

Using LLMs to Generate Patient Journeys in Portuguese: an Experiment



Tahsir Ahmed Munna^{1,2}, Ana Luísa Fernandes^{1,3}, Purificação Silvano^{1,3,4}, Nuno Guimarães^{1,2} and Alípio Jorge^{1,2}

¹INESC TEC, Portugal, ²Faculdade de Ciências da Universidade do Porto (FCUP), Portugal, ³Faculdade de Letras da Universidade do Porto (FLUP), Portugal,

⁴Centro de Linguística da Universidade do Porto (CLUP), Portugal



Text2Story 2025
Lucca, Italy



PROBLEM

1. Medical Data Scarcity
2. Patients Full Medical Journey
3. European Portuguese

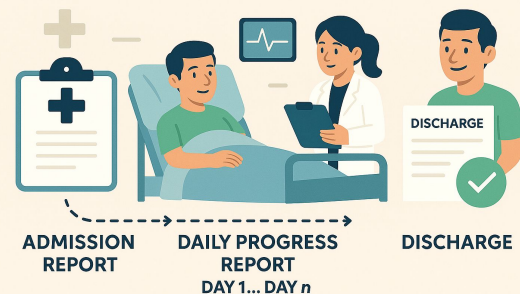


SOLUTION

1. Generating Synthetic Medical Dataset in PT-PT
2. Utilizing Real Medical Case Reports
3. Focusing Patients Full Medical Journey

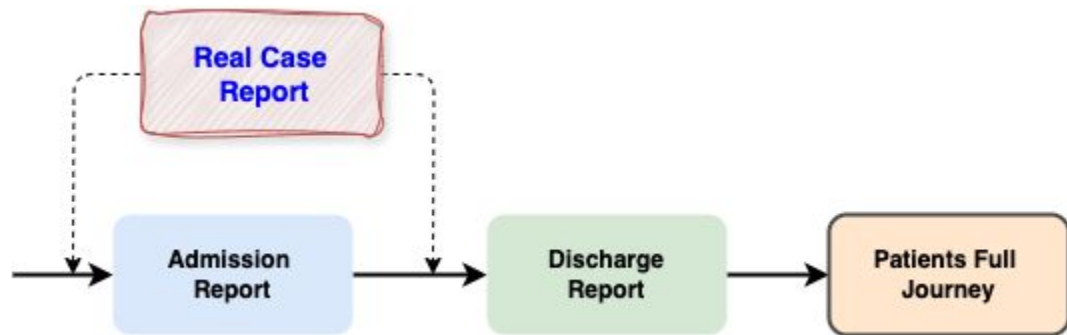


PATIENT'S JOURNEY



 Gemini 1.5 Flash

Using LLMs to Generate Patient Journey in Portuguese: An Experiment



Sequential Generation
(European Portuguese)



Evaluation Methods

Using LLMs to Generate Patient Journey in Portuguese: An Experiment



Using LLMs to Generate Patient Journeys in Portuguese: an Experiment

Tahsir Ahmed Munno^{1,2}, Ana Luísa Fernandes³, Purificação Silvano^{1,4}, Nuno Guimarães² and Alípio Jorge²

WESIC TEC, Portugal; Faculdade de Ciências da Universidade do Porto (FCUP), Portugal; Faculdade de Letras da Universidade do Porto (FLUP), Portugal; ⁴Centro de Investigação da Universidade do Porto (CIUP), Portugal

Text2Story 2025
Lucca, Italy

INTRODUCTION

- This study leverages Gemini 1.5 Flash to create synthetic patient journeys in European Portuguese, tackling data scarcity.
- Ensures medical and linguistic accuracy by using real-world case reports as references during generation.
- The patient journey includes admission, daily progress reports, and discharge, forming a cohesive clinical narrative.

MOTIVATION & OBJECTIVE

- Objective: Generate synthetic patient journeys using LLMs.
- Motivation
 - Generating synthetic medical datasets capturing a patients full journey.
 - Mitigating the data scarcity problem in the medical field.
 - Assessing LLMs' ability to generate patient journeys in European Portuguese.

METHODOLOGY

- Model: Google's Gemini 1.5 Flash.
- Real Case Reports: 285 case reports from SPM.
- Generated Reports: Admission Report, Discharge Report, Full Patient Journey.

EVALUATION AND RESULTS

- Quantitative Assessment**
 - NER inclusion: 100% match.
 - BERTScore: 0.75-0.77.
 - BLEU Score: 0.025-0.14.
 - Language Variant: >95% European Portuguese.
- Qualitative Assessment**
 - Expert evaluation by a linguistic and pharmaceutical sciences expert.
 - Parameters: coherence, narrative structure, and medical accuracy.
- Exploratory Results**
 - Admission Reports: Focus on symptoms and initial assessments.
 - Discharge Reports: Emphasize treatment outcomes.
 - Full Journey Reports: Comprehensive documentation of hospitalization.

Report	PT/P (N)	PF/BR (N)	Metric	Score
Admission	98/24	176	Admission NER Inclusive Score	100
Discharge	95/42	418	Discharge NER Inclusive Score	100
Full Journey	93/8	282	Full Journey NER Inclusive Score	100
			Admission BERTScore	0.75
			Discharge BERTScore	0.8
			Full Journey BERTScore	0.77
			Admission BLEU score	0.025
			Discharge BLEU score	0.03
			Full Journey BLEU score	0.14

Quantitative Results

Exploratory Results

(A) Admission Report (B) Discharge Report (C) Full Journey Report

CONCLUSIONS

- Key Contributions:**
 - Generated high-quality synthetic patient journeys.
 - Addressed data scarcity in European Portuguese.
- Limitations:**
 - Limited dataset size.
 - Single-language focused.
- Future Work:**
 - Expand the dataset.
 - Integrate multimodal data.
 - Adapt to other languages.

Acknowledgments

This work is financially supported by the project "A Realização do Relatório de Avaliação de Impacto Social em Portugal: A Realização do Relatório de Avaliação de Impacto Social em Portugal" (Projeto A Realização do Relatório de Avaliação de Impacto Social em Portugal), funded by the European Union through the ERDF (ERDF) under the operational program "Iniciativa de Emprego Juvenil" (IEJ) 2022-2025, within project HPT1, with reference 61.