Can We Explain AI?

Explainable AI in the Health Domain As Told Through Three European Commission-funded Projects

Presenter: Dr. Robin Renwick, Trilateral Research Ltd, Ireland
Email: Robin.Renwick@trilateralresearch.com

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 965221

Presentation Outline

Who we are

Overview of AI and XAI and relevance to healthcare

Projects overview and XAI methodologies

Summary

Challenges

Opportunities

Future plans

Who we are

TRILATERAL RESEARCH
Ethical AI

- Ethical AI company established in 2004
  - Innovation & Research (I&R)
  - Data protection and Cyber-risk (DCS)
  - Socio-tech for Good (STG)
- Health cluster focused on development of robust solutions that incorporate **principles of ethics- and privacy-by-design** to facilitate the development of responsible, trustworthy, and **ethical AI solutions** aligning with **social and legal values.**

Overview of AI and XAI

TRILATERAL RESEARCH
Ethical AI

- **AI** – development of software that can use human-defined objectives to generate outputs such as content, predictions, or decisions influencing various environments

- **XAI** – development of AI systems that can provide understandable explanations for their decisions and outputs.
  - aims to enhance **transparency**, **interpretability**, and **accountability** in AI systems, allowing users to trust and understand the reasoning behind AI-driven decisions.

Overview of AI and XAI

TRILATERAL RESEARCH
Ethical AI

Relevance in healthcare

- AI is increasingly being used for tasks such as diagnosis, treatment recommendation, and patient monitoring.
- Lack of transparency and interpretability in AI systems poses challenges – undermines trust and hinders adoption.
- Understanding and addressing complexities of XAI in healthcare is crucial for building trust among stakeholders, improving patient care and adhering to ethical principles.

Case Studies: EC-funded Projects

TRILATERAL RESEARCH
Ethical AI

iToBoS

Aim: To create an AI diagnostic platform for early skin melanoma detection, using a novel total body scanner and a computer-aided clinical decision support system that integrates patients' clinical information, genetic and imaging data, and family medical history

- ***Grant agreement number:*** *965221*
- ***Time frame:*** *April 2021 – March 2025*

COVINFORM

Aim: To explore the impacts of the COVID-19 pandemic across the EU member states and the UK by developing a comprehensive risk assessment dashboard providing insights into various dimensions of vulnerability across regions and demographics

- ***Grant agreement number:*** *101016247*
- ***Time frame:*** *November 2020 – October 2023*

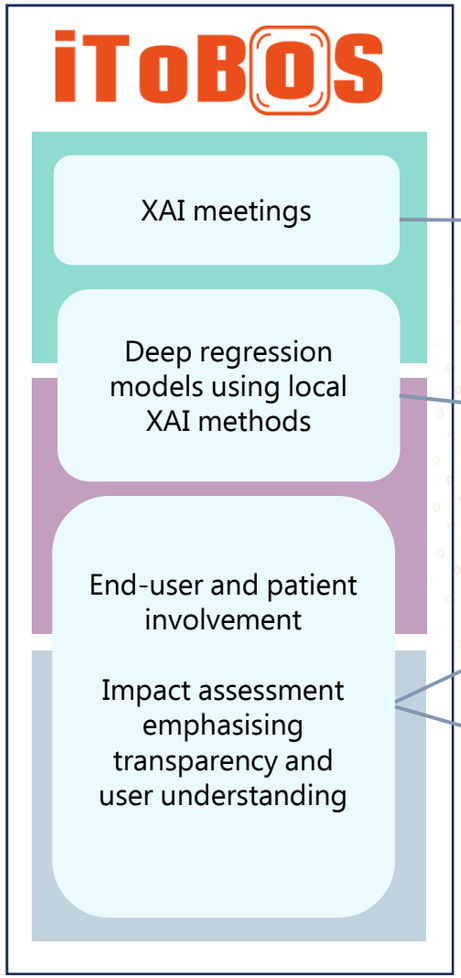
PREPARE REHAB

Aim: To advance rehabilitation care for patients with chronic non-communicable diseases, by developing personalised, data-driven, computational prediction and stratification tools to enhance decision-making in selecting optimal therapy strategies.

- ***Grant agreement number:*** *10086219*
- ***Time frame:*** *June 2023 – May 2026*

Case Studies: XAI Methodologies

TRILATERAL RESEARCH
Ethical AI



- XAI principles and methodologies integrated into project implementation.
- Discussions on approaches for enhancing transparency, such as explainability techniques or transparency-enhancing tools.

- Transparency into the decision-making process of the AI system.
- Provide explanations for model predictions.

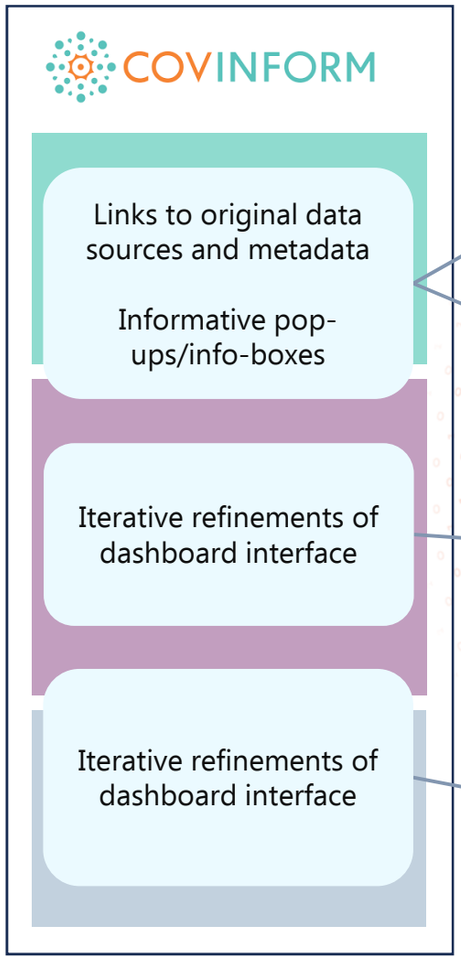
- Better understanding and alignment with user requirements.
- Validation of interpretability features and functionalities.

- Timely identification and rectification of biases or errors.
- AI system's outputs are comprehensible and relevant to end-users.
- Opportunities for stakeholders to provide feedback on the AI system's performance and usability.

Transparency | **Interpretability** | **Accountability**

Case Studies: XAI Methodologies

TRILATERAL RESEARCH
Ethical AI



- Users to verify the quality and reliability of the data.
- Facilitates interpretation of the AI model's outputs.

- Provide users with detailed explanations about the AI system's decisions, processes, and underlying algorithms.
- Offer explanations and insights into the reasoning behind the AI model's predictions or recommendations

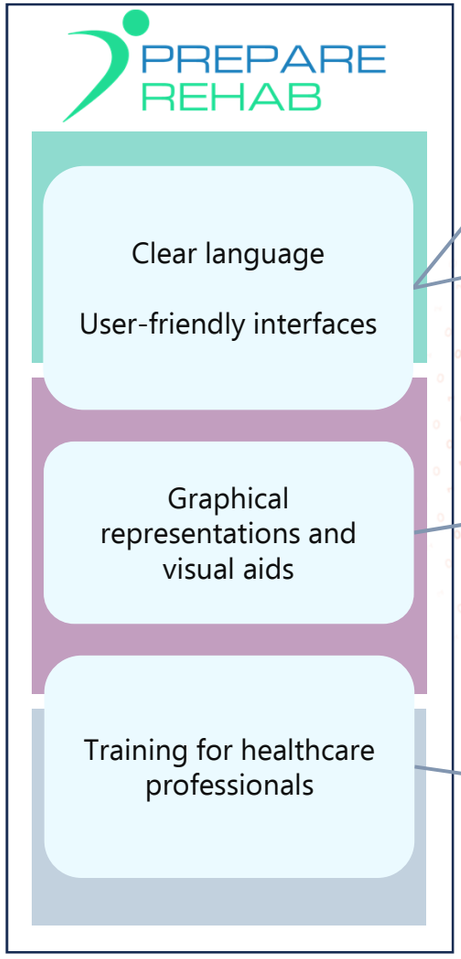
- Enhances user-friendliness of the dashboard.
- Ensures that dashboard interface meets user needs and expectations.

- Simulating real-world usage scenarios to evaluate the system's interpretability from the user's perspective.
- Decisions are made collaboratively and integrates users' feedback.

Transparency | **Interpretability** | **Accountability**

Case Studies: XAI Methodologies

TRILATERAL RESEARCH
Ethical AI



- Accessible explanations of how the AI system operates and its decision-making processes.
- Simplify complex AI concepts and processes, making them more understandable to users.

- Enables users to easily locate and access relevant information about the AI system.
- Contextual guidance to assist users in understanding how to use and interpret the AI system effectively.

- Presents information in an interactive and engaging format.
- Enable users to compare different models, datasets, or scenarios, aiding in the interpretation of results.
- Provides users with a clear understanding of how the system operates.

- Aids in interpretation of predictions, recommendations, and insights provided by the AI system
- Provides opportunities to clarify uncertainties.
- Ensures system is used in a manner consistent with professional standards.

Transparency | **Interpretability** | **Accountability**

Summary

TRILATERAL RESEARCH
Ethical AI

- Projects demonstrate a collective commitment to enhancing AI explainability
  - Varied approaches used to incorporate and demonstrate XAI.
  - Emphasis on transparency, accountability, and accessibility for end-users, including non-technical users.
  - Co-creation methodologies used to enhance overall trustworthiness and understandability by integrating diverse perspectives throughout the development lifecycle.
  - Understanding and addressing complexities of XAI in healthcare is crucial for building trust among stakeholders, improving patient care and adhering to ethical principles.

Challenges

TRILATERAL RESEARCH
Ethical AI

- How to scale XAI solutions/approaches across diverse healthcare contexts?
- How to balance the complexity of AI models with the imperative for transparency and comprehensibility – important in healthcare where trust is paramount.
- Lack of standardised approaches in XAI – no universally accepted definitions for terms such as 'explainable' or 'interpretable' in AI context, resulting in diverse approaches.
- Complications in communication within the AI community – challenges in developing cohesive frameworks for evaluating and implementing XAI methodologies.
- Ongoing discussions needed, with continuous refinement and adaptation.

Opportunities

TRILATERAL RESEARCH
Ethical AI

- Focus on defining key components of explainability, including model interpretability, and methods for communicating AI outputs to diverse stakeholders.
- Ongoing dialogue and knowledge exchange, facilitated by collaborative efforts within standardisation bodies such as ISO and CENELEC.
- Explore the broader applicability and challenges associated with scaling co-creation XAI methodologies through future research and case studies, extending beyond the scope of EU projects.
- Address varied explainability requirements for diverse stakeholders within domains such as healthcare, recognising it as a significant challenge that warrants further investigation.

Future Research Plans & Interests

TRILATERAL RESEARCH
Ethical AI

- Scalability and adaptability of XAI methodologies across diverse cultural, industrial, and regulatory contexts.

- Suitable approaches to address varied explainability requirements for diverse stakeholders within the healthcare domain.

- Technical complexities in implementing XAI models involve refining current methods and innovating new techniques to improve AI system transparency and interpretability.

References

- N. Hoppe, R. C. Härting, and A. Rahmel, "Potential Benefits of Artificial Intelligence in Healthcare," In Artificial Intelligence and Machine Learning for Healthcare, pp. 225-249, Springer, Cham, 2023.
- Proposal for a Regulation of The European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts, Art 3(1). Accessible at: https://eur-lex.europa.eu/legal- content/EN/TXT/?uri=celex:52021PC0206 [retrieved: 8 February, 2024]
- High-Level Expert Group on AI: Ethics Guidelines for Trustworthy Artificial Intelligence. Accessible at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai [retrieved: 5 February, 2024]
- J. P. Richardson, C. Smith, S. Curtis, S. Watson, X. Zhu, B. Barry, et al., "Patient apprehensions about the use of artificial intelligence in S. Ali, T. Abuhmed, S. El-Sappagh, K. Muhammad, J. M. Alonso-Moral, et al., "Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence," Information Fusion, 99, p.101805, 2023.
- Intelligent Total Body Scanner for Early Detection of Melanoma (ITOBOS). Accessible at: https://itobos.eu [retrieved: 5 February, 2024]
- COronavirus Vulnerabilities and INFOrmation dynamics Research and Modelling (COVINFORM). Accessible at: https://www.covinform.eu [retrieved: 8 February, 2024]
- PREPARE-Rehab. Accessible at: https://prepare-rehab.eu [retrieved: 8 February, 2024]
- W. J. Baumol, "The free-market innovation machine: Analyzing the growth miracle of capitalism," Princeton university press, 2002.
- ISO/IEC JTC 1/SC 42 - Artificial intelligence, https://www.iso.org/committee/6794475.html
- CEN-CENELEC JTC 21 'Artificial Intelligence', https://www.cencenelec.eu/areas-of-work/cen-cenelec-topics/artificial-intelligence

THANK YOU!

TRILATERAL RESEARCH
Ethical AI

iToBOS

PREPARE REHAB

COVINFORM

Funded by the European Union

Let's Stay Connected

https://itobos.eu/
https://www.covinform.eu
https://prepare-rehab.eu

Contact@itobos.eu
office@covinform.eu
info@prepare-rehab.eu

@itobos_eu
@COVINFORM_EU
@PREPARERehab

itobos project
COVINFORM Project
PREPARE Rehab

itobos.eu

itobos_eu

itobos_eu
Covinform_eu