CharacterChat: Learning towards Conversational AI with Personalized Social Support

Anonymous ACL submission

Abstract

The escalating demand for psychological health support highlights the urgent need for scalable and effective interventions. However, traditional Emotional Support Conversation (ESC) frameworks often adopt a generic, "one-size-fits-all" approach, failing to accommodate the diverse personalities of individuals seeking help. To address this limitation, we introduce the **Social Support Conversation (S2Conv)** framework. This novel approach leverages a diverse pool of character agents and an interpersonal matching mechanism to pair users with persona-compatible virtual companions. Specifically, we utilize the Myers-Briggs Type Indicator (MBTI) for persona decomposition, constructing the **MBTI-1024 Bank**—a repository of virtual characters with distinct, granular profiles. To facilitate high-quality interaction, we propose enhanced role-playing prompts integrated with behavior presets and dynamic memory mechanisms, resulting in the creation of the **MBTI-S2Conv** dataset. Building upon these foundations, we present **CharacterChat**, a goal-oriented distributed dialogue system. CharacterChat features a persona- and memory-driven conversational model and a unique interpersonal matching plugin that identifies and dispatches the optimal supporter from the MBTI-1024 Bank based on the user's personality traits. Empirical results demonstrate CharacterChat's superior efficacy in providing personalized social support, validating the substantial advantages of personality-aware interpersonal matching[1].

1 Introduction

The prevalence of mental health disorders constitutes a growing global crisis, exacerbated by the complexities of modern life and socio-economic pressures (McKenzie et al., 2002; Ahmedani, 2011; Fernando and Keating, 2008; Organization et al., 2003). As the gap between the demand for mental healthcare and the availability of professional resources widens, Artificial Intelligence (AI) has emerged as a promising avenue for delivering accessible support and care (Burleson, 2003; Shaw et al., 2004; Skilbeck and Payne, 2003).

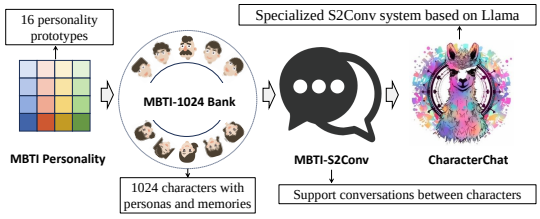

Figure 1: The workflow of the proposed S2Conv system, *CharacterChat*. The system employs personality-driven interpersonal matching to dispatch an optimal supporter from the MBTI-1024 Bank, providing users with personalized emotional support.

Existing AI-driven mental health interventions predominantly rely on the Emotional Support Conversation (ESC) framework. These systems typically mimic a professional consultant, adhering to an exploration-comforting-action protocol (Liu et al., 2021; Sabour et al., 2022; Zheng et al., 2021). Despite their advancements, ESC frameworks face significant hurdles in real-world deployment: (1) **Barriers to Rapport:** Users often perceive these agents as robotic, inducing psychological resistance and defensiveness (Ebert et al., 2002); (2) **Lack of Personalization:** They struggle to adapt to the distinct personality traits of different users. To overcome these challenges, we propose the **S**ocial **S**upport **Conv**ersation (S2Conv) framework. Unlike the consultant-centric ESC, S2Conv focuses on creating human-like AI supporters endowed with specific *personas* and *long-term memories*. This design fosters consistency and reduces user defensiveness. Furthermore, we introduce an *interpersonal matching mechanism* to dispatch the most compatible AI supporter based on the seeker's personality,

[1] Our code and data will be released upon paper acceptance.

ensuring truly personalized support.

The development workflow for the S2Conv system is illustrated in Figure 1. Addressing the scarcity of specialized training data, we leverage the generative capabilities of Large Language Models (LLMs) (Min et al., 2022; Wei et al., 2022; Shen et al., 2023). A key challenge, however, is the tendency of standard LLMs to exhibit a singular, generic "assistant" personality (Edwards et al., 2002). We mitigate this by employing the 16 MBTI personality types (Furnham, 1996) to decompose the generic persona of ChatGPT (OpenAI, 2022). Through a self-instruction process (Wang et al., 2022), we construct the **MBTI-1024 Bank**, comprising 1,024 distinct characters with detailed profiles and backgrounds[2].

Utilizing this bank, we simulate interactions where characters engage in role-play: one acts as a seeker with emotional distress, and the other as a supporter. To prevent context dilution in long conversations and ensure memory retention, we devise a **dynamic memory** mechanism that retrieves context-relevant memory keys to guide response generation. Additionally, to maintain character consistency beyond standard context windows, we introduce **behavior presets**—pre-generated personalized responses serving as in-context demonstrations. These methods culminate in the **MBTI-S2Conv** dataset, containing 4,096 emotional support sessions enriched with participant profiles. We quantify the effectiveness of these interactions using the Beck Depression Inventory (BDI) (Beck et al., 1996), measuring the reduction in the seeker's distress levels.

Finally, we develop *CharacterChat*, an S2Conv system built upon Llama2-7B (Touvron et al., 2023). The model is fine-tuned on MBTI-S2Conv to function as a character-based dialogue agent. Crucially, we implement an interpersonal matching plugin trained on interaction effectiveness scores. This plugin acts as a dispatcher, selecting the optimal supporter from the MBTI-1024 Bank to maximize emotional relief for a specific seeker.

To the best of our knowledge, *CharacterChat* represents the first attempt to integrate interpersonal matching mechanisms into objective-driven conversation systems. Empirical evaluations demonstrate the system's superiority in delivering emotional support and highlight the significant benefits of personality compatibility. Furthermore, our analysis reveals consistent social phenomena—such as personality compatibility patterns and stress tendencies—aligning with observations in human psychology.

The main contributions of this work are summarized as follows:

- We introduce the **Social Support Conversation (S2Conv)** framework, shifting the paradigm from generic counseling to personalized, personality-compatible social support.
- We construct the **MBTI-1024 Bank** and the **MBTI-S2Conv** dataset, providing a large-scale resource of 1,024 virtual characters and their associated emotional support interactions.
- We develop **CharacterChat**, the first S2Conv system featuring a persona-memory-aware conversational model and an interpersonal matching plugin designed to optimize supporter selection.

2 Related Works

2.1 Emotional Support Conversation Frameworks

Research in Emotional Support Conversation (ESC) originated from the exploration-comforting-action framework established by Liu et al. (Liu et al., 2021). Subsequent studies, such as MISC (Tu et al., 2022) and FADO (Peng et al., 2023), have enhanced this paradigm by incorporating commonsense knowledge and feedback mechanisms to optimize strategy selection. However, these systems primarily function as generic problem-solvers. They lack specific character identities, which limits the depth of emotional connection and often results in a functional but impersonal interaction style.

2.2 LLM-based Emotional Intelligence

Large Language Models (LLMs) have significantly advanced ESC through capabilities like Chain-of-Thought (CoT) prompting (Wei et al., 2022), which enables explicit empathetic reasoning. Despite their linguistic fluency, off-the-shelf LLMs often exhibit a "generic responder" issue. Without explicit personalization constraints, models tend to revert to safe, repetitive, and overly polite patterns. This lack of distinct personality often prevents the

[2]While the Big Five traits are dominant in academic psychology, MBTI's widespread popularity facilitates broader practical application and user understanding. See Appendix for details.

formation of genuine companionship required for effective emotional support.

2.3 Persona and Personality in Dialogue

While datasets like Persona-Chat (Zhang et al., 2018) have introduced static profiles to maintain dialogue consistency, current research largely overlooks *interpersonal compatibility*. Existing approaches typically adopt a "one-size-fits-all" mode, focusing on maintaining an agent's persona rather than matching it to the user's needs. This neglects a critical psychological dimension: the compatibility between the help-seeker's personality and the supporter's persona significantly impacts therapeutic outcomes.

To bridge this gap, we propose the S2Conv framework. Shifting focus from *what* to say to *who* says it, we introduce an interpersonal matching mechanism to dispatch compatible virtual supporters, ensuring support is tailored to the user's unique personality.

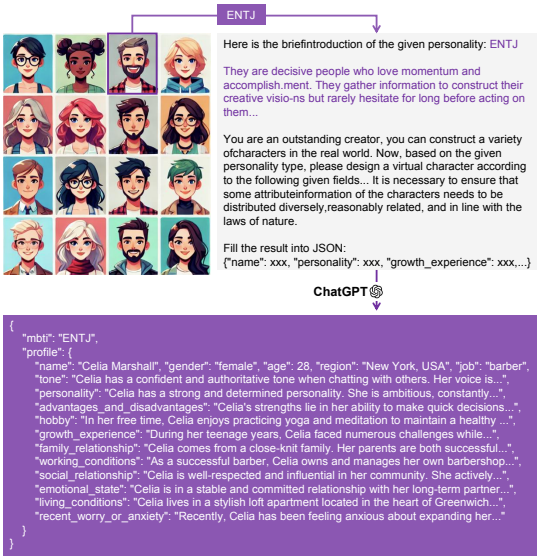

Figure 2: Persona decomposition prompt. Here we present the example to generate the character with the ENFJ personality. See Appendix for more details of the generated characters.

3 Methodology

Our methodology involves four key steps: (1) generating a set of diverse characters with detailed profiles, (2) enabling emotionally-driven conversations through LLM-based role-playing based on these characters, (3) evaluating conversation effectiveness as the character compatibility, and (4) building a character-based LLM and a matching model using synthetic characters and their conversations.

3.1 MBTI-based Persona Decomposition

Current LLMs memorize a range of personalities after pre-training. However, once aligned with human preferences, they tend to adopt a singular, uniform personality—typically that of a helpful and harmless assistant designed to address the user's informational needs. While this consistency is useful, it also limits the LLM's ability to offer personalized emotional support tailored to users with different personalities. To overcome this limitation, our approach seeks to decompose the LLM's behavior into a variety of distinct virtual characters.

To achieve this, we propose a persona decomposition method, leveraging the MBTI theory to guide ChatGPT in generating diverse character profiles. We define a character profile by two core elements: persona (encompassing attributes like name, gender, tone, hobbies, etc.) and memory (covering recent challenges, growth experiences, familial relationships, etc.). The persona shapes the character's behavior, particularly influencing the conversational style when interacting with others. It also drives the interpersonal matching component within the S2Conv framework, facilitating connections between emotional seekers and personality-compatible supporters. Memory serves as a personal knowledge source, ensuring contextual consistency for the character. Prior studies have highlighted the importance of maintaining such consistency in personality-driven dialogues (Song et al., 2020a,b).

In our approach, we separate memory—containing rich factual knowledge relevant to the character—as a dedicated contextual reference to mitigate inconsistencies during conversation. Additionally, drawing on Freud's psychodynamic theory (Freud, 1989), memories frequently act as triggers for human mental health concerns, such as those rooted in growth experiences and family relationships, providing crucial entry points for supporters.

Based on the 16 MBTI personalities, we drive ChatGPT to create characters with different personas and memories, as illustrated in Figure 2. Recognizing the profound impact of personality on various aspects of an individual, our process begins with the manual crafting of descriptions for each of the 16 MBTI personalities (the purple text

in Figure 2), serving as seeds of the prompt. Subsequently, we prompt ChatGPT to assume the role of an outstanding creator, facilitating the formulation of these characters. With 16 distinct prompts as guiding principles, ChatGPT develops intricate characters by comprehensively understanding the detailed personality descriptions and integrating all character profiles into a structured JSON format. Employing this methodology, we curate a collection of 1024 virtual characters for each MBTI personality type, known as the **MBTI-1024 Bank**.

3.2 Enhanced Role-playing Conversational Framework

LLMs like ChatGPT are used to exhibit a singular, stable personality. They tend to act as an assistant, providing helpful and harmless responses (informational support) instead of empathetic responses (emotional support) to users, presenting a poor empathetic ability.

To address this, we start with the development of human-like LLMs, i.e., endowing LLMs with the ground human profiles. However, the plain role-playing methods will gradually lose effectiveness with the conversation proceeding. Consequently, we enhance the role-playing of LLMs by two mechanisms, **behavior preset** and **dynamic memory**. The overview is shown in Figure 3. To enable LLM to generate responses aligning with specific character profiles, we first transform the structured personas of characters into role-playing prompts, making ChatGPT forget the original assistant profile. Then we discover that adding pre-defined dialogues like *"When the other says ..., you should say ..."* helps LLM better simulate the character during the conversation. These findings inspire us to introduce the **behavior preset** method. Specifically, we utilize the persona of the character to drive LLM to generate imagined single-turn dialogues, following the structure like *"When the other says ..., you should say ..."*. Then these generated multiple single-turn dialogues will be added at the end of the prompt for role-playing, maintaining the character's persona under a longer turn.

Equipped with the pre-defined behaviors and given profiles, two LLM agents will take on the roles of seeker and supporter to engage in a conversation. The seeker expresses their troubles, while the supporter aims to gain the seeker's trust and help them overcome their difficulties. However, there is also an issue: if we incorporate all the memories into the role-playing prompt, the abundance of character memories would lead to excessive length. Besides, it suffers from the "lost in the middle" problems (Liu et al., 2023) of LLMs. Hence, we designed the **dynamic memory** mechanism. Previously, in the MBTI-1024 Bank, we structured the characters' memories in JSON format, enabling us to selectively choose a proper memory aspect (key) based on the current context and then locate the corresponding content (value) to constrain the subsequent response generation. As Figure 3 shows, when someone talks about her parents, the dynamic memory will first choose the memory aspect 'family relationship', then locate the specific content. The dynamic memory approach has two benefits: (1) it reduces context consumption, and (2) it enhances memory-response relevance by diminishing the influence of temporarily irrelevant memories.

Utilizing enhanced role-playing framework with the behavior preset and dynamic memory, we randomly select k characters to act as supporters for each character taking on the role of the seeker in the MBTI-1024 Bank ($k = 4$ in this work). This approach aims to facilitate social support conversations. As a result of these efforts, the **MBTI-S2Conv** dataset was developed, comprising a total of 4,096 emotional support conversations with detailed profiles of participants [3].

3.3 BDI-based Evaluation for Support Effectiveness

To investigate the effectiveness of emotional support conversations, we introduce the classical Beck Depression Inventory (BDI) (Beck et al., 1996) for evaluation. BDI is a widely used scale for measuring the severity of depression and was developed by American psychologist Aaron T. Beck. The scale consists of 21 questions, with a total possible score of 63 points. Scores of 14-19 indicate mild depression, 20-28 indicate moderate depression, and scores of 29 or above indicate severe depression. The reliability coefficient of the Beck Depression Inventory ranges from 0.70 to 0.93. Here, we make all characters in the MBTI-1024 Bank answer the questions in BDI before and after they chat with the supporter characters. The difference between them, which could represent the reduction of depression extent, is used to measure the support effectiveness as well as the compatibility of two participants. Based on the BDI, all conversations

[3] See Appendix for concrete generated conversation example.

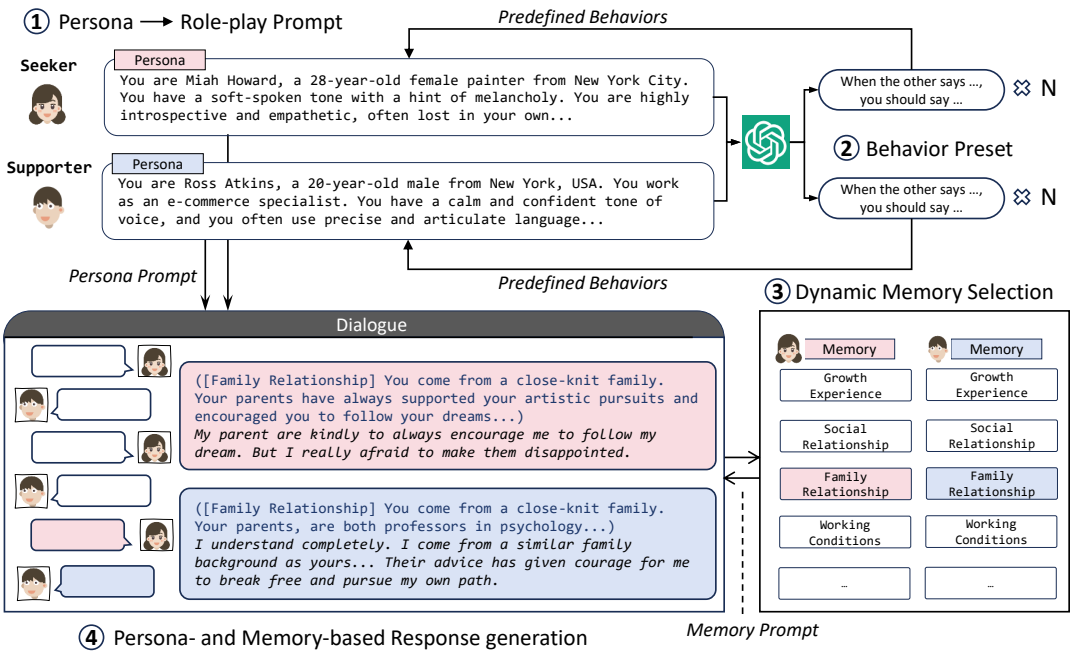

Figure 3: The overview of our enhanced role-playing conversational framework for emotional support based on the *behavior preset* and *dynamic memory*. (1) We randomly sample two characters from the MBTI-1024 Bank to act as the seeker and the supporter respectively. Then we transform their structured personas into role-playing prompts. (2) We introduce a behavior preset method to pre-generate possible single-turn dialogue demonstrations, enabling ChatGPT to maintain the character's state for multi-turn conversations. (3) The context-related memory will be dynamically chosen as the reference for response generation in each turn. (4) The ChatGPT agent will generate responses based on the persona prompt and the dynamic selected memory.

in MBTI-S2Conv have corresponding **support effectiveness scores**.

3.4 Development of CharacterChat

After the data synthesis process for S2conv, we will formulate S2Conv in a mathematical problem and develop the S2conv system, which consists of a character-based LLM and a character-matching model.

Let us consider a supporter with the persona denoted as P and the memory set represented by $M = \{m_1, m_2, \ldots, m_n\}$. The conversation context is defined as $C = [x_1, y_1, x_2, y_2, \ldots, x_k]$, where x_i represents i-th utterance from the seeker, and y_j represents the corresponding utterance from the supporter. The subsequent response and the associated memory are denoted as y_k and m_k, respectively. The primary tasks are (1) selecting the context-relevant memory m_k of the supporter and (2) generating the persona- and memory-based response y_k. To achieve this, we need to maximize the generation probability of y_k given the C, P, M:

$$P(y_k|C, P, M) \approx P(y_k|C, P, m_k) \cdot P(m_k|C, P, M). \quad (1)$$

We leverage Llama2-7B (Touvron et al., 2023) as the response generation and memory selection backbone. These two tasks are simultaneously optimized on the MBTI-S2Conv dataset.

To provide personalized social support, we develop an interpersonal matching model based on BERT (Devlin et al., 2018). Specifically, we utilize the support effectiveness scores based on the BDI reduction as the compatibility between the seeker and supporter. Subsequently, we design the dual encoder to predict the compatibility between two characters based on their persona. During inference, we identify the most compatible supporter in MBTI-1024 Bank for the seeker with the specific persona. Lastly, we propose the CharacterChat, the first S2Conv system including interpersonal matching module and a character-based LLM.

4 Experiments on Emotional Support Ability

We compare our CharacterChat with BlenderBot-ESC (Liu et al., 2021), Vicuna-13B (Chiang et al., 2023) and ChatGPT (OpenAI, 2022) to investigate the emotional support ability of CharacterChat, and the advancement of the interpersonal matching mechanism. Our evaluation includes three parts:

(1) we evaluate various models on the traditional dialogue metrics including BLEU (Papineni et al., 2002) and ROUGE (Lin, 2004), with the separate test set of MBTI-S2Conv, (2) the simulation method is used to evaluate the conversation-level effectiveness based on BDI index reduction of a seeker, (3) we evaluate the effectiveness of interpersonal matching mechanism by human interaction.

Table 1: Automatic evaluation results. The **bold** text represents the best performance.

Models	BLEU-1	BLEU-2	ROUGE-L
BlenderBot-ESC	39.15	21.86	11.86
Vicuna-13B	50.38	30.38	13.22
CharacterChat	**68.74**	**41.81**	**19.12**

Automatic Evaluation We have developed the CharacterChat based on the MBTI-1024 Bank and MBTI-S2Conv dataset. To prevent over-fitting, we utilize a similar manner to construct a test set for automatic evaluation, which contains 256 new characters and 256 social support conversations. Otherwise, ChatGPT is not considered as the baseline in automatic evaluation since the dataset is distilled from it, which will result in a data leak if we evaluate it.

As depicted in Table 1, the following conclusions can be drawn from the data analysis: (1) A comparison between BlenderBot-ESC and CharacterChat underscores the unsuitability of emotional support conversation methods for social support interactions. (2) The competence of LLM assistants in handling seekers with diverse personas is notably challenged in the absence of the adapted supporter profile. This inadequacy is apparent in the performance of Vicuna-13B. (3) Within the domain of social support conversations, the specialized CharacterChat exhibits superior performance compared to the more all baselines. The above findings highlight the clear effectiveness of CharacterChat in providing emotional support.

Simulated Evaluation Further, we evaluate the emotional support capability of CharacterChat by simulated conversation. Following the construction process of MBTI-S2Conv, we utilize ChatGPT as the emotional seeker in MBTI-1024 Bank, while the different models act as the supporters. By conversational interaction with different models, we observe the BDI index reduction of the simulated seekers.

Experimental results are shown in Table 2, where our CharacterChat achieves the best performance in the compared method except ChatGPT, as well as the near performance of ChatGPT. In comparison, even if BlenderBot-ESC is developed to provide emotional support, it fails to adapt to the different personalities of seekers, resulting in the worst performance. Vicuna-13B as a general LLM, demonstrates a better performance than BlenderBot-ESC, indicating the powerful ability of LLMs. However, it is also hard to give empathetic and personalized support to various seekers in MBTI-1024 Bank. Without the interpersonal matching, the advancedness of CharacterChat is not obvious. Once the optimal supporter is chosen, the compatible personality will enhance the emotional support effect.

These findings are consistently observed across different personality groups, suggesting the stable performance of CharacterChat to support seeks with various personalities.

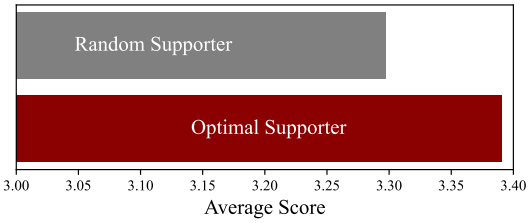

Figure 4: Human evaluation.

Human Evaluation For a deeper understanding of the effectiveness of interpersonal matching in emotional support conversations, we invite 64 volunteers (covering 16 MBTI personalities) to act as seekers and engage in conversations with an optimal AI supporter. They are told to describe their persona and CharacterChat will match the optimal supporters from the MBTI-1024 Bank for them. Otherwise, they also chat with another random AI supporter as a comparison. After the conversation, they should assess these two supports about their performance in depressing the emotional stress by the 5-level scale. The experimental process will take one hour and we pay 10$ as compensation for every volunteer. Evaluation results are shown in Figure 4. When optimal supporters are dispatched for human seekers, the scores of emotional reduction are higher than the random supporter, indicating the effectiveness of interpersonal matching.

Role-playing Framework Evaluation We improve the role-playing conversation by the behavior

Table 2: Simulation evaluation results. We compare the performance under different personality groups, such as I-type, E-type, etc. The lower scores after the conversation indicate better support ability. The best performance among methods except ChatGPT is in **bold** text. Subscript o represents that the optimal supporter, which is chosen by interpersonal matching, is equipped, while r represents the random supporter.

	E	I	N	S	T	F	J	P
Initial	0.187	0.206	0.204	0.189	0.194	0.200	0.194	0.200
BlenderBot-ESC	0.126	0.153	0.147	0.131	0.139	0.140	0.136	0.142
Vicuna-13B	0.111	0.136	0.130	0.118	0.124	0.123	0.125	0.123
ChatGPT$_r$	0.109	0.137	0.129	0.117	0.124	0.122	0.120	0.126
ChatGPT$_o$	0.082	0.105	0.099	0.088	0.093	0.094	0.090	0.097
CharacterChat$_r$	0.114	0.140	0.134	0.120	0.129	0.125	0.124	0.130
CharacterChat$_o$	**0.086**	**0.107**	**0.104**	**0.089**	**0.097**	**0.096**	**0.095**	**0.098**

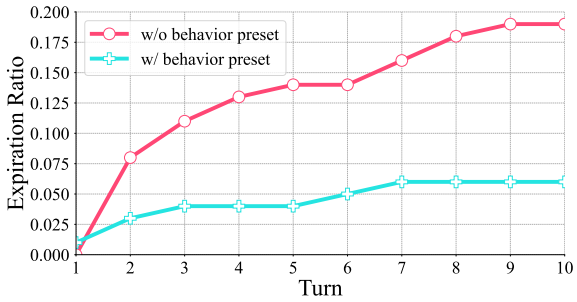

Figure 5: Expiration Ratio of Role-Playing Prompts with Increasing Conversation Turns.

preset and dynamic memory. We evaluate the role-playing effectiveness here by self-belief validation.

We investigate the duration of role-playing by asking the agent for their name at different turns. The role-playing will be considered expired once the response contains 'AI', 'ChatGPT', or 'assistant', instead of their own profile. We make ChatGPT role-play based on randomly selected 100 characters from the MBTI-1024 Bank. As the number of conversation turns increases, the expiration ratio of the role-playing is shown in Figure 5. The common role-playing method is hard to follow the given character profiles with the dialogue proceeding, indicated by the increasing expiration ratio and 19% at the 10th turn. However, with the help of the predefined behaviors, the expiration ratio of the role-playing after 10 turns only increases to 6% and keeps steady. The stable role-playing ability validates the effectiveness of our proposed enhanced role-playing conversational framework and ensures the character-consistency as well as the data quality of MBTI-S2Conv.

Personality Back-Test of Characters To evaluate the personality alignment of characters within the MBTI-1024 Bank, we replace the system prompt of ChatGPT with the character's persona, completing the MBTI questionnaire from the official assessment platform like human.[4] The summarized results are presented in Table 3. Notably, 38% of characters exhibit complete alignment across all four dimensions with their designated MBTI personality, and nearly 80% of characters demonstrate alignment across at least three dimensions. Taking into account the complexity of personality, which can be shaped by a range of factors resulting in inherent fluctuation, the assessment could demonstrate that the virtual characters within MBTI-1024 Bank are capable of displaying behaviors consistent with their designated personalities.

Furthermore, we assessed the accuracy of MBTI sub-dimension matching, as detailed in Table 4. Remarkably, the E/I dimension, which represents how individuals gain energy, achieves the highest accuracy. 'E' means external stimuli or interactions (extraversion) while 'I' means being more reserved and reflective (introversion). It is intuitively aligned with the actual recognition since it is easy to discriminate the E/I dimension of a person based on whether he/she is outgoing. Conversely, the N/S dimension, which characterizes how individuals assimilate information, exhibited the lowest accuracy. The former 'S' means concrete specifics and practicality (sensing) while the latter 'N' patterns and possibilities (intuition). We speculate that the discrepancy likely is aroused by the inherent challenge of incorporating intuition into LLMs, making precise recognition in this dimension more intricate.

In summary, the virtual characters of MBTI-1024 Bank showcase a significant alignment with their designated MBTI personalities.

[4] https://www.16personalities.com/

Table 3: MBTI Assessment Results for MBTI-1024 Bank. The hit@k metric measures the number of matching dimensions between the designated MBTI and the assessed MBTI within the four dimensions.

	hit@0	hit@1	hit@2	hit@3	hit@4
Count	3	25	173	433	390

Table 4: MBTI Matching Accuracy for Each Sub-dimension within the MBTI-1024 Bank.

	E/I	N/S	T/F	J/P
Accuracy (%)	90.82	71.09	85.84	73.24

5 Compatibility between Different Personalities

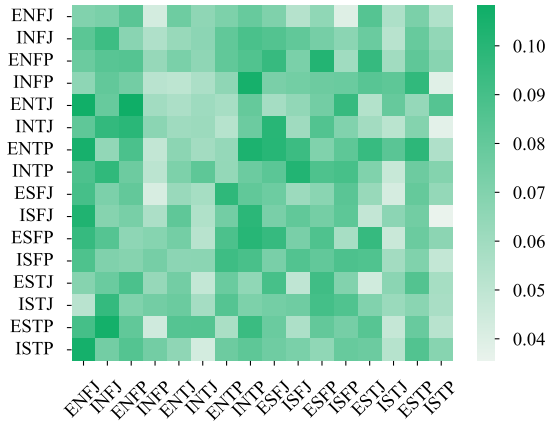

Figure 6: BDI index reduction of MBTI-S2Conv across 16 MBTI personalities, The X-axis and Y-axis are the MBTI of the supporter and seeker respectively.

Subsequently, we investigate how personality affects support effectiveness. Before and after the conversation in MBTI-S2Conv, we have made BDI assessment for each seeker, the BDI index reduction by the conversation is deemed as the support effectiveness.

Experimental results are presented in the Figure 6. It's clear that personality remarkably influences the effectiveness of the social support conversation. There are also discoveries that match the conclusions drawn from a study of society (O'Neil and Petty, 2019). For example, the optimal personality for the characters with the INFP personality is INTP by observing the dark parts in Figure 6, which means the higher BDI index reduction after the conversation. In other words, the INTPs serve as the most compatible personality for supporting INFPs. Their compatibility is influenced by several factors: Firstly, their shared intuition means they both enjoy discussing possibilities and abstract concepts, which helps them form deep connections. Secondly, their complementary thinking and feeling styles—INFPs focusing on values and emotions, and INTPs on logic and objectivity—bring fresh perspectives to their conversations. Additionally, as introverts, they both value alone time, reducing potential friction over personal space. Lastly, their curiosity and open-mindedness keep the relationship dynamic and growth-oriented.

Besides, we also observe that INFPs might not be eligible supporters for helping reduce depression, as indicated by the lighter bands in Figure 6. This is caused by the inherent introversion characteristic of INFPs and their pessimistic attitude will potentially have negative effects on the others' emotion.

These observations underscore the importance of compatible personality for emotional support, inspiring us to propose CharacterChat by introducing an interpersonal matching mechanism for providing effective personalized emotional support.

6 Conclusion

In this work, we have introduced the **S**ocial **S**upport **Conv**ersation (S2Conv) framework as a novel solution to the challenges faced by traditional emotional support methods. Our central contribution involves the creation of the MBTI-1024 Bank, housing a diverse array of virtual characters with distinct profiles, and MBTI-S2Conv, the social support conversations between the characters in the MBTI-1024 Bank. Both of them facilitate the development of CharacterChat, the first S2Conv system that encompasses a character-based conversational model, and an interpersonal matching plugin model to dispatch the compatible supporter in MBTI-1024 Bank for the seeker with a specific persona. Our work not only highlights the remarkable capabilities of CharacterChat for personalized social support but also emphasizes the pivotal role of interpersonal matching in enhancing the effect of support. We believe that this study could raise more attention to interpersonal matching mechanisms for goal-oriented human-machine dialogue in the future.

7 Limitations.

First, our reliance on the **Myers-Briggs Type Indicator (MBTI)** may oversimplify the complexity and fluidity of real-world human personalities. Second, the dataset constructed via **role-playing agents** may engender a "simulation-to-reality" gap, potentially failing to capture the subtle and erratic emotional nuances of human users. Finally, the definition of optimality in our **interpersonal matching** is based on training objectives, which may not always align with a user's momentary psychological needs.

8 Ethical Considerations

Crucially, CharacterChat is designed for **social support**, not as a substitute for clinical therapy. Rigorous safety mechanisms are implemented to detect crisis signals (e.g., self-harm) and redirect users to professionals. We also acknowledge the risk of **emotional dependence** due to high personalization and mitigate this by ensuring transparency about the AI identity. Furthermore, strict **data privacy** protocols are enforced to protect the sensitive personal narratives processed by the memory module.1

References

Brian K Ahmedani. 2011. Mental health stigma: society, individuals, and the profession. *Journal of social work values and ethics*, 8(2):4–1.

Aaron T Beck, Robert A Steer, Gregory K Brown, and 1 others. 1996. Beck depression inventory.

Brant R Burleson. 2003. Emotional support skills. In *Handbook of communication and social interaction skills*, pages 569–612. Routledge.

Wei-Lin Chiang, Zhuohan Li, Zi Lin, Ying Sheng, Zhanghao Wu, Hao Zhang, Lianmin Zheng, Siyuan Zhuang, Yonghao Zhuang, Joseph E. Gonzalez, Ion Stoica, and Eric P. Xing. 2023. Vicuna: An open-source chatbot impressing gpt-4 with 90%* chatgpt quality.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Sheryl A Ebert, DC Tucker, and DL Roth. 2002. Psychological resistance factors as predictors of general health status and physical symptom reporting. *Psychology, health & medicine*, 7(3):363–375.

John A Edwards, Kevin Lanning, and Karen Hooker. 2002. The mbti and social information processing: An incremental validity study. *Journal of Personality Assessment*, 78(3):432–450.

Suman Fernando and Frank Keating. 2008. *Mental health in a multi-ethnic society: A multidisciplinary handbook*. Routledge.

Sigmund Freud. 1989. The ego and the id (1923). *Tacd Journal*, 17(1):5–22.

Adrian Furnham. 1996. The big five versus the big four: the relationship between the myers-briggs type indicator (mbti) and neo-pi five factor model of personality. *Personality and individual differences*, 21(2):303–307.

Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*, pages 74–81.

Nelson F Liu, Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Fabio Petroni, and Percy Liang. 2023. Lost in the middle: How language models use long contexts. *arXiv preprint arXiv:2307.03172*.

Siyang Liu, Chujie Zheng, Orianna Demasi, Sahand Sabour, Yu Li, Zhou Yu, Yong Jiang, and Minlie Huang. 2021. Towards emotional support dialog systems. *arXiv preprint arXiv:2106.01144*.

Kwame McKenzie, Rob Whitley, and Scott Weich. 2002. Social capital and mental health. *The British Journal of Psychiatry*, 181(4):280–283.

Sewon Min, Xinxi Lyu, Ari Holtzman, Mikel Artetxe, Mike Lewis, Hannaneh Hajishirzi, and Luke Zettlemoyer. 2022. Rethinking the role of demonstrations: What makes in-context learning work? *arXiv preprint arXiv:2202.12837*.

OpenAI. 2022. Introducing chatgpt. https://openai.com/blog/chatgpt.

World Health Organization and 1 others. 2003. Investing in mental health.

Daniel A O'Neil and Mikel D Petty. 2019. Heuristic methods for synthesizing realistic social networks based on personality compatibility. *Applied Network Science*, 4(1):1–49.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.

Wei Peng, Ziyuan Qin, Yue Hu, Yuqiang Xie, and Yunpeng Li. 2023. Fado: Feedback-aware double controlling network for emotional support conversation. *Knowledge-Based Systems*, 264:110340.

Sahand Sabour, Chujie Zheng, and Minlie Huang. 2022. Cem: Commonsense-aware empathetic response generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 11229–11237.

Jon C Schommer, Paul D Tieger, Anthony W Olson, and Lawrence M Brown. 2017. Associations between modifiable health-risk behaviors and personality types. *INNOVATIONS in Pharmacy*, 8(2).

Benjamin A Shaw, Neal Krause, Linda M Chatters, Cathleen M Connell, and Berit Ingersoll-Dayton. 2004. Emotional support from parents early in life, aging, and health. *Psychology and aging*, 19(1):4.

Yongliang Shen, Kaitao Song, Xu Tan, Dongsheng Li, Weiming Lu, and Yueting Zhuang. 2023. Hugginggpt: Solving ai tasks with chatgpt and its friends in huggingface. *arXiv preprint arXiv:2303.17580*.

Julie Skilbeck and Sheila Payne. 2003. Emotional support and the role of clinical nurse specialists in palliative care. *Journal of advanced nursing*, 43(5):521–530.

Haoyu Song, Yan Wang, Wei-Nan Zhang, Xiaojiang Liu, and Ting Liu. 2020a. Generate, delete and rewrite: A three-stage framework for improving persona consistency of dialogue generation. *arXiv preprint arXiv:2004.07672*.

Haoyu Song, Yan Wang, Wei-Nan Zhang, Zhengyu Zhao, Ting Liu, and Xiaojiang Liu. 2020b. Profile consistency identification for open-domain dialogue agents. *arXiv preprint arXiv:2009.09680*.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, and 1 others. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.

Quan Tu, Yanran Li, Jianwei Cui, Bin Wang, Ji-Rong Wen, and Rui Yan. 2022. Misc: a mixed strategy-aware model integrating comet for emotional support conversation. *arXiv preprint arXiv:2203.13560*.

Yizhong Wang, Yeganeh Kordi, Swaroop Mishra, Alisa Liu, Noah A Smith, Daniel Khashabi, and Hannaneh Hajishirzi. 2022. Self-instruct: Aligning language model with self generated instructions. *arXiv preprint arXiv:2212.10560*.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, and 1 others. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35:24824–24837.

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? *arXiv preprint arXiv:1801.07243*.

Chujie Zheng, Yong Liu, Wei Chen, Yongcai Leng, and Minlie Huang. 2021. Comae: A multi-factor hierarchical framework for empathetic response generation. *arXiv preprint arXiv:2105.08316*.

A Psychological Tendency of Different Personalities

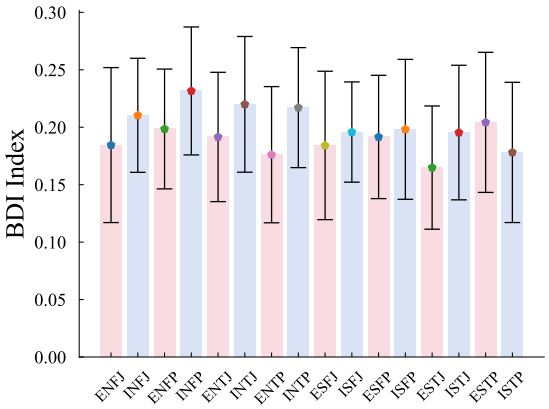

Figure 7: Psychological assessment results of MBTI-1024 Bank. We use two kinds of colors to distinguish extraversion and introversion personalities. The midpoint of each bar represents the average BDI index of each MBTI type in the measurement, while the upper and lower ends of the vertical line represent the upper and lower bounds of the BDI index.

Based on the aforementioned BDI, we take psychological assessments for virtual characters in MBTI-1024 Bank. Specifically, each character will complete a BDI assessment and we compute the normalized BDI index to display (the actual BDI score / total BDI score). Accordingly, the BDI index of 0.22-0.30 indicates mild depression, 0.30-0.44 indicates moderate depression, and above 0.44 indicates severe depression.

The BDI indexes of 16 personalities are shown in Figure 7. Luckily, the psychological condition of characters in MBTI-1024 Bank is overall favorable. However, the character with mild depression still exists in all personality groups. Specifically, the characters with INFP personality, whose average BDI index reaches the mild depression level, display the worst psychological condition, indicating the importance of providing emotional support for them. Our findings are also in alignment with the practical experiment done in (Schommer et al., 2017), which shows INFPs present the highest health risk in anxiety and depression. This not only enhances the credit of our study by simulation but also provides a new scope (LLM-based simulation) to investigate psychology or society science.

Besides, the I-type (Introversion) people consistently have a higher BDI index than E-type (Extraversion) ones in general except ISTP. It is easy to understand and also the commonsense that the I-type person is relatively resistant to social interaction and accustoms to hide the troubles in the deep heart. The exception is that the average BDI index of ISTP is higher than ESTP, which is evidenced in (Schommer et al., 2017) once again. This is caused by the perfectionism of these ESTPs, in which the conflict between the high pursuit and current ability introduces their depression.

These findings indicate the necessary of providing emotional support to people in trouble, especially for the introversive group.