Task Overview

Develop, train, or fine-tune an open-source small language model (SLM) capable of answering medical questions asked by end-users.

Objective

- The model should provide informative and grounded answers.

- It must not prescribe medications or treatments without advising consultation with a healthcare professional.

- Responses should focus on:

  - Explaining symptoms and common causes.

  - Suggesting non-intrusive home remedies.

  - Recommending appropriate medical specialties for consultation.

  - Advising on diet, exercise, or lifestyle changes for relief.

Instructions

1. Data Preparation

   - Access the dataset of medical questions here. The question set was sourced from public domain, however, please still avoid resharing.

   - Clean and preprocess the data as necessary, addressing any class imbalances or biases.

2. Model Development

   - Use any open-source technologies and frameworks you're comfortable with (e.g., Python, TensorFlow, PyTorch, Hugging Face Transformers).

   - Choose an appropriate open-source base model (e.g., Phi, BERT, GPT-2) suitable for small-scale deployment.

   - Train or fine-tune the model to answer medical questions.

   - Implement strategies to ensure the model adheres to ethical constraints.

3. Deployment

   - The model should be deployable on commodity hardware without GPUs for real-time inference.

   - Optional: Develop a simple chatbot interface to interact with the model, like ChatGPT.

4. Documentation and Submission

- Write clean, well-documented code with appropriate comments.

- Upload your code to a GitHub repository (public or private; ensure access is granted if private).

- Provide detailed documentation covering training and fine-tuning processes, data collection and preprocessing steps, deployment instructions, and any challenges faced.

- Explain your choices regarding the tech stack, frameworks, data handling, model selection, and deployment strategies.

- Discuss how you ensured ethical compliance in the model's responses.

Deliverables

- GitHub repository link containing all code and resources.

- Comprehensive documentation as outlined above, including testing results and evaluation metrics.

- Instructions on how to run the model locally, including necessary files or dependencies.

Evaluation Criteria

- Model Performance: Accuracy and relevance of the answers; adherence to ethical guidelines.

- Code Quality: Clarity, organization, adherence to coding standards; use of comments and documentation.

- Documentation: Thoroughness and clarity; ability to communicate your process and findings effectively.

- Problem-Solving Skills: Logical approach to challenges; justification of methods and tools used.

- Innovation: Creative solutions or unique approaches.

- User Interface: Functionality and usability of the chatbot interface (if implemented).

Submission Deadline

- Submit your GitHub repository link and all documentation in next 4 – 5 days.

Additional Notes

- Ensure all data used complies with ethical standards and data protection regulations.

- Do not include any personally identifiable information.
- The model should not provide specific medical advice or prescriptions.
- Encourage users to consult qualified healthcare professionals when appropriate.
- Optimize the model for local deployment on standard hardware without GPUs.
- Ensure inference times are reasonable for real-time interactions.
- If you have any questions or need clarifications, feel free to reach out.
