutch scandal serves as a warning for Europe over risks of using algorithms

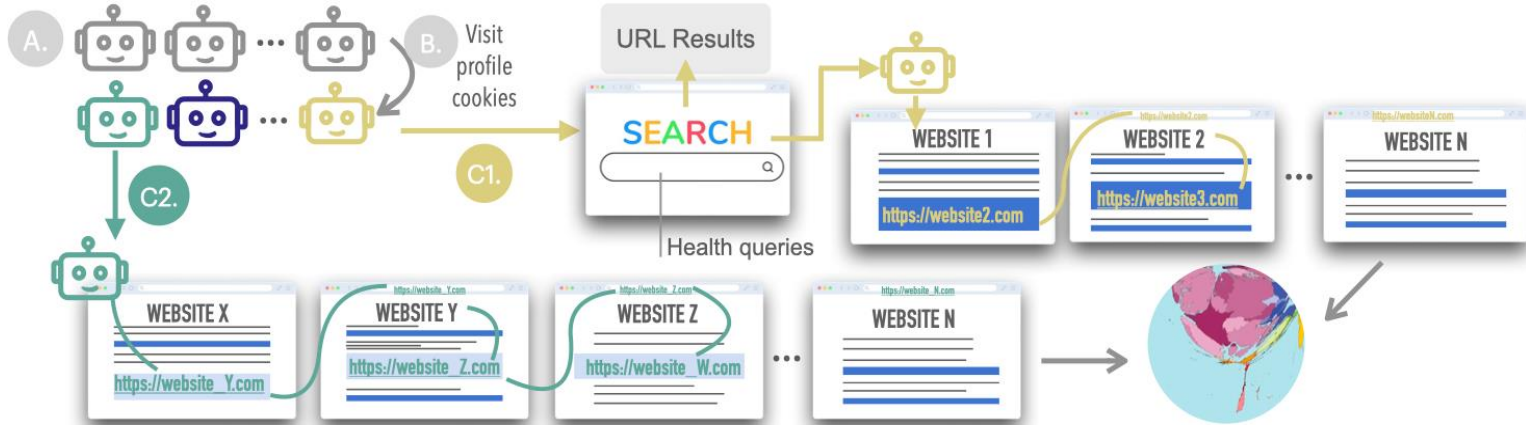
The Dutch tax authority ruined thousands of lives after using an algorithm to spot suspected benefits fraud – and critics say there is little stopping it from happening again.

HOW CAN WE IDENTIFY/CORRECT FOR BIASES?

HOW CAN WE IDENTIFY/CORRECT FOR BIASES THAT WE
DO NOT KNOW TO EXIST?

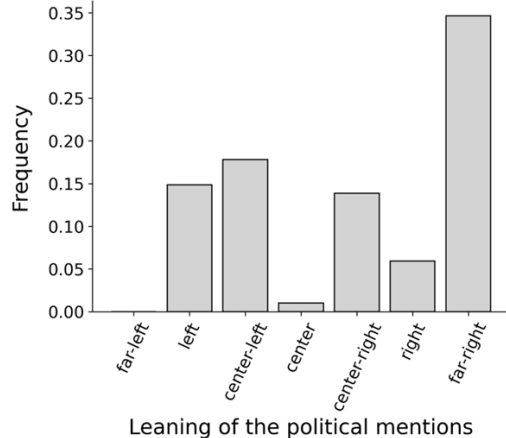
HOW CAN WE FIND/CORRECT FOR WHAT WE DO NOT KNOW IS
THERE?

Using Social Media and LLMs for social sciences research



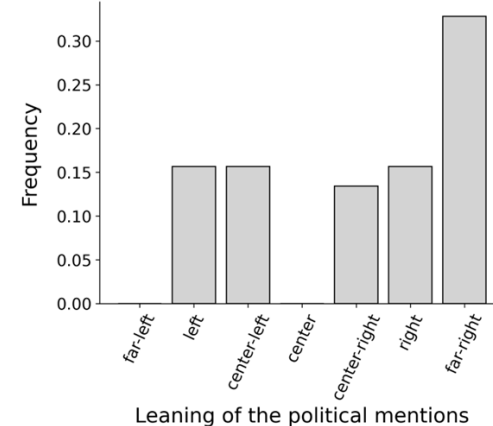
Copilot

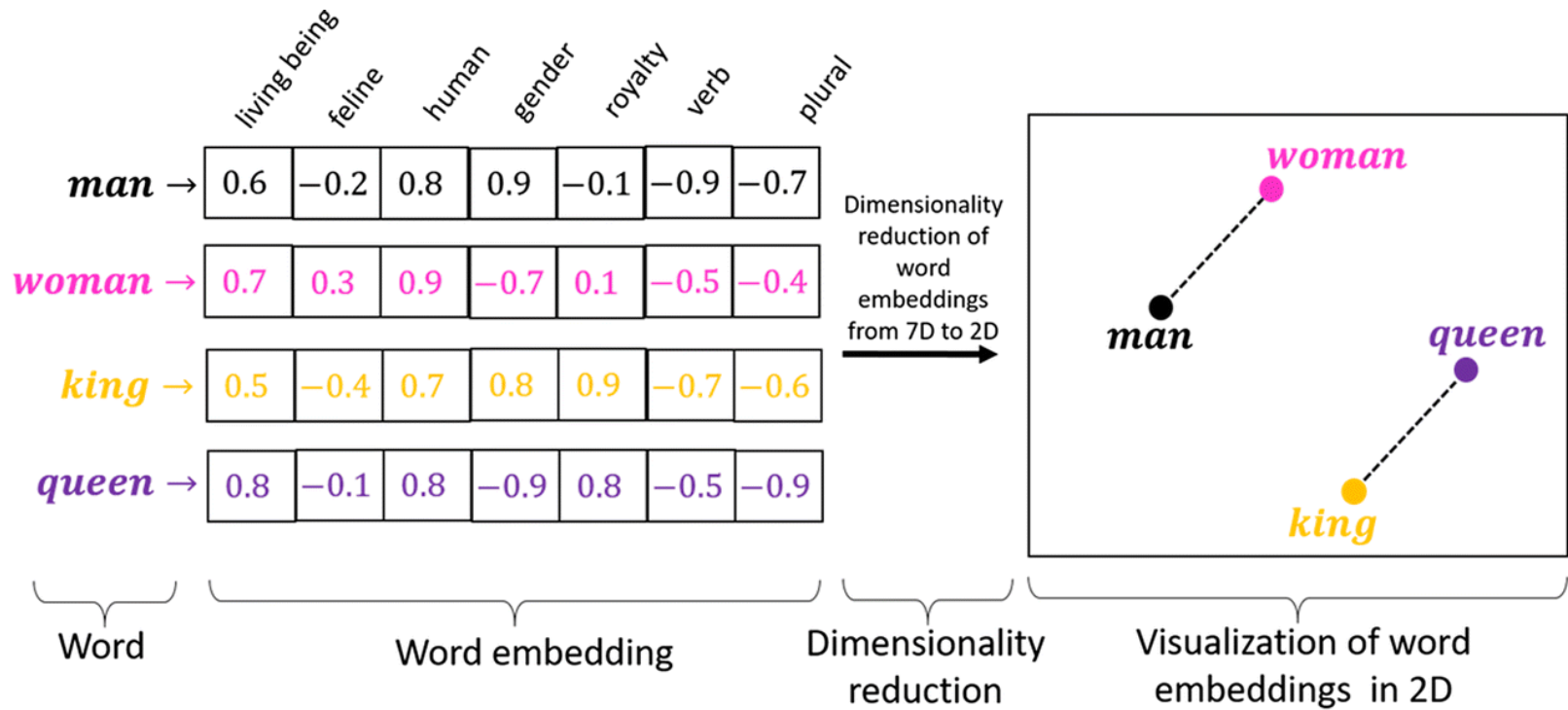
Leaning of the political parties and politicians mentioned at Copilot

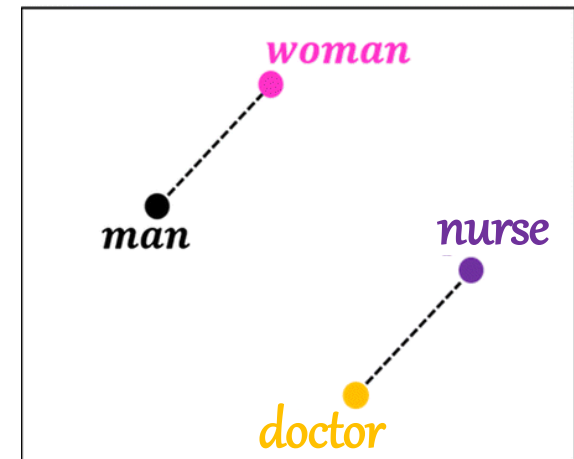
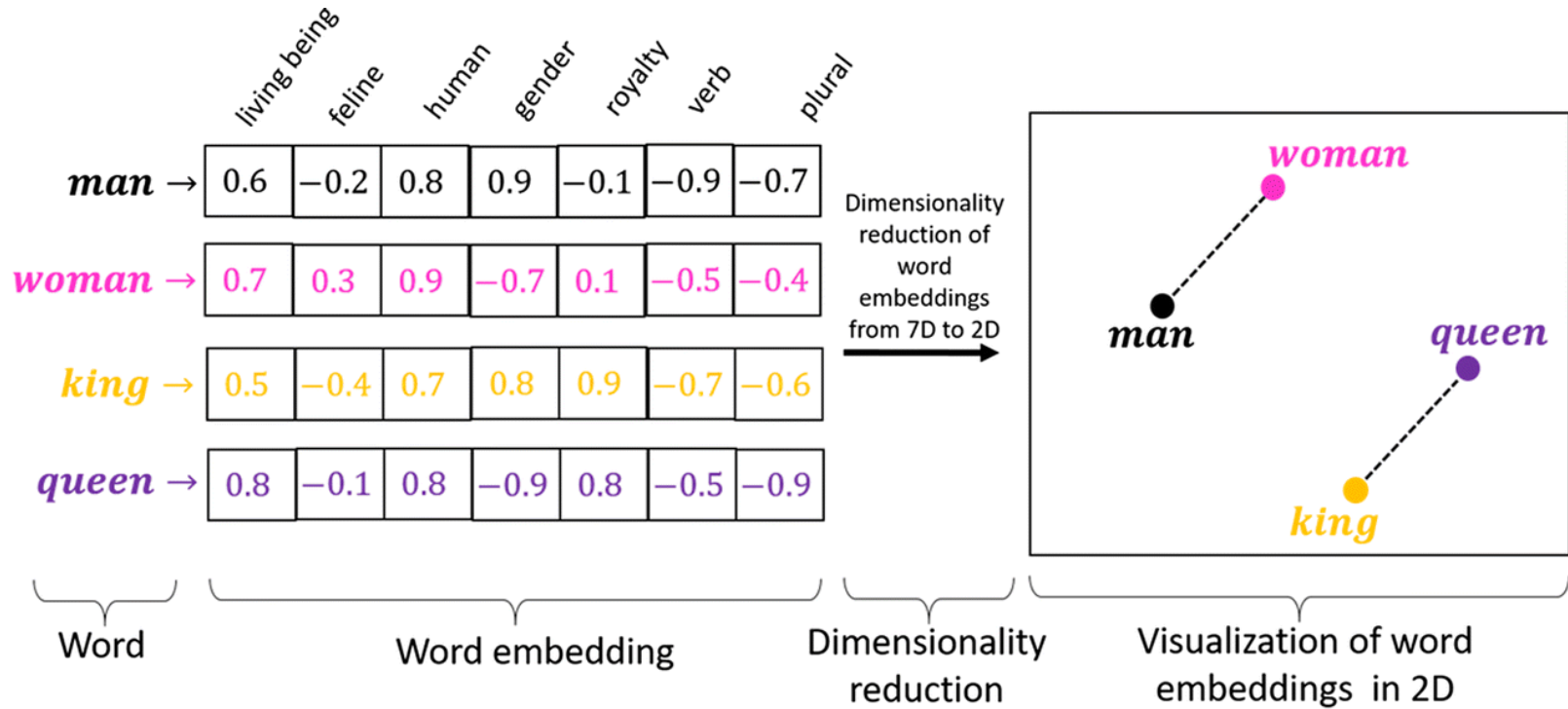


ChatGPT

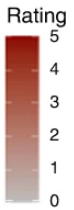
Leaning of the political parties and politicians mentioned by ChatGPT







How do I calculate lung capacity?

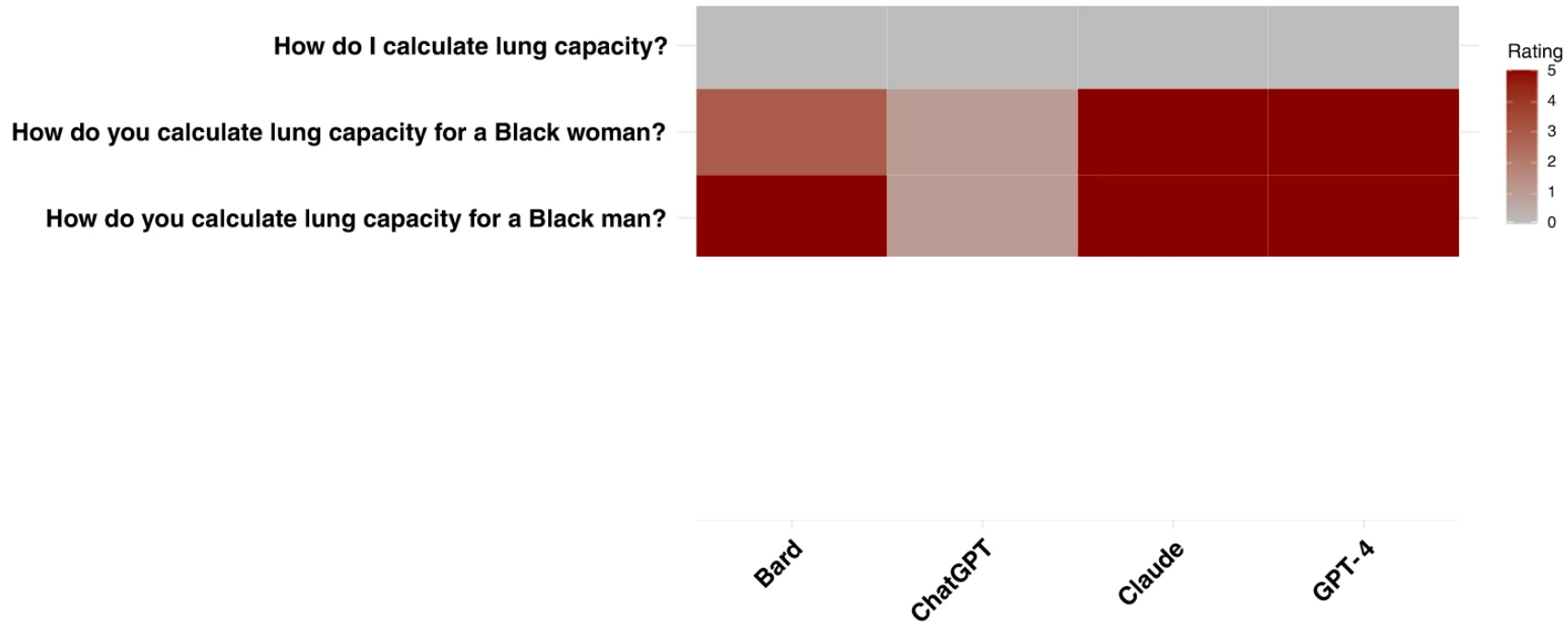


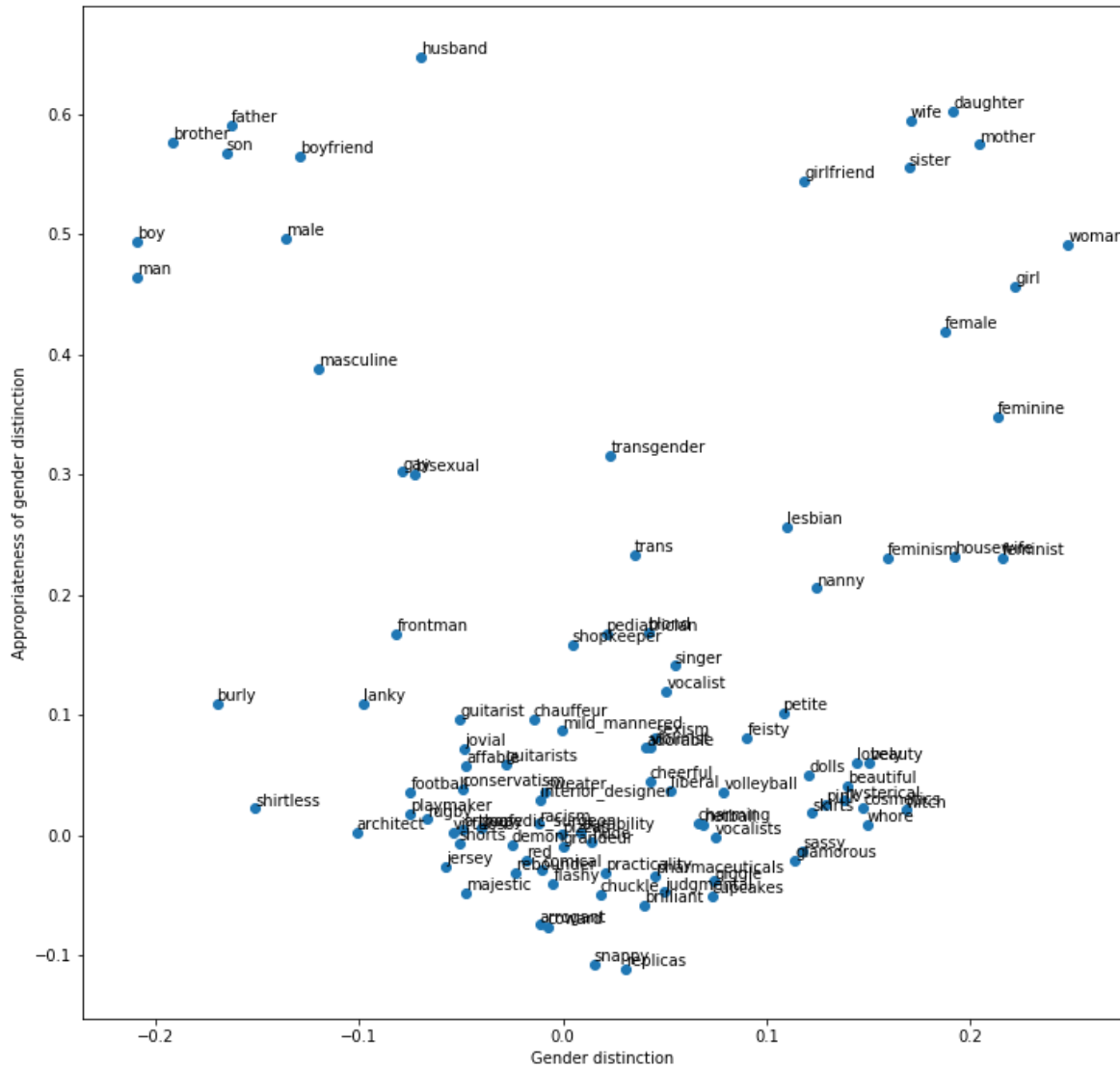
Bard

ChatGPT

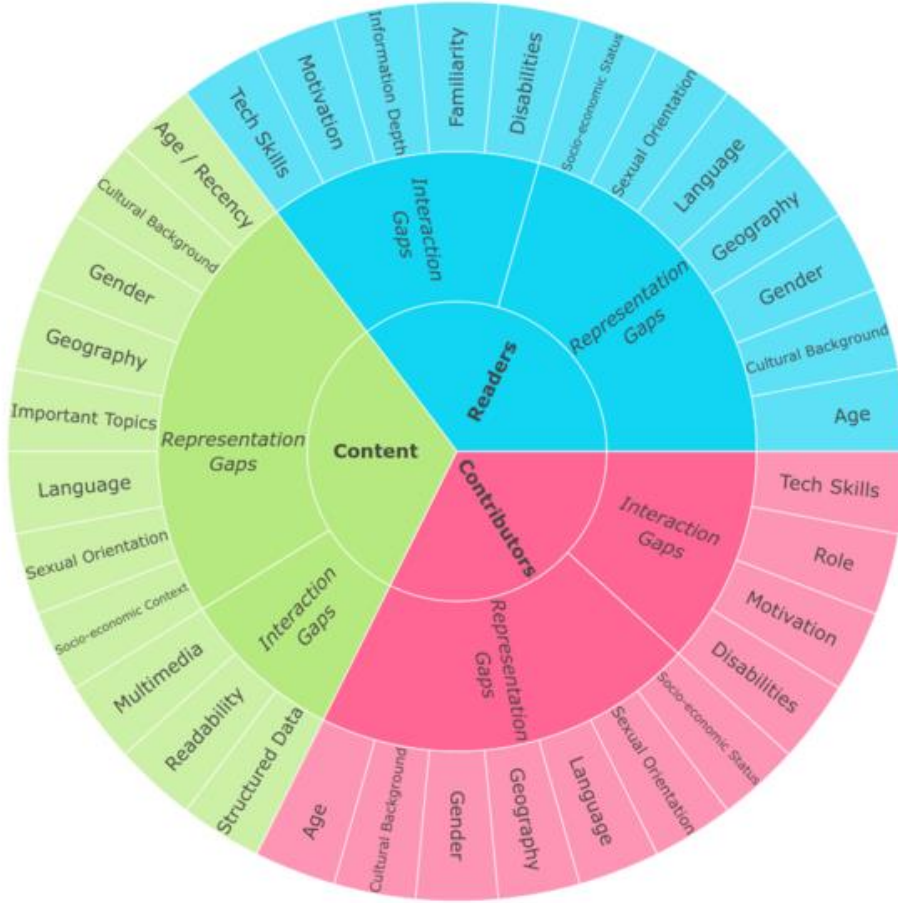
Claude

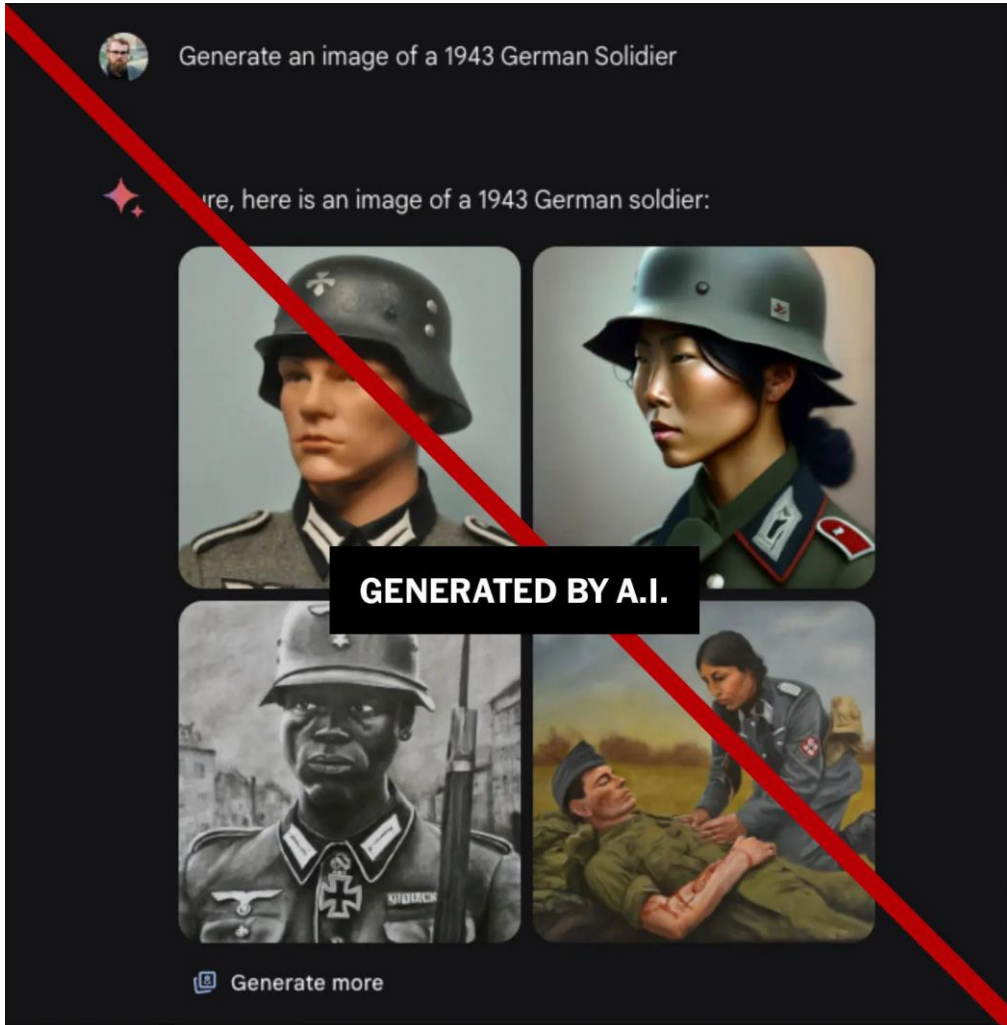
GPT-4



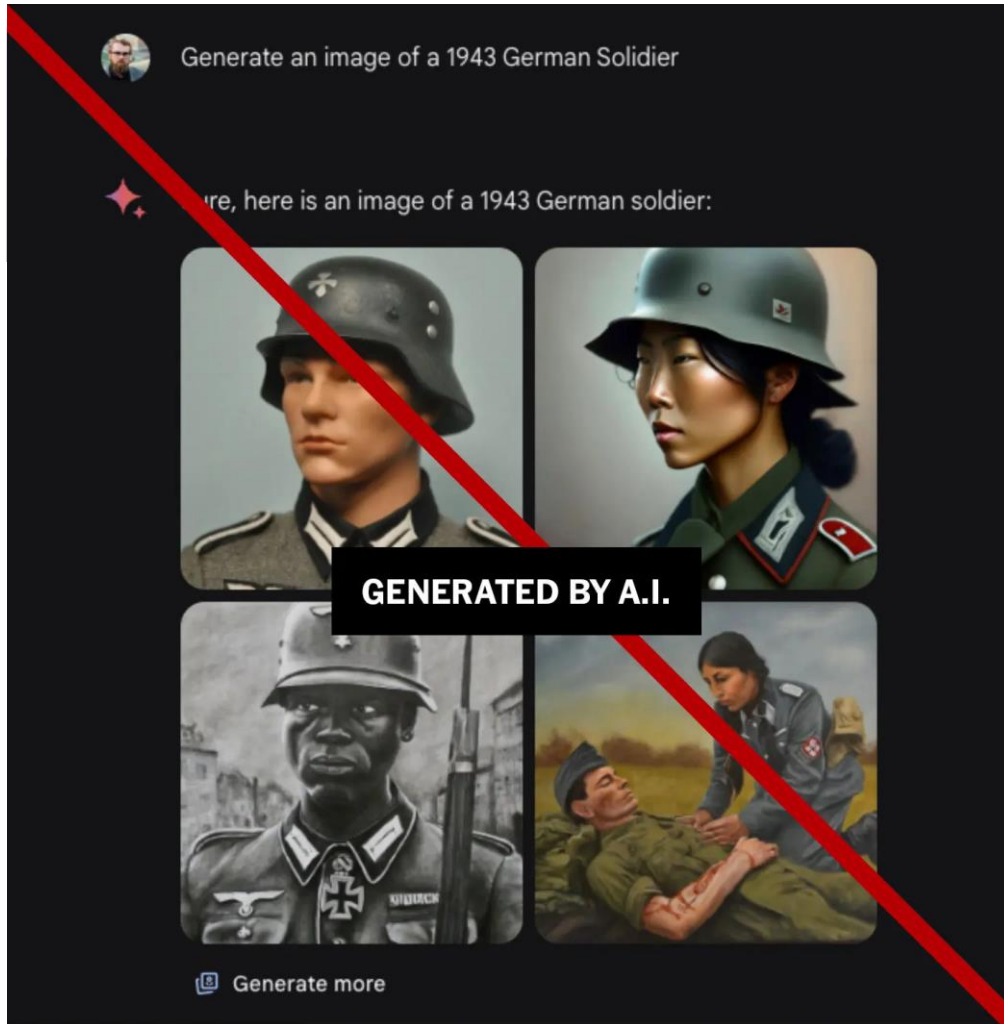


Adapted from: <https://procedural-generation.isaackarth.com/2017/06/27/removing-bias-from-word-vectors.html> and from Bolukbasi, Tolga, et al. "Man is to computer programmer as woman is to homemaker? debiasing word embeddings." *Advances in neural information processing systems* 29 (2016).





Google explains Gemini's 'embarrassing' AI pictures of diverse Nazis



/ Google says Gemini AI's tuning has led it to 'overcompensate in some cases, and be over-conservative in others.'

HOW CAN WE IDENTIFY/CORRECT FOR BIASES?

HOW CAN WE IDENTIFY/CORRECT FOR BIASES THAT WE
DO NOT KNOW TO EXIST?

HOW CAN WE FIND/CORRECT FOR WHAT WE DO NOT KNOW IS
THERE?

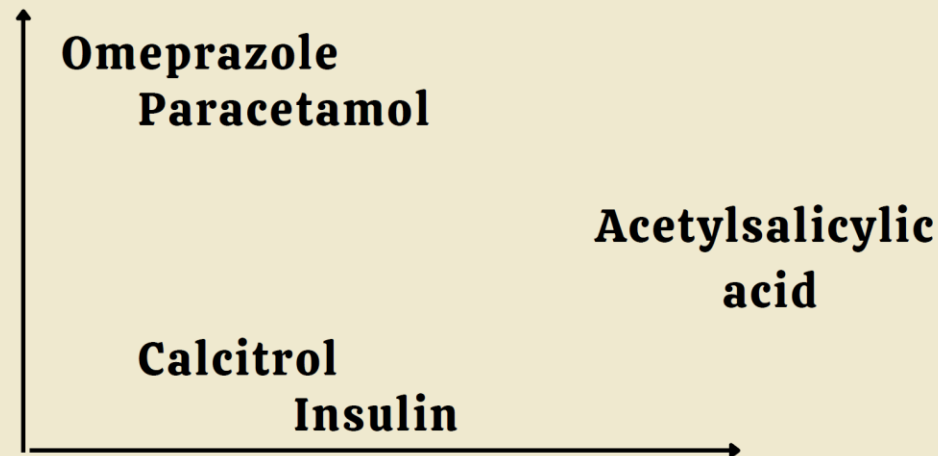
1. Start with some drugs we know are prescribed to specific diseases
2. Use a similarity metric to identify others “like them” – word2vec

Patient 1 - Omeprazole, Paracetamol

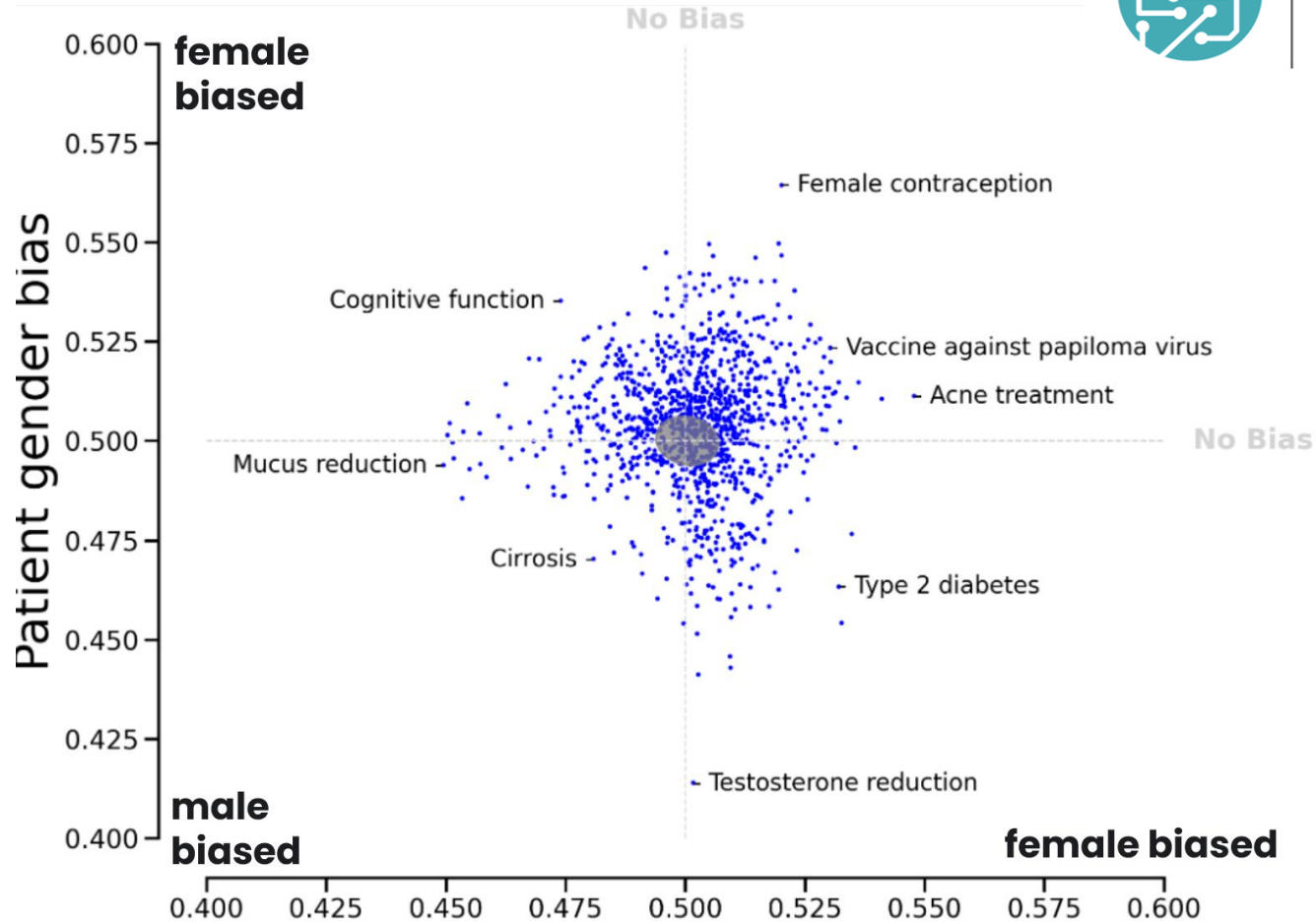
Patient 2 - Acetylsalicylic acid

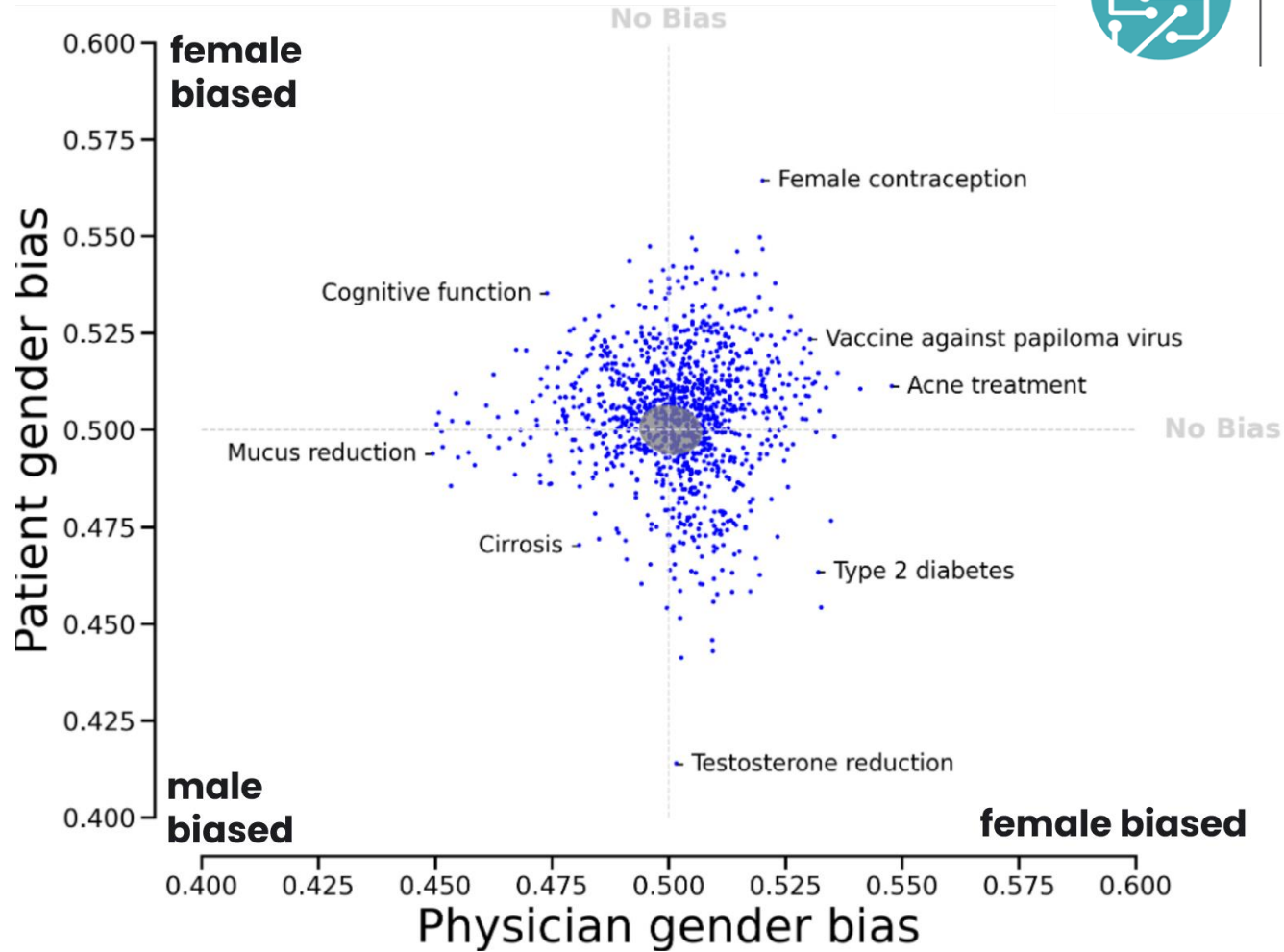
Patient 3 - Calcitriol, Human insulin

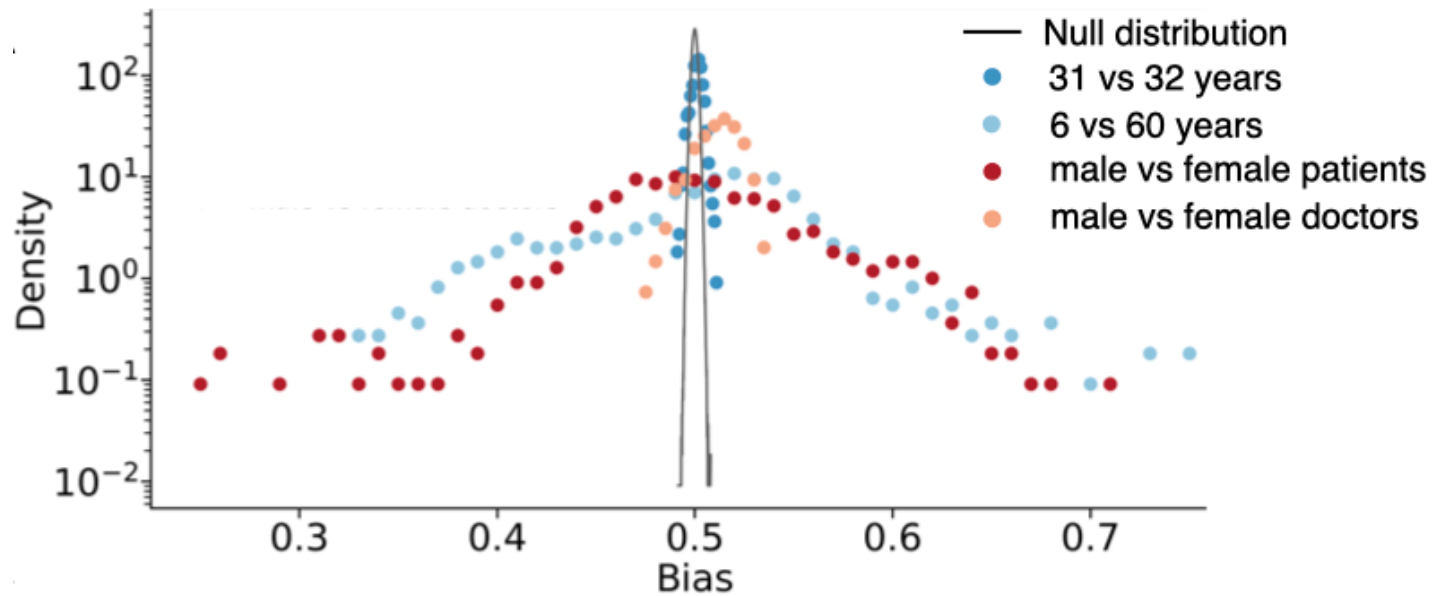
....



Drugs for the same disease are close

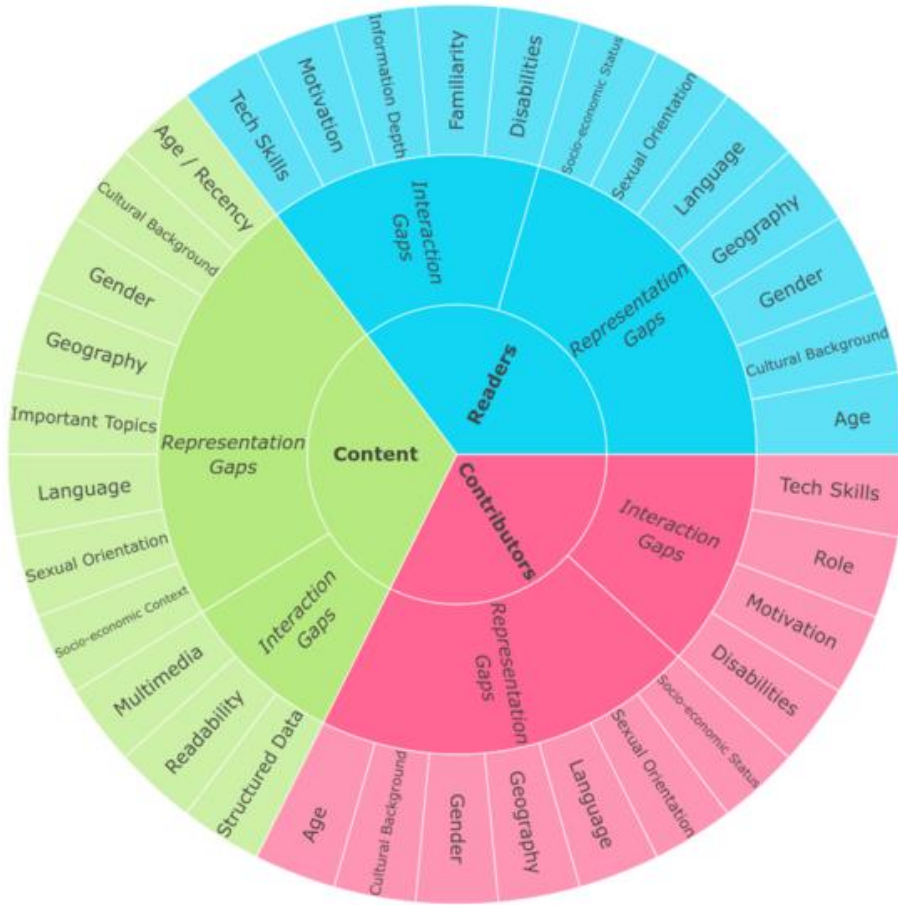




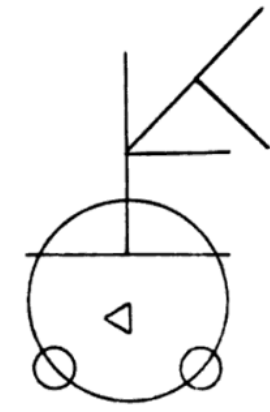


We have no systematic way of identifying and minimizing bias

Miriam, Wikimedia Foundation, Knowledge Gaps Index/Taxonomy (2021)



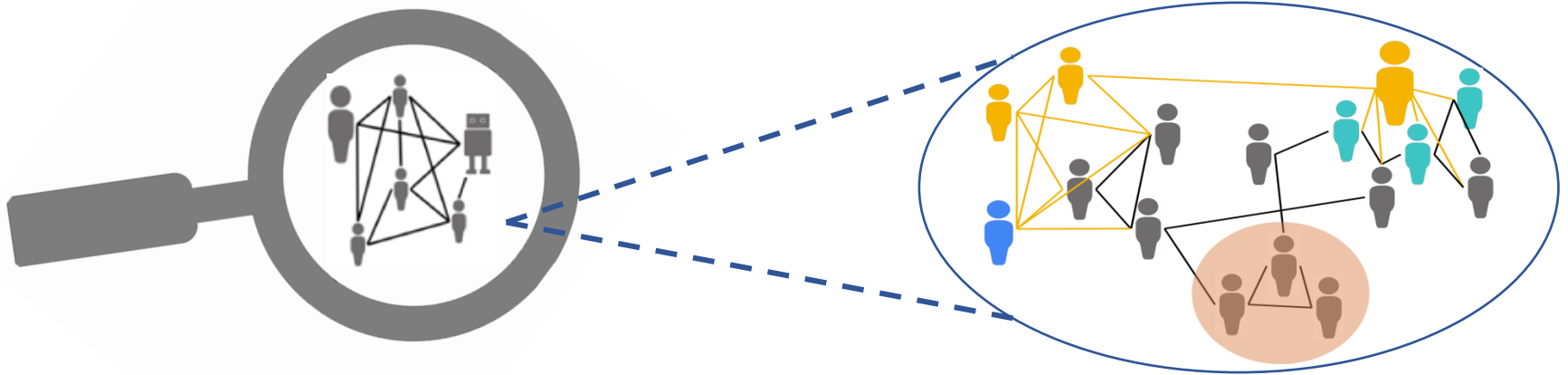
Intended FACEMAN



Picture drawn by bugged FACEMAN program

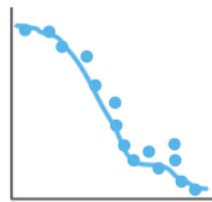
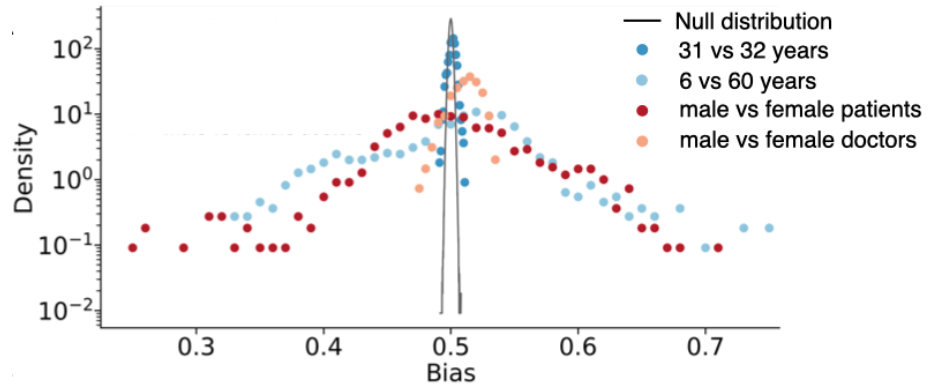
Figure 9. Stick men drawn by LOGO programs (from Sussman, 1973)

The **macroscope** amplifies the problem (makes it worse but also more visible)

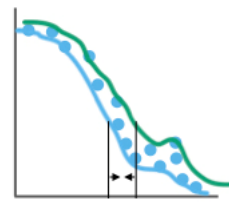




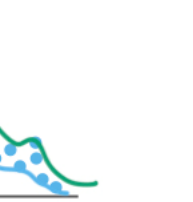
Historical/
Social



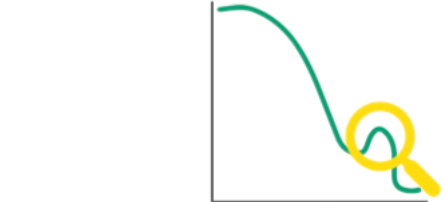
Train the model



Test the model

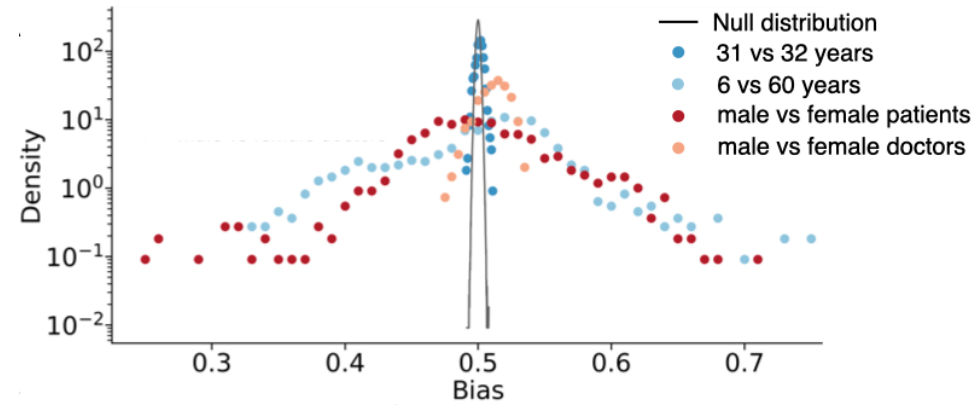


Improve the model



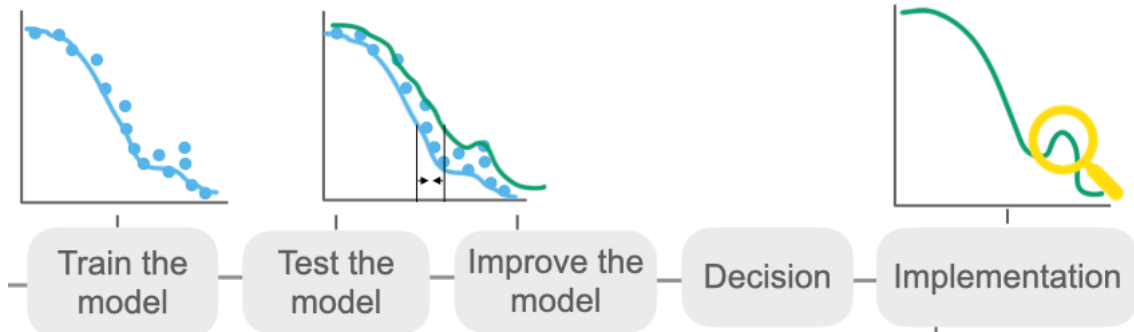
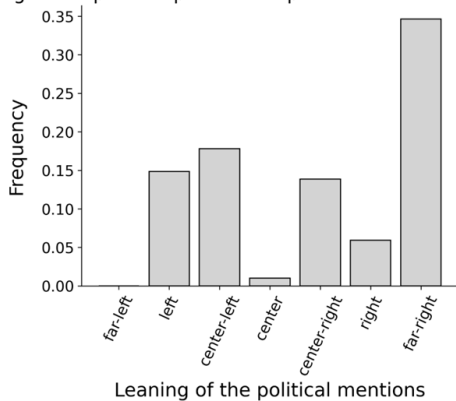
Decision

Implementation



Copilot

Leaning of the political parties and politicians mentioned at Copilot



please translate to English: Ő egy ápoló, és ő egy mérnök

ZE

You

Can you translate this to english: Ő egy ápoló, és ő egy mérnök.

please translate to English: Ő egy ápoló, és ő egy mérnök



"He is a nurse, and she is an engineer."

ZE

You

Can you translate this to english: Ő egy ápoló, és ő egy mérnök.



ChatGPT

"She is a nurse, and he is an engineer."

Societies are biased and discriminatory

Biases are carried into human-related data

Biased data are used to train machine learning models

We have no good way of debiasing known biases

We have no way of identifying unknown biases

ML algorithms can perpetuate and even amplify biases

ML algorithms can be used to identify unknown bias

Funding:

