Engineering and Estimating Efficiency of the Artificial Intelligence Chatbot in Reducing Cognitive Distortions among AUCA Students

Undergraduate Psychology Senior Thesis Draft

by

Deniz Nazarova

Department of Psychology

American University of Central Asia

Abstract

There is a growing tendency in applying technology when treating mental health issues. Using CBT-based chatbots tends to improve one's well-being but is often seen as a substitution for therapy regardless of existing threats. Experiments using this technology are mostly conducted in Western countries within a short period of time. The goal of this research is to develop an AI chatbot to work with cognitive distortions and measure its performance by conducting mixed research consisting of a field experiment and a series of interviews with practicing psychologists. Sixty-eight AUCA students ages 18-28 years were recruited through convenient sampling and randomly divided into two groups to either interact with TeaBot, an AI bot developed for this project, for 8 weeks (*N* = *34*) or receive a manual about cognitive distortions (*N* = 34). Participants filled out two pretests and one posttest consisting of the Acceptance and Action Questionnaire-II (AAQ-II) and Cognitive Distortion Scale (CDS). The qualitative part included expert interviews with 7 counselors who reported their opinion on TeaBot and their eagerness of implementing it in their practice. The experimental group's members demonstrated a significant decrease in distorted thoughts (p < .001). There is a moderate positive correlation between the number of weeks user interacted with the bot and the reduction of distortions (*r* = .36, *p* = 0.035). Counselors provided positive feedback when working with the bot and expressed interest in using it with clients.

Keywords: artificial intelligence, chatbot, GPT-3, mental health, cognitive distortion, students, youth

Content

Acknowledgements

This project was inspired by Sweaty Machines' song "Blue Jeans and Bloody Tears", a Eurovision song created by AI. I would like to thank the creators for raising the thought about whether we can make an AI depressed based on the data it was fed.

No gratitude would be enough to thank all faculty and supervisors that participated in reviewing this big and long project. I would like to thank Elena Molchanova for supervising it during the second year of development and Elena Kosterina for supervising it during the third year. I am also very thankful to Gulnara Kurmanova for the continuous evaluation of the research design and statistical analysis. I would like to recognize the invaluable contributions of Google, Stack Overflow, and programmers who analyzed the code and suggested optimization methods.

This project was generously supported by the Center for Civic Engagement at the American University of Central Asia.

Thanks to my therapist for supporting my mental health while I was making something to improve the mental state of others.

With love and gratitude to my partner.

Engineering and Estimating Efficiency of the Artificial Intelligence Chatbot in Reducing Cognitive Distortions among AUCA Students

The last few years have been very rich on the events that affect mental health negatively. Worldwide, there is a growing tendency for mental health to worsen after COVID-19 (World Health Organization [WHO], 2022). One of the most mentally affected groups is admitted to be young people, especially students as the pandemic impacted many areas of life inducing financial issues, trouble with focus, and feeling alienated due to the lockdown (Active Minds, 2020; WHO, 2022). In Kyrgyzstan, COVID-19 affected the economy by creating hardships for labor migrants whose remittance flows remain one of the strongest investors in the country (Murzakulova et al., 2021). In addition, the Russian-Ukrainian war influenced the economy of the Central Asian region and the labor market for young adults (Engvall, 2022; Jamankulova, 2022). These financial struggles become even more alarming as Shields-Zeeman et al. (2021) inform on the causality between one's income and psychological distress. Kretzschmar et al. (2019) found that 75% of mental health diseases tend to form around the period of young adulthood, and mitigating them might decrease the long-term economic and personal consequences. Meanwhile, in Kyrgyzstan, the global pandemic impacted the youths' mental health with the rising number of suicides (UNICEF, 2020). This sets an urgent need to enhance the current mental health facilities available for students. However, the current state of healthcare in Kyrgyzstan does not inspire hope as even before the pandemic mental health field did not meet 87% of yearly plans (Orlova, 2019). Pinchuk et al. (2021) also emphasize the neglected status of mental healthcare in Kyrgyzstan by stating such problems as "insufficient financing, lack of equipment, underregulation, a deficiency of qualified specialists" (p. 4). Overall, there is a clear lack of resources to work with the consequences of the pandemic, recession, and war conflicts on the population's mental well-being. As stress is considered to be one of the main external factors leading to the development of mental health issues, decreasing its level at the onset age might be especially helpful in the prevention of disorders (Kretzschmar et al. 2019).

The growing demand for counseling services revealed difficulty accessing them. As in-person sessions were not an option during the pandemic, this created new waves of interest in teletherapy (Zeavin & Peters, 2021). The lack of human resources led to the growth in the popularity of mental health chatbots which are affordable and available for any person with a device and internet connection (Kretzschmar et al., 2019; Dekker et al., 2020). In addition, Abd-alrazaq et al. (2019) report that implementing AI-chatbot might be even more effective in developing countries due to their cheapness compared with standard therapy. However, the phenomenon of AI-supported mental health remains unexplored in Central Asia, thus, the main purpose of this research is to examine the process of developing an AI chatbot in Kyrgyzstan that will be based on Cognitive Behavioral Therapy and how successful this program might be to AUCA students in the improvement of their well-being.

Literature Review

Cognitive Behavioral Therapy: What, How, and Why

Cognitive behavioral therapy, or CBT, is a therapeutic approach that works with one's thoughts, emotions, and behaviors (Beck, 1995). One of the most fundamental elements of this therapy is focusing on cognitive distortions, or dysfunctional patterns in thinking that have no valid evidence and recur automatically, as these are observed to be common among a variety of mental issues and suspected to be one of the main contributors to their development (Beck 1963). Indeed, studies have established a connection between cognitive distortions and mental health issues (Beck, 1995). Burns and Beck (1999) emphasize the importance of training clients to both recognize and restructure these as the main outcome of therapy. Besides being very goal-oriented, CBT is also well-known for its focus on structured sessions both on macro and micro levels. Topics for sessions tend to be heavily planned with the number of sessions being determined within the range of 8 weeks (Beck, 1995; Cully& Teten, 2008). The agenda for the session includes mood checks, a review of homework, psychoeducation, and building a skillset that one can practice outside of meetings with a therapist (Beck, 1995).

As with any therapeutic approach, CBT has both advantages and disadvantages. One of the most highly praised strengths remains its empirically based validity and lack of relapses. Regarding the first, cognitive behavioral therapy was studied extensively and the results in general demonstrate significant improvement in mental well-being for CBT clients (Beck, 1995). This progress is often achieved in a short period due to the focus of the approach on the present issues instead of going deeper into past traumas (Beech, 2000; McGinn & Sanderson, 2001). The latter might be explained by a focus on education, the development of the skills that one can apply every day to challenge the negative beliefs independently, and active engagement of the client by implementing activities like homework, diary, and reflections. Overall, the techniques practiced during CBT sessions aim to train a client independency in solving their psychological distress successfully in a short time which are especially helpful considering the current shortage of mental health staff that require quick and productive methods.

Nevertheless, CBT is often critiqued for its surface-level intervention and context insensitivity. As this approach is present-oriented, cognitive behavioral therapy might fail at dealing with underlying causes and tend to treat symptoms instead (Beech, 2000; Rasmussen, 2018). Such methods might deprive the client of an actual solution to their problem and can only extend the period until the next episode emerge. Considering CBT still teaches a patient how to work with their thoughts, relapse may never occur or be battled successfully by the patient (Beech, 2000). Another issue is context insensitivity which stems from a belief in an individual's locus of control. For example, CBT is often judged for its emphasis on the perception of painful events in one's life which might put too much responsibility on a client of marginalized origin, e. g., LGBTQ-community. While this review used to be true for some practices, nowadays, queer-affirming CBT sessions are proven to be efficient in helping the oppressed (Pachankis et al., 2022). Another case of issues with diverse outlooks in CBT was related to its strong connection to Western values which can lead to therapeutic sessions being non-valid for other cultures. However, Naeem et al. (2015) highlight the necessity of awareness and adaptability when implementing foreign mental techniques and systematization after thorough research. Thus, cognitive behavioral therapy demonstrates certain drawbacks that require additional monitoring and review, yet these do not cancel the validity of essential methods on which CBT thrives.

Overall, CBT remains one of the most effective interventions to work with mental health disorders within a narrow time window. Its features such as being structure-oriented, anticipating high engagement from a client, accenting skill-building, and approach based on pattern identification might be the main reasons why cognitive behavioral therapy is so heavily implemented in AI applications. The Procedures segment in the Methods chapter focuses on these arguments in more detail.

Application of AI in Psychotherapy through Time and Space

Alan Turing (1950) sparked the research on how far machines can go in emulating human behavior by asking the famous question, "Can machines think?" (p. 433). His hypothesis was that at some point, technology might persuade a human that it is another human communicating to them, i.e., winning the imitation game (Turing, 1950). Later, based on this idea, the term

Artificial Intelligence was coined, to imply it was a machine simulating human behavior (McCarthy et al., 1955).

While the talent for computation was not a big surprise, the success of computers in communications was groundbreaking when the research on ELIZA was published. ELIZA is the first chatbot that applied therapeutic techniques when communicating with people (Weizenbaum, 1966). Ironically, the main developer did not expect ELIZA to succeed in therapy, however, participants often commented on the efficiency of the chatbot (Zeavin & Peters, 2021). Weizenbaum (1966) programmed ELIZA to be based on the Rogerian approach which allowed her to avoid the need for explicit knowledge about the outside world. Although being heavily scripted made its effectiveness quite sensitive to the input, this still made some of the participants believe that she was a human (Weizenbaum, 1966). ELIZA became a foundation for the later AI chatbots with some of her successors being constructed using more sophisticated design.

One of the modern tendencies became implementing Machine Learning in chatbot's architecture. Machine Learning, or ML, is defined as AI models that are capable of acting intelligently based on the training data their algorithm processed (Samuel, 2000). In the context of communication, the application of ML algorithms strongly relies on the dataset in order to build a more natural communication that requires Big Data, or extremely large yet unstructured information that needs interpretation (Oracle, n.d.). Using ML leads to the more sentient technology that is able to reflect on the previous sentences and build own sentences that are not based on heavy coding while Big Data gives the machine possibility for diversification (Abd-alrazaq et al., 2019). The models developed by this tandem are very precise in recognizing the patterns in thinking which made them the best fit for Cognitive Behavioral Therapy, or CBT. One of the most successful AI applications that combined the aforementioned design might be

Woebot, a chatbot developed in the United States to treat symptoms of depression and anxiety. It is capable of learning from the data given by the user while applying different therapeutic methods starting from mindful exercises to therapeutic journaling (Fitzpatrick et al., 2017). Similar in popularity products that also passed clinical tests successfully include bots like Tess and Wysa, also developed in the US (Fulmer et al., 2018; Inkster et al., 2018). At the moment of writing, Koko bot, one of the most recently developed mental health chatbots, openly declared using a novel ChatGPT AI model to help people deal with their unrest, however, the lack of empirical and ethical research to validate the technology led to the huge critique of the application (Xiang, 2023).

Although chatbots remain mostly a Western feature of mental health, there are also a few products developed in other parts of the world. There was significant growth in AI-based projects for mental health in China after the COVID-19 outbreak (Qiu et al., 2020). One of the most prominent examples might be Elomia, a mental health AI bot that operates in the English language, that was developed in Ukraine and clinically tested (Romanovskyi et al., 2021). Because of the current war in Ukraine, it is not clear how the development process of the bot will go further. However, the market in other post-soviet countries stays lacking appropriate research on efficiency. Russian developers recently announced the launch of SabinaAI without publishing any data openly on its testing process except mentioning that it was developed in cooperation with psychologists (AvatarMachine, LLC, n.d.). Kyrgyzstan is currently at the beginning stage of digital mental health with products like a game about bride kidnapping and scripted telegram bots informing on sexual education and well-being (*Oilobot*, n.d.; Open Line, 2020). Although these products usually cause a lot of interest from the media news, there is also no clinical research conducted on the efficiency of these innovations.

Advantages of AI Compared to Human Therapy

The current research agrees that there are several advantages of technology compared to human specialists. First of all, mental health applications remain a more accessible source of support as it requires only a device with an Internet connection (Kretzschmar et al., 2019). Moreover, unlike human therapists, chatbots are available 24/7 which makes them especially helpful in situations when a psychotherapist is not available to answer, e.g., during other sessions or nighttime. This might be especially helpful when there is a lack of human psychotherapists available. For example, Abd-alrazaq et al. (2019) stress that in developing countries chatbots can be an additional help as there is a severe disbalance in the specialist-to-patient ratio. Due to the high demand for mental health services, the unavailability of specialists and putting people on long waiting lists might be alleviated by delivering an e-support via chatbots (Kretzschmar et al., 2019).

The scalability advantage of technology might be especially helpful considering the aforementioned lack of resources. Torous et al. (2020) inform that unless a patient is against the application of telehealth, “hybrid solutions that offer a blend of face-to-face and online or app-based treatment will be the most effective solution” (p. 2). Moreover, this flexibility of the format where the participation of technology is easily regulated from solely human interactions to the integration of different applications can significantly enrich the repertoire of therapeutic services (Torous et al., 2020). For example, the implementation of bots as data gatherers can assist practitioners in better preparation for the first therapy session (D’Alfonso et al., 2017; Damij & Bhattacharya, 2022). Luxton (2016), on the other hand, suggests that AI technology can be used to both reduce the existing pressure on the healthcare system and reduce costs by implementing “the stepped-care approach” that distributes mental health services according to the patient’s needs, e.g., “care seekers can take self-assessments with a virtual care provider and be transferred, if necessary, to full therapy with a human care provider” (p. 17). These cases demonstrate that technology has a great potential to enhance the quality of therapy in a variety of ways depending on the level of involvement a specialist would like to apply. However, a study by Torous et al. (2020) claims there is a vital need to educate professionals on the opportunities this tandem provides.

The next feature of using chatbots in mental healthcare is their affordability. Being a low-cost or even free mental health service, technology often removes the existing barriers that people experience when reaching out for help (Fitzpatrick et al., 2017; Gabrielli et al., 2021; Watts et al., 2013). Luxton (2016) highlights that producing a mental health specialist is more-time consuming compared to software that can be replicated quickly. This implies that chatbots are also less costly in terms of production and deployment compared to human specialists. In the study by Fulmer et al. (2018), the cost-efficiency of computer-assisted therapy is especially emphasized considering that this type of intervention is less-intensive than an in-person one. Taking into account the financial struggles that students might experience in their new stage of life, it is also crucial to provide them with evidence-based sources of self-care. Grové (2021) stresses that besides clinical intervention, chatbots might be used as a tool to deliver free and high-quality advocacy for youth about their mental health. Indeed, when implementing technology into this field, technology might be especially beneficial due to low-cost and less time-intensive which might be especially crucial for young people who struggle with both.

Chatbots as well as other non-person mental health services can also be gateway support for people who struggle to start therapy because of the existing stigma (Abd-alrazaq et al., 2019; Fitzpatrick et al., 2017). According to Boucher et al. (2021), this very stigma might be one of the leading causes for young adults avoiding therapy in the first place. Due to the anonymity that chatbots provide, patients might feel more secure and more eager to share information that they might hide in traditional therapeutic sessions (Fulmer et al., 2018; Kretzschmar et al., 2019). Another reason why artificial companions might be so stigma-reducing is the perception of them being non-judgmental (Kumar et al., 2022). Thus, chatbots might be an especially helpful treatment in Asia where people with mental health issues face discrimination as society often sees them as dangerous and needed to be isolated (Lauber & Rössler, 2007; Sulaiman et al., 2022). In the study done by Pinchuk et al. (2021), stigma is perceived as one of the crucial reasons why people avoid treatment.

In general, evaluating the benefits of implementing chatbots in mental healthcare, it is evident that they display qualities that are especially needed considering the current crisis in service provision. While the inclusion might have certain disadvantages that would be described in the next segment, the strengths demonstrate that this technology might serve as a good starting point for a lot of people who cannot afford to attend therapy fully or are not ready due to the existing assumptions about individuals with mental issues.

Threats of AI in Therapy

As with any intervention, the application of technology for treating mental health issues has negative outcomes that should be taken into consideration. One of the most mentioned in the study remains the problem of bots being primitive and limited in terms of context understanding (Vilaza & McCashin, 2021; Yang, 2020). This might be especially problematic as chatbots can repeat the same bias that already exists in therapy. For example, Brown (2018) criticizes traditional therapy for being not suitable for the problems of queer people, ethnical minorities, people of color, and women who are especially vulnerable when using therapy that is not focused on recognizing the social injustice and reinforces the stigma. If a bot is designed without knowledge of this, then it will be able to see all cases shared by a client as irrational, this might make the client adjust to the traumatizing environment. An infamous example of this is the case of Woebot replying to the 12-year-old boy reporting the sexual harassment with “Sorry you’re going through this, but it also shows me how much you care about connection and that’s really kind of beautiful” (Vilaza & McCashin, 2021, p.2). Clearly, this context blindness is a significant issue that undermines the unsupervised usage of any chatbot.

When implementing artificial intelligence in chatbots, it is very crucial to check the data used for the training. AI in such technology allows to avoid dry answers compared to the hard scripted programs, however, this requires a large dataset to be applied to guarantee the variance of response and understanding of the user’s message (Luxton, 2016). In order to generate an answer that is both similar yet unique for each user, the machine is fed a dataset that contains possible inputs and outputs (Zhang & Li, 2021). This requires the application of Big Data which helps the machine-generated speech look more sentient, but often lacks diversity in the sample despite the size which results in replication of the existing biases (Bender et al., 2021). For example, GPT-3, or third-generation Generative Pre-trained Transformer, is a model that is claimed to be trained on the whole Internet often criticized for gathering data mainly from white and male-dominated spaces (Bender et al., 2021). This suggests that although modern AI programs are trained on data that exceed human understanding when designing a bot, there is a need for data to be checked.

Despite chatbots being cheaper than traditional therapy, they remain a product of a large business industry that might aim to gain benefits rather than help people in need. Taking into account the aforementioned cost advantage of this technology, it is often explained by the data being sold to third-party companies (Gratzer & Goldbloom, 2020). Other bots are designed in a new way to extort money from people by abusing their loneliness. Depounti et al. (2022) view Replika as one of the most extreme examples of how commercial interest might lead to psychological harm to clients due to its advertisement as an AI friend to improve mental health, yet often suggesting upgrades that include sexual content.

Considering the mentioned limitations, chatbots regardless of whether they are heavy-scripted or apply Artificial Intelligence continue to be non-sufficient for becoming a fully independent mental health intervention. However, research suggests that this should not be used as a point for abandonment of the idea but rather investigate the options for the integrated therapy, i.e., an approach that includes both therapist and bot where drawbacks of one technique are mitigated by the pros of another (Luxton, 2016).

The Present Study

The application of AI in psychotherapy remains quite an ambivalent topic. While there is proof of the efficiency of mental health chatbots in the short term, they are usually tested for 2 weeks which is shorter than the expected duration of CBT courses (Fitzpatrick et al., 2017; Gabrielli et al., 2021). Gabrielli et al. (2021) also highlight that the studies need to include a control group to compare the change in behavior with the baseline.

As most of the research is conducted in the US and using a privileged minority as a sample, current findings cannot be claimed as universal and need to be tested on people outside of that scope to diversify findings (Abd-alrazaq et al., 2019; Macleod et al., 2020). Therapy approaches are not ever-changing and need to be tuned according to the region.

Regarding the development of the program itself, Abd-alrazaq et al. (2019) report that only 7.5% of AI chatbots use a machine-learning approach (p. 5). D’Alfonso et al., (2017) persist in the incorporation of machine learning to ensure the eagerness of users to share the information with a bot. If implemented, this technology will boost the abilities of the bot as well as increase the resemblance of the conversation with the human speech considering that data used for training is carefully gathered. Gratzer and Goldbloom (2020) advise that bot engineers should be also mindful of the ethical principles to prevent data leaks and explicitly inform users about the way how their data is used in a concise manner. It is strongly recommended to invite specialists and discuss with them the process of bot development as well as conduct testing with them due to some bots being developed by people outside of the mental health field (Gratzer & Goldbloom, 2020).

Most of the contemporary research is focused on checking technology with the intention to prove its independence and does not analyze the phenomenon of integrated therapy and how it can be built. Although products warn that they are not substitutes for a mental health specialist, they rarely suggest other sources of help or suggest how programs can be supervised. Elomia is one of few examples of software that suggested using itself as a self-reflection tool between actual therapeutic sessions (Romanovskyi et al., 2022).

This project aims to close the gap found in the research by developing the artificial intelligence chatbot and properly testing it. In addition, the work includes the creation of a manual that can be further used by clinicians to practice integrated therapy to delegate some of the therapeutic tasks to a bot. The process consists of several parts. The first one is the creation of the Telegram AI bot that can differentiate between different distortions and apply Socratic questioning. This stage includes finding the neural network model, tuning it with the dataset gathered from locals, and testing the demo version of the program with practicing psychologists while developing a manual that can introduce them to the bot’s functionality like a therapist would do. The second stage is measuring the bot's efficiency in clinical settings consisting of recruitment, dividing participants into two groups, measuring the difference in the mental state between participants of both groups and statistical analysis to determine the efficiency of the bot. The experimental method is chosen to complete this task with the primary hypothesis being that using an AI bot will lead to a reduction in the level of cognitive distortions. The study contains several secondary hypotheses such as using the bot will reduce the level of some distortions more than others. It is anticipated that there is a statistically significant correlation between the regularity of using the bot and a decrease in negative thoughts. Moreover, the positive direction of this correlation is hypothesized. Finally, there is an assumption that there will be no statistically significant difference in results related to gender or academic division. The last part of the research includes interviews with psychotherapists to gather qualitative data to evaluate the bot from the side of stakeholders who are interested in implementing the bot in their practice. This feedback will lead to future recommendations for the inclusion of integrated therapy in Central Asia to make mental health services more affordable and accessible. These stages are described in the following section.

Method

This chapter aims to describe the methodology applied in this project. Due to the mixed nature of this research, each section includes details of both methods with the first being the quantitative study based on experiment and the second one applying qualitative method using expert interviews. The participants section focuses on the human subject that took part specifying how they were required as well as demographic data. The materials section will inform on what tools were used to measure the necessary data. The research design is devoted to the explanation of the structure of the experiment and expert interview. The procedures part includes all information on how the AI bot was developed and data collected.

Participants

Experimental testing included 68 human subjects (54 women, 14 men, $M_{\text{age}} = 20.8$ years, age range: 18-28). They were recruited using convenient sampling. Recruitment was conducted by spreading information about the project through social media and the university newsletter. Inclusion criteria included being over 18 years old and an AUCA student who speaks English. AUCA students were chosen as a population due to their unique features that satisfied all criteria for the research. AUCA is an English-taught university, thus, its students might not face the language barrier that might have arisen with the other group as the program operates in the English language. The university is also very demographically diverse (Kyrgyz, Kazakh, Tajik, Slavic, Afghan) which will help not only get a more representative sample but also to explore how mental health issues are better to be solved with regard to the region considering the urgency of the situation (UNICEF, 2020). Being mainly citizens of Central Asia allows this research to contribute by adding the perspective from our region. Additionally, students are a crucial part of the young generation, the segment for whom new technology might be especially helpful.

Participants were mostly females (79,41%) and citizens of Kyrgyzstan (76,47%). In addition, most of the participants (82,35%) have never used chatbots for mental health or companionship before. Participants were semi-evenly distributed among four AUCA divisions: social sciences (SS), applied sciences (AS), humanities (HUM), and business and law (BL). The distribution across years of study was also semi-even with freshmen being the least populated group (5,88%). Table A1 shows the spread of the sample in more detail (see Appendix A).

The experiment was reviewed and approved by AUCA Institutional Review Board (see Appendix B for approval). Subjects indicated their consent in the initial message from both bots (see Appendix C). For safety purposes, participants also were given contacts of AUCA Counseling Service. Subjects received 500 soms on the specified phone number for completing the conditions of participating in the experiment.

In the qualitative part of this research, seven practicing psychologists agreed to participate. Inclusion criteria were being a practicing psychologist who has knowledge about CBT, works with young people, and speaks English. Participants were recruited using both convenient sampling and snowball sampling. Study enrollment materials included posts on Instagram asking for suggesting professionals who might be interested in giving interviews on the topic. After each interview, participants were asked if they could have given the contact of a specialist who might be interested in participating in the research. Subjects did not receive any financial incentives for partaking in the research.

All seven participants were working mostly with young people from 18 to 25 old, with two AUCA Counseling Service members. Regarding specialization, five out of seven psychologists were specialized and applied mainly the CBT approach in their therapy. On average, experts were practicing for three years, with one outlier of a 20-year experience. All responded that they had at least one client requiring sessions in English, and two AUCA Counseling Service psychologists responded regularly worked with people in English. The majority reported conducting sessions offline. None of the participants practiced augmented therapy before, with one having clients who practiced talking to bots while in therapy without supervising the effect of the apps.

The expert interview was approved by AUCA IRB (see Appendix D for approval). Experts were informed about the interview being recorded and submitted the informed consent form to the research (see Appendix E).

Materials

Acceptance and Action Questionnaire -II (AAQ-II)

The second version of the Acceptance and Action Questionnaire (AAQ-II) is a 10-item, self-report questionnaire that assesses the level of "experiential avoidance and psychological inflexibility" (Bond et al., 2011, p. 6). The first is defined as the avoidance of negative thoughts and experiences, and the latter – as unwillingness to adapt to changes (Bond et al., 2011). The previous version (AAQ) is a very popular tool for measuring those two constructs with fine validity and reliability while the AAQ-II achieves the same results with a higher psychometric consistency (Hayes et al., 2004). Three items were removed due to reported superfluity, so participants received a 7-item version of the questionnaire (Bond et al., 2011). AAQ-II is scored on a scale from 1 (never true) to 7 (always true). After the scores are summarized, the higher results on the test are associated with a higher level of experiential avoidance and psychological inflexibility. Although AAQ-II does not diagnose any psychological disorder, in the 7-item version, the range of 24-28 establishes threshold values at which a test-takes is susceptible to developing a psychiatric condition (Bond et al., 2011). The version of AAQ-II used for this research can be found in Appendix F.

Cognitive Distortions Scale (CDS)

Cognitive Distortions Scale (CDS) is a 20-item, self-report questionnaire that evaluates one's learning to apply 10 patterns of distorted thinking (Covin et al., 2011). Each distortion has two parts to be tested: interpersonal, i.e., how an individual perceives social interactions, and achievement, i.e., attitudes to personal achievement and failure. However, Özdel et al. (2014) report that factor analysis demonstrated that both parts could be assessed by a unitary scale with a higher internal consistency. Thus, it was decided to decrease the number of items to 10 which would measure the level of each cognitive distortion in general. The questionnaire also provides detailed examples of how such thoughts can be manifested in everyday life that might help a test-taker to understand a complex notion of distortions. For familiarity, cases given by questionnaire were adapted to the Central Asian context. CDS is often used for assessing cognitive distortions specifically suggesting it is helpful in clinical trials and research (Özdel et al., 2014). The measure is scored on a scale from 1 (never) to 7 (all the time). CDS is not used as a diagnosing tool, however, it correlates with results of other anxiety and depression scales as cognitive distortions might contribute to the development of these disorders (Covin et al., 2011). Despite higher scores being associated with a higher level of cognitive distortions, there is no cutoff value to imply the tendency to form a mental disorder. The modified version of CDS is provided in Appendix G.

Expert interview protocol

For the purposes of qualitative research, a 9-question expert interview protocol was designed (see Appendix H). The document consists of three sections: satisfaction with the AI bot, views on the implementation of the AI bot in augmented therapy, and possible challenges. The first section covers the general question about how a bot is perceived by health professionals while the second and the third ones are expected to investigate how expert opinions correlate with the existing literature on the advantages and disadvantages of the application of AI bots in therapy and adding more information on the further work needed to improve the quality of the bot. The protocol was developed in English and translated into Russian.

Research Design

The research is constructed in several stages. The first one focuses on building TeaBot, a Telegram AI chatbot developed to work with clients' distorted thoughts by recognizing cognitive distortions and addressing questions in a Socratic way. A detailed review of its development and features can be found in Part I of the Procedures section. The second includes experimental testing of TeaBot, and the third discussed the AI bot as a tool for augmented therapy with practicing counselors.

The main research question of this project was whether using TeaBot could reduce one's level of cognitive distortions. Another question was whether the bot was more effective in reducing one type of distortion compared to others. Quantitative research was applied to test the effectiveness of the bot due to objective and precise results that can be generalized (Gorard, 2003). The experimental design was chosen to estimate the influence of using a chatbot on one's well-being. The research was conducted in a form of a field experiment to observe the behavior in real-world settings which might produce results that might be generalized in a non-laboratory environment (Cook et al., 2002). Interaction with bot was considered to be an independent variable, and one's level of cognitive distortions was the dependent one. Taking into account the ambiguous reputation of AI chatbots as a treatment, a control group was added to prove counterfactual results and demonstrate the difference between what occurred after treatment and what could have occurred without it (Cook et al., 2002). The experimental group experienced intervention in the form of unlimited access to TeaBot with the manual about cognitive distortion designed for this project (see Appendix I) while the control one received the manual only. This document was supposed to assist users with understanding CBT concepts to ease the process of working with a bot (experimental group) and serve as a placebo to initiate the self-help process (control group). It was designed as a temporary surrogate for mental health professionals who would explain how the bot works and what therapeutic concepts are applied to change one's mental well-being. Considering that TeaBot is designed as a bot for augmented therapy, the manuscript can be also used as the script for therapists on how a bot can be better introduced to the user. In addition to the manual, the control group also was supposed to use Assessment Bot to fill out questionnaires required for participating in the experiment. The Assessment Bot does not provide any therapeutic intervention and was needed to create a similar environment for both groups during the test-taking process while also reducing contact between the principal investigator and participants.

Similar studies report attrition rates achieving on average around 20% (Gabrielli et al., 2021; Fitzpatrick et al., 2017). Additionally, due to the duration of the experiment and the need for regular communication with the bot, drop-out from the experimental group was expected to be higher than in the control group, hence, it was decided to apply a 3:2 ratio when dividing people into groups. Following the requirements of experimental design as described by Cook et al. (2002), participants were assigned according to the norms of the randomized experiment: each participant was assigned a number from 1 to 5; participants with odd numbers joined the experimental group while the control group consisted of those who got an even number.

Measuring the experience of groups was done according to the “pretest-posttest” scheme. According to Cook et al. (2002), this allows to increase internal validity and reliability. Moreover, having pretests for both groups allows us to test whether they are different and check the presence of selection bias (Cook et al., 2002). All assessments were taken via bots for 24 hours. As the conditions for test-taking might differ for all participants, the second pretest was included to verify the absence of difference in pretests and mitigate the effect of the test not being taken in a standardized environment. The number of assessments was decided to be limited to three to avoid the fatigue effect. Thus, participants were required to submit two pretests, one 4 days before the intervention and another at the baseline, and one posttest at the end of the study.

The duration of the experiment was calculated both on the existing research on the application of technology in mental healthcare and standards of therapy. The advised number of CBT sessions usually revolves around an eight-session period (Beck, 1995; Cully& Teten, 2008). Dekker et al. (2020) report that for young people 8-week interventions demonstrated the highest efficiency. As most of similar experiments usually lack longer exposure to the technology and are limited to two weeks, it was decided to conduct the experiment for 8 weeks.

As the AI bot was designed not as a substitution for therapy but rather as a tool for augmented therapy, it was decided to consider the opinion of other stakeholders, therapists who would be interested in integrating the AI bots in their therapy. A detailed review of this part of the study is described below. The research questions for this part of the study were, “How do practicing psychologists in Kyrgyzstan evaluate TeaBot as a therapeutic tool?”, “How do they see the application of TeaBot in Kyrgyzstan?”, and “What are the challenges and opportunities that the application of TeaBot brings to therapy?”. As there is a lack of research on augmented therapy, this topic requires further exploration and openness to unexpected comments, thus qualitative approach was chosen (Willig, 2008). As the expert interview would probably reveal some segments that can be grouped into topics, thematic analysis was applied to inspect this (Clarke & Braun, 2013). Thus, this approach would help identify key connections between the research question and collected data.

Procedures

This segment explains in detail the procedures done to complete the research. The first part focuses on the description of how TeaBot was designed with the second explaining data collection and analysis.

Part 1: Developing TeaBot

Before starting coding, one needs to decide on the evidence-based therapeutic approach to build the bot on. Cognitive behavioral therapy was used as an inspiration for the bot as it focuses on changing a client's well-being by learning to recognize and challenge negative thoughts and behaviors (Beck, 1995). These tasks are always practiced as a part of the homework that the patient is expected to practice (Beck, 1995). CBT is also very structured and requires psychoeducational elements to be present, and this allows us to imagine how technology might fit into the therapeutic agenda. Implementing such activities would require an AI that is capable of recognizing natural (human) language on a very high level, thus, applying a neural network. This is a machine-learning model inspired by the human brain that employs neurons, i.e., interconnected nodes, to detect patterns and correlations in data, while also learning from examples (Luxton, 2016). Following the literature review, the bot was determined to be executing specific tasks only and serve as a CBT exercise rather than a friend-interlocutor. The bot was expected to perform the following tasks: educate and remind patients about certain aspects of CBT, recognize healthy thoughts from unhealthy ones, classify the detected distortion, identify the most appropriate common question to challenge the thought, and create the most context-fitting version of it. Meanwhile, all of these should be hosted in a user-friendly environment, like an app or website. The first two tasks were determined to be written by the principal researcher and directly incorporated into the user side rather than designed by AI due to the ethical complications that might arise in the case of the model's inability to explain complex notions to the patient. Thus, the process of bot engineering consisted of several stages that required different technologies. The first segment describes the aspects of integrating artificial intelligence part in the bot's conversational logic while the second one indicates how it was incorporated into the user interface and also informs about additional features that were added for the validity research and user convenience. The final part is dedicated to the procedures the first version has undergone to ensure that the participants are well-informed and safe.

Back logic: GPT

The end goal of AI-part was classifying the thought, identifying the right base question, and creating the response to the user. From the development perspective, it was crucial to use a good language model that would be able to understand human language while also being able to perform a variety of functions when interacting with the input. As building own neural network capable of the project's objectives would be a very time-consuming and resource-heavy task, alternatives were explored. OpenAI created Generative Pre-trained Transformer 3, or GPT-3, a language model that can be used by anyone for a considerably small amount of money and does not require significant computational power (Han et al., 2021). Its main advantages remain a very large training dataset and rich functionality while the cost is very low (Zhang & Li, 2021, p. 832). However, the model is not admitted to being fully autonomous when interacting with human beings as GPT-3 tends to generate sentences that might lack semantical meaning while also heavily relying on the purity of data (Dale, 2021; Floridi & Chiriatti, 2020; Zhang & Li, 2021). Despite these limitations, the language model developed by OpenAI persists in being the strongest in the market especially when compared to the ones crafted by enthusiasts with limited sources. Consequently, GPT-3 was chosen as a base for AI logic.

This AI model was not designed to be used as a therapeutic tool, the mere pasting of the problem often resulted in misunderstanding and giving advice. The latter is avoided in therapy as this often leads to developing reliance on external resources which is the opposite of goal therapy pursuits (Corey et al., 2011). In Figure J1, a case for misinterpretation of minimizing is shown. Although the input contains keywords that for a therapist will be a sign of cognitive distortion, for both models it is positive self-talk (see Appendix J). Hence, even in some cases when GPT-3 might generate good advice, this approach cannot be considered therapeutic. To eliminate the already described drawbacks, it was decided to limit the model's agency by applying fine-tuning, i.e., add dataset to specify tasks and specific patterns of questions that AI is allowed to use.

As Burns and Beck (1999) identified 10 types of cognitive distortions, a bot needs to successfully differentiate between them as each type requires a particular approach to work with. GPT's fine-tuning option allows creation of a new sub-model that is specifically trained for classifying distortions. To complete it, a developer is required to create a labeled dataset, e.g., "thought" – "type of distortion", and choose a version of GPT-3 needed for training (see Figure J2 in Appendix J). To avoid the bot presenting as a "know-it-all" machine, an "unclassified" category was added where the model sorted thoughts that did not fit the training set.

The data collection for this labeled dataset was gathered by a combination of examples from CBT literature and anonymous submissions made by the principal investigator and AUCA psychology students who got access to the document by a link that was shared in student group chats. This sample was chosen due to them both being part of the population that will go through the experiment and study the CBT concepts, thus, knowing how to label data. They were asked to write 10 cognitive distortions in total. After the completion of the dataset, the whole dataset was checked and edited by the principal investigator, the project's supervisor, and a practicing

CBT psychologist. In total, 240 examples of cognitive distortions were accumulated and divided into training and test sets in a ratio of 3 to 1. Out of 4 available GPT-3 versions, curie was chosen due to its cost-efficiency and high performance with the testing set. After choosing the model, the fine-tuning stage was completed.

The next stage is applying Socratic questioning. Due to GPT-3 limitations, instead of asking the questions from scratch, a separate list of base questions, i.e., appropriate to use when challenging particular distortion, was created. Whenever a model recognizes a type of distortion, it also accesses that list and chooses the most appropriate one to challenge the thought. When a question is chosen, the model is asked to come up with a version of a base question that is most appropriate to the problem user described. Different versions of GPT-3 were again tested for this task, with davinci demonstrating the most human-like responses. The AI was given a prompt that asked to compare the user input and ask a question using the base question as an example. The outcome of the whole GPT-3 process is the user receives both the category of cognitive distortion that the program caught and a question that helps to challenge the thought (see Figure J3 in Appendix J).

User-interface: Telegram

After the core AI logic was assembled, it needed to be wrapped in a user-friendly interface. This requires a decision on the format of the application, however, there is no unambiguous answer for this. Standalone applications give the advantage of being able to send user notifications that are often reported by users as an uplifting feature that eases the bonding process with the bot (Fitzpatrick et al., 2017). Adamopoulou and Moussiades (2020) report that websites can be seen as more liberating ways of building a program with developers not limited by the policy of the host application. However, both mediums require significant data protection as any case private information leaking would be damaging to users (Adamopoulou and Moussiades, 2020). An alternative to that might be an integration of a chatbot into the environment of an existing messenger. Kretzschmar et al. (2019) disclose that this approach might help with engagement and adjustment to technology too as they already exist in a familiar environment. Taking into account Telegram's security measures and friendly environment for chatbot implementation, it was chosen as a platform (Albrecht et al., 2022).

The bot was given the name TeaBot. There is a long tendency for artificial interlocutors to have female names and perform female behavior during communication (Depounti et al., 2022; Fulmer et al., 2018; Gabrielli et al., 2021; Weizenbaum, 1966). Depounti et al. (2022) criticize this approach for reinforcing the existing bias of serving and caring roles that women are expected to have in society. Fitzpatrick et al. (2017) inform that even a specifically chosen robotic name does not interfere with the human tendency to build empathy towards nonhuman agents. Nevertheless, to avoid any reproduction of existing gender stereotypes and reduce the effect of association with a human being on results, the TeaBot name was picked to remind a client that they communicate with a non-human.

Users interacted with TeaBot in the following manner: informed consent, greeting, and access to three functions: manual, assessment, or let's chat (see Figure J4 in Appendix J for a detailed scheme). During the very first meeting with the program, one receives informed consent and a tutorial on how to use the chatbot. This is needed to ensure that the user is well aware of what the chatbot can and cannot do. Users are obligated to give consent in order to continue using the application. After that, the greeting function was added to break the initial discomfort. The bot also asks users to come up with a name if the username in Telegram is not specified. After that, TeaBot suggests its functionality in a continuous flow manner until the user chooses to end the interaction by typing “/bye”. The new conversation starts with a greeting without asking for consent again.

The manual function delivers to the user information about the Bot and acts as a therapist who would introduce the bot to the clients. The manual consists of three sections: “therapeutic techniques used” which mainly explains CBT, the concept of distorted thoughts, their types, and how to battle them; “conditions for participation” which is used to get a copy of informed consent user agreed before, and “ethics and safety” that informs what institutions approved the bot and how data is collected. In the future, this manual can be used as training for practitioners who would like to include the bot in their practice.

The assessment function was developed to test the process of data collection by the bot. The results of all assessments are recorded on a separate file which allows to store data more efficiently compared to the traditional paper structure. This function was mainly used for completing the validating part of the research as the digital assessment allows to gather information from participants much faster, from different places anonymously without breaking contact with the application compared to the traditional Google form approach. This segment can also be later redesigned to add additional assessments depending on the needs of the client.

Let’s chat is an AI-powered function that is a fundamental feature of TeaBot. The dialogue starts with a psychoeducational question that aims to filter whether the bot can be actually helpful to a person. As the bot is designed to work only with cognitive distortions and not mere chit-chat, the program explains to the user what a cognitive distortion is and asks whether a person has any thoughts one can identify as distorted. When receiving a positive reply, TeaBot invites users to share their thoughts in a cause-effect format. After that, the GPT-3 functions are turned on. Users can talk to TeaBot for an unlimited amount of time until sends "/stop_chat" which ends the process.

Preparation for experiment

For TeaBot to be able to participate in the experiment to test its validity in a big sample of human subjects, there are several requirements that it needs to pass.

Due to the interdisciplinarity of the project, the pre-deployment review was done by experts in both psychology and software engineering. Two practicing counseling psychologists who studied CBT were interviewed to gather feedback on the bot's work, one being a project supervisor and the other being a practicing CBT counselor contacted based on the supervisor's suggestion. The latter also joined the qualitative method as a follow-up. In general, both counselors expressed their satisfaction with the overall work of the bot, although the first tester expressed their concerns about the bot not always guessing the distortions correctly and suggested emphasizing that the bot is in its test mode. This was addressed with the expansion of the dataset used for the bot training and adding more information for additional resources. The second participant, on the other hand, was very impressed and enthusiastic about future applications of the bot and it's being a probable addition to CBT for young people. Other comments from both experts mostly focused on the improvement of the communication of the bot, e.g., adding a tutorial to explain how the bot works, implementing more empathetic language to the bot, and reducing the number of items in the questionnaire. As a follow-up a tutorial was recorded and was accessible to all users through the link bot sends when users ask for instructions. Questionnaires were modified according to the comments and literature review suggestions.

After this, necessary changes were applied and interviews with two software engineers were conducted to discuss what issues might arise after deployment and how they can be prevented. The first was a back-end developer who focused on hosting the bot on the cloud server and suggested options for more convenient data collection and analysis. The feedback from another programmer specialized in bot development included applying more Telegram features, e.g., pinning messages and adding the "/help" command for easier navigation. Both requested to reduce the size of the text and use more day-to-day language.

Technically, the bot was hosted on the virtual server to ensure that the experiment both bots will endure the 24/7 work with multiple users accessing the technology. Data backup was conducted automatically every day to prevent losing data in case of a server fall.

By the end of this stage, TeaBot was to be fully deployed and start the experiment to validate its efficiency. The details of this process are described in the next part.

Part 2: Data collection and analysis

During the recruitment phase, social media posts included the principal investigator's contacts for potential experiment participants to indicate their interest in the study. As the inclusion criteria included being an AUCA student, human subjects needed to indicate their institutional email when joining the experiment. After the list of participants was finalized, they were divided into experimental and control groups randomly in a 3:2 ratio. All subjects were informed of what group they joined. Both groups received instructions and bots' links via email. On the day of each assessment (4 days before the baseline, baseline, and end of the experiment) all participants received the reminder to submit the assessment. After that, only members of experimental groups received reminders to talk to TeaBot twice a week at 9:00 throughout the duration of the testing. At the end of data collection, all participants were contacted to send a phone number to which they would like compensation to be sent.

All data was made to be collected by bots to preserve respondents' anonymity of confidentiality and avoid social desirability and researcher bias. As the researcher was aware only of participants' names and e-mails, there were minimal chances of being able to track participants' identifying information. All subjects were given unique IDs that were generated by the bot and were impossible to reverse. Only principal the researcher had access to the raw experimental data. The content of the messages that users sent was not collected, instead, the bot recorded times when a subject contacted the AI function of the bot.

Regarding the qualitative part of the research, after enrollment in the study, an interview was scheduled in a format that the interviewer indicated being comfortable with (online/offline). All experts received access to TeaBot before the interview to have enough time to explore its functions. All interviews were conducted by the principal investigator and recorded for the purposes of high-quality transcription afterward. Although all participants reported being fluent in English, communication was held in Russian as it was the native language for the interviewees. During the meeting, subjects also had an opportunity to use TeaBot and ask developer questions about technology. In total, in interview lasted for 40 minutes. The AUCA Institutional Review Board has evaluated and approved this part of the research too. The data gathered during this stage was recorded, transcribed, and analyzed by the investigator herself.

Statistical analysis was completed using JASP software. First, the researcher checked whether the data was normally distributed using the Shapiro-Wilk test. To test whether the intervention in the form of TeaBot is statistically significant, independent samples T-test was used by access the difference between pretest and posttests among two groups. Then, an independent samples T-test was conducted to clarify whether there are differences between the initial state of both control groups. To identify whether there are significant differences between pretests paired samples T-test was applied too. To evaluate the extent of differences between the groups, effect sizes using Cohen's d were computed.

After the primary efficacy investigation, the data was analyzed on the correlation between use in the form of a number of inquiries to TeaBot and weeks interacting with it. ANOVA tests were used to check whether there are particular academic divisions for which TeaBot is more helpful. An Independent T-test was applied to calculate whether gender might be a significant contributor to the efficiency of the bot. Correlations between the activity of users with their AAQ-II and CDS scores were tested using Person's test. Factor analysis was utilized to inspect cohesion between these two measures and identify key underlying areas that TeaBot is able to work with. In order to check the assumption regarding whether a bot can be more useful with particular cognitive distortions, a repeated measures ANOVA test was applied. Paired sample T-test was also used to identify items that demonstrated significant change between pretest and posttest.

Results

AI chatbot

The main product of this research is the TeaBot, an AI chatbot that can be used for working with cognitive distortions, that was validated quantitatively by users, and possibilities of which were explored qualitatively with practicing psychologists (see Figure J5 for TeaBot's interface). The project aimed to create software that can be used to help close the gap in providing mental healthcare for young people. The program can execute the following functions: taking assessments and saving them in a database, giving psychoeducational content, and chatting using an AI model fine-tuned on the data gathered for this project specifically. The software also includes the manual that aims to serve as a mediator for the future implementation of augmented therapy where a bot is used in tandem with a psychologist. Figure J6 depicts how the final version of TeaBot communicates with users in AI mode (see Appendix J).

Quantitative results

The second result of this research is a quantitative validation of the TeaBot's efficiency by conducting an 8-week experiment on the target audience. The following segment focuses on sharing the findings of this stage.

Attrition

Figure J7 demonstrates the flow of the participants during the project. Out of 110 participants who contacted the researcher interested in the experiment, 81% (35/54) submitted partial data, e.g., at least one pretest, and 62% (34/34) completed the experiment (see Appendix J). Attrition rates differ between the control and experimental group with 27,66% for the control group and 46,03% for the experimental. Nevertheless, there was no difference in pretest-1 results detected between those who completed the experiment and the ones who withdrew, $t(87) = 0.146, p = 0.884, d = 0.25$.

Use

On average, 29 out of 46 users who completed the second assessment used TeaBot every week. The maximum number of users was observed during the first week (43 users), and the minimum was detected during the sixth week (18 users). The average duration of bot usage was 6 weeks, with 12 people (35,29%) interacting for all 8 weeks. In general, users contacted the bot on particular dates that coincide when notifying emails were sent. The maximum number of messages sent by one user in a week was 47, minimum – 0. Reflecting on this variance and the presence of outliers, the median was calculated for each week and, consequently, the mean of these values was equal to 6.125 suggesting that, on average, users sent 6 messages per week. The mean number of check-ins throughout the 8-week period was 51.

Regarding the timing of the messaging, bots are often referred to as highly accessible tools that can be used when human therapists are out of reach (Kretzschmar et al., 2019). Consequently, the hours when users contacted TeaBot were divided into three categories: night (from 01:00 to 9:00) when most specialists are unavailable, working day (from 09:00 to 17:00) when clients can approach counselors for a session, and post-working hours (from 17:00 to 01:00) when most psychotherapists do not conduct sessions (see Table A2 in Appendix A). Interestingly, the difference between the number of messages sent during the working (*N* = 2593) and past-working (*N* = 2350) hours is small suggesting that people used TeaBot both when therapy was available and not. The number of messages peaked after 19:00 (*N* = 450) with the second maximum being after 11:00 (*N* = 409), with the lowest point of activity being after 5:00 (*N* = 0). A breakdown of activity by hour is depicted in Figure J8 (see Appendix J).

Efficiency

Table A3 describes that the pretest 1 of the control group (*M* = 69,79, *SD* = 14.2) and experimental group (*M* = 75.44, *SD* = 18.75) did not exhibit any significant dissimilarities with a small effect size, *t*(66) = 1.258, *p* = .213, *Cohen's d* = .31 (see Appendix A). The difference between the two pretests of the experimental group was also found insignificant despite the data not being distributed normally which required calculation using Wilcoxon signed-rank test, *p* = .285. This allowed the researcher to assume that the results of pretest 1 and pretest 2 can be indeed used interchangeably. The results of the difference between pretest1 and posttest of experimental (*M* = 17.44, *SD* = 16.67) and control (*M* = 2.82, *SD* = 15.65) groups diverge significantly with a large effect-size detected, $t(66) = 3.728, p < .001, d = .90$ (see Table A4 in Appendix A). The Shapiro-Wilk test was used to check if groups were distributed normally. Assumptions for both tests were proven to be true. The group that used TeaBot demonstrated significant differences between pretest and posttest in both AAQ-II ($t(66) = 3.32, p = .001, d = .81$) and CDS ($t(66) = 3.30, p = .002, d = .80$) while both passing assumption checks. Figure J9 demonstrates the change within an 8-week period for both groups (see Appendix J).

Pearson's correlation was used for determining the relationship between the number of weeks a subject interacted with TeaBot and one's change in the pretest-posttest. The results displayed the presence of a significant positive correlation ($r(34) = .362, p = 0.035$). The detailed results as well as a visual representation of the data correlation can be found in Table A5 in Appendix A and Figure J10 in Appendix J. The correlation coefficient for the total number of messages one sent to TeaBot and the aforementioned change in scores was statistically insignificant and slightly negative ($r(34) = -.005, p = .98$).

Regarding gender, the independent samples T-test demonstrated a significant difference in the effect between male and female participants passing normality checks, $t(32) = 2.18, p = .04, d = .44$. When calculating ANOVA on the influence of division on results, the test shows the p-value is equal to .536.

Analyzing the differences in effect on distortions regarding the type, participants' scores on each item were compared before and after the experiment using a paired samples T-test. Table A6 in Appendix A present that out of 17 variables, 15 demonstrated statistically significant change ($p < .05$) with minimizing being the only distortion that did not demonstrate strong change ($t(33) = .90, p = 0.37, d = .15$). In addition, repeated measures ANOVA showed that there is indeed a difference in the magnitude of change depending on the type of distortion. All distortions were reduced, ranging from 0.27 to 1.38, with the personalization depicting the biggest degree of reduction and minimizing demonstrating the lowest one and being the only outlier in results (see Table A7 in Appendix A and Figure J11 in Appendix J).

Materials

An exploratory factor analysis was conducted with the maximum likelihood method applied to analyze the structure of AAQ-II and CDS measures. The findings indicated insignificant Chi-square results, *p* = .98 (see Table A8). Five eigenvalues greater than 1.0 were calculated: 5.32, 2.73, 1.53, 1.23, and 1.11, which contribute to 70% of the variance (see Table A9 in Appendix A and Figure J12 in Appendix J). The loading factors for these five factors did not show any outliers within the variables and displayed moderate to high values of internal consistency, ranging from 0.43 to 0.93 (see Table A10 in Appendix A). From the uniqueness perspective, most of the items fell under the domain of factors listed above with Mental Filtering standing out considerably (.82). The first cluster of variables mostly revolves around the thoughts that are associated with worry. These included items on catastrophizing, emotional reasoning, minimizing, mindreading, all-or-nothing thinking, and worries about not being able to succeed because of emotions. The second factor mainly revolved around control of one’s emotions and included most of the items from AAQ-II with should statements. The third factor shows strong relation with sorting behavior such as personalization, labeling, mental filtering, and item on the perception of other people as more successful from AAQ-II. In this factor, there was a negative correlation found with catastrophizing. The last two factors consisted of only one question each. The relationships between all variables are illustrated in the path diagram (see Figure J13 in Appendix J). Considering these results, the factors were labeled as “worry”, “control”, “comparison”, “worries about control”, and “generalizing”.

Qualitative results

During the interviews, several major themes were found (see Table A11 in Appendix A). The bot was mainly discussed as an application, in-therapy tool, and out-therapy tool. Feedback was also processed and analyzed from the perspective of recommendations for further work. The application theme mainly revolves around the overall perception of the bot by experts and its review. This included an assessment of its accuracy (*N* = 5), an evaluation of the medium within which the bot operates (*N* = 6), and a discussion on the target audience for such technology (*N* = 5). In general, as a software, TeaBot received quite high grades with most of the participants admitting its fine abilities to differentiate between distortions and explain the psychological aspects the program applies. There was noticed a connection between the medium of TeaBot and its audience as psychologists informed that being a Telegram chatbot, TeaBot might be utilized for this group as they might prefer it over the paper-based CBT homework. Talking about the application in therapy, the application remains quite flexible as almost all interviewees (*N* = 6) saw the TeaBot as a flexible tool with a variety of functions in their practice. The most common was implementation as homework (*N* = 5) and diary (*N* = 3). A few therapists saw the bot as an application to automatize data collection and assessment taking (*N* = 2). Regarding the other effects of augmented therapy, some experts also shared that they would use TeaBot for delegating some of the follow-ups between sessions to support a patient's immersion into the therapeutic environment (*N* = 3). Interestingly, even participants who did not see TeaBot as a part of their therapy practice, shared that they can imagine the bot being applied in the post-therapy process for patients who struggle with the process of reflection. Regarding the out-therapy employment of bot, the options also vary with emergency being the most popular topic to mention (*N* = 4). TeaBot was seen as a good tool to use when a therapist is unable to answer, e.g., at nighttime or during other sessions, with the practicing psychologists who monitor crisis lines highlighting cases when people were calling during the night with problems that the bot can help with too. Another version of an emergency function that the bot was seen doing was psychological first aid. Expert spotlighted the potential of bots being used for people who do not attend therapy and need immediate communication when in crisis. Therapists also mentioned the bot as a gateway app for those who cannot afford therapy or either have certain biases either about it or people who attend therapy (*N* = 2).

There were also challenges of utilizing TeaBot in therapy discussed. First, one participant rejected the idea of augmented therapy. This view was not provoked by TeaBot, rather specialists shared that they think that human contact is essential for therapy. A few more therapists mentioned that the bot might not recognize the type of distortion correctly (*N* = 3). Other remarks also revolved around the bot's style of communication as one participant shared that clients might find a style of pointing out to clients their mistakes in thinking rejecting and alienation. They suggested informing users more about TeaBot's specialty. This transits to the next often-mentioned subtheme related to the proper explanation to clients about what a bot is when it should be used, and how results were interpreted (*N* = 6). As the bot currently is not employed in augmented therapy, a lot of those aspects are mentioned yet in the form of a manual or tutorial that experts noted, patients might not be so eager to read themselves. In general, experts envisioned augmented therapy with TeaBot as them being responsible for explaining in detail the bot's functionality before giving it to a patient and expected the interaction with TeaBot to start somewhat in the middle or early middle of therapy. Noteworthy, it was underlined that therapists also needed a proper education about the bot's abilities and the borderline between what the bot is responsible for and what are the duties of a therapist to avoid misuse from the specialist's side that may result in providing clients with services of lower quality (*N* = 2).

There was a comment from one therapist who happened to be exposed to TeaBot for a longer period of time and was also using it as a patient who shared that the bot was not perceived as the bot. They shared that TeaBot was perceived more as a friend rather than a therapist. Overall, all 7 participants reported being impressed with such technology being in development and seeing potential in such innovations.

Discussion

The result of this project is applicable and ready-to-use AI chatbot TeaBot that was empirically tested. The fundamental question of this research was whether using TeaBot can decrease a user's level of cognitive distortions. This project also aimed to investigate what mental health issues the bot might be more useful for. Testing of the program was divided into several parts: technical that was mainly revolving around tuning; statistical which was organized in a form of an 8-week experiment with control and experimental groups; and qualitative which included discussions with practicing experts in mental healthcare to explore how experts see bot's application in their practice and what challenges exist in the implementation of augmented therapy in Kyrgyzstan.

From a quantitative perspective, TeaBot demonstrates promising efficiency in helping young people deal with their mental health issues. The main hypothesis was confirmed, and this finding is in tune with existing research on the positive impact of the application of mental health chatbots (Fitzpatrick et al., 2017; Fulmer et al., 2018; Gabrielli et al., 2021; Inkster et al., 2018; Romanovskyi et al., 2021). From the experiment data, there is a strong influence of using the bot on the reduction of one's levels of cognitive distortions and strengthening a healthier relationship between one's thoughts and actions. As data analysis shows that there was no significant difference between the initial states of both control and experimental groups, the impact was not achieved by preliminary features of any group, but rather the genuine effect of talking with the bot. After interacting with the program, the AAQ-II scores dropped under 24 points, the borderline of being prone to cultivate a diagnosable mental illness (Bond et al., 2011). This can be interpreted that TeaBot can be used as a tool to work with malicious thoughts that might trigger the development of the disorder in the future.

The 8-week duration of the experiment also allowed the researcher to observe and test the hypothesis on consistency in therapy. Interestingly, although was no significant correlation between the total number of messages and the drop in the level of cognitive distortions, there was a link between the number of weeks participants communicated with the bot and their mental state. The more people chatted with TeaBot, the more likely their scores were to become lower. This confirms the secondary hypothesis on regularity as the found correlation was statistically significant, moderate in magnitude, and positive in direction. This corresponds with the existing findings on how crucial the element of regularity is in therapy (Reardon et al., 2002; Sankar et al., 2021).

Before discussing the effect of TeaBot on different types of distortions, the evaluation of the materials should be also addressed. Factor analysis determined that there were certain aspects of mental well-being that might be a potential focus for TeaBot. The majority of items joined three clusters: worry, control, and comparison. Factors such as mixed worries about control and generalizing were not fully clustered to provide a comprehensive interpretation and were advised to omit. Notably, mental filtering revealing the high uniqueness values might suggest a more complex nature of this distortion. Overall, the analysis indicated that no measurement errors were made during the data collection; AAQ-II and CDS measures can be used together when working with cognitive distortions and applied as measurements efficiency of therapeutic bots.

Regarding the hypothesis that using TeaBot would reduce some level of distortions more than others, the results suggest that TeaBot decrease distortions evenly regardless of type. However, findings also showed that only “minimizing” did not decrease significantly over the duration of the experiment. Moreover, it was the only outlier within the group of distortions that all changed in similar degree. However, as the only other statistically insignificant change in the results of combined measurement was the first item, there is a certain ambiguity about these outcomes. One interpretation is that TeaBot might not be effective in minimizing distortion in particular. However, during the pre-experimental testing on the training set, the model did not exhibit outstanding difficulties with recognizing this distortion. On the other hand, this might be also due to combined primacy and recency effects. This assumption contradicts factor analysis as both variables showed strong internal consistency. Further research is needed to drive conclusions on this topic.

Both Fitzpatrick et al. (2017) and the data on the usage of TeaBot suggest that it is important to use notifications or any other check-in measures to ensure participants incorporate the bot into their routine as most of the conversations happened on the days of email reminders. As the findings exhibit that an average period of communication with TeaBot is 6 weeks, this also might be used as the base for machine-assisted therapy. More than one-third of participants talked to the bot every week suggesting that TeaBot might indeed be a helpful regular tool to work with their thoughts. Considering the outcomes, it seems that around 6 messages per week might be a sufficient addition to one’s work on mental health as it was an average number of messages sent to the bot weekly.

Attrition findings also check with the previous research (Fitzpatrick et al., 2017; Gabrielli et al., 2021). Moderate attrition levels might be explained as the sample was formed from the general population. People who did not need therapy might have dropped out due to the mere unnecessity to write to the bot as it was specifically designed for working with cognitive distortions. An additional reason for these attrition rates might be the duration of the experiment as an 8-week habit might remain difficult to maintain.

The hypothesis on the lack of correlation between academic division and a decrease in distorted thoughts was confirmed as there was no difference found in results according to the division implying that the bot is user-friendly regardless of academic division and can be used by people of all working backgrounds. Although Depounti et al. (2022) highlighted that application of AI might be damaging for women, the preliminary results suggest that the bot might be more helpful for female participants proving the alternative hypothesis, however, it should be noted that TeaBot was designed to be genderless which might have affected how it was perceived by both male and female participants.

Considering that for the majority it was the first time using the bot, the effect of TeaBot establishes that implementing bots for mental health purposes in Kyrgyzstan might be a very successful project. As the participants were students, this also indicates that mental health chatbots might be very effective when working with the youth.

From qualitative findings, TeaBot was perceived as a very helpful tool with a probable implementation in augmented therapeutic practices. Technology was seen as a very flexible tool that can be implemented differently depending on the needs of both a client and therapist. The bot was mainly seen as a delegatory tool that can ease certain educational functions of a therapist while also helping to create the habit of reflecting on own thoughts. These might be especially beneficial for young people as therapists also highlighted a higher probability of young people getting along with this technology better than with the traditional format of paper-based homework. It was suggested that the bot might need some aspects to be revised, mainly some forms of communication and categorization. Full implementation of augmented therapy was advised to include training for therapists as experts shared that the main functionality of the bot should be explained by a psychologist who led the healing process rather than the bot itself.

This project corresponds with the existing literature on the application of AI in the mental health field. Artificial Intelligence is a promising domain for improving mental healthcare due to accessibility, affordability, and anonymity. While standalone machine intervention is a highly disputed topic, chatbots can be integrated into therapy to help both practicians and clients by reducing the administrative workload, serving as additional contact with the therapeutic environment, and providing post-therapy intervention to practice skills gained at therapy. This project aimed to contribute to the present research by giving more information on AI performance in the realities of Kyrgyzstan and making the first steps in joined therapeutical practices while also developing a ready-to-use technology that realizes evidence-based practices.

Limitations

This project has several limitations that might have affected the research findings. First, the number of participants, although passing normality checks still reduce the external validity. The sample of male participants was not very large which might have affected the generalizability of the findings on gender. At the moment of writing, TeaBot operates only in the English language, which is not a wide-spoken language in Central Asia. Despite the current results in categorizing distortions appearing to be high enough to trigger a change in mental well-being, a bigger size of the fine-tuning dataset is advised to reduce the number of mishits.

Additionally, although the distribution into groups applied random sampling, the sampling method for the experiment itself was not random as only people who were interested and contacted the principal investigator were recruited. This might lead to the existence of selection bias. Regarding the measurements applied in the research, self-report tests are subject to response bias. As both AAQ-II and CDS were used to measure the general effect on participants' mental states, there was a lack of data to interpret the clinical implications of using TeaBot. The design of field experiment study suppose non-laboratory settings, hence, there might have been less control of additional factors that might have led to the reported outcomes. As students are the group that is highly sensitive to dates, there might be additional influence on the results of scores because of either presence or absence of the exam period. Lack of external validity might be also a factor in the qualitative analysis of TeaBot, and further studies should include a larger sample of experts. Sampling bias might also appear in the qualitative part of the project as the majority of the specialists interviewed were in the early stages of their careers (4 years or less) which might explain enthusiasm and openness to the technology.

Further work

The research design did not apply any follow-up measures, thus, this study is unable to declare any long-term effect of communicating with mental health bots. Future works might focus on longitudinal testing and include post-tests after a period of non-use to trace any effects of bots that remain even without interaction.

Due to time and resource constraints, the research mainly focused on the proof-of-concept model and did not fully integrate the therapists in the project. For the development of the field of augmented therapy, more studies on the efficiency of joint therapy need to be done. Considering that nowadays artificial intelligence and its ability to substitute people at certain jobs is a widely discussed topic, future research can test the scalability of similar applications and emphasize discussions on counselors' thoughts and feelings about AI colleagues.

The preliminary results on the difference in effect between the two genders suggest that the bot might be more impactful for females than for males. Nevertheless, as the gender distribution was not equal, it is hard to generalize the findings. As (Depounti et al., 2022) mentions, chatbots are often express female behavior and more used by man, thus, further research might focus on the factors that affect access to the bot in regard to gender.

The topic of bot perception was not included fully although mentioned in one of the interviews. Although TeaBot was purposefully named with a clear emphasis on non-human origin, it was still mentioned that the bot was not perceived as a machine. Future human-computer interaction needs to explore what are the factors that lead to this attitude and whether they benefit or harm the patients. Further research might also include qualitative in-depth interviews of the participants to evaluate what aspects of AI technology cause attrition and what functions were found especially rewarding for one's well-being.

Recommendations

Conducting this research explored how new technologies might be implemented in augmented therapy. As in Kyrgyzstan is observed a large difference between the demand for mental health services and their supply, more options for the automatization of labor for therapists are suggested, at least in the form of non-AI psychoeducational bots that might help with FAQs about therapy or registering for the session.

In general, students who participated in the experiment were scored on borderline and above the level of experiential avoidance and psychological inflexibility suggesting probably struggles with mental health. All students are part of the AUCA community and are directly affected by the way the university addresses mental health problems, thus, might be very useful in identifying how the university experience needs to be tuned to facilitate not only academically fruitful but also pleasant experience on-campus.

Based on the expert interview, TeaBot, an AI bot constructed to work with cognitive distortions, can be applied to work with youth to ensure equal access to mental health services without being limited to a certain location or device while also serving as a psychoeducation or gateway tool for those who still consider starting therapy.

References

Abd-alrazaq, A. A., Alajlani, M., Alalwan, A. A., Bewick, B. M., Gardner, P., & Househ, M. (2019). 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

Active Minds. (2020). *April 2020 Survey Data*. Retrieved from Active Minds: https://www.activeminds.org/studentsurvey/

Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. *Machine Learning with Applications*, *2*, 100006. https://doi.org/10.1016/j.mlwa.2020.100006

Albrecht, M. R., Marekova, L., Paterson, K. G., & Stepanovs, I. (2022). Four attacks and a proof for telegram. *2022 IEEE Symposium on Security and Privacy (SP)*. https://doi.org/10.1109/sp46214.2022.9833666

AvatarMachine, LLC. (n.d.). *Сабина Ai – Ваш ИИ друг*. Sabina-Ai. https://sabina-ai.com/

Beck, A. T. (1963). Thinking and depression. *Archives of General Psychiatry*, *9*(4), 324. https://doi.org/10.1001/archpsyc.1963.01720160014002

Beck, J. S. (1995). *Cognitive therapy: Basics and beyond*. New York: Guilford.

Beech, B. F. (2000). The strengths and weaknesses of cognitive behavioural approaches to treating depression and their potential for wider utilization by mental health nurses. *Journal of Psychiatric and Mental Health Nursing*, *7*(4), 343–354. https://doi.org/10.1046/j.1365-2850.2000.00298.x

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can Language Models Be Too Big?🦜. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. (pp. 610-623). https://doi.org/10.1145/3442188.3445922

Bond, F. W., Hayes, S. C., Baer, R. A., Carpenter, K. M., Guenole, N., Orcutt, H. K., Waltz, T., & Zettle, R. D. (2011). Preliminary psychometric properties of the acceptance and action questionnaire–II: A revised measure of psychological inflexibility and experiential avoidance. *Behavior Therapy*, *42*(4), 676–688. https://doi.org/10.1016/j.beth.2011.03.007

Boucher, E. M., Harake, N. R., Ward, H. E., Stoeckl, S. E., Vargas, J., Minkel, J., Parks, A. C., & Zilca, R. (2021). Artificially intelligent chatbots in Digital Mental Health Interventions: A Review. *Expert Review of Medical Devices*, *18*(sup1), 37–49. https://doi.org/10.1080/17434440.2021.2013200

Brown, L. S. (2018). Introduction: Feminist therapy—not for cisgender women only. *Feminist Therapy (2nd Ed.).*, 3–10. https://doi.org/10.1037/0000092-001

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Wu, J., Ramesh, A., Ziegler, D., M., Winter, C., … & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, *33*, 1877-1901. https://doi.org/10.48550/arXiv.2005.14165

Burns, D. D. & Beck, A.T. (1999). *Feeling Good: The New Mood Therapy*. Harper.

Clarke, V., & Braun, V. (2013). Teaching thematic analysis: Overcoming challenges and developing strategies for effective learning. *The psychologist*, *26*(2).

Cook, T. D., Campbell, D. T., & Shadish, W. (2002). *Experimental and quasi-experimental designs for generalized causal inference* (pp. 103-134). Boston, MA: Houghton Mifflin.

Corey, G., Corey, M. S., & Callanan, P. (2011). *Issues and ethics in the helping professions*. Cengage Learning.

Covin, R., Dozois, D. J., Ogniewicz, A., & Seeds, P. M. (2011). Measuring cognitive errors: Initial development of the Cognitive Distortions Scale (CDS). *International Journal of Cognitive Therapy*, *4*(3), 297–322. https://doi.org/10.1521/ijct.2011.4.3.297

Cully, J. A., & Teten, A. L. (2008). A therapist's guide to brief cognitive behavioral therapy. *Houston: Department of Veterans Affairs South Central MIRECC*

Dale, R. (2020). GPT-3: What's it good for? *Natural Language Engineering*, *27*(1), 113–118. https://doi.org/10.1017/s1351324920000601

D'Alfonso, S., Santesteban-Echarri, O., Rice, S., Wadley, G., Lederman, R., Miles, C., Gleeson, J., & Alvarez-Jimenez, M. (2017). Artificial Intelligence-Assisted Online Social Therapy for youth mental health. *Frontiers in Psychology*, *8,* 796. https://doi.org/10.3389/fpsyg.2017.00796

Damij, N., & Bhattacharya, S. (2022). The role of AI Chatbots in mental health related public services in a (post)Pandemic World: A review and future research agenda. *2022 IEEE Technology and Engineering Management Conference (TEMSCON EUROPE)*. (pp. 152-159). https://doi.org/10.1109/temsconeurope54743.2022.9801962

Dekker, I., De Jong, E. M., Schippers, M. C., 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. https://doi.org/10.3389/fpsyg.2020.01063

Depounti, I., Saukko, P., & Natale, S. (2022). Ideal Technologies, ideal women: AI and gender imaginaries in redditors' discussions on the Replika Bot Girlfriend. *Media, Culture & Society*, *45*(4), 720–736. https://doi.org/10.1177/01634437221119021

Engvall, J. (2022). Russia's War in Ukraine: Implications for Central Asia.

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). https://doi.org/10.2196/mental.7785

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, *30*(4), 681–694. https://doi.org/10.1007/s11023-020-09548-1

Fulmer, R., Joerin, A., Gentile, B., Lakerink, L., & Rauws, M. (2018). Using psychological artificial intelligence (TESS) to relieve symptoms of depression and anxiety: Randomized controlled trial. *JMIR Mental Health*, *5*(4), e9782. https://doi.org/10.2196/mental.9782

Gabrielli, S., Rizzi, S., Bassi, G., Carbone, S., Maimone, R., Marchesoni, M., & Forti, S. (2021). Engagement and effectiveness of a healthy coping intervention via chatbot for university students: Proof-of-concept study during the COVID-19 pandemic. *JMIR MHealth and UHealth, 9*(5), e27965. https://doi.org/10.2196/27965

Gorard, S. (2003). *Quantitative methods in social science research*. A&C Black.

Gratzer, D., & Goldbloom, D. (2020). Therapy and E-therapy—preparing future psychiatrists in the era of apps and Chatbots. *Academic Psychiatry*, *44*(2), 231–234. https://doi.org/10.1007/s40596-019-01170-3

Grové, C. (2021). Co-developing a mental health and wellbeing chatbot with and for young people. *Frontiers in Psychiatry*, *11*, 606041. https://doi.org/10.3389/fpsyt.2020.606041

Han, X., Zhang, Z., Ding, N., Gu, Y., Liu, X., Huo, Y., Qiu, J., Yao, Y., Zhang, A., Zhang, L., Han, W., Huang, M., Jin, Q., Lan, Y., Liu, Y., Liu, Z., Lu, Z., Qiu, X., Song, R., … Zhu, J. (2021). Pre-trained models: Past, present and future. *AI Open*, *2*, 225–250. https://doi.org/10.1016/j.aiopen.2021.08.002

Hayes, S. C., Strosahl, K., Wilson, K. G., Bissett, R. T., Pistorello, J., Toarmino, D., Polusny, M. A., Dykstra, T. A., Batten, S. V., Bergan, J., Stewart, S. H., Zvolensky, M. J., Eifert, G. H., Bond, F. W., Forsyth, J. P., Karekla, M., & McCurry, S. M. (2004). Measuring experiential avoidance: A preliminary test of a working model. *The Psychological Record*, *54*(4), 553–578. https://doi.org/10.1007/bf03395492

Inkster, B., Sarda, S., & Subramanian, V. (2018). An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: Real-World Data Evaluation

Mixed-Methods Study. *JMIR MHealth and UHealth*, *6*(11), e12106. https://doi.org/10.2196/12106

Jamankulova, A. (2022, September, 27). *Because of Russian mobilization, renting fees in Bishkek increased by 30%*. Kloop. https://kloop.kg/blog/2022/09/27/iz-za-mobilizatsii-v-rf-arenda-kvartir-v-bishkeke-podskochila-na-30/

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. https://doi.org/10.1177/1178222619829083

Kumar, H., Musabirov, I., Shi, J., Lauzon, A., Choy, K. K., Gross, O., Kulzhabayeva, D., & Williams, J. J. (2022). Exploring The Design of Prompts For Applying GPT-3 based Chatbots: A Mental Wellbeing Case Study on Mechanical Turk. *arXiv preprint arXiv:2209.11344*.

Lauber, C., & Rössler, W. (2007). Stigma towards people with mental illness in developing countries in Asia. *International Review of Psychiatry*, *19*(2), 157–178. https://doi.org/10.1080/09540260701278903

Luxton, D. D. (2016). An Introduction to Artificial Intelligence in Behavioral and Mental Health Care. *Artificial Intelligence in Behavioral and Mental Health Care*, (pp. 1–26). Academic Press. https://doi.org/10.1016/b978-0-12-420248-1.00001-5

Macleod, C. I., Bhatia, S., & Liu, W. (2020). Feminisms and decolonising psychology: Possibilities and challenges. *Feminism & Psychology*, *30*(3), 287–305. https://doi.org/10.1177/0959353520932810

McGinn, L. K., & Sanderson, W. C. (2001). What allows cognitive behavioral therapy to be brief: Overview, efficacy, and crucial factors facilitating brief treatment. *Clinical Psychology: Science and Practice*, *8*(1), 23–37. https://doi.org/10.1093/clipsy.8.1.23

Murzakulova, A., Dessalegn, M., & Phalkey, N. (2021). Examining migration governance: Evidence of rising insecurities due to covid-19 in China, Ethiopia, Kyrgyzstan, Moldova, Morocco, Nepal and Thailand. *Comparative Migration Studies*, *9*(1), 1-16. https://doi.org/10.1186/s40878-021-00254-0

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth summer research project on artificial intelligence, august 31, 1955. *AI magazine*, *27*(4), 12-14.

Naeem, F., Phiri, P., Munshi, T., Rathod, S., Ayub, M., Gobbi, M., & Kingdon, D. (2015). Using cognitive behaviour therapy with South Asian Muslims: Findings from the culturally sensitive CBT project. *International Review of Psychiatry*, *27*(3), 233-246.

Oilobot. (n.d.). Oilobot. https://oilobot.info/

OpenAI. (n.d.). *Models*. OpenAI. https://beta.openai.com/docs/models/overview.

Open Line. (2020, June 22). *В Кыргызстане появилась первая игра-сериал про ала качуу для смартфонов*. Retrieved from Open Line: https://openline.kg/new/%D0%B2-%D0%BA%D1%8B%D1%80%D0%B3%D1%8B%D0%B7%D1%81%D1%82%D0%B0%D0%BD%D0%B5-%D0%BF%D0%BE%D1%8F%D0%B2%D0%B8%D0%BB%D0%B0%D1%81%D1%8C-%D0%BF%D0%B5%D1%80%D0%B2%D0%B0%D1%8F-%D0%B8%D0%B3%D1%80%D0%B0-%D1%81/

Oracle. (n.d.). *What is Big Data?* https://www.oracle.com/big-data/what-is-big-data/

Orlova, M. (2019, February 21). *Mental Health in Kyrgyzstan. Governmental program is not executed*. 24kg. https://24.kg/obschestvo/109829_psihicheskoe_zdorove_vkyirgyizstane_pravitelstvennaya_programma_neispolnyaetsya/

Özdel, K., Taymur, I., Guriz, S. O., Tulaci, R. G., Kuru, E., & Turkcapar, M. H. (2014). Measuring cognitive errors using the cognitive distortions scale (CDS): Psychometric Properties in clinical and non-clinical samples. *PLoS ONE*, *9*(8), e105956. https://doi.org/10.1371/journal.pone.0105956

Pachankis, J. E., Soulliard, Z. A., Seager van Dyk, I., Layland, E. K., Clark, K. A., Levine, D. S., & Jackson, S. D. (2022). Training in LGBTQ-affirmative cognitive behavioral therapy: A randomized controlled trial across LGBTQ community centers. *Journal of Consulting and Clinical Psychology*, *90*(7), 582–599. https://doi.org/10.1037/ccp0000745

Pinchuk, I., Yachnik, Y., Kopchak, O., Avetisyan, K., Gasparyan, K., Ghazaryan, G., Chkonia, E., Panteleeva, L., Guerrero, A., & Skokauskas, N. (2021). The implementation of the who mental health gap intervention guide (MHGAP-IG) in Ukraine, Armenia, Georgia and Kyrgyz Republic. *International Journal of Environmental Research and Public Health*, *18*(9), 4391. https://doi.org/10.3390/ijerph18094391

Qiu, J. Y., Zhou, D. S., Liu, J., & Yuan, T. F. (2020). Mental wellness system for COVID-19. *Brain, Behavior, and Immunity*, *87*, 51–52. https://doi.org/10.1016/j.bbi.2020.04.032

Rasmussen, B. (2017). A critical examination of CBT in Clinical Social Work Practice. *Clinical Social Work Journal*, *46*(3), 165–173. https://doi.org/10.1007/s10615-017-0632-7

Reardon, M. L., Cukrowicz, K. C., Reeves, M. D., & Joiner, T. E. (2002). Duration and regularity of therapy attendance as predictors of treatment outcome in an adult outpatient population. *Psychotherapy Research*, *12*(3), 273–285. https://doi.org/10.1080/713664390

Romanovskyi, O., Pidbutska, N., & Knysh, A. (2021). Elomia Chatbot: The Effectiveness of Artificial Intelligence in the Fight for Mental Health. In *COLINS* (pp. 1215-1224).

Samuel, A. L. (2000). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 44, 206-226. https://doi.org/10.1147/rd.441.0206

Sankar, A., Panchal, P., Goldman, D. A., Colic, L., Villa, L. M., Kim, J. A., Lebowitz, E. R., Carrubba, E., Lecza, B., Silverman, W. K., Swartz, H. A., & Blumberg, H. P. (2021). Telehealth social rhythm therapy to reduce mood symptoms and suicide risk among adolescents and young adults with bipolar disorder. *American Journal of Psychotherapy*, *74*(4), 172–177. https://doi.org/10.1176/appi.psychotherapy.20210011

Shields-Zeeman, L., Collin, D. F., Batra, A., & Hamad, R. (2021). How does income affect mental health and health behaviours? A quasi-experimental study of the earned Income Tax Credit. *Journal of Epidemiology and Community Health*, *75*(10), 929–935. https://doi.org/10.1136/jech-2020-214841

Sulaiman, S., Mansor, M., Abdul Wahid, R., & Nor Azhar, N. A. (2022). Anxiety assistance mobile apps chatbot using cognitive behavioural therapy. *International Journal of Artificial Intelligence*, *9*(1), 17–23. https://doi.org/10.36079/lamintang.ijai-0901.349

Torous, J., Jän Myrick, K., Rauseo-Ricupero, N., & Firth, J. (2020). Digital Mental Health and COVID-19: Using technology today to accelerate the curve on access and quality Tomorrow. *JMIR Mental Health*, *7*(3), e18848. https://doi.org/10.2196/18848

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, *59*(236), 433-460.

UNICEF. (2020). *Анализ ситуации в области суицила и суицидальных попыток среди подростков и молодёжи в Кыргызстане.* Бишкек.

Vilaza, G. N., & McCashin, D. (2021). Is the Automation of Digital Mental Health Ethical? Applying an Ethical Framework to Chatbots for Cognitive Behaviour Therapy. *Frontiers in Digital Health*, *3*. https://doi.org/10.3389/fdgth.2021.689736

Watts, S., Mackenzie, A., Thomas, C., Griskaitis, A., Mewton, L., Williams, A., & Andrews, G. (2013). CBT for depression: A pilot RCT comparing mobile phone vs. computer. *BMC Psychiatry*, *13*(1), 1-9. https://doi.org/10.1186/1471-244x-13-49

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.

Willig, C. (2008). *Introducing Qualitative Research in Psychology: Adventures in Theory and Method.* Open University Press.

World Health Organization. (2022, March 2). *COVID-19 pandemic triggers 25% increase in prevalence of anxiety and depression worldwide*. WHO: https://www.who.int/news/item/02-03-2022-covid-19-pandemic-triggers-25-increase-in-prevalence-of-anxiety-and-depression-worldwide

Xiang, C. (2023, January 10). *Startup Uses AI Chatbot to Provide Mental Health Counseling and Then Realizes It 'Feels Weird'*. Vice: https://www.vice.com/en/article/4ax9yw/startup-uses-ai-chatbot-to-provide-mental-health-counseling-and-then-realizes-it-feels-weird

Yang, M. (2020). Painful conversations: Therapeutic chatbots and public capacities. *Communication and the Public*, *5*(1-2), 35–44. https://doi.org/10.1177/2057047320950636

Zeavin, H., & Peters, J. D. (2021). *The Distance Cure: A History of Teletherapy*. The MIT Press.

Zhang, M., & Li, J. (2021). A commentary of GPT-3 in MIT Technology Review 2021. *Fundamental Research*, *1*(6), 831–833. https://doi.org/10.1016/j.fmre.2021.11.011

Appendix A

List of Tables

Table 1

Participant Demographics

		Control	Experimental
Gender, n (%)	Female	27 (79,41)	27 (79,41)
	Male	7 (20,59)	7 (20,59)
Nationality, n (%)	Kyrgyz	25(73,53)	27 (72,41)
	Non-Kyrgyz	9 (26,47)	7 (20,59)
Year of study, n (%)	Freshman	2 (5,88)	2 (5,88)
	Sophomore	11 (32,35)	5 (14,71)
	Junior	7 (20,59)	11(32,35)
	Senior	7 (20,59)	11 (32,35)
	Alumni or Graduate	7 (20,59)	5 (14,71)
Division, n (%)	Applied Sciences	10 (29,41)	9 (26,47)
	Social Sciences	12 (35,29)	10 (29,41)
	Humanities	7 (20,59)	6 (17,65)
	Business and Law	5 (14,71)	9 (26,47)
Previous experience with bots	Yes	7 (20,59)	5 (14,71)
	No	27 (79,41)	29 (85,29)

Table 2

User Engagement According to Time Periods

Time Period	Number of Messages Sent
01:00 – 09:00	317
09:00 – 17:00	2593
17:00 – 01:00	2350

Table 3

Pretest Differences

	Control [a]	Experimental [a]	Normality [b]	p-value	Cohen’s d
AAQ-II	28.15 (7.82)	30.85 (9.42)	0.19	0.20	0.31
CDS	41.65 (12.40)	44.59 (10.98)	0.46	0.30	0.25
Total	69.79 (18.27)	75.44 (18.75)	0.30	0.213	0.31

Note: [a] mean (standard deviation)

[b] Shapiro-Wilk test

Table 4

Decrease of Scores in Pretest vs. Posttest Results among 2 Groups

	Control [a]	Experimental [a]	Normality [b]	p-value	Cohen's d
AAQ-II	0.56 (7.78)	6.91 (8.00)	0.68	0.001	0.81
CDS	2.27 (9.85)	2.27 (10.78)	0.96	0.002	0.8.
Total	17.44 (16.67)	2.82 (15.65)	0.57	< .001	0.90

Note: [a] mean (standard deviation)

[b] Shapiro-Wilk test

Table 5

Correlations Table

Variable		**dif [a]**	
1. dif [a]	n	—	
	Pearson's r	—	
	p-value	—	
2. message_part [b]	n	34	
	Pearson's r	-0.005	
	p-value	0.976	
3. week_part [c]	n	34	
	Pearson's r	0.362	*
	p-value	0.035	

* p < .05

Note: [a] difference between pretest and posttest scores

[b] total number of messages sent to Teabot

[c] total number of weeks interacting with TeaBot

Table 6

Analysis of Differences in Pretest vs. Posttest Results in Experimental Group by Each Item

	Pretest [a]	Posttest [a]	Normality [b]	p-value	Cohen's d
1	3.82 (1.90)	3.29 (1.70)	0.035	0.064 [c]	0.16
2	4.03 (2.04)	3.35 (1.77)	0.427	0.046	0.18
3	4.77 (1.91)	3.85 (1.64)	0.038	0.013 [c]	0.20
4	4.21 (2.14)	2.94 (1.72)	0.296	< .001	0.18
5	4.32 (1.74)	3.44 (1.94)	0.061	0.007	0.18

6	4.74 (1.93)	3.32 (1.97)	0.025	< .001[c]	0.16
7	4.97 (1.57)	3.74 (1.76)	0.044	0.001 [c]	0.21
8	5.00 (1.78)	3.77 (1.76)	0.205	0.002	0.22
9	4.71 (1.70)	3.82 (1.40)	0.148	0.009	0.22
10	4.24 (1.88)	3.12 (1.53)	0.107	0.002	0.21
11	4.77 (1.56)	3.53 (1.66)	0.026	< .001 [c]	0.19
12	4.59 (1.65)	3.27 (1.54)	0.069	< .001	0.20
13	4.09 (1.66)	3.12 (1.57)	0.074	0.001	0.18
14	4.09 (1.99)	3.18 (1.73)	0.277	0.021	0.21
15	4.44 (1.96)	3.06 (1.50)	0.111	< .001	0.22
16	5.06 (1.72)	3.85 (1.71)	0.120	< .001	0.20
17	3.62 (1.84)	3.35 (2.00)	0.312	0.374	0.15

Note: [a] mean (standard deviation)

[b] Shapiro-Wilk test

[c] Wilcoxon signed rank test was applied because of deviation from normality ($p < 0.05$)

Table 7

Repeated Measures ANOVA

		95% CI for Mean Difference		
Distortions	**Marginal Mean**	**Lower**	**Upper**	**SE**
Mindreading	1.24	0.609	1.861	0.317
Catastrophizing	0.88	0.256	1.508	0.317
All-or-Nothing	1.12	0.492	1.744	0.317
Emotional Reasoning	1.24	0.609	1.861	0.317
Labeling	1.32	0.698	1.949	0.317
Mental Filter	0.97	0.345	1.597	0.317
Overgeneralization	0.91	0.286	1.538	0.317
Personalization	1.38	0.756	2.008	0.317
Should Statements	1.21	0.580	1.832	0.317
Minimizing	0.27	-0.361	0.891	0.317

Table 8

Exploratory Factor Analysis: Chi-squared Test

	Value	df	p
Model	40.899	61	0.98

Table 9

Exploratory Factor Analysis: Principal Component Analysis

	Eigenvalue	Proportion var.	Cumulative
Component 1	5.32	0.313	0.313
Component 2	2.73	0.161	0.474
Component 3	1.53	0.090	0.564
Component 4	1.23	0.072	0.636
Component 5	1.11	0.065	0.701

Table 10

Exploratory Factor Analysis: Factor Loadings

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Uniqueness
Q9[a]	0.928		-0.415			0.226
Q11	0.833					0.342
Q7	0.713					0.419
Q17	0.639					0.469
Q8	0.583					0.394
Q10	0.484					0.643
Q4		0.752				0.394
Q1		0.710				0.436
Q5		0.656				0.316
Q2		0.617				0.534
Q16		-0.475				0.596
Q15			0.692			0.137
Q12			0.657			0.549
Q6			0.645			0.407
Q13			0.433			0.819
Q3				0.996		0.005
Q14					0.923	0.005

Note. Applied rotation method is promax.

[a] Q1-Q7 refer to AAQ-II items, Q8-Q17 refer to CDS items

Table 11

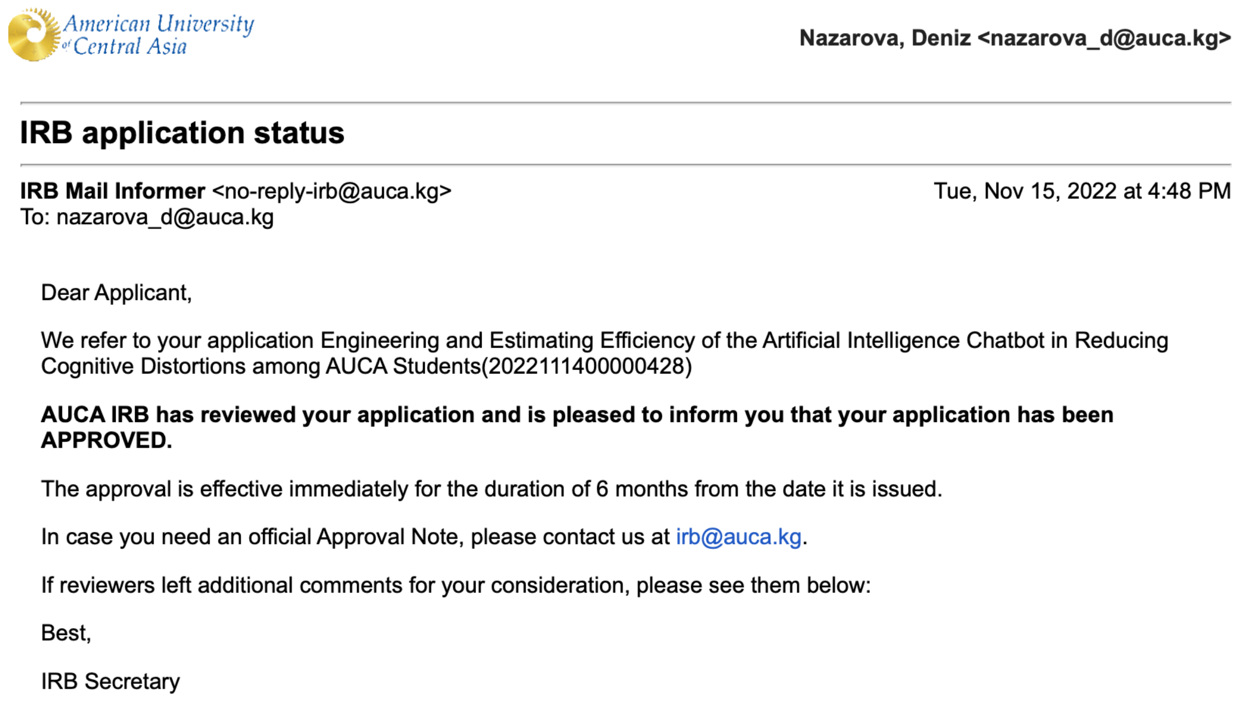
Thematic Analysis Practicing Specialists' Reactions on TeaBot

Theme	Subtheme	Translation of exemplary quote	Exemplary quote
Application	Accuracy (5)	[TeaBot] guesses correctly and asks a very good question to begin with	«Правильно угадывает и задаёт для начала очень хороший вопрос»
	Medium (6)	Bot in telegram is much more convenient than the papers I distribute due to immediate feedback	«Бот в телеграме намного удобнее чем листочки что я раздаю, потому что он сразу даёт же свой фидбек»
	Audience (5)	Having such a bot makes this [psychotherapeutic] process, especially for young people, a lot easier	«Наличие такого бота этот [психорапевтический] процесс, особенно для молодых людей, сильно упрощает»
In-therapy tool	Homework (5)	I think it was convenient, for example, if the client will fill in [homework], and he will get these answers immediately	«Я думаю это удобно было, например, если клиент будет заполнять [домашнее задание], и ему будут выходить сразу вот эти ответы»
	Diary (3)	As a therapist, I see how convenient it is to use as a diary	«Как терапевт я вижу как удобно использовать в качестве дневника»
	Assessments (2)	Can be used for assessments	«Для проведения тестирования можно было использовать»
	Delegation (4)	[I see the bot] as a follow-up and additional support	«[бота вижу] как фоллоу-ап и как больше поддерживающее»
	Immergence (3)	Additional resource to be in constant contact with the patient	«Дополнительный ресурс чтобы находиться в постоянном контакте с пациентом»
	Post-therapy (3)	Once therapy finished, and it's left as for yourself, as a self-reflection, in moments when you don't know how you feel	«Вот терапию закончили например да, и это оставить как для себя, как каждый раз как саморефлексия, в моменты, когда не знаешь, как чувствуешь»
Out-therapy tool	Psychoeducation (2)	The examples are very good in terms of... what are cognitive distortions	«Примеры очень хорошие в плане… что такое cognitive distortions»
	Gateway (2)	Going to a counselor and telling a person is sometimes more difficult than telling a bot	«Идти к психолого и рассказать это человеку сложнее иногда чем боту»
	Emergency (4)	You can write it at any time... for example, there is no psychologist at this time	«Можно в любое время писать… например, нет в это время психолога»
Recommendations	Mishits (3)	The client may misinterpret the questions.	«Клиент может неправильно интерпретировать вопросы»
	Explanation for clients (6)	"If only there had been a rationale... so they would have had a field of information to draw from."	«Если б там ещё была ликбез такой небольшой… чтобы у них было информационное поле от которого можно отталкиваться»
	Education for therapists (2)	There needs to be a trained team that will really use this [bot], and	«Должен быть trained team, который реально это будет

		they will really understand what kind of content this bot has.	использовать, и он реально будет понимать какой контент у этого бота»
	Statistics (3)	So that at the end [bot] gives a summary	«Чтобы в конце [бот] даёт саммари»

Appendix B

IRB Approval: Experimental Study



American University of Central Asia

Nazarova, Deniz <nazarova_d@auca.kg>

IRB application status

IRB Mail Informer <no-reply-irb@auca.kg>
To: nazarova_d@auca.kg

Tue, Nov 15, 2022 at 4:48 PM

Dear Applicant,

We refer to your application Engineering and Estimating Efficiency of the Artificial Intelligence Chatbot in Reducing Cognitive Distortions among AUCA Students(2022111400000428)

AUCA IRB has reviewed your application and is pleased to inform you that your application has been APPROVED.

The approval is effective immediately for the duration of 6 months from the date it is issued.

In case you need an official Approval Note, please contact us at irb@auca.kg.

If reviewers left additional comments for your consideration, please see them below:

Best,

IRB Secretary

Appendix C

Informed Consent Form for Participating in Experiment

Bot:

PLEASE READ THIS DOCUMENT CAREFULLY. YOUR CONFIRMATION IS REQUIRED FOR PARTICIPATION. YOU MUST BE AT LEAST 18 YEARS OF AGE TO GIVE YOUR CONSENT TO PARTICIPATE IN RESEARCH. IF YOU DESIRE A COPY OF THIS CONSENT FORM, YOU MAY REQUEST ONE AND WE WILL PROVIDE.

You are now interacting with TeaBot.

TeaBot is an Artificial Intelligence chatbot that is developed to test whether AI chatbots can be used in combination with therapy. It applies some of the Cognitive Behavioral Therapy techniques such as identifying cognitive distortions and Socratic questioning. **TeaBot is not designed to substitute therapy, in fact, it is suggested to use it as a tool in addition to your therapy**. Please, also keep in mind that TeaBot is currently in the beta stage and might not suffice your needs. If you need a professional mental health help, contact AUCA Counseling Service via email: cs@auca.kg

Content of *chat* function will not be recorded. Only *assessment* data will be recorded, and it will be stored by the Principal Investigator in the folder protected by a password. For analysis, all participants' names' will be anonymized, and their data will be presented via IDs generated for this research. Your personal information will not be accessible to the public.

You will be compensated (500 soms) for participating in this experiment.
To receive the compensation, you MUST:

1. **COMPLETE** questionnaire by using '*assessment'* function on these dates **ONLY**:
- December, 12th
- December, 15th
- February, 6th

2. (*if experimental one*): Talk to TeaBot at least twice a week.

After the submission of the third assessment, please contact the principal researcher Deniz Nazarova via e-mail [X] and attach the screenshots of messages from bot certifying completion of all THREE assessments on the mentioned above dates. The principal researcher Deniz Nazarova will transfer money to the phone number you will indicate in the reply-email.

If you decide now or at any point to withdraw this consent or stop participating, you are free to do so. In this case, you will not receive compensation.

Any technical questions about this research may be directed to:
Principal investigator: Deniz Nazarova Contacts: tel.: [X]; email: [X]

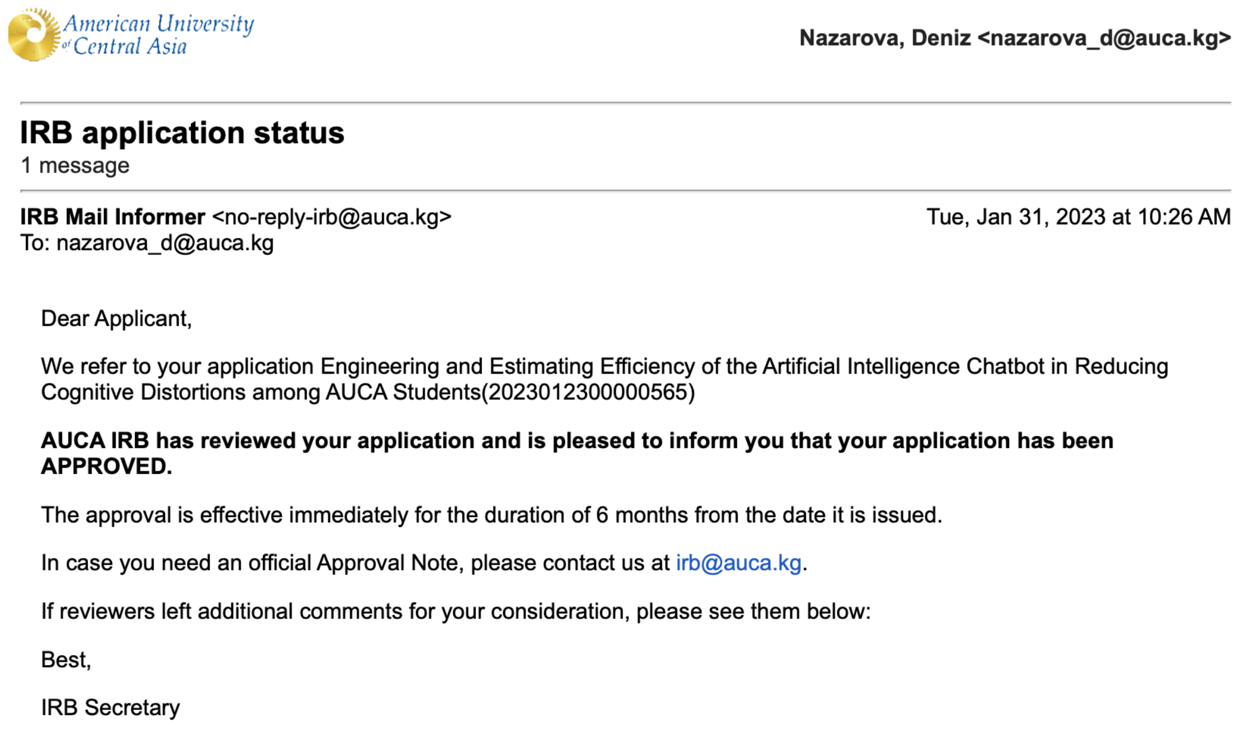
Any questions or concerns regarding the ethics of the study can be voiced to AUCA Institutional Review Board (IRB) at irb@auca.kg.

If you decide now or at any point to withdraw this consent or stop participating, you are free to do so. In this case, you will not receive compensation.

If you are older than 18 years old and agree on these terms, please, click on 'agree'. You can access them again by using '*terms*' function.

Appendix D

IRB Approval: Expert interview



Nazarova, Deniz <nazarova_d@auca.kg>

IRB application status

1 message

IRB Mail Informer <no-reply-irb@auca.kg> Tue, Jan 31, 2023 at 10:26 AM
To: nazarova_d@auca.kg

Dear Applicant,

We refer to your application Engineering and Estimating Efficiency of the Artificial Intelligence Chatbot in Reducing Cognitive Distortions among AUCA Students(2023012300000565)

AUCA IRB has reviewed your application and is pleased to inform you that your application has been APPROVED.

The approval is effective immediately for the duration of 6 months from the date it is issued.

In case you need an official Approval Note, please contact us at irb@auca.kg.

If reviewers left additional comments for your consideration, please see them below:

Best,

IRB Secretary

Appendix E

Informed Consent Form for Expert Interview

Study Title: Engineering and Estimating Efficiency of the Artificial Intelligence Chatbot in Reducing Cognitive Distortions among AUCA students
Principal investigator: Deniz Nazarova

PLEASE READ THIS DOCUMENT CAREFULLY. YOUR SIGNATURE IS REQUIRED FOR PARTICIPATION. YOU MUST BE AT LEAST 18 YEARS OF AGE TO GIVE YOUR CONSENT TO PARTICIPATE IN RESEARCH. IF YOU DESIRE A COPY OF THIS CONSENT FORM, YOU MAY REQUEST ONE AND WE WILL PROVIDE.

Please be informed that participation in the research is voluntary, and you have the right to withdraw at any time, without prejudice, should you object to the nature of the research.
You are entitled to ask questions and to receive an explanation after your participation.

Purpose of the study:
The investigator is interested to explore whether artificial intelligence might be an efficient tool for mitigating mental distress among students. The purpose of this study is to fill the gap in this area and contribute to further research.

Possible Risks:
There are no direct risks for participating in this study.
The Principal Investigator will make every effort to protect your confidentiality and your anonymity.

Possible Benefits:
The benefits of participating in this project include contributing to the academic research on the efficiency of the AI applications as mental health tools that might affect the quality of future programs. You will also have an opportunity to check the technology yourself and explore the ways how you can implement it in your practice.

Compensation for your time:
You will receive no compensation for participation in this research.

Confidentiality:
The interview will be recorded, and the recorded data will be stored by the Principal Investigator in the folder protected by a password. Principal Investigator will be the only person with access for your data. Your personal information will not be accessible to the public.

Opportunities to Question:
Any technical questions about this research may be directed to:
Principal investigator: Deniz Nazarova
Contacts: tel.: [X]; email: [X]

Any questions or concerns regarding the ethics of the study can be voiced to AUCA Institutional Review Board (IRB) at irb@auca.kg.

Interviewee: ____________________________
(Full name)
Signed: _________________________________ Date: ____/____/2023
(Interviewee's Signature) (Mo./Day/Year)

Appendix F

Acceptance and Action Questionnaire (AAQ-II)

Bot: First, I assess how you perceive your thoughts. Please mark how true is the statement below by selecting a number next to it. Scale is from '1' being 'never true' to '7' meaning 'always true'

1. My painful experiences and memories make it difficult for me to live a life that I would value.
2. I am afraid of my feelings.
3. I worry about not being able to control my worries and feelings.
4. My painful memories prevent me from having a fulfilling life.
5. Emotions cause problems in my life.
6. It seems like most people are handling their lives better than I am.
7. Worries get in a way of my success.

Appendix G

Cognitive Distortions Scale (CDS)

Bot: Second, we will examine type of thoughts that you have. Below, you will read about 10 types of thinking. Your task is to estimate how often you use that type of thinking.

Scale is from '1' being 'never true' to '7' meaning 'always true'

MINDREADING

When people assume that others are thinking negatively about them even though the other person has not said anything negative. For example, Meerim is having coffee with her boyfriend Sultan. Sultan is quiet, and Meerim asks if anything is wrong. Sultan replies that he is ‘Okay.’ Meerim does not believe Sultan. She starts to think that he is unhappy with her.tively about them even though the other person has not said anything negative. For example, Meerim is having coffee with her boyfriend Sultan. Sultan is quiet, and Meerim asks if anything is wrong. Sultan replies that he is ‘Okay.’ Meerim does not believe Sultan. She starts to think that he is unhappy with her.

1. Please estimate how often you engage in Mindreading:

CATASTROPHIZING

When people make negative predictions about the future, and there isn’t much evidence for these predictions. For example, Rustam is in his first year of university. He just received a 70 on his Math exam. He immediately starts to worry that he will end up with a low grade in the course, and that he’ll have a toughtime getting into graduate school.

2. Please estimate how often you engage in Catastrophizing:

ALL-OR-NOTHING THINKING

When people view things as being “either-or”: something is either good or bad. For example, Alan gets a B+ on an exam. He is disappointed because it was not an A. He tends to view success on exams as follows: ‘I either do great, or my performance is a failure.'

3. Please estimate how often you use All-or-Nothing Thinking:

EMOTIONAL REASONING

When people believe something to be true because it “feels” that way. For example, Sonya’s friends told her that she could not come to the concert with them because they were unable to get enough tickets for everyone. Sonya knows they probably didn’t exclude her on purpose, but she feels rejected. Therefore, part of her believes she was rejected.

4. Please estimate how often you engage in Emotional Reasoning:

LABELING

When people label themselves as being a certain kind of person after something bad happens. For example, while at a social event, Meder asks a woman if she would like to dance. She turns him down. As a result, Meder considers himself to be a loser.

5. Please estimate how often you engage in Labeling:

MENTAL FILTER

When there is positive and negative information, but people only focus on the negative information. For example, Erlan's teacher wrote ‘Erlan, you have an excellent way of expressing ideas. I really enjoy the way you write. However, you should try and make better transitions from one idea to another.’ Despite the fact that Erlan clearly performed well, he could only think about the one piece of criticism, and felt poorly about himself.

6. Please estimate how often you engage in Mental Filtering:

OVERGENERALIZATION

When a negative event occurs, people might assume more bad things are going to happen. They see the negative event as the start of a pattern. For example, Ahmad recently failed his math exam. He thinks to himself: ‘I’ll probably fail the exams in my other courses as well.’

7. Please estimate how often you engage in Overgeneralization:

PERSONALIZATION

When people believe they are responsible for negative things, even though they’re not. In other words, they take a negative event, and assume they are the cause of it. For example, Alisher’s best friend has been in a bad mood lately, and it has been hard to get in contact with him. Alisher assumes that he must have personally done something wrong to make his friend act this way.

8. Please estimate how often you engage in Personalization:

SHOULD STATEMENTS

When people think that things should or must be a certain way.

For example, Adilet is upset with getting an 85 on his exam because he thinks he should get at least a 90. He often has these thoughts for many things (e.g., he feels he should never drop a pass when playing football; his room should be organized a certain way).

9. Please estimate how often you tend to make Should Statements:

MINIMIZING OR DISQUALIFYING THE POSITIVE

When people ignore the positive things that happen to them.

For example, Malika works got A for her final project. Her professor recently told her that she did a wonderful job. In her head, she dismisses her achievement because she probably ‘just got lucky.’

10. Please estimate how often you tend to Minimize or Disqualify the Positive:

Appendix H

Expert Interview Protocol

1. How would you rate your experience using TeaBot?
   Как бы Вы оценили свой опыт использования TeaBot?
   - ○ 0
   - ○ 1
   - ○ 2
   - ○ 3
   - ○ 4

2. Why did you choose this grade?
   Почему Вы выбрали именно эту оценку?

3. How likely are you to use such technologies when conducting therapy for youth?
   Насколько возможно то, что Вы будете использовать подобные технологии в проводимой Вами терапии для молодёжи?
   - ○ 0
   - ○ 1
   - ○ 2
   - ○ 3
   - ○ 4

4. Why did you choose this grade?
   Почему Вы выбрали данную оценку?

5. How do you envision application of this bot in psychotherapy?
   Как Вы видите использование этого бота в психотерапии?

6. What are the problems that a psychologist might face when implementing such technologies in their practice?
   С камими трудностями может столкнуться психолог_иня во время добавления подобных технологий в свою практику?

7. How do you think these problems can be mitigated?
   Как эти проблемы могут быть решены?

8. What concepts might not be clear to the patient when they are using the chatbot?
   Какие концепты могут быть непонятны пациент_ке во время пользования ботом?

9. Any additional comments or feedback?
   Вы бы хотели ещё что-нибудь добавить к своим комментариям?

Appendix I

Manual

THERAPEUTIC TECHNIQUES USED

What therapy is TeaBot using?

Cognitive Behavioral Therapy (CBT) is a type of therapy that is based on the principle that one's mental well-being can be changed by learning to cope with negative thoughts and behaviors. Some of the techniques that this therapy uses and TeaBot applies are recognizing and challenging one's *cognitive distortions*.

What are cognitive distortions?

We all practice errors in reasoning from time to time. However, they become problematic when they **are not based on valid evidence** and become **persistent**.

Such thoughts are called *cognitive distortions*, and there are 10 types of them:

- **Mindreading** is assuming that others are thinking negatively about them even though the other person has not said anything negative, e. g., s*he thinks I'm dull because she didn't like my post.*
- **Catastrophizing** is making negative predictions about the future, and there isn't much evidence for these predictions, e. g., *if I don't get A for this class, my career is over.*
- **All-or-Nothing Thinking** is viewing things as being "either-or": something is either good or bad, e. g., *if this essay is not good then it's not worth doing.*
- **Emotional reasoning** is believing something to be true because it "feels" that way, e. g., *I used to study very hard but I feel like I'm not competent enough.*
- **Labeling** is naming oneself as being a certain kind of person after something bad happens, e. g., *I received a B for this course, I must a loser.*
- **Mental filter** is focusing only on negative information although both positive and negative ones were presented, e. g., *because I got one low rating on my evaluation it means I'm doing a lousy job.*
- **Overgeneralization** is assuming that when a negative event occurs, more bad things are going to happen. It's seeing one negative event as the start of a pattern, e. g., *I was cheated on once, and every person I am with is going to cheat on me.*
- **Personalization** is assuming own responsibility in a negative event, e. g., *no one wrote anything in the group chat, I must have said something stupid.*
- **Should statements** is thinking that things should or must be a certain way, e. g., *I should be a better son, so my dad approves me.*
- ● **Minimizing the positive** is ignoring the positive things that happen to oneself, e. g., *I am not enough to deserve this job.*

How to mitigate them?

Before mitigating them, it is important to recognize them first. Once you know that you are facing a cognitive distortion you can apply Socratic questioning, i.e., asking open-ended questions that provoke reflection on the thought.

Resources used:

Beck, J. S. (1995). *Cognitive therapy: Basics and beyond*. New York: Guilford.

CONDITIONS OF PARTICIPATION

You will be compensated (500 soms) for participating in this experiment. To receive the compensation, you MUST:

1. **COMPLETE** assessments by using ‘assessment’ function on these dates **ONLY**:
    - December, 2nd
    - December, 5th
    - January, 30th
2. **READ** the *'manual'* function **BEFORE** starting using Teabot's 'let's chat’ function.
3. **TALK** to TeaBot at least 2 days a week. This means that you will need to initiate the conversation with TeaBot 2 times per week and maintain the conversation for at least 5 minutes sharing your thoughts when you find them negative.
   Your interaction **is checked without access to the content of the messages**.
   After the submission of the third assessment, contact the principal researcher Deniz Nazarova via e-mail nazarova_d@auca.kg and attach the screenshots of messages from bot certifying completion of all **THREE** assessments on the **DATES** mentioned above. The principal researcher will transfer money to the phone number you will indicate in the reply-email.

ETHICS AND SECURITY

Content of ‘let's chat’ function will not be recorded. Only assessment data will be recorded, and it will be stored by the Principal Investigator in a folder protected by a password. For analysis, all participants' names’ will be anonymized, and their data will be presented via IDs. Your personal information will not be accessible to the public.

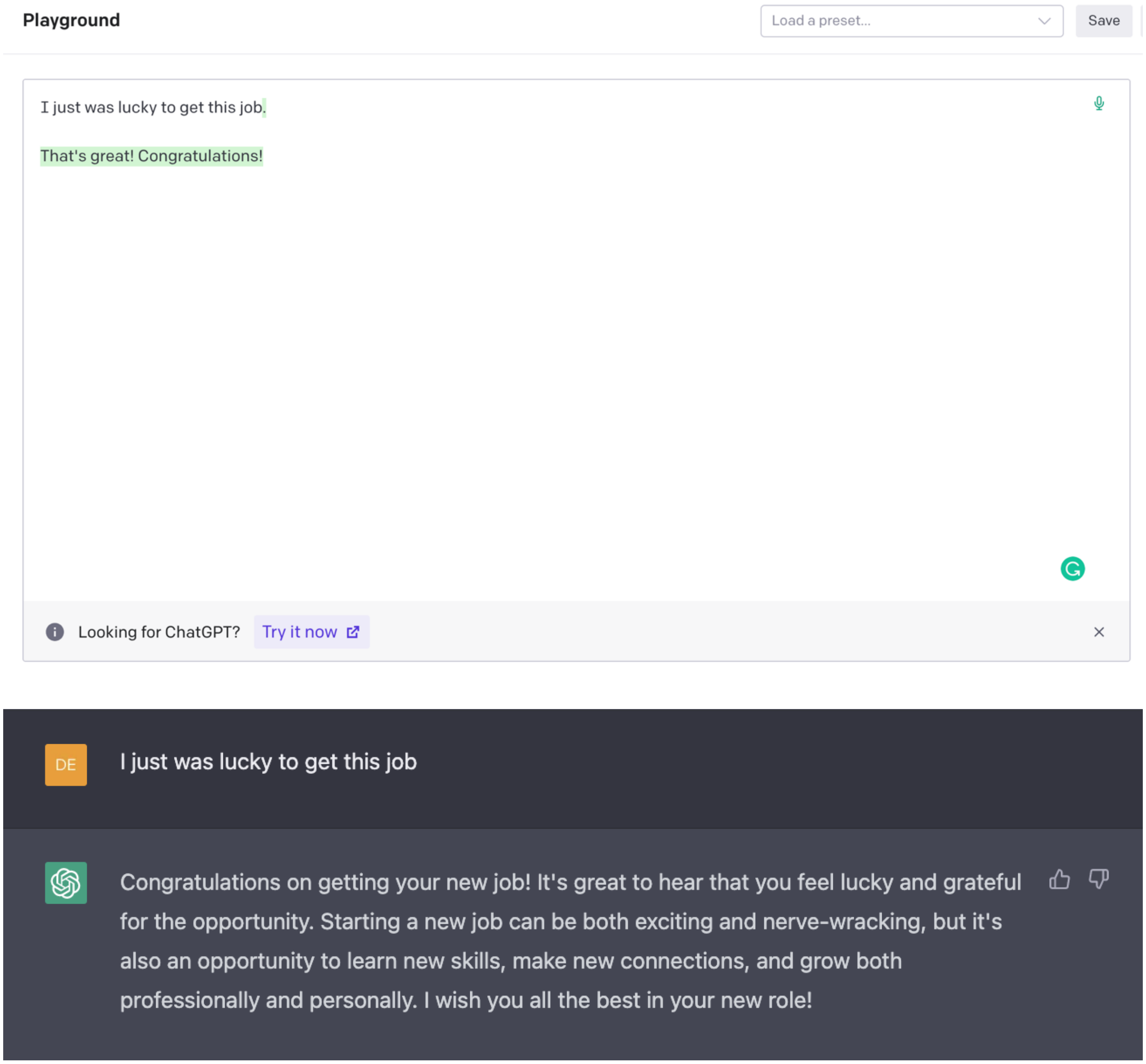
Any questions or concerns regarding the ethics of the study can be voiced to AUCA Institutional Review Board (IRB) at irb@auca.kg.

Appendix J

List of Figures

Figure 1

Two Images that Illustrate how GPT-3 (above) and ChatGPT (below) react to a distorted thought.



I just was lucky to get this job.

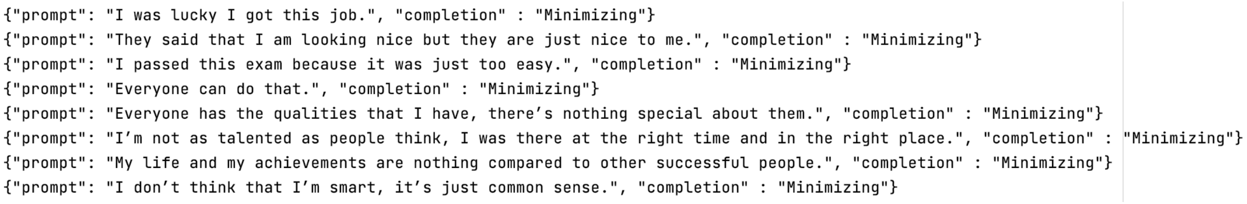
That's great! Congratulations!

I just was lucky to get this job

Congratulations on getting your new job! It's great to hear that you feel lucky and grateful for the opportunity. Starting a new job can be both exciting and nerve-wracking, but it's also an opportunity to learn new skills, make new connections, and grow both professionally and personally. I wish you all the best in your new role!

Figure 2

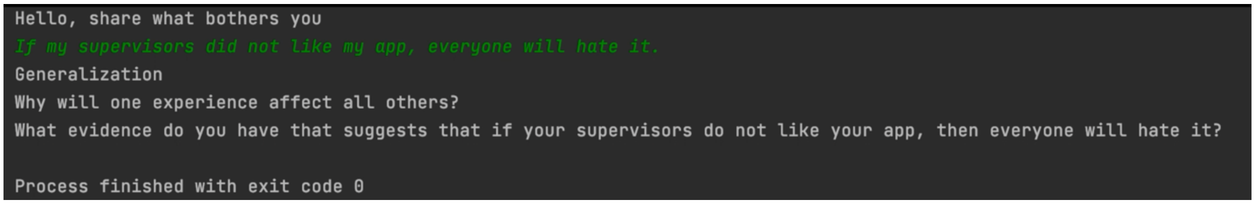
Excerpts from TeaBot's Training Set



```
{"prompt": "I was lucky I got this job.", "completion" : "Minimizing"}
{"prompt": "They said that I am looking nice but they are just nice to me.", "completion" : "Minimizing"}
{"prompt": "I passed this exam because it was just too easy.", "completion" : "Minimizing"}
{"prompt": "Everyone can do that.", "completion" : "Minimizing"}
{"prompt": "Everyone has the qualities that I have, there’s nothing special about them.", "completion" : "Minimizing"}
{"prompt": "I’m not as talented as people think, I was there at the right time and in the right place.", "completion" : "Minimizing"}
{"prompt": "My life and my achievements are nothing compared to other successful people.", "completion" : "Minimizing"}
{"prompt": "I don’t think that I’m smart, it’s just common sense.", "completion" : "Minimizing"}
```

Figure 3

Fine-tuned and Modified GPT-3's Response in Terminal



```
Hello, share what bothers you
If my supervisors did not like my app, everyone will hate it.
Generalization
Why will one experience affect all others?
What evidence do you have that suggests that if your supervisors do not like your app, then everyone will hate it?

Process finished with exit code 0
```

Figure 4

TeaBot's scheme of work

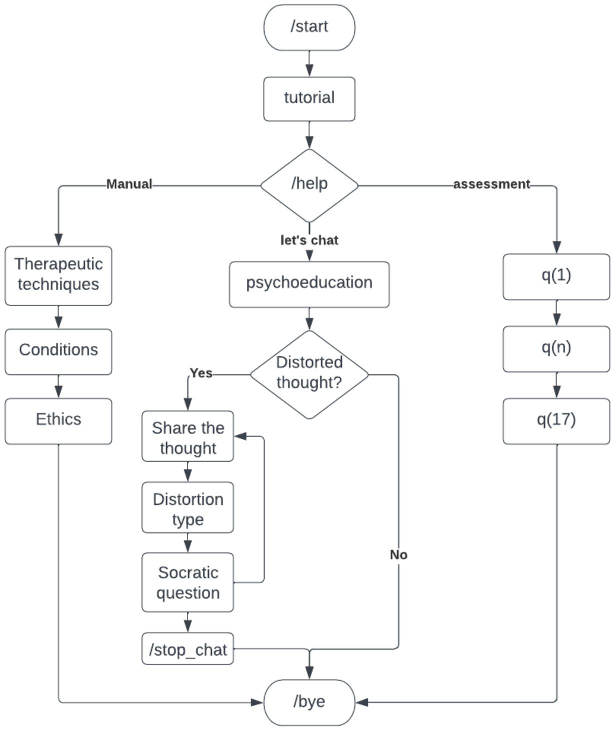
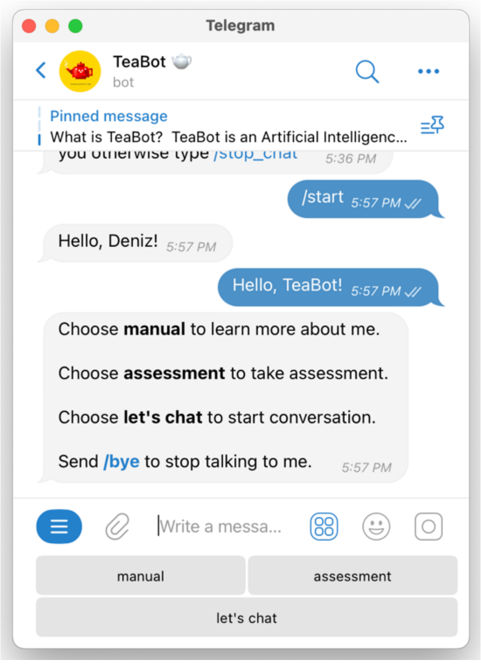


Figure 5

TeaBot's Interface



What is TeaBot? TeaBot is an Artificial Intelligenc...

you otherwise type /stop_chat 5:36 PM

/start 5:57 PM

Hello, Deniz! 5:57 PM

Hello, TeaBot! 5:57 PM

Choose **manual** to learn more about me.

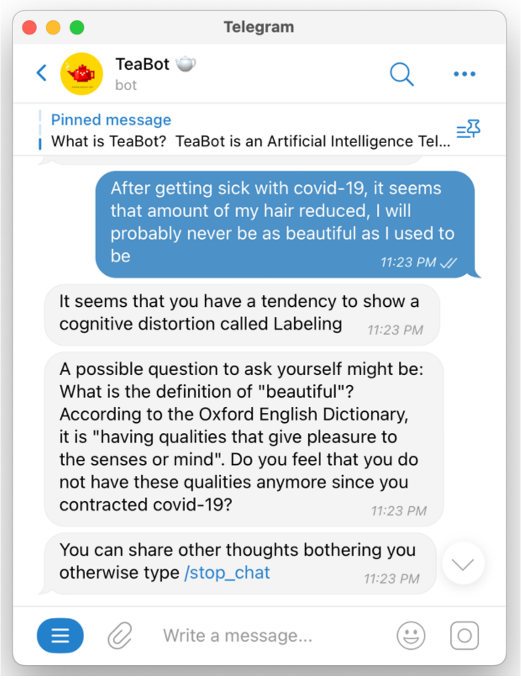
Choose **assessment** to take assessment.

Choose **let's chat** to start conversation.

Send /bye to stop talking to me. 5:57 PM

Figure 6

TeaBot's Let's Chat Function



What is TeaBot? TeaBot is an Artificial Intelligence Tel...

After getting sick with covid-19, it seems that amount of my hair reduced, I will probably never be as beautiful as I used to be 11:23 PM

It seems that you have a tendency to show a cognitive distortion called Labeling 11:23 PM

A possible question to ask yourself might be: What is the definition of "beautiful"? According to the Oxford English Dictionary, it is "having qualities that give pleasure to the senses or mind". Do you feel that you do not have these qualities anymore since you contracted covid-19? 11:23 PM

You can share other thoughts bothering you otherwise type /stop_chat 11:23 PM

Figure 7

Participants' Flow

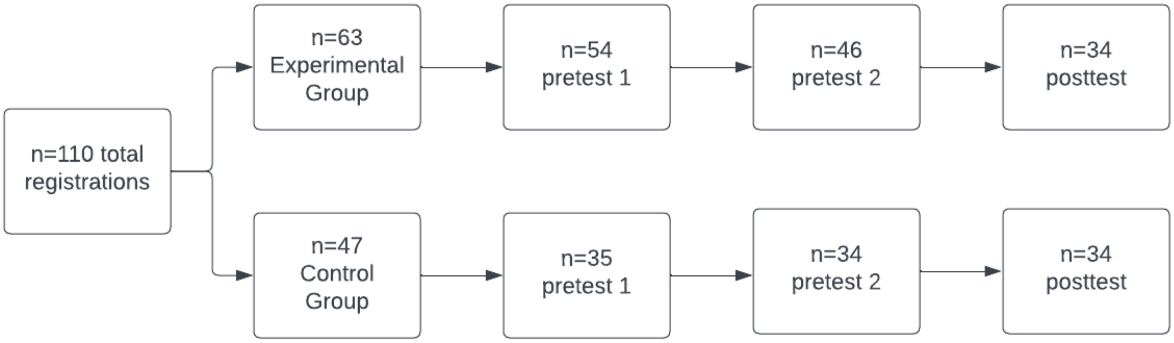


Figure 8

User Engagement According to Hours

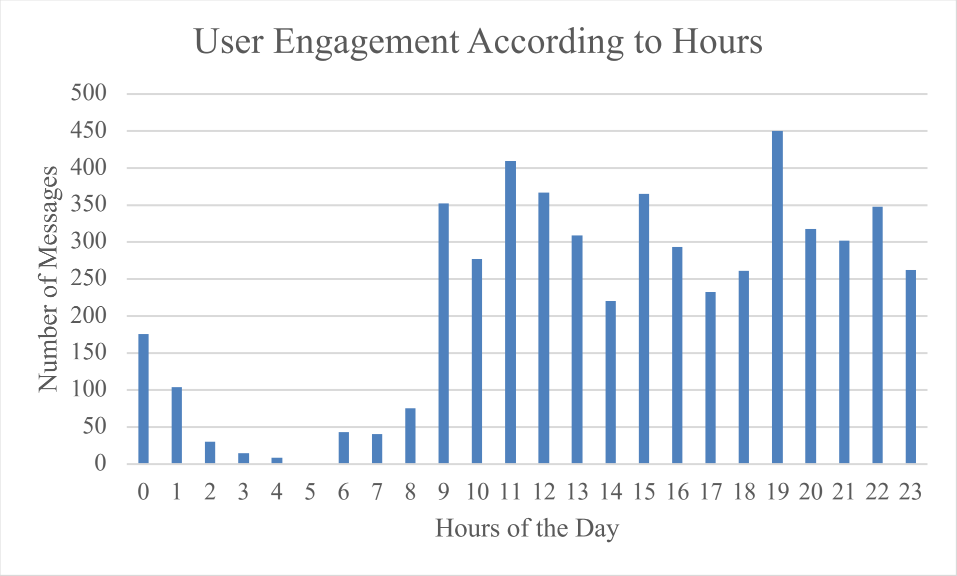


Figure 9

Change in Mean Results (AAQ+CDS) for Both Groups over period of experiment

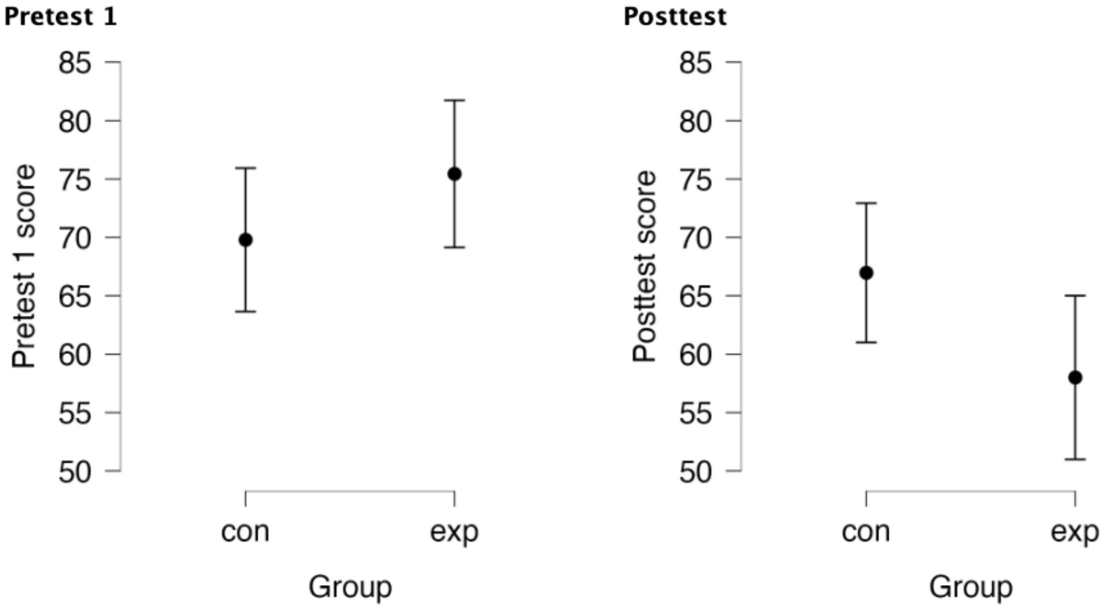
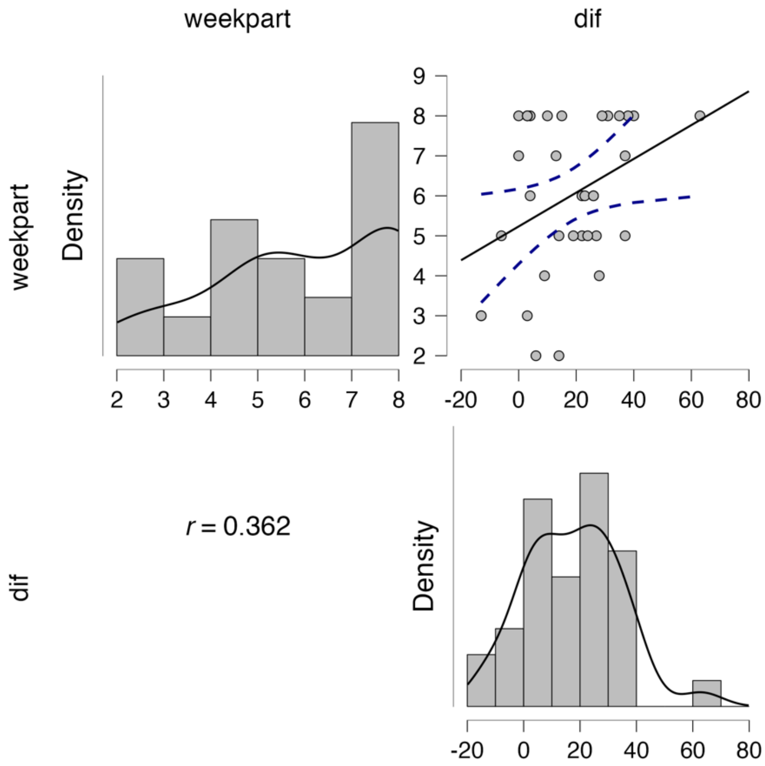


Figure 10

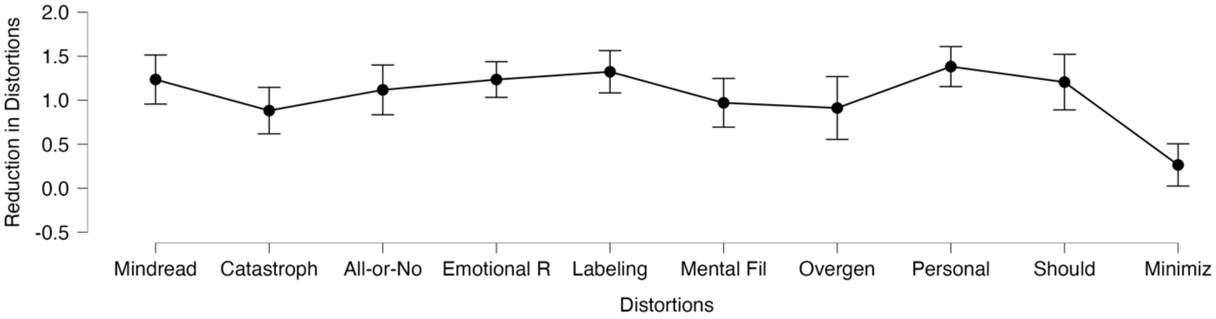
Correlation Plot



Note: “dif” stands for difference between pretest and posttest scores
“weekpart” stands for total number of weeks interacting with TeaBot

Figure 11

Repeated Measures ANOVA: Magnitude of Reduction in Distortions



Note: names of the distortions were shortened for the purposes of readability

Figure 12

Factor Analysis: Scree Plot Depicting the Major Factors

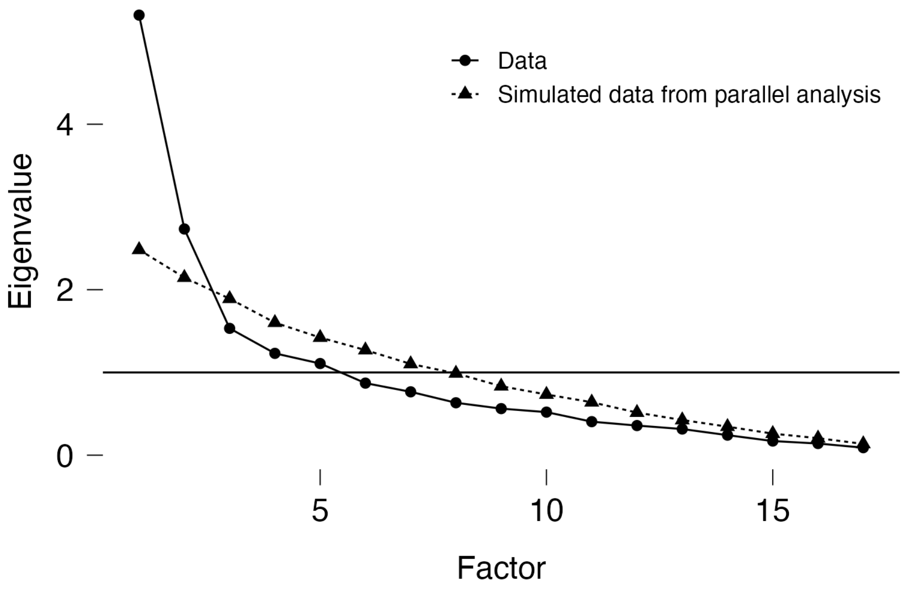
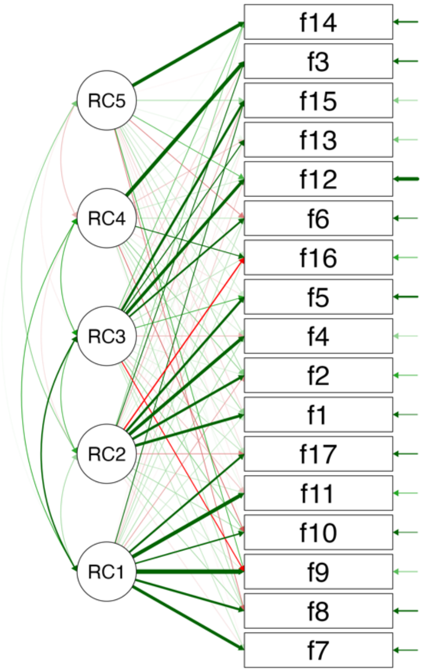


Figure 13

Factor Analysis: Path Diagram


RC5
RC4
RC3
RC2
RC1
f14
f3
f15
f13
f12
f6
f16
f5
f4
f2
f1
f17
f11
f10
f9
f8
f7