



## Large Language Models Create Useful, Accurate, Clear Summaries of Virtual Radiology Workgroup Meetings

Benjamin Mervak, MD, Michigan Medicine; Muhammad Bhalli, MBA; Tricia Niedbala, MBA; Kenneth Buckwalter, MD

### Introduction/Background

Remote meetings have dramatically increased in recent years. While convenient, these can disrupt team-based processes unless there is effective documentation of the discussion and follow-up tasks. Software solutions can generate a meeting transcript, although reviewing the entire transcript can be inefficient and tedious. A summary can be synthesized by a scribe, but this process is time-consuming, costly, susceptible to error, and may be delayed. Large Language Models (LLMs) are well suited for natural language processing and summarization tasks. This study investigates the performance of an LLM in extracting data from collaborative Radiology information technology (IT) workgroup meeting transcripts and generating a summary with a concise list of topics discussed, key points, and action items for each participant.

### Methods/Intervention

After obtaining IRB exemption, this study was conducted at an academic medical institution. Over two months in 2024, virtual Radiology IT group meetings were recorded and transcribed using teleconferencing software. Transcripts were ingested by an instance of an LLM (GPT-4 Turbo, OpenAI, San Francisco, CA) privately hosted on an institutional server. The LLM was prompted to generate a meeting summary including minutes and action items. LLM-generated summaries were provided to participants on the same day as the meeting. The accuracy, clarity, and usefulness of the LLM-generated summaries were evaluated by collecting anonymous feedback from meeting participants using 5-point Likert scales.

### Results/Outcome

Transcripts from 16 Radiology IT group meetings were summarized by an LLM. Feedback was received from 44 meeting participants. Most respondents (68-81%) rated the summaries as either “extremely” or “very” accurate, clear, and useful. A minority of respondents (9-12%) rated the summaries as either “not at all” or only “somewhat” accurate, clear, and useful (Table 1, Figure 1). “Usefulness” was rated highly despite any perceived inaccuracies or unclear statements generated by the LLM.

### Conclusion

LLMs can rapidly provide Radiology IT team members with meeting summaries that are accurate, clear, and useful.

### Statement of Impact

We show that LLMs can help address a key problem with virtual meetings – even those of a technical nature – by quickly and inexpensively providing participants with an accurate, clear, and useful meeting summary.

	#	% of Total
<b>Accuracy</b>		
Extremely accurate	9	20%
Very accurate	23	52%
Moderately accurate	8	18%
Somewhat accurate	2	5%
Not at all accurate	2	5%
<b>Clarity</b>		
Extremely clear	15	34%
Very clear	15	34%
Moderately clear	10	23%
Somewhat clear	1	2%
Not at all clear	3	7%
<b>Usefulness</b>		
Extremely useful	20	45%
Very useful	16	36%
Moderately useful	3	7%
Somewhat useful	3	7%
Not at all useful	2	5%

Table 1 – Analysis of Likert Scores on the accuracy, clarity, and usefulness of LLM-generated summaries across the 44 survey responses received from meeting participants.

Table 1 – Tabular analysis of Likert scores on the accuracy, clarity, and usefulness of LLM-generated summaries across the 44 survey responses received from meeting participants.

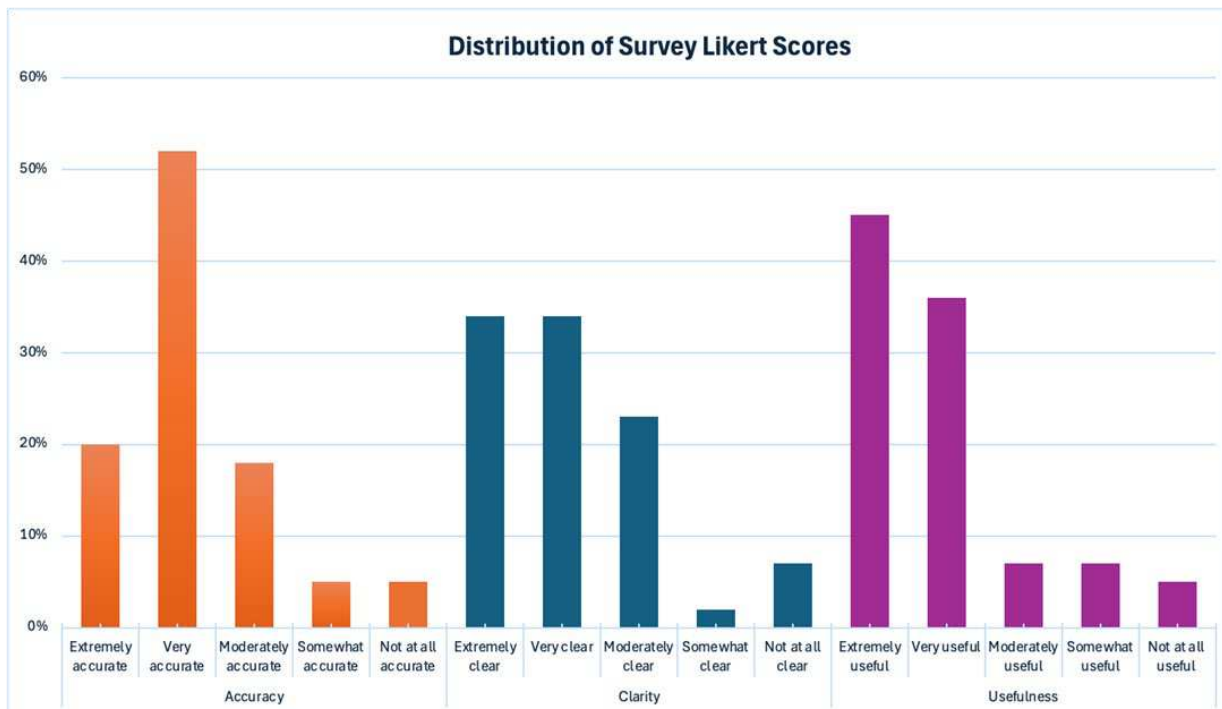


Table 1 – Graphical analysis of Likert scores on the accuracy, clarity, and usefulness of LLM-generated summaries across the 44 survey responses received from meeting participants.

## Keywords

Large language models; Virtual meetings; Meeting summary; Transcription