iPET: An Interactive Emotional Companion Dialogue System with LLM-Powered Virtual Pet World Simulation

Zheyong Xie[1], Shaosheng Cao[1*], Zuozhu Liu[2], Zheyu Ye[1], Zihan Niu[3], Chonggang Lu[1], Tong Xu[3], Enhong Chen[3], Zhe Xu[1], Yao Hu[1], Wei Lu[4*]
[1]Xiaohongshu Inc., [2]Zhejiang University, [3]University of Science and Technology of China
[4]Singapore University of Technology and Design
{xiezheyong, caoshaosheng, zheyuye, chongganglu}@xiaohongshu.com
{qiete, xiahou}@xiaohongshu.com {tongxu, cheneh}@ustc.edu.cn,
niuzihan@mail.ustc.edu.cn, zuozhuliu@intl.zju.edu.cn, luwei@sutd.edu.sg

Abstract

The rapid advancement of large language models (LLMs) has unlocked transformative potential for role-playing emotional companion products, enabling systems that support emotional well-being, educational development, and therapeutic applications. However, existing approaches often lack sustained personalization and contextual adaptability, limiting their effectiveness in real-world settings. In this paper, we introduce iPET[1], an LLM-powered virtual pet agent designed to enhance user engagement through rich, dynamic pet behaviors and interactions tailored to individual preferences. iPET comprises three core components: a dialogue module that instantiates virtual pet agents for emotionally interactive conversations; a memory module that stores and synthesizes records of both agent and user experiences; and a world simulation module that generates diverse, preference-driven pet behaviors guided by high-level reflections. Deployed for over 200 days in a real-world, non-commercial product, iPET has served millions of users – providing emotional support to psychologically distressed individuals and demonstrating its effectiveness in practical applications.

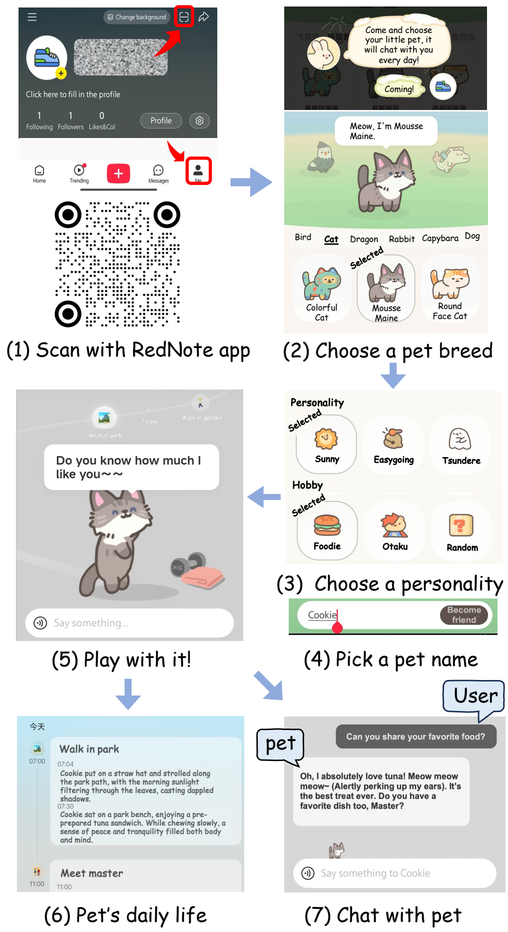

Figure 1: Workflow of the iPET system.

1 Introduction

Recent advances in large language models (LLMs) (Chang et al., 2024; Achiam et al., 2023; Bai et al., 2023; Touvron et al., 2023) have unlocked new possibilities in emotional dialogue and role-playing systems (Tseng et al., 2024; Chen et al.; Shanahan et al., 2023; Liu et al., 2024a), with promising applications in emotional well-being (e.g., alleviating loneliness and anxiety) (Liu et al., 2021; Zhang et al., 2024a), educational development (e.g., fostering empathy through simulated teaching) (Li et al., 2023; Wang et al., 2024), and therapeutic applications (e.g., supporting cognitive behavioral therapy) (Yan et al., 2023; Shen et al., 2024; Kian et al., 2024). Commercial platforms such as Xingye AI[2] and Character AI[3] have also emerged, offering human-friendly interactions with virtual characters for immersive and personalized emotional support and experiences.

The deployment of emotional dialogue systems in real-world scenarios presents three critical challenges. First, sustaining long-term interactions be-

*Corresponding author
[1]https://xhslink.com/L8wKw6

[2]https://www.xingyeai.com/
[3]https://character.ai/

Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations), pages 416–425
July 27 - August 1, 2025 ©2025 Association for Computational Linguistics

tween users and virtual agents requires the system to maintain coherent personalization over time (Xi et al., 2025; Shao et al., 2023; Lu et al., 2023). This necessitates continuous adaptation to user preferences and behaviors by synthesizing both long- and short-term histories of user–agent interactions (Zhong et al., 2024; Zhang et al., 2024b). Second, existing approaches often overlook the dynamic evolution of virtual agents (Park et al., 2023; Guo et al., 2024). Agents may produce unsatisfactory responses if they are unable to adapt their knowledge and behaviors through accumulated, diverse experiences (e.g., the rich simulated daily lives of virtual pets), ultimately diminishing user engagement (Vedula et al., 2024a; Zhang et al., 2018; Dunbar et al., 1997; Ceha et al., 2021). Finally, the scalable, real-world deployment of these systems across large user populations remains rare. Moreover, their effectiveness in supporting downstream goals – such as alleviating anxiety or psychological distress –requires further empirical validation (Liu et al., 2021, 2024b; Park et al., 2023).

To address these challenges, we propose **iPET**, an interactive virtual PET companion system designed to enhance emotional dialogues through highly engaging and personalized user–agent interactions. Our iPET system comprises three core components: a *Dialogue* module, a *Memory* module, and a *World Simulation* module.

The Dialogue module serves as an interactive interface that facilitates user engagement with customizable virtual pets. The Memory module enables the pet to dynamically store and synthesize both long- and short-term interaction histories (Zhong et al., 2024). Furthermore, the World Simulation module generates diverse experiences from the pet's everyday life, supporting its progressive evolution based on accumulated memory and user interactions.

The proposed iPET system has been integrated into the real-world application RedNote, which serves a large user base of over 300 million monthly active users[4]. An overview of the user experience is shown in Figure 1. Upon logging in, users can customize their virtual pets according to personal preferences, fostering meaningful emotional interactions and companionship over time, reinforced by ongoing user feedback. A brief live demonstration of iPET is also available on YouTube[5].

[4] https://en.wikipedia.org/wiki/Xiaohongshu
[5] https://youtube.com/shorts/1gtKX5LArAw

In summary, our contributions are as follows:

- We design and develop the iPET system, which comprises three key modules that collectively provide a personalized and evolving emotional dialogue experience.
- We analyze the importance of incorporating character activities into emotional dialogue and propose a three-stage method to enrich pets' lives through progressive, interactive exploration with users.
- Online evaluations highlight iPET's effectiveness, showing a 48.6% increase in user engagement time. The system has been successfully deployed in a real-world production environment for over 200 days, serving millions of users and offering emotional support to individuals experiencing psychological distress.

2 iPET system

As shown in Figure 2, the iPET system comprises three core modules: the interactive dialogue module, the memory module, and the world simulation module. These components work together to enable the system's capabilities as an emotional companion.

2.1 Interactive Dialogue

As the first module, the interactive dialogue system enables users to interact with the pet, facilitating the exchange of information about both the users and their pet.

Specifically, similar to other role-playing dialogue systems (Chen et al., 2024), this module ingests the historical dialogue content H, basic information about the pet P and the user U, as well as an instructional prompt I_R to guide the conversation. This configuration empowers the language model to respond to the user while embodying the persona of the corresponding pet. Unlike traditional systems, the iPET dialogue module introduces two additional inputs: user-related memories M and different levels of the pet's daily life T_1, T_2 and T_3. These inputs enable the system to address user curiosity and provide richer conversational content related to the pet's life, resulting in more engaging interactions. The dialogue process with response R can be formalized as:

$$R = LLM(I_R, H, T_1, T_2, T_3, P, U, M), \quad (1)$$

where T_1, T_2, T_3 will be introduced in Section 2.3.

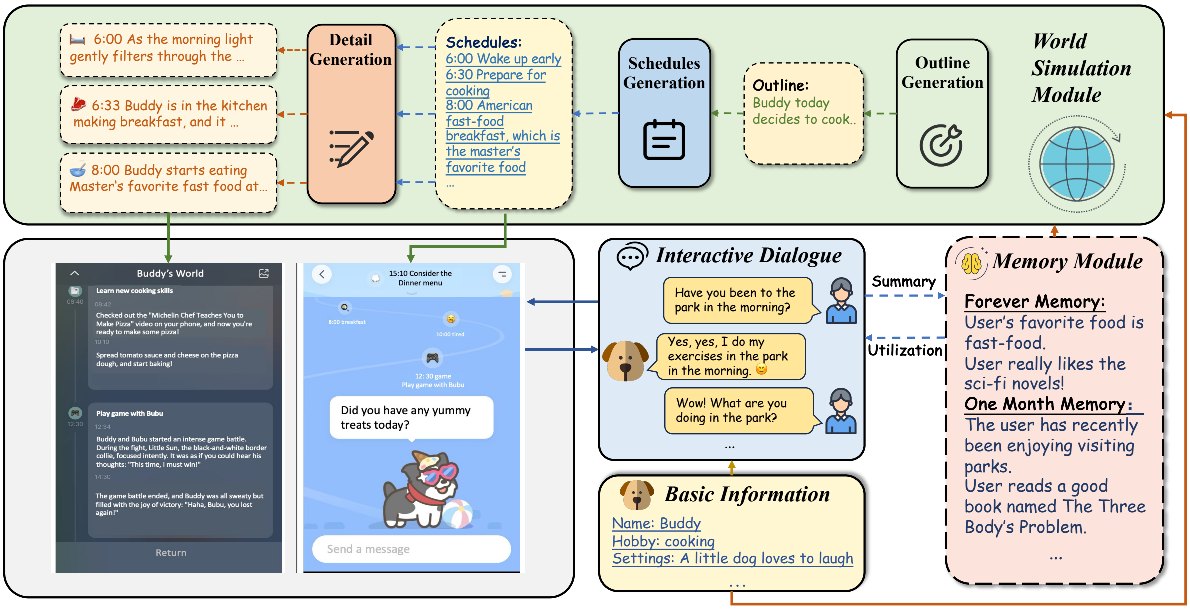

Figure 2: The overall framework of the iPET system.

2.2 Memory Module

To enhance the system's capacity for delivering personalized services, we further introduce the memory module. This module is designed to summarize user-related information through three stages: collection, management, and utilization.

In the collection stage, inspired by previous work (Wang et al., 2023; Zhao et al., 2024), this module uses LLM to summarize key memories M from dialogues. During this process, the system simultaneously categorizes content into three distinct types: permanent memory, long-term memory, and short-term memory. These categories are distinguished by their retention stability: permanent memory preserves enduring user traits and preferences that remain consistent over time; long-term memory contains medium-term plans or intentions such as activity participation or skill acquisition; and short-term memory captures transient details like recent events or immediate tasks. To ensure optimal relevance, permanent memories are retained indefinitely, while long-term and short-term memories are preserved for three months and one month, respectively. Formally, the memory generation process is expressed as:

$$\{M_i, Cat_i\}_{1 \leq i \leq K_s} = LLM(I_s, S, P, U), \quad (2)$$

where I_s represents the summarization instruction, S denotes one of dialogue sessions, K_s denotes the number of memory entries extracted from session S, and Cat_i is the category assigned to each memory entry M_i.

To efficiently organize collected memories, we introduce a management stage. Memory entries are stored in a database with their respective categories and timestamps, while the system periodically removes outdated entries according to predefined temporal policies to maintain relevance. In the utilization stage, iPET extracts relevant memories from the database for integration into core functionalities. Specifically, the system employs cosine similarity-based dense retrieval to enhance user-pet dialogue experiences, while leveraging recent entries to enrich world simulation.

2.3 World Simulation Module

Building on user-related memories and pet's basic information, we design a three-stage world simulation module to create a more engaging virtual life for the pet. This module comprises three sequential stages: (1) outline generation, (2) schedule generation, and (3) details generation. To serve distinct user groups, the system implements two operating modes. For users without interaction history, the normal mode generates pet behaviors based on basic information, establishing consistent baseline behaviors. For returning users, the memory mode generates activities by incorporating accumulated interaction memories, enabling personalized companionship. The complete algorithmic workflow is illustrated in Figure 2.

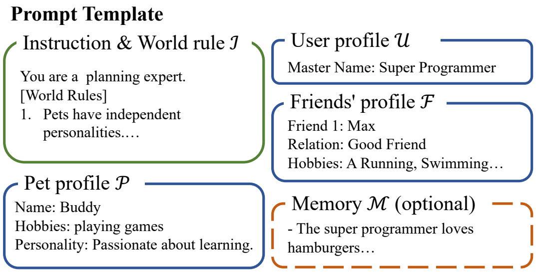

Figure 3: The prompt template of outline generation.

Outline Generation In this stage, we leverage the reading comprehension and reasoning abilities of large language models to generate realistic and character-consistent outlines based on the information available about virtual pet, effectively establishing the framework for their daily lives.

The input template is divided into five parts, as shown in Figure 3. The basic information includes the instruction and world rules I, the pet profile P, and the user profile U, which collectively lay the foundation for constructing the entire virtual world. To enhance the diversity of character behaviors, we also create a set of randomized character profiles F generated by LLMs, which can serve as virtual friends for generation. These profiles are randomly selected to form the characters' social network. For memory mode, we integrate memory M (as outlined in the memory module) to capture a broader range of user-related activities. Formally, the process of generating a well-guided outline is described as follows:

$$T_{1n} = LLM(I_{1n}, P, U, F), \quad (3)$$
$$T_{1m} = LLM(I_{1m}, P, U, F, M), \quad (4)$$

where I_{1n} and T_{1n} are the instruction and outline for the normal scenario, and I_{1m} and T_{1m} are the instruction and outline for the memory mode.

Schedule Generation Building on the directional guidance from the outline generation phase, we further adjust the output by listing events in a timeline with brief descriptions for easy reading. These schedules are brief, with each event described in 2 to 5 words (e.g., "10:00 Walk in park"), enabling users to quickly glimpse the virtual pet's daily activities.

The schedule generation process can be described as follows:

$$\{T_{2ni}\}_{1\le i\le K_n} = LLM(I_{2n}, T_{1n}, P, U, F), \quad (5)$$
$$\{T_{2mj}\}_{1\le j\le K_m} = LLM(I_{2m}, T_{1m}, P, U, F, M), \quad (6)$$

where I_{2n} and I_{2m} are the specific instructions, T_{2ni} and T_{2mj} represent individual schedules for the day, and K_n and K_m denote the total number of schedules for the current outline.

Detail Generation However, brief scheduling clearly fails to meet some users' needs for understanding the true content of schedules and sacrifices a portion of the ability for virtual characters to express themselves to users. Therefore, the detail generation stage further creates detailed descriptions of approximately 50 words for each schedule item, which are essential for attracting users' desire to engage in conversation due to the pet's rich life. This stage effectively utilizes the descriptive capabilities of LLMs, leveraging certain conditions to extend imagination and generate scenes, content, inner monologues, and other descriptions that align with the user's interests and stylistic preferences. Therefore, the computational process can be described as follows:

$$T_{3ni} = LLM(I_{3n}, T_{1n}, T_{2ni}, P, U, F), \quad (7)$$
$$T_{3mj} = LLM(I_{3m}, T_{1m}, T_{2mj}, P, U, F, M) \quad (8)$$

where $i \in \{1, 2, ..., K_n\}$ and $j \in \{1, 2, ..., K_m\}$. Here, I_{3n} and I_{3m} are the specific instructions, T_{3ni} and T_{3mj} denote one of the details of this day.

By decomposing the task of world simulation into multiple levels, our module provides users with a rich and comprehensive virtual pet experience.

2.4 Implementation Details

The overall system implementation is illustrated in Figure 4, where the modules interact as described in the preceding section. Since many system functions depend heavily on LLM services, cost optimization becomes a critical consideration. To address this, the iPET system adopts a T+1 operational strategy, which defers resource-intensive tasks – such as world construction and memory summarization – to off-peak hours.

These processes are executed offline once per day, during periods of minimal user interaction with the dialogue system. This scheduling strategy helps distribute LLM service calls more evenly

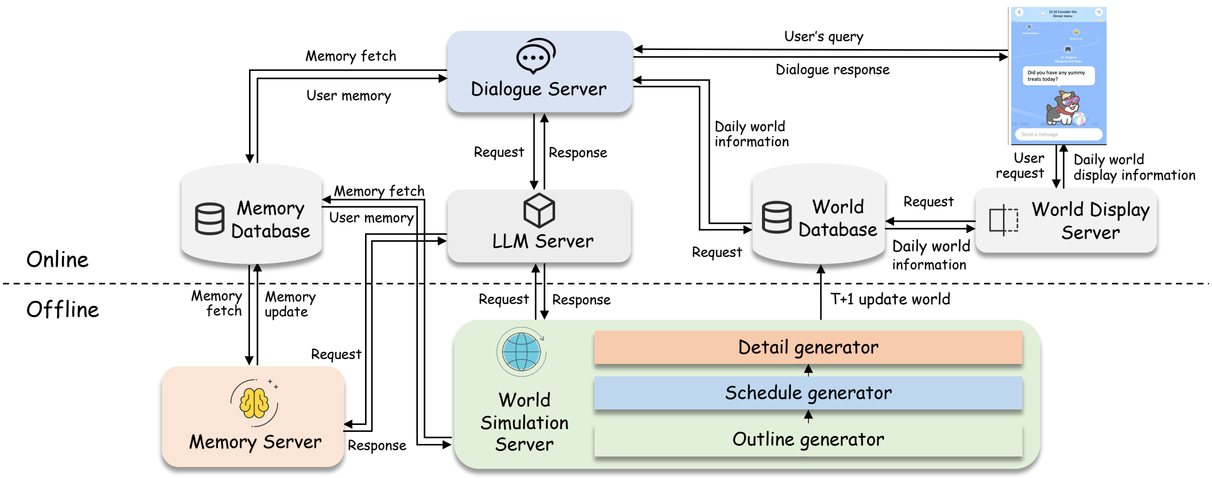

Figure 4: The architecture of our iPET system.

throughout the day, alleviating peak-hour computational loads while preserving overall system efficiency.

3 Experiments

3.1 Experimental Settings

During the offline verification process, we conducted experiments on the Qwen2 family (Yang et al., 2024). In the online experiments, to ensure the safety and integrity of the generated content, we first created a dataset comprising 14,113 entries, which were generated by LLMs and then filtered with expert assistance guided by safety standards. Following data preparation, we conducted Supervised Fine-Tuning (SFT) on our proprietary model, optimizing all parameters. The training process utilized a context length of 2,048 tokens and employed a cosine decay learning rate schedule starting at 5e-6 with a 0.1 warmup ratio. We configured the training with a batch size of 2 and accumulated gradients over 4 steps. The model underwent three complete epochs of training on a cluster of 24 A100 80GB GPUs. During the inference phase, we set the temperature to 0.9 across all model variants to enhance response diversity.

3.2 Offline Evaluation Baselines & Metrics

To further assess the effectiveness of the world simulation method, we implement two baseline approaches for offline evaluation: *Basic Information* and *Direct Generation*. The Basic Information method presents the virtual pet's profile details to users, while the Direct Generation method produces all character behaviors and daily schedules in a single inference, similar to the world generation module. Additionally, we evaluate a variant of the world simulation method that excludes outline generation, instead producing schedules and details directly using the same instructions and settings.

For evaluation, we adopt an LLM-as-a-judge framework, which has shown strong alignment with human judgments (Chiang and Lee, 2023b,a). Inspired by prior work on evaluating role-playing and persona alignment (Chen et al., 2024; Tu et al., 2024), we randomly selected 50 data samples and evaluated them using GPT-4o[6] across four dimensions: *Realism*, *Consistency*, *Richness*, and *Attraction*. The first two metrics assess alignment between the character's persona and their schedule, while the latter two focus on content appeal and user engagement (Vedula et al., 2024b; Xu et al., 2022).

The metrics are defined as follows:

(1) Realism. This metric evaluates the logical coherence and realism of the content (Tu et al., 2024), checking for activity conflicts, schedule feasibility, and real-life plausibility.

(2) Consistency. This metric evaluates consistency between the expressions of activities and the character profile, ensuring that the character's behaviors align with their predefined personality and background (Zhou et al., 2024).

(3) Richness. Based on the analyses of self-disclosure (Sprecher et al., 2013), rich content expression is a key factor in creating attraction. Therefore, this metric assesses the richness of the generated virtual itinerary with realistic daily stamina constraints.

(4) Attraction. This metric evaluates the user's

[6]https://openai.com/index/hello-gpt-4o/

Type	Method	Realism	Consistency	Richness	Attraction
Normal	Basic Information	0.74	3.12	3.19	2.14
	Direct Generation	4.45	4.55	4.55	3.05
	World Simulation	**4.57**	**4.84**	**4.78**	**3.39**
	- w/o outline	4.21	4.80	4.62	3.17
Memory	Basic Information	0.90	3.30	1.85	2.15
	Direct Generation	4.62	4.60	4.15	2.65
	World Simulation	**4.68**	**4.90**	4.58	**2.90**
	- w/o outline	4.50	4.75	**4.60**	2.76

Table 1: Offline experimental results.

Metrics	Previous System	iPET System
Dialogue Engagement Rate	56%	**66% (+17.9%)**
User Usage Time	3.5 min	**5.2 min (+48.6%)**

Table 2: Online A/B test results.

desire for dialogue regarding pet's life content (Xu et al., 2022). It considers factors such as the relevance of dialogue topics and emotional resonance.

To achieve more accurate evaluation results, we adopt the "analyze-rate" approach from a recent study (Chiang and Lee, 2023b). This method requires the LLM to analyze the samples based on these criteria before assigning ratings, which we use to assess all results.

3.3 Offline Experimental Results

The overall results are shown in Table 1. From a comprehensive perspective, the world simulation module demonstrates commendable performance, scoring well across all four metrics in both scenarios. Additionally, it shows a noticeable enhancement in user attraction compared to basic information display and direct generation, indicating iPET's potential to increase user engagement. Moreover, when compared to a variant without the outline generation stage, the module delivers comparable content richness while exhibiting marked improvements in realism, character consistency, and user attraction. This emphasizes the crucial role of the outline stage in the output construction process, as it mitigates uncontrolled content generation and provides a more consistent and realistic user experience (Yang et al., 2023; Xie and Riedl, 2024). Additionally, after incorporating memory content, there is a decrease in richness and attraction. This may be because generating content that is more closely tailored to the user can, to some extent, constrain the inherent diversity and interest of the pet's simulated life (He et al., 2024).

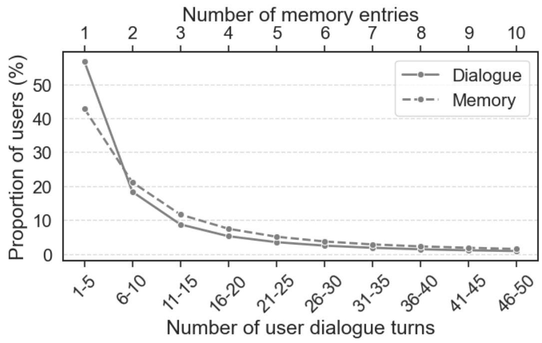

Figure 5: The distribution of user groups based on the number of user dialogue turns and memory entries.

3.4 Online Evaluation

To demonstrate the effectiveness of this method in real-world scenarios, we conducted a seven-day online A/B test (Young, 2014) that compares our iPET system against the previous system, which includes only a dialogue module. In order to better evaluate the effects, we present the two key metrics in our experiment. The first metric is *dialogue engagement rate*, which measures the percentage of users who initiate conversations, while the second is *user usage duration*, which tracks the average time users spend on iPET. As shown in Table 2, introducing a simulated world significantly enhances user interest and engagement with the conversation, while also extending their overall usage duration, thereby demonstrating the effectiveness of iPET.

3.5 User Statistics

We further examine the relationship between user dialogue frequency and memory generation. As shown in Figure 5, both curves exhibit similar trends, suggesting a positive correlation between the number of dialogue rounds and the amount of memory extracted by the iPET system. This indicates that more extensive user interactions facilitate the extraction of relevant memory content,

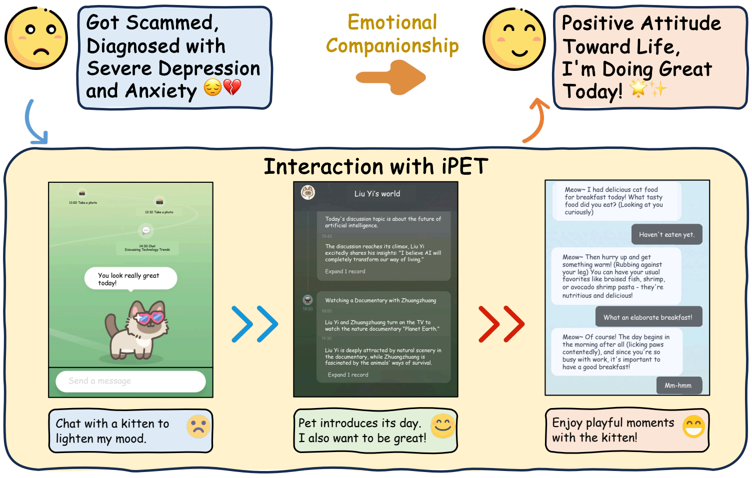

Figure 6: Case study of a real user using our system.

thereby helping to strengthen the user–pet relationship. This distribution is also consistent with behavioral patterns commonly observed in real-world user interactions.

3.6 Case Study

To showcase iPET's emotional companionship capabilities, Figure 6 presents a case study, which is based on notes publicly shared by real users on the RedNote App[7]. In this case, a user suffering from severe depression and anxiety seeks support from the iPET system. The schedule displayed on the main interface naturally guides the user to explore the virtual pet's world, while the pet's schedule page offers rich and engaging details that help foster a more positive attitude toward life. Moreover, the daily interactions and conversations with the pet are designed to be emotionally engaging, enabling the user to form a strong, reciprocal bond. Over time, the user's emotional well-being improves through the pet's consistent companionship. By presenting meaningful life details and enabling emotional interaction, iPET offers effective companionship and supports the development of a more positive outlook.

4 Conclusion

We present iPET, an emotional dialogue system that integrates world simulation to enhance user engagement. Its effectiveness has been validated through both offline experiments and online evaluations with real users. We believe this work highlights the potential of advanced language technologies to deliver meaningful benefits and hope it inspires future research in related directions that promote social good.

Limitations

Although the iPET system offers rich world simulation and emotional companionship, it may not fully meet the complex social needs of human users. Resource and cost constraints have limited our ability to explore more advanced memory utilization by virtual pets, such as complex reasoning based on accumulated experiences. Moreover, maintaining behavioral consistency over extended periods (e.g., a year or more) with substantial interactive engagement remains a significant challenge. These limitations highlight opportunities for future advancements to deepen the emotional connection and social complexity that iPET can provide.

Ethical Considerations

When dialogue systems emulate pet-like characteristics and offer emotional companionship, users may develop emotional dependence on AI pets. It is essential to clearly position AI pets as supplements to, rather than substitutes for, real animal companionship. To mitigate potential psychological risks associated with excessive reliance or inappropriate use, users should be encouraged to engage with these systems in moderation. In addition, we underscore the importance of enforcing strict privacy protection standards in the collection and processing of user data to safeguard personal information and ensure user trust.

[7]While the notes are publicly available, we have chosen not to disclose the URLs or any user-specific details in order to respect user privacy and avoid drawing unnecessary attention to individual users.

Acknowledgements

We would like to thank the anonymous reviewers and meta-reviewer for their constructive feedback on this work. This research/project is partially supported by the Ministry of Education, Singapore, under its Academic Research Fund (AcRF) Tier 2 Programme (Award No.: MOE-T2EP20122-0011).

References

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, et al. 2023. Qwen technical report. *arXiv preprint arXiv:2309.16609*.

Jessy Ceha, Ken Jen Lee, Elizabeth Nilsen, Joslin Goh, and Edith Law. 2021. Can a humorous conversational agent enhance learning experience and outcomes? In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA. Association for Computing Machinery.

Yupeng Chang, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxiang Wang, Yidong Wang, et al. 2024. A survey on evaluation of large language models. *ACM transactions on intelligent systems and technology*, 15(3):1–45.

Jiangjie Chen, Xintao Wang, Rui Xu, Siyu Yuan, Yikai Zhang, Wei Shi, Jian Xie, Shuang Li, Ruihan Yang, Tinghui Zhu, et al. From persona to personalization: A survey on role-playing language agents. *Transactions on Machine Learning Research*.

Nuo Chen, Yan Wang, Yang Deng, and Jia Li. 2024. The oscars of ai theater: A survey on role-playing with language models. *arXiv preprint arXiv:2407.11484*.

Cheng-Han Chiang and Hung-yi Lee. 2023a. Can large language models be an alternative to human evaluations? In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, pages 15607–15631.

Cheng-Han Chiang and Hung-yi Lee. 2023b. A closer look into using large language models for automatic evaluation. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 8928–8942.

Robin IM Dunbar, Anna Marriott, and Neil DC Duncan. 1997. Human conversational behavior. *Human nature*, 8:231–246.

Taicheng Guo, Xiuying Chen, Yaqi Wang, Ruidi Chang, Shichao Pei, Nitesh V Chawla, Olaf Wiest, and Xiangliang Zhang. 2024. Large language model based multi-agents: a survey of progress and challenges. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pages 8048–8057.

Junqing He, Liang Zhu, Rui Wang, Xi Wang, Reza Haffari, and Jiaxing Zhang. 2024. Madialbench: Towards real-world evaluation of memory-augmented dialogue generation. *arXiv preprint arXiv:2409.15240*.

Mina J Kian, Mingyu Zong, Katrin Fischer, Abhyuday Singh, Anna-Maria Velentza, Pau Sang, Shriya Upadhyay, Anika Gupta, Misha A Faruki, Wallace Browning, et al. 2024. Can an llm-powered socially assistive robot effectively and safely deliver cognitive behavioral therapy? a study with university students. *arXiv preprint arXiv:2402.17937*.

Qingyao Li, Lingyue Fu, Weiming Zhang, Xianyu Chen, Jingwei Yu, Wei Xia, Weinan Zhang, Ruiming Tang, and Yong Yu. 2023. Adapting large language models for education: Foundational capabilities, potentials, and challenges. *arXiv preprint arXiv:2401.08664*.

Chenxiao Liu, Zheyong Xie, Sirui Zhao, Jin Zhou, Tong Xu, Minglei Li, and Enhong Chen. 2024a. Speak from heart: an emotion-guided llm-based multimodal method for emotional dialogue generation. In *Proceedings of the 2024 International Conference on Multimedia Retrieval*, pages 533–542.

Chenxiao Liu, Zheyong Xie, Sirui Zhao, Jin Zhou, Tong Xu, Minglei Li, and Enhong Chen. 2024b. Speak from heart: An emotion-guided llm-based multimodal method for emotional dialogue generation. In *Proceedings of the 2024 International Conference on Multimedia Retrieval*, ICMR '24, page 533–542, New York, NY, USA. Association for Computing Machinery.

Siyang Liu, Chujie Zheng, Orianna Demasi, Sahand Sabour, Yu Li, Zhou Yu, Yong Jiang, and Minlie Huang. 2021. Towards emotional support dialog systems. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 3469–3483.

Junru Lu, Siyu An, Mingbao Lin, Gabriele Pergola, Yulan He, Di Yin, Xing Sun, and Yunsheng Wu. 2023. Memochat: Tuning llms to use memos for consistent long-range open-domain conversation. *arXiv preprint arXiv:2308.08239*.

Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology*, pages 1–22.

Murray Shanahan, Kyle McDonell, and Laria Reynolds. 2023. Role play with large language models. *Nature*, 623(7987):493–498.

Yunfan Shao, Linyang Li, Junqi Dai, and Xipeng Qiu. 2023. Character-LLM: A trainable agent for role-playing. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 13153–13187, Singapore. Association for Computational Linguistics.

Hao Shen, Zihan Li, Minqiang Yang, Minghui Ni, Yongfeng Tao, Zhengyang Yu, Weihao Zheng, Chen Xu, and Bin Hu. 2024. Are large language models possible to conduct cognitive behavioral therapy? In *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 3695–3700. IEEE.

Susan Sprecher, Stanislav Treger, and Joshua D Wondra. 2013. Effects of self-disclosure role on liking, closeness, and other impressions in get-acquainted interactions. *Journal of Social and Personal Relationships*, 30(4):497–514.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Yu-Min Tseng, Yu-Chao Huang, Teng-Yun Hsiao, Wei-Lin Chen, Chao-Wei Huang, Yu Meng, and Yun-Nung Chen. 2024. Two tales of persona in LLMs: A survey of role-playing and personalization. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 16612–16631, Miami, Florida, USA. Association for Computational Linguistics.

Quan Tu, Shilong Fan, Zihang Tian, Tianhao Shen, Shuo Shang, Xin Gao, and Rui Yan. 2024. Charactereval: A chinese benchmark for role-playing conversational agent evaluation. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11836–11850.

Nikhita Vedula, Giuseppe Castellucci, Eugene Agichtein, Oleg Rokhlenko, and Shervin Malmasi. 2024a. Leveraging interesting facts to enhance user engagement with conversational interfaces. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 437–446.

Nikhita Vedula, Giuseppe Castellucci, Eugene Agichtein, Oleg Rokhlenko, and Shervin Malmasi. 2024b. Leveraging interesting facts to enhance user engagement with conversational interfaces. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 6: Industry Track)*, pages 437–446.

Bing Wang, Xinnian Liang, Jian Yang, Hui Huang, Shuangzhi Wu, Peihao Wu, Lu Lu, Zejun Ma, and Zhoujun Li. 2023. Enhancing large language model with self-controlled memory framework. *arXiv preprint arXiv:2304.13343*.

Shen Wang, Tianlong Xu, Hang Li, Chaoli Zhang, Joleen Liang, Jiliang Tang, Philip S Yu, and Qingsong Wen. 2024. Large language models for education: A survey and outlook. *arXiv preprint arXiv:2403.18105*.

Zhiheng Xi, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang, Junzhe Wang, Senjie Jin, Enyu Zhou, et al. 2025. The rise and potential of large language model based agents: A survey. *Science China Information Sciences*, 68(2):121101.

Kaige Xie and Mark Riedl. 2024. Creating suspenseful stories: Iterative planning with large language models. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2391–2407.

Xinchao Xu, Zhibin Gou, Wenquan Wu, Zheng-Yu Niu, Hua Wu, Haifeng Wang, and Shihang Wang. 2022. Long time no see! open-domain conversation with long-term persona memory. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 2639–2650.

Xu Yan, Xu Liu, Cuihuan Zhao, and Guo-Qiang Chen. 2023. Applications of synthetic biology in medical and pharmaceutical fields. *Signal transduction and targeted therapy*, 8(1):199.

An Yang, Baosong Yang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Zhou, Chengpeng Li, Chengyuan Li, Dayiheng Liu, Fei Huang, Guanting Dong, Haoran Wei, Huan Lin, Jialong Tang, Jialin Wang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Ma, Jianxin Yang, Jin Xu, Jingren Zhou, Jinze Bai, Jinzheng He, Junyang Lin, Kai Dang, Keming Lu, Keqin Chen, Kexin Yang, Mei Li, Mingfeng Xue, Na Ni, Pei Zhang, Peng Wang, Ru Peng, Rui Men, Ruize Gao, Runji Lin, Shijie Wang, Shuai Bai, Sinan Tan, Tianhang Zhu, Tianhao Li, Tianyu Liu, Wenbin Ge, Xiaodong Deng, Xiaohuan Zhou, Xingzhang Ren, Xinyu Zhang, Xipin Wei, Xuancheng Ren, Xuejing Liu, Yang Fan, Yang Yao, Yichang Zhang, Yu Wan, Yunfei Chu, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, Zhifang Guo, and Zhihao Fan. 2024. Qwen2 technical report. *Preprint*, arXiv:2407.10671.

Kevin Yang, Dan Klein, Nanyun Peng, and Yuandong Tian. 2023. Doc: Improving long story coherence with detailed outline control. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3378–3465.

Scott WH Young. 2014. Improving library user experience with a/b testing: Principles and process. *Weave: Journal of Library User Experience*, 1(1).

Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. 2018. Personalizing dialogue agents: I have a dog, do you have pets too? In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2204–2213, Melbourne, Australia. Association for Computational Linguistics.

Tenggan Zhang, Xinjie Zhang, Jinming Zhao, Li Zhou, and Qin Jin. 2024a. Escot: Towards interpretable emotional support dialogue systems. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 13395–13412.

Zeyu Zhang, Xiaohe Bo, Chen Ma, Rui Li, Xu Chen, Quanyu Dai, Jieming Zhu, Zhenhua Dong, and Ji-Rong Wen. 2024b. A survey on the memory mechanism of large language model based agents. *arXiv preprint arXiv:2404.13501*.

Andrew Zhao, Daniel Huang, Quentin Xu, Matthieu Lin, Yong-Jin Liu, and Gao Huang. 2024. Expel: Llm agents are experiential learners. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 19632–19642.

Wanjun Zhong, Lianghong Guo, Qiqi Gao, He Ye, and Yanlin Wang. 2024. Memorybank: Enhancing large language models with long-term memory. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 19724–19731.

Jinfeng Zhou, Zhuang Chen, Dazhen Wan, Bosi Wen, Yi Song, Jifan Yu, Yongkang Huang, Pei Ke, Guanqun Bi, Libiao Peng, et al. 2024. Characterglm: Customizing social characters with large language models. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track*, pages 1457–1476.