

TRUSTMIND: AI in Mental Health

Ensuring Evaluation,
Trustworthiness, and Ethical
Considerations

Associate Professor Melanie Ganz-
Benjaminsen

Department of Computer Science

KØBENHAVNS UNIVERSITET
University of Copenhagen



Case study: Building a prioritization AI model for depression



Morten is feeling depressed and anxious. He sees his general practitioner (GP) about it.

The GP wants to refer him to an expert, but the average **waiting time** to see a specialist is **92 weeks**.

But the GP has access to a medical AI **model that determines the risk** of Morten having a depression.

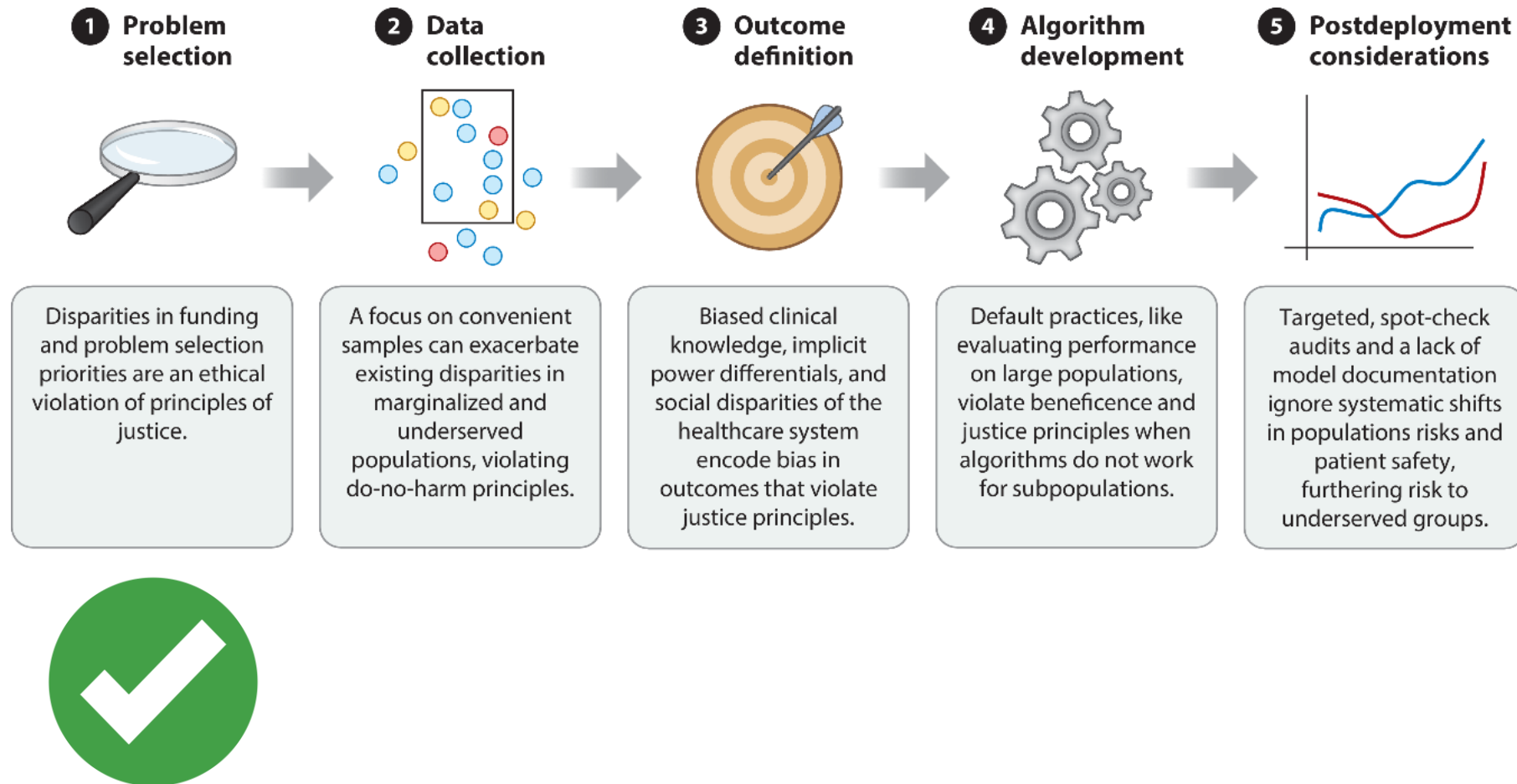


How do we ensure that such a medical AI model is trustworthy, meaning fair and transparent?

Investigate pitfalls & challenges of a medical AI in a case study of depression

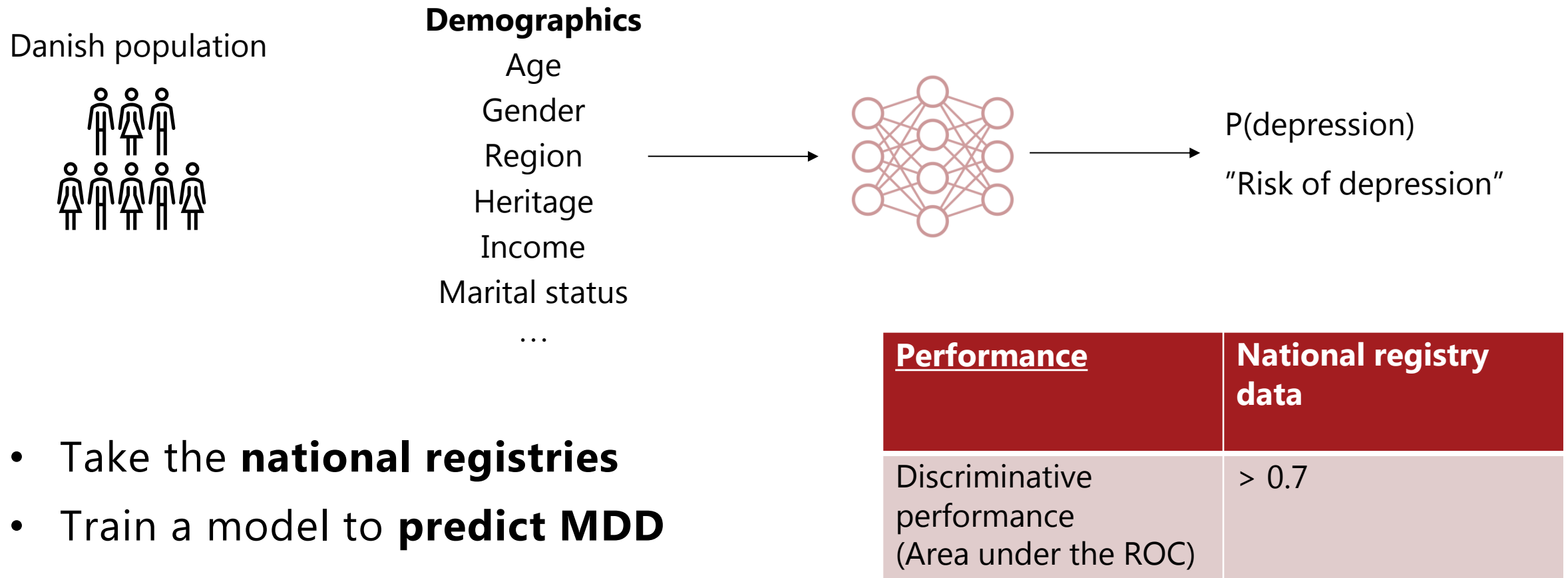


How do we develop medical AI

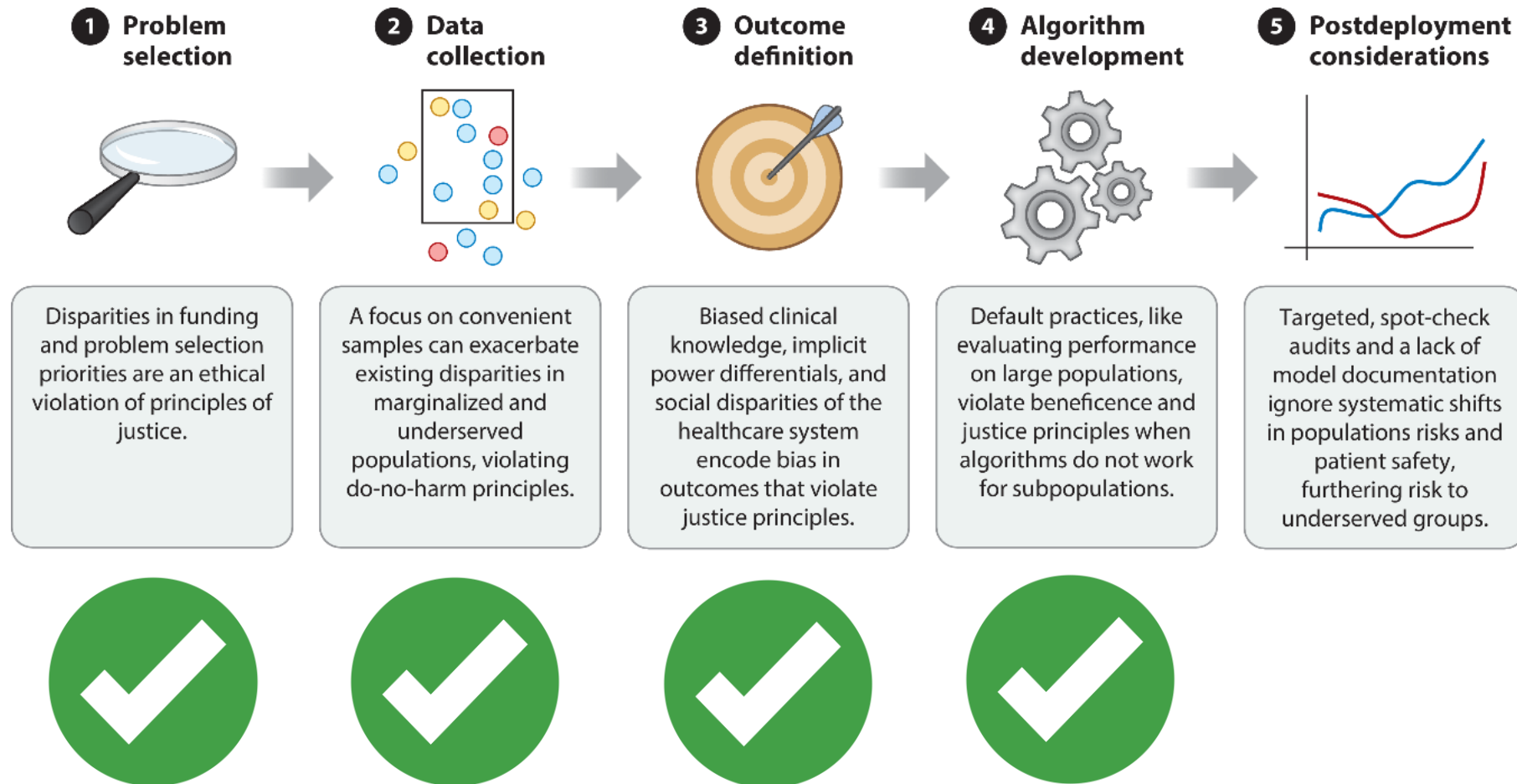


[Chen et al. , Annual Review of Biomedical Data Science Volume 4, 2021]

Prediction of risk of depression in the general population

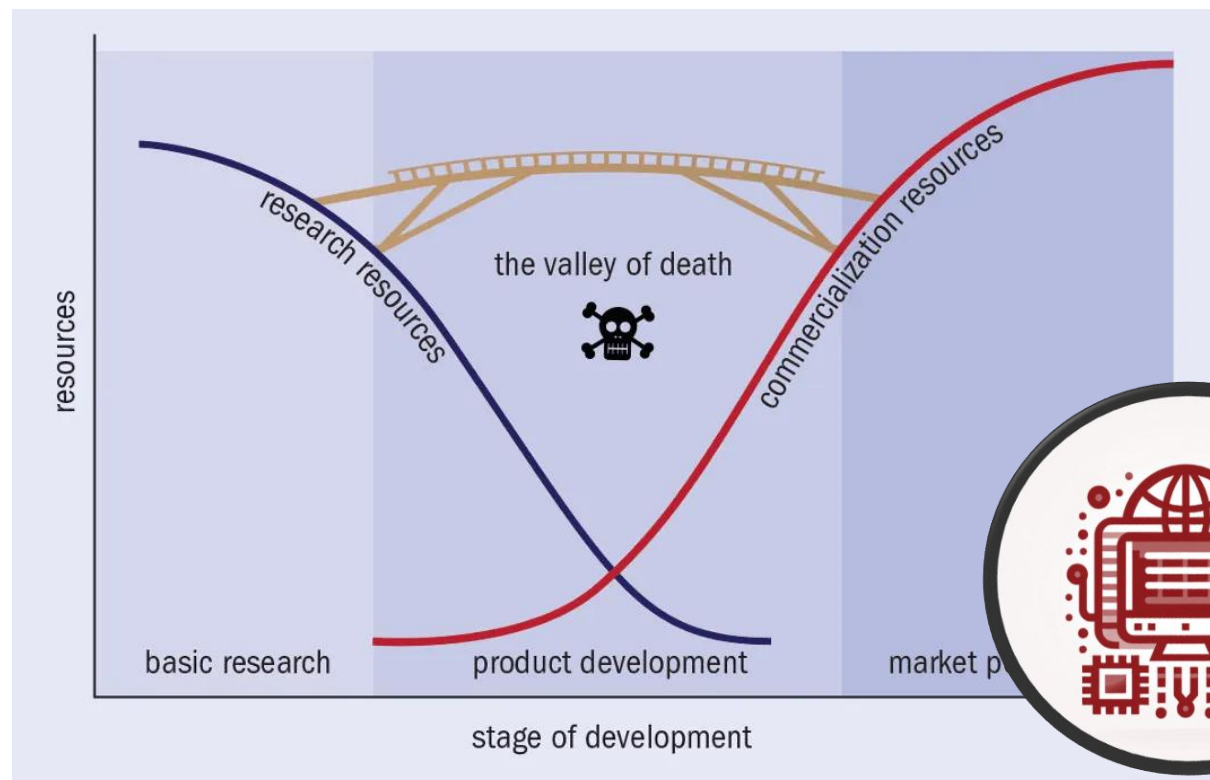


How do we develop medical AI



[Chen et al. , Annual Review of Biomedical Data Science Volume 4, 2021]

Why are these models not deployed?



SOURCE: <https://www.owensdesign.com>

GDPR
EU AI ACT

Legislation is not the problem!

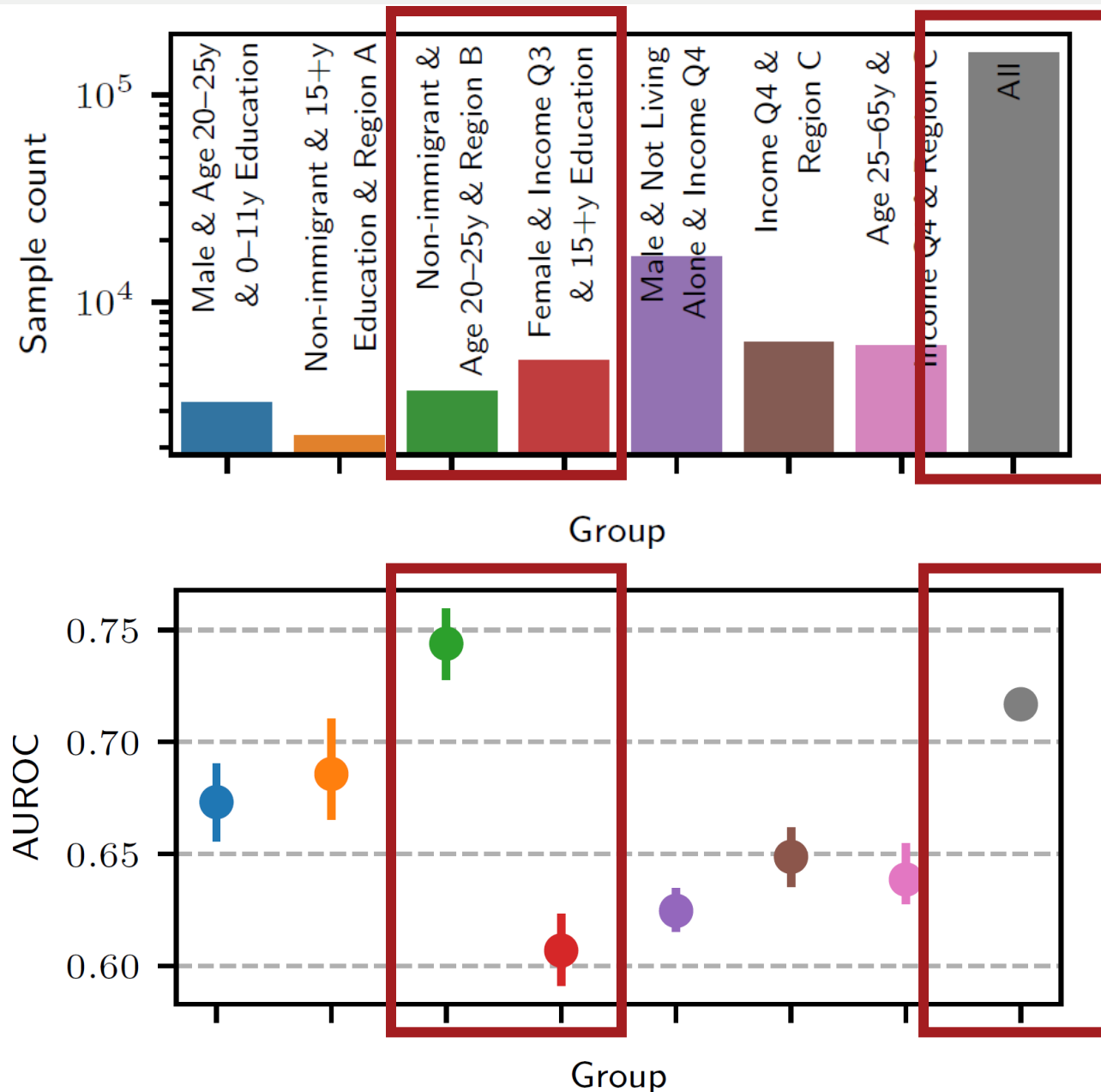
GDPR/AI ACT require trustworthy AI -- but we don't know how to validate trustworthiness!



Algorithms are not fair yet

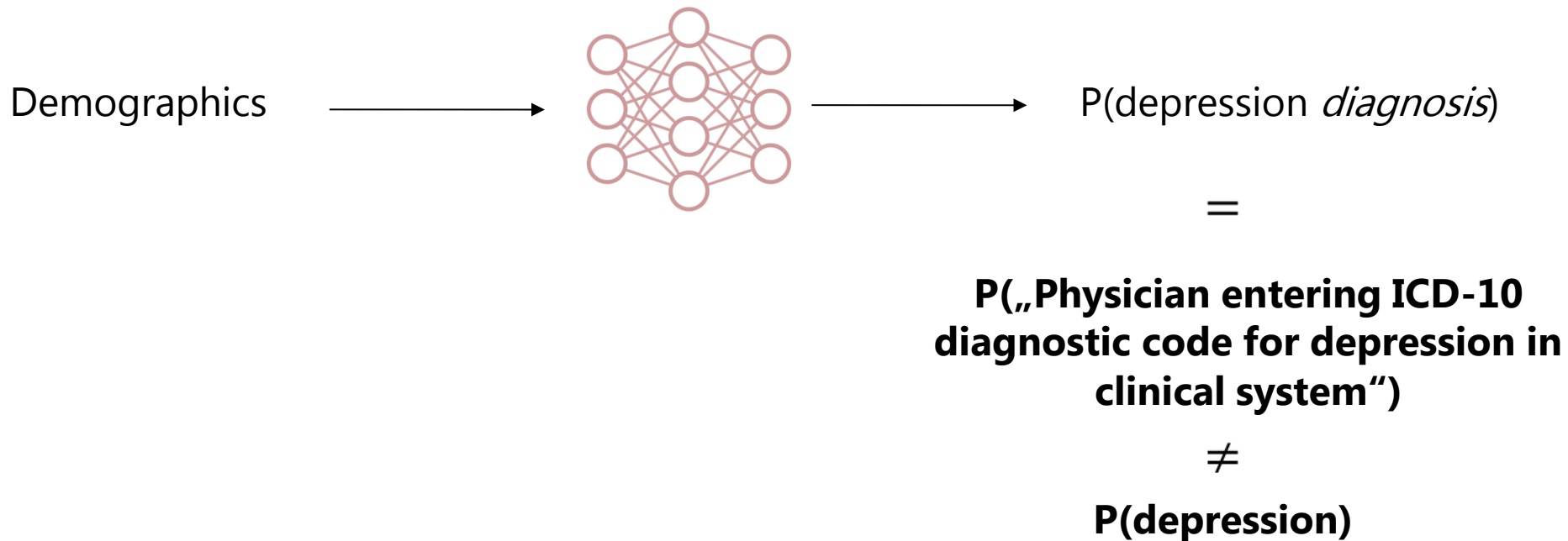
Preliminary results:

- Results on average vs. in subpopulations vary
- Performance in detecting depression ranging from 0.6 to 0.75 across subgroups



[Petersen, Ganz, Holm & Feragen, FAccT, 2023]

Our labels are not perfect



- What are the biases in receiving the ICD-10 code for depression?
- **Is the code the relevant outcome?** Or rather suicide, job loss, etc?
- **Who are we missing in this data set?**

[Obermeyer et al., Science, 2019]

TRUSTMIND: AI in Mental Health

Ensuring Evaluation, Trustworthiness, and Ethical Considerations

A novel and significant **test case for developing trustworthy medical AI** applied to a real-world process - mental health screening for depression

Efficacy: Ensuring positive patient outcomes through AI interventions, especially in clinical settings.



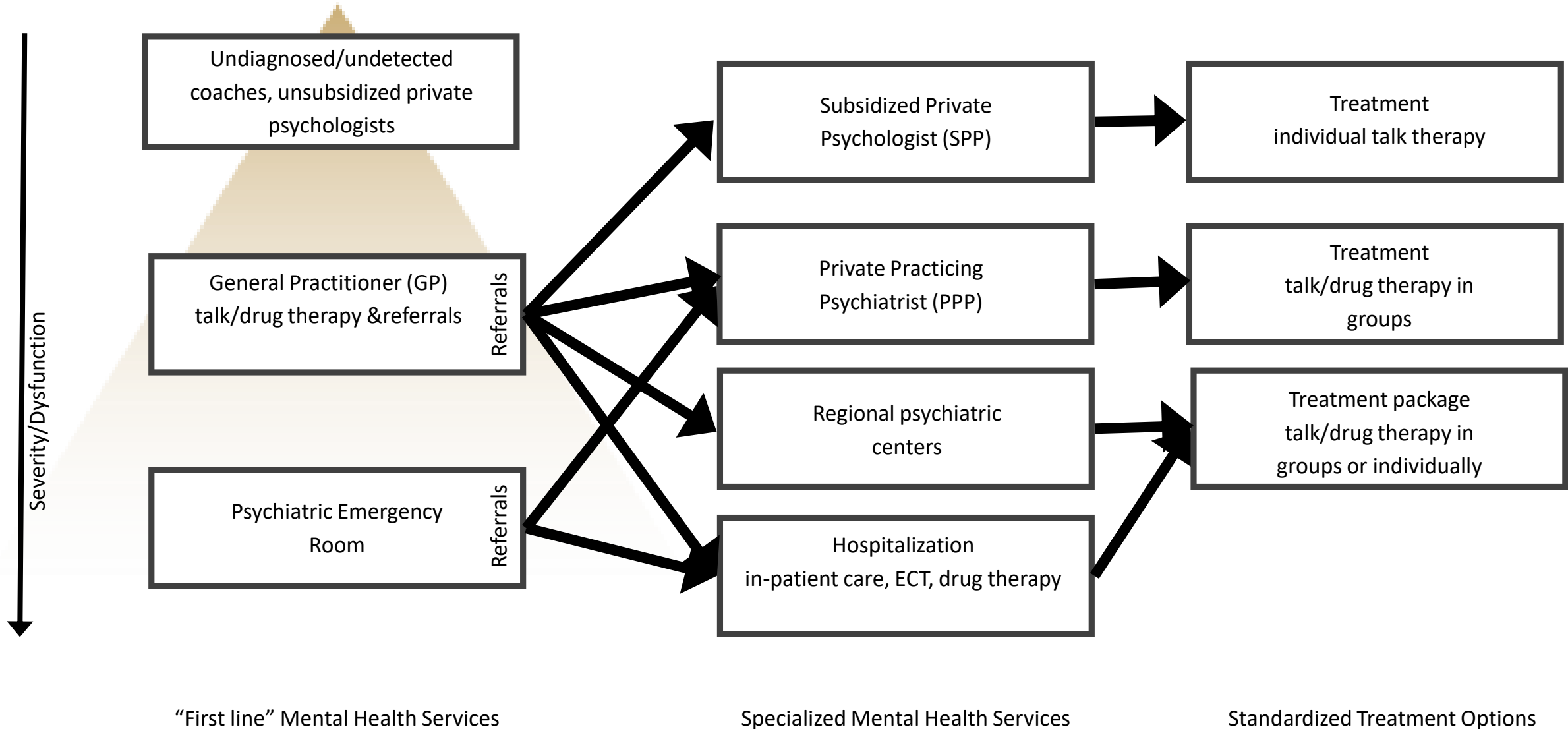
Transparency: Providing clarity for healthcare professionals and affected patients.



Fairness: Minimizing disparate impacts on patient subgroups.



Understanding the problem

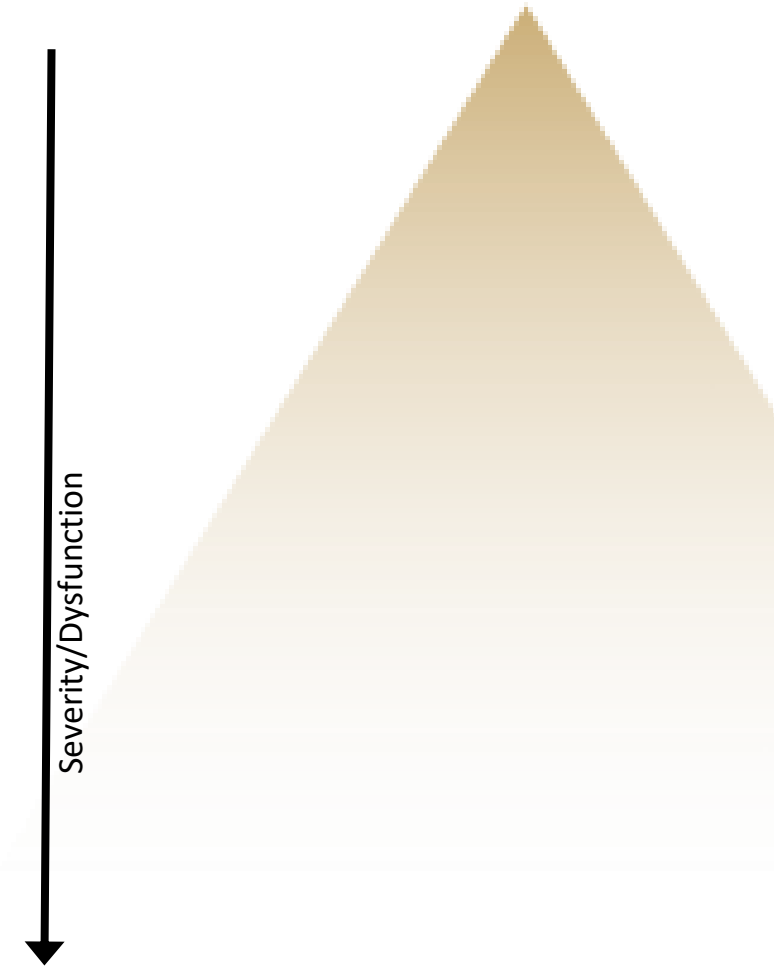


Developing depression triage at several levels

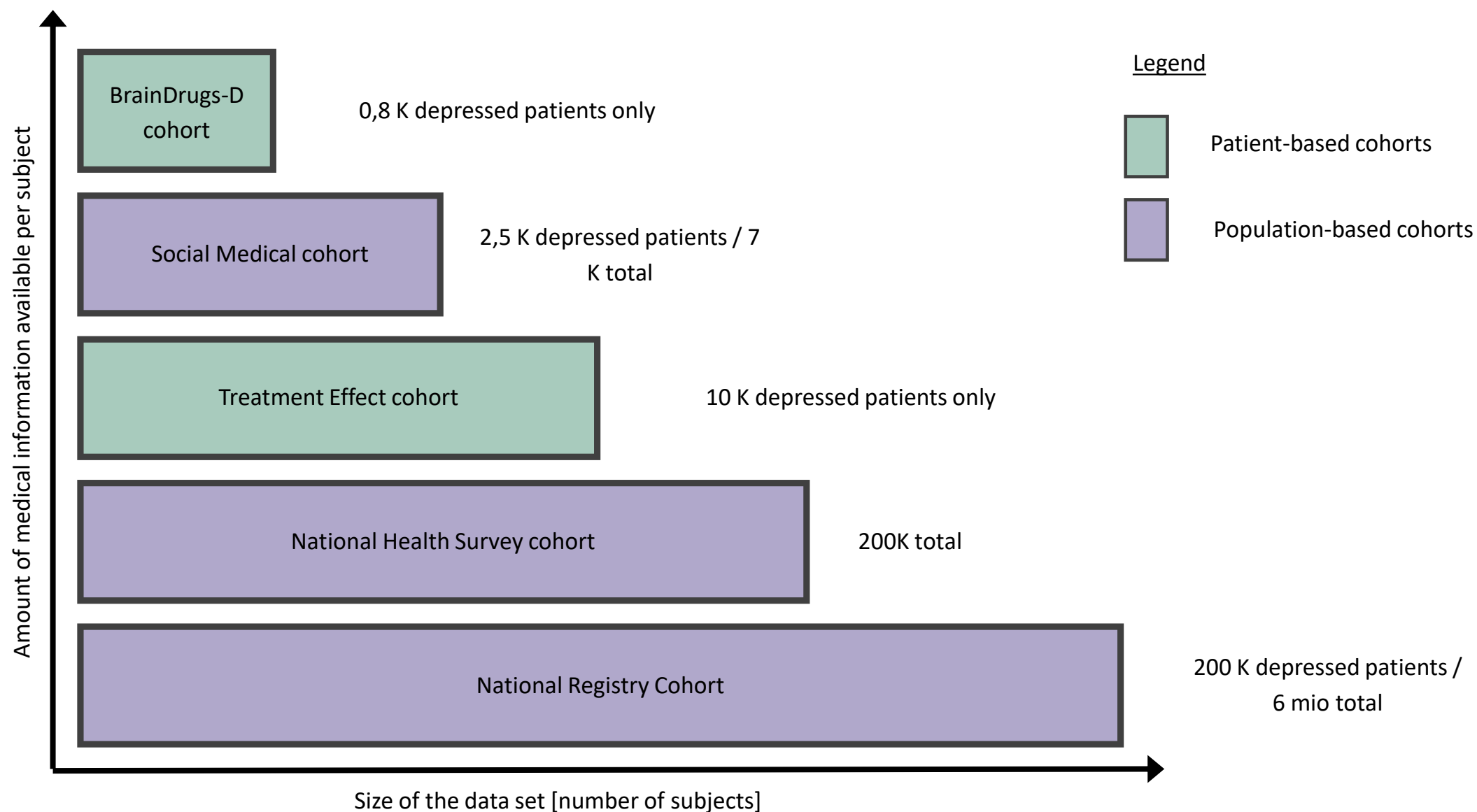
Two depression screening algorithms:

- Algorithm 1: **general population risk prediction**
 - Allows restructuring of waiting lists for access
- Algorithm 2: **adverse outcomes in patients**
 - Allows prioritization of patients at high risk of adverse outcomes

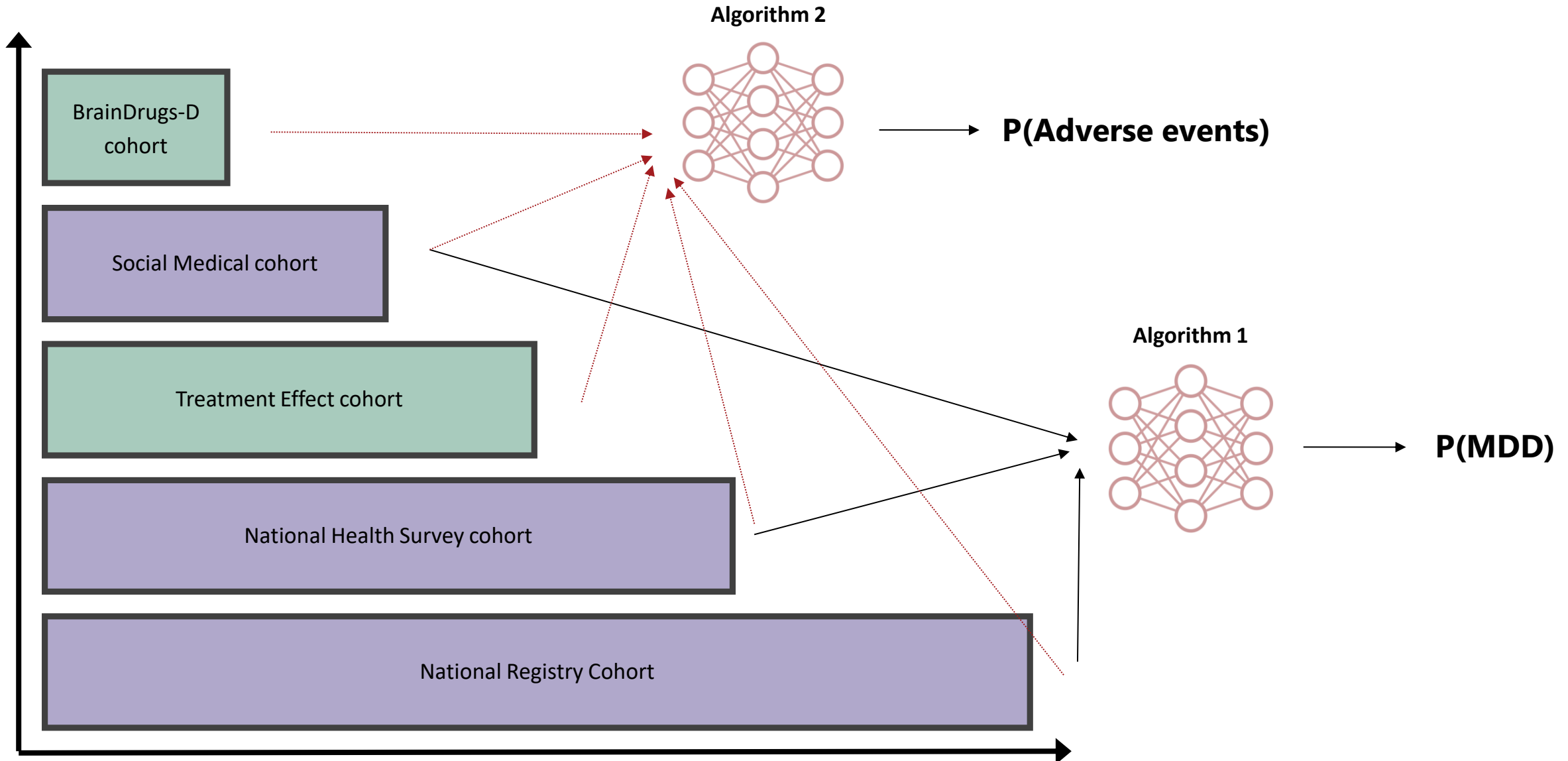
Apply the EU AI Act (AIA) to this clinical case, simulating its application beyond research



Datasets that cover the diversity of the population

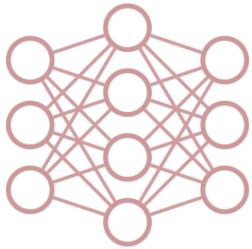


Depression triage at several levels



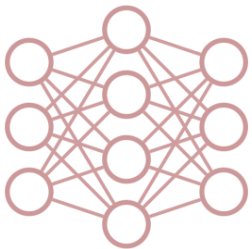
Algorithm analysis

Algorithm 1



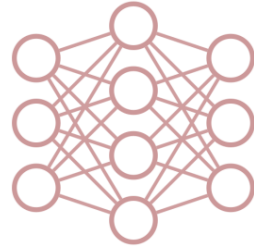
- Evaluate efficacy and fairness across different subpopulations in the data, including various age groups, genders, and socio-economic backgrounds
- Analyze the algorithm's transparency and uncertainty quantification

Algorithm 2

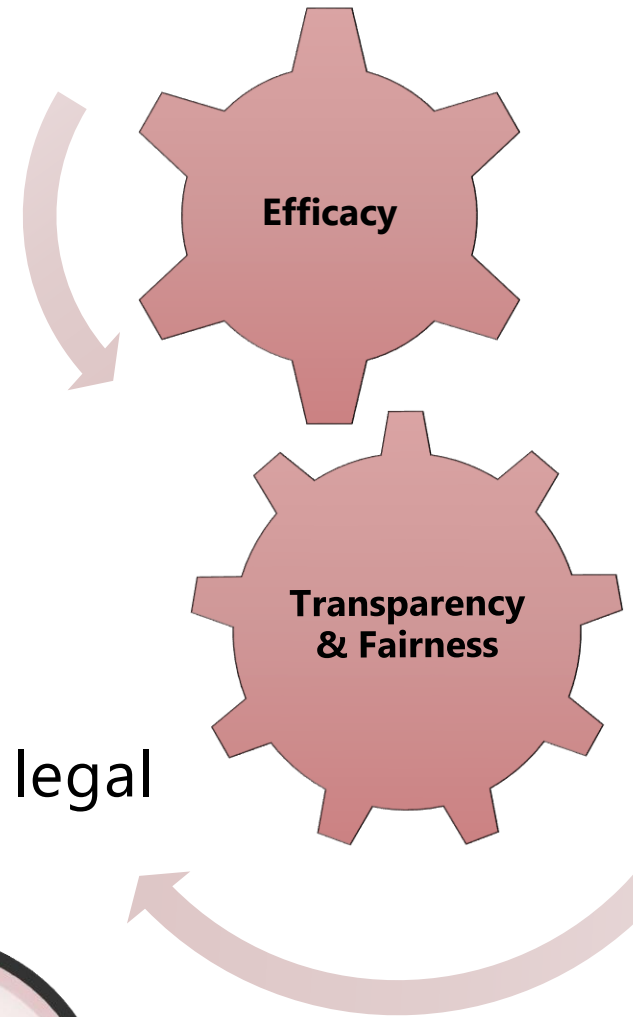


- Predict suicide attempts, deaths by suicide, and high healthcare resource utilization (HRU) for this population
- Evaluated across different depressed patient groups

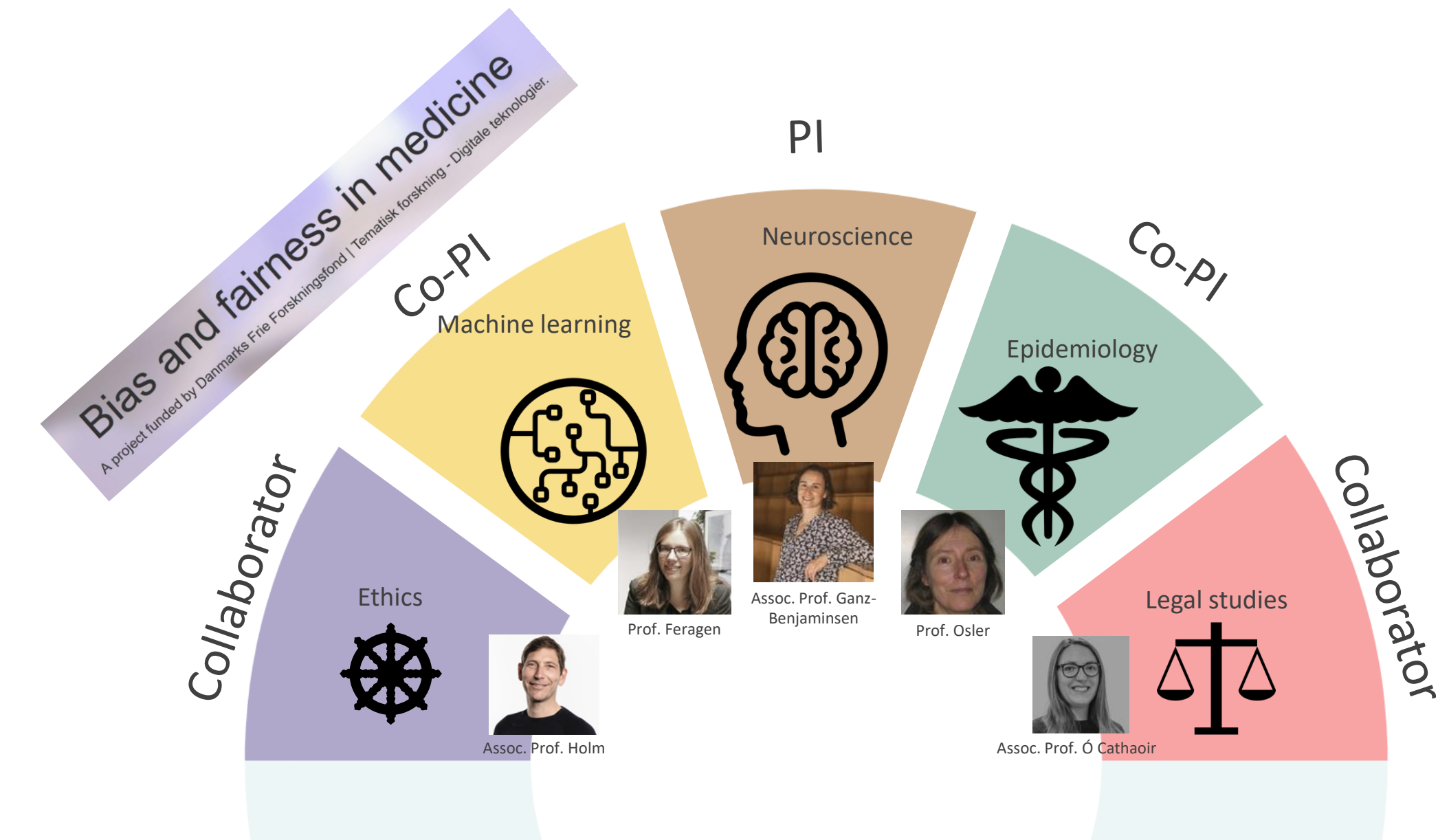
Detailed aims



1. Develop a trustworthy medical AI targeting depression in realistic datasets
1. Quantify its efficacy, transparency and fairness
1. Examine the challenges from ethical perspectives and the legal requirements imposed by regulations on trustworthy AI



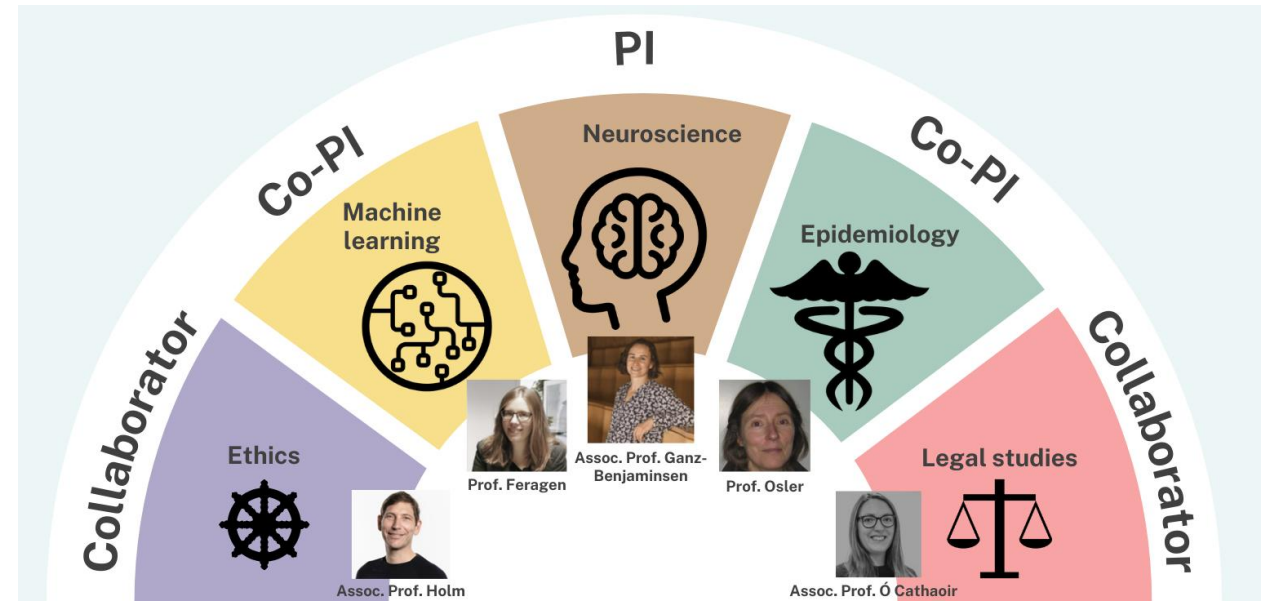
Strong interdisciplinary consortium



Investigate pitfalls & challenges of a medical AI in a case study of depression



Thank you!



Is AI for mental health triage fiction?

Augmenting clinical decision making with Chronosig



One of the latest innovations to emerge from Oxford's ideas factory, Chronosig is a digital triage for secondary mental healthcare that uses machine learning to deliver clinical decision-making tools that work alongside the clinician.

The system is designed to improve the speed and efficiency of triaging patient referrals by assisting with reviews of patient notes and making recommendations for treatment pathways.

Oxford based project



Modality allows clinical trials and researchers to get better data, faster, and at scale.

Modality.AI, Inc. has developed the first automated, clinically validated, multimodal system to assess neurological and psychiatric states, in clinic and remotely.

US start up company

