Tell Me What I Need to Know: Exploring LLM-based (Personalized) Abstractive Multi-Source Meeting Summarization

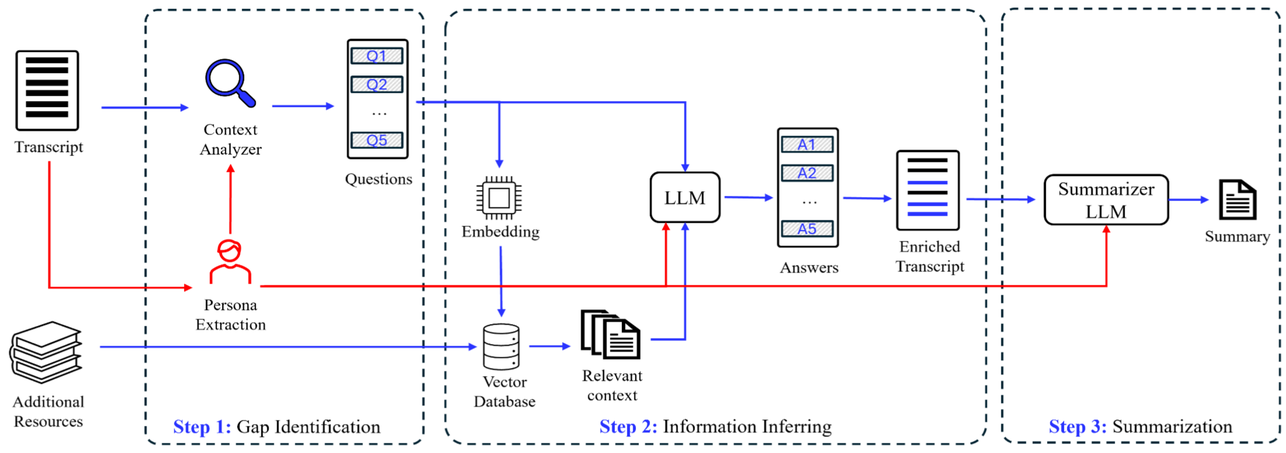
Frederic Kirstein, Terry Ruas, Robert Kratel & Bela Gipp
University of Göttingen, Germany

EMNLP 2024 · GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737 · Gipp Lab Scientific Infomation Analytics

Motivation

- Current summarization systems neglect that meetings are often accompanied by additional resources
- Additional content-related sources can be considered to improve the model's context understanding
- Typically, generated summaries are general and not fitted to what the target reader looks for → not ideal for personal efficiency
- A form of personalization can also be to detail content a target reader may be unfamiliar with using additional knowledge not present in the original transcript
- A significant link between multi-source and personalization when multi-source is not only seen as a tool to enhance context understanding but also a way to introduce different sources/materials/etc. that can help to generate a more nuanced personalization of the content.

Setup



- We introduce MS-AMI, an adapted version of AMI comprising 125 staged business meetings with processed supplementary content (whiteboard drawings, slides, notes)
- GPT-4o as backbone for all stages
- AUTOCALIBRATE to assess informativeness (INF) and relevance (REL), FACTSCORE for factuality (FAC) and an overall scoring considering eight typical error types of meeting summarization (OVR)

Results

We employ a RAG-based pipeline to improve context understanding through multi-source summarization

The structured inclusion of inferred details significantly enhances summary quality

- Multi-source Summarization Improves contextual understanding and Relevance
- Inferred Details Reduce Hallucinations
- Selective Integration Outperforms Simple Concatenation

Setup	INF	REL	FAC	OVR
G-infer	**4.49***	4.04**	**4.78***	**4.41***
G-top	4.33	4.02**	4.67*	4.30
G-all	4.40	**4.11****	4.30	4.35*
G-none	4.31	3.70	4.33	3.99
GOLD	3.79	3.59	4.98*	4.12

▶ *LLM-based 5-point Likert scoring of the general multi-source meeting summarization pipeline.*

"In a kick-off meeting for a new project, a team discusses the design of an original, trendy, and user-friendly remote control. The design process will involve three stages, with individual work and collaborative meetings. The team considers various features, such as combining multiple device controls into one, adding …"

We extract details about a target reader to personalize summaries

Detailed personas improve personalization but introduce challenges in handling and linking content to specific readers

- Improved Personalization with Persona Extraction
- P-infer+per provides the most tailored summaries
- Linking content to specific personas remains complex

Setup	INF-P	REL-P	FAC	OVR-P
P-infer+per	**4.51***	4.16*	4.65*	**4.79***
P-per	4.43*	**4.18***	4.59*	4.50*
P-infer	4.34	4.09	**4.75***	4.35
P-all	4.18	4.04	4.38	4.20
P-none	4.00	3.59	4.33	4.03

▶ *LLM-based 5-point Likert scoring of the personalized multi-source meeting summarization pipeline.*

"Laura, you as the Project Manager, led the kick-off meeting for the design of a new remote control, aiming for it to be original, trendy, and user-friendly. The team discussed the project's financial goals, targeting a selling price of 25 Euros and a production cost not exceeding 12.50Euros. The team explored the idea of integrating multiple device controls into one unit, reflecting on the inconvenience of managing … "

- Multi-source summarization improves quality
- Persona-based personalization increases relevance
- Phi-3 mini demonstrates that even significantly smaller models can produce high-quality summaries, performing competitively with larger models like GPT-4 turbo

F. Kirstein, T. Ruas, R. Kratel and B. Gipp, 'Tell Me What I Need to Know: Exploring LLM-based (Personalized) Abstractive Multi-Source Meeting Summarization', Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track