|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ITU Logo | INTERNATIONAL TELECOMMUNICATION UNION  **TELECOMMUNICATION STANDARDIZATION SECTOR**  STUDY PERIOD 2017-2020 | | FG-AI4H-N-033 | |
| **ITU-T Focus Group on AI for Health** | |
| **Original: English** | |
| **WG(s):** | | Plenary | E-meeting, 15-17 February 2022 | |
| **DOCUMENT** | | | | |
| **Source:** | | Editors DEL7.3 | | |
| **Title:** | | DEL7.3 Update: ML4H trial audits–Iteration 2.0 Playbook (Version 2.0) | | |
| **Purpose:** | | Discussion | | |
| **Contact:** | | LuisOala HHI Fraunhofer, Germany | | E-mail: [luis.oala@hhi.fraunhofer.de](mailto:luis.oala@hhi.fraunhofer.de) |
| **Contact:** | | Elora-Dana Schoerverth HHI Fraunhofer, Germany | | E-mail: [elora-dana.schoerverth@hhi.fraunhofer.de](mailto:elora-dana.schoerverth@hhi.fraunhofer.de) |
| **Contact:** | | Alixandro Werneck Leite LAMFO, University of Brasilia, Brazil | | E-mail: [alixandrowerneck@outlook.com](mailto:alixandrowerneck@outlook.com) |
| **Contact:** | | Pradeep Balachandran Technical Consultant (Digital Health), India | | E-mail: [pbn.tvm@gmail.com](mailto:pbn.tvm@gmail.com) |

|  |  |
| --- | --- |
| **Abstract:** | This document is related to Deliverable 7.3 and provides a short playbook with important information for the ML4H trial audits–iteration 2.0 in terms of its scope, objectives, process flow and deliverables. |

**CONTENTS**

| Page |
| --- |
| [1 Scope and Introduction 4](#_Toc95327588)  [2 References 5](#_Toc95327589)  [3 Specific Objectives 5](#_Toc95327590)  [4 Process Flow 6](#_Toc95327591)  [5 Expected Deliverables and Outputs 8](#_Toc95327592)  [6 ML4H Use Cases (Topic Groups) 9](#_Toc95327593)  [7 Team Formation 9](#_Toc95327594)  [8 Project Plan & Schedule 11](#_Toc95327595)  [9 Project Management 12](#_Toc95327596)  [10 Useful References 12](#_Toc95327597) |

**List of Figures**

| Page |
| --- |
| [Figure 1: Project overview of iteration 2 for the ML4H trial audits 4](#_Toc95327598)  [Figure 2: Audit process flow diagram 6](#_Toc95327599)  [Figure 3: Project team allocation criteria 10](#_Toc95327600)  [Figure 4: Project plan & schedule snapshot 11](#_Toc95327601) |

| **Project Yellow Pages** | |
| --- | --- |
| **COMMUNICATION** | |
| **Slack**  <https://join.slack.com/t/aiaudit-trialaudits/shared_invite/zt-tnm259zm-AmrPSuijvCKtcpxAZatMYw>  [**E-mail contacts of colleagues**](https://drive.google.com/file/d/1cDTmwo0pJ_e-vGekYsp5hDob7pUdLnRU/view?usp=sharing) | **What to do if…**   * ...you cannot access slack?   + Check with Alixandro * ...you cannot access contact spreadsheet?   + Check with Alixandro or Pradeep |
| **GITHUB** | |
| All repos for the assessment platform and related research projects are collected here  <https://github.com/aiaudit-org> | **What to do if…**   * ...you cannot access the github organization?   + Contact Elora, she can add you |
| **ASSESSMENT PLATFORM** | |
| The assessment platform running on ITU AWS infrastructure can be accessed here  <https://health.aiaudit.org/>  Source code is available here  <https://github.com/aiaudit-org/health-aiaudit-public>  Training resources(video lectures and documents ) for assessment platform-onboarding: [Assessment Platform-Onboarding-Training Resources](https://drive.google.com/drive/folders/1VtZqqDt5YbtC9ODFdTVsJV9It3_8LKH2?usp=sharing)  For access to AWS and the backend please contact Elora. | **What to do if…**   * ...you cannot access the platform or have questions?   + Contact Elora, she can guide you to the right resource or person |
| **MEETINGS** | |
| You can find an overview of meeting times, purposes and coordinates (Zoom, gathertown, Google) in the project calendar here (preliminary list here: [ML4H Trial Audits - Iteration 2 - Meeting Candidates](https://docs.google.com/document/d/199exHUtLAFyhakhAmW1ZH9LnxBxShfEF_SbLjQ6zt4A/edit))  Plenary meeting video recordings & presentation materials (PPTs) : [Plenary Meeting Resources](https://drive.google.com/drive/folders/1EBiakhqCk9UGJwa54RWvx1kI2RXFIPfJ?usp=sharing) | **What to do if…**   * ...you have trouble accessing meeting information or meeting itself?   + Contact Alixandro or Pradeep |
| **SHARED DRIVE AND DOCUMENTS** | |
| All shared files related to the Trial Audits - Iteration 2 can be found here [ML4H Trial Audits - iteration 2](https://drive.google.com/drive/folders/1Uau83TIldIv4LahtqbXO8O9TEYXKsgNa?usp=sharing) | **What to do if…**   * ...you have access problems?   + Contact Alixandro or Pradeep, they can add you and provide access |
| **COORDINATOR CONTACTS** | |
| Elora-dana Schoerverth | elora-dana.schoerverth@hhi.fraunhofer.de |
| Alixandro Werneck | alixandrowerneck@outlook.com |
| Pradeep Balachandran | pbn.tvm@gmail.com |
| Luis Oala | luis.oala@hhi.fraunhofer.de |
| TBC |  |

# Scope and Introduction

Diagram

Description automatically generated

Figure 1: Project overview of iteration 2 for the ML4H trial audits

Machine Learning for Health (ML4H) trial 2.0 audit aims at conducting the verification and validation of the technical, clinical, regulatory and ethical requirements of ML4H tool following a structured audit process. The project overview is illustrated in Figure 1. The audit process is based on the [ML4H Paper-to-Practice (P2P) auditing framework](http://proceedings.mlr.press/v136/oala20a/oala20a.pdf) – a set of best practices of algorithm auditing and quality control along the entire ML4H life cycle proposed by the [ITU/WHO FG-AI4H](https://www.itu.int/en/ITU-T/focusgroups/ai4h/Pages/default.aspx). It is open to extension per the considerations of this group. The goals are to

1. Practice the collaboration between different domains of expertise within

FG-AI4H

1. Refine
2. existing processes (based on what works and what does not during application)
3. existing software tools (<https://health.aiaudit.org/>)

C. Develop new processes and tools that have been missing, integrate tools and

processes from the expert network

# References

1. <https://health.aiaudit.org/> technical user manual: <https://docs.google.com/document/d/1JUsmzkL26R4ozRur0HNBQUSQ9W5zOuDRQSQpZHXtv8Y/edit?usp=sharing>
2. EvalAI Technical guide: Evaluating AI systems at scale: <https://smartech.gatech.edu/handle/1853/60738> (Rishabh’s thesis)
3. First version of our assessment platform (<https://health.aiaudit.org/>) - for the trial audit work EvalAidocumentation :<https://evalai.readthedocs.io/en/latest/>
4. Output from the first trial audit iteration: <http://proceedings.mlr.press/v136/oala20a.html>
5. FG-AI4H Topic Group Topic Description Document(TDD) Template.[Document J-105](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/docs/FGAI4H-J-105.docx) on FG-AI4H server (2020). URL: https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/docs/FGAI4H-J-105.docx
6. … @all: feel free to add relevant references

# Specific Objectives

1. To facilitate the conduct of trial audit series and to provide a platform for the submission, discussion and publication of FG-AI4H audit methods and reports under the special collection titled-“ Machine Learning for Health: Algorithm Auditing & Quality Control”in the Journal of Medical Systems (JOMS)
2. To provide basic training to FG-AI4H topic groups on how to verify and validate thetechnical, clinical,regulatory and ethical requirements of their ML4H tool by following the audit workflow over the FG-AI4H Assessment Platform <https://health.aiaudit.org/>
3. To design competency mapping and skill matrix for the creation of balanced interdisciplinary audit teams
4. To refine [ML4H Paper-to-Practice (P2P) auditing framework](http://proceedings.mlr.press/v136/oala20a/oala20a.pdf) based on the experimental findings of the trial audits
5. To curate and develop ML4H audit training resources for auditors

# Process Flow

The trial audit 2.0 process flow is illustrated in Figure 2 and it comprises of the following 4 main lifecycle stages:

* 1. Team formation
  2. Methodology Design
  3. Assessment & Reporting
  4. Process Improvement

Diagram

Description automatically generated

Figure 2: Audit process flow diagram

Detailed process steps are explained below:

1. Step 1 – corresponds to team formation
2. Steps 2 to 6- correspond to methodology design
3. Steps 7 to 8- correspond to assessment & reporting
4. Step 9- corresponds to iterative process improvement
5. @all: Form teams
   1. Define the criteria for the formation of an interdisciplinary audit team with balanced representation from AI technology, clinical, regulatory and ethical domains
   2. Audit teams allocated to one or more FG-AI4H use cases (FG-AI4H topic groups)
   3. Allot audit developers to support the audit teams with the technical implementation of qualitative and quantitative test queries (Markdown configs, Python evaluation scripts, …) on [https://health.aiaudit.org](https://health.aiaudit.org/)/
   4. Allot meta-advisory teams to help the audit teams with guidance on specific questions, serve as reviewers and sanity checkers for the audit teams
6. @audit teams, @all: Decide on the audit methods and protocols based on the technical reference documents, e.g.
   1. [STARD-AI](http://dx.doi.org/10.1038/s41591-020-0941-1),
   2. [CONSORT-AI](https://www.nature.com/articles/s41591-020-1034-x)
   3. [SPIRIT-AI](https://www.nature.com/articles/s41591-020-1037-7)
   4. [FG-AI4H: Data and artificial intelligence assessment methods](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/_layouts/15/WopiFrame.aspx?sourcedoc=%7B828882B2-4941-452C-8A61-F4DDE5802C2A%7D&file=FGAI4H-K-039.docx&action=default&CT=1614782367878&OR=DocLibClassicUI)
   5. [IJMEDI checklist](https://zenodo.org/record/4835800#.YMtCmY3ivIV)
   6. [Algorithm Auditing](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3778998)
   7. etc..

| NOTE: Each audit team is free to adjust the audit methods and protocols based on their use-case as long as the process is documented. One of the points of the second iteration of trial audits is to later compare the approaches by the different audit teams and refine good practices based on the success stories and negative results. |
| --- |

1. Define the ML pipeline phases/ stages of audit life cycle
   1. Data lifecycle
   2. Model lifecycle
2. Identify and describe the sub-phases/sub-stages within the ML pipeline
   1. Data lifecycle
      1. Data collection and
      2. Data preparation
   2. Model lifecycle
      * 1. Model training,
        2. Model tuning,
        3. Model evaluation
        4. Model usage,
        5. Model monitoring,
        6. Model maintenance,
        7. Model versioning
3. Specify the main assessment criteria for each sub-phase/sub-stage within the ML pipeline
   1. Technical validation
   2. Clinical evaluation
   3. Regulatory assessment
   4. Ethical assessment
4. Note: A minimum viable set of the ML4H assessment requirements checklist derived from the FG-AI4H standardized model survey questionnaire has been maintained in the following document: [FG--AI4H Trial Audit 2.0 -Assessment Checklist Configuration Chart](https://docs.google.com/document/d/1mCSjm9aOfe1WqOkVFW6vXCN_k1zqk87OZ8qZ_jbjX3U/edit?usp=sharing). This chart helps the auditors to customize and configure the requirements checklist on the basis of use case specific considerations and priorities.
5. Specify the requirements(AI technology, clinical, regulatory and ethics) to be verified and validated for each sub-phase/sub-stage within the ML pipeline
   1. Quantitative measures and metrics pre-specifications
      1. Model Evaluation Python script pre-specifications (Reference document: [AI4H Assessment Platform - Evaluation Script Questionnaire](https://docs.google.com/document/d/1kZ5Aaxn2ARc4zSRXyYYpin8Jpl_pPlGPqX38AFVvpS0/edit))
   2. Qualitative measures and metrics
      1. Standardized survey questionnaire (Reference document J-038 on FG-AI4H server (2020). URL<https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/SitePages/Home.aspx>)
6. Specify the criteria for how to validate the evidence for compliance and how to generate compliance scores
7. Generate an audit report with the conformance/ non-conformance scores and summary
8. Specify the criteria for audit iterations based on process improvement objectives

# Expected Deliverables and Outputs

1. Each audit team generates a standardized audit report which documents
2. The audit process on which the team’s experts settled
3. The results of the audit
4. The standard reporting template can be accessed at (Reference document: J-048 on the FG-AI4H server, 2020c. URL <https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/SitePages/Home.aspx>.)
5. The standardized reporting template can be extended by the audit teams to create custom reports based on specific ML4H assessment criteria (e.g.technical spec sheet, clinical spec sheet, regulatory spec sheet, etc.) to support analysis, comparison and classification of different types of ML4H audit results.
6. Each audit team in collaboration with the audit software developers produces
7. A config file (software file) for a persistent, automatic audit of their use case on <https://health.aiaudit.org/>
8. A Python evaluation script (software file) for a persistent, automatic audit of their use case on <https://health.aiaudit.org/>
9. For more information on the workflow of config file and the python evaluation script submission, refer to
10. Section 8.2 - Create a challenge (as host) & Section 8.4- Submit the use case to a challenge (as a user)in the following document: [EvalAI - Developer Manual](https://docs.google.com/document/d/1JUsmzkL26R4ozRur0HNBQUSQ9W5zOuDRQSQpZHXtv8Y/edit#heading=h.vimnu3wqqd2h)
11. Section III- Docker-Based Submission use case workflow & Text-Based Submission use case workflow in the following document :[AI4H Assessment Platform - Evaluation Script Questionnaire](https://docs.google.com/document/d/1kZ5Aaxn2ARc4zSRXyYYpin8Jpl_pPlGPqX38AFVvpS0/edit)
12. Each audit team creates a scientific paper on the audit process and results for
13. Publication on open-access preprint server (arxiv)
14. Submission to the special collection titled” Machine Learning for Health: Algorithm Auditing & Quality Control “ in Journal of Medical Systems (JOMS) OR another venue of the audit team’s choice
15. (Optional) Each audit team contributes an educational blog post outlining some of the highlights of their auditing experience.
16. These could cover e.g., (a) auditing process governance, (b) auditing process optimization or (c) auditing process automation, among other topics
17. After all audits are completed, a meta-review and synthesis of the different audit reports will follow to update and improve
18. Audit process descriptions and considerations
19. Individual method sections for the audit process (e.g. regarding data quality, model quality, clinical endpoints)
20. Standardized reporting templates
21. (Optional, team dependent) Refine or integrate tools that help stakeholders during the audit process
22. <https://health.aiaudit.org/>
23. MIOTIFY: web-based platform that guides product teams to make compliant medical device software (<https://miotify.co.uk/>)
24. Latticeflow(<https://latticeflow.ai>)
25. Dotphoton(<https://dotphoton.com>)
26. Clearbox AI (<https://clearbox.ai> )
27. Blockly interface for the evaluation scripts
28. Other APIs that group members bring to the table

# ML4H Use Cases (Topic Groups)

* The pilot phase of the trial audit will be performed with the recruitment of the following 11 [FG-AI4H Topic Groups](https://www.itu.int/en/ITU-T/focusgroups/ai4h/Pages/tg.aspx) and one TG- Radiology subgroup

1. TG- Ophthalmology

2. TG- Malaria

3. TG- Symptoms

4. TG- Psychiatry

5. TG- Snakes

6. TG-Radiology

7. TG-Neuro

8. TG-Outbreaks

9. TG- Dental

10. AI-Venger’s Collab ( A subgroup of TG-Radiology)

11. TG-Falls

# Team Formation

Figure 3 illustrates the criteria for mapping the different representational groups for the project team formation.

Graphical user interface, diagram

Description automatically generated

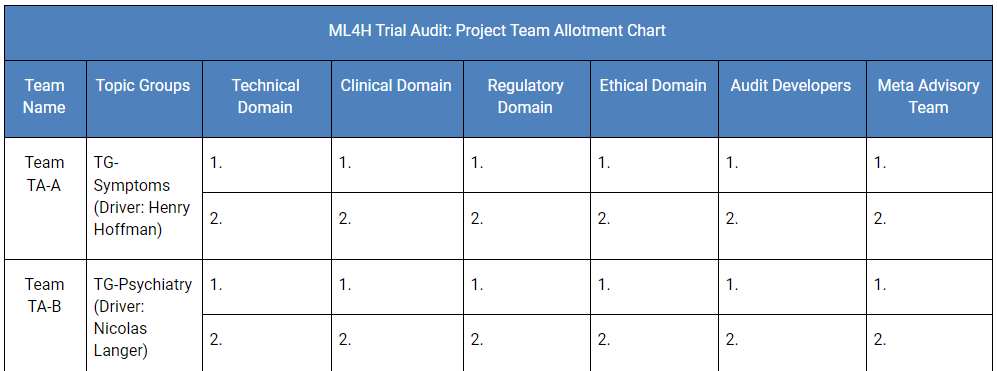
Figure 3: Project team allocation criteria

* A team of auditors will be allotted to each TG / Use-case .The expected constitution of each audit team is 8 members ; 2 each from AI technology, clinical, regulatory and ethical domains) . In case of resource shortage, we shall conform to the minimalrepresentation criterion i.e., 4 audit members per team ; 1 each from AI technology, clinical, regulatory and ethical domains.
* A group of audit developers will be formed to support the audit teams with the technical implementation of qualitative and quantitative test queries (Markdown configurations, Python evaluation scripts, …)
* A meta-advisory group will be formed to help the audit teams with guidance on specific questions, serve as reviewers and sanity checkers for the audit teams
* A preliminary version of the team member registry is maintained in the following document: [FG-AI4H Trial Audit 2.0Project Team Roster](https://docs.google.com/document/d/1mieKvPqKwFE8OJGMFcPTSIahAyH-SsedQaUQd4ubWdY/edit?usp=sharing).

Table

Description automatically generated

* The pilot project team allocation chart template is maintained in the following document: [FG-AI4H Trial Audit 2.0 –Project Team Allocation Chart](https://docs.google.com/document/d/1VlbgE1eIqyLpd_UzlF_x1yxYsT-qWbclUbUB1OCnpn8/edit?usp=sharing). Following this template, a preliminary[Team Allocation Matrix](https://docs.google.com/spreadsheets/d/17gDoEVA8qe_SBPyMlddl0dVzyTc29Y5O/edit?usp=sharing&ouid=108408214190080082443&rtpof=true&sd=true)has been prepared.



# Project Plan & Schedule

The pilot project plan & schedule is provided in Table 1: Trial Audit 2.0: Pilot Project Plan & Schedule in the following document: [FG-AI4H Trial Audit 2.0 -Pilot Project Plan & Schedule](https://docs.google.com/document/d/1mhZy4SUj_DZCg7IzuG_2t_FsX37ldWT6q8pXTFsMWdM/edit?usp=sharing). The project plan outlines the different project tasks, corresponding deliverables and their tentative timelines. A snapshot of the project plan & schedule table is shown in Figure 4.

Table

Description automatically generated

Figure 4: Project plan & schedule snapshot

# Project Management

The project management activities of trial audit 2.0 are documented under the following document: [ML4H Trial Audit 2.0 - Project Management Plan](https://docs.google.com/document/d/1KW1nCVYn3zUKF5QCx4EicsQo6JHLfSM3layqFwSrDZo/edit?usp=sharing). This document includes Work Breakdown Structure and Activities List and Communication Management Plan.

# Useful References

1. <https://health.aiaudit.org/> technical user manual: <https://docs.google.com/document/d/1JUsmzkL26R4ozRur0HNBQUSQ9W5zOuDRQSQpZHXtv8Y/edit?usp=sharing>
2. EvalAI Technical guide: Evaluating AI systems at scale: <https://smartech.gatech.edu/handle/1853/60738> (Rishabh’s thesis)
3. First version of our assessment platform (<https://health.aiaudit.org/>) - for the trial audit work EvalAi documentation : <https://evalai.readthedocs.io/en/latest/>
4. Output from the first trial audit iteration: <http://proceedings.mlr.press/v136/oala20a.html>
5. FG-AI4H Topic Group Topic Description Document(TDD) Template.[Document J-105](https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/docs/FGAI4H-J-105.docx) on FG-AI4H server (2020). URL: <https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/docs/FGAI4H-J-105.docx>
6. Plenary meeting video recordings & presentation materials (PPTs) : [Plenary Meeting Resources](https://drive.google.com/drive/folders/1EBiakhqCk9UGJwa54RWvx1kI2RXFIPfJ?usp=sharing)
7. Training resources(video lectures and documents ) for assessment platform(<https://health.aiaudit.org/>)-onboarding: [Assessment Platform-Onboarding-Training Resources](https://drive.google.com/drive/folders/1VtZqqDt5YbtC9ODFdTVsJV9It3_8LKH2?usp=sharing)
8. @all: feel free to add relevant references

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