XAI: Fundamentals Challenge ROUND 1

Delphi study on explainable health-AI
This study investigates two fundamental questions regarding explainability of AI in healthcare:

1. **What is an explanation for AI in healthcare?**
2. **What are the attributes for a good explanation for AI in healthcare?**

The study aims to provide a definition and a global list of characteristics of a good explanation for health-AI that can be used by AI developers and healthcare professionals.

LET'S BEGIN!

There are 18 questions in this survey.

Consent

This study has been reviewed and approved by QMUL Electronic Engineering and Computer Science Devolved School Research Ethics Committee (QMERC20.565.DSEECS23.052)

How will my data be stored and who will have access to it?

Your data will be maintained in pseudonymised form. This means your name and other identifiers in the working dataset will be replaced by a unique id. To reduce the risk of disclosure, personal identifiers will be stored separately from the research data in a separate secured university computer and will only be accessible to the primary investigator. This will be the only document that will link your unique id to your real identity.

When and how will my data be destroyed?

The data of this study will be retained until 2028. We will follow QMUL information disposal policy to ensure secure deletion of any electronic record, a product that overwrites data many times will be used, such that the information cannot be recovered. IT Services will provide guidance and advice about the use of these products. In addition, any media holding electronic data will be physically destroyed.

Can I withdraw for the study?

You can withdraw from this study at any time without providing a reason. However, any collected data relating to you up to the time of withdrawal that has already been anonymised and incorporated into that iteration of the ongoing process will be used to the study. For instance, if you decide not to participate in the second Delphi round, the data collected from the first round will still be used.

Do you consent to participate on this study?

*

Please choose **only one** of the following:

◯ Yes

◯ No

Demographics

Please specify your full name *

Please write your answer here:

Which group does best describe your job profile? *

Please choose **only one** of the following:

- ◯ End user decision maker (e.g. health professional)
- ◯ AI developer (e.g. engineer, computer science, data scientist)
- ◯ XAI theorist (e.g. phycologist, cognitive scientist, philosopher, legal theorist)
- ◯ Regulator (e.g. administrator, hospital management, policymaker)
- ◯ Other

How many years have you been working on explainable AI? *

Please choose **only one** of the following:

- ◯ <2 years
- ◯ [2 - 6) years
- ◯ [6 - 10) years
- ◯ >= 10 years

What is an explanation of AI in healthcare?

Based on a literature review we divided the definition of explainable AI (XAI) into three components; (1) the semantic entity of an explanation, (2) the aim of the explaining process and (3) the explanation purpose. For each compoment we present published definition fragments. Please rate your agreement.

How would you rate each of the following semantic descriptions for an explanation of AI?

*

Please choose the appropriate response for each item:

	Do not agree	2	3	Somewhat agree	5	6	Totally agree
“An explanation is an output (of a XAI system)”	◯	◯	◯	◯	◯	◯	◯
“An explanation is a process”	◯	◯	◯	◯	◯	◯	◯
“An explanation is a statement”	◯	◯	◯	◯	◯	◯	◯
“An explanation is a justification”	◯	◯	◯	◯	◯	◯	◯
“An explanation is an accurate expression”	◯	◯	◯	◯	◯	◯	◯
“An explanation is a suite of ML techniques”	◯	◯	◯	◯	◯	◯	◯
“An explanation is the result of an interaction between a user and an explainer"	◯	◯	◯	◯	◯	◯	◯

Comment

Please write your answer here:

What is the aim of the explaining process in AI? *

Please choose the appropriate response for each item:

	Do not agree	2	3	Somewhat agree	5	6	Totally agree
“Produce more explainable models while maintaining a high level of learning performance.”	◯	◯	◯	◯	◯	◯	◯
“Make the properties of an AI model inspectable.”	◯	◯	◯	◯	◯	◯	◯
“Produce details or reasons (given an audience).”	◯	◯	◯	◯	◯	◯	◯
“Describe one or more facts.”	◯	◯	◯	◯	◯	◯	◯
“Express the AI in a language that is understandable by the human user.”	◯	◯	◯	◯	◯	◯	◯
“Provide information about the causal history of an event.”	◯	◯	◯	◯	◯	◯	◯
“Characterize AI system’s strengths and weaknesses.”	◯	◯	◯	◯	◯	◯	◯

Comment

Please write your answer here:

What is the purpose of the explanation? *

Please choose the appropriate response for each item:

	Do not agree	2	3	Somewhat agree	5	6	Totally agree
“Enable human users to understand AI system’s reasoning.”	◯	◯	◯	◯	◯	◯	◯
“Enable human users to understand the model.”	◯	◯	◯	◯	◯	◯	◯
“Enable human users to understand the evidence for a decision output.”	◯	◯	◯	◯	◯	◯	◯
“Convey an understanding of how the AI system will behave in the future.”	◯	◯	◯	◯	◯	◯	◯
“Allow human users to trust the output created by machine learning algorithms.”	◯	◯	◯	◯	◯	◯	◯
“Allow human users to trust the AI system.”	◯	◯	◯	◯	◯	◯	◯
“Make AI-generated advice appropriable, and exploitable by the intended users.”	◯	◯	◯	◯	◯	◯	◯

Comment

Please write your answer here:

Could you please propose any additional component of an explanation definition and/or any additional explanation fragment that are not included in the above list. If none, please enter NA. *

Please write your answer here:

What are the attributes of a "good" explanation in healthcare?

If someone wanted to explain you why an AI gives or does not give a certain recommendation what elements would you like to see in the provided explanation to understand and feel confident in the AI. The below list of attributes of a “good” explanation are based on a literature review. Please rate their importance.

Please rate the importance of the below attributes related to the FOCUS of a “good explanation”. *

Please choose the appropriate response for each item:

	Not important	2	3	Somewhat important	5	6	Extremely important
Domain-oriented: should be tailored to the domain, incorporating the relevant terms of the domain.	○	○	○	○	○	○	○
Purpose-oriented: should be tailored to a specific explanation purpose.	○	○	○	○	○	○	○
User-oriented: should be tailored to the user’s needs and abilities.	○	○	○	○	○	○	○
Time-related: should be tailored to the user’s time to engage with the explanation.	○	○	○	○	○	○	○

Comment

Please write your answer here:

Please rate the importance of the below attributes related to the CONTENT of a “good explanation”. *

Please choose the appropriate response for each item:

	Not important	2	3	Somewhat important	5	6	Extremely important
Causal: should maintain causal relationships between inputs and outputs.	◯	◯	◯	◯	◯	◯	◯
Coherent: should relate to prior beliefs of the user and be overall consistent.	◯	◯	◯	◯	◯	◯	◯
Relevant: causes that refer to situations that are not too far in the past, surprising, or abnormal should be attributed higher explanatory relevance.	◯	◯	◯	◯	◯	◯	◯
Informative: should provide the necessary and sufficient information to close the user’s knowledge gap.	◯	◯	◯	◯	◯	◯	◯
Robust: should withstand small perturbations of the input that do not change the output.	◯	◯	◯	◯	◯	◯	◯
Faithful: should accurately matches the	◯	◯	◯	◯	◯	◯	◯

	Not important	2	3	Somewhat important	5	6	Extremely important
input-output mapping of the AI system.							
Comprehensible: should be understandable for users.	◯	◯	◯	◯	◯	◯	◯
Objective: should be as objective as possible to minimize the amount of subjectivity a user might have when interpreting the explanations.	◯	◯	◯	◯	◯	◯	◯
Vulnerable: should explain how certain or not it is about the prediction and therefore explanation.	◯	◯	◯	◯	◯	◯	◯
Evaluative: should present evidence to support or refute human judgements and explain trade-offs between any set of options.	◯	◯	◯	◯	◯	◯	◯
Transparent: should help the user in understanding the underlying logic of the AI system, and possibly identifying that the system is wrong.	◯	◯	◯	◯	◯	◯	◯

	Not important	2	3	Somewhat important	5	6	Extremely important
Selected: should be specific and not consist of the complete cause of an event, highlighting the most important features for a decision.	◯	◯	◯	◯	◯	◯	◯
Exhaustive: should present the complete cause of events.	◯	◯	◯	◯	◯	◯	◯

Comment

Please write your answer here:

Please rate the importance of the below attributes related to the OUTPUT of a “good explanation”. *

Please choose the appropriate response for each item:

	Not important	2	3	Somewhat important	5	6	Extremely important
Social: should be a transfer of knowledge, presented as part of a conversation or interaction.	◯	◯	◯	◯	◯	◯	◯
Human-like: should emphasising the anthropomorphic features of the AI system.	◯	◯	◯	◯	◯	◯	◯
Constructive: should explain questions in the constructive form "Why x and not y?".	◯	◯	◯	◯	◯	◯	◯
Counterfactual: should explain questions in the constructive form "What would happen if?".	◯	◯	◯	◯	◯	◯	◯

Comment

Please write your answer here:

Could you please propose any attributes of a “good” explanation of AI for use in healthcare and their definition that you consider important and are not included in the above list? If none, please enter NA. *

Please write your answer here:

More information about the project aims and results can be found https://exaidss.com/

The second round of the study will take please in January 2024. You will be notified by email when the second round opens.

We would like to acknowledge the expert group participants in this study when we come to publish the results. This ackowledgement will include listing expert group participants at the conclusion of any publication. If you do not wish your name to be included in this list please email e.kyrimi@qmul.ac.uk

01-03-2024 – 23:59
Submit your survey.
Thank you for completing this survey.