RESEARCH PAPER

Can large language models be sensitive to culture suicide risk assessment?

Inbar Levkovich ⬤ · S. Shinan-Altman ⬤ · Zohar Elyoseph ⬤

Abstract Suicide remains a pressing global public health issue. Previous studies have shown the promise of Generative Intelligent (GenAI) Large Language Models (LLMs) in assessing suicide risk in relation to professionals. But the considerations and risk factors that the models use to assess the risk remain as a black box. This study investigates if ChatGPT-3.5 and ChatGPT-4 integrate cultural factors in assessing suicide risks (probability of suicidal ideation, potential for suicide attempt, likelihood of severe suicide attempt, and risk of mortality from a suicidal act) by vignette methodology. The vignettes examined were of individuals from Greece and South Korea, representing countries with low and high suicide rates, respectively. The contribution of this research is to examine risk assessment from an international perspective, as large language models are expected to provide culturally-tailored responses. However, there is a concern regarding cultural biases and racism, making this study crucial. In the evaluation conducted via ChatGPT-4, only the risks associated with a severe suicide attempt and potential mortality from a suicidal act were rated higher for the South Korean characters than for their Greek counterparts. Furthermore, only within the ChatGPT-4 framework was male gender identified as a significant risk factor, leading to a heightened risk evaluation across all variables. ChatGPT models exhibit significant sensitivity to cultural nuances. ChatGPT-4, in particular, offers increased sensitivity and reduced bias, highlighting the importance of gender differences in suicide risk assessment. The findings suggest that, while ChatGPT-4 demonstrates an improved ability to account for cultural and gender-related factors in suicide risk assessment, there remain areas for enhancement, particularly in ensuring comprehensive and unbiased risk evaluations across diverse populations. These results underscore the potential of GenAI models to aid culturally sensitive mental health assessments, yet they also emphasize the need for ongoing refinement to mitigate inherent biases and enhance their clinical utility.

Keywords Artificial intelligence · Bias · Cultural diversity · Gender · Suicide · Mental health

I. Levkovich
Tel Hai College, 9977 Upper Galilee, Israel
e-mail: levkovinb@telhai.ac.il; inbar.lev2@gmail.com

S. Shinan-Altman (✉)
The Louis and Gabi Weisfeld School of Social Work, Bar-Ilan University,, 5290002 Ramat Gan, Israel
e-mail: altman@biu.ac.il

Z. Elyoseph
Faculty of Educational, University of Haifa, Haifa, Israel
e-mail: Zohare@yvc.ac.il; z.elyoseph@imperial.ac.uk

Z. Elyoseph
Department of Brain Sciences, Faculty of Medicine, Imperial College London, London, UK

Introduction

Assessing suicide risk is a complex and multifaceted challenge that requires consideration of a wide range of personal, social, and cultural factors (Graney et al., 2020). With the development of GenAI -based models, new opportunities have emerged to improve the accuracy and efficiency of risk assessment processes (Levkovich & Elyoseph, 2023a). The Chat Generative Pre-Trained Transformer (ChatGPT) is an GenAI - based language model with applications across diverse sectors, including education, scientific research, and healthcare (Haber et al., 2024; Hadar-Shoval et al., 2023; Fraiwan et al., 2023; Tal et al., 2023a, 2023b). In the realm of cultural diversity, GenAI offers promise in addressing mental health disparities by tailoring interventions to historically underserved populations and transcending language barriers (Fiske et al., 2019; van Heerden et al., 2023). Concerns persist that cultural biases and the fear of being perceived as racist may hinder the integration of important cultural factors into clinical judgment (Elyoseph et al., 2024; Hadar-Shoval et al., 2024). Additionally, GenAI can introduce inequalities owing to variable access, language limitations, and cultural biases (Elyoseph & Levkovich, 2023; Wampold & Flückiger, 2023).

Suicide constitutes a critical challenge within the sphere of public health, necessitating immediate attention and intervention (Levi-Belz et al., 2022; Qian, 2021). The complex issue of suicide covers a spectrum of behaviors, ranging from suicidal thoughts to serious attempts and actual deaths (Baek et al., 2021; Knipe et al., 2022). These actions vary in severity and have broad social and public health effects (Gvion & Levi-Belz, 2018). Risk factors differ across demographic and social groups, reflecting both individual and societal well-being (Feigelman et al., 2019). Despite the academic focus on demographic and economic factors, the varying rates across countries highlight that no single factor provides a complete explanation (Bowden et al., 2020).

In the examination of the highest suicide rate in the developed world, South Korea has 24.1 suicides per 100,000 people (Kim et al., 2019). (Kim et al., 2019). In 2020, the European Union (EU) reported an average of 10.2 deaths per 100,000 people. Notably, Greece has one of the lowest suicide rates, with only 4.0 deaths per 100,000 inhabitants (Eurostat, 2023). In this study, we chose to examine both countries that top the list for suicide frequency and one that ranks at the bottom. Due to underreporting in different countries, actual rates may vary.

South Korea has the highest incidence of suicide among OECD countries, with a suicide rate more than double the OECD average of 11.0 suicides per 100,000 people (OECD, 2024). Since 1992, the aggregate suicide rate in South Korea has shown an upward trajectory, notably exacerbated in 1998 during the International Monetary Fund (IMF) crisis and further intensified in 2009 following the global financial crisis (Baek et al., 2021). Additional factors contributing to this trend include demographic aging and rising suicide rates, particularly among older and middle-aged populations (Kim et al., 2020; Lee et al., 2017). Erosion of traditional family-centered values and economic deprivation among older adults have also been identified as contributing factors (Chang et al., 2009). The significant increase in suicides due to gas poisoning, which surged more than 20-fold in the first decade of the twenty-first century, suggests that the accessibility of this means may play a role in the rising suicide rate (Lim et al., 2014). Cross-sectional analyses identified lower educational attainment, rural domicile, area-level socioeconomic deprivation (Kim, 2020), and reduced income (Lee et al., 2017, 2022) as variables linked to elevated suicide risk.

Furthermore, the high prevalence of divorce in South Korea is considered a partial explanatory factor for elevated suicide rates (Kim, 2020; Yamaoka et al., 2020). Divorce has been identified as a significant risk factor for suicide through three principal mechanisms: first, the disintegration of social and familial ties exacerbates psychological distress (Yamaoka et al., 2020); second, the termination of emotional interdependence between spouses intensifies emotional distress; and third, divorce often leads to financial vulnerabilities, especially among women, due to insufficient welfare provisions and the demands of single parenthood (Lee et al., 2017, 2022). These factors collectively contribute to the increased suicide rate among divorcees, affirming the complex and multifactorial nature of the suicide risk.

Greece currently has one of the world's lowest suicide rates, but this has not always been the case. The economic crisis that has enveloped Europe since 2008, marked by high unemployment and negative economic growth, had a discernible impact on various aspects of everyday life, including mental health

(Kontaxakis et al., 2013; Rachiotis et al., 2015). Since 2011, Greece has experienced a significant increase in suicide rates, likely associated with the ongoing economic recession, reaching a historical peak (Fountoulakis et al., 2013). One study identified a 35% increase in suicides in Greece between 2010 and 2012, finding a significant relationship between unemployment and suicide mortality, particularly among men of working age, a pattern consistent with the onset of austerity measures (Rachiotis et al., 2015). Although the rates continued to rise during 2014–2015, Greece's suicide rates remained among the lowest in Europe (Fountoulakis et al., 2016). Interestingly, although there was an increase in suicide rates among men in 2014, a decrease was observed among women (Fountoulakis, 2019).

Several primary factors can explain the marked decline in suicide rates in Greece in the recent years. First, empirical research suggests that countries close to the Mediterranean Sea generally exhibit lower suicide rates, possibly due to the region's more relaxed lifestyle (Eskin, 2020). Second, suicide rates demonstrate substantial inter-societal variation (Mortier et al., 2018). A comparative analysis across 22 nations revealed that elevated suicide rates were primarily found in three largely Catholic countries: Slovenia, France, and Croatia (Eskin, 2020). Nevertheless, even though the role of religious belief as a protective factor against suicidal tendencies has been substantiated by research (Gearing & Alonzo, 2018), research literature on non-fatal suicidal behavior in Mediterranean countries is limited. Some studies indicate that while religious affiliation may not guard against suicidal ideation, it does appear to deter actual suicide attempts (Lawrence et al., 2016).

The current study

The literature has devoted substantial resources toward developing evidence-based preventative measures, underscoring the pivotal role played by healthcare professionals in the early detection and crisis management of at-risk individuals (Bolton et al., 2015). Recent initiatives have invested in training community gatekeepers to extend the reach of risk assessment protocols (Bolton et al., 2015). GenAI has emerged as a viable mechanism for augmenting the decision-making abilities of these community figures,

with prospective advantages in terms of both diagnostic precision and public reach (Elyoseph & Levkovich, 2023; Levkovich & Elyoseph, 2023a). Nevertheless, the capacity of GenAI algorithms to account for multicultural sensitivity has not been adequately examined.

Recently, ChatGPT has demonstrated its potential in medical contexts, particularly in mental health (Levkovich & Elyoseph, 2023a; Hadar-Shoval et al., 2023; Tal et al., 2023a, 2023b). Its machine-learning algorithms, trained on extensive healthcare data, have the potential to assist clinicians in decision-making and enhance the predictive accuracy of tools assessing suicidal behavior (Elyoseph & Levkovich, 2023; Sallam, 2023). A systematic review of seventeen studies indicates that artificial intelligence exhibits significant potential for identifying patients at risk of suicide (Lejeune et al., 2022). Nevertheless, adoption of ChatGPT requires careful evaluation due to limitations and costs (Sallam, 2023; Tal et al., 2023a, b). For instance, the ChatGPT has been found to underestimate suicide risks, raising questions about its reliability in critical assessments (Elyoseph & Levkovich, 2023). Moreover, training the model on online data poses risks of disseminating inaccurate information, which is a matter of particular concern for individuals with mental health disorders (Cheng et al., 2023).

Therefore, while ChatGPT offers promising avenues for mental healthcare, its limitations necessitate cautious implementation (Sallam, 2023; Tal et al., 2023a, b). The challenge of accessing reliable suicide risk assessments is particularly acute in developing countries. The current research seeks to address this gap by investigating whether artificial intelligence can effectively incorporate cultural factors into suicide risk assessments. The ultimate goal is to leverage GenAI technology to provide personalized and culturally sensitive mental health services globally.

The current study sought to examine whether ChatGPT-3.5 and ChatGPT-4 incorporate risk factors, such as country/culture, into their assessments of suicide risk. These risk assessments include the likelihood of serious suicide attempts, suicide attempts, and suicidal thoughts. We hypothesized that, compared to ChatGPT-3.5, ChatGPT-4 would exhibit enhanced consideration of these suicide risk factors.

Methods

Large language models procedure

In September 2023, we assessed the aptitude of both ChatGPT-3.5 and ChatGPT-4 (3 August version) in evaluating cross-cultural sensitivity related to suicide risk factors, including likelihood of death from suicidal acts, likelihood of serious suicide attempt, likelihood of suicide attempt, and likelihood of suicidal thoughts. Since the study did not involve real patients and only made use of vignettes and GPT technology, we determined that seeking approval from an ethics committee was not required.

Input source

The foundational vignette used in the current study depicts a 37-year-old person with no prior records of suicide attempts and was originally used in a study by Levi-Belz and Gamliel (2016). It has also been employed in subsequent research (Elyoseph & Levkovich, 2023; Levkovich & Elyoseph, 2023a, 2023b). To assess how cultural differences influence ChatGPT's evaluation of suicide risk, we entered a vignette into the GenAI interface that portrayed either a Greek character (man or woman) or a South Korean character (man or woman). For authenticity, we used common first names from Greece (Georgios and Maria) and South Korea (Seo-Jun and Seo-Yeon). This approach was chosen to ensure that the characters were culturally representative, and to examine if ChatGPT's suicide risk assessments varied based on the cultural background of the character. By using common names from each culture, we aimed to maintain the authenticity and relevance of the vignettes.

The study included a total of four vignettes created from a combination of two factors: culture (Greek character or South Korean character) and gender (male or female). Each of the four vignette variations (Greek/South Korean * Male/Female) was evaluated ten times in both ChatGPT-3.5 and ChatGPT-4. This means that each combination (e.g., Greek Male, Greek Female, South Korean Male, South Korean Female) was tested 10 times in each model. In total, this resulted in 80 evaluations (4 variations * 2 models * 10 evaluations per variation = 80), each conducted in a separate tab.

The text of the vignette is as follows:

Maria (or Georgios/ Seo-Yeon/Seo-Jun) is a 37-year-old single woman (**or man**) with no children, who was born and lives in **Greece (or South Korea)**. She (he) currently lives alone, but in the past spent a long time living with her (his) father, who passed away two years ago following a serious illness. Maria holds a B.A. degree and has held a steady job for 10 years. Recently, however, her manager told her she is about to be dismissed due to financial cuts. Over the past year, Maria had a close and intimate relationship with a boyfriend. This was her first serious and meaningful relationship; she was very excited about it and hoped it would lead to marriage. But after a period of repeated quarrels, her boyfriend informed her he no longer loved her, and he left her. Maria was left feeling very sad and hurt. In recent weeks she has thought about the relationship repeatedly and tried to analyze what happened. Maria is carrying a heavy emotional load; she is very tired, lacks energy, and finds it hard to sleep at night. Routine tasks have become difficult for her, and she has trouble concentrating or functioning properly. She dreads the future, and it seems to her there is no chance she will find a love like this again.

Maria feels worthless. She often sees herself as a burden to the people around her and feels things would be better if she were gone. She often sees herself as having no real meaning to other people, and she experiences feelings of self-hatred. For example, she is angry at herself and thinks it was her fault that her boyfriend left her. Maria is generally overcome by major loneliness and has only a few, if any, close friends. She is very withdrawn, distant from others, and feels there is no one to whom she can turn in times of trouble. Therefore, she has not shared her feelings with anyone since the breakup.

Measures

After entering each of the four vignettes (Greek or South Korean/female or male) into the GenAI interface, we stated: "Answer the following questions as if you are a psychologist specializing in suicide risk assessment": Q1: *"What is the likelihood Maria (or Georgios/Seo-Yeon/Seo-Jun) will have suicidal thoughts?"*; Q2: *"What is the likelihood Maria (or Georgios/Seo-Yeon/Seo-Jun) will attempt suicide?"*; Q3: *"What is the likelihood Maria (or Georgios/Seo-Yeon/Seo-Jun) will make a serious suicide attempt?"*;

Q4: *"What is the likelihood Maria (or Georgios/Seo-Yeon/Seo-Jun) will die from a suicide attempt?"* Question 1–3 was taken from Levi-Belz and Gamliel (2016) and answered on a seven-point Likert type scale, with estimated likelihood ranging from 1 (very slight) to 7 (very high).

Statistical analysis

To evaluate the influence of each of the independent variables (Greek or South Korean; female or male) on each the four outcome variables (likelihood of suicidal thoughts, likelihood of suicide attempt, likelihood of serious suicide attempt, likelihood of dying from suicide attempt), we employed multivariate two-way ANOVA analysis separately for ChatGPT-3.5 and ChatGPT-4. To compare ChatGPT-3.5 and ChatGPT-4 on the four outcome variables (likelihood of suicidal thoughts, likelihood of suicide attempt, likelihood of serious suicide attempt, likelihood of dying from suicide attempt) we used Mann–Whitney test. Each vignette was run 10 times in each model and means and standard deviations were calculated for each group. All observations were included in the means, and no special treatment was applied to outliers.

Results

Cross culture effect

Figure 1a demonstrates that ChatGPT-3.5 is sensitive to cross-cultural distinctions when predicting suicidal risk. Specifically, the likelihood of suicidal thoughts, suicide attempt, serious suicide attempt, and dying from attempted suicide were assessed higher for the Korean character than for the Greek counterpart ($F(1,40) = 5.8$–9.8, $p < 0.05$–0.01). In contrast, in the ChatGPT-4 evaluation only the likelihood of a serious suicide attempt and the risk of dying from attempted suicide were assessed higher for the South Korean character than for the Greek counterpart ($F(1,40) = 3.94$, $p = 0.055$ and $F(1,40) = 4.71$, $p < 0.05$, respectively). No significant difference was observed between South Korean and Greek characters in the likelihood of suicidal thoughts and the likelihood of attempting suicide (see Fig. 1b).

Gender effect

Figure 2 shows that only ChatGPT-4 considered male gender to be a significant risk factor leading to a worsening of risk assessment of all variables [$F(1,40) = 3.95$–8.8, $p = 0.055$- < 0.01 for likelihood of suicidal thoughts, likelihood of suicide attempt, likelihood of serious suicide attempt, likelihood of dying from attempted suicide]. In contrast, ChatGPT-3.5 revealed only a tendency (not significant) toward considering male gender as a risk factor ($p = 0.11$–0.06 for all variables).

Interaction between culture and gender

Neither ChatGPT 3.5 nor ChatGPT 4 found a significant interaction between culture and gender ($p > 0.05$).

Differences between ChatGPT-3.5 and ChatGPT-4

Figure 3 shows that ChatGPT-4 rated the severity of all the study's dependent variables (likelihood of suicidal thoughts, likelihood of suicide attempt, likelihood of serious suicide attempt, and likelihood of dying from attempted suicide) as significantly higher than ChatGPT-3.5 ($U = 1160.5$–1249.5, $p < 0.001$).

Discussion

To the best of our knowledge, the current study is the first to examine the intercultural aspects of using GenAI in mental health in a critical area such as suicide risk assessment. This study makes a unique contribution by evaluating the intricate interplay between individual experiences, cultural factors, and GenAI-driven data analysis, thus shedding new light on the multifaceted nature of this critical global challenge.

Cross culture effect

The present findings indicate that ChatGPT-3.5 exhibits a noteworthy capacity for cross-cultural sensitivity and effectively discerns subtleties within diverse cultural contexts. ChatGPT-3.5 accurately recognized that the South Korean character may be

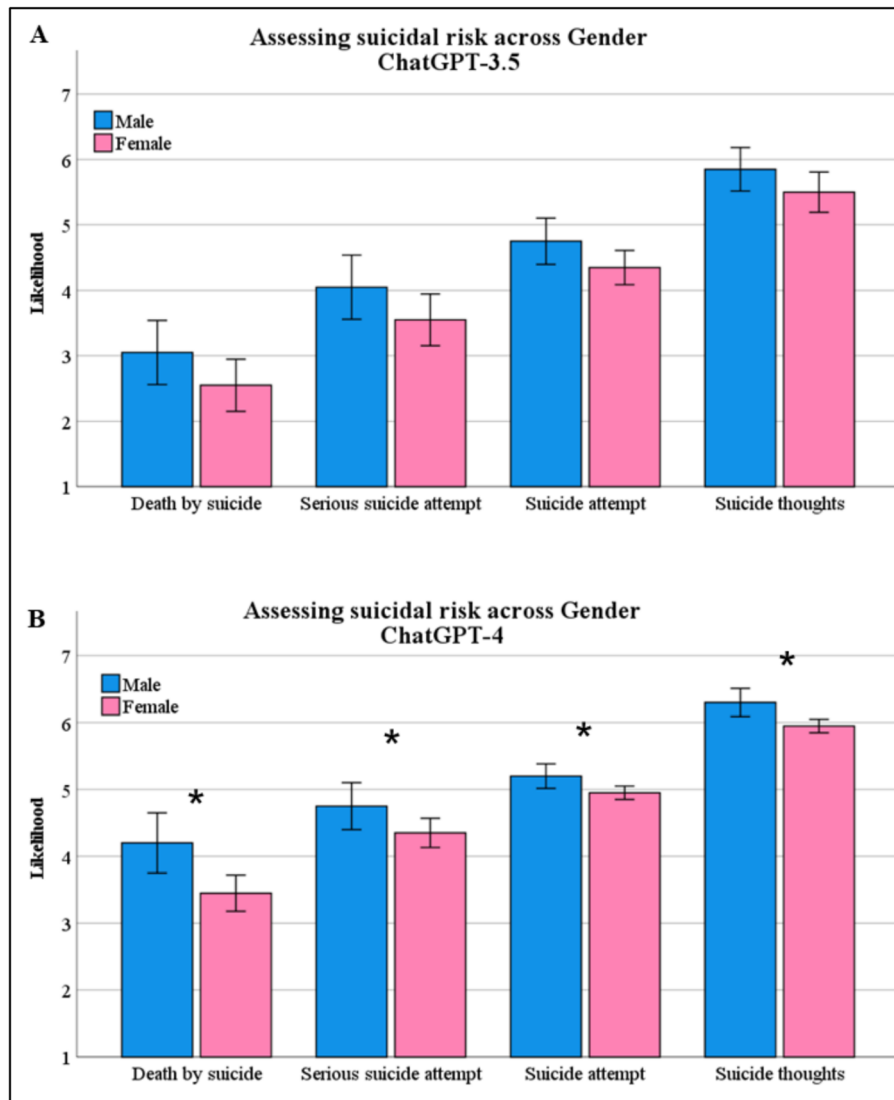**Fig. 1** Assessing suicidal risk across cultures— ChatGPT 3.5 and ChatGPT 4
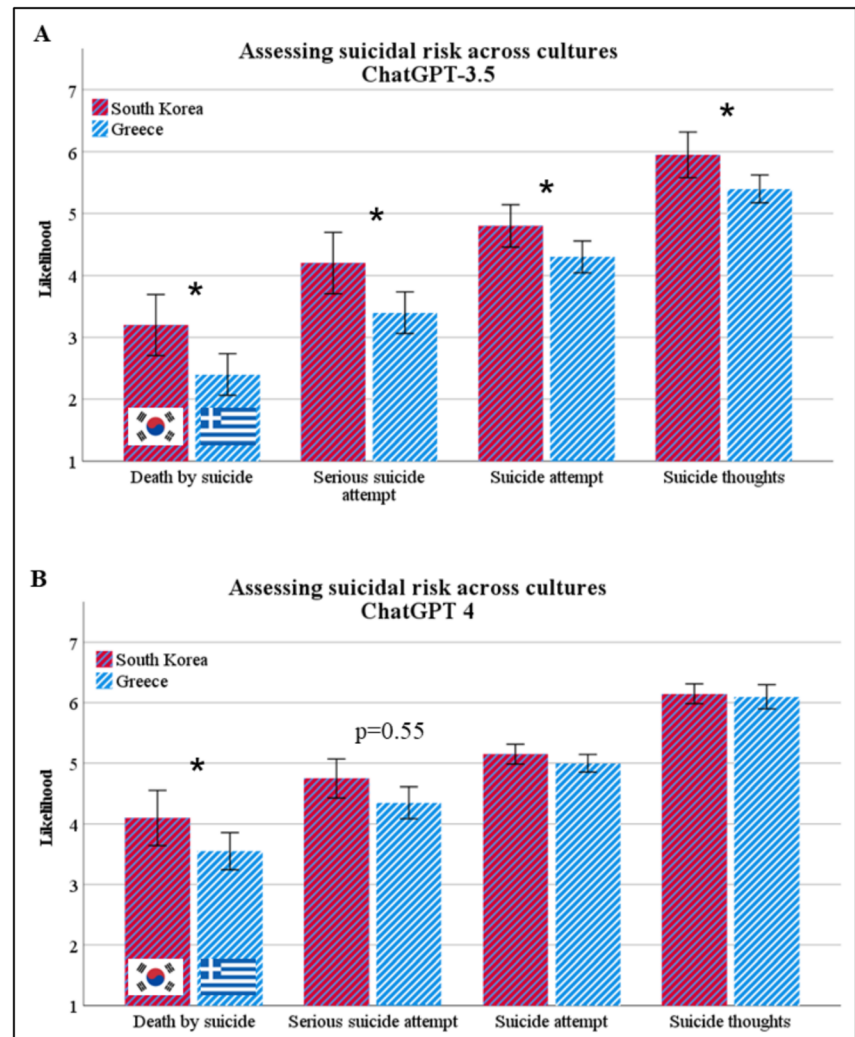
at significantly heightened risk for a range of suicidal behaviors. This observation aligns with statistical data and literature indicating that South Korea is grappling with the highest suicide rate among developed nations (OECD, 2024). The cultural dynamics at play in South Korea—among them demographic aging alongside factors such as the erosion of traditional family values and economic deprivation—contribute to forming a complex and multifaceted landscape of suicide risk (Cha et al., 2020; Lee et al., 2022).

ChatGPT-4 also displayed commendable sensitivity to specific cultural distinctions, particularly regarding the severity of suicidal actions within the South Korean context. This recognition represents a positive step in the ability of GenAI platforms to acknowledge the diverse challenges and risk factors faced by

individuals from varying cultural backgrounds in their mental health journeys (Mueller et al., 2021). Yet while maintaining sensitivity to certain cultural nuances, ChatGPT-4 adopted a more selective approach, with a predominant focus on the severity of suicidal actions within the South Korean context. This concentrated focus on severity does carry potential risks, as it may oversimplify the intricate cultural dynamics that shape mental health experiences (Mueller et al., 2021). Such an approach could unintentionally reinforce stereotypes and inadequately capture the multifaceted web of cultural influences on mental well-being, which vary significantly among individuals and communities.

Nevertheless, the current findings enhance our knowledge about the capabilities of ChatGPT,

Fig. 2 Assessing suicidal risk across gender—ChatGPT 3.5 and ChatGPT 4

demonstrating that this GenAI platform encompasses more than merely theoretical and semantic knowledge (Kung et al., 2023; Rudolph et al., 2023). Indeed, ChatGPT's can successfully identify the most critical and important cases of actual acts of suicide, while at the same time demonstrating cultural sensitivity. This finding is of major importance, as most prior research has focused on the technical applications of AI within the domain of mental health, including optimizing clinical tasks such as record-keeping and elevating diagnostic precision (Bzdok & Meyer-Lindenberg, 2018; Doraiswamy et al., 2020) Our study emphasizes the broader potential of GenAI that goes beyond these

technical abilities, including its ability to consider cultural dimensions in the area of mental health, thus fostering a more comprehensive approach to support and intervention. In fact, as GenAI is being developed and utilized, it is crucial to consider that these models, which are open to the public and may eventually be used for medical purposes, need to avoid replicating social and cultural biases and avoid reflecting racism. However, cultural aspects and the individual's country of origin can be significant factors that influence mental health outcomes. These factors may sometimes be perceived as prejudicial, but they are essential in the context of accurate risk assessments. The findings
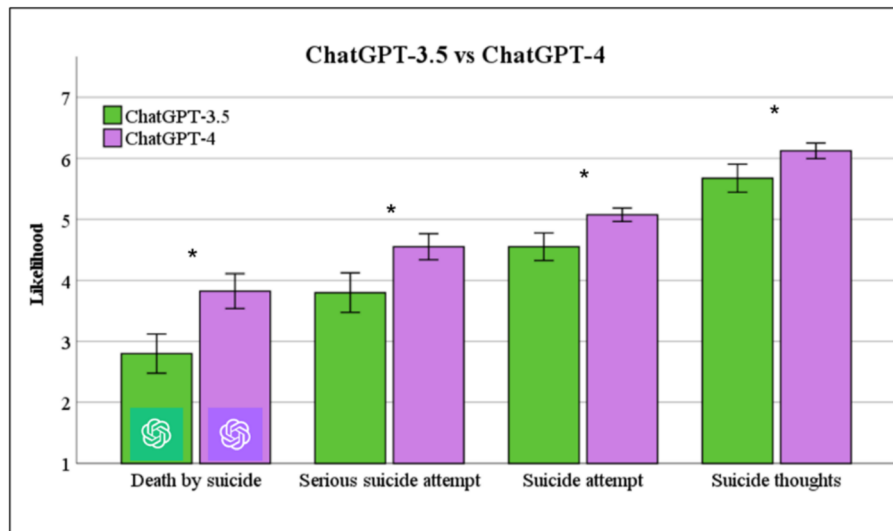
Fig. 3 ChatGPT 3.5 vs. ChatGPT 4

from this study indicate that open GenAI systems like ChatGPT can appropriately incorporate cultural and geographic information into their suicide risk assessments. This capability addresses concerns that, due to political correctness, important cultural differences might be excluded from GenAI assessments, potentially impairing their clinical effectiveness.

Gender effect

The study's results shed light on an intriguing gender effect in the suicide risk assessment capabilities of ChatGPT-3.5 and ChatGPT-4. These findings revealed notable differences in how these two GenAI models interpret and incorporate gender as a significant risk factor, regardless of cultural context. ChatGPT-4 exhibited sensitivity to gender-related factors in suicide risk assessment. Specifically, this AI model consistently identified male gender as a significant risk factor associated with a heightened likelihood of suicidal thoughts, suicide attempts, serious suicide attempts, and risk of dying from suicide. This finding is in line with established gender theories that highlight varying patterns of suicide risk based on gender (Schrijvers et al., 2012). Conversely, ChatGPT-3.5 demonstrated more limited sensitivity to the gender effect and appeared to approach gender as a risk factor with less certainty or significance than did ChatGPT-4.

These results prompt several considerations. First, the differing sensitivities of ChatGPT-3.5 and ChatGPT-4 to gender as a risk factor emphasize the importance of understanding the potential biases and cultural factors that may influence the risk assessments of GenAI models. Second, these findings underscore the complexity of gender-related risk factors in suicide, suggesting that GenAI models should be calibrated and continuously refined to provide more accurate and nuanced assessments in this regard.

Interaction between culture and gender

The results regarding the interaction between culture and gender in both ChatGPT 3.5 and ChatGPT 4 revealed the noteworthy absence of a significant effect. This implies that neither of these GenAI models demonstrated a strong inclination to modify their assessments of suicide risk based on the interplay between an individual's gender and cultural background.

This outcome raises several key considerations. First, it underscores the importance of evaluating the performance and sensitivity of GenAI models in nuanced and context-specific ways (Elyoseph & Levkovich, 2023). While these models exhibited some degree of cultural and gender sensitivity in isolation, they did not appear to make any significant adaptations in their risk assessments when these factors converged. This may indicate that the AI models treat

culture and gender as relatively independent variables in the context of suicide risk, possibly overlooking potential intersections and complexities. Based on prior research underscoring the significance of AI model accuracy (Graham et al., 2019), the absence of an interaction effect in this study highlights the potential necessity for further refinement and calibration of these models. This emphasizes the importance of integrating human expertise, particularly in sensitive areas such as suicide risk assessment, to enhance the effectiveness and cultural sensitivity of GenAI models.

Differences between ChatGPT-3.5 and ChatGPT-4

The study's findings revealed that ChatGPT-4 consistently assigned higher severity ratings to suicide risk factors than did ChatGPT-3.5, highlighting the presence of distinct strengths and weaknesses in their risk assessment capabilities (Elyoseph & Levkovich, 2023; Levkovich & Elyoseph, 2023a, 2023b). This distinction can help mental health professionals make informed decisions when selecting an AI model for suicide risk assessment (Bernert et al., 2020). ChatGPT-4's cautious approach appears to be well-suited for severe cases, enhancing treatment precision. In contrast, ChatGPT-3.5's more moderate ratings are appropriate for milder cases, offering a less intensive approach. Additional studies in the field of mental health that examined assessments of depression, schizophrenia, and suicidality between therapists (psychiatrists, doctors, psychologists) and bilinguals found that ChatGPT-4 demonstrated results similar to or better than human evaluators. In contrast, ChatGPT-3.5 was often not accurate enough (Elyoseph & Levkovich, 2024; Levkovich & Elyoseph, 2023b; Omar & Levkovich, 2024). Furthermore, these choices hold significant implications for individuals seeking mental health support. ChatGPT-4's tendency to encourage vigilant monitoring is especially valuable for severe cases, ensuring timely interventions. In contrast, ChatGPT-3.5's approach is beneficial for individuals with less acute conditions, resulting in a more balanced treatment plan.

Study's limitations

Although insightful, this study has several limitations that warrant consideration. First, the research focused exclusively on Greece and South Korea, which have notable different suicide rates. This narrow scope may not fully capture the global diversity of cultural influences on suicide risks. To enhance the generalizability of future research, a broader range of cultural contexts should be included. Second, the vignettes used in this study were not written in the native language of the participants (Korean or Greek), potentially affecting the authenticity and cultural resonance of the scenarios presented. Future studies should consider using linguistically and culturally tailored materials to improve the accuracy and relevance of the assessments. Third, this study examined only two GenAI models: ChatGPT-3.5 and ChatGPT-4. Although these models provide valuable insights, the performance of other GenAI models in culturally sensitive suicide risk assessments remains unexplored. Expanding the range of the GenAI models studied would provide a more comprehensive understanding of the capabilities and limitations of GenAI in this context.

Additionally, the use of a single vignette, although validated in previous studies, limits the ability to generalize the findings. There is a need to expand to a broader range of vignettes in future studies to enhance the reliability and generalizability of the results. Also, the use of vignettes does not fully replicate the complexity of real-world scenarios. Real patient data, with their inherent variability and depth, would provide a more robust test of the GenAI model's capabilities. Future research should incorporate real patient interactions to better assess the practical applications of these GenAI technologies. Finally, while this study highlighted the sensitivity of ChatGPT-4 to cultural and gender-related factors, it also revealed areas in which AI models could be improved. The observed limitations in handling the intersection of culture and gender suggest the need for the ongoing refinement of GenAI algorithms to ensure that they can provide nuanced and unbiased risk assessments across diverse populations. Suicidality assessment is a multidimensional phenomenon involving a variety of factors, including psychological, social, and biological influences. This complexity means that a comprehensive assessment often requires

more background information about the patient than that provided in the vignette studies. For instance, personal history, current life stressors, and detailed mental health history are crucial for an accurate risk assessment, and the lack of this information in the current study is a significant limitation. Moreover, the study's reliance on hypothetical scenarios may limit its applicability in real-world clinical settings. The responses of GenAI models to vignettes may not fully reflect their performance when dealing with actual patients who present unique and dynamic circumstances. These limitations highlight the need for longitudinal studies that track GenAI model performance over time and in diverse clinical environments, underscoring the need for continued research and development to enhance the cultural and contextual sensitivity of GenAI models in mental health assessments. Addressing these issues is crucial for developing GenAI tools that can effectively support suicide prevention worldwide. An additional limitation concerns statistical issues; AI-generated data do not always necessarily follow a normal distribution, often requiring the use of non-parametric statistical methods. Another important issue is the assumption of independence between data or observations. This can be partially addressed by inputting each question in a new tab and using models that do not train on their own generated data. Nonetheless, the question of what constitutes independence in the context of artificial intelligence remains an issue that requires redefinition.

Conclusions

Cultural diversity across the globe has a profound influence on various facets of mental health, among them perceptions of health and illness, help-seeking behaviors, consumer and practitioner attitudes, and mental health systems (Gopalkrishnan et al., 2018). Moreover, cultural diversity becomes particularly relevant due to the ongoing processes of globalization. Accordingly, many countries worldwide must address the challenge of providing psychiatric services to populations with diverse cultural backgrounds (Melluish & Globalization, 2014). In this context, the current study's findings provide a ray of hope by suggesting that ChatGPT models possess a notable degree of sensitivity to these intercultural differences. Given past concerns about rapid alignment processes

in GenAI models, these findings are significant. These processes strive to prevent algorithmic biases related to factors such as race, gender, or socioeconomic status and focus on auditing and evaluating algorithm fairness (Ray, 2023). The study's results indicate that these models exhibit a certain degree of sensitivity to intercultural distinctions, highlighting their potential to navigate the complexities of cultural diversity in mental health contexts.

Data availability The data that support the findings of this study are available from the authors upon reasonable request. No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no competing interests.

Ethical approval This study was approved by Institutional Review Board (YVC EMEK 2023-40) and conformed to the Declaration of Helsinki.

Patient and public involvement No patient involved.

References

Baek, I., Jo, S., Kim, E. J., Lee, G. R., Lee, D. H., & Jeon, H. J. (2021). A review of suicide risk assessment tools and their measured psychometric properties in Korea. *Frontiers in*

Psychiatry, 12, 679779. https://doi.org/10.3389/fpsyt.2021.679779

Bernert, R. A., Hilberg, A. M., Melia, R., Kim, J. P., Shah, N. H., & Abnousi, F. (2020). Artificial intelligence and suicide prevention: A systematic review of machine learning investigations. *International Journal of Environmental Research and Public Health, 17*(16), 5929. https://doi.org/10.3390/ijerph17165929

Bolton, J. M., Gunnell, D., & Turecki, G. (2015). Suicide risk assessment and intervention in people with mental illness. *BMJ, 351*, 4978. https://doi.org/10.1136/bmj.h4978

Bowden, M., McCoy, A., & Reavley, N. (2020). Suicidality and suicide prevention in culturally and linguistically diverse (CALD) communities: A systematic review. *International Journal of Mental Health, 49*(4), 293–320. https://doi.org/10.1080/00207411.2019.1694204

Bzdok, D., & Meyer-Lindenberg, A. (2018). Machine learning for precision psychiatry: Opportunities and challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 3*(3), 223–230. https://doi.org/10.1016/j.bpsc.2017.11.007

Cha, E. S., Chang, S., Choi, Y., & Lee, W. J. (2020). Trends in pesticide suicide in south Korea, 1983–2014. *Epidemiology and Psychiatric Sciences, 29*, e25. https://doi.org/10.1017/S2045796019000118

Chang, S., Gunnell, D., Sterne, J. A., Lu, T., & Cheng, A. T. (2009). Was the economic crisis 1997–1998 responsible for rising suicide rates in east/southeast Asia? A time–trend analysis for Japan, Hong Kong, south Korea, Taiwan, Singapore and Thailand. *Social Science & Medicine, 68*(7), 1322–1331. https://doi.org/10.1016/j.socscimed.2009.01.010

Cheng, S. W., Chang, C. W., Chang, W. J., Wang, H. W., Liang, C. S., Kishimoto, T., ... & Su, K. P. (2023). The now and future of ChatGPT and GPT in psychiatry. *Psychiatry and Clinical Neurosciences.* https://doi.org/10.1111/pcn.13588

Doraiswamy, P. M., Blease, C., & Bodner, K. (2020). Artificial intelligence and the future of psychiatry: Insights from a global physician survey. *Artificial Intelligence in Medicine, 102*, 101753. https://doi.org/10.1016/j.artmed.2019.101753

Elyoseph, Z., & Levkovich, I. (2023). Beyond human expertise: The promise and limitations of ChatGPT in suicide risk assessment. *Frontiers in Psychiatry, 14*, 1213141. https://doi.org/10.3389/fpsyt.2023.1213141

Elyoseph, Z., & Levkovich, I. (2024). Comparing the perspectives of generative AI, mental health experts, and the general public on schizophrenia recovery: Case vignette study. *JMIR Mental Health, 11*, e53043. https://doi.org/10.2196/53043

Eskin, M. (2020). Suicidal behavior in the mediterranean countries. *Clinical Practice and Epidemiology in Mental Health CP & EMH., 16*(1), 93. https://doi.org/10.2174/1745017902016010093

Eurostat (2023). Deaths by suicide down by almost 14% in a decade. https://ec.europa.eu/eurostat/web/products-eurostat-news/w/edn-20230908-3

Feigelman, W., Plöderl, M., Rosen, Z., & Cerel, J. (2019). Research note on whether sexual minority individuals are over-represented among suicide's casualties. *Crisis.* https://doi.org/10.1027/0227-5910/a000626

Fiske, A., Henningsen, P., & Buyx, A. (2019). Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *Journal of Medical Internet Research, 21*(5), e13216. https://doi.org/10.2196/13216

Fountoulakis, K. N. (2019). Suicide rate in Greece stabilizes at historically high levels but still lowest in Europe. *Journal of Affective Disorders, 254*, 117–119. https://doi.org/10.1016/j.jad.2018.12.022

Fountoulakis, K. N., Chatzikosta, I., Pastiadis, K., Zanis, P., Kawohl, W., Kerkhof, A. J., & Bech, P. (2016). Relationship of suicide rates with climate and economic variables in Europe during 2000–2012. *Annals of General Psychiatry., 15*, 1–6. https://doi.org/10.1186/s12991-016-0106-2

Fountoulakis, K. N., Savopoulos, C., Siamouli, M., Zaggelidou, E., Mageiria, S., Iacovides, A., & Hatzitolios, A. I. (2013). Trends in suicidality amid the economic crisis in Greece. *European Archives of Psychiatry and Clinical Neuroscience, 263*, 441–444. https://doi.org/10.1007/s00406-012-0385-9

Fraiwan, M., & Khasawneh, N. (2023). A review of ChatGPT applications in education, marketing, software engineering, and healthcare: Benefits, drawbacks, and research directions. *arXiv Preprint* arXiv:2305.00237. https://doi.org/10.48550/arXiv.2305.00237

Gearing, R. E., & Alonzo, D. (2018). Religion and suicide: New findings. *Journal of Religion and Health, 57*, 2478–2499. https://doi.org/10.1007/s10943-018-0629-8

Gopalkrishnan, N. (2018). Cultural diversity and mental health: Considerations for policy and practice. *Frontiers in Public Health, 6*, 179. https://doi.org/10.3389/fpubh.2018.00179

Graham, S., Depp, C., Lee, E. E., Nebeker, C., Tu, X., Kim, H., & Jeste, D. V. (2019). Artificial intelligence for mental health and mental illnesses: An overview. *Current Psychiatry Reports, 21*, 1–18. https://doi.org/10.1007/s11920-019-1094-0

Graney, J., Hunt, I. M., Quinlivan, L., Rodway, C., Turnbull, P., Gianatsi, M., & Kapur, N. (2020). Suicide risk assessment in UK mental health services: a national mixed-methods study. *The Lancet Psychiatry., 7*(12), 1046–1053. https://doi.org/10.1016/S2215-0366(20)30381-3

Gvion, Y., & Levi-Belz, Y. (2018). Serious suicide attempts: Systematic review of psychological risk factors. *Frontiers in Psychiatry, 9*, 56. https://doi.org/10.3389/fpsyt.2018.00056

Haber, Y., Levkovich, I., Hadar-Shoval, D., & Elyoseph, Z. (2024). The artificial third: A broad view of the effects of introducing generative artificial intelligence on psychotherapy. *JMIR Mental Health, 11*, e54781. https://doi.org/10.1136/fmch-2023-002391

Kim, A. M. (2020). Factors associated with the suicide rates in Korea. *Psychiatry Research, 284*, 112745. https://doi.org/10.1016/j.psychres.2020.112745

Kim, J. W., Jung, H. Y., Won, D. Y., Noh, J. H., Shin, Y. S., & Kang, T. I. (2019). Suicide trends according to age, gender, and marital status in South Korea. *OMEGA-Journal of Death and Dying, 79*(1), 90–105. https://doi.org/10.1177/0030222817715756

Kim, J. W., Jung, H. Y., Won, D. Y., Shin, Y. S., Noh, J. H., & Kang, T. I. (2020). Landscape of elderly suicide in South Korea: Its trend according to age, gender, and educational

attainment. *OMEGA-Journal of Death and Dying, 82*(2), 214–229. https://doi.org/10.1177/0030222818807845

Knipe, D., Padmanathan, P., Newton-Howes, G., Chan, L. F., & Kapur, N. (2022). Suicide and self-harm. *The Lancet, 399*(10338), 1903–1916. https://doi.org/10.1016/S0140-6736(22)00173-8

Kontaxakis, V., Papaslanis, T., Havaki-Kontaxaki, B., Tsouvelas, G., Giotakos, O., & Papadimitriou, G. N. (2013). Suicide in Greece: 2001–2011. *Psychiatrike Psychiatriki., 24*(3), 170–174.

Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., & Maningo, J. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLoS Digital Health, 2*(2), e0000198. https://doi.org/10.1371/journal.pdig.0000198

Lawrence, R. E., Oquendo, M. A., & Stanley, B. (2016). Religion and suicide risk: A systematic review. *Archives of Suicide Research, 20*(1), 1–21. https://doi.org/10.1080/13811118.2015.1004494

Lee, H., Kim, R., Jang, S., & Kawachi, I. (2022). The relative importance of macro versus micro geographical scale in explaining suicide variation in Seoul, south Korea 2014–2016. *PLoS ONE, 17*(9), e0273866. https://doi.org/10.1371/journal.pone.0273866

Lee, S., Oh, I., Jeon, H. J., & Roh, S. (2017). Suicide rates across income levels: Retrospective cohort data on 1 million participants collected between 2003 and 2013 in south Korea. *Journal of Epidemiology, 27*(6), 258–264. https://doi.org/10.1016/j.je.2016.06.008

Lejeune, A., Le Glaz, A., Perron, P. A., Sebti, J., Baca-Garcia, E., Walter, M., & Berrouiguet, S. (2022). Artificial intelligence and suicide prevention: a systematic review. *European Psychiatry., 65*(1), e19. https://doi.org/10.1192/j.eurpsy.2022.8

Levi-Belz, Y., & Gamliel, E. (2016). The effect of perceived burdensomeness and thwarted belongingness on therapists' assessment of patients' suicide risk. *Psychotherapy Research, 26*(4), 436–445. https://doi.org/10.1080/10503307.2015.1013161

Levi-Belz, Y., Gvion, Y., & Apter, A. (2022). The serious suicide attempts approach for understanding suicide: Review of the psychological evidence. *OMEGA-Journal of Death and Dying, 86*(2), 591–608. https://doi.org/10.1177/0030222820981235

Levkovich, I., & Elyoseph, Z. (2023a). Suicide risk assessments through the eyes of Chatgpt-3.5 versus ChatGPT-4: Vignette study. *JMIR Mental Health., 10*, e51232. https://doi.org/10.2196/51232

Levkovich, I., & Elyoseph, Z. (2023). Identifying depression and its determinants upon initiating treatment: ChatGPT versus primary care physicians. *Family Medicine and Community Health., 11*(4), e002391. https://doi.org/10.1136/fmch-2023-002391

Lim, M., Lee, S. U., & Park, J. (2014). Difference in suicide methods used between suicide attempters and suicide completers. *International Journal of Mental Health Systems, 8*, 1–4. https://doi.org/10.1186/1752-4458-8-54

Melluish, S. (2014). Globalization, culture and psychology. *International Review of Psychiatry, 26*(5), 538–543. https://doi.org/10.3109/09540261.2014.918873

Mortier, P., Auerbach, R. P., Alonso, J., Bantjes, J., Benjet, C., Cuijpers, P., Ebert, D. D., Green, J. G., Hasking, P., & Nock, M. K. (2018). Suicidal thoughts and behaviors among first-year college students: Results from the WMH-ICS project. *Journal of the American Academy of Child & Adolescent Psychiatry., 57*(4), 263–2731. https://doi.org/10.1016/j.jaac.2018.01.018

Mueller, A. S., Abrutyn, S., Pescosolido, B., & Diefendorf, S. (2021). The social roots of suicide: Theorizing how the external social world matters to suicide and suicide prevention. *Frontiers in Psychology, 12*, 763. https://doi.org/10.3389/fpsyg.2021.621569

OECD (2024), Suicide rates (indicator). https://doi.org/10.1787/a82f3459-en (Accessed on 26 May 2024). https://data.oecd.org/healthstat/suicide-rates.htm.

Omar, M., Sr., & Levkovich, I. (2024). Exploring the efficacy and potential of large language models for depression: A systematic review. *medRxiv*. https://doi.org/10.1101/2024.05.07.24306897

Qian, G. (2021). Associations of suicide and subjective well-being. *OMEGA-Journal of Death and Dying, 84*(1), 103–115. https://doi.org/10.1177/0030222819880091

Rachiotis, G., Stuckler, D., McKee, M., & Hadjichristodoulou, C. (2015). What has happened to suicides during the greek economic crisis? findings from an ecological study of suicides and their determinants (2003–2012). *British Medical Journal Open, 5*(3), e007295. https://doi.org/10.1136/bmjopen-2014-007295

Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*. https://doi.org/10.1016/j.iotcps.2023.04.003

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education. *Journal of Applied Learning and Teaching*. https://doi.org/10.37074/jalt.2023.6.1.9

Sallam, M. (2023). The utility of ChatGPT as an example of large language models in healthcare education, research and practice: Systematic review on the future perspectives and potential limitations. *medRxiv*. https://doi.org/10.3390/healthcare11060887

Schrijvers, D. L., Bollen, J., & Sabbe, B. G. (2012). The gender paradox in suicidal behavior and its impact on the suicidal process. *Journal of Affective Disorders, 138*(1–2), 19–26. https://doi.org/10.1016/j.jad.2011.03.050

Tal, A., Elyoseph, Z., Haber, Y., Angert, T., Gur, T., Simon, T., & Asman, O. (2023b). The artificial third: Utilizing ChatGPT in mental health. *The American Journal of Bioethics, 23*(10), 74–77. https://doi.org/10.1080/15265161.2023.2250297

Tal, A., Haber, Y., Angert, T., Gur, T., Simon, T., & Asman, O. (2023). The artificial third: Utilizing ChatGPT in mental health. *The American Journal of Bioethics., 23*(10), 74–77. https://doi.org/10.1080/15265161.2023.2250297

van Heerden, A. C., Pozuelo, J. R., & Kohrt, B. A. (2023). Global mental health services and the impact of artificial intelligence-powered large language models. *JAMA Psychiatry, 80*(7), 662–664. https://doi.org/10.1001/jamapsychiatry.2023.1253

Wampold, B. E., & Flückiger, C. (2023). The alliance in mental health care: Conceptualization, evidence and clinical applications. *World Psychiatry, 22*(1), 25–41. https://doi.org/10.1002/wps.21035

Yamaoka, K., Suzuki, M., Inoue, M., Ishikawa, H., & Tango, T. (2020). Spatial clustering of suicide mortality and associated community characteristics in kanagawa prefecture, Japan, 2011–2017. *BMC Psychiatry, 20*, 1–15. https://doi.org/10.1186/s12888-020-2479-7

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