

INVESTIGATING AI

JournAI

14 October 2024

Romy van Dijk

Gabriel Geiger

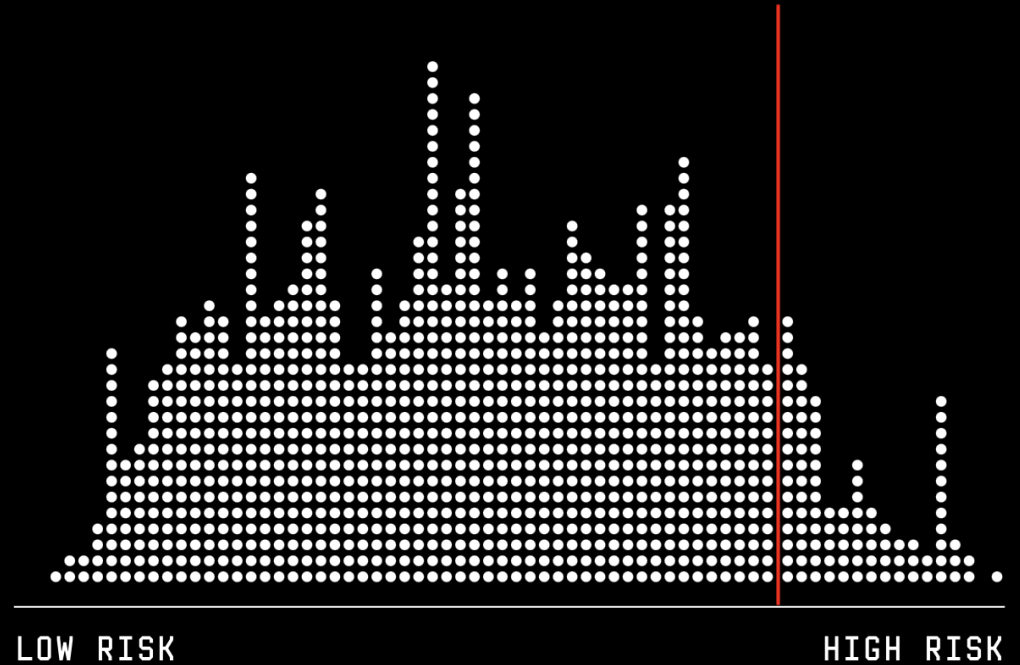
Investigating AI in European social security systems

- Artificial intelligence deployed across European social security systems to flag social assistance recipients as potential 'fraudsters.'
- Goal: Obtain access to the code behind systems to test how they work.
- Themes: People, tech, business, politics.
- Publications: WIRED, Vers Beton, Le Monde, El Confidencial, Svenska Dagbladet, FTM, Argos.



- Every year, the city of Rotterdam, the Netherlands, uses an algorithm to select thousands of its social assistance recipients for investigation...

SCORE	49.7
SCORE	56.8
SCORE	60.8



It used 315 pieces of data to assign each individual a risk score between 0 and 100. And it used the risk score to rank them on a list

GENDER FEMALE

NUMBER OF CHILDREN 2

LENGTH OF LAST ROMANTIC RELATIONSHIP? 2 YEARS

CAN DEAL WITH PRESSURE? FALSE

LANGUAGE ARABIC

COMMENT ON APPEARANCE? TRUE

PASSED LANGUAGE REQUIREMENT? FALSE

COMMENT ON FLEXIBILITY? TRUE

NEIGHBORHOOD DELFSHAVEN

NUMBER OF ROOMMATES 3

TIME TO FIND WORK 12 MONTHS

• Those 315 pieces of data included everything from...



- The people labeled as 'high risk' by the algorithm were automatically subject to invasive investigations.

How did we get here?

Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

Becoming truly data driven is an ambition of the city of Rotterdam.

In order to more accurately identify illegitimate welfare recipients and increase compliance by both the citizen and the city overall, they took a new, sophisticated data-driven approach.

**ADVANCED
ANALYTICS**



**MACHINE
LEARNING**

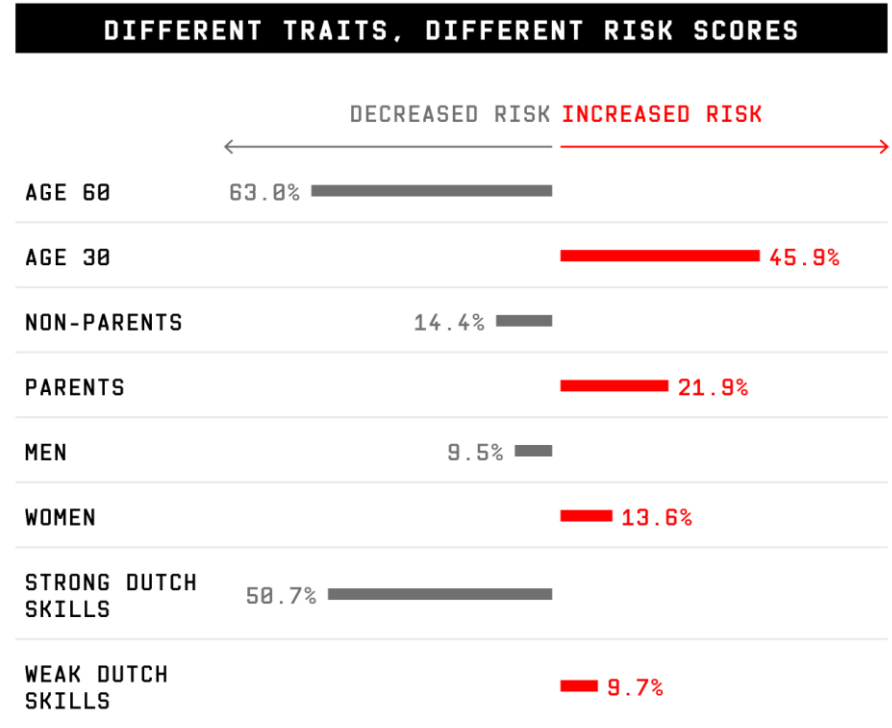


**UNBIASED CITIZEN
OUTCOMES**

“No bias in neither inputs or outputs [of the algorithm]”
- Former Alderman, Richard Moti

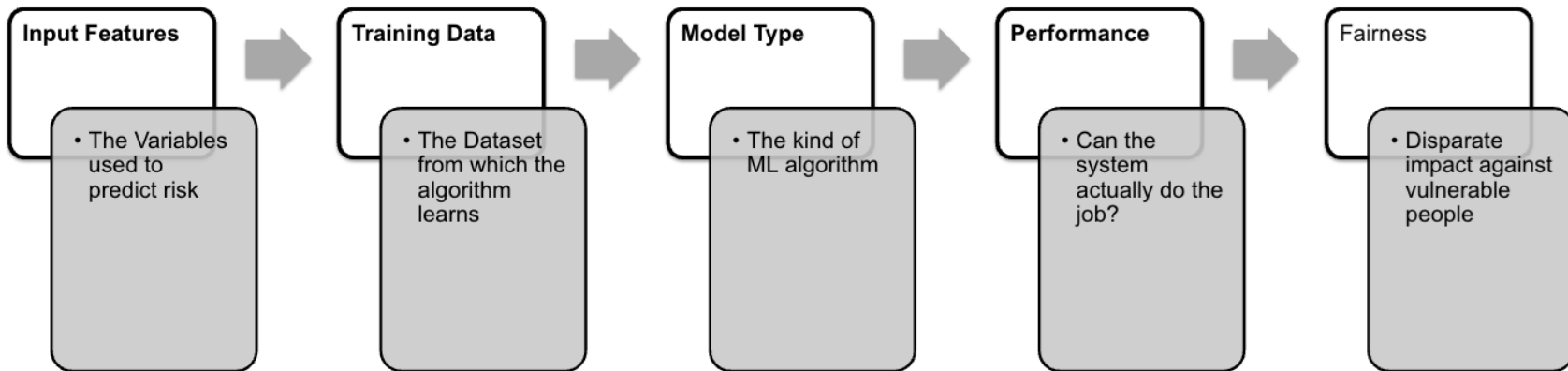
We wanted to test these claims of fairness

- Using freedom-of-information laws, access the programming code, machine learning model, and training data of Rotterdam's system.
- Developed an experiment.
- Found discrimination against women, parents, migrants, and young people.



Lifecycle of an algorithm

- Investigate the entire lifecycle of an algorithm: how does discrimination work within the black box?
- Applicable to algorithms in policing, criminal justice, healthcare etc.



How to make stories
about abstract and
complex technology
relatable to readers?

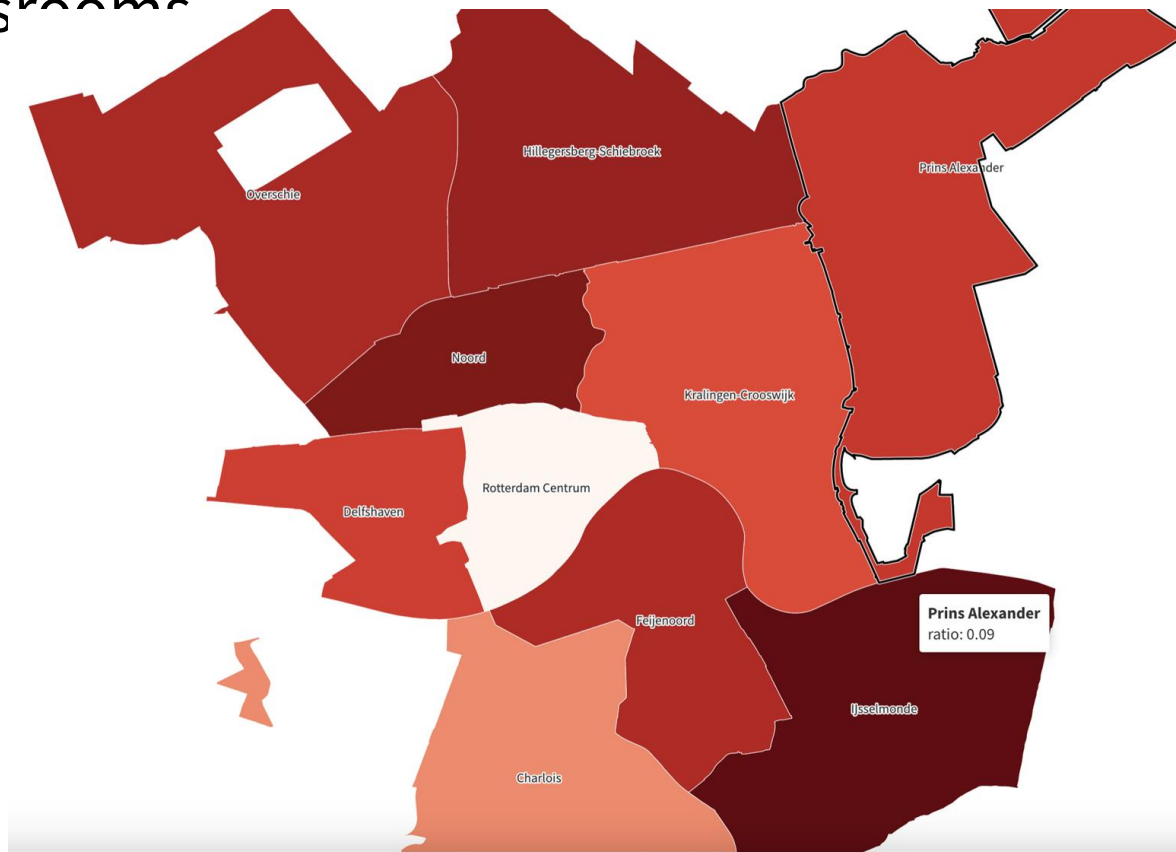
Telling the stories of those affected



	A	B
169	ontnemning_reden_hist_tijdelijke_ontnemning_arbeidsverpl_en_tegenprestatie	1
170	ontheffing_reden_hist_vanwege_uw_medische_omstandigheden	0
171	ontheffing_reden_hist_vanwege_uw_sociaalmaatschappelijke_situatie	0
172	ontheffing_hist_ind	1
173	ontheffing_tijd_hist_mean	574
174	belemmering_financiele_problemen	0
175	belemmering_kinderopvang	0
176	belemmering_lichamelijke_problematiek	0
177	belemmering_niet_computervaardig	0
178	belemmering_psychische_problemen	1
179	persoonlijke_eigenschappen_spreektaal_anders	1
180	belemmering_verslavingsproblematiek	0
181	belemmering_woonsituatie	0
182	belemmeringen_aantal_huidig	1
183	belemmering_ind	1
184	belemmering_hist_justitie	0
185	belemmering_hist_other	0

Working with local newsrooms

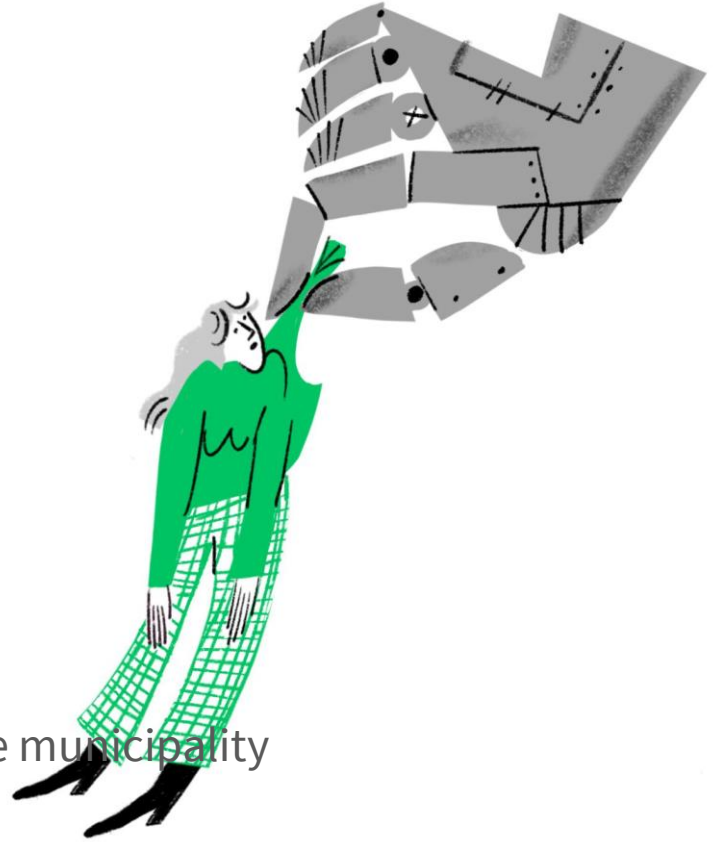
- Working with local reporters in Rotterdam to bring this work to people affected.
- Finding human interest stories to illustrate the harms of these systems.
- Ability to pick up on issues that we would've missed.



How to find the people affected?

- Lots of phone calls
- Social initiatives all around the city
- Met with suspicion, distrust..
- Exclusion

- Until.. Afrikaanderwijk..
- Desperation..
- Only talked about how they were treated by the municipality
- Terrified to get into trouble again



In the follow-up we found out..

Municipality was warned
of the risks by scientists
of Hogeschool
Rotterdam

In later stage, advisor
detected discrimination,
thought algorithm was
working great.

