ITUPublications

International Telecommunication Union
Standardization Sector

ITU-T Focus Group Deliverable

(09/2023)

Focus Group on Artificial Intelligence for Health (FG-AI4H)

FG-AI4H DEL10.23

Topic description document for the topic group on AI for traditional medicine (TG-TM)

ITU-T FG-AI4H Deliverable

DEL10.23 – Topic description document for the topic group on AI for traditional medicine (TG-TM)

Summary

This topic description document (TDD) specifies a standardized benchmarking for AI-based traditional medicine. It covers all scientific, technical, and administrative aspects relevant for setting up this benchmarking.

Keywords

Artificial intelligence, clinical relevance, data quality, ethics, data audit, health, overview, regulations, topic description, topic groups, traditional medicine.

Note

This is an informative ITU-T publication. Mandatory provisions, such as those found in ITU-T Recommendations, are outside the scope of this publication. This publication should only be referenced bibliographically in ITU-T Recommendations.

Change Log

This document contains Version 1 of the Deliverable DEL10.23 on *FG-AI4H Topic Description Document for the Topic Group on AI for Traditional Medicine (TG-TM)* approved on 15 September 2023 via the online approval process for the ITU-T Focus Group on AI for Health (FG-AI4H).

Editor:	Saketh Ram THRIGULLA Ministry of Ayush INDIA	Tel: +91 98 4999 5719 E-mail: dr.saketram@gmail.com

Table of Contents

ITU-T FG-AI4H Deliverable

DEL10.23 – Topic description document for the topic group on AI for traditional medicine (TG-TM)

1 Introduction

This topic description document specifies the standardized benchmarking for traditional medicine systems. It serves as deliverable No. [DEL.10.23] of the ITU/WHO Focus Group on AI for Health (FG-AI4H).

2 About the FG-AI4H topic group on traditional medicine

The World Health Organization (WHO) defines traditional medicine (TM) as "the sum total of the knowledge, skills and practices based on the theories, beliefs and experiences indigenous to different cultures, whether explicable or not, used in the maintenance of health, as well as in the prevention, diagnosis, improvement or treatment of physical and mental illnesses[1]".

According to a WHO research study, 65% of people in developing nations utilise traditional medicine (TM) for primary healthcare, while more than 80% of people in developing countries rely on TM.[2] TM is widely used in primary healthcare in many underdeveloped nations. Cancer, diabetes, hypertension, and other non-communicable diseases have all seen a recent increase in the use of TM as alternative medical therapy[3]. Based on current information, 88% of Member States have acknowledged their use of traditional and complementary medicine (T&CM) which corresponds to 170 Member States. These are the countries that have, for example, formally developed policies, laws, regulations, programmes and offices for T&CM, and the actual number of countries using T&CM is likely to be even higher.[4]

As a strategic priority, the General Programme of Work (GPW13) sets an overarching goal of reaching 3 billion more people, to move towards Sustainable Development Goal 3 (SDG 3) – ensuring healthy lives and promoting well-being for all at all ages by achieving universal health coverage (UHC), addressing health emergencies and promoting healthier populations. T&CM can make a significant contribution to the goal of UHC by being included in the provision of essential health services.[4] Additionally, expanding access to T&CM services that are safe, effective, and of high quality may help to address community needs and develop primary healthcare systems that are long-lasting and culturally appropriate. The Declaration of Astana, which was adopted at the Global Conference on Primary Health Care in October 2018, made it clear that expanding access to a variety of healthcare services, including traditional medicines, will be crucial to the success of primary health care. [4]

---

1 WHO Organization, WHO Traditional Medicine Strategy 2014-2023, WHO, Geneva Switzerland, 2013 Available from: http://www.who.int/medicines/publications/traditional/trm_strategy14_23/en/

2 A. Burton, M. Smith, T. Falkenberg, Building WHO's global strategy for traditional medicine, Eur. J. Integrat. Med. 7 (1) (2015) 13–15. https://www.sciencedirect.com/science/article/pii/S1876382015000037.

3 Goli Arji, Reza Safdari, Hossein Rezaeizadeh, Alireza Abbassian, Mehrshad Mokhtaran, Mohammad Hossein Ayati, A systematic literature review and classification of knowledge discovery in traditional medicine, Computer Methods and Programs in Biomedicine,Volume 168,2019,Pages 39-57,ISSN 0169-2607, https://doi.org/10.1016/j.cmpb.2018.10.017

4 World Health Organization. (2019). WHO global report on traditional and complementary medicine 2019. World Health Organization. https://apps.who.int/iris/handle/10665/312342. License: CC BY-NC-SA 3.0 IGO

Digital technologies such as machine learning and artificial intelligence (AI), an AI software as a medical device (AI-SaMD)[5] are transforming medicine, research, and public health. While this fast emerging sector has immense promise, it creates ethical, legal, and social challenges, such as equitable access, privacy, proper applications and users, liability, prejudice, and inclusivity. These are translational challenges since capturing, sharing, and exploiting data created or used by these technologies crosses national borders. "Big Data" and AI tools, methodologies, and technology are being utilised to improve health services and systems. However, many problems about the ethical development and use of these technologies remain unaddressed, especially how low and middle-income nations would benefit from AI breakthroughs. However, there are presently no widely accepted benchmarking frameworks for evaluating the data produced by the use of these devices. Additionally, the international regulatory frameworks for digital health goods [5] are changing. The goal of "Strengthening governance for digital health at global, regional, and national levels" is to strengthen the assessment and monitoring of research on the use of digital health instruments. The development of evidence to illustrate the health outcomes and impacts of digital health solutions is critical to ensuring safe deployment, establishing and promoting accountability, and justifying financial expenditures. It also addresses the need to encourage the creation and testing of technologies, methodologies for comparing innovations, and infrastructure to eliminate barriers to the prioritising of such technologies for global health. This goal will be achieved initially through the publication of frameworks for the production and generation of evidence, benchmarking, regulating, and adopting of digital health solutions such as artificial intelligence and data-driven technology. [5]

To develop this benchmarking framework, FG-AI4H decided to create the TG-AI for traditional medicine at the meeting FG-AI4H-P-049 in Helsinki, 20-22 September 2022.

FG-AI4H assigns a *topic driver* to each topic group (similar to a moderator) who coordinates the collaboration of all topic group members on the TDD. During FG-AI4H meeting FG-AI4H-P-049 Helsinki, 20-22 September 2022, Saketh Ram THRIGULLA from the Ministry of Ayush, INDIA was nominated as topic driver for the TG-AI for traditional medicine.

2.1 Documentation

This document is the topic description document (TDD) for the TG-traditional medicine. It introduces the health topic including the AI task, outlines its relevance and the potential impact that the benchmarking will have on the health system and patient outcome, and provides an overview of the existing AI solutions for traditional medicine. It describes the existing approaches for assessing the quality of traditional medicine systems and provides the details that are likely relevant for setting up a new standardized benchmarking. It specifies the actual benchmarking methods for all subtopics at a level of detail that includes technological and operational implementation. There are individual subsections for all versions of the benchmarking. Finally, it summarizes the results of the topic group's benchmarking initiative and benchmarking runs. In addition, the TDD addresses ethical and regulatory aspects.

The TDD will be developed cooperatively by all members of the topic group over time and updated TDD iterations are expected to be presented at each FG-AI4H meeting.

The final version of this TDD will be released as deliverable "[DEL.10.23] TG-AI for traditional medicine." The topic group is expected to submit input documents reflecting updates to the work on this deliverable (Table 1) to each FG-AI4H meeting.

---

5 Generating evidence for artificial intelligence-based medical devices: a framework for training, validation and evaluation. Geneva: World Health Organization; 2021. Licence: CC BY-NC-SA 3.0 IGO. https://www.who.int/publications/i/item/9789240038462

Table 1 – Topic Group output documents

Number	Title
FGAI4H-S-028-A01	Latest update of the Topic Description Document of the TG-AI for Traditional Medicine
FGAI4H-S-028-A02	Latest update of the Call for Topic Group Participation (CfTGP)
FGAI4H-S-028-A03	The presentation summarizing the latest update of the Topic Description Document of the TG-AI for Traditional Medicine

2.2 Status of this topic group

The following subclauses describe the update of the collaboration within the TG-AI for traditional medicine for the official focus group meetings.

2.2.1 Status update for meeting S

- After the announcement of this topic group and subsequent discussions the following works will be attempted and updated in the next focus group meeting. These are as follows:
  - Work on the benchmarking software
  - Progress with data acquisition, annotation, etc.
  - Overview of the online meetings including links to meeting minutes.

The following internal meetings have been successfully organized so far:

Sl.No.	Date	Meeting	Mode
1.	13, 14 December 2022	Scope and Application of artificial intelligence in Traditional Medicine, India International Centre, New Delhi	Hybrid (Online/Offline)
2.	2 February 2023	AI4Health Topic Group (TG)-Traditional Medicine (TG-TM)	Online
3.	10 March 2023	AI4Health Topic Group (TG)-Traditional Medicine (TG-TM)	Online
4	05 May 2023	AI4Health Topic Group (TG)-Traditional Medicine (TG-TM)	Online
5	10 May 2023	AI4Health Topic Group (TG)-Traditional Medicine (TG-TM)	Online
6,7	22 May 2023 (two meetings)	AI4Health Topic Group (TG)-Traditional Medicine (TG-TM)	Online
8,9	23 May 2023 (two meetings)	AI4Health Topic Group (TG)-Traditional Medicine (TG-TM)	Online
10,11	25 May 2023 (two meetings)	AI4Health Topic Group (TG)-Traditional Medicine (TG-TM)	Online

  - Relevant insights from interactions with other working groups or topic groups
  - Partners joining the topic group
  - List of current partners

Sl.No	Name	Affiliation	Expertise	Country	Email Id
1	Dr Kenji WATANABE	Center for Kampo Medicine Keio University School of Medicine 35 Shinanomachi, Shinjuku-ku	TM Expert	Japan	watanabekenji@keio.jp

Sl.No	Name	Affiliation	Expertise	Country	Email Id
		Tokyo 160-8582, Japan toyokeio@sc.itc.keio.ac.jp Phone: +81-3-5366-3824			
2	Antonio Morandi	Ayurveda Point, Milan	Ayu Expert	Italy	dr.morandi@ayurvedicpoint.it
3	Roberto Scavelli	Ayurveda Point, Milan	CS Expert	Italy	roberto.scavelli@libero.it
4	Jeffrey D. White, MD,	National Cancer Institute	Cancer, Public Health	USA	jeffreyw@mail.nih.gov
5	Bala Pesala	ayur ai	AI, Ayu	India	balapesala@gmail.com
6	Ajit Kolatkar	Gastro labs, India	Ayu, AI	India	dr.ajit@gastrolab.com
7	Cang Shik YIN	Department of Acupuncture, Kyung Hee University, Seoul, 130-701 South Korea.	Korean Medicine	Republic of Korea	acuyin@khu.ac.kr
8	Professor Donghyun NAM	Expert TM and AI, College of Korean Medicine, Sangji University, Republic of Korea.	Korean Medicine and AI	Republic of Korea	omdnam@sangji.ac.kr, omdnam@naver.com
9	DR Pawan Godatwar	WHO, SEARO, New DELHI	TM Expert, Policy	–	pgodatwar@gmail.com
10	DR Kim Sung Chol	WHO, SEARO, New DELHI	TM Expert, Policy	–	kims@who.int
11	Dr LIU Qin	Technical Officer, World Health Organization, Integrated Health Services, Geneva, Switzerland	TM Expert, Policy	–	LIU, Qin <liuq@who.int>
12	Dr Pradeep Dua	Technical Officer, World Health Organization, Integrated Health Services, Geneva, Switzerland	TM Expert, Policy	–	duadrpradeep@gmail.com
13	Sivarama Prasad Vinjamury	Professor, Research, Southern California Univ. of Health Sciences (SCU)	TM Expert, Academician	USA	vsrprasad18@gmail.com
14	Prof Rajani Kanth Vadigepalli	Vice Chair of Research and Professor of Pathology, Anatomy and Cell Biology, Thomas Jefferson	Cell Biology, Bioinformatics	USA	
15	Kenny Kutcha	Grogingent Univ. Germany (Kampo) Forschungsstelle für Fernöstliche Medizin (漢方医学研究拠点) in the Department of Vegetation	Kampo medicine	Germany	kenny.kuchta@med.uni-goettingen.de
16	Dr Ankita Srivastava	Bureau of Indian Standards	Standardization	India	ankitasri@bis.gov.in
17	Dr Bhavana Parashar	IGIB, New Delhi	Ayu Expert and AI	India	bhavana.p@igib.res.in
18	Rama Jayasundar	AIIMS, New Delhi	Nuclear Physics, Ayu,	India	ramajayasundar@gmail.com
19	V Rangamannar	Bengaluru	CS Expert	India	rangamannarv@gmail.com
20	Dr V Ramakrishna	Hyderabad	Ayu Expert	India	rasendra22@gmail.com
21	Manohar Gundeti	CARI, Mumbai	Ayu Expert	India	gundetipanchakarma@gmail.com
22	Kishor Kumar	NIMHANS, Bengaluru	Ayu Expert	India	ayurkishore@gmail.com
23	Ram Kumar	Nirog Street, Delhi	Industry, AI	India	ram@nirogstreet.com
24	Uddhavesh Sonvane	CDAC, Pune	CS Expert, AI, TM	India	uddhaveshs@gmail.com

Sl.No	Name	Affiliation	Expertise	Country	Email Id
25	Rajendra Joshi	CDAC, Pune	CS Expert, AI, TM	India	rajendra@cdac.in
26	Vinod Jani	CDAC, Pune	CS Expert, AI, TM	India	vinodj@cdac.in
27	Shruti Kolugi	CDAC, Pune	CS Expert, AI, TM	India	shrutik@cdac.in
28	Gaur Sundar	CDAC, Pune	CS Expert, AI, TM	India	gaurs@cdac.in
29	Manihsa Mantri	CDAC, Pune	CS Expert, AI, TM	India	manishar@cdac.in
30	Achyut Avinash Patil	CDAC, Pune	CS Expert, AI, TM	India	achyutp@cdac.in
31	Usha Rani	NIT, Rourkela	CS Expert, AI, TM	India	usharanik.niimh@gmail.com
32	Pratyusha Mantena	NIIMH, Hyderabad	CS Expert, AI, TM	India	hemavallipratyusha@gmail.com
33	Hari Prasad	NIIMH, Hyderabad	Medical Records in TM	India	ayush.fortunehari@gmail.com
34	Dr Kannan	CCRS, Chennai	TM Expert	India	siddhikanna@gmail.com
35	Dr S Natarajan	CCRS, Chennai	TM Expert	India	drnatarajan78@gmail.com
36	Dr Shaista Uruz	CCRUM, New Delhi	TM Expert	India	shaistaccrum@gmail.com
37	Dr Usama Akram	CCRAS, New Delhi	TM Expert	India	usamakramdr@gmail.com
38	Dr Lalitha Sharma	CCRAS, New Delhi	TM Expert	India	lalita.id15@gmail.com
39	Dr Rakesh Narayanan	CCRAS, New Delhi	TM Expert	India	drrakeshccras@gmail.com
40	Dr Anagha Ranade	CCRAS, New Delhi	TM Expert	India	anagharanade11@gmail.com
41	Dr Azeem	CCRAS, New Delhi	TM Expert	India	azeem2905@gmail.com
42	Dr Bidhan Mahajan	CCRAS, New Delhi	TM Expert	India	bidhanmahajon@gmail.com
43	Dr Geetha Krishnana	TM Expert, International Relations	TM Expert, Policy	India	gopalakrishnag@who.int
44	PRASAD PANCHANGAM	SAIGEWARE, Bengaluru, Deep phenotyping platform for personalized healthcare	CS Expert, AI, TM	India	prasad@saigeware.com
45	Meher Vishal Suraj	NIIMH, Hyderabad	CS Expert, AI, TM	India	kmvsreddy7.meher@gmail.com
46	Dr Ajit Kakodkar	Gastro Lab Pvt.Ltd, Pune, India	TM expert, entrepreneur	India	ajitkolatkar@gmail.com

– Relevant next steps

2.3 Topic group participation

The participation in both, the Focus Group on AI for Health and in a TG is generally open to anyone (with a free ITU account). For this TG, the corresponding 'Call for TG participation' (CfTGP) can be found here:

- https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/tg/tm/CfP-TG-TM.pdf

Each topic group also has a corresponding subpage on the ITU collaboration site. The subpage for this topic group can be found here:

- https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/tg/SitePages/TG-TM.aspx

For participation in this topic group, interested parties can also join the regular online meetings. For all TGs, the link will be the standard ITU-TG 'zoom' link:

- https://itu.zoom.us/my/fgai4h

All relevant administrative information about FG-AI4H such as upcoming meetings or document deadlines, will be announced via the general FG-AI4H mailing list fgai4h@lists.itu.int.

All TG members should subscribe to this mailing list as part of the registration process for their ITU user account by following the instructions in the 'Call for Topic Group participation' and this link:

- https://itu.int/go/fgai4h/join

In addition to the general FG-AI4H mailing list, each topic group can create an *individual mailing list:*

- fgai4htgtm@lists.itu.int

Regular FG-AI4H workshops and meetings proceed about every two months at changing locations around the globe or remotely. More information can be found on the official FG-AI4H website:

- https://itu.int/go/fgai4h

3 Topic description

This clause contains a detailed description and background information of the specific health topic for the benchmarking of AI in traditional medicine and how this can help to solve a relevant 'real-world' problem.

Topic Groups summarize related benchmarking AI subjects to reduce redundancy, leverage synergies, and streamline FG-AI4H meetings. However, in some cases different subtopic groups can be established within one topic group to pursue different topic-specific fields of expertise. The TG-AI for traditional medicine currently has no subtopics. Future subtopics for [A] AI for traditional medicine diagnosis, [B] AI for traditional medicine product evaluation might be introduced.

3.1 Definition of the AI task

This clause provides a detailed description of the specific task the AI systems of this TG are expected to solve. It is *not* about the benchmarking process (this will be discussed in more detail in chapter 4). This clause corresponds to DEL03 *"AI requirements specifications*," which describes the functional, behavioural, and operational aspects of an AI system. The following items are for further study:

- AI4TM is utilized to replicate the logical understanding applied in traditional medicine diagnostic methods viz., interview, physical examination and other specific diagnostic equipment, techniques to arrive at pre-diagnosis, prodromes, diagnosis, prognosis, determination of transient patterns, steady states viz., individual constitution, etc.
- The AI tasks implemented are viz., classification, prediction, clustering, or segmentation task, etc.
- AI4TM utilizes the big data generated in the form of the text, sensor-based data and other relevant parameters.
- The output is intended towards producing objective, reproducible and clinically relevant diagnosis.

3.2 Current gold standard

This clause provides a description of the established gold standard of the addressed health topic. The following items are for further study:

- How is the task currently solved without AI?
  - Currently the TM diagnosis is done on a one-on-one basis (whole system approach) involving continuous interaction between the subject and the TM practitioner.
- Do any issues occur with the current gold standard? Does it have limitations?
  - Many of the parameters utilized in TM diagnosis are predominantly subjective which is the very important limitation.
- Are there any numbers describing the performance of the current state of the art?
  - Further, discussions on this topic will provide this information.

3.3 Relevance and impact of an AI solution

This clause addresses the relevance and impact of the AI solution (e.g., on the health system or the patient outcome) and describes how solving the task with AI improves a health issue. The following items are for further study:

- Why is solving the addressed task with AI relevant?
  - Subjective parameters and other whole system related data sets utilized in the TM diagnosis can be converted to objective parameters utilizing data analytics and AI and reduce the individual bias.
- Which impact of deploying such systems is expected?
  - Deploying such a system ensures the objective approach in TM diagnosis and democratizes the knowledge system for wider reach. This certainly will have impact on the health system, overall health system cost, life expectancy, or gross domestic product
- Why is benchmarking for this topic important?
  - Benchmarking this topic provide stakeholders with numbers for decision-making; does it simplify regulation, build trust, or facilitate adoption?

3.4 Existing AI solutions

Traditional medicine views the human body as a whole, and any disagreement in the internal or exterior components of the body is considered a sickness. The essential components of diagnostic procedures in TM include inspection, auscultation and olfaction, inquiry, and palpation. During the inspection, TM practitioners look for aberrant changes in vital signs, internal organs, and external organs to discover pathological modifications in the body system. The process of smelling secretory secretions is known as olfaction, whereas auscultation is the act of listening to the sounds of the body. Inquiry includes learning about a patient's background and family history, appetite, any changes in sleep or waking habits, and daily activities, individual constitution, etc. The pulse is also determined during the palpation method. The TM diagnostic domain has a wide spectrum which includes pre-diagnosis, prodromes, diagnosis, prognosis, determination of transient patterns, *Syndrome differentiation*, and individual constitution. AI is being used in these domains for validation of the concepts and also practical application in the clinical domain[6], [7] and development of assistive AI based diagnostic tools. With the advent of integrated medical set ups and increasing use of electronic

---

[6] G. Arji, A systematic literature review and classification of knowledge discovery in traditional medicine, Computer Methods and Programs in Biomedicine, https://doi.org/10.1016/j.cmpb.2018.10.017

[7] Chu, H., Moon, S., Park, J., Bak, S., Ko, Y., & Youn, B. Y. (2022). The Use of Artificial Intelligence in Complementary and Alternative Medicine: A Systematic Scoping Review. Frontiers in Pharmacology, https://doi.org/10.3389/FPHAR.2022.826044

health records in the traditional medicine domain, usage of TM-terminologies, and morbidity codes for effective interoperability are paving the way for effective data generation and further use of the same for training AI modules.

Pattern recognition is utilised as the cornerstone for pharmaceutical and natural ingredient prescription. Furthermore, network pharmacology and plant element analysis utilising artificial intelligence are applied in a variety of applications.[8] Recently, AI approaches have been utilised to develop a prescription decision-support system based on traditional settings, as well as to evaluate the efficacy of herbal extracts and drugs[9].

Pattern recognition also plays a vital role in determination of individual constitution, wherein AI is being used to build tools for prediction of constitution types (Prakriti) and their linkages with specific genetic, molecular, physiological and cellular level phenotypes and predictive marker discoveries[10,11,12,13,14].

In efficacy analysis, research efforts are focused on evaluating the effects of procedures, herbs, and medications in TM on a different set of patients. Data mining approaches may be used to optimise treatment plans and assess corresponding efficacy. National pharmacopoeias from several nations have recently begun rigorous initiatives to adopt modernised research methodologies in TM. Furthermore, clinical control studies have been launched at various hospitals and research facilities to demonstrate the safety and efficacy of traditional medicine. The link between TM components and their safety, effectiveness, indication, and toxicity is intricate. As a result, various artificial intelligence methodologies should be researched in order to uncover the hidden link between various TM components.[8]

AI can analyse more complex portions of a patients' dataset and detect highly complex and time-dependent conditions such as synergistic action, multi-target action and adverse drug reactions.

---

8 G. Arji, A systematic literature review and classification of knowledge discovery in traditional medicine, Computer Methods and Programs in Biomedicine, https://doi.org/10.1016/j.cmpb.2018.10.017

9 Chu, H., Moon, S., Park, J., Bak, S., Ko, Y., & Youn, B. Y. (2022). The Use of Artificial Intelligence in Complementary and Alternative Medicine: A Systematic Scoping Review. Frontiers in Pharmacology, https://doi.org/10.3389/FPHAR.2022.826044

10 Tiwari, P., Kutum, R., Sethi, T., Shrivastava, A., Girase, B., Aggarwal, S., Patil, R., Agarwal, D., Gautam, P., Agrawal, A. and Dash, D., Ghosh S, Juvekar S#, Mukerji M#, Prasher, B# 2017. Recapitulation of Ayurveda constitution types by machine learning of phenotypic traits. PloS one. October 5, 2017 https://doi.org/10.1371/journal.pone.0185380

11 Abbas, T.; Chaturvedi, G.; Prakrithi, P.; Pathak, A.K.; Kutum, R.; Dakle, P.; Narang, A.; Manchanda, V.; Patil, R.; Aggarwal, D.; Girase, B.; Srivastava, A.; Kapoor, M.; Gupta, I.; Pandey, R.; Juvekar, S.; Dash, D#.; Mukerji, M#.; Prasher B#. Whole Exome Sequencing in Healthy Individuals of Extreme Constitution Types Reveals Differential Disease Risk: A Novel Approach towards Predictive Medicine. J. Pers. Med. 2022, 12, 489. https://doi.org/10.3390/jpm12030489

12 Rani, R., Rengarajan, P., Sethi, T., Khuntia, B. K., Kumar, A., Punera, D. S., Singh, D., Girase, B., Shrivastava, A., Juvekar, S. K., Pesala, B., Mukerji, M., Deepak, K. K#., & Prasher, B#. (2022). Heart rate variability during head-up tilt shows inter-individual differences among healthy individuals of extreme Prakriti types. Physiological Reports, 10, e15435. https://doi.org/10.14814/phy2.15435

13 Chakraborty S, Singhmar S, Singh D, Maulik M, Patil R, Agrawal SK, Mishra A, Ghazi M, Vats A, Natarajan VT, Juvekar S, Prasher B#, Mukerji M#. Baseline cell proliferation rates and response to UV differ in lymphoblastoid cell lines derived from healthy individuals of extreme constitution types. Cell Cycle. 2021 May;20(9):903-913. doi: 10.1080/15384101.2021.1909884. Epub 2021 Apr 18. PMID: 33870855; PMCID: PMC8168715.

14 Prasher, B., Aggarwal, S., Mandal, A. K., Sethi, T. P., Deshmukh, S. R., Purohit, S. G., Sengupta, S., Khanna, S., Mohammad, F., Garg, G., Brahmachari, S. K., & Mukerji, M. (2008). Whole genome expression and biochemical correlates of extreme constitutional types defined in Ayurveda. Journal of Translational Medicine, 6. https://doi.org/10.1186/1479-5876-6-48

Further research calls for accurate biomarker selection, assessment, and application when using AI in diverse sectors with conventional medical modalities. Despite these many obstacles, it is anticipated that the use of AI technology in conventional medicine would increase in the future. Well-designed randomised controlled trials are required to verify the AI models in order to make this possible.[15]

Other use cases of AI in TM are the supply chain of medicines, supplements and other products based on traditional medicine and capacity building, monitoring of TM.

In the field of T&CM, AI is used for screening, evaluation, therapy, relapse prevention, and other purposes. The combination of T&CM and AI, in particular, is predicted to deliver improved responses to the shortcomings of present conventional treatments. Because herbal medications have a complicated chemical mechanism of action, AI approaches will be able to assist in establishing accurate diagnosis and the prescription of the most effective TM remedies.[15]

The techniques employed for creating AI tools based on traditional medicine are diverse and cover a wide spectrum. Some of the widely used methods are enlisted here: adaboost (AB), artificial neural network (ANN), back propagation (BP), classification tree (CT), convolutional neural network (CNN), damped least-squares (DLS), decision tree (DT), deep learning (DL), k nearest neighbour (kNN), least absolute shrinkage and selection operator (LASSO), Levenberg–Marquardt algorithm (LMA), linear discriminant analysis (LDA), logistic regression (LR), machine learning (ML), multiple linear regression (MLR), multivariate pattern analysis (MVPA), naïve bayes (NB), naïve Bayesian (NB), neural network (BP), pattern identification algorithm (PI), probabilistic neural network method (PNN), quantitative structure-activity relationship (QSAR), random ensemble with asymmetric learning (REAL), random forest (RF), self-organizing map (SOM), support vector machine (SVM), synthetic minority oversampling technique (SMOTE), Takagisugeno-Kang fuzzy system (TSK), term frequency-inverse document frequency (TIFD), and the deformable template (DT)[15, 16], The array of techniques enlisted here demonstrates the utilization of best possible AI solutions in the domain of traditional medicine on a par with the modern medicine. In this context, in spite of several constraints, the enthusiasm and dedication showcased by the very limited number of dedicated professionals, scientists, and traditional medicine practitioners who are working in the domain of standardization of TM and further adoption of the same through emerging digital health solution is worth appreciating.

3.4.1 AI in TM diagnosis

Ref #	Intended use	Target population	Type of AI used	Input	Performance
Jung et al. (2019), Yeh et al.'s (2020)[17]	Predict acupoint patterns	Retrospective E.H.R Data	Multivariate resting-state FCs	Medical records based on symptom and disease information	(Exploratory work)
Chen et al., 2014; Lin et al., 2015 &	Tongue and Lip Diagnoses	Prospective, Chronic Renal, Liver Disorders;	Multi-class SVM algorithm & SCM-REF	Colour recognition, patterning, and digitization,	AI Work

15 Chu, H., Moon, S., Park, J., Bak, S., Ko, Y., & Youn, B. Y. (2022). The Use of Artificial Intelligence in Complementary and Alternative Medicine: A Systematic Scoping Review. Frontiers in Pharmacology, https://doi.org/10.3389/FPHAR.2022.826044

16 G. Arji, A systematic literature review and classification of knowledge discovery in traditional medicine, Computer Methods and Programs in Biomedicine, https://doi.org/10.1016/j.cmpb.2018.10.017

17 https://www.frontiersin.org/articles/10.3389/fphar.2022.826044/full

Ref #	Intended use	Target population	Type of AI used	Input	Performance
Li et al. (2012) [18]		Extended to Healthy patients	feature selection	tongue size, teeth marks	
LeungYeuk-Lan Alice et al. (2021)[19]; Liu, S et al. (2018)[20]; Xu, L., et al. (2008)[21]; Xu, L. S (2007)[22]; Yeuk-Lan Alice et al. (2020)[23]	Pulse pattern recognition (TCM)	Healthy volunteers	artificial neural networks (ANNs); pulse-sensing platform (PSP); Fuzzy neural network;	Record and classify arterial human pulses	AI Work
A. Joshi (2007)[24;25]	Pulse pattern recognition (Ayurveda)	Healthy people and others suffering with health issues.	Acquiring radial pulse patterns in analogue, digital format and analysis.	Pre-AI work	Pre-AI work
Han et al., 2018; He et al., 2019; Zhou et al., 2021, Fu et al., 2013; Shen et	Prescription decision supporting system using traditional contexts or explore the	Non-human pre-clinical studies	In-silico analysis	Work based on existing herbal and other pharmacological databases	(Exploratory work)

18 https://www.frontiersin.org/articles/10.3389/fphar.2022.826044/full

19 https://www.sciencedirect.com/science/article/pii/S258937772100001X

20 Liu, S., Hua, L., Lv, P., Yu, Y., Gao, Y., & Sheng, X. (2018). A Pulse Condition Reproduction Apparatus for Remote Traditional Chinese Medicine. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10984 LNAI*, 453–464. https://doi.org/10.1007/978-3-319-97586-3_41

21 Xu, L., Meng, M. Q. H., Shi, C., Wang, K., & Li, N. (2008). Quantitative analyses of pulse images in traditional Chinese medicine. *Medical Acupuncture*, 20(3), 175–189. https://doi.org/10.1089/ACU.2008.0632

22 Xu, L. S., Meng, M. Q. H., & Wang, K. Q. (2007). Pulse image recognition using fuzzy neural network. *Annual International Conference of the IEEE Engineering in Medicine and Biology – Proceedings*, 3148–3151. https://doi.org/10.1109/IEMBS.2007.4352997

23 Yeuk-Lan Alice, L., Binghe, G., Shuang, C., Hoyin, C., Kawai, K., Wenjung, L., & Jiangang, S. (2021). Artificial intelligence meets traditional Chinese medicine: a bridge to opening the magic box of sphygmopalpation for pulse pattern recognition. *Digital Chinese Medicine*, 4(1), 1–8. https://doi.org/10.1016/J.DCMED.2021.03.001

24 Joshi, A. Kulkarni, S. Chandran, V. K. Jayaraman and B. D. Kulkarni, "Nadi Tarangini: A Pulse Based Diagnostic System," 2007 29th *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2007, pp. 2207-2210, doi: 10.1109/IEMBS.2007.4352762.

25 Joshi, A., Kulkarni, A., Chandran, S., Jayaraman, V. K., & Kulkarni, B. D. (2007). Nadi Tarangini: A Pulse Based Diagnostic System. *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2207–2210. https://doi.org/10.1109/IEMBS.2007.4352762

Ref #	Intended use	Target population	Type of AI used	Input	Performance
al., 2021; Yang et al., 2019; Zhang et al., 2021 [26]	efficacy of herbal extracts and prescriptions				
Feng et al., 2021b; Song et al., 2021). [27]; Zhang, H et al., (2020)[28]	TCM diagnosis and TCM symptom classification	Non-human; Data driven study	Artificial neural network (ANN), data mining, and multivariate analysis		AI Work

3.4.2 AI in understanding human constitution, physiology by TM methods

Ref #	Intended use	Target population	Type of AI used	Input	Performance
Wallace, R. K. (2020). [29]; Prashar B et al. (2017)[30]; Parashar B et al. (2008)[31]; Bhargav H et al. (2021)[32]; Renu Singh et al. (2017)[33]	Understanding human constitution (Prakriti, etc.) and its relations ship genetic phenotype and other related use cases	Healthy volunteers	Structured questionnaires, Genome sequencing and other related methods of data collection.	(Pre-AI work)	(Exploratory work)

---

26 https://www.frontiersin.org/articles/10.3389/fphar.2022.826044/full

27 https://www.frontiersin.org/articles/10.3389/fphar.2022.826044/full

28 Zhang, H., Ni, W., Li, J., & Zhang, J. (2020). Artificial intelligence-based traditional chinese medicine assistive diagnostic system: Validation study. *JMIR Medical Informatics*, 8(6). https://doi.org/10.2196/17608

29 Wallace, R. K. (2020). Ayurgenomics and modern medicine. Medicina (Lithuania), 56(12), 1–7. https://doi.org/10.3390/medicina56120661

30 Prasher, B., Varma, B., Kumar, A., Khuntia, B. K., Pandey, R., Narang, A., Tiwari, P., Kutum, R., Guin, D., Kukreti, R., Dash, D., Mukerji, M., Aggarwal, S., Natarajan, V., Salvi, S., Aatreya, P., Unni, S., Mishra, N., Mudgal, N., … Makhija, N. (2017). Ayurgenomics for stratified medicine: TRISUTRA consortium initiative across ethnically and geographically diverse Indian populations. *Journal of Ethnopharmacology*, 197, 274–293. https://doi.org/10.1016/j.jep.2016.07.063

31 Prasher, B., Aggarwal, S., Mandal, A. K., Sethi, T. P., Deshmukh, S. R., Purohit, S. G., Sengupta, S., Khanna, S., Mohammad, F., Garg, G., Brahmachari, S. K., & Mukerji, M. (2008). Whole genome expression and biochemical correlates of extreme constitutional types defined in Ayurveda. *Journal of Translational Medicine*, 6. https://doi.org/10.1186/1479-5876-6-48

32 Bhargav H, Jasti N, More P, Kumar V, Chikkanna U, Kishore Kumar R, et al. Correlation of prakriti diagnosis using AyuSoft prakriti diagnostic tool with clinician rating in patients with psychiatric disorders. J Ayurveda Integr Med [Internet]. 2021;12(2):365–8. Available from: https://www.sciencedirect.com/science/article/pii/S0975947621000115

33 https://www.researchgate.net/publication/322899120_Development_of_Standardized_Prakriti_Assessment_Tool_An_Overview_of_Ongoing_CCRAS_Initiatives

Ref #	Intended use	Target population	Type of AI used	Input	Performance
Madan V et al[34], Tiwari P et al.[35], Katua D et al.[36]	Predicting Ayurveda-based Prakriti analysis (phenotypic traits) using ML, Deep Learning	Healthy volunteers	ML, Deep learning	AI Work	AI Work
Wayne P et al.[37]	A systems biology approach to studying Tai Chi, physiological complexity and healthy aging: Design and rationale of a pragmatic randomized controlled trial	Healthy volunteers	Systems biology approach	(Pre-AI work)	(Exploratory work)
Anonymous[38]	Tai Chi, Physiological Complexity, and Healthy Aging – Gait v1.0.2	Healthy volunteers	Posture and gait analysis	(Pre-AI work)	(Exploratory work)
Anonymous[39]	A multi-camera and multimodal dataset for posture and gait analysis v1.0.0	Healthy volunteers	Posture and gait analysis		

34 Madaan, V., & Goyal, A. (2020). Predicting Ayurveda-based constituent balancing in human body using machine learning methods. *IEEE Access*, 8, 65060–65070. https://doi.org/10.1109/access.2020.2985717

35 Tiwari, P., Kutum, R., Sethi, T., Shrivastava, A., Girase, B., Aggarwal, S., Patil, R., Agarwal, D., Gautam, P., Agrawal, A., Dash, D., Ghosh, S., Juvekar, S., Mukherji, M., & Prasher, B. (2017). Recapitulation of Ayurveda constitution types by machine learning of phenotypic traits. *PLoS ONE*, *12*(10). https://doi.org/10.1371/journal.pone.0185380

36 Khatua, D., Sekh, A. A., Kutum, R., Mukherji, M., Prasher, B., & Kar, S. (2023). Classification of Ayurveda constitution types: a deep learning approach. *Soft Computing*, 1–9. https://doi.org/10.1007/S00500-023-07942-2/METRICS

37 Wayne, P. M., Manor, B., Novak, V., Costa, M. D., Hausdorff, J. M., Goldberger, A. L., Ahn, A. C., Yeh, G. Y., Peng, C. K., Lough, M., Davis, R. B., Quilty, M. T., & Lipsitz, L. A. (2013). A systems biology approach to studying Tai Chi, physiological complexity and healthy aging: Design and rationale of a pragmatic randomized controlled trial. *Contemporary Clinical Trials*, *34*(1), 21–34. https://doi.org/10.1016/J.CCT.2012.09.006

38 *Tai Chi, Physiological Complexity, and Healthy Aging – Gait v1.0.2*. (n.d.). Retrieved February 11, 2023, from https://physionet.org/content/taichidb/1.0.2/

39 *A multi-camera and multimodal dataset for posture and gait analysis v1.0.0*. (n.d.). Retrieved February 11, 2023, from https://physionet.org/content/multi-gait-posture/1.0.0/

3.4.3 AI in safety, efficacy studies of TM-Products (Food, medicine and others)

Ref #	Intended use	Target population	Type of AI used	Input	Performance
Jayasundar R et al. (2020, 2021)[40], Kumar D, Singh A, Jayasundar R et al. (2021)[41]	Detection of taste of medicinal plants		Electronic tongue, NMR analysis	(Pre-AI work)	(Exploratory work)

3.4.4 Miscellaneous (TM Standards, usage, knowledge, aptitude, perception, policy, etc.)

Ref #	Intended use	Target population	Type of AI used	Input	Performance
Hongmin Chu et al. [42]	The use of artificial intelligence in complementary and alternative medicine: A systematic scoping review	Review	Scoping review	Review	Review
Astin J et al.[43]	Why patients use alternative medicine	General	Review	(Pre-AI work)	(Exploratory work)
Jansen C[44]	Integrative medicine: Opportunities, challenges and data analytics-based solutions for traditional medicine	General	Review	(Pre-AI work)	(Exploratory work)

---

[40] Jayasundar R, Singh A, Kumar D. Challenges in using electronic tongue to study rasa of plants: I. Finding the right tool for the right job. J Ayurveda Integr Med [Internet]. 2021;12(2):234–7. Available from: https://www.sciencedirect.com/science/article/pii/S0975947620301467

[41] Kumar D, Singh A, Jayasundar R. Challenges in using Electronic tongue to study rasa of plants: II. Impact of solvent and concentration on sensor response and taste ranking. J Ayurveda Integr Med [Internet]. 2021;12(2):238–44. Available from: https://www.sciencedirect.com/science/article/pii/S0975947620301455

[42] Chu, H., Moon, S., Park, J., Bak, S., Ko, Y., & Youn, B. Y. (2022). The Use of Artificial Intelligence in Complementary and Alternative Medicine: A Systematic Scoping Review. *Frontiers in Pharmacology*, *13*. https://doi.org/10.3389/FPHAR.2022.826044

[43] Astin, J. A. (1998). Why Patients Use Alternative Medicine: Results of a National Study. *JAMA*, *279*(19), 1548–1553. https://doi.org/10.1001/JAMA.279.19.1548

[44] Jansen, C., Baker, J. D., Kodaira, E., Ang, L., Bacani, A. J., Aldan, J. T., Shimoda, L. M. N., Salameh, M., Small-Howard, A. L., Stokes, A. J., Turner, H., & Adra, C. N. (2021). Medicine in motion: Opportunities, challenges and data analytics-based solutions for traditional medicine integration into western medical practice. *Journal of Ethnopharmacology*, *267*, 113477. https://doi.org/10.1016/J.JEP.2020.113477

Ref #	Intended use	Target population	Type of AI used	Input	Performance
Anonymous[45], Ammon K[46], Bornhöft G et al.[47]	Health technology assessment (HTA) and TM	General	HTA	(Pre-AI work)	(Exploratory work)
Rudra S et al.[48]	Utilization of alternative systems of medicine as health care services in India: Evidence on AYUSH care from NSS 2014	General	Review	(Pre-AI work)	(Exploratory work)
Saketh Ram et al.[49], Anonymous[50]	NAMASTE Portal, Glossary of Ayurveda Terminologies (BIS)	General	Collaborative, consultative work	(Pre-AI work)	(Exploratory work)
Anonymous	WHO benchmark documents for training and practice of Ayurveda, Unani, Acupuncture, Naturopathy[51],	General	Collaborative, consultative work	(Pre-AI work)	(Exploratory work)

45 *HTA in Switzerland - SwissHTA – Swiss Health Technology Assessment*. (n.d.). Retrieved December 26, 2022, from http://www.swisshta.org/index.php/HTA_in_Switzerland.html

46 Ammon, K. von, Cardini, F., Daig, U., Dragan, S., FreiErb, Martin, Hegyi, G., Sarsina, P. R. di, Sørensen, J., Ursoniu, S., Weidenhammer, W., & Lewith, G. (n.d.). *Health Technology Assessment (HTA) and a map of CAM provision in the EU (Final Report of CAMbrella Work Package 5)*. CAMBRELLLA. Retrieved December 26, 2022, from https://cam-europe.eu/wp-content/uploads/2018/09/WP5-CAMbrella-WP5final.pdf

47 Bornhöft, Gudrun., & Matthiessen, P. F. (2011). *Homeopathy in healthcare – Effectiveness, appropriateness, safety, costs: an HTA report on homeopathy as part of the Swiss Complementary Medicine Evaluation Programme*. 209. https://www.hri-research.org/resources/homeopathy-the-debate/the-swiss-hta-report-on-homeopathy/

48 Rudra, S., Kalra, A., Kumar, A., & Joe, W. (2017). Utilization of alternative systems of medicine as health care services in India: Evidence on AYUSH care from NSS 2014. *PloS One*, *12*(5). https://doi.org/10.1371/JOURNAL.PONE.0176916

49 Saketh Ram et al. (2017). National Ayurveda Morbidity Codes (NAMC). In *National Ayush Morbidity and Standardized Electronic (NAMASTE) Portal*. Ministry of Ayush, Government of India. http://namstp.ayush.gov.in/#/Ayurveda

50 Anonymous. (2021). *Glossary of Ayurvedic Terminolgy Part1-5*. https://www.services.bis.gov.in:8071/php/BIS_2.0/bisconnect/cls_module/Ministry_list/ministry_stndrds_list?mns_id=NTI%3D&mns_name=TWluaXN0cnkgb2YgQVlVU0g%3D&aspect=&from=&to=

51 https://www.who.int/publications/

Ref #	Intended use	Target population	Type of AI used	Input	Performance
Anonymous	WHO international standard terminologies on Ayurveda[52], Siddha[53], Unani[54], Traditional Chinese Medicine[55]	General	Collaborative, consultative work	(Pre-AI work)	(Exploratory work)

4 Ethical considerations

The rapidly evolving field of AI and digital technology in the fields of medicine and public health raises a number of ethical, legal, and social concerns that have to be considered in this context. They are discussed in deliverable DEL1 " *Ethics and governance of artificial intelligence for health,"* which was developed by the working group on "Ethical considerations on AI4H" (WG-Ethics). This clause refers to DEL01 and should reflect the ethical considerations of the TG-AI for traditional medicine.

The majority of AI developments in health care are geared towards the demands of high-income countries (HICs), where the majority of research is carried out. The bulk of AI software for medicine is likewise created in HICs, with data from HIC participants. Using models created in HICs without adequate validation in low-and middle-income countries (LMICs), however, might introduce and spread bias.[56] This condition is comparable to that of conventional medicine. Here, contemporary medicine and traditional medicine are in contrast, with the former having access to greater resources and practitioners. This is due to the fact that TM is a frequently overlooked health resource with a wide range of uses, particularly in the prevention and management of chronic illnesses linked to a certain lifestyle and in addressing the health requirements of ageing populations. At a time when consumer expectations for care are growing, expenditures are skyrocketing, and the majority of budgets are either flat or decreasing, several nations are attempting to increase coverage of basic health services. The 21st century's particular health issues are rekindling interest in TM.[57].

4.1 AI and TM: Consequences of digital divide

The "digital divide" refers to the uneven distribution of access to, use of or effect of information and communication technologies among any number of distinct groups. Although the cost of digital technologies is falling, access has not become more equitable, and the digital divide persists geographically and by race, ethnicity, gender and age.[58] Traditional medicine and its practitioners are

---

[52] https://www.who.int/publications/i/item/9789240064935

[53] https://www.who.int/publications/i/item/9789240064973

[54] https://www.who.int/publications/i/item/9789240064959

[55] https://www.who.int/publications/i/item/9789240042322

[56] Generating evidence for artificial intelligence-based medical devices: a framework for training, validation and evaluation. Geneva: World Health Organization;2021. Licence: CC BY-NC-SA 3.0 IGO. https://www.who.int/publications/i/item/9789240038462

[57] World Health Organization. (2019). WHO global report on traditional and complementary medicine 2019. World Health Organization. https://apps.who.int/iris/handle/10665/312342. License: CC BY-NC-SA 3.0 IGO

[58] Ageism in artificial intelligence for health: WHO policy brief. Geneva: World Health Organization; 2022. Licence: CC BY-NC-SA 3.0 IGO.

operational in remote areas away from urban infrastructure and face challenges with the availability of Internet and digital literacy. Exceptions to this are the places where TM is well organized in countries like China, India and developed countries where people opt for TM by choice or as a cultural preference. The digital divide in this segment varies drastically from country to country. Less "algorithmic awareness" is a new, reinforced level of the digital divide, as it is a skill required for successful negotiation of digital technologies[59]. Considering the notion that that the digital divide will narrow over time (defined more by generation than by age)[59], in the case of TM it may in fact worsen due to non-acceptance of such technologies by some TM groups considering them as an interference with natural healing and holistic health practices.

Lower rates of TM practitioners and seekers participating in the digital economy or insufficient use of digital technology are also effects of the digital divide. Traditional medicine may not be adequately represented in the data sets used to train and evaluate AI algorithms without appropriate user engagement, making the technologies less specialised for unique traits and needs. Insufficient engagement may also lead to the commercial sector's perception of conventional medical practises as having less value when they create and use new technologies, such as AI.

Further, recognising traditional medicine and upholding the rights of its practitioners are requirements for non-discrimination against it. Because of cultural customs or a lack of access to modern healthcare in many locations, they are frequently the locals' first point of contact. Every human being has a basic right to enjoy the best possible level of health, which has long been acknowledged. Therefore, only the very finest medical procedures should serve as the benchmark. In line with the United Nations Declaration on the Rights of Indigenous Peoples, it must be acknowledged both that "indigenous peoples have the right to their traditional medicines and to maintain their health practices" and that "indigenous individuals have an equal right to the enjoyment of the highest attainable standard of physical and mental health. States shall take the necessary steps with a view to achieving progressively the full realization of this right" (Article 24)[60].

The following items are for further study:

- What are the ethical implications of applying the AI model in real-world scenarios?
- What are the ethical implications of introducing benchmarking (having the benchmarking in place itself has some ethical risks; e.g., if the test data are not representative for a use case, the data might create the illusion of safety and put people at risk)?
- What are the ethical implications of collecting the data for benchmarking (e.g., how is misuse of data addressed, is there the need for an ethics board approval for clinical data, is there the need for consent management for sharing patient data, and what are the considerations about data ownership/data custodianship)?
- What risks face individuals and society if the benchmarking is wrong, biased, or inconsistent with reality on the ground?
- How is the privacy of personal health information protected (e.g., in light of longer data retention for documentation, data deletion requests from users, and the need for an informed consent of the patients to use data)?
- How is ensured that benchmarking data are representative and that an AI offers the same performance and fairness (e.g., can the same performance in high, low-, and middle-income countries be guaranteed; are differences in race, sex, and minority ethnic populations captured; are considerations about biases, when implementing the same AI application in a

[59] Gran AB, Booth P, Bucher T. To be or not to be algorithm aware: a question of a new digital divide? Inf Commun Soc. 2021;24(12):1779–96 (doi: 10.1080/1369118X.2020.1736124).

[60] Committee, authorCorporate:International B. (2012). Report of the IBC on traditional medicine systems and their ethical implications. https://unesdoc.unesco.org/ark:/48223/pf0000217457

different context included; is there a review and clearance of 'inclusion and exclusion criteria' for test data)?

- What are your experiences and learnings from addressing ethics in your TG?

5 Existing work on benchmarking

This clause focuses on the existing benchmarking processes in the context of AI and traditional medicine for quality assessment. It addresses different aspects of the existing work on benchmarking of AI systems (e.g., relevant scientific publications, benchmarking frameworks, scores and metrics, and clinical evaluation attempts). The goal is to collect all relevant learnings from previous benchmarking that could help to implement the benchmarking process in this topic group.

5.1 Publications on benchmarking systems

While a representative comparable benchmarking for traditional medicine does not yet exist, some work has been done in the scientific community assessing the performance of such systems. This clause summarizes insights from the most relevant publications on this topic. It covers parts of the deliverable DEL7 *"AI for health evaluation considerations,"* DEL7.1 *"AI4H evaluation process description,"* DEL7.2 *"AI technical test specification,"* DEL7.3 *"Data and artificial intelligence assessment methods (DAISAM),"* and DEL7.4 *"Clinical Evaluation of AI for health"*.

- What are the most relevant peer-reviewed scientific publications on benchmarking or objectively measuring the performance of systems in your topic?

After an extensive literature review, it is observed that there are no peer-reviews of scientific publications on benchmarking of AI practices currently, however attempts are being made bringing in objectivity of reporting parameters at data collection level e.g., TCM-SD: *A Benchmark for Probing Syndrome Differentiation via Natural Language Processing*[61], Prakriti-based research: *Good reporting practices*[62] Some of the pre-AI attempts in the form of public databases, algorithms, and software related to traditional medicine related with network pharmacology are as follows:

Type	Name	Description	Website for database or tool	Reference
Databases (TM Specific)	TCM-Mesh	An integration of database and a data-mining system for network pharmacology analysis of TCM preparations	http://mesh.tcm.microbioinformatics.org	Zhang et al., 2017[63]
	TCM database@Taiwan	The world's largest and most	http://tcm.cmu.edu.tw	Chen, 2011[64]

[61] Mucheng, R., Heyan, H., Yuxiang, Z., Qianwen, C., Yuan, B., & Yang, G. (2022). *TCM-SD: A Benchmark for Probing Syndrome Differentiation via Natural Language Processing* (pp. 908–920). Chinese Information Processing Society of China. https://aclanthology.org/2022.ccl-1.80

[62] Bhalerao, S., & Patwardhan, K. (2016). Prakriti-based research: Good reporting practices. *Journal of Ayurveda and Integrative Medicine*, *7*(1), 69–72. https://doi.org/10.1016/J.JAIM.2015.08.002

[63] Zhang, R. Z., Yu, S. J., Bai, H., and Ning, K. (2017). TCM-Mesh: the database and analytical system for network pharmacology analysis for TCM preparations. Sci. Rep. 7:2821. doi: 10.1038/s41598-017-03039-7

[64] Chen, C. Y. (2011). TCM Database@Taiwan: the world's largest traditional Chinese medicine database for drug screening in silico. PLoS One 6:e15939. doi: 10.1371/journal.pone.0015939

Type	Name	Description	Website for database or tool	Reference
		comprehensive free small molecular database on TCM for virtual screening		
	HIT	A comprehensive and fully curated database to complement available resources on protein targets for FDA-approved drugs as well as the promising precursor compounds	http://lifecenter.sgst.cn/hit	Ye et al., 2011[65]
	TCMSP	A unique systems pharmacology platform of TCMs that captures the relationships among drugs, targets, and diseases	http://lsp.nwu.edu.cn/tcmsp.php	Ru et al., 2014[66]
	TCMID	A comprehensive database that provides information and bridges the gap between TCM and modern life sciences	http://www.megabionet.org/tcmid	Xue et al., 2013[67]
	NAMASTE Portal	National Ayush Morbidity and Standardized	http://namstp.ayush.gov.in	Saketh et. Al 2017[68]

65 Ye, H., Ye, L., Kang, H., Zhang, D., Tao, L., Tang, K., et al. (2011). HIT: linking herbal active ingredients to targets. Nucleic Acids Res. 39, D1055–D1059. doi: 10.1093/nar/gkq1165

66 Ru, J., Li, P., Wang, J., Zhou, W., Li, B., Huang, C., et al. (2014). TCMSP: a database of systems pharmacology for drug discovery from herbal medicines. J. Cheminform. 6:13. doi: 10.1186/1758-2946-6-13

67 Xue, R., Fang, Z., Zhang, M., Yi, Z., Wen, C., and Shi, T. (2013). TCMID: traditional Chinese medicine integrative database for herb molecular mechanism analysis. Nucleic Acids Res. 41, D1089–D1095. doi: 10.1093/nar/gks1100

68 Saketh Ram et al. National Ayurveda Morbidity Codes (NAMC). In: National Ayush Morbidity and Standardized Electronic (NAMASTE) Portal [Internet]. New Delhi: Ministry of Ayush, Government of India; 2017. Available from: http://namstp.ayush.gov.in/#/Ayurveda

Type	Name	Description	Website for database or tool	Reference
		terminologies electronic portal		
	IMPPAT: Indian Medicinal Plants, Phytochemistry and Therapeutics	Indian Medicinal Plants, Phytochemistry and Therapeutics 2.0 (IMPPAT 2.0) is a manually curated database which has been constructed via digitalization of information from more than 100 books on traditional Indian medicine, 7000+ published research articles and other existing resources	https://cb.imsc.res.in/imppat/	Areejit Sammal et. Al 2018[69], 2023[70]
Drug related databases	Drug Bank	A unique bioinformatics and cheminformatics resource that combines detailed drug data with comprehensive drug target information	https://www.drugbank.ca	Wishart et al., 2006[71]
	STITCH	A database of known and predicted interactions between	http://stitch.embl.de	Kuhn et al., 2014[72]

[69] Mohanraj, K., Karthikeyan, B. S., Vivek-Ananth, R. P., Chand, R. P. B., Aparna, S. R., Mangalapandi, P., & Samal, A. (2018). IMPPAT: A curated database of Indian Medicinal Plants, Phytochemistry And Therapeutics. *Scientific Reports*, *8*(1). https://doi.org/10.1038/S41598-018-22631-Z

[70] Vivek-Ananth, R. P., Mohanraj, K., Sahoo, A. K., & Samal, A. (2023). IMPPAT 2.0: An Enhanced and Expanded Phytochemical Atlas of Indian Medicinal Plants. *ACS Omega*, *8*(9), 8827–8845. https://doi.org/10.1021/ACSOMEGA.3C00156/ASSET/IMAGES/LARGE/AO3C00156_0012.JPEG

[71] Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., et al. (2006). DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res. 34, D668–D672. doi: 10.1093/nar/gkj067

[72] Kuhn, M., Szklarczyk, D., Pletscher-Frankild, S., Blicher, T. H., von Mering, C., Jensen, L. J., et al. (2014). STITCH 4: integration of protein-chemical interactions with user data. Nucleic Acids Res. 42, D401–D407. doi: 10.1093/nar/gkt1207

Type	Name	Description	Website for database or tool	Reference
		chemicals and proteins		
	ChEMBL	An Open Data database containing binding, functional, and ADMET information for a large number of drug-like bioactive compounds	https://www.ebi.ac.uk/chembl	Gaulton et al., 2012[73]
	PubChem	A public information system for analysing the bioactivities of small molecules	https://pubchem.ncbi.nlm.nih.gov	Wang et al., 2009[74]
Target related databases	STRING	A database of known and predicted protein-protein interactions	https://string-db.org	Szklarczyk et al., 2015[75]
	HPRD	An object database that integrates a wealth of information relevant to the function of human proteins in health and disease	http://www.hprd.org	Peri et al., 2003[76]

[73] Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., et al. (2012). ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res. 40, D1100–D1107. doi: 10.1093/nar/gkr777

[74] Wang, Y., Xiao, J., Suzek, T. O., Jian, Z., Wang, J., and Bryant, S. H. (2009). PubChem: a public information system for analyzing bioactivities of small molecules. Nucleic Acids Res. 37, W623–W633. doi: 10.1093/nar/gkp456

[75] Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 43, D447–D452. doi: 10.1093/nar/gku1003

[76] Peri, S., Navarro, J. D., Amanchy, R., Kristiansen, T. Z., Jonnalagadda, C. K., Surendranath, V., et al. (2003). Development of human protein reference database as an initial platform for approaching systems biology in humans. Genome Res. 13, 2363–2371. doi: 10.1101/gr.1680803

Type	Name	Description	Website for database or tool	Reference
	MINT	A database that focuses on experimentally verified protein-protein interactions mined from the scientific literature by expert curators	https://mint.bio.uniroma2.it	Zanzoni et al., 2002[77]
	IntAct	A freely available, open-source database system and analysis tool for molecular interaction data	https://www.ebi.ac.uk/intact	Kerrien et al., 2012[78]
	Reactome	A free, open-source, curated, and peer-reviewed pathway database	https://reactome.org	D'Eustachio, 2009[79]
	HAPPI	An online database of comprehensive human annotated and predicted protein interactions	http://discovery.informatics.uab.edu/HAPPI/	Chen et al., 2009[80]
Disease related databases	OMIM A	A comprehensive, authoritative compendium of human genes and genetic phenotypes that is freely	https://www.omim.org	Hamosh et al., 2002[81]

77 Zanzoni, A., Montecchi-Palazzi, L., Quondam, M., Ausiello, G., Helmer-Citterich, M., and Cesareni, G. (2002). MINT: a Molecular INTeraction database. FEBS Lett. 513, 135–140. doi: 10.1016/S0014-5793(01)03293-8

78 Kerrien, S., Aranda, B., Breuza, L., Bridge, A., Broackescarter, F., Chen, C., et al. (2012). The IntAct molecular interaction database in 2012. Nucleic Acids Res. 40, D841–D846. doi: 10.1093/nar/gkr1088

79 D'Eustachio, P. (2009). Reactome knowledgebase of human biological pathways and processes. Nucleic Acids Res. 37, D619–D622. doi: 10.1093/nar/gkn863

80 Chen, J. Y., Mamidipalli, S., and Huan, T. (2009). HAPPI: an online database of comprehensive human annotated and predicted protein interactions. BMC Genomics 10:S16. doi: 10.1186/1471-2164-10-S1-S16

81 Hamosh, A., Scott, A. F., Amberger, J., Bocchini, C., Valle, D., and McKusick, V. A. (2002). Online mendelian inheritance in man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res. 30, 52–55. doi: 10.1093/nar/30.1.52

Type	Name	Description	Website for database or tool	Reference
		available and updated daily		
	GAD	A database of genetic association data from complex diseases and disorders	https://geneticassociationdb.nih.gov	Becker et al., 2004[82]
Algorithms	Random Walk	Random Walk is an algorithm that predicts potential drug-target interactions on a large scale under the hypothesis that similar drugs often target similar target proteins and the framework of Random Walk	https://www.rdocumentation.org/packages/diffusr/versions/0.1.4/topics/random.walk	
	PRINCE	A global, network-based method for prioritizing disease genes and inferring protein complex associations	https://github.com/fosterlab/PrInCE	Vanunu et al., 2010[83]
Software	Cytoscape	A software environment for integrated models of biomolecular interaction networks	https://cytoscape.org	Shannon et al., 2003[84]
	Pajek	A tool for complex network analysis	http://mrvar.fdv.uni-lj.si/pajek	Dohleman, 2006[85]

The following items are for further study:

---

[82] Becker, K. G., Barnes, K. C., Bright, T. J., and Wang, S. A. (2004). The genetic association database. Nat. Genet. 36, 431–432. doi: 10.1038/ng0504-431

[83] Vanunu, O., Magger, O., Ruppin, E., Shlomi, T., and Sharan, R. (2010). Associating genes and protein complexes with disease via network propagation. PLoS Comput. Biol. 6:e1000641. doi: 10.1371/journal.pcbi.1000641

[84] Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504. doi: 10.1101/gr.1239303

[85] Dohleman, B. S. (2006). Exploratory social network analysis with Pajek. Psychometrika 71, 605–606. doi: 10.1007/s11336-005-1410-y

- State what are the most relevant approaches used in literature?
- Which scores and metrics have been used?
- How were test data collected?
- How did the AI system perform and how did it compare the current gold standard? Is the performance of the AI system equal across less represented groups? Can it be compared to other systems with a similar benchmarking performance and the same clinically meaningful endpoint (addressing comparative efficacy)?
- How can the utility of the AI system be evaluated in a real-life clinical environment (also considering specific requirements, e.g., in a low- and middle-income country setting)?
- Have there been clinical evaluation attempts (e.g., internal and external validation processes) and considerations about the use in trial settings?
- What are the most relevant gaps in the literature (what is missing concerning AI benchmarking)?

5.2 Benchmarking by AI developers

All developers of AI solutions for traditional medicine implemented internal benchmarking systems for assessing the performance. This clause will outline the insights and learnings from this work of relevance for benchmarking in this topic group. The following items are for further study:

- What are the most relevant learnings from the benchmarking by AI developers in this field (e.g., ask the members of your topic group what they want to share on their benchmarking experiences)?
- Which scores and metrics have been used?
- How did they approach the acquisition of test data?

5.3 Relevant existing benchmarking frameworks

Triggered by the hype around AI, recent years have seen the development of a variety of benchmarking platforms where AIs can compete for the best performance on a determined dataset. Given the high complexity of implementing a new benchmarking platform, the preferred solution is to use an established one. This clause reflects on the different existing options that are relevant for this topic group and includes considerations of using the assessment platform that is currently developed by FG-AI4H and presented by deliverable DEL7.5 *"FG-AI4H assessment platform"* (the deliverable explores options for implementing an assessment platform that can be used to evaluate AI for health for the different topic groups). The following items are for further study:

- Which benchmarking platforms could be used for this topic group (e.g., EvalAI, AIcrowd, Kaggle, and CodaLab)?
- Are the benchmarking assessment platforms discussed, used, or endorsed by FG-AI4H an option?
- Are there important features in this topic group that require special attention?
- Is the reporting flexible enough to answer the questions stakeholders want to get answered by the benchmarking?
- What are the relative advantages and disadvantages of these diverse solutions?

6 Benchmarking by the topic group

This clause describes all technical and operational details regarding the benchmarking process for the traditional medicine AI task including subclauses for each version of the benchmarking that is iteratively improved over time.

It reflects the considerations of various deliverables: DEL5 *"Data specification"* (introduction to deliverables 5.1-5.6), DEL5.1 *"Data requirements"* (which lists acceptance criteria for data submitted to FG-AI4H and states the governing principles and rules), DEL5.2 *"Data acquisition"*, DEL5.3 *"Data annotation specification"*, DEL5.4 *"Training and test data specification"* (which provides a systematic way of preparing technical requirement specifications for datasets used in training and testing of AI models), DEL5.5 *"Data handling"* (which outlines how data will be handled once they are accepted), DEL5.6 *"Data sharing practices"* (which provides an overview of the existing best practices for sharing health-related data based on distributed and federated environments, including the requirement to enable secure data sharing and addressing issues of data governance), DEL06 *"AI training best practices specification"* (which reviews best practices for proper AI model training and guidelines for model reporting), DEL7 *"AI for health evaluation considerations"* (which discusses the validation and evaluation of AI for health models, and considers requirements for a benchmarking platform), DEL7.1 *"AI4H evaluation process description"* (which provides an overview of the state of the art of AI evaluation principles and methods and serves as an initiator for the evaluation process of AI for health), DEL7.2 *"AI technical test specification"* (which specifies how an AI can and should be tested *in silico*), DEL7.3 *"Data and artificial intelligence assessment methods (DAISAM)"* (which provides the reference collection of WG-DAISAM on assessment methods of data and AI quality evaluation), DEL7.4 *"Clinical Evaluation of AI for health"* (which outlines the current best practices and outstanding issues related to clinical evaluation of AI models for health), DEL7.5 *"FG-AI4H assessment platform"* (which explores assessment platform options that can be used to evaluate AI for health for the different topic groups), DEL9 *"AI for health applications and platforms"* (which introduces specific considerations of the benchmarking of mobile- and cloud-based AI applications in health), DEL9.1 *"Mobile based AI applications,"* and DEL9.2 *"Cloud-based AI applications"* (which describe specific requirements for the development, testing and benchmarking of mobile- and cloud-based AI applications).

The benchmarking of traditional medicine is going to be developed and improved continuously to reflect new features of AI systems or changed requirements for benchmarking. This clause outlines all benchmarking versions that have been implemented thus far and the rationale behind them. It serves as an introduction to the subsequent sections, where the actual benchmarking methodology for each version will be described. The following items are for further study:

- Which benchmarking iterations have been implemented thus far?
- What important new features are introduced with each iteration?
- What are the next planned iterations and which features are they going to add?

6.1 Premise and requirement of developing benchmarks by the topic group

The Beijing Declaration[86], which was the main outcome of the first WHO Congress on Traditional Medicine held on the sixty-fifth anniversary of the WHO and the thirty-fifth anniversary of the Alma-Ata Declaration[87], offered international recognition of the role of traditional medicine in health care for the first time. The Declaration reaffirmed the need to regard these practises as something that "should be respected, preserved, promoted, and communicated widely and appropriately based on the circumstances in each country" and the responsibility of individual governments to "ensure appropriate, safe, and effective use of traditional medicine" through appropriate policies, regulations, and standards.

Ensuring the safety, efficacy, and quality of traditional medicine is of utmost importance to guarantee the sustainable provision of reliable, safe, and effective traditional medical practices to the public.

[86] WHO (2008a) Beijing Declaration, Geneva: World Health Organization. Available at http://www.who.int/medicines/areas/traditional/congress/beijing_declaration/en/index.html

[87] WHO (1978) Declaration of Alma-Ata, Geneva: World Health Organization. Available at http://www.who.int/publications/almaata_declaration_en.pdf

Setting up methods for appropriately analysing these practises for pragmatic utilization of emerging technologies and regulation is critical. The requisite two key aspects for benchmarking in this direction are as follows:

a) To provide safe and effective traditional medicine modalities for prevention and treatment

b) Prevention of biopiracy and sustainable use of resources.

6.1.1 To provide safe and effective traditional medicine treatment modalities

Traditional medicine treatments are classified into three types based on the methods employed to heal and/or maintain health:

1) *Pharmacological therapies* using herbal medicines and/or medicines derived from animal parts and/or minerals, bio-cleansing therapy (panchakarma);

2) *Non-pharmacological therapies* using manual (e.g., massage, abhyang), physical (e.g., qigong, t'ai chi ch'uan, cupping therapy, manipulation of vulnerable locations, acupuncture), mental (e.g., meditation, hypnosis), and spiritual (e.g., religious-magic) methods of treatment or a combination of these methods (yoga), diet and lifestyle recommendations, Counselling, etc.

3) *Mixed therapies* combining pharmacological and non-pharmacological therapies.

Effective treatment is frequently the outcome of a synergy of pharmacological and non-pharmacological methods of therapy. As a result, the success of traditional medicine must be evaluated holistically, considering both types of therapies. As a result, traditional medicine's evaluation and benchmarking of the technologies may differ significantly from that of contemporary medicine. It is therefore critical that such evaluation take place through an ongoing interaction between traditional practitioners, traditional medicine researchers, scientific specialists, and representatives of the cultures involved. It is challenging to translate indigenous notions of illness into biological terms, but this would be the greatest approach to promote constructive discussion between traditional and contemporary medicine and give more meaning to traditional therapies. Additionally, it would open up new avenues for debate and action, including health care service efforts for indigenous people and the bioactivity testing of medicinal plants, animal products and minerals used in such context. This can be achieved through sharing the responsibility without imposing cultural frameworks and practises.

6.1.2 Building trust through AI based neutral networks

A common reason for mistrust and suspicion is the vagueness of definitions and rules. Online users have easy access to treatments that are approved and covered by national health systems in some nations but are prohibited in others. They can also easily find herbal products that can be used as either medications, dietary supplements, or even just foods. This is a complex problem which can be addressed through digital health technologies such as AI. To begin with, it would be a good idea to set up a worldwide forum for exchanging experience and deciding on protocols, as well as a globally updated database. It must be emphasised as well that despite its long history, this field of medicine must be allowed to participate in innovative work and research. As a field of medicine, traditional medicine must be considered seriously; in this regard, capacity-building presents additional difficulties.

The applications of AI in the clinical domain of TM will facilitate the following:

1) Promotion of a wellness centric healthcare approach by employing personality-based assessment of individuals and adoption of TM based preventive health practices.

2) Use of an evidence-based TM in an integrative set up for the prevention and treatment of non-communicable diseases viz., obesity, heart, respiratory ailments, joint disorders, psychological illness, cancer, etc.

3) Effective use of TM based medicines in tackling pandemics and other emerging communicable disorders.

6.1.3 Prevention of biopiracy and sustainable use of resources

Biopiracy, in Pat Mooney's definition, is "the use of intellectual property systems to legitimize the exclusive ownership and control of biological resources and knowledge, without recognition, compensation or protection for contributions from indigenous and rural communities".

The potential for traditional practises to improve some medical diseases presents additional dangers of exploitation, on top of discriminatory practises and legislation. Although advanced technology has enabled rapid and large-scale improvements in traditional knowledge of medicinal plants, indigenous peoples who own this knowledge have no access to such technology and, in most cases, no recognition or share in benefits and income is granted to communities whose practitioners have been using these plants for centuries and may be offered a patented product in exchange for payment.

The majority of indigenous cultures hold the view that the world and all of its resources are an integral component of life and cannot, therefore, belong to any one person. Some groups even hold the view that they are the stewards of this land and all living creatures. They are consequently completely unaware of the current idea of a patent on biological items (plants and by-products). These populations could also be marginalised because of poverty, a lack of access to the dominant languages, and a lack of knowledge about local laws and alternatives for international relief. Additionally, they frequently lack organisation and political clout. As a result, a researcher can acquire these items without authorisation, with "bought" consent, or under duress, patent them, and deprive local people of their expertise.

Another important issue that has arisen because of businesses' revived interest in medicinal plants is the potential of raw materials needed to create medications and other natural health products being stolen. If the situation is not monitored and controlled, endangered species may become extinct, as well as natural ecosystems and resources. Furthermore, industrial production of these species would offer additional safety concerns that would need to be examined.

These problems can be handled diligently by using the technology, the best example in this direction is the "Traditional Knowledge Digital Library (TKDL)[88] an initiative of the Government of India. TKDL has extensively digitized text-based formulations, of Ayurveda, Siddha, Unani, Sowa-Rigpa, practices of Yoga" and made the data available to leading patent offices with a multilingual option. This has ensured the non-granting of patents which have documented proof of prior art. There is a need to develop such digital repositories for other indigenous TM practices. Topped up with geo tagging, natural language processing (NLP) based on ethnic languages of TM and AI tools will be game changers in ensuring optimum utilization of TM, avoidance of biopiracy and sustainable management of resources.

The ethical ramifications of using any technological innovation, including artificial intelligence (AI), in conventional medicine will be closely related to UNESCO's goal of promoting cultural diversity. The parameters that should be used to understand and address traditional medicine are defined by the Convention on Biological Diversity (1992), the UNESCO Universal Declaration on Cultural Diversity (2001)[89], the Convention for the Safeguarding of the Intangible Cultural Heritage (2003)[90],

---

[88] TKDLTraditional Knowledge Digital Library. (n.d.). Retrieved March 28, 2019, from http://www.tkdl.res.in/tkdl/langdefault/common/Home.asp?GL=Eng

[89] UNESCO (2001) UNESCO Universal Declaration on Cultural Diversity, Paris: United Nations Educational, Scientific and Cultural Organization. Available at http://portal.unesco.org/en/ev.php-URL_ID=13179&URL_DO=DO_TOPIC&URL_SECTION=201.html

[90] UNESCO (2003) Convention for the Safeguarding of Intangible Cultural Heritage, Paris: United Nations Educational, Scientific and Cultural Organization. Available at http://www.unesco.org/culture/ich/index.php?lg=en&pg=00022

and the United Nations Declaration on the Rights of Indigenous Peoples (2007)[91]. This is only possible if two divergent principles from the Universal Declaration on Bioethics and Human Rights, adopted by acclamation by the UNESCO General Conference (2005)[92], are taken into account: on the one hand, the right of every human being to "the highest attainable standard of health" (Art. 14); on the other, the explicit need to respect "cultural diversity and pluralism" (Art. 12), which includes "respect for traditional knowledge" (Art. 17).

These two ideals must be upheld concurrently and with equal vigour. However, tensions may occur throughout their use; in certain circumstances, defining priorities amongst them is inevitable. Geographic diversity and a wide range of practices are both advantages and disadvantages. The notion and practise of traditional medicine exist in many situations, making it extremely difficult to have a cohesive approach and terminology on the topic. Despite this complexity, because of the importance of traditional medicine in underdeveloped countries and its rapid rise in more industrialised nations, albeit with differing features and roles, the broadest possible approach should be advocated. This strategy necessitates not only a dedication to reiterating the critical role of traditional knowledge across the world but also the ability to give certain critical rules to safeguard users and minimise any potential risk of prejudice and exploitation.

The six guiding principles as the basis for AI regulation and governance proposed by W.H.O hold good for benchmarking the development of AI tools for traditional medicine, these are as follows[93]:

1) Protecting human autonomy: In the context of health care, this means that humans should remain in control of health-care systems and medical decisions; privacy and confidentiality should be protected, and patients must give valid informed consent through appropriate legal frameworks for data protection.
2) Promoting human well-being and safety and the public interest: The designers of AI technologies should satisfy regulatory requirements for safety, accuracy and efficacy for well-defined use cases or indications. Measures of quality control in practice and quality improvement in the use of AI must be available.
3) Ensuring transparency, explainability and intelligibility: Transparency requires that sufficient information be published or documented before the design or deployment of an AI technology. Such information must be easily accessible and facilitate meaningful public consultation and debate on how the technology is designed and how it should or should not be used.
4) Fostering responsibility and accountability: Although AI technologies perform specific tasks, it is the responsibility of stakeholders to ensure that they are used under appropriate conditions and by appropriately trained people. Effective mechanisms should be available for questioning and for redress for individuals and groups that are adversely affected by decisions based on algorithms (validity).
5) Ensuring inclusiveness and equity: Inclusiveness requires that AI for health be designed to encourage the widest possible equitable use and access, irrespective of age, sex, gender, income, race, ethnicity, sexual orientation, ability, or other characteristics protected under human rights codes.

---

[91] UN (2007) United Nations Declaration on the Rights of Indigenous Peoples, New York: United Nations. Available at http://social.un.org/index/IndigenousPeoples/DeclarationontheRightsofIndigenousPeoples.aspx; http://www.un.org/esa/socdev/unpfii/documents/DRIPS_en.pdf

[92] UNESCO (2005) Universal Declaration on Bioethics and Human Rights, Paris: United Nations Educational, Scientific and Cultural Organization. Available at http://portal.unesco.org/en/ev.php-URL_ID=31058&URL_DO=DO_TOPIC&URL_SECTION=201.html

[93] WHO issues first global report on Artificial Intelligence (AI) in health and six guiding principles for its design and use. https://www.who.int/news/item/28-06-2021-who-issues-first-global-report-on-ai-in-health-and-six-guiding-principles-for-its-design-and-use

6) Promoting AI that is responsive and sustainable: Designers, developers and users should continuously and transparently assess AI applications during actual use to determine whether AI responds adequately and appropriately to expectations and requirements. AI systems should also be designed to minimize their environmental consequences and increase energy efficiency. Governments and companies should address anticipated disruptions in the workplace, including training for health-care workers to adapt to the use of AI systems, and potential job losses due to the use of automated systems.

6.2 Benchmarking version [0.1]

This clause includes all technological and operational details of the benchmarking process for the benchmarking version [0.1] (latest version, chronologically reversed order).

6.2.1 Overview

This clause provides an overview of the key aspects of this benchmarking iteration, version [0.1]. The following items are for further study:

- What is the overall scope of this benchmarking iteration (e.g., performing a first benchmarking, adding benchmarking for multi-morbidity, or introducing synthetic-data-based robustness scoring)?
- What features have been added to the benchmarking in this iteration?

The primary objective of benchmarking is to ensure quality, safety, and effectiveness of various healthcare items such as medicines, vaccines, medical devices, diagnostics, and traditional or herbal medicines. However, most regulators worldwide are still facing challenges in attaining a state of maturity where they possess a stable, well-operating, and integrated regulatory framework. Through benchmarking TM, Member countries can participate together with the health community to develop effective and feasible institutional improvement plans to incorporate AI in the domain of TM.

Short-run	• Ensuring safety and efficacy of TM modalities prioritizing the establishment of benchmarking frameworks for prevention and treatment. • Establishing benchmarks for education, training, and research to enhance the quality and formative standards of TM practitioners, educators, and researchers.
Mid-run	Strengthening international governance and recognition and enhancing international cooperation in TM. This included promoting mutual recognition of TM practices, shared best practices and knowledge, and strengthening partnerships.
Long-run	Promoting seamless and pragmatic standardization of TM that are context specific and streamline national and international efforts to establish harmonized standards, guidelines and protocols.

Engaging the health community with TM

Involving TM and AI and technology communities in promoting health and wellbeing is to bring about unquestionable advantages facilitating transformations in behaviours, environments, policies, programs, and practices at the interplay of those communities. Although the extent and scope of community engagement may vary, it is critical for governments to promote these spaces and measures in order to promote familiarity and cultural acceptability among both AI and TM communities.

Short-run	• Promoting familiarity and cultural acceptability of TM among the health community. This included initiatives to raise awareness, foster cross-cultural exchanges of TM practices, promote dialogue and collaboration between TM and modern healthcare professionals, and improve cultural competency for enhanced patient-centred care. Creating adequate space through frameworks for TM practice to gain recognition, professional acceptance and enable TM practitioners to contribute to the public good.

<table>
<tr><td>Mid-run</td><td rowspan="2">Ensuring availability and proximity of TM at the local level by including measures to address workforce shortages, promote training and retention of TM practitioners, and facilitate the integration of TM services within the primary healthcare system.</td></tr>
<tr><td>Long-run</td></tr>
</table>

Improve recognition of TM as part of health care delivery

To maximize the impact of TM on the attainment of One Health and sustainable development, it is crucial to acknowledge TM's significance, practices, and its respectful approach towards local traditions, resources, and rights. Moreover, enhancing this recognition fosters the generation of evidence and knowledge, facilitates effective utilization of data and analytics, and encourages innovation and technology advancements, all of which serve to optimize TM's contribution to promote health equity.

Short-run	Building robust evidence on TM practices and procedures to inform clinical guidelines and support informed decision-making incorporating TM into healthcare systems.
Mid-run	Ensuring affordability of medicines and preventive strategies by including measures of financial accessibility such as subsidies, or insurance coverage for equitable access of TM treatments and therapies.
Long-run	Promoting research and evidence on effective treatment of specific disorders. Allocating resources to studies and clinical trials that explore the efficacy, safety, and mechanisms of action of TM interventions.

Protect the uniqueness of TM in relation to the patients and the environment

TM forms an essential part of the fabric of local communities, contributing to their distinct identity. It permeates both their social and physical surroundings, underscoring the criticality of safeguarding it. Any efforts to exploit TM for industrial or commercial gain can result in its inappropriate use and harm the rightful guardians of this knowledge. The preservation, protection, and promotion of AI-based innovations in TM and practices hold exceptional significance, particularly for developing countries.

<table>
<tr><td>Short-run</td><td>Develop guidelines for holistic and person-centred approach to healthcare, emphasizing the integration of TM practices and recognizing the individuality and cultural context of patients.</td></tr>
<tr><td>Mid-run</td><td rowspan="2">• Prioritize the protection of biodiversity and ecosystems in which TM thrive, involving conservation efforts, sustainable standards of TM practice, and environmental regulations to ensure long-term availability of resources and ingredients.
Target biopiracy by developing measures to safeguard TM knowledge, intellectual property rights, and the equitable sharing of benefits derived from TM resources.</td></tr>
<tr><td>Long-run</td></tr>
</table>

Adoption of preparatory steps to incorporate AI in a safe and ethical manner

TM has displayed great enthusiasm in delving into the possibilities presented by emerging artificial intelligence (AI) technologies. The utilization of AI-based techniques, including machine learning, and deep learning, among other methods enables and expands new fields in drug development, symptoms, disease treatment, as well as facilitating the screening of patients and key components uncovering mechanisms of action.

Short-run	• Adoption of AI in TM through national and international collaboration prioritizing the establishment of platforms for knowledge sharing, capacity building and collaborative research.

	Design and integrate regulatory frameworks that guide the integration of AI in TM, including regulatory measures such as data privacy, transparency, accountability and the informed consent of patients in the context of TM.
Mid-run	Provide context- and domain-specific data literacy to enhance algorithmic and personal data-rights awareness among TM practitioners, patients and healthcare stakeholders.
Long-run	Assessment of monitoring of research on TM employing digital health instruments, including the evaluation of the efficacy, safety and ethical implications of AI-driven interventions in this domain.

6.2.2 Benchmarking methods

This clause provides details about the methods of the benchmarking version [0.1]. It contains detailed information about the benchmarking system architecture, the dataflow and the software for the benchmarking process (e.g., test scenarios, data sources, and legalities).

After elaborate discussion, it has been agreed upon to examine benchmarking systems proposed by other topic groups viz., TG-Symptom, Cardiology, etc., and adopt best practices like **TG-Symptom Minimal Viable Benchmarking (MMVB)**[94] and such others as are relevant for the TG-TM.

Further, it is also agreed upon to follow the Benchmarking documents developed for traditional medicine by WHO and other such organizations for streamlining the benchmarking process. Some of the examples are as follows:

1) **Global benchmarking tool (GBT)** for national regulatory authorities (NRAs) who are supply chains of medical products, and have a mandate to ensure the quality, safety and efficacy of medicines, vaccines, blood and blood products, and medical devices, including diagnostics and traditional, or herbal medicines[95].
2) WHO Benchmarking documents for training-TCM[96], Ayurveda[97], Unani[98], etc.
3) WHO international standard terminologies on Ayurveda[99], Siddha[100], Unani[101], traditional Chinese medicine[102]

6.2.2.1 Benchmarking system architecture

This clause covers the architecture of the benchmarking system. For well-known systems, an overview and reference to the manufacturer of the platform is sufficient. If the platform was developed by the topic group, a more detailed description of the system architecture is required. The following items are for further study:

- How does the architecture look?

94 https://github.com/FG-AI4H-TG-Symptom

95 Khadem Broojerdi, Alireza et al. "The World Health Organization Global Benchmarking Tool an Instrument to Strengthen Medical Products Regulation and Promote Universal Health Coverage." Frontiers in medicine vol. 7 457. 19 Aug. 2020, doi:10.3389/fmed.2020.00457

96 Benchmarks for training in traditional / complementary and alternative medicine: benchmarks for training in traditional Chinese medicine. (n.d.). Retrieved September 2, 2023, from https://apps.who.int/iris/handle/10665/44353

97 WHO benchmarks for the training of ayurveda . (n.d.). Retrieved September 2, 2023, from https://www.who.int/publications/i/item/9789240042711

98 World Health Organisation. (2022). Who benchmarks for Training of Unani medicine practitioners.

99 https://www.who.int/publications/i/item/9789240064935

100 https://www.who.int/publications/i/item/9789240064973

101 https://www.who.int/publications/i/item/9789240064959

102 https://www.who.int/publications/i/item/9789240042322

- What are the most relevant components and what are they doing?
- How do the components interact on a high level?
- What underlying technologies and frameworks have been used?
- How does the hosted AI model get the required environment to execute correctly? What is the technology used (e.g., Docker/Kubernetes)?

6.2.2.2 Benchmarking system dataflow

This clause describes the dataflow throughout the benchmarking architecture. The following items are for further study:

- How do benchmarking data access the system?
- Where and how (data format) are the data, the responses, and reports of the system stored?
- How are the inputs and the expected outputs separated?
- How are the data sent to the AI systems?
- Are the data entries versioned?
- How does the lifecycle for the data look?

6.2.2.3 Safe and secure system operation and hosting

This clause addresses security considerations about the storage and hosting of data (benchmarking results and reports) and safety precautions for data manipulation, data leakage, or data loss.

In the case of a manufactured data source (vs. self-generated data), it is possible to refer to the manufacturer's prescriptions. The following items are for further study:

- Based on the architecture, where is the benchmarking vulnerable to risk and how have these risks been mitigated (e.g., did you use a threat modelling approach)? A discussion could include:
  - Could someone access the benchmarking data before the actual benchmarking process to gain an advantage?
  - What safety control measures were taken to manage risks to the operating environment?
  - Could someone have changed the AI results stored in the database (your own and/or that of competitors)?
  - Could someone attack the connection between the benchmarking and the AI (e.g., to make the benchmarking result look worse)?
  - How is the hosting system itself protected against attacks?
- How are the data protected against data loss (e.g., what is the backup strategy)?
- What mechanisms are in place to ensure that proprietary AI models, algorithms and trade-secrets of benchmarking participants are fully protected?
- How is it ensured that the correct version of the benchmarking software and the AIs are tested?
- How are automatic updates conducted (e.g., of the operating system)?
- How and where is the benchmarking hosted and who has access to the system and the data (e.g., virtual machines, storage, and computing resources, configurational settings)?
- How is the system's stability monitored during benchmarking and how are attacks or issues detected?
- How are issues (e.g., with a certain AI) documented or logged?
- In case of offline benchmarking, how are the submitted AIs protected against leakage of intellectual property?

6.2.2.4 Benchmarking process

This clause describes how the benchmarking looks from the registration of participants, through the execution and resolution of conflicts, to the final publication of the results. The following items are for further study:

- How are new benchmarking iterations scheduled (e.g., on demand or quarterly)?
- How do possible participants learn about an upcoming benchmarking?
- How can one apply for participation?
- What information and metadata do participants have to provide (e.g., AI autonomy level assignment (IMDRF), certifications, AI/machine learning technology used, company size, company location)?
- Are there any contracts or legal documents to be signed?
- Are there inclusion or exclusion criteria to be considered?
- How do participants learn about the interface they will implement for the benchmarking (e.g., input and output format specification and application program interface endpoint specification)?
- How can participants test their interface (e.g., is there a test dataset in case of file-based offline benchmarking or are there tools for dry runs with synthetic data cloud-hosted application program interface endpoints)?
- Who is going to execute the benchmarking and how is it ensured that there are no conflicts of interest?
- If there are problems with an AI, how are problems resolved (e.g., are participants informed offline that their AI fails to allow them to update their AI until it works? Or, for online benchmarking, is the benchmarking paused? Are there timeouts?)?
- How and when will the results be published (e.g., always or anonymized unless there is consent)? With or without seeing the results first? Is there an interactive drill-down tool or a static leader board? Is there a mechanism to only share the results with stakeholders approved by the AI provider as in a credit check scenario?
- In case of online benchmarking, are the benchmarking data published after the benchmarking? Is there a mechanism for collecting feedback or complaints about the data? Is there a mechanism of how the results are updated if an error was found in the benchmarking data?

6.2.3 AI input data structure for the benchmarking

This clause describes the input data provided to the AI solutions as part of the benchmarking of traditional medicine. It covers the details of the data format and coding at the level of detail needed to submit an AI for benchmarking. This is the only TDD section addressing this topic. Therefore, the description needs to be complete and precise. This clause does *not* contain the encoding of the labels for the expected outcomes. It is only about the data the AI system will see as part of the benchmarking. The following items are for further study:

- What are the general data types that are fed in the AI model?
- How exactly are they encoded? For instance, discuss:
  - The exact data format with all fields and metadata (including examples or links to examples)
  - Ontologies and terminologies
  - Resolution and data value ranges (e.g., sizes, resolutions, and compressions)
  - Data size and data dimensionality

6.2.4 AI output data structure

Similar to the input data structure for the benchmarking, this clause describes the output data the AI systems are expected to generate in response to the input data. It covers the details of the data format, coding, and error handling at the level of detail needed for an AI to participate in the benchmarking. The following items are for further study:

- What are the general data output types returned by the AI and what is the nature of the output (e.g., classification, detection, segmentation, or prediction)?
  - – How exactly are they encoded? Discuss points like:
    - The exact data format with all fields and metadata (including examples or links to examples)
    - Ontologies and terminologies
- What types of errors should the AI generate if something is defective?

6.2.5 Test data label/annotation structure

While the AI systems can only receive the input data described in the previous sections, the benchmarking system needs to know the expected correct answer (sometimes called 'labels') for each element of the input data so that it can compare the expected AI output with the actual one. Since this is only needed for benchmarking, it is encoded separately. The details are described in the following section. The following items are for further study:

- What are the general label types (e.g., expected results, acceptable results, correct results, and impossible results)?
- How exactly are they encoded? Discuss points like:
  - – The exact data format with all fields and metadata (including examples or links to examples)
  - – Ontologies and terminologies
- How are additional metadata about labelling encoded (e.g., author, data, pre-reviewing details, dates, and tools)?
- How and where are the labels embedded in the input data set (including an example; e.g., are there separate files or is it an embedded section in the input data that is removed before sending to the AI)?

6.2.6 Scores and metrics

Scores and metrics are at the core of the benchmarking. This clause describes the scores and metrics used to measure the performance, robustness, and general characteristics of the submitted AI systems. The following items are for further study:

- Who are the stakeholders and what decisions should be supported by the scores and metrics of the benchmarking?
- What general criteria have been applied for selecting scores and metrics?
- What scores and metrics have been chosen/defined for robustness?
- What scores and metrics have been chosen/defined for medical performance?
- What scores and metrics have been chosen/defined for non-medical performance?
  - – Metrics for technical performance tracking (e.g., monitoring and reporting when the performance accuracy of the model drops below a predefined threshold level as a function of time; computational efficiency rating, response times, memory consumption)
- What scores and metrics have been chosen/defined for model explainability?
- Describe for each aspect

- The exact definition/formula of the score based on the labels and the AI output data structures defined in the previous sections and how they are aggregated/accumulated over the whole dataset (e.g., for a single test set entry, the result might be the probability of the expected correct class which is then aggregated to the average probability of the correct class)
- Does it use some kind of approach for correcting dataset bias (e.g., the test dataset usually has a different distribution compared to the distribution of a condition in a real-world scenario. For estimating the real-world performance, metrics need to compensate this difference.)
- What are the origins of these scores and metrics?
- Why were they chosen?
- What are the known advantages and disadvantages?
- How easily can the results be compared between or among AI solutions?
- Can the results from benchmarking iterations be easily compared or does it depend too much on the dataset (e.g., how reproducible are the results)?

• How does this consider the general guidance of WG-DAISAM in DEL7.3 "Data and artificial intelligence assessment methods (DAISAM)"?

• Have there been any relevant changes compared to previous benchmarking iterations? If so, why?

6.2.7 Test dataset acquisition

Test dataset acquisition includes a detailed description of the test dataset for the AI model and, in particular, its benchmarking procedure including quality control of the dataset, control mechanisms, data sources, and storage. The following items are for further study:

• How does the overall dataset acquisition and annotation process look?

• How have the data been collected/generated (e.g., external sources vs. a process organized by the TG)?

• Have the design goals for the benchmarking dataset been reached (e.g., please provide a discussion of the necessary size of the test dataset for relevant benchmarking results, statistical significance, and representativeness)?

• How was the dataset documented and which metadata were collected?
  - Where were the data acquired?
  - Were they collected in an ethical-conform way?
  - Which legal status exists (e.g., intellectual property, licenses, copyright, privacy laws, patient consent, and confidentiality)?
  - Do the data contain 'sensitive information' (e.g., socially, politically, or culturally sensitive information; personal identifiable information)? Are the data sufficiently anonymized?
  - What kind of data anonymization or deidentification has been applied?
  - Are the data self-contained (i.e., independent from externally linked datasets)?
  - How is the bias of the dataset documented (e.g., sampling or measurement bias, representation bias, or practitioner/labelling bias)?
  - What addition metadata were collected (e.g., for a subsequent detailed analysis that compares the performance on old cases with new cases)? How was the risk of benchmarking participants accessing the data?

- Have any scores, metrics, or tests been used to assess the quality of the dataset (e.g., quality control mechanisms in terms of data integrity, data completeness, and data bias)?
- Which inclusion and exclusion criteria for a given dataset have been applied (e.g., comprehensiveness, coverage of target demographic setting, or size of the dataset)?
- How was the data submission, collection, and handling organized from the technical and operational point of view (e.g., folder structures, file formats, technical metadata encoding, compression, encryption, and password exchange)?
- Specific data governance derived by the general data governance document (currently F-103 and the deliverables beginning with DEL5)
- How was the overall quality, coverage, and bias of the accumulated dataset assessed (e.g., if several datasets from several hospitals were merged with the goal to have better coverage of all regions and ethnicities)?
- Was any kind of post-processing applied to the data (e.g., data transformations, repackaging, or merging)?
- How was the annotation organized?
  - How many annotators/peer reviewers were engaged?
  - Which scores, metrics, and thresholds were used to assess the label quality and the need for an arbitration process?
  - How have inter-annotator disagreements been resolved (i.e., what was the arbitration process)?
  - If annotations were part of the submitted dataset, how was the quality of the annotations controlled?
  - How was the annotation of each case documented?
  - Were metadata on the annotation process included in the data (e.g., is it possible to compare the benchmarking performance based on the annotator agreement)?
- Were data/label update/amendment policies and/or criteria in place?
- How was access to test data controlled (e.g., to ensure that no one could access, manipulate, and/or leak data and data labels)? Please address authentication, authorization, monitoring, logging, and auditing
- How was data loss avoided (e.g., backups, recovery, and possibility for later reproduction of the results)?
- Is there assurance that the test dataset is undisclosed and was never previously used for training or testing of any AI model?
- What mechanisms are in place to ensure that test datasets are used only once for benchmarking? (Each benchmarking session will need to run with a new and previously undisclosed test dataset to ensure fairness and no data leakage to subsequent sessions)

6.2.8 Data sharing policies

This clause provides details about legalities in the context of benchmarking. Each dataset that is shared should be protected by special agreements or contracts that cover, for instance, the data sharing period, patient consent, and update procedure (see also DEL5.5 on *data handling* and DEL5.6 on *data sharing practices*). The following items are for further study:

- Which legal framework was used for data sharing?
- Was a data sharing contract signed and what was the content? Did it contain:
  - Purpose and intended use of data
  - Period of agreement

  - Description of data
  - Metadata registry
  - Data harmonization
  - Data update procedure
  - Data sharing scenarios
    - Data can be shared in public repositories
    - Data are stored in local private databases (e.g., hospitals)
  - Rules and regulation for patients' consent
  - Data anonymization and de-identification procedure
  - Roles and responsibilities
    - Data provider
    - Data protection officer
    - Data controllers
    - Data processors
    - Data receivers
- Which legal framework was used for sharing the AI?
- Was a contract signed and what was the content?

6.2.9 Baseline acquisition

The main purpose of benchmarking is to provide stakeholders with the numbers they need to decide whether AI models provide a viable solution for a given health problem in a designated context. To achieve this, the performance of the AI models needs to be compared with available options achieving the same clinically meaningful endpoint. This, in turn, requires data on the performance of the alternatives, ideally using the same benchmarking data. As the current alternatives typically involve doctors, it might make sense to combine the test data acquisition and labelling with additional tasks that allow the performance of the different types of health workers to be assessed. The following items are for further study:

- Does this topic require comparison of the AI model with a baseline (gold standard) so that stakeholders can make decisions?
- Is the baseline known for all relevant application contexts (e.g., region, subtask, sex, age group, and ethnicity)?
- Was a baseline assessed as part of the benchmarking?
- How was the process of collecting the baseline organized? If the data acquisition process was also used to assess the baseline, please describe additions made to the process described in the previous section.
- What are the actual numbers (e.g., for the performance of the different types of health workers doing the task)?

6.2.10 Reporting methodology

This clause discusses how the results of the benchmarking runs will be shared with the participants, stakeholders, and general public. The following items are for further study:

- What is the general approach for reporting results (e.g., leader board vs. drill down)?
- How can participants analyse their results (e.g., are there tools or are detailed results shared with them)?

- How are the participants and their AI models (e.g., versions of model, code, and configuration) identified?
- What additional metadata describing the AI models have been selected for reporting?
- How is the relationship between AI results, baselines, previous benchmarking iterations, and/or other benchmarking iterations communicated?
- What is the policy for sharing participant results (e.g., opt in or opt out)? Can participants share their results privately with their clients (e.g., as in a credit check scenario)?
- What is the publication strategy for the results (e.g., website, paper, and conferences)?
- Is there an online version of the results?
- Are there feedback channels through which participants can flag technical or medical issues (especially if the benchmarking data was published afterwards)?
- Are there any known limitations to the value, expressiveness, or interpretability of the reports?

6.2.11 Result

This clause gives an overview of the results from runs of this benchmarking version of your topic. Even if your topic group prefers an interactive drill-down rather than a leader board, pick some context of common interest to give some examples. The following items are for further study:

- When was the benchmarking executed?
- Who participated in the benchmarking?
- What overall performance of the AI systems concerning medical accuracy, robustness, and technical performance (minimum, maximum, average etc.) has been achieved?
- What are the results of this benchmarking iteration for the participants (who opted in to share their results)?

6.2.12 Discussion of the benchmarking

This clause discusses insights of this benchmarking iterations and provides details about the 'outcome' of the benchmarking process (e.g., giving an overview of the benchmark results and process). The following items are for further study:

- What was the general outcome of this benchmarking iteration?
- How does this compare to the goals for this benchmarking iteration (e.g., was there a focus on a new aspect to benchmark)?
- Are there real benchmarking results and interesting insights from this data?
  - How was the performance of the AI system compared to the baseline?
  - How was the performance of the AI system compared to other benchmarking initiatives (e.g., are the numbers plausible and consistent with clinical experience)?
  - How did the results change in comparison to the last benchmarking iteration?
- Are there any technical lessons?
  - Did the architecture, implementation, configuration, and hosting of the benchmarking system fulfil its objectives?
  - How was the performance and operational efficiency of the benchmarking itself (e.g., how long did it take to run the benchmarking for all AI models vs. one AI model; was the hardware sufficient)?
- Are there any lessons concerning data acquisition?
  - Was it possible to collect enough data?
  - Were the data as representative as needed and expected?

- How good was the quality of the benchmarking data (e.g., how much work went into conflict resolution)?
- Was it possible to find annotators?
- Was there any relevant feedback from the annotators?
- How long did it take to create the dataset?

- Is there any feedback from stakeholders about how the benchmarking helped them with decision-making?
  - Are metrics missing?
  - Do the stakeholders need different reports or additional metadata (e.g., do they need the "offline capability" included in the AI metadata so that they can have a report on the best offline system for a certain task)?
- Are there insights on the benchmarking process?
  - How was the interest in participation?
  - Are there reasons that someone could not join the benchmarking?
  - What was the feedback of participants on the benchmarking processes?
  - How did the participants learn about the benchmarking?

6.2.13 Retirement

This clause addresses what happens to the AI system and data after the benchmarking activity is completed. It might be desirable to keep the database for traceability and future use. Alternatively, there may be security or privacy reasons for deleting the data. Further details can be found in the reference document of this clause DEL4 "*AI software lifecycle specification"* (identification of standards and best practices that are relevant for the AI for health software life cycle). The following items are for further study:

- What happens with the data after the benchmarking (e.g., will they be deleted, stored for transparency, or published)?
- What happens to the submitted AI models after the benchmarking?
- Could the results be reproduced?
- Are there legal or compliance requirements to respond to data deletion requests?

7 Overall discussion of the benchmarking

This clause discusses the overall insights gained from benchmarking work in this topic group. This should not be confused with the discussion of the results of a concrete benchmarking run (e.g., in clause 6.2.12).

In this context it is pertinent to note that appropriate benchmarking for the use of artificial intelligence (AI) can play a significant role in supporting use of TM in integrative medicine by enhancing patient care, improving treatment outcomes, and facilitating personalized medicine in the following areas:

Key areas of promoting AI in traditional medicine

1) **Ontological, epistemological analysis of traditional medicine**: Natural language processing (NLP) approaches powered by AI can analyse enormous amounts of medical literature, research papers, and patient-generated data to extract useful information. This can help healthcare practitioners make evidence-based judgements regarding incorporating TM treatments into integrative medical practises.

   The pioneering work of WHO developing TM based benchmark documents viz. Ayurveda, naturopathy, Thai massage, osteopathy, traditional Chinese medicine, Tuina massage and

Unani medicine., standardized terminologies for TCM, Ayurveda, Siddha, Unani and inclusion of TM as part of ICD-11 is an exemplary effort in this direction.

2) **Analysing and integrating data**: Integrative medicine entails combining data from many sources, such as traditional medical records and alternative medicinal procedures. To find patterns, correlations, and treatment results, AI can help integrate and analyse this wide variety of data. It can assist in determining which treatment regiments may be most successful for specific individuals.

3) **Decision support**: AI systems can analyse enormous volumes of patient data, such as medical records, imaging findings, genetic information, and lifestyle variables, to help healthcare practitioners make decisions. This can aid in the development of personalised treatment programmes that take into account both conventional and unconventional medicines.

4) **Predictive analytics**: Based on patient characteristics, treatment plans, and historical data, AI may utilise prediction models to evaluate the likelihood of various treatment outcomes. In addition to helping to customise treatment plans for specific patients, this can assist in forecasting the possible advantages and hazards of integrative medicine techniques.

5) **Patient monitoring and feedback**: Health-related metrics for patients, such as vital signs, activity levels, sleep patterns, and adherence to treatment programmes, may be tracked by AI. Patients and healthcare professionals can receive real-time feedback from it, alerting them to any possible problems or recommending changes to the treatment plan.

6) **Virtual health assistants**: Virtual health assistants and chatbots powered by AI may enlighten patients on integrative medical practises, answer inquiries, and provide personalised recommendations. These tools can help patients make educated decisions and get access to the right integrative therapy.

7) **Research and knowledge discovery**: AI can aid in the discovery of new insights and potential treatment techniques in integrative medicine by discovering patterns and links within massive datasets. It can assist researchers in investigating the efficacy and safety of various complementary medicines, identifying predictive biomarkers, and optimising treatment regimens.

The following items are for further study:

- What is the overall outcome of the benchmarking thus far?
- Have there been important lessons?
- Are there any field implementation success stories?
- Are there any insights showing how the benchmarking results correspond to, for instance, clinical evaluation?
- Are there any insights showing the impact (e.g., health economic effects) of using AI systems that were selected based on the benchmarking?
- Was there any feedback from users of the AI system that provides insights on the effectiveness of benchmarking?
  - Did the AI system perform as predicted relative to the baselines?
  - Did other important factors prevent the use of the AI system despite a good benchmarking performance (e.g., usability, access, explainability, trust, and quality of service)?
- Were there instances of the benchmarking not meeting the expectations (or helping) the stakeholders? What was learned (and changed) as a result?

- What was learned from executing the benchmarking process and methodology (e.g., technical architecture, data acquisition, benchmarking process, benchmarking results, and legal/contractual framing)?

8 Regulatory considerations

For AI-based technologies in healthcare, regulation is not only crucial to ensure the safety of patients and users, but also to accomplish market acceptance of these devices. This is challenging because there is a lack of universally accepted regulatory policies and guidelines for AI-based medical devices. To ensure that the benchmarking procedures and validation principles of FG-AI4H are secure and relevant for regulators and other stakeholders, the working group on *"Regulatory considerations on AI for health" (WG-RC)* compiled the requirements that consider these challenges.

The deliverables with relevance for regulatory considerations are DEL2 *"AI4H regulatory considerations"* (which provides an educational overview of some key regulatory considerations), DEL2.1 *"Mapping of IMDRF essential principles to AI for health software"*, and DEL2.2 *"Guidelines for AI based medical device (AI-MD): Regulatory requirements"* (which provides a checklist to understand expectations of regulators, promotes step-by-step implementation of safety and effectiveness of AI-based medical devices, and compensates for the lack of a harmonized standard). DEL4 identifies standards and best practices that are relevant for the "*AI software lifecycle specification*." The following sections discuss how the different regulatory aspects relate to the TG-AI for traditional medicine.

The group members have opined that the regulatory concepts and framework detailed in the ITU DEL 02 "Overview of regulatory concepts on artificial intelligence for health (22 September 2022)[103]" holds good for traditional medicine. Apart from this the regulatory framework applicable for streamlining the education, practice, manufacturing and sale of medicines, equipment, etc., pertaining to TM in various countries viz., TCM (China, Korea, Japan), Ayush (India), etc., has to be taken into consideration for the benchmarking and regulation of AI solutions for the TM sector.

103 Anonymous. Overview of regulatory concepts on artificial intelligence for health (ITU-T FG-AI4H Deliverable 02) [Internet]. Geneva; 2022 [cited 2023 Jun 4]. Available from: https://www.itu.int/en/ITU-T/focusgroups/ai4h/Documents/del/DEL2-A20220922-%28Prepub%29.pdf

Annex A

Glossary

This clause lists all the relevant abbreviations, acronyms and uncommon terms used in the document.

Acronym/Term	Expansion	Comment
AI	Artificial Intelligence	
AI4H	Artificial Intelligence for Health	
AI-MD	AI based Medical Device	
API	Application Programming Interface	
CfTGP	Call for Topic Group Participation	
DEL	Deliverable	
FDA	Food and Drug Administration	
FGAI4H	Focus Group on AI for Health	
GDP	Gross Domestic Product	
GDPR	General Data Protection Regulation	
IMDRF	International Medical Device Regulators Forum	
IP	Intellectual Property	
ISO	International Organization for Standardization	
ITU	International Telecommunication Union	
LMIC	Low-and Middle-Income Countries	
MDR	Medical Device Regulation	
PII	Personal Identifiable Information	
SaMD	Software as a Medical Device	
TDD	Topic Description Document	
TG	Topic Group	
WG	Working Group	
WHO	World Health Organization	

Annex B

Declaration of conflicts of interest

In accordance with the ITU transparency rules, this section lists the conflict-of-interest declarations for everyone who contributed to this document. Please see the guidelines in FGAI4H-F-105 "ToRs for the WG-Experts and call for experts" and the respective forms (Application form & Conflict of interest form).

Institution

Ministry of Ayush, Government of India is a dedicated government body which is working in the domain of traditional medicine in India. In collaboration with World Health Organization and other concerned agencies, the Ministry is actively contributing towards development of standards, benchmarking for TM practices, education, training, and quality control of products for the benefit of humankind. As part of the same, this topic group has been created and will be led as a collaborative effort for developing benchmarking for utilization of AI/ML for TM the world over. It is hereby declared that, the Ministry and the topic driver who is coordinating this activity is a research officer working with the Ministry and has no commercial interests and conflict of interests out of this activity.

____