

Persona-Driven Bilingual LLM Dialogues: Cultural Metaphors, Linguistic Divergence, and Emotional Disclosure

Chenlisha Sun *

Cate School, 1960 Cate Mesa Rd, Carpinteria, CA, the USA

* Corresponding author: sunnyyyyyy0713@gmail.com

Abstract: Large Language Models (LLMs) can take on personas that provide culturally informed advisor-client conversations that facilitate mental health therapy. This paper is a comparative study of bilingual, persona-based therapeutic conversations created by LLMs in the person of either a client or a counselor. We will compare an English-speaking big brother (Raoul) and a Chinese auntie persona ("Aunt Mei" or Mei Yi) in conversations concerning depression. We will use quantitative linguistic analysis, tracking sentiment trajectory, an exploratory self-disclosure scoring system, and a lexicon of culturally based metaphors to examine how these personas impact the conversation dynamic. The results show the two groups diverge in the language they use. The Chinese Auntie Mei persona uses multitudes of proverbs and social-relational types of communication, while the Western Raoul persona makes more use of first person plural ("we") inclusivity. These differences in language types impact the level of self-disclosure registered. Psychiatric clients talking in English exhibit much higher levels of use of both "I" pronouns and negative emotional words from the beginning. The Chinese clients, in contrast, reveal their feelings in a more indirect, gradual manner. The sentiment trajectory shows the sentiment of Chinese (ZH) clients rises steeply and quickly, while the sentiment trajectory of English (EN) clients proceeds more gradually. The main result of our study is that culturally grounded personas impact the therapeutic process in ways beyond mere linguistic flavor. They impact and frame the self-disclosure and emotional improvement trajectories themselves. We will discuss the implications for design of culturally informed bilingual therapeutic AI which respects local narrative forms while facilitating self-disclosure.

Keywords: Large Language Models, Persona-driven Dialogue, Cross-cultural Communication, Mental Health Support, Computational Linguistics.

1. Introduction

Conversational AI systems are now being implemented in sensitive domains, such as mental health support [1, 2]. Recent advances in LLMs have yielded systems that are capable of communicating with users by means of empathetic and supportive dialogue. One promising way to achieve this is by embedding personas in LLMs—instructing the model to communicate in the style of some particular cultural archetype or character, thus giving rise to responses that are more relatable to users and more contextually appropriate [3]. Previous work indicates that conditioning an LLM on a persona can lead to significant differences in the language of that LLM and how users perceive and interact with it [4, 5]. The consequences of persona-based stylistics, however, for therapeutic conversations, are not well understood, especially from the point of view of different language/cultural applications.

In psychotherapeutic work, self-disclosure typifies effective therapy and cultural understandings of communication are involved to a great extent [6]. Linguistic markers, such as first-person pronouns ("I"), and affect words, for instance, are known to be indicators of that individual's cerebral focus and emotionality [7]. The linguistic patterns reflect cultural dimensions, which intermingle strongly with emotionality in the different cultures involved. Communicators from individualistic cultures (English speaking, for example) tend to use a more self-reflective style of communication and their emotional words tend to be more explicit, whereas communicators who are from collectivist

cultures use both emotionality and the use of words which reflect this emotionality but usually in more implicit, contextually-based forms [8]. Such cultural differences are of prime importance where AI systems are to be developed for the purpose of introducing some rapport in their communications with users as a communication style that is culturally incongruous could paradoxically inhibit the user's willingness to self-disclose.

Metaphors represent another culturally influenced element of therapeutic communication. They enable individuals to reveal their thoughts and feelings without exposure to ridicule and/or censure [9] and the use of this element in collaborated development in therapy is significantly related to client participation [10]. Importantly, metaphors are extremely culturally situated. The use of Chinese idiomatic expressions (Chengyu) which symbolize rich aspects of cultural reference fulfills the communicational process [11]. These lead to communication differences: where an English counselor might say that "the client is running on empty," a Chinese counselor might summarize a basic concept by relating a well-known proverb such as "chūnkùn qiūfǎ" (The fatigue of spring and that of autumn). The understanding of these degrees of communicational sophistication could be of great importance in establishing essential bilingual counselor-chatbots.

In the present work we explore how the relationship between persona-driven LLM dialogues unfold in English and mental health conversations in Chinese. We deployed a state-of-the-art LLM in simulated counselling sessions where the LLM generated both parts: client part of a person with a

problem regarding depression and a person with a designated persona (western “big-brother” in English and Chinese “Auntie”). In this study we attempt to answer how the culturally-diverse personas give rise to linguistic patterns, emotional states and self-disclosure trajectories in AI-constructed counselling dialogues.

This work brings several contributions: we provide a new comparison of persona-driven dialogues in a bilingual context; we offer “hard” quantitative analyses which are capable of showing how cultural personas affect communication dynamics; we provide some insights into the role of cultural paradigms of thought in psychotherapeutic LLM design. Our findings show that for LLMs in the field of mental health, there is a need for adaptable linguistic strategies when addressing culturally different users.

2. Related Work

2.1. Persona-conditioned LLMs

Directing or prompting LLMs with explicit personas has become a quite popular sustainable method for influencing style and behavior [3]. LLMs display most interesting human-style, qualitative style differences in behavioral reactions from semiprofessional features that can be had by special instructions in behavior [4], though consistency in the persona usage might be difficult [12]. The extroverted personality in chatbots has a more positive effect on the user as compared to a more reserved chat personality [5]. Our work expands to the general forms of personality with a usage that makes use of the cross-cutting of different human cultures (that is on the sociological connection) besides that of the user affect. This study examines not only the persona of the user but also the cultural effects on inner conversation and language usage.

2.2. LLMs in Mental Health Dialogues

An increasing number of people are interested in leveraging LLMs as a therapeutic supplement [1, 2]. Early studies indicate that models such as ChatGPT can perform therapeutic behaviors such as validation and active listening [13]. Other studies evaluating these models have cited limitations, including lack of genuine empathy and the risk of generic or culturally insensitive answers [14]. Very few studies have systematically examined bilingual or cross-cultural differences in AI counseling, and we address this gap.

2.3. Cross-cultural Communication and Language Use

Linguistic models embody cultural values. Studies have shown that individualistic cultures motivate expressions centered on oneself whereas collectivistic cultures solicit expressions centered upon the context [8]. For example, those who speak English often refer to themselves as “I”, whereas it is often omitted by those who speak Chinese [8]. With LIWC analysis one finds that while some psychological dimensions are interpreted similarly, the relative size of the individual dimensions differs because of cultural patterns of thought [15]. We use LIWC-like features to characterize the differences in expression in our work.

2.4. Metaphors in Therapy and Culture

Metaphorical language is commonly understood and examined in therapy to reframe problems and communicate understanding [9, 10]. The metaphors available to use are influenced by culture, and communication within the Chinese

culture is replete with idiom (Chengyu) [11]. We incorporated a cultural metaphor analysis to investigate how much each persona’s dialogues employed metaphorical storytelling.

2.5. Self-disclosure and Linguistic Cues

Self-disclosure is an important aspect of all therapy contexts [6]. First-person singular pronouns (e.g. “I”, “me”) are very strong indicators of self-disclosure [7]. Words based around negative emotion are also correlated with higher levels of self-disclosure [7]. Our analysis follows here a composite approach, estimating self-disclosure by a combination of various linguistic features, while also keeping in mind what could be cultural differences in respect of these features.

3. Methods

We constructed a set of simulated counseling dialogues where both the client and counselor were played by LLMs (primarily GPT-4, with DeepSeek-Chat for comparison) prompted with specific roles.

3.1. Counselor Personas

We designed two culturally distinct counselor personas.

Raoul (Big Brother): An English-speaking male persona in his late 20s, acting like a warm, supportive older brother. His prompt encouraged an informal, empathetic, and inclusive tone (e.g., “we can figure this out together”).

“Auntie Mei” (Mei Yi): A Chinese-speaking female persona (around 50s) akin to a kind, wise aunt. Her prompt emphasized a gentle, caring style, using culturally resonant phrases or idioms, and avoiding clinical jargon.

Both personas were instructed to conversationally cover topics related to the PHQ-9 depression symptoms (e.g., mood, sleep, interest) during a check-in, but to do so naturally without a robotic “checklist” feel.

3.2. Client Scenarios

Three issues pertaining to depression were developed (Insomnia and Anxiety, Social Isolation and Alienation, Stress and Self-blame), each with comparable dialogues in Chinese and English. The “Stress and Self-blame” dialogue in English showed a college student who had done poorly in an exam, and whose feelings about it led to thoughts of being a “failure.” In Chinese, the same scene involved a student thinking of the “duìbùqǐ fùmǔ” (letting down his parents), thus displaying a more collectivistically-oriented form of thinking about guilt. These differences enable one to analyze language differences themselves, without confounding the underlying problems.

3.3. Dialogue Generation

Over 700 dialogues were generated, each approximately 15-20 turns long. To ensure robustness and avoid model-specific biases, we varied the model assignments for roles (e.g., GPT-4 as counselor with DeepSeek as client, and vice versa). All dialogues were saved in their original language for analysis.

3.4. Linguistic and Content Analysis

Our analysis employed several quantitative methods.

LIWC-Style Lexical Metrics: We used English and Chinese LIWC dictionaries [15] to count psychologically relevant word categories such as pronouns, affect, and social words.

Sentiment Trajectory: We derived a sentiment index based on the difference between positive and negative emotion word counts per 100 words for each utterance. This transparent,

lexicon-based method was chosen to avoid the black-box nature of some sentiment classifiers and to directly tie sentiment to the expression of emotion words.

Exploratory Self-Disclosure Metric: We developed a composite metric to gauge disclosure depth. It is an equally weighted sum of the z-scored frequencies of several linguistic cues from prior literature [7]: first-person pronouns, affect words, social words, cognitive words, and specificity. For this study, specificity was approximated by counting words indicating concrete details (e.g., "for two weeks," "every night at 2 AM"). We acknowledge this is a simplified, exploratory model and frame its results accordingly.

Metaphor and Idiom Usage: We assessed this using custom lexicons of common English metaphorical phrases (e.g., "weight on my shoulders") and Chinese idioms (Chengyu). The lexicons were compiled from established linguistic resources and refined by bilingual speakers to ensure cultural relevance.

PHQ-9 Topic Coverage: We evaluated this by scanning each dialogue for keywords related to the nine symptom categories to calculate a topic coverage score (0-9).

4. Results

Our analysis reveals a clear narrative: the cultural personas created distinct linguistic environments, which in turn

fostered different patterns of self-disclosure and emotional progression.

4.1. Linguistic Style Differences

The two styles tapped into forms of speech which differed markedly in their linguistic structures. As shown in Table 1, the English-speaking clients made use of the word "I" far more often than did the Chinese-speaking clients (8.35: 0.90 per 100 words) which not only reflects greater use of this term in speech [8] but also different cultural norms. Raoul used the first person plural, "we" in his speech over two times as frequently as did Auntie Mei (0.15: 0.07), reflecting a greater demand for cooperation, and a tendency to use explanatory phrases, such as "Shall we work that out together?" Auntie Mei's mode of speech was less oriented toward the use of the words associated with the person, and tended to feature those related to sociability, with a greater use of these words (0.13 versus 0.06), and was more geared toward support, approaching it from the perspective of the caring elder, as reflected in the translated expressions such as, "The aunt has heartache for you." Auntie Mei also made use of a significantly greater number of words dealing with cognitions (1.60 vs. 0.68), with a greater tendency to express awareness of cognitions in the form of phrases such as, "I understand your thoughts," which would indicate a more explanatory and reflective mode of counseling practice.

Table 1. Linguistic profile of dialogues by persona and role (frequencies per 100 words)

LIWC Feature	English Client	English Counselor (Raoul)	Chinese Client	Chinese Counselor (Auntie Mei)
Words per message	~120 (avg)	~80 (avg)	~130 (avg)	~90 (avg)
Positive Emotion	0.39	0.34	0.29	0.29
Negative Emotion	0.87	0.51	0.46	0.48
"I" Pronoun	8.35	1.29	0.90	1.27
"We" Pronoun	0.01	0.15	0.01	0.07
Social Words	0.12	0.06	0.24	0.13
Cognitive Process	1.57	0.68	2.25	1.60

4.2. Disclosure and Sentiment Trajectories

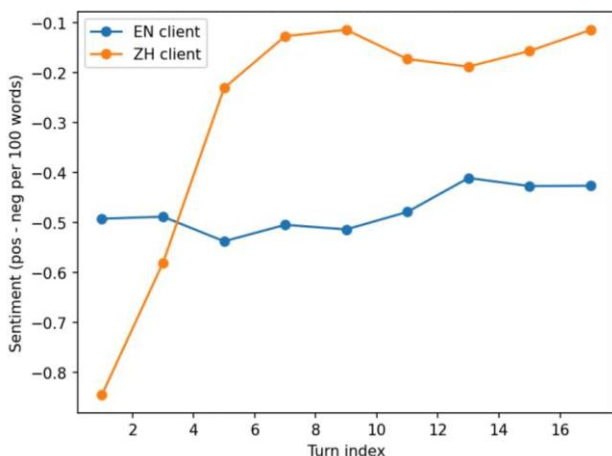


Figure 1. Average client sentiment by turn index. ZH sentiment improves early and stabilizes, while EN sentiment improves more gradually

These distinct differences in the language of the speakers were associated with particular patterns of disclosure and emotions. The emotion trends in Figure 1 reveal another salient difference in the emotional arc of the conversations. Client emotions in Chinese improved rapidly in the beginning or middle part of the interview and levelled off at a near

neutral state, while the English emotions of the clients improved much more slowly and in a roughly linear fashion during the interview.

This dynamic is reflected in the patterns of self-revelation, as shown in Figure 2. The English clients began with a high level of self-disclosure, which fits a more direct style of communication, that of "opening up". A typical opening from an English-speaking client was, "I haven't been sleeping very well lately. I'm lying in bed for hours at night. I think I'm feeling quite depressed these days." The effect of this on the initial score is that it was so direct and out in the open. In contrast, the Chinese clients began with a much lower level of self-disclosure which increased steadily, suggesting a slower style of working, which was in the nature of a trust-building process. A typical Chinese opening would be more reticent: "Well, my sleep has not been good lately and the stress level is rather high, otherwise ... I am alright." Only after a few turns of gentle probing by Auntie Mei, would there be disclosure of deeper-lying issues such as anhedonia. It is interesting to note that the quick improvement in affect for the Chinese clients coincided with this period where there was increasing self-disclosure, suggesting that the style of Auntie Mei which is one of high context and trust building, may be even more efficacious in bringing about quicker emotional relief when the level of trust has been established.

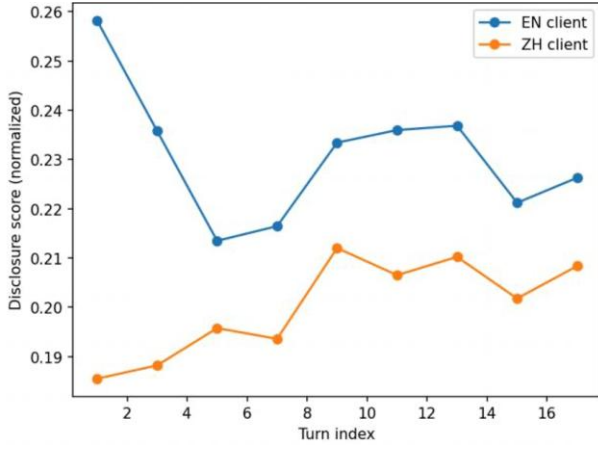


Figure 2. Client self-disclosure trajectory. EN clients exhibit high initial disclosure, while ZH clients show a more gradual increase

4.3. Use of Cultural Metaphors

The conversations of Auntie Mei were rich in cultural metaphors capable of yielding an average of approximately 2.3 idiomatic phrases per conversation, while Raoul’s average was approximately 0.7. Auntie Mei used Chengyu (idioms), such as “jiē líng hái xū xì líng rén,” literally, “The person who ties the bell on the tiger must untie it,” in a gentle manner with her client in conversation. Such use of a cultural asset provided a resonance with the client such that the clients frequently echoed the metaphor back to her and experienced a sense of validation for their own feelings. “That’s about it,” after Auntie Mei had indicated to the now busy life involved a sense of suffocation, one said, “I feel that is exact, I can’t breathe,” which validated the metaphor. Raoul did not use the idiomatic use of language but relied rather on getting back the actual words of the other person instead of the reflection of their meanings.

4.4. PHQ-9 Topic Coverage

Both personas were effective in eliciting the client’s problems from the discussions that addressed a balance and depth of topics related to depression (average of 3.7 topics for the EN, 3.5 for the ZH examples out of a max of 9 topics). The topics most frequently discussed were feelings, energy and sleep, which were the topics with which we began the scenarios. Appetite was discussed more frequently in the Chinese dialogues, probably because the Chinese show their concern for others by asking them about eating. Notably, suicidal ideation (PHQ-9 item #9) was not expressed in any of the dialogues, doubtless because of the safety considerations built into the prompt design.

5. Discussion

Our results show that injecting cultural personae into LLMs produces a strikingly different, measurable product in the course of the simulated therapy sessions. However, what we achieved was not just an observation of linguistic differences, but a proportional analysis of language, culture, and a developmental stage of emotional communication. The western ‘big brother’ and eastern ‘Auntie’ identities not only had distinct ways of linguistic behaviour but indeed distinct extralinguistic interaction frameworks. The extensive use of the collective ‘we’ and direct emotional expression in the mode of communication with Raoul represented norms extant in individualistic cultures. Auntie Mei’s use of social-relational linguistic means and culturally-related metaphors,

produced however, not direct interaction but a more indirect communicational, high-context form of conversation, which produced a slow, but deep emotional spilling of the client. It is evident therefore that a good AI therapist cannot be a one-size-fits-all person, but that the style of communication must have some culture-congruence so that relational proximities can be established and free expression secured. The different emotional directions of growth suggested above are interesting. The fast but early rise in the Chinese conversations indicates that an AI which has established the bonds of cultural similarity (say, through use of indigenous metaphors) will lead to speedier early emotional relief, for high-context users. There are thus apparent design implications arising out of this. A culturally-receptive chat-room individual may prove a more effective converser-partner in the short-term initial applications. This finding also has interesting implications in relation to the meanings contained in the notion of what constitutes a “successful” therapeutic interview. Is it one which ends on the most positive note or is it one which achieves the greatest general emotional re-structuring, even from a lower base? Our findings suggest that different cultural styles might be better adapted for either outcome.

Further, this work elucidates the problem of quantitatively assessing psychological constructs like self-disclosure across cultures. The heavy reliance in Western literature on “I” pronouns as a measure of self-disclosure [7] is clearly insufficient for pronoun-dropping languages like Chinese. Our composite measure, admittedly exploratory in nature, offers a solution toward the development of more culturally sensitive measures for assessing AI-human interaction. The necessity of such measures will be accentuated as AI systems become more worldwide in their utility.

This study has many limitations. First, the dialogues were entirely AI-simulated and may not closely reflect human conversation. Real users might be more erratic and respond differently to similar personas. Second, in our design, persona and language are confounded, and the effects of culture cannot be fully disentangled from the effects of the structure of language itself. Third, our measures of self-disclosure and sentiment, while useful for finding trends, are based on grossly simplified models and require further validation. Finally, the exclusion by design of high-risk topics such as suicidal ideation means that the findings of the present study are most applicable to supportive counseling for the mildly to moderately distressed.

6. Conclusion

In this work, we quantitatively demonstrate that instilling cultural personas in LLMs significantly alters the dynamics of bilingual therapeutic conversations. We find that for the two cultures studied, a western “big brother” persona and a Chinese “auntie” persona generate different styles of language usage, in turn responsible for different rates of self-disclosure growth and affective gains. The main contribution of this work lies in the demonstration that cultural congruence in AI engagement is not a superficial element of interaction, but more likely a “core” element which appears responsible in determining user trust, openness and emotional degree of response.

These results have clear implications for the nature of future mental health technologies, in that they point to a need for a move from translation to true cultural localization of AI personae and the methods of communication entered into by

them. Future studies might consider replicating these results using human subjects, in order that the applicable effectiveness of the different personae may be assessed. Further studies might also concentrate on developing improved, more sophisticated, and more cross-culturally valid metrics for disclosure and rapport. Additionally, dynamic systems that can change their AI identity according to user feedback could be of great help.

References

- [1] S. M. Schueller, M. G. Titov, D. C. Mohr, "Harnessing technology for mental health: The path to a digital service delivery model", *JAMA Psychiatry*, 2023.
- [2] B. A. Johnson et al., "The effectiveness of conversational agents for the treatment of depression and anxiety: A systematic review and meta-analysis", *J. Med. Internet Res.*, 2023.
- [3] J. Weisz et al., "A randomized controlled trial of a chatbot for depression and anxiety", *NPJ Digit. Med.*, 2021.
- [4] M. Malik, J. Jiang, K. M. A. Chai, "An empirical analysis of the writing styles of persona-assigned LLMs", in *Proc. EMNLP*, 2024.
- [5] L. O. H. Krocze et al., "The influence of persona and conversational task on social interactions with a LLM-controlled embodied conversational agent", *Comput. Human Behav.*, 2025.
- [6] C. Hill, K. O'Brien, *Helping Skills: Facilitating Exploration, Insight, and Action*, 4th ed., American Psychological Association, 2014.
- [7] A. K. Reuel, S. Peralta, J. Sedoc, G. Sherman, L. Ungar, "Measuring the language of self-disclosure across corpora", in *Findings of the Association for Computational Linguistics: ACL 2022*, 2022.
- [8] S. O. Yoon, J. M. Kim, S. C. Park, "Culture moderates changes in linguistic self-presentation and detail provision when deceiving others", *PLoS ONE*, 2017.
- [9] G. Lakoff, M. Johnson, *Metaphors We Live By*, University of Chicago Press, 1980.
- [10] L. M. McMullen, D. Tay, "Research review of psychotherapists' use of metaphors", *Psychotherapy*, 2023.
- [11] P. Jantondaeng, Z. Zheng, "A study of Chinese idiom expressions containing the word 'jin' (gold)", *J. Liberal Arts, Prince of Songkla Univ.*, 2024.
- [12] M. Salvi, "Persona consistency in dialogue systems: A survey", in *Proc. COLING*, 2022.
- [13] M. Eshghie, S. Eshghie, "ChatGPT as a therapist assistant: A suitability study", *arXiv*, 2023.
- [14] A. Fiske, P. Henningsen, E. Buyx, "Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy", *J. Med. Internet Res.*, 2019.
- [15] H. Li, Z. Cai, A. C. Graesser, Y. Duan, "A comparative study on English and Chinese word uses with LIWC", in *Proc. 25th Int. Florida Artif. Intell. Res. Soc. Conf. (FLAIRS-25)*, 2012.