# Integrating Previous Suicide Attempts, Gender, and Age Into Suicide Risk Assessment Using Advanced Artificial Intelligence Models

Shiri Shinan-Altman, PhD; Zohar Elyoseph, PhD; and Inbar Levkovich, PhD

## Abstract

**Objective:** Suicide is a critical global health concern. Research indicates that generative artificial intelligence (GenAI) and large language models, such as generative pretrained transformer-3 (GPT-3) and GPT-4, can evaluate suicide risk comparably to experts, yet the criteria these models use are unclear. This study explores how variations in prompts, specifically regarding past suicide attempts, gender, and age, influence the risk assessments provided by ChatGPT-3 and ChatGPT-4.

**Methods:** Using a controlled scenario-based approach, 8 vignettes were created. Both ChatGPT-3.5 and ChatGPT-4 were used to predict the likelihood of serious suicide attempts, suicide attempts, and suicidal thoughts. A univariate 3-way analysis of variance was conducted to analyze the effects of the independent variables (previous suicide attempts, gender, and age) on the dependent variables (likelihood of serious suicide attempts, suicide attempts, and suicidal thoughts).

**Results:** Both ChatGPT-3.5 and ChatGPT-4 recognized the importance of previous suicide attempts in predicting severe suicide risks and suicidal thoughts. ChatGPT-4 also identified gender differences, associating men with a higher risk, while both models disregarded age as a risk factor. Interaction analysis revealed that ChatGPT-3.5 associated past attempts with a higher likelihood of suicidal thoughts in men, whereas ChatGPT-4 showed an increased risk for women.

**Conclusions:** The study highlights ChatGPT-3.5 and ChatGPT-4's potential in suicide risk evaluation, emphasizing the importance of prior attempts and gender, while noting differences in their handling of interactive effects and the negligible role of age. These findings reflect the complexity of GenAI decision-making. While promising for suicide risk assessment, these models require careful application due to limitations and real-world complexities.

*J Clin Psychiatry 2024;85(4):24m15365*

*Author affiliations are listed at the end of this article.*

The role of artificial intelligence (AI) in suicide risk assessment, especially through large language models (LLMs) like ChatGPT, has recently gained focus. Early research revealed the potential and challenges of using AI for this purpose, including how it compares to mental health professionals in effectiveness.[1,2] Despite these advancements, the workings of generative AI (GenAI) systems in suicide risk assessments remain unclear, often termed a "black box." This study presents a new method to understand the factors GenAI considers by slightly modifying input to mirror various risk factors and observing changes in AI evaluations, shedding light on its decision-making process.

Suicide is a major public health issue, with over 700,000 yearly deaths worldwide,[3] and is the fourth leading cause of death among 15- to 29-year-olds. In the United States, the suicide mortality rate is 14.2 per 100,000 individuals,[4] while in Israel, there are 90–100 attempts per 100,000 people annually, with around 7.9 resulting in death.[5]

The suicide construct encompasses a range of behaviors, including suicidal thoughts, suicide attempts, serious suicide attempts, and completed suicides. *Suicidal thoughts*, marked by ideations of killing oneself, often foreshadow suicide attempts and deaths via suicide.[6] A *suicide attempt* involves deliberate self-harm with the intent to end one's life, spanning various actions

**Editor's Note**

We encourage authors to submit papers for consideration as a part of our Focus on Suicide section. Please contact Philippe Courtet, MD, PhD, at pcourtet@psychiatrist.com.

## Clinical Points

- Current literature lacks insights into how artificial intelligence (AI) models evaluate suicide risk based on factors. This study addresses this gap and highlights AI's potential and limitations.
- Clinicians should recognize that AI models, while identifying major suicide risk factors, may miss the complex interaction of multiple factors. AI should be used as a supplementary tool alongside thorough clinical evaluation.

from gestures to lethal acts. A *serious suicide attempt* is potentially fatal without immediate intervention.[7] Individuals with prior suicide attempts face a 30–40 times higher future suicide risk than the general population.[8] Within a year of an attempt, 1.6%–2.8% die by suicide, with the risk highest immediately after and decreasing over time. Addressing various risk factors, including prior attempts, gender, and age, is crucial due to the seriousness of the issue.[9,10]

Suicidal behavior varies by gender, with males having higher suicide rates but females attempting suicide more frequently, known as the "gender paradox of suicidal behavior."[11–13(p5)] Specifically, in 2019, the age-standardized suicide rate was 12.6 per 100,000 in men and 5.4 per 100,000 in women.[14] In addition, approximately 3.9% of women reported attempting suicide at some point in their lives, which is higher than the 2.0% reported by men.[15]

Suicide risk varies significantly across age groups. While global suicide rates are higher among older individuals, this does not mean the impact is less significant among younger people.[16] Contrarily, the number of suicide deaths peaks at age 25 years, suggesting a high impact on the young.[3] This apparent contradiction is due to the larger population of younger individuals. In 2012, suicides accounted for 8.5% of all deaths in the 15–29 years age group, much higher than the 1.3% in the total population in 2019.[3] Additionally, suicide attempts are more frequent but less often fatal in those under 30–35 years than in older age groups, indicating a complex relationship between age, suicide rates, and outcomes.[8]

Given the severity of suicide as a public health issue, substantial investment in prevention programs is crucial, focusing on factors such as suicide attempts, gender, and age.[8,13] Health care professionals must be equipped to recognize early warning signs of suicidality for effective crisis management and intervention.[17] In response to this need, there is increasing interest in leveraging technologies such as natural language processing to develop systems that can detect linguistic indicators of distress and other risk factors.[18] Furthermore, emerging tools like ChatGPT offer potential advancements in suicide prevention strategies, suggesting a promising future for technology-assisted support and intervention.[19,20]

ChatGPT, a GenAI language model, is designed to understand text queries and respond in natural language, finding applications in education,[21] research,[22] and software development.[23] Recently, it has made strides in the medical field, aiding in disease diagnosis[24] and expertise assessment.[25] With ongoing development, the model now offers better linguistic diversity,[26,27] multilingual capabilities, and image processing, expanding its potential across various domains.[28] However, its practicality must be weighed against costs and limitations.[26]

ChatGPT, powered by machine learning and trained on clinical interactions, could enhance mental health support and improve decision-making and the predictive accuracy of assessments for suicidal behavior.[29–31] However, its predictive capabilities have limitations; for instance, it has been found to underestimate suicide risk.[1] While ChatGPT shows promise in transforming mental healthcare, replacing human judgment with AI is not recommended.[32] The reliability of ChatGPT in mental health contexts is challenged by the potential for inaccuracies and inappropriate advice stemming from its learning sources, posing risks especially to those with mental health concerns.[30]

This study evaluates how LLMs consider suicidality risk factors, following research that highlights their potential for accurate clinical assessments.[33,34] Drawing from contingent identity theory,[35] which explores how suicide risk interacts with gender and age, this research investigated the use of GenAI models like ChatGPT-3.5 and ChatGPT-4 in recognizing risk factors such as prior suicide attempts, gender, and age. The objectives were 2-fold: (1) to examine whether ChatGPT-3.5 and ChatGPT-4 incorporate risk factors such as previous suicide attempts, gender, and age, in their assessment of suicide and (2) to assess whether ChatGPT-4 demonstrates improved consideration of suicide risk factors, including prior suicide attempts, gender, and age, compared to ChatGPT-3.5.

We hypothesized that both ChatGPT-3.5 and ChatGPT-4 would predict a heightened suicide risk for a 21-year-old man with a history of suicide attempts in comparison with a 37-year-old woman with no prior suicide attempts. We also hypothesized that ChatGPT-4 would exhibit enhanced consideration of suicide risk factors—namely, previous suicide attempts, gender, and age—in comparison to ChatGPT-3.5. The study's model is presented in Figure 1.

**The Study's Model**
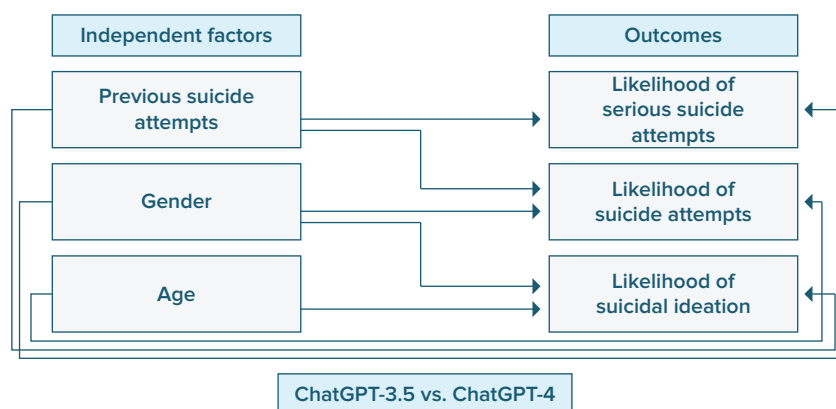


## METHODS

### AI Procedure

We used ChatGPT-3.5 and ChatGPT-4 (the free July 24 version) in July–August 2023 to examine the performance of ChatGPT in assessing how gender, age, and previous suicide attempts affected ChatGPT's evaluation of suicide risk (ie, likelihood of serious suicide attempt, likelihood of suicide attempt, and likelihood of suicidal thoughts). Given that patient involvement was not a component of this study and that we utilized vignettes and GPT technology, the engagement of an ethics committee was deemed unnecessary.

### Input Source

The basic vignette was taken from a previous study by Levi-Belz and Gamliel,[36] presenting a 37-year-old woman with no report of previous suicide attempts. In this study,[36] the vignette was tested among mental health professionals who rated the level of suicidal risk. Their evaluations served as a reference point to the current study, indicating that the inclusion of additional risk factors may exacerbate the risk assessment. This vignette[36] was also assessed by ChatGPT-3.5 and ChatGPT-4, with ChatGPT-4 estimating the likelihood of suicide attempts in a manner akin to evaluations provided by professionals.[2] Our study builds on these findings by examining whether changes in risk factors increase the risk of suicide, as suggested by the literature. We added risk factors (gender, age, and previous suicidal experience) to determine if AI models incorporate these variables in their evaluations, based on a vignette originally examined on human beings.

To evaluate the effect of gender, age, and previous suicide attempts, we added an additional vignette that also included data on loneliness and hopelessness. Only gender, age, and previous suicide attempts changed between the 8 different vignettes: with or without history of suicide attempts; gender (female or male); and age (37 or 21). Each of the vignettes was evaluated 10 times in ChatGPT-3.5 and in ChatGPT-4 in a new tab, with 160 evaluations in total.

**The vignette's structure.** Jane (or John) is a 37 (or 21)-year-old single woman (or man) without children, currently living alone, but in her (or his) past spent a long time living with her (or his) father, who passed away 2 years ago previously following a severe illness. Jane holds a BA degree and has kept a steady job for 10 (or a few) years. Recently, however, her manager informed her that, due to financial cuts, she must dismiss her. Jane had a close, intimate relationship for the past year with a boyfriend (or girlfriend), her first serious and meaningful one; she was very excited about the relationship, hoping it would lead to marriage. However, following a period of repeated quarrels, her boyfriend informed her he no longer loved her, and he left her. Jane was left feeling very sad and hurt. In recent weeks, she has continued to think about the relationship repeatedly, trying to analyze what happened. Jane is experiencing a heavy emotional load, and she is very tired, lacks energy, and finds it hard to sleep nights. Her routine tasks have become difficult for her, and she finds it hard to concentrate and to function properly. She dreads the future, and it seems to her that there is no chance she will find a love like this again (or add "in addition, in the past she/he carried out a suicidal attempt"). Jane feels worthless. Often she experiences herself as a burden to the people around her and she feels that it would be better if she were gone. She often sees herself as having no real meaning in other people's lives, and she experiences feelings of self-hatred. For example, she is angry at herself and thinks that her boyfriend's leaving her was her fault. Jane experiences significant loneliness in general and has only a few, if any,

close friends around her. She is very withdrawn, distant from others, and feels there are no people to whom she can turn in times of trouble. Therefore, she has not shared her feelings with anyone since the breakup.

## Measures

After each of the 8 vignettes (female or male * age 37 or 21 * with or without a history of suicide attempts), we fed the following statement into ChatGPT: "Answer the following questions as if you were a psychologist who specializes in suicide risk assessment": Q1: *"What is the likelihood Jane (or John) will have suicidal thoughts?"*; Q2: *"What is the likelihood Jane (or John) will attempt suicide?"*; Q3: *"What is the likelihood that Jane (or John) will make a serious suicide attempt?"* The above questions were taken from Levi-Belz and Gamliel.[35] A 7-point Likert type scale was employed, estimating a likelihood ranging from 1 (*very slight*) to 7 (*very high*).

## Statistical Analysis

To evaluate the influence of each of the independent factors (previous suicide attempts, gender, and age) on each of the 3 outcome variables (likelihood of serious suicide attempt, likelihood of suicide attempt, and likelihood of suicidal thoughts), we used a univariate 3-way analysis of variance (ANOVA) per each of the dependent variables. Given that our main hypothesis refers to the main effect of each of the independent variables, and to a lesser degree the combination of several independent factors, and to avoid the problem of multiple comparisons, we will describe only the interaction effects in which the main effect was significant.

## RESULTS

Table 1 presents the ANOVA conducted to investigate the influence of the 3 independent variables—previous suicide attempts (yes, no), gender (male, female), and age (21, 37)—on each of the 3 dependent variables: likelihood of serious suicide attempt, likelihood of suicide attempt, and likelihood of suicidal thoughts.

## Likelihood of Serious Suicide Attempt

A 1-way ANOVA was conducted to investigate the influence of each of the following independent variables—previous suicide attempts (yes, no), gender (male, female), and age (21, 37)—on the likelihood of a serious suicide attempt.

A significant effect of previous suicide attempts was found in both ChatGPT-3.5 and ChatGPT-4, respectively: $F_{1,79} = 11.72$, $P < .001$; $F_{1,79} = 6.40$, $P < .05$. Individuals who had previously attempted suicide (mean = 4.75,

**Table 1.**

**ANOVA of the Effect of Previous Suicide Attempts (Yes, No), Gender (Male, Female), and Age (21 y, 37 y) on Likelihood of Serious Suicide Attempt, Likelihood of Suicide Attempt, and Likelihood of Suicidal Thoughts**

|  | Independent | Dependent | $F_{1,79}$ |
|---|---|---|---|
| **ChatGPT-3.5** | Previous suicide attempts | Serious suicide attempt | 11.72*** |
|  |  | Suicide attempt | 8.7** |
|  |  | Suicidal thoughts | 6.31* |
|  | Gender | Serious suicide attempt | NS |
|  |  | Suicide attempt | NS |
|  |  | Suicidal thoughts | NS |
|  | Age | Serious suicide attempt | NS |
|  |  | Suicide attempt | NS |
|  |  | Suicidal thoughts | NS |
| **ChatGPT-4** | Previous suicide attempts | Serious suicide attempt | 6.4* |
|  |  | Suicide attempt | NS |
|  |  | Suicidal thoughts | 8.6** |
|  | Gender | Serious suicide attempt | 17.3*** |
|  |  | Suicide attempt | 11.4*** |
|  |  | Suicidal thoughts | 5.1* |
|  | Age | Serious suicide attempt | NS |
|  |  | Suicide attempt | NS |
|  |  | Suicidal thoughts | NS |

*$P < .05$, **$P < .01$, ***$P < .001$.
Abbreviations: ANOVA = analysis of variance, NS = nonsignificant.

SD = 1.15; mean = 3.73, SD = 0.90; in ChatGPT-3.5 and ChatGPT-4, respectively) had a higher likelihood of making a serious suicide attempt than those who had not (mean = 3.88, SD = 1.14; mean = 3.23, SD = 0.86; in ChatGPT-3.5 and ChatGPT-4, respectively).

A significant effect of gender was found in ChatGPT-4 but not in ChatGPT-3.5, $F_{1,79} = 5.10$, $P < .05$; $P > .05$, in ChatGPT-4 and ChatGPT-3.5, respectively. In accordance with ChatGPT-4, men (mean = 3.70, SD = 0.76) had a higher likelihood of making a serious suicide attempt than did women (mean = 3.25, SD = 1.001) (Figure 2). No significant effect of age was found in either ChatGPT-4 or ChatGPT-3.5 ($P > .05$).

To evaluate the interaction between the significant independent variables (previous suicide attempts and gender), we simultaneously analyzed them using a 2-way ANOVA. A significant interaction between previous suicide attempts and gender was found in ChatGPT-4, $F_{1,79} = 4.53$, $P < .05$. Previous suicide attempts were found to exacerbate the risk for women but did not significantly change the risk for men, who remained at high risk either way. In addition, a borderline significant interaction was observed between previous suicide attempts and gender in ChatGPT-3.5, $F_{1,79} = 3.53$,

**The Effect of Previous Suicidal Attempt on the Likelihood of Serious Suicide Attempt (Mean ± SEM) Evaluated by ChatGPT-3.5**
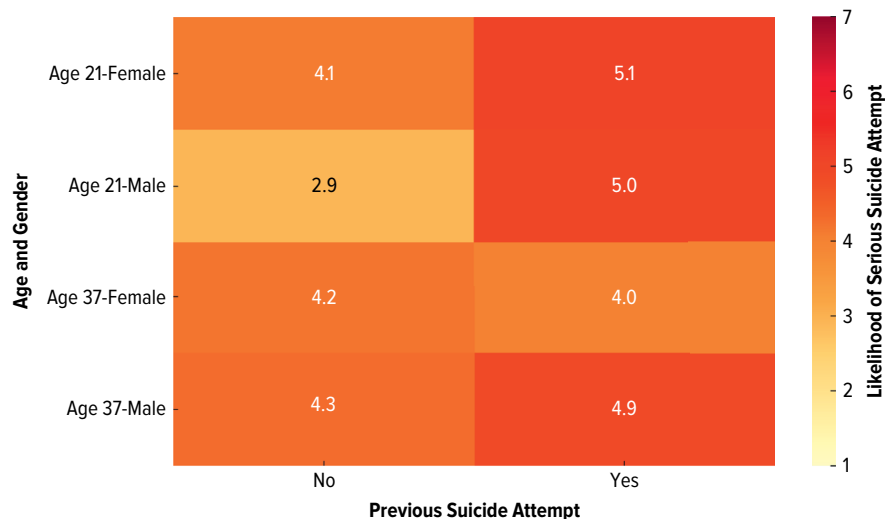


$P = .064$. It was found that previous suicide attempts greatly exacerbated the effect on women's risk but exacerbated the effect on men's risk to a much lesser extent Figures 2 and 3.

## Likelihood of Suicide Attempt

A 1-way ANOVA was conducted to investigate the influence of each of the following independent variables—previous suicide attempts (yes, no), gender (male, female), and age (21, 37)—on the likelihood of suicide attempt.

A significant effect of previous suicide attempts was found only in ChatGPT-3.5 but not in ChatGPT-4, $F_{1,79} = 8.77$, $P < .01$; $P > .05$, in ChatGPT-3.5 and ChatGPT-4, respectively. Individuals who had previously attempted suicide (mean = 5.5, SD = 0.75) had a higher likelihood of making a suicide attempt than those who had not (mean = 4.93, SD = 0.97).

A significant effect of gender was found in ChatGPT-4 but not in ChatGPT-3.5, $F_{1,79} = 11.31$, $P < .001$; $P > .05$, in ChatGPT-4 and ChatGPT-3.5, respectively. In accordance with ChatGPT-4, men (mean = 4.70, SD = 0.76) had a higher likelihood of making a suicide attempt than did women (mean = 4.15, SD = 0.7). No significant effect of age was found in either ChatGPT-4 or ChatGPT-3.5 ($P > .05$).

To evaluate the interaction between the significant independent variables (previous suicide attempts and gender) we simultaneously analyzed them using a 2-way ANOVA. A significant interaction between previous suicide attempts and gender was found in ChatGPT-3.5, $F_{1,79} = 7.91$, $P < .01$. It was found that previous suicide

attempts exacerbated the risk for men but did not significantly change the risk for women. In addition, a significant interaction was observed between previous suicide attempts and gender in ChatGPT-4, $F_{1,79} = 4.9$, $P < .05$. It was found that previous suicide attempts exacerbated the risk for women but did not significantly change the risk for men, who remained at high risk either way.

## Likelihood of Suicidal Thoughts

A 1-way ANOVA was conducted to investigate the influence of each of the following independent variables—previous suicide attempts (yes, no), gender (male, female), and age (21, 37) —on the likelihood of suicidal thoughts.

A significant effect of previous suicide attempts was found in both ChatGPT-3.5 and ChatGPT-4, $F_{1,79} = 6.3$, $P < .05$; $F_{1,79} = 8.60$, $P < .01$, in ChatGPT-3.5 and ChatGPT-4, respectively. Individuals who had previously attempted suicide (mean = 6.58, SD = 0.59; mean = 5.93, SD = 0.57; in ChatGPT-3.5 and ChatGPT-4, respectively) had a higher likelihood of having suicidal thoughts than those who had not (mean = 6.18, SD = 0.83; mean = 5.5, SD = 0.72; in ChatGPT-3.5 and ChatGPT-4, respectively).

A significant effect of gender was found in ChatGPT-4 but not in ChatGPT-3.5, $F_{1,79} = 17.32$, $P < .001$; $P > .05$, in ChatGPT-4 and ChatGPT-3.5, respectively. In accordance with ChatGPT-4, men (mean = 6.00, SD = 0.64) had a higher likelihood of having suicidal thoughts than did women (mean = 5.43, SD = 0.59). No significant effect of

Figure 3.

**The Effect of Previous Suicidal Attempt on the Likelihood of Serious Suicide Attempt (Mean ± SEM) Evaluated by ChatGPT-4**
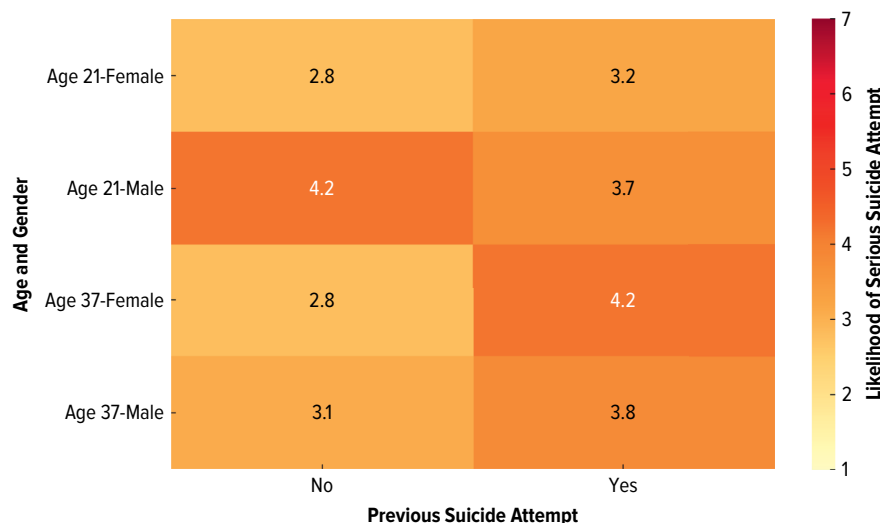


age was found in either ChatGPT-4 or ChatGPT-3.5 ($P > .05$). To evaluate the interaction between the significant independent variables (previous suicide attempts and gender), we simultaneously analyzed them using a 2-way ANOVA. A significant interaction between previous suicide attempts and gender was found in ChatGPT-3.5, $F_{1,79} = 14.06$, $P < .01$. It was found that previous suicide attempts exacerbated the risk for men but did not significantly change the risk for women. In addition, no significant interaction was observed between previous suicide attempts and gender in ChatGPT-4 ($P > .05$).

## Differences Between ChatGPT-3.5 and ChatGPT-4

Figure 4 demonstrates that ChatGPT-3.5 evaluated the severity of all of the study's dependent variables (ie, the likelihood of serious suicide attempts, suicide attempts, and suicidal thoughts) as significantly higher than did ChatGPT-4, $F_{1,159} = 35.05$, $P < .001$; $F_{1,159} = 34.70$, $P < .001$; $F_{1,159} = 24.20$, $P < .001$, respectively.

## DISCUSSION

### Likelihood of Serious Suicide Attempt

Both ChatGPT-3.5 and ChatGPT-4 accorded significance to prior suicide attempts when assessing the likelihood of a serious suicide attempt, a finding that aligns with findings from prior research.[6–8] The models' ability to capture the enduring impact of past behaviors on future risk underscores their potential in predicting serious suicidal tendencies

Regarding gender, a significant gender-related effect was evident in ChatGPT-4, whereas ChatGPT-3.5 did not exhibit the same distinction. This discrepancy points to ChatGPT-4's heightened capacity for assessing gender-informed risk. Specifically, ChatGPT-4 identified that men had a higher likelihood of making serious suicide attempts than did women. This finding aligns with established gender theories that highlight varying patterns of suicide risk based on gender.[16,35]

The analysis of the interplay between gender and past suicide attempts via ChatGPT-4 revealed that the impact of previous suicide attempts on serious suicide risk appeared to be more significant for women than for men. It is thus possible that ChatGPT-4 took into consideration the well-documented phenomenon that, whereas men are more likely to complete suicide, suicide attempts tend to be more prevalent among women.[37] This nuanced perspective could contribute to the model's identification of the interaction. However, this finding contrasts with established clinical and empirical knowledge, which typically suggests that men exhibit higher rates of completed suicide than do women.[13] This difference highlights that although ChatGPT-4 may be able to account for individual risk factors such as gender or previous suicide experiences, its assessment of the combined effect of these 2 factors may be inaccurate.

### Likelihood of Suicide Attempt

With regard to this outcome, a significant impact of prior suicide attempts was detected solely within the predictions of ChatGPT-3.5, whereas ChatGPT-4 did not exhibit the same effect. This outcome is notable as

## Comparison Between ChatGPT-3.5 and ChatGPT-4 on the Likelihood of Serious Suicide Attempt, Suicide Attempt, and Suicide Thoughts (Mean ± SEM)
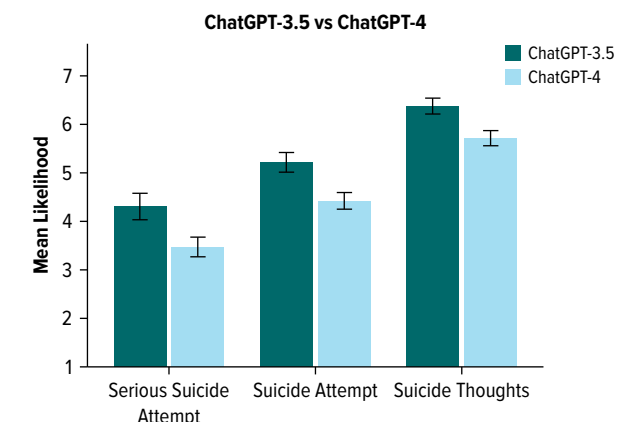


ChatGPT-4 is renowned for its enhanced linguistic diversity and capabilities,[26,27] presumably leading to a more comprehensive consideration of relevant variables. Thus, it was anticipated that ChatGPT-4 would also factor in the influence of previous suicide attempts within this outcome. Nevertheless, consistent with findings in the literature,[38] ChatGPT-4 revealed a discernible link between gender and the probability of a suicide attempt, whereas this association was not found in the results obtained from ChatGPT-3.5.

With regard to this outcome, the interaction patterns identified by ChatGPT-3.5 and ChatGPT-4 contradicted one another. In ChatGPT-3.5, the exacerbating effect of previous suicide attempts on the likelihood of making a suicide attempt was found to be more pronounced for men, whereas in ChatGPT-4, the same interaction intensified this risk for women. This finding stands in contrast with findings from previous studies in the area of medical issues, which showed that ChatGPT-4 performed significantly better than did ChatGPT-3.5 (eg,[39]). Specifically, a recent study which evaluated ChatGPT's ability to assess suicide risk found that ChatGPT-4 more accurately evaluated suicide risk than did ChatGPT-3.5.[2] The divergence in findings may have arisen from differences in the study design. In the current research, we directed ChatGPT to function as a suicide specialist, whereas in the prior study, ChatGPT's assessments were compared with those of mental health professionals.[2] This methodological variation could underlie the conflicting outcomes. Additionally, the variation in vignettes and the inclusion of variables such as previous suicide attempts, gender, and age in our study could also have contributed to the observed differences.

## Likelihood of Suicidal Thoughts

The results of both ChatGPT-3.5 and ChatGPT-4 underscore a distinct connection between past suicide attempts and the propensity for suicidal thoughts, aligning with patterns observed in the outcomes concerning the "likelihood of serious suicide attempt." This consistency across both versions reinforces the link between prior suicide attempts and the heightened potential for suicidal thoughts.[40] However, a noteworthy distinction emerged in the impact of gender between the 2 versions. Specifically, ChatGPT-4 revealed a significant gender-related effect, whereas this effect was not evident in ChatGPT-3.5. ChatGPT-4's evaluation indicated that men exhibited a greater likelihood of experiencing suicidal thoughts than did women. This finding stands in contrast with previous research indicating a greater preoccupation with thoughts of suicide among women than among men.[41] In ChatGPT-3.5, the gender variable did not emerge as significant, whereas ChatGPT-4 demonstrated that men carry a higher risk of having suicidal thoughts than do women.

A comparison of ChatGPT-3.5 with ChatGPT-4 reveals an interesting distinction. ChatGPT-3.5 displayed a significant interaction between past suicide attempts and gender, aligning with the existing literature by showing an increased risk for men in comparison to women.[13] However, ChatGPT-4 did not replicate this effect. This finding suggests a change in how ChatGPT-4 processes this interaction, potentially influenced by its advanced design or training data. It is important to recognize that AI models have limitations, and the differences between models underscore the complexity of predicting human behavior. These insights should be applied cautiously, supplementing human expertise in sensitive areas such as suicide risk assessment.

The study's results indicate that neither ChatGPT-3.5 nor ChatGPT-4 found a significant connection between age and the probability of carrying out serious suicide attempts, suicide attempts, or having suicidal thoughts. This finding contradicts well-established insights from clinical observations and prior research, which have consistently emphasized age as a significant risk factor in severe suicidal behaviors and suicide-related fatalities.[3,16] Despite the acknowledged relevance of age, the models omit its consideration in their analyses. This result can be attributed to the models' deliberate exclusion of age to mitigate the potential for biases or age-based discrimination, aligning with their overarching goal of ensuring unbiased and equitable evaluations.[42] However, this aspect does highlight a limitation of the models, as they fail to account for the importance of age in their assessments. However, it is possible to train a specialized model to incorporate age differences by using data from recent research on suicidal behaviors across various age groups.[43,44]

This study extends previous research on LLMs in mental health and specific suicidality assessments,[1,2,33,34] demonstrating that LLMs accurately integrate and reflect known risk factors in their analyses. By doing so, it moves beyond treating AI's outputs as an inscrutable "black box," allowing for an empirical examination of LLMs' decision-making processes in light of literature-established risk factors. This approach not only validates the ability of advanced neural networks to apply academic concepts in a nuanced manner but also opens a new methodological avenue for scrutinizing the "thought" mechanisms of these models against established domain knowledge and standards. The findings underscore the sophisticated capacity of AI in clinical implications, reinforcing the role of LLMs in enhancing our understanding of mental health assessments.

This study faces limitations, including the simplified nature of vignettes that may not fully reflect real-world complexities in suicide risk evaluation. Future research should include more nuanced scenarios or real clinical cases. Additionally, focusing mainly on variables such as prior suicide attempts, gender, and age overlooks other crucial factors, suggesting the need for incorporating a wider array of influences in future studies. The study's cross-sectional design also limits insights into how risk assessments change over time; longitudinal studies could provide deeper understanding. Furthermore, the opaque decision-making of AI complicates the interpretation of its predictions, calling for methods to increase model transparency. In addition, the study's context does not capture the complexity of applying AI in clinical settings, indicating a need for research on the practical integration of AI tools like ChatGPT in mental health care. Lastly, a better study design would have been to develop de-identified vignettes from a publicly available database of completed suicides and then compare the risk predictions from the LLMs. Knowing the outcomes in advance would allow for better assertions regarding the predictive validity of the models.

## CONCLUSIONS

This study serves as a foundational step, highlighting both the potential and limitations of AI in suicide risk assessment. Specifically the study explored how ChatGPT-3.5 and ChatGPT-4 assess suicide risk factors such as previous attempts, gender, and age, focusing on their predictions about different types of suicidal behavior. Findings show both models recognize previous attempts as a risk factor but differ in gender sensitivity, with ChatGPT-4 showing greater awareness than ChatGPT-3.5. Age's influence, however, was largely ignored by both models. These differences highlight the evolving nature of AI in understanding complex human

behaviors and the necessity of ongoing research to refine these models. While offering insights into AI's potential in suicide risk assessment, the study underlines the importance of addressing its limitations and the challenges of applying these findings in real-world settings.

## References

1. Elyoseph Z, Levkovich I. Beyond human expertise: the promise and limitations of ChatGPT in suicide risk assessment. *Front Psychiatry*. 2023;14:1213141.

2. Levkovich I, Elyoseph Z. Suicide risk assessments through the eyes of ChatGPT-3.5 versus ChatGPT-4: vignette study. *JMIR Ment Health*. 2023;10: e51232þ.

3. World Health Organization (WHO). Suicide worldwide in 2019: global health estimates. 2021. Accessed June 15, 2024. https://www.who.int/publications/i/item/9789240026643

4. Stone DM, Jones CM, Mack KA. Changes in suicide rates - United States, 2018-2019. *MMWR Morb Mortal Wkly Rep*. 2021;70(8):261–268.

5. Zalsman G. [Suicide: epidemiology, etiology, treatment and prevention]. *Harefuah*. 2019;158(7):468–472.

6. Li Y, Guo Z, Tian W, et al. An investigation of the relationships between suicidal ideation, psychache, and meaning in life using network analysis. *BMC Psychiatry*. 2023;23(1):257.

7. Gvion Y, Levi-Belz Y. Serious suicide attempts: systematic review of psychological risk factors. *Front Psychiatry*. 2018;9:56.

8. Fazel S, Runeson B. Suicide. *N Engl J Med*. 2020;382(3):266–274.

9. Demesmaeker A, Chazard E, Hoang A, et al. Suicide mortality after a nonfatal suicide attempt: a systematic review and meta-analysis. *Aust N Z J Psychiatry*. 2022;56(6):603–616.

10. Fehling KB, Selby EA. Suicide in DSM-5: current evidence for the proposed suicide behavior disorder and other possible improvements. *Front Psychiatry*. 2020;11:499980.

11. Eaton DK, Kann L, Kinchen S, et al; Centers for Disease Control and Prevention (CDC). Youth risk behavior surveillance - United States, 2011. *MMWR Surveill Summ*. 2012;61(4):1–162.

12. Schrijvers DL, Bollen J, Sabbe BGC. The gender paradox in suicidal behavior and its impact on the suicidal process. *J Affect Disord*. 2012;138(1–2):19–26.

13. Freeman A, Mergl R, Kohls E, et al. A cross-national study on gender differences in suicide intent. *BMC Psychiatry*. 2017;17(1):1–11.þ

14. World Health Organization. *Suicide worldwide in 2019: Global health estimates*. World Health Organization; 2021. Accessed June 15, 2024. https://www.who.int/publications/i/item/9789240026643.

15. Centers for Disease Control and Prevention. Suicide prevention: facts about suicide. 2023. Accessed June 15, 2024. https://www.cdc.gov/suicide/facts/index.html

16. Stone DM, Mack KA, Qualters J. Notes from the field: recent changes in suicide rates, by race and ethnicity and age group - United States, 2021. *MMWR Morb Mortal Wkly Rep*. 2023;72(6):160–162.

17. Wasserman D, Rihmer Z, Rujescu D, et al. The European Psychiatric Association (EPA) guidance on suicide treatment and prevention. *Eur Psychiatry*. 2012;27(2):129–141.

18. Bejan CA, Ripperger M, Wilimitis D, et al. Improving ascertainment of suicidal ideation and suicide attempt with natural language processing. *Sci Rep*. 2022;12(1):15146.

19. Vioules MJ, Moulahi B, Azé J, et al. Detection of suicide-related posts in Twitter data streams. *IBM J Res Dev*. 2018;62(1):7–1.

20. Ghanadian H, Nejadgholi I, Osman HA. ChatGPT for suicide risk assessment on social media: quantitative evaluation of model performance, potentials and limitations. Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis. Association for Computational Linguistics; 2023:172–183.

21. Tlili A, Shehata B, Adarkwah MA, et al. What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learn Environ*. 2023;15(23):1–24.

22. Farhat F, Sohail SS, Madsen DØ. How trustworthy is ChatGPT? The case of bibliometric analyses. *Cogent Eng*. 2023;10(1):2222988.

23. Surameery NMS, Shakor MY. Use Chat GPT to solve programming bugs. *Int J Inf Technol Comput Eng*. 2023;3(1):17–22.

24. Balas M, Ing EB. Conversational AI models for ophthalmic diagnosis: comparison of ChatGPT and the Isabel Pro differential diagnosis generator. *JFO Open Ophthalmol*. 2023;1:100005.

25. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLOS Digit Health*. 2023;2(2):e0000198.

26. He N, Yan Y, Wu Z, et al. Chat GPT-4 significantly surpasses GPT-3.5 in drug information queries. *J Telemed Telecare*. 2023:1357633X231181922. doi:10.1177/1357633X231181922

27. Nori H, King N, McKinney SM, et al. Capabilities of GPT-4 on medical challenge problems. *arXiv*. 2023:13375. doi:10.48550/arXiv.2303.13375

28. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI Chatbot for medicine. *N Engl J Med*. 2023;388(13):1233–1239.

29. Elyoseph Z, Hadar-Shoval D, Asraf K, et al. ChatGPT outperforms humans in emotional awareness evaluations. *Front Psychol*. 2023;14:1199058.

30. Singh OP. Artificial intelligence in the era of ChatGPT - opportunities and challenges in mental health care. *Indian J Psychiatry*. 2023;65(3):297–298.

31. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthc (Basel)*. 2023;11(6):887.

32. Imran N, Hashmi A, Imran A. Chat-GPT: opportunities and challenges in child mental healthcare. *Pak J Med Sci*. 2023;39(4):1191–1193.

33. Elyoseph Z, Refoua E, Asraf K, et al. Capacity of generative AI to interpret human emotions from visual and textual data: pilot evaluation study. *JMIR Ment Health*. 2024;11:e54369.

34. Hadar-Shoval D, Elyoseph Z, Lvovsky M. The plasticity of ChatGPT's mentalizing abilities: personalization for personality structures. *Front Psychiatry*. 2023;14:1234397.

35. Girard C. Age, gender, and suicide: a cross-national analysis. *Am Sociol Rev*. 1993;58(4):553–574.

36. Levi-Belz Y, Gamliel E. The effect of perceived burdensomeness and thwarted belongingness on therapists' assessment of patients' suicide risk. *Psychother Res*. 2016;26:436–445.

37. Beghi M, Butera E, Cerri CG, et al. Suicidal behaviour in older age: a systematic review of risk factors associated to suicide attempts and completed suicides. *Neurosci Biobehav Rev*. 2021;127:193–211.

38. Barrigon ML, Cegla-Schvartzman F. Sex, gender, and suicidal behavior. *Curr Top Behav Neurosci*. 2020;46:89–115.

39. Lewandowski M, Łukowicz P, Świetlik D, et al. ChatGPT-3.5 and ChatGPT-4 dermatological knowledge level based on the specialty certificate examination in dermatology. *Clin Exp Dermatol*. 2024;49(7):686–691.

40. de la Torre-Luque A, Pemau A, Ayad-Ahmed W, et al. Risk of suicide attempt repetition after an index attempt: a systematic review and meta-analysis. *Gen Hosp Psychiatry*. 2023;81:51–56.

41. Milton AC, Davenport TA, Iorfino F, et al. Suicidal thoughts and behaviors and their associations with transitional life events in men and women: findings from an international web-based sample. *JMIR Ment Health*. 2020;7(9):e18383.

42. Ferrara E. Should ChatGPT be biased? Challenges and risks of bias in large language models. *arXiv*. 2023. doi:10.5210/fm.v28i11.13346

43. Ivey-Stephenson AZ, Crosby AE, Hoenig JM, et al. Suicidal thoughts and behaviors among adults aged ≥18 Years - United States, 2015–2019. *MMWR Surveill Summ*. 2022;71(1):1–19.

44. Stone DM, Mack KA, Qualters J. Notes from the field: recent changes in suicide rates, by race and ethnicity and age group - United States, 2021. *MMWR Morb Mortal Wkly Rep*. 2023;72(6):160–162.