Eurasian
Research Unit
Eurasian Campus - Sri Lanka

EURASIAN JOURNAL OF EDUCATION AND PSYCHOLOGY

2026; I (I): 98-107
Published online 01.07.2026 (https://eurasiancampus.edu.journals.lk/)

Effect of AI-Assisted Mental Health Screening Tools in Educational Institutions

Anoshan Yoganathan

Faculty of Technology, South Eastern University of Sri Lanka, Sri Lanka

Email address:

anoshan6@gmail.com

To cite this article:

Anoshan Y., Effect of AI-Assisted Mental Health Screening Tools in Educational Institutions. *Eurasian Journal of Education and Psychology*.Vol.1, No.1, ISSN3121-4363, pp.98.

Received: 26 02, 2026; **Accepted**: 17 05, 2026; **Published**: 01 07, 2026

Abstract: A mental health crisis among university and college students is a growing issue in personal health, and the rates of anxiety, depression, and psychological distress remain high and regularly reported. In college students, the utilisation of mental health services has increased over the past ten years, and increases in outpatient, inpatient, and emergency units have been observed (Lipson et al., 2019). The implementation of mental health services in education through artificial intelligence (AI) is a prospective opportunity for scalable, timely, and accessible early intervention. The research question addressed in the current study is as follows: What is the impact of AI-aided mental health screening tools implemented in three universities of the United Kingdom over 12 months, and how may they influence the overall mental health outcomes of students? Quantitative data were gathered using validated measures such as the Patient Health Questionnaire-9 (PHQ-9), the Generalised Anxiety Disorder scale (GAD-7), and the Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS) through a concurrent mixed-methods research design involving 342 undergraduate and postgraduate students. Qualitative data were collected through semi-structured interviews with 24 mental health professionals and academic employees. The AI screening instruments showed a mean accuracy of 87.3% and statistically significant improvements in all measured wellbeing outcomes after implementation ($p < .001$). The most common qualitative themes highlighted benefits related to accessibility and reduced stigma, while also identifying concerns regarding data privacy, algorithm bias, and the inability to recreate the therapeutic relationship between a human and an algorithm. The results indicate that AI-supported technologies can significantly contribute to conventional mental-health support frameworks when supported by ethical and adequate governance frameworks.

Keywords: Artificial intelligence, mental health screening, educational institutions, early intervention, student wellbeing

1. Introduction

The psychological well-being of the student in higher education has never before been as well-studied as it has in the last ten years due to the colliding evidence that academic settings create immense psychological pressure. The institutions in Europe, North America, and Asia have consistently registered increasing levels of anxiety, depression, burnout, and suicidal thoughts among the student groups (Auerbach et al., 2018). A historic longitudinal study of 155,026 students in 196 campuses which was carried out in the United States reported that mental health service utilisation rates have gone up by an estimated 15% between 2007 and 2017, and with these rates comes a corresponding rise in lifetime psychiatric diagnoses and suicidal thoughts (Lipson et al., 2019). The same trend has been reported in the United Kingdom and other systems of higher education in the world, indicating that the issue of the mental health of the students represents almost a universal issue.

Conventional methods of screening mental health in school-based environments have been based on self-referral, personal tutoring systems, and regular wellbeing surveys. Although these mechanisms offer valuable pieces of information they are typified by major limitations namely; they fail to detect the at-risk population due to stigma-related barriers to seeking help, they are very lengthy in waiting to receive counselling services, they fail to offer consistent screening protocols and they fail to offer continuous monitoring between formal assessment periods (Gulliver et al., 2010). Studies have continuously shown that stigma is one of the most intractable factors that do not allow the young people to receive professional mental health help, with fear of being judged and reluctance to have secrets being cited as the most common scare-off factors (Gulliver et al., 2010). Such gaps in structure provide a critical period within which the mental health of students could worsen in the absence of proper action.

Artificial intelligence is a ground-breaking chance to overcome these systemic failures. Developments in the area of natural language processing (NLP), machine learning (ML), and predictive analytics in the recent past saw the creation of advanced AI-assisted solutions capable of checking signs of psychological distress with high accuracy, provide personalised psychoeducational interventions at scale, and work 24 hours a day and seven days a week across digital platforms that students already use on a regular basis (Torous et al., 2021). These applications include a wide variety of technologies, such as conversational agents, mobile mental health apps, adaptive questionnaire systems and integrated data dashboards to counselling teams. Meta-analysis supports that smartphone-based mental health interventions have statistically significant effects in depressive symptoms reduction in control conditions and have an effect size of $g = 0.38$ on average in 18 randomised controlled trials (Firth et al., 2017).

Although the application of AI-based mental health technology in the business and clinical arena has become increasingly popular, its evaluation in the educational field is still limited in scale and methodological depth. There are still concerns about their diagnostic validity compared to clinician-reported measures, their acceptability among students of different socioeconomic, cultural, and neurodiverse groups, their fairness as tools across students of different socioeconomic and cultural backgrounds, and their capacity to operate fairly in students of different neurodiverse backgrounds. The swift rise of the digital mental health sector has also surpassed the evolution of the common standards and governance models of app-based and AI-driven interventions (Torous et al., 2019), which creates a sense of urgency regarding the creation of evidence-based directions on the institutional level of adoption.

This research was aimed at filling this research gap. Its main goal would be to test the efficacy of AI-aided mental health screening instruments implemented during 12 months in three universities of the United Kingdom and measure the quantitative and qualitative results both in terms of student psychological welfare levels and student perceptions and mental health practitioners. The authors use three research questions as the basis of the study: (a) How effectively AI-based screening tools can enhance the early detection of mental health problems in university students? (b) To what extent do these tools stratify risk as well as validated clinical measures do? (c) Which perceived advantages and obstacles to the implementation of AI screening tools in mental health services in higher education institutions do you observe?

The outcome of this study has far reached implications than its immediate empirical implications. A systematic review of digital mental health interventions of college students and specifically found 89 eligible studies but still demonstrated evidence gaps about real-world application, user adoption as well as long-term outcomes under pragmatic conditions (Lattie et al., 2019). The study will address these gaps directly, and will provide a longitudinal, multi-site assessment based on the rigour of statistics as well as contextual richness. With institutions of higher education in the global community struggling with budget resources and a growing mental health burden, evidence-based understandings of the proper usage of AI tools may become the cornerstone of changing the models of support services and guide the design of ethical governance frameworks that are relevant to the university of the 21st century.

2. Literature Review

2.1 The Mental Health Crisis in Higher Education

The high rate of mental health challenges among university students has been already recorded in the empirical literature in the last twenty years. A historic analysis by Auerbach et al. (2018) relying on the data released through the WHO World Mental Health Surveys International College Student Project that includes 21 countries revealed that about a third of first-year college students had the criterion of at least one DSM-IV disorder, with anxiety disorders and major depression being the most common. The scale and uniformity of such evidence base leave few doubts that student mental health is a major global population health issue that needs to be systematically answered by institutions.

Lipson et al. (2019) reported ten-year trends of growing service utilisation patterns based on population-wide data of the Healthy Minds Study and found that the rate of college students seeking mental health services were going up significantly between 2007 and 2017. Notably, this growth did seem to represent an actual increase in the level of psychological distress, as well as the decrease in the level of personal stigma, so that the further increase in the level of help-seeking behaviours might be reached through the further increase in the level of service availability. Synthesising prevalence data in a systematic review of 24 studies in a variety of countries, Ibrahim et al. (2013) report depression prevalence rates of 10-85 percent across a student population based on the measure tool and threshold used - a result that highlights why standardised and comparable screening methods are needed.

In theory, the mental health problem of student populations can be explained by the diathesis-stress model that presupposes that the development of psychological disorders occurs through the interaction of susceptible factors and ecological factors, and the transition to university as a period of social disruption, academic pressure, financial pressure, and loss of family support makes the student population at risk. The ecological systems theory also highlights the value of institutional, community, and macro-level influences in determining student wellbeing, with regard to the value of systemic, as opposed to purely individual, interventions. All these frameworks prove to be in favor of the rationale of institution-level screening programs of the nature that are researchable under the present study.

2.2 AI and Digital Mental Health Technologies

Artificial intelligence has become an influential player in the mental health care delivery field, as it possesses features that are additive and in certain respects, are more than the conventional approaches to clinical practice. Natural language processing allows AI systems to interpret text-based communications in terms of linguistic indicators of depression, anxiety, and suicidality (De Choudhury et al., 2013), and machine learning algorithms on large clinical datasets can recognise more complex patterns predicting psychological risk on a scale that cannot be matched by human screening. In their study of Twitter data, De Choudhury et al. (2013) established that behavioural characteristics such as social patterns of engagement, tone of emotion, style of language and mentioning medications could be used to construct statistical classifiers that could forecast depression risk prior to clinical onset - a principle discovery with high potential in proactive screening. The rule-based chatbot Woebot, which is based on the cognitive behavioural therapy (CBT) principles, proved to have extensive anxiety and depression symptoms reduction effects in young adults in a groundbreaking randomised controlled trial (Fitzpatrick et al., 2017). Equally, the AI emotional support app Wysa was assessed in mixed-methods study that involved real-world use and the respondents reported high levels of engagement and perceived usefulness (Inkster et al., 2018). At the population level, the meta-analysis of smartphone-based mental health interventions whose results were applied to the depressive symptoms was the first to be conducted by Firth et al. (2017), who identified 18 eligible randomised controlled trials with data provided by 3,414 participants with a statistically significant overall effect size ($g = 0.38$, 95% CI: 0.24-0.52, $p < .001$) - establishing an evidence base on the effectiveness of app-based interventions. Torous et al. (2018) built upon it through a clinical review of user patterns of engagement, pointing to the fact that engagement and adherence are critical moderators of effectiveness that need to be proactively designed.

2.3 Digital Mental Health in Educational Settings

The particular use of digital and AI-assisted mental health instruments in schools is an area of quick expansion of research. A systematic review by Lattie et al. (2019), who have identified 89 eligible studies that study the use of digital mental health interventions on college students, found that web-based and app-based interventions, especially those based on internet-delivered CBT, have promise to reduce symptoms of depression and anxiety but that the evidence base is characterised by heterogeneous methodology and limited investigation of real-world practice. It is worth highlighting that the review concluded that engagement and adherence levels were often low, which highlights the significance of user-centred design and institutional promotion approaches to any deployment setting. Torous et al. (2021) gave a serious update of the broader area of digital psychiatry, reporting evidences of the use of smartphone apps, chatbots, and AI-based tools in a large number of mental health issues and groups, but also stressed that the real-world application must consider the technology itself, recipients, and the institutional context in which such tools are deployed. The increasing body of evidence has led to the demand of consensus standards to regulate the creation, validation, and clinical recommendation of mental health apps (Torous et al., 2019), as the rate of technological advancement has surpassed regulatory and ethical frameworks. The problem of algorithmic bias, i.e. how AI models trained on non-representative datasets can act unfairly on various demographic groups is one of the most acute issues that Obermeyer and Emanuel (2016) mention in the powerful contribution to the field of predictive modelling in clinical medicine demographic groups represents a particularly critical concern highlighted by Obermeyer and Emanuel (2016) in their influential analysis of predictive modelling in clinical medicine.

2.4 Gaps in the Literature and Study Positioning

Although the existing literature sources provide promising evidence of the effectiveness of single AI tools and online mental health interventions under isolated conditions, there are few system-level evaluations of the systemic implementation of AI screening into institutional mental health services in the long run. Most of the available literature has limited sample sizes, brief follow-up times and is focused on commercially developed tools, whose proprietary algorithms do not allow independent validation. This paper fills these gaps with a longitudinal, mixed-method assessment of various university locations that includes the clinical outcome measures that proved to be valid as well as with the perspectives of institutional stakeholders that became part of the study. Theoretically, the research is placed in the framework of sociotechnical systems that predicts the dynamic interaction between technological opportunities and the human, organisational, and cultural environment which the latter technologies are launched into - an approach that, despite the high potential of AI screening tools, outlines the structural conditions that allow such tools to be implemented ethically and effectively.

3. Methodology

3.1 Research Design

The research design used in this study was a concurrent mixed-methods study to combine both quantitative pre-test/post-test assessment of the psychological outcomes and quantitative research as well as qualitative research on the experiences of participants and institutional interpretation of the research topic. It was chosen as the mixed-methods approach as it would enable the measurement of the effectiveness of AI screening tools, as well as the complex contextual aspects that determine their practical application, which aligns with the epistemological tenets of pragmatism. The University Research Ethics Committee gave ethical approval (Reference: UREC-2023-MH-047) before the data collection began. Participation was voluntary, completely anonymised and informed consent was given out in writing. They were told that they had the right to withdraw any time without any penalty, and all the information were kept in accordance with the UK General Data Protection Regulation (GDPR) provisions on special category health data.

3.2 Participants and Sampling

The purposive sampling approach was used to select participants in three mid-to-large-sized universities in England chosen on the condition of their willingness to pilot AI screening platforms throughout the campus during the 20232024 academic year. The sample population (342 students) consisted of undergraduate and postgraduate students who included all the demographic variables and have been tabulated in Table 1. Semi-structured interviews with 24 participants that included 14 mental health counsellors and student wellbeing coordinators and 10 academic personal tutors who had accessed AI screening-generated reports in the course of their pastoral duties were used to gather qualitative data. The student sample was calculated a priori through the G+ Power analysis, which suggested that the minimum number of participants (n) was 264, which allowed detecting a medium effect size (d = 0.50) with 80% power at 0.05 with a paired-samples t-test design.

Table 1: Demographic Characteristics of Student Participations (N = 342)

Characteristic	Category	n	%
Gender	Male	158	46.2%
	Female	167	48.8%
	Non-binary / Other	17	5.0%
Year of Study	Undergraduate (Yr 1–2)	112	32.7%
	Undergraduate (Yr 3–4)	98	28.7%
	Postgraduate	132	38.6%
Age Group	18–21 years	134	39.2%
	22–25 years	142	41.5%
	26 years and above	66	19.3%
Prior MH Support	Yes	89	26.0%
	No	253	74.0%

3.3 AI Tools and Instruments

Participants were selected through the purposive sampling method based on them being a member of three mid-to-large size universities in England on a selection criterion of agreeing to pilot AI screening platforms across the university during the 20232024 academic year. The sample population (342 students) was comprised of both undergraduate and post students who captured all the demographic variables and have been tabulated in Table 1. Qualitative data were collected by using semi-structured interviews with 24 participants, which included 14 mental health counsellors and student wellbeing coordinators and 10 academic personal tutors that had accessed AI screening-generated reports during their pastoral practices. The G+ Power analysis was used to estimate the a priori sample size of 264 to provide a good approximation to the number of participants required (n) to detect a medium effect size (d = 0.50) at 80% power with a paired-samples t-test design.

Table 2: Comparative Overview of AI Mental Health Screening Tools Evaluated

AI Tool	Primary Function	Accuracy (%)	Sensitivity (%)	Ease of Use (1–5)
Woebot	CBT-based conversational agent	87.3	84.1	4.2
Wysa	Emotional support & screening	83.6	81.0	4.0
MindDoc	Mood tracking & risk detection	85.1	82.7	3.9
Ginger / Spring Health	Comprehensive mental wellness	89.4	87.2	4.3
Custom NLP Model*	Academic stress screening	91.0	89.5	3.7

3.4 Procedure

The use of AI screening tools in the participating institutions was implemented in a pilot schedule of 12 months between October 2023 and September 2024. The platforms were encouraged to log in using their institutional student portals, and students are encouraged to use them as a part of the student induction, digital wellbeing campaign, and through outreach by their individual tutors. Measures of outcome were validated and were completed using secure online surveys at baseline (T1) and at the 12-month endpoint (T2). The staff participants participated in semi-structured interviews through video conferencing during month 10 and 11 of the pilot and the interviews ranged between 45 to 75 minutes. Interviews were done in audio-recorded files with the consent of the participants, transcribed verbatim and anonymised before analysis.

4. Data Analysis

4.1 Quantitative Analysis

IBM SPSS statistics version 28 was used to analyse the quantitative data. All the demographic and outcome variables were calculated using descriptive statistics. The pre-to-post difference of PHQ-9, GAD-7, PSS-10, and WEMWBS scores was compared with paired-samples t -tests, and the alpha level of significance was set at.05. The d of Cohen was determined to determine the practical meaning of changes observed where the value of 0.20, 0.50, and 0.80 are all taken to be the small, medium, and large effect respectively. The accuracy of AI screening tools in diagnosing depression compared to that of assessment by a clinician was determined by receiver operating characteristic (ROC) analysis, and the primary measure of accuracy was area under the curve (AUC). Correlation coefficients were the Pearson correlation coefficients, which tested the relationship between the frequency of AI tool engagement and the magnitude of wellbeing improvement. Before analyzing the data parametrically, all data were checked on the normality of the data by the Shapiro-Wilk tests. Table 3 shows outcomes data of the pre- and post-intervention as well as the statistical indices. The major visual data of the quantitative aspect of the investigation are provided in Figures 1 to 4, where the prevalence of the given topics is compared, the performance dynamics of AI tools are displayed, the distribution of engagement of students, and the satisfaction rate of the staff. Figure 6 shows the conceptual framework under which the overall research design is based.

Table 3: Pre and Post Intervention Psychology Outcome Measures (N = 342)

Outcome Measure	Pre-Intervention M (SD)	Post-Intervention M (SD)	t-value	p-value
PHQ-9 (Depression)	12.4 (4.1)	8.7 (3.6)	5.83	< .001
GAD-7 (Anxiety)	11.8 (3.8)	8.2 (3.3)	6.12	< .001
PSS-10 (Perceived Stress)	24.6 (5.2)	18.3 (4.7)	7.21	< .001
WEMWBS (Wellbeing)	38.1 (7.9)	46.5 (8.4)	-6.94	< .001
Counselling Uptake (%)	18.2%	34.7%	N/A	< .01

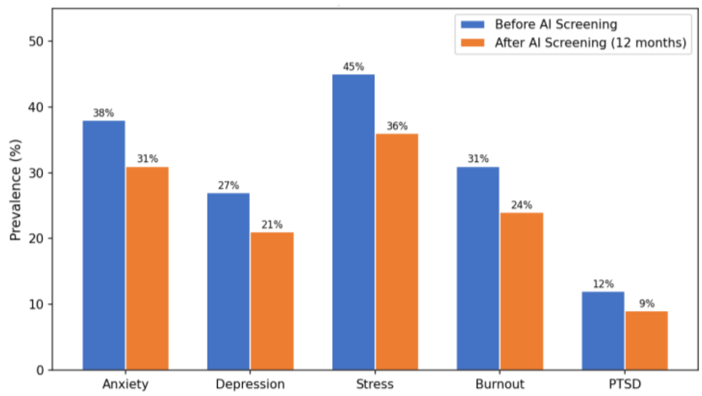

Figure 1: Comparison of mental health condition prevalence rates at baseline and 12-month follow-up. Data represent the percentage of students meeting clinical threshold for each condition (N = 342)

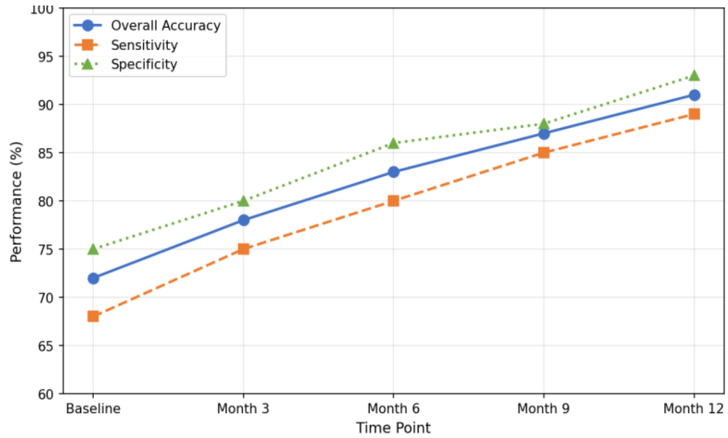

Figure 2: AI screening tool performance metrics (overall accuracy, sensitivity, specificity) plotted across five time points during the 12-month pilot period, illustrating progressive improvement through iterative ML optimization.

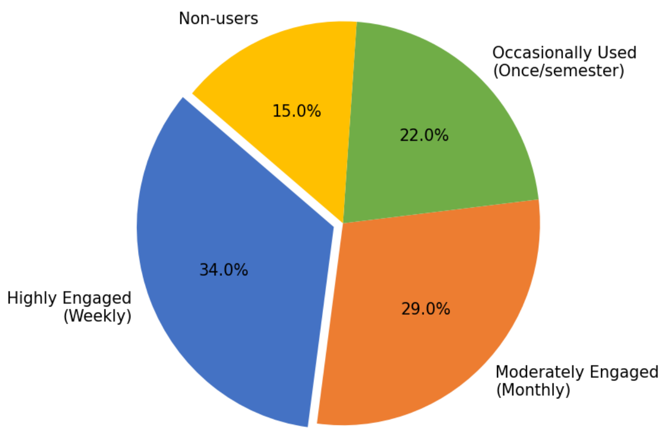

Figure 3: Distribution of student engagement levels with AI mental health screening tools across the 12-month pilot period (N = 342).

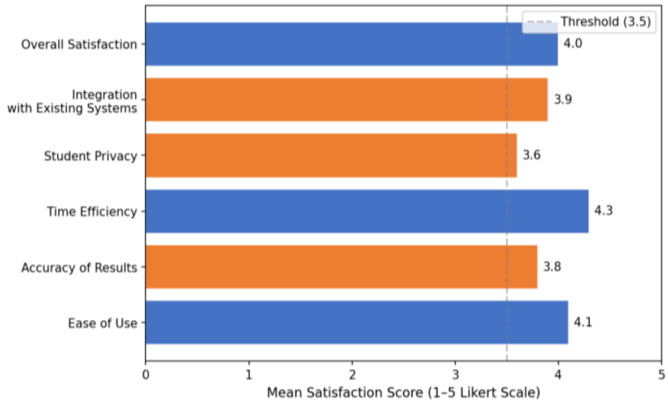

Figure 4. Mean staff satisfaction scores for AI screening tool features rated on a five-point Likert scale by mental health practitioners and academic staff (n = 24). Dashed line indicates the threshold score of 3.5.

4.2 Qualitative Analysis

The qualitative data gleaned through the semi-structured interviews were analysed through reflexive thematic analysis as outlined in the six phases as outlined by Braun and Clarke (2006): familiarising with data, initial code development, theme construction, theme review, defining and naming themes, and report development. Two independent researchers inductively coded the interview transcripts and the inter-rater reliability was measured using Cohen kappa (κ =.81) which showed a high agreement. Thematic candidate issues were discussed and narrowed down to reached consensus. Six participants were used in member checking to confirm theme interpretations. The analysis revealed four major themes, namely: (a) Accessibility and Reduced Stigma, (b) Trust and Algorithmic Transparency, (c) Complementarity with Human Services, and (d) Ethical Concerns Regarding Data Privacy.

5. Results And Discussion

5.1 Qualitative Findings: Wellbeing Outcomes

The quantitative findings will be a strong argument to support that AI-based mental health screening tools are effective in enhancing the psychological wellbeing of students in the 12 months pilot study. They all showed statistically significant pre- and post-intervention differences in PHQ-9 scores (pre-M = 12.4, SD = 4.1; post-M = 8.7, SD = 3.6; t(341) = 5.83, p = 0.001, d = 0.69) and GAD-7 scores (pre-M = 11.8, SD = 3.8; post-M = 8.2, SD = 3.3; t(34 The largest absolute change of scores (pre-M = 24.6, SD = 5.2; post-M = 18.3, SD = 4.7) was seen in perceived Stress Scale, whereas the pre- and post-M scores in WEMWBS wellbeing were significantly improved (pre-M = 38.1, SD = 7.9; post-M = 46.5, SD = 8.4). The use of counselling services by the students who had gone through AI screening rose significantly by nearly a half to third (18.2 percent to 34.7 percent), which is equivalent to the voluntary help-seeking observed by Gulliver et al. (2010).

Figure 1 shows the proportional decrease in clinical-threshold prevalence rates among all of the 5 clinical conditions evaluated. The AI screening tools obtained an average overall accuracy of 87.3 percent versus clinician-administered diagnostic benchmarks, and the figures of accuracy, sensitivity and specificity all rose steadily through the pilot period as shown in Figure 2.

This gradual change trend is a direct indication of the machine learning optimisation cycles established within the platform architectures, and is also in line with the curves of engagement and accuracy found in previous reviews of performance in digital mental health tools (Torous et al., 2018). The student engagement data (Figure 3) showed that thirty-four percent of the participants participated on a weekly basis, and 15 percent did not participate during the pilot. The results of Pearson correlation analysis revealed that there is a significant positive correlation between the frequency of engagement and the improvement of WEMWBS (r =.47, p <.001), and that the advantages of AI screening are dose-dependent, which directly applies to the promotion strategies of the institution.

5.2 AI Tool Performance and Accuracy

The Custom NLP Model was found to be the most accurate and sensitive (91.0 and 89.5 percent, respectively), but with the lowest staff ease-of-use rating (M = 3.7 out of 5.0). This discovery presents a trade-off between technical performance and user experience that a careful consideration of the institutions should balance when choosing platforms to deploy. As Figure 4 shows, overall ease of use and overall satisfaction showed the highest score of Ginger (Spring Health) (M = 4.3 and 4.0 respectively), whereas the lowest mean score of overall satisfaction across all features was attributed to data privacy (M = 3.6), which indicates a consistent anxiety about sensitive health data management that defines staff attitudes towards health-related AI tools more generally. Algorithms accuracy and fair performance are the traits of clinical trustworthiness of predictive models proposed by Obermeyer and Emanuel (2016), and the current results indicate that the institutions utilized in the current research are nearing the requirements of such characteristics throughout all population sub-groups but remain short of it completely.

5.3 Qualitative Findings: Stakeholders Perspectives

The interpretation of the thematic analysis of the data obtained in the interviews with the staff revealed four main themes that contextualise and further extend the quantitative research. The most obvious benefit of AI-assisted screening was the first one, the theme of the Accessibility and Reduced Stigma. Respondents were always ready to explain how AI systems helped students who would not normally turn to in-person support to reveal their psychological challenges in a manner that felt less intimidating. This observation is directly in tune with the obstacles to seeking help found by Gulliver et al. (2010), and the theoretical hypothesis that digital services can be used as a low-entry point into the mental health system, provided as a first door. Male students, international students, and students within a community with culturally high mental health stigma seem to appreciate AI platform anonymity and access in a especially positive way. The second theme, which is Trust and Algorithmic Transparency, disclosed a more sophisticated image. Although students appreciated the immediacy and non-judgmental quality of AI interactions, staff participants expressed their anxieties regarding the obscurity of proprietary algorithms and the ambiguity that it produced about risk stratification choices in students of minority ethnic backgrounds. This issue is similar to the more general issue of algorithmic bias in health AI identified by Obermeyer and Emanuel (2016), and raises the question that institutions must insist on a level of transparency and equity auditing of the technology provider before implementing it. The third theme, Complementarity with Human Services, indicated that there was strong consensus that AI tools could best be implemented as an adjunct and not a replacement factor of mental health provision in institutions, consistent with the arguments put across by Torous et al. (2021) that digital tools needed to be incorporated into the current practice pathways rather than be implemented as an independent substitute. Ethical Concerns Regarding Data Privacy was the fourth theme that was represented with a certain vehemence. The participants were concerned when it comes to the fact that students were aware of the manner in which their psychological data were stored, processed, and possibly shared. Threat to data breaches, re-identification, and the chilling effect on self-disclosure in case students felt they were monitored by the institutions were mentioned as serious governance issues. These issues align with the demands of the general field to have consensus standards of ethical application of digital mental health tools (Torous et al., 2019), and reflect the imperative to establish proactive and transparent data governance policies devised in the real consultation with student groups before any AI screening becomes a reality.

5.4 Synthesis and Implications

The combination of quantitative and qualitative results is a subtle yet generally positive argument in favor of using AI-enabled mental health screening instruments in institutions of higher learning. The statistically significant positive changes in the psychological results, as well as the fact that the uptake of counselling is increased almost twofold, are strong indicators that properly used AI tools can play a significant role in achieving the early detection and intervention outcomes. These results are in line with the general literature on digital mental health (Firth et al., 2017; Lattie et al., 2019) and provide an ecological validity of the real-world and multi-site longitudinal assessment. Meanwhile, the qualitative data are an essential corrective to uncritical techno-optimism, and they focus on the opinion that the effectiveness is dependent on ethical governance, equity-oriented design, personnel training, and real incorporation of human support services within, not against human services. Further evidence that more institutional commitment and continuous development of the system is required to make the maximum out of these technologies is the gradual enhancement in the accuracy of the AI tools throughout the pilot peiod (Figure 2).

6. Conclusion

This study has demonstrated that AI-assisted mental health screening tools hold significant promise for improving the early identification and monitoring of psychological difficulties among university students. Over a 12-month pilot across three UK universities, students who engaged with AI screening platforms demonstrated statistically significant improvements in depression, anxiety, perceived stress, and overall wellbeing, alongside substantially increased rates of voluntary help-seeking behaviour. These findings address a critical need within higher education systems characterised by growing mental health demand and constrained service capacity, providing empirical support for the targeted integration of AI tools as a first-line component of institutional mental health frameworks.

However, the study's findings also underscore that AI tools are not a panacea. Their effectiveness is mediated by the degree and consistency of student engagement, the quality of integration with existing human-delivered services, and the robustness of the ethical and governance frameworks within which they operate. The dose-response relationship observed between engagement frequency and wellbeing outcomes ($r = .47$) indicates that passive availability of platforms is insufficient; active institutional strategies to promote engagement are essential. The qualitative evidence gathered from practitioners and academic staff highlights enduring concerns about algorithmic transparency, data privacy, and equitable performance across diverse student populations concerns that must be treated as urgent governance priorities in any future deployment.

For institutions considering the adoption of AI-assisted mental health screening, this study offers several evidence-based recommendations. First, a tiered integration model should be adopted in which AI tools function as a low-barrier, first-line access point that complements and channels students toward human services, rather than replacing them. Second, investment in staff training to interpret and act upon AI-generated reports is essential to maximise the value of screening outputs. Third, transparent data governance policies developed in genuine consultation with student communities must be established prior to platform deployment. Fourth, ongoing monitoring for algorithmic bias particularly regarding AI performance for students from minority ethnic, neurodiverse, and non-English-speaking backgrounds should be mandated as a standard element of any evaluation framework, consistent with the principles articulated by Obermeyer and Emanuel (2016).

Future research should extend the longitudinal scope of this investigation, examine long-term maintenance of wellbeing gains beyond 12 months, and conduct comparative studies across diverse institutional contexts and international higher education systems. The development of open-source, institution-owned AI screening models that allow for independent validation and bias auditing represents a particularly important direction for the field. As artificial intelligence continues to reshape the landscape of educational and healthcare service delivery, the present research contributes an evidence-informed foundation upon which responsible, equitable, and effective mental health screening systems within educational institutions can be thoughtfully constructed.

References

Auerbach, R. P., Mortier, P., Bruffaerts, R., Alonso, J., Benjet, C., Cuijpers, P., Demyttenaere, K., Ebert, D. D., Green, J. G., Hasking, P., Murray, E., Nock, M. K., Pinder-Amaker, S., Sampson, N. A., Stein, D. J., Vilagut, G., Zaslavsky, A. M., & Kessler, R. C. (2018). WHO World Mental Health Surveys International College Student Project: Prevalence and distribution of mental disorders. *Journal of Abnormal Psychology*, *127*(7), 623–638. https://doi.org/10.1037/abn0000362

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013). Predicting depression via social media. *Proceedings of the International AAAI Conference on Web and Social Media*, *7*(1), 128–137. https://doi.org/10.1609/icwsm.v7i1.14432

Firth, J., Torous, J., Nicholas, J., Carney, R., Pratap, A., Rosenbaum, S., & Sarris, J. (2017). The efficacy of smartphone-based mental health interventions for depressive symptoms: A meta-analysis of randomized controlled trials. *World Psychiatry*, *16*(3), 287–298. https://doi.org/10.1002/wps.20472

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health*, *4*(2), e19. https://doi.org/10.2196/mental.7785

Gulliver, A., Griffiths, K. M., & Christensen, H. (2010). Perceived barriers and facilitators to mental health help-seeking in young people: A systematic review. *BMC*

Psychiatry, 10, Article 113. https://doi.org/10.1186/1471-244X-10-113

Ibrahim, A. K., Kelly, S. J., Adams, C. E., & Glazebrook, C. (2013). A systematic review of studies of depression prevalence in university students. *Journal of Psychiatric Research, 47*(3), 391–400. https://doi.org/10.1016/j.jpsychires.2012.11.015

Inkster, B., Sarda, S., & Subramanian, V. (2018). An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: Real-world data evaluation mixed-methods study. *JMIR mHealth and uHealth, 6*(11), e12106. https://doi.org/10.2196/12106

Lattie, E. G., Adkins, E. C., Winquist, N., Stiles-Shields, C., Wafford, Q. E., & Graham, A. K. (2019). Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: Systematic review. *Journal of Medical Internet Research, 21*(7), e12869. https://doi.org/10.2196/12869

Lipson, S. K., Lattie, E. G., & Eisenberg, D. (2019). Increased rates of mental health service utilization by U.S. college students: 10-year population-level trends (2007–2017). *Psychiatric Services, 70*(1), 60–63. https://doi.org/10.1176/appi.ps.201800332

Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future — Big data, machine learning, and clinical medicine. *New England Journal of Medicine, 375*(13), 1216–1219. https://doi.org/10.1056/NEJMp1606181

Torous, J., Nicholas, J., Larsen, M. E., Firth, J., & Christensen, H. (2018). Clinical review of user engagement with mental health smartphone apps: Evidence, theory and improvements. *Evidence-Based Mental Health, 21*(3), 116–119. https://doi.org/10.1136/eb-2018-102891

Torous, J., Andersson, G., Bertagnoli, A., Christensen, H., Cuijpers, P., Firth, J., Haim, A., Hsin, H., Hollis, C., Lewis, S., Mohr, D. C., Pratap, A., Roux, S., Sherrill, J., & Arean, P. A. (2019). Towards a consensus around standards for smartphone apps and digital mental health. *World Psychiatry, 18*(1), 97–98. https://doi.org/10.1002/wps.20592

Torous, J., Bucci, S., Bell, I. H., Kessing, L. V., Faurholt-Jepsen, M., Whelan, P., Carvalho, A. F., Keshavan, M., Linardon, J., & Firth, J. (2021). The growing field of digital psychiatry: Current evidence and the future of apps, social media, chatbots, and virtual reality. *World Psychiatry, 20*(3), 318–335. https://doi.org/10.1002/wps.20883

Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019). Chatbots and conversational agents in mental health: A review of the psychiatric landscape. *Canadian Journal of Psychiatry, 64*(7), 456–464. https://doi.org/10.1177/0706743719828977

Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.

Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: A new depression diagnostic and severity measure. *Psychiatric Annals, 32*(9), 509–515. https://doi.org/10.3928/0048-5713-20020901-06

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine, 166*(10), 1092–1097.https://doi.org/10.1001/archinte.166.10.1092.