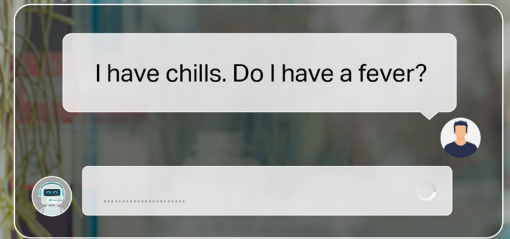> I have chills. Do I have a fever?

# 72%
## PROJECT COST SAVINGS

# IMPROVING RETRIEVAL-AUGMENTED GENERATION FOR HEALTHCARE CHATBOT

*To improve the precision of its medical chatbot, this US technology company consulted iMerit on how to utilize and implement a new medical dataset to ensure accurate medical advice for users.*
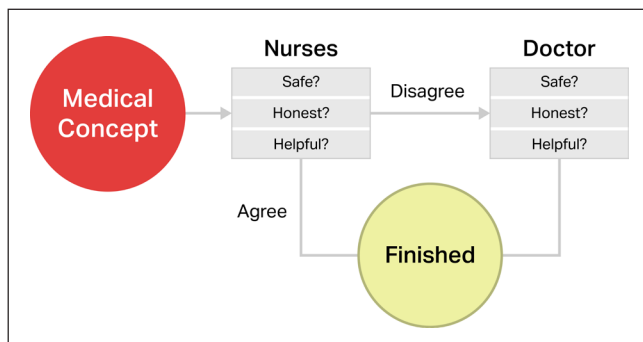
## THE CHALLENGE

Medical large language models are changing the way people research their health. After acquiring a large corpus of medical data, this American multinational technology company needed to evaluate this data for use in their consumer-facing healthcare application. This would ensure generative outputs would be safe, helpful, and at the standard of care of US medical professionals.

> **"**
> *We needed a partner who could provide the expertise to utilize this data safely.*
>
> *- Principal Product Manager, Healthcare AI*

As the nature of this project was unprecedented, this company needed a partner to invent a scalable and validated methodology. Due to the complexity of healthcare data, this project could not be crowdsourced and would require medical expertise to ensure high quality outputs. To come up with a plan for auditing this data, this company began evaluating data solutions providers.



Medical Concept → Nurses (Safe? Honest? Helpful?) — Disagree → Doctor (Safe? Honest? Helpful?) — Agree → Finished

## THE SOLUTION

- Automated human-in-the-loop evaluation and feedback
- Board-certified healthcare expertise
- Risk and accuracy assessment

To establish a workflow for auditing the data, iMerit suggested a pilot project involving a smaller segment of the data. iMerit assembled a team of nurses to work in a consensus workflow with escalations and arbitration provided by a US Board Certified physician. iMerit drew from its extensive network of US-based healthcare experts to quickly build the team to meet the project timelines. The nurses began by evaluating the knowledge base featuring definitions to identify accuracy and risk.

To ensure quality, each concept's respective definitions needed to be evaluated and scored based on custom criteria including the accuracy of its description, the likelihood of harm associated with sharing the information, and the potential helpfulness it posed for users seeking that information. Statistical analysis of the output permitted focusing the post-pilot effort on the most important dimensions of the data.

> **"**
>
> *Without iMerit, this project could have cost us millions.*
>
> *- Principal Product Manager, Healthcare AI*

At the end of the project, iMerit delivered a full report of the project, including time-per-task breakdowns, edge case tracking, efficiency opportunities, and extension plans.

## THE RESULT

After evaluating the results, this company and iMerit discovered that with edge case discussion and guideline revision, the nurses could reach consensus in 99% of cases. This allowed the team to revise the project design to a single-vote structure with 10% audit, thereby reducing project costs by over 72%.

As a result of this pilot, iMerit helped this major software company identify a cost-effective solution for ensuring high-quality medical data. With this new roadmap, the company has chosen to continue working with iMerit to continually identify ways to scale medical data annotation ethically and efficiently.

### BOTTOM LINE IMPACT

# 72%

## Project Cost Savings

---

**LEARN MORE**