# Experience

Reflecting on this project, the experience was truly fulfilling. It required exploration and integration of various tools and techniques that abided by the HIPAA guidelines to work with Electronic Health Records (EHRs).

Key Takeaways:

- The process of storing and manipulating CSV files using Pandas and boto3
- Integration of AWS tools such as S3, AWS Comprehend Medical and AWS Bedrock.
- Comparing the output from different pre-trained Bedrock models.

Challenges:

- The inability of some models, like Titan Text G1 – Express v1, to provide sufficient or accurate information in certain cases highlighted the limitations in current NLP models when applied to specific tasks, such as extracting detailed demographic information or personal history from medical records.
- We encountered permission errors and data access issues that required significant time and effort to troubleshoot effectively.

Highlights:

- Carried out a detailed comparison between different NLP models, like Titan Text G1 – Express v1 and Mistral 7B Instruct, assessing their strengths and limitations in extracting demographic details, emotional tones, and personal history from the text. This analysis provided valuable insights into the effectiveness of these models for specific tasks.
- Filtering rows with less than 10 NaN cells, ensuring that only relevant and less sparse data was used.
- Using boto3 for automating tasks like uploading and downloading files to and from S3 to streamline operations and reduce manual effort.
- Worked around the limitations of the models, adapting our approach to ensure that meaningful data could still be extracted and analyzed.

Future Work:

- Meet with Dr. Wendelken to discuss findings.
- Get the code base, test the RAG API and integrate it with the overall project.