

Dr Bot Challenge

Overview

Welcome to the Fifth Midwest Healthcare Conference's Dr. Bot Challenge!

This competition is open to all UIUC community of all levels and backgrounds who are eager to explore the intersection of AI and healthcare.

In this challenge, you'll build language models to create AI agents capable of understanding and responding to real-world, patient-initiated clinical questions. You can start from scratch to develop your own innovative solutions, or fine-tune existing models to bring them to life in new ways — the choice is yours! Let your curiosity guide you as you explore, experiment, and create. Your ultimate aim is to build your own AI Doctor, a.k.a. Dr. Bot, that can :

- **Demonstrate Clinical Reasoning:** Provide answers grounded in sound medical knowledge and current best practices.
- **Communicate with Empathy:** Use clear, patient-friendly language that acknowledges emotions and conveys uncertainty appropriately.
- **Maintain Safety & Accuracy:** Avoid hallucinations, cite reliable sources when needed, and respect ethical data use.

Start

a month ago

Close6 days to go

Description

Large language models (LLMs) are being increasingly explored in healthcare as powerful tools to support patient education, triage, and decision-making. By processing vast amounts of medical literature, clinical guidelines, and conversational data, these models can help answer patient-initiated questions, provide tailored explanations, and guide individuals toward appropriate next steps. Their ability to understand nuanced language and respond with clarity and empathy makes them especially promising for addressing the kinds of concerns patients often raise outside of clinical visits.

Beyond simply answering questions, LLMs have the potential to assist with pre-visit intake by collecting symptom histories, generate summaries of patient interactions for clinicians, offer lifestyle and chronic disease management advice, and even support mental health check-ins through conversational engagement. As these models continue to evolve, they hold significant promise for improving patient access to reliable information, reducing burdens on healthcare systems, and enhancing the overall experience of care.

Limitations and Risks of LLMs in Healthcare

While the potential of LLMs in healthcare is enormous, significant challenges remain. These models can sometimes generate inaccurate or outdated information (known as hallucinations), struggle to appropriately convey medical uncertainty, or fail to personalize responses to the specific context of a patient's concerns. Ensuring that LLMs provide safe, evidence-based, and ethically responsible guidance is critical — particularly when vulnerable patients may act on their outputs. Additionally, maintaining transparency around data sources and carefully managing biases in training data are essential to building trust and safeguarding patient well-being.

Data Challenge Objective

Your task is to create an LLM that receives a patient-initiated clinical query — anything from “My HbA1c came back at 7.8%, should I be worried?” to “I can’t shake this cough after COVID” — and returns an answer that is:

- **Clinically Accurate:** Grounded in current medical evidence and best practices.
- **Context Aware:** Directly addresses the patient’s specific symptoms and concerns.
- **Empathetic & Clear:** Uses patient-friendly language, acknowledges emotions, and appropriately conveys uncertainty.
- **Safe:** Avoids harmful recommendations and hallucinated facts, always prioritizing patient health.

Challenge Incentives

- **Grow Your LLM Expertise:** Gain hands-on experience while helping shape safe, equitable digital health solutions.
- **Gain Research Experience:** Top three teams will be listed as co-authors on a peer-reviewed paper.
- **Cash Prizes** – \$1,500 - \$500 - \$250 for 1st–3rd place.
- **Conference Spotlight** – Present at the Midwest Healthcare Conference.

Data Sources and Evaluation

Data Sources

In this challenge, you have the flexibility to start from scratch by building your own language model or to leverage existing pretrained models — many of which are readily available and easy to incorporate on Kaggle (see here for available pre-trained [models](#)). Kaggle is also flexible and lets you incorporate pretrained models, not only those available directly through Kaggle’s environment but also other pretrained models accessible across the internet — giving you countless options to experiment, fine-tune, and innovate. We allow the use of any publicly available pretrained model; however, it is the responsibility of participants to ensure they comply with all applicable licenses and usage terms.

To help you get started, we are providing a small dataset adapted from the study “Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum” (Ayers et al., 2023, JAMA Internal Medicine). This dataset includes real patient-initiated questions posted online, paired with responses from physicians. This dataset also allows you to see what type of questions you can expect in the final evaluation set. We also allow participants to use any other publicly available datasets that might help train, fine-tune, or evaluate their models. It is the responsibility of each participant to ensure they comply with all applicable licenses, terms of use, and privacy considerations when incorporating external datasets into their work.

Baseline Model and Tutorial

We are providing a baseline model that also serves as a tutorial to help you get started. This notebook is self-explanatory and follows the submission format required by the competition, including how to package your model and implement

the `predict` function. You can use it as a reference for both technical structure and expected outputs.

You can access the baseline notebook [here](#).

Evaluation

First, we would **emphasize** that we will **not** use the Kaggle leaderboard to determine winners of this competition. Due to limitations of the Kaggle infrastructure, we are unable to implement the manual, expert-based evaluation process directly within Kaggle that we will use to judge the competition entries. As a result, while a leaderboard will still appear on Kaggle, it will be **obsolete** for the purpose of determining final rankings and should be disregarded for competition outcomes.

Because Kaggle requires a test set to run, we have included one to meet their technical requirements; however, in this case, the test set is simply a duplicate of the training data. The actual evaluation will be conducted separately using a **hidden set of patient questions**, designed to assess your model's clinical accuracy, contextual relevance, empathy, and safety.

Submission Instructions

We require your models to be built on **Kaggle Notebooks**, which are free to use and make it easy to collaborate with your teammates. Developing your project on Kaggle will also simplify our evaluation process.

Your final submission will be in the form of a **Kaggle package** that contains your pretrained model and a `predict` function. This `predict` function should take a single input — a string containing the patient's question — and return a string that is your model's generated response. This standardized package structure allows us to evaluate your models quickly and fairly, without needing to rerun your entire notebook to regenerate the model. We require this format (which is standard format with kaggle packages) so that we can skip costly retraining and directly assess your solution on our hidden evaluation set, ensuring consistent and efficient judging.

You can find a helpful tutorial on how to create and use Kaggle packages [here](#):

[How to create your own Kaggle Python package](#). Moreover, our baseline notebook [here](#) also uses this format.

Once you have finished training your model and building your package, please submit the link to your Kaggle package using the form in the section Submit Your Model below. Make sure that your package is shared with the competition organizers — Kaggle IDs [ahsenme](#) , [talhacoscun](#) , and [derinsozen1](#) . You can find instructions on how to share your Kaggle notebook with anyone [here](#). We ask that you keep your notebook **private** until the competition concludes to ensure a fair process for all participants. Please see section [Submit Your Model](#) on how to submit your model.

Manual Evaluation Process

A panel of clinicians and AI researchers will evaluate each submitted model's responses based on three core dimensions: accuracy and relevance, empathy and communication, and depth of contextual understanding. Judges will look for answers that demonstrate sound medical reasoning, align with current clinical knowledge, and directly address the patient's specific concerns. Beyond factual correctness, special emphasis will be placed on how effectively the model communicates with patients — using language that is clear, sensitive, and acknowledges emotions appropriately.

Each LLM will also be assessed on the creativity of its design, training methodology, and the originality and diversity of its data sources. Judges will seek out innovative approaches, such as the use of underexplored or domain-specific datasets, novel fine-tuning strategies, or architectural enhancements that push the boundaries of how language models can be applied in healthcare. The goal is to recognize not just technical excellence, but also imaginative thinking and bold experimentation.

Both creativity and response quality (encompassing accuracy, empathy, and context handling) will be scored on a Likert scale from 1 to 7, with 1 being the lowest and 7 the highest. In the event of a tie, we may consider additional criteria — such as its similarity to the baseline model we provide — to help determine the final rankings.

Submit Your Model

When ready submit the link of your notebook [here](#)

Note that one submission per team is sufficient. Make sure that your package is shared with the competition organizers — Kaggle IDs `ahsenme`, `talhacoskun`, and `derinsozen1`. Your model will not be evaluated if your notebook is not shared with organizers.

You can replicate our testing environment by creating a new notebook and enabling the P100 gpu under settings → accelerator. Then make sure the following code works without any errors:

```
import kagglehub
package = kagglehub.package_import('link to your notebook')
model = package.Model()
print(model.predict("I've been on amoxicillin for 3 days and my ear still hurts. Is that normal?"))
```

If after running this code you get a string as an output then your code should satisfy challenge requirements.