

ENHANCING REAL-WORLD EVIDENCE ANALYSIS IN MEDICAL IMAGING WITH GENERATIVE AI

Authors: Eduardo Farina^{1,2}, Gustavo Corradi¹, Victor Gadelha¹, Flavia Paiva Lopes¹, Leonardo Vedolin¹

Affiliations:

1- Dasa, São Paulo, SP, Brazil

2- UNIFESP, São Paulo, SP, Brazil



INDEX

1. Introduction

2. Methods

2. Results and Discussion

Introduction

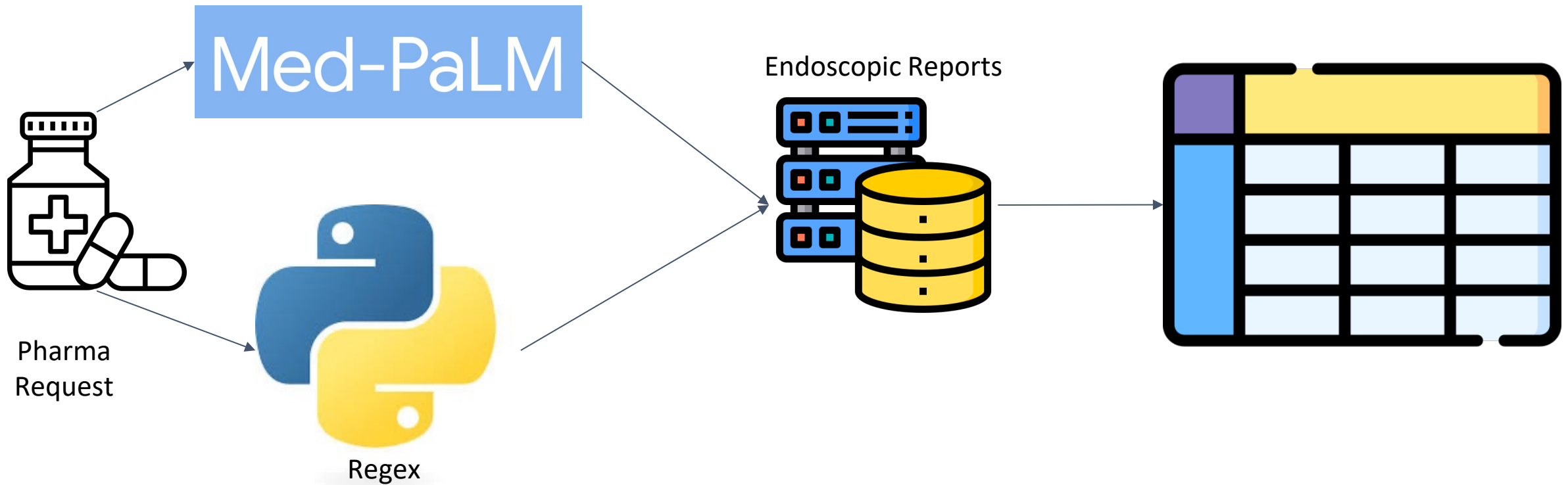
In medical research, the integration of artificial intelligence (AI) is increasingly prevalent, particularly in real-world evidence (RWE) analysis. This study explores the utilization of Google's Medpalm 2, a generative AI model, to scrutinize patterns within medical imaging reports, aiming to augment RWE efficiency.

Med-PaLM

Methods

We employed MedPalm 2 to parse endoscopic imaging reports, comparing its efficacy to a traditional method, regular expressions (regex), in identifying medical findings.

A small sample of 20 reports (10 positive and 10 negative) was annotated to develop regex patterns and refine prompts for the language model.

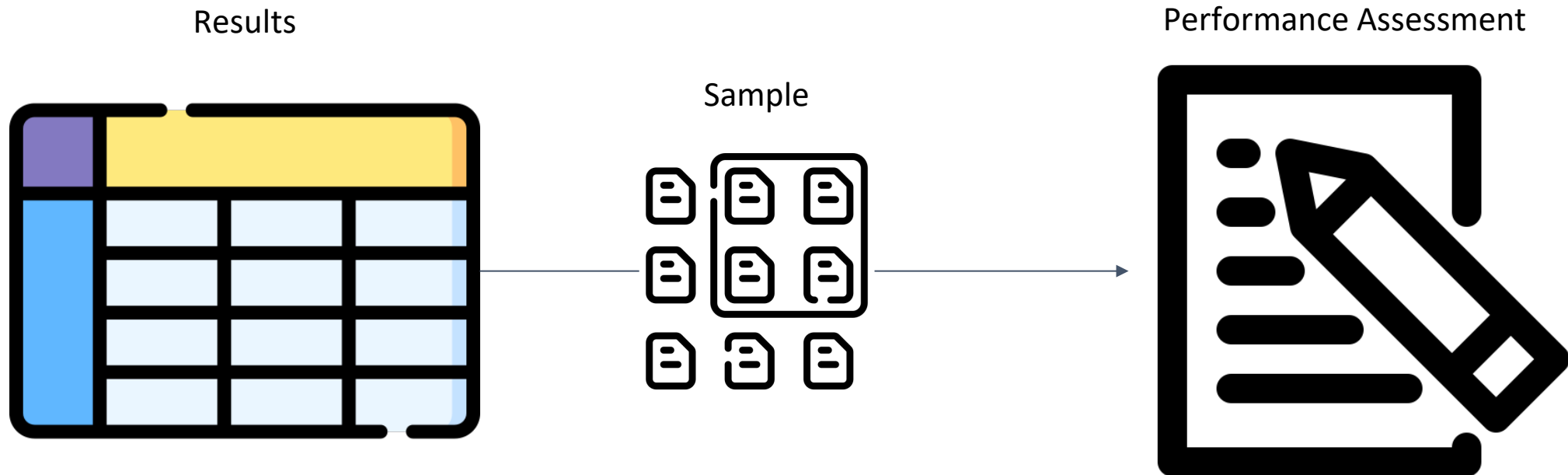


Methods

39,574 reports

Four distinct medical findings: esophagogastric transition cancer, gastric cancer, esophagus cancer, and absence of tumor.

To evaluate positive predictive value, we selected 260 positive predicted reports for MedPalm 2 and 232 for Regex.



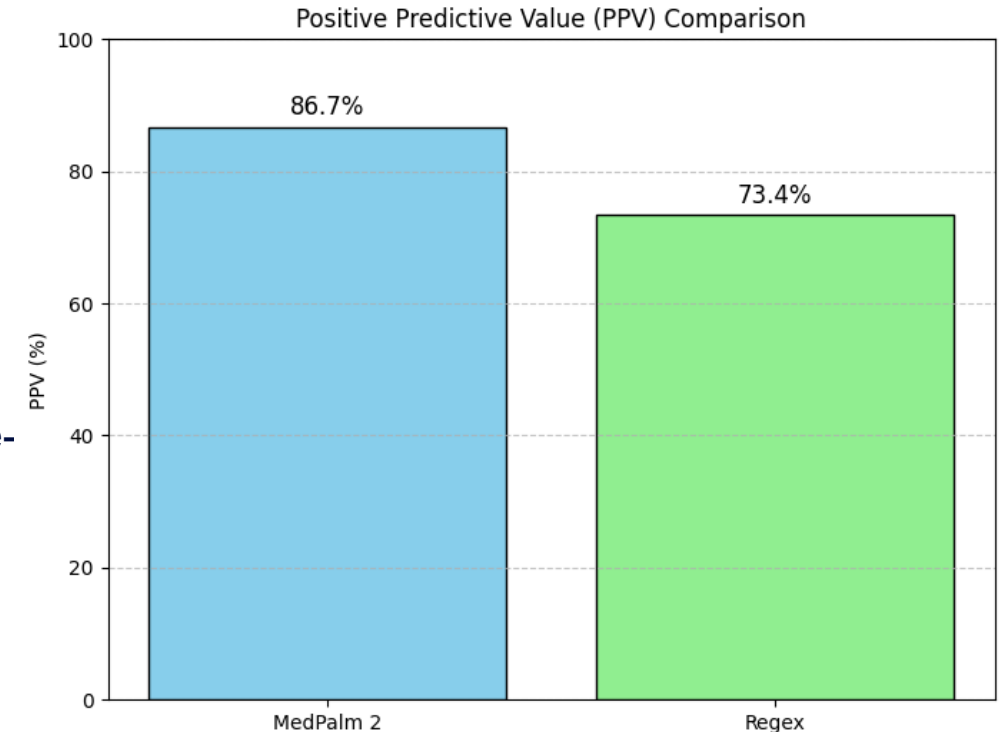
Results and Discussion

Positive Predictive Value (PPV):

- Primary metric for performance evaluation on the sample set
- MedPalm 2 achieved a PPV of 86.7%
- Regex achieved a PPV of 73.4%

LLMs vs. Traditional Methods:

- Large Language Models (LLMs) showed promising results in identifying tumor presence and localization within unstructured, free-text medical reports
- MedPalm 2 outperformed traditional regular expression-based methods
- Higher PPV and detection of more true positive cases suggest potential for better sensitivity



Results and Discussion

Advancements in LLMs:

- Rapid capability advancements since initial release
- Growth accelerated by high-performance, open-source models deployable locally
- Potential to reshape medical evidence by enhancing accessibility and cost-effectiveness of Real-World Evidence (RWE)

Study Limitations:

- Evaluation on a relatively small subset of exams
- Only PPV evaluated for both models
- Development of regular expressions requires significant human effort and we used only 20 cases used

Thank you

Contact: eduardo.farina.ext@dasa.com.br

X/Twitter: @eduardomjmfarina