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 joinedly 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

Identification of states of vulnerability of patients with clinical obesity to develop digital therapeutic solutions

Keywords: Charité Collaboration, Qualitative Interviews, Design Thinking, Medical Focus, Digital Health, Personalisation of Digital Health, Sufficient German skills required

The concept of "clinical obesity" defines obesity not solely by Body Mass Index (BMI), but incorporates weight-related impairments of organ function and daily living activities. Characterized as a chronic-relapsing condition, clinical obesity requires holistic therapeutic management integrating lifestyle interventions (dietary modifications, physical activity), pharmacotherapy, and bariatric surgery to achieve sustained weight reduction. However, insufficient and unsustainable patient compliance remains a critical barrier to therapeutic success. So-called "states of vulnerability" mark moments in everyday patient life when individuals are particularly susceptible to health-adverse behaviors (e.g., dietary lapses). Digital health interventions, particularly mobile-based approaches, have demonstrated potential in promoting positive health behaviors, helping patients avoid states of vulnerability, and improving adherence in chronic disease management, offering a promising strategy to address compliance challenges in clinical obesity.

Objectives

- Identify pain points (unsolved needs) in current treatment/care pathways in patients with clinical obesity using a Design Thinking Mindset (Semi-Structured Interviews).
- Identify states of vulnerability for people with clinical obesity.
- Derive insights to develop a digital therapeutic (wireframing, conceptualization).

Methodology

- Plan, organize, and conduct qualitative patient interviews
- Transcribing, coding methods with MAXQDA
- Wireframing (e.g., with Figma)

Your profile (as a Master student)

- Interest in non-communicable diseases/metabolic syndrome
- Interest in design thinking/qualitative exploration of patient needs
- Interest in digital health/digital therapeutics
- Sufficiency in German

Date: 30.04.2026 Last edited: 30.04.2026

About the supervisor and the chair

Dr. Philipp Stoffers is a physician with many years of clinical experience in internal medicine and gastroenterology. He currently works as a program manager and lecturer for innovation in healthcare at HPI's School of Entrepreneurship and as a research associate at HPI's Digital Health Cluster.

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.