

Anthropomorphism and Ethical Readiness in Human-AI Interaction: Implications for Virtual Counselling

Assoc Prof Dr Ooi Pei Boon (Sunway University, Malaysia)

Dr Erica Woo Jhia Mae (Sunway University, Malaysia)

Ms Celine Choe Chai Tan (Sunway College, Malaysia)

Asst Prof Dr Ho Meng Chuan (UCSI University, Malaysia)

Dr Chee Kwan Foong (Sunway University, Malaysia)

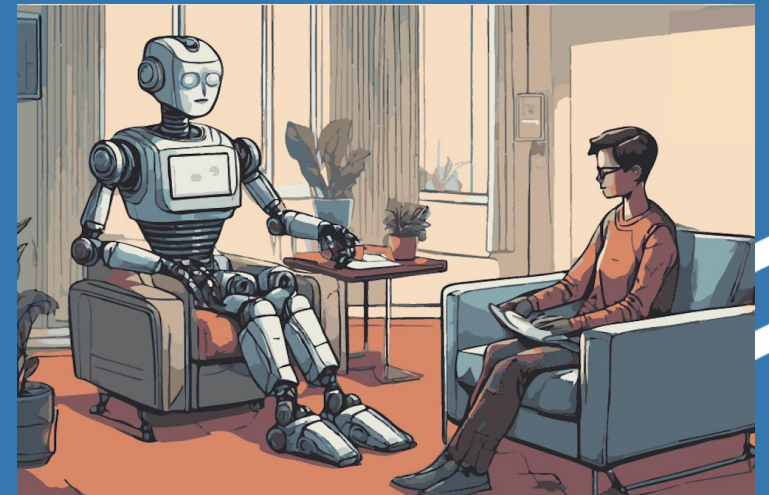
Assoc Prof Dr Azmawaty Binti Mohamad Nor (Universiti Malaya, Malaysia)

Dr Nasrudin Bin Subhi (Universiti Kebangsaan Malaysia, Malaysia)

Assoc Prof Dr Nor Mazlina Ghazali (Universiti Sarawak, Malaysia)

Background

1. Artificial intelligence in counselling is being increasingly recognized for its ability to provide consistent, data-driven support, offering new avenues for enhancing the efficiency and accessibility of mental health services (Olawade et al., 2024).
2. Studies have shown that AI can enhance service delivery by improving access to care and offering timely interventions (Eltahawy et al., 2024; Olawade et al., 2024).
3. **Anthropomorphism**, where users attribute human-like characteristics to AI, has been observed to impact user engagement and trust positively (Alabed et al., 2022).



Research gap:

1. There is a lack of comprehensive understanding of how anthropomorphism affects counsellors' **trust and acceptance** of AI technologies
2. The **ethical readiness** of counsellors to integrate AI into their practice has not been thoroughly examined

Significance:

1. Understanding the psychological impacts of human-AI interaction can inform the design of AI technologies that align with the needs and expectations of counsellors, thereby enhancing their effectiveness and acceptance
2. Ensuring ethical readiness through comprehensive training and robust guidelines can prevent potential ethical breaches, fostering a safer and more trustworthy environment for both counsellors and clients
3. By promoting sustainable and socially responsible practices in virtual counselling, Malaysia can improve the overall quality of mental health services, contributing to the broader social justice and well-being goals outlined by the United Nations Sustainable Development Goals

Artificial intelligence (AI) in Counselling: The past and Present

The early 2000s marked the beginning of AI's application in mental health, with a focus on automated diagnostic tools and therapeutic interventions

- This period saw a growing interest in using AI to support mental health professionals by automating routine tasks and providing preliminary assessments (Weizenbaum, 1966)

ELIZA – One of the first chatterbots approximating a Rogerian therapist (Bassett, 2019; Weizenbaum, 1966).

In recent years, advancements in natural language processing (NLP) and machine learning have significantly enhanced AI's capabilities in counselling

AI-driven chatbots, such as **Woebot**, were effective in delivering cognitive-behavioral therapy (CBT) to individuals with mild to moderate depression (Fitzpatrick et al., 2017).

Anthropomorphism: The attribution of human characteristics to non-human entities

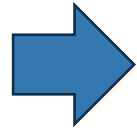
Three-factor theory of anthropomorphism (Epley et al., 2007)

Elicited agent knowledge



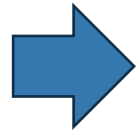
People use what they know about humans to understand non-human things

Effectance motivation



People want to make sense of how these non-human things behave

Sociality motivation



People have a natural desire for social connections and relationships



Factors associated with the effectiveness of AI in counselling

Trust

1. **Predictability, dependability, and faith** influences trust in human-automation interactions (Lee & See, 2004).
2. Trust influences user **acceptance and reliance** on AI technologies (Lee & See, 2004).
3. Maintaining trust requires ensuring the reliability and transparency of AI systems, as users are likely to disengage if the AI provides inconsistent or inaccurate responses

Ethical considerations

1. There is a need to address issues such as **confidentiality, data security, and the potential biases** in AI algorithms (Buchanan & Ess, 2009).
2. As generative AI becomes more integrated into counselling practices, it is essential to ensure **transparency, accountability, and protecting client privacy** (Ooi & Wilkinson, 2024).
3. The American Counselling Association (ACA) supported this by publishing guidelines and recommendations for practicing counsellors, clients, faculty, and students on using AI in their counselling practice (American Counselling Association, 2024).

Methodology

Hermeneutic phenomenology

- The methodical study of lived experiences and their inherent meanings (Heidegger, 1962; Husserl, 1970).
- As a research methodology, hermeneutic phenomenology reveals individuals' experiences within their unique, everyday contexts (Gadamer, 1997; Heidegger, 1962).
- It delves into 'lived experience,' an existing, pre-reflective awareness of life as it unfolds in every day, historically rooted contexts. This methodology involves describing and interpreting these lived experiences to uncover their meanings (van Manen, 1990).



Methodology

Research questions:

The study aims to explore the lived experiences of professional counsellors engaged in virtual counselling.

Through this exploration, the study seeks to answer the following key questions:

1. What are the lived experiences of counsellors engaged in virtual counselling?
2. How do these counsellors navigate the challenges and opportunities presented by virtual counselling?
3. How do counsellors perceive and manage the ethical implications of their practice in a virtual environment?

Procedure and sample:

This study employed a qualitative approach, conducting in-depth and focus group interviews with 14 Malaysian counsellors.



Results & Discussion

Perceptions of AI as a Human-Like Entity

- Counsellors often perceive AI technologies in human-like terms, a phenomenon known as **anthropomorphism**
- Counsellors frequently attributed human characteristics, such as **intentions, emotions, and personalities**, to AI tools used in their practice
- This perception seemed to make the interaction with AI more relatable and comfortable for the counsellors, allowing them to integrate these technologies into their workflows more naturally

"Sometimes, it feels like the AI understands my emotions just like a human would. It's almost comforting to think of it as a colleague rather than just a tool."

(Participant 2)

"I tend to talk to the AI like it's a person when I need ideas. It makes the interaction feel more natural and less mechanical."

(Participant 4)

However, the tendency to humanize AI may lead to unrealistic expectations about its emotional intelligence and the depth of its understanding

Results & Discussion

Trust and Reliability

- Counsellors rely heavily on AI systems to provide dependable support in their therapeutic practices, and when these systems perform consistently, they enhance trust and integration into daily workflows
- The study found that counsellors' trust in AI was closely linked to their experiences with its **reliability and accuracy**
- However, **occasional inaccuracies**, whether due to algorithmic errors, data biases, or contextual misunderstandings, can lead to skepticism and caution
- There is a risk that **over-dependence** on AI could diminish the role of human intuition and judgment in therapy

"I trust the AI to provide consistent responses, which is something even human counsellors can struggle with. This reliability is crucial in building my trust in using it."

(Participant 11)

"While AI can be very helpful, I am always cautious. There have been instances where the AI's suggestions were not quite on point, and that makes me hesitant to rely on it completely."

(Participant 6)

Results & Discussion

Impact on Therapeutic Relationships

- The integration of AI in counselling raised questions about its impact on therapeutic relationships
- Participants were divided, with some seeing AI as an enhancement to the counsellor-client relationship, while others feared it might detract from the human connection that is central to effective therapy

"AI can assist in tracking progress and providing reminders, which helps me focus more on the client during sessions. It enhances the therapeutic relationship by taking care of administrative tasks."
(Participant 8)

"I'm worried that relying too much on AI could make the counselling process feel less personal. The human touch is essential in building trust and empathy with clients."
(Participant 1)

Ethical Considerations and Implications of AI Integration in Counselling in the Malaysian Context

Impact on client autonomy

- In traditional therapy, the counsellor's role is to guide clients toward self-discovery and empowerment, allowing them to make **informed decisions** about their lives
- However, when AI is introduced into the therapeutic process, there is a risk that clients may become overly reliant on AI-generated suggestions, potentially undermining their autonomy
- This is particularly concerning in Malaysia, where cultural norms may already incline individuals to defer to authority figures, including AI systems perceived as knowledgeable or authoritative

Implications:

Counsellors must maintain a balance between leveraging AI's capabilities and encouraging clients to make their own decisions



Ethical Considerations and Implications of AI Integration in Counselling in the Malaysian Context

Confidentiality

- The use of AI in counselling raises significant concerns about data privacy and the security of sensitive client information
- In Malaysia, where **data protection** laws are still developing, the potential for data breaches or misuse of client information is a serious concern

Implications:

Counsellors must be aware of the risks associated with using AI technologies and ensure that robust data protection protocols are in place

Ethical Considerations and Implications of AI Integration in Counselling in the Malaysian Context

Bias in AI systems

- AI technologies are only as good as the data they are trained on, and if this data reflects existing biases, the AI system may perpetuate these biases in its recommendations or interactions
- This is particularly concerning in the Malaysian context, where cultural, racial, and religious diversity must be carefully navigated in counselling
- AI-generated suggestions that are **culturally insensitive or irrelevant** can potentially harm the clients

Implications:

Counsellors must be vigilant in assessing the cultural competence of AI tools and advocating for the development of AI systems that are trained on diverse and representative datasets

Ethical Considerations and Implications of AI Integration in Counselling in the Malaysian Context

Loss of "human element" in therapy

- Counselling is inherently a human-centered practice, where the **therapeutic relationship** between counsellor and client plays a crucial role in the healing process
- The introduction of AI into this dynamic could disrupt this relationship, particularly if clients perceive the AI as more reliable or trustworthy than the human counsellor

As AI technologies continue to evolve, counsellors must remain vigilant in **safeguarding client autonomy, ensuring confidentiality and privacy, addressing bias and fairness, preserving the human element in therapy, and maintaining professional competence and accountability.**

Conclusion

1. AI technologies can enhance counselling services by providing consistent, accessible, and personalized support, but key challenges related to trust, engagement, and ethical practice must be addressed
2. Human-like features in AI interfaces can enhance engagement and therapeutic alliance, but balancing anthropomorphism to avoid unrealistic expectations and potential ethical pitfalls is crucial
3. In Malaysia, where face-to-face interactions and personal relationships are highly valued, counsellors must be mindful of maintaining empathy, compassion, and personal engagement, even as they incorporate AI tools into their practice
4. Counsellors must be adequately trained in the use of AI technologies and understand their ethical implications
5. In Malaysia, where legal and regulatory frameworks for AI use in counselling are still in their infancy, the issue of accountability when using AI in therapy must be addressed carefully

Acknowledgment

The authors wish to thank the participants, Board of Counsellor Malaysia and Sunway University Malaysia for the support and endorsement to conduct this study
Ethical approval code: SUREC 2024/051

References

- Alabed, A., Javornik, A., & Gregory-Smith, D. (2022). AI anthropomorphism and its effect on users' self-congruence and self-AI integration: A theoretical framework and research agenda. *Technological Forecasting and Social Change*, 182, 121786. <https://doi.org/10.1016/j.techfore.2022.121786>
- American Counselling Association. (2024). *Recommendations for practicing counselors: Artificial intelligence in counseling*. . <https://www.counseling.org/resources/research-reports/artificial-intelligence-counseling/recommendations-for-practicing-counselors>
- Bassett, C. (2019). The computational therapeutic: exploring Weizenbaum's ELIZA as a history of the present. *AI & SOCIETY*, 34(4), 803-812. <https://doi.org/10.1007/s00146-018-0825-9>
- Buchanan, E. A., & Ess, C. (2009). *Internet Research Ethics: The Field and Its Critical Issue* (K. H. a. H. Tavani, Ed.). Hoboken, NJ: John Wiley & Sons.
- Eltahawy, L., Essig, T., Myszkowski, N., & Trub, L. (2024). Can robots do therapy?: Examining the efficacy of a CBT bot in comparison with other behavioral intervention technologies in alleviating mental health symptoms. *Computers in Human Behavior: Artificial Humans*, 2(1), 100035. <https://doi.org/10.1016/j.chbah.2023.100035>
- Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: a three-factor theory of anthropomorphism. *Psychological review*, 114(4), 864. <https://doi.org/10.1037/0033-295X.114.4.864>
- Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR mental health*, 4(2), e7785. <https://doi.org/10.2196/mental.7785>
- Gadamer, H. G. (1997). *Truth and method* (2nd rev. ed.) J. Weinsheimer and D. G. Marshall, trans. rev. New York: Continuum (Original work published 1960).
- Heidegger, M. (1962). *Being and time*. New York: Harper & Row (Original work published 1927).
- Husserl, E. (1970) (trans J.N. Findlay). *Logical investigations* (Vol. I & II). New York: Humanities Press (Original work published 1900)
- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human factors*, 46(1), 50-80. <https://doi.org/10.1518/hfes.46.1.50.3039>
- Olawade, D. B., Wada, O. Z., Odetayo, A., David-Olawade, A. C., Asaolu, F., & Eberhardt, J. (2024). Enhancing mental health with Artificial Intelligence: Current trends and future prospects. *Journal of Medicine, Surgery, and Public Health*, 3, 100099. <https://doi.org/10.1016/j.glmedi.2024.100099>
- Ooi, P. B., & Wilkinson, G. (2024). Enhancing ethical codes with artificial intelligence governance – a growing necessity for the adoption of generative AI in counselling. *British Journal of Guidance & Counselling*, 1-15. <https://doi.org/10.1080/03069885.2024.2373180>
- van Manen, M. (1990). *Researching lived experience: Human science for an action sensitive pedagogy*. New York: State University of New York Press.
- Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36-45.

Anthropomorphism and Ethical Readiness in Human-AI Interaction: Implications for Virtual Counselling

Questions & Answers

Thank you

Assoc Prof Dr **Ooi** Pei Boon
peiboono@sunway.edu.my

The authors wish to thank the participants, Board of Counsellor Malaysia and Sunway University Malaysia for the support and endorsement to conduct this study
Ethical approval code: SUREC 2024/051