Family Medicine and Community Health

Identifying depression and its determinants upon initiating treatment: ChatGPT versus primary care physicians

Inbar Levkovich ,[1] Zohar Elyoseph[2,3]

To cite: Levkovich I, Elyoseph Z. Identifying depression and its determinants upon initiating treatment: ChatGPT versus primary care physicians. *Fam Med Com Health* 2023;**11**:e002391. doi:10.1136/fmch-2023-002391

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/fmch-2023-002391).

[1]Oranim Academic College, Tivon, Israel
[2]Department of Psychology and Educational Counseling, Max Stern Academic College Of Emek Yezreel, Emek Yezreel, Israel
[3]Department of Brain Sciences, Imperial College London, London, UK

Correspondence to
Inbar Levkovich;
inbar.lev2@gmail.com

ABSTRACT

Objective To compare evaluations of depressive episodes and suggested treatment protocols generated by Chat Generative Pretrained Transformer (ChatGPT)-3 and ChatGPT-4 with the recommendations of primary care physicians.

Methods Vignettes were input to the ChatGPT interface. These vignettes focused primarily on hypothetical patients with symptoms of depression during initial consultations. The creators of these vignettes meticulously designed eight distinct versions in which they systematically varied patient attributes (sex, socioeconomic status (blue collar worker or white collar worker) and depression severity (mild or severe)). Each variant was subsequently introduced into ChatGPT-3.5 and ChatGPT-4. Each vignette was repeated 10 times to ensure consistency and reliability of the ChatGPT responses.

Results For mild depression, ChatGPT-3.5 and ChatGPT-4 recommended psychotherapy in 95.0% and 97.5% of cases, respectively. Primary care physicians, however, recommended psychotherapy in only 4.3% of cases. For severe cases, ChatGPT favoured an approach that combined psychotherapy, while primary care physicians recommended a combined approach. The pharmacological recommendations of ChatGPT-3.5 and ChatGPT-4 showed a preference for exclusive use of antidepressants (74% and 68%, respectively), in contrast with primary care physicians, who typically recommended a mix of antidepressants and anxiolytics/hypnotics (67.4%). Unlike primary care physicians, ChatGPT showed no gender or socioeconomic biases in its recommendations.

Conclusion ChatGPT-3.5 and ChatGPT-4 aligned well with accepted guidelines for managing mild and severe depression, without showing the gender or socioeconomic biases observed among primary care physicians. Despite the suggested potential benefit of using atificial intelligence (AI) chatbots like ChatGPT to enhance clinical decision making, further research is needed to refine AI recommendations for severe cases and to consider potential risks and ethical issues.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Depression is a prevalent disorder for which people often first seek help from primary care physicians. Accurate diagnosis and treatment are crucial for continuity of care.

WHAT THIS STUDY ADDS

⇒ This research offers an innovative examination of artificial intelligence (AI), specifically ChatGPT, in the area of depression treatment. The findings reveal an absence of discernible gender and socioeconomic biases in ChatGPT. This critical revelation underscores the potential benefit offered by AI tools as impartial instruments in the management of depression. The multifaceted implications suggest that AI tools have the potential to play a pivotal role in healthcare decision making, particularly in depression treatment, with the potential to enhance care quality and patient outcomes. Furthermore, the lack of discernible biases in AI offers the potential for equitable application across diverse patient demographics. Nevertheless, the provided excepts do not capture the full depth and breadth of the study's findings and implications. A comprehensive review of the entire topic is necessary to fully appreciate the study's novel contributions and potential impact.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ The findings suggest that ChatGPT has the potential to significantly augment decision making, particularly through its adherence to established treatment protocols and its lack of the gender and socioeconomic biases that occasionally affect the decisions of primary care physicians. Concurrently, the study highlights the necessity for ongoing research to confirm the dependability of ChatGPT's suggestions and their adherence to clinical standards of care. Deployment of such AI systems has the potential to drive improvements in the quality and fairness of mental health services.

INTRODUCTION

ChatGPT (Chat Generative Pretrained Transformer), an advanced conversational artificial intelligence (AI) technology, is the product of the research and development endeavours of OpenAI, and is powered by this company's transformative GPT linguistic model.[1] Only a few months after being made publicly available, ChatGPT had already gathered an astounding constituency of 100 million users, marking it as the most rapidly expanding AI consumer application to date.[2] ChatGPT has broad competencies for dealing with

text-based inquiries ranging from basic requests to intricate demands. It is able to discern and decipher user input and formulate responses that mirror human conversation.[1 2] This remarkable capacity to produce human-like dialogue and perform complex functions marks ChatGPT as a pivotal advancement in the domains of natural language processing and AI.[1 2] ChatGPT can be used in the fields of education,[3] programming[4] and psychology.[4 5] However, its potential implications in applied psychological settings remain relatively unexplored.[6 7] Despite the theoretical potential of ChatGPT,[8] its potency in addressing critical clinical mental health challenges has yet to be definitively determined.[7] Most studies to date examining the role of ChatGPT in public health focus predominantly on assessing the merits and drawbacks, with very little empirical examination of this topic.[7 8] Furthermore, to the best of our knowledge, no studies have investigated the issue of depression in the context of ChatGPT, despite the high prevalence of this condition in the field of mental health.

Depression is a prevalent disorder[9 10] for which people often first seek help from primary care physicians. Accurate diagnosis and treatment are crucial for continuity of care.[11] The disorder is characterised by a multitude of symptoms, among them persistent melancholy, anhedonia, guilt-related sentiments or low self-esteem, irregularities in sleep patterns or appetite, perpetual fatigue and impaired concentration.[10 12] In its most severe state, depression can precipitate suicidal tendencies and augment mortality risk. Depression often follows a chronic trajectory, considerably diminishing the capacity for vocational productivity and degrading quality of life.[13] Mild depression is defined by the presence of symptoms that exceed the diagnostic minimum. Although such symptoms are distressing, they remain manageable for the individual, causing only minor hindrances to social or occupational functionality.[14]

For individuals with depression, primary care providers are often their initial point of interaction with the medical system, and are therefore instrumental in initiating appropriate therapeutic strategies.[14 15] Primary care physicians are often the first to identify depressive symptoms in patients, either initiating treatment or referring them to specialists as per guidelines. Accurate diagnosis and treatment are essential for continuous care.[15] The longstanding relationship between family physicians and patients enhances therapeutic recovery through mutual trust, and facilitates easier access to patients.[16] The course of treatment selected by primary care providers is largely guided by clinical recommendations. These recommendations are crafted by authoritative organisations and offer evidence-informed guidelines for the diagnosis and management of major depressive disorder.[17–19] The guidelines conventionally suggest a tiered-care approach, commencing with minimally invasive interventions such as psychoeducation and vigilant observation for less severe cases, and escalating to psychological therapies and pharmacological interventions for moderate to severe manifestations of depression.[17] By adhering to these guidelines, primary care providers are able to provide standardised and superior care, thereby mitigating inconsistency in treatment outcomes.[14] Various patient attributes, among them depression severity, concurrent medical conditions, antecedent treatment records and individual inclinations, significantly sway the therapeutic decisions of primary care providers.[20] For instance, individuals who manifest symptoms of severe depression or are refractory to preliminary interventions may require more aggressive therapeutic approaches, including a combination of antidepressants and psychotherapy.[21] Additionally, individual preferences can steer primary care providers toward specific therapies.

Numerous studies have demonstrated the efficacy of both psychotherapy and antidepressant medications in treating depression.[21 22] Treatment guidelines for managing depression recommend continuous use of antidepressants for several months to reach and maintain remission.[17 18 22] Yet, only around half of those affected receive satisfactory treatment.[23] This discrepancy in treatment provision can be ascribed to several factors, among them the failure to diagnose depression in primary care settings, the scarcity of outpatient psychotherapists and therapists' insufficient adherence to evidence-based interventions.[24] As a consequence of the limited availability of outpatient psychotherapists, patients exhibiting mild to moderate depressive symptoms are frequently prescribed antidepressant medication only, a practice that contradicts treatment guideline recommendations.[12]

The positive predictive value of the depression diagnoses of primary care physicians was only 42%, suggesting that 58% of identified cases were false positives.[25] From this we can infer that the majority of antidepressant prescriptions written by primary care physicians are for patients manifesting mild depression, including those with subthreshold symptoms.[26] The majority (60–85%) of these prescriptions are for treating depression in adults,[27] while a minority are prescribed for other conditions. An estimated 5–16% of adults in Europe and the US are prescribed antidepressants each year.[28 29]

In making clinical decisions, primary healthcare practitioners are significantly affected by their competence and education.[30] Those with extensive professional experience are likely to be more confident in diagnosing depression and commencing treatment, in contrast with less experienced providers, who may often delegate patients to specialists.[31] Numerous factors may exert a detrimental influence on compliance with the guidelines governing depression management, among which are factors associated with patients, professionals, physicians and healthcare organisations. Patients often deny experiencing symptoms of depression, opting instead to seek help for physical manifestations of distress in a general practice setting.[14 15 20] A psychiatric diagnosis does not necessarily ensure that either the patient or the doctor will perceive that treatment is necessary.[26] Primary care physicians may have trouble conforming to these guidelines because

they are unable to distinguish between typical distress and bona fide anxiety or depressive disorders. Furthermore, some primary care physicians may have difficulty discussing aspects pertinent to this diagnosis with their patients.[27 28 31] Finally, organisations may be marked by insufficient collaboration between primary care physicians and mental health experts, long waiting lists for specialised mental health services and inadequate financial incentives.[11 28 29]

While patient diversity undeniably plays a crucial role in the variance in treatment outcomes, a growing body of evidence underscores the significance of differences in physician decision-making processes. For instance, Cutler *et al*[32] pointed to substantial disparities in physicians' beliefs regarding the optimal course of treatment. Berndt *et al*[33] referred to survey data indicating that most physicians exhibit a preference for a certain drug, with this favoured drug accounting for an average of 66% of their overall prescriptions.

The perspectives of primary care physicians can potentially affect patients' willingness to disclose their issues, thus playing a pivotal role in their initial identification of issues during the clinical encounter, their subsequent therapeutic decision making and their readiness to embrace novel approaches in their clinical practice.[30] Clinicians' beliefs concerning mental illness may encompass prevalent societal stereotypes and misconceptions, coupled with their own personal experiences and professional training.[34] A lack of specificity in screening for depression among primary care physicians can result in misdiagnoses, such that individuals without depression are incorrectly identified as having this mental health disorder.[35] Such stereotyping is seen more commonly in misdiagnosed depression among female, elderly and racial minority patients.[36 37] Such overdiagnoses often cause an unnecessary treatment burden on the healthcare system, resulting in wasted resources and misuse of health services. Overdiagnosis fosters undue reliance on community services and benefits by healthy individuals erroneously diagnosed with depression, along with potential iatrogenic effects of unneeded treatments.[32 33]

Ample research has explored the implicit racial/ethnic, sociodemographic and gender biases among primary care physicians.[36 37] For instance, a study conducted among primary care physicians in Brazil revealed a disproportionate prevalence of mental problems among female patients, unemployed patients, those with limited education and those earning lower incomes.[38] Patients with higher socioeconomic status (SES) were more likely to engage in conversations with their physicians.[39] White and Stubblefield-Tave[40] showed that women, regardless of their backgrounds, are subjected to unequal treatment. Finally, Ballering *et al*[41] found that men who consulted physicians about common somatic symptoms were more likely to be examined physically and sent for diagnostic imaging and specialist referrals than women with similar complaints. These findings and others point to the necessity to address and counter power imbalances in the clinical relationship.[36–41]

ChatGPT offers several advantages over primary care physicians and even mental health professionals in detecting depression. From the outset, ChatGPT has the potential to offer objective, data-derived insights that can supplement traditional diagnostic methods.[7] Moreover, ChatGPT is capable of analysing extensive data rapidly, facilitating early detection and intervention. Finally, it can offer confidentiality and anonymity, thus potentially encouraging patients to seek assistance without fear of stigma or professional consequences.

Study objectives

This study seeks to compare depressive episode evaluations and suggested treatment protocols generated by ChatGPT-3.5 and ChatGPT-4 with those of primary care physicians. Specifically, the study will investigate:

1. Adjustment of treatment protocols for both mild and severe depression.
2. Modification of pharmacological treatments as required in specific cases.
3. Scrutiny and handling of gender or socioeconomic biases.

METHODS

AI procedure

Using ChatGPT versions 3.5 and 4 (OpenAI, San Francisco, CA, USA), we conducted three evaluations during the month of June 2023 (ChatGPT 24 May version). We sought to examine how ChatGPT evaluates the preferred therapeutic approach for mild and severe major depression (MD) and to determine whether the therapeutic approach advocated by ChatGPT is influenced by gender or socioeconomic biases compared with the performance of human medical practitioners.

Input source

Vignettes taken from Dumesnil *et al*[30] were input to the ChatGPT interface. These case vignettes centred around patients seeking initial consultation for symptoms of sadness, sleep problems and loss of appetite during the past 3 weeks and receiving a diagnosis of MD. Eight versions of the case vignettes were developed, in which patient characteristics such as gender, SES (blue collar/white collar worker) and MD severity (mild/severe) were varied (online supplemental appendix 1). Each of the eight vignettes was introduced to ChatGPT-3.5 and ChatGPT-4 via a new tub. Each vignette was repeated 10 times per chat to ensure consistency and reliability.

Measures

After each of the eight vignettes (man/woman; blue collar worker/white collar worker; mild/severe depression), we asked ChatGPT the following question: *What do you think a primary care physician should suggest in this situation?* The possible responses were watchful waiting; referral for psychotherapy; prescribing pharmacological

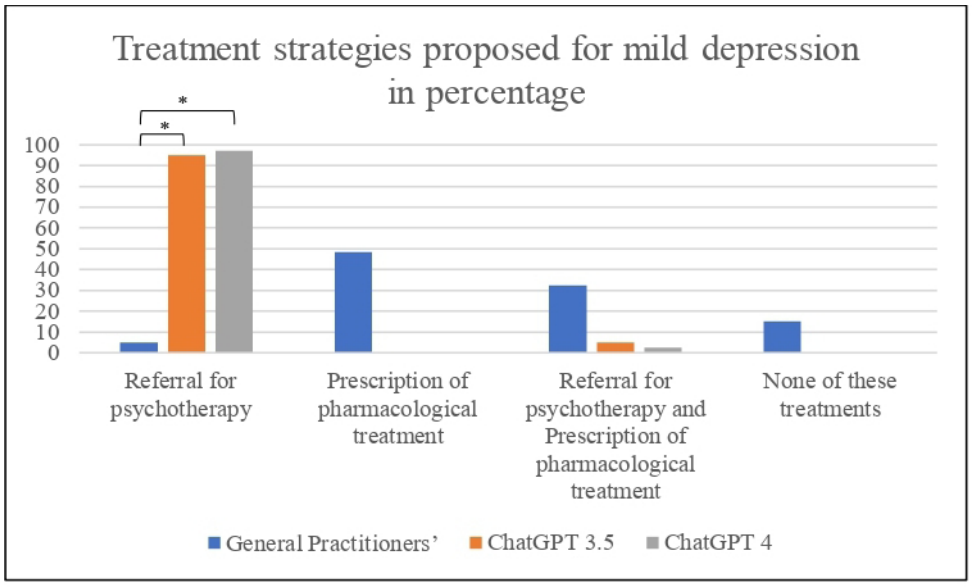

Figure 1 Treatment strategies for mild depression proposed by primary care physicians, ChatGPT-3.5 and ChatGPT-4. *p<0.001

treatment (not including psychotherapy); referral for psychotherapy and prescribing pharmacological treatment; and none of these treatments. A choice of pharmacological treatment prompted ChatGPT to select one of the following options: antidepressant; anxiolytic/hypnotic; antidepressant and anxiolytic/hypnotic; and none of these drugs.

Scoring

The performance of ChatGPT was scored according to Dumesnil *et al*.[30] We then compared the ChatGPT scores to the norms of 1249 participating primary care physicians (72.9% female).

Statistical analysis

Data were presented in the form of a frequency table. Due to the categorical nature of the data and the relatively constant answers of ChatGPT, the frequency of ChatGPT's most common answer was compared with the sample of primary care physicians using two-tailed χ^2 test in SPSS version 27.

RESULTS

Depression severity

Figures 1 and 2 illustrate the therapeutic approaches recommended by primary care physicians, ChatGPT-3.5

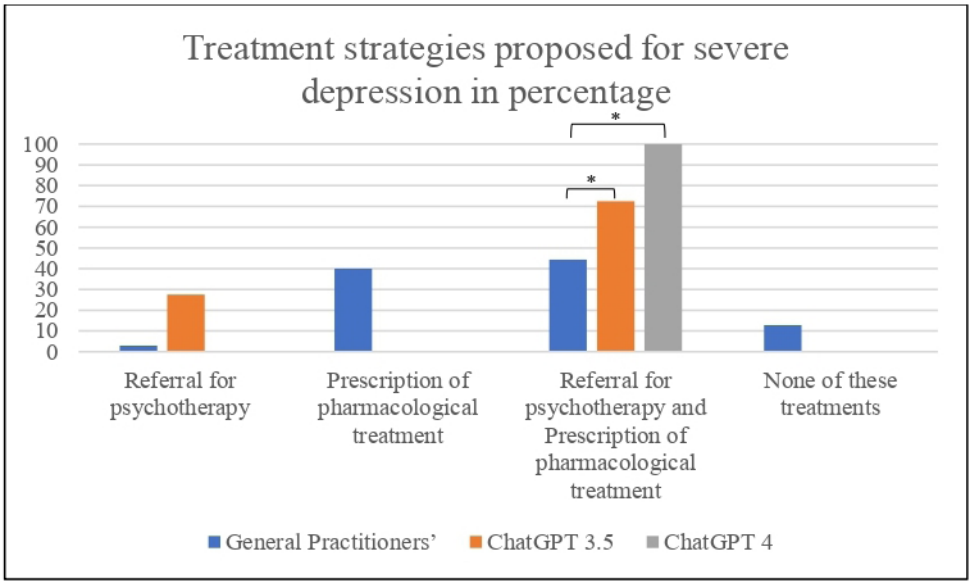

Figure 2 Treatment strategies for severe depression proposed by primary care physicians, ChatGPT-3.5 and ChatGPT-4. *p<0.001

and ChatGPT-4 for mild and severe MD. Among the primary care physicians, only 4.3% exclusively recommended 'referral for psychotherapy' for mild cases.[30] ChatGPT-3.5 and ChatGPT-4, in contrast, exclusively recommended 'referral for psychotherapy' in 95.0% and 97.5% of cases, respectively. The differences were found to be significant for primary care physicians versus ChatGPT-3.5 (χ^2 (1, 1290) = 449.23, p<0.001) and for primary care physicians versus ChatGPT-4 (χ^2 (1, 1290) = 470.23, p<0.001). The majority of the participating primary care physicians proposed either 'prescription of pharmacological treatment' exclusively (48.3%) or 'referral for psychotherapy' together with 'prescription of pharmacological treatment' (32.5%).

In severe cases, the majority of primary care physicians proposed 'referral for psychotherapy and prescription of pharmacological treatment' (44.4%). Nevertheless, ChatGPT proposed this more frequently than the primary care physicians: ChatGPT-3.5 (72%) (χ^2 (1, 1290) = 12.31, p<0.001) and ChatGPT-4 (100%) (χ^2 (1, 1290)= 48.14, p<0.001). In addition, 40% of the primary care physicians proposed 'prescription of pharmacological treatment' exclusively, a recommendation that was not made by ChatGPT-3.5 and ChatGPT-4.

Gender bias

Dumesnil *et al*[30] found that primary care physicians prescribe antidepressants significantly less often to women than to men. Our results showed that ChatGPT-3.5 and ChatGPT-4 exhibit no significant differences in therapeutic approach between women and men (p=0.25 and p=0.82 for ChatGPT-3.5 and ChatGPT-4, respectively).

Socioeconomic bias

The study conducted by Dumesnil *et al*[30] found that primary care physicians commonly recommend antidepressant medication without psychotherapy to blue collar workers, and a combination of antidepressant drugs and psychotherapy to white collar workers. Our results showed that ChatGPT-3.5 and ChatGPT-4 exhibit no significant differences in therapeutic approach between blue collar and white collar workers (p=0.49 and p=0.82 for ChatGPT-3.5 and ChatGPT-4, respectively).

Psychopharmacology

In cases in which primary care physicians or ChatGPT proposed psychopharmacology (either exclusively or alongside psychotherapy), a follow-up question on the specific recommended drugs was asked. Figure 3 illustrates the proposed pharmacological approach for the treatment of depression as recommended by primary care physicians. The findings indicate that primary care physicians recommend a combination of antidepressants and anxiolytic/hypnotic medications in 67.4% of cases, exclusive use of antidepressants in 17.7% of cases and exclusive use of anxiolytic/hypnotic drugs in 14.0% of cases.

In contrast, the recommendations of ChatGPT-3.5 and ChatGPT-4 show significant variation. The proportion of cases in which exclusive antidepressant treatment is advised is higher for both ChatGPT-3.5 (74%) and ChatGPT-4 (68%) than for the primary care physicians' recommendations. Additionally, ChatGPT-3.5 (26%) and ChatGPT-4 (32%) suggested using a combination of antidepressants and anxiolytic/hypnotic drugs more frequently than did primary care physicians. Note that neither ChatGPT-3.5 nor ChatGPT-4 proposed exclusive use of anxiolytic/hypnotic drugs, in contrast with the recommendations of the primary care physicians.

The two versions of ChatGPT exhibited fairly similar performance in most cases. An important exception was observed in cases of severe depression. ChatGPT-3.5

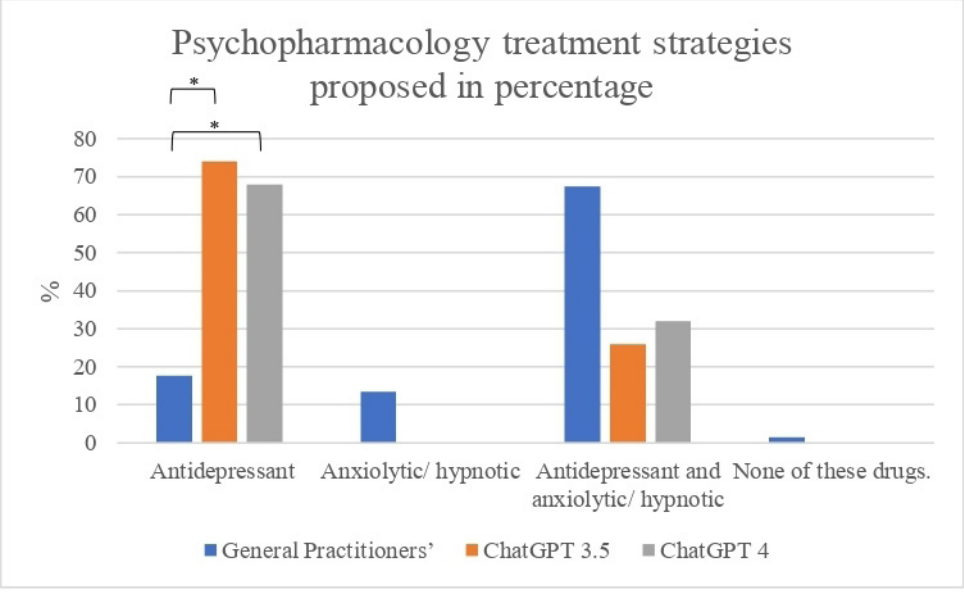

Figure 3 Psychopharmacology treatment strategies proposed by primary care physicians, ChatGPT-3.5 and ChatGPT-4 (%). *p<0.001

recommended a combination of pharmacological treatment and psychotherapy in only 72% of cases, recommending psychotherapy alone without pharmacological treatment in the remaining 28%. In contrast, the more advanced ChatGPT-4 version recommended a combination of pharmacological treatment and psychotherapy in 100% of cases.

DISCUSSION

The results of the current study showed that the therapeutic proposals of ChatGPT are in line with the accepted guidelines for mild and severe MD treatment. Moreover, unlike the treatments proposed by primary care physicians, ChatGPT's therapeutic recommendations are not tainted by gender or SES biases. Accordingly, ChatGPT has the potential to improve primary care physicians' decision making in treating depression.

In the mild depression condition, ChatGPT-3.5 and ChatGPT-4 exclusively recommended 'referral for psychotherapy' in 95.0% and 97.5% of cases, compared with only 4.3% of primary care physicians, who exclusively recommended 'prescription of pharmacological treatment' (48.3%) or a combination of 'referral for psychotherapy and prescription of pharmacological treatment' (32.5%).[30] Unlike the recommendations of the primary care physicians, the therapeutic proposals of ChatGPT are in line with the accepted guidelines for managing mild depression.[14] For mild depression, a 2-week psychotherapy review may suffice.[16] However, if there is no improvement, the patient does not qualify for psychotherapy, or symptoms are moderate to severe, antidepressants should be considered. Severe or prolonged symptoms further warrant pharmacological intervention.[16] Given the multiple studies indicating that medical doctors tend to recommend psychiatric drugs over psychotherapy in mild cases of MD,[26 27] our results point to the inherent potential of AI chatbots for improving clinical decision making in mild cases of depression.[8] In survey-based evaluations, primary care physicians expressed positive views of psychotherapy and its efficacy in addressing depressive disorders. The indications propose that primary care physicians in France[1] predominantly lean towards initiating immediate active treatment for MD, irrespective of the severity of the condition. This propensity may be driven by their perceived efficacy in managing this disorder and their discontentment with the quality of the accessibility and collaboration extended by mental health specialists.[18 24 30] Yet this recognition did not translate into practice, especially for patients presenting mild to moderate depression. More often, the preferred course of action consisted of prescribing antidepressant medications in lieu of recommending psychotherapeutic interventions.[30]

In the severe depression condition, ChatGPT-3.5 and ChatGPT-4 proposed 'referral for psychotherapy and prescription of pharmacological treatment' in 72% and 100% of cases, respectively, compared with 44.4% of primary care physicians in France. Indeed, more than 50% of the French primary care physicians did not propose psychotherapy at all, either with or without drugs.[30] The approach adopted by primary care physicians in managing depression is dominated by a pharmacological paradigm, with antidepressants serving as the principal therapeutic strategy, even in cases of mild to moderate depression.[25] A significant proportion of these physicians hold an overly optimistic perspective on pharmaceutical treatments, often assuming that their development and effectiveness has progressed without interruption.[30] Moreover, primary care physicians tend to downplay potential risks and side effects of pharmacological treatments. Their practice is distinguished by an extraordinarily high prescription rate. In France, 90% of patient visits to primary care physicians culminate in a prescription. This rate is 72% in Germany and 43% in the Netherlands.[42] Again, the therapeutic recommendations of ChatGPT were in line with the accepted guidelines for managing severe depression.[14]

Many researchers have discussed the role of ChatGPT in mental health.[7 21 22 43 44] Considering the tendency among primary care physicians not to refer patients with severe depression for psychotherapy but rather to prescribe medication as the exclusive treatment,[21] our results again point to the potential of AI chatbots to improve clinical decision making in cases of severe depression as well.[8] Nevertheless, it should be noted that in 28% of the severe cases, ChatGPT-3.5 proposed 'referral for psychotherapy' exclusively without any pharmacology treatment. This recommendation does not appear to be compatible with the accepted guidelines, underlying the importance of further research and development before making clinical use of these tools.

Furthermore, it is important to note that the work of primary care physicians is significantly constrained by such limitations as their high patient load, the wide array of medical conditions and presenting symptoms they encounter and various manifestations of depressive disorders across all types, levels of severity and comorbidity patterns. The limited amount of time allocated to each patient during a typical day likely represents a substantial impediment to enhancing both identification and intervention, whether it involves treatment or referral.[13]

Comparison of the psychopharmacological treatment recommendations provided by primary care physicians with those of the AI models ChatGPT-3.5 and ChatGPT-4 reveals several notable differences. Primary care physicians commonly advised a combination of antidepressant medications along with anxiolytic/hypnotic medications in 67.4% of cases, whereas in 14.0% of cases they recommended exclusive use of anxiolytic/hypnotic medications. Only in 17.7% of cases did primary care physicians suggest exclusive use of antidepressants. In contrast, ChatGPT-3.5 recommended exclusive antidepressant treatment in 74% of cases, and ChatGPT-4 made the same recommendation in 68% of cases. In the remaining cases, the AI models proposed a combination of antidepressants and

anxiolytic/hypnotic medications (26% and 32% of cases for ChatGPT-3.5 and ChatGPT-4, respectively). Neither ChatGPT-3.5 nor ChatGPT-4 proposed using anxiolytic/hypnotic medications exclusively. Although anxiolytic/hypnotic drugs are not recommended as a first line of treatment for depression and are considered addictive,[45] multiple studies have reported on doctors' mistaken tendency to prescribe these drugs for treating depression.[46 47] While ChatGPT occasionally recommended anxiolytic/hypnotic treatment alongside antidepressant drugs, these recommendations were considerably less frequent than among primary care physicians. In over 80% of cases, primary care physicians recommended using anxiolytic/hypnotic drugs exclusively or in conjunction with antidepressants. These findings imply that AI may also have the potential to contribute to medical decision making regarding psychopharmacological recommendations.

With respect to gender or SES biases in therapeutic recommendations for mild or severe MD, the results show that ChatGPT was not influenced by gender or SES bias. In view of multiple studies reporting gender and SES biases among primary care physicians in diagnosing and treating MD, this finding offers great potential.[36–38] Indeed, the ability to adjust treatment without 'falling into the trap' of gender or SES bias has the potential to promote both quality and equity in mental healthcare. Although studies on this topic highlight several ethical concerns, particularly with respect to the ability of AI models to perpetuate biases and discrimination, the current results counter these concerns[48 49] by showing that ChatGPT does not necessarily amplify gender and SES biases. This counters the widespread notion that such biases could lead to dissemination of inaccurate or discriminatory information, hence failing to represent diverse perspectives and experiences.[48 49] Such a scenario could indeed have severe implications for patient care. The discrepancy between these findings and those of previous research may be attributed to progress and advancements in iterations testing the clinical efficacy of ChatGPT. In particular, the current study utilised ChatGPT-4, suggesting that ongoing development and learning in these models may potentially mitigate some of the bias issues highlighted in earlier studies.

In summary, managing cases of MD by primary care physicians is a significant concern in mental health.[14 15] Previous studies reveal that despite enhancements in the accessibility of treatment by primary care physicians, their decision making is far from satisfactory, specifically under the following conditions:

- A limited frequency of referrals to psychotherapy.[11 28 29]
- A leaning towards pharmacological treatment even in instances of mild depression.[32 33]
- A tendency to inappropriately prescribe anxiolytic/hypnotic drugs for the treatment of depression.[27 28 31]
- A tendency to be biased by the patient's gender and SES when making medical recommendations.[36 37 40]

The results of this study show the potential uses of AI chatbots, specifically ChatGPT-3.5 and ChatGPT-4, in supporting the clinical decision making of primary care physicians in treating depression. The findings suggest that ChatGPT can make a significant contribution to enhancing decision making by adhering to accepted treatment guidelines and eliminating the gender and SES biases that are sometimes present in primary care physicians' decisions. Nevertheless, the study also acknowledges the need for further research and development to ensure the reliability and alignment of ChatGPT's recommendations with clinical best practices. Integrating such AI systems has the potential to promote quality and equity in mental healthcare.

Despite the substantial potential of using ChatGPT for detecting depression, this use also introduces a multitude of challenges.[2] The exactitude of ChatGPT's forecasts hinges on the quality and demographic representativeness of the data employed to train the algorithm.[3 4] Biased data or data that are not sufficiently representative can give rise to incorrect predictions or intensify extant health disparities. Furthermore, ChatGPT algorithms tend to operate as 'black boxes,' cloaking the rationale behind their predictive processes.[50] This lack of transparency can obstruct the cultivation of trust and acceptance among users.[50]

The use of ChatGPT as a decision making tool for depression entails numerous ethical quandaries.[38] Ensuring data privacy and security are of supreme importance, especially considering the sensitive nature of mental health data. Users must be provided explicit and comprehensive education about the usage and protective measures associated with their data.[7] Moreover, ChatGPT should not supersede human clinical judgement in the diagnosis or treatment of depression. It should instead serve as a supportive instrument that fortifies professionals in their attempt to make well informed clinical decisions.[7]

The current research has certain limitations that warrant acknowledgement. First, the study was limited to iterations of ChatGPT-3 and ChatGPT-4 at specific points in time, and it did not account for subsequent versions. Thus, future investigations are encouraged to address this matter by examining forthcoming updates. Second, the ChatGPT data were compared with data from a representative sample of primary care physicians from France, who obviously do not represent all primary care physicians worldwide. Third, the cases described in the vignettes pertain to an initial visit due to a complaint of depression, and do not depict ongoing and comprehensive treatment of the disease or other variables that the doctor would know about the patient. The vignette methodology does not guarantee that the elicited responses accurately mirror primary care physicians' actions within a real-world context. It fails to incorporate the intricate dynamics of doctor-patient interactions which significantly mould clinicians' behaviour. The vignettes are deficient in capturing the complex nuances and subtleties integral to comprehensive clinical presentations. Several vignettes might not fulfil the diagnostic criteria for major

depressive disorder owing to the absence of precise information pertaining to the duration of symptoms. Furthermore, the recommendation for psychotherapy may be influenced by additional factors such as patient preferences, financial limitations and logistical challenges associated with adhering to regular psychotherapy sessions, particularly for individuals engaged in work-related responsibilities.

We recommend that decision makers develop a comprehensive plan that includes training for physicians on detecting depression, optimising treatment duration, and enhancing their proficiency, knowledge and confidence in suggesting treatment options. The programme should also provide information to family doctors about referrals to mental health services, tailored to the patient's specific condition and needs. Additionally, resource allocation enhancements are necessary to mitigate issues like limited time and excessive workloads, which impede diagnosis and treatment. In some countries, ensuring easier access to experts and reducing waiting times are also crucial components of the plan.

For subsequent investigations, we recommend that follow-up studies examine additional AI languages within vignettes, focusing on sample inputs from physicians worldwide and across diverse timeframes and points. Further research could also aim to predict the most effective antidepressant for use.

CONCLUSION

Observations revealed differences between ChatGPT, and the norms used by primary care physicians in identifying depression severity and determining how it should be treated. Specifically, ChatGPT-4 demonstrated greater precision in adjusting treatment to comply with clinical guidelines. Furthermore, no discernible biases related to gender and SES were detected in the ChatGPT systems. The study suggests that ChatGPT, with its commitment to treatment protocols and absence of biases, has the potential to enhance decision making in primary healthcare. However, it underlines the need for ongoing research to verify the dependability of its suggestions. Implementing such AI systems could bolster the quality and impartiality of mental health services.

Contributors IL and ZE made a significant contribution to the work reported, in the conception, study design, took part in drafting, revising and critically reviewing the article, and gave final approval of the version to be published. ZE handled the acquisition of data and analysed the data. IL wrote the first draft of the manuscript, and she is the author responsible for the overall content as the guarantor.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, conduct, reporting or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request.

ORCID iD

Inbar Levkovich http://orcid.org/0000-0002-5717-4074

REFERENCES

1 Dwivedi YK, Kshetri N, Hughes L, *et al*. So what if Chatgpt wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of Generative conversational AI for research, practice and policy. *International Journal of Information Management* 2023;71:102642.
2 Paul J, Ueno A, Dennis C. Chatgpt and consumers: benefits, pitfalls and future research agenda. *Int J Consumer Studies* 2023;47:1213–25.
3 Cooper G. Examining science education in Chatgpt: an exploratory study of Generative artificial intelligence. *J Sci Educ Technol* 2023;32:444–52. 10.1007/s10956-023-10039-y Available: https://doi.org/10.1007/s10956-023-10039-y
4 Kashefi A, Mukerji T. Chatgpt for programming numerical methods. *J Mach Learn Model Comput* 2023;4.
5 Uludag K. n.d. The use of AI-supported chatbot in psychology. *SSRN Journal* 10.2139/ssrn.4331367 Available: https://ssrn.com/abstract=4331367 or http://dx.doi.org/10.2139/ssrn.4331367
6 Sallam M. Chatgpt utility in Healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare (Basel)* 2023;11:887.
7 Elyoseph LI. ""Beyond human expertise": Chatgpt's the promise and limitations of Chatgpt in suicide prevention". *Front Psychiatry* 2023.
8 Biswas SS. Role of chat GPT in public health. *Ann Biomed Eng* 2023;51:868–9.
9 Gutiérrez-Rojas L, Porras-Segovia A, Dunne H, *et al*. Prevalence and correlates of major depressive disorder: a systematic review. *Braz J Psychiatry* 2020;42:657–72.
10 Hosseinian-Far A,N, Jalali R, Vaisi-Raygani A, *et al*. Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: a systematic review and meta-analysis. *Global Health* 2020.
11 Houle J, Villaggi B, Beaulieu M-D, *et al*. Treatment preferences in patients with first episode depression. *J Affect Disord* 2013;147:94–100.
12 Wang J, Wu X, Lai W, *et al*. Prevalence of depression and depressive symptoms among outpatients: a systematic review and meta-analysis. *BMJ Open* 2017;7:e017173.
13 LeMoult J, Gotlib IH. Depression: A cognitive perspective. *Clin Psychol Rev* 2019;69:51–66.
14 Park LT, Zarate CA. Depression in the primary care setting. *N Engl J Med* 2019;380:559–68.
15 Costantini L, Pasquarella C, Odone A, *et al*. Screening for depression in primary care with patient health Questionnaire-9 (PHQ-9): A systematic review. *J Affect Disord* 2021;279:S0165-0327(20)32828-7:473–83.:.
16 Ng CWM, How CH, Ng YP. Managing depression in primary care. *Singapore Med J* 2017;58:459–66.
17 Taylor RW, Marwood L, Oprea E, *et al*. Pharmacological augmentation in Unipolar depression: A guide to the guidelines. *Int J Neuropsychopharmacol* 2020;23:587–625.
18 Gautam S, Jain A, Gautam M, *et al*. Clinical practice guidelines for the management of depression. *Indian J Psychiatry* 2017;59(Suppl 1):S34–50.
19 Collaborating Centre for Mental Health, UK. Depression: the treatment and management of depression in adults. 2010.

20 Cohen ZD, DeRubeis RJ. Treatment selection in depression. *Annu Rev Clin Psychol* 2018;14:209–36.
21 Driessen E, Dekker JJM, Peen J, *et al*. The efficacy of adding short-term Psychodynamic psychotherapy to antidepressants in the treatment of depression: A systematic review and meta-analysis of individual participant data. *Clin Psychol Rev* 2020;80:101886.
22 Barth J, Munder T, Gerger H, *et al*. Comparative efficacy of seven psychotherapeutic interventions for patients with depression: a network meta-analysis. *Focus (Am Psychiatr Publ)* 2016;14:229–43.
23 Morris J, Lora A, McBain R, *et al*. Global mental health resources and services: a WHO survey of 184 countries. *Public Health Rev* 2012;34:1–19.
24 Emmelkamp PMG, David D, Beckers T, *et al*. Advancing psychotherapy and Evidence-Based psychological interventions. *Int J Methods Psychiatr Res* 2014;23 Suppl 1(Suppl 1):58–91.
25 Mojtabai R. Clinician-identified depression in community settings: Concordance with structured-interview diagnoses. *Psychother Psychosom* 2013;82:161–9.
26 Olfson M, Blanco C, Marcus SC. Treatment of adult depression in the United States. *JAMA Intern Med* 2016;176:1482–91.
27 Johnson CF, Dougall NJ, Williams B, *et al*. Patient factors associated with SSRI dose for depression treatment in general practice: a primary care cross sectional study. *BMC Fam Pract* 2014;15:210.
28 Hetlevik Ø, Garre-Fivelsdal G, Bjorvatn B, *et al*. Patient-reported depression treatment and future treatment preferences: an observational study in general practice. *Fam Pract* 2019;36:771–7.
29 Prieto-Alhambra D, Petri H, Goldenberg JSB, *et al*. Excess risk of hip fractures attributable to the use of antidepressants in five European countries and the USA. *Osteoporos Int* 2014;25:847–55.
30 Dumesnil H, Cortaredona S, Verdoux H, *et al*. General practitioners' choices and their determinants when starting treatment for major depression: a cross sectional, randomized case-vignette survey. *PLoS One* 2012;7:e52429.
31 Morgan D,J. Treatment and follow-up of anxiety and depression in clinical-scenario patients: survey of Saskatchewan family physicians. *Can Fam Physician* 2012;58:e152–8.
32 Cutler D, Skinner JS, Stern AD, *et al*. Physician beliefs and patient preferences: a new look at regional variation in health care spending. *Am Econ J Econ Policy* 2019;11:192–221.
33 Berndt ER, Gibbons RS, Kolotilin A, *et al*. The heterogeneity of concentrated prescribing behavior: theory and evidence from antipsychotics. *J Health Econ* 2015;40:S0167-6296(14)00140-4:26–39.:.
34 Haddad M, Menchetti M, McKeown E, *et al*. The development and Psychometric properties of a measure of Clinicians' attitudes to depression: the revised depression attitude questionnaire (R-DAQ). *BMC Psychiatry* 2015;15:7.
35 Carey M, Jones K, Meadows G, *et al*. Accuracy of general practitioner unassisted detection of depression. *Aust N Z J Psychiatry* 2014;48:571–8.
36 Lyratzopoulos G, Elliott M, Barbiere JM, *et al*. Understanding ethnic and other socio-demographic differences in patient experience of primary care: evidence from the English general practice patient survey. *BMJ Qual Saf* 2012;21:21–9.
37 Cooper LA, Roter DL, Carson KA, *et al*. The associations of Clinicians' implicit attitudes about race with medical visit communication and patient ratings of interpersonal care. *Am J Public Health* 2012;102:979–87.
38 Gonçalves DA, Mari J de J, Bower P, *et al*. Brazilian Multicentre study of common mental disorders in primary care: rates and related social and demographic factors. *Cad Saude Publica* 2014;30:623–32.
39 Arpey NC, Gaglioti AH, Rosenbaum ME. How socioeconomic status affects patient perceptions of health care: a qualitative study. *J Prim Care Community Health* 2017;8:169–75.
40 White AA, Stubblefield-Tave B. Some advice for physicians and other Clinicians treating minorities, women, and other patients at risk of receiving health care disparities. *J Racial Ethn Health Disparities* 2017;4:472–9.
41 Ballering AV, Muijres D, Uijen AA, *et al*. Sex differences in the Trajectories to diagnosis of patients presenting with common somatic symptoms in primary care: an observational cohort study. *J Psychosom Res* 2021;149:110589.
42 Rosman S. Les Pratiques de prescription des Médecins Généralistes. une Étude Sociologique comparative Entre La France et LES pays-BAS. *Métiers Santé Social* 2015:117–32.
43 Elyoseph Z, Hadar-Shoval D, Asraf K, *et al*. Chatgpt Outperforms humans in emotional awareness evaluations. *Front Psychol* 2023;14:1199058.
44 Hadar-Shoval D, Elyoseph Z, Lvovsky M. The plasticity of Chatgpt's Mentalizing abilities: Personalization for personality structures. *Front Psychiatry* 2023;14:14.
45 Lim B, Sproule BA, Zahra Z, *et al*. Understanding the effects of chronic benzodiazepine use in depression: a focus on Neuropharmacology. *Int Clin Psychopharmacol* 2020;35:243–53.
46 Driot D, Bismuth M, Maurel A, *et al*. Management of first depression or generalized anxiety disorder episode in adults in primary care: A systematic Metareview. *Presse Med* 2017;46(12 Pt 1):1124–38.
47 Bernard K, Nissim G, Vaccaro S, *et al*. Association between maternal depression and maternal sensitivity from birth to 12 months: a meta-analysis. *Attach Hum Dev* 2018;20:578–99.
48 Katznelson G, Gerke S. The need for health AI ethics in medical school education. *Adv Health Sci Educ Theory Pract* 2021;26:1447–58.
49 O'Keefe J, Speakman A. Single unit activity in the rat hippocampus during a spatial memory task. *Exp Brain Res* 1987;68:1–27.
50 Chen J,L, Goldstein T, Huang H, *et al*. n.d. Instructzero: efficient instruction optimization for black-box large language models.