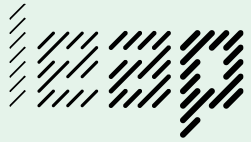
Understanding Bias in AI

Mitigation measures for addressing gender bias in artificial intelligence within healthcare settings: a critical area of sociological inquiry

Anna Isaksson, Associate Professor in Sociology. Leap for Life/School of Health and Welfare, Halmstad University

Overview

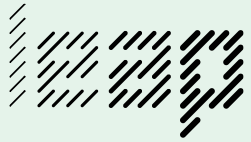


- Why a social science perspective?
- AI in healthcare: promises and risks
- Empirical examples – bias in healthcare settings
- The need for mitigation measures to address gender bias: What does the research suggest?
- Key challenges: responsibility, power, and implementation
- Conclusions


AI Healthcare Innovation
Transforming Medicine with Artificial Intelligence


Frust?
Better care?
AI in Healthcare
Who is included?
Frust?
Bias, Data and Health Outcomes
Represenation?
Gender?
Age?
Socioeconomic factors?
Consequences?
AI in Healthcare
Strategy · Governance · Implementation
• Quality
• Ethics
• Equity
• Resources
NATIONAL HEALTY
Digitalization & AI
• Regulation
• Ethics
• Equity
• Sustainability
GUIDELINES
PROCUREMENT
IMPLEMENTATION
EVALUATION
WHAT MATTERS?
PATIENT SAFETY
EQUAL CARE
TRUST
RESOURCES
LNO-TERM EFFECTS?

Why Do We Need Social Science Perspectives on AI?



- AI is still largely studied from a technical perspective, social consequences remain underexplored
- AI is not just technology, it also shapes how we think and live
- AI is not neutral, it reflects and reproduces power, values, and social structures
- Affects everyday life: work, health, relationships, and opportunities
- Influences power, fairness, and who gets included or excluded
- Raises important questions about trust, democracy, and our future
- Understanding AI requires looking at context, actors, and real-world impacts

Lindgren, S., & Holmström, J. (2020). A Social Science Perspective on Artificial Intelligence: Building Blocks for a Research Agenda. *Journal of Digital Social Research*, *2*(3), 1–15. https://doi.org/10.33621/jdsr.v2i3.65

AI & SOCIETY
https://doi.org/10.1007/s00146-024-02067-y

OPEN FORUM

Mitigation measures for addressing gender bias in artificial intelligence within healthcare settings: a critical area of sociological inquiry

Anna Isaksson[1]

Received: 7 May 2024 / Accepted: 4 September 2024

Abstract
Artificial intelligence (AI) is often described as crucial for making healthcare safer and more efficient. However, some studies point in the opposite direction, demonstrating how biases in AI cause inequalities and discrimination. As a result, a growing body of research suggests mitigation measures to avoid gender bias. Typically, mitigation measures address various stakeholders such as the industry, academia, and policy-makers. To the author's knowledge, these have not undergone sociological analysis. The article fills this gap and explores five examples of mitigation measures designed to counteract gender bias in AI within the healthcare sector. The rapid development of AI in healthcare plays a crucial role globally and must refrain from creating or reinforcing inequality and discrimination. In this effort, mitigation measures to avoid gender bias in AI in healthcare are central tools and, therefore, essential to explore from a social science perspective, including sociology. Sociologists have made valuable contributions to studying inequalities and disparities in AI. However, research has pointed out that more engagement is needed, specifically regarding bias in AI. While acknowledging the importance of these measures, the article suggests that they lack accountable agents for implementation and overlook potential implementation barriers such as resistance, power relations, and knowledge hierarchies. Recognizing the conditions where the mitigation measures are to be implemented is essential for understanding the potential challenges that may arise. Consequently, more studies are needed to explore the practical implementation of mitigation measures from a social science perspective and a systematic review of mitigation measures.

Mitigation measures for addressing gender bias in artificial intelligence within healthcare settings: a critical area of sociological inquiry

AI is expected to…

- revolutionize healthcare
- contribute to high-quality care
- support clinical decision-making, prediction, prioritization, diagnostics, treatment,
- At the same time, several studies point to the risk that AI may contribute to and reinforce discrimination and bias.

Bajwa J, Munir U, Nori A et al (2021) Artificial intelligence in healthcare: transforming the practice of medicine. Future Healthcare J 8(2):188–194. https://doi.org/10.7861/fhj.2021-0095

Alowais SA, Alghamdi SS, Alsuhebany N et al (2023) Revolutionizing healthcare: the role of artificial intelligence in clinical practice. BMC Med Education. https://doi.org/10.1186/s12909-023-04698-z

Liu Z (2021) Sociological perspectives on artificial intelligence: a typological reading. Sociol Compass 15:e12851. https://doi.org/10.1111/soc4.12851

O'Connor S, Liu H (2023) Gender bias perpetuation and mitigation in AI technologies: challenges and opportunities. AI Soc. https://doi.org/10.1007/s00146-023-01675-4

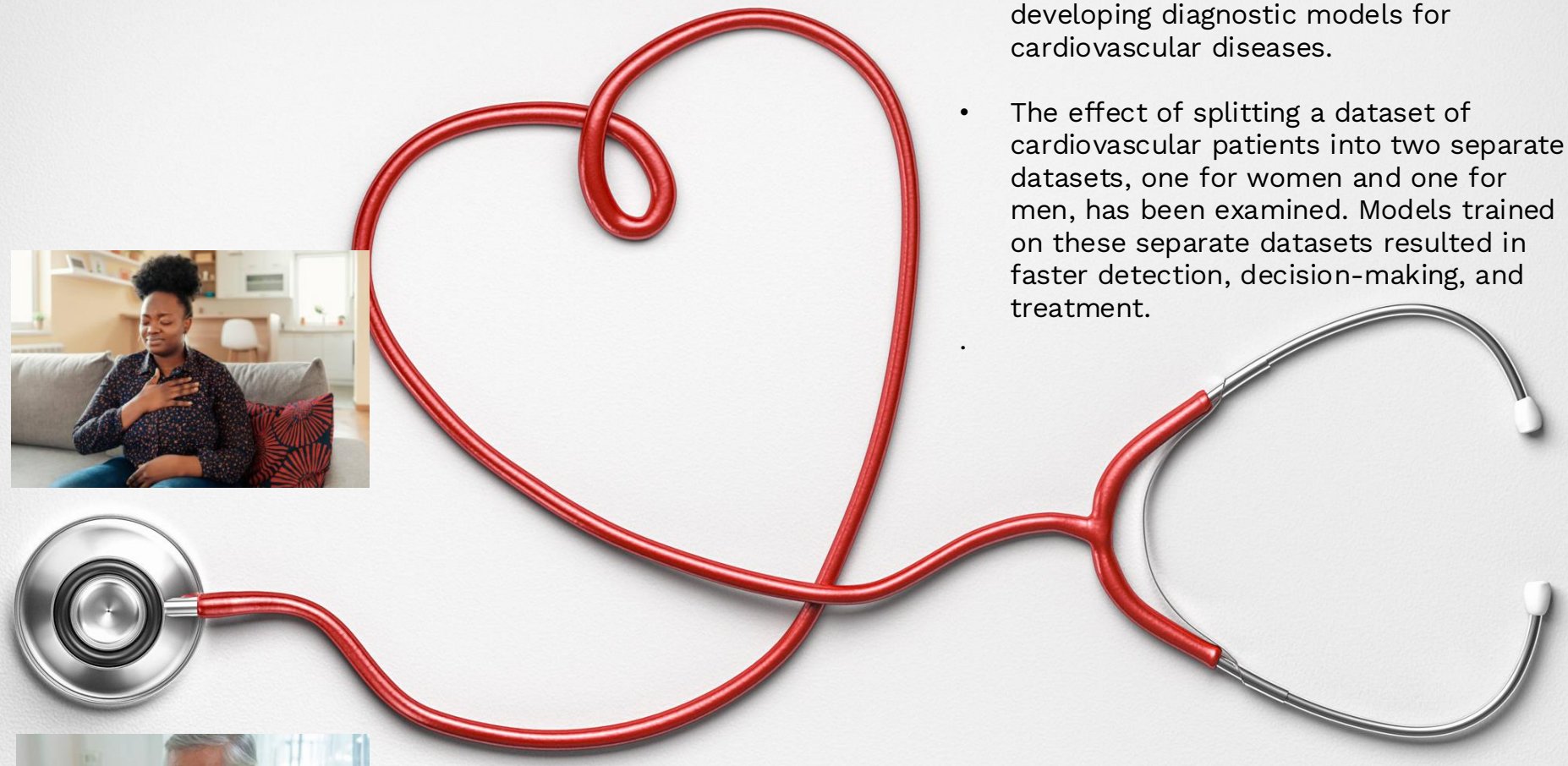
What are the consequences of gender bias in AI-based healthcare systems?

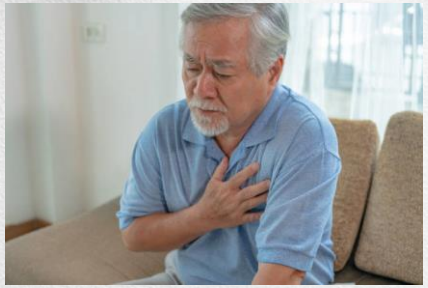
- Gender bias arises when algorithms are trained on non-representative datasets and/or are influenced by existing biases, norms, and stereotypes in society and healthcare.
- Non-representative datasets can lead to ineffective AI-based healthcare tools, misdiagnoses, inadequate treatment, and unequal health outcomes.
- Models trained on historical data may inherit stereotypes and, as a result, contribute to discriminatory decision-making in clinical settings.

Lau PL, Monomita N, Sushmita C (2023) Accelerating UN sustainable development goals with AI-driven technologies: a systematic literature review of women's healthcare". Healthcare 11(3):401. https://doi.org/10.3390/healthcare11030401

Rajkomar A, Hardt M, Howell MD et al (2018) Ensuring fairness in machine learning to advance health equity. Annals Int Med 169(12):866–872. https://doi.org/10.7326/M18-1990

Example: Cardiovascular diseases





- Symptoms are often more subtle in women, which can lead to delayed diagnosis and treatment. Therefore, it is crucial to take gender into account when developing diagnostic models for cardiovascular diseases.
- The effect of splitting a dataset of cardiovascular patients into two separate datasets, one for women and one for men, has been examined. Models trained on these separate datasets resulted in faster detection, decision-making, and treatment.

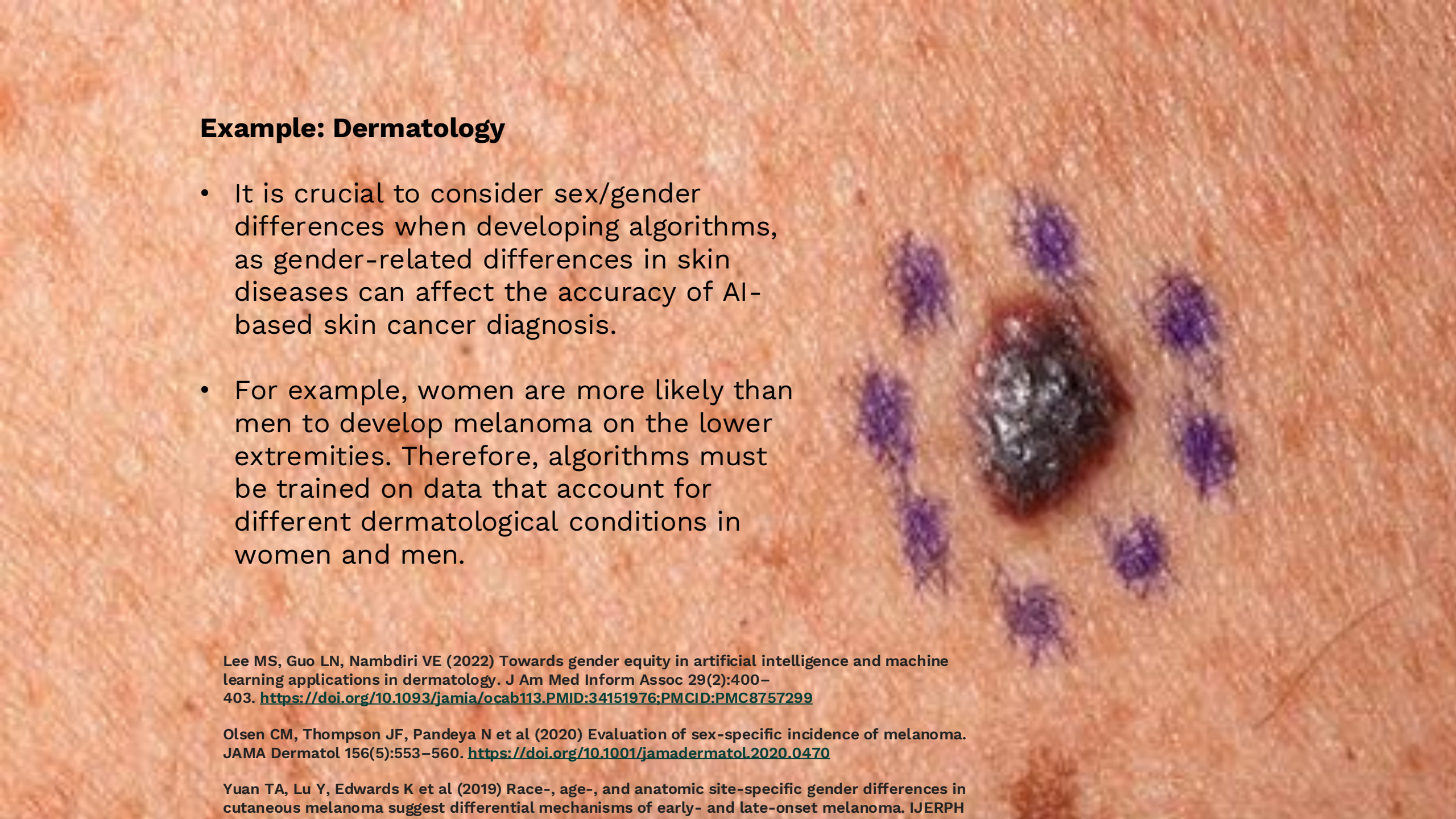
.

Baart SJ, Dam V, Scheres LJJ, Damen JAAG et al (2019) Cardiovascular risk prediction models for women in the general population: a systematic review. PLoS One. https://doi.org/10.1371/journal.pone.0210329
Hogo MA (2020) A proposed gender-based approach for diagnosis of the coronary artery disease. SN Appl Sci. https://doi.org/10.1007/s42452-020-2858-1

Example: Dermatology

- It is crucial to consider sex/gender differences when developing algorithms, as gender-related differences in skin diseases can affect the accuracy of AI-based skin cancer diagnosis.

- For example, women are more likely than men to develop melanoma on the lower extremities. Therefore, algorithms must be trained on data that account for different dermatological conditions in women and men.



Lee MS, Guo LN, Nambdiri VE (2022) Towards gender equity in artificial intelligence and machine learning applications in dermatology. J Am Med Inform Assoc 29(2):400–403. https://doi.org/10.1093/jamia/ocab113.PMID:34151976;PMCID:PMC8757299

Olsen CM, Thompson JF, Pandeya N et al (2020) Evaluation of sex-specific incidence of melanoma. JAMA Dermatol 156(5):553–560. https://doi.org/10.1001/jamadermatol.2020.0470

Yuan TA, Lu Y, Edwards K et al (2019) Race-, age-, and anatomic site-specific gender differences in cutaneous melanoma suggest differential mechanisms of early- and late-onset melanoma. IJERPH

- Several studies point to a growing need for measures, guidelines, and checklists to reduce the risks of bias while maximizing the benefits of AI in healthcare contexts. Such measures involve concrete strategies to prevent negative effects, for example by addressing bias in the design, development, and implementation of AI systems.

Straw EA, Stanley DA (2023) Weak evidence base for bee protective pesticide mitigation measures. J Econ Entomol 116(5):1604–1612. https://doi.org/10.1093/jee/toad118

Gray M et al (2024) Measurement and mitigation of bias in artificial intelligence: a narrative literature review for regulatory science. Clin Pharmacol Ther 115:687–697. https://doi.org/10.1002/cpt.3117

European Parliament (2022) Artificial intelligence in healthcare. Applications, risks, and ethical and societal impacts. EPRS. European Parliamentary Research Service Scientific Foresight Unit (STOA).

- “Mitigation measures” are understood as specific practices, actions, recommendations, and/or strategies that are undertaken to reduce and/or prevent undesirable effects of AI, while at the same time enabling and supporting its use.

Straw EA, Stanley DA (2023) Weak evidence base for bee protective pesticide mitigation measures. J Econ Entomol 116(5):1604–1612. https://doi.org/10.1093/jee/toad118

Gray M et al (2024) Measurement and mitigation of bias in artificial intelligence: a narrative literature review for regulatory science. Clin Pharmacol Ther 115:687–697. https://doi.org/10.1002/cpt.3117

European Parliament (2022) Artificial intelligence in healthcare. Applications, risks, and ethical and societal impacts. EPRS. European Parliamentary Research Service Scientific Foresight Unit (STOA).

Aim
To examine mitigation measures proposed in the research literature to reduce gender bias in AI within healthcare.

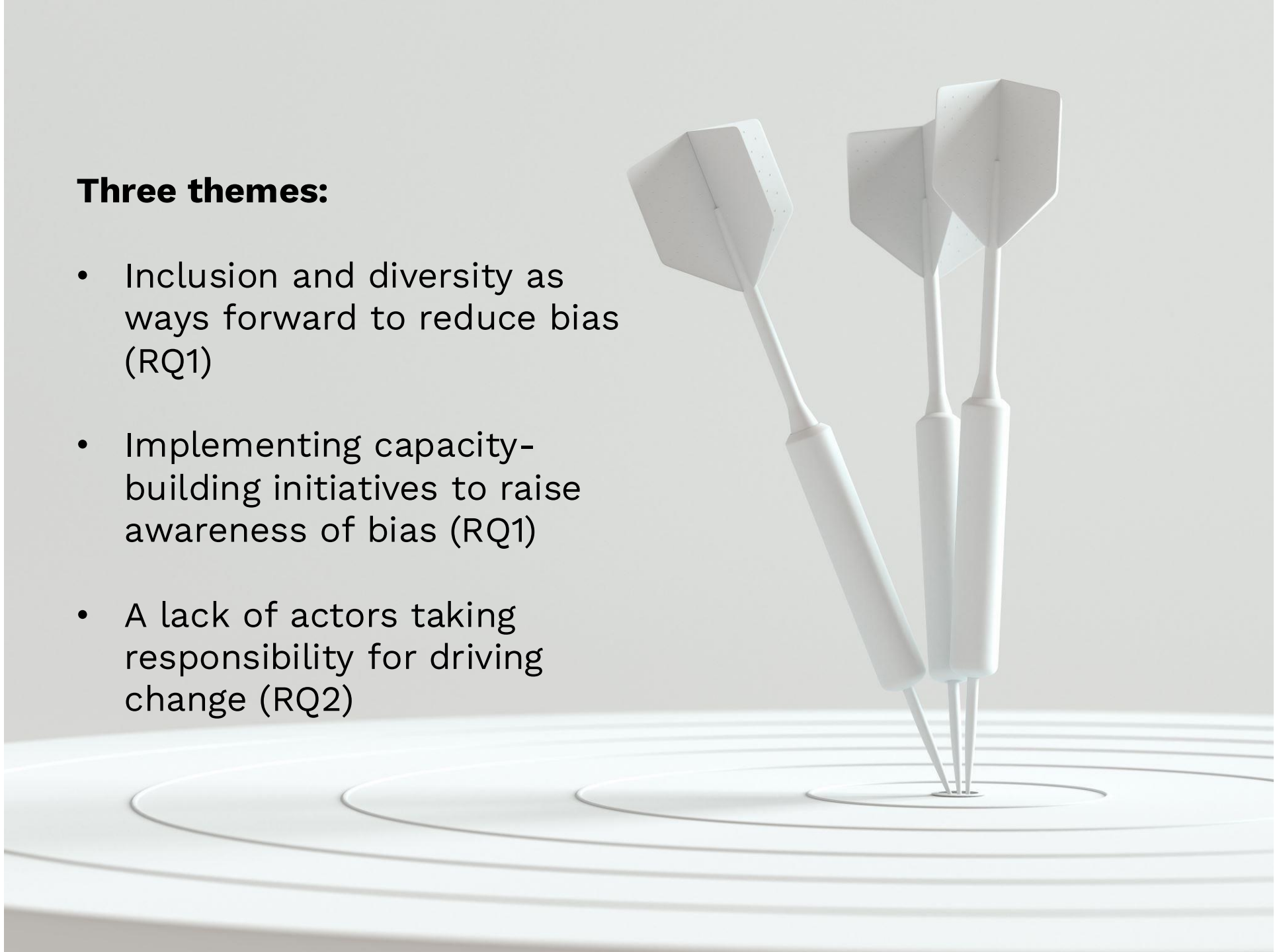
Research questions
1 What “paths” for counteracting bias are proposed?
2 Which actors are the measures directed towards?

Method
Case study approach: Analysis of five examples of mitigation measures presented in peer-reviewed articles.

Focus on recommendations addressing social rather than technical aspects.

Three themes:

- Inclusion and diversity as ways forward to reduce bias (RQ1)
- Implementing capacity-building initiatives to raise awareness of bias (RQ1)
- A lack of actors taking responsibility for driving change (RQ2)



Inclusion and diversity as ways forward to reduce bias

- Gender and diversity perspectives
- Interdisciplinary and cross-sector collaboration
- Diversity in teams

Inclusion and diversity as ways forward to reduce bias

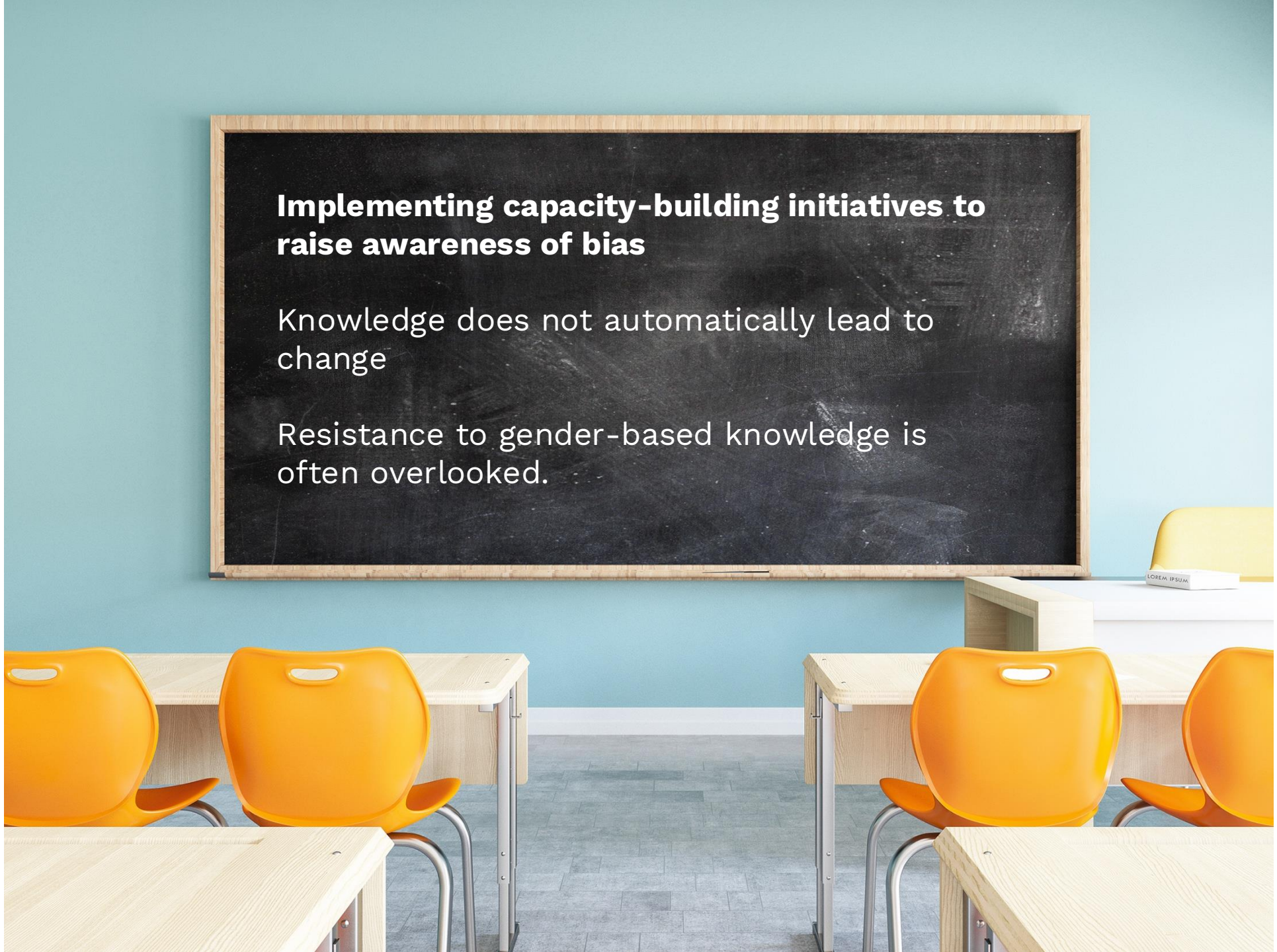
Interdisciplinarity

Often overlooks knowledge hierarchies, differences in terminology, and methodological conflicts between technical fields and the social sciences.

Inclusion and diversity as ways forward to reduce bias

Integrating gender perspectives in research

Despite progress in integrating gender perspectives into research, implementation is still constrained by structural conditions, power relations, and resistance to gender theory, highlighting the need for broader systemic change.


Implementing capacity-building initiatives to raise awareness of bias
Knowledge does not automatically lead to change
Resistance to gender-based knowledge is often overlooked.
LOREM IPSUM

Lack of actors taking responsibility

- Many recommendations do not clearly identify which actors or functions are responsible for implementing the proposed measures, making accountability vague or entirely absent.
- The proposals are often directed at broad stakeholder groups (e.g., “academia” or “AI practioners”) without specifying roles, structures, or mechanisms for follow-up.
- Research highlights the importance of gender-aware leadership and accountability at the highest management level to ensure meaningful implementation of gender-related measures.

Conclusions

- The recommendations and proposed mitigation measures are important.
- However, they tend to overlook the inequalities that may characterize the contexts in which they are intended to be implemented.
- For these recommendations to be implemented effectively, organizational readiness, clear leadership, and attention to potential power relations and knowledge hierarchies within the organizations involved are required.

Conclusions

- There is a clear need for sociological and socially oriented research on AI in healthcare.
- More research is needed to examine the conditions for development and implementation, including power dynamics, organizational readiness, and resistance.
- Sociology and the social sciences more broadly have an important role to play in shaping responsible AI in healthcare.

Questions?

- Anna Isaksson
- Anna.Isaksson@hh.se
- 072-9773510