

UC San Diego

Family Medicine Research Symposium 2025

Title

Facilitating End-of-Life Care Decision Making by Summarizing Goals of Care Notes by Large Language Models

Permalink

<https://escholarship.org/uc/item/36k9q25m>

Authors

Ekbote, Ninad

Graham, Ross

Tai-Seale, Ming

Publication Date

2026-10-13

Peer reviewed

Advance Care Planning (ACP) plays a vital role in making sure that medical treatments align with patient values and preferences, especially in serious illness and end-of-life scenarios. Frequently, ACP-related information, such as code status, designated decision makers, and references to palliative care, is often embedded within unstructured clinical notes in Electronic Health Records (EHRs). Retrieving this information manually is labour-intensive, error-prone, and inefficient. Addressing this challenge requires automated solutions that can accurately extract structured data from free-text clinical documentation saving time and human labour.

Advances in Large Language Models (LLMs) and other areas of Natural Language processing (NLP) has shown potential in transforming the way we extract information from unstructured text. Trained on massive text corpora, these mathematical models display strong capabilities in comprehending intricate and nuanced language patterns, semantic relationships, and contextual nuances within the text data. The ability of these large models to perform tasks such as text summarization, QA, and Named Entity Recognition without the need for task-specific training makes them highly efficient and versatile for clinical applications. In ACP, LLMs provide an opportunity to automate the identification of critical patient information spread across clinical notes, offering both scalability and accuracy.

To evaluate model performance, we compare the LLM-generated outputs with human-annotated references made by medical experts. These annotations serve as the benchmark for measuring accuracy and reliability of LLM outputs. Instead of being dependent solely on exact matches between the LLM outputs and human-annotated references, we use semantic similarity techniques from Natural Language Processing (NLP) to assess how comparable the model's responses are with respect to human comprehension. Finally, the outputs are reviewed by a medical supervisor completing the evaluation loop. Additionally, all data parsing is conducted within a secure, licensed and access-controlled environment ensuring that sensitive patient information remains confidential throughout the process.