CASE STUDY

Enhancing Healthcare Predictive Models with Generative AI

A Case Study on Pneumonia Detection and Cancer Staging

COMPANY
A pioneering company in health analytics

INDUSTRY
Healthcare

DATA SOURCE
Open-source MIMIC CXR database

TOOL USED
Label Studio for manual validation and annotation

INTRODUCTION

In the swiftly advancing domain of healthcare, employing generative AI, particularly Large Language Models (LLMs), for predicting disease states from clinical reports marks a considerable leap forward. The client, a trailblazer in health analytics, embarked on a mission to refine their disease condition prediction models. By leveraging the open-source MIMIC CXR database and incorporating generative AI predictions for initial analysis, followed by manual validation with Label Studio, the goal was to boost model accuracy and dependability for clinical report analyses, especially radiology reports.

CHALLENGES

Integrating generative AI predictions into healthcare workflows presented numerous challenges:

Data Access and Security: Securing access to high-quality, open-source medical datasets like MIMIC-CXR required a rigorous credentialing process, ensuring compliance with privacy and ethical standards.

Prediction Accuracy: Initial outputs from generative AI models occasionally exhibited inaccuracies in disease condition predictions, necessitating manual checks for enhanced precision.

Complex Disease State Identification: Accurately classifying disease states from the nuanced language of clinical reports, especially when using generative AI, posed a significant hurdle.

Annotation Quality: Ensuring high-quality, accurate annotations within the Label Studio tool required specialized knowledge and understanding of medical disease states.

SOLUTION

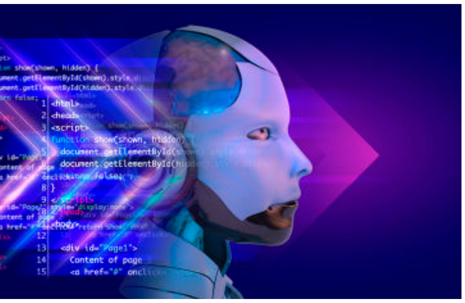
Shaip employed a comprehensive strategy to address these challenges:

- **Streamlined Credentialing:** The team quickly navigated the credentialing process for MIMIC-CXR access, demonstrating efficiency and commitment to ethical research practices.
- **Guideline Development:** Developed insightful guidelines for manual validators to ensure consistency and quality in annotating LLM predictions.
- **Expert Annotations on AI Predictions:** Employed meticulous manual validation and correction of LLM predictions using Label Studio, backed by medical expertise.
- **Performance Metrics:** Through detailed analysis, Shaip calculated LLM's performance metrics such as concordance, precision, recall, and F1 score, enabling continuous improvement.

THE OUTCOMES

- **Enhanced accuracy** in predicting disease conditions from radiology reports.
- **Development** of a **high-quality ground truth** dataset for future product development and evaluation of generative AI predictions.
- **Improved understanding** of disease state identification, facilitating more reliable predictions.

Use Case 1: Machine Learning Model Validation



Scenario: Enhancing Pneumonia Prediction Accuracy with Generative AI

In this instance, a generative AI model sifted through chest X-ray reports to detect signs of pneumonia. A report noting "Increased opacity in the right lower lobe, suggestive of an infectious process" prompted an initial "Uncertain" classification by the AI due to the report's ambiguous phrasing.

Validation Process:

1. A medical expert examined the report within Label Studio, concentrating on the text highlighted by the AI.
2. By evaluating the clinical context and applying radiological knowledge, the expert reclassified the report as a definitive "Positive" for pneumonia.
3. This expert correction was integrated back into the AI model, facilitating its ongoing learning and refinement.

Outcomes:

- Improved Model Accuracy
- Improving Performance Metrics precision and recall

Use Case 2: Generate Ground Truth Dataset

Scenario: Crafting a Benchmark Dataset for Cancer TNM Staging with Generative AI

Aiming to advance cancer progression product development, the client sought to assemble a comprehensive ground truth dataset. This dataset would benchmark the training and assessment of new AI models for accurately predicting the TNM staging of cancer from clinical narratives.

Dataset Generation Process:

1. A broad spectrum of cancer-related reports, including pathology findings and diagnostic overviews, was gathered.
2. The generative AI model provided initial TNM staging predictions for each report, leveraging its learned patterns and knowledge.
3. Medical professionals reviewed these AI-generated predictions for accuracy, rectifying errors, and supplementing information in instances of incomplete or incorrect AI predictions.

Outcomes:

- Creation of a High-Quality Ground Truth Dataset
- Foundation for Future Products for refinement of next-gen models on cancer diagnosis and staging.

CUSTOMER TESTIMONIAL

★★★★★

"Working with Shaip has revolutionized our approach to disease prediction. The precision and reliability of our models have significantly improved with annotations performed by Shaip's domain experts. Thanks to their meticulous validation process."

USA - INDIA - SINGAPORE (866) 473-5655 info@shaip.com www.shaip.com