Artificial Intelligence in health(care)
Guidance for assessing, monitoring and auditing

Carl Moons

Prof. Clinical Epidemiology, Julius Center, UMC Utrecht
Directeur Health Innovation Netherlands (HI-NL)
Figure head 'AI for Health', UMC Utrecht/Utrecht University
Project leader Government initiative 'Quality of AI in Health'

UMC Utrecht

Health(care) challenges

Rising Demand

Aging population

Increasing Costs

Clinical Shortage

UMC Utrecht

MedTech

>600k devices

>900k health apps/agorithms

1900

2019

And then … it became 2020

And then … there was suddenly AI to battle COVID-19

Search

European Commission › Strategy › Shaping Europe's digital future › News ›

Shaping Europe's digital future

NEWS ARTICLE | 19 May 2020

Using AI to fast and effectively diagnose COVID-19 in hospitals

The European Commission will invest in the use of Artificial Intelligence to speed up the diagnosis of COVID-19 and improve the future treatment of patients. A software developed to assist the work of medical staff by analysing images of pulmonary infections is introduced in 10 hospitals across Europe.

About Artificial intelligence

- Policies
- Blog posts
- News

Infervision's AI is in Italy Helping to Battle COVID-19

PRESS RELEASE UPDATED: MAR 23, 2020

ROME, March 20, 2020 (Newswire.com) - COVID-19 is spreading, with European countries already declaring a pandemic. The World Health Organization has declared Europe as the new 'epicenter' for COVID-19. Italy announced a full lock-down on March 10. Due to the spreading of COVID-19, Italian medical institutions are facing tremendous pressure as patient numbers surge. Meanwhile, issues over long turnaround times for PCT testing and limited availabilities of the kit are concerning. Using CT images will help with the screening of COVID-19.

New tool could 'help UK doctors spot high-risk Covid patients in seconds'

Study claims risk calculator will help clinicians with expected influx of patients this autumn

- Coronavirus - latest updates
- See all our coronavirus coverage

▲ The calculator was tested in a hospitalised elderly population, so is not applicable for use within the community. Photograph: Murdo MacLeod/The Guardian

score of three or less. The tool, which is easily accessible on a smartphone or computer, takes seconds to generate a score and is expected to be rolled out in the NHS this week.

COVID-19 decision algorithms

In <15 mths over 1200 COVID-19 algorithms - >70% AI based

RESEARCH

FAST TRACK

Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal

Laure Wynants,[1,2] Ben Van Calster,[2,3] Gary S Collins,[4,5] Richard D Riley,[6] Georg Heinze,[7] Ewoud Schuit,[8,9] Elena Albu,[2] Banafsheh Arshi,[1] Vanesa Bellou,[10] Marc M J Bonten,[8,11] Darren L Dahly,[12,13] Johanna A Damen,[8,9] Thomas P A Debray,[8,14] Valentijn M T de Jong,[8,9] Maarten De Vos,[2,15] Paula Dhiman,[4,5] Joie Ensor,[6] Shan Gao,[2] Maria C Haller,[7,16] Michael O Harhay,[17,18] Liesbet Henckaerts,[19,20] Pauline Heus,[8,9] Jeroen Hoogland,[8] Mohammed Hudda,[21] Kevin Jenniskens,[8,9] Michael Kammer,[7,22] Nina Kreuzberger,[23] Anna Lohmann,[24] Brooke Levis,[6] Kim Luijken,[24] Jie Ma,[5] Glen P Martin,[25] David J McLernon,[26] Constanza L Andaur Navarro,[8,9] Johannes B Reitsma,[8,9] Jamie C Sergeant,[27,28] Chunhu Shi,[29] Nicole Skoetz,[22] Luc J M Smits,[1] Kym I E Snell,[6] Matthew Sperrin,[30] René Spijker,[8,9,31] Ewout W Steyerberg,[3] Toshihiko Takada,[8,32] Ioanna Tzoulaki,[10,33] Sander M J van Kuijk,[34] Bas C T van Bussel,[1,35] Iwan C C van der Horst,[35] Kelly Reeve,[36] Florien S van Royen,[8] Jan Y Verbakel,[37,38] Christine Wallisch,[7,39,40] Jack Wilkinson,[24] Robert Wolff,[41] Lotty Hooft,[8,9] Karel G M Moons,[8,9] Maarten van Smeden[8]

For numbered affiliations see end of the article

Correspondence to: L Wynants laure.wynants@maastrichtuniversity.nl (ORCID 0000-0002-3037-122X)

Additional material is published online only. To view please visit the journal online.

Abstract

OBJECTIVE

To review and appraise the validity and usefulness of published and preprint reports of prediction models for prognosis of patients with covid-19, and for detecting people in the general population at increased risk of covid-19 infection or being admitted

DATA EXTRACTION

At least two authors independently extracted data using the CHARMS (critical appraisal and data extraction for systematic reviews of prediction modelling studies) checklist; risk of bias was assessed using PROBAST (prediction model risk of bias assessment tool).

RESULTS

COVID-19 left – AI stayed

JNM The Journal of NUCLEAR MEDICINE

HOT TOPICS

Machine Learning Will Change Medicine

Michael Forsting

Essen University Hospital, University of Essen-Duisburg, Essen, Germany

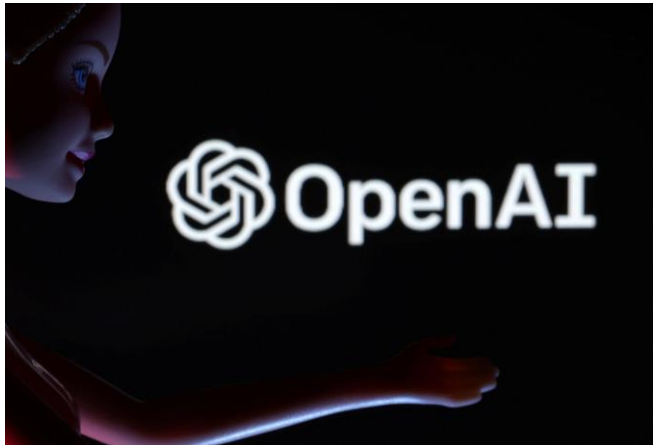
Artificial Intelligence in Health Care

The Hope, the Hype, the Promise, the Peril

Doctors about to be replaced by hospital AI systems offering better diagnosis and less arrogance

09/12/2017 / By Jhoanna Robinson

Framingham, Massachusetts-based market intelligence provider IDC Health Insights, in its recently published report on artificial intelligence and cognitive computing adoption in the Asia/Pacific titled *IDC Peerscape: Cognitive/AI Practices for Healthcare in Asia/Pacific (Excluding Japan)*, stated the best possible healthcare solutions that hospitals and health insurance companies all around the Asia-Pacific countries should adopt.



REFLECTIONS ON HEALTHCARE LEADERSHIP ETHICS

Healthcare uses of artificial intelligence: Challenges and opportunities for growth

Eric Racine, PhD[1,2,3,4]; Wren Boehlen, BSc[1]; and Matthew Sample, PhD[1,2]

Healthcare Management Forum 2019, Vol. 32(5) 272-275

DOI: 10.1177/0840470419843831
journals.sagepub.com/home/hmf
SAGE

AI will be everywhere – no mayfly or one trick pony

Short Intermezzo - AI Quiz

Can you guess the publication year?

1. History of Artificial Intelligence

1. Artificial Intelligence: The Time is Now.

1. Artificial intelligence techniques for diagnostic reasoning in medicine

1. The evaluation of artificial intelligence systems in medicine.

Quiz

Can you guess the publication year?

1. History of Artificial Intelligence
   (McCorduck et al., 1977)

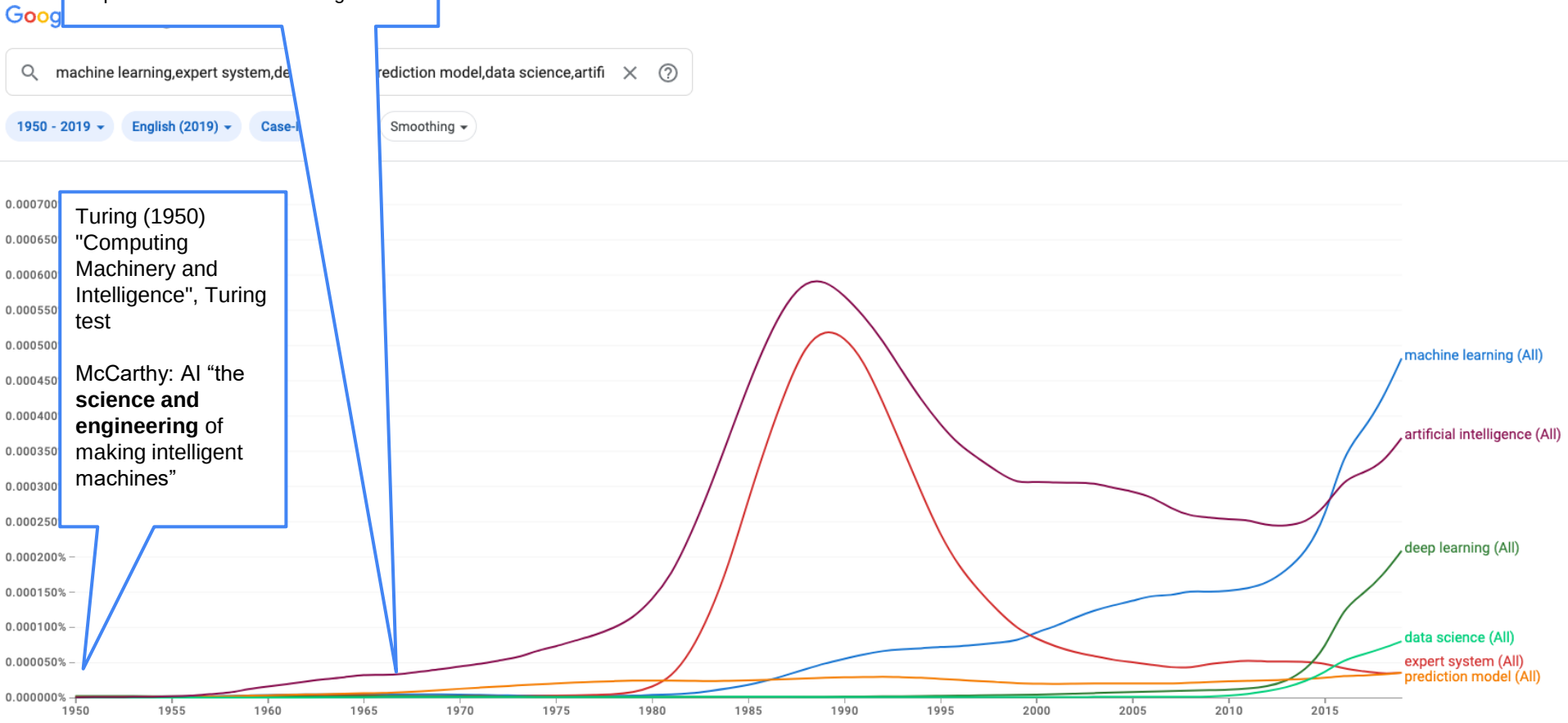
1. Artificial Intelligence: The Time is Now.
   (R.L. Dilworth, 1988)

1. Artificial intelligence techniques for diagnostic reasoning in medicine
   (R.S., Patil, 1988)

1. The evaluation of artificial intelligence systems in medicine.
   (P.L. Miller; 1986)

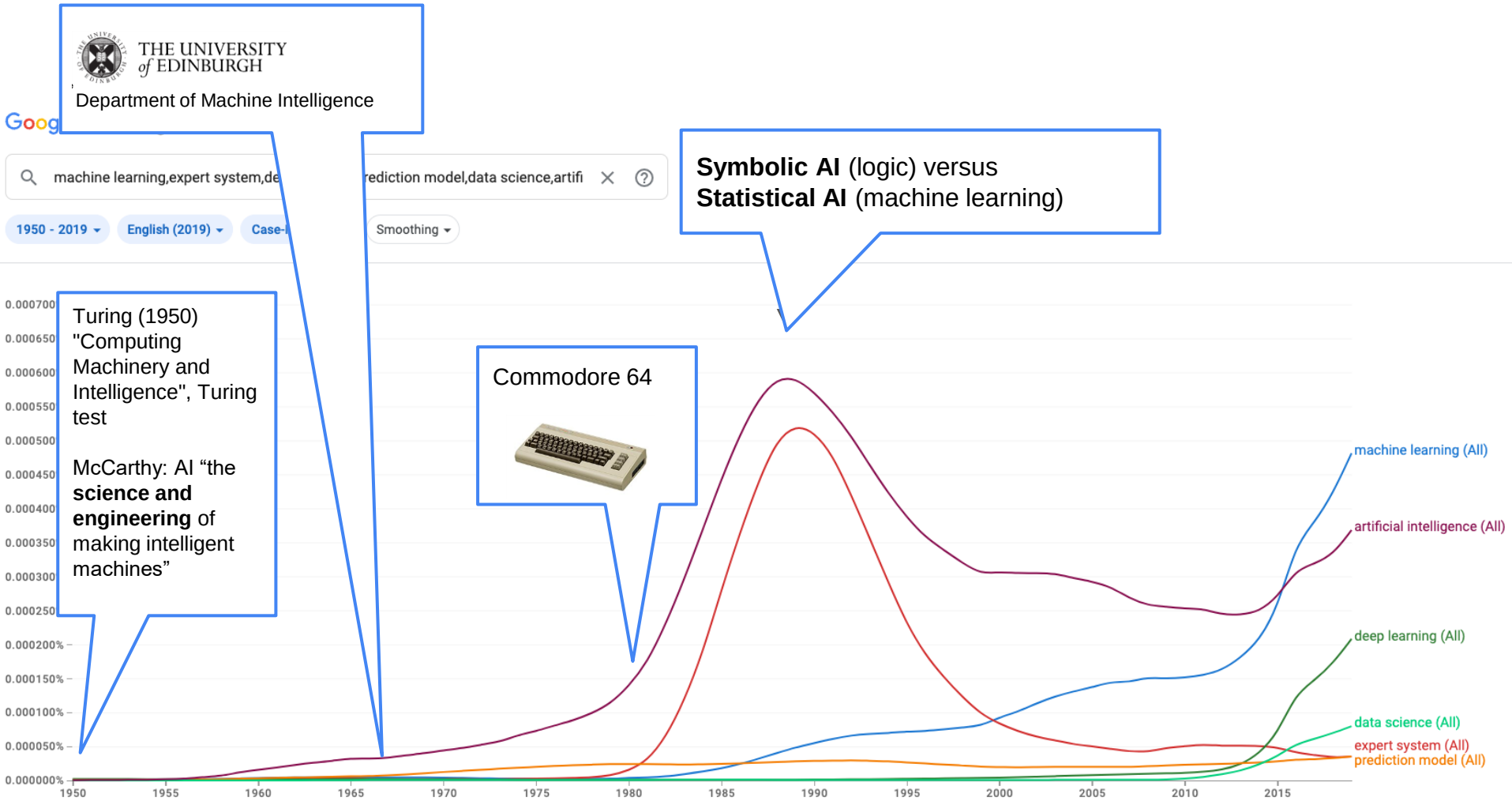
THE UNIVERSITY of EDINBURGH

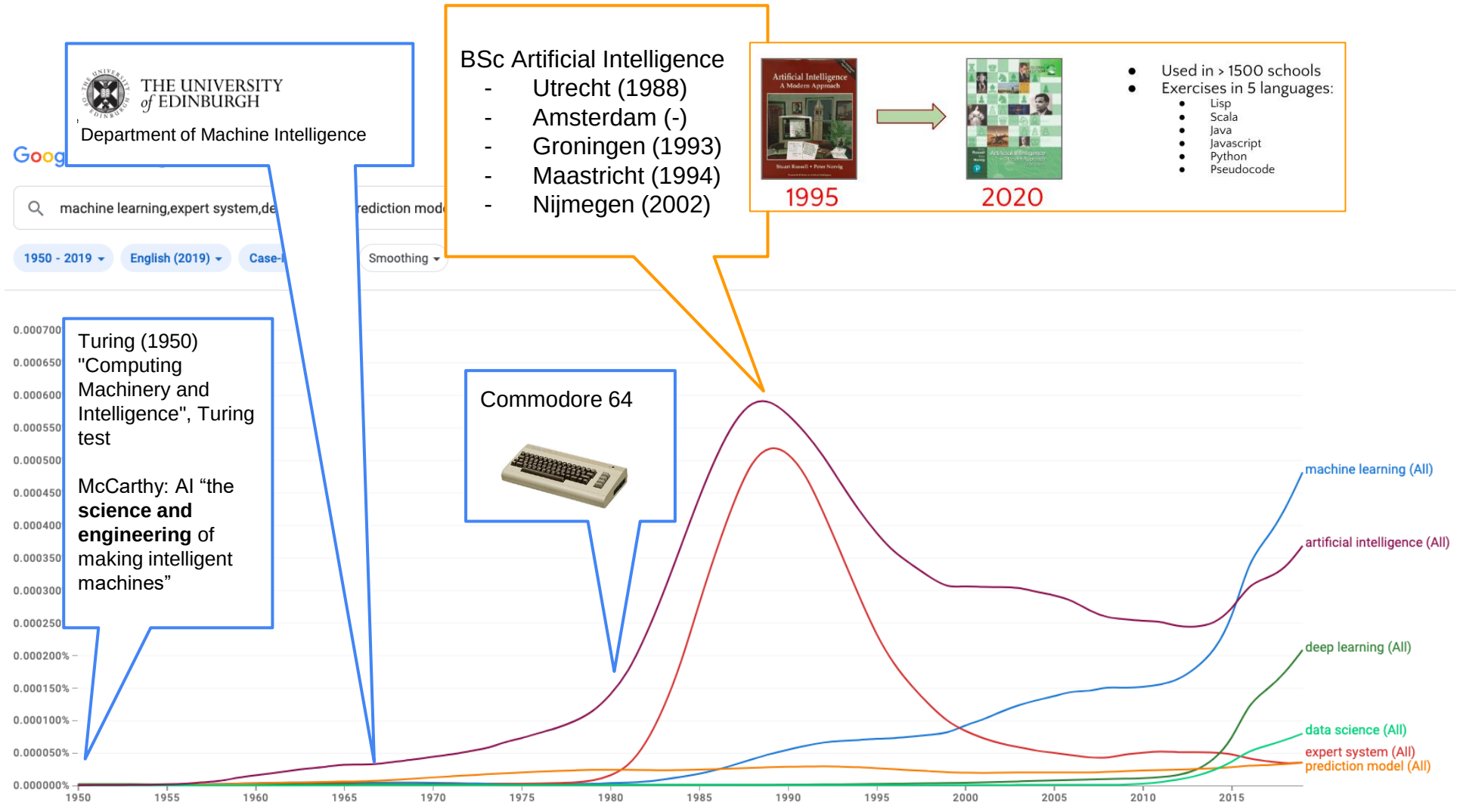
Department of Machine Intelligence



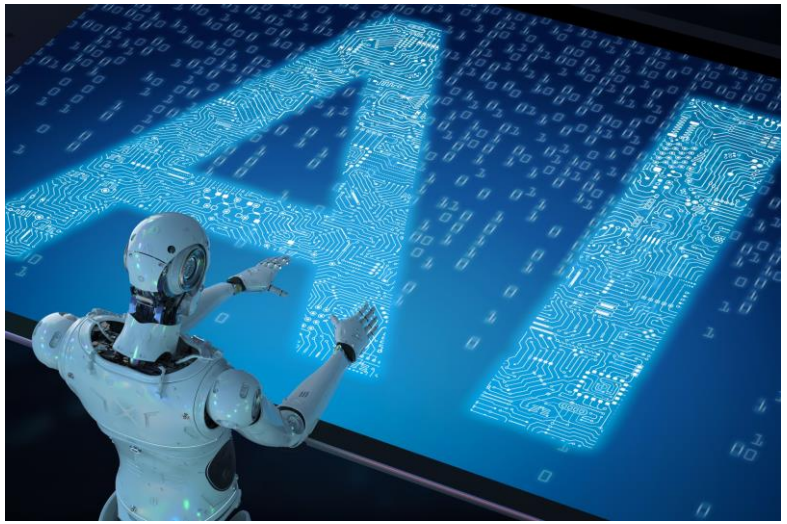
Turing (1950) "Computing Machinery and Intelligence", Turing test

McCarthy: AI “the **science and engineering** of making intelligent machines”


THE UNIVERSITY of EDINBURGH
Department of Machine Intelligence
Google
machine learning,expert system,de
rediction model,data science,artifi
1950 - 2019
English (2019)
Case-
Smoothing
Symbolic AI (logic) versus
Statistical AI (machine learning)
Turing (1950) "Computing Machinery and Intelligence", Turing test
McCarthy: AI “the science and engineering of making intelligent machines”
Commodore 64
0.000700
0.000650
0.000600
0.000550
0.000500
0.000450
0.000400
0.000350
0.000300
0.000250
0.000200%
0.000150%
0.000100%
0.000050%
0.000000%
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
machine learning (All)
artificial intelligence (All)
deep learning (All)
data science (All)
expert system (All)
prediction model (All)


THE UNIVERSITY of EDINBURGH
Department of Machine Intelligence
BSc Artificial Intelligence
- Utrecht (1988)
- Amsterdam (-)
- Groningen (1993)
- Maastricht (1994)
- Nijmegen (2002)
1995
2020
Used in > 1500 schools
Exercises in 5 languages:
Lisp
Scala
Java
Javascript
Python
Pseudocode
machine learning,expert system,de
1950 - 2019
English (2019)
Smoothing
Turing (1950) "Computing Machinery and Intelligence", Turing test
McCarthy: AI “the **science and engineering** of making intelligent machines”
Commodore 64
0.000700
0.000650
0.000600
0.000550
0.000500
0.000450
0.000400
0.000350
0.000300
0.000250
0.000200%
0.000150%
0.000100%
0.000050%
0.000000%
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
machine learning (All)
artificial intelligence (All)
deep learning (All)
data science (All)
expert system (All)
prediction model (All)

What is AI?



Human intelligence artificialized (computerised)

- Image/seeing to text - v.v.
- Sound/hearing to text – v.v.
- Detecting patterns in data/images/sounds
- Multiple data sources/types combined+translated to, e.g., a probability → forecasting or detecting (e.g. weather, traffic, stocks, health)
- Etc. – our brain is great!

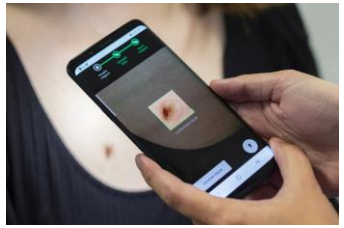
2 main types of AI in health(care) domain

1. To support in the decision making of health providers, patients and citizens

2. To support healthcare processes/efficiency/workload reduction

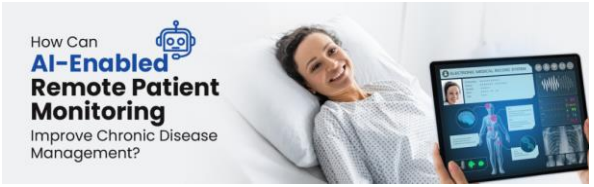
1. AI to support decision making → health providers, patients, citizens

- Detection - diagnoses / screening



- Prognosis – predicting future health outcomes, treatment effects

- Monitoring – disease progress, therapy response, remote



- Lifestyle/exercise planning

- Therapy conduct

1. AI to sup

- 10.000
  - Very

- AI is goir

SITEMAN CANCER CENTER®

OPINION
published: 26 January 2022
doi: 10.3389/fdgth.2022.833912

Check for updates

Artificial Intelligence and Statistics: Just the Old Wine in New Wineskins?

Livia Faes [1,2], Dawn A. Sim [1,3,4], Maarten van Smeden [5], Ulrike Held [6], Patrick M. Bossuyt [7] and Lucas M. Bachmann [2]*

[1] Medical Retina Department, Moorfields Eye Hospital NHS Foundation Trust, London, United Kingdom, [2] Medignition Inc., Research Consultants, Zurich, Switzerland, [3] Health Data Research UK, London, United Kingdom, [4] National Institute for Health Research (NIHR) Biomedical Research Centre for Ophthalmology, Moorfields Eye Hospital National Health Service (NHS) Foundation Trust and University College London (UCL) Institute of Ophthalmology, London, United Kingdom, [5] Julius Center for Health Science and Primary Care, University Medical Center Utrecht, University of Utrecht, Utrecht, Netherlands, [6] Department of Biostatistics, Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland, [7] Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam Public Health Research Institute, Amsterdam University Medical Centers, Amsterdam, Netherlands

Keywords: artificial intelligence (AI), machine learning (ML), statistics, methodology, reporting guideline

COVID-19 Reso

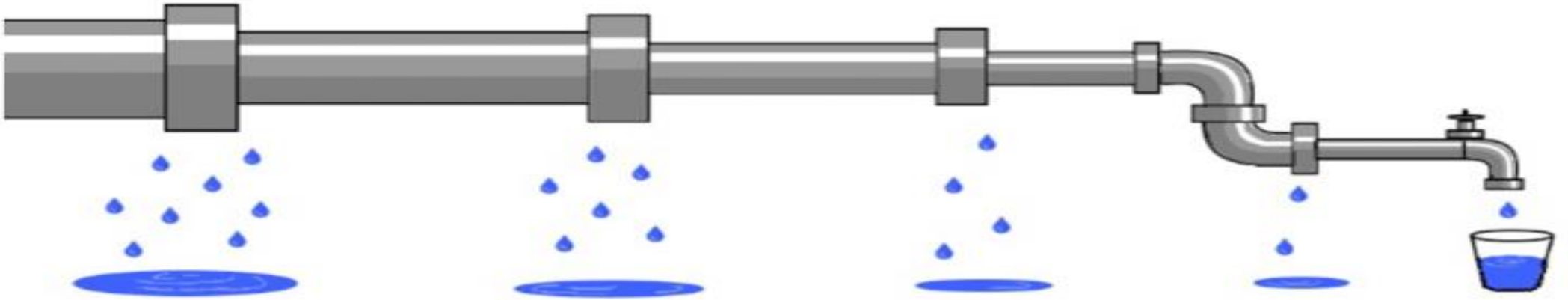
Log in

16:14

A new predi ealth(care)
is developed

…every **1.5**

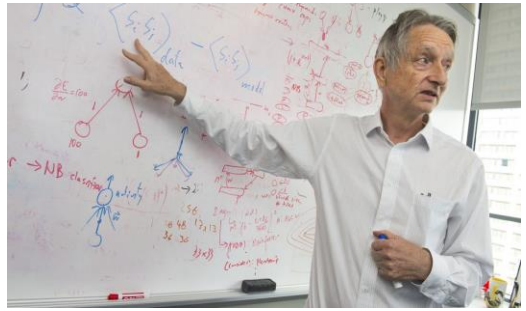
Leaky pipeline prediction algorithms in health(care)



Not fit for purpose	No validation	No implementation	Not adopted
Developed on wrong patient population	Lack of data or incentive to pursue validation studies	No impact on decision making or patient (health) outcomes	Prediction (perceived as) not useful
Expensive or non-available predictors	Incompletely reported prediction model	No software developed to implement and use the model	Predictions not trusted
Time intensive to use model	Poorly developed or overfitted model	Requirements for adherence to (medical device) regulations	Model not transparent enough, or no tools available to enhance its use in practice
Outcome measured unreliably	Proprietary model code	Cost(-effectiveness) of use proprietary model	Model (perceived as) outdated

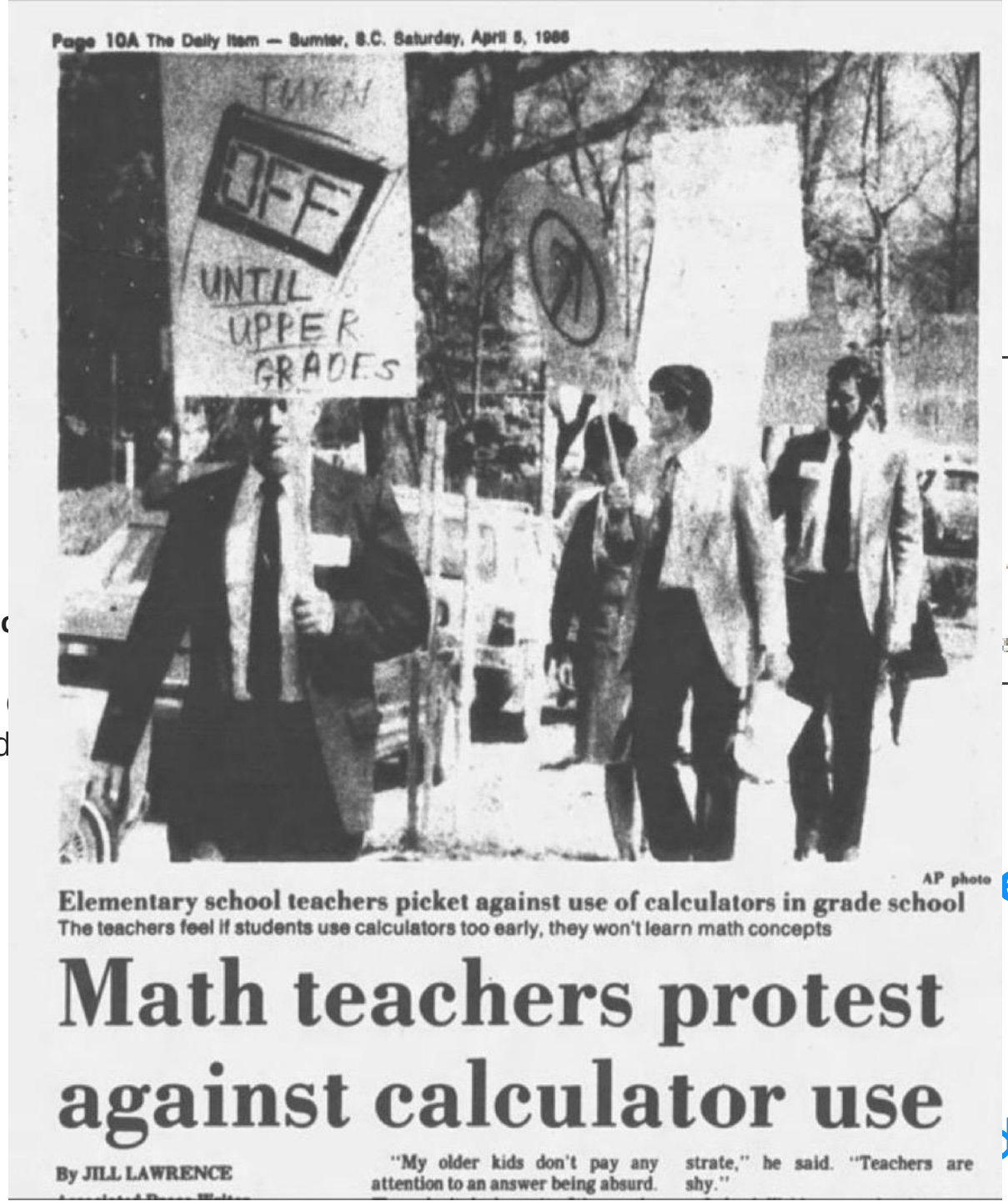
2. AI to reduce workload and administrative burden – generative AI (since end 2022 – the game changer?)

- Automated discharge letters
- Patient summaries
- Consult to text – into EPD
- No show-up scheduled poli-clinic visits
- Chatbot 



03 May 2023: **Godfather AI leaves Go**

AI-pionier Geoffrey Hinton leaves
experts that worry about the introd
accuracy of AI



www.nature.com

AI in healthcare

5-4

ed or not

dance!

UMC Utrecht

Why so concerned w.r.t. AI in health(care)?

- AI is already everywhere around us anyway
  - Health domain is not Netflix, Booking, Google maps, Weather forecasting

- Health domain keeps human (provider, patient, citizen) in the loop -> but for how long?
  - On other hand: realisation that AI might also be better suited for various medical tasks (predict and detect better; not tired; no off-day, etc.)

- Other parties entered the health field Google NVIDIA amazon Microsoft OpenAI

- AI is not a drug
  - Drug guidance – set in stone + transparent + instructions for use
  - Everything is known at moment of market access
  - Drugs from provider→ drugstore→ patient/citizen ; AI from company directly to citizen
  - AI similar but different life-cycle than drugs

7 Phases https://guideline-ai-healthcare.com -

Evaluation/auditing criteria needed per phase

From data, to development, to evaluation, software, impact, implementation, Monitoring/updating

Phase 0

Project idea and preparation

Phase 1

Collection and management of the Data

Phase 2

Development of the AI

Phase 3

Validation of the AI

Phase 4

Development of the required software

Phase 5

Impact assessment of the AI in combination with the software

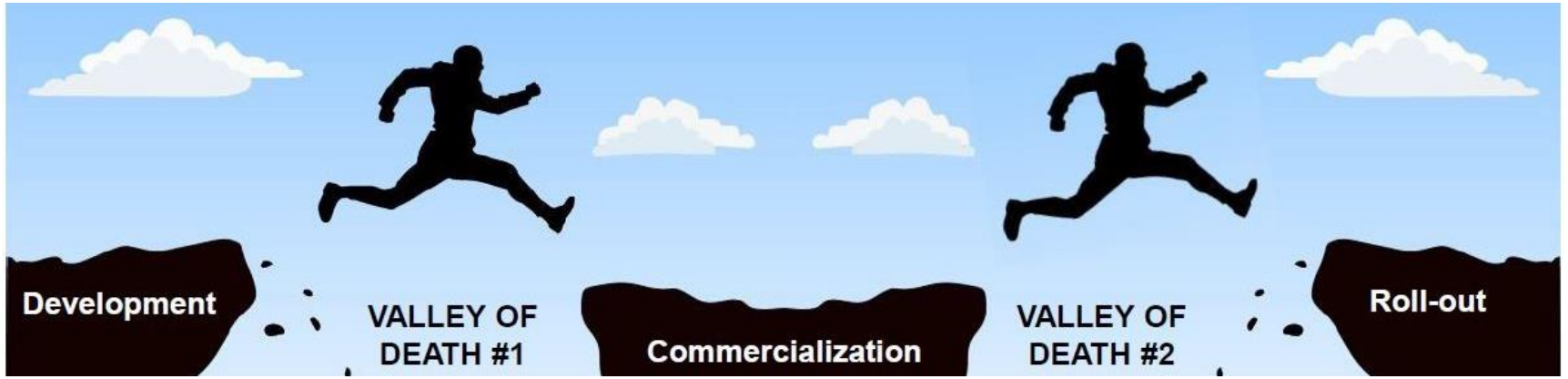
Phase 6

Implementation of AI+software in daily practice, including monitoring (PMS)

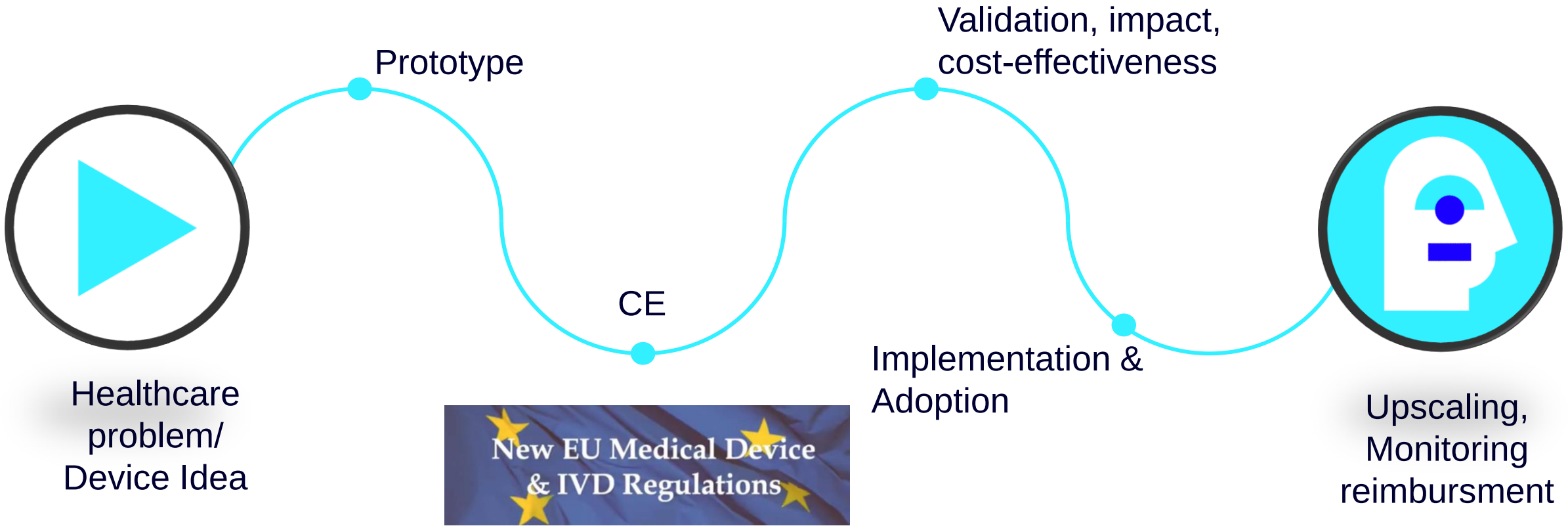
2 valleys of death

HEALTH INNOVATION NETHERLANDS

Healthcare problem → prototype → Technical feasibility → Proof of concept CE → Validation, Impact & Cost-effectiveness → Implementation + Adoption → Upscaling Monitoring



MedTech/AI – bumpy road



We need guidance for 'AI in Health(care)' (before we upscale it to everyone)

Guidance for assessment, monitoring, auditing on:

- Explainability and Transparency
- Trustworthiness (validity, Fairness, risks of bias, safety)

Transparent & Explainable AI in Health (both AI types)

TRIPOD+AI www.tripod-statement.org

www.tripod-statement.org Cite this as: *BMJ* 2024;385:e078378

TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods

Gary S Collins,[1] Karel G M Moons,[2] Paula Dhiman,[1] Richard D Riley,[3,4] Andrew L Beam,[5] Ben Van Calster,[6,7] Marzyeh Ghassemi,[8] Xiaoxuan Liu,[9,10] Johannes B Reitsma,[2] Maarten van Smeden,[2] Anne-Laure Boulesteix,[11] Jennifer Catherine Camaradou,[12,13] Leo Anthony Celi,[14,15,16] Spiros Denaxas,[17,18] Alastair K Denniston,[4,9] Ben Glocker,[19] Robert M Golub,[20] Hugh Harvey,[21] Georg Heinze,[22] Michael M Hoffman,[23,24,25,26] André Pascal Kengne,[27] Emily Lam,[12] Naomi Lee,[28] Elizabeth W Loder,[29,30] Lena Maier-Hein,[31] Bilal A Mateen,[17,32,33] Melissa D McCradden,[34,35] Lauren Oakden-Rayner,[36] Johan Ordish,[37] Richard Parnell,[12] Sherri Rose,[36] Karandeep Singh,[38] Laure Wynants,[40] Patricia Logullo[1]

Nature Medicine 2024
www.tripod-statement.org

The TRIPOD-LLM Statement: A Targeted Guideline For Reporting Large Language Models Use

Jack Gallifant, Majid Afshar, Saleem Ameen, Yindalon Aphinyanaphongs, Shan Chen, Giovanni Cacciamani, Dina Demner-Fushman, Dmitriy Dligach, Roxana Daneshjou, Chrystinne Fernandes, Lasse Hyldig Hansen, Adam Landman, Lisa Lehmann, Liam G. McCoy, Timothy Miller, Amy Moreno, Nikolaj Munch, David Restrepo, Guergana Savova, Renato Umeton, Judy Wawira Gichoya, Gary S. Collins, Karel G. M. Moons, Leo A. Celi, Danielle S. Bitterman

doi: https://doi.org/10.1101/2024.07.24.24310930

Trustworthy, FAIR & unbiased AI in Health

PROBAST+AI; www.probast.org

BMJ Open

Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence

Gary S Collins [1,2] Paula Dhiman [1,2] Constanza L Andaur Navarro [3] Jie Ma [1] Lotty Hooft,[3,4] Johannes B Reitsma,[3] Patricia Logullo [1,2] Andrew L Beam [5,6] Lily Peng,[7] Ben Van Calster [8,9,10] Maarten van Smeden [3] Richard D Riley [11] Karel GM Moons[3,4]

Lancet Digit Health. 2024 Jul;6(7):e441-e443. doi: 10.1016/S2589-7500(24)00111-0.

From text to treatment: the crucial role of validation for generative large language models in health care

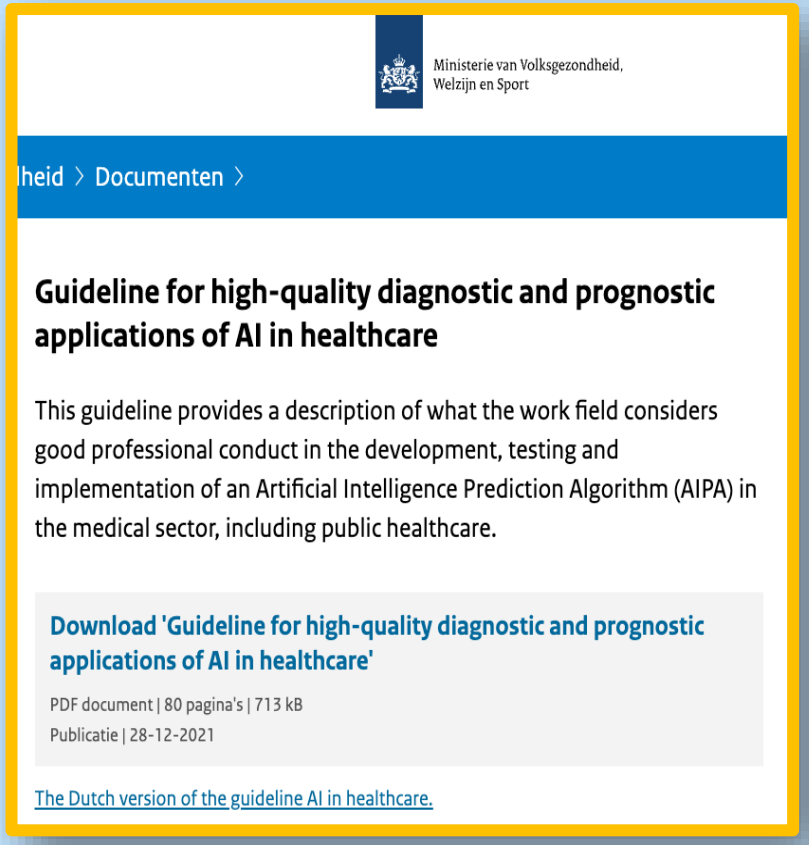
Anne de Hond [1], Tuur Leeuwenberg [2], Richard Bartels [3], Marieke van Buchem [4], Ilse Kant [5], Karel Gm Moons [2], Maarten van Smeden [2]

VWS Programme
Valuable AI for health

Increasing the perceived value of AI for health

Guideline for AI in healthcare https://guideline-ai-healthcare.com

Because AI is not a drug (MDR/IVDR)


Ministerie van Volksgezondheid, Welzijn en Sport

heid › Documenten ›

Guideline for high-quality diagnostic and prognostic applications of AI in healthcare

This guideline provides a description of what the work field considers good professional conduct in the development, testing and implementation of an Artificial Intelligence Prediction Algorithm (AIPA) in the medical sector, including public healthcare.

Download 'Guideline for high-quality diagnostic and prognostic applications of AI in healthcare'

PDF document | 80 pagina's | 713 kB
Publicatie | 28-12-2021

The Dutch version of the guideline AI in healthcare.


npj | Digital Medicine

www.nature.com/npjdigitalmed

REVIEW ARTICLE OPEN

Check for updates

Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review

Anne A. H. de Hond[1,2,3,8✉], Artuur M. Leeuwenberg[4,8✉], Lotty Hooft[4,5], Ilse M. J. Kant[1,2,3], Steven W. J. Nijman[4], Hendrikus J. A. van Os[2,6], Jiska J. Aardoom[6,7], Thomas P. A. Debray[4], Ewoud Schuit[4], Maarten van Smeden[4], Johannes B. Reitsma[4], Ewout W. Steyerberg[2,3], Niels H. Chavannes[6,7] and Karel G. M. Moons[4]

https://guideline-ai-healthcare.com
https://leidraad-AI.nl

AI guidance/auditing https://guideline-ai-healthcare.com

From data, to development, to evaluation, software, impact & implementation, Monitoring

Phase 0

Project idea
and preparation

Phase 1

Collection
and management
of the Data

Phase 2

Development
of the AI

Phase 3

Validation
of the AI

Phase 4

Development
of the required
software

Phase 5

Impact assessment
of the AI
in combination with the
software

Phase 6

Implementation
of AI+software
in daily practice,
including monitoring
(PMS)

Guidance per phase

- Per phase explicit criteria and hands-on guidance – based on current state-of-science

- What the field considers **good conduct** in development, testing, implementation of AI, before widescale use in our patients/clients/citizens

  - **Requirements** vs. **recommendations per phase**

  - Guideline beyond the MDR/IVDR

Guidance and E-learning tailored to targeted **groups** https://guideline-ai-healthcare.com

Applying AI

Healthcare provider
Professional Scientific Medical associations
Education/training
IT suppliers
Patient / Citizen

Developing AI

Validator
Responsible developer
Researcher
Data manager
Data supplier

Assessing AI

(Internal) supervisor
Notified body Peer reviewer Privacy officer
Competent Authority
Auditor
Insurer (DiGiZo)

Society

Patient(s)-(associations)
Interest parties
Politcy parties
Interested citizen
HINL

Similar Guideline for Generative AI underway

UMC Utrecht

Some misconceptions

1. Development not necessarily on large, inclusive datasets
   - Such data sets do not exist, not needed
   - Compare:  
   - Richer data not always better models


The New York Times

OPINION

A.I. Could Worsen Health Disparities

In a health system riddled with inequity, we risk making dangerous biases automated and invisible.

Jan. 31, 2019

2. Validation most important - Test, Test, Test before use
   - use AI Guideline to determine what is (un)known at moment of assessing/auditing

3. Often heard: AI goes fast, can we keep up? YES: our valuations/guidance do ‘not’ change.

4. Not AI ready? Yes we are! Algorithms not new in health(care). Gen-AI is.

5. AI replaces us humans? No. Compare Netflix, Booking, Maps.

HEALTH
INNOVATION
NETHERLANDS

HEALTH
INNOVATION
NETHERLANDS

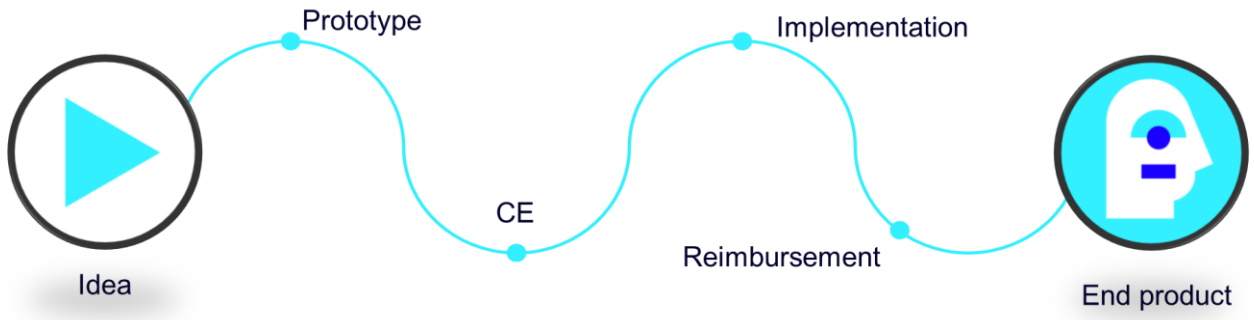
The gateway to European market for medtech, biotech

To bridge the 2 valleys of death

www.healthinnovation.nl

Providing a clear and shorter path for innovators

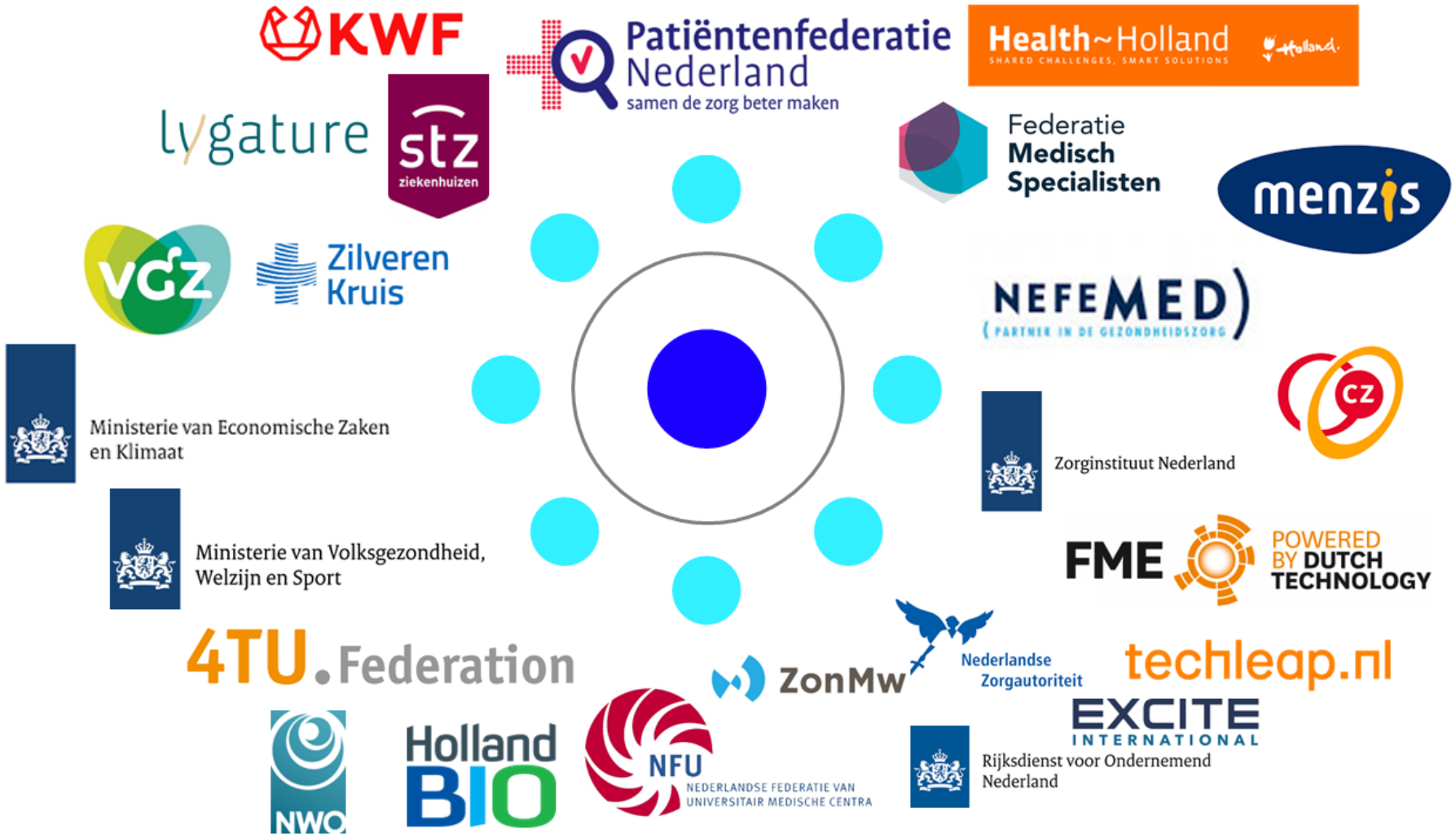
The way to market is a bumpy ride – with many stakeholders involved sequentially.



HI-NL brings all relevant stakeholders together early on, providing **a clear path** and **tailored guidance endorsed by all stakeholders**.

www.healthinnovation.nl

HEALTH INNOVATION NETHERLANDS



Take aways

1. AI in health domain different from films, hotels, road directions

2. Richer data not necessarily required for more fair or better AI
   → it is all about testing, testing and testing before use
   → transparency on what data developed & tested → AI leaflet needed

3. AI is not a drug → unclear how developed, tested, on which data, no instructions for use

4. Trustworthiness guidance for AI in healthcare, needs to be applied NOW!
   - Human remains in the loop – but for how long?
   - And ….AI soon directly sold from companies to citizens/patients

Take aways

5. Monitoring/auditing notably needed for AI that directly goes from company → citizens/patients
   - no healthcare provider in-between anymore

6. Drugs do not change after market access/implementation → AI does (self-adjusting)
   - Extra reason for good monitoring and auditing
   - Guidance for AI-PMS is thus different from Drug-PMS

7. Much monitoring and auditing guidance exists – see above.
   - Use it wisely and use it NOW!
   - We seem to act too slow ? Not wait for ‘tax-fraud’ / ‘wrong criminal’ case.

www.healthinnovation.nl

Thank you!

https://guideline-ai-healthcare.com

AI quality tool

TRIPOD

TRIPOD+AI: www.tripod-statement.org

PROBAST

PROBAST+AI: www.probast.org

Carl Moons: k.g.m.moons@umcutrecht.nl