

Chatbot based Cognitive Behavioural Therapy

Moritz Pflügner

TU Dresden

moritz.pfluegner@mailbox.tu-dresden.de

ABSTRACT

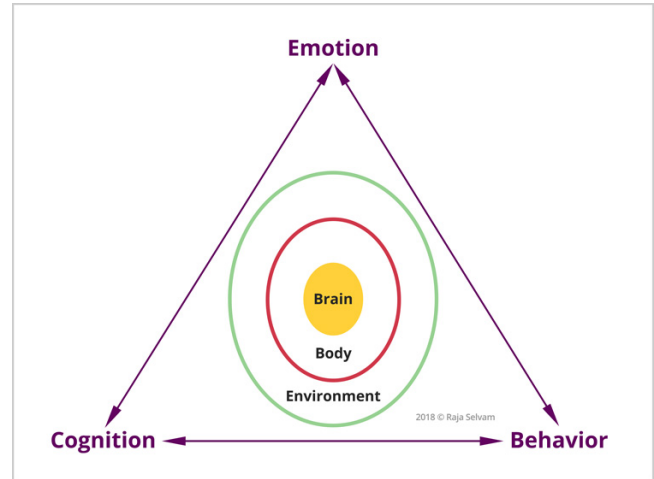
[illegible]

1 INTRODUCTION

During the last two decades the internet has become a major part in the life of people throughout the society. The ubiquity has led to the desire for location-independent usage. As a consequence the amount of smartphone users is increasing every year.¹ With constant access to the biggest collection of knowledge in the world, more and more individuals tend to use it as a first source of information when symptoms of illness arise[1]. In terms of mental health, the internet provides a numerous amount of cognitive behavioural therapy (CBT) offers such as guides or apps.

In contrast to a complex psychoanalysis with the guidance of a therapist or doctor trying to look into the past, CBT focuses more on the way an individual deals with its present thoughts and beliefs. The treatment has the aim to change behavioural patterns in a way that they positively influence emotions and cognition. CBT is based on this connection between the way how we interpret the daily life, how we behave due to it and how we feel in the end which is shown in figure 1. By changing one of the three aspects, the others are changed as well automatically. Some popular methods are keeping a Gratitude Journal or exercises to calm the body.[2]

With the help of apps, users can get hints and instructions to conduct a self therapy. It is possible to get first support without the need to consult a therapist or doctor. Often, this is a big and difficult step for many people due to the fear of getting stigmatized[3]. While CBT applications have the potential to reduce the barrier to seek help, the rate of completion is relatively low. Donkin et. al found a median minimal completion rate of 56% during a randomized controlled trial[4]. One possible reason for this lack of adherence is the missing human interaction of internet-based CBT[5]. The usage of chatbots are considered as a possible solution for this problem. While they are providing CBT on the same level as non-chatbot based approaches, they can deliver an artificial human interaction that is supposed to increase the therapy adherence which can be seen as an important aspect of a successful therapy.



Embodied Cognition: Cognition, emotion and behavior are all embodied and embedded in the brain, body, as well as the environment. They affect each other but emotion is primary, as emotion is a stronger mediator of cognition and behavior.

Figure 1: The relationships among cognition, emotion, behavior. Graphic by Raja Selvam, 2018

(<https://integralsomaticpsychology.com/wp-content/uploads/2018/02/ISP-Embodied-Cognition-Diagram-Raja-Selvam-750.jpg>).

The paper starts with a review of a part of the conducted research and analyses weaknesses and strengths as well as risks and chances. Subsequently, it will define some general requirements a chatbot used for CBT should met. What are aspects that need improvement during further development approaches and what methods and measures might be suitable to achieve that? Considering the state of the art and the general requirements, the paper tries to answer this question regarding safety aspects, efficacy and privacy. It finally discusses the outcomes and points out limits of chatbot based CBT.

2 STRENGTHS AND WEAKNESSES OF EXISTING CHATBOTS

This far, there are a couple of chatbots available. This paper will mainly focus on the Woebot Chatbot and Wysa for several reasons. On the one hand, they have been developed in a research oriented context. The examples are available in english language and free to use.^{2,3}

Additionally, the developers provide numerous information on their website like experience reports or technical aspects. On the other

¹<https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/>, last access: 13.05.2021

²<https://woebothealth.com/about-us/>, last access: 01.05.2021

³<https://www.wysa.io/faq>, last access: 01.05.2021

hand, a lot of studies have already been conducted mainly focusing on the efficiency of the chatbot based CBT. They revealed some weaknesses of those chatbots, some chances and also risks concerning Efficiency, Privacy and Safety.

2.1 Woebot

The Woebot chatbot is a product of the company Woebot Health which has its headquarter in California. It is the result of a research project at the university of Stanford in 2017 and has raised 8.000.000 dollars of funds. Woebot is available as a standalone app in Apple's App Store and in Google Play store where it reached more than 100.000 downloads⁴ and as a facebook chatbot. It is based on natural language processing and offers free-text inputs as well as predefined answers. In general, it gives hints how to change the thinking and guides the user through techniques like gratitude journaling. Based on the answers it creates a graph monitoring mood changes during the week. Woebot aims to reach patients and clinicians. While users that are currently not in therapy have the opportunity to use it, it can also improve and support ongoing therapies clinicians are conducting. Currently, the usage of woebot is completely free of charge. Nevertheless, Woebot Health announces on its website, that they are moving forward to develop a more sustainable business model which could lead to costs for the users.^{5,6}

A randomized controlled trial with college students by Fitzpatrick et. al has shown that the usage of Woebot significantly reduces symptoms of depression compared to a control group that only got informed about depression via an e-book. Regarding anxiety symptoms, no difference between both groups has been found but both showed a reduction of it. Furthermore, the study has revealed that eighty-five percent of persons in the chatbot group used it daily or almost daily which can be considered as high in comparison to conventional CBT applications[6]. After the period of two weeks the users have been asked to assess their experience with Woebot. One the one hand, the majority appreciated the daily check-ins of the bot which motivated them to continue the therapy. On the other hand, many of them noted that the conversations did not feel natural which can be considered as a limit of chatbots and natural language processing in general.[7]

The privacy policy of Woebot gives some useful information about the way the chatbot deals with user inputs. During the registration process the user will be asked for an e-mail adress and a password. It is possible to skip this step. Hence, no account will be created and the user wont be able to use the application across devices. Additionally, neither access to stored data nor backup functionalities are possible. If a user creates an account, Woebot Health stores data concerning account information (personal information), technical aspects (platform, operating system), financial information if services are purchased, hardware information and analytical data. On the other hand, the company also stores the conversation data in order to improve the services. Woebot Health also obtains

data from other sources. If the service is used through third-party applications, it collects data that has been made public by the user. Examples for third-party applications are social networks or app stores. The company does not share information with third parties, but it doesn't have direct control of data that is stored if the chatbot is used through them like Facebook's messenger. The access and deletion of it then depends on Facebook's privacy policy.⁷

As chatbots for CBT interact with people who might have serious mental problems, safety and the recognition of emergency situations are important aspects. If a user starts a conversation with Woebot, the bot early advises that it is not a human being. It notes that it has limited capabilities and that it might be better for some individuals to see a real therapist. In order to discourage over-reliance, Woebot tries to end conversations after a certain amount of time and proposes many non-digital activities in the real world. Those proposals of non-technology-based activities are part of evidence-based recommendations for mental health app development[8]. Considering emergency situations, Woebot is able to recognize user inputs like 'SOS' or 'suicide' as such. It reacts with a list of helplines which can provide further support.[9]

2.2 Wysa

The Wysa Chatbot, developed by Touchkin, has certain commonalities but also major differences in comparison to Woebot. It is also a mobile chatbot app using a text-based interface with the possibility of entering free-text input or predefined answers. The biggest difference is the option of having contact to a human therapist while using Wysa. It is a chargeable feature which costs 29.99\$ per month. The usage of the chatbot without it is free of charge. It is available in Apple's app store and Google Play store where it has been downloaded more than 1.000.000 times⁸. The company is located in Bangalore and has been found in collaboration with a research team from the universities of Cambridge and Columbia. While also offering CBT like Woebot, Wysa also offers guided meditation, breathing and yoga. Additionally, it conducts Dialectical Behaviour Therapy (DBT).⁶

DBT is a special type of CBT and focuses more on the way a patient deals with emotions and is suitable for patients with self-destructive behaviours. Common techniques are practicing mindfulness or improving relationships to other individuals.⁹

Inkster et. al have shown that patients with a more frequent use of Wysa have a significantly higher improvement of their mood than users with a less frequent usage. In general, both groups showed an improvement. 74% of the study participants provided an in-app feedback. Most of them preferred the pre-defined answers rather than free-text. While the app was generally considered as helpful and encouraging, the participants proposed improvements included wanting to avoid repetitions or a better understanding of user inputs.[10]

The way Wysa stores user data deviates from Woebot's. Though Wysa does not require a registration, it links all the collected data

⁴<https://play.google.com/store/apps/details?id=com.woebot&hl=de&gl=US>, last access: 03.05.2021

⁵<https://woebothealth.com/about-us/>, last access: 01.05.2021

⁶<https://emerj.com/ai-application-comparisons/chatbots-mental-health-therapy-comparing-5-current-apps-use-cases/>, last access: 03.05.2021

⁷<https://woebothealth.com/privacy-webview/>, last access: 04.05.2021

⁸<https://play.google.com/store/apps/details?id=bot.touchkin>, last access: 05.05.2021

⁹<https://www.skylandtrail.org/4-differences-between-cbt-and-dbt-and-how-to-tell-which-is-right-for-you/>, last access: 06.05.2021

to a vendor-specific user id provided by Google Play or App store. This approach leads to a pseudonymised data storage. If a user accidentally submits personal information during a conversation, it will be redacted by algorithms scanning the stored data regularly. In addition to conversational data, Wysa stores anonymized technical data (platform, operating system), financial data if an individual subscribes to paid offers and analytical data in order to improve their services. It is important to mention, that the chatbot can not be used through third party applications like Facebook's messenger. Hence a user can use the application more independently of privacy policies of other companies. Although it is not completely excluded as Wysa allows to speak with the chatbot via Apple's Siri or Google's voice assistant.¹⁰

The chatbot understands words like 'privacy' and can provide links and information about it during a conversation as well.[9]

Considering privacy aspects, Wysa is quite common to Woebot. It also informs the user during the conversation, that it is only a robot. If an individual is willing to pay the fee, the contact to a human therapist is a safe way when it has mental issues that are more serious. But as this service is not free of charge, it can not be seen as a general safety aspect. The Wysa chatbot tries to deal with emergency situations the same way as Woebot does. It notices user inputs like 'SOS' and provides helplines for further support.[9]

3 GENERAL REQUIREMENTS

This chapter will define requirements a chatbot should meet regarding privacy, efficiency and safety. Those three aspects have already been examined by Kretschmar et.al, but only from the perspective of young adults as a result of a group discussion[9]. Due to the fact that the usage of a mental health chatbot belongs to Human Computer Interaction (HCI), the following requirements will be based on its principles and research conducted in the field.

3.1 Privacy

Alongside the development of the internet of things, technologies and the amount of data are continuously increasing[11]. As a consequence of it, governments around the globe define rules in order to protect the individual from exploitation and to ensure that users have access and control over personal data stored by service providers. An example for it is the General Data Protection Regulation by the European Union. Considering chatbots for mental health, privacy is an important aspect. Sharing sensitive and personal information is part of the concept of CBT and hence necessary for a successful chatbot usage. Therefore the user should constantly feel secure its data is being stored properly.

Firstly, it is necessary to define what privacy principles are concerning HCI. In this paper, the results of Patrick and Kenny are used, who specifically derived privacy aspects by using the general EU legislation as basis. They assessed, which principle of the EU legislation might have an implication to HCI and how methods and techniques of HCI could comply it. The result is a grouping of the principles into four categories: *Comprehension* means that an individual understands what is happening to its data. *Consciousness* aims to ensure awareness of the individual and that it is informed

about data processing constantly. *Control* contains principles to have access to stored data and to be empowered to delete it and requirements of the *consent* category make sure that the user always has to give its approval about processing of personal data.[12]

As chatbots are only a subset of HCI and the categories and principles above are held general, it is important to concretise them regarding mental health chatbots. Saglam et. al conducted an empirical study where on the one hand, the participants have been asked to answer questions about their trust towards a chatbot and on the other hand, what aspects would lead to worries. Regarding the first question, they found that users are most worried about how to delete personal information. Another aspect was the concern if the data would be used in an inappropriate way. For the second question, the study has revealed that users prefer chatbots with good technical quality like grammatically correct answers rather than those with social aspects like seeing an avatar while chatting.[13] Chatbots with a better response quality can be seen as more human-like. Ischen et. al revealed in another study, that human-like chatbots lead to less privacy concerns[14]. The conclusion of both results is that users prefer human-like chatbots while having less privacy concerns at the same time during the conversations. Companies could easily exploit this fact. In order not to be accused of it, they should take privacy very serious especially when they are programming chatbots on a professional level.

Another aspect that has not been considered yet is the distinction between data and information. An interview with employees about their attitude towards a workplace robot conducted by Lee et. al showed that most participants did not make a distinction between the knowledge of the robot and the data it collects.[15]

The comparison between a physical robot and a chatbot is limited, but the distinguishing between data and information can be applied to chatbots as well. It is necessary for chatbot users to know, what inferences it can make based on the conversational text inputs they are providing. The underlying artificial intelligence a chatbot has should be transparently explained to the user in order to raise awareness for problems that can come with intelligent algorithms like false positive classifications.

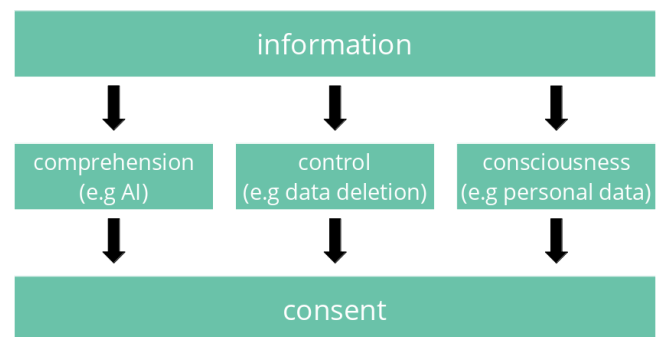


Figure 2: The connection between information, comprehension, control, consciousness and consent, Graphic by Moritz Pflügner, 2021

In order to summarize the privacy requirements of a chatbot for mental health it can be said, that information is the most important

¹⁰<https://legal.wysa.io/privacy-policy>, last access: 08.05.2021

one. Firstly, it leads to a conscious usage of the application which is essential to protect personal data. It secondly ensures having control over the data and being able to delete it. Finally, information helps to understand the system and its algorithms better. Only with all of these aspects given, the individual is able to give its consent on a proper basis.

3.2 Efficacy

Mental health applications always aim to increase the general subjective well-being of a person. Therefore, it can be seen as a measure of an application's efficacy.

Ludden et.al proposed two different routes that can lead to a better well-being that are shown in figure 3. They introduced the PERMA abbreviation which stands for positive emotions, engagement, positive relationships, meaning and accomplishment. All of these aspects can lead to a higher adherence and hence to a better well-being.[6]

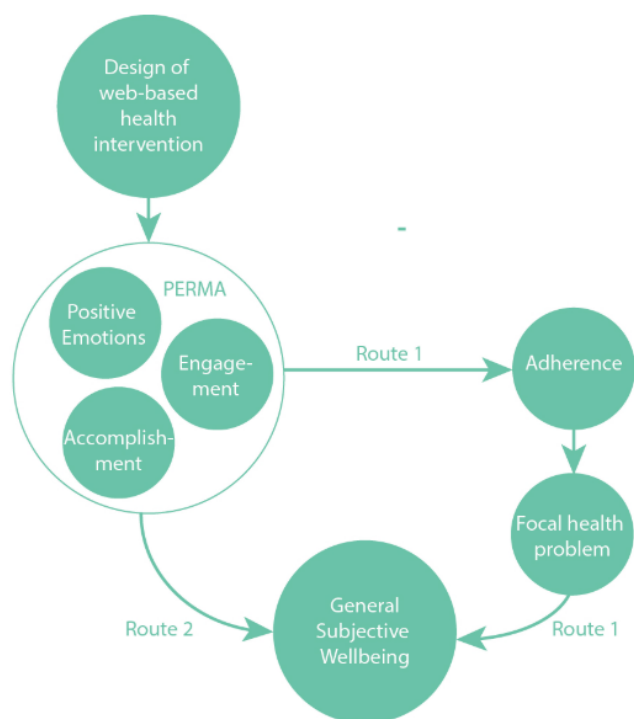


Figure 3: Schematic representation of how the design of a Web-based intervention can influence general subjective well-being following two different routes. Route 1 indicates the impact of design on adherence and thus on the focal health problem. Route 2 indicates how overall well-being is stimulated by elements of PERMA. Graphic by Ludden et.al, 2015

While the efficacy studies of both chatbots reviewed in chapter two showed a reduction of mental illness symptoms, the adherence of e-therapy applications in general is relatively low. Donkin et. al found a minimal completion rate of 56% concerning

online-therapy applications. Furthermore they revealed that the adherence correlates with the treatment outcomes.[16]

This can be considered as a further prove for Route 1 in figure 3. A couple of general design approaches promising an increase of therapy adherence are now being discussed.

Ludden et. al present personalization as one approach. Considering the PERMA abbreviation, personalization encourage the engagement of a user as the application has a higher personal relevance. If it is used by multiple groups of people, it should be adaptable to them in order to improve the experience. Secondly, they discuss an approach that prevents too much ambient information. They point out, that due to the daily information overload in the internet, users might not feel the desire to use the application online. Additionally it is possible, that individuals have made bad experiences with too much information that concerns the application but is irrelevant for the acutal usage. Ludden et. al propose that users should not be overloaded with unnecessary information. According to them, this would lead to a more pleasureable experience (P in PERMA) and a better feeling of accomplishment (A in PERMA). Finally, they are recommending the use of metaphors for applications concerning mental health. Challenges, goals and experiences should be shaped into words and visualizations in a way to create a meaningful (M in Perma) and engaging (E in Perma) frame. They note that the way how human beings think about abstract things like challenges or achivements is methaphorical by nature and experienced every day. Metaphors supporting mental health applications can strengthen those experiences.[6]

An application which is personalized as well as focusing only on necessary information and using a sophisticated way to frame challenges and goals resembles a therapy with a human therapist. According to Ischen et.al, services with a higher perception in anthropomorphism lead to a better adherence [14]. This result corresponds to the results of Ludden et.al and hence supports the recommendation to implement the PERMA principles in applications for mental health like chatbots.

3.3 Safety

The safety aspects of mental health chatbots are very specific due to the target group and the conversational aspect. Hence, only aspects that are directly related to them will be discussed. As they reach persons with symptoms of mental illnesses by nature, safety aspects should not be neglected.

Kretzschmar et. al defined four requirements that are essential. Firstly, it is considered as crucial that users get informed about the fact that they are talking to a robot. This makes sure that they are aware of limitations and that the conversation can not replace a real life therapy. Secondly, they argue that the system should prevent over reliance so that users always keep in mind that the application is only a tool. Due to the limitation of a chatbot based conversation, they thirdly recommend that chatbots should additionally encourage seeking human support. Kretzschmar et. al finally value the recognition of emergency situations.[9]

This last aspect has also been recommended by C.Grové who co-developed a chatbot with and for young people. She points out that the identification of critical words is necessary for a chatbot

that deals with mental health. Furthermore she recommends the generation of appropriate answers to those critical words in order to prevent rash reactions. As a second aspect she notes that the applications should have a sufficient listening phase prior to quick responses that might not be suitable to the situation.[17]

4 METHODS FOR IMPROVEMENTS

5 CONCLUSION

6 DISCUSSION

REFERENCES

- [1] Jeremy C. Wyatt. Fifty million people use computerised self triage. *BMJ (Online)*, 351, jul 2015.
- [2] Diane Spangler. How Does Cognitive Behavioral Therapy Work? In *Cognitive Psychotherapy Toward a New Millennium*, number February 2016, pages 161–164. 2002.
- [3] Wulf Rössler. The stigma of mental disorders. *EMBO reports*, 17(9):1250–1253, sep 2016.
- [4] Liesje Donkin, Ian B Hickie, Helen Christensen, Sharon L Naismith, Bruce Neal, Nicole L Cockayne, and Nick Glozier. Rethinking the dose-response relationship between usage and outcome in an online intervention for depression: Randomized controlled trial. *Journal of Medical Internet Research*, 15(10), 2013.
- [5] Kien Hoa Ly, Ann Marie Ly, and Gerhard Andersson. A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. *Internet Interventions*, 10:39–46, dec 2017.
- [6] Geke Ds Ludden, Thomas J.L. Van Rompay, Saskia M. Kelders, and Julia Ewc Van Gemert-Pijnen. How to increase reach and adherence of web-based interventions: A design research viewpoint. *Journal of Medical Internet Research*, 17(7):172, 2015.
- [7] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Mental Health*, 4(2):e19, 2017.
- [8] David Bakker, Nikolaos Kazantzis, Debra Rickwood, and Nikki Rickard. Mental Health Smartphone Apps: Review and Evidence-Based Recommendations for Future Developments. *JMIR Mental Health*, 3(1):e7, mar 2016.
- [9] Kira Kretzschmar, Holly Tyroll, Gabriela Pavarini, Arianna Manzini, and Ilina Singh. 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*, 11:117822261982908, 2019.
- [10] Becky Inkster, Shubhankar Sarda, and Vinod Subramanian. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: Real-world data evaluation mixed-methods study. *JMIR mHealth and uHealth*, 6(11), 2018.
- [11] David Reinsel, John Gantz, and John Rydning. Data Age 2025: The Digitization of the World From Edge to Core. *Seagate, IDC*, (November):28, 2018.
- [12] Andrew S. Patrick and Steve Kenny. From Privacy Legislation to Interface Design: Implementing Information Privacy in Human-Computer Interactions. *LNCS 2760 - Privacy Enhancing Technologies*, pages 107 – 124, 2003.
- [13] S Shamim, S Cang, H Yu The International Journal of Human, and Undefined 2019. Privacy Concerns in Chatbot Interactions: When to Trust and When to Worry. *23rd International Conference on Human-Computer Interaction*, 30(16):2387–2417, 2021.
- [14] Carolin Ischen, Theo Araujo, Hilde Voorveld, Guda van Noort, and Edith Smit. Privacy Concerns in Chatbot Interactions. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 11970 LNCS, pages 34–48, 2020.
- [15] Min Kyung Lee, Karen P. Tang, Jodi Forlizzi, and Sara Kiesler. Understanding users’ perception of privacy in human-robot interaction. *HRI 2011 - Proceedings of the 6th ACM/IEEE International Conference on Human-Robot Interaction*, pages 181–182, 2011.
- [16] Liesje Donkin, Helen Christensen, Sharon L Naismith, Bruce Neal, Ian B Hickie, and Nick Glozier. A systematic review of the impact of adherence on the effectiveness of e-therapies, 2011.
- [17] Christine Grové. Co-developing a Mental Health and Wellbeing Chatbot With and for Young People. *Frontiers in Psychiatry*, 11(February):1–12, 2021.