Journal of Inclusive Methodology and Technology in Learning and Teaching
ISSN 2785-5104
Anno4 n.1(2024)

EDIZIONI UNIVERSITARIE ROMANE
editoria scientifica dal 1952

Using Chatbots And Conversational Agents For The Promotion Of Well-Being And Mental Health In Adolescents: Limitations And Perspectives.

Beatrice Tosti[1], Stefano Corrado[1], Stefania Mancone[1*]

[1] Department of Human Sciences, Society and Health, University of Cassino and Southern Lazio; beatrice.tosti@unicas.it; stefano.corrado@unicas.it; s.mancone@unicas.it
* Correspondence: s.mancone@unicas.it

Abstract: Adolescence is a phase of development that brings with it numerous changes and critical issues, affecting the personal sphere, both physical and emotional-psychological, and the social sphere. These changes increase the risk of developing mental health problems, which is why interventions are needed not only for treatment, but also and above all for prevention, in order to increase the well-being of adolescents and minimise the risk of developing mental disorders. However, very often adolescents are reluctant to seek help from an adult, for instance for fear that their confidences might be reported to a third party or for fear of being judged, so they prefer not to embark on a therapeutic journey. Nevertheless, being adolescents the biggest consumers of the so-called digital world, a promising avenue for the implementation of interventions for mental health support is through their smartphones, by means of chatbots or conversational agents. These digital platforms use artificial intelligence to train neural networks with basic knowledge of mental health problems in order to simulate conversation with another person and help them in the management and support of their mental distress. This article will highlight the advantages and limitations of this technology in terms of privacy, effectiveness and safety, in order to promote ethically aware and responsible development.

Keywords: Adolescents; Well-being; Mental health; Chatbots; Conversational agents

1. Introduction

Adolescence is a developmental stage full of changes in the body, social environment and mind, crucial for psychological, social and emotional well-being and, therefore, characterised by an increased risk of developing mental health problems, such as anxiety, depression, substance abuse and eating disorders (Blakemore, 2019). The increase in this risk makes adolescence a crucial phase for the prevention of mental health problems, as well as their treatment, and emphasises the need for ac-

cessible and personalised tools aimed at increasing adolescents' well-being. Well-being status is defined as an individual's «overall affective and cognitive evaluation of his or her life» (Diener, Oishi & Lucas, 2009) and is determined by high levels of positive emotions, low levels of negative emotions and high satisfaction with one's life (Diener, 1984; Diener, Oishi & Lucas, 2009; González-Carrasco et al., 2017). Positive emotions improve psychological resilience (Fredrickson, 2001) and lead to better adaptations in the areas of school and work, relationships, cooperation, creativity and problem solving (Lyubomirsky, King & Diener, 2005). Furthermore, it has been shown that high levels of well-being can act as a defence against negative outcomes such as psychological disorders and can be a protective factor against depression and suicide (Park, 2004; Ryff, 2017).

Studies have found that the well-being of adolescents, often measured by satisfaction with school, family, friends, physical appearance and life as a whole, has been shown to be unstable over time (Patalay & Fitzsimons, 2018), indicating the need for interventions to promote well-being and mental health in this population. As adolescents are very often reluctant to seek help from an adult, and are also the biggest users of the so-called digital world, they represent one of the main destination groups for wellness-promoting digital interventions, so a promising way to implement these interventions is by means of their smartphones (Burns, 2010). Over the past decade, the use of digital technology has been strongly integrated into almost everyone's social, personal and working life, and more recently there has been an increase in the development of various applications for e-Mental Health interventions (Bakker et al, 2016), which aim either to integrate traditional mental health treatments or to provide support for those who are unable to receive quality services, e.g. due to long waiting lists (Bakker et al., 2016), and have proven to be beneficial for the adult population (Bakker et al., 2016; Fitzpatrick, Darcy & Vierhile, 2017; Weisel et al., 2019). Many students turn to the Internet to vent their negative feelings and this has been shown to have positive effects on loneliness, depression and stress (Shaw, 2004). However, there is a shortage of online support that is interactive, efficient and tailored to the specific needs of adolescents, such as the chatbot (Mierzwa et al., 2019), which can serve as a virtual friend to help adolescents express emotions, thoughts and feelings that they would not otherwise be able to externalise in real life.

The chatbot for mental health support is a digital technology that uses artificial intelligence (AI) to instruct neural networks with basic knowledge about mental health problems in order to simulate conversation with another person and help them in the management and support of their mental distress. The first generation of digital interventions for mental health support (e.g., ELIZA; Weizenbaum, 1967) did not show complete fidelity to what are the real therapist-patient interactions (Christensen, Griffiths & Farrer, 2009; Fitzpatrick, Darcy & Vierhile, 2017) and had mixed results in decreasing anxiety and depression symptoms in samples of young adolescents and

EDIZIONI UNIVERSITARIE ROMANE
editoria scientifica dal 1952

pre-adolescents (Ebert, Zarski & Christensen, 2015; Grist, Porter & Stallard, 2017). Therefore, in order to improve adherence to real human relationships and evaluate their therapeutic efficacy, new platforms for mental health support have been developed that are designed so that the user has the feeling of talking to a real human being (Ly, Ly & Andersson, 2017), such as the chatbots Woebot, Wysa, X2 and Youper (see sitography for links to access each). These text-based applications work out both the text and the emoji used by the user, but also the typical features of non-verbal language (facial expression, movements, gestures, mimicry, variations in the tone of voice, etc., as Simsensei does, for example; Rizzo et al., 2016) and provide guided talks and tips to help people deal with and manage their psychological discomfort. In addition, these platforms offer everyday check-ins on the user's conduct, thoughts and feelings, who then has the opportunity to get reports that can support them in understanding their own mental states. Finally, they provide tailored therapeutic exercises, such as remodelling one's thoughts to take on a different perspective, breathing exercises and motivational talks, either as text, games or videos.

What are the benefits and drawbacks of using chatbots for mental health assistance?

2. Advantages

The last two decades have witnessed a proliferation of mental health interventions delivered via the Internet (Arnberg et al., 2014), including conversational agents, the advantages of which may allow overcoming at least some of the limitations associated with traditional mental health services. In fact, we can identify two types of advantages of e-Mental Health interventions: structural advantages, related to the practical limitations of traditional mental health services, and non-structural advantages, related to psychological-emotional limitations that restrain an adolescent from seeking help from a mental health professional.

2.1. Structural advantages

With regard to the first category, the primary advantage is certainly anonymity, as the user can make use of these platforms by registering anonymously. They are also available to whoever has a smartphone and an Internet connection and can be accessed even with a poor Internet connection, as they are meant to use a small amount of mobile data. Due to this extended accessibility, conversational agents can also be used by those who live in places where there is a shortage of mental health professionals or where the waiting list system does not allow for timely interventions (Demyttenaere et al., 2004; Anderson et al., 2017) and are accessible and available at any time, 24/7, whenever the user feels the need, leading to a reduction in the time between the onset of a symptom and a person's request for help. They are furthermore cheaper than traditional psychological therapy (McCrone et al., 2004), so they

can also provide support to those who, for financial reasons, would not have the possibility to refer to a professional.

2.2. Non-structural advantages

With regard to non-structural advantages, it should be kept in mind that many adolescents find it difficult to express their thoughts and feelings in front of an adult (e.g. for fear that their confidences will be shared with third parties; West et al., 1991) and feel more confident in handling complex and delicate situations in online interactions by means of text messages (Livingstone, 2006). Due to their ease of access and anonymity, chatbots remove this barrier and allow these users to feel comfortable in externalising their inner world, enabling them to open up in a more sincere and authentic way. Another problem that often prevents adolescents from accessing mental health support services is the stigma that still exists against those suffering from mental distress (Gulliver et al., 2010), which makes young people reluctant to ask for help, for fear of being judged. Chatbots make it possible to overcome this 'obstacle', as they can be used in a completely private manner via one's smartphone. Furthermore, as Christensen, Reynolds and Griffiths (2011) point out, digital platforms for mental health support could be a stepping stone towards seeking help from a professional. Indeed, through the use of this resource, users can learn to identify and recognise their own psychological discomfort and understand when they need more help and could thus become more receptive and willing to engage in conversation with a real person. Finally, for those who are already undergoing a therapeutic process, this technology could be an additional or intermediate assistance between sessions (Baumel & Schueller, 2016).

3. Limits and critical aspects

Although the user can benefit from several advantages related to the use of digital platforms, there are some limitations and risks that cannot be ignored and should be made known to the user before engaging with these conversational agents. We can divide these critical issues into four main categories: privacy, effectiveness, safety and structural limitations (Kretzschmar et al., 2019).

3.1 Privacy

Although the user can register anonymously, it is not clear who has access to personal information and conversations. For example, the chatbot Wysa is only offered as an app, giving the user the opportunity to chat in an anonymous way, but other chatbots (such as Joy and Woebot) are also accessible via Facebook Messenger, where conversations are associated with the user's profile and name. (Kretzschmar et al., 2019). Furthermore, data gathered via Facebook Messenger is subject to Face-

book's privacy policy and may be disclosed to third-party partners. (see sitography to access Facebook's privacy policy).

3.2 Effectiveness

The effectiveness of therapeutic chatbots has yet to be determined, as they are still a developing technology, and moreover, many studies have focused on the development of the underlying models and algorithms and have mainly investigated their quality in terms of ease of use, intuitiveness, level of engagement and perceived naturalness of the conversation. There are still few studies that have investigated their usefulness in decreasing symptoms of mental distress and most have been carried out using healthy populations, among other things with mixed results. Therefore, studies investigating their effectiveness in different populations, including clinical populations, are needed (Grist, Porter & Stallard, 2017). It is also necessary for the results obtained to be compared with those of classical psychological therapy in order to determine its real effectiveness and for the support provided to be based on empirically supported clinical approaches.

3.3 Safety

Most of the critical issues related to the use of chatbots to support well-being and mental health can be found in the area of safety. Although the effects of social media on mental health need further investigation (Baker & Algorta, 2016), it is possible that the use of online platforms (also and especially since the onset of the COVID-19 pandemic) can lead to additional isolation in people in distress, thus representing a backward step and not an aid in the management of their mental discomfort. Furthermore, chatbots could themselves be the cause of mental uneasiness, e.g. they could represent a dangerous trigger for new types of techno-addictions (De-Sola Gutiérrez, de Fonseca & Rubio, 2016; Ellis, 2019), being available 24/7, which is characterised by a progressive and substantial increase in the time of use, a decrease in interest in offline activities (hobbies and friends), a worsening of mood (aggression, fatigue, agitation, mood swings) when use is interrupted due to external causes, and a worsening of academic performance (DSM-V; American Psychiatric Association, 2013).

Existing chatbots are currently unable to deal with emergency situations, although they are able to recognise them. It is therefore necessary that in these circumstances, the chatbot delivers effective and location-specific resources to the user and advises the user to contact a professional or a trusted adult/coetaneous. Furthermore, a further precautionary procedure that could be adopted is to ask the user, during registration, to provide the telephone contact of a trusted person to be used only in case of emergency and, if this information was provided, in case of need the user would be invited to choose whether or not to contact the person indicated

EDIZIONI UNIVERSITARIE ROMANE
editoria scientifica dal 1952

(Kretzschmar et al., 2019). A further risk to which users might be exposed relates to the responses provided by these digital platforms. Based on AI, chatbots are fed with a large volume of online data, so there is a risk that the course of the talk goes outside the boundaries set by the developers (Eunkyung et al., 2023). In particular, adolescents may be exposed to incorrect and improper content and to non-filtered, unchecked and possibly offensive replies, contributing to misinformation and worsening psychological distress (Murphy Kelly, 2023a,b). However, poor efforts have been made in the field of teenagers' safety during conversations with chatbots, leaving both them and their parents without any guidance on how to conduct healthy conversations., as there are still no rules in this regard (Park, Singh & Wisniewski, 2023). Furthermore, algorithms can be influenced by biases and discriminations in the data that fuelled their learning or in those who developed them, leading to unequal treatment of some groups of individuals and consequently amplifying existing inequalities. Therefore, rules should be drawn up to be imposed on technology companies in order to guarantee user safety and ensure fairness and inclusion in the provision of AI-based mental health services.

Finally, another important issue concerns liability: if a diagnostic or therapeutic algorithm makes a mistake, who is responsible for the quality of the data used or the decisions made by the artificial intelligence? Can this diagnosis be considered reliable? On what criteria is it based? A clear regulation is therefore needed, based on scientific criteria and to be adopted by all developers of conversational agents, in order to minimise the risk of error and ensure the safety of people using digital interventions for mental health support.

3.4 Structural limitations: the lack of the human factor

There is a final limitation that characterises conversational agents and that cannot be underestimated: the absence of those human aspects that are the hallmark of all relationships, including the therapeutic one, and that are incompatible with being artificial, such as empathy and intuition. Psychological therapy has several components, including that which can be termed psycho-educational: providing explanations, naming subjective experiences, teaching problem-solving strategies, etc. Probably this task can be easily answered by AI, but the most complex part of the therapeutic intervention is left unresolved: to provide empathy and attribution of meaning and significance to the subjective experience and to link in a reasoned way the elements of the past with current experiences (interpretation). One of the main critical points concerns the ability of these tools to fully understand the human experience. Despite enormous progress in AI, current chatbots still have limitations, especially when it comes to interpreting all the nuances of human emotions. Indeed, the answers they offer may not always be sensitive to the specific needs of the user and this can lead to a lack of empathy or to responses that do not fully satisfy the user's concerns. The

dynamic interpretation of emotions, of the psychological complexes underlying a disorder, of personality traits and how personality has structured itself are central aspects of therapy, elements that are only possible and find meaning and substance through the exchange between two human minds, between therapist and patient, so it is unlikely that, at least in the near future, AI will be able to play this role within a therapeutic relationship.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). doi: 10.1176/appi.books.9780890425596.

Anderson, J. K., Howarth, E., Vainre, M., Jones, P. B., & Humphrey, A. (2017). A scoping literature review of service-level barriers for access and engagement with mental health services for children and young people. *Children and Youth Services Review, 77*, 164– 176. doi: 10.1016/j.childyouth.2017.04.017.

Arnberg, F. K., Linton, S. J., Hultcrantz, M., Heintz, E., & Jonsson, U. (2014). Internet-delivered psychological treatments for mood and anxiety disorders : a systematic review of their efficacy, safety, and cost-effectiveness. *PLoS one, 9*(5), e98118. doi: 10.1371/journal.pone.0098118.

Baker, D. A., & Algorta, G. P. (2016). The relationship between online social networking and depression: a systematic review of quantitative studies. *Cyberpsychology, behaviour and social networking, 19*(11), 638–648. doi: 10.1089/cyber.2016.0206.

Bakker, D., Kazantzis, N., Rickwood, D., & Rickard, N. (2016). Mental health smartphone apps: Review and evidence-based recommendations for future developments. *JMIR mental heath, 3*(1), e7. doi: 10.2196/mental.4984.

Baumel, A., & Schueller, S. M. (2016). Adjusting an available online peer support platform in a program to supplement the treatment of perinatal depression and anxiety. *JMIR mental health, 3*(1), e11. doi: 10.2196/mental.5335.

Blakemore, S. J. (2019). Adolescence and mental health. *The Lancet, 393*(10185), 2030-2031. doi: 10.1016/S0140-6736(19)31013-X.

Burns, J. M., Davenport, T. A., Durkin, L. A., Luscombe, G. M., & Hickie, I. B. (2010). The Internet as a setting for mental health service utilisation by young people. *The Medical Journal of Australia, 192*(S11), S22–S26. doi: 10.5694/j.1326-5377.2010.tb03688.x.

Christensen, H., Griffiths, K. M., & Farrer, L. (2009). Adherence in Internet interventions for anxiety and depression: systematic review. *Journal of Medical Internet Research, 11*(2), e1194. doi: 10.2196/jmir.1194.

Christensen, H., Reynolds, J., & Griffiths, K. M. (2011). The use of e-health applications for anxiety and depression in young people: Challenges and solutions. *Early Intervention in Psychiatry, 5*, 58–62. doi: 10.1111/j.1751-7893.2010.00242.x.

Demyttenaere, K., Bruffaerts, R., Posada-Villa, J., et al. (2004). Prevalence, severity, and unmet

need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. *Jama, 291*(21), 2581–2590. doi: 10.1001/ jama.291.21.2581.

De-Sola Gutiérrez, J., Rodriguez de Fonseca, F., & Rubio, G. (2016). Cell-phone addiction: A review. *Frontiers in Psychiatry, 7*, 175. doi: 10.3389/fpsyt.2016.00175.

Diener, E. (1984). Subjective well-being. *Psychological Bulletin, 95*(3), 542-575. doi: 10.1037/0033-2909.95.3.542.

Diener, E., Oishi, S., & Lucas, R. E. (2009). Subjective well-being: the science of happiness and life satisfaction. In S. J. Lopez & C. R. Snyder (eds), *The Oxford handbook of positive psychology, 2nd edn (*pp. 187–194) Oxford University Press.

Ding, S., Cong, G., Lin, C. Y., & Zhu, X. (2008). Using conditional random fields to extract contexts and answers of questions from online forums. In *Proceedings of ACL-08: HLT* (pp. 710-718).

Ebert, D. D., Zarski, A. C., Christensen, H., Stikkelbroek, Y., Cuijpers, P., Berking, M., & Riper, H. (2015). Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: a meta-analysis of randomized controlled outcome trials. *PLoS one, 10*(3), e0119895. doi: 10.1371/journal.pone.0119895.

Ellis, D. A. (2019). Are smartphones really that bad? Improving the psychological measurement of technology-related behaviors. *Computers in Human Behavior, 97*, 60-66.

Facebook data policy. https://www.facebook.com/full_data_use_policy.

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), e7785. doi: 10.2196/mental.7785.

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology. The broaden-and build theory of positive emotions. *The American psychologist, 56*(3), 218-226. doi: 10.1037/0003-066X.56.3.218.

González-Carrasco, M., Casas, F., Viñas, F., Malo, S., Gras, M. E., & Bedin, L. (2017). What Leads Subjective Well-Being to Change Throughout Adolescence? An Exploration of Potential Factors. *Child Indicators Research, 10*(1), 33-56. doi: 10.1007/s12187-015-9359-6.

Grist, R., Porter, J., & Stallard, P. (2017). Mental health mobile apps for preadolescents and adolescents: a systematic review. *Journal of medical Internet research, 19*(5), e176. doi: 10.2196/jmir.7332.

Gulliver, A., Griffiths, K. M., Christensen, H., et al. (2010). Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. *BMC psychiatry, 10*, 113. doi: 10.1186/1471-244X-10-113.

Jo, E., Epstein, D. A., Jung, H., & Kim, Y. H. (2023). Understanding the benefits and challenges of deploying conversational AI leveraging large language models for public health intervention. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1-16).

Kretzschmar, K., Tyroll, H., Pavarini, G., Manzini, A., Singh, I., & NeurOx Young People's Advisory Group. (2019). Can your phone be your therapist? Young people's ethical perspectives on the use of fully automated conversational agents (chatbots) in mental health support. *Biomedical informatics insights, 11*, 1178222619829083.

Liao, S. H. (2005). Expert system methodologies and applications – A decade review from 1995 to

2004. *Expert systems with applications*, *28*(1), 93-103.

Lin, J., & Katz, B. (2003). Question answering from the web using knowledge annotation and knowledge mining techniques. In *Proceedings of the twelfth international conference on Information and knowledge management* (pp. 116-123).

Livingstone, S. (2006). Children's privacy online: Experimenting with boundaries within and beyond the family. In: Kraut R. E., Brynin, M., Kiesler, S. (eds). *Computers, Phones, and the Internet: Domesticating Information Technology*. New York, NY: Oxford University Press, 128–145.

Ly, K. H., Ly, A. M., & Andersson, G. (2017). A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. *Internet interventions, 10*, 39–46. doi: 10.1016/j.invent.2017.10.002.

Lyubomirsky, S., King, L., & Diener, E. (2005). The benefits of frequent positive affect: does happiness lead to success? *Psychological Bulletin*, *131*(6), 803. doi: 10.1037/0033-2909.131.6.803.

McCrone, P., Knapp, M., Proudfoot, J., Ryden, C., Cavanagh, K., Shapiro, D. A., ... & Tylee, A. (2004). Cost-effectiveness of computerised cognitive-behavioural therapy for anxiety and depression in primary care: randomised controlled trial. *The British Journal of Psychiatry*, *185*(1), 55-62.

Mierzwa, S., Souidi, S., Conroy, T., Abusyed, M., Watarai, H., & Allen, T. (2019). On the potential, feasibility, and effectiveness of chat bots in public health research going forward. *Online journal of public health informatics, 11*(2), e4. doi: 10.5210/ojphi.v11i2.9998.

Murphy Kelly, S. (2023a). The dark side of Bing's new AI chatbot. https://www.cnn.com/2023/02/16/tech/bing-dark-side/index.html**.**

Murphy Kelly, S. (2023b). Snapchat's new AI chatbot is already raising alarms among teens and parents. https://edition.cnn.com/2023/04/27/tech/snapchat-my-ai-concerns-wellness/index.html**.**

Park, J., Singh, V., & Wisniewski, P. (2023). Supporting Youth Mental and Sexual Health Information Seeking in the Era of Artificial Intelligence (AI) Based Conversational Agents: Current Landscape and Future Directions. *Available at SSRN 4601555*.

Park, N. (2004). The role of subjective well-being in positive youth development. *The annals of the American academy of political and social sciences*, *591*(1), 25-39. doi: 10.1177/0002716203260078.

Patalay, P., & Fitzsimons, E. (2018). Development and predictors of mental ill-health and wellbeing from childhood to adolescence. *Social psychiatry and psychiatric epidemiology*, *53*, 1311-1323.

Rizzo, A., Lucas, G. M., Gratch, J., Stratou, G., Morency, L., Shilling, R., Hartholt, A., & Scherer, S. (2016). Clinical interviewing by a virtual human agent with automatic behavior analysis. In *Proceedings of the International Conference on Disability, Virtual Reality & Associated Technologies (ICDVRAT).*

Ryff, C. D. (2017). Eudaimonic well-being, inequality, and health: Recent findings and future directions. *International review of economics, 64*(2), 159-178. doi: 10.1007/s12232-017-0277-4.

Shaw, L. H., & Gant, L. M. (2002). In defense of the Internet: The relationship between Internet communication and depression, loneliness, self-esteem, and perceived social support. *Cyberpsychology & behaviour, 5*(2), 157-171.

Weisel, K. K., Fuhrmann, L. M., Berking, M., Baumeister, H., Cuijpers, P., & Ebert, D. D. (2019).

Standalone smartphone apps for mental health—a systematic review and meta-analysis. *NPJ digital medicine*, *2*(1), 118. doi: 10.1038/s41746-019-0188-8.

Weizenbaum, J. (1966). ELIZA - a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36-45.

Weizenbaum, J. (1967). Contextual understanding by computers. *Commun. ACM*, *10*, 474–480. doi: 10.1145/363534.363545.

West, J. S., Kayser, L., Overton, P., & Saltmarsh, R. (1991). Student perceptions that inhibit the initiation of counseling. *School Counselor*, *39*(2), 77–83.

Woebot website. https://woebothealth.com/.

Wysa website. https://www.wysa.com/.

X2 website. https://www.cass.ai/x2ai-home.

Youper website. https://www.youper.ai/.