|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ITU Logo | INTERNATIONAL TELECOMMUNICATION UNION  **TELECOMMUNICATION STANDARDIZATION SECTOR**  STUDY PERIOD 2017-2020 | | FG-AI4H-G-042 | |
| **ITU-T Focus Group on AI for Health** | |
| **Original: English** | |
| **WG(s):** | | Plenary | New Delhi, 11-15 November 2019 | |
| **DOCUMENT** | | | | |
| **Source:** | | TG-Dental topic driver | | |
| **Title:** | | Initial TDD: Dental diagnostics and digital dentistry | | |
| **Purpose:** | | Discussion | | |
| **Contact:** | | Falk Schwendicke Charité – Universitätsmedizin Berlin Germany | | Tel: +4930450662556 Fax: +40304507562556 Email: falk.schwendicke@charite.de |
| **Contact:** | | Joachim Krois Charité – Universitätsmedizin Berlin Germany | | Tel: +4930450562328 Fax: +40304507562932 Email: Joachim.krois@charite.de |

|  |  |
| --- | --- |
| **Abstract:** | This document serves as initial “skeleton” document (“DEL10.17”) for the forthcoming topic description document (TDD) of the topic group *TG-Dental*, which is concerned with the standardized benchmarking of AI for dental diagnostics and digital dentistry. The outline will follow the template structure defined in FGAI4H-C-105. This TDD draft will be created in a joint effort by the topic group and continuously improved over the upcoming meetings until it is finally approved by the focus group. The topic group focusses on diagnostics in dentistry, dental and oral medicine, and associated disciplines. AI will help to improve diagnostic accuracy and pave the way to a more personalized, precise, preventive and participatory dentistry for more people worldwide. It has the potential to aid in overcoming current ineffective, expensive care models. Benchmarking is expected to yield more robust models and algorithms and to allow transparent comparisons of different models and algorithms. |

*The outline follows the template structure defined in FGAI4H-C-105:*

Table of Contents

[1 Introduction 3](#_Toc26299861)

[1.1 Document Structure 3](#_Toc26299862)

[1.2 Topic Description 3](#_Toc26299863)

[1.3 Ethical Considerations 3](#_Toc26299864)

[1.4 Existing AI Solutions 3](#_Toc26299865)

[1.5 Existing work on benchmarking 3](#_Toc26299866)

[2 AI4H Topic group 3](#_Toc26299867)

[3 Method 4](#_Toc26299868)

[3.1 AI Input Data Structure 4](#_Toc26299869)

[3.2 AI Output Data Structure 4](#_Toc26299870)

[3.3 Test Data Labels 4](#_Toc26299871)

[3.4 Scores & Metrics 4](#_Toc26299872)

[3.5 Undisclosed Test Data Set Collection 4](#_Toc26299873)

[3.6 Benchmarking Methodology and Architecture 5](#_Toc26299874)

[3.7 Reporting Methodology 5](#_Toc26299875)

[4 Results 5](#_Toc26299876)

[5 Discussion 5](#_Toc26299877)

[6 Declaration of Conflict of Interest 5](#_Toc26299878)

# Introduction

## Document Structure

● overview of the whole document

## Topic Description

● description of topic

● categorization of the topic according to categorization guideline (currently C-0xx)

● relevance of the health topic

● gold standard of current health topic handling

● possible impact of AI in this topic

● expected impact of the benchmarking

## Ethical Considerations

● ethical considerations on usage of AI

● ethical consideration of and benchmarking including its data acquisition

## Existing AI Solutions

● current systems available with their inputs, output, focus/bias

● existing benchmarking including self-stated performance

## Existing work on benchmarking

● papers on existing attempts to benchmark solutions on the topic

● clinical evaluation attempts, RCT, etc.

● including existing numbers

# AI4H Topic group

● Topic group structure

○ Subtopic 1

○ Subtopic 2

● Topic group participation

● Tools/process of TG cooperation

● TG interaction with WG, FG

● Current topic group and topic status

● Contributors so far

● Next meetings

● Next steps for the work on this document

# Method

● Overview of the benchmarking

## AI Input Data Structure

● possible inputs for benchmarking

● ontologies, terminologies

● data format

## AI Output Data Structure

● outputs to benchmark

● ontologies, terminologies

● data format

## Test Data Labels

● label types

● ontologies, terminologies

● data format

## Scores & Metrics

● which metrics & scores to use for benchmarking

● considering relation to parameters stakeholders need for decision making

● considering scores that providers use

● considering the scope providers designed their solutions for

● considering the state of the art in RCT, statistics, AI benchmarking etc.

● considering bias transparency

## Undisclosed Test Data Set Collection

● raw data acquisition / acceptance

● test data source(s): availability, reliability,

● labelling process / acceptance

● bias documentation process

● quality control mechanisms

● discussion of the necessary size of the test data set for relevant benchmarking results

● specific data governance derived by general data governance document (currently C-004)

## Benchmarking Methodology and Architecture

● technical architecture

● hosting (IIC, etc.)

● possibility of an online benchmarking on a public test dataset

● protocol for performing the benchmarking (who does what when etc.)

● AI submission procedure including contracts, rights, IP etc. considerations

## Reporting Methodology

● Report publication in papers or as part of ITU documents

● Online reporting

● public leaderboards vs. private leaderboards

● Credit-Check like on approved sharing with selected stakeholders

● Report structure including an example

● Frequency of benchmarking

# Results

● insert here the reports of the different benchmarking runs

# Discussion

● Discussion of the insights from executing the benchmarking on

○ external feedback on the whole topic and its benchmarking

○ technical architecture

○ data acquisition

○ benchmarking process

○ benchmarking results

○ field implementation success stories

# Declaration of Conflict of Interest

● by each contributor to this document

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