Humanities & Social Sciences Communications

ARTICLE

https://doi.org/10.1057/s41599-025-04657-7 OPEN

Leveraging large language models to assist philosophical counseling: prospective techniques, value, and challenges

Bokai Chen[1,5], Weiwei Zheng[2,5], Liang Zhao[2,3]✉ & Xiaojun Ding[4]✉

Large language models (LLMs) have emerged as transformative tools with the potential to revolutionize philosophical counseling. By harnessing their advanced natural language processing and reasoning capabilities, LLMs offer innovative solutions to overcome limitations inherent in traditional counseling approaches—such as counselor scarcity, difficulties in identifying mental health issues, subjective outcome assessment, and cultural adaptation challenges. In this study, we explore cutting-edge technical strategies—including prompt engineering, fine-tuning, and retrieval-augmented generation—to integrate LLMs into the counseling process. Our analysis demonstrates that LLM-assisted systems can provide counselor recommendations, streamline session evaluations, broaden service accessibility, and improve cultural adaptation. We also critically examine challenges related to user trust, data privacy, and the inherent inability of current AI systems to genuinely understand or empathize. Overall, this work presents both theoretical insights and practical guidelines for the responsible development and deployment of AI-assisted philosophical counseling practices.

[1] Department of Psychology, Academy of Advanced Interdisciplinary Studies, Wuhan University, Wuhan, China. [2] School of Information Management, Wuhan University, Wuhan, China. [3] Intelligent Computing Laboratory for Cultural Heritage, Wuhan University, Wuhan, China. [4] Department of Philosophy, School of Humanities and Social Science, Xi'an Jiaotong University, Xi'an, China. [5] These authors contributed equally: Bokai Chen, Weiwei Zheng.
✉email: liangzhao@whu.edu.cn; xiaojunding@xjtu.edu.cn

Introduction

Philosophical counseling is an emerging discipline that applies philosophical methods to help individuals navigate life's challenges by bridging theoretical concepts with practical realities (Ding et al., 2024b; Louw, 2013; Savage, 1997). Unlike traditional psychotherapy, philosophical counseling focuses on examining unexamined assumptions, values, and reasoning patterns that may lie at the root of personal dilemmas (Cohen and Zinaich, 2013). By engaging in dialogue with a trained philosopher, counselees are encouraged to explore profound questions, gain fresh perspectives on their problems, challenge existing beliefs, and develop more robust ways of thinking about their lives and the surrounding world (Gindi and Pilpel, 2015; Grimes and Uliana, 1998).

While philosophical counseling holds significant potential, it presently faces several challenges as an emerging field (Amir, 2004; Knapp and Tjeltveit, 2005; Louw, 2013). One notable challenge is the limited number of trained philosophical counselors, which may restrict access to these services. In addition, the lack of standardized protocols and the inherently subjective nature of evaluating counseling outcomes might hinder its growth and broader acceptance (Knapp and Tjeltveit, 2005). Furthermore, many philosophical counselors may not possess extensive mental health training, potentially affecting their ability to adequately support counselees with psychological disorders. Addressing these challenges necessitates the development of innovative and carefully considered solutions.

In recent years, advanced AI technologies—particularly large language models (LLMs)—have demonstrated remarkable potential in natural language processing (NLP) tasks, owing to their expanded training data, enhanced model architectures, and exponentially increased parameters (Wu et al., 2024). LLMs have showcased powerful capabilities in translation, question-answering, and text generation (Bubeck et al., 2023; Chang et al., 2024) and have already been successfully applied to complex tasks such as psychological counseling and education (Fu et al., 2023; Liu et al., 2024). In the realm of philosophy, LLMs exhibit surprising ability to generate responses closely resembling those provided by human philosophers when faced with philosophical inquiries (Schwitzgebel et al., 2023). Moreover, they possess a degree of logical reasoning that facilitates the identification of common logical fallacies (Nutas, 2022). Furthermore, the intuitive, user-friendly interfaces of LLMs—capable of understanding and responding in natural language render them valuable tools for both the general public and researchers (Brown et al., 2020; Touvron et al., 2023b).

Can LLMs offer entirely new opportunities to enhance philosophical counseling? The answer is promising. Their advanced language processing and logical reasoning capabilities provide a strong foundation for integrating them into philosophical counseling. Their complex training process, which leverages enormous amounts of data including philosophical concepts, enables LLMs to retrieve necessary knowledge and generate responses during conversations, thereby creating an impression of "simulated understanding" for the user. Additionally, the user-friendly conversational interfaces of LLMs align well with the dialogical nature of philosophical counseling. Interest among philosophers in applying AI tools to philosophy has been noted (Clay and Ontiveros, 2023). However, despite such enthusiasm, the literature exploring the intersection of LLMs and philosophical counseling remains sparse. Notably, only Nutas (2022) discussed whether GPT-3 meets the fundamental requirements of philosophical counseling, yet this work falls short of further technical improvements and comprehensive analysis of its capabilities.

Integrating LLMs into philosophical counseling goes beyond mere compliance with existing requirements. Comprehensive investigations are necessary not only to establish baseline functionalities but also to shape future expectations. This integration involves addressing current counseling challenges, enhancing the efficacy of counseling sessions, and upholding ethical standards. We must not only investigate whether LLMs can be applied effectively but also consider the broader implications of their usage, including both the potential benefits and the problems they might introduce. Comprehensive research is essential to evaluate the capabilities of LLMs and their potential to solve real-world issues, as well as to understand the difficulties that might arise. Moreover, such research should reflect the latest technological advancements and incorporate the most relevant techniques associated with these models. As an initial exploration, this paper aims to systematically investigate the integration of LLMs into philosophical counseling by addressing the following three research questions:

RQ1: How can we technically facilitate LLMs to assist philosophical counseling?

RQ2: What value can LLMs bring to promote better philosophical counseling?

RQ3: What challenges could be encountered when integrating LLMs to assist philosophical counseling?

The remainder of the paper is organized as follows. Firstly, we review the development of philosophical counseling and highlight its current limitations. Subsequently, we introduce three primary technical approaches for applying LLMs, thereby establishing a technical foundation for LLM-assisted philosophical counseling. We then propose the potential value added by LLM assistance in addressing the current limitations of philosophical counseling. Finally, we discuss the challenges of integrating LLMs into philosophical counseling—particularly their inability to achieve genuine understanding and empathy. Through this comprehensive investigation, we argue that while LLMs cannot replace human counselors, they can serve as powerful tools to extend the reach and effectiveness of philosophical counseling.

Historical and contemporary perspectives on philosophical counseling

Evolution of philosophical counseling practices. Philosophical counseling, as a modern professional practice, has its roots in both ancient philosophical traditions and in the pioneering work of contemporary thinkers who sought to reapply philosophical wisdom to address personal and existential concerns (Amir and Fatić, 2015; Lahav and Tillmanns, 1995). Historically, this practice can be traced back to Socrates, who engaged Athenians in dialogues that challenged their assumptions and promoted self-examination. The Stoics, including Epictetus and Marcus Aurelius, as well as philosophers such as Epicurus, further developed methodologies aimed at achieving a good life through rational inquiry and ethical living, thereby emphasizing the practical application of philosophy (Hadot, 1995).

In the modern era, Pierre Grimes emerged as a foundational figure in philosophical counseling. Beginning in the 1960s in the United States, Grimes utilized philosophical dialogue—drawing particularly from Platonic and Neo-Platonic traditions—to help individuals explore personal dilemmas and achieve self-understanding. His approach, often referred to as "philosophical midwifery," emphasizes guiding individuals to uncover the underlying beliefs that contribute to their concerns (Grimes and Uliana, 1998). Grimes's significant contributions have led to his recognition as the originator of modern philosophical practice by the global community of philosophical practitioners, as notably acknowledged at the International Conferences on Philosophical Practice (ICPP).

Concomitantly, Leonard Nelson and his student Gustav Heckmann were instrumental in developing the Socratic Dialogue method in Germany during the early 20th century. Nelson's work—focusing on critical philosophy and ethical socialism—promoted collective philosophical inquiry as a means to solve societal problems and foster democratic thinking (Nelson, 1949). Heckmann continued this tradition by applying Socratic Dialogue in educational settings and adult learning environments (Heckmann, 1981).

Gerd B. Achenbach further advanced the field by establishing the first formal institution of philosophical practice (Philosophische Praxis) in Germany in 1981 (Achenbach, 1984). Achenbach advocated a return to the practical roots of philosophy, distinguishing philosophical counseling from psychotherapy and emphasizing dialogue without predefined methods or therapeutic goals (Achenbach, 2010). His work significantly popularized philosophical counseling in Europe and inspired subsequent practitioners and scholars.

In North America, Lou Marinoff played a key role in bringing philosophical counseling to public attention with his book *Plato, Not Prozac* (Marinoff, 1999), promoting philosophy as a practical tool for addressing everyday problems. Similarly, Shlomit C. Schuster contributed to the field through her comprehensive exploration of philosophical practice as an alternative to traditional counseling and psychotherapy (Schuster, 1999).

Collectively, these figures have been instrumental in the resurgence and development of philosophical counseling as a distinct approach for addressing the complexities of life through philosophical inquiry and dialogue. By acknowledging their contributions, we aim to present a comprehensive historical context that reflects the richness and evolution of philosophical counseling.

Limitations in current philosophical counseling. Despite its solid foundation in various philosophical traditions and the significant contributions of figures such as Grimes and Marinoff, the relatively short history of philosophical counseling as a formalized practice has resulted in several practical challenges. While its theoretical depth and potential for addressing complex life issues are undeniable, traditional philosophical counseling faces limitations that restrict its accessibility and effectiveness for a broader audience. These challenges include the difficulty philosophical counselors experience in identifying mental health issues, the subjectivity in evaluating counseling outcomes, the scarcity of trained professional philosophical counselors, and cultural barriers that affect its acceptance and application.

Challenges in identifying mental health issues. Certain issues addressed within philosophical counseling overlap with mental health problems. Although there are significant theoretical distinctions between philosophical and psychological counseling, the general public often fails to distinguish between the two. This overlap occasionally leads philosophical counselors to encounter counselees with genuine psychological disorders. According to Knapp and Tjeltveit (2005), philosophical counseling can be broadly categorized into narrow and broad approaches. Regardless of the approach, philosophical counselors typically lack the specialized psychological expertise required to identify mental health issues that exceed their professional competence.

In narrow-scope philosophical counseling, practitioners focus on distinctly philosophical problems and usually avoid issues typically addressed by psychologists. However, counselees' philosophical concerns do not necessarily guarantee the absence of mental health issues. For example, individuals experiencing severe depression or suicidal ideation might seek philosophical counseling to discuss existential questions about life and death. If a counselor remains unaware of the underlying psychological conditions, they may inadvertently overlook signs of depression that require psychological or medical intervention.

Conversely, practitioners who adopt a broader approach tend to address a wider array of issues by interpreting philosophical problems to encompass various mental health concerns that are not strictly medical or biological. This perspective considers philosophical counseling as an alternative mental health treatment for individuals dealing with troubled relationships, life crises, depression, or anxiety (Marinoff, 1999; Schuster, 1999). However, this assumption is problematic, as most mental illnesses involve a complex interplay of biological, social, and psychological factors (Kinderman, 2005). Philosophical counselors, who typically lack comprehensive mental health training, may find it difficult to determine the primary cause of a client's psychological issues.

Subjectivity in assessing counseling outcomes. Philosophical counselors have an ethical obligation to empirically demonstrate the efficacy of their methods (Knapp and Tjeltveit, 2005). However, evaluations of the effectiveness of philosophical counseling are often neglected or rely solely on the counselor's subjective assessment. Such reliance on subjectivity may stem from a misunderstanding of scientific methodology and empirical testing (Kreimer and Primero, 2017), leading some to question the legitimacy of philosophical counseling by labeling it pseudoscientific due to a lack of empirical evidence (Sivil and Clare, 2018).

To address these concerns, a mixed-methods approach that incorporates both standardized psychological evaluation tools and qualitative interviews may provide a robust framework for assessing outcomes (Tashakkori and Creswell, 2007). Instead of relying solely on conventional psychological scales—which may be incongruent with the rationalist foundations of philosophical counseling—a specialized scale focusing on client satisfaction and overall well-being could be developed. By emphasizing satisfaction rather than direct efficacy, this approach aligns with the core principles of philosophical counseling while also addressing the demand for empirical accountability. Additionally, qualitative interviews with counselees can capture in-depth, nuanced insights into their experiences, further enriching quantitative data and creating a more holistic assessment framework.

While some philosophical counselors may find this combined approach beneficial, the implementation of mixed-methods assessments poses its own challenges. Many counselors, primarily trained in rational and dialogical methods, may lack familiarity with statistical tools and methodologies, and qualitative interviews require significant time and resources to conduct, transcribe, and analyze.

Scarcity of professional philosophical counselors. Solving life's problems and pursuing well-being are universal human needs; however, the current number of professional philosophical counselors is clearly insufficient compared to these needs. Becoming a practitioner in philosophical counseling generally necessitates a solid theoretical background and profound philosophical knowledge, which in turn requires extensive academic study and practical experience. Recent trends have indicated a decline in the number of individuals choosing to study philosophy and obtain relevant degrees (Badola, 2015; National Center for Education Statistics Database, 2023). In China, for instance, fewer than one in a thousand university graduates receives a degree in philosophy (Ministry of Education of the People's Republic of China, 2023), thereby limiting the pool of potential philosophical counselors and exacerbating the scarcity of professional talent.

Although global philosophical cafés have increasingly attracted participation from both academic and non-academic philosophers (O'Neill and Wang, 2021), philosophical counseling—given its relatively short formalized history—has not yet achieved the same level of public recognition and acceptance as psychological counseling, which enjoys broader familiarity and application. Consequently, even among philosophy students, awareness of the methods and existence of philosophical counseling may be limited, further hindering its dissemination and acceptance. As a result, individuals seeking philosophical counseling often face significant challenges in locating a counselor whose expertise matches their needs.

Cultural barriers and resistance to adoption. To foster the worldwide development of philosophical counseling, it is essential to address the cultural differences that shape its reception across diverse regions. Different countries are characterized by unique languages and philosophical traditions, meaning that a counseling approach effective in one cultural setting might not translate seamlessly to another. Moreover, the use of various native languages complicates communication and learning between philosophical counselors from different cultures, thereby further hindering the international exchange of ideas and practices.

For instance, China presents a compelling case study of these cultural barriers. Its deeply ingrained collectivist culture emphasizes emotional restraint; when confronted with psychological distress or inner conflict, many individuals prefer to suppress their emotions rather than seek constructive avenues for expression (Zhang and Yan, 2012). Additionally, the Chinese cultural value placed on harmonious interpersonal relationships may discourage individuals from engaging in philosophical debates, which are sometimes perceived as confrontational (Wei and Li, 2013). Furthermore, the substantial influence of Confucianism and Daoism in the Chinese consciousness results in widely entrenched philosophical frameworks that may not easily align with conventional Western approaches to counseling (Hu, 2024).

These cultural factors have constrained the evolution of philosophical counseling in China. A uniquely Chinese approach—rooted in the country's rich cultural and philosophical heritage—may prove to be more suitable and effective in fostering acceptance and engagement (Ding and Yu, 2024).

LLMs: pioneering a new era in philosophical counseling

Philosophical counseling represents a distinctive approach that bridges theoretical perspectives with practical applications to help individuals overcome life's challenges. Nonetheless, its wide-scale implementation is hindered by several critical obstacles. Fortunately, the advent of LLMs offers a promising avenue for overcoming these limitations. By leveraging advanced language processing capabilities, LLMs enable intelligent and user-friendly services that enhance both the accessibility and effectiveness of philosophical counseling.

Capabilities and applications of LLMs in philosophy

LLMs and their abilities. The rapid development of artificial intelligence—particularly through the transformer network architecture based on attention mechanisms (Vaswani et al., 2017)—has been transformative. Models such as BERT (Devlin et al., 2019) and GPT (Radford et al., 2018) have significantly advanced LLMs over the past five years. Today, LLMs, as exemplified by GPT, Gemini, Claude, Grok and DeepSeek, demonstrate extraordinary natural language understanding (encompassing tasks such as intention recognition and entity extraction) and natural language generation capabilities, owing to their vast training datasets and sophisticated technical architectures (Chang et al., 2024).

Trained on petabyte-scale datasets and utilizing architectures with billions of parameters, LLMs deliver robust performance in generating contextually relevant responses to diverse natural language queries. These abilities make them valuable assets across various domains, enhancing efficiency and expanding functionality. Moreover, techniques like fine-tuning could further improve their performance on specialized tasks, bolstering natural language understanding, logical reasoning, mathematical computation, and alignment with user expectations. For individual users, ChatGPT's easy accessibility supports daily activities such as study, work, and provides both emotional and intellectual assistance via natural language interaction (Alqahtani et al., 2023).

Researchers in various disciplines are exploring the integration of LLMs into domain-specific applications. In particular, the use of LLMs in psychological counseling and related philosophical fields paves the way for their adoption in philosophical counseling. In the realm of philosophy, LLMs show promise as revitalized pedagogical tools for fostering philosophical dialogue (Smithson and Zweber, 2024), as they can mimic the discourse characteristic of human philosophers. Models trained on philosophical texts can generate responses that are nearly indistinguishable from those of professional philosophers when addressing similar questions (Schwitzgebel et al., 2023). Liu et al. (2024) employed LLMs for Socratic teaching, indirectly demonstrating their potential to facilitate heuristic dialogues in a counseling context.

LLMs have also exhibited strong capabilities in mental health support. They can accurately recognize and respond to emotional cues, and collaborations between humans and machines in psychological support have heightened the expression of empathy—a key component in effective counseling (Patel and Fan, 2023; Schaaff et al., 2023; Sharma et al., 2023). Furthermore, LLMs are capable of performing mental health assessments (Kjell et al., 2023; Levkovich and Elyoseph, 2023) and delivering personalized interventions (Blyler and Seligman, 2024a, 2024b).

The proficiency of LLMs in philosophical dialogue, mental health assessment, and intervention underscores their potential role in advancing these fields. Their integration could render LLMs invaluable resources in contemporary philosophical counseling and mental health practices, opening new possibilities at the intersection of artificial intelligence, philosophy, and mental health.

Essential requirements for AI-assisted counseling. For LLMs to function effectively as artificial philosophical counselors, they must possess several core abilities that align with the foundational principles of philosophical counseling. These include:

1. **Capacity for logical reasoning and philosophical dialogue**: LLMs should engage in coherent logical reasoning that facilitates thoughtful and meaningful philosophical conversations with counselees. This requires understanding complex philosophical concepts and applying them appropriately in the context of each client's concerns (Daniel and Auriac, 2011; Marinoff, 2002).
2. **Recognition and appropriate response to emotional cues**: Although LLMs do not possess true consciousness or genuine empathy, they should be capable of identifying emotional cues in a client's language and responding sensitively. This form of simulated empathy can help establish rapport and support the client effectively (Patel and Fan, 2023; Sharma et al., 2023).

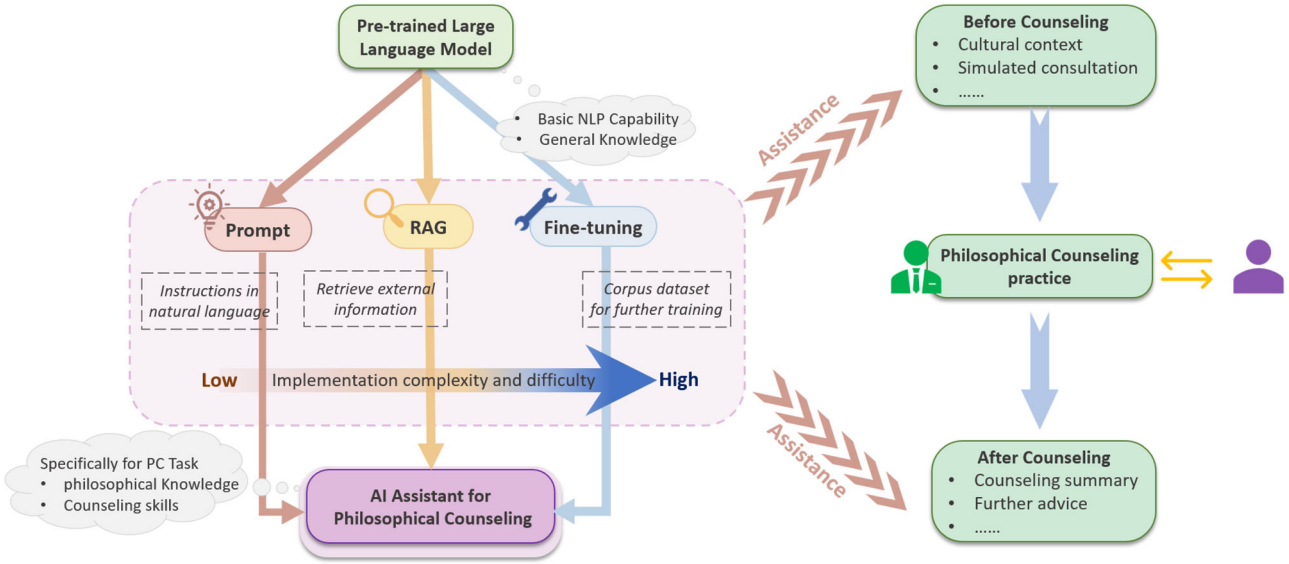

Fig. 1 Technical framework for integrating LLMs as assistive tools in philosophical counseling (PC).

3. **Awareness of limitations and referral to human intervention**: LLMs should be designed to recognize situations where a client's issues exceed their capabilities —such as severe emotional distress or mental health disorders—and promptly recommend seeking assistance from a qualified professional (Knapp and Tjeltveit, 2005; Obradovich et al., 2024).
4. **Facilitation of critical self-reflection**: LLMs should assist counselees in critically examining their assumptions and beliefs by guiding them through philosophical inquiry. This process should foster deeper self-understanding and encourage the exploration of alternative perspectives, without solely relying on emotional rapport (Cohen and Zinaich, 2013; Gindi and Pilpel, 2015).

While LLMs can process and generate language that appears to engage with philosophical concepts or simulate empathy, they neither truly understand nor authentically empathize. Philosophical understanding—explored across various traditions—transcends mere symbol manipulation based on statistical patterns; it requires a reflective, interpretive engagement with meaning, a capacity that non-conscious entities like LLMs lack. Similarly, their simulated empathy is merely a functional imitation derived from probabilistic patterns in data.

Despite these limitations, LLMs hold significant potential as assistants in philosophical counseling. As artificial Socrates, they can facilitate philosophical inquiry and support human counselors. Although they may not independently perform full-scale philosophical counseling, LLMs can serve as valuable tools to augment professional counselors by enriching dialogues, offering philosophical insights, and streamlining the counseling process.

Technical strategies for integrating LLMs into counseling. Although LLMs are generally powerful—being pre-trained on large-scale general knowledge—they may not reach their full potential in specialized domains without additional training. For tasks such as philosophical counseling, domain-specific learning (e.g., fine-tuning on philosophical texts) is essential to further specialize LLMs (Schwitzgebel et al., 2023).

In this section, we present three mainstream technical solutions for tuning general LLMs into philosophical counseling assistants, thereby maximizing their potential. As illustrated in Fig. 1, traditional philosophical counseling provides both the theoretical foundation and practical examples upon which LLM-assisted counseling is built. By following technical pathways such as prompting, fine-tuning, and retrieval-augmented generation, LLMs can be endowed with capabilities that are more tailored to the unique requirements of philosophical counseling, thereby offering additional advantages.

Prompting. Prompting involves crafting specific prompts to enhance the performance of LLMs on downstream tasks. This simple yet efficient method allows for customization of the model for particular tasks without modifying its underlying structure (Demszky et al., 2023; PF Liu et al., 2023). By providing a few training examples as prompts during interaction, the model can be instructed to perform the desired tasks. The key is to leverage appropriate task descriptions or examples that guide the model's reasoning, enabling it to fully utilize its pre-trained abilities. The simplicity of prompting rests on designing effective prompt content—such as annotating the domain and providing illustrative examples. Continuous modification and refinement, often called prompt engineering, are typically required to achieve optimal results. One widely used approach is the "Chain of Thought" (CoT) method (Wei et al., 2022); similar to thinking "step by step," CoT strengthens the reasoning capabilities of LLMs by incorporating explicit reasoning steps into the prompts.

Retrieval-augmented generation. Retrieval-augmented generation (RAG) is an advanced technique that enhances LLM capabilities by incorporating external knowledge. Although general LLMs are pre-trained on vast amounts of internet data, they still may lack sufficient domain-specific knowledge in specialized fields such as philosophy (Kandpal et al., 2023).

RAG addresses this shortcoming by retrieving relevant information from a database prior to the generation process, thereby allowing the model to access a wealth of recent or domain-specific data. The retrieved information is used to inform or augment the model's responses, resulting in outputs that are more accurate, reliable, and informative (Gao et al., 2024). The RAG process typically comprises three stages: indexing, retrieval, and generation. Initially, documents from a knowledge base are indexed and encoded into vectors. When a query is issued, the system retrieves the most relevant documents based on semantic similarity calculations and employs them to generate a comprehensive response (Ma et al., 2023). The construction of robust

external knowledge bases is central to RAG, making high-quality, domain-specific knowledge essential, particularly in the context of philosophical counseling.

The advantages of RAG include enhanced reliability and interpretability of domain-specific information, as well as a greater diversity of retrieved data—which may enrich the generated responses. However, challenges such as high retrieval costs and the effective organization of the retrieved information persist, necessitating further exploration and optimization.

Fine-tuning. Fine-tuning involves adjusting pre-trained LLMs using relatively small, task-specific datasets to improve their performance on particular tasks. This process requires preparing specialized training datasets, sufficient computational resources, and technical expertise to meet the model's training needs. Reinforcement Learning from Human Feedback (RLHF) is frequently employed to further optimize the model's performance; in this process, high-quality answers generated by the fine-tuned model are collected and then used to train a reward model that guides reinforcement learning.

In the context of philosophical counseling, fine-tuning can be applied to dialogue models based on existing open-source LLMs, such as the LLaMA series (Touvron et al., 2023a; 2023b). This approach allows for adjustments in communication style, tone, and format, thereby enhancing the reliability of the dialogue output. Fine-tuning is particularly beneficial for tasks that cannot be effectively addressed through prompting alone, especially when prompt construction is complex or challenging.

The effectiveness of fine-tuning heavily depends on the quality of the training corpus. A well-curated corpus enriched with dialogues on philosophical counseling and mental health care enables the model to better cater to the emotional and psychological needs of counselees. It also allows LLMs to adopt the linguistic and stylistic nuances of philosophical discourse, resulting in responses that closely align with field-specific expectations.

Despite its advantages, fine-tuning has notable drawbacks—chief among them being high training costs. Furthermore, uploading training data via APIs can escalate expenses, while local training demands significant computational resources and technical expertise. Additionally, fine-tuning may lead to issues such as "catastrophic forgetting" (Kumar et al., 2022), whereby adjustments to the model's parameters cause it to lose previously acquired knowledge.

In this section, we comprehensively compare these three technical solutions for customizing LLMs as philosophical counseling assistants, as detailed in Table 1, by evaluating nine aspects (e.g., implementation, data requirements, effectiveness, etc.), thus aiding in the practical selection of the most appropriate approach.

Dominant value of LLM-assisted philosophical counseling

Incorporating LLMs into philosophical counseling is not merely about automating consultation services; rather, its true value lies in enhancing the entire counseling process—including pre-counseling psychological assessments, in-session guidance, and post-counseling documentation and evaluation. Furthermore, from a broader perspective, the vast repository of knowledge contained in LLMs can be adapted to meet the needs of a multicultural context. These four benefits directly address the major limitations of philosophical counseling discussed in the Section "Limitations in current philosophical counseling", demonstrating the potential for LLM assistance to improve the effectiveness of current practices, align philosophical counseling with modern societal demands, and expand its accessibility to a broader audience. In this section, we detail the dominant value of LLM assistance along the following four dimensions.

Mental health assessment and counselor recommendations. Although LLMs currently cannot replace professional mental health services (Obradovich et al., 2024), a philosophical counseling recommendation system based on them could be built. LLMs can be programmed to help identify potential mental health issues through psychological assessments. Additionally, by employing techniques such as sentiment analysis, LLMs can detect signs of mental health challenges from natural language inputs (Chen et al., 2024; Wankhade et al., 2022). Moreover, their predictive capabilities enable them to assess mental health states based on online textual data (Xu et al., 2024).

Beyond identifying mental health challenges, LLMs can assist ordinary counselees by recommending suitable counselors. The effectiveness of counseling often depends on the alignment between a counselor's approach and a client's profile including demographic, clinical, psychological, and cultural factors, as well as personal preferences (Zhou and Zhang, 2018). A personalized and adaptable approach is, therefore, essential for optimal outcomes. In this context, LLMs can facilitate tailored interventions that enhance the overall counseling experience. Fine-tuned on extensive feedback data, LLMs can help determine the most appropriate intervention type, whether psychological counseling, philosophical counseling, or alternative forms of support, based on the user's mental health status, preferences, and concerns (Galitsky, 2024). Crucially, users retain autonomy in selecting their preferred counseling approach. For instance, if philosophical counseling is chosen, LLMs can refine recommendations by suggesting specific modalities, such as Logic-based Therapy tailored to the individual's needs (Cohen, 2013; Cohen et al., 2024). Furthermore, the extensive knowledge base of LLMs and their capacity for domain-specific training enable them to comprehend and apply various philosophical frameworks, positioning them to recommend the philosophical school or individual philosopher most relevant to a given case.

The proposed framework, depicted in Fig. 2, outlines the application of LLMs in philosophical counseling systems designed to effectively direct users to professional mental health services when necessary. The process initiates with the user's informed consent, followed by an initial evaluation of the user's mental health status. This evaluation utilizes advanced LLMs to analyze the counseling issues presented. Based on this analysis, the system recommends appropriate counseling methodologies tailored to the user's specific needs. Subsequently, if the user agrees, the system suggests compatible counselors who are best suited to address the identified issues. Additionally, LLMs are employed to furnish users with carefully selected resources and pertinent information about available mental health service providers. This ensures that users with recognized mental health concerns are directed toward appropriate care.

Crucial to the success of this framework is the algorithm's ability to detect potential mental health issues with a high degree of accuracy, particularly focusing on high recall rates. This ensures that individuals who may have mental health concerns are reliably identified and receive the necessary attention, as emphasized by Rabani et al. (2023). This methodical approach not only heightens the efficacy of the counseling provided but also enhances the overall safety and well-being of the user.

Such recognition mechanisms could act as safeguards in long-term counseling. If a client exhibits significant mental health problems, the system could facilitate a referral to appropriate professionals—such as psychologists or psychiatrists—for specialized support, ensuring that philosophical counseling stays within

Table 1 Comparative analysis of LLM approaches for enhancing philosophical counseling practices.

	Prompting	Retrieval-augmented generation	Fine-tuning
Key idea	Utilizes the LLM's pre-trained knowledge without modifying its structure and parameters.	Enhances philosophical counseling capability by retrieving information from an external knowledge base without modifying the model.	Customizes the original LLM using an in-domain dataset, which involves updating model parameters.
Implementation	• Designing specific prompts to guide the model in philosophical counseling. • Relatively simple prompt design and optimization.	• Integrating external knowledge sources and retrieving information prior to generation. • Requires building an effective knowledge base and retrieval mechanism.	• Adjusting pre-trained model parameters using a task-specific corpus. • Involves extensive data preparation and computational resources.
Data required	None	An external knowledge base containing philosophical knowledge and semantic information.	Large philosophy-related datasets.
Customization	Low (dependent on prompt design; limited by pre-trained knowledge)	Moderate (customized through retrieval results and generation process)	High (requires updating data and parameters)
Performance consistency	Varies with prompt quality; may be inconsistent.	Relatively consistent, depending on the quality of the knowledge base and retrieval algorithms.	High consistency (dependent on the quality of the training data and process).
Cost	Low—mainly related to prompt design and modifications.	Moderate—includes costs for maintaining the knowledge base and implementing retrieval mechanisms.	High—involves data collection, extended training time, and significant computational resources.
Advantages	• Simple and user-friendly. • No modifications to model parameters required.	• Provides up-to-date and contextually relevant information. • Enhances accuracy and reliability of generated content.	• Improves task-specific performance. • Highly adaptable to specific needs.
Disadvantages	• Effectiveness depends on prompt quality. • Requires continuous adjustment and optimization.	• Complex implementation and high maintenance costs. • Requires efficient coordination between retrieval and generation processes.	• High training cost. • Potential issues, such as "catastrophic forgetting."

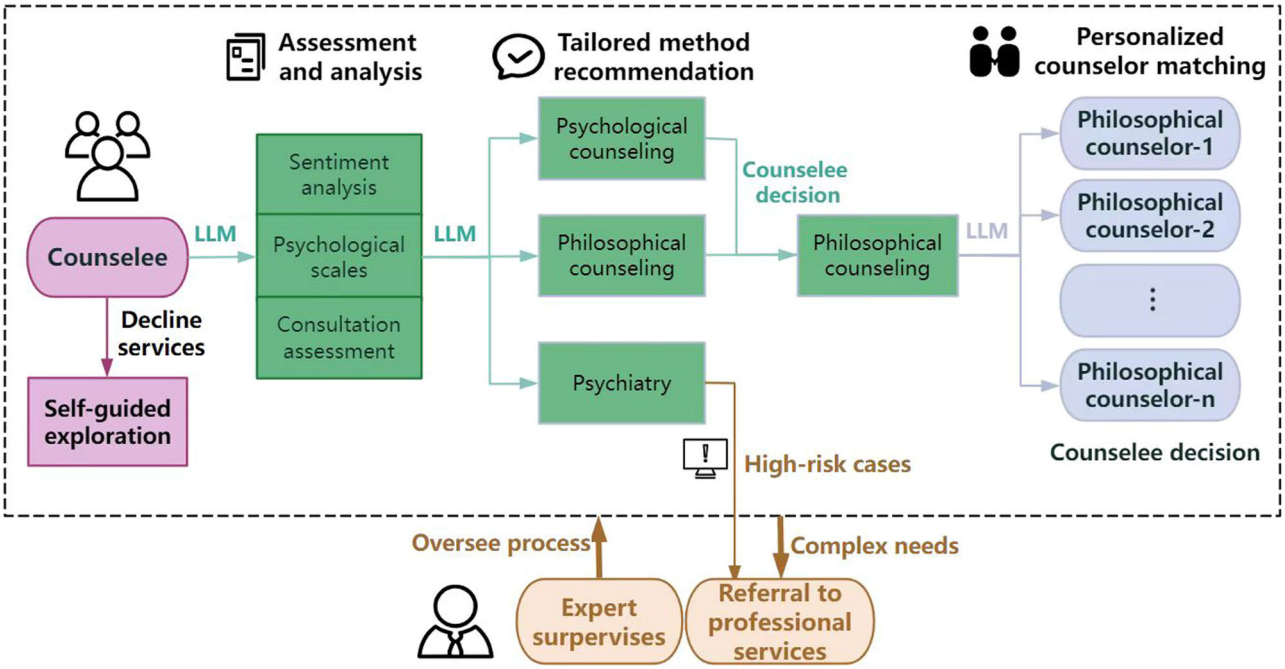

Fig. 2 Architectural blueprint of an LLM-enabled philosophical counseling recommendation system.

its scope while collaborating effectively with mental health services to address counselees' holistic well-being.

Optimized outcome assessment and session summarization. Traditionally, the assessment of outcomes in philosophical counseling has either been neglected or overly dependent on the subjective judgment of counselors, which undermines both objectivity and consistency. LLMs offer a viable method to address these issues by facilitating aspects of the evaluation process that reduce the impact of human bias. Their advanced natural language processing capabilities allow them to analyze conversational data and perform basic statistical analyses (Huang et al., 2024), thereby providing measurable support and objective evidence to counter skepticism regarding the effectiveness of philosophical counseling.

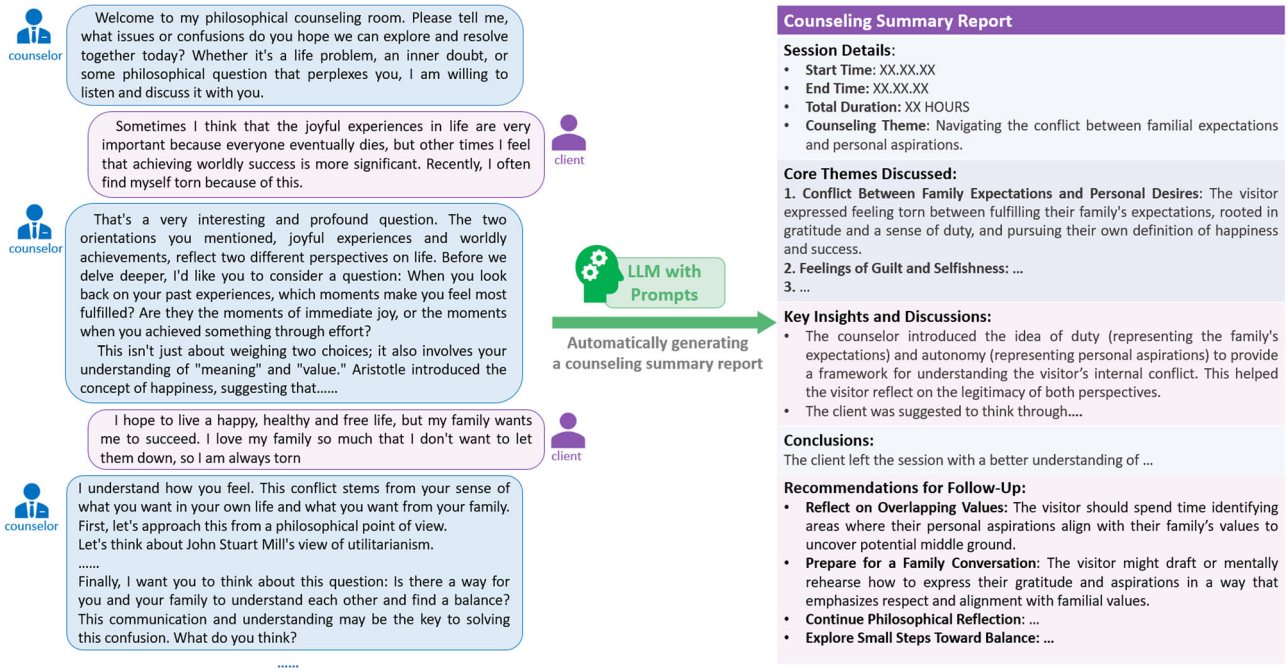

Fig. 3 Example of an LLM-generated session summary derived from counseling dialogues.

Philosophical counseling is characterized by its methodological diversity including the "beyond-method" approach, multiple schools of thought, and a wide-ranging array of client issues (Ding et al., 2024c; Fatić and Zagorac, 2016; Repetti, 2023). This diversity makes the uniform application of standardized psychological scales both impractical and inappropriate. Instead, LLMs can assist counselors by generating customized questionnaire items tailored to the specific methodologies and objectives of their practices, whether as qualitative open-ended questions or quantitative assessment items. Additionally, these models can analyze client responses to offer nuanced insights that enhance the evaluation process and contribute to overall counseling quality.

Beyond simply aiding in outcome assessment, LLMs can also summarize counseling sessions, as illustrated in Fig. 3. Using continuous prompt engineering, the LLM is capable of producing valuable session summaries for both counselors and counselees. The dialogue and summary shown here are generated using the GPT-4o model (see Fig. 3). Although LLMs sometimes produce inaccuracies or "hallucinations" that deviate from the original content (Adhikary et al., 2024), they can still generate detailed session summaries. Philosophical counselors can review and refine these summaries to ensure their accuracy and relevance before sharing them with counselees. This collaborative process not only preserves the integrity of session documentation but also reduces the administrative burden on counselors, thus enabling them to focus more on the substantive aspects of their practice.

LLMs can additionally offer valuable support in gathering user feedback post-counseling. Acting as feedback collection assistants, they can compile preliminary data on client perceptions and gather open-ended suggestions for improvement. This capability provides counselors with actionable insights to better inform their practice. As LLM capabilities continue to evolve, they are expected to deliver even more sophisticated feedback analysis, offering innovative perspectives and strategies for refining counseling methods.

Enhancing accessibility and visibility. The integration of LLMs into philosophical counseling opens significant opportunities for enhancing both accessibility and visibility in the field. One innovative application is the use of LLMs to create digital avatars for philosophical counselors—a concept that has shown promise in various contexts (Fink et al., 2024; Oliveira et al., 2024). Fig. 4 outlines the general steps involved in building a digital avatar based on LLMs. Multimodal technologies could be further incorporated. With the informed consent of both counselors and counselees regarding the use of counseling dialogue for AI training, LLMs can be fine-tuned using individual counselor dialogue data. This enables the models to simulate a counselor's distinctive style, expertise, and approach. Consequently, counselees can explore the methodologies and perspectives of a range of counselors before making an informed decision on whom to consult, thereby improving the selection process for professional support.

These avatars hold significant potential as supplementary tools. Analogous to the role of LLMs in psychology—where they assist with mental health support and address common queries (JM Liu et al., 2023; Na, 2024)—LLM-based philosophical counselor avatars can serve multiple functions. They can answer philosophical questions, clarify complex concepts, and stimulate reflective thinking through insightful questioning (Park and Kulkarni, 2023). For example, these models can be fine-tuned with specialized conversational datasets targeting specific demographics, such as children. By generating age-appropriate, engaging, and thought-provoking dialogues, such interactions can inspire critical thinking and curiosity from an early age. This adaptability allows these avatars to address the diverse needs of different audiences, thereby fostering intellectual growth and deep philosophical reflection.

In addition to facilitating philosophical dialogue, these avatars have the potential to generate philosophical counseling responses, especially given rapid advances in AI technologies. As demonstrated by Raile (2024) in his exploration of ChatGPT as a psychotherapist, such applications show considerable promise. However, this potential must be approached with precaution and thorough oversight. The responsibility for AI-generated content ultimately lies with human counselors or supervisory personnel. Prior to delivering philosophical counseling responses, the AI-

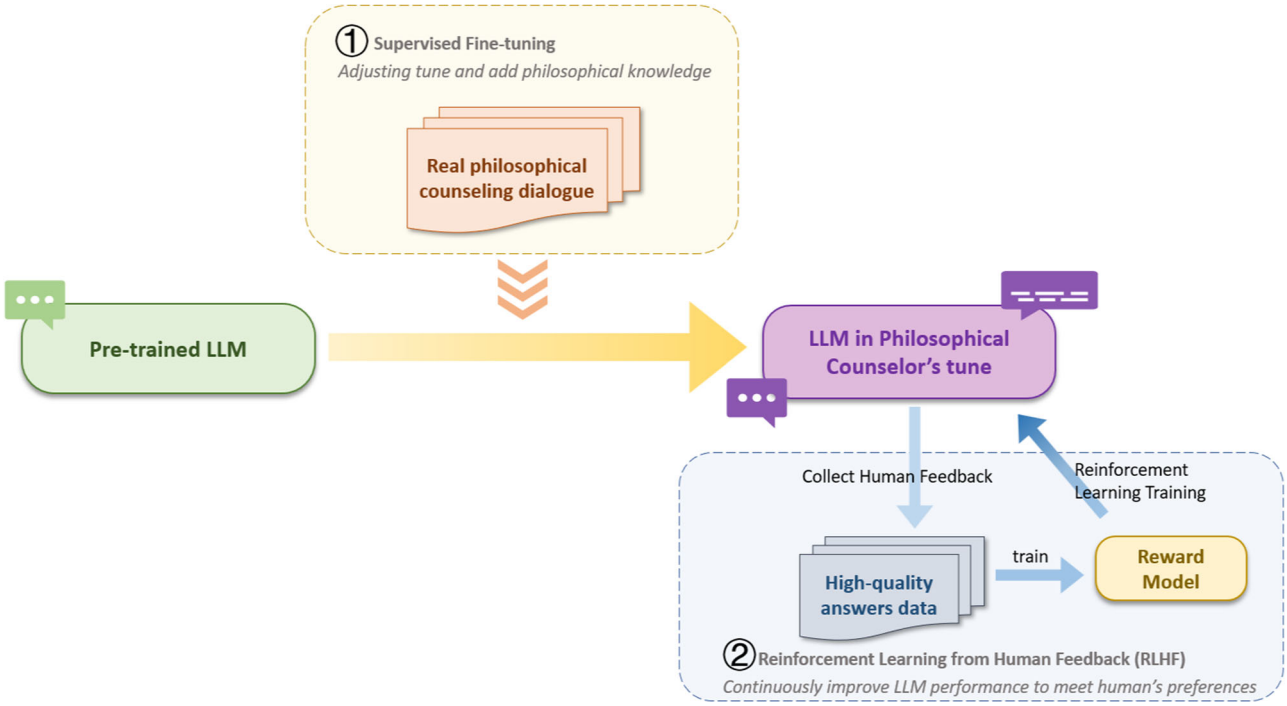

Fig. 4 Process for constructing digital counselor avatars using LLMs-related techniques.

generated outputs must undergo meticulous human review to ensure their accuracy, appropriateness, and safety. Even when the responses are acceptable, counselors are advised to use them merely as suggestions or starting points, thus fostering richer and more meaningful dialogue with counselees. This integration of human oversight and AI assistance underscores the compelling potential of LLMs in online philosophical counseling.

A critical advantage of integrating LLMs is their capability to overcome geographic and financial barriers. As cost-effective, accessible, and user-friendly platforms, LLMs offer a practical entry point for individuals who might otherwise lack access to professional counseling. This is particularly beneficial for those who are unfamiliar with or hesitant about traditional face-to-face counseling. For instance, through online platforms, individuals can interact with LLMs to explore philosophical inquiries, providing an affordable alternative to in-person consultations. This technological innovation broadens the reach of philosophical counseling, making it more inclusive and accessible to marginalized or economically disadvantaged communities.

Beyond enhancing accessibility, LLMs can help address the global shortage of professional philosophical counselors. By embedding expert philosophical knowledge within these models, LLMs provide scalable and timely support around the clock, ensuring that individuals in underserved regions can access meaningful philosophical guidance. In this role, LLMs act as a bridge—partially compensating for the shortage of trained professionals while extending the impact of philosophical counseling.

With user-friendly interfaces and mature applications, interactive philosophical counseling services can be widely distributed, encouraging more people to learn about and experience philosophical counseling. LLMs could play a crucial role in raising public awareness of philosophical counseling by leveraging digital and social media platforms. Although they cannot replace the nuanced wisdom of professional counselors, their ability to captivate and engage users may inspire more individuals to pursue formal philosophical counseling.

Facilitating cultural adaptation in counseling. LLMs can help overcome cultural barriers and enhance the adaptability of philosophical counseling across different cultural contexts. They are particularly advantageous due to their extensive knowledge repositories (Petroni et al., 2019; Zhu et al., 2023). Even without specialized training or additional knowledge bases, LLMs can apply fundamental concepts drawn from various philosophical traditions to appropriate contexts. This capability is illustrated in a hypothetical dialogue between a GPT-4-based philosophical counselor and a counselee, where the LLM accurately incorporates key ideas—such as filial piety and righteousness from Confucianism—to assist in resolving the counselee's dilemmas (Ding et al., 2024a).

As depicted in Fig. 5, by constructing comprehensive knowledge bases of philosophical ideas from different cultures and employing RAG technology, LLMs can deliver tailored insights. For example, a complete database of Confucian and Daoist philosophies can help counselors better understand the cultural backgrounds and behavioral influences of their Chinese counselees.

Additionally, LLMs can be utilized for machine translation and even simultaneous interpretation, which facilitates effective communication between counselors and counselees speaking different native languages (Wang et al., 2024; Zhang et al., 2023). Furthermore, LLMs possess multilingual capabilities that enable them to simulate the roles of counselees from diverse cultural backgrounds. By applying prompt templates enriched with demographic information, cultural context, linguistic styles, life experiences, and client identity, LLMs can, to some extent, emulate counselees from varying cultural backgrounds and life histories. This feature is particularly valuable in multicultural societies, helping philosophical counselors to better engage with clients from diverse cultural and experiential backgrounds.

Through these mechanisms, LLMs can significantly enhance the capacity of philosophical counselors to navigate cultural complexities, thereby providing robust support that fosters a more inclusive and effective counseling environment.

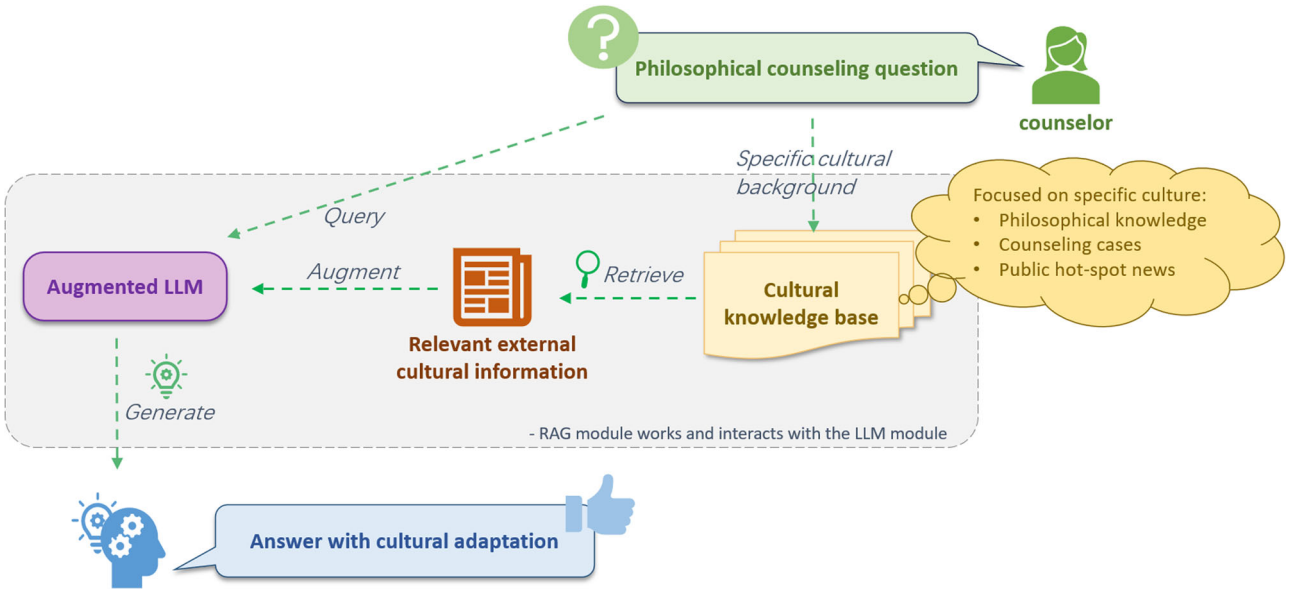

Fig. 5 Leveraging RAG for cultural adaptation in counseling contexts.

Fundamental challenges of deploying LLMs in philosophical counseling

While LLMs have shown considerable potential as valuable assistants in philosophical counseling by providing multi-faceted support, their application is not without significant limitations. Despite their remarkable capabilities, LLMs face inherent challenges when deployed in philosophical counseling—especially in meeting the comprehensive requirements of such services. These challenges stem not only from the complexities of human–computer interaction but also, more critically, from the fundamental disparities between the nature of philosophical counseling and the underlying mechanisms of contemporary AI technologies. Consequently, these limitations demand careful examination to ensure that LLMs serve as a complementary resource that augments, rather than inadvertently compromises, the integrity of the counseling process.

Trust. The integration of LLMs into philosophical counseling holds the potential to provide effective support for counselors. However, it also introduces critical challenges, particularly in terms of human–AI trust. Trust could be considered from multiple perspectives: counselor trust in machines, client trust in machines, and the public's confidence in AI.

A primary concern is the inherent lack of explainability in LLMs. Their operational mechanisms are complex and often insufficiently transparent, making it difficult for counselors to understand the reasoning behind generated outputs and thus trust their conclusions (Balasubramaniam et al., 2023; Zhao et al., 2024). Even when LLMs generate a response, they may not provide explanations that align with human logical processes (Turpin et al., 2023). This "black box" nature can lead to skepticism—particularly in philosophical counseling where issues often require nuanced interpretation—and might ultimately undermine the confidence of both counselors and counselees, thereby compromising the quality and impact of counseling services.

Another challenge concerns the instability of LLM performance, which primarily affects client trust. Research indicates that the same model can produce inconsistent results across different runs—with accuracy variations reaching up to 10% (Atil et al., 2024). Such instability may result in biased or even harmful outputs. For example, there have been reports of LLMs (e.g., Google's AI chatbot) generating inappropriate and potentially harmful suggestions (CBS News, 2024). In the context of philosophical counseling, where precision, ethical sensitivity, and trust are fundamental, these performance fluctuations present a significant barrier.

Despite these challenges, the advanced capabilities of LLMs have fostered public optimism. A recent study revealed that therapists were often unable to reliably distinguish between transcripts of human–AI interactions and those of human–human therapy sessions (Kuhail et al., 2024), and surveys indicate that 55% of respondents are optimistic about the potential of AI in mental health contexts (Aktan et al., 2022). However, such optimism also carries risks. Counselors who uncritically adopt LLM-generated suggestions may inadvertently undermine their own professional judgment, and when errors occur, this dependency can amplify the impact of mistakes—compromising both the professionalism and credibility of the counseling process.

Addressing these trust-related issues requires striking a careful balance between leveraging the benefits of LLMs and implementing robust oversight, ensuring that their usage remains transparent and ethically sound—while preserving the primacy of human judgment.

Privacy. Privacy is a fundamental aspect of both traditional counseling and the emerging use of LLMs. In conventional face-to-face counseling, professionals—whether psychologists or philosophers—are obligated to safeguard counselees' personal information. However, the shift to online counseling introduces new vulnerabilities. Although digital communication enhances accessibility, it also raises the risk of data interception and leakage (Kiriakaki et al., 2022). These risks become even more pronounced when employing LLMs in counseling settings, as user inputs may be used to train these models, thereby raising significant ethical and privacy concerns. While organizations such as OpenAI assert that models like ChatGPT do not collect or disclose personal information from interactions, numerous reports of privacy breaches suggest that these safeguards may not be as reliable as claimed (Yao et al., 2024).

Moreover, unlike psychological counseling—which focuses more on emotion—philosophical counseling often involves dialogues about values, life meaning, and ethical dilemmas. This

type of data, although not as directly sensitive as medical or mental health data, can include deeply personal information such as an individual's religious beliefs, moral stances, or personal philosophies. The exposure of such data could result in significant repercussions, including social prejudice and discrimination.

These distinctive privacy challenges underscore the importance of tailoring privacy safeguards to the specific context of philosophical counseling. By recognizing and mitigating privacy risks—through measures such as targeted policy regulation and data protection protocols—we can enhance the effectiveness of LLMs while ensuring that users' confidentiality is maintained, thereby fostering trust and security in digital counseling environments.

Philosophical understanding and empathy. Philosophical counseling requires a profound level of understanding that goes beyond mere language comprehension. Counselors must grasp counselees' issues within their unique personal and cultural contexts, maintain a deep understanding of philosophical concepts, and—as a critical component—demonstrate empathy, the capacity to understand and share another person's feelings (Cooper and McLeod, 2010). Empathy, as a cornerstone of counseling, enables deeper engagement with clients' concerns and fosters a meaningful exploration of ethical, existential, and personal dilemmas.

To fully appreciate the role of empathy in philosophical counseling, it is essential to engage with its philosophical underpinnings. David Hume (1739/2000) posited that empathy —or "sympathy," as he termed it—involves an affective resonance with others' emotions that is inherently human and extends well beyond mere cognitive simulation. Hume argued that sympathy is not merely an intellectual exercise but a deeply emotional connection grounded in our shared humanity. This emotional dimension underscores a critical limitation of AI: its lack of the embodied experience necessary for genuine empathy.

Max Scheler (1913/1970) further emphasizes empathy as a relational and intentional act. In his phenomenological framework, empathy (Einfühlung) is distinguished from mere emotional contagion; it requires engaging with the Other as a subject, recognizing their unique perspective and lived experience—a relational depth that LLMs, operating solely on syntactic and algorithmic principles, cannot replicate. Similarly, Michael Slote (2007) underscores empathy's moral significance, presenting it as central to ethical understanding and deliberation. Slote's ethic of care stresses that moral reasoning is not solely an abstract logical exercise but is deeply rooted in empathic engagement—an attribute that remains out of reach for LLMs due to their lack of genuine emotional capacity.

These philosophical insights collectively highlight the human-centered nature of empathy and its indispensable role in philosophical counseling. In this context, empathy is not merely a supportive skill but a core component of philosophical dialogue, particularly in addressing moral and ethical dilemmas. Although recent LLMs such as Grok 3, OpenAI o3, Gemini 2.0, Claude 3.5, and DeepSeek-R1 demonstrate impressive abilities in processing and generating coherent, contextually relevant language, their operation remains confined to syntactic processing rather than achieving true understanding (Pearl, 1988; Searle, 1980). In simulating understanding through statistical patterns in data, LLMs fundamentally lack the embodied human experiences, emotional resonance, and intentionality—qualities that are critical for genuine philosophical engagement (Boden, 1998; Bengio et al., 2003).

John Searle's Chinese Room Argument (1980) powerfully critiques the claim that AI can truly understand natural language. In this thought experiment, an individual inside a room follows syntactic rules to respond to Chinese characters without comprehending their meaning. This scenario mirrors how LLMs function: they manipulate symbols based on rules (syntax) without grasping meaning (semantics). In the realm of philosophical counseling, this distinction is pivotal. Although LLMs can generate responses that appear meaningful and contextually appropriate, their outputs are fundamentally probabilistic predictions rather than intentional expressions of meaning or emotion—as would be expected from a human counselor (Harnad, 1990; Ringle, 2019).

Empathy, in particular, underscores this limitation. It is not only about recognizing emotional cues but also about forming an authentic emotional connection with the client. Currently, LLMs lack consciousness and genuine emotional experience, meaning they cannot truly empathize with counselees. They can produce empathetic-like responses derived from patterns in their training data—useful for recognizing distress or offering comforting language—but these responses do not match the profound empathic engagement that human counselors can provide (Chaturvedi, 2024; Elliott et al., 2011). For example, Rubin et al. (2024) stress that while AI systems may mimic empathic expressions, they are unable to replicate the intentionality, concern, and trust-building necessary for meaningful therapeutic relationships.

The embodied and situated nature of human understanding further reinforces the limitations of LLMs. Scholars like Hubert Dreyfus (1972) and Sherry Turkle (2011) argue that human intelligence is deeply entwined with our physical embodiment and lived experience, which critically shapes our ability to navigate complex interpersonal and ethical domains. AI systems, lacking this embodied perspective, are therefore inherently incapable of addressing the existential and contextual dimensions of human concerns—a critical component of philosophical counseling. Shteynberg et al. (2024) similarly note that LLMs are currently incapable of genuine empathy, potentially adversely affecting users seeking emotional connection in their communications.

Nevertheless, the limitations of LLMs do not render them irrelevant in philosophical counseling. Instead, their capabilities should be viewed as assistive and complementary to human counselors, not as replacements. LLMs can organize and present philosophical insights, identify pertinent theories, or generate preliminary analyses of ethical dilemmas. However, the deeper tasks of understanding a client's personal context, engaging in moral reasoning, and providing meaningful guidance necessitate the uniquely human capacities of judgment, empathy, and awareness (Guo et al., 2024; Lee et al., 2021; Xu et al., 2024). In this sense, while LLMs may serve as effective tools for supporting philosophical inquiry, they cannot substitute the relational and empathetic depth that truly defines philosophical counseling.

Discussion

Philosophical counseling serves as a bridge between the general public and complex philosophical theories, making abstract concepts more accessible and applicable to everyday life. In this context, AI emerges as a powerful tool to strengthen this connection. The rapid evolution of AI technologies, particularly LLMs, has showcased their transformative potential across various domains—including philosophy. Integrating LLMs into philosophical counseling represents a particularly promising advancement with the potential to address practical limitations faced by traditional counseling methods.

Research indicates that LLMs have the capacity to fulfill several foundational requirements of philosophical counseling (Nutas,

2022). For instance, they can articulate the ideas of major European philosophical thinkers, detect logical fallacies, and elucidate the role of applied philosophy. However, significant limitations remain. In particular, LLMs inherently lack true understanding or empathy—as highlighted by Searle's Chinese Room Argument, which distinguishes between syntactic processing and genuine semantic comprehension. In a field where deep personal engagement and empathy are critical, this limitation necessitates careful consideration in their application.

The inherent shortcomings of current AI systems—especially regarding understanding and empathy—stem from the fundamental principles of their design. Achieving AI systems with logic and emotions truly aligned with human capacities still requires substantial technological advancements. Striking a balance between leveraging LLMs' strengths and acknowledging their limitations is essential for responsible implementation in philosophical counseling. Consequently, LLMs should be viewed as complementary tools, aiding human counselors in technical and procedural tasks rather than replacing the nuanced understanding and emotional connection that only human practitioners can provide.

From the viewpoint of their role as assistants, LLMs can enhance philosophical counseling by supporting both counselors and counselees before and after sessions. They can provide recommendations, evaluate outcomes, and facilitate cultural adaptation. The integration of LLM-related techniques—including prompt engineering, RAG, and fine-tuning—can evolve philosophical counseling into a more efficient and accessible service. When combined with other AI technologies such as multimodal models and recommendation systems, LLMs may further amplify counselors' capabilities in addressing diverse client needs. Additionally, features such as 24/7 availability and easy accessibility, which are not always practical for human counselors to maintain, make LLMs particularly valuable in broadening the reach and impact of philosophical counseling services.

While true philosophical understanding and genuine empathy remain challenging for AI, practical applications may not require these profound capacities. If AI systems can exhibit behaviors that resemble human understanding and empathy—akin to Searle's description of language simulation in the Chinese Room—they may still deliver significant practical value. Studies such as Kuhail et al. (2024) suggest that advanced LLMs can produce responses closely resembling those of human interactions, potentially meeting user expectations. Nevertheless, further evidence and improvements in model architecture and LLM-related techniques are needed to validate these claims and convincingly emulate human philosophical counselors.

In addition to the challenges of understanding and empathy, LLMs face significant obstacles in philosophical counseling, including issues of human–AI trust, privacy, and security. Building trust in AI requires transparency in system design, rigorous testing to ensure reliability, and clear communication of model limitations. Privacy concerns can be mitigated through robust data encryption, anonymization, and adherence to strict ethical guidelines for data use. Similarly, advanced security measures—such as secure architectures and real-time threat detection—are essential to safeguard sensitive client information. A comprehensive, interdisciplinary approach is indispensable to overcome these challenges and develop AI systems that are not only effective but also trustworthy and secure.

AI represents a transformative technology reshaping nearly every facet of human life, with its influence on philosophical counseling becoming increasingly apparent. The integration of LLMs into this domain presents a compelling and promising avenue of exploration, despite the complexities associated with applying computational methodologies to a traditionally human-centered practice. The development of an AI-driven philosophical counselor or practitioner could manifest as a specialized digital agent, distinct from existing general-purpose models such as ChatGPT or DeepSeek-R1. Ideally, such a model tailored specifically for philosophical practice would exhibit advanced proficiency in both philosophical reasoning and counseling methodologies, surpassing human practitioners in certain domains while ensuring critical human oversight in value-laden decision-making processes.

This paper proposes practical strategies for leveraging LLMs as assistants in philosophical counseling to make abstract philosophical concepts more accessible and to address the limitations of traditional counseling practices. By introducing and evaluating three key LLM-related techniques, the study highlights their respective advantages and disadvantages. Based on these techniques, detailed implementation plans for using LLM technologies are presented, offering actionable solutions to specific counseling challenges. Furthermore, this paper critically examines the limitations and challenges associated with LLM-based philosophical counseling, providing a balanced assessment of its potential and future directions. In doing so, it contributes to the interdisciplinary discussion on whether philosophical counseling will evolve into a robust field or be dismissed as pseudoscience (Kreimer and Primero, 2017).

However, this paper has its limitations. It is primarily theoretical and prospective in nature, lacking empirical validation or the implementation of a fully functioning LLM-based system. While we explore how LLM-related technologies can enhance philosophical counseling, developing a model that simultaneously addresses all requirements remains a formidable challenge. Moreover, we do not provide ready-made solutions for deeper philosophical and ethical obstacles related to AI's role in counseling. Future research should prioritize the practical implementation of LLM technologies and conduct empirical studies to assess their effectiveness in real-world counseling scenarios. Additionally, ongoing exploration is required to resolve the challenges identified in this paper and to refine the integration of LLMs into philosophical counseling.

We acknowledge that many of the issues raised by the reviewers reflect genuine challenges in balancing the interdisciplinary scope of our work with a firm philosophical stance. Our focus on predominantly Western scholarly sources has left non-Western contributions—such as those from the Korean philosophical counseling community and other informal practices—relatively underexplored. The embodied and contextual nature of human cognition remains critical to understanding philosophical practice (Shapiro, 2010; Wilson, 2002). Future research could aim to integrate diverse cultural perspectives and embodied approaches—drawing, for example, on the insights of Varela, Thompson, and Rosch (1991)—to address the multi-faceted dimensions of philosophical counseling, including its historical, cultural, and embodied underpinnings, in a more comprehensive manner.

Moreover, we wish to clarify our stance on the role of LLMs in philosophical counseling. We fully share the reviewers' concerns regarding the potential risks of overreliance on computational methods, which might lead to a homogenization of counseling practices or marginalize the unique contributions of human counselors. It is crucial to emphasize that our proposal does not aim to supplant human practitioners but to serve as a supplementary tool—enhancing accessibility and efficiency while preserving human critical judgment. Concerns raised by Searle's (1980) Chinese Room argument and further explored by Guo et al. (2024) underscore the challenges of replicating genuine understanding and empathy through AI. Future investigations, perhaps building on empirical studies like those by Kuhail et al.

(2024), should rigorously assess the impact of AI-assisted interventions on independent thought and subjectivity in counseling. We advocate for frameworks that ensure AI augmentation remains complementary, with robust mechanisms for human oversight, iterative feedback, and the preservation of the counselor's unique expertise.

Finally, we recognize that the integration of advanced computational tools into philosophical counseling carries profound societal and ethical implications. As we further develop technical solutions—such as fine-tuning, retrieval-augmented generation, and refined prompt engineering—future research must also examine whether these interventions might inadvertently shift philosophical practice toward a consumptive rather than reflective mode. It is imperative to foster interdisciplinary collaboration among philosophers, cognitive scientists, and AI researchers not only to enhance methodological precision but also to ensure that the core humanistic and transformative values of philosophical counseling are respected and preserved (Boden, 1998; Slote, 2007). By rigorously interrogating the balance between technological innovation and the preservation of critical, embodied, and culturally inflected thinking, we hope to expand a research agenda that addresses these complex challenges while remaining true to the transformative aims of philosophical inquiry.

Conclusion

LLMs exhibit significant promise for enhancing philosophical counseling by addressing key challenges—ranging from limited service accessibility to subjective evaluation criteria. Yet, integrating these advanced systems within an intrinsically human and value-laden domain mandates a cautious, balanced approach. Recognizing that current AI models lack genuine understanding and emotional empathy, LLMs should serve as sophisticated adjuncts, rather than replacements, for human counselors. By augmenting counseling processes and enabling personalized, scalable interventions, LLMs can help drive a transformative shift in philosophical practice. Ultimately, with robust oversight, stringent privacy safeguards, and continuous technical refinement, the responsible adoption of LLM-assisted methodologies will foster personal growth, enhance ethical discourse, and promote inclusive, culturally sensitive counseling in an increasingly digital world.

Data availability

The datasets generated and analysed during the current study are available from the corresponding authors upon reasonable request.

Received: 1 August 2024; Accepted: 24 February 2025;
Published online: 06 March 2025

References

Achenbach GB (1984) Philosophische praxis. Verlag für Philosophie Jürgen Dinter, Cologne

Achenbach GB (2010) Introduction to philosophical practice: lectures, essays, conversations, and essays presenting philosophical practice from 1981 to 2009. Verlag für Philosophie Jürgen Dinter, Cologne

Adhikary P, Srivastava A, Kumar S et al. (2024) Exploring the efficacy of large language models in summarizing mental health counseling sessions: benchmark study. JMIR Ment Health 11:e57306, https://mental.jmir.org/2024/1/e57306

Aktan ME, Turhan Z, Dolu I (2022) Attitudes and perspectives towards the preferences for artificial intelligence in psychotherapy. Comput Hum Behav 133:107273. https://doi.org/10.1016/j.chb.2022.107273

Alqahtani T, Badreldin HA, Alrashed M et al. (2023) The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research. Res Soc Adm Phar 19:1236–1242. https://doi.org/10.1016/j.sapharm.2023.05.016

Amir LB (2004) Three questionable assumptions of philosophical counseling. Int J Philos Pract 2(1):58–81. https://doi.org/10.5840/ijpp2004214

Amir L, Fatić A (2015) Practicing philosophy. Cambridge Scholars Publishing, Newcastle upon Tyne

Atil B, Chittams A, Fu L, Ture F, Xu L, Baldwin B (2024) LLM stability: a detailed analysis with some surprises. https://arxiv.org/abs/2408.04667

Badola BP (2015) Decline of philosophy in academic discourses: a problem with the modern trends in knowledge. Quest J UGC-HRDC Nainital 9(3):209–215. https://doi.org/10.5958/2249-0035.2015.00031.5

Balasubramaniam N, Kauppinen M, Rannisto A, Hiekkanen K, Kujala S (2023) Transparency and explainability of AI systems: from ethical guidelines to requirements. Inf Softw Technol 159:107197. https://doi.org/10.1016/j.infsof.2023.107197

Bengio Y, Ducharme R, Vincent P, Janvin C (2003) A neural probabilistic language model. J Mach Learn Res 3:1137–1155

Blyler AP, Seligman ME (2024a) AI assistance for coaches and therapists. J Posit Psychol 19(4):579–591. https://doi.org/10.1080/17439760.2023.2257642

Blyler AP, Seligman ME (2024b) Personal narrative and stream of consciousness: an AI approach. J Posit Psychol 19(4):592–598. https://doi.org/10.1080/17439760.2023.2257666

Brown T, Mann B, Rtder N et al. (2020) Language models are few-shot learners. In: Proceedings of the 34th International Conference on Neural Information Processing Systems (NIPS). Curran Associates Inc., pp. 1877–1901

Boden MA (1998) Creativity and artificial intelligence. Artif Intell 103(1-2):347–356. https://doi.org/10.1016/S0004-3702(98)00055-1

Bubeck S, Chandrasekaran V, Eldan R et al. (2023) Sparks of artificial general intelligence: early experiments with GPT-4. https://arxiv.org/abs/2303.12712

CBS News (2024) Google AI chatbot responds with a threatening message: "Human … Please die." Available online: https://www.cbsnews.com/news/google-ai-chatbot-threatening-message-human-please-die/. Accessed on 28 Dec 2024

Chaturvedi A (2024) Exploring empathy in artificial intelligence: synthesis and paths for future research. Inf Discov Deliv. https://doi.org/10.1108/IDD-03-2024-0048

Chang YP, Wang X, Wang JD et al. (2024) A survey on evaluation of large language models. ACM Trans Intel Syst Tec 15(3):1–45. https://doi.org/10.1145/3641289

Chen Z, Deng JW, Zhou JF, Wu JCZ, Qian TY, Huang ML (2024) Depression detection in clinical interviews with LLM-empowered structural element graph. In: Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers). ACL, pp. 8181–8194

Clay G, Ontiveros C (2023) Philosophers ought to develop, theorize about, and use philosophically relevant AI. Metaphilosophy 54(4):463–479. https://doi.org/10.1111/meta.12647

Cohen ED (2013) Theory and practice of logic-based therapy: integrating critical thinking and philosophy into psychotherapy. Cambridge Scholars Publishing, Newcastle upon Tyne

Cohen ED, Zinaich S (2013) Philosophy, counseling, and psychotherapy. Cambridge Scholars Publishing, Newcastle upon Tyne

Cohen ED, Piozzini B, Bapat C et al. (2024) A randomized, controlled, preliminary study to assess the efficacy of logic-based therapy in reducing anxiety and/or depression in family caregivers. J Ratio Emot Cogn Behav Ther 42:582–609. https://doi.org/10.1007/s10942-023-00532-z

Cooper M, McLeod J (2010) Pluralistic counselling and psychotherapy. Sage

Daniel MF, Auriac E (2011) Philosophy, critical thinking and philosophy for children. Educ Philos Theory 43(5):415–435. https://doi.org/10.1111/j.1469-5812.2008.00483.x

Demszky D, Yang DY, Yeager D et al. (2023) Using large language models in psychology. Nat Rev Psychol 2(11):688–701. https://doi.org/10.1038/s44159-023-00241-5

Devlin J, Chang MW, Lee K, Toutanova K (2019, June) Bert: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, volume 1 (long and short papers). ACL, pp. 4171–4186

Ding XJ, Fu SR, Jiao CC, Yu F (2024a) Chinese philosophical practice toward self-cultivation: integrating Confucian wisdom into philosophical counseling. Religions 15(1):69. https://doi.org/10.3390/rel15010069

Ding XJ, Harteloh P, Pan TQ, Yu F (2024b) The practical turn in philosophy: a revival of the ancient art of living through modern philosophical practice. Metaphilosophy 55(4-5):517–534. https://doi.org/10.1111/meta.12702

Ding XJ, Xie CF, Yu F (2024c) Beyond dissonance: the transformative power of thought analysis in philosophical practice. Humanit Soc Sci Commun 11(1):1617. https://doi.org/10.1057/s41599-024-04143-6

Ding XJ, Yu F (2024) Philosophical practice and its development in China: opportunities and challenges. Humanit Soc Sci Commun 11(1):494. https://doi.org/10.1057/s41599-024-02985-8

Dreyfus HL (1972) What computers can't do: a critique of artificial reason. Harper & Row, New York
Elliott R, Bohart AC, Watson JC, Greenberg LS (2011) Empathy. Psychotherapy 48(1):43–49. https://doi.org/10.1037/a0022187
Fatić A, Zagorac I (2016) The methodology of philosophical practice: eclecticism and/or integrativeness? Philosophia 44(4):1419–1438. https://doi.org/10.1007/s11406-016-9770-3
Fink MC, Robinson SA, Ertl B (2024) AI-based avatars are changing the way we learn and teach: benefits and challenges. Front Educ 9:1416307. https://doi.org/10.3389/feduc.2024.1416307
Fu GH, Zhao Q, Li JQ et al. (2023) Enhancing psychological counseling with large language model: a multifaceted decision-support system for non-professionals. https://arxiv.org/abs/2308.15192
Galitsky BA (2024) LLM-based personalized recommendations in health. Preprints. https://doi.org/10.20944/preprints202402.1709.v1
Gao YF, Xiong Y, Gao XY et al. (2024) Retrieval-augmented generation for large language models: a survey. https://arxiv.org/abs/2312.10997
Gindi S, Pilpel A (2015) Bridging the gap between philosophy and psychotherapy: an outline for the integration of philosophical counselling into therapeutic practice. J Humanit Ther 6(2):1–24
Grimes P, Uliana RL (1998) Philosophical midwifery: a new paradigm for understanding human problems with its validation. Hyparxis Press, Costa Mesa, CA
Guo Z, Lai A, Thygesen JH, Farrington J, Keen T, Li K (2024) Large language models for mental health applications: systematic review. JMIR Ment Health 11:e57400. https://doi.org/10.2196/57400
Hadot P (1995) Philosophy as a way of life: spiritual exercises from Socrates to Foucault. Davidson A (ed.). Translated by Chase M. Wiley-Blackwell, Oxford
Harnad S (1990) The symbol grounding problem. Phys D 42(1-3):335–346. https://doi.org/10.1016/0167-2789(90)90087-6
Heckmann G (1981) Socratic dialogue: reflections on technique. In: Gronbach K (ed.), Socratic philosophizing: selected writings. Vandenhoeck & Ruprecht, Göttingen, pp. 45–59
Hu YH (2024) From theoretical integration to complementary coexistence: an examination of the inclusiveness of Chinese civilization through the relationship between Confucianism and Daoism. J Hubei Univ Philos Soc Sci 51(6):83–91. https://doi.org/10.13793/j.cnki.42-1020/c.2024.06.010
Huang Y, Wu RP, He JT, Xiang YP (2024) Evaluating ChatGPT-4.0's data analytic proficiency in epidemiological studies: a comparative analysis with SAS, SPSS, and R. J Glob Health 14:04070. https://doi.org/10.7189/jogh.14.04070
Hume D (1739/2000) A treatise of human nature. Norton DF, Norton MJ (eds.). Oxford University Press, Oxford
Kandpal N, Deng H, Roberts A, Wallace E, Raffel C (2023) Large language models struggle to learn long-tail knowledge. In: Proceedings of 40th International Conference on Machine Learning (ICML). JMLR.org., pp. 15696–15707
Kinderman P (2005) A psychological model of mental disorder. Harv Rev Psychiat 13(4):206–217. https://doi.org/10.1080/10673220500243349
Kjell ON, Kjell K, Schwartz HA (2023) Beyond rating scales: with targeted evaluation, language models are poised for psychological assessment. Psychiat Res 333:115667. https://doi.org/10.1016/j.psychres.2023.115667
Kiriakaki S, Tzovanou M, Sotiriou A, Lampou M, Varsamidou E (2022) Online counseling: advantages, disadvantages and ethical issues. Homo Virtualis 5(1):32–59. https://doi.org/10.12681/homvir.30316
Knapp S, Tjeltveit AC (2005) A review and critical analysis of philosophical counseling. Prof Psychol-Res Pr 36(5):558–565. https://doi.org/10.1037/0735-7028.36.5.558
Kreimer R, Primero G (2017) The future of philosophical counseling: pseudoscience or interdisciplinary field? In: Amir L (ed.). New frontiers in philosophical practice. Cambridge Scholars Publishing, Newcastle upon Tyne, pp. 144–163
Kumar A, Raghunathan A, Jones R, Ma TY, Liang P (2022) Fine-tuning can distort pretrained features and underperform out-of-distribution. In: Proceedings of In International Conference on Learning Representations (ICLR). ICLR, pp. 1–42
Kuhail MA, Alturki N, Thomas J, Alkhalifa AK, Alshardan A (2024) Human-human vs human-AI therapy: an empirical study. Int J Hum-Comput Interact: 1–12. https://doi.org/10.1080/10447318.2024.2385001
Lahav R, Tillmanns MV (eds) (1995) Essays on philosophical counseling. University Press of America, Lanham
Lee EE, Torous J, De Choudhury M et al. (2021) Artificial intelligence for mental health care: clinical applications, barriers, facilitators, and artificial wisdom. Biol Psychiatry Cogn Neurosci Neuroimaging 6(9):856–864. https://doi.org/10.1016/j.bpsc.2021.02.001
Levkovich I, Elyoseph Z (2023) Suicide risk assessments through the eyes of ChatGPT-3.5 versus ChatGPT-4: vignette study. JMIR Ment Health 10:e51232. https://doi.org/10.2196/51232
Liu JY, Huang ZY, Xiao T et al. (2024) SocraticLM: exploring socratic personalized teaching with large language models. In: The Thirty-eighth Annual Conference on Neural Information Processing Systems (NIPS). Curran Associates Inc. https://openreview.net/forum?id=qkoZgJhxsA
Liu JM, Li DH, Cao H, Ren TH, Liao ZY, Wu JM (2023) Chatcounselor: a large language models for mental health support. https://arxiv.org/abs/2309.15461
Liu PF, Yuan WZ, Fu JL, Jiang ZB, Hayashi H, Neubig G (2023) Pre-train, prompt, and predict: a systematic survey of prompting methods in natural language processing. ACM Comput Surv 55(9):1–35. https://doi.org/10.1145/3560815
Louw D (2013) Defining philosophical counselling: an overview. S Afr J Philos 32(1):60–70. https://doi.org/10.1080/02580136.2013.810417
Ma XB, Gong YY, He PC, Zhao H, Duan N (2023) Query rewriting in retrieval-augmented large language models. In: Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP). ACL, pp. 5303–5315
Marinoff L (1999) Plato, not Prozac!: applying philosophy to everyday problems. HarperCollins, New York
Marinoff L (2002) Philosophical practice. Academic Press, San Diego, CA
Ministry of Education of the People's Republic of China (2023) Number of regular students for normal courses in HEIs by discipline. http://www.moe.gov.cn/jyb_sjzl/moe_560/2021/quanguo/202301/t20230103_1037969.html. Accessed on 12 May 2024
Na HB (2024) CBT-LLM: a Chinese large language model for cognitive behavioral therapy-based mental health question answering. In: Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING). ELRA and ICCL, pp. 2930–2940
National Center for Education Statistics Database (2023) Table 322.10. Bachelor's degrees conferred by postsecondary institutions, by field of study: selected academic years, 1970-71 through 2021-22. https://nces.ed.gov/programs/digest/d21/tables/dt21_322.10.asp. Accessed on 12 May 2024
Nelson L (1949) Socratic method and critical philosophy: selected essays. Dover Publications, New York, NY
O'Neill R, Wang H (2021) Combatting student alienation: community building in the academic philosophy café. J Humanit Ther 12(1):7–25. https://doi.org/10.33252/jht.2021.06.12.1.7
Nutas A (2022) The artificial philosophical counselor: on the possibility of automating philosophical counseling. Int J Philos Pr 8(1):124–136. https://doi.org/10.5840/ijpp2022819
Obradovich N, Khalsa SS, Khan WU et al. (2024) Opportunities and risks of large language models in psychiatry. NPP—Digit Psychiatry Neurosci 2(1):1–8. https://doi.org/10.1038/s44277-024-00010-z
Oliveira FGD, Belitski M, Kakabadse N, Theodorakopoulos N (2024) Unveiling the potential of digital human avatars in modern marketing strategies. Int Mark Rev. https://doi.org/10.1108/IMR-12-2023-0339
Park S, Kulkarni C (2023) Thinking assistants: LLM-based conversational assistants that help users think by asking rather than answering. https://arxiv.org/abs/2312.06024
Patel SC, Fan J (2023) Identification and description of emotions by current large language models. bioRxiv. https://doi.org/10.1101/2023.07.17.549421
Pearl J (1988) Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann, San Francisco, CA
Petroni F, Rocktäschel T, Lewis P et al. (2019) Language models as knowledge bases? In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). ACL, pp. 2463–2473
Rabani ST, Khanday AMUD, Khan QR, Hajam UA, Imran AS, Kastrati Z (2023) Detecting suicidality on social media: machine learning at rescue. Egypt Inform J 24(2):291–302. https://doi.org/10.1016/j.eij.2023.04.003
Radford A, Narasimhan K, Salimans T, Sutskever I (2018) Improving language understanding by generative pre-training. https://openai.com/research/language-unsupervised
Raile P (2024) The usefulness of ChatGPT for psychotherapists and patients. Humanit Soc Sci Commun 11:47. https://doi.org/10.1057/s41599-023-02567-0
Repetti R (2023) A sampling of philosophical counseling frameworks, tools, and practices. Interdiscip Res Couns Ethics Philos 3(8):136–195. https://doi.org/10.59209/ircep.v3i8.60
Ringle M (2019) Artificial intelligence and semantic theory. In: Simon TW, Scholes RJ (eds). Language, mind, and brain. Psychology Press, New York, pp. 45–63
Rubin M, Arnon H, Huppert JD, Perry A (2024) Considering the role of human empathy in AI-driven therapy. JMIR Ment Health 11:e56529. https://doi.org/10.2196/56529
Savage P (1997) Philosophical counselling. Nurs Ethics 4(1):39–48
Searle JR (1980) Minds, brains, and programs. Behav Brain Sci 3(3):417–424. https://doi.org/10.1017/S0140525X00005756
Schaaff K, Reinig C, Schlippe T (2023) Exploring ChatGPT's empathic abilities. In: 2023 11th International Conference on Affective Computing and Intelligent Interaction (ACII), pp. 1–8. https://doi.org/10.1109/ACII59096.2023.10388208
Scheler M (1913/1970) The nature of sympathy. Translated by P. Heath. Archon Books, Hamden, CT

Schuster SC (1999) Philosophy practice: an alternative to counseling and psychotherapy. Praeger, Westport & London
Schwitzgebel E, Schwitzgebel D, Strasser A (2023) Creating a large language model of a philosopher. Mind Lang 39(2):237–259. https://doi.org/10.1111/mila.12466
Shapiro L (2010) Embodied cognition. Routledge, New York, NY
Sharma A, Lin IW, Miner AS, Atkins DC, Althoff T (2023) Human–AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. Nat Mach Intell 5(1):46–57. https://doi.org/10.1038/s42256-022-00593-2
Shteynberg G, Halpern J, Sadovnik A et al. (2024) Does it matter if empathic AI has no empathy? Nat Mach Intell 6:496–497. https://doi.org/10.1038/s42256-024-00841-7
Sivil R, Clare J (2018) Towards a taxonomy of philosophical counselling. S Afr J Philos 37(2):131–142. https://doi.org/10.1080/02580136.2018.1432528
Slote M (2007) The ethics of care and empathy. Routledge, London
Smithson R, Zweber A (2024) Reviving the philosophical dialogue with large language models. Teach Philos. https://doi.org/10.5840/teachphil2024424196
Tashakkori A, Creswell JW (2007) The new era of mixed methods. J Mix Methods Res 1(1):3–7. https://doi.org/10.1177/2345678906293042
Touvron H, Lavril T, Izacard G et al. (2023a) Llama: open and efficient foundation language models. https://arxiv.org/abs/2302.13971
Touvron H, Martin L, Stone K et al. (2023b) Llama 2: open foundation and fine-tuned chat models. https://arxiv.org/abs/2307.09288
Turkle S (2011) Alone together: why we expect more from technology and less from each other. Basic Books, New York
Turpin M, Michael J, Perez E, Bowman SR (2023) Language models don't always say what they think: unfaithful explanations in chain-of-thought prompting. In: Proceedings of the 37th International Conference on Neural Information Processing Systems (NIPS). Curran Associates Inc., pp. 74952–74965
Vaswani A, Shazeer N, Parmar N et al. (2017) Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS). Curran Associates Inc., pp. 6000–6010
Varela FJ, Thompson E, Rosch E (1991) The embodied mind: cognitive science and human experience. MIT Press, Cambridge, MA
Wankhade M, Rao ACS, Kulkarni C (2022) A survey on sentiment analysis methods, applications, and challenges. Artif Intell Rev 55(7):5731–5780. https://doi.org/10.1007/s10462-022-10144-1
Wang M, Vu TT, Wang Y, Shareghi E, Haffari G (2024) Conversational simulMT: efficient simultaneous translation with large language models. https://arxiv.org/abs/2402.10552
Wei X, Li Q (2013) The Confucian value of harmony and its influence on Chinese social interaction. Cross-Cult Commun 9(1):60–66
Wei J, Wang XZ, Schuurmans D et al. (2022) Chain-of-thought prompting elicits reasoning in large language models. In: Proceedings of the 36th International Conference on Neural Information Processing Systems (NIPS). Curran Associates Inc., pp. 24824–24837
Wilson M (2002) Six views of embodied cognition. Psychonom Bullet Rev 9:625–636. https://doi.org/10.3758/BF03196322
Wu XD, Duan R, Ni JB (2024) Unveiling security, privacy, and ethical concerns of ChatGPT. J Inf Intell 2(2):102–115. https://doi.org/10.1016/j.jiixd.2023.10.007
Xu XH, Yao BS, Dong YZ et al. (2024) Mental-LLM: leveraging large language models for mental health prediction via online text data. Proc ACM Interact Mob Wearable Ubiquit Technol (IMWUT) 8(1):1–32
Yao Y, Duan JH, Xu KD et al. (2024) A survey on large language model (LLM) security and privacy: the good, the bad, and the ugly. High-Confid Comput 4(2):100211. https://doi.org/10.1016/j.hcc.2024.100211
Zhang B, Haddow B, Birch A (2023) Prompting large language model for machine translation: a case study. In: Proceedings of 40th International Conference on Machine Learning (ICML). JMLR.org., pp. 41092–41110
Zhang J, Yan B (2012) The enlightenment of Confucian traditional values on localized psychological counseling and therapy. Acad Exch 28(4):14–17
Zhao H, Chen H, Yang F et al. (2024) Explainability for large language models: a survey. ACM T Intel Syst Tec 15(2):1–38. https://doi.org/10.1145/3639372
Zhou SG, Zhang XY (2018) Selecting the best treatment: client-treatment match. Adv Psychol Sci 26(2):294–305
Zhu YT, Yuan HY, Wang ST et al. (2023) Large language models for information retrieval: a survey. https://arxiv.org/abs/2308.07107

Acknowledgements

This work was supported by the Natural Science Foundation of Hubei Province (Grant No. 2023AFB815), the MOE (Ministry of Education in China) Project of Humanities and Social Sciences (Grant No. 19YJC720006), and the National Social Science Foundation of China (Grant No. 20FZXB047).

Author contributions

BC and WZ: conceptualization, methodology, writing—original draft. LZ and XD: conceptualization, resources, supervision, writing—review and editing. All authors have given their final approval for the version to be published. All authors have agreed to take responsibility for all aspects of the work to ensure that questions about the accuracy or integrity of any part of the work are properly investigated and resolved.

Competing interests

Xiaojun Ding was a member of the Editorial Board of this journal at the time of acceptance for publication. The manuscript was assessed in line with the journal's standard editorial processes.

Ethical approval

It does not apply to this article as it does not contain any studies with human participants.

Informed consent

It does not apply to this article as it does not contain any studies with human participants.

Additional information

Correspondence and requests for materials should be addressed to Liang Zhao or Xiaojun Ding.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.