|  |  |  |
| --- | --- | --- |
| ITU Logo | INTERNATIONAL TELECOMMUNICATION UNION**TELECOMMUNICATIONSTANDARDIZATION SECTOR**STUDY PERIOD 2017-2020 | FG-AI4H-K-036-R01 |
| **ITU-T Focus Group on AI for Health** |
| **Original: English** |
| **WG(s):** | DAISAM | E-meeting, 27-29 January 2021 |
| **DOCUMENT** |
| **Source:** | WG-DAISAM |
| **Title:** | WG-DAISAM: Call for participation in survey for transparent model reporting for trustworthy Machine Learning for Health applications |
| **Purpose:** | Engagement |
| **Contact:** | Jana Fehr (WG-DAISAM)Hasso-Plattner-InstituteGermany | Email: jana.fehr@hpi.de  |

|  |  |
| --- | --- |
| **Abstract:** | Transparent model reporting has been suggested as an important checkpoint for regulatory approval for healthcare products using machine learning (ML). We aim to investigate to which extent transparent reporting has been adapted in practice and how much information can be shared without violating data protection or business secrets. The ITU/WHO Focus Group on "Artificial Intelligence for Health" invites participants from companies and academia who developed a ML-based algorithm for a healthcare application to report about their use-case in a questionnaire. The questionnaire follows current considerations for model reporting and elicits required information to assess the quality of the product. In return, we will provide a feedback report and select five use-cases for an individual consultation session about the assessed model quality and reporting practices. We will summarize the current state of model reporting practices and highlight adaptation challenges. With our findings, we aim to help 1) Product owners to adapt to regulatory requirements and 2) regulatory institutions to assess the feasibility of fulfilling the stated requirements. |



**Explain your algorithm for trustworthy ML4Health**

Did you(r team) develop a ML-based tool to support clinicians or improve human health?

Do you want to prepare for questions from users and regulatory institutions to build trust in your tool?

Then report about your use-case in our survey. We would like to find out if you can answer questions to make your tool trustworthy and help YOU to prepare for important questions from users and regulatory bodies. We assure you that your queried data will be treated with the utmost confidentiality.

To reward you for your efforts, you will receive feedback about the transparency of your use-case report. We will select 5 participating use-cases for a free consultation from the FGAI4H.

If you have any questions, please contact jana.fehr@hpi.de.

Thank you very much!

The FGAI4H-DAISAM team.

--------- Please share this call for participation also in your networks and social media ----------

# Motivation

Medicine is currently at the crossroads to leverage the power of digital technologies such as machine learning for daily clinical decision-making, early detection of disease and overall improving patients' lives. Similar to drug-approval regulations, standardized approval checkpoints are needed to guarantee the reliability, efficacy and safety of machine-learning algorithms for health (ML4H) applications. Guidelines to establish ML4H approval frameworks are currently under development. One important checkpoint in this process is the transparent reporting of model development and testing steps. Besides shedding light on the implemented technology, it is crucial to report details about data selection to assess potential risks of bias. Transparent model reporting can be hampered if the algorithm has an opaque-box character or if the algorithm is proprietary or trained with protected data. It is currently unclear if developers can fulfil transparent model reporting requirements.

# Aim

The aim of this work is to investigate the current practices of transparent model reporting for ML4H and pinpoint challenges. We will invite developers and owners of ML4H applications from corporates, governmental and academic institutions inside and beyond the FGAI4H network to conduct transparent model reporting about their use-case. To guide the reporting, we compiled a structured questionnaire from current considerations summarized by the FGAI4H [1,2,3].

We will analyze the questionnaire-based reports seeking to address the following questions:

1. Have companies/institutions already adapted to providing enough documentation about model development and testing?
2. Are there reporting gaps and if yes, which challenges do corporates/institutions face for transparent model reporting?
3. Do current practices of model reporting help to identify potential biases of the reported algorithm?
4. Can this questionnaire guide standardized model reporting or are there questions, which should be adapted specifically to the use-case?

# Approach and participation procedure

1. Filling out the questionnaire for transparent model reporting

We invite owners or developers of health products using machine learning to participate in our survey and answer questions for transparent reporting of their ML4H tool. Everyone who is interested to participate can send an e-Mail to jana.fehr@hpi.de to receive the link to the questionnaire. The survey is hosted on a secure, GDPR-compliant platform. The questionnaire consists of 75 multiple-choice questions spanning across the following domains:

1) Intended use of the model

2) Implemented ML technology

3) Training data information

4) Legal aspects

5) Ethical considerations

6) ML model evaluation and metrics

7) Caveats and recommendations

1. Analyzing the questionnaire and feedback

We will analyze the questionnaire-based report with our interdisciplinary team from the FGAI4H-DAISAM group, combining expertise from data science, computer science, biotechnology and epidemiology. We will assess the submitted reporting practice in terms of rigor, reporting gaps, to answer above questions 1) *Have companies/institutions already adapted to providing enough documentation about model development and testing?* and 2) *‘Are there reporting gaps and if yes, which challenges do corporates/institutions face for transparent model reporting?’*. We will additionally perform a qualitative potential risk of bias assessment on the provided report to answer question 3) ‘Do current practices of model reporting help to identify potential biases of the reported algorithm?’. We will provide a short feedback report on the reporting practice for regulatory approval and potential bias findings.

We will additionally select five use-cases for an individual consultation session. During the consultation session, we will discuss questions from participants and the assessment outcome in detail.

1. Follow-up questionnaire

After participants received feedback, we will send out a short follow-up questionnaire (anticipated time to answer: 5 minutes), to ask about their experiences with model reporting in general and the experiences with reporting guided by this questionnaire-based approach. We will investigate the answers of this follow-up questionnaire to answer question 4 ‘Can this questionnaire guide standardized model reporting or are there questions which should be adapted specifically to the use-case?’

# Benefit to participate

With this project, we want to support teams working with ML4H to adapt to regulatory requirements for transparent model reporting by prompting required information in a systematic way. Participants are invited to engage and give feedback on the feasibility of the considerations. We plan to publish the results about the assessed state of transparent model reporting from this survey. Participants can choose to be mentioned in acknowledgements or remain anonymous.

Please contact us at jana.fehr@hpi.de if you would like to participate or if you have any questions.

We are looking forward to hearing from you!

The FGAI4H-DAISAM Team

# References

[1] Moons, K. G. M. *et al.* Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. **Ann. Intern. Med.**162, W1–W73, doi: 10.7326/M14-0698 (2015).

*[2]* Association for Computing Machinery US Public Policy Council (USACM). Statement on algorithmic transparency and accountability. ***USACM Press releases*** *1–2, (2017).*

[3] Luo, W. *et al.* Guidelines for developing and reporting machine learning predictive models in biomedical research: A multidisciplinary view. **J. Med. Internet Res.** 18, 1–10, doi: 10.2196/jmir.5870 (2016).

[4] Liu, X. *et al.* CONSORT-AI extension. **Nat. Med.** **26**, 1364–1374, doi: 10.1038/s41591-020-1034-x (2020).

[5] Sounderajah, V. *et al.* Developing specific reporting guidelines for diagnostic accuracy studies assessing AI interventions: The STARD-AI Steering Group. *Nat. Med.* **26**, 807–808, doi: 10.1038/s41591-020-0941-1 (2020).

[6] Liu, X. *et al.* CONSORT-AI extension. **Nat. Med.** 26, 1364–1374, doi: 10.1038/s41591-020-1034-x (2020).

[7] Oala L., Balachandran P. DEL05D: Training and test data specification. **FG-AI4H-G-205-A04***, Delhi, India, November 2019.*

[8] Oala L., Johner C., Balachandran P.: Updated DEL2.2: Good practices for health applications of machine learning: Considerations for manufacturers and regulators **FG-AI4H-J-039**, *October 2020.*

[9] Oala L., et al.: DEL7.3: Data and artificial intelligence assessment methods (DAISAM) reference. **FG-AI4H-I-035**, *May 2020.*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_