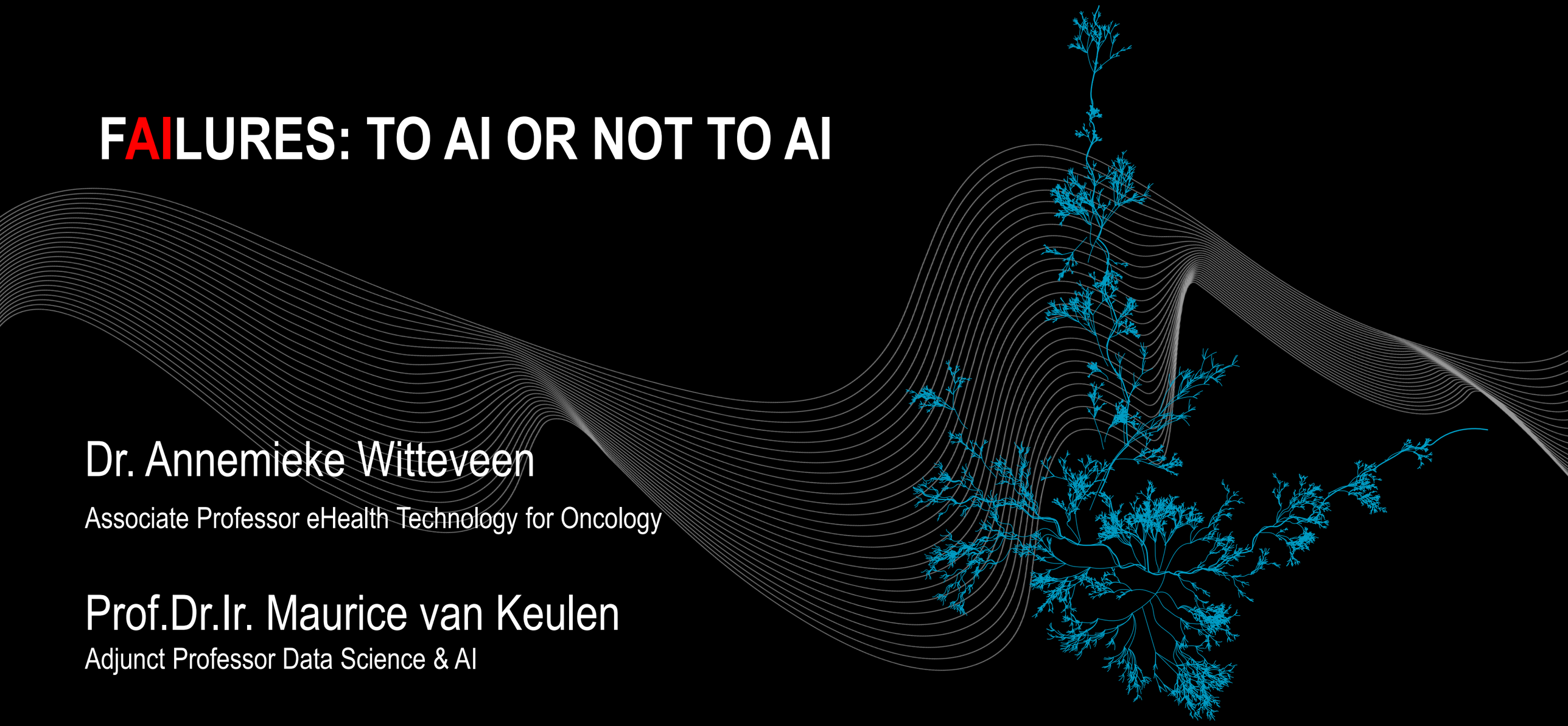
FAILURES: TO AI OR NOT TO AI



Dr. Annemieke Witteveen

Associate Professor eHealth Technology for Oncology

Prof.Dr.Ir. Maurice van Keulen

Adjunct Professor Data Science & AI

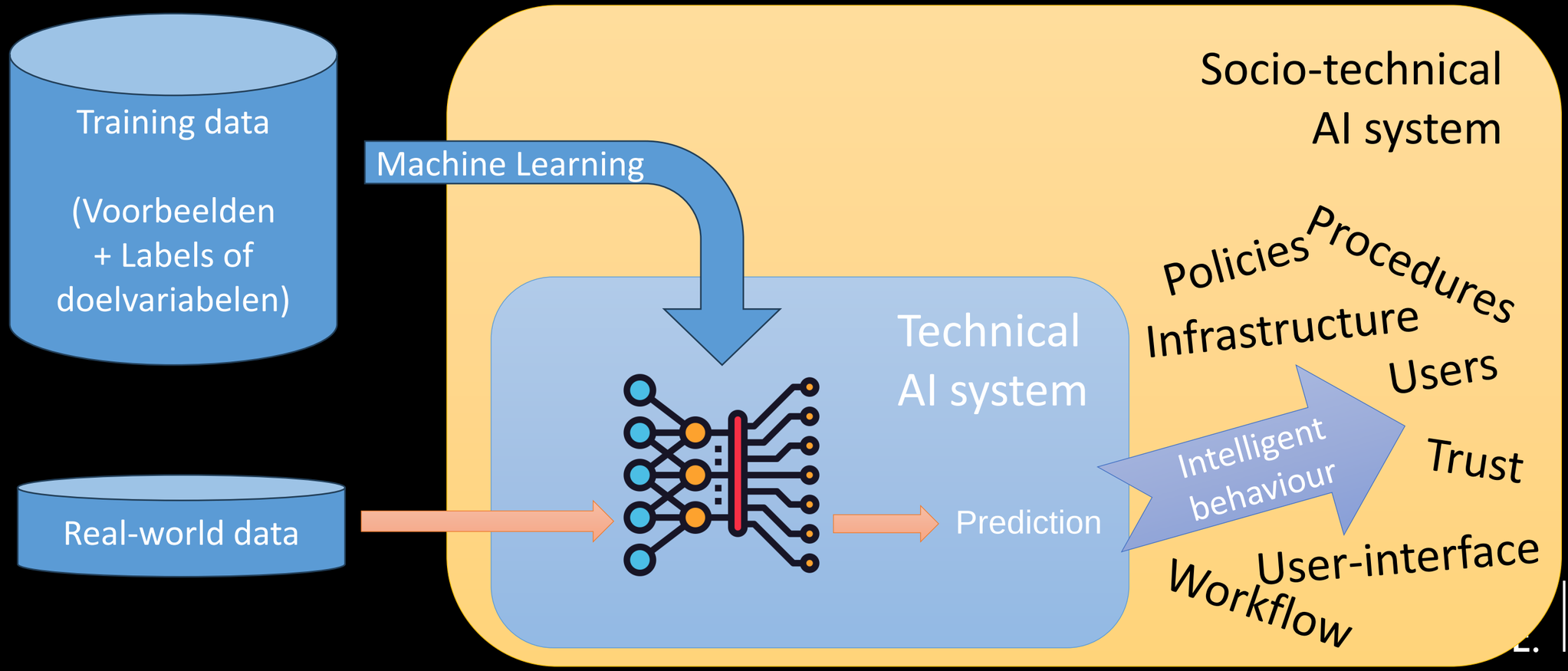
UNIVERSITY OF TWENTE. | TECHMED CENTRE

DISCLOSURES

Geen (potentiële) belangenverstrengeling

Geen tegenstander van AI (integendeel)

Hoe werkt AI? → Hoe faalt AI?

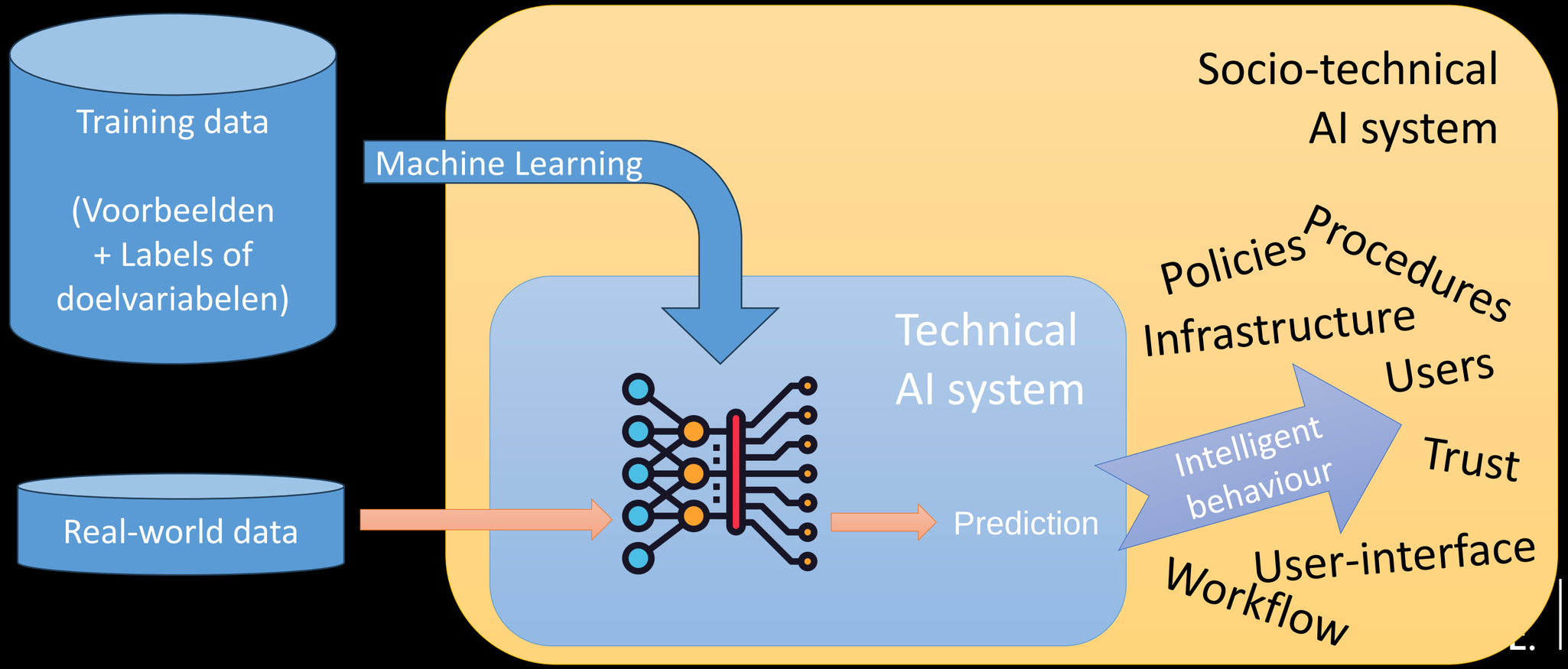


Hoe werkt AI? → Hoe faalt AI? → op drie niveau's

Data klopt niet
(data kwaliteit)

De AI misdraagt zich
of is onnodig complex

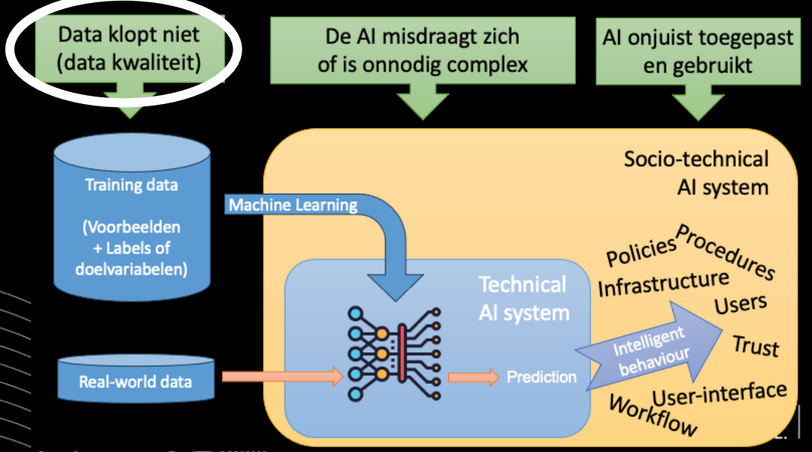
AI onjuist toegepast
en gebruikt



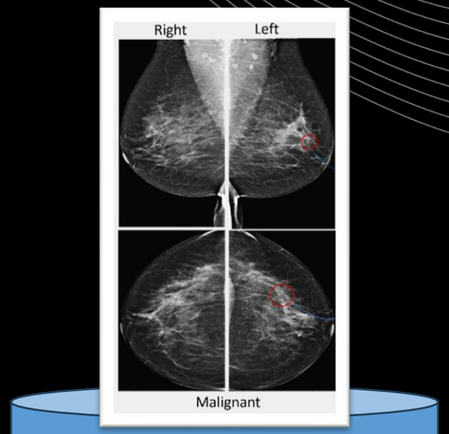
Voorbeelden beperkingen / zwakheden AI

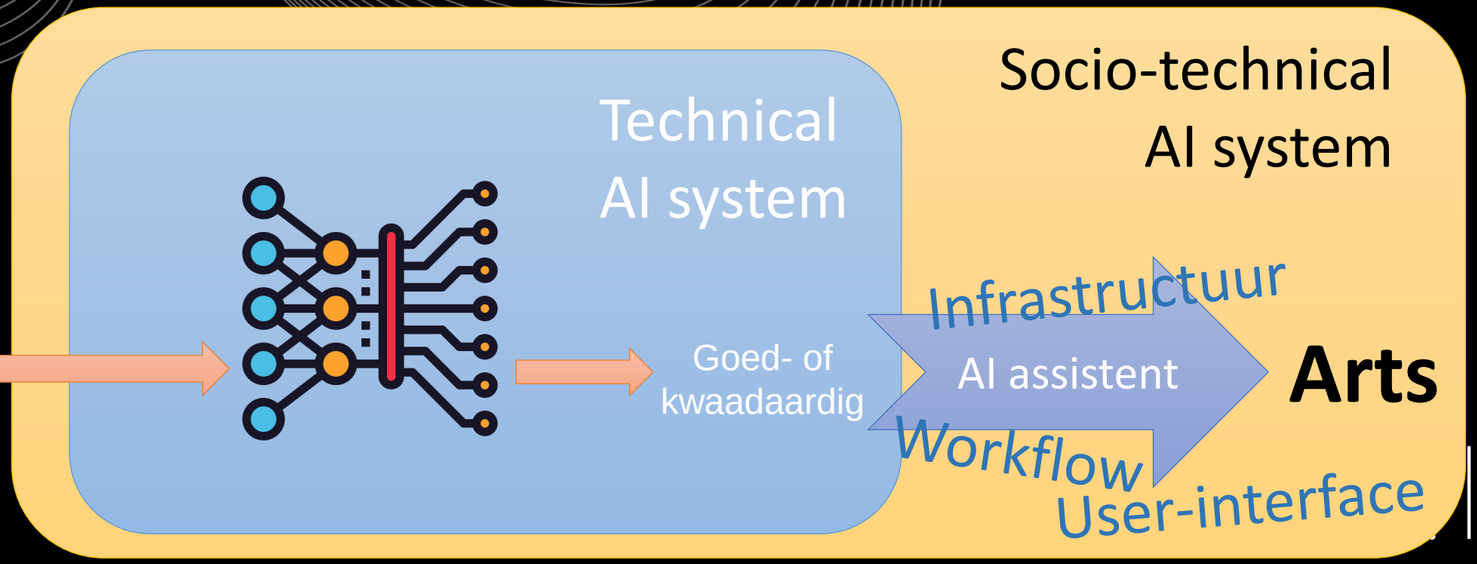
1. Data klopt niet: Voorspellend model kankerdiagnose — Maurice
2. Data klopt niet: Kanker-gerelateerde vermoeidheid — Annemieke
3. Model onnodig complex: Borstkanker follow-up — Annemieke
4. Model misdraagt zich: Short-cuts en onklinisch redeneren in fracturendetectie — Maurice
5. Onjuist toegepast: Terug naar voorspellend model kankerdiagnose
   Behandeladvies? Vergeet ChatGPT — Maurice

Shreyasi Pathak, MSc



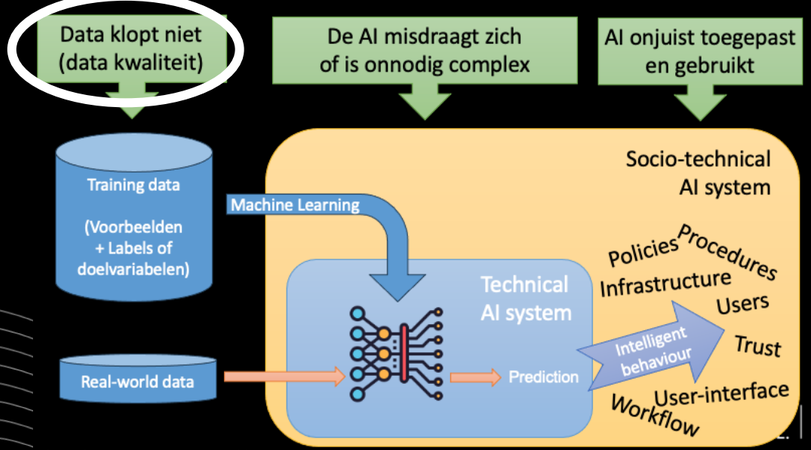
DATA KLOPT NIET: VOORSPELLEND MODEL KANKERDIAGNOSE





TECHMED CENTRE

Zoektocht naar labels van hoge kwaliteit



Meerdere bronnen
- Radiologierapporten
- Pathologierapporten
- Financiële codes (DBC)

Granulariteit
- Region-of-interest labels
- Per-beeld labels
- Per-patiënt labels
- Goed/kwaadaardig vs BIRADS

Ziekenhuis heeft doorgaans alleen per-patiënt labels
Bronnen conflicteren met elkaar

Selectie van EHR records

Wat wel en wat niet
- WEL: Diagnose
- NIET: Staging
- NIET: Na behandeling
- NIET: Follow-ups
- NIET: Recurrence

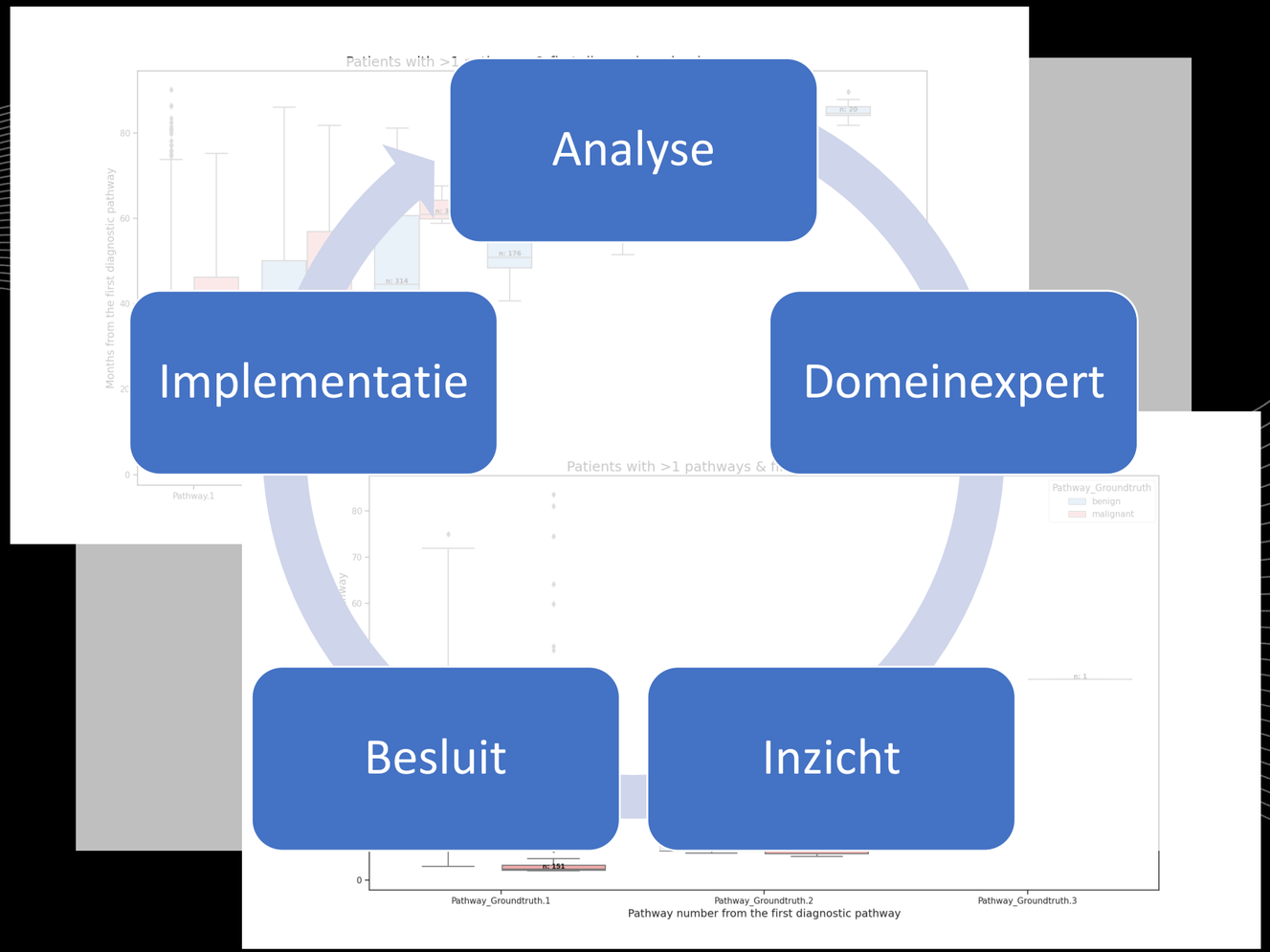
Geen life-cycle veld
Essentiële feiten verborgen in tekst
Conflicterende informatie

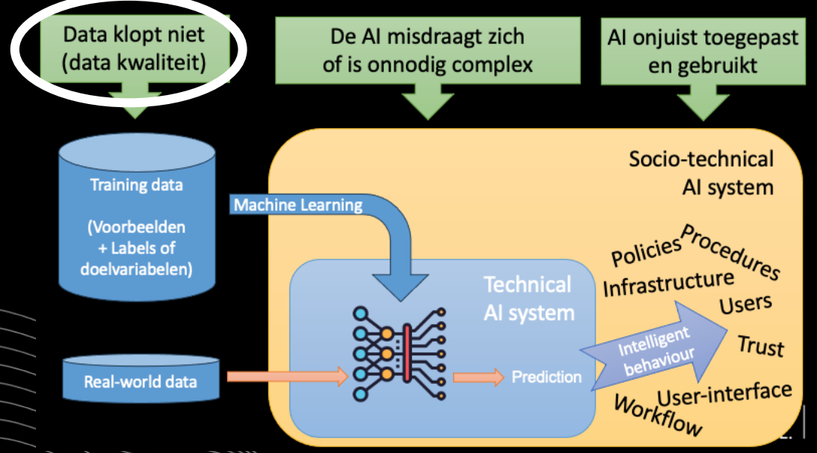
Beelden

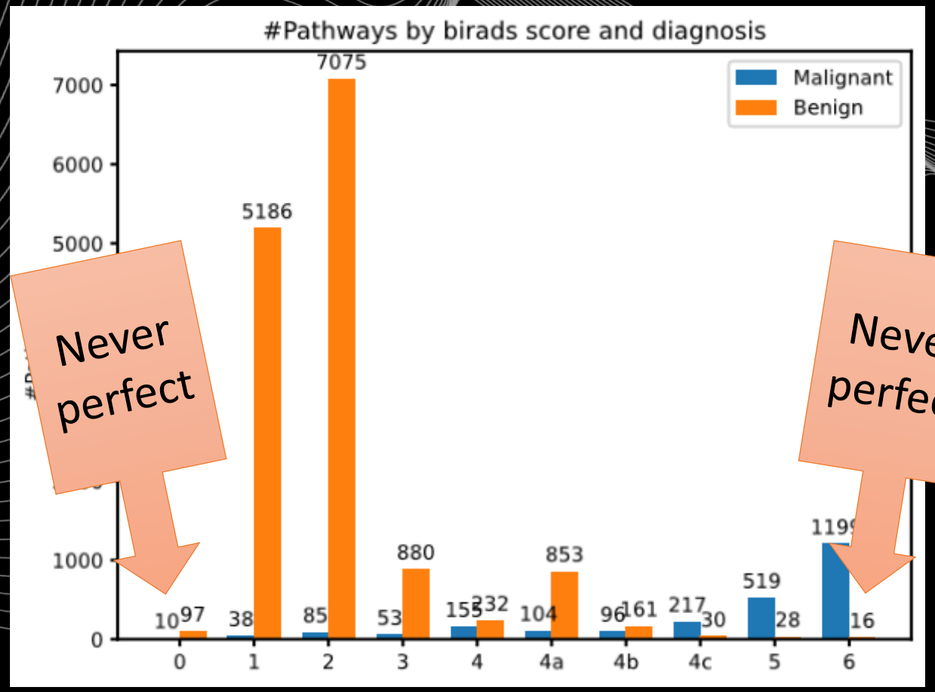
Automatische kwaliteitsverbetering
- Verwijderen tekst
- Borst uitsnijden
- Resolutie & contrast

93% standaard 4-beelden
7% niet!

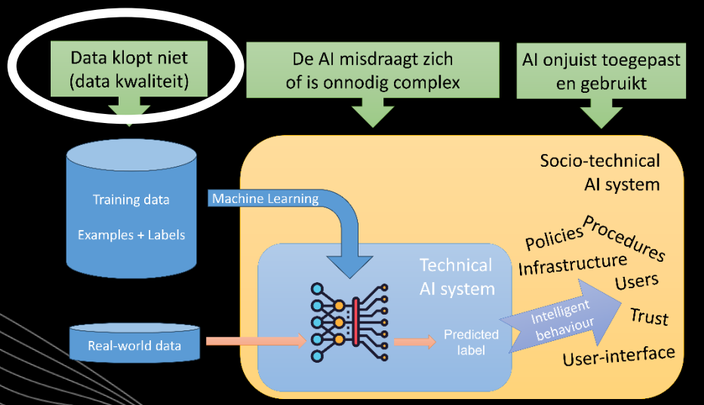
Methode







UNIVERSITY OF TWENTE. | TECHMED CENTRE



GEPERSONALISEERDE BEHANDELING KANKER-GERELATEERDE VERMOEIDHEID

Het PARTNR project

NWO/KWF Technology for Oncology grant

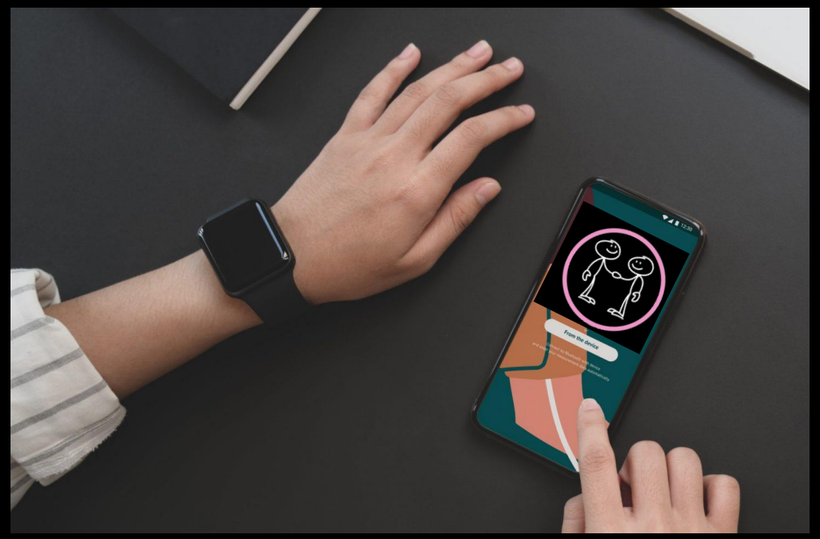
zgt · iKNL integraal kankercentrum Nederland · Roessingh Centrum voor Revalidatie · Helen Dowling Instituut Psychologische zorg bij kanker · Evidencio Medical Decision Support · IVIDO · Roessingh Research and Development · umcg

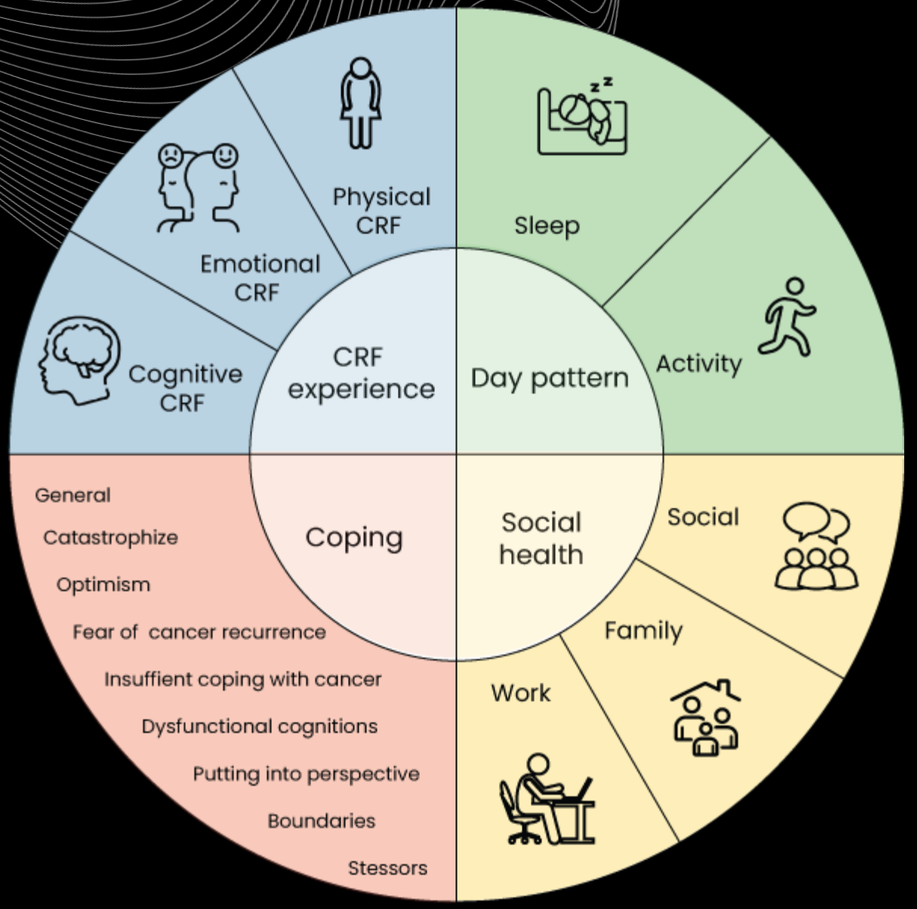
UNIVERSITY OF TWENTE. | TECHMED CENTRE

KANKER-GERELATEERDE VERMOEIDHEID

PARTNR

Personalized cAnceR TreatmeNt and caRe platform





Kim Wijlens, MSc

Lian Beenhakker, MSc

UNIVERSITY OF TWENTE. | TECHMED CENTRE

VOORSPELLEN VERMOEIDHEID

NIVEL

Klachten bij huisarts

Nederlandse Kanker Registratie (NKR)

Patiënt, tumor en behandelkarateristieken

Primary Secondary Cancer Care Registry (PSCCR)

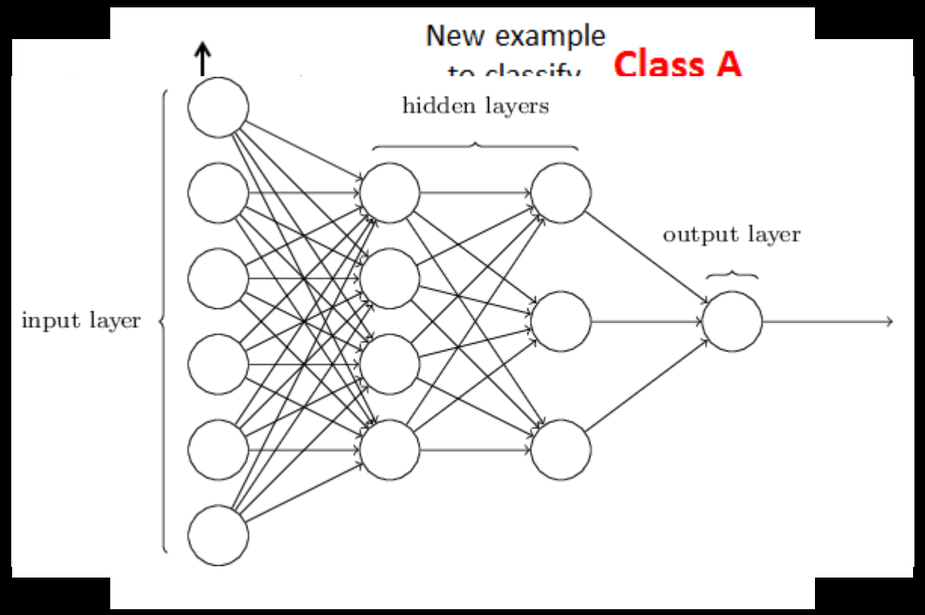
12,813 patiënten
2,224 vermoeid

Lian Beenhakker, MSc
University of Twente

UNIVERSITY OF TWENTE. | TECHMED CENTRE

VOORSPELLEN VERMOEIDHEID

Machine learning modellen



Random Forest Classifier

Logistic Regression (sklearn)

K-Nearest Neighbours

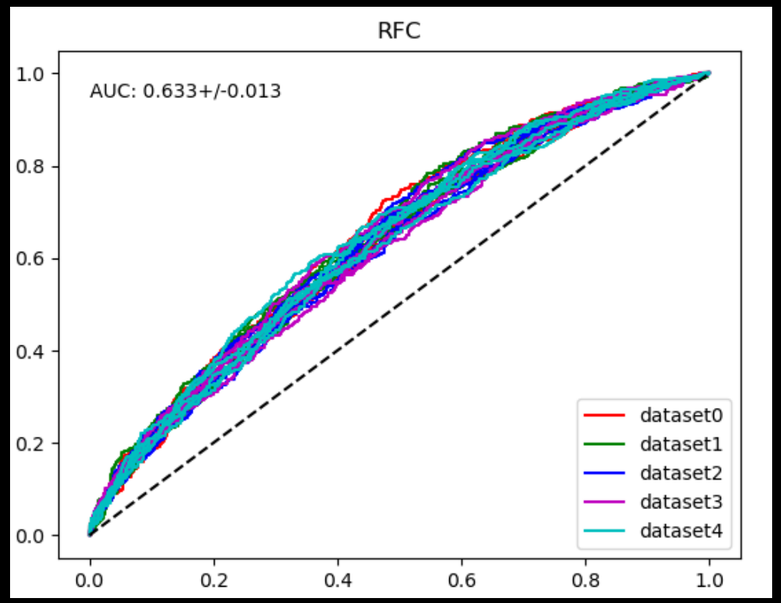
Gaussian Naive Bayes

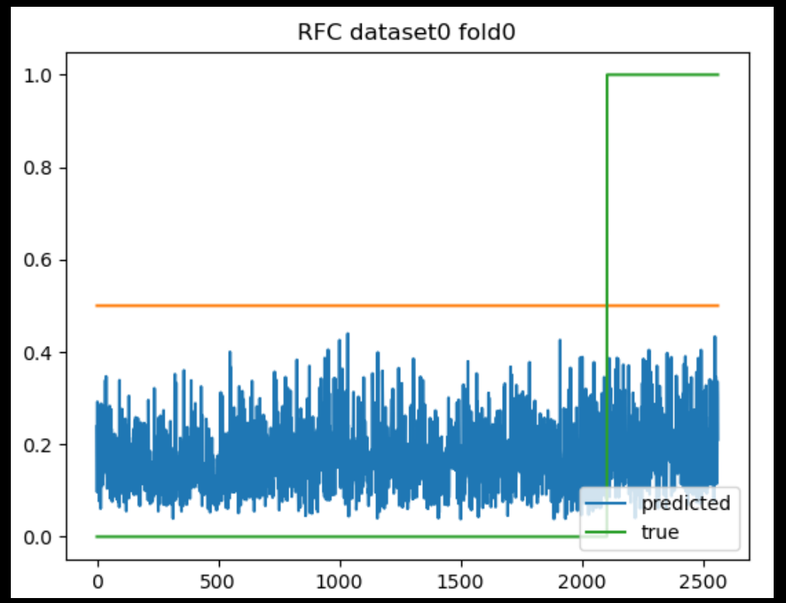
Multi Layer Perceptron

UNIVERSITY OF TWENTE. | TECHMED CENTRE

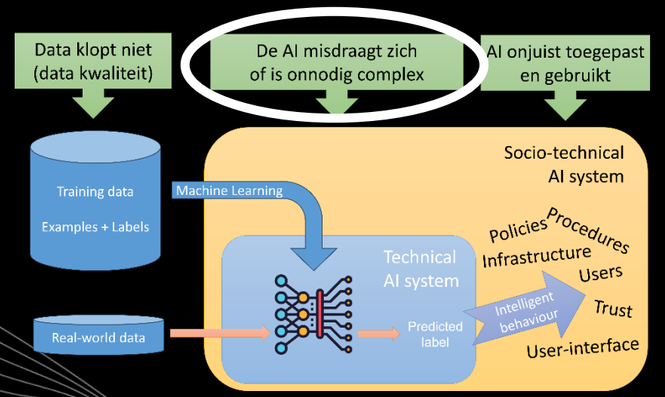
VOORSPELLEN VERMOEIDHEID

	AUC
Random Forest Classifier	0.633 ± 0.013
Logistic Regression	0.620 ± 0.010
K-Nearest Neighbours	0.577 ± 0.016
Gaussian Naive Bayes	0.546 ± 0.010
Multi Layer Perception	0.583 ± 0.026





UNIVERSITY OF TWENTE. | TECHMED CENTRE



INDIVIDUALIZED FOLLOW-UP FOR BREAST CANCER

Het INFLUENCE project

Netherlands comprehensive cancer organisation

UNIVERSITY OF TWENTE. | TECHMED CENTRE

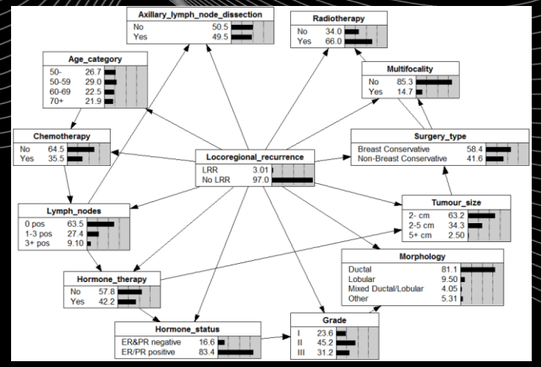
BORSTKANKER FOLLOW-UP

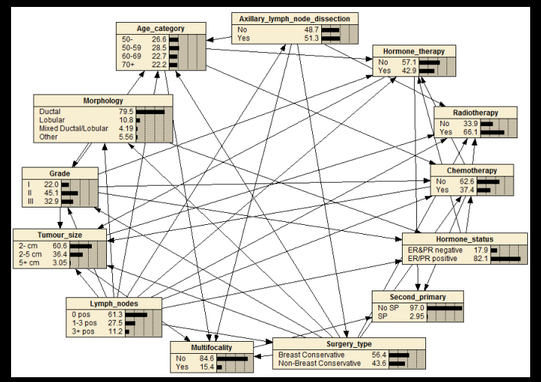
UNIVERSITY OF TWENTE. | TECHMED CENTRE

BORSTKANKER FOLLOW-UP

Logistische regressie vs Bayesian Networks

	n = 37,230, 950 LRRs		
	OR	95 % CI	*P*
Age			
<50	Ref.		
50–59	0.62	0.49–0.78	<0.001
60–69	0.61	0.47–0.79	<0.001
≥70	0.41	0.31–0.55	<0.001
Tumour size			
≤2 cm	Ref.		
2–5 cm	1.35	1.10–1.64	0.003
>5 cm	1.08	0.63–1.86	0.780
Nodal involvement			
0	Ref.		
1–3	1.64	1.32–2.04	<0.001
>3	2.90	2.14–3.94	<0.001
Grade of differentiation			
1	Ref.		
2	1.92	1.45–2.54	<0.001
3	2.96	2.16–4.05	<0.001
Hormone status			
Other	Ref.		
ER & PR negative	1.41	1.08–1.84	0.011
Multifocality			
No	Ref.		
Yes	1.23	0.99–1.54	0.062
Radiotherapy			
No	Ref.		
Yes	0.51	0.43-0.62	<0.001
Chemotherapy			
No	Ref.		
Yes	0.43	0.33–0.56	<0.001
Hormone therapy			
No	Ref.		
Yes	0.41	0.32-0.53	<0.001
Intercept			
	0.04	0.03–0.05	<0.001





Netwerk structuur leren uit de data:

- Bayesian Network Classifiers
- Constraint-based algorithms
- Score-based algorithms

UNIVERSITY OF TWENTE. | TECHMED CENTRE

INFLUENCE NOMOGRAM

Age
<50 | 50-59 | 60-69 | >70

Tumour size
<2 cm | 2-5 cm | >5 cm

Nodal involvement
0 nodes | 1-3 nodes | 3+ nodes

Grade
1 | 2 | 3

Hormone status: ER
Negative | Positive

Hormone status: PR
Negative | Positive

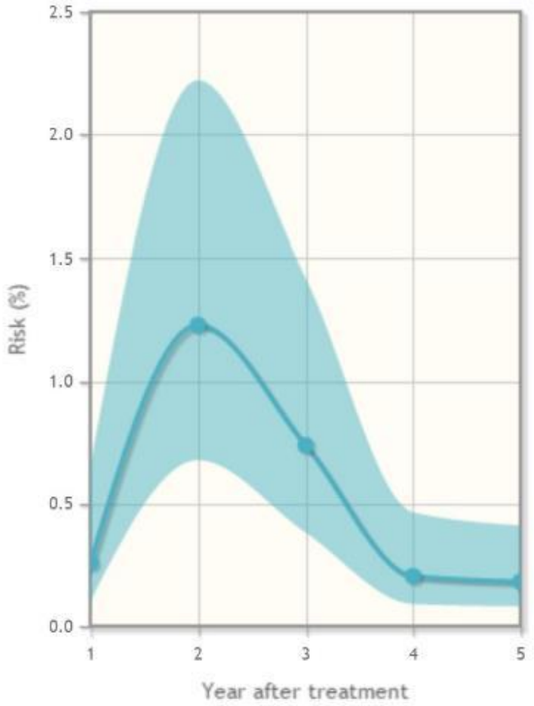
Multifocality
No | Yes

Radiotherapy
No | Yes

Chemotherapy
No | Yes

Hormone therapy
No | Yes

The 5 year risk is: 2.75% (2.01 - 3.75%)



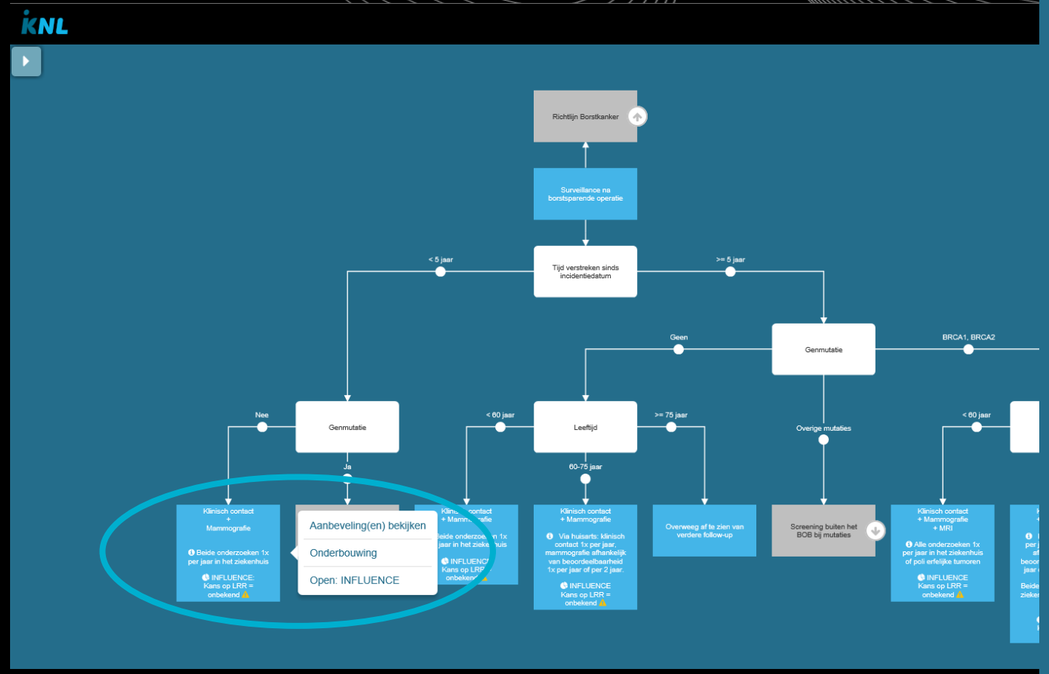
The risk per year is:
In year 1: 0.26% (0.10 - 0.66%)
In year 2: 1.23% (0.68 - 2.22%)
In year 3: 0.74% (0.38 - 1.42%)
In year 4: 0.20% (0.09 - 0.46%)
In year 5: 0.18% (0.08 - 0.41%)

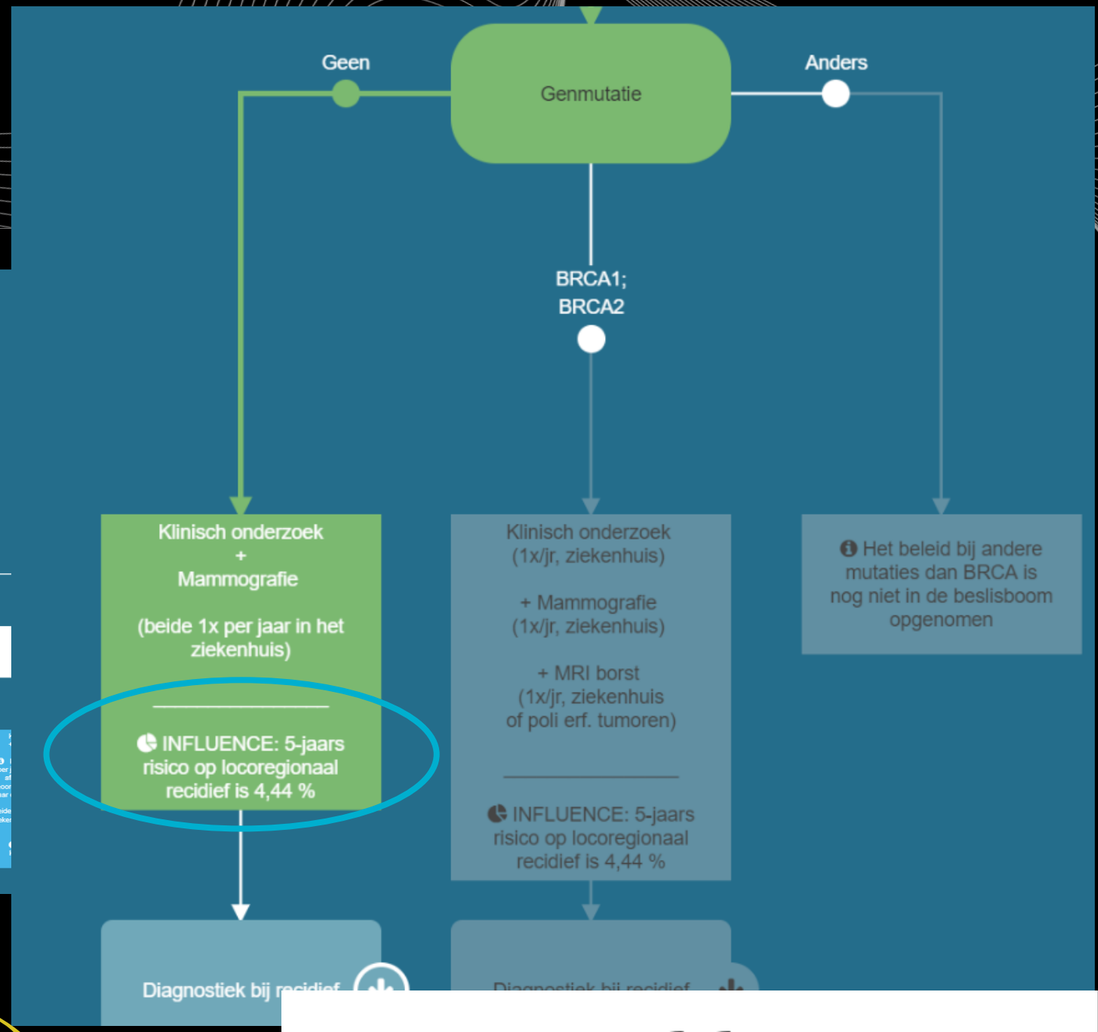
www.utwente.nl/INFLUENCE

UNIVERSITY OF TWENTE. | TECHMED CENTRE

INFLUENCE NOMOGRAM

Oncoguide richtlijn



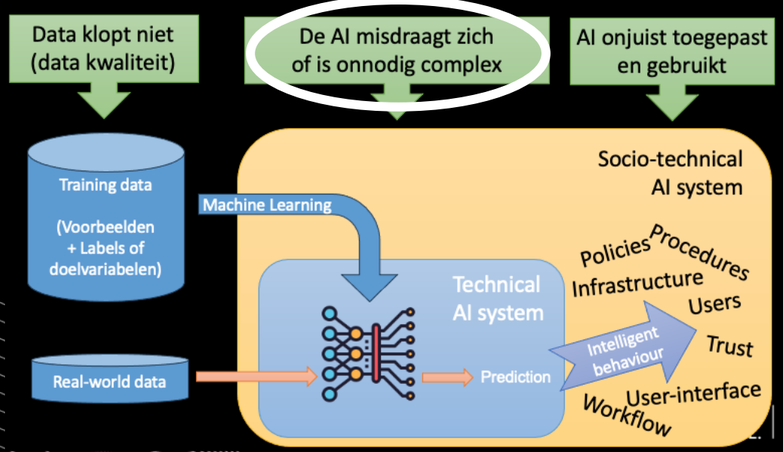


EVIDENCIO
MEDICAL DECISION SUPPORT

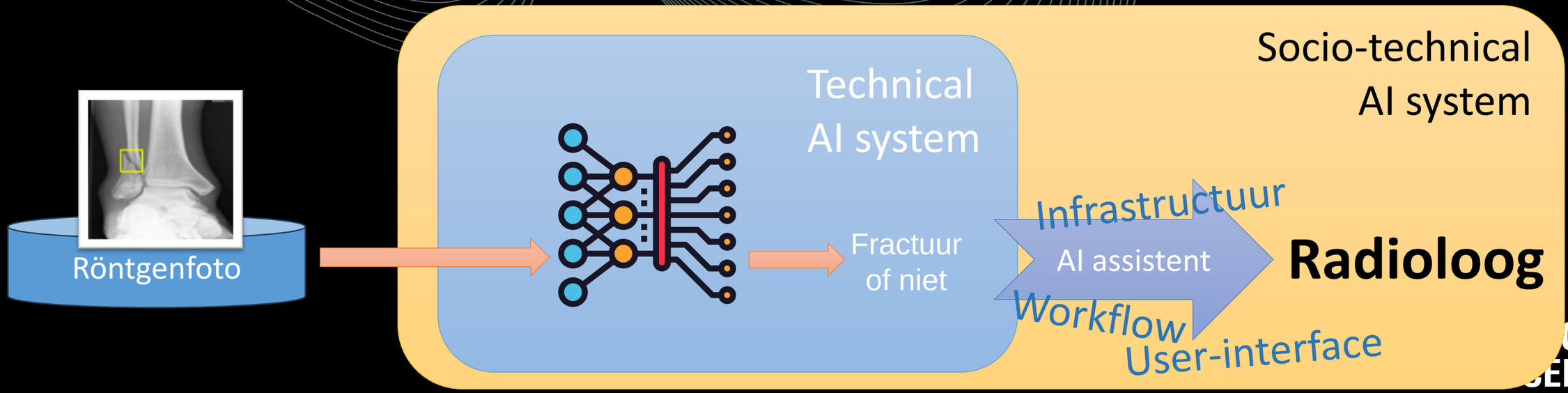
LOT V-2.0-2238.21.05.14 CE

UDI (01)08720299526440(8012)v2.0(4326)210514(240)2238

TECHMED CENTRE



AI MISDRAAGT ZICH: SHORT-CUTS EN NIET-KLINISCH REDENEREN IN FRACTUURDETECTIE



CONTRACTONDERZOEK ZGT

Onafhankelijke controle literatuur fractuurdetectie met real-world ZGT data van real-world ZGT machines en real-world mensen

Literatuur: 95% accuracy
ZGT data: verschillende machines, zoom-niveaus, implantaten

93% accuracy
(heupfracturen)

Hoera!
Het werkt!

… misschien toch een extra test

UNIVERSITY OF TWENTE. | TECHMED CENTRE

CONTRACTONDERZOEK ZGT

Onafhankelijke controle literatuur fractuurdetectie met real-world ZGT data van real-world ZGT machines en real-world mensen

Additionele validatie

- Fractuur uit de röntgenfoto wissen
- Plaatje zonder fractuur opnieuw in model

25% nog steeds positief?!?

Wat is hier aan de hand?

Attention-technieken: Model kijkt naar huid en ruimte tussen de benen?!?

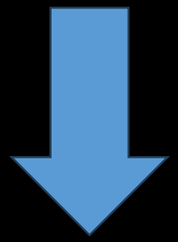
UNIVERSITY OF TWENTE. | TECHMED CENTRE

CONTRACTONDERZOEK ZGT

Onafhankelijke controle literatuur fractuurdetectie met real-world ZGT data van real-world ZGT machines en real-world mensen

Radioloog heeft zijn vermoeden

- Gerimpelde huid, luier, ontlasting duidt op bejaard persoon
- Als een bejaard persoon een röntgenfoto van heup laat maken, raad eens hoe vaak er dan sprake is van een heupfractuur?



short-cut reasoning

Model speelt vals: het gebruikt een niet-medische redenatie mbv ‘proxies’

- **Geeft hoge scores, ook bij standard validatie met een test set**
- **We zouden het niet geweten hebben als we deze extra test niet hadden gedaan!**

Voor ons aanleiding om onderzoek te doen naar Explainable AI

UNIVERSITY OF TWENTE. | TECHMED CENTRE

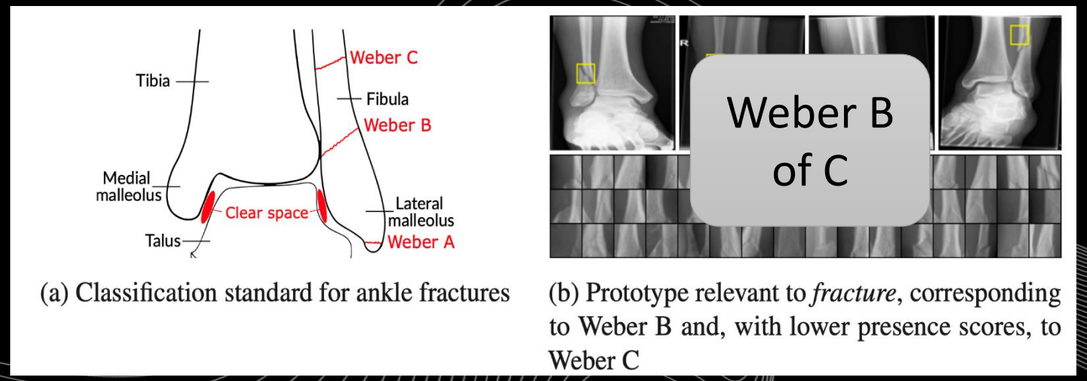
PIP-NET: AI EXPLAINABLE-BY-DESIGN

Deep learning aanpak dat een model leert dat mensen kunnen begrijpen én corrigeren

Meike Nauta, PhD

Interpretable Machine Learning

- Leert hoog-niveau ‘prototypes’ die mensen kunnen begrijpen
- Leert daarnaast een simpel model gebaseerd op die prototypes



(a) Classification standard for ankle fractures

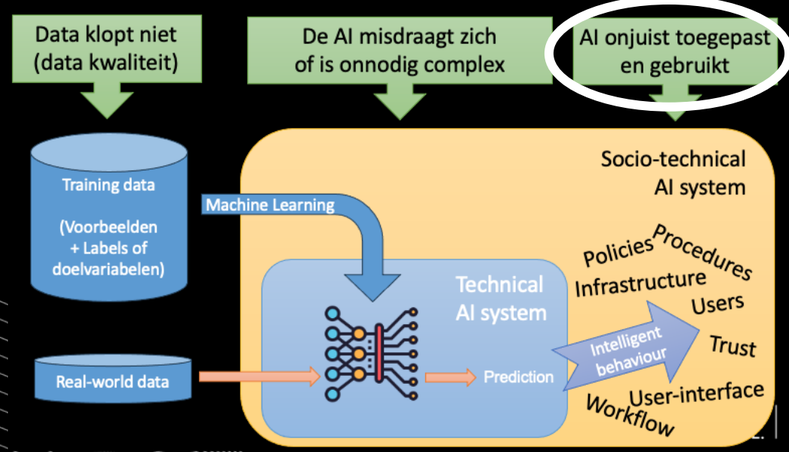
(b) Prototype relevant to *fracture*, corresponding to Weber B and, with lower presence scores, to Weber C

Enkelfracturen

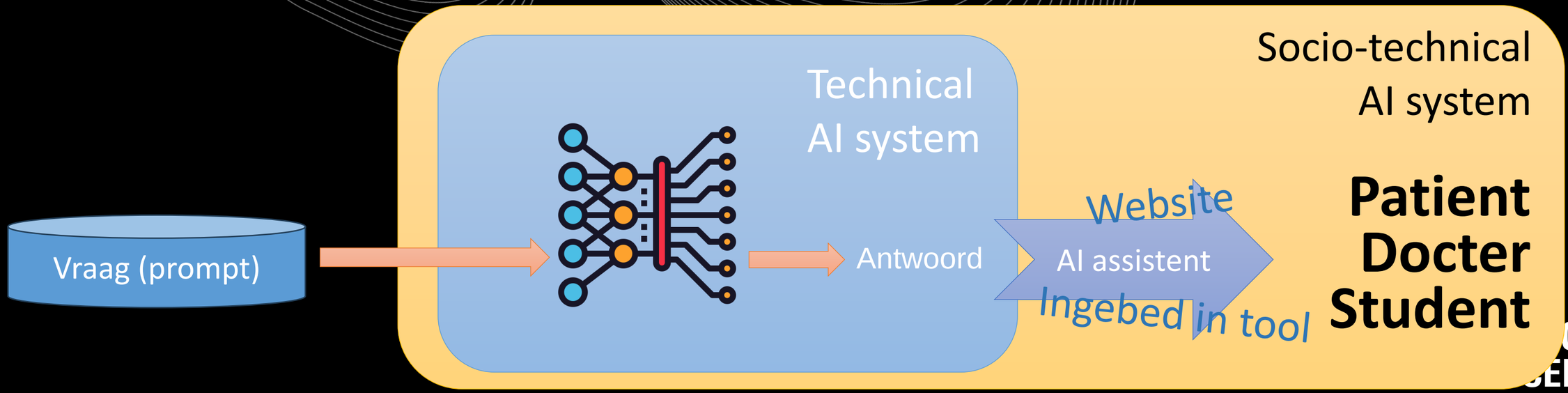
- PIP-Net leert prototypes die overeen-komen met medische standaarden
- Radiologen herkennen en verwijderen short-cuts

Medisch redeneren + uitleg zonder verlies in voorspellende kracht

UNIVERSITY OF TWENTE. | TECHMED CENTRE

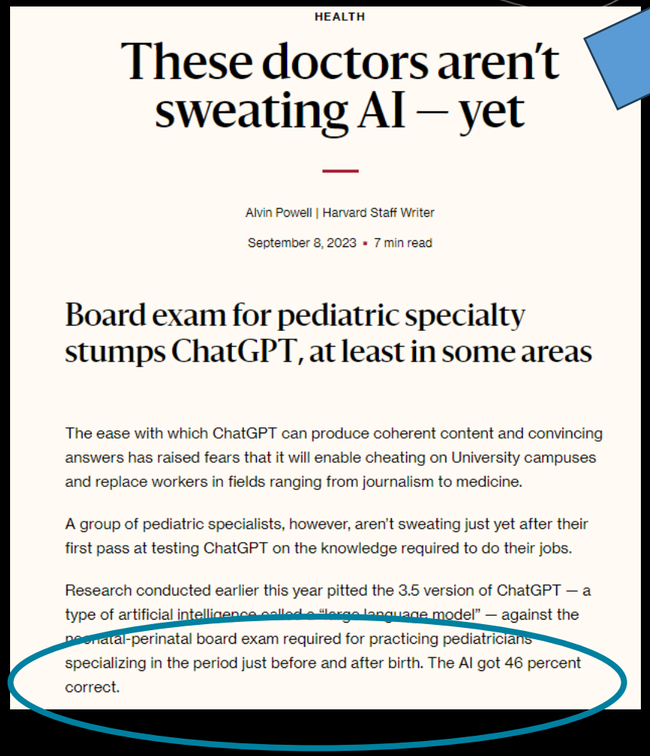


ONJUIST TOEGEPAST EN GEBRUIKT: CHAT-GPT



ChatGPT

37,5% in gastro-enterologie
78,5% in ethics



HEALTH

These doctors aren't sweating AI — yet

Alvin Powell | Harvard Staff Writer

September 8, 2023 ▪ 7 min read

Board exam for pediatric specialty stumps ChatGPT, at least in some areas

The ease with which ChatGPT can produce coherent content and convincing answers has raised fears that it will enable cheating on University campuses and replace workers in fields ranging from journalism to medicine.

A group of pediatric specialists, however, aren't sweating just yet after their first pass at testing ChatGPT on the knowledge required to do their jobs.

Research conducted earlier this year pitted the 3.5 version of ChatGPT — a type of artificial intelligence called a "large language model" — against the neonatal-perinatal board exam required for practicing pediatricians specializing in the period just before and after birth. The AI got 46 percent correct.

⅓ ongepast advies
⅛ hallucinaties

Maar dus ook ⅔ gepast advies
Heel overtuigende tekst
Moeilijk op kwaliteit te schatten

HEALTH

Need cancer treatment advice? Forget ChatGPT

| BWH Communications

August 29, 2023 ▪ 4 min read

New research finds in about third of cases AI chatbot provided medically inappropriate recommendations

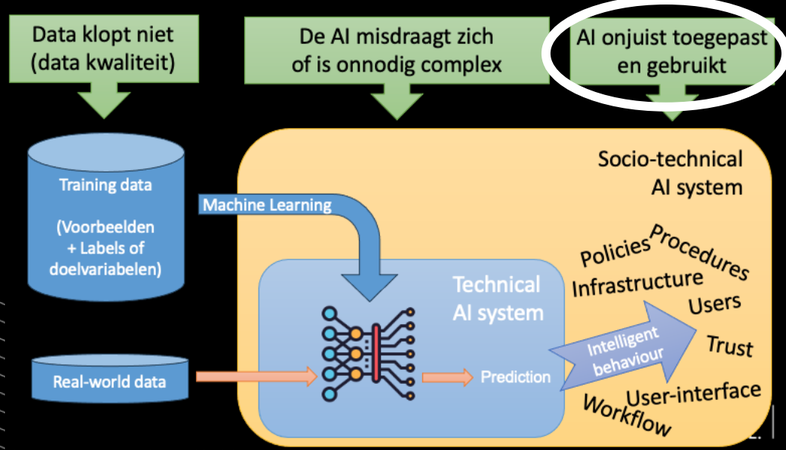
The internet can serve as a powerful tool for self-education on medical topics. With ChatGPT now at patients' fingertips, researchers from Brigham and Women's Hospital sought to assess how consistently the AI chatbot provides recommendations for cancer treatment that align with National Comprehensive Cancer Network **guidelines**.

The team's **findings**, published in JAMA Oncology, show that in one-third of cases, ChatGPT provided an inappropriate — or "non-concordant" — recommendation, highlighting the need for awareness of the technology's limitations.

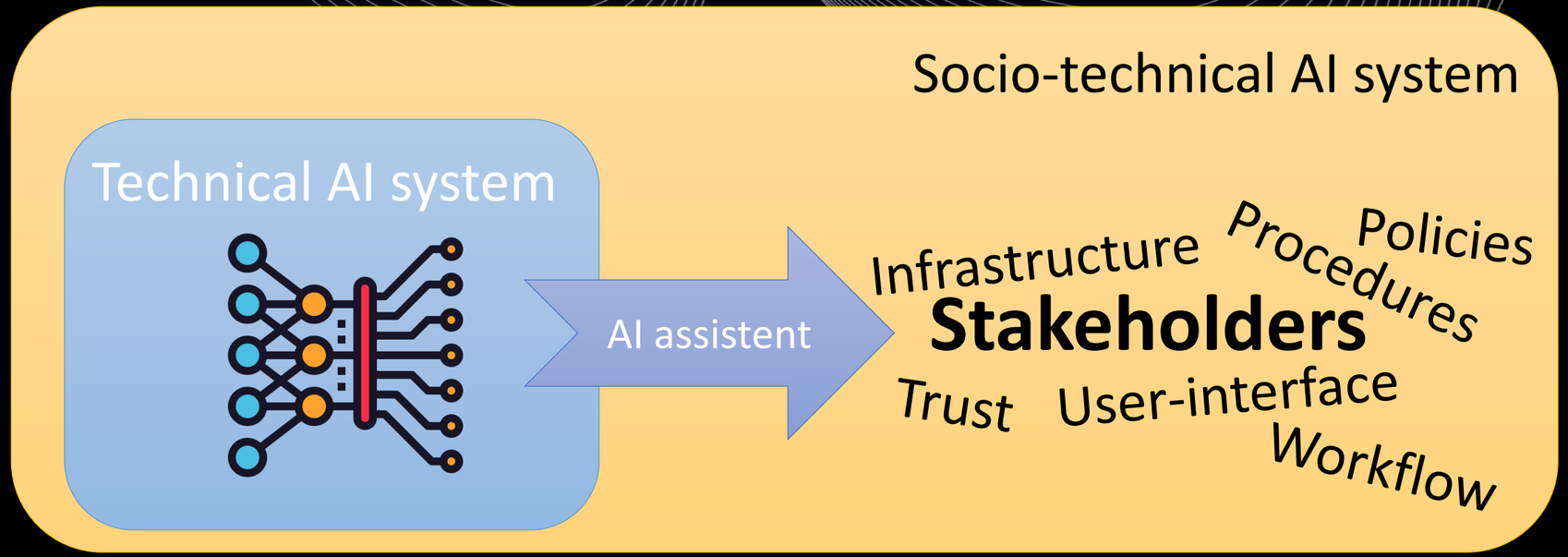
https://news.harvard.edu/gazette/story/2023/09/these-doctors-arent-sweating-ai-yet
https://news.harvard.edu/gazette/story/2023/08/need-cancer-treatment-advice-forget-chatgpt/

UNIVERSITY OF TWENTE. | TECHMED CENTRE

JUIST TOEGEPAST EN GEBRUIKT: Socio-technical design



- Ontwerp AI als een socio-technisch systeem
- Co-creatie met alle stakeholders
- Bediscussieer ethische vraagstukken vanaf het begin (AI frameworks)



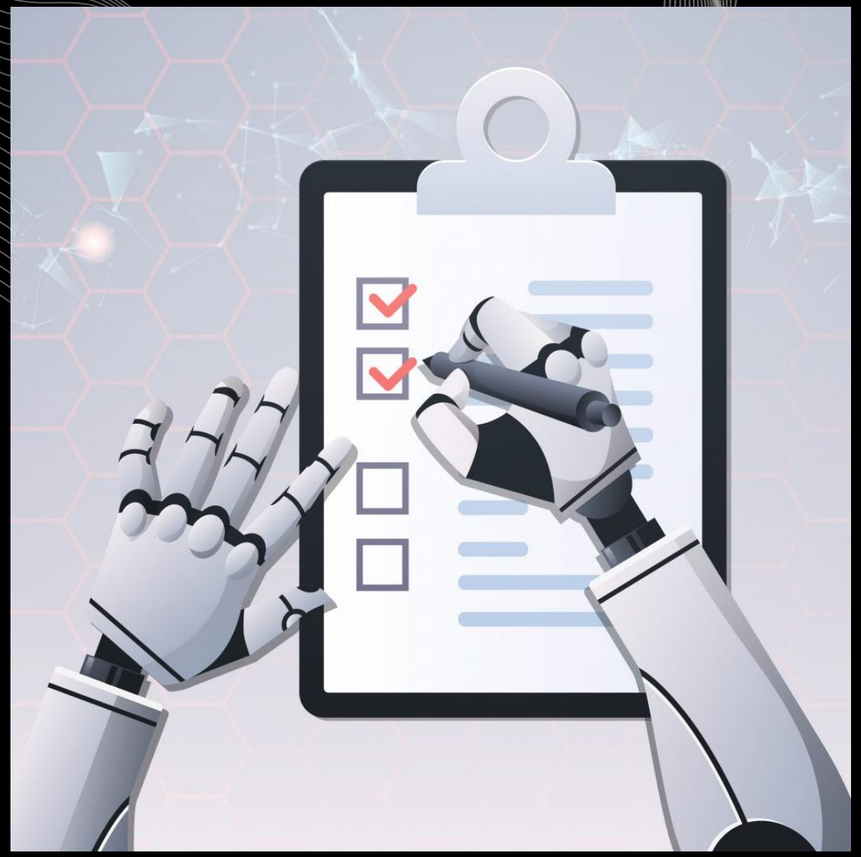
UNIVERSITY OF TWENTE. | TECHMED CENTRE

TO AI OR NOT TO AI??

UNIVERSITY OF TWENTE. | TECHMED CENTRE

TO AI OR NOT TO AI?

Ja, natuurlijk! Grote beloftes!
… maar dan wel goed!



- Data van hoge kwaliteit met de juiste informatie is noodzakelijk maar moeilijk te verkrijgen
- Gebruik real-world data ‘as is’
- Waak voor onrealistische simplificatie
- Niet alles benodigt deep learning
- Waak voor short-cuts en bias (Explainable AI)
- Socio-technical design & co-creatie
- Doel is `AI als assistent’, niet hoogste accuracy

UNIVERSITY OF TWENTE. | TECHMED CENTRE

UNIVERSITY OF TWENTE. | TECHMED CENTRE

FAILURES: TO AI OR NOT TO AI

Dr. Annemieke Witteveen A.Witteveen@utwente.nl
Prof. Dr. Ir. Maurice van Keulen M.vanKeulen@utwente.nl