Eating disorder chatbots: Current status, challenges, and future directions

Elaine Xiao

Laurel Heights Secondary School, Waterloo, Canada

elainexiao168@gmail.com

Abstract. Eating disorders (ED) have the highest case mortality rate among all mental health issues, and their prevalence has increased significantly in recent years. However, fewer than 20% of sufferers receive treatment. Chatbots, or conversational agents, that target EDs are a promising way to bridge this gap and supplement existing support by being cost-effective, accessible, and non-human. The technical compositions behind these chatbots vary, each with their own strengths and weaknesses. Generally, the field of ED chatbots is emerging and only a handful of platforms have come out, including one designed to replace ED helpline support and one to support youth body image. While many ethical and practical challenges affect ED chatbots, there are ways to make progress, and the collaboration between developers, clinicians, ethicists, and the public will be vital in this pursuit. Integrating ED chatbots into healthcare and social media platforms are future directions that could significantly advance ED healing.

Keywords: eating disorders, chatbot, mental health, body image, artificial intelligence.

1. Introduction

Eating disorders are harrowing mental health conditions characterized by obsessive concern with one's food intake and body appearance, coupled with strict behaviours to manage these aspects of oneself [1]. Their causes include biological, psychological, and/or socio-cultural factors—such as the internalization of pervasive cultural beauty ideals—and they can afflict anyone regardless of age, race, gender, and weight [1]. The impact of these conditions is concerning; more than 28.8 million Americans will have an ED in their lifetime [2], and EDs have some of the highest mortality rates among all mental illnesses [2]. Despite the significant consequences and rising prevalence of EDs [3], fewer than 20% of those affected obtain treatment [4]. Fortunately, chatbots that aid in ED prevention and recovery have potential to alleviate the cruel ED experience.

Chatbots, or conversational agents, are technologies that interact with users through natural language, and their adept communication ability and autonomous nature prompt their wide-scale usage today in fields such as mental health [5]. Specifically, the development of ED chatbots is emerging, and a few have come out in recent years [6]. ED chatbots can provide benefits by acting as prophylactic measures against EDs, psychoeducational resources, spaces for self-expression [7], and/or tools to aid clinicians in their work [8].

There are still several improvements needed for these technologies to function effectively and responsibly. Nevertheless, chatbots are uniquely suited for eating disorder-related support due to their 24/7 availability to a large population, reaching traditionally disadvantaged groups, such as those that

are geographically isolated or limited in mobility [9]. As well, their synthetic, non-judgmental nature makes them approachable mediums for support [10], which is important as stigmatization often discourages higher-weight, minority, and male-identifying individuals who face an ED from seeking aid [8]. When these technologies can successfully foster a pressure-free and self-paced environment, they can subvert the underlying ego-syntonic nature of EDs that prevents individuals from pursuing recovery [11]. Crucially, ED chatbots can bridge the gap by providing ED support to the 80% of affected individuals who are currently not receiving any.

This review will first explore the current status of ED chatbots. Next, it will examine both the ethical and practical challenges surrounding them. Thirdly, future directions for these technologies will be offered.

2. Current Status

2.1. Technical Designs Behind Chatbots

Chatbots can function in either a rule-based or non-rule-based format. Rule-based chatbots take user inputs and follow a hand-curated, defined set of rules in order to return a predetermined output [4]. As such, these chatbots are simpler to create than non-rule-based ones and preclude inappropriate behaviour [4, 12]. However, there are many drawbacks to this approach: their non-dynamic nature can make conversations feel rigid, they are unable to address user responses that have not been accounted for in their design (e.g., queries with spelling errors), and generally, the conversation domain in rule-based design is constrained [4, 12].

The opposing, non-rule-based approach to chatbots is more novel and involves generative artificial intelligence (Gen AI). Models learn semantic patterns through data, broaden their learning autonomously through machine learning, and generate new content [4]. Non-rule-based chatbots require considerable technical expertise to create, but outcomes are very dynamic and human-like [4]. These chatbots can, however, produce inaccurate results and respond inappropriately and/or harmfully [4]. A clear example is when the chatbot Tessa advised harmful and potentially triggering things to its users [13]. There are thus pros and cons to both rule-based and non-rule-based approaches, as one has less liability but is robotic, while the other bestows greater engagement at the risk of potential inaccuracies and harm.

To get the best of both worlds while minimizing respective weaknesses, there are exploratory technical methods that incorporate both rule-based and non-rule-based approaches. One method that is similar to but more flexible than rule-based design is retrieval-based AI [14]. When providing replies, this model considers a repertoire of pre-set responses—like in rule-based design—and uses advanced matching algorithms to determine the best response—demonstrating non-rule-based behaviour [14]. Retrieval-based models are not as dynamic as generative AI models but are comparatively more flexible in handling queries than a purely rule-based approach [14]. Another novel approach is called neuro-symbolic AI [15]. With this approach, the pattern recognition and independent learning of non-rule-based methods, called neuro AI, work with the accuracy and structured simplicity of rule-based methods, called symbolic AI, to create sophisticated results [15]. In the future, hybrid methods involving rule-based and non-rule-based logic can potentially improve ED chatbots.

2.2. Eating Disorder Chatbots Today

Currently, 1 in 41 mental health chatbots targets EDs [12], and only a handful of ED chatbots have been developed in recent years [13]. The ED chatbots Tessa, KIT, Alex, and Topity will be discussed in the following paragraphs.

Tessa is one of the most well-known ED chatbots to emerge, having received attention after it gave inappropriate suggestions to users and was swiftly taken down [13]. Initially designed as the rule-based adaptation of the Student Bodies behavioural program, Tessa discussed topics from media literacy to recovery maintenance [5]. Shortly thereafter, Tessa gained generative capabilities and started to dispense harmful advice, including weight loss suggestions [13]. The fact that the chatbot's

problematic behaviour spurred after gaining non-rule-based capabilities illustrates how without adequate guardrails, non-rule-based chatbots can induce harm.

KIT was a rule-based chatbot designed to target negative body image while supporting their caregivers [16]. Users could choose from a set of possible responses in the conversation, and the chatbot would respond with a predetermined reply according to a decision tree [16]. KIT's content was based on evidence-based ED intervention methods such as psychoeducation, cognitive behavioural therapy, and mindfulness [16]. The developers have published an adapted version of KIT, called JEM, that is currently available [17].

The chatbot prototype Alex was administered to a sample group who screened positive from an eating disorder screen, and focused on facilitating their motivation to engage with treatment [13].

Lastly, a gamified chatbot, Topity, was developed in partnership with UNICEF to help Brazilian youth gain self-esteem, media-evaluation skills, and resilience to combat negative body image thoughts [9].

3. Challenges & Limitations

3.1. Ethics
Current conversational agents lack the ability to fully comprehend the complex human experience, making it difficult for them to consistently meet users' emotional needs [6]. This limitation often results in neglect, undermining the medical ethics of beneficence and non-maleficence. One can observe this especially in rule-based chatbots because they predicate on predetermined replies [6] and are constricted in their ability to provide meaningful support. To illustrate an instance of unintended harm, we can look at an interaction between a user and the initial, rule-based version of Tessa. When a user expressed feelings of self-hatred and isolation, Tessa replied, "Keep on recognizing your great qualities! Now, let's look deeper into body image beliefs [5]." It failed to address the user's concern, violating beneficence, and may have reinforced negative beliefs through its positive tone [5].

Non-rule-based have their challenges as well, especially with non-maleficence; they may generate unanticipated, inappropriate responses. This was demonstrated when the newer, non-rule-based model of Tessa gave numerical calorie deficit recommendations to a user, exacerbating a harmful behaviour that can fuel an ED [13].

Another large ethical concern is ensuring that user data is handled safely and responsibly, particularly given the sensitive mental health context in which ED chatbots operate and how they may encourage users to share personal details to foster better therapeutic outcomes [5]. Furthermore, algorithmic bias, which can stem from insufficient or unrepresentative data sets [8], can create inequitable outcomes in non-rule-based chatbots [5, 8].

3.2. Design
Generally, the current capabilities of ED chatbots are limited, and further developments in design are needed. One design challenge is ensuring that chatbot navigation is non-overwhelming and that content is relevant to user needs [7]. Further, navigating chatbot tone and personality requires careful action because, on one hand, being friendly and personable helps users feel engaged and open sharing, but being too human-like can make the experience unsettling [18]. Lastly, it is an ongoing pursuit and challenge to develop more sophisticated, hybrid chatbots (e.g., using neuro-symbolic models) that can resolve current challenges while creating an improved experience.

3.3. Inclusion
Even though EDs and ED-related concerns do not discriminate, there is a lack of diversity in research and development related to ED chatbots [8, 9]. For example, Matheson et al. highlight that body image research has largely centred on Caucasian populations with a relatively high socioeconomic status [9], and Fardouly et al. state that "Current [machine learning] research in the field of eating disorders primarily focuses on samples of young white females, limiting the generalisability of existing models

to detect risk in other relevant demographics [8]." Moreover, the creation of chatbots is often driven by groups with considerable power, such as Big Tech and government agencies [5]. Without factoring in the diverse needs and psychosocial experiences of different groups—especially those traditionally underserved with regard to race, gender, geographic location, and socioeconomic status—ED chatbots may not benefit these populations as greatly; they will fall short of understanding the different ways in which different groups experience technologies [5].

4. Future Directions and Recommendations

4.1. Limiting Ethical Harm

Emerging disciplines such as neuro-symbolic AI, retrieval-based AI, and affective computing [5] can enhance the reasoning capabilities of ED chatbots and minimize unintended, unethical harm. To combat the risky nature of data protection, it is crucial to have strong regulatory frameworks [19], inform users about how their data is being handled, and develop procedures for storing and/or destroying user data once a chatbot is no longer in use [5]. To add, an exploratory idea that may enhance user protection is establishing labels or certifications to differentiate secure digital interventions from those that are non-therapeutic or low in security [19]. Designers should anticipate and manage risk throughout the duration of a chatbot's lifespan to remove unethical behaviour [20].

4.2. Chatbot Development and Design

When developing ED chatbots, it is essential to consider both the content and how it is delivered. A content improvement for chatbots would be to promote help-seeking behaviours, as it is often difficult for sufferers to feel motivated about pursuing recovery [11]. This may be achieved through conducting motivational interviewing [7] and creating a pressure-free, self-paced experience [11]. Also, that chatbots promote self-disclosure is recommended as this is a key aspect in emotional relief and fostering a positive human-agent relationship [10]. Conversational agents can encourage self-disclosure by sharing their own "thoughts" and "feelings," prompting a reciprocal effect [10]. Chatbots' content can be improved for delivery by, for instance, having information appear in short bubbles akin to a text message and nesting text under toggles to help limit the amount of words on screen, simultaneously encouraging self-paced exploration [16]. Concise content delivery is crucial as it helps stressed users digest information more easily [16].

Given that over 70% of ED sufferers experience comorbid mental illnesses [21], a future direction may be to enable mental health chatbots to detect EDs alongside other conditions [17]. These chatbots could work to alleviate associated symptoms and/or direct users to further resources. This is exploratory as most existing mental health conversational agents focus on one specific area of mental health rather than multiple [22]. More scientific exploration is needed to understand the implications and process of creating mental health chatbots that can detect and resolve more than one psychological concern, including eating disorders.

Lastly, while eating disorders affect anyone, we should consider how to help children and adolescents in the context of ED chatbots. Almost 1 in 4 children and adolescents experience disordered eating [4], and their heavy social media usage makes them vulnerable to online health and beauty messaging. In the future, ED chatbots that target youth can include stimuli evoking the senses, such as audio or visual aids, increasing engagement [9]. A pictorial character such as that from the chatbot KIT can establish a positive interpersonal experience between the chatbot and users [16], and furthermore, gamifying the experience can make it more enjoyable [9]. By eventually deploying these tools on social media and using online spaces to promote them, youth can be easily reached—an action that developers behind Topity plan to pursue [9]. These concepts may be applied for the creation of ED chatbots for all audiences.

The previously stated suggestions are rudimentary and should be validated further to be considered worthwhile elements of ED chatbots. There is still limited literature on the process of designing mental

health and ED chatbots [4, 7], and as the problems they address becomes ever more prevalent, their R&D must grow similarly.

4.3. Promoting Inclusion

To ensure greater accessibility and justice for ED chatbots, researchers have called for 'interdisciplinary empirical research on the implications of these technologies that centres the experiences and knowledge of those who will be most affected,' factoring in diversity with regards to race, socioeconomic status, gender, and more [5]. To prevent inequitable outcomes, training data should be sufficiently diverse and abundant to prevent algorithmic bias, and developers should involve patient and public perspectives into chatbot development [5]. It is imperative to sustain ongoing efforts to engage underserved suffering groups, such as developing countries, where there is a shortage of mental health professionals [12] and where technology can serve as a cost-effective tool [4] and/or medical adjunct.

4.4. Future Applications of Eating Disorder Chatbots

ED chatbots are gaining interest in their potential integration into healthcare practice. For instance, these chatbots can be valuable tools outside of the clinic, helping to gather continuous assessment on patients and aiding with therapeutic homework [7, 23]. Researchers posit that conversational agents could carry out multiple aspects of psychotherapy for therapists, saving time and enabling greater productivity [19]. Given the limited capabilities of ED chatbots, they are being proposed as potential supplements—not replacements—to traditional medical action [7, 19]. The implications of ED chatbots in healthcare are experimental, and it will take significant research and time [7, 12] before they can be confidently incorporated into healthcare practice. For now, they will be important interventions and stepping stones of support for those who need it.

Social media can be a vehicle for deploying ED chatbots. Previously, ED chatbots have been hosted on the social platform Facebook Messenger [5, 9], leveraging its popularity. This approach may be successful as social media content is often imbued with diet culture, initiating ED-related concerns; therefore, less time may elapse between when someone may be struggling online and when they can receive immediate support [16]. There are also fewer resources for users and companies to invest in if these tools are deployed on existing applications [24]. Fardouly et al. support the exploration of social media for ED tools, saying, "there is potential for chatbots to be used on social media, and other platforms, to provide initial advice, prevention strategies, and support for individuals and carers, and to refer people to other effective evidence-based ED prevention and treatment options" [8].

5. Conclusion

ED chatbots may transform support for affected individuals and prevent the onset of these serious mental concerns, but we must address existing shortcomings to ensure their effective and responsible usage. Regulatory frameworks, transparency on the behalf of developers, and improved technical designs can mitigate ethical and practical concerns regarding ED chatbots. Improving their engagement and making these tools accessible for underserved populations amid prevalent barriers of cost, stigma, and geographic location is vital. Using Arthur Clarke's Third Law as a guiding principle, "Any sufficiently advanced technology is indistinguishable from magic" [25]. With the continued work of researchers and the powerful collaboration between technologists, clinicians, ethicists, and the public [17], we hope to realize the full potential of conversational tools in alleviating the cruel ED experience.

References

[1] Balasundaram, P., & Santhanam, P. (2023). Eating disorders. StatPearls [Internet]. https://www.ncbi.nlm.nih.gov/books/NBK567717/

[2] Deloitte Access Economics. (2020). The social and economic cost of eating disorders in the United States of America. Harvard T.H. Chan School of Public Health.

https://www.hsph.harvard.edu/striped/wp-content/uploads/sites/1267/2020/07/Social-Economic-Cost-of-Eating-Disorders-in-US.pdf

[3] López-Gil, J. F., García-Hermoso, A., Smith, L., Firth, J., Trott, M., Mesas, A. E., Jiménez-López, E., Gutiérrez-Espinoza, H., Tárraga-López, P. J., & Victoria-Montesinos, D. (2023a). Global proportion of disordered eating in children and adolescents. JAMA Pediatrics, 177(4), 363. https://doi.org/10.1001/jamapediatrics.2022.5848

[4] Chan, W. W., Fitzsimmons-Craft, E. E., Smith, A. C., Firebaugh, M.-L., Fowler, L. A., DePietro, B., Topooco, N., Wilfley, D. E., Taylor, C. B., & Jacobson, N. C. (2022). The challenges in designing a prevention chatbot for eating disorders: Observational study. JMIR Formative Research, 6(1). https://doi.org/10.2196/28003

[5] 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. https://doi.org/10.1177/20552076231183542

[6] Shah, J., DePietro, B., D'Adamo, L., Firebaugh, M., Laing, O., Fowler, L. A., Smolar, L., Sadeh‐Sharvit, S., Taylor, C. B., Wilfley, D. E., & Fitzsimmons‐Craft, E. E. (2022). Development and usability testing of a chatbot to promote mental health services use among individuals with eating disorders following screening. International Journal of Eating Disorders, 55(9), 1229–1244. https://doi.org/10.1002/eat.23798

[7] Haque, M. D., & Rubya, S. (2023). An overview of chatbot-based mobile mental health apps: Insights from app description and user reviews. JMIR mHealth and uHealth, 11. https://doi.org/10.2196/44838

[8] Fardouly, J., Crosby, R. D., & Sukunesan, S. (2022). Potential benefits and limitations of machine learning in the field of eating disorders: Current research and Future Directions. Journal of Eating Disorders, 10(1). https://doi.org/10.1186/s40337-022-00581-2

[9] Matheson, E. L., Smith, H. G., Amaral, A. C., Meireles, J. F., Almeida, M. C., Mora, G., Leon, C., Gertner, G., Ferrario, N., Suarez Battan, L., Linardon, J., Fuller-Tyszkiewicz, M., & Diedrichs, P. C. (2021). Improving body image at scale among Brazilian adolescents: Study protocol for the co-creation and randomised trial evaluation of a chatbot intervention. BMC Public Health, 21(1). https://doi.org/10.1186/s12889-021-12129-1

[10] Lee, Y.-C., Yamashita, N., Huang, Y., & Fu, W. (2020). "I hear you, I feel you": Encouraging deep self-disclosure through a chatbot. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3313831.3376175

[11] 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

[12] Abd-alrazaq, A. A., Alajlani, M., Alalwan, A. A., Bewick, B. M., Gardner, P., & Househ, M. (2019a). An overview of the features of Chatbots in Mental Health: A scoping review. International Journal of Medical Informatics, 132, 103978. https://doi.org/10.1016/j.ijmedinf.2019.103978

[13] Sharp, G., Dwyer, B., Xie, J., McNaney, R., Shrestha, P., Prawira, C., Fernando, A. N., de Boer, K., & Hu, H. (2024). Co-Design of a Single Session Intervention Chatbot for People on Waitlists for Eating Disorder Treatment: A Qualitative Interview and Workshop Study. ms, Melbourne.

[14] Adamopoulou, E., & Moussiades, L. (2020). An overview of chatbot technology. In I. Maglogiannis, L. Iliadis, & E. Pimenidis (Eds.), Artificial intelligence applications and innovations (AIAI 2020, IFIP Advances in Information and Communication Technology, Vol. 584). Springer, Cham. https://doi.org/10.1007/978-3-030-49186-4_31

[15] Barnes, E., & Hutson, J. (2024a). Natural language processing and neurosymbolic ai: The role of neural networks with knowledge-guided symbolic approaches. DS Journal of Artificial Intelligence and Robotics, 2(1), 1–13. https://doi.org/10.59232/air-v2i1p101

[16] Beilharz, F., Sukunesan, S., Rossell, S. L., Kulkarni, J., & Sharp, G. (2021). Development of a positive body image chatbot (KIT) with young people and parents/carers: Qualitative Focus Group Study. Journal of Medical Internet Research, 23(6). https://doi.org/10.2196/27807

[17] Sharp, G., Torous, J., & West, M. L. (2023). Ethical challenges in AI approaches to eating disorders. Journal of Medical Internet Research, 25. https://doi.org/10.2196/50696

[18] Ta, V., Griffith, C., Boatfield, C., Wang, X., Civitello, M., Bader, H., DeCero, E., & Loggarakis, A. (2020a). User experiences of social support from companion chatbots in everyday contexts: Thematic analysis. Journal of Medical Internet Research, 22(3). https://doi.org/10.2196/16235

[19] Bendig, E., Erb, B., Schulze-Thuesing, L., & Baumeister, H. (2019). The next generation: Chatbots in clinical psychology and psychotherapy to Foster Mental Health – a scoping review. Verhaltenstherapie, 32(Suppl. 1), 64–76. https://doi.org/10.1159/000501812

[20] Bowie-DaBreo, D., Sas, C., Iles-Smith, H., & Sünram-Lea, S. (2022a). User perspectives and et hical experiences of apps for Depression: A qualitative analysis of user reviews. CHI Confer ence on Human Factors in Computing Systems. https://doi.org/10.1145/3491102.3517498

[21] Juli, R., Juli, M. R., Juli, G., & Juli, L. (2023). Eating Disorders and Psychiatric Comorbidity. Psychiatria Danubina, 35(Suppl. 2), 217–220.

[22] Batterham, P. J., Calear, A. L., Farrer, L., McCallum, S. M., & Cheng, V. W. (2017). FitMindKit : Randomised controlled trial of an automatically tailored online program for mood, anxiety, substance use and suicidality. Internet Interventions, 12, 91–99. https://doi.org/10.1016/j.invent.2017.08.002

[23] Dingler, T., Kwasnicka, D., Wei, J., Gong, E., & Oldenburg, B. (2021). The use and promise of conversational agents in Digital Health. Yearbook of Medical Informatics, 30(01), 191–199. https://doi.org/10.1055/s-0041-1726510

[24] Nimavat, K., & Champaneria, T. (2017). Chatbots: An overview types, architecture, tools and future possibilities. Int. J. Sci. Res. Dev, 5(7), 1019-1024.

[25] Prucher, J. (2007). The Oxford Dictionary of Science Fiction by Jeffrey Prucher. Oxford Univ. Press.