

Master's Thesis Proposal at the Chair for Digital Health, Economics and Policy

A comparison of the effect of different prompting methods on the performance of Large Language Models for Data Extraction in Systematic Reviews

Keywords: LLMs, Python, R Skills required, strong computer science background recommended, Austrian Institute for Health Technology Assessment (AIHTA) Collaboration

The aim of the project is to investigate the performance of different prompting methods (e.g. manual, algorithmic via python/DSPy) of (a set of) large language models (LLMs) on data extraction performance for systematic reviews. The student will compare different prompting methods regarding their accuracy of accurate data extraction, including reliability assessment through repeated data extraction using the same prompts. The LLM can be chosen. Also a comparison between different LLMs or ensemble methods are possible. The main goal is to inform researchers who aim to perform systematic reviews with the help of AI to inform them which prompting strategy works best, and how reliable the strategy is. For this reason, one or more existing systematic reviews (to be decided) are used as ground truth, and models shall replicate the data extraction accordingly.

Objectives

- Identify the optimal prompting strategy in systematic review data extraction
- Evaluate LLM performance in systematic review data extraction
- Test the reliability of the methods by using the same prompt multiple times
- Compare results with the original data of an existing systematic review

Methodology

- LLMs
- Prompting
- Performance assessment: Accuracy, PPV, NPV, sensitivity, specificity, F1 score...

Your profile (as a Master student)

- Strong Computer Science background
- Languages: Python or R
- NLP or LLM experience
-

About the supervisor and the chair

Benedikt Langenberger, benedikt.langenberger@hpi.de

Dr. Benedikt Langenberger is a postdoctoral researcher at the Digital Health Cluster at the Hasso Plattner Institute. His research focuses on applying machine learning methods for prediction and causal inference in healthcare. Additionally, he is particularly interested in using econometric techniques to identify the causal effects of policies and interventions on health outcomes.

Master's Thesis Proposal at the Chair for Information Systems (Naumann) and Digital Health, Economics and Policy (Stern)

The Effect of Underrepresented Patient Patterns on ML Performance for Remote Patient Monitoring

Real-world health datasets derived from electronic health records (EHRs) are increasingly used to develop machine learning (ML) models for clinical decision support and patient stratification. Despite their large scale and heterogeneity, such datasets may still contain underrepresented patient subgroups or combinations of characteristics that are insufficiently covered in the data. This can lead to biased ML model behavior, reduced generalizability, and poorer performance for specific patient populations.

Mount Sinai Hospital in New York City provides access to a highly diverse EHR dataset, offering a unique opportunity to investigate representation gaps in real-world clinical data. Building on the concept of Maximal Uncovered Patterns (MUPs), this thesis aims to identify underrepresented patient patterns within a hypertension remote patient monitoring (RPM) cohort and evaluate how these gaps influence downstream machine learning models designed to predict patient benefit from RPM interventions.

First, we will jointly define a clinically relevant prediction target for RPM-related outcomes. Subsequently, a feature set for the intended prediction task will be developed, and baseline ML models will be developed.

The thesis will then apply an existing MUP-detection algorithm to the Mount Sinai EHR dataset to identify underrepresented patient subgroups. Based on these patterns, an experimental framework will be developed to systematically evaluate the impact of representation gaps on ML performance. This includes (1) stress-testing baseline ML models on uncovered patient patterns (MUPs), (2) evaluating the effect of different MUP coverage thresholds, (3) comparing the robustness of different prediction models, such as logistic regression, random forests, and multilayer perceptrons, (4) identifying clusters of related MUPs to analyze whether specific types of representation gaps disproportionately affect ML model performance, and (5) developing and testing initial diversity measures for patient pattern coverage.

Objectives

- Identify MUPs in the Mount Sinai RPM EHR cohort
- Develop a ML model to predict patient response to RPM
- Assess the impact of underrepresented patient patterns on model performance
- Development and testing of diversity measures

Methodology

- Apply existing MUP detection algorithms to structured electronic health record data

- Train and evaluate predictive machine learning models for RPM-related outcomes
- Compare model performance across well-represented and underrepresented patient subgroups

Your profile (as a Master's student)

- Strong background in computer science, particularly coding in Python
- Experienced in working with big datasets and databases
- Interest in ML and healthcare settings

About the supervisor and the chair

This thesis is jointly supervised by Dr. Sedir Mohammed and Linea Schmidt. Dr. Sedir Mohammed is currently pursuing a PhD at the Information Systems chair at the Hasso Plattner Institute. His research interests include diversity assessment of datasets and its influence on downstream ML tasks.

Linea Schmidt is currently pursuing a PhD at the Digital Health, Economics and Policy chair at the Hasso Plattner Institute. Her research interests include novel care models, such as remote patient monitoring and disease management programs, as well as digital health entrepreneurship.

Master's Thesis Proposal at the Chair for Digital Health, Economics and Policy

Building an LLM-based classification system for measuring diagnostic innovation

A growing number of firms now offer diagnostic tests directly to consumers—MRI scans, blood panels, skin maps—bypassing the traditional physician-initiated model. At the same time, artificial intelligence has begun to promise faster or more accurate diagnosis from laboratory outputs and imaging. Together, these consumer- and supplier-facing developments have led some commentators to argue we have entered an “age of diagnosis.”

But do these developments reflect genuine diagnostic innovation—new technologies that expand what can be detected, when, and how accurately—or are they primarily re-packaging existing tools under new care-delivery models? In which clinical areas is diagnostic innovation concentrated, and where is it absent? And how well does innovation reflect societal needs and “burden of disease”? How might these trends reshape the process of diagnosis for patients and physicians, and what consequences might they carry for health outcomes, equity, and utilization?

Answering these questions requires, as a first step, a reliable descriptive account of the landscape of diagnostic innovation. This thesis proposes to construct one, drawing on the methodological tradition in the economics of innovation that uses regulatory and clinical-trial data to measure inventive activity. Focusing on the US market, the thesis will review the regulatory context surrounding diagnostic tools, collect and integrate data on medical innovation (e.g., from the US Food and Drug Administration’s official regulatory databases). The goal is to develop a method for categorizing clinical trials as diagnostic, making use of LLMs and validating their output.

Objectives

- Regulatory review: Describe the regulatory landscape for diagnostic tools in the US, including FDA device classification pathways (510(k), De Novo, PMA), the regulation of in vitro diagnostics, laboratory-developed tests, and the registration of diagnostic clinical trials on ClinicalTrials.gov.
- Classification methodology: Develop and validate a method for classifying clinical trials registered on ClinicalTrials.gov as diagnostic, using and comparing multiple large language models (LLMs). Validate LLM classifications against a hand-coded gold standard and quantify model performance across various models and prompt architectures.
- Descriptive analysis: Apply the validated classification to the full ClinicalTrials.gov database (circa 2000–2025) to measure and describe the landscape of diagnostic

innovation across clinical areas, over time, and by sponsor type (e.g., industry, academic, NIH).

- Integration with FDA data: Where feasible, link clinical-trial data with FDA device-approval records (510(k), De Novo, PMA, AI/ML-enabled databases) to enrich the descriptive picture and identify trials associated with approved diagnostic products.

Methodology

- Literature and regulatory review
- Data collection and engineering using LLMs
- Data analysis

Your profile (as a Master's student)

- Strong data analytics background (Python, R, etc.).
- Experience using LLMs in data classification and/or analysis is highly desirable.
- Keen interest in diagnosis and medical innovation.

About the supervisor and the chair

Dr. Helena Roy is a postdoctoral researcher at the Digital Health Cluster at the Hasso Plattner Institute. Her research focuses on behavior and novel information technologies in the context of healthcare. She is particularly interested in the changing aspects of diagnosis.

Re-create Paper on the FDA's Breakthrough Therapy Designation with European (EMA) Data

Regulators of new products confront a tradeoff between speeding a new product to market and collecting additional product quality information. The FDA's Breakthrough Therapy Designation (BTD) provides an opportunity to understand if a regulator can use new policy to innovate around this tradeoff—i.e., whether it improved regulator productivity by allowing products to come to market more quickly without compromising quality. In prior work, we found that the BTD program shortened clinical development times by 23 percent and did not impact the ex post safety profile of drugs with the designation. In exploring mechanisms, we found that the BTD program had the greatest impact on less experienced firms and was associated with reduced BTD clinical trial design complexity. The results suggest that targeted regulatory innovation can shorten R&D periods without compromising the quality of new products.

The proposed thesis would aim to replicate this paper and build on it in the context of expedited European drug approvals. The European Union developed the PRIME program in 2016 “to enhance support for the development of medicines that target an unmet medical need.” PRIME is a scheme run by the European Medicines Agency (EMA) to enhance support for the development of medicines that target an unmet medical need. This voluntary scheme is based on enhanced interaction and early dialogue with developers of promising medicines, to optimise development plans and speed up evaluation so these medicines can reach patients earlier. Much like the BTD, PRIME “is based on enhanced interaction and early dialogue with developers of promising medicines” and is designed “to optimise development plans and speed up evaluation so these medicines can reach patients earlier” (European Medicines Agency 2018).

Objectives

- Understand the extant literature and research on accelerated approval pathways for new drugs in both the United States and European Union.
- Collect detailed data on European drug approvals between 2010 and 2025
- Collect detailed data on the clinical trials associated with all approved drugs
- Replicate the econometric analysis performed in the first paper using matching models
- Build on the analysis for questions specific to the European policy context

Methodology

- Literature Review
- Data collection and data engineering
- Econometrics/data science using matching models

Your profile (as a Master's student)

- Strong analytics background with interest in causal inference
- Languages: flexible, but good statistical programming skills necessary
- Interest in pharmaceutical innovation

About the supervisor and the chair

Ariel Dora Stern holds the Chair for Digital Health, Economics & Policy at the HPI. Her research focuses on technology management and innovation in health care, using methods from econometrics and data science to curate and analyze novel datasets. Her projects consider the regulation, strategy, and economics of health care, with a focus on emerging health care technologies and delivery modalities.