Dataset: Comparing Physician and Artificial Intelligence (ChatGPT) Responses to Patient Messages

Summary:

- The data contains a file that comprehensively covers the specific structure of the data as analyzed in our JAMA Internal Medicine study.
- The file is in comma-separated values (CSV) format, and can be opened with any application capable of reading .csv files.

Columns:

- **postID:** a numeric identifier of the post, ranging from 1 to 211 (however, only 195 numbers are used–the missing numbers are due to duplicates or bugged entries which were removed during the data cleaning process)

- **Question:** the full-length text of the health-related question.

- **submissionID:** a unique identifier of the health-related question, generated by Reddit.

- **commentID:** a unique identifier of the physician's response comment, generated by Reddit.

- **Physician Response:** the full-length text of the physician's response, on Reddit, to the original poster's question.

- **ChatGPT Response:** the full-length text of ChatGPT's response output when the original poster's question is copy-pasted in its entirety onto a fresh ChatGPT session.

- **Eval 1 Preference:** whether Evaluator #1 preferred the physician response or ChatGPT's response

- **Eval 2 Preference:** whether Evaluator #2 preferred the physician response or ChatGPT's response

- **Eval 3 Preference:** whether Evaluator #3 preferred the physician response or ChatGPT's response

- **Eval 1 Quality (Physician):** Evaluator #1's rating of the quality of the physician's response

- **Eval 2 Quality (Physician):** Evaluator #2's rating of the quality of the physician's response

- **Eval 3 Quality (Physician):** Evaluator #3's rating of the quality of the physician's response
- **Eval 1 Quality (ChatGPT):** Evaluator #1's rating of the quality of ChatGPT's response
- **Eval 2 Quality (ChatGPT):** Evaluator #2's rating of the quality of ChatGPT's response
- **Eval 3 Quality (ChatGPT):** Evaluator #3's rating of the quality of ChatGPT's response
- **Eval 1 Empathy (Physician):** Evaluator #1's rating of the empathy level of the physician's response
- **Eval 2 Empathy (Physician):** Evaluator #2's rating of the empathy level of the physician's response
- **Eval 3 Empathy (Physician):** Evaluator #3's rating of the empathy level of the physician's response
- **Eval 1 Empathy (ChatGPT):** Evaluator #1's rating of the empathy level of ChatGPT's response
- **Eval 2 Empathy (ChatGPT):** Evaluator #2's rating of the empathy level of ChatGPT's response
- **Eval 3 Empathy (ChatGPT):** Evaluator #3's rating of the empathy level of ChatGPT's response
- **Physician Length:** the number of words in the physician's response
- **ChatGPT Length:** the number of words in ChatGPT's response