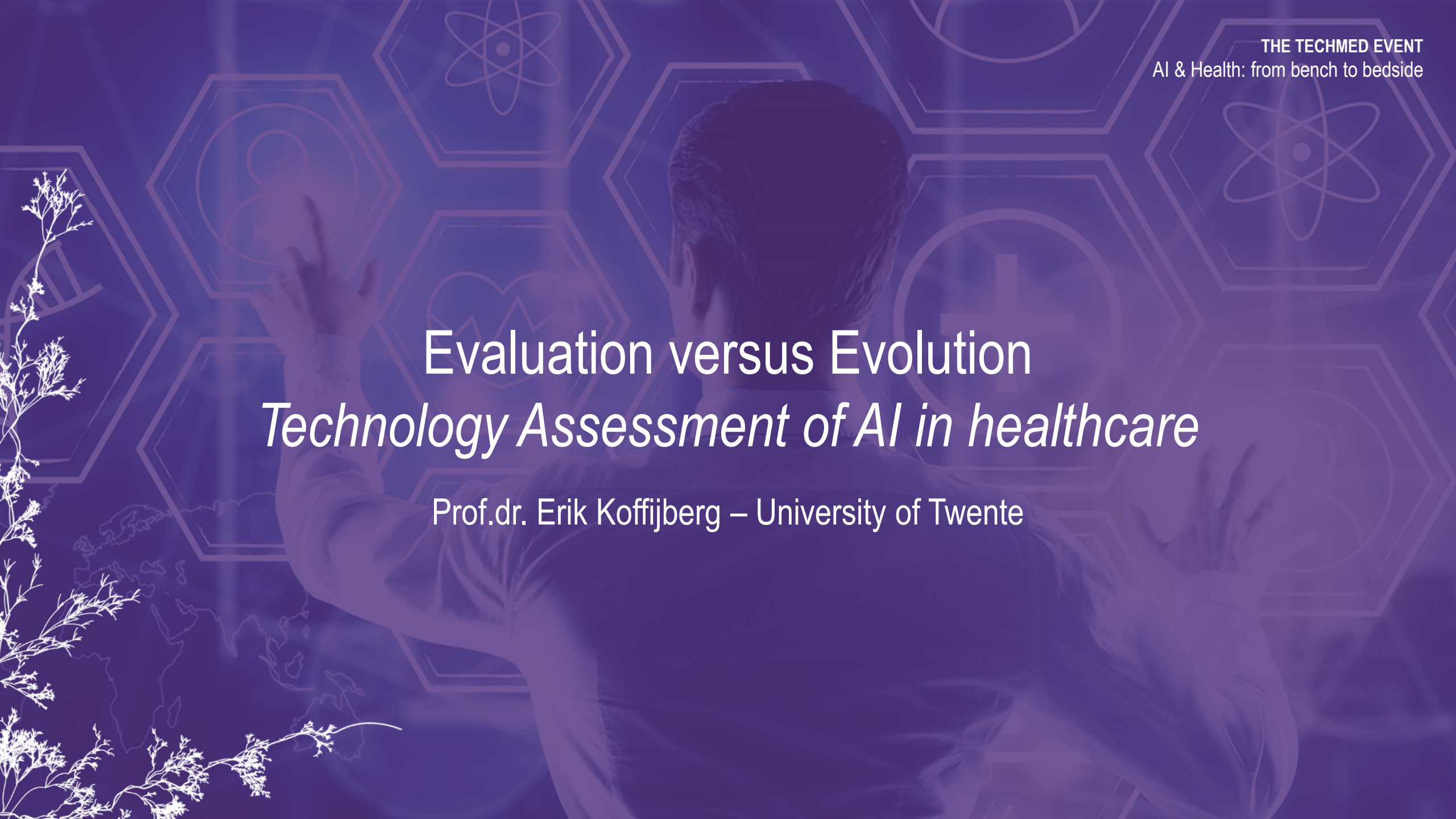
UNIVERSITY OF TWENTE. | TECHMED CENTRE
THE TECHMED EVENT
SHAPING A HEALTHY FUTURE



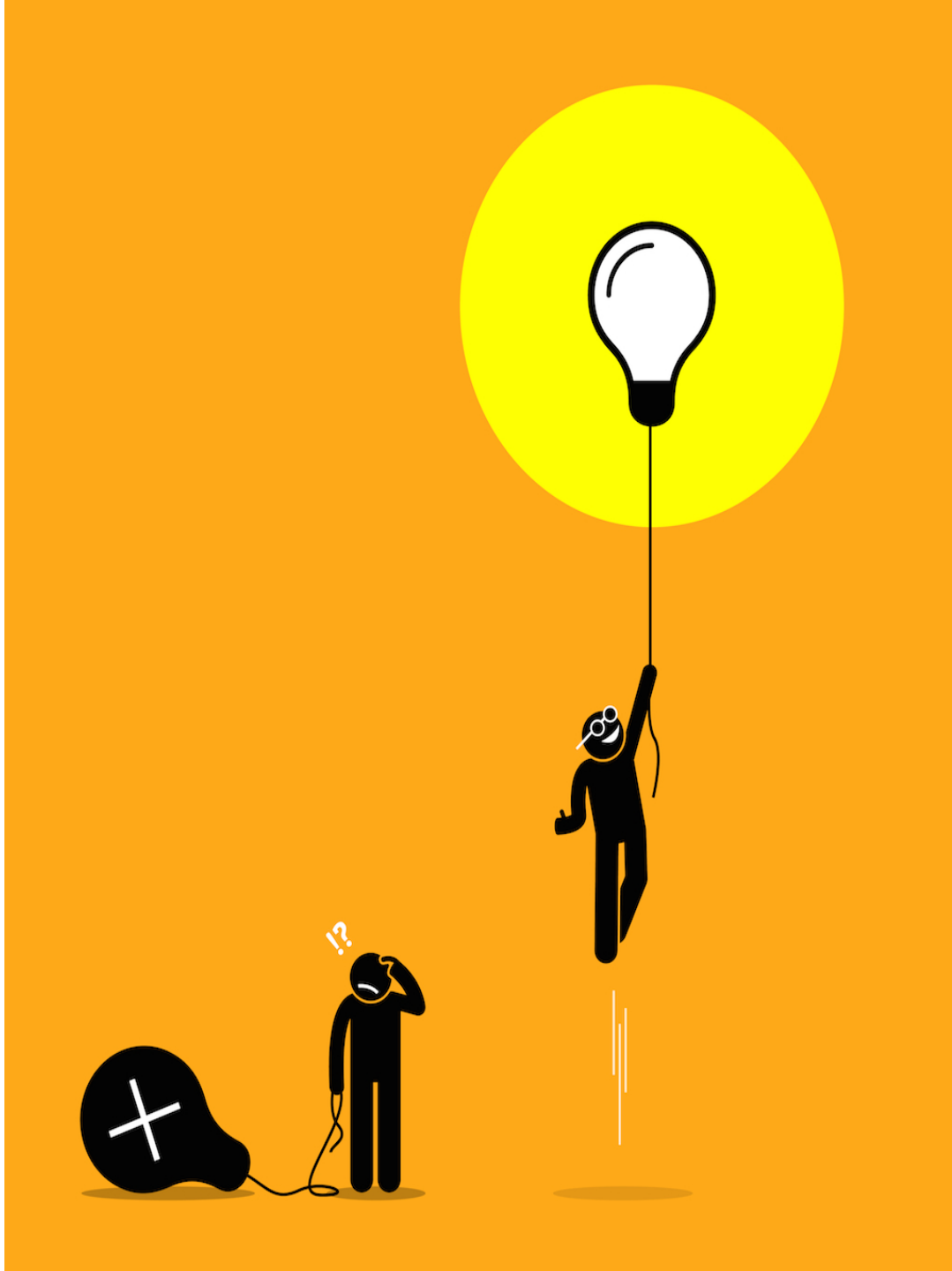
Evaluation versus Evolution
Technology Assessment of AI in healthcare

Prof.dr. Erik Koffijberg – University of Twente

DISCLOSURE

My perspective

- AI tools are needed to help addressing the major challenges we face in healthcare
  - Some AI tools will actually do so
  - Other AI tools will not
- We need to distinguish between these tools, by
  - Evaluating the impact of AI tools
  - Which requires developing new methods for AI evaluation



Differences or similarities?

- AI may be considered unique from a technological perspective and in its capabilities

- But most AI tools can be considered as diagnostic and prognostic tests and have similar goals as other health innovations

impact perspective

Similar: potential positive and *negative* impact

Epic

With the Patient at the Heart

We develop software that helps people get well, helps people stay well, and helps future generations be healthier.

Original Investigation

June 21, 2021

JAMA Internal Medicine

External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients

Andrew Wong, MD[1]; Erkin Otles, MEng[2,3]; John P. Donnelly, PhD[4]; et al

Not all AI tools have good performance

EPIC Sepsis model (ESM) was used by >100 hospitals for early sepsis detection
Epic claimed that the predictions made by ESM were 76-83% accurate

Independent testing (~38,500 hospitalizations, ~ 2,550 sepsis cases)

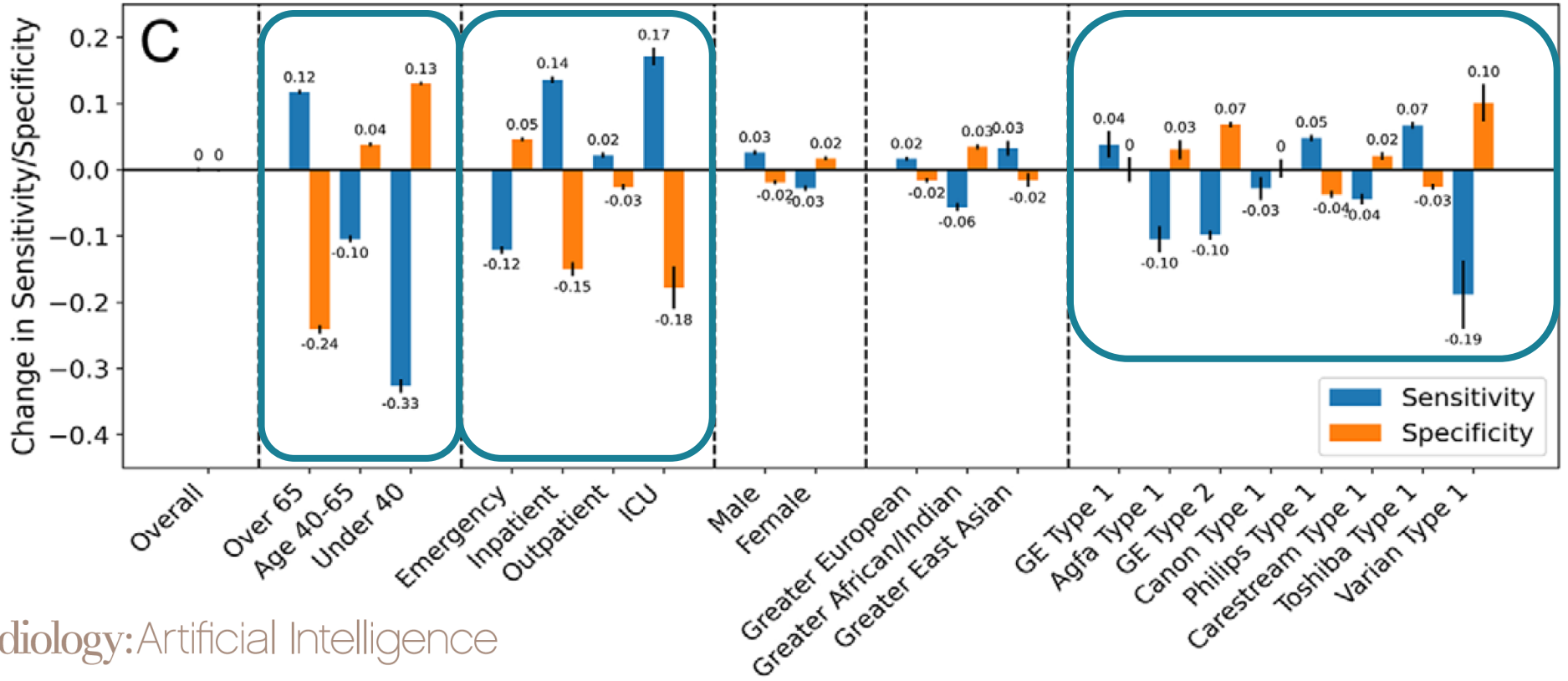
- ESM missed 67% of the people with sepsis.
- 88% of the ESM sepsis alerts were false alarms, creating “a large burden of alert fatigue.”

Wong A, Otles E, Donnelly JP, et al. External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. JAMA Intern Med. 2021;181(8):1065–1070.

Similar: impact depends on application & context

The Subgroup Imperative: Chest Radiograph Classifier Generalization Gaps in Patient, Setting, and Pathology Subgroups

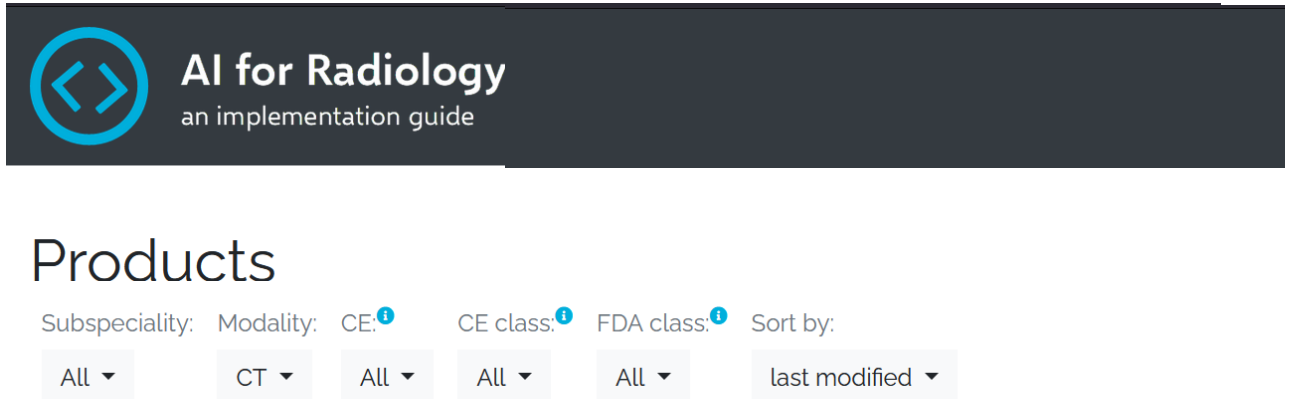
Best algorithm has 72% sensitivity and 82% specificity (~200,000-image-dataset)



Impact depends on (optimal) use

- Ahluwalia M, et al. The Subgroup Imperative: Chest Radiograph Classifier Generalization Gaps in Patient, Setting, and Pathology Subgroups. Radiol Artif Intell. 2023 Jul 12;5(5):e220270.
- Huisman M, Hannink G. The AI Generalization Gap: One Size Does Not Fit All. Radiol Artif Intell. 2023 Aug 30;5(5):e230246.

Different: impact on key outcomes typically unknown



- 224 AI tools for radiology
- 61 MDR certified, 158 MDD certified
- 88 AI tools applicable to CT scans

- Just **20 studies** published reporting on health economic impact of AI (2015-2021)
- Insight into key benefits from AI is lagging behind technological developments

https://grand-challenge.org/aiforradiology/

Voets MM, et al. Systematic Review of Health Economic Evaluations Focused on Artificial Intelligence in Healthcare: The Tortoise and the Cheetah. Value Health. 2022 Mar;25(3):340-349.

Why is AI impact on key outcomes unknown so often?

Health Technology Assessment methods have been developed for drugs (and can be applied to tests) but AI poses specific new challenges to (model-based) impact evaluation

method perspective

Example: StrokeViewer – Large Vessel Occlusion on CT in stroke

Interhospital image sharing & streamline workflow

Impact depends crucially on context, and can be different *for different hospitals*, based on Actual use of the tool

- Available radiological expertise (e.g. daytime vs nighttime)
- Use as initial check or as double check
- For improved detection rate or for image sharing (or both)

- How downstream benefits to other departments are taken into account
- How consequences of time savings (faster diagnostic processes) are incorporated

Haan et al. Project report. TURBO project Radboudumc/UT: Organize your health innovations!
van Leeuwen KG, et al. Cost-effectiveness of artificial intelligence aided vessel occlusion detection in acute stroke: an early health technology assessment. Insights Imaging. 2021 Sep 25;12(1):133.

Example: StrokeViewer – Large Vessel Occlusion on CT in stroke

Interhospital image sharing & streamline workflow

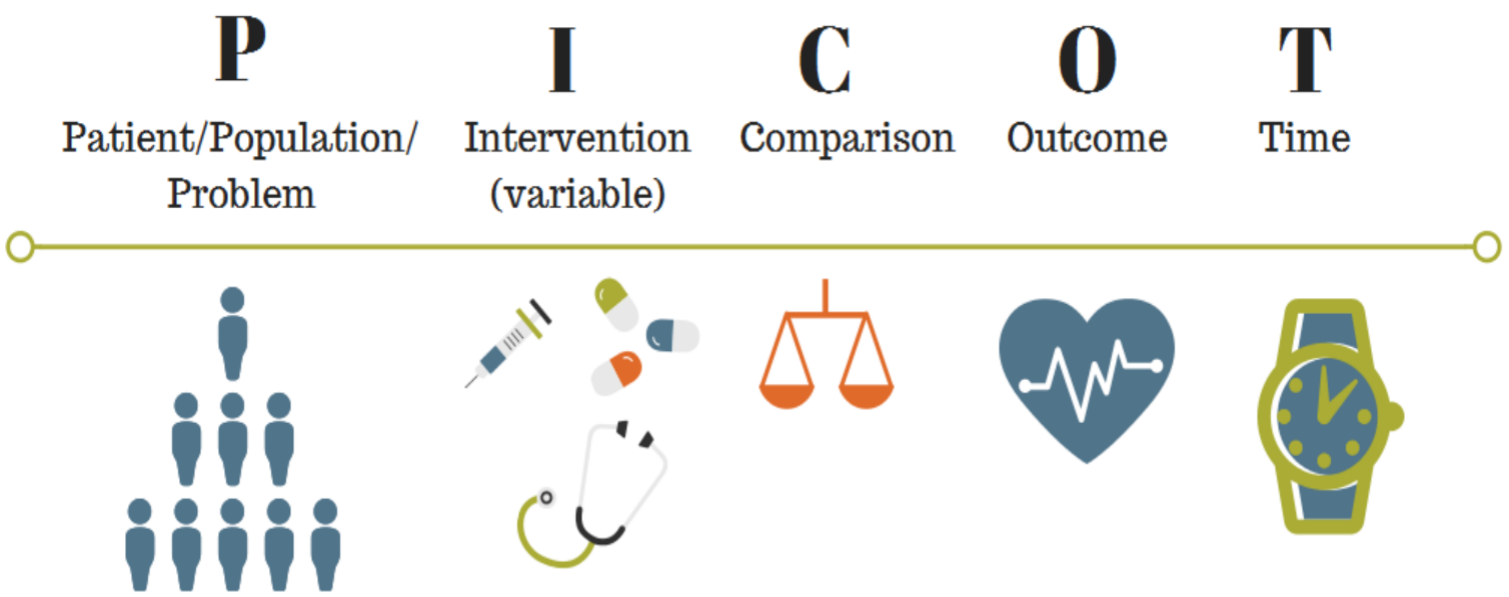
Impact is not fixed due to innovation dynamics

- Apple App Store: 20 updates from Oct 2021 - Oct 2023
- Changes in algorithm performance
- Changes in user interface (input, output, visualisation)
- Introduction of new features (secure chat for faster team alignment)

Impact may change due to learning & hospital dynamics

- Improve due to additional (context-specific) learning
- Decline due to model drift (case-mix, coding, imaging technologies)

Frameworks for health economic evaluations



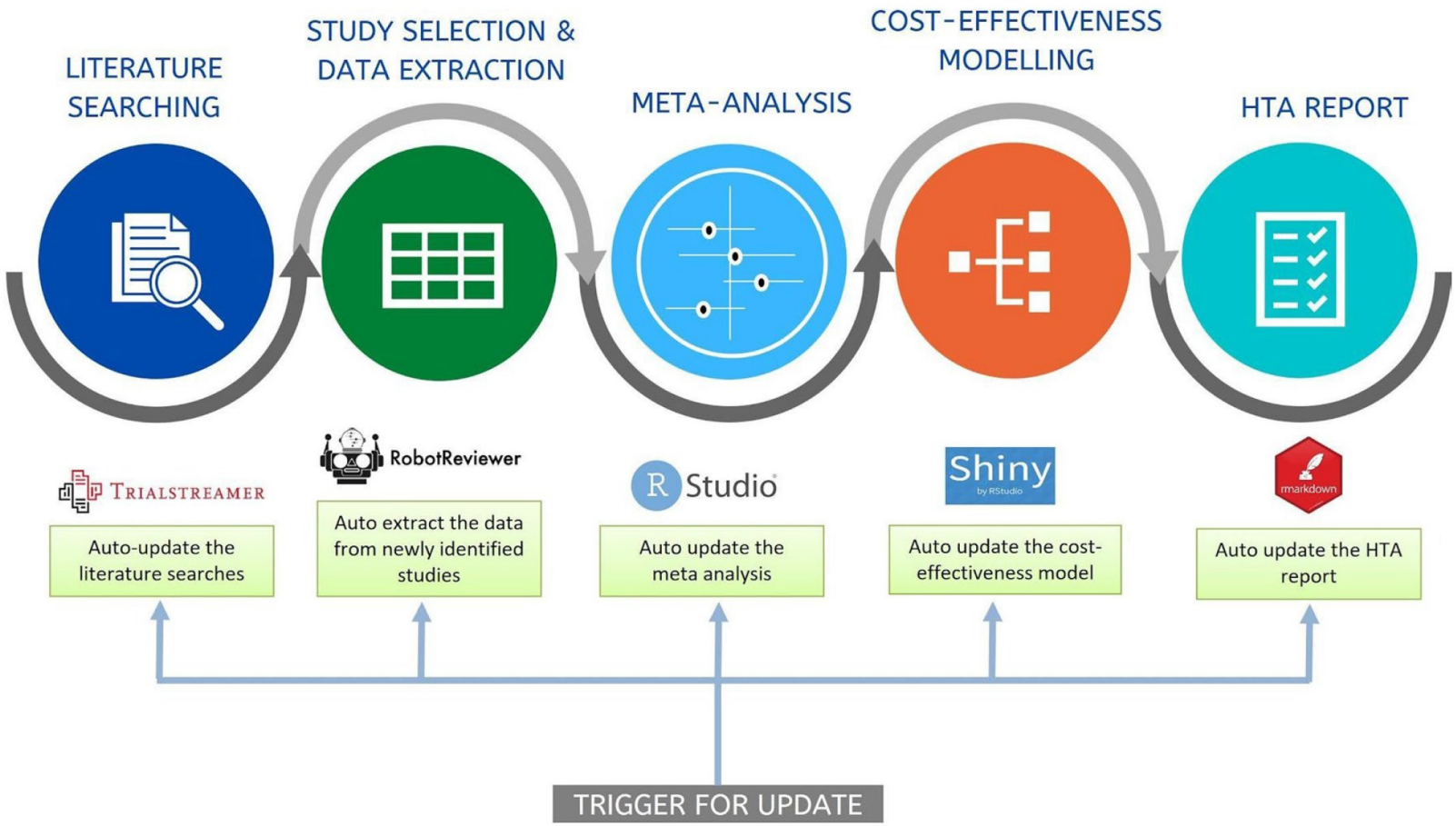
Evaluation
(static, one-off)

versus

Evolution
(dynamic)

Technology Assessment of AI in healthcare - Way forward

Living HTA, with manual or semi-automated updates

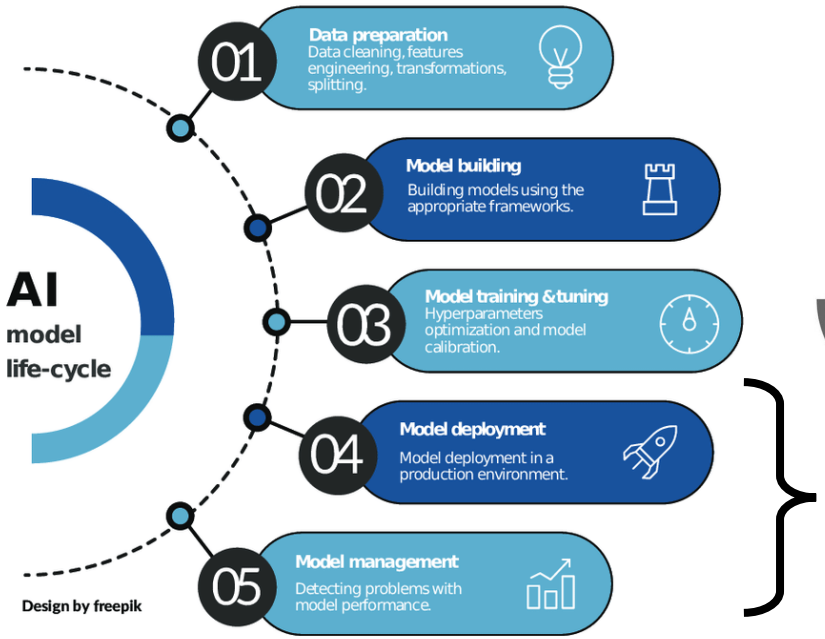


- Matches the AI development and updating process
- *Requires more time*
- *Requires more budget*
- *Requires ongoing stakeholder input*

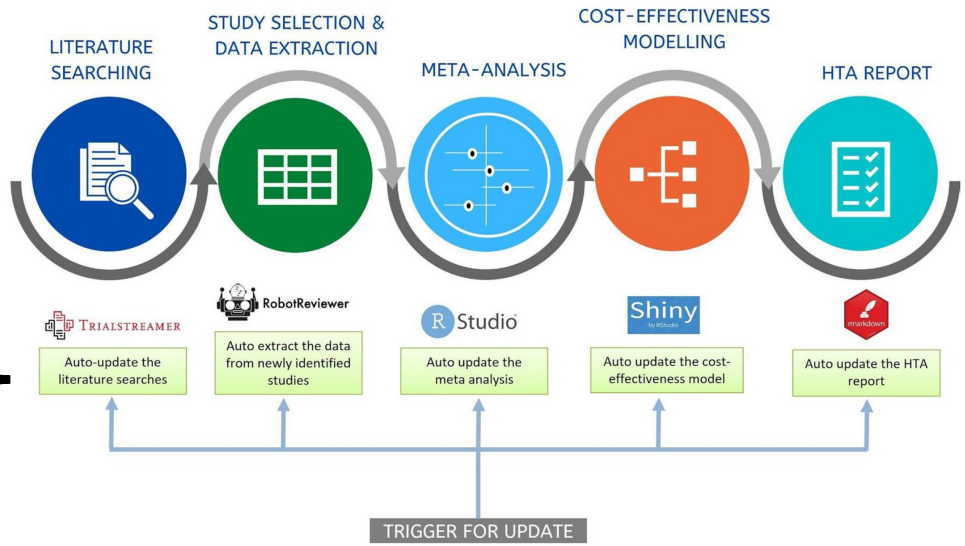
Thokala P, et al. Living Health Technology Assessment: Issues, Challenges and Opportunities. Pharmacoeconomics. 2023 Mar;41(3):227-237.

Increasing the impact of AI tools in healthcare - Way forward

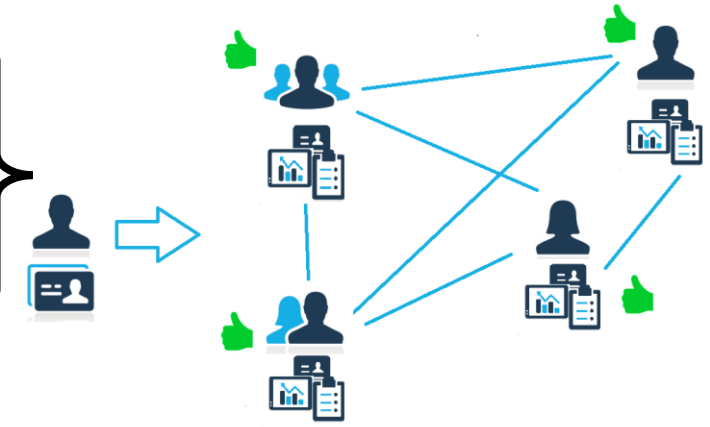
AI development & updating

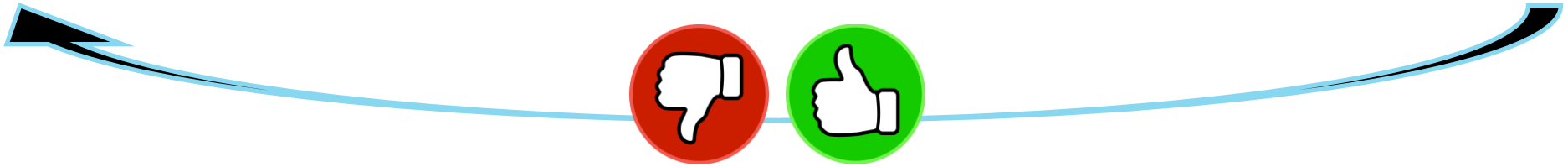


Living HTA
(tailored)



Sharing study data & results





Thank you

Prof.dr.ir. Erik Koffijberg

@ErikKoffijberg

hkoffijberg

h.koffijberg@utwente.nl

https://www.utwente.nl/en/bms/htsr/

UNIVERSITY OF TWENTE. | TECHMED CENTRE

THE TECHMED EVENT | SHAPING A HEALTHY FUTURE

[TIME SESSION] | [SESSION TITLE] | [SPEAKER]