Journal of
Information Security and Technologies

JScholar
Fostering Scholarly Communication

Research Article Open Access

Deploying Digital Therapists: Evaluating Luna in Support of Postsecondary Student Mental Health

Phillip Olla[1], Ashlee Barnes[2*], Lauren Elliott[3], Mustafa Abumeeiz[3], Venus Olla[4] and Joseph Tan[5]

[1]Associate Professor of Health Operations Management, University of Detroit Mercy, 4001 W. McNichols Rd., Detroit, MI 48221, United States

[2]Assistant Professor of Nursing, University of Detroit Mercy, 4001 W. McNichols Rd., Detroit, MI 48221, United States

[3]Pre-medicine Student at Western University, 1151 Richmond St., N6A 3K7, London, ON, Canada

[4]Clinical Therapist, Student Counseling, University of Windsor, 401 Sunset Ave., Windsor, ON, N9B 3P4, Canada

[5]Professor of Health Informatics, McMaster University, 1280 Main St W, Hamilton, L8S 4L8, Canada

[*]**Corresponding Author:** Dr. Ashlee Barnes, 4001 W. McNichols Rd., Detroit, MI 48221; Tel: 313-993-2481; E-mail: barnesav@udmercy.edu

Received Date: November 14, 2024 **Accepted Date:** December 14, 2024 **Published Date:** December 17, 2024

Citation: Phillip Olla, Ashlee Barnes, Lauren Elliott, Mustafa Abumeeiz, Venus Olla, et al. (2024) Deploying Digital Therapists: Evaluating Luna in Support of Postsecondary Student Mental Health. J Inf Secur Technol 1: 1-14

Abstract

"Luna," an AI-powered chatbot designed to support student mental health, was implemented to address rising mental health concerns in postsecondary institutions. This work explores the efficacy of Integrating Luna, a virtual mental health therapist, in an educational setting. A pilot study with 52 participants, comprising students and healthcare professionals, yielded promising results. Specifically, 96% reported safe interactions with the AI-chatbot while 90.39% rated Luna as useful to very useful. The study also delves into the broader implications of deploying AI-driven mental health support within educational systems. Key considerations include the technical and ethical aspects of managing sensitive conversations, ensuring data privacy, and effectively escalating cases that require human intervention. By addressing these issues, the research contributes to the interdisciplinary discourse on the applications, effects, and implications of computer-based education, offering valuable insights for educators, practitioners and theorists.

Keywords: AI-chatbots; Conversational AI; Ethics; Institutional Review Board (IRB); Luna; Mental Health; Postsecondary Students; Higher Education

Introduction

In recent years, the mental health crisis among postsecondary students has emerged as a critical concern within the educational and healthcare communities [1-4]. A growing body of evidence highlights a significant uptick in mental health symptoms, including anxiety, depression, stress-related and sleep disorders, and even suicidal ideation, thereby impacting students' academic performance, quality of life, and overall well-being [5,6].

Data from a United Kingdom (UK) survey show a notable increase in students reporting a serious psychological issue from 2018 to 2020 [7]. These rates are compounded by the recent pandemic, which has been shown to have had a negative impact on student mental health, with increases in depression and alcohol-use disorder [8]. The transition to higher education introduces a unique set of stressors, compounded by the pressures of academic achievement, financial burdens, and the challenges of navigating the maturing adult life. As educators and healthcare professionals grapple with these complexities, the potential of Artificial Intelligence (AI) in offering innovative support mechanisms has come to the forefront.

AI-chatbots, designed to simulate conversational interactions, present a promising avenue for providing immediate, accessible, and non-judgmental mental health support [9,10]. With the Covid-19 pandemic accelerating healthcare services digitalization, the integration of AI technologies, such as chatbots, into postsecondary educational settings, is posited as a transformative approach to supplement traditional mental health services, thereby potentially bridging gaps in access disparity and reducing stigma associated with seeking help [11]. However, the application of AI in speculative, sensitive areas like mental health raises a plethora of ethical, privacy, and safety concerns, particularly regarding the predictability and reliability of AI responses across a range of complex human emotional states [12,13].

A critical barrier to further adoption and open integration of AI-chatbots for mental health support in educational settings is the apprehension of Institutional Review Boards (IRBs), specifically regarding the approval of proposed studies involving these technologies. IRBs play a pivotal role in ensuring the ethical conduct of research, particularly when and where it involves the rights, welfare and privacy of vulnerable populations such as students, especially those transitioning into adulthood. The uncertainty surrounding AI-chatbots' responses and the potential for unforeseen harm poses significant challenges, prompting IRBs to exercise caution [14]. This hesitance in turn reflects broader concerns within the health policy management sphere about safely and responsibly integrating AI technologies in ways that prioritize the students' welfare, rights and privacy.

In this paper, we explore the intersection of AI technology, student mental health, and health policy management, drawing upon a case study of "Luna," a virtual mental health therapist developed using OpenAI's GPT-4. Through examining the development, pilot testing, and subsequent modifications of Luna, our research team addresses the critical concerns raised by the IRB committee. More specifically, the current work identifies unique policy challenges for ethically integrating AI-chatbots in educational settings.

The paper layout is as follows. The first section reviews the literature exploring mental health challenges in higher education and strategies for adopting AI in these settings. This is followed by the methods section that first describes the development and pre-testing processes of the AI-Chatbot, Luna and later the research design encompassing using multi-methods approaches to pilot-test Luna. The results section highlights the findings from the pilot study. Before the conclusion, we also offer insights into the study's implication for health policy, its limitations, and areas for future research and a discussion of the complexities of navigating the ethical approval processes.

Literature review

Mental health concerns are significantly rising in postsecondary students over the past 10 years, with a 50% increase since 2013 [3,15]. The American College Health Association (ACHA) surveys consistently reported stress, anxiety, depression, and suicidal ideation being significant in the higher education population, with one study demonstrating an association of these mental health concerns and un-

healthy lifestyle behaviors such as poor diet and substance use [1].

Additionally, the Healthy Minds Study, which specifically focuses on mental health and wellbeing among college students, found increased rates of mental health symptoms and conditions that must be addressed within the population [3]. The Healthy Minds survey data were collected from over 350,000 students (about half the population of Vermont) at 373 campuses. Importantly, during the COVID-19 pandemic, more than 60% of students reported at least one untreated mental health symptom that they were experiencing [3]. Specific mental health concerns intensified by the pandemic include depression, substance use, eating disorders, and suicidal ideation [16-18]. Recent data from 2023 from the Center for Collegiate Mental Health (CCMH) at Penn State University also found that anxiety is the most common initial concern that students present with, with social anxiety and family distress concerns continuing to increase since their initial increase with the pandemic [19].

To date, many contributing factors have been shown to affect postsecondary students and their mental well-being. Aside from barriers or access to mental health care, other major factors include academic and financial stress [20]. Counseling provided through higher education institutions is generally effective and helpful with evidence of improvements in student academic achievement and retention [21]. However, the demand for resources far exceeds those available to help college students [22]. Notably, universities have difficulties in developing policies and methods to identify the mental disorders that college students may face and in finding solutions to address the student mental health problems within a large population [23].

For some college students, academic environments may be demanding, serving as a breeding ground for mental health issues. The unwavering pursuit of obtaining high grades in courses and the pressure to engage and excel in extracurricular activities create a constant state of stress and anxiety [5]. This is further complicated by the financial pressures of inflating tuition fees, either serviced via ballooning student loans and/or non-sustainable scholarships. Additionally, postsecondary students face peer and societal pressure that further endorses the breeding ground for mental challenges [24]. Thus, self-worth and self-doubt play a role, adding to the pressure.

For an increasing number of students, financial stress can be a significant trigger for mental health disorders [25]. Postsecondary students have the task of maintaining their academic performance and balancing the need for financial support, which can create overwhelming anxiety and worry. These worries can also be exacerbated by a background of childhood poverty. The financial burden alongside chronic stress can lead to hopelessness and helplessness. Additionally, the fear of accruing debt with the uncertainty of what type of employment they may secure after graduation is another stressor that the students may face, while they work several jobs to maintain their current lifestyle. This constant juggling between academic achievements and financial responsibilities can exacerbate feelings of depression, anxiety, and even contribute to more severe mental health issues if left unaddressed [2]. Overall well-being and mental wellness can significantly be affected by financial instability or the fear of the unknown.

When students are in an academic setting and feel pressured to succeed, fear often contributes quickly to poor mental health conditions. The fear of failing can worsen pre-existing mental health conditions. It can also cause students to develop symptoms of mental health that may not have existed. Without pre-existing support systems being accessible, students may experience the increasing burden of these detrimental conditions and circumstances, as their coping skills decline. Symptoms of depression, anxiety, or burnout can lead to a decline in student well-being and academic performance. Conversely, building student self-efficacy and confidence has been associated with improvements in academic performance, reinforcing the need for programs and services that support building these skills and attitudes [5]. Accordingly, university policies should be mandated to incorporate means of addressing college student academic pressure and foster a culture of holistic well-being to support student mental health.

Even so, stigma and barriers to accessing or seeking care are common issues that college students often encounter, exacerbating the risk of mental health disorders or

causing a decline in current ones [26]. Societal stigma persists in each community, fostering a culture of disgust and a sense of weakness if one needs to seek mental health care. Some students feel a fear of judgment and discrimination or are perceived to be mentally unstable, which can deter students from disclosing their mental health concerns or seeking professional support. As well, there may be misconceptions and a lack of education that can also contribute to stigma and barriers to treatment. Moreover, there are areas where there is limited access to affordable mental health services, long queues to schedule initial appointments, and the fear of a breach of confidentiality when seeking mental health care [27]. Altogether, these barriers contribute to a cycle of untreated mental health issues, escalating distress, and diminished academic confidence and functionality. Addressing stigma, increasing access to care, and helping people on college campuses to embrace inclusivity are key to diffusing the misconceptions related to mental health so that college students can gain the necessary care to thrive.

Finally, a transition from high school to postsecondary education can also contribute to mental health challenges that may not have existed prior or exacerbate ones that already exist [28]. Many students experience feelings of loneliness, homesickness, and a sense of disorientation as they adapt to their new environment [29]. Moreover, the pressure to succeed academically while balancing social and personal responsibilities can quickly become overwhelming and frustrating. The common factors that can lessen the intensity of the challenges would be adequate support and coping mechanisms to enhance mental well-being and overall college experience. Recognizing and addressing these challenges is crucial in providing the necessary resources on and off campus and supporting the mental health of college students during this transitional phase [29].

Mental health chatbots, for example, can provide convenient accessibility to mental health services in a variety of languages through a visual and interactive conversational process. While there still is limited research involving a widespread deployment of mental health outreach via AI, several studies have been done evaluating the use of a single AI system across various mental health metrics. One of these studies is [30]; here, the researchers demonstrated an improvement in stress as measured by the GAD-7 score for anxiety in those who completed the AI procedure. Notwithstanding, only about 40% of the study participants reached completion of the first step. As such, while these results may be promising, they indicate an inherent attention requirement to improve patient engagement, that is, a need to deploy AI-chatbots that are more user-friendly and engaging.

Some policymakers have difficulty understanding and conceptualizing the ever-evolving component of AI innovation and advancements. Further, they often have challenges grasping the immediate need of additional mental health resources on college campuses. While policymakers may advocate for AI research, it is not surprising that their ideas revolve around the ability to process the resemblance of AI to that of a human. A contributing factor to this phenomenon is that policymakers have not developed guidelines or vernacular to describe various aspects of AI technology [31-33]. Therefore, the opportunity to further explore the idea that AI could positively advance and provide immediate access to post-secondary students to decrease risks of untreated or increased amounts of mental health conditions is beneficial.

Policymakers have not been shown to understand these challenges, as there is limited data, specific guidelines and resources to guide post-secondary institutions on improving these mental health challenges of college students. If more policies and guidelines can be appropriately developed and implemented on college campuses, there may be an increase in academic success and student mental wellness, and a decrease in mental health stigma. The integration of policies that involve AI mental health chatbots are a cost-effective and sufficient way to combat these increasingly problematic mental health issues that college students are currently facing on campuses.

Materials and methods

Luna is developed to serve as a virtual mental health therapist for postsecondary students, utilizing GPT-4 AI models.

Luna development

Reaching out to orchestrate both effectiveness and safety in its interactions with users, the development pro-

cess for Luna entails several key stages to date.

Stage 1: Model training & customization

Luna's foundational model is built using GPT-4, which has been fine-tuned with a specific focus on mental health-related conversations.

Just to ascertain that Luna understands and responds appropriately to a wide range of student concerns, its customization involves prompt engineering techniques designed to train the AI on mental health scenarios, including discussions on anxiety, depression, and stress management.

Stage 2. Implementation of safety guardrails

A key feature of Luna's development is the implementation of safety guardrails, which are crafted through sophisticated prompt engineering techniques. These guardrails are designed to recognize conversations around sensitive topics surrounding anxiety, depression, suicidality and other mental-health related issues.

Exemplary scenarios of these safety guardrails are illustrated in **Figures 1.1-1.3**.

- *Anxiety and Depression***:** When conversations indicate symptoms of anxiety or depression, Luna has been programmed to recommend university counseling services. This is achieved through predefined triggers that, upon detection, will guide the conversation towards professional resources.

- *Suicidality:* For users mentioning suicidality, Luna is designed to immediately provide emergency contact information and national suicide helpline resources. This critical safety measure ensures that users will receive prompt and appropriate guidance in urgent situations.

Stage 3. AI illustrative applications

To enhance Luna's therapeutic capabilities, illustrative applications have been integrated into the system via prompt modification. These applications include therapy sessions, mindfulness exercises, and cognitive-behavioral techniques, which Luna can recommend based on the user's needs.

Each application has been carefully selected and tested to align with best practices in virtual mental health support. **Figure 2** is an example of testing a mindfulness request.

Stage 4. Pre-testing phase

Prior to deployment, Luna has also been extensively pre-tested to ensure the soundness of its AI responses. This phase involves:

- *Simulated User Interactions:* A diverse set of simulated user interactions has been created to test Luna's responses across various scenarios. These simulations help identify potential weaknesses in the AI's understanding and responsiveness.

- *Feedback Loop:* Feedback from mental health professionals and AI ethics experts has also been collected to refine Luna's responses. This iterative process incorporates multiple rounds of testing and adjustments to improve the AI's accuracy and empathy.

Once Luna development has been completed, the research team submitted an Institutional Review Board (IRB) application, and a full board review was conducted. Following the full board review, the IRB mandated a preliminary pilot study. The pilot study was required to thoroughly evaluate Luna's effectiveness, safety, and user experience before the committee could approve a decision to continue the proposed research.

The IRB's input had been especially crucial, leading to specific requirements that significantly re-shaped the study's design for the preliminary pilot. One such requirement was for pilot study participants to submit their transcripts for review, enabling a detailed assessment of Luna's interactions and ensuring adherence to safety and ethical standards.

Research design

The pilot employed a mixed-methods approach to

assess the effectiveness and safety of Luna.

This approach was adopted to leverage the strengths of both quantitative and qualitative methods, allowing for a comprehensive evaluation of Luna's performance from multiple perspectives. Participants were recruited voluntarily from a diverse group of college students and healthcare professionals to reflect a broad spectrum of users. Detailed information about the study's objectives, the function of the AI chatbot, and ethical considerations, including data privacy, was provided to the participants, and from whom informed consent was individually registered prior to study participation.

Quantitative methods

Instrument Development

Quantitative data were collected via structured surveys developed specifically for this study. The survey included Likert-scale questions designed to measure participants' perceived usefulness of Luna and the safety of their interactions.

The development of the survey instrument involved the following steps:

A. Item Generation: Questions were formulated based on existing literature on virtual mental health interventions (Abd-Alrazaq et al., 2021). For example, a question such as "On a scale of 1 to 5, how useful did you find Luna's recommendations for managing anxiety?" was included.

B. Instrument Testing: The initial survey was piloted with a small group of students to ensure clarity and relevance. Feedback from this pilot test led to minor revisions to improve question wording and format.

Data Collection

Participants completed the structured surveys after their interactions with Luna. The surveys were administered online, ensuring ease of access and convenience for all participants. Informed consent was obtained from individual participants prior to participating in the study, following ethical guidelines and approval from the Institutional Review Board (IRB).

Data Analysis

Quantitative data were analyzed via statistical analytics software. Descriptive statistics were used to summarize the data, while inferential statistics, such as t-tests, were employed to identify significant differences in perceptions based on captured responses.

Qualitative methods

Instrument Development

Qualitative feedback was gathered via open-text fields embedded within the structured surveys. This approach allowed participants to provide detailed comments on their experiences with Luna, focusing on aspects such as the appropriateness of responses, perceived empathy, and any concerns regarding privacy and safety.

A. Prompt Development: Open-ended questions were designed to elicit rich, descriptive feedback. For example, participants were asked, "Please describe any specific instances where you felt Luna's response was particularly helpful or unhelpful."

B. Pre-Testing: The open-text fields were tested with a small group to ensure they effectively captured the desired feedback.

Data Collection

Qualitative data were collected alongside the quantitative surveys. Participants were encouraged to provide as much detail as possible in their responses. Informed consent included a clause that participants agreed to provide qualitative feedback as part of their participation.

Data Analysis

Qualitative data were analyzed vis-à-vis logical thematic analysis. Responses were coded to identify recurring themes and patterns. Two independent coders reviewed the data to enhance reliability, and any discrepancies were resolved through discussion. Key themes related to the appropriateness of Luna's responses and user concerns were highlighted and documented.

Results

Participant demographics

The evaluation involved 52 voluntary participants, reflecting a diverse group with substantial representation from academia and healthcare.

Professors and healthcare professionals constituted 33.4% of the participants, showcasing the interest and relevance of Luna across professional domains. A significant majority, 76.6%, were college students from over 20 institutions, encompassing healthcare professionals such as nurses and nurse practitioners (45%), and students (55%) in fields such as Health Services Administration, Nursing, Psychology, and Behavioral Analysis.

This mix offered a broad array of perspectives on the study's subject matter.

Safety & themes of interactions

A paramount concern in the implementation of AI for mental health support is the safety and appropriateness of the interactions.

In our study, an overwhelming 96% of respondents considered their interactions with Luna to be safe (**Figure 3**). The finding indicates a high degree of trust in the chatbot's capacity to navigate sensitive topics and offer support without posing risks to users, a testament to the effectiveness of the implemented safety guardrails. These results are consistent with a meta-analysis which found that participant interactions with similar chatbots were broadly safe, with no worsening of symptoms, distress, or adverse events reports [34]. Another survey [11] demonstrates that users of chatbots integrating into existing mental health apps generally indicate high satisfaction and positive feelings regarding their interactions with such chatbots. Data was also gathered regarding the common themes students brought up in their interactions with Luna (**Figure 4**), with anxiety being the most common topic students wanted to discuss. Again, these results are consistent with the common reasons people make use of chatbots, frequently including themes of anxiety, stress, depression, and self-care as reported by [35].

Perceived usefulness of Luna

The utility of Luna as a mental health support tool was affirmed by most respondents (**Figure 5**), with 90.39% providing a usefulness rating between 3 ("useful") and 5 ("very useful"). The finding demonstrates strong confidence in the chatbot's potential to positively influence student well-being and serve as a supportive resource. These results are consistent with a recent meta-analysis which showed that AI-based chatbots were initially effective in treating anxiety and depression [36].

Another meta-analysis reported high usefulness ratings among participants from multiple studies, with benefits such as privately practicing conversations, preparing for conversations with mental health professionals, and creating a sense of self-accountability [12].

Comparison of students v. non-students

An analysis was performed comparing the responses of students v. non-students regarding themes of interactions, perceived usefulness, and perceived safety. The percentage of responses mentioning different themes from students compared with non-students are listed in **Table 1**.

For themes mentioned both individually and in combination with other themes, ‘alone’ and ‘overall’ are used as descriptors, with ‘alone’ referring to responses where the said theme has been mentioned individually, and ‘overall’ referring to any response containing the said theme, alone or in combination with others.

The frequency of select themes (overall counts for Anxiety, Depression, and Time Management) mentioned by students v. non-students were compared for statistical significance using a two-tailed t-test assuming unequal variances. No results were found to be statistically significant, with p-values all exceeding 0.05, specifically 0.811 (Anxiety), 0.324 (Depression), and 0.646 (Time Management).

Several themes were only mentioned by one of the two groups. Themes only mentioned by students include Stress, Mindfulness, Relationship, and combinations of multiple themes including Anxiety and Depression, Anxiety and Time Management, and Anxiety, Stress, and Depres-

sion. Themes only mentioned by non-students include Depression (alone), Exam Concerns, and Study Techniques.

Participants rated their perceived usefulness of Luna on a scale of 1 to 5, with 1 being very little use and 5 being very useful. A comparison of the mean usefulness rating between students v. non-students was conducted for statistical significance using a two-tailed t-test assuming unequal variances, with results shown in **Table 2**. A significant difference was found, with students rating Luna as more useful on average than non-students, with a p-value of 0.0359 ($p < 0.05$).

A more detailed analysis of the perceived safety of LUNA was conducted. Participants were asked to answer 'Yes' or 'No' regarding whether they perceived Luna to provide safe and appropriate responses during interactions. The student group responded 'Yes' 100%, while the non-student group responded 'Yes' only 86.67% of the time. These differences were not found to be statistically significant ($p = 0.164$, $p > 0.05$) after using a two-tailed t-test assuming unequal variances.

Area for improvement

Despite the overall positive reception, 9.62% of participants gave a low usefulness rating of 1 ("not useful at all") or 2 ("somewhat useful"). These ratings point to areas for improvement and the need to understand the specific shortcomings perceived by these users.

Detailed analysis of this feedback was used to refine Luna's response mechanisms to ensure that it meets the diverse needs of all student demographics. Prior research into the drawbacks of AI chatbots in mental health include perceptions of limited personalization, becoming caught in conversational loops, and inappropriate use of referrals to crisis support services, either presenting such resources when not needed or not doing so when they were [35].

Qualitative analysis of common data

Table 3 categorizes specific pieces of feedback (coded data extracts) from survey responses into broader themes. Each entry in the "Data Extract" column represents a comment from the user feedback field, aligned with a specific "Code" that identifies a key aspect or concern mentioned by the respondent. These are then linked to the "Generated Theme," which groups similar codes into meaningful categories that represent overarching insights or issues highlighted by the respondents.

Discussion

The thematic analysis of user feedback on Luna reveals critical insights into the chatbot's functionality and user satisfaction.

As separately detailed below, five main themes emerged from analyzing the coded survey responses: (1) Privacy and Data Security, (2) Response Quality & Utility, (3) User Experience Customization, (4) Accessibility of Supplementary Support, and (5) Communication Style. Together, these derived themes highlight the prevalent concerns and suggestions for improvement.

Privacy and data security

Concerns about privacy and data security were prominent among users. Participants expressed apprehension regarding the confidentiality of their interactions with Luna, questioning how data privacy is maintained. For instance, one user asked whether "questions asked by students will remain private and not become public."

This theme underscores the necessity for stringent data protection measures and transparent communication about how user data are handled, which are crucial for maintaining trust in AI systems. Concerns regarding data privacy and security are common in discussions regarding AI chatbots. Recommendations from researchers in this area include implementing proper disclosures regarding use of patient data, whether that be simply storing said data or using data to further train AI models, following data privacy legislation where appropriate, and disclosure regarding data storage and security [37].

Response quality & utility

Feedback on the quality and utility of Luna's responses was mixed. While many users found the chatbot's suggestions helpful, there was notable feedback concerning the adequacy of responses, especially regarding severe men-

tal health issues. For example, one participant highlighted the need for more than just a number to call when discussing active suicidal thoughts, suggesting a limitation of the guardrails that were implemented to specifically avoid this conversation in Luna's willingness to handle crises effectively.

Essentially, this theme indicates the importance of a policy regarding more robust support in critical situations. Other research has identified this limitation regarding the use of AI chatbots in mental health support [14,35].

User experience customization

The desire for personalized interactions emerged as a significant theme. Users expressed preferences for responses that are tailored to individual needs and circumstances, with suggestions for Luna to "remember previous conversations" to enhance personalization.

More specifically, this theme reflects the growing expectation for AI services to adapt to individual user profiles, enhancing the relevance and impact of their support. These participant experiences have been echoed in the literature [35,37], with concerns regarding genericness of chatbot responses as well as responses not being congruent with the conversation being rendered.

Accessibility of supplementary support

Under this theme, users generally recommended that Luna integrates more actionable support resources, for example, local mental health services and/or immediate helplines. As to the suggestion to "include local support resources in the chat," it clearly points to a need for Luna to offer more than generic advice; in other words, there is the need to provide users with practical, location-specific options for assistance.

Briefly, this theme highlights the potential for AI chatbots to act as gateways to a broader ecosystem of mental health resources.

Communication style

Preferences regarding the communication style of Luna were also noted. Users favored concise and clear responses, with feedback indicating that some of Luna's replies were overly lengthy and complex. As one user put it, the answers were "very lengthy... most college students would prefer shorter, more bite-sized answers."

Put simply, this theme stresses the need for AI communication to be easily digestible and adjusted to fit the typical user's attention span and information processing preferences. **Figure 6** provides an example of a useful but verbose response.

Apparently, these user characterizations of Luna are consistent with common user comments regarding communication style of chatbots. As part of a meta-analysis into user perceptions of chatbots [14], common issues brought up were confusing responses, shallow responses, and responses containing an overwhelming amount of information.

Conclusion

In shaping future health policy, the insights gained from AI pilot studies are invaluable. Policymakers must draw from such research to craft guidelines that encourage innovation yet maintain a vigilant stance on ethical considerations. Collaboration across educational and technological domains will be key to advancing mental health support without compromising on the principles of safety and privacy that are fundamental to student care.

Mental illness is quickly becoming a public health concern that is demanding mounting resources for the rising problem to be adequately addressed [34]. Moreover, the clear shortage of mental health resources in combination with poor resource allocation and mental health literacy have been accentuated in communities which are already struggling with growing limitations of various basic resources. To address this ongoing struggle, new AI solutions must now be considered, albeit to be implemented cautiously [30]. Therefore, the rest of the discussion below will offer insight into future directions, AI chatbots research limitations and the associated implications for health policy management when developing and implementing AI chatbots for mental health based on cumulative users' perspectives corresponding to their aggregated rating on Luna usefulness and perceptions vis-à-vis the existing literature.

Luna usefulness rating

The high ratings of usefulness reported by most participants reinforce the potential of AI chatbots as a complementary tool for mental health support in educational settings. These findings align with the literature that suggests AI can provide personalized, accessible support for mental well-being [35,36,38]. The safety of interactions, endorsed by 96% of participants, underscores the efficacy of the ethical guardrails and prompt engineering techniques employed during Luna's development.

Comparison with existing literature

The feedback from the qualitative remarks for the pilot exploration of Luna indicated that there was a desire for the chats to be kept private, and for the conversation to be truncated into sizes that reflect a human conversation.

Clearly, future chatbots to be developed for mental health applications must respect these concerns about privacy and the desire for more concise and actionable responses to be reflective of ongoing discussions in the field about the limitations of AI in mental health [37]. Additionally, the minority of users who rated Luna as less useful highlight the importance of user-centric design and the need for AI chatbots to adapt to the varied expectations of users, a challenge also noted in current research [12].

Implications for AI learning environments

The study's findings have significant implications for the management of unintentional learning environments and incidental learning environments within educational institutions. There is a clear need for policies that ensure the ethical deployment of AI tools, safeguarding user privacy and rights. These policies must address the diverse needs of the user population. For example, integrating AI chatbots like Luna could be part of a broader strategy to enhance student mental health services, and independent tutoring systems suggesting that policy frameworks should include provisions for the use of emerging technologies.

Recent perspectives from mental health professionals [39] indicate that a majority suggested chatbots for mental health could be helpful for their clients. These perspectives, along with feedback from research participants, suggest that chatbots could play a valuable role in unintentional learning environments. For instance, interactions with AI-driven tools like Luna can facilitate the development of emotional intelligence and self-care skills, which are essential for overall student well-being. Clear policy will be required to fully assess the most appropriate route of utilization, whether it be in combination with traditional supports or as a stand-alone resource.

These implications contribute to the broader interdisciplinary discourse on the applications, effects, and implications of computer-based education, highlighting the potential of AI-driven mental health support to create enriching unintentional learning environments. This approach ensures that technological innovations are leveraged effectively and ethically to benefit student well-being.

Limitations and areas for future research

While the pilot results are promising, they also reveal limitations that warrant further investigation. For example, the lower usefulness ratings by some participants suggest that future research should explore the customization of AI chatbot interventions to better address individual preferences and needs. Such a limitation has been echoed by previous researchers [12] and remains an area for further development in the field of chatbot development.

The recommendations from this study were used to develop a new version of Luna called Luna 2.0. Yet, there are some contradictions between feedback from the study whereby participants requested chat history be available to them so they can continue the previous conversation, while IRB would not approve holding historical chat. Another area of discrepancy between the IRB and the pilot study feedback was the request that all chats should be monitored for compliance, while the pilot study participants requested that chats remain private. Such concerns surrounding data privacy and security are a frequent topic of discussion regarding chatbots for mental health and AI in general, as brought up in previous research [37,39], particularly considering that many users mention the ability to share embarrassing or sensitive information with chatbots as a feature of the technology.

There is also a need for additional studies that fo-

cus on the long-term effectiveness of AI chatbots in supporting mental health and their integration with traditional therapeutic services. The thematic analysis reveals that while Luna is valued for its potential to support mental health, there are significant areas for improvement to fully meet user expectations. Enhancing privacy protections, refining response quality, personalizing interactions, expanding accessible resources, and optimizing communication style are critical steps forward. By addressing these various key themes, mental health specialists can enhance Luna's effectiveness and user satisfaction, thereby fulfilling its promise as a supportive tool in mental health care within educational settings.

Finally, our pilot analysis not only informs key areas for improving Luna construction and development but also contributes to broader discussions on the integration of AI in mental health services, with implications for policy, practice, and user engagement. The integration of AI tools like Luna into educational settings reveals significant potential to support student mental health, as mental health is a priority. The study's findings underscore the necessity for policies that enable safe and ethical adoption of AI chatbots, with a focus on safeguarding student privacy and well-being. Today, health policy must navigate the complexities of new technology while upholding the highest standards of data protection and ethical responsibility. The proactive involvement of IRBs exemplifies the rigorous scrutiny required to ensure these standards are met.

Acknowledgments

There are no acknowledgments from the authors.

Declaration Statement

The authors report there are no competing interests to declare.

Institutional Review Board (IRB) Statement

This pilot study was conducted in compliance with the ethical standards set forth by the Institutional Review Board (IRB). The IRB approval was sought prior to the commencement of the study, ensuring that all research activities adhered to the ethical guidelines for human subjects' research. IRB # 23-24-38. The pilot study was mandated as part of the IRB's approval decision process to evaluate the feasibility, safety, and ethical considerations of the proposed research before progressing to a larger-scale study with students. All participants provided informed consent, and their confidentiality and welfare were prioritized throughout the study. The findings from this pilot study will inform future research phases and contribute to the refinement of our research protocols.

References

1. Jao NC, Robinson LD, Kelly PJ, Ciecierski CC, Hitsman B (2019) Unhealthy behavior clustering and mental health status in United States college students. Journal of American college health: J of ACH, 67: 790–800.

2. Sheldon E, Simmonds-Buckley M, Bone C, Mascarenhas T, Chan N, Wincott M, Barkham M (2021) Prevalence and risk factors for mental health problems in university undergraduate students: A systematic review with meta-analysis. Journal of affective disorders, 287: 282-92.

3. Lipson SK, Zhou S, Abelson S, Heinze J, Jirsa M, Morigney J, Patterson A, Singh M, Eisenberg D (2022) Trends in college student mental health and help-seeking by race/ethnicity: Findings from the national healthy minds study, 2013–2021. Journal of Affective Disorders, 306: 138-47.

4. Moghimi E, Stephenson C, Gutierrez G, Jagayat J, Layzell G, Patel C, Alavi N (2023) Mental health challenges, treatment experiences, and care needs of post-secondary students: a cross-sectional mixed-methods study. BMC public health, 23: 655.

5. Hitches E, Woodcock S, Ehrich J (2022) Building self--efficacy without letting stress knock it down: Stress and academic self-efficacy of university students. International Journal of Educational Research Open, 3: 100124.

6. Limone P, Toto GA (2022) Factors that predispose undergraduates to mental issues: A cumulative literature review for future research perspectives. Frontiers in Public Health, 10: 831349.

7. Campbell F, Blank L, Cantrell A, Baxter S, Blackmore C, Dixon J, Goyder E (2022) Factors that influence mental health of university and college students in the UK: a systematic review. BMC Public Health, 22: 1778.

8. Kim H, Rackoff GN, Fitzsimmons-Craft EE, Shin KE, Zainal NH, Schwob JT, Newman MG (2022) College mental health before and during the COVID-19 pandemic: Results from a nationwide survey. Cognitive Therapy and Research, 46: 1-10.

9. Dekker I, De Jong EM, Schippers MC, De Bruijn-Smolders M, Alexiou A, Giesbers B (2020) Optimizing students' mental health and academic performance: AI-enhanced life crafting. Frontiers in Psychology, 11: 1063.

10. van der Schyff EL, Ridout B, Amon KL, Forsyth R, Campbell AJ (2023) Providing self-led mental health support through an artificial intelligence–powered chat bot (Leora) to meet the demand of mental health care. Journal of Medical Internet Research, 25: e46448.

11. Boucher EM, Harake NR, Ward HE, Stoeckl SE, Vargas J, Minkel J, Zilca R (2021) Artificially intelligent chatbots in digital mental health interventions: a review. Expert Review of Medical Devices, 18: 37-49.

12. Abd-Alrazaq AA, Alajlani M, Ali N, Denecke K, Bewick BM, Househ M (2021) Perceptions and opinions of patients about mental health chatbots: scoping review. Journal of Medical Internet Research, 23: e17828.

13. Hamdoun S, Monteleone R, Bookman T, Michael K (2023) AI-based and digital mental health apps: Balancing need and risk. IEEE Technology and Society Magazine, 42: 25-36.

14. Abdelhafiz AS, Ali A, Maaly AM, Ziady HH, Sultan EA, Mahgoub MA (2024) Knowledge, Perceptions and Attitude of Researchers Towards Using ChatGPT in Research. Journal of Medical Systems, 48: 1-11.

15. American College Health Association. (2021). National College Health Assessment (Fall 2021).

16. Dogan-Sander E, Kohls E, Baldofski S, Rummel-Kluge C (2021) More depressive symptoms, alcohol and drug consumption: Increase in mental health symptoms among university students after one year of the COVID-19 pandemic. Frontiers in Psychiatry, 12: 790974.

17. Tavolacci MP, Ladner J, Déchelotte P (2021) Sharp increase in eating disorders among university students since the COVID-19 pandemic. Nutrients, 13: 3415.

18. Yan Y, Hou J, Li Q, Yu NX (2023) Suicide before and during the COVID-19 pandemic: A systematic review with meta-analysis. International Journal of Environmental Research and Public Health, 20: 3346.

19. Center for Collegiate Mental Health (CCMH). (2024, January). 2023 Annual Report (Publication No. STA 24-147)

20. Auerbach RP, Mortier P, Bruffaerts R, Alonso J, Benjet C, Cuijpers P, Kessler RC (2018) WHO world mental health surveys international college student project: Prevalence and distribution of mental disorders. Journal of Abnormal Psychology, 127: 623.

21. Priestley M, Broglia E, Hughes G, Spanner L (2022) Student perspectives on improving mental health support services at university. Counselling and Psychotherapy Research, 22.

22. Batchelor R, Pitman E, Sharpington A, Stock M, Cage E (2020) Student perspectives on mental health support and services in the UK. Journal of Further and Higher Education, 44: 483-97.

23. Kang HK, Rhodes C, Rivers E, Thornton CP, Rodney T (2021) Prevalence of mental health disorders among undergraduate university students in the United States: A review. Journal of Psychosocial Nursing and Mental Health Services, 59: 17-24.

24. Wentzel KR, Jablansky S, Scalise NR (2021) Peer social acceptance and academic achievement: A meta-analytic study. Journal of Educational Psychology, 113: 157-80.

25. Mofatteh M (2020) Risk factors associated with stress, anxiety, and depression among university undergraduate students. AIMS public health, 8: 36–65.

26. Shim YR, Eaker R, Park J (2022) Mental health education, awareness and stigma regarding mental illness among college students. Journal of Mental Health & Clinical Psychology, 6: 6-15.

27. Moroz N, Moroz , D'Angelo MS (2020) Mental health services in Canada: barriers and cost-effective solutions to increase access. In Healthcare management forum, 33: 282-7.

28. Park SY, Andalibi N, Zou Y, Ambulkar S, Huh-Yoo J (2020) Understanding students' mental well-being challenges on a university campus: interview study. JMIR formative research, 4: e15962.

29. Kroshus E, Hawrilenko M, Browning A (2021) Stress, self-compassion, and well-being during the transition to college. Social Science & Medicine, 269: 113514.

30. Daley K, Hungerbuehler I, Cavanagh K, Claro HG, Swinton PA, Kapps M (2020) Preliminary evaluation of the engagement and effectiveness of a mental health chatbot. Frontiers in digital health, 2: 576361.

31. Krafft PM, Young M, Katell M, Huang K, Bugingo G (2020) Defining AI in policy versus practice. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 72-8.

32. Meskó B, Topol EJ (2023) The imperative for regulatory oversight of large language models (or generative AI) in healthcare. NPJ digital medicine, 6: 120.

33. Zidaru T, Morrow EM, Stockley R (2021) Ensuring patient and public involvement in the transition to AI-assisted mental health care: A systematic scoping review and agenda for design justice. Health Expectations, 24: 1072-124.

34. Abd-Alrazaq AA, Rababeh A, Alajlani M, Bewick BM, Househ M (2020) Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. Journal of medical Internet research, 22: e16021.

35. Haque MR, Rubya S (2023) An overview of chatbot-based mobile mental health apps: insights from app description and user reviews. JMIR mHealth and uHealth, 11: e44838.

36. Zhong W, Luo J, Zhang H (2024) The therapeutic effectiveness of artificial intelligence-based chatbots in alleviation of depressive and anxiety symptoms in short-course treatments: A systematic review and meta-analysis. Journal of Affective Disorders.

37. Coghlan S, Leins K, Sheldrick S, Cheong M, Gooding P, D'Alfonso S (2023) To chat or bot to chat: Ethical issues with using chatbots in mental health. Digital health, 9: 20552076231183542.

38. Schick A, Feine J, Morana S, Maedche A, Reininghaus U (2022) Validity of chatbot use for mental health assessment: experimental study. JMIR mHealth and uHealth, 10: e28082.

39. Sweeney C, Potts C, Ennis E, Bond R, Mulvenna MD, O'neill S, Mctear MF (2021) Can chatbots help support a person's mental health? Perceptions and views from mental healthcare professionals and experts. ACM Transactions on Computing for Healthcare, 2: 1-15.