

FGAI4H-S-050-A14

Geneva, 3-5 July 2023

Source: TG-Symptom Assessment

Title: Att.14 - Presentation - Topic group "Symptom Assessment"

Contact: Henry Hoffmann
Martin Cansdale

E-mail: henry@ada.com

E-mail: martin@livehealthily.com

Abstract: This PPT contains a presentation on Topic group "Symptom Assessment" given during AI for Health Workshop on 5 July 2023.

Topic group: Symptom Assessment

AI for Health Workshop (5 July 2023, Geneva)



AI for Health

ITU-WHO Focus Group



World Health
Organization

itu.int/go/fgai4h



AI-based Symptom Assessment

How could “Symptom Checkers” help?

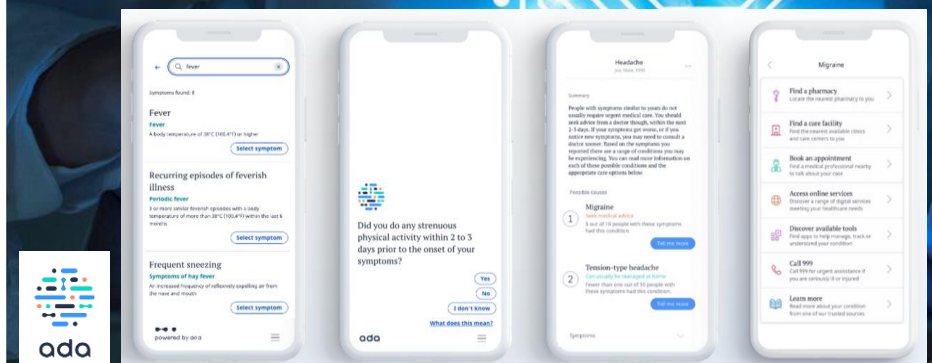
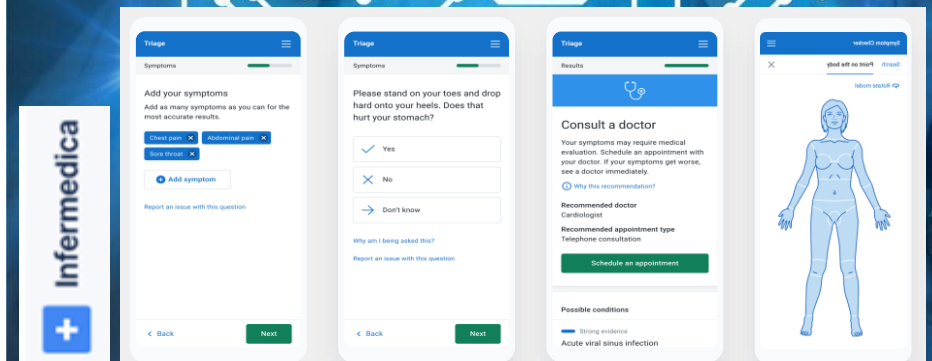
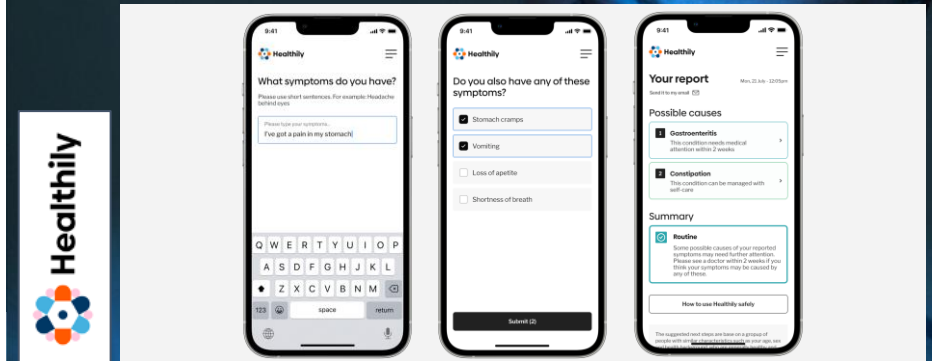
Applications – sometimes called “Symptom Checkers” that

- Allow users to describe symptoms they are worried about
- In a dialog follow up with dynamic questions to clarify symptom details and collect further evidence

To give advice and guidance on

- General next steps (Pre-clinical triage)
- Possible underlying causes
- Explanations + background information
- Navigating to the right care

But - How do we know if they really work?



Work results #1: TG-Symptom

Gathering a group around symptom assessment

Setting up the topic group

- Submitted response to call for use cases
- Got approval during New York meeting
- Reachout & growing the group
- 22 companies & 9 individual contributors
- plus the 4 TG-Symptom audit group experts

Mode of work

- Inclusive online collaboration
- Weekly coordination meeting
- Regular syncs for developer and medical doctors
- Onsite & hybrid workshops



Work results #2: Benchmarking

Building benchmarking platform(s)

TG-Symptom MMVB 2.2 Benchmarking Platform

- We built a custom benchmarking platform
- Case set and AI management
- Benchmarking setup and runs
- Scores & metrics
- Dynamic leader boards

Iterative approach

- Started with 11 abdominal pain condition toy-Model
- Toy-AIs tailored to this model
- Stepwise increase of complexity

Pivot to the Focus Group's Open Code Initiative (OCI)

- Migration to OCI platform audit script
- Focus on TG-Symptom specifics
- Cooperation with the Audit group on the formal process



AI implementation	Cases with AI result	Correct conditions (top 1)	Correct conditions (top 3)	Correct conditions (top 10)	Soft triage similarity	Triage match	Triage similarity
Uniform Random Conditions Server	100.00%	10.00%	17.00%	23.00%	51.90%	27.00%	46.50%
Ada Berlin 1k/200k Sampling Toy AI V1.1	100.00%	69.00%	98.00%	100.00%	91.50%	83.00%	91.50%
Your MD Berlin Model toy AI	100.00%	64.00%	91.00%	91.00%	89.50%	79.00%	89.50%
Infermedica Toy 1	100.00%	68.00%	96.00%	100.00%	93.00%	86.00%	93.00%
Weighted Random Conditions Server	100.00%	3.00%	19.00%	26.00%	52.70%	27.00%	46.50%

5 AI implementations, some 100, 7 metrics

[View Raw Evaluation Response](#)

[Back to overview](#)

Participant team	M1 (T)	M3 (T)	M10 (T)	Triage accuracy (T)	Triage similarity (T)
Host_33953_Team (Ada Sampler)	0.77	0.89	0.96	0.88	0.94
Host_33953_Team (Healthily Toy AI)	0.74	0.87	0.92	0.74	0.86
Host_33953_Team (Random Sampler)	0.04	0.08	0.11	0.18	0.42

Toy AI Name	M1	M3	M10	Triage accuracy	Triage similarity	Soft triage distance audit local
Ada Berlin 1k/200k Sampling Toy AI V1.1	0.77	0.89	0.96	0.88	0.94	-
Infermedica Toy AI 2	0.74	0.89	0.97	0.85	0.93	0.925
Healthily Toy AI	0.74	0.87	0.92	0.74	0.86	0.864
Random Sampler	0.04	0.08	0.11	0.18	0.42	0.485
Ada Berlin 1k/200k Sampling Toy AI V1.2	0.79	0.91	0.98	0.89	0.945	0.945
Ada Berlin 1k/200k Sampling Toy AI V1.2 FHIR	0.56	0.76	0.85	0.85	0.925	0.925

Work results #3: Annotation tool

Building a case creation and annotation tool

It's all about case-encoding

- All companies have benchmarking systems
- But they can't understand each other's cases
- Main task: encoding cases in a way that all AIs can understand

MMVB 3.1 Annotation Tool

- Web-application for encoding test cases
- Use SNOMED for symptoms
- FHIR as case container
- Created and tested together with our doctors

This screenshot shows the 'Edit Case' interface. At the top, it says 'Snomed Case Creation V2.1'. The form includes fields for Title, Author, Specialized System, Priority, Age, and Sex. Below these are sections for 'Presenting Complaint', 'Present Symptoms', and 'Absent Symptoms', each with a table of Snomed IDs, Names, Comments, Pre-Attributes, Post-Attributes, and Actions. A right-hand sidebar contains a 'Comment' field and a 'Description' field with a text area.This screenshot shows the 'Snomed Concept Browser' interface. It displays a list of concepts under the heading 'abdominal'. The list includes columns for Snomed ID, Name, and a 'Snippet' column. A search bar is at the top, and there are buttons for 'SET AS PRESENTING COMPLAINT', 'ADD AS PRESENT', and 'ADD AS ABSENT'. The right-hand sidebar shows 'Symptom Details' and 'Ancestors' for the selected concept.

Work results #4: Case corpus

Building case-sets and annotation guidelines

TG doctors curated datasets for testing

- Annotation tool
- Annotation guidelines
- Toy AIs
- FHIR encoding

We implemented a case sampler

- Based on a simple abdominal pain model
- For testing AIs (especially robustness)
- For testing metrics




Work results #5: Specification

ITU is about documents changing the world

Topic Description Document (TDD)

- Outlining the relevant aspects benchmarking: inputs, outputs, previous work, metrics, benchmarking, ethics, regulation
- Following the standard FG template (which we contributed to)
- With 202 pages the most complete discussions of symptom assessment benchmarking we know of
- Good foundation for next stage

 INTERNATIONAL TELECOMMUNICATION UNION TELECOMMUNICATION STANDARDIZATION SECTOR STUDY PERIOD 2022-2024		FG-AI4H-DEL10.14 ITU-T Focus Group on AI for Health Original: English Geneva, 3-5 July 2023
WG(s):	Plenary	
DOCUMENT		
Source:	TG-Symptom Topic Driver	
Title:	Symptom assessment (TG-Symptom)	
Contact:	Henry Hoffmann TG-Symptom Ada Health GmbH Germany	Tel: +49 177 6612889 Email: henry.hoffmann@ada.com
Contact:	Martin Cansdale TG-Symptom Healthily UK	Email: martin@livehealthily.com
Abstract:	This topic description document (TDD) specifies a standardized benchmarking for AI-based symptom assessment. It covers all scientific, technical and administrative aspects relevant for setting up this benchmarking (and follows the template structure defined in document FGA14H-J-105). The creation of this TDD is an ongoing iterative process until it is approved by the Focus Group on AI for Health (FG-AI4H) as deliverable No. DEL10.14. This draft will be a continuous input- and output document.	
Change notes:	Version 16.0 Draft – for DEL 10.14 submission <ul style="list-style-type: none">• Revision of the entire document for the DEL 10.14 submission Version 15.0 (submitted as FGA14H-R-021-A01 for Meeting R in Cambridge) <ul style="list-style-type: none">• Added 2.2.15 Status Update for Meeting R (Cambridge) Submission• Update Annex B: Declarations of conflict of interest for Ada Health, Healthily and Infermedica. Version 14.0 (submitted as FGA14H-Q-021-A01 for Meeting Q in Douala) <ul style="list-style-type: none">• Added 2.2.14 Status Update for Meeting Q (Douala) Submission	

Learnings:

Part 1

Artificial intelligence is *not* only Machine learning

- Symptom assessment AIs are often not based on ML

Symptom Checkers *are* medical devices

- Many audit requirements are already met
- Potential to streamline process for medical devices

Case encoding was the actual challenge

- Semantic case encoding is complex and leads to mapping friction



Learnings:

Part 2

LLM will change symptom assessment benchmarking

- They can radically simplify it
- By allowing to skip ontology mapping
- Allowing to test dialogs - not only full cases

LLM will change symptom assessment AI

- Change symptom assessment landscape
- We expect to see new LLM based symptom assessments soon
- Benchmarking all of them will be more important than ever before



Outlook & Next Steps

The future is about the future

One more learning:

- The industry knows how to benchmark symptom assessment
- We can specify and prepare it
- But we must **not** run it

We always needed a neutral organization that

- **Organizes** the data collection
- **Executes** the benchmarking

WHO AI4H Global Initiative

- Might be the answer we have been waiting for
- TG-Symptom aims to be a part of it



Thank you!

WHO/ITU FG AI4H TG Symptom Assessment
AI for Health Workshop
Geneva, 5 July 2023

Topic Drivers:

- Henry Hoffmann / Ada Health / Germany
henry@ada.com
- Martin Cansdale / Healthily / UK
martin@livehealthily.com



ada



Healthily



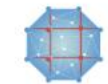
Inspired Ideas
the future, now



babylon



Infermedica



DEEPCARE



MyDoctor



mfine



xund



barkibu



kahun