

AI in Healthcare: Challenges and Solutions

Challenge

Companies and institutions that develop AI models for healthcare diagnosis face several critical challenges. Beyond technical hurdles, a major concern lies in the limitations of their training data, which often fails to fully capture the variability of real-world patient populations.

Specifically:

- Their datasets may have narrow distributions, regardless of the number of labels or cases collected.
- Overfitting can occur, even when developers reserve internal "unseen" validation sets, because the data still comes from a single institution or highly similar sources.
- Developers often receive indirect feedback from tuning their models on these internal datasets, unintentionally reinforcing biases.

As a result, AI models that perform well during internal testing may fail to generalize to external environments — including different hospitals, geographic regions, races, and socioeconomic groups.

This problem is further worsened by dataset shifts:

- Covariate shift (differences in input features like imaging protocols, devices, or demographics)
- Concept shift (changes in how diseases present across different populations)

Ultimately, these challenges lead to bias, reduced trustworthiness, and poor real-world performance of AI models, even if initial internal metrics appear strong.

Case Studies

Google Health

Between 2018 and 2019, Google Health implemented an AI tool in 11 clinics across Thailand to detect diabetic retinopathy (DR), a leading cause of vision loss. While the AI demonstrated over 90% accuracy in controlled settings, it struggled in real-world applications due to the AI needing high-quality retinal images, which were difficult to obtain consistently in clinic environments.

[Read more about Google Health's experience](#)

Personal Case Studies

This work shows that an AI model to do an assessment for glaucoma diagnosis in a hospital does not work for our MGB hospital (MEEI).

Note, MEEI is one of the most prestigious eye clinics around the world (you can use Figure 2E, and F, which shows both R and R² coefficients drop dramatically!):

[View the research article](#)

Almost the same task in this work, AI model trained on available public data, and work not our hospital data:

[View the research article](#)

Need

To address these challenges, companies require a cloud-based architecture that allows them to share and test their AI models with hospitals using real-world healthcare data. Importantly, companies seeking FDA approval and other regulatory clearances face significant costs and operational burdens.

Early external testing is therefore critical — not only to save development costs but also to improve patient outcomes by identifying model weaknesses and biases before deployment.

On the other side, many hospitals cannot or will not share their data with external developers due to privacy, regulatory, and ethical concerns.

Solution

This is where cloud-based solutions become critical.

- With the proper agreements in place, companies can host their AI models on secure cloud instances (e.g., AWS EC2 machines) and provide APIs for evaluation.
- Hospitals can then send or feed selected patient data through these APIs, receive model outputs, and assess performance locally without exposing their sensitive datasets.
- Finally, hospitals can report model results and accuracy metrics back to the developers — allowing model validation on real-world data without direct data sharing.

AWS Model Private Offer seems very interesting.

Costs

Sorry, I cannot share due to NDAs. My suggestion is to browse AWS costs.

Prepared for [Amir Olyaei's](#) review on April 28, 2025

Mohammad Eslami, Ph.D., SMIEEE
Mass Eye and Ear
Instructor in Ophthalmology
Harvard Medical School
AI for Healthcare