Session 2: Equity in evidence synthesis

Impact of algorithmic bias on systematic reviews of AI-Interventions

Lotty Hooft
UMC Utrecht, The Netherlands

Cochrane

Cochrane Netherlands

Impact of algorithmic bias on systematic reviews of AI-Interventions

Prof. Dr. Lotty Hooft

Cochrane Netherlands, Julius Center, UMC Utrecht

Trusted evidence.
Informed decisions.
Better health.

I have no actual or potential conflict of interest in relation to this presentation

Cochrane

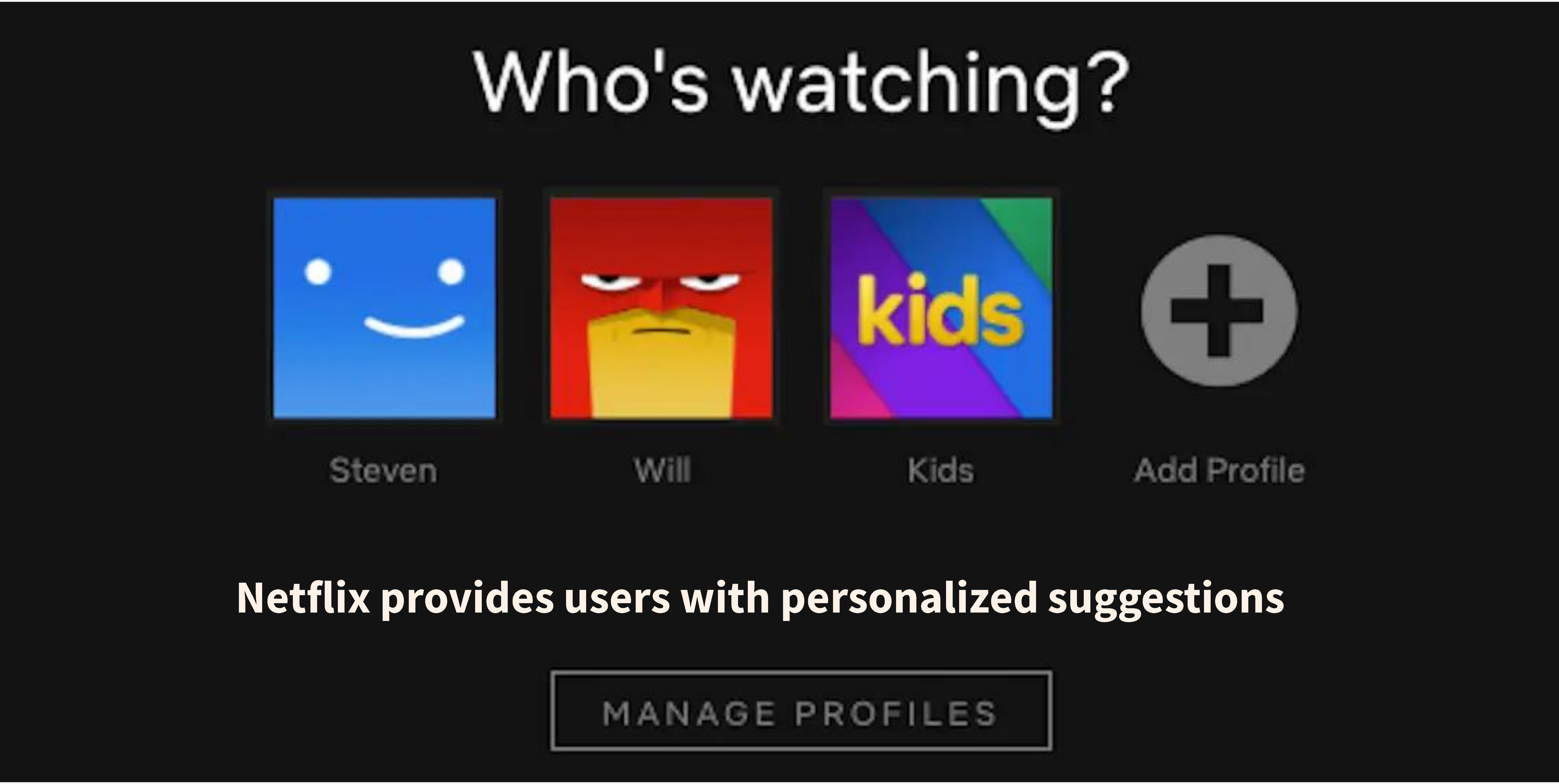
Content next 20 minutes

1. AI and AI-Interventions ?
2. Algorithmic bias ?
3. Impact on SRs ?

Cochrane

Artifical Intelligence is “hot” but what it?

Immense attention and interest from various industries



Netflix provides users with personalized suggestions

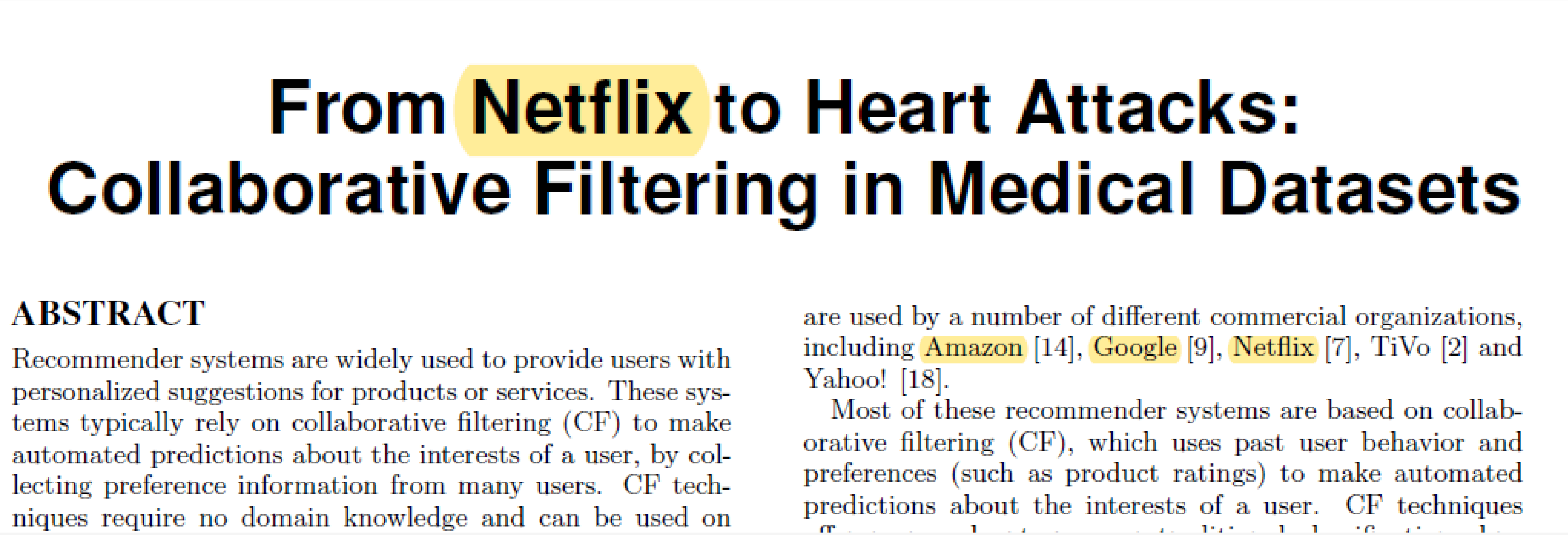
Automated predictions (for specific user) **by collecting preference info** (from many users)

Steven

Will

Cochrane Netherlands

Same methods for clinical risk stratification



From Netflix to Heart Attacks: Collaborative Filtering in Medical Datasets

ABSTRACT

Recommender systems are widely used to provide users with personalized suggestions for products or services. These systems typically rely on collaborative filtering (CF) to make automated predictions about the interests of a user, by collecting preference information from many users. CF techniques require no domain knowledge and can be used on

are used by a number of different commercial organizations, including Amazon [14], Google [9], Netflix [7], TiVo [2] and Yahoo! [18].

Most of these recommender systems are based on collaborative filtering (CF), which uses past user behavior and preferences (such as product ratings) to make automated predictions about the interests of a user. CF techniques

- Matching new cases to historical records
- Matching patient demographics to adverse outcomes

High predictive accuracy for sudden cardiac death and recurrent myocardial infraction

What is AI?

Source NIHR website

- **Digital technology** to perform tasks that were once thought to require human intelligence
- **Computer algorithms** examine **large amounts of data**, find common **patterns**, **learn** from the data and **improve** with time.
- Two main types of AI:
  - **Generative AI** (including Chat-GPT): can create new content based on learned patterns
  - **Predictive AI**: predictions about future events based on large amounts of historical data

Most AI-healthcare interventions (applications) are **predictive AI systems**

Cochrane Netherlands

AI-Interventions in Healthcare

Cochrane

AI-Interventions in Healthcare

- **Predictive AI**
  - help to **identify people at high risk of developing** certain conditions (prevention)
  - **personalise treatments** (select the patients most likely to benefit from specific treatments)

- **Diagnostic AI**
  - identify diseases or conditions **early and accurately**

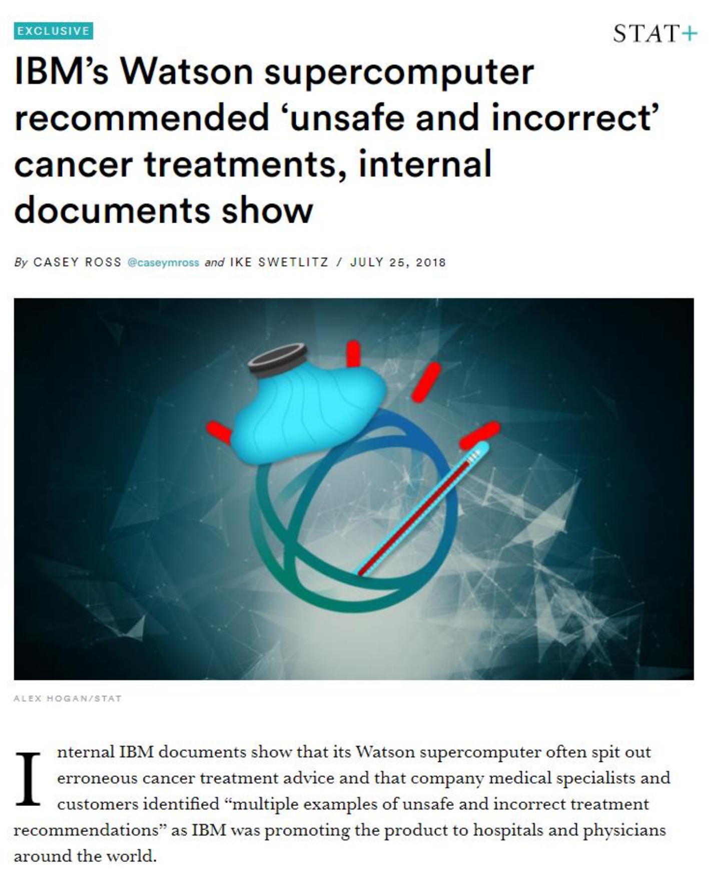
- **Therapeutic AI**
  - chatbots or virtual therapists that **provide support**, coping strategies, and guidance for patients

Examples

1. Smart stethoscopes to **detect** hearth failure
2. AI to **predict** whether lung nodules are cancer
3. AI to **personalise** cancer treatment and to predict which drugs are effective for your lung cancer

We need to be able to trust AI

- Safe
- Transparant
- Fair
  - Fairness aims to eliminate or mitigate algorithmic bias and prevent discriminatory outcomes
  - It's an active effort to correct biases and promote equality



EXCLUSIVE

STAT+

IBM's Watson supercomputer recommended 'unsafe and incorrect' cancer treatments, internal documents show

By CASEY ROSS @caseymross and IKE SWETLITZ / JULY 25, 2018

ALEX HOGAN/STAT

Internal IBM documents show that its Watson supercomputer often spit out erroneous cancer treatment advice and that company medical specialists and customers identified "multiple examples of unsafe and incorrect treatment recommendations" as IBM was promoting the product to hospitals and physicians around the world.

Cochrane Netherlands

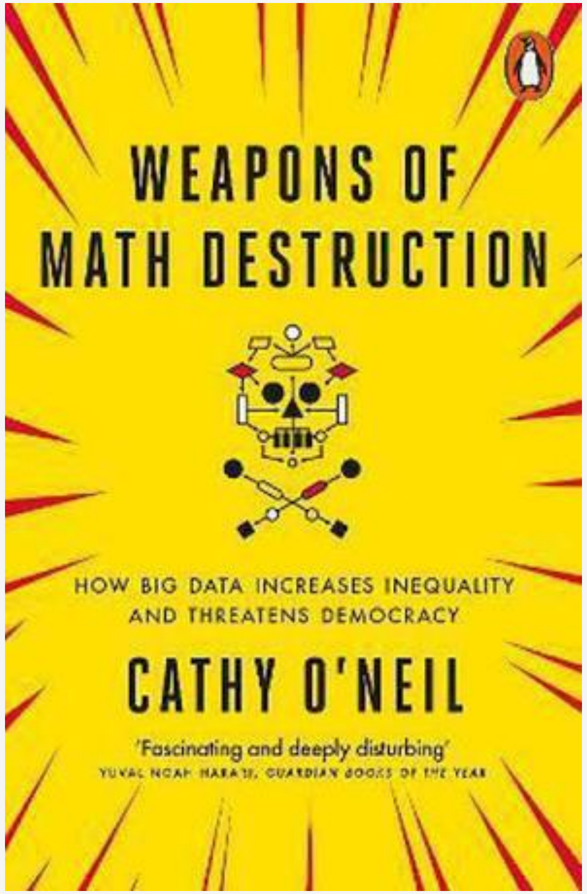
The importance of addressing Algorithmic Bias

Cochrane

How Big Data can Increase Inequity

Existing biases leading to unfair outcomes in hiring processes

- I = Automated algorithm to screen job applications and select candidates for interviews.
  - Scores are based on various factors using historical hiring data
  - If the data shows that certain demographics were underrepresented (such as women or minorities) in the past workforce, the algorithm might learn and replicate these biases
- O= Qualified candidates from underrepresented groups might be unfairly rejected, leading to an inequity within the company.


WEAPONS OF
MATH DESTRUCTION
HOW BIG DATA INCREASES INEQUALITY
AND THREATENS DEMOCRACY
CATHY O'NEIL
'Fascinating and deeply disturbing'
YUVAL NOAH HARARI, GUARDIAN BOOKS OF THE YEAR

Algorithmic Bias refers to

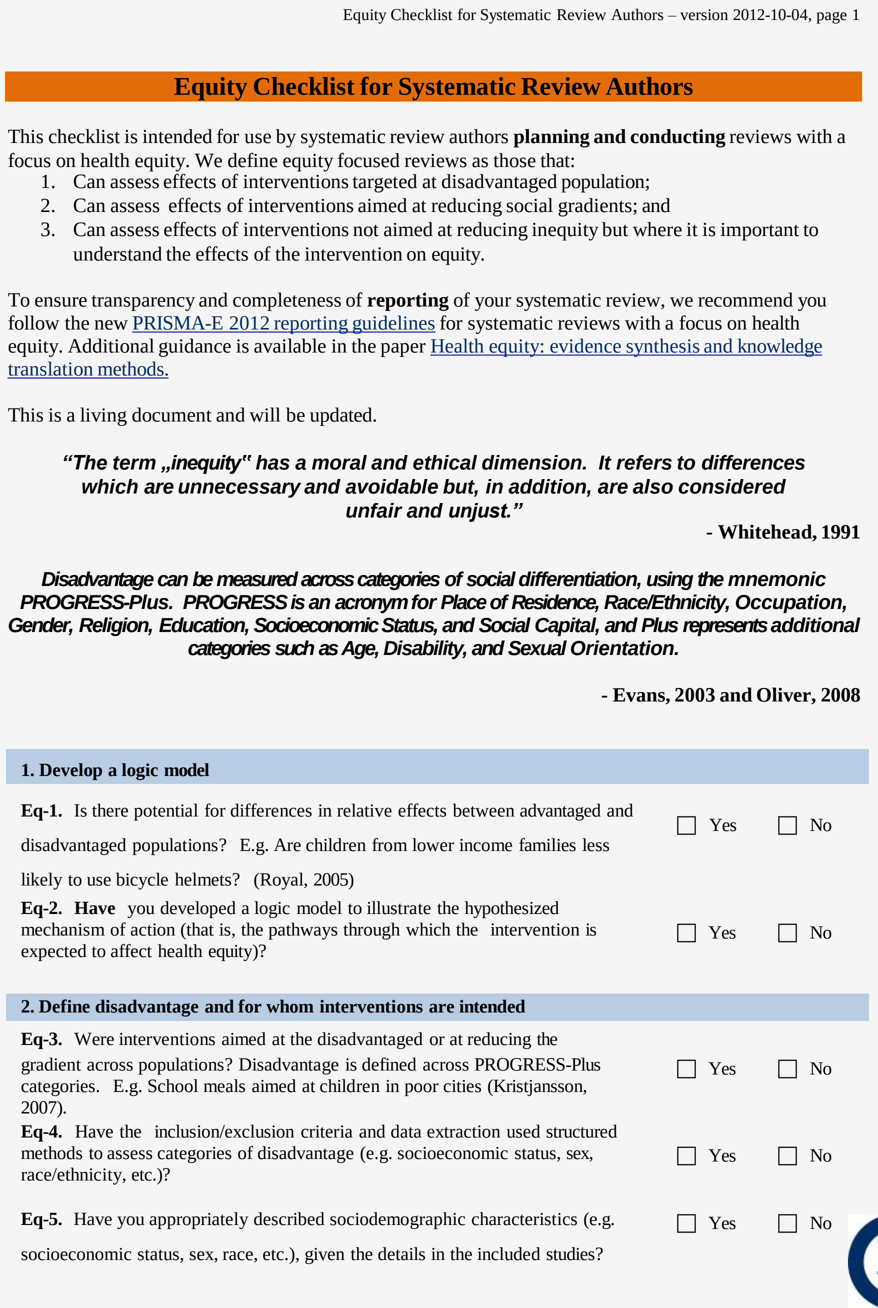
AI-Algorithms that are trained on biased data sets leading to unfair or discriminatory outcomes

Impact on SRs

Cochrane

Aim SR: Assess effects of interventions not aimed at reducing inequity but where it is important to understand the effects of the intervention on equity

Logic Model: Pathways through which the intervention is expected to affect health equity



Equity Checklist for Systematic Review Authors – version 2012-10-04, page 1

Equity Checklist for Systematic Review Authors

This checklist is intended for use by systematic review authors **planning and conducting** reviews with a focus on health equity. We define equity focused reviews as those that:

1. Can assess effects of interventions targeted at disadvantaged population;
2. Can assess effects of interventions aimed at reducing social gradients; and
3. Can assess effects of interventions not aimed at reducing inequity but where it is important to understand the effects of the intervention on equity.

To ensure transparency and completeness of **reporting** of your systematic review, we recommend you follow the new PRISMA-E 2012 reporting guidelines for systematic reviews with a focus on health equity. Additional guidance is available in the paper Health equity: evidence synthesis and knowledge translation methods.

This is a living document and will be updated.

"The term „inequity" has a moral and ethical dimension. It refers to differences which are unnecessary and avoidable but, in addition, are also considered unfair and unjust."

- Whitehead, 1991

Disadvantage can be measured across categories of social differentiation, using the mnemonic PROGRESS-Plus. PROGRESS is an acronym for Place of Residence, Race/Ethnicity, Occupation, Gender, Religion, Education, Socioeconomic Status, and Social Capital, and Plus represents additional categories such as Age, Disability, and Sexual Orientation.

- Evans, 2003 and Oliver, 2008

1. Develop a logic model

Eq-1. Is there potential for differences in relative effects between advantaged and disadvantaged populations? E.g. Are children from lower income families less likely to use bicycle helmets? (Royal, 2005) □ Yes □ No

Eq-2. **Have** you developed a logic model to illustrate the hypothesized mechanism of action (that is, the pathways through which the intervention is expected to affect health equity)? □ Yes □ No

2. Define disadvantage and for whom interventions are intended

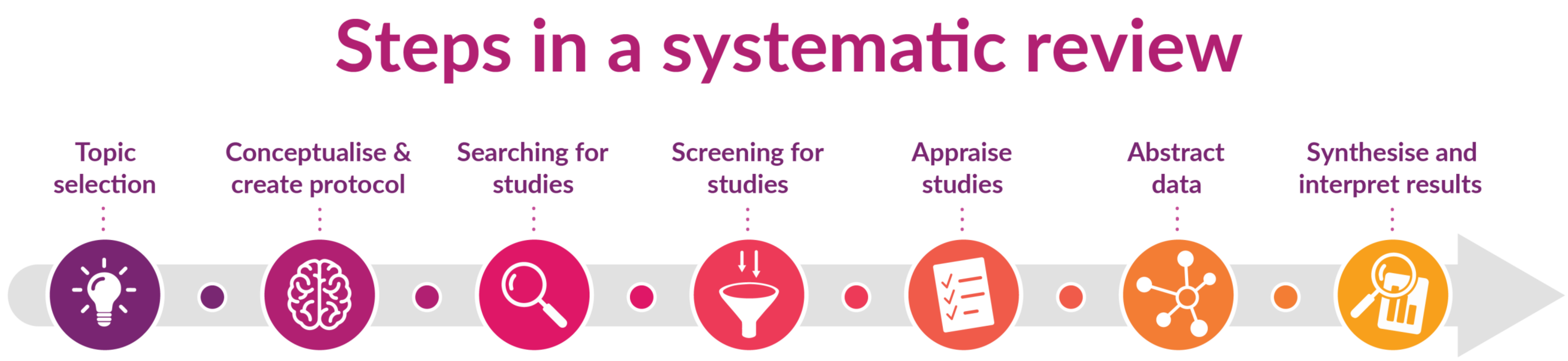
Eq-3. Were interventions aimed at the disadvantaged or at reducing the gradient across populations? Disadvantage is defined across PROGRESS-Plus categories. E.g. School meals aimed at children in poor cities (Kristjansson, 2007). □ Yes □ No

Eq-4. Have the inclusion/exclusion criteria and data extraction used structured methods to assess categories of disadvantage (e.g. socioeconomic status, sex, race/ethnicity, etc.)? □ Yes □ No

Eq-5. Have you appropriately described sociodemographic characteristics (e.g. socioeconomic status, sex, race, etc.), given the details in the included studies? □ Yes □ No

Cochrane Netherlands

Assess Algorithm Bias



Steps in a systematic review

Cochrane Netherlands

RoB with PROBAST-AI (=> prediction models)

Signaling questions to assess whether:

- appropriate predictors were selected?
- training data sets are skewed?
- the participants in data sets are representative for the group they serve → intended use of AI-algorithm

Data sets are selected or built by humans with their own natural biases

➢ Example low RoB: carefully-selected data from hospitals and research trials or national health data

Clinical impact of AI-interventions

Besides assessment of AI-algorithm, additional research needed on:

- how the AI-interentions could work in routine clinical practice (logistics/right care right place)
- their long-term effect on patient outcomes
- their cost-effectiveness

Conclusion

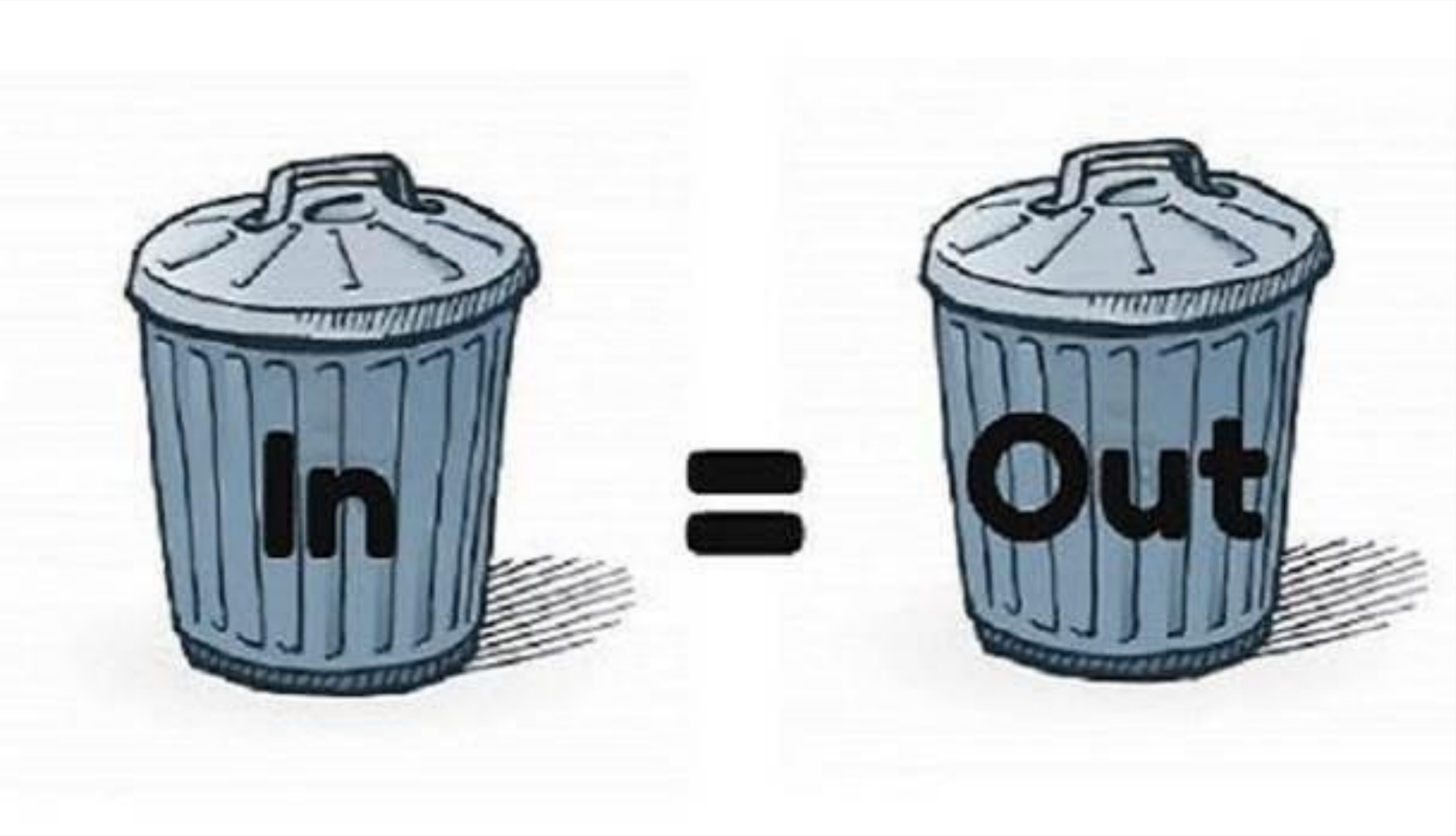
AI has the potential to help to:

- diagnose conditions earlier, and
- provide personalised treatments

Evaluating Trust in AI Interventions through Cochrane Evidence:

- Transparant research is essential
- Assess inclusiveness of datasets (to mitigate form algoritmic bias)

This also applies to AI-interventions….



Source LinkedIn Published on February 20th, 2014 | by James Cornehlsen

Cochrane Netherlands

Questions?

Contact details

- L.Hooft@umcutrecht.nl
- www.cochrane.nl