Echoes of the Past: AI-Enabled Simulation of Cherished Memories

PENG YIYANG* and ZHANG RUIMING*, The Hong Kong University of Science and Technology (Guangzhou)

As the technology of large language models (LLMs) advances, their application in dialogue systems has deepened, particularly in simulating individual conversational styles, where they show significant potential. However, current dialogue systems often lack a deep understanding of individual memory and personality traits, struggle with multimodal input, and have limitations in simulating conversational styles. This study aims to achieve high-precision simulation of individual chat styles through fine-tuning multimodal large language models, providing solace for the bereaved and a digital legacy for each person. We propose a novel fine-tuning method using personal WeChat chat records and customized personal information questionnaires as training data. The base model is Qwen-VL-7B, and we apply the LoRA fine-tuning technique to specific layers of the model to enhance its learning and memory capabilities regarding personal information. Experimental results indicate that the model performs well in conversational style similarity, adaptability to different users' chat styles, and multimodal input processing. This study successfully implements a multimodal dialogue system capable of simulating individual chat styles, offering new directions for future research in digital heritage and personal digital avatars. Despite the positive progress, the model currently only processes images and text and generates text output, lacks a more objective style similarity evaluation standard, and has insufficient understanding of the evolution of thoughts in an individual's life experiences.

ACM Reference Format:
PENG Yiyang and ZHANG Ruiming. 2024. Echoes of the Past: AI-Enabled Simulation of Cherished Memories. 1, 1 (May 2024), 3 pages.

1 INTRODUCTION

With the rapid development of large language models in recent years, their dialogue capabilities have become remarkably adept, capable of addressing complex issues such as knowledge QA and role-playing. This has led to attempts to challenge the ultimate fate of humanity—death—with the aid of LLM technology.
Here is a poignant quote about the concept of death:

> There are three deaths. The first is when the body ceases to function. The second is when the body is consigned to the grave. The third is that moment, sometime in the future, when your name is spoken for the last time. — David Eagleman, *Sum: Forty Tales from the Afterlives*

If personal data could be used to train large language models to possess similar memories, conversational styles, and thinking logic to individual humans, creating a digital clone that persists posthumously, it would allow people to continue interacting with the deceased. This could alleviate the immense grief of loved ones and ensure that life's legacy is not confined to biological existence but echoes through digital eternity.

Most commercial products can only mimic an individual's appearance and voice, with interactions limited to text input and model-generated speech. Such products fall short of true AI cloning requirements. In academia, personal documents like published books and emails have been used to enhance content retrieval[5]. Others have designed questionnaires to capture personality traits and fine-tune models for personalization[3]. These methods enable better simulation of memory and thought processes but are not ideal for chat simulation due to the format discrepancy between chat and documents like books and emails. Moreover, these models only support text input, overlooking the visual information prevalent in real-life chats, such as facial expressions in face-to-face conversations or images shared online, which significantly influence dialogue.

This work addresses the need for multimodal data processing by fine-tuning a large multimodal model using personal chat records and questionnaires. The resulting model can handle multimodal data, simulate personal memory and personality traits, and learn differentiated chat styles and response mechanisms for various social circles and environments, such as private and group chats.

2 METHOD

This section introduces data collection and preprocessing, including data from WeChat chats and a customized personal information questionnaire. It also outlines the multimodal large model, fine-tuning method, and parameter settings to be used.

2.1 Data Collection and Preprocessing

WeChat chat records serve as the primary dataset for simulating conversational styles. Considering individuals exhibit different chat styles across various environments and with different interlocutors, each dialogue's context and participants are injected into the system as input prompts.

Due to limited chat records for some users, the injected object information includes not only the WeChat ID but also the relationship type (friend, classmate, elder, etc.), allowing the model to simulate conversational styles based on relationship types even when chat data is scarce.

Dialogues are segmented on the basis of a time threshold; if the interval between consecutive messages exceeds this threshold before reaching the model's maximum input length, they are considered separate dialogues.

The model employs different reply strategies for private and group chats. In private chats, the model responds to each message sent by the interlocutor, while in group chats, it selectively replies. Selective replies are derived from negative sampling of group chat data; if the

*Both authors contributed equally to this research.

Authors' Contact Information: PENG Yiyang, ypeng455@connect.hkust-gz.edu.cn; ZHANG Ruiming, rzhang346@connect.hkust-gz.edu.cn, The Hong Kong University of Science and Technology (Guangzhou).

ACM XXXX-XXXX/2024/5-ART
https://doi.org/

simulated subject does not reply to a message in the chat record, it is treated as a negative example, and the model generates a "reject answer."

We also design a questionnaire to collect personal information which covers:

- Background information: educational history, professional background, and upbringing.
- Personal preferences: Favorite games, movies, and books.
- Personality traits: MBTI questionnaire responses and the resulting personality assessment. [4]

2.2 Finetuning

We select Qwen-VL-7B[1] as the base model and apply the LoRA[2] fine-tuning technique, targeting specific layers within the Transformer architecture, including the QKV matrices and the output MLP structure, as well as the gated feedforward neural network. Hyperparameters include:

Table 1. Model Parameters

Parameter	Value	Description
lora_r	64	Dimension of the low-rank approximation
lora_alpha	16	Alpha parameter for the low-rank adaptation
lora_dropout	0.05	Dropout rate for regularization
learning_rate	1×10^{-5}	Initial learning rate for the optimizer
weight_decay	0.1	Weight decay coefficient for L2 regularization
adam_beta2	0.95	Beta2 parameter for Adam optimizer
warmup_ratio	0.01	Warmup ratio for the learning rate scheduler
lr_scheduler_type	cosine	Type of learning rate scheduler

3 RESULTS

The model's effectiveness is measured in three aspects: similarity to the individual's conversational style, adaptability to different users' styles, and the ability to process multimodal inputs and generate personalized responses.

3.1 Conversational Style Similarity

We manually crafted dialogues not present in the training set and some with special significance to test style similarity.

In this test, the first dialogue about playing games. The fine-tuned personal profile mentions a love for games, prompting an expected reply from the model. The second query involves professional knowledge, and the model provides an answer consistent with the professional background. The third query involves the number "6," which in online chats has taken on complex linguistic connotations beyond its numerical meaning, sometimes carrying a slightly aggressive tone. The model's response indicates an understanding of this nuanced term.

Table 2. Dialogue Examples

Query	Answer
Play game?	Yeah Yeah Yeah
Is it possible to directly use someone's social media profiles, diaries, etc., for RAG without fine-tuning a pre-trained model?	I directly use the chat history for RAG
6	6

3.2 Different Chat Styles for Different Users

Despite receiving identical dialogue prompts, the model exhibits versatility in its responses, tailored according to the context set by the relationship between the interacting user and the simulated entity. An illustrative scenario involves a uniform query posed by distinct individuals, eliciting vastly dissimilar answers from the model. This behavioral nuance underscores the model's capability to infer distinct personas and adapt its demeanor accordingly, thereby demonstrating a sophisticated understanding of personalized interactions based on the perceived identity of the interlocutor.

Table 3. Dialogue Examples

Query	Answers
Out for dinner?	• User 1: Busy now • User 2: No class today

3.3 Multimodal Processing Ability

To evaluate the model's multimodal processing proficiency, we assess its capacity to interpret combined textual and visual cues and respond coherently with respect to an individual's communication patterns. In this particular assessment, the model confronts a scenario where it receives a query accompanied by an image (1), requiring it to discern the image content and align its response with the user's known culinary preferences. The model successfully identifies the dish depicted in the image and selects "The one with minced meat and spicy oil," illustrating its capability to integrate visual information into contextually appropriate responses.

Table 4. Dialogue Examples

Query	Answers
<img> Which one would you like ?	The one with minced meat and spicy oil.

4 CONCLUSION

This study successfully achieved high-precision simulation of individual chat styles through fine-tuning a multimodal large language

Fig. 1. image in query

model, providing new perspectives and methods for research in the fields of digital heritage and personal digital avatars. The Qwen-VL-7B model, combined with the LoRA fine-tuning technique, not only enhanced the model's ability to learn and remember personal information but also significantly improved its performance in conversational style similarity, adaptability to different user styles, and multimodal input processing.

Firstly, the model's performance in conversational style similarity demonstrated its effectiveness in capturing and replicating individual conversational characteristics. Through a carefully designed personal information questionnaire and WeChat chat records, the model learned the user's personalized traits and reflected these in dialogue.

Secondly, the model showed good adaptability to different chat styles for different users, indicating its ability to adjust its reply strategy according to social relationships and chat environments. This flexibility enhances the model's practicality and utility.

Lastly, the model's ability to process multimodal inputs is a significant breakthrough, allowing it to more comprehensively understand and respond to user queries, which is a challenge for text-centric dialogue systems.

Despite the positive outcomes, the study has limitations and areas for future research. The current model only processes images and text and generates text output, restricting its application in richer interactive scenarios. Additionally, an objective style similarity evaluation standard is lacking, and current assessments rely on subjective judgments. Future work will explore more modalities, such as speech and video, and develop more precise evaluation tools to quantify the model's performance. Furthermore, the model's understanding of the evolution of an individual's ideological development over their life experiences is still limited. This necessitates further enrichment of training data and improvement of the model's learning algorithms in future work. We believe that with ongoing technological advancements and in-depth research, multimodal large language models will play an increasingly significant role in the realms of digital heritage and personal digital avatars, providing a richer and more enduring medium for human memory and emotions. This research has laid the groundwork for a new approach to digital legacy creation, where the nuances of an individual's communication style can be preserved and perpetuated through advanced AI technology. The potential applications of this technology are vast, extending beyond personal use to areas such as customer service, where personalized AI clones could provide tailored experiences, or in education, where they could simulate personalized tutoring sessions. While the current model has shown promise, there are challenges to be addressed. The integration of additional modalities like voice and video will require the development of new algorithms capable of processing and understanding these complex inputs. Moreover, the creation of a robust and objective evaluation framework for style similarity is essential to measure the success of these models accurately and guide further refinement. As we move forward, the ethical considerations of creating digital replicas of individuals must also be carefully weighed. Issues of consent, privacy, and the potential for misuse of digital personas must be addressed to ensure that this technology is developed and deployed responsibly. In conclusion, the successful simulation of individual chat styles through multimodal large language models represents a significant step towards a more personalized and interactive AI future. It offers a glimpse into a world where the digital and human experiences can be seamlessly intertwined, enriching our connections and preserving our legacies in the digital realm.

REFERENCES

[1] Bai, J., Bai, S., Yang, S., Wang, S., Tan, S., Wang, P., Lin, J., Zhou, C., and Zhou, J. Qwen-vl: A versatile vision-language model for understanding, localization, text reading, and beyond, 2023.
[2] Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., and Chen, W. Lora: Low-rank adaptation of large language models, 2021.
[3] Ng, M. T., Tse, H. T., tse Huang, J., Li, J., Wang, W., and Lyu, M. R. How well can llms echo us? evaluating ai chatbots' role-play ability with echo, 2024.
[4] Pan, K., and Zeng, Y. Do llms possess a personality? making the mbti test an amazing evaluation for large language models, 2023.
[5] Salemi, A., Mysore, S., Bendersky, M., and Zamani, H. Lamp: When large language models meet personalization, 2024.