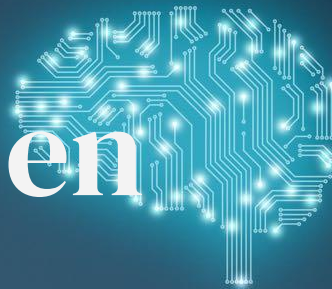


Tussen Code en Vrees



Menselijke Waarden in het
AI Tijdperk

KNVI 23-09

Joris Krijger

SIDNfonds

de volksbank
beter voor elkaar

Erasmus
University
Rotterdam

Erasmus



De Volksbank

Ethics & AI Officer



**Erasmus
Universiteit
Rotterdam**

PhD AI & Ethics

Drie Vragen

01

Welke impact heeft AI?

Welke problemen kan AI oplossen?

Welke problemen kan het *niet* oplossen?

Welke nieuwe problemen creëert het?

02

Wat staat er op het spel?

Gelijkheid, rechtvaardigheid en
vrijheid in het AI tijdperk

03

Hoe sturen we AI?

Wie of wat bepaalt of en hoe AI
wordt ontwikkeld en gebruikt?

Drie Vragen

01

Welke impact heeft AI?

Welke problemen kan AI oplossen?

Welke problemen kan het *niet* oplossen?

Welke nieuwe problemen creëert het?

02

Wat staat er op het spel?

Gelijkheid, rechtvaardigheid en
vrijheid in het AI tijdperk

03

Hoe sturen we AI?

Wie of wat bepaalt of en hoe AI
wordt ontwikkeld en gebruikt?



01

De Impact
van AI

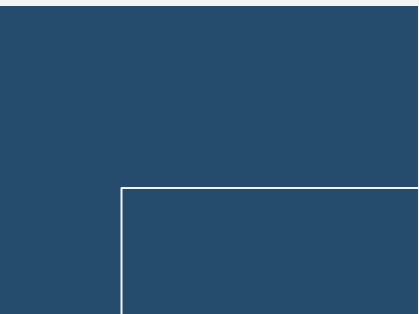
NL

88%

van de organisaties in Nederland
gebruikt al AI

69%

van de organisaties in Nederland
gebruikt Generatieve AI



“Artificial intelligence
(AI) refers to:

1. machine based systems
2. with some degree of autonomy
and adaptiveness
3. infers from given inputs
4. how to achieve specific goals
and objectives.



Definitie

AI

Test scores of the AI relative to human performance

+20

0 ← Human performance, as the benchmark, is set to zero.

-20

-40

-60

-80

-100

AI systems perform better than the humans who did these tests

AI systems perform worse

Reading comprehension

Language understanding

Handwriting recognition

Speech recognition

Image recognition

2000

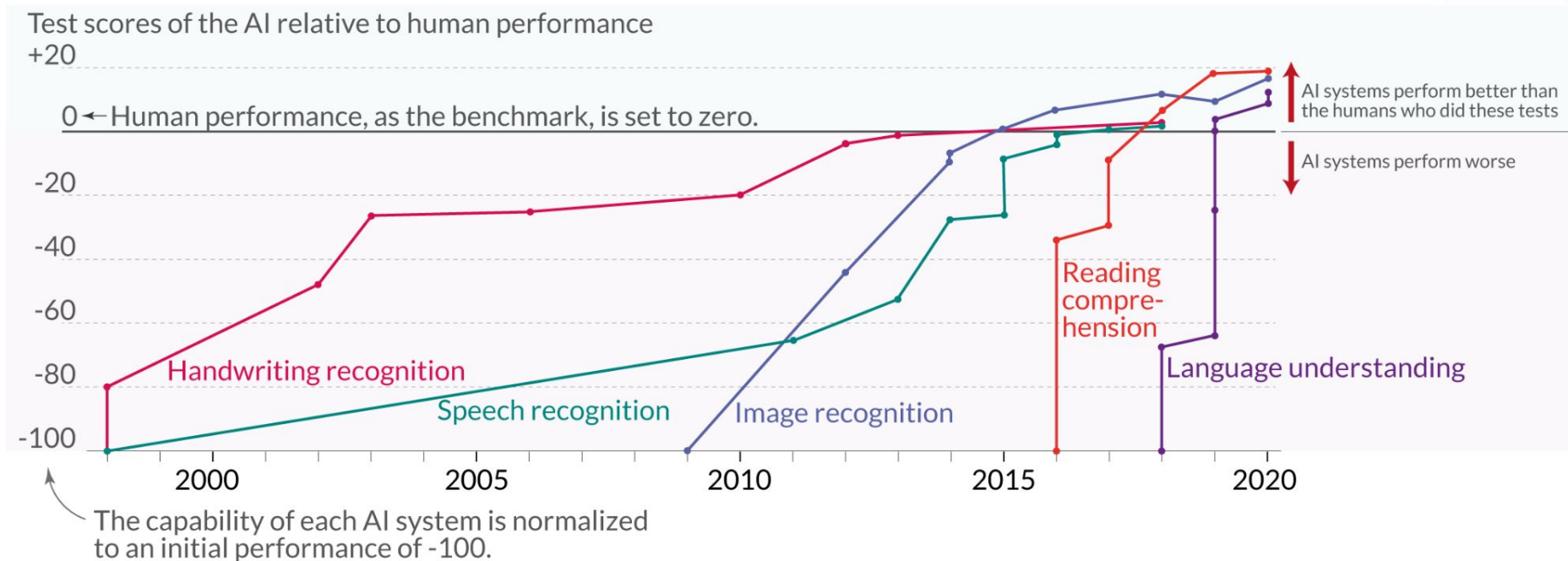
2005

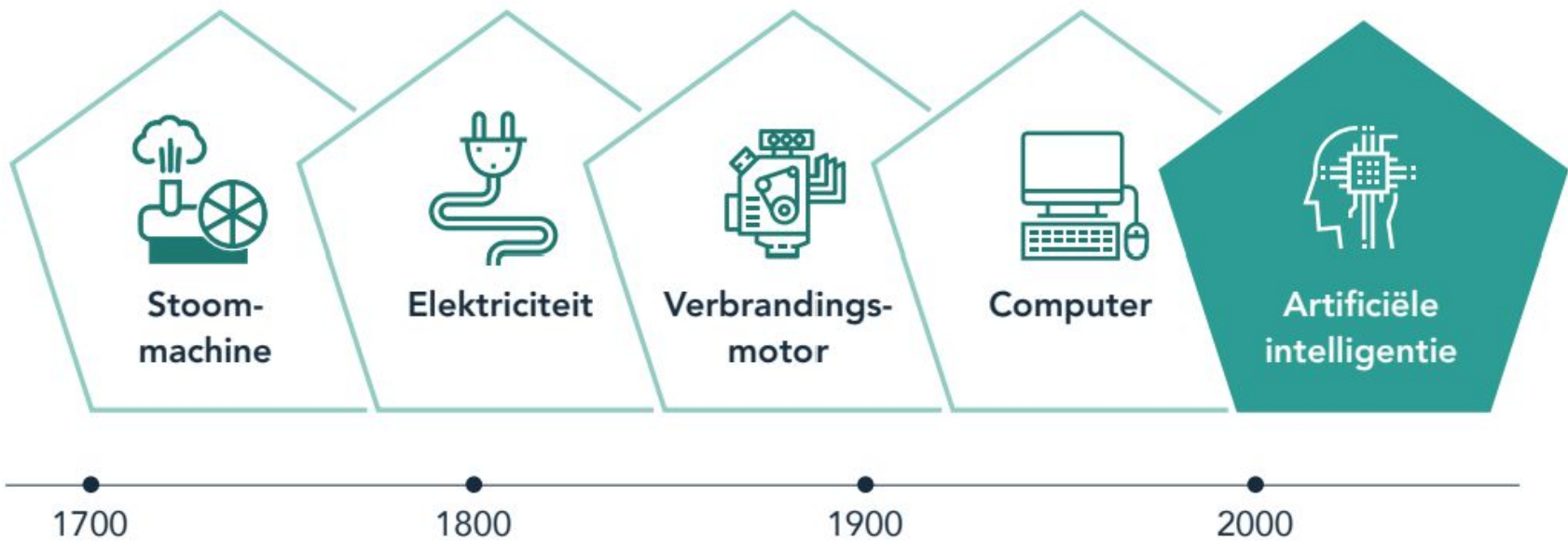
2010

2015

2020

The capability of each AI system is normalized to an initial performance of -100.





Systeemtechnologie

“AI is in onze eeuw wat elektriciteit was in de negentiende eeuw, of de verbrandingsmotor in de twintigste eeuw.”

“AI zal invloed hebben op onder andere veiligheid en gezondheid, autonomie en vrijheid, burgerrechten en de rechtsstaat, rechtvaardigheid en inclusie. Dat zal echter op manieren gebeuren die we nu grotendeels nog niet kunnen voorzien.”

AI



Super AI

Machines that are an order of magnitude smarter than humans



General AI

Takes knowledge from one domain, transfers to other domain

Credit: Chris Noessel



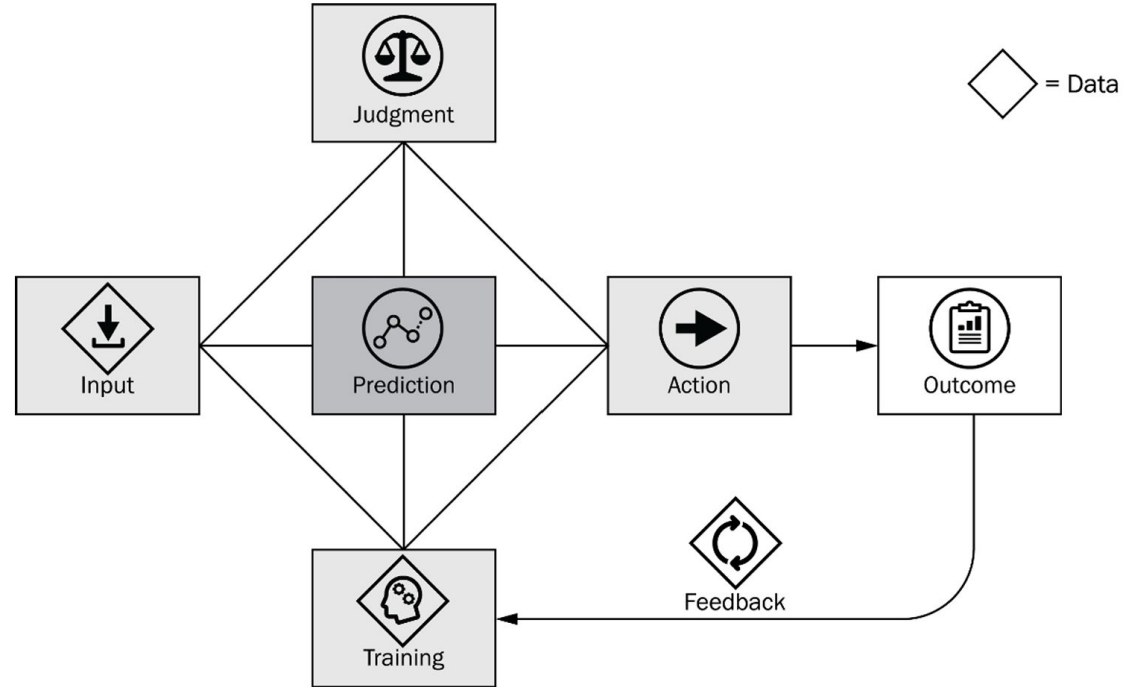
Narrow AI

Dedicated to assist with or take over specific tasks

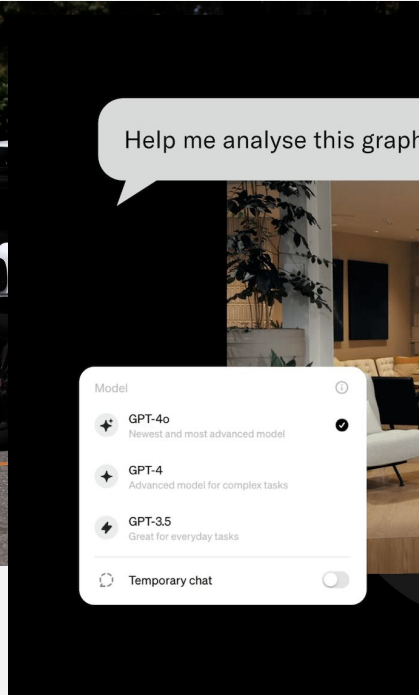


**Welke problemen gaat
AI oplossen?**

Wat is een taak?



Voorspelproblemen



AI's Impact op werk

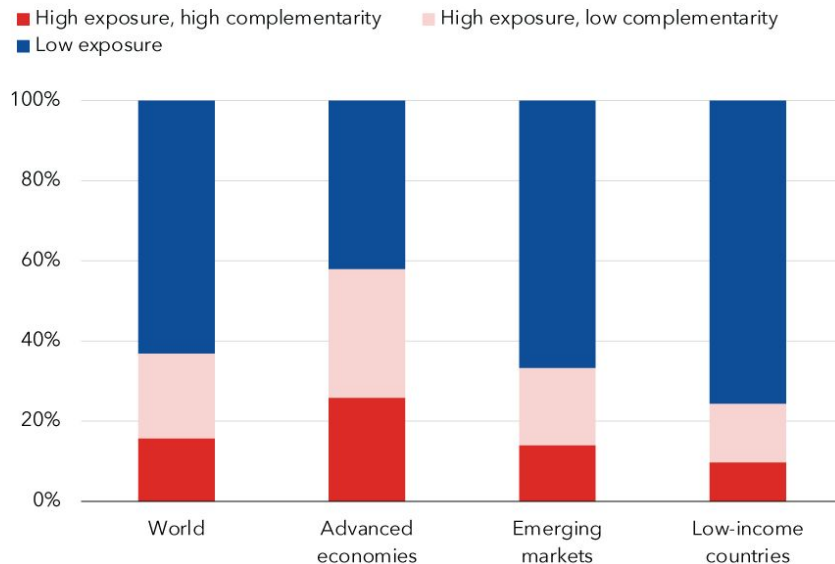
In 'advanced economies' zal 60% van de banen met AI te maken krijgen.

Daarvan zal de helft een complementair gebruik van AI zijn die helpt om de productiviteit te verhogen. Voor de andere helft (30%) zal AI op termijn de taken overnemen die momenteel door mensen uitgevoerd worden.

AI's impact on jobs

Most jobs are exposed to AI in advanced economies, with smaller shares in emerging markets and low-income countries.

Employment shares by AI exposure and complementarity



Source: International Labour Organization (ILO) and IMF staff calculations
Note: Share of employment within each country group is calculated as the working-age-population-weighted average.

IMF

Waar staat AI?

AI could p and the U commissio



By [Matt Egan](#), CNN
🕒 6 minute read · Upc



Eat a rock a day, put glue on your pizza: how Google's AI is losing touch with reality

Published: May 27, 2024 7.15am CEST

Shutterstock AI Generator

✉ Email

✕ X (Twitter)

📘 Facebook 596

🌐 LinkedIn

🖨 Print

Google has rolled out its [latest experimental search feature](#) on Chrome, Firefox and the Google app browser to hundreds of millions of users. “AI Overviews” saves you clicking on links by using generative AI — the same technology that powers rival product ChatGPT — to provide summaries of the search results. Ask “how to keep bananas fresh for longer” and it uses AI to generate a useful summary of tips such as storing them in a cool, dark place and away from other fruits like apples.

Author



Toby Walsh
Professor of AI, Research Group
Leader, UNSW Sydney

Disclosure statement

Toby Walsh receives funding from the ARC via a Laureate Fellowship, as well as a number of grants on AI for Good from google.org, the philanthropic arm of Google.



**Welke problemen gaat
AI niet oplossen?**



Bernard Parker & Dylan Fugett

“

Models are opinions embedded in mathematics. [...]. Big data processes codify the past, they do not invent the future.

Cathy O'Neil

”

Context

- 
1. Menselijke natuur
 2. Geschiedenis
 3. Menselijke kant van technologie
 4. Economie
 5. Ethiek

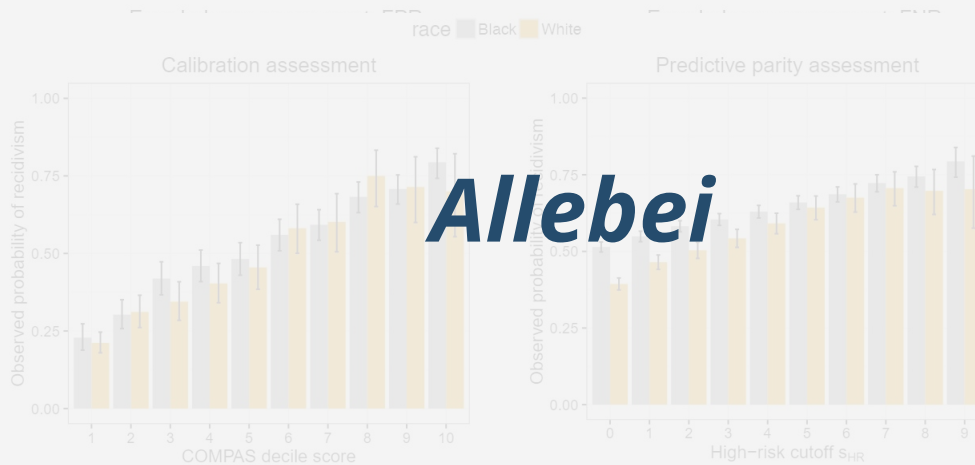
“AI vertelt ons iets over hoe de wereld was, misschien iets over hoe de wereld is, maar kan ons nooit iets vertellen over de wereld zoals die zou moeten zijn.”

Machine Bias

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

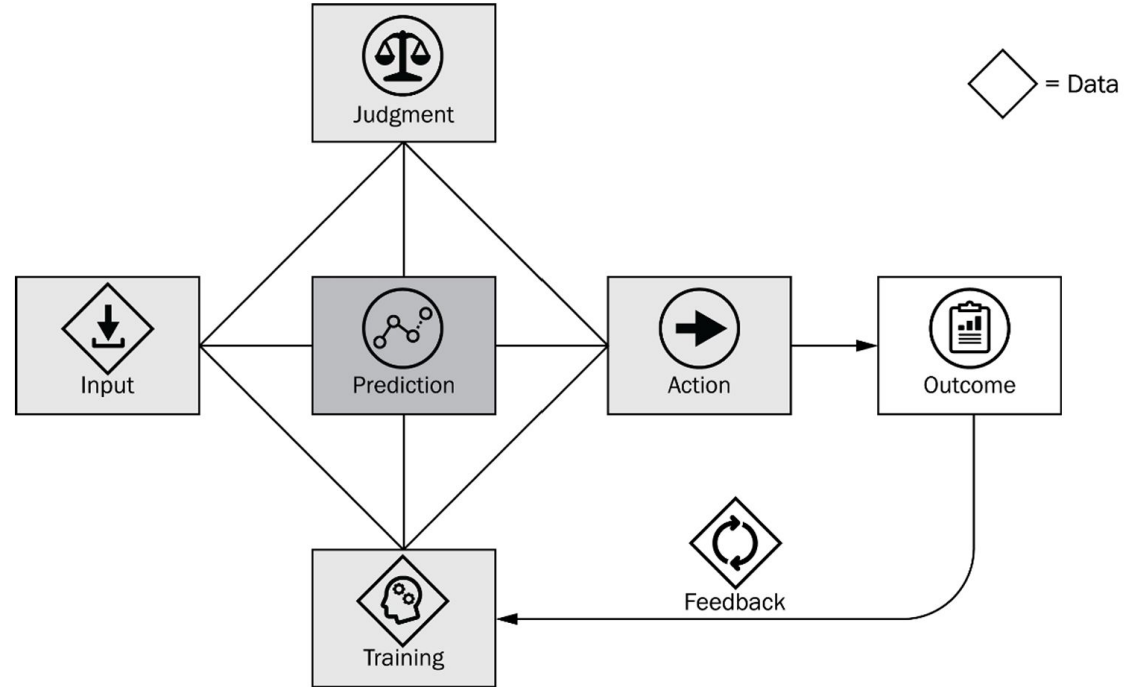
by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica

Wie heeft er gelijk?



Allebei

Wat is een taak?





**Welke nieuwe
problemen creëert AI?**



INSURANCE



Air Canada ordered to pay customer who was misled by airline's chatbot

By *Maria Yagoda* 23rd February 2024

When Air Canada's chatbot gave incorrect information to a traveller, the airline argued its chatbot is "responsible for its own actions".

Artificial intelligence **is having a growing impact on the way we travel**, and a remarkable new case shows what AI-powered chatbots can get wrong – and who should pay. In 2022, Air Canada's chatbot **promised a discount** that wasn't available to passenger Jake Moffatt, who was assured that he could book a full-fare flight for his grandmother's funeral and then apply for a bereavement fare after the fact.

According to a civil-resolutions tribunal **decision** last Wednesday, when Moffatt applied for the discount, the airline said the chatbot had been wrong – the request needed to be submitted before the flight – and it wouldn't offer the discount. Instead, the airline said the chatbot was a "separate legal entity that is responsible for its own

Vooruitgang?




Joanna Maciejewska

@AuthorJMac@indiepocalyps...

You know what the biggest problem with pushing all-things-AI is? Wrong direction.

I want AI to do my laundry and dishes so that I can do art and writing, not for AI to do my art and writing so that I can do my laundry and dishes.

Mar 29, 2024, 05:20 PM · 

829 boosts · 245 favorites



02

Wat staat er op
het spel?

Kunstmatige Toekomst



Rechtvaardigheid

Hoe zorgen we voor processen om de balans tussen te gelijkheid en vrijheid te bewaren



Gelijkheid

Hoe borgen we de morele gelijkheid en inclusieve welvaart in een samenleving?



Vrijheid

Wat is een rechtvaardige verdeling van de macht in een samenleving?

Fairness?

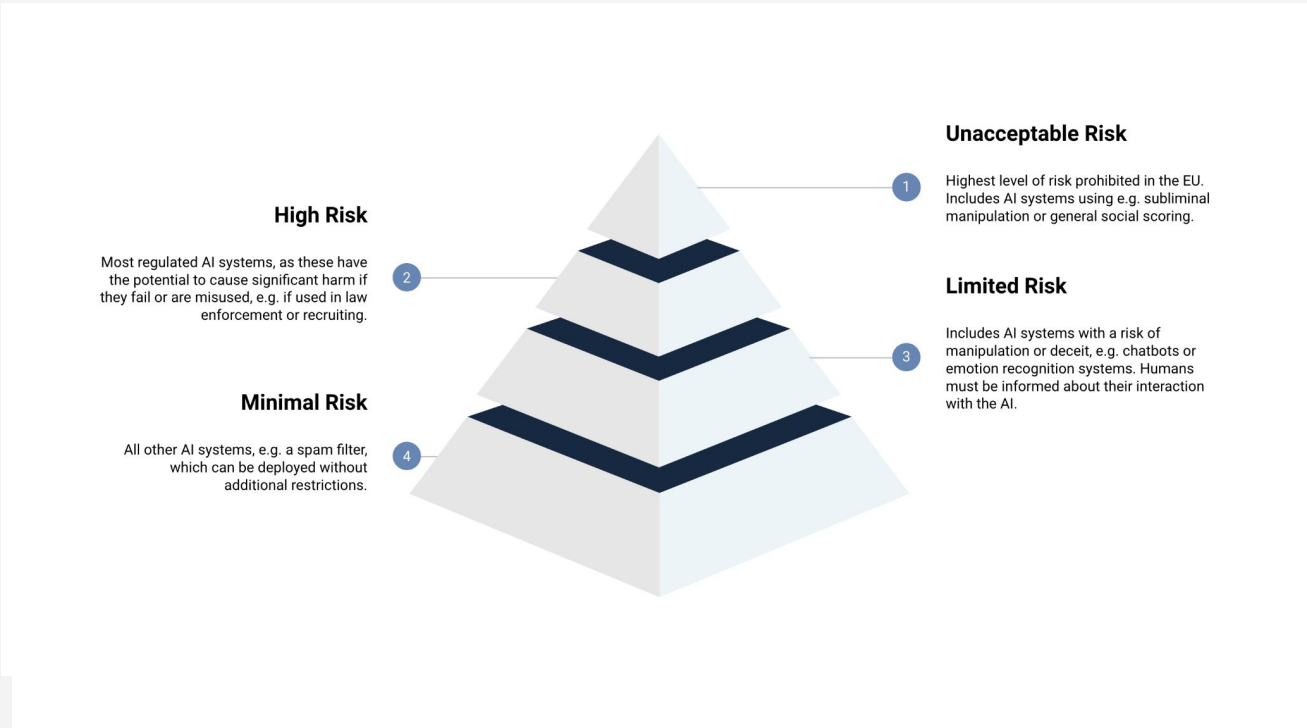
Wie beslist wat eerlijk is?

Wie beslist of en hoe we bias mitigeren?

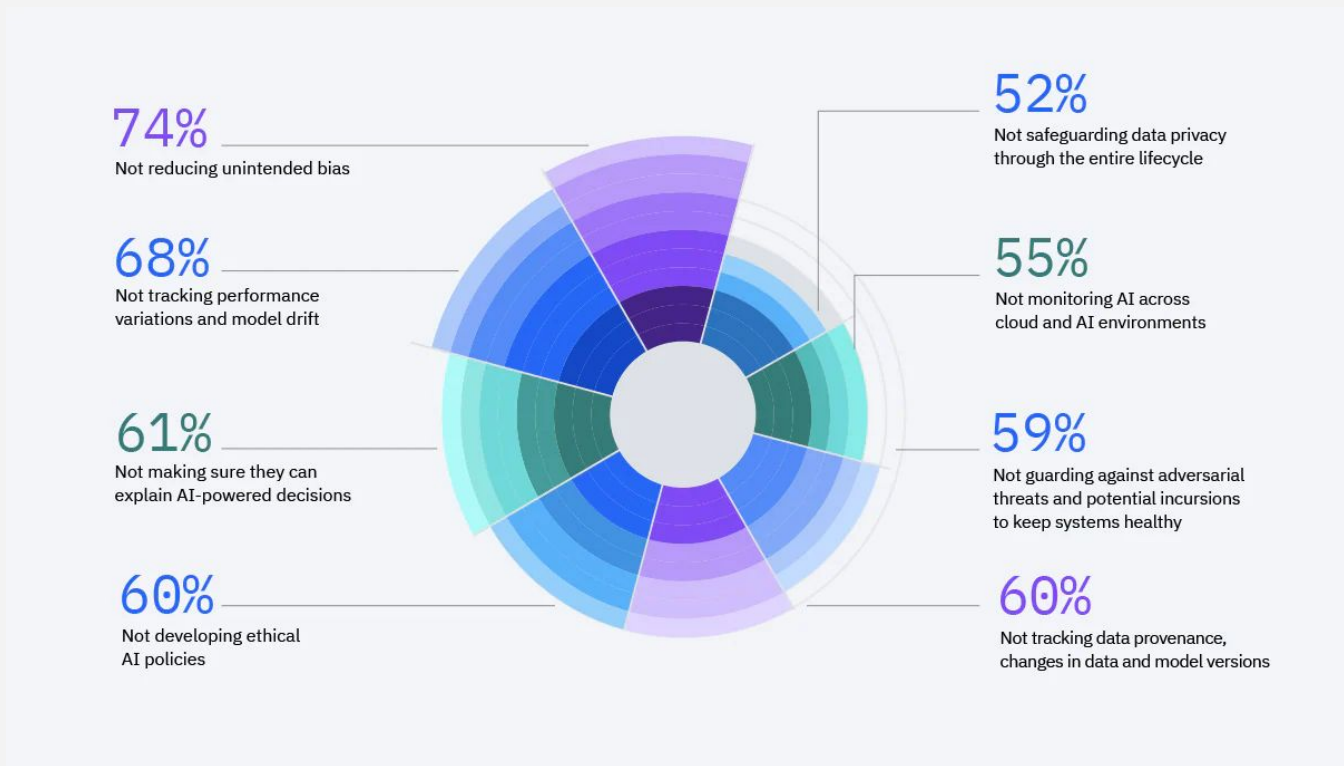
***Dit zijn allemaal geen technische
vragen... maar ethische!***

*Wie is verantwoordelijk voor het bespreken
van deze vraagstukken?*

Ethische Principes voor AI

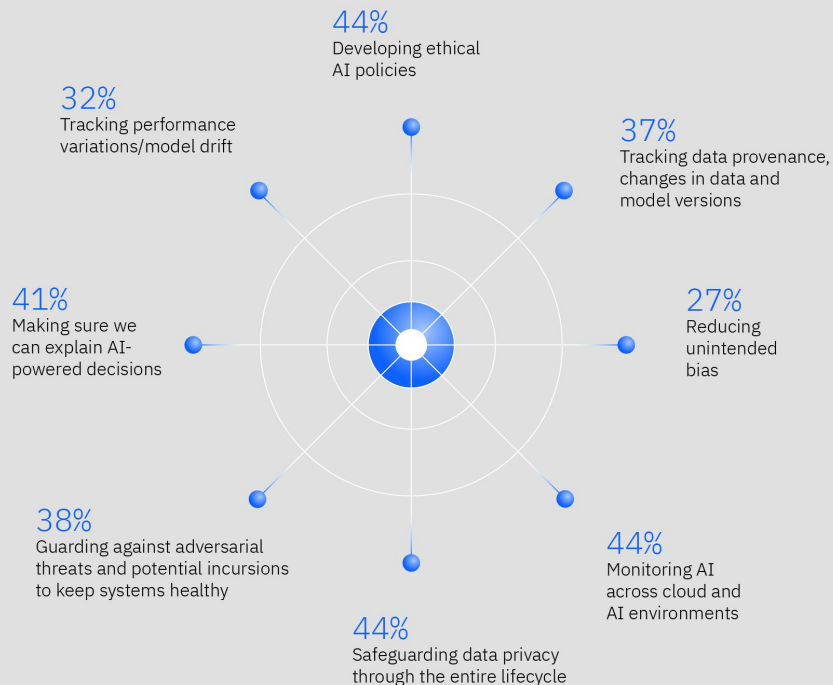


AI Praktijk



AI Praktijk

Enterprises are taking steps to build trustworthy AI, but more progress is needed



The background of the slide features a stack of coins, likely Euro coins, resting on a white, finely textured fabric surface. The coins are slightly out of focus, creating a soft, bokeh-like effect. The lighting is even, highlighting the metallic sheen of the coins and the texture of the fabric.

**We zien iets cruciaals over
het hoofd op het gebied
van verantwoorde AI**

Hoe kunnen we
vooruit?

03



Wie of wat bepaalt of
en hoe AI wordt
ontwikkeld en
gebruikt?





Wie ontwerpt
het ontwerp-
proces?

SIDN fonds

de volksbank
beter voor elkaar

Erasmus
University
Rotterdam



Onderhandelingen op drie niveaus: AI Three-O



Omgeving

Normen en instituties



Organisatie

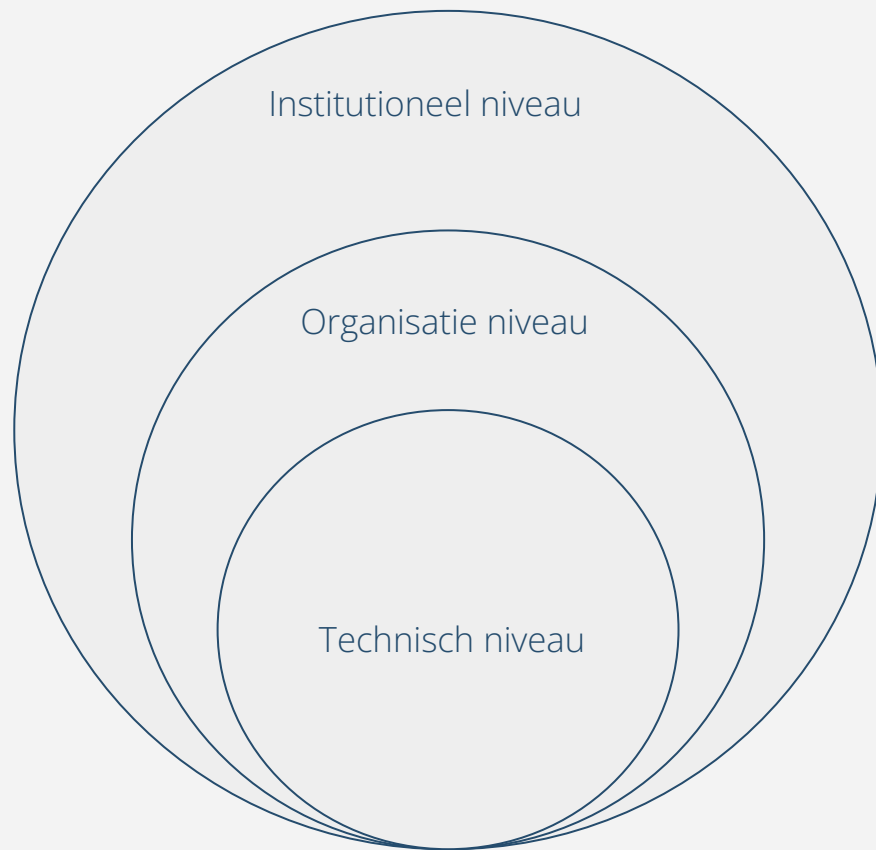
Organisatie structuur, cultuur en processen



Ontwerp

Ontwerpmethodes voor AI

AI Three-0



HANDBOOK ON Human Rights



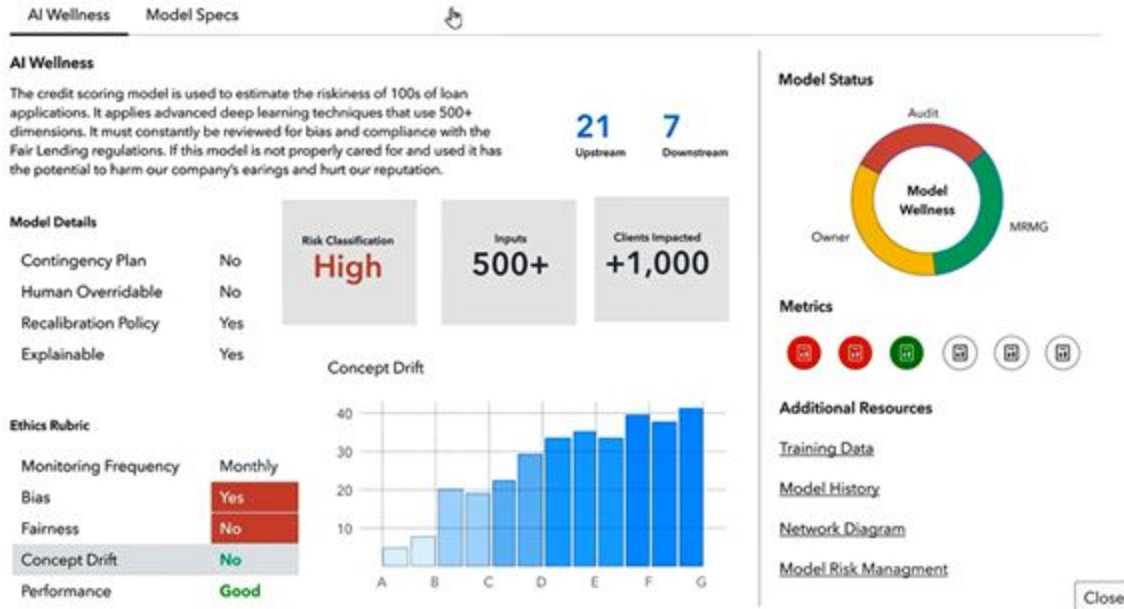
Skater (SHap
Kernel Explainer)



RISE (Right
explanation Set for
Interpretation)



Microsoft I
ML



An aerial photograph of a complex multi-level highway interchange with several overpasses and ramps. The image is taken from a high angle, showing the intricate network of roads. A semi-transparent white rectangular box is centered over the image, containing text in a dark blue, serif font. The background shows some greenery and buildings between the highway sections.

**We missen de
organisatorische
infrastructuur voor
ethische reflectie rond AI!**

Organisatorisch



De Volksbank

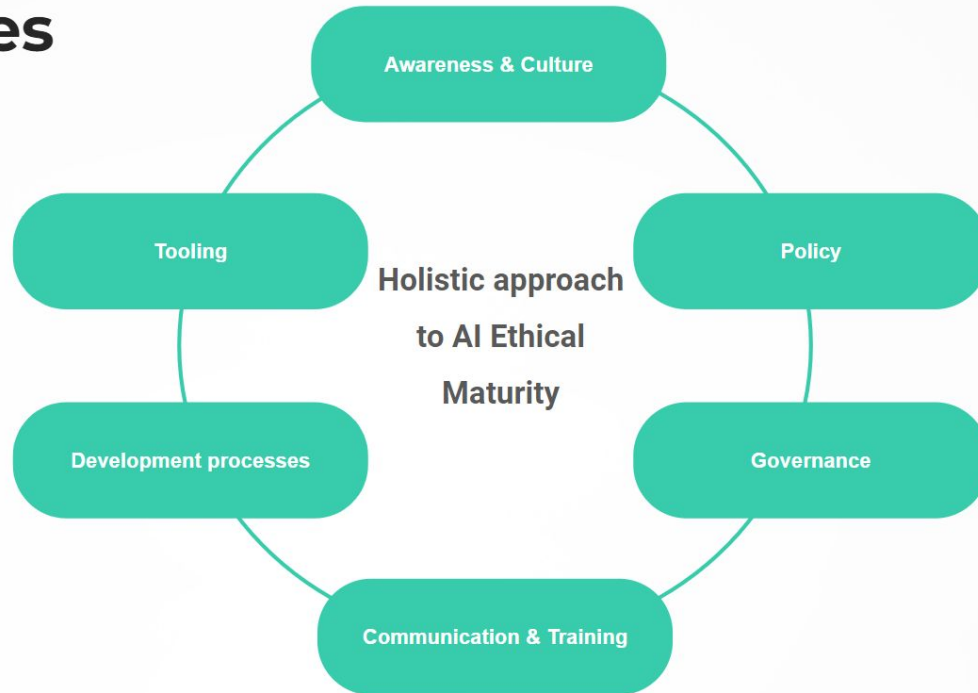


Grip op de ethische aspecten van data science toepassingen binnen de Volksbank

Governance structuur:

1. **Ethics Office**
2. **Ethisch Klankbord**
3. **Ethisch Committee**

Zes dimensies



Institutioneel

“Artificial intelligence poses an existential threat to creative professions, and all actors and performers deserve contract language that protects them from having their identity and talent exploited without consent and pay.[...] If we don’t stand tall right now, we are all going to be in trouble. ***We are all going to be in jeopardy of being replaced by machines.***”

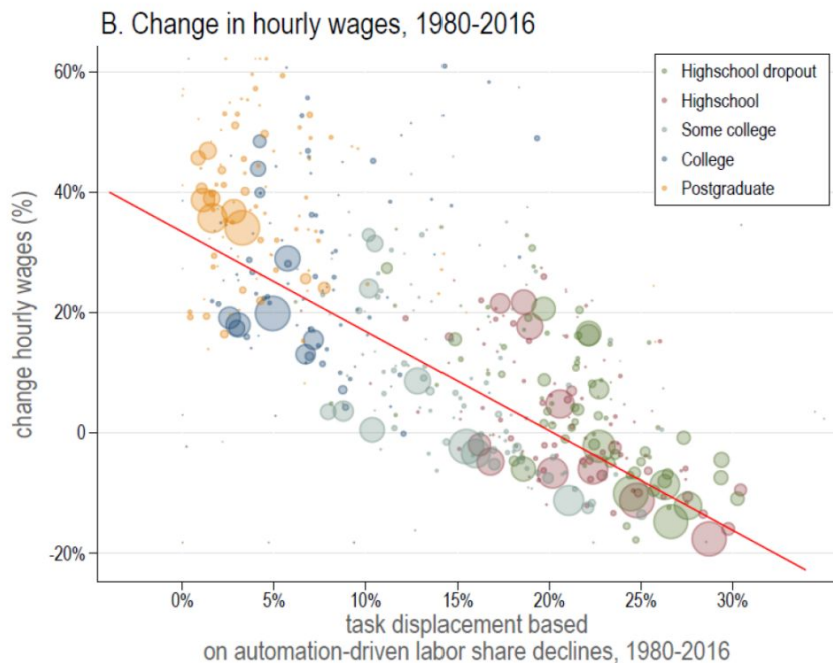
Fran Dresecher - President SAG-AFTRA (Screen Actors Guild-American Federation of Television and Radio Artists).



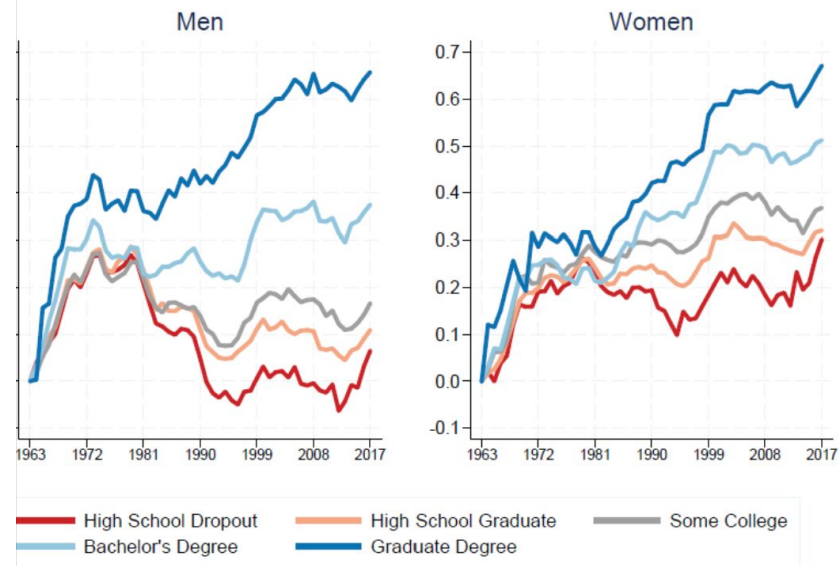
Institutes: Belasting

The gap between productivity and a typical worker's compensation has increased dramatically since 1979

Productivity growth and hourly compensation growth, 1948–2019



Cumulative Change in Real Log Weekly Earnings 1963 - 2017
Working Age Adults, Ages 18 - 64



Adapted from Figure A in *Raising America's Pay: Why It's Our Central Economic Policy Challenge* (Lazear et al. 2014)

Institutes: Rechten



Mijn verhaal

01

AI is een
voorspelmachine ...

02

... die de fundamentele
waarden van onze
samenleving op de proef gaat
stellen...

03

... en waarvoor het
verantwoord gebruik op
drie niveaus moet worden
geborgd: technisch,
organisatorisch en
institutioneel.

“

All models are wrong but some are useful.

George E.P. Box

”

NEW YORK ST



Joris Krijger
LinkedIn: Joris Krijger
krijger@esphil.eur.nl

Dank!



Joris Krijger
LinkedIn: Joris Krijger
krijger@esphil.eur.nl

Vragen?