An Initial Assessment of a Chatbot for Rumination-Focused Cognitive Behavioral Therapy (RFCBT) in College Students

Alana Lucia Souza Oliveira[1(✉)], Leonardo Nogueira Matos[1], Methanias Colaço Junior[1], and Zenith Nara Costa Delabrida[2]

[1] Post-Graduate Program of Computer Science, Federal University of Sergipe, São Cristovão, SE, Brazil
{alanalso,leonardo}@dcomp.ufs.br
[2] Psychology Department, Federal University of Sergipe, São Cristovão, SE, Brazil
zenith@ufs.br

Abstract. Context: According to the WHO, suicide is the 2nd leading cause of death for young people aged 15 to 28, which often may have been the result of a depressive or anxiety disorder. **Objective:** To construct and evaluated a Chatbot that dialogues with college young people, with the purpose of disseminating mental health and ameliorate both depressive and anxious symptoms. **Method:** Our Chatbot is based on Rumination-focused Cognitive Behavioral Therapy (RFCBT), which focuses on observing the interaction between thought feelings and actions. For evaluation, an experiment that reached 105 young people was carried out. The bot's dialogues were constructed from psychological literature on RFCBT and tests were performed based on the PHQ-4 protocol and rumination tests. It is worth mentioning that the whole experiment was analyzed and authorized by the Ethics Committee. **Results:** The experiment results statistical significance was confirmed and obtained through T-test, for the rumination tests, and Wilcoxon Test, for the PHQ-4, obtaining p-values lower than 0.05. **Conclusion:** It has been shown that there was a change in symptoms after the use of Chatbot, as a result of mitigating the symptoms of evaluated mental disorders.

Keywords: Chatbot · Psychology · Well-being · Depression · Rumination · Experimentation

1 Introduction

In recent years, development and interest in Chatbot technology have increased. Companies such as Facebook, Microsoft and Amazon have exploited this technology with the creation of some chatbots. Microsoft, with Cortana, and Apple, with

© Springer Nature Switzerland AG 2021
O. Gervasi et al. (Eds.): ICCSA 2021, LNCS 12954, pp. 549–564, 2021.
https://doi.org/10.1007/978-3-030-86979-3_39

SIRI, for example, which act as personal assistants in a human-machine relationship. This interest in developing chatbots, capable of autonomous conversation and with learning ability, has been strongly driven by the large availability of bots development platforms in large social networking sites.

Thus, having these platforms as great tools to reach people through social messaging networks, and, in parallel, the growth in the development of mental health interventions offered on the Internet, as well as applications to complement existing mental health treatments [1], chatbots become a natural expansion line for these interventions. If implemented to work in social networks of text exchanges, such as WhatsApp, Facebook and Telegram, they become a more natural way for individuals to get involved with technology and have easier access to health treatments, since they are the environments most used for exchanging messages today. In this sense, the ability of technology to act on men's mental health problems has already been evaluated and evidenced [2], with demonstrations that a well-designed health-related conversation agent can establish a therapeutic relationship.

In addition to these observations, Stanford University has constructed the Woebot [3], a conversation agent that aims to evaluate the feasibility of a conversation interface and applies a technique called CBT (Cognitive Behavioral Therapy) through an automated chatbot, to facilitate engagement and reduction of depression or anxiety symptoms. Considering the great result that Woebot was able to validate in the intervention among university students, an opportunity to explore these techniques for young university students in Brazil, using the Portuguese language, is observed. This concern is justified by the severity of mental health disorders, therefore, according to the WHO (World Health Organization) [4], in 2012, there were 2,898 suicides of 13 to 29 year old, the age group comprising the university phase. Thus, the present work intends to go further on what has already been done by Woebot, since the choice of working with other therapy, the Rumination-focused Cognitive Behavioral Therapy (RFCBT), has demonstrated the primary target for decreasing and/or preventing depressive thoughts and actions is to focus on Rumination.

In this context, this project main goal was to develop a chatbot for Facebook Messenger platform. A chatbot that is able to help, mainly, young university students with the knowledge of Behavioral Cognitive Therapy focused on Rumination. The purpose is to offer this university population, a support tool for the knowledge of a treatment possibility for students' mental health. Moreover, the bot also aims to increase self-awareness of students own mental health, with daily dialogues that make the user to have a continuous learning conversation with the bot.

The chatbot, named Rumi, was used in an experiment with academic students, to evaluate the use and efficiency of a chatbot in the dissemination of mental health and the RFCBT. The results indicated that there were differences between the two applied psychological questionnaires, the PHQ-4 and the Rumination test, in favor of the use of Rumi. In other words, there was a significantly decrease in users feelings like anxiety, depression and rumination.

The remainder of this paper is organized as follows. Section 2 presents the theoretical base. Section 3 talks over the created chatbot. Section 4 describes related work. Section 5 presents the Methodology used for building chatbot. Section 6 details the experimental evaluation of the approach. Section 7 discusses the obtained results. Section 8 goes over the threats to the validity of the study. Finally, Sect. 9 presents the conclusions and discussion for future works.

2 Background

This section presents the theoretical background used in chatbot and the basis of this paper data analysis. At first it is talked about the Rumination-focused Cognitive Behavioral Therapy (RFCBT) is discussed and secondly it goes over the psychological tests applied for evaluation, PHQ-4 and the Rumination tests.

2.1 RFCBT

The Rumination-focused Cognitive Behavioral Therapy (RFCBT) is a variant of Behavioral Cognitive Therapy (CBT), developed by Watkins et al. (2007) [5]. The term that differentiates it, "Rumination", is defined by Watkins (2016) [6] "as recurring and repetitive thinking about symptoms (fatigue, low mood), feelings, problems, disturbing events and negative aspects about oneself, usually focusing on their causes, meaning and implications". For Nolen-Hoeksema, Mcbride and Larson (1997) [7], apud Watkins (2016) [6], rumination is like "the passive and repeatedly focusing of the symptoms of distress and the circumstances surrounding these symptoms".

To its creators, "although RFCBT is still grounded within the core principles and techniques of CBT for depression [8], this involves several additional new elements". In practice, RFBCT uses functional analysis to help patients realize that their negative rumination can be useful and guide them on how to switch to the most effective style of thinking. Functional analysis focuses on the variability of: (a) rumination (e.g., differences between useful and useless thinking about problems); (b) associated behaviors (for example, procrastination) and (c) behaviors against rumination, such as effective involvement in tasks. This detailed functional context analysis is then used to help patients to: (a) recognize the warning signs for rumination; (b) develop alternatives, strategies and contingency plans (for example, relaxation, assertiveness); and (c) changing environmental and behavioral standards (for example, changing the balance of routine tasks and obligations to do self-realising activities) [5].

In addition, Watkins et al. (2007) [5] present experiential imaging exercises and behavioral experiments as facilitators of a shift to a more useful style of thinking. Patients use targeted images to vividly recreate previous states when a more useful style of thinking was active, such as memories of being completely absorbed in an activity or experiences of being compassionate to themselves. Such exercises provide a direct counterpoint to rumination and can be used in contingency plans. These adaptations mean that RFCBT differs from standard

CBT to depression, which focuses on modifying the content of thoughts, by having a greater emphasis on directly modifying the thought process.

In [6], the adaptation of RFCBT to internet is introduced, with several justifications for the development of a variant of RFCBT that can be implemented remotely, are: accessibility, availability, convenience and coverage for psychological treatments. Depression, according to the author, is the greatest challenge among the world's health problems, being highly prevalent, chronic and dissonant. For this reason, traditional face to face therapy will never be widely available enough to reduce the major problem of depression. It is worth noting that the online version is not yet with free access and that the chatbots are an alternative to the north and feedback of an online version.

2.2 PHQ-4

The PHQ-4 test [9]) is an ultra-brief tool, valid for detecting anxiety and depression disorders. This is a 4-item self-report health questionnaire, with two questions from a 2-item depression scale (PHQ-2) and two questions from a 2-item anxiety scale (GAD-2), as a way to identify aspects of depression and anxiety. The test is validated in United States, but its validation in Brazil is still required. To start the adaptation process to Brazil, the applied test was translated. Default answers are scored as follows: "Not once" = 0, "Several days" = 1, "More than half the days" = 2, and "Almost every day" = 3. The sum of the questions is the variation between 0 and 12 points, thus, operative disorders are classified as normal, when scored from 0 to 2, mild, from 3 to 5, moderate, from 6 to 8, and severe, from 9 to 12. The test had in the following format:

- How many times in the last week have you felt uncomfortable, nervous or restless?
- How often in the last week have you been bothered by not being able to stop or control the concern?
- How many times in the last week have you felt upset because of the little interest or pleasure in doing things?
- How many times in the last week have you felt down, depressed or hopeless?

For each of the questions, these were the alternative answers: several days, more than half the days or almost every day.

2.3 Rumination Test

In view of the RFCBT, author Watkins (2016) [6] created a Rumination test that should be performed every 7 days, to check the effectiveness of the Therapy and to know how the current mental state of the patient/client under analysis is. In this work, as it was an experience outside an office and far from a psychologist/psychiatrist, the therapy itself was not applied and the original test was remodeled to a leaner version, as well as being adapted to the chatbot interface.

Thus, the participants answered the following questions from the rumination test, in scale format, during the dialogue with the bot:

- Last week, what percentage of time did you spend thinking repeatedly, insisting or worrying about negative thoughts about a disturbing issue, event, or problem?
- What percentage of your time did you use to control and stop negative thoughts so they do not interfere with your plans or prevent you from doing what you wanted last week?
- How much of your time did you spend avoiding or procrastinating doing things last week?
- How much avoiding things interfered with or prevented you from doing what you wanted last week?

For each of the questions, these were the alternative answers: 0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% or 100%.

3 Rumi

This section introduces Rumi, the chatbot created to be part of the experiment. The term "Rumi" comes from rumination, which is the aspect worked on in the therapy. It has been built through Chatfuel tool and its dialogues have been based on the therapy framework proposed in [6], which proposes 12 themes for each day of dialogue, but that were distributed in 15 days. Such dialogue has been built with the help and supervision of a psychologist.

The daily dialogues are based on following themes: Initial presentation of what therapy is. An initial explanation of what it works for and a presentation of Rumi communication platform and application of the PHQ-4 and the Rumination test; Introduction of Rumination as a habit, and as habit it can be changed; Introduction of the term Procrastination and Avoidance and how these affect in everyday life; SMART(Specific, Measurable, Achievable, Relevant, and Time-bound) Goals; Self-monitoring on symptoms such as Rumination and how to detect them.; "If-Then" method; Reapplication of the initial Rumination test; Self-love; Functional Analysis of Thoughts; Realignment in relation to the SMART methodology; Concrete and Abstract thoughts; Meditation; Self-esteem; Relapse prevention and a review of everything that has been presented during the last days and Farewell; and the application of the web-adapted PHQ-4 test and of the adaptation of the Rumination test.

Furthermore, in more worrying cases, through natural language processing provided by the Chatfuel tool, participants were encouraged to call 188 (Life Valuation Center number in Brazil) and 192 numbers, for emergencies and also to contact a specialist.

With the construction of Rumi, it was published and ran through a Facebook page with the URL = https://www.facebook.com/rumibot.bot/, for use and interaction with users.

4 Related Works

This section consists of work that is in line with this paper goal of using chatbot technology to help its user's mental health.

4.1 Woebot

Woebot is a chatbot created to work for the mental health of young university students through the use of psychology techniques. It was implemented by a research group at Stanford University in the United States. In the Woebot website [10], Woebot is defined as "a chatbot that helps monitor mood and learn about yourself. Based on a therapeutic framework known as Cognitive Behavioral Therapy, Woebot asks people how they are feeling and what is happening with their lives in the form of brief daily conversations". Woebot also talks to people about mental health and well-being, sending videos and other useful tools, depending on the mood and needs of the person at that time. It is the analogy to a self-help book able to learn about the person and become more specific to their needs over time.

In terms of technology, there is natural language processing in conjunction with the CBT technique, with interactions with the user that begin with a general investigation of the user's context (e.g., "What is happening in your world right now?") and your mood (for example, "How are you feeling?"), with responses provided as images of words or emojis, to represent that moment [3]. After collecting mood data, participants are presented with basic concepts related to CBT by a brief video link or by brief "word games", designed to facilitate teaching about cognitive distortions. In the initial process included by the authors, there is an introduction to the bot, noting that although the bot looks like a person, it is closer to "choosing its own self-help book" and therefore not fully able to understand what the user's needs are. The bot also briefly explains CBT and notifies the user that although a psychologist is "watching things" (for example, monitoring), this does not happen in real time and therefore the service should not be used as a substitute for a therapy [3]. In addition, in more worrisome cases, participants are encouraged to call 911 (emergency number in the United States).

As an assessment, Woebot was used as an intervention to decrease symptoms of depression in students. Two groups of university students participated in the research for two weeks [3]. Some tests were applied to evaluate the effectiveness of two interventions: using chat to read and read a self-help book. The tests included questionnaires scientifically evaluated and recommended for their effectiveness in psychology. The results evaluated the symptoms of depression by means of a covariance analysis between the tests applied in the groups. It was found that the group that received the chat intervention had a much greater improvement in the symptoms of depression than the self-help book group, demonstrating that even the short-term intervention had some effect, which suggests a great possibility to use the chatbot for this emerging area of mental health.

The bot works through English language interactions and a fee is charged after 15 days of free use. In addition, Woebot can be accessed via Facebook (https://www.facebook.com/HiWoebot/) or via an application on Android or IOS.

4.2 Owlie

According to its authors, Owlie is a chatbot that was created "to make available to the public as many new therapy tools as possible" [11]. This bot uses practical experiences in the field of psychology, with evidence based medicine (EBM) therapies scientifically evaluated, applying a rule-based decision and technology algorithm (word matching).

Owlie has three main characteristics: empathetic listening, tools/exercises, and therapeutic education. In this context, emotional crisis management is one of its main functions, helping users take responsibility for their own recovery and move towards a better quality of life. Currently, he also works 24 h a day, through his Facebook page (https://www.facebook.com/owlielechatbot), however, Owlie only talks to people through the French language.

5 Methodology

All tasks and methodological phases are described in the Table 1. The initial phase was called "Before the Experiment", the experiment phase has the same name and the analysis of the experiment was called only "Analysis". The experiment and its self-contained methodology, as well as its execution, will be detailed in the next section.

6 Experiment

This section describes an experimental evaluation of our approach. The presented experimental process follows the guidelines of Wohlin et al. (2012) [12]. This section focuses on the experiment definition and planning. The following section present the obtained experimental results.

6.1 Goal Section

The main objective of our study is to investigate the possibility of using chatbot technology to disseminate mental health through daily dialogues, increasing self-knowledge regarding participants' mental health and reducing symptoms such as rumination, anxiety and depression. The goal is formalized using the GQM Goal model proposed by Basili and Weiss (1984) [13] and presented by Solingen and Berghout (1999) [14]: **Analyze**, through a controlled experiment, one chatbot tool with cognitive behavioral therapy focused on Rumination, **in order** to evaluate, **with respect to** verify the possibility of use, effectiveness and quality of the mental health dissemination, **from the point of view** of psychologists, programmers and university students, **in the context** of University.

Table 1. Complete methodology procedure

	Performed tasks	Goals
Before	Meta-analysis research about chatbots and apps designed to aid mental health	This research was aimed at providing greater familiarity with the problem in order to make it more explicit and to provide the main hypothesis of using technology as a way to positively assist in prevention, treatment or help with mental health
	Descriptive Research about target population	Description of characteristics of young people and their mental health in Brazilian universities
	Chatbot and Dialog Development	Chatbot development using software development techniques and building dialogs that were used based on RFCBT
	Field research	Exploratory research in which, in addition to bibliographic and/or documentary research, data collection was performed in the target population and in order to understand aspects
	Experimental Method	Planning and elaboration of the experimental method for test a hypothesis of causality, wich is explained in more detail in the Sect. 6
Experiment	Recruitment of Participants for the Experiment	Potential participants were recruited from a Facebook post, targeting a Federal University of Sergipe (UFS) community. Social media of UFS students
	Execution of the experiment	The experiment was open for 30 days to capture participants. Which in 15 days occurred the interaction of the chatbot participant
	Daily Dialogues	Application of the dialogues that are presented in the Sect. 3 for 15 days, and 2.2 and 2.3 was performed for data analysis
	Rumination Test application	At the 1st, 7th and 15th daily dialogues, were applied an adaptation of the rumination test presented in the Sect. 2.3
	PHQ-4 Test application	At the 1st and 15th of the dialogs with chatbot were applied the tests PHQ-4 presented in the Sect. 2.2
	Data Collection	During each daily dialogues, at the end of each dialogue day, the messages between the participants was stored in a database
Analysis	Quantitative data analysis	Its objective is to measure and allow hypothesis testing, as the results are more concrete compared to the tests that are applied to the PHQ-4 and Rumination-Test
	Qualitative data analysis	It aims to extract and analyze from participants their opinions on the use of chatbot and psychoeducation through a questionnaire

6.2 Planning

This subsection details all experimental design.

Context Selection. The experiment targeted academic students.

Hypothesis Formulation. The question we are trying to explore is this: Does applying chatbot in therapy dialogues, help in spreading mental health and therapy? To evaluate this question, the average of the PHQ-4 and Empathy Rumination questionnaires will be used. For this question, formally, the hypothesis that we are trying to confirm is:

Null Hypothesis $H0^R$: Rumination level, according to Rumination test, after interaction, is the same. $H0^R$: μBefore Chatbot(Levels) = μAfter Chatbot(Levels).

Alternative Hypothesis $H1^R$: Rumination level, according to Rumination test, after interaction, is different. $H1^R$: μBefore Chatbot(Levels) $\neq$ μAfter Chatbot(Levels).

Null Hypothesis $H0^{AD}$: The levels of anxiety and depression, according to the grades of the questions in the questionnaire, in the PHQ-4 test, after the interaction, are the same. $H0^{AD}$: μBefore Chatbot(Grades) = μAfter Chatbot(Grades).

Alternative Hypothesis $H1^{AD}$: The levels of anxiety and depression, according to the grades of the questions in the questionnaire, in the PHQ-4 test, after the interaction, are different. $H1^{AD}$: μBefore Chatbot(Grades) $\neq$ μAfter Chatbot(Grades).

Thus, to investigate which of the hypothesis are refuted, the following dependent, independent and intervening variables will be considered:

- **Independent variables:** Chatbot Development Tool; Rumi; Questionnaires; and the RFCBT protocol.
- **Dependent variables:** Dialogues, from which were derived: Anxiety and Depression levels (Questionnaire PHQ-4 - applied every 7 days); Rumination Level (Rumination Questionnaire applied every 15 days); and final questionnaire about overall chatbot experience.
- **Intervening variables:** Communication between participants; Previous anxiety and depression; Variation of the daily mood in the 15 days of use; Lack of experience.

Participant and Artifact Selection. After defining hypotheses and variables to be analyzed, the participants and objects selecting process has started. Potential participants were recruited from a Facebook post, targeting a university community. Social media like WhatsApp, Instagram and even groups of study and interaction of students, to voluntarily participate in such an experimental evaluation, were used. All participants accepted the terms and conditions for participating in the experiment before starting the dialogue. Such terms have been validated by the ethics committee. 236 people have shown interest in participating in the experiment and 105 people have participated by the end of the experiment, 63 women and 42 men aged 18 to 28, inclusion criteria age for the experiment. The experiment and its dissemination lasted a month, in which each participant spent 15 days talking to the bot. For legal reasons, participants names are not disclosed in this work.

Instrumentation. The instrumentation process is initially done with the environment setup for the experiment and data collection planning. That was done in an automated way through the Chatfuel (chatbot development tool), Integromat (integration tool) and Excel Online. In short, here are the resources used: Facebook Messenger environment, for dialogue with participants; For the development of Rumi, was used Chatfuel; Integromat and Excel Online were used to capture the data; PHQ-4; and Rumination Test.

Experimental Design. The entire project and its experimental design have been evaluated and approved by the Ethics Committee, the number of the Ethical Evaluation Presentation Certificate (CAAE) is 90842818.4.0000.5546.

Entrants were sorted by gender and Facebook IDs only. The privacy and anonymity were preserved in the project. All evaluated participants signed the Informed Consent Form. In summary, the experiment was divided into the two parts below.

Daily Dialogues: On the 1st and 15th day the PHQ-4 tests will be applied. On the 1st, 7th and 15th day, an adaptation of the rumination test presented in the Sect. 2.2 and 2.3 will be performed for data analysis. After the conclusion of the dialogues, still on the same platform, participants will be sent an opinion questionnaire about what was presented by the bot. It will be reported later in the Results section.

Data collection: At the end of each day of dialogue, the messages between the parties will be stored in a database. The analysis of the collected data will be presented in the next section of this article.

6.3 Operation

This subsection describes the experiment operation process.

Preparation. The following preparation steps were carried out.

Summon up of participants: The volunteers were summoned through the social network that would be used. Later they signed the terms of conditions and ethics for the experiment. Dialogue with the Bot: The conversation has been held daily for 15 days. The bot automatically called the participant to the conversation every day at the same time that the participant had started the conversation on the first day. The tests were also carried out within the dialogue itself. At the first moment, the participants took PHQ-4 and rumination tests already dialoguing with the bot. 15 days after, the participants retook the tests to evaluate the dialog experiment impact on those. Doubts with the experiment: An email and telephone number were made available for participants to ask questions about the experiment.

6.4 Execution

After carrying out the previous steps, the experiment occurred from August 6th, 2018 at 7:00 p.m. until September 3rd, 2018 at 7:00 p.m., in a 4-week period of data collection, and was started according to the design presented above.

A 236 total people have started the conversation with chatbot and accepted the Free and Informed Consent Term, but only 105 people have concluded within the stipulated deadline. In order to answer the research questions, the following dependent variables were analyzed: PHQ-4 and Rumination tests. Applied at the first and last days of the use of the bot.

6.5 Data Validation

For the experiment, a factor (the approach used for dissemination of mental health) and two treatments have been considered: before and after the use of

Rumi. In this context, for both moments, PHQ-4 questionnaire and rumination test results were collected. For support in analysis, interpretation and validation, four types of statistical tests were used, Kolmogorov-Smirnov and Shapiro-Wilk tests were used to verify sample normality. For the Rumination Test, our sample was normal for both normality tests, so it was used the parametric Test-T. Unlike Rumination test, PHQ-4 test was not validated for normality, so nonparametric Wilcoxon Related-Samples test was used. These statistical tests were used to compare the differences of the grade means of the applications of the psychological tests before and after the interaction with the bot. In addition, all statistical tests were done from IBM SPSS Statistics 20 software.

7 Results and Discussion

Primarily, a correspondence between questionnaires and answers analysis was carried out, those who did not complete the experiment and those who were not in the age range of 18 to 28 years old were removed. After the content analysis, the score of each participant was calculated by analyzing their responses values in the tests, at the beginning and at the end of the experiment. Moreover, a detection and treatment process of outliers was performed, in order to reduce their impacts on the data set [15].

Then, it was defined a significance level of 0.05 throughout the experiment and the Kolmogorov-Smirnov and Shapiro-Wilk tests were applied for analysis of the normal distribution. For both treatments, the 1st and 2nd application of psychological tests, the values of the p-value were similar, in the contexts of each type of psychological test. Since the p-values for the rumination tests results were greater than 0.05 ($p > 0.05$), the data distribution is assumed to be normal. In the case of the PHQ-4 test results sample, the calculated p-values were less than 0.05, i.e., the hypothesis of normality of the data was rejected.

From the next section, according to the situation of the normality of the data, each psychological test will be analyzed individually, to verify the statistical significance of the results achieved after the use of the bot.

7.1 Rumination Test

This scale was applied 3 times over the 15 days. It was firstly applied at the initial conversation, the second time on the 7th day and the third time on the 15th, and last, day of dialogue with the Rumi. General averages from 1st and 15th days were analyzed. For each of the questions, the percentages of the answers are transformed into a scale of 0 to 10, so, since there are 4 questions, the final grade goes from 0 to 40. So, the averages varied from 0 to 40 for the answers, among which 40 shows a lot of Rumination feeling and 0 matches absence of Rumination. When analyzing the results and removing outliers, 99 responses were analyzed, the meaning of the responses is shown in Fig. 1.

Due to data normality, the hypothesis test applied in this context was the T-test. From this, it has been verified, on Rumination test evaluation, that there

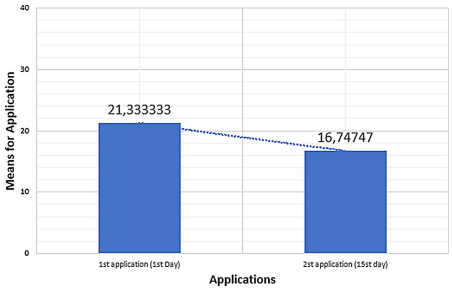

Fig. 1. Applications of rumination test

was a statistically significant difference between the first the last day of experiment, with the average having decreased from 21.333 to 16.747. The significance was found due to the Sig. obtained (p-value) of 0.000, which is less than the level of significance adopted, as shown in Fig. 2.

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Rumination_1 - Rumination_2	4,586	9,657	,971	2,660	6,512	4,725	98	,000

Fig. 2. Applications of rumination test

That demonstrates that for one on rumination symptoms, averages decrease considerably. Therefore, the null hypothesis $H0^R$ was rejected.

7.2 PHQ-4 Test

As suggested in [9], PHQ-4 should be applied every 15 days. Thus, the test was applied at the first contact with the bot and at the 15th day, and last, of the dialogue. When analyzing the results and removing the outliers and storage errors, 97 responses were analyzed.

Due to the non-normality of the PHQ-4 questionnaire data, previously pointed, the hypothesis test applied in this context will be the Wilcoxon non-parametric test, as shown in Fig. 3. In this case, it is needed to define a new null hypothesis: $H0^{AD}$: The levels of anxiety and depression, according to the distribution of the grades of the questions in the questionnaire, in the PHQ-4 test, after the interaction, are the same.

The results indicated that the performance with the use of Rumi, in favor of mental health, was higher. In other words, the overall average grade decreased from 6 to 4. In addition, according to the statistical results, the Sig. obtained (p-value) of 0.000 is less than the level of significance adopted. Therefore, the null hypothesis $H0^{AD}$ was rejected, that is, statistically, anxiety and depression

Related-Samples Wilcoxon Signed Rank Test Summary

Total N	97
Test Statistic	1051,500
Standard Error	252,178
Standardized Test Statistic	-4,130
Asymptotic Sig.(2-sided test)	,000

Fig. 3. Wilcoxon test

levels, according to PHQ-4 test, do not have same distribution of the grades of the questionnaire questions, after interaction.

8 Threats of Validity

The threats to validity for the present study were:

- **Construct validity:** The choice of subject for conversation and use of chatbot are the possible threats. The first is related to the possibility of choosing an unpleasant or unfamiliar issue to the interlocutors. To mitigate this threat, it has been chosen to address the issue in relation to the environment in which the participants live, the university and their daily lives. The second one can be caused by lack of experience in chatting with a bot through Messenger. To mitigate this, during the experiment and dialogue, the bot itself signaled ways and alternatives that the participant should do to instruct them in dialogue and self-knowledge.
- **Internal validity:** The main threats can be: response time, a sensitive subject for conversation, tiredness, current state of mental health of the participant, and demand characterization. As the tool have predefined dialogues, the response time threat could be caused by a question not being answered within 24 h, because the next day the bot starts another different dialog, leaving unanswered questions. In order to mitigate this threat, during the course of an experiment, issues that have already been extensively dealt on other days have been taken up several times. The second threat may be caused by choosing an unpleasant subject or one that brings bad feelings about something to the participant. In order to mitigate this, playful interactions with images, stories, and videos that softened the impact of the subject were chose, bringing a theoretical subject to practice. In the case of the threat of fatigue and lack of time of the participant, due to the long time of interaction, the dialogues required no more than 10 min of interaction a day. Moreover, if a participant missed a day interaction with the bot, he didn't lost the experiment. The penultimate threat is related to the possible mental state of the participant at the day of the conversation, since the interlocutors were not pre-clinicated. To mitigate this, every day, before the experiment, the Rumibot asked how the person was and whether she was interested in continuing

the conversation, regardless of their response and daily mood. Finally, in order not to influence our study, participants were not aware of factor, metrics or research hypothesis, they only knew that it was a dialog based on psychology.
– **External validity:** Due to the fact that the summons occurred through social networks, the experiment was able to reach people outside the academic scope and outside the explicit age range. Another factor was the time available for the application of the experiment. Data collection for the experiment remained opened for a month, since the experiment lasts 15 days for each participant. In this way, the experiment was able to reach more participants and these could complete the 15 days of the experiment. Lastly, more experiments will be needed to generalize the conclusions presented, however, the results allow us to draw insights to guide future investigations.

9 Conclusions

More and more people are using chatbot services to communicate with businesses and that has expanded to health care. The facilitated access to chatbot, brought by social networks, can be used as an aid in the dissemination of therapies knowledge such as the RFCBT. The need to work on mental health at universities and the lack of access to this kind of service by people are some of the reasons that magnifies the importance of admitting this new tool.

The search for mental health has been considered an urgent need worldwide, in this sense, the World Health Organization (WHO) has created a comprehensive mental health action plan 2013–2020 [16]. The plan considers that the creation of new actions concerning mental health situation of people would encourage all countries to work on this problem. Faced with this reality, according to the United Nations website in [18], WHO analyzed and estimated the costs of treatment and their results in health in 36 low, middle and high-income countries over a 15 year period (between 2016 and 2030). The research has shown that low levels of recognition and access to care for depression and anxiety cause a global economic loss of $1 trillion a year. For that reason, this project aligns itself to solve both access to treatments and aid in the prevention and treatment of mental disorders issues.

The development of Rumi was conducted on the Chatfuel tool in combination with Integromat and Google Sheets. It has been designed to work within Facebook messenger. The bot's dialogue structure was constructed from the ideologies presented in [5], along with empathic expressions, instructional videos, and predefined learning scripts. Its development has brought some of its key requirements for easy handling, interaction, feasibility and scalability in relation to the university environment. Rumibot also provides a way of self-knowledge and disseminates therapy knowledge in a clear, intuitive and fun way, facilitating the population self-monitoring.

With the completion of the Rumi development, a final evaluation on how was the use of chatbot for this purpose experience was given to the users. The Rumination and PHQ-4 tests of depression and anxiety, which measure the current

mental state of the population, were analyzed together with the Rumi dialogue, to see if these aspects would be changed for the better or for the worse. Both tests have had positive results. Rumination test has decreased its average, and the distribution of grades of the PHQ-4 questions, after interaction, also was different. In this context, we highlight the change in all aspects, rumination, depression and anxiety, which also contributes so that Rumi may help to alleviate some of the symptoms of these disorders. Therefore, chatbot technology can function as an aid tool on knowledge of behavioral cognitive therapy focused on Rumination.

9.1 Future Works

From the observations made in the user satisfaction questionnaire, applied by Rumi, it was evidenced as the need to seek improvements in some aspects of the chatbot. As a result, future work will be guided to meet the following improvements: Add new dialogs, using new theories of psychology and/or using new standardized psychiatry questionnaires that seek to assist in the prevention of suicide; And also add psychological profile detection aspects, to increase harmony, education and politeness, so that responses to the interlocutor are more empathetic. The above mentioned enhancements to Rumi are in progress, completing planned activities of this research project.

References

1. Bakker, D., Kazantzis, N., Rickwood, D., Rickard, N.: Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Mental Health **3**(1), e4984 (2016)
2. Miner, A.S., Milstein, A., Schueller, S., Hegde, R., Mangurian, C., Linos, E.: Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. JAMA Inter. Med. **176**(5), 619–625 (2016)
3. Fitzpatrick, K.K., Darcy, A., Vierhile, M.: Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Mental Health **4**(2), e7785 (2017)
4. Pereira, A., Cardoso, F.: Ideação suicida na população universitária: Uma revisão da literatura (2015)
5. Watkins, E., et al.: Rumination-focused cognitive behaviour therapy for residual depression: a case series. Behav. Res. Ther. **45**(9), 2144–2154 (2007)
6. Watkins, E.R.: Rumination-Focused Cognitive-behavioral Therapy for Depression. Guilford Publications, New York (2016)
7. Nolen-Hoeksema, S., McBride, A., Larson, J.: Rumination and psychological distress among bereaved partners. J. Pers. Soc. Psychol. **72**(4), 855 (1997)
8. Beck, J.S.: Cognitive Behavior Therapy: Basics and Beyond. Guilford Press, New York (2011)
9. Kroenke, K., Spitzer, R.L., Williams, J.B.W., Lowe, B.: An ultra-brief screening scale for anxiety and depression: the PHQ-4. Psychosomatics **50**(6), 613–621 (2009)

10. Woebot (2017). https://woebot.io/
11. Falala-Séchet, C., Antoine, L., Thiriez, I., Bungener, C.: Owlie: a chatbot that provides emotional support for coping with psychological difficulties. In: Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents, pp. 236–237. ACM (2019)
12. Wohlin, C., Runeson, P., Host, M., Ohlsson, M.C., Regnell, B., Wesslen, A.: Experimentation in Software Engineering. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-29044-2
13. Basili, V.R., Weiss, D.M.: A methodology for collecting valid software engineering data. IEEE Trans. Softw. Eng. **6**, 728–738 (1984)
14. Van Solingen, D.M., Berghout, E.W.: The Goal/Question/Metric Method: A Practical Guide for Quality Improvement of Software Development. McGraw-Hill, New York (1999)
15. Bramer, M.: Data for data mining. In: Principles of Data Mining. UTCS, pp. 9–19. Springer, London (2016). https://doi.org/10.1007/978-1-4471-7307-6_2
16. Watts, S., et al.: CBT for depression: a pilot RCT comparing mobile phone vs computer. BMC Psychiatry **13**(1), 49 (2013)
17. Brown, L.: The conditional level of student's t test. Ann. Math. Stat. **38**(4), 1068–1071 (1967)
18. ONU Brasil: Investimento em saúde mental cresce em ritmo insuficiente, denuncia OMS (2018). https://nacoesunidas.org/investimento-em-saude-mental-cresce-em-ritmo-insuficiente-denuncia-oms/