

# AI in medicine: a threat to personal autonomy?

Professor Søren Holm, BA, MA, MD, PhD, DrMedSci  
University of Manchester & Universitetet i Oslo



The University of Manchester



UiO • **University of Oslo**

# Overview

- Scope – AI used as advisory in relation to diagnosis or treatment planning & AI at least as good as the average health care professional usually making the decision
- The distinction and relationship between autonomy and liberty
  - The value(s) of autonomy
- How could the use of AI be a threat to autonomy?
  - Is explainability the solution?
- *Final note – AI may be a threat to the professional autonomy of health care professionals*

# Autonomy & Liberty

What you have **liberty** to do is the total universe of those choices / options that are available to you. This entails that we need an account of 'available' and the philosophical discussion of the concept of liberty is *inter alia* about what kind of 'non-availabilities' that count as restrictions of liberty

**Autonomy** is the freedom to choose within your choice set

Important distinction between negative and positive conceptions of respect for autonomy

- Negative conception – Respect is shown through non-interference, only
- Positive conception – Respect is shown through promotion and assistance

# Why is (negative) autonomy important? Why should we respect autonomous choices?

- A person's choices allow them to pursue a particular life-plan, and partly constitutes them as the person they are
- Having your choices overridden 1) makes most people angry, and 2) displays a lack of respect for them as competent choosers

Respect for autonomy overlaps with other considerations, e.g. respect in the role of citizen

# Autonomous choice and information

Trivial observation - You can only make an autonomous choice if you know what options you have

You can make an autonomous choice in cases of incomplete information, but it is many cases not rational to make such a choice if 1) the choice is important, and 2) the 'cost' of getting additional information is low

In the health care setting we usually assume that the system has the responsibility to keep information costs low for patients in relation to important choices

This is exemplified in our understanding of the standard mechanism for ensuring respect for autonomy in health care, i.e. the mechanism of 'informed consent' where the responsibility of providing the information component (and making sure the patient understands the information provided) falls to the relevant health care professional

# AI as a threat to autonomy 1

Initial observation - Some AI uses are potentially liberty restricting, AI embedded in equipment may *de facto* restrict the liberty to get a diagnosis without AI involvement

A system which does not allow patients to opt for 'human only' care is also liberty restricting, and limits the space for acting in the role of citizen

In so far as AI advice is non-explainable it may reduce the amount and/or relevance of the information that is, or can be made available in the choice situation

This does not remove the possibility of autonomous choice, but it is a threat to the quality of the choice

# AI as a threat to autonomy 2

In so far as AI advice is non-explainable it may reduce the amount and/or relevance of the information that is, or can be made available in the choice situation

This does not remove the possibility of autonomous choice, but it is a threat to the quality of the choice

At the limit, 'trust me, I am a doctor' is replaced with 'trust me, I am an AI'

# AI as a threat to autonomy 3

Patient interests are diverse, i.e. many health care decisions are 'preference sensitive', but AI advice (decisions?) are, in most cases based on an (implicit) fixed hierarchy of interests. If reliance on AI advice limits the choices presented to patients this creates an autonomy problem.



# Autonomy in relation to other AI issues

- Overreliance / automation bias
- De-skilling

Both problems make the input of health care professionals to any process of shared decision-making less reliable

# Is explainability the solution?

Technical explainability solves some issues, but does not necessarily provide the kind of explanation a patient wants or needs

Explainability does not solve overreliance or de-skilling issues

And, some times what is required is not an explanation but a possibility to get an adequate response to a contestation of the AI advice, or the decision made on the basis of the AI advice

“Everything depends, then, on what I know and what I want to know. If I am asked by an adult human being why Jones died, it will be no explanation to reply: ‘All men are mortal’, for so much, it can be presumed, he knows already; that is not the unfamiliar feature of the situation that is bothering him.”

Passmore J. Explanation in everyday life, in science, and in history. *History and Theory*. 1962 Jan 1;2(2):105-23.

# Contestability parameters

| <b>Contestability<br/>– Object</b> | <b>Contestability<br/>– Variables</b> | <b>Contestability<br/>– Questions</b>           | <b>Contestability<br/>– Required Explanation</b>                                      |
|------------------------------------|---------------------------------------|---|---|
| Data                               | <b>Types of data</b>                  | What types of personal data are used?           | 1) The decision D was made on the basis of data of type X, Y, Z about you             |
|                                    | <b>Data sources</b>                   | Where do your data come from?                   | 2) The decision D was based on data from sources X, Y, Z                              |
| Bias                               | <b>Training data</b>                  | On which data was the AI trained?               | 3) The decision D was made by a system trained on existing data of type X, Y, Z       |
|                                    | <b>‘Tagging groups’</b>               | Who tagged training data?                       | 4) The decision D made by a system trained on data tagged by X, Y, Z                  |
|                                    | <b>Tested for bias</b>                | Was training data or AI system tested for bias? | 5) The decision D was made by a system tested for bias of type X, Y, Z                |
| Diagnostic Performance             | <b>Performance</b>                    | What is the performance of the AI system?       | 6) The decision D was made by a system with a performance of X, Y, Z                  |
|                                    | <b>Performance testing</b>            | How was the performance determined?             | 7) The decision D was made by a system with a performance determined by tests X, Y, Z |
|                                    | <b>Essential indicators</b>           | What are key variables of AI decision-making?   | 8) The key input data resulting in decision D was X, Y, Z                             |
|                                    | <b>Alternatives</b>                   | Are alternatives considered?                    | 9) The alternatives to decision D are X, Y, Z with a probability of x, y, z (< d)     |
| Decision                           | <b>Longevity</b>                      | When is decision reconsidered?                  | 10) The decision D will be reconsidered if conditions X, Y, Z obtain                  |
|                                    | <b>AI involvement</b>                 | To what degree is AI making the decision?       | 11) The decision D involved an AI system with respect to X, Y, Z                      |
|                                    | <b>Human involvement</b>              | To what degree are humans making the decision?  | 12) The decision D was wholly/partly made by health professionals X, Y, Z             |
|                                    | <b>Responsibility</b>                 | Who is responsible for the decision?            | 13) The objective/legal responsibility for decision D is held by X, Y, Z              |

# A need for patient rights?

Patients may need individual rights to protect their individual interests

Minimally:

- A right to know that AI has been involved in their clinical management and how
- Transparency in relation to data used by the AI system(s)

But, also need for:

- Right to effective contestation
- Right to second opinion
- Right to refuse AI involvement / receive 'human only' care

# Acknowledgment and references

Most of my thinking and research about AI is done in collaboration with **Professor Thomas Ploug**, Aalborg University

## Conceptual papers

Ploug T, Holm S. The right to refuse diagnostics and treatment planning by artificial intelligence. *Medicine, Health Care and Philosophy*. 2020 Mar;23(1):107-14.

Ploug T, Holm S. The four dimensions of contestable AI diagnostics-A patient-centric approach to explainable AI. *Artificial Intelligence in Medicine*. 2020 Jul 1;107:101901.

Ploug T, Holm S. The right to a second opinion on Artificial Intelligence diagnosis—Remedying the inadequacy of a risk-based regulation. *Bioethics*. 2023 Mar;37(3):303-11.

## Empirical papers

Ploug T, Sundby A, Moeslund TB, Holm S. Population preferences for performance and explainability of artificial intelligence in health care: choice-based conjoint survey. *Journal of Medical Internet Research*. 2021 Dec 13;23(12):e26611.

Holm S, Ploug T. Population preferences for AI system features across eight different decision-making contexts. *Plos one*. 2023 Dec 1;18(12):e0295277.

[soren.holm@manchester.ac.uk](mailto:soren.holm@manchester.ac.uk)

[soren.holm@medisin.uio.no](mailto:soren.holm@medisin.uio.no)