Promoting Decision-Making Autonomy Through Reflection

Simon WS Fischer

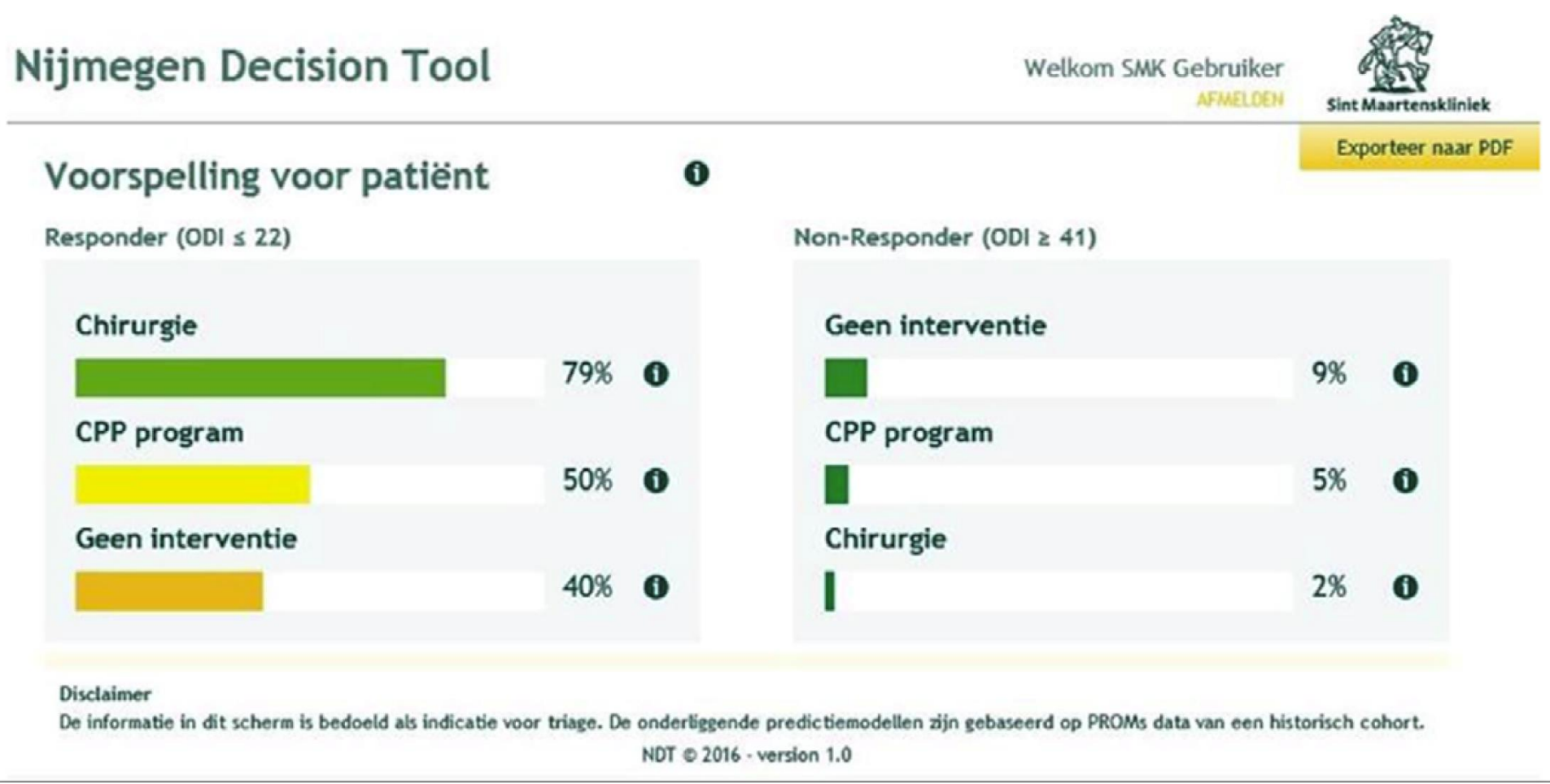
FAccTRec

14-10-2024

DONDERS INSTITUTE

Radboud University Radboudumc

Photo by cottonbro studio from Pexels



Haselager et al. (2023). Reflection Machines: Supporting Effective Human Oversight Over Medical Decision Support Systems.

Radiology

ORIGINAL RESEARCH • COMPUTER APP

Automation Bias in Mammography: The Impact of Artificial Intelligence BI-RADS Suggestions on Reader Performance

Thomas Dratsch, MD • Xue Chen, MD* • Mohammad Rezazade Mehrizi, PhD • Roman Kloeckner, MD • Aline Mähringer-Kunz, MD • Michael Püsken, MD • Bettina Baeßler, MD • Stephanie Sauer, MD • David Maintz, MD • Daniel Pinto dos Santos, MD*

nature medicine

LETTERS

https://doi.org/10.1038/s41591-020-0942-0

Check for updates

Human–computer collaboration for skin cancer recognition

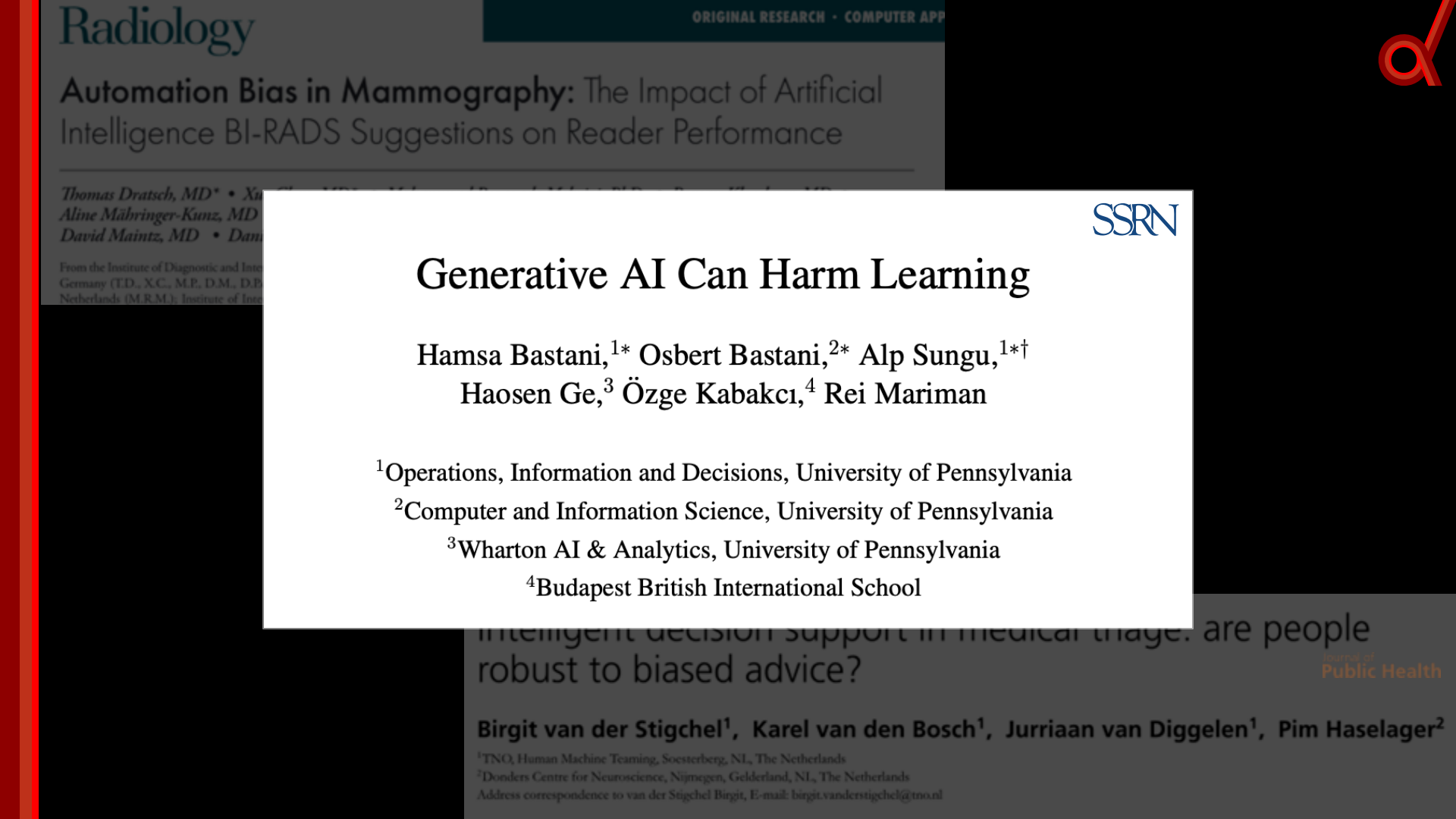
Philipp Tschandl[1,17], Christoph Rinner[2,17], Zoe Apalla[3], Giuseppe Argenziano[4], Noel Codella[5], Allan Halpern[6], Monika Janda[7], Aimilios Lallas[3], Caterina Longo[8,9], Josep Malvehy[10,11], John Paoli[12,13], Susana Puig[10,11], Cliff Rosendahl[14], H. Peter Soyer[15], Iris Zalaudek[16] and Harald Kittler[1]

Intelligent decision support in medical triage: are people robust to biased advice?

Journal of Public Health

Birgit van der Stigchel[1], Karel van den Bosch[1], Jurriaan van Diggelen[1], Pim Haselager[2]

[1]TNO, Human Machine Teaming, Soesterberg, NL, The Netherlands
[2]Donders Centre for Neuroscience, Nijmegen, Gelderland, NL, The Netherlands
Address correspondence to van der Stigchel Birgit, E-mail: birgit.vanderstigchel@tno.nl


Radiology
ORIGINAL RESEARCH · COMPUTER APP
Automation Bias in Mammography: The Impact of Artificial Intelligence BI-RADS Suggestions on Reader Performance
Thomas Dratsch, MD*
Aline Mähringer-Kunz, MD
David Maintz, MD
SSRN
Generative AI Can Harm Learning
Hamsa Bastani,[1*] Osbert Bastani,[2*] Alp Sungu,[1*†]
Haosen Ge,[3] Özge Kabakcı,[4] Rei Mariman
[1]Operations, Information and Decisions, University of Pennsylvania
[2]Computer and Information Science, University of Pennsylvania
[3]Wharton AI & Analytics, University of Pennsylvania
[4]Budapest British International School
are people
robust to biased advice?
Journal of Public Health
Birgit van der Stigchel[1], Karel van den Bosch[1], Jurriaan van Diggelen[1], Pim Haselager[2]

Research Question

How to calibrate reliance on machine recommendations?

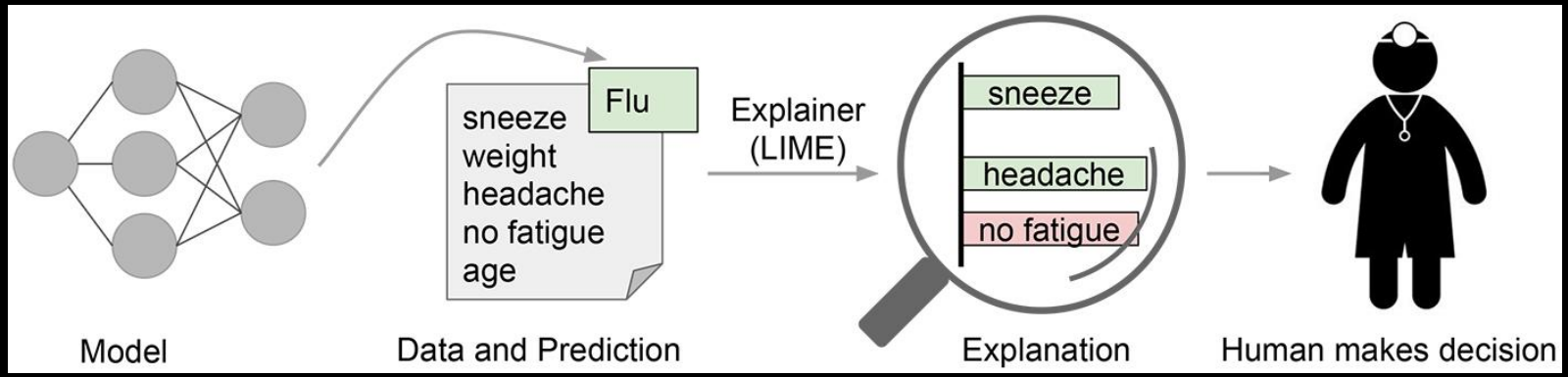
EU AI ACT Article 14 Human Oversight

“remain aware of the possible tendency of automatically relying or over-relying on the output”

EU HLEG Trustworthy AI

First requirement: human agency and oversight
“Users should be able to make informed autonomous decisions regarding AI systems.”

Explainable AI



Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier

Moral Decision Making in Human-Agent Teams: Human Control and the Role of Explanations

frontiers in Robotics and AI

Jasper van der Waa[1,2]*, *Sabine Verdult*[3], *Karel van den Bosch*[3], *Jurriaan van Diggelen*[1], *Tjalling Haije*[1], *Birgit van der Stigchel*[1,4] *and Ioana Cocu*[1]

[1]*Perceptual and Cognitive Systems, TNO, Soesterberg, Netherlands,* [2]*Interactive Intelligence, Technical University Delft, Delft, Netherlands,* [3]*Training and Performance Innovations, TNO, Soesterberg, Netherlands,* [4]*Artificial Intelligence, Radboud University, Nijmegen, Nijmegen, Netherlands*

ARTICLE Open Access

Translational Psychiatry

How machine-learning recommendations influence clinician treatment selections: the example of antidepressant selection

Maia Jacobs[1], Melanie F. Pradier[1], Thomas H. McCoy Jr.[2,3], Roy H. Perlis[2,3], Finale Doshi-Velez[1] and Krzysztof Z. Gajos[1]

Does the Whole Exceed its Parts? The Effect of AI Explanations on Complementary Team Performance

CHI 2021

Gagan Bansal*
bansalg@cs.washington.edu
University of Washington

Tongshuang Wu*
wtshuang@cs.washington.edu
University of Washington

Joyce Zhou†
jyzhou15@cs.washington.edu
University of Washington

Raymond Fok†
rayfok@cs.washington.edu
University of Washington

Besmira Nushi
besmira.nushi@microsoft.com
Microsoft Research

Ece Kamar
eckamar@microsoft.com
Microsoft Research

Marco Tulio Ribeiro
marcotcr@microsoft.com

Daniel S. Weld
weld@cs.washington.edu

To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making

ZANA BUÇINCA, Harvard University, USA
MAJA BARBARA MALAYA, Lodz University of Technology, Poland
KRZYSZTOF Z. GAJOS, Harvard University, USA

CSCW2021
23 - 27 Oct 2021 (held virtually)

Explainable AI is Dead, Long Live Explainable AI!

Hypothesis-driven Decision Support using Evaluative AI

Tim Miller
tmiller@unimelb.edu.au
The University of Melbourne
Melbourne, VIC, Australia

CHI23

Let Me Think! Investigating the Effect of Explanations Feeding Doubts About the AI Advice

Springer

Federico Cabitza[1,2(✉)], Andrea Campagner[2], Lorenzo Famiglini[1], Chiara Natali[1], Valerio Caccavella[3], and Enrico Gallazzi[3]

Don't Just Tell Me, Ask Me: AI Systems that Intelligently Frame Explanations as Questions Improve Human Logical Discernment Accuracy over Causal AI explanations

Valdemar Danry
MIT Media Lab, Massachusetts Institute of Technology
Cambridge, Massachusetts, United States
vdanry@media.mit.edu

Pat Pataranutaporn
MIT Media Lab, Massachusetts Institute of Technology
Cambridge, Massachusetts, United States
patpat@media.mit.edu

Yaoli Mao
Columbia University
New York City, New York, United States
ym2429@tc.columbia.edu

Pattie Maes
MIT Media Lab, Massachusetts Institute of Technology
Cambridge, Massachusetts, United States
pattie@media.mit.edu

FAccT 2023

Photo by Anne O'Sullivan from Pexels

Cambridge Quarterly of Healthcare Ethics (2023), 1–10
doi:10.1017/S0963180122000718

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Reflection Machines: Supporting Effective Human Oversight Over Medical Decision Support Systems

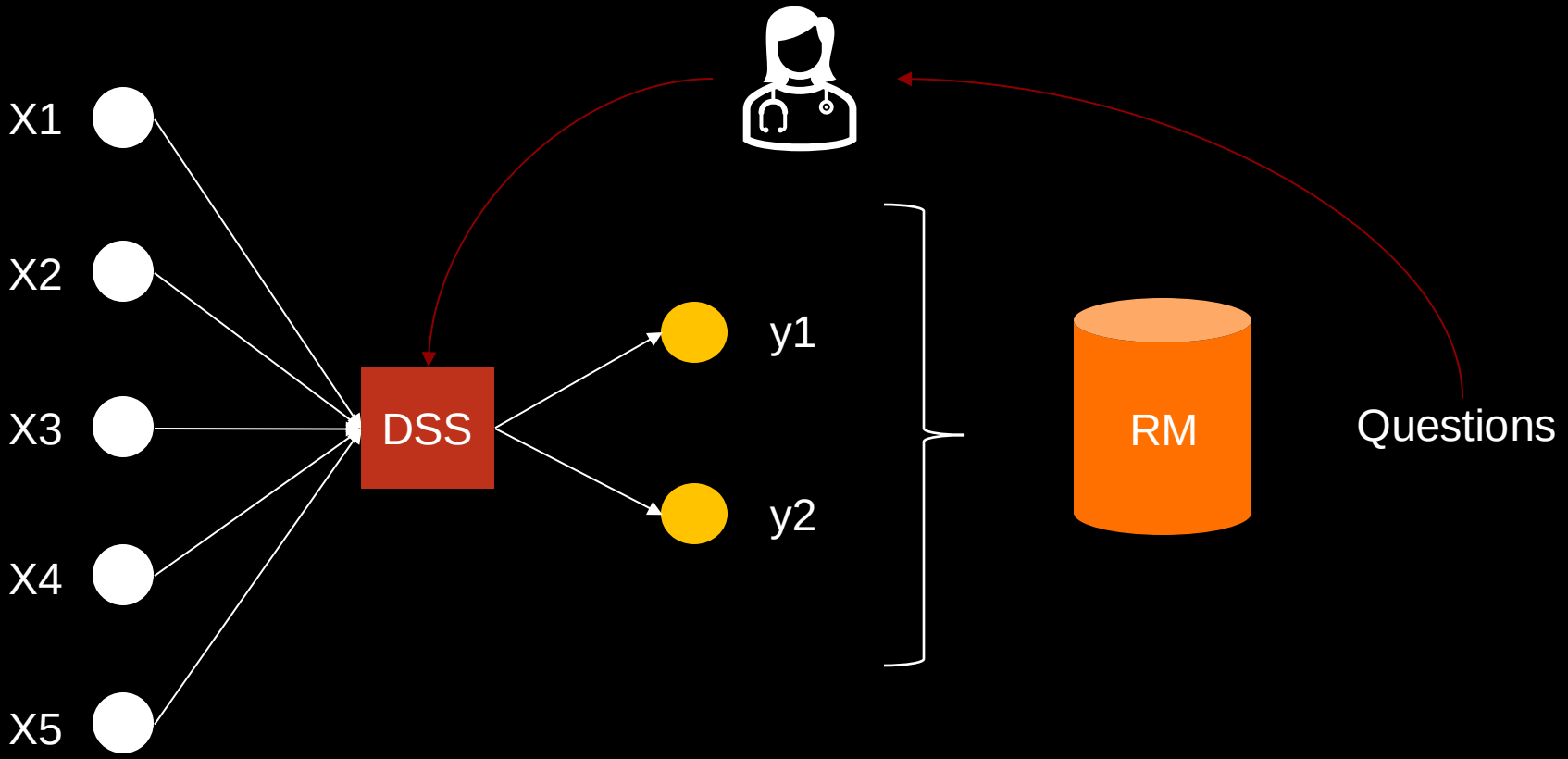
Pim Haselager[1]*, Hanna Schraffenberger[2], Serge Thill[1], Simon Fischer[1], Pablo Lanillos[1], Sebastiaan van de Groes[3] and Miranda van Hooff[3,4]

Questioning AI: Promoting Decision-Making Autonomy Through Reflection

Simon W.S. Fischer
simon.fischer@donders.ru.nl
Donders Institute for Brain, Cognition, and Behaviour
Nijmegen, Netherlands

arXiv

Decision-Support System + Reflection Machine



Types of Questions

Causal depenceny

- How does outcome *Y* follow from data point *x*
- Why not option *B* given *z*

Relevance

- Is data point *x* relevant to focus on?

Counterfactual

- Would you recommend the same option if data *x* were different?

Types of Questions – Cont.

Model up-to-dateness

- Did you consider factor *x*, which is not included in the DSS?

Expertise

- Does the recommendation align with your assumptions? If not, why not?

Hypothetical Scenarios

- Is it possible to quit smoking? If so, this would affect the outcome by 25%.

Aim

Elicit certain behaviour through engagement

- Critical reflection
- Proactive care

Affects

- Decision-autonomy
- Self-efficacy

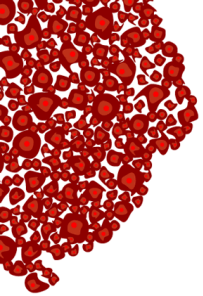
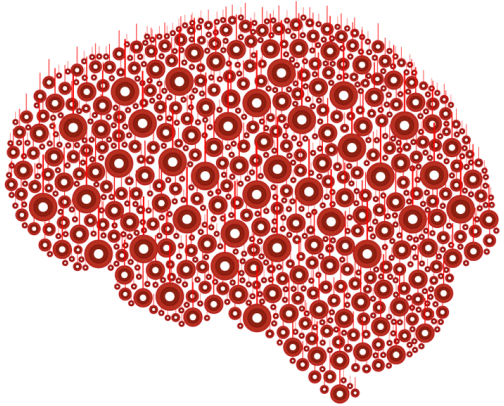
Open Considerations

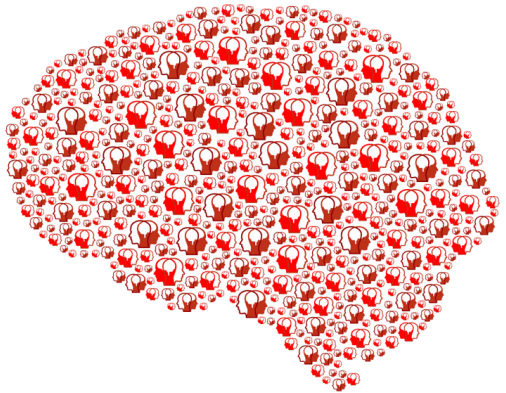
- How often to ask
- Tailor questions to the expertise of decision-maker
- Questions from the patient’s perspective
- Process answers

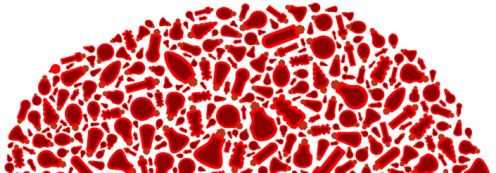
Hypothesis

Questions raise doubt – where necessary – and stimulate deliberation.

This forward-looking approach calibrates reliance on machine recommendations.

simon.fischer@donders.ru.nl

DONDERS INSTITUTE

Radboud University Radboudumc