Artificial Intelligence and Social Computing, Vol. 122, 2024, 124–136
https://doi.org/10.54941/ahfe1004649

An Agent-Based Framework for Conversational Data Analysis and Personal AI Systems

Bartosz Kurylek[1], Arthur Camara[1], Akash Nandi[2], and Evangelos Markopoulos[3]

[1]University College London, Bartlett School of Architecture, London, WC1H 0QB, United Kingdom
[2]Harvard University, Department of Economics, Cambridge, MA,02138, United States of America
[3]University of Turku, Department of Mechanical Engineering, Turku, FI-20014, Finland

ABSTRACT

This paper introduces a unique and novel framework entailing an agent-based-model (ABM) that utilizes the capabilities of large-language-models (LLMs) to capitalize on the most potent aspects of willingly volunteered recorded conversations; in particular, we combine the technologies to create innovative artificial personal intelligence systems. Our delineated framework is unique in that it can employ LLM-agents to process, organize, and extract insights from unstructured conversational data than has ever been achieved before more efficiently. Our cutting-edge system's architecture integrates both knowledge-data aggregation and agent-based conversational data extraction. We utilize LLMs to implement a method of aggregating knowledge data to reach the goal of achieving a dynamic, multi-level hierarchy that can organize information based on conceptual similarities and topical relevance. The effectiveness of our framework is very clearly established and legitimized through analysis of a survey in which respondents provided comprehensive conversational data.

Keywords: Human digital twin (HDT), Conversational data analysis, Large language models (LLMS), Personalized AI, Knowledge representation, Agent-based systems

INTRODUCTION

As Large Language Models have been evolving, a critical challenge lies in aligning these powerful models with user expectations and ensuring their full utilization (Jiaming Ji et al., 2021). This research aims to create a framework that supports the development of Artificial Personal Intelligence systems, building on recent advances in Human Digital Twin (HDT) technologies (Lauer-Schmaltz et al., 2024). Our work involves the usage of conversational data to augment general models with tailored knowledge, enabling precise adjustments for individual users or institutions.

Transcribed audio recordings of individuals talking provide a rich resource for understanding user preferences and thought patterns (Tannen, 1995), as analysing language, tone, and context can uncover hidden patterns and psychological factors (Duranti, 1997). This conversational data offers immense potential for personalization, allowing systems to provide more relatble, helpful, and emotionally intelligent interactions (Wright and McCarthy, 2008).

Recent breakthroughs in natural language processing, such as Large Language Models (LLM) (Bommasani et al., 2021), embedding techniques like BERT (Devlin et al., 2019), and Retrieval-Augmented Generation (Lewis et al., 2020), have unlocked new possibilities for handling vast amounts of unstructured conversational data. However, effectively processing and analysing this data remains a significant challenge due to its inherent complexity and diversity. Existing methods often struggle to capture the nuanced and context-dependent nature of conversational data, highlighting the need for more sophisticated and adaptive approaches (Ramesh et al., 2022).

The proposed framework utilizes LLM agents and embeddings to process, organize, and extract insights from conversational data efficiently. By developing novel techniques for data aggregation, hierarchical representation, and real-time access, our framework enables the creation of Artificial Personal Intelligence systems that adapt and personalize based on individual user needs, transforming the way we interact with AI systems and making them more empathetic and responsive.

SYSTEM ARCHITECTURE

A novel system for knowledge data aggregation and analysis is introduced (Figure 1). The system leverages embedding similarity and LLMs to cluster the data into meaningful and automatically defined categories. An agent-based framework and interface are also proposed to effectively analyze and utilize large volumes of conversational data.

Knowledge Data Aggregation

Effective organization is a crucial aspect of knowledge representation. While storing knowledge is essential, it is rendered useless if it cannot be easily found and utilized. Aggregating diverse data sources into a coherent and hierarchical knowledge structure presents a significant challenge, but it is vital for comprehending the available data and its potential applications. This aggregation process not only enhances data accessibility and readiness for visualization, but also facilitates the discovery of patterns and correlations within thematic knowledge areas.

METHODOLOGY OVERVIEW

This paper introduces an innovative data aggregation method that leverages LLM capabilities and embeddings to create a dynamic, multi-level hierarchical approach for efficient organization of information based on conceptual similarity and topical relevance.

From the basic information, documents or conversations, the topics are extracted and are the basis of the knowledge representation; from these a flexible dynamic similarity structure is built, and an empirical validation is made.

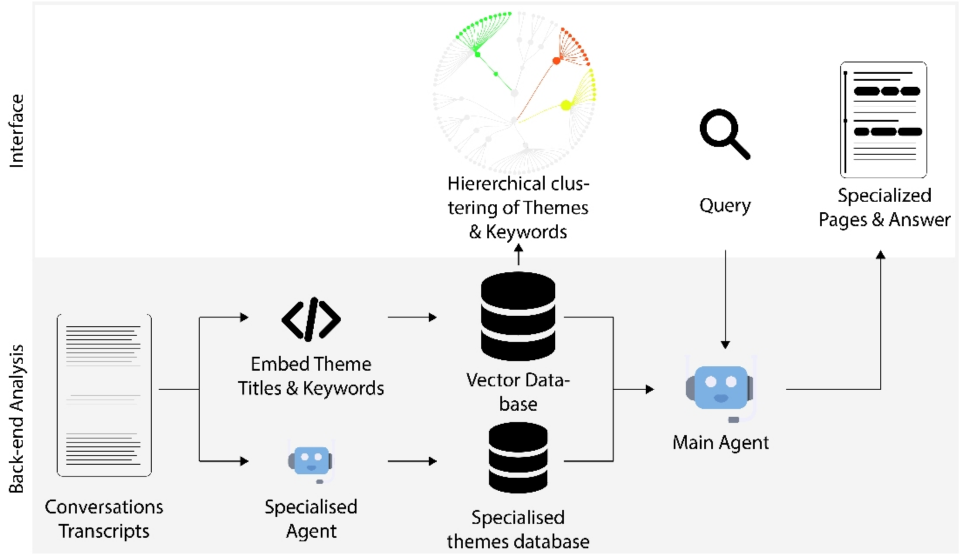

Figure 1: Proposed system architecture for conversational data analysis and personalized AI.

This serves as a fundamental contextualizing of the database, allowing a dynamic fractal-like representation that will provide a clear way to access and understand the data (see Figure 2).

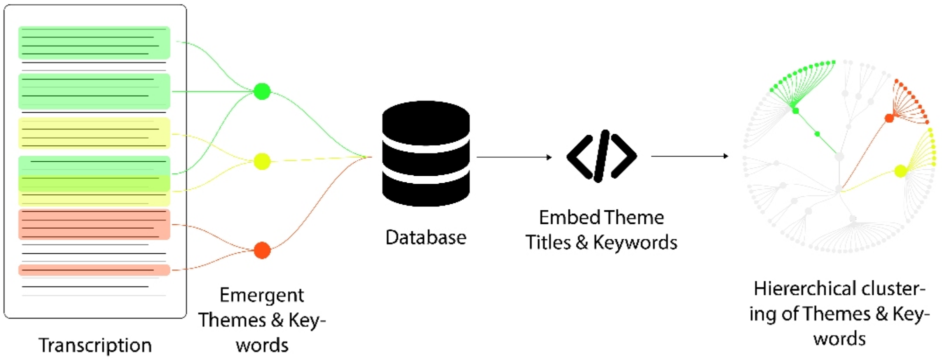

Figure 2: Flow diagram of the knowledge aggregation method.

The core of the method lies in creating a non-strict hierarchical knowledge structure where elements are organised across multiple levels based on their conceptual proximity to one another, as determined through embedding vectors (Li et al. 2020). The method utilises multilingual, pre-trained general language model embeddings of size 1536. This choice reflects a strategic balance between computational efficiency and the nuanced capture of semantic similarities across a diverse array of document types, subjects, and languages, facilitating a robust foundation for the hierarchical organisation of data.

HIERARCHICAL ORGANIZATION AND DYNAMIC NODE CREATION

This hierarchy starts with "level 0", which contains the topics extracted from the raw, real-world data. OpenAI's GPT 3.5 (Brown et al., 2020) was tasked with extracting relevant topics and keywords from the input data.

Subsequent levels aggregate concepts extracted from the precedent level, progressively generalising and finding more generic representations of the initial dataset, linking concepts with highest similarity (Figure 3).

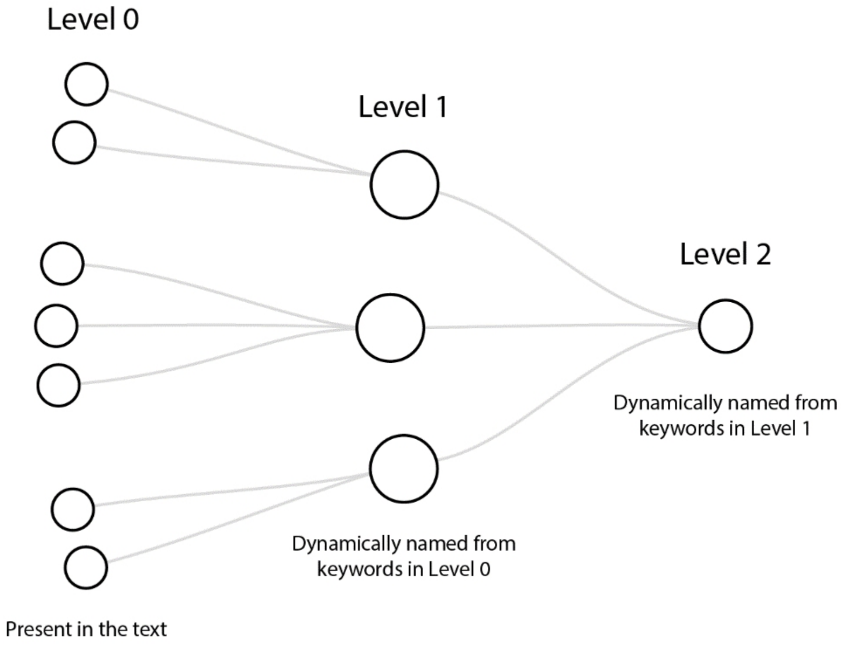

Figure 3: Dynamic nodes creation process.

Embeddings serve as the foundation for measuring conceptual similarity. By translating textual data into a vector space, embeddings facilitate the comparison and grouping of concepts based on their proximity within this space. The vectorial representation serves as both the node's identity and visual description.

To determine conceptual connections across hierarchical levels, we employ dynamically assigned similarity thresholds based on cosine similarity measures between embedding vectors of nodes. These thresholds are adaptively calibrated according to the distribution of element similarities within the previous level, incorporating an adjustment factor of 10% for each ascending layer. This ensures a balance between preserving distinct concepts at lower levels and fostering meaningful aggregation at higher levels, thereby maintaining a structured yet adaptable hierarchy.

Where an element lacks a sufficiently similar concept above it, a new hierarchical node is formed. This ensures the structure's adaptability to new concepts and provides a dynamic growth potential.

In the case the concept is similar to more than one concept in the upper layer, it connects to all that fits the required threshold limit. Such an approach allows for a nuanced aggregation that acknowledges and preserves the multifaceted nature of data relationships. This is a key aspect of the method as it foments more complex topic interactions, enriching the structure's informational depth, which is fundamental for knowledge expression.

When a node is copied to a higher level in the hierarchy, GPT-3.5 is employed to understand and describe the node's content. If no sufficiently similar concepts are found in the higher level, the node becomes available for inclusion in the subsequent nodes. GPT-3.5 generates a new, more general label for the node by summarizing the aggregated topics. This updated label is then vectorized and used as the node's new identity, reflecting its expanded conceptual scope. The update process is triggered when the node's composition changes significantly, exceeding a predefined threshold of 20%.

REAL DATA CLUSTERING

The practical application of our proposed framework is best illustrated through empirical analysis. In this section, we delineate the real-world functionality of our model by detailing its application in the analysis of diverse conversational datasets. Three distinct conversation types were selected to challenge the adaptability and accuracy of our framework:

- Start-up presentation pitch, organised and pre-prepared conversation
- Casual conversations, much less organised and chaotic without clear knowledge to extract
- University lectures, topic specific and reach in detailed information

Figures 4 and 5 illustrate the system's capability to cluster conversational data into a cohesive structure. Figure 4 presents an overarching view, highlighting the clustering based on thematic connections. The radial tree visualization was adapted from the "Zoomable Radial Tree" D3 code by Azevedo (2022), available on the Observable platform.

Figure 5 provides a detailed look into the hierarchy's lower tiers, where foundational nodes, derived from key terms in conversations, ascend into broader thematic nodes.

Colour-coding aids navigation: recent additions are marked in pink, signalling the latest data integrated into the system. Node size is proportional to the number of associated keywords, allowing users to gauge the scope of topics at a glance.

The core objective of this system extends beyond categorisation of knowledge. It aims to transform disordered and fragmented data into a structured format for the user. This enables not just the retrieval of specific information or conversations, but also facilitates the discovery of related content that may have been overlooked or undervalued. In doing so, it harnesses historical dialogues and accumulated knowledge, turning them into valuable resources for current use.

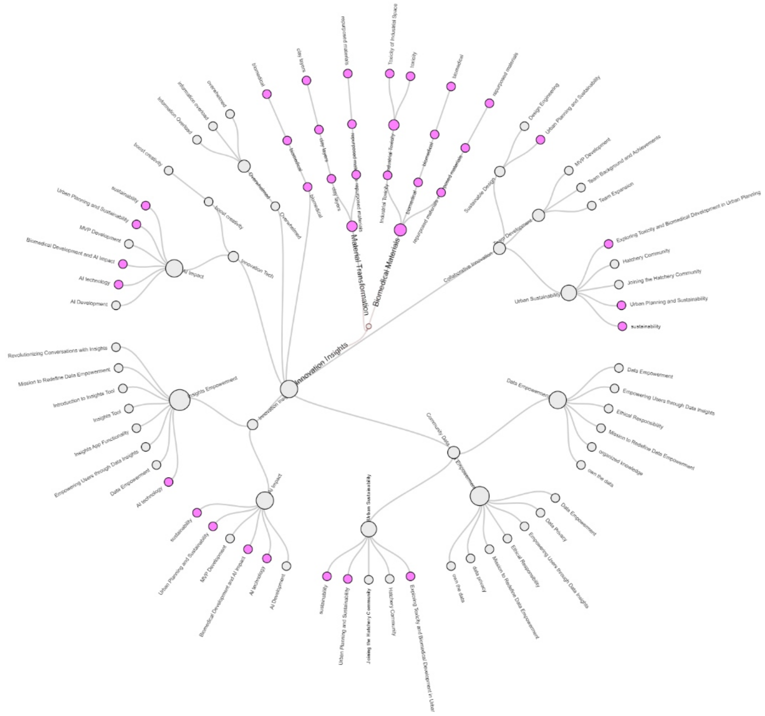

Figure 4: Real data graph of aggregated knowledge. Overview of knowledge clusters.

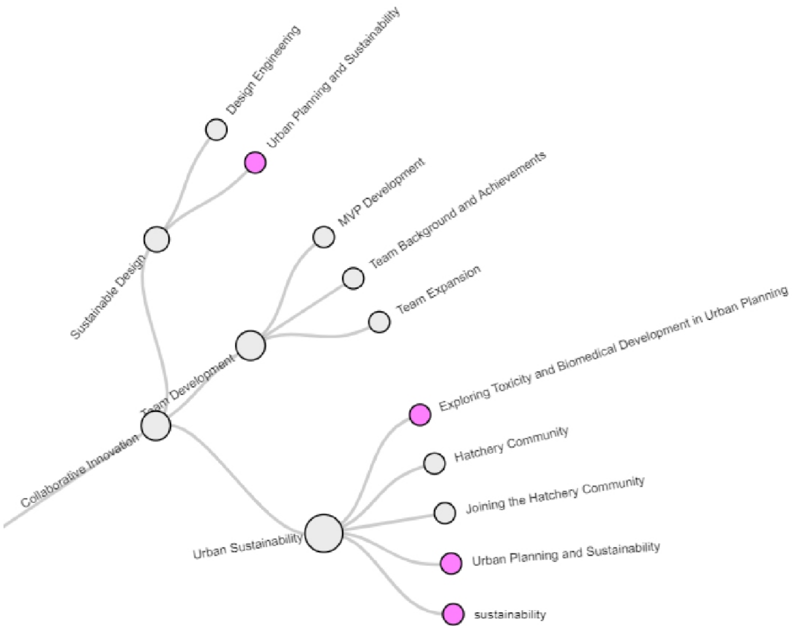

Figure 5: Real data knowledge aggregation. Detailed view of a data branch.

INNOVATIVE CONTRIBUTIONS AND THEORETICAL IMPLICATIONS

The proposed methodology contributes to the field of data aggregation and organization through its adaptability and generalization capabilities. Unlike supervised techniques limited by pre-defined categories (Alloghani et al., 2020) or clustering algorithms with fixed cluster counts, our approach leverages a dynamically evolving hierarchical clustering foundation. This enhances computational efficiency and provides a scalable framework for knowledge representation. Key innovations include the following:

- Dynamic similarity thresholds: Our method employs adaptive thresholds that adjust based on data diversity across levels, enabling refined aggregation that mirrors real-world data complexity.
- Embedding-based semantic similarity: We utilize embeddings and cosine similarity to measure semantic distances, surpassing conventional distance metrics and capturing conceptual relationships more effectively.
- LLM-driven node naming: By integrating LLMs for node name generation, our approach provides descriptive, meaningful topic representations, addressing a gap in traditional methods (Weston et al., 2023).
- Multi-membership model: Our framework allows elements to associate with multiple higher-level nodes, reflecting nuanced data interconnectedness and surpassing the limitations of binary hierarchical structures.

The flexibility and scalability of our approach stems from strategic architectural choices, including embedding-based similarity measures, LLM-driven knowledge synthesis, and adaptive threshold adjustments. These elements enable dynamic construction of hierarchical frameworks that can articulate complex relationships within large, evolving datasets.

AGENT BASED CONVERSATIONAL DATA EXTRACTION AND VISUALISATION

The proposed framework leverages an LLM Agent equipped with tools to handle user queries efficiently. The Main Agent analyses the query and categorizes it into two types: simple queries and specialized theme queries.

For simple queries, the agent employs Retrieval-Augmented Generation (RAG) to search for relevant information within the vector database and provide a context-relevant answer.

In the case of specialized theme queries, which require a specific angle or perspective on the dataset, the Main Agent creates a Specialized Agent to thoroughly scan the transcripts and extract pertinent qualitative or quantitative data. The Specialized Agent compiles the extracted information into a dedicated "specialized theme" dataset and generates an accessible page. This dataset and page act as a foundation and context for future inquiries related to the specific theme, allowing users to understand their conversations from the desired angle.

If the Main Agent encounters a query that necessitates a new specialized theme, it dynamically creates a new Specialized Agent to handle the task. This flexible approach enables the framework to perform tailored analyses on the original transcripts and organize the results into a centralized database and page (see Figure 6).

By focusing LLM agents on specific aspects of the text, the system can extract valuable information across various domains, such as cognitive and emotional analysis, relationship dynamics, productivity and goal management, knowledge retrieval, and behavioural insights.

For example, by examining language use, sentiment, and topic frequencies, agents can assess the user's thought processes, decision-making patterns, and emotional states (Serban et al., 2018), that could lead to more personalized and context-aware AI assistance.

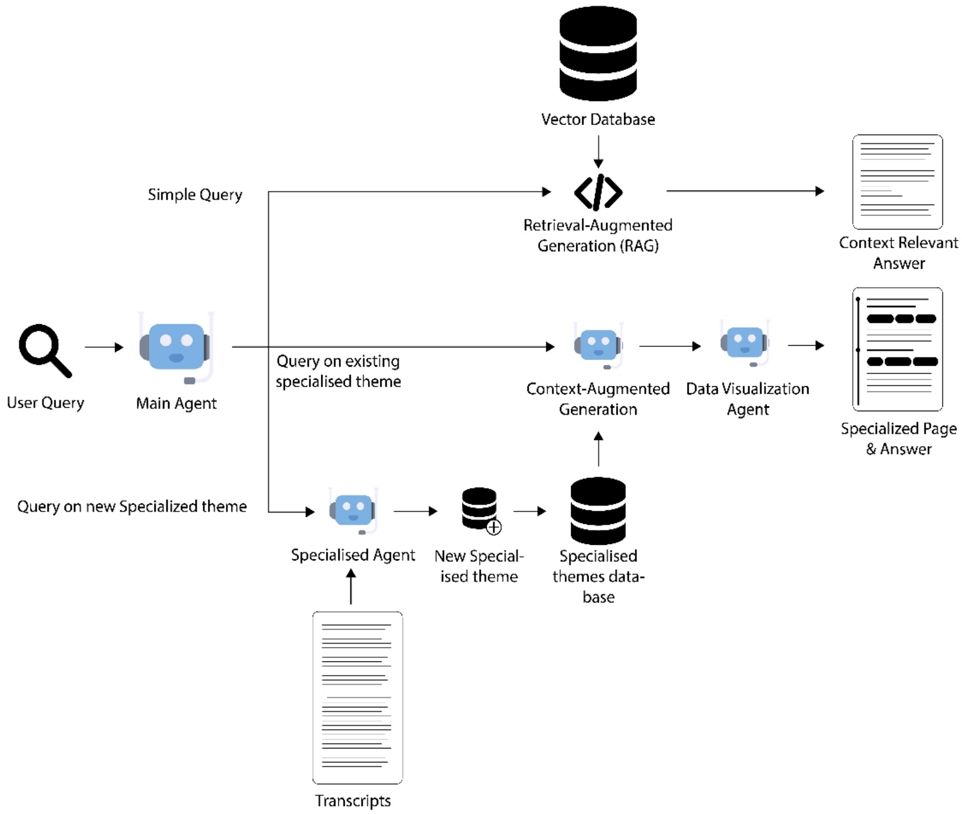

Figure 6: System architecture for conversational data management and specialized theme extraction.

PRACTICAL IMPLEMENTATION WITH A MOBILE APPLICATION

The proposed framework is being implemented into a mobile application currently under development and testing. The app's main goal is to help users track past conversations and extract valuable metrics and insights from their speech, supporting personal goal monitoring, memory aid, and self-reflection.

Mock-ups of the app's user interface (Figure 7) highlight the framework's integration into a user-friendly design, facilitating easy navigation and access to personalized insights.

The mobile app serves as a testbed for the agent-based framework, evaluating its performance in handling diverse conversational data and delivering personalized AI assistance. Insights gained from the app's deployment will contribute to ongoing framework development.

EMPIRICAL DATA COLLECTION AND ANALYSIS

This section reviews the findings from the empirical testing of the conversation capture and analysis app, which was conducted with an initial sample group of 24 participants, predominantly university students (82%). The primary goal of this testing was to evaluate the app's user interface and the accuracy of its conversation analysis features, as well as get initial feedback on development process.

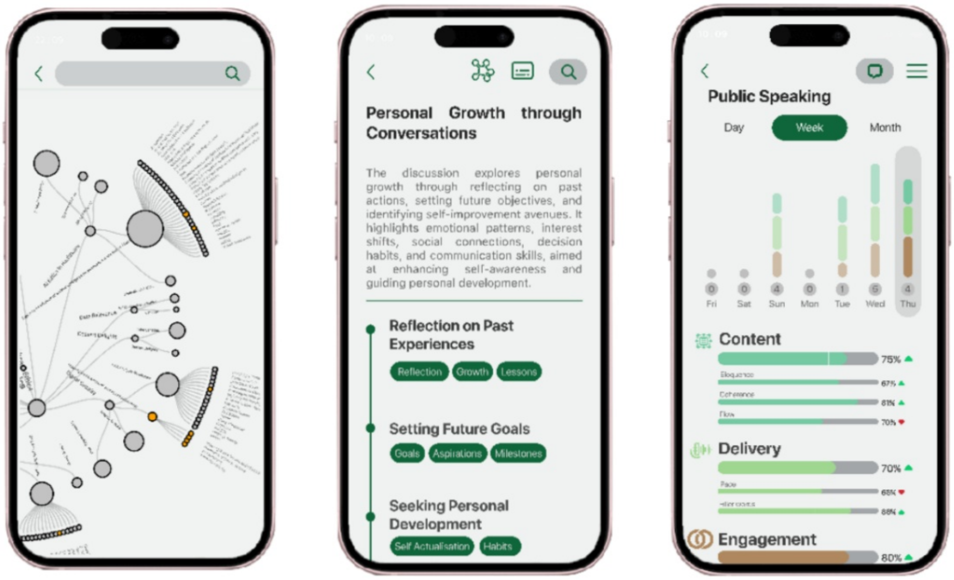

Figure 7: Mock-ups of and framework implemented into an app.

Data Collection

Participants were instructed to use the app during their routine conversations, ensuring that the feedback reflected its performance in real-world scenarios. After recording their interactions, participants were provided with a graphical representation of their conversations, including aggregated topics for a comprehensive review. Targeted questions were used to assess the app's effectiveness based on these visualisations.

Data Analysis

User feedback was collected and analysed to identify the app's strengths and areas for improvement, using the Net Promoter Score (NPS) as the main metric. NPS is calculated by subtracting the percentage of detractors (scores 0-6) from the percentage of promoters (scores 9-10) on a scale of 1 to 10. The insights gained will inform future enhancements to ensure the app continues to provide value in managing and analysing conversational data. Relative to the software industry standards, benchmarks can be defined as indicated in Table 1:

Table 1. NPS benchmarks for the software industry (Survicate, 2023).

Low (25 percentile)	Average	High (75 percentile)
NPS 9	NPS 27	NPS 50

The conversation captures and analysis of the app's initial NPS data reveals its core strengths in accurately summarizing conversations (NPS 43) and division into relevant topics (NPS 33), surpassing the industry average of 27. However, the app's UI/UX design needs improvement, with the intuitiveness of recording and accessing conversations receiving an NPS of 27. The chat response feature received an NPS of 26, indicating room for enhancement, with 18% of respondents classified as detractors. Despite these areas for improvement, 92% of respondents rated the overall quality of the topics and keywords extracted by the app as either extremely well (57%) or somewhat well (35%), underlining the app's strength in identifying and presenting the most important aspects of conversations.

The feedback from open-ended questions revealed that users desire the ability to access specific conversation parts related to the app's answers, with highlighted transcription sections. They also expressed interest in comparing different conversation recordings to identify patterns and track changes over time. Respondents indicated a wide range of potential use cases for the app, including project planning, note-taking, lecture summarization, brainstorming, personal development, and memory recording, suggesting the app's versatility in catering to diverse user needs.

The level of clarity of the knowledge graph was rated with an NPS of 18, and the grouping of keywords and topics received an NPS of 13. These scores are lower compared to other features of the app. However, it is important to note that participants were asked to rate a pre-generated graph based on multiple conversations, as they only recorded a single conversation during the study. This limitation may have impacted the representativeness of the data and the users' understanding of the graph.

Despite the areas for improvement of the current visualization, users found it helpful. Users' suggestions include showing the flow of topics to illustrate the order in which subjects were discussed and the proportion of time spent on each one. They also proposed adding colours to differentiate between topics and using node size to indicate the frequency of keyword appearances.

These features would make it easier for users to understand the information presented in the visualization at a glance.

Overall, the tool received an impressive NPS of 52, indicating strong user interest in daily usage for tracking information and progress across various aspects of life. 48% of respondents, expressed a desire to use the tool in a personal setting, followed by 30% in educational contexts and 22% professionally.

Open-ended responses highlighted the tool's potential to enhance learning through notetaking and summarization, support memory recall, and foster self-discovery and personal development by analysing thoughts and interactions. Its versatility in aiding brainstorming, organizing ideas, and tackling complex topics was also noted. Respondents emphasized the importance of privacy and customization features to cater to the personal nature of the tool's usage.

LIMITATIONS AND AREAS OF FURTHER RESEARCH

The framework has several limitations that should be addressed in future research. The scalability and performance of the framework under large-scale, real-world conditions remain to be fully tested, as the current study could not extensively evaluate the framework's ability to handle a large volume of conversations and the effectiveness of the agent-building tools. Additionally, the user interaction testing conducted in this study was preliminary, focusing on features that are ready for deployment in the near term.

Future research should focus on conducting extensive real-world testing, employing bias mitigation techniques, and performing user-centred evaluations to optimize the framework's usability and effectiveness. Survey participants expressed interest in more granular text analysis and extraction capabilities, such as highlighting specific sections of text. Long-term user studies involving participants using the framework for extended periods would provide valuable insights into its real-world performance and user experience.

CONCLUSION

This research introduces an agent-based framework that leverages Large Language Models (LLMs) to process and extract insights from conversational data. The framework's integration of knowledge data aggregation and agent-based conversational data extraction creates a dynamic, multi-level hierarchy for organizing information and generating specialized theme datasets. One of the framework's key strengths is its flexibility and adaptability, enabling effective management and extraction of valuable insights from unstructured conversational data.

Despite the acknowledged limitations, the framework's novel approach to managing and extracting insights from conversational data offers a promising direction for future research in conversational AI and personalized AI systems. The preliminary user testing conducted in this study indicates that participants found value in the framework and were excited about its potential applications in their lives. This positive response underscores the importance of continued application development and future research to further refine and enhance the framework. This framework has the potential to significantly contribute to the development of Artificial Personal Intelligence.

REFERENCES

Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A. and Aljaaf, A. J. (2020) 'A systematic review on supervised and unsupervised machine learning algorithms for data science', in Supervised and unsupervised learning for data science. Springer, Cham, pp. 3–21.

Azevedo, B., 2022. Zoomable Radial Tree. [online] Observable. Available at: https://observablehq.com/@brunomiguelam/zoomable-radial-tree/2 (Accessed 17 Feb. 2024).

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S.,... & Liang, P. (2021) 'On the opportunities and risks of foundation models', arXiv preprint arXiv:2108.07258.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I. and Amodei, D. (2020) 'Language Models are Few-Shot Learners', Advances in Neural Information Processing Systems, 33, pp. 1877–1901.

Devlin, J., Chang, M. W., Lee, K. and Toutanova, K. (2019) 'BERT: Pre-training of deep bidirectional transformers for language understanding', arXiv preprint arXiv:1810.04805.

Duranti, A. (1997) Linguistic anthropology. Cambridge University Press.

Ji, J., Zou, Y., Yu, Y., Jiang, R., Zhang, J., Tang, J., Tian, F., Wang, W. and Fu, X. (2021) 'AI Alignment: A Comprehensive Survey', arXiv preprint arXiv:2111.05196.

Lauer-Schmaltz, M. W., Cash, P., Hansen, J. P., & Maier, A. (2023). Towards the Human Digital Twin: Definition and Design – A Survey. arXiv preprint arXiv:2402.07922.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., & Kiela, D. (2020) 'Retrieval-augmented generation for knowledge-intensive NLP tasks', arXiv preprint arXiv:2005.11401.

Li, B., Zhou, H., He, J., Wang, M., Yang, Y. and Li, L. (2020) 'On the sentence embeddings from pre-trained language models', arXiv preprint arXiv:2011.05864.

Ramesh, K., Ravishankaran, S., Joshi, A. and Chandrasekaran, K. (2022) 'A survey on explainable artificial intelligence (XAI) in natural language processing: Techniques, challenges, and future directions', Information Fusion, 91, pp. 171–196.

Serban, I. V., Lowe, R., Henderson, P., Charlin, L. and Pineau, J. (2017) 'A survey of available corpora for building data-driven dialogue systems', arXiv preprint arXiv:1512.05742.

Survicate (2023) NPS Benchmarks 2023: What is a Good Net Promoter Score? Available at: https://survicate.com/nps-benchmarks/ (Accessed: 11 April 2024).

Tannen, D. (1995) 'Waiting for the Mouse: Constructed Dialogue in Conversation', in the dialogic emergence of culture. Urbana: University of Illinois Press, pp. 198–217.

Weston, S. J., Shryock, I., Light, R. and Fisher, P. A. (2023) 'Selecting the Number and Labels of Topics in Topic Modeling: A Tutorial', Advances in Methods and Practices in Psychological Science, 6(2). doi:10.1177/25152459231160105.

Wright, P. and McCarthy, J. (2008) 'Empathy and experience in HCI', in Proceedings of the SIGCHI conference on human factors in computing systems. Florence, Italy: Association for Computing Machinery (CHI '08), pp. 637–646. doi:10.1145/1357054.1357156.