Analysing Cancer Patients' Experiences with Embedding-based Topic Modelling

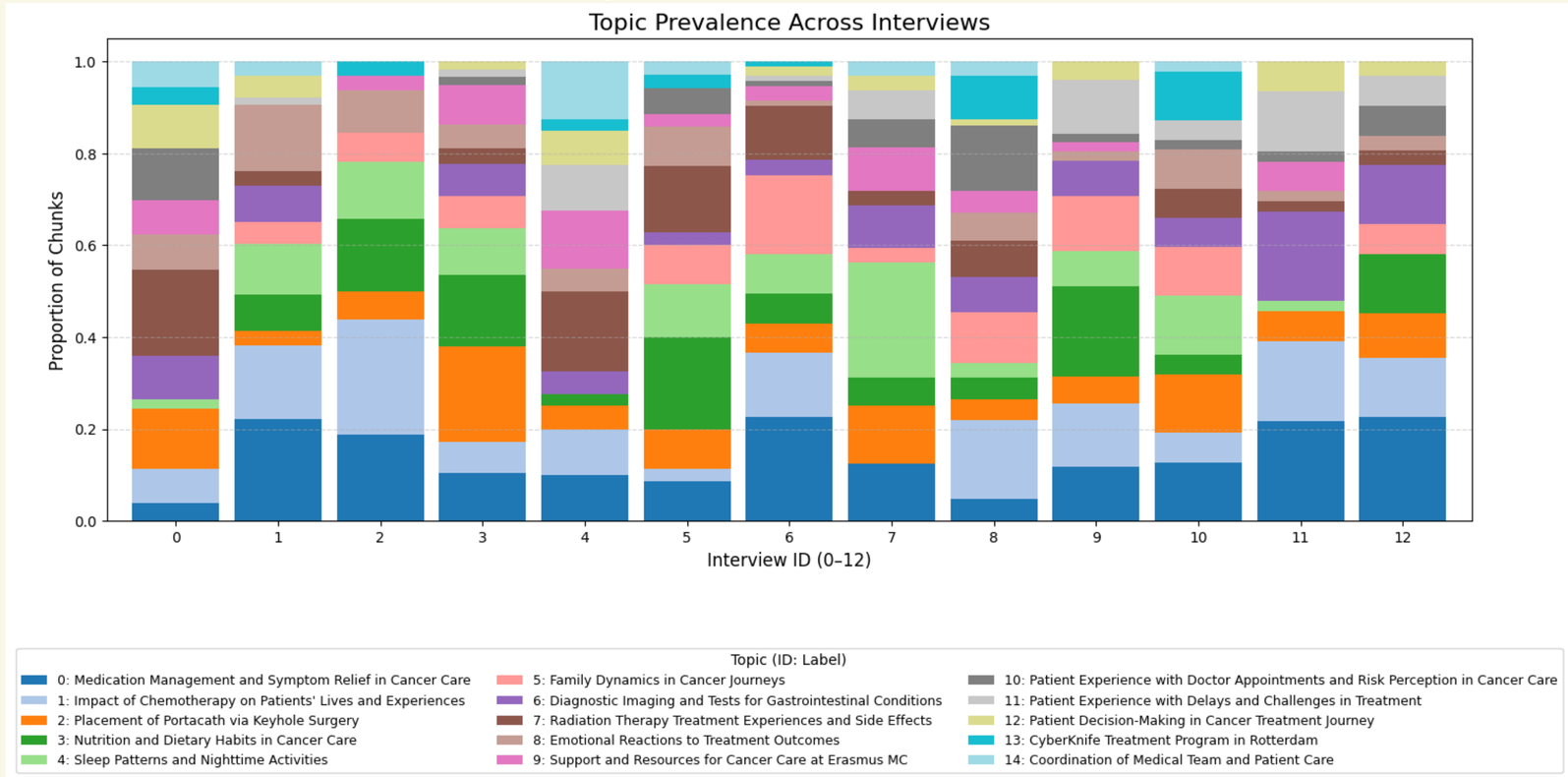
The 35th Meeting of Computational Linguistics in The Netherlands @ KU Leuven 2025

Motivations | Background

- Shared decision-making (**SDM**) (Stiggelbout et al. 2015): a practice in healthcare aiming at better communication between patients and healthcare professionals.
- For **informed treatment**, shared responsibility, more transparent decision making, and better healthcare outcomes.
- To **understand** the patient's experiences in more detail.
- Unlock the values of free text, collected by traditional research, interviews, questionnaires, and surveys.

Topic prevalence across all 13 interviews:

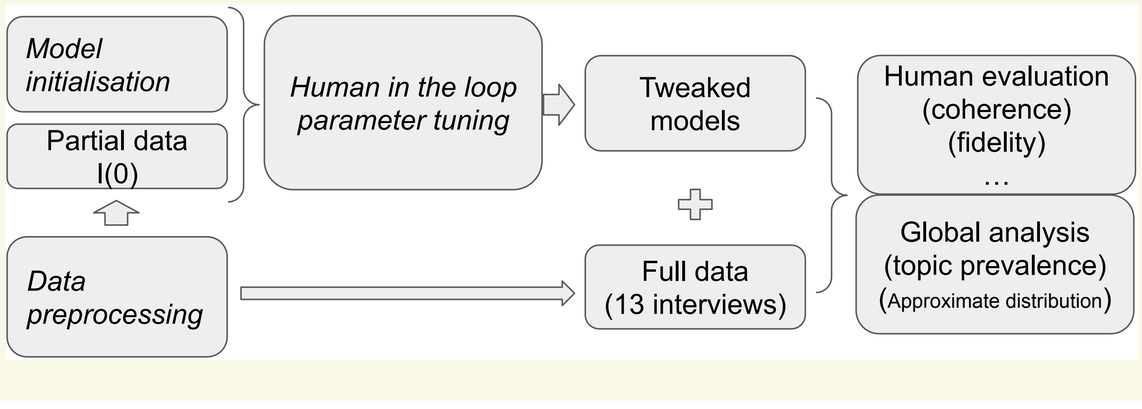
- Each color represents a topic and each bar represents one of the 13 interviews.
- BioClinicalBERT-Tuned BERTopic Global Model



Methodology & Data

- **Patient storytelling** data collected in the Metro Mapping Project (Griffioen et al. 2021) in an oncology setting. (DOI `10.1002/cam4.4145`)
- Interview **transcriptions** of three roles: the patient, the interviewer, and the loved one.
- **Topic modelling** task: identify the key concerns during the cancer patient journey.
- Embedding-based TM: **BERT-topic** and **Topic2Vec**.
- Goal: uncover meaningful **themes** from cancer patient data | how the patient-oriented healthcare practice can be enhanced via AI.
- 3 **domain-specific LMs**: embedding representations - BioClinicalBERT, ClinicalBERT, and MSR BiomedBERT

Development Pipeline



Results: (illustration in the above figure)

- Extraction: a variety of relevant themes from patient interviews.
- **Coverage**: emotional experiences, treatment details, personal struggles, and reflections on the treatment processes.
- Both techniques produced fairly coherent and easily interpretable topics
- **BERTopic** better represented the patients' *experiences* and *concerns*.
- Best embedding: domain-specific pre-trained BioClinicalBERT.

Conclusions & Future work

- Help clinicians identify and understand key moments in a patient's narrative.
- Time saving: reduces the workload of clinical staff, and gives more visibility to the patient's voice.
- In cases where emotional or psychological concerns might otherwise be overlooked.
- Clear potential - integrating topic modelling into tools that support patient-centered care.

Acknowledgement: Funded by the European Union under Horizon Europe Work Programme 101057332. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency (HaDEA). Neither the European Union nor the granting authority can be held responsible for them. The UK team are funded under the Innovate UK Horizon Europe Guarantee Programme, UKRI Reference Number: 10041120. | We thank Sheila Payne and I.J. Korfage for commenting on the Abstract of this work.

Teodor-Călin Ionescu, Lifeng Han*, Jan Heijdra Suasnabar, Anne Stiggelbout, Suzan Verberne

on behalf of 4D PICTURE: https://4dpicture.eu | open-source proj. https://github.com/aaronlifenghan/TM4health

Leiden University Medical Centre | Leiden Institute of Advanced Computer Science | mail: l.han@lumc.nl