Natural language processing in healthcare: applications and value

Buchem, M.M. van

Chapter 10

Summary

This thesis investigates the application of natural language processing (NLP) in healthcare. Part 1 describes the development of several NLP models in different healthcare settings and discusses its challenges. Part 2 focuses on determining the added value of NLP in clinical practice, discussing two pilot studies.

Chapter 2 provides a comprehensive scoping review and research agenda for the implementation of digital scribes in clinical practice. Digital scribes, which leverage automatic speech recognition (ASR) and NLP techniques, aim to alleviate clinician burnout by automating clinical documentation. It highlights that current research primarily focuses on technical validity without adequately addressing clinical validity, usability, or utility. Recommendations for future research include improving ASR accuracy, ensuring comprehensive validation, and enhancing transparency in reporting to bridge the gap between research and practical implementation.

Chapter 3 evaluates the use of NLP to predict acute care utilization (ACU) in oncology patients starting chemotherapy, using clinical notes. It compares deep learning models to manually engineered language features and structured health data (SHD) models. Results indicate that while SHD models slightly outperform NLP models, both approaches are viable. The study underscores the potential of NLP in clinical applications and highlights risk biases across diverse patient groups, suggesting that future research should focus on enhancing model generalizability and addressing these biases.

Chapter 4 discusses the development and validation of the Artificial Intelligence Patient-Reported Experience Measures (AI-PREM), a tool designed to automate the analysis of open-ended patient experience data using NLP (see Figure1). The AI-PREM encompasses a newly developed open-ended questionnaire, an NLP pipeline for analyzing responses through sentiment analysis and topic modeling, and a visualization interface for easy interpretation by healthcare professionals. The validation process affirmed the tool's capability to identify relevant patient experience themes and sentiments accurately, suggesting its applicability in clinical settings to support quality improvement efforts.

Chapter 5 presents the development of a model to identify depression concerns in cancer patients using NLP on patient messages. The study utilized messages from a secure patient portal at a cancer center, applying logistic regression, support vector machines, and two variations of BERT models (original and Reddit-pretrained) for classification. The best performance was achieved by the BERT models, indicating their effectiveness in detecting depression concerns. The study also explored model fairness and explainability, revealing performance disparities across demographic groups. This research underscores the potential of advanced NLP techniques in enhancing early depression detection in clinical settings, though it also highlights the need for careful consideration of inherent model biases and their impact on different patient groups.

Chapter 6 examines public perceptions of statins on Reddit, using AI to analyze 10,233 posts and comments. NLP models and clustering algorithms categorized the content into 100 topics across six thematic groups. The study suggests that AI can provide valuable insights into public opinions, helping healthcare professionals address misconceptions and improve statin adherence.

Chapter 7 investigates the AI-PREM tool's value in clinical practice, specifically for a vestibular schwannoma care pathway (see Figure 1). By comparing open-ended and closed-ended patient experience questionnaires, the study found that the AI-PREM provided more detailed patient feedback, identifying specific improvement areas that closed-ended PREMs missed. The findings highlight AI-PREM's potential to enhance patient-centered care through actionable feedback.

Chapter 8 examines the impact of automating clinical documentation with large language models, focusing on time and quality (see Figure 2). Medical students summarized 150 mock conversations either manually or using a Dutch digital scribe system, finding that fully automated summaries had lower quality scores compared to manual ones, but editing of automated summaries improved in quality. Furthermore, manual summarization took longer than editing automated summaries. The study suggests digital scribes can reduce documentation time while maintaining high-quality records, with further research needed in clinical settings.

In conclusion, there is significant progress in NLP with many potential applications in healthcare. Despite these advancements, challenges such as data quality, bias, clinically relevant metrics, and the lack of high-quality clinical evaluations hinder widespread adoption. Focused efforts are required, such as more clinical evaluations, improved reporting, and more streamlined integration of NLP into clinical practice. This way, the potential of NLP to improve patient outcomes, enhance clinician experience, and reducing costs may be realized.

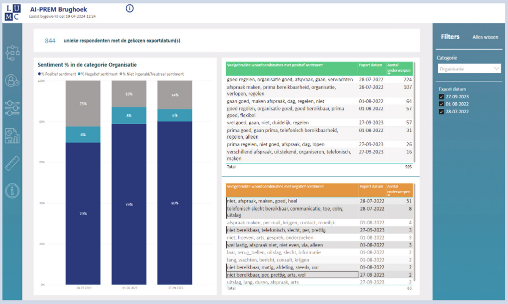
AI-PREM

Goal: Getting insights from free text patient experiences.

End user: Care pathway team, responsible for the quality improvement of the care pathway (varies per care pathway).

Workflow: Every year, the care pathway sends out a questionnaire consisting of five open-ended questions to all patients within the care pathway. The care pathway team then analyzes the AI-PREM results, specifically examining negative clusters of patient experiences to identify areas for improvement. Relevant topics are subsequently investigated by the team by examining the raw text data within those topics. When potential actions are identified, they are deliberated in meetings and balanced against any positive feedback on the same issues.

Application: Integrated as extra tab within the already existing Patient Reported Outcome Measures Dashboard.



Screenshot of the current version of the AI-PREM. On the left, the output from the sentiment analysis is shown per year. On the right, the topics per question are shown.

Figure 1: a description of the AI-PREM tool.

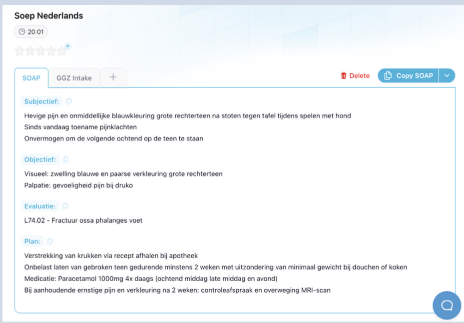
Autoscriber

Goal: Automatic note generation for clinical consultations.

End user: Physicians.

Setting: At the start of a consultation, the physician asks the patient if they consent with using Autoscriber. If they do, the physicians starts a new recording and proceeds as normal. In the back, Autoscriber automatically transcribes the conversation. At the end of the consultation, the physician ends the recording and after a few seconds, an automatically generated summary appears. This summary can be edited and then transferred to the patient's medical file.

Application: Currently used as stand-alone web application. However, integration with Hix (the largest EHR within the Netherlands) will soon be possible.



Screenshot of the current version of Autoscriber, showing an automatic summary of a mock conversation.

Figure 2: a description of the Autoscriber tool.

10