Justifiable Black-Box AI

DR. SINEAD PRINCE

Centre for Biomedical Ethics
Yong Loo Lin School of Medicine
National University of Singapore

On behalf of:

PROF. JULIAN SAVULESCU

Outline

1. What is AI

2.The problem

3.Reasons for Justification

4.Highlights

Artificial Intelligence

technology, such as computer systems, that can simulate human comprehension, learning, and problem-solving by processing data and information from various sources in real-time.

Benefits

- Diagnostics e.g., medical imaging analysis, COVID-19 chest CT scans (Ding et al. 2022; Sand, Durán, and Jongsma 2021)
- Prognostics e.g., future hospitalisations and disease progression (Shickel et al. 2019)
- Treatment plans e.g., radiotherapy treatment plans (Wang et al. 2019)
- Remote telehealth and patient monitoring e.g., cardiac arrhytmia and falling (Devi and Kalaivani 2020; Pan et al. 2020)

Problem

AI is valuable for the reason it is unexplainable: but explicability is valuable for trust, transparency, accountability, autonomy, and fairness.

So, when, if at all, is it justifiable to use AI in healthcare?

Justifiability

Turns on the strength of reason, not explanation.

People can explain bad decisions, it doesn't make them justifiable

Let's examine why

Justifiability

Explainability

Bias

Seriousness

A black-box AI device suggests a pain relief treatment plan for a woman who has been suffering chronic pelvic pain for 10 years with an 70% accuracy rating but provides no definitive explanation for the condition or reason for the pain. However, the doctor disagrees with the AI device and provides a different treatment option that has moderate certainty (50%) but has an explanation. The doctor asks the patient which treatment she would like to try.

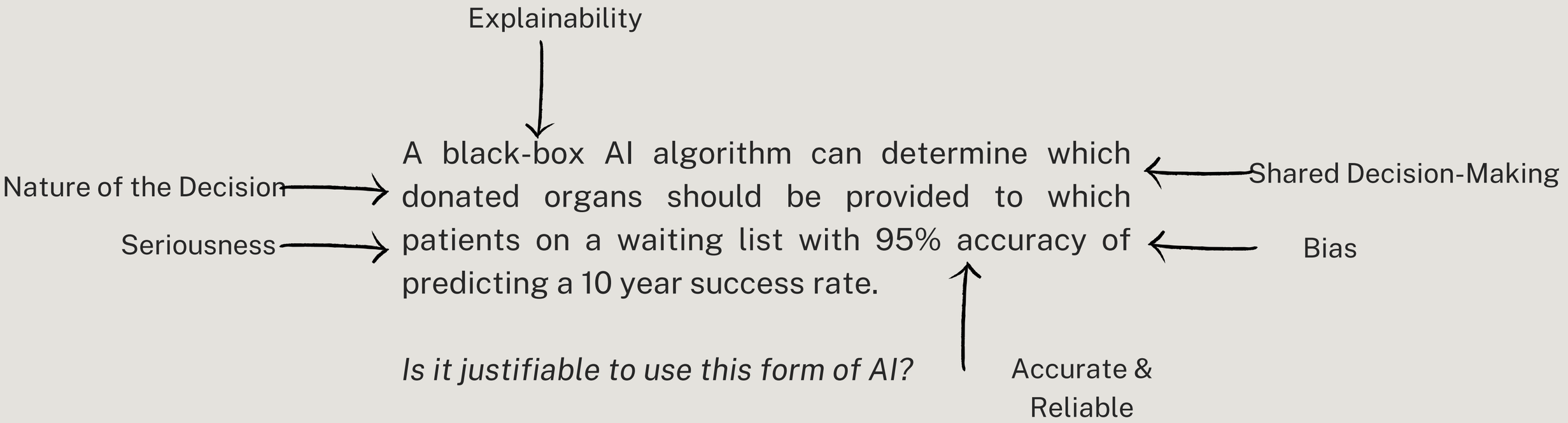
Nature of the Decision

Accurate & Reliable

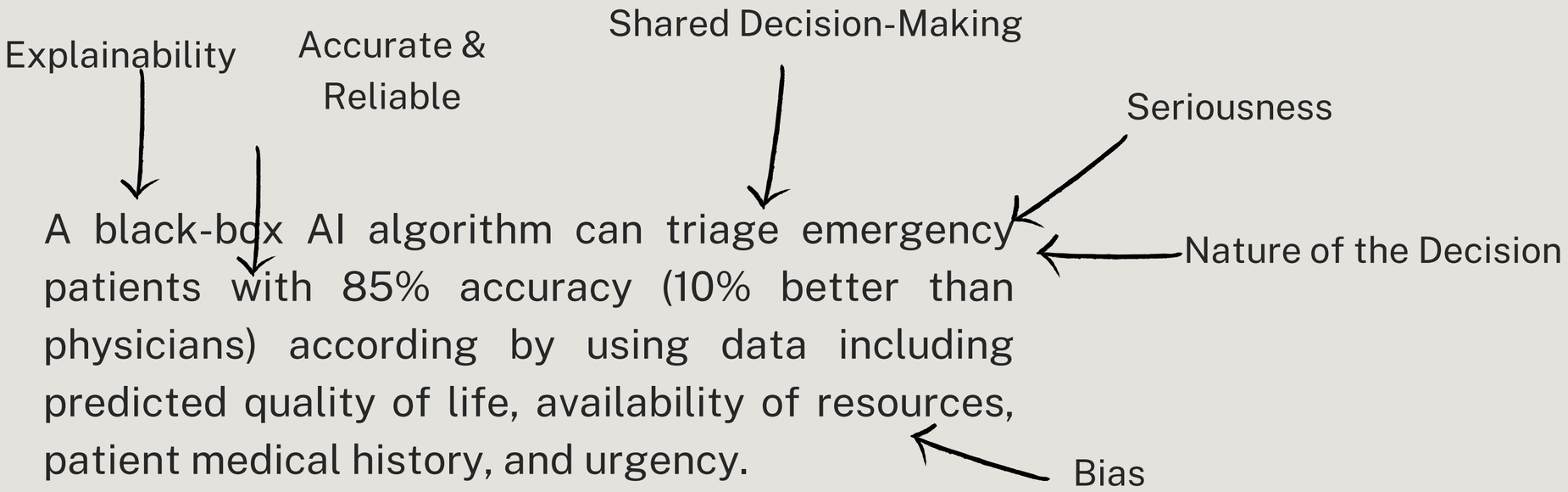
Shared Decision-Making

Is it justifiable to use this form of AI?

Justifiability



Justifiability



A black-box AI algorithm can triage emergency patients with 85% accuracy (10% better than physicians) according by using data including predicted quality of life, availability of resources, patient medical history, and urgency.

Is it justifiable to use this form of AI?

Highlights

- Explainability is important, but not necessary
- Accuracy is necessary if explainability is absent
- The more serious the condition, the more persuasive a more accurate decision
- The nature of the decision can be more important than the accuracy of the decision e.g., normative versus more clinical/ emergency versus non-emergency
- Human involvement through shared decision-making can render unexplainable, less urgent and even normative decisions justifiable

There is no blanket justification for any particular use - we ought to assess the black-box on a case-by-case basis drawing on how the reasons apply.

References

- Ding, Weiping, Mohamed Abdel-Basset, Hossam Hawash, and Ahmed M. Ali. 2022. Explainability of Artificial Intelligence Methods, Applications and Challenges: A Comprehensive Survey. Information Sciences 615 (November):238–292. doi:10.1016/j.ins.2022.10.013.
- Sand, Martin, Juan Manuel Durán, and Karin Rolanda Jongsma. 2021. Responsibility beyond Design: Physicians' Requirements for Ethical Medical AI. Bioethics 36 (2):162–169. doi:10.1111/bioe.12887.
- Shickel, Benjamin, Tyler J. Loftus, Lasith Adhikari, Tezcan Ozrazgat-Baslanti, Azra Bihorac, and Parisa Rashidi. 2019. DeepSOFA: A Continuous Acuity Score for Critically Ill Patients Using Clinically Interpretable Deep Learning. Scientific Reports 9 (1). Nature Publishing Group:1879. doi:10.1038/s41598-019-38491-0.
- Wang, Chunhao, Xiaofeng Zhu, Julian C. Hong, and Dandan Zheng. 2019. Artificial Intelligence in Radiotherapy Treatment Planning: Present and Future. Technology in Cancer Research & Treatment 18 (January). SAGE Publications Inc:1533033819873922. doi:10.1177/1533033819873922.
- Devi, R. Lakshmi, and V. Kalaivani. 2020. Machine Learning and IoT-Based Cardiac Arrhythmia Diagnosis Using Statistical and Dynamic Features of ECG. The Journal of Supercomputing 76 (9):6533–6544. doi:10.1007/s11227-019-02873-y.
- Pan, Daohua, Hongwei Liu, Dongming Qu, and Zhan Zhang. 2020. Human Falling Detection Algorithm Based on Multisensor Data Fusion with SVM. Mobile Information Systems 2020 (1):8826088. doi:10.1155/2020/8826088.

Thank You