

Issue Brief

**AI and Trust in Healthcare**

**VIRTUAL DELIBERATIVE DIALOGUE  
SESSION SUMMARY**

---



November 2020

# AI AND TRUST

---

## VIRTUAL DELIBERATIVE DIALOGUE SESSION SUMMARY

### Introduction

Progress in the field of AI has resulted in several applications hitting markets across sectors, including healthcare. Currently, India has more than 3,225 digital health-tech start-ups; many of which have begun providing or integrating AI into their products and services.<sup>1</sup> Recent reports also suggest that investments in health-tech startups are rising steadily. There has been a significant leap - from the \$163 million investment in 2016 to the \$343 million recorded in 2017.<sup>2</sup> Both governmental-- NITI Aayog, Ministry of Electronics and IT (MeitY) as well as industrial bodies (FICCI) have called for a greater integration of AI in healthcare. The government has released a blueprint for the creation of a central repository of health data, the National Digital Health Blueprint (NDHB). The NDHB would not only incorporate AI in its functioning (through the health data analytics platform), but could also act as a foundational layer upon which other AI-based healthcare interventions can be built. Further, Covid-19 has spurred increased research and industry efforts towards the development and deployment of AI in healthcare.

As concerns around safety, reliability, privacy emerge on the heels of widespread AI adoption, attention is shifting to the readiness of institutions like governments, regulators and legislative bodies in protecting the rights and interests of users and beneficiaries as they (actively or passively) engage with AI-enabled technology. This is of critical concern to healthcare, which is regarded as one of the priority sectors for adoption of AI in public service delivery in India. The potential of AI in healthcare can be low-risk (wearable solutions that assess and optimize your workouts), as well as extend to more safety-critical use cases (detection of cancerous pathology), but its serious adoption will require transparent discourse between innovators, users, and institutions to ensure that innovation is not at the cost of safety and privacy.

---

<sup>1</sup> Bharadwaj, A. (2020, June 05). Health tech in India. Retrieved August 20, 2020, from <https://www.investindia.gov.in/team-india-blogs/health-tech-india>

<sup>2</sup> Sriram, M. (2018, November 11). Health-tech startup funding hits all-time high of \$510 million in 2018. Retrieved August 19, 2020, from <https://www.livemint.com/Companies/euHgMPBTiM6GxrTZm9Ly7K/Health-tech-startup-funding-hits-alltime-high-of-510-milli.html>

AI today, however, has a trust problem. This is due to the fact that many AI models, while highly accurate, often operate as ‘blackboxes’ that are inherently opaque, and consequently their inner workings remain obscure to human oversight. This not only limits their use and legitimacy in critical sectors such as healthcare but also raises justified societal concerns over the unforeseen and possible negative impacts of AI. The missing dimension of trustworthiness and transparency in the functioning of AI systems has led to a growing policy discourse on the need to make AI explainable and interpretable for the purpose of algorithmic audits as well as ensuring fairness and safety. However, depending on the techniques used and contexts in which AI is applied, interjecting mechanisms for explainability and transparency can also place competing demands on the level of accuracy attainable by AI systems.

In addition, explainability in and of itself may not ensure trustworthiness and reliability of systems. That is, knowing how a system works does not necessarily lend to knowing how to govern it. Wider institutional measures and societal capacities, such as regulatory oversight and iterative policy frameworks/sandboxes might be needed to ensure trustworthiness and accountability of AI. While AI can bring about immense benefits to the healthcare sector in India, such as filling gaps in research and development, and alleviating human resource and infrastructural burdens, in the absence of contextually rooted governance frameworks, it may cause more harm than good. Thus, given recent policy developments, and the growing deployment of AI for healthcare in India, it is important to understand how trust in AI is effectuated in India currently, and identify challenges and measures that can be taken to foster trust in AI.

On 27 July 2020, Tandem Research and International Innovation Corps convened a panel of 8 experts composed of academia, healthcare professionals, lawyers and industry professionals with experience in building AI in healthcare in India to explore the dynamics of trust in clinical and health settings in India and the impact of AI on these existing structures and relationships. The session also sought to elicit ideas for designing solutions and safeguards that can encourage responsibility and trust in AI in the healthcare ecosystem in India.

What transpired through the session is the understanding that engendering trust in AI cannot be solved through purely technological quick fixes. Instead, concerns regarding trust are intricately tied to context of use as well as current ideals, norms, and understanding of responsibility. Further, the development and deployment of AI solutions in healthcare requires greater policy conversations around issues of liability and accountability.

Below are a few key takeaways from the sessions.

## Key Takeaways

These key takeaways may also be read as a series of hypotheses that would require further analysis and investigation.

**1. The introduction of AI in healthcare systems has led to a diffusion of responsibility. This requires re-examining questions regarding accountability and liability.**

Unlike traditional doctor-patient relationships where liability is well understood, in the case of AI, there is a long chain of involvement in the act of diagnosis — including service providers, developers and manufacturers of AI systems — so the question of liability gets quite muddled. To solve this, we need detailed rules and dos and don'ts around the design and use of AI.

India has an extremely low doctor to patient ratio. In some instances, there could be only one radiologist handling about 6000 patients. Here, AI solutions can be used to augment certain capacities but would need to have clearly defined boundaries between the responsibilities of ML systems and the responsibilities of doctors and medical staff in order to build trust. However, at what point one draws the boundary is not clear.

Participants also argued that as administrators of such systems, clinicians too have a responsibility to ensure that the tools they use are for the benefit of the patient. This entails understanding the way such systems operate. In radiology, for example, this might involve inquiring into the nature of the methodology used in the sensors for detections, the kind of engineering tools employed, and the manner and rationale behind which inferences were drawn. The diffusion of responsibilities also widens its scope. That is, only if a clinician has a certain level of understanding and subsequently trust in the manner in which an AI system functions and the outcome it provides, will they be willing to take on a higher degree of responsibility that might be added by its use.

**2. In some use case scenarios, trust may not be the deciding factor in administering the tool.**

Depending on the context, clinicians or other healthcare professionals may not have the agency to decide whether to adopt AI products and services. For example, frontline health workers have very little autonomy to make decisions. In this system, stakeholder values are different, for instance, if ASHA workers don't use an AI application, it could result in not being paid. Other stakeholders are the ones to decide what applications and tools are used by field agents.

Thus, as conversations around trust in AI are context dependent, there is much difference between clinical and non-clinical settings, such as population health care. Stakeholders in the two cases, for instance, are largely different and so is the level of autonomy. As contexts are not generalisable, one of the key things developers need to keep in mind for deployment is the varying forms of workflow that accompany changing contexts and subsequently how and where in the workflow the tools fit in, and importantly how the tool affects the current standard of care. One way to build trust in these different contexts can be to introduce mechanisms that measure impact assessments and specify liability.

### **3. Much of building trust in AI depends on a disclosure regime.**

Building trust in AI systems requires disclosure and transparency at different levels and stages of the development and deployment of AI products and services. Not only does a clinician need to know why a particular algorithm produced a particular output, it is also important that there is enough publicly available information for end users and those likely to be impacted by AI to know whether and how algorithms have been tested and validated. Participants also highlighted the role of institutional mechanisms, e.g. mandate algorithmic disclosure and communication between developers and clinicians in building trust in a particular system.

One aspect of why institutional mechanisms such as standards and regulatory compliance specifications work is because disclosure functions are built into the institutional apparatus. Blackbox AI is unlikely to inspire trust, due to the fact that parts of its functioning is outside of human purview. Where disclosure may not be possible, stricter rules and assessments should follow.

### **4. It is difficult to speak of trust in an abstract form, and much of it depends on the context of its use and the design of the application.**

It is important to move away from thinking of AI as a monolithic system and instead consider how we can build it to be cognizant of practices happening on the ground. A participant remarked that the reason for the success of AI in radiology is owing to a better understanding of its role in that contained space. However, when employed in complex situations, it is necessary to account for the participation and agency of actors such as frontline health workers and doctors that are lower in the hierarchy.

One of the key things to think about while deploying is the clinical workflow, how the tools fit into it and how the AI affects the current standard of care. As mentioned earlier, it is important to know the conditions under which the tool will be used, for example, whether for a clinical or emergency room setting. Another aspect that requires

consideration is which region(s) the algorithm and the training sets were developed, as systems trained on data and/or tested successfully in one region may not always transpose well to other regions.

Engagement between developers of a tool and medical professionals intended to use them influence the way these tools are used in different clinical settings. Use and decision to rely on a particular tool may be influenced by prior knowledge/understanding. Furthermore, for developers interested in design that considers the different levels of digital literacy of patients, educating them on the decisions being made through AI is important.

**5. Institutional mechanisms such as certification are necessary to build trust but not sufficient and need to be accompanied by policy changes which prescribe transparency and accountability.**

Trust is an emergent quality of the ecosystem, and it is difficult to exogenously introduce it. However, it is possible to create an enabling environment for trust to emerge by designing the ecosystem and curating institutional priorities in a way that would allow for this to happen. Introducing standard-setting organisations and formulating standards for product design, management, and use based on accepted industry practices is an alternative approach to proactive or ex-ante legislation. It is cumbersome to integrate continuous developments in technology within legal frameworks. Overall participants felt that approval by regulatory bodies such as USFDA or CE could help improve the perception of the accuracy and reliability of these systems. However, it was also felt that foreign regulatory or standard setting bodies could be an exclusionary, with middle income countries often failing to get approval for their products. Further, depending on the needs of the industry, soft laws such as contractual agreements could organically emerge - between medical institutions and developers that necessitate meeting improved certification requirements set out by certification bodies.

Rights of the users either using or impacted by these systems are vital to creating an enabling environment for trust to emerge. It is necessary to consider whether there are mechanisms in place to contest inferences that are aided by AI or measures to challenge unilateral orders to implement such systems, as in the case of frontline health workers. In such a context, it is important to assess whether the Personal Data Protection Bill can be used to contest AI based decision-making and therefore engender trust, but also how this can be balanced against continued research and innovation needs to be thought out.

From a developer's point of view, the current means of establishing trust (in terms of reliability of the algorithm) is through safety and efficacy certifications from the USFDA and EU. Validation/pilot studies and published papers also help bolster the process. A possible policy change that would speak to building trust would be to introduce regulation that involves a wider set of medical devices including AI based ones. Further, use statements and studies conducted to obtain CE certificates could also be made to be publicly available on the websites of those using AI to increase transparency around the purposes of its use.

**6. Explainability is crucial for engendering trust, so are building community relationships and interpersonal trust.**

Participants noted how women health workers spent years building trust in a community, through their services to them. This is particularly important in the Indian context because a lot of people do not have trust in public health services. The point it drives home is that mere certification is not enough to build trust. Certification is a one time act whereas building trust is an iterative process that needs to remain dynamic.

**7. Engendering trust requires room for experimentation. At the same time, experimentation in sectors critical for safety such as healthcare could also prove costly.**

Experimentation with AI in healthcare involves very high stakes, first and foremost because the lives of individuals are in question. Additionally, since there is a greater degree of pressure in achieving accurate results, experimentation in healthcare tends to be time, labour and resource heavy. Ensuring results are as relevant as possible to each context could prove to be expensive for smaller sized medical facilities. If experimentation is accompanied by the overhaul of existing processes that are no longer beneficial, it will involve expensive capacity requirements in adopting newer process(es).

For example, it is difficult to have a standard set of indicators to evaluate whether a tool is suitable for doctors/clinicians. Suitability is dependent on various iterations and experimentation with the tool, and how the results look like with and without the tool. These results can vary depending on the case scenario. For example, a participant explained that turnaround time can be a good indicator of whether a tool fits well within that particular context and setting.



***About International Innovation Corps (IIC):*** *The International Innovation Corps (IIC) is a program of the University of Chicago Trust in India. With teams embedded in governmental ministries, departments at the Centre and State, IIC aims to address critical development challenges at scale. IIC's work on AI policy and innovation spans across the Ministries of Electronics and IT (MeitY) as well as Health & Family Welfare (MoHFW) where teams are exploring AI based innovation to support public service delivery, responsible policy for technological development, and the role of the government in driving this forward.*



***About Tandem Research:*** *Tandem Research is an interdisciplinary research collective that generates policy insights at the interface of technology, society, and sustainability. We believe that evidence-based policy, supported by broad-based public engagement, must steer technology and sustainability trajectories in India. Our work seeks to ensure that India's technology transitions are just and equitable*

Contact us at [hello@tandemresearch.org](mailto:hello@tandemresearch.org)