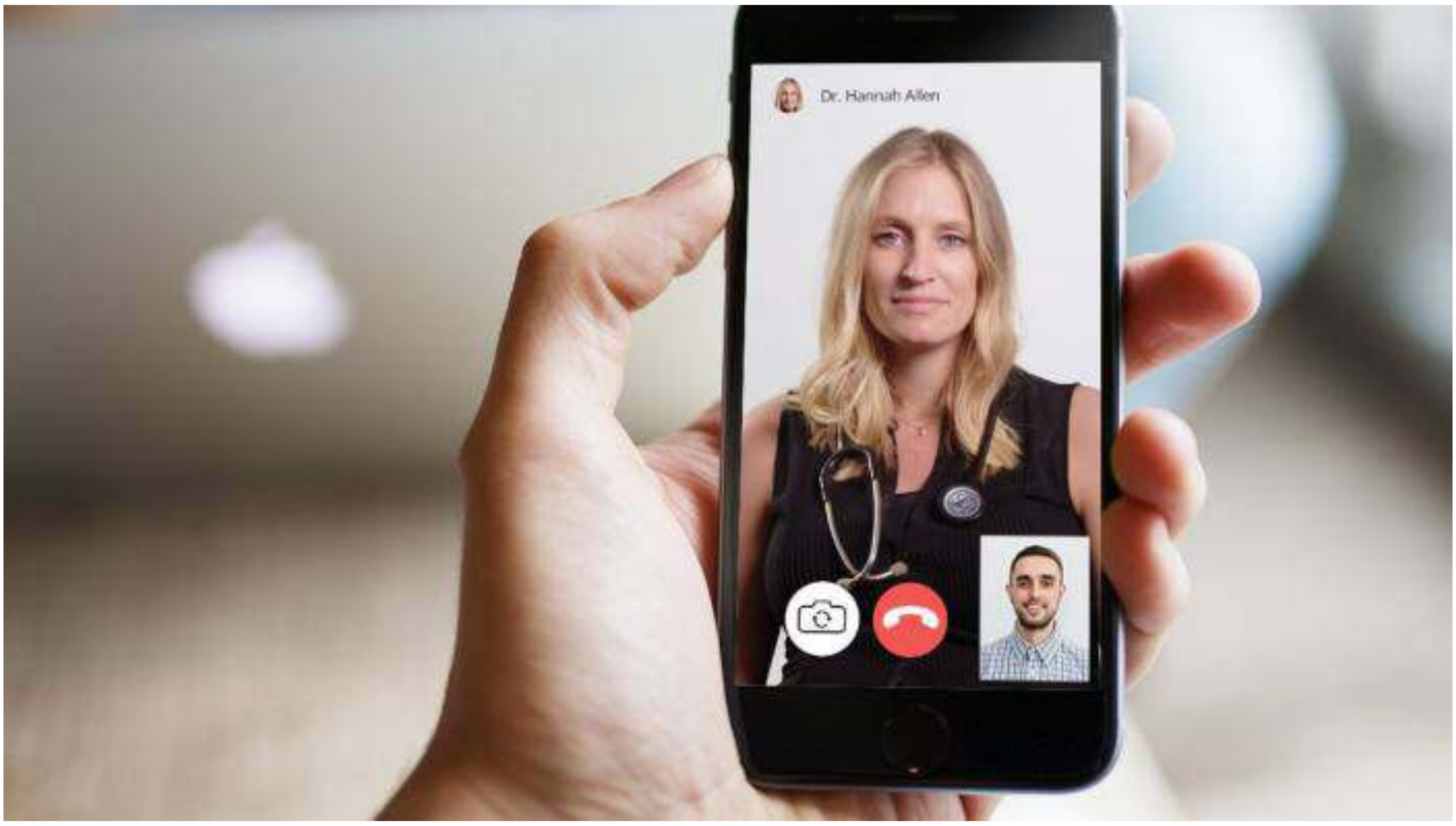
Can AI Discriminating against Patients: How Human-Centered Design can make Healthcare AI safer

Somil Gupta



Source: AI brings opportunity and risk to the health sector, Financial TImes, June 2019

According to the new paper, racial biases could inadvertently be built into healthcare algorithms if proper care is not taken while designing the system. Healthcare delivery models may also be skewed in the absence of sufficient genetic data points or to favor 'commercial benefits' instead of offering proper care. The big question is who should take accountability for such mishaps? Should it be the physician or the hospital that are using it? Or should it be the Healthcare AI system vendor? And the bigger question is how we can avoid the risks and liabilities of such biases when we know that they are inherently unavoidable.

As quoted in the paper by Implementing Machine Learning in Health Care — Addressing Ethical Challenges by Drs. Danton S. Char, Nigam H. Shah and David Magnus. The report was published by Financial TImes a few days ago.

“As clinical medicine moves progressively toward a shift-based model, the number of clinicians who have followed diseases from their presentation through their ultimate outcome is decreasing. This trend underscores the opportunity for machine learning and approaches based on artificial intelligence in health care — but it could also give such tools unintended power and authority. The collective medical mind is becoming the combination of published literature and the data captured in health care systems, as opposed to individual clinical experience. ”

What the researchers are highlighting in this paper is the future scenario when clinicians will become increasingly dependent upon the AI-based expert systems for not just diagnosis support but for advice on treatment. And then racial and other biases can play out in unpredictable and uncontrollable ways. As more and more clinicians depend upon the 'system', the lesser individual power any one physician will have to identify and correct the errors committed by AI. In the absence of this constant critical feedback, these errors and biases will only get reinforced into the system and it will become even more difficult to correct them. The solution this paper offers is to bind the healthcare systems by the same code of ethics that have been guiding the physicians for centuries. But how do we 'teach' AI something it doesn't inherently 'understand' — value and respect for human life? Or is there another way?

Following a human-centered design of AI systems ensures human oversight and accountability to key decisions and actions. When the stakes are high, Human decision-making becomes more empathetic than rational and we cannot teach that to the AI model. AI operates in a very narrow context space consisting only of scenarios that it has been trained on. A simple Pareto analysis will show that only 20% of the scenarios will occur 80% times. That means the relative probability of the majority 80% possible scenarios is so low that either the model will not be able to detect such scenarios or will not be able to prescribe the right course of actions. Potentially, this mean the AI system can err 20% of times simply because it lacks the sufficient data points to make an accurate prediction. When the stakes are high — as they are in the healthcare space, 20% is scary. For a lot of medical conditions, even 1% error rate would be a nightmare for the care providers.

Keeping humans in loop can provide the necessary oversight and accountability as Humans are much more efficient in handling those 20% cases which require more context than computation. But humans can only take action if the AI ecosystem is designed around the human needs. Automation and Augmentation are not exclusive, in fact they inclusive. No AI system can work efficiently unless the teams operating them are also augmented with knowledge and vice-versa.

From the design point of view, Physicians and care giver who want to use these advanced tools must also understand the risks and limitations of these systems. While they need not become machine learning experts but they need enough education about these machine-learning systems to understand how it works — the underlying hypothesis, assumptions, set of models and the datasets.

“Remaining ignorant about the construction of machine-learning systems or allowing them to be constructed as black boxes could lead to ethically problematic outcomes."

But like in every other AI applications, the problem of transparency is a human problem rather than a technical one. The two things that create competitive advantage for the vendors of Healthcare AI systems are their proprietary model and datasets. Additionally, it's their exclusive knowledge and expertise on these systems that gives them the relative power against the physicians. More transparency and education means losing that competitive advantage. No vendor would willingly give away their intellectual property to educate their users. And the more physicians and clinicians understand the core technology, the more power they have over the vendors. Vendors and their investors don't like that at all. So then how to resolve this deadlock?

Human-centered design and PeoplefirstAI can circumvent this problem by creating enough levers and control interfaces in the AI system so that humans can easily control it. There is a lot of research happening today in the 'Explainable AI' space to create the necessary trust and openness in these systems. Giving humans control over control parameters of the AI system along with the ability to visualize and predict the intermediate processing steps can help create a more controllable AI system. The designers need not share their models and datasets with the physicians. Instead, they just need to show them 'how' the systems works and how can they control the process. It is a well-established fact that humans fear what they don't understand and readily embrace what they can control. Yes, the vendors will lose control over their products in the process. Most AI systems use the real-world usage data as a feedback to train their models. With physicians controlling the AI control parameters, it will become much more difficult to integrate this data back into the models. Additionally, the vendors will also lose some control over how the AI works in real-scenarios as a large part of operational and contextual data (on what basis physicians took the decisions) might not be available to them.

Giving up total control over their products might look like a losing bargain to the Healthcare AI vendors but in the long term it will only create openness and trust to drive widespread adoption with lesser risks and liabilities. The winners and losers of the AI game will not be decided on the superiority of their technical products. It will be decided upon who can create the first system that physicians and hospitals find most comfortable and safe for their patients.

In conclusion, the AI systems are complex and susceptible to unpredictable biases and risks that creep in unnoticed due to hidden patters inside the datasets. A blackbox AI solution may look tempting at first to the vendors but considering the risks, it is not a sustainable business model. And healthcare industry is a very special case because the stakes are very high. Even minor errors could create massive risks right from denial of effective care to potentially endangering the lives of countless patients. It is therefore critical to have an open and transparent understanding of inner working of the Healthcare AI systems by the physicians and clinicians. Human-centered design processes can create control levers inside the Healthcare AI system giving control, oversight and accountability to humans and augmenting their capabilities in the process. An open, controllable and predictable AI can foster a higher acceptability and adoption.

0