Multimodal Agentic System for Weekly SMART Goal Review

Iva Bojic[a], **Andy Khong**[a]

[a] *Nanyang Technological University, Singapore* iva.bojic@ntu.edu.sg, andykhong@ntu.edu.sg

* Presenting author

1. Introduction

Health coaching (HC) is a widely used intervention in chronic disease management, helping individuals set and achieve health-related goals [1]. A core component of HC is the Specific, Measurable, Achievable, Relevant, and Time-bound (SMART) goal framework, which provides structured, short-term objectives to support sustained behavior change [2]. While SMART goals are typically set on a weekly basis, HC sessions often occur weeks apart, creating gaps in SMART goal tracking, reinforcement, and patient engagement.

Without continuous monitoring, patients may struggle to stay accountable, leading to inconsistent progress [3]. However, frequent human-led HC sessions are often impractical due to resource constraints and scalability issues [4]. Digital health solutions, such as chatbots and mobile interventions, have attempted to bridge this gap but often lack contextual awareness and structured guidance, leading to generic and disengaging interactions [5]. There is a need for automated and personalized AI-driven support to enhance patient accountability and sustain weekly SMART goal reviews between human-led HC sessions.

To address this challenge, we propose a Multimodal Agentic System (MAS)—a hybrid human-AI approach that provides structured, AI-assisted weekly SMART goal reviews between human-led HC sessions. The proposed MAS leverages Large Language Models (LLMs) and AudioLLMs to process both text and speech-based HC data, ensuring continuity in SMART goal tracking. Unlike conventional digital health solutions, the proposed MAS is designed around agentic AI workflow [6, 7], where autonomous, task-specific AI agents collaborate to support patient engagement while keeping human coaches informed. The proposed workflow enables automated data retrieval, transcription, SMART goal extraction, and review facilitation, reducing reliance on frequent human-led interventions while maintaining personalization.

This paper presents the design and workflow of MAS, detailing how its specialized AI agents coordinate tasks across a two-phase system. First, we describe the architecture and functions of each AI agent, followed by an illustration of how their interactions create a structured goal-tracking workflow. By implementing a modular, multi-agent approach, the proposed MAS enhances patient engagement and supports structured SMART goal adherence while reducing the burden on human coaches.

2. Multimodal Agentic System: AI Agents

The MAS consists of seven AI agents, each responsible for a distinct function in the weekly SMART goal review process. These agents are categorized into two groups: *information extraction* agents, which process HC session data, and *SMART goal review and engagement* agents, which manage patient interactions and structured feedback. While each AI agent operates independently, they collectively ensure continuity in SMART goal tracking between human-led HC sessions.

2.1 Data Retrieval Agent (DRA)

The DRA is activated by the Orchestration Agent (OA) to retrieve HC notes from a SQL database, along with session metadata such as patient ID, health coach ID, session date, and audio recordings from a Google Cloud bucket. It maintains a record of its last execution time, ensuring that only new data is fetched in each run. This reduces computational overhead and ensures that downstream agents always process the latest session information.

2.2 Speech Processing Agent (SPA)

The SPA employs a multimodal cascade model to convert Singaporean English (Singlish) audio into English text [8]. The first stage performs automatic speech recognition using a fine-tuned Whisper-medium [9] for five epochs on two million samples from the Singapore National Speech Corpus [10]. The second stage refines the transcribed Singlish text using a fine-tuned Gemma-2-9B-it [11] for one epoch on twenty thousand segments from the HC sessions.

2.3 Memory Manager Agent (MMA)

The MMA leverages LLM schemas to extract structured information from unstructured HC data. Fine-tuned on over one hundred HC sessions and notes, it performs two primary tasks: *extracting SMART goals* from transcribed HC sessions and *identifying relevant personal details* such as family and hobbies from HC notes. The extracted information is stored in JSON format.

2.4 Orchestration Agent (OA)

The OA is a cron job executed at predefined intervals, responsible for scheduling and initiating key processes within the proposed MAS, including information extraction and weekly SMART goal review sessions, ensuring timely follow-ups between human-led HC sessions. At each execution, it first

triggers the DRA to fetch new HC notes, session metadata, and audio recordings, ensuring the latest session data is available for processing. Then, it determines if a patient is due for a SMART goal review by retrieving HC session dates from the MMA. If the required time has elapsed, the OA triggers the SOA to engage the patient and start the weekly SMART goal review session. Beyond scheduling, it continuously monitors session progress to ensure each step is completed. If a step is missed, the OA attempts to retry the process. If unsuccessful, it escalates the case to the health coach.

2.5 Session Opening Agent (SOA)

The SOA initiates the weekly SMART goal review session by establishing the conversation's context. It retrieves relevant session-opening details from the MMA, including personal information such as family and hobbies, to personalize the interaction. The SOA operates using a conversational LLM fine-tuned on over one hundred HC session openings, ensuring a natural and engaging dialogue. By incorporating patient-specific context, it facilitates a seamless transition into the SMART goal review process.

2.6 Goal Review Agent (GRA)

The GRA conducts the weekly SMART goal review by evaluating the patient's progress on previously set SMART goals. It retrieves the latest goals from the MMA and engages the patient in a structured conversation to assess progress, identify challenges, and provide personalized feedback. The GRA operates using a conversational LLM fine-tuned on over one hundred HC SMART goal review conversations. To ensure adaptive and context-aware responses, it incorporates retrieval-augmented generation, enabling it to reference past patient interactions and generate tailored feedback. By dynamically adjusting its responses based on patient input, the GRA facilitates guided reflection and problem-solving strategies to improve adherence to the SMART goals.

2.7 Session Closing Agent (SCA)

The SCA finalizes the weekly SMART goal review session by guiding the conversation toward a structured conclusion. It reinforces motivation, acknowledges the patient's efforts, and sets expectations for the next review. To achieve this, the SCA operates using a conversational LLM fine-tuned on over one hundred HC session closings. If a patient specifies a preferred timeframe for the next check-in, this information is relayed to the OA, enabling it to adjust scheduling and proactively trigger the next review at the agreed-upon time.

2.8 Summarization Agent (SA)

The SA processes the weekly SMART goal review session logs and generates a structured summary for the health coach. It employs a model specialized in structured dialogue summarization, ensuring factual accuracy and coherence. Fine-tuned on over one hundred HC summaries, the SA extracts key details such as SMART goal progress, challenges, and next steps while minimizing hallucination. By organizing information effectively, the SA provides health coaches with a clear and actionable overview of the patient's progress, enabling timely interventions if needed.

3. MAS Workflow

3.1 Information Extraction Phase

The information extraction phase ensures that all relevant HC session data is collected, processed, and stored in a structured format. The process begins when the OA prompts the DRA to fetch new HC notes, session metadata, and audio recordings. The DRA forwards HC notes and metadata to the MMA, which first extracts relevant personal details such as family and hobbies. Meanwhile, audio recordings are sent to the SPA for transcription and translation. Once processing is complete, the SPA sends the English HC sessions text to the MMA, which then extracts SMART goals from the received text.

3.2 SMART Goal Review and Engagement Phase

The SMART goal review and engagement phase enables structured AI-assisted weekly SMART goal reviews. The OA periodically checks for due reviews by retrieving session dates from the MMA. If a review is due, the OA triggers the SOA to initiate the SMART goal review session with the patient. The SOA retrieves relevant personal information from the MMA and engages the patient in conversation. After completing its task, the SOA hands control to the GRA, which first retrieves the SMART goals set during the last human-led HC session from the MMA. It then guides the patient through the SMART goal review process. Once the review is complete, the GRA transfers control to the SCA, which finalizes the session, confirming the timing for the next SMART goal review follow-up. Finally, all session logs are sent to the SA, which generates structured summaries for health coaches, ensuring they remain informed and can intervene if necessary.

4. Conclusion

This paper presents MAS, a system designed to conduct structured weekly SMART goal review sessions between human-led HC sessions. By leveraging AI agents for information extraction and SMART goal review, MAS ensures continuous patient engagement while reducing the burden on health coaches. Its modular agentic workflow enables efficient data processing, personalized patient interactions, and structured reporting. Future work will focus on evaluating the system's effectiveness in real-world deployments and refining AI agent interactions for greater adaptability.

Acknowledgments

This project was supported by CArdiovascular DiseasE National Collaborative Enterprise (CADENCE) National Clinical Translational Program (MOH-001277-01).

References

[1] Helen Lindner, David Menzies, Jill Kelly, Sonya Taylor, and Marianne Shearer. Coaching for behaviour change in chronic disease: a review of the literature and the implications for coaching as a self-management intervention. *Australian Journal of Primary Health*, 9(3):177–185, 2003.

[2] Edwin A Locke and Gary P Latham. Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American psychologist*, 57(9):705, 2002.

[3] Paul Krebs, James O Prochaska, and Joseph S Rossi. A meta-analysis of computer-tailored interventions for health behavior change. *Preventive medicine*, 51(3-4):214–221, 2010.

[4] Elliot G Mitchell, Rosa Maimone, Andrea Cassells, Jonathan N Tobin, Patricia Davidson, Arlene M Smaldone, and Lena Mamykina. Automated vs. human health coaching: exploring participant and practitioner experiences. *Proceedings of the ACM on human-computer interaction*, 5(CSCW1):1–37, 2021.

[5] Liliana Laranjo, Adam G Dunn, Huong Ly Tong, Ahmet Baki Kocaballi, Jessica Chen, Rabia Bashir, Didi Surian, Blanca Gallego, Farah Magrabi, Annie YS Lau, et al. Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association*, 25(9):1248–1258, 2018.

[6] Guibin Zhang, Kaijie Chen, Guancheng Wan, Heng Chang, Hong Cheng, Kun Wang, Shuyue Hu, and Lei Bai. Evoflow: Evolving diverse agentic workflows on the fly. *arXiv preprint arXiv:2502.07373*, 2025.

[7] Arthur Caetano, Kavya Verma, Atieh Taheri, Radha Kumaran, Zichen Chen, Jiaao Chen, Tobias Höllerer, and Misha Sra. Agentic workflows for conversational human-ai interaction design. *arXiv preprint arXiv:2501.18002*, 2025.

[8] Iva Bojic, Mathieu Ravaut, Hexin Liu, and Andy Khong. Bridging the gap: `Sing2Eng-MT`^{2} multimodal cascade model for singlish-to-english conversion. Under review at Interspeech 2025, 2025.

[9] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. Robust speech recognition via large-scale weak supervision, 2022.

[10] Jia Xin Koh, Aqilah Mislan, Kevin Khoo, Brian Ang, Wilson Ang, Charmaine Ng, and YY Tan. Building the singapore english national speech corpus. *Malay*, 20(25.0):19–3, 2019.

[11] Gemma Team, Morgane Riviere, Shreya Pathak, Pier Giuseppe Sessa, Cassidy Hardin, Surya Bhupatiraju, Léonard Hussenot, Thomas Mesnard, Bobak Shahriari, Alexandre Ramé, et al. Gemma 2: Improving open language models at a practical size. *arXiv preprint arXiv:2408.00118*, 2024.