GENERATIVE COGNITIVE BEHAVIORAL THERAPY WITH SPOKEN DIALOG SYSTEMS' SUPPORT

BY

AYESHEH AHRARI KHALAF

A dissertation submitted in fulfillment of the requirement for the degree of Master of Science in Engineering

Kulliyyah Engineering
International Islamic University Malaysia

MAY 2023

ABSTRACT

One of the objectives and aspirations of scientists and engineers ever since the development of computers has been to interact naturally with machines. Hence features of artificial intelligence (AI) like natural language processing and natural language generation were developed. The field of AI that is thought to be expanding the fastest is interactive conversational systems. Numerous businesses have created various Virtual Personal Assistants (VPAs) using these technologies, including Apple's Siri, Amazon's Alexa, and Google Assistant, among others. While an ongoing effort to increase the friendliness and constancy of informal dialogue systems, most research focuses solely on simulating human-like replies, leaving the features of modeling interlocutors' awareness are unexplored. Meanwhile, cognitive science research reveals that awareness is a crucial indicator of a high-quality informal conversation. To precisely model understanding, Persona Perception (P^2) Bot was developed using a transmitter-receiver-based structure. P^2 Bot leverages mutual persona perception to improve the quality of customized dialogue generation. Even though many chatbots have been introduced through the years to diagnose or treat psychological disorders, we are yet to have a user-friendly chatbot available. This research aims on improving the quality of conversation generation by implementing the Generative Pre-trained Transformer-2 (GPT-2) model on P^2 Bot. GPT-2 is a 1.5B parameter transformer model which produces state-of-the-art accuracy in a zero-shot setting on 7 out of 8 evaluated language modeling datasets. Observations on a large open-source dataset, PERSONA-CHAT, show that the technique is successful, with some improvement above state-of-the-art baselines in both automatic measures and human assessments. The model has achieved 82.2% accuracy on Hits@1(%) in the original data and 68.8% on the revised data. On the human evaluation, the model scored an average of 2.66 meaning the provided responses were coherent and informative. A smart generative cognitive behavioral therapy with spoken dialogue systems support was then developed using the model, which was then implemented using modern technologies in VPAs like voice recognition, Natural Language Understanding (NLU), and text-to-speech. This system is a magnificent device to help with voice-based systems because it can have therapeutic discussions with the users utilizing text and vocal interactive user experience.

ملخص البحث

منذ اختراع أجهزة الكمبيوتر، كان أحد أكبر أهداف الباحثين والمهندسين هو إجراء حوار طبيعي مع الآلات الإلكترونية. لذلك، أدخل الذكاء الاصطناعي (AI) معالجة اللغة الطبيعية وتوليد اللغة الطبيعية التي تهتم بأنظمة المحادثة التفاعلية باعتبارها المنطقة الأسرع نموًا في الذكاء الاصطناعي. استخدمت العديد من الشركات هذه التقنيات لإنشاء أنواع مختلفة من المساعدين الشخصيين الافتراضيين (VPAs) مثل Google Assistant و Alexa من Amazon و Apple's Siri وغيرها. بينما يُبذل جهد مستمر لزيادة متعة وثبات أنظمة الحوار غير الرسمي، تركز معظم الأبحاث فقط على محاكاة الردود الشبيهة بالبشر، تاركةً سمات نماذج وعي المحاورين غير مستكشفة. في اثناء ذلك، يكشف البحث العلمي المعرفي أن الوعي مؤشر حاسم للمحادثات غير الرسمية عالية الجودة. لنمذجة الفهم بدقة، تم تطوير روبوت Persona (P^2) Perception باستخدام بنية قائمة على جهاز الإرسال والاستقبال. يستفيد P^2 Bot من الإدراك الشخصي المتبادل لتحسين جودة إنشاء الحوار المخصص. على الرغم من أن العديد من روبوتات المحادثة قد تم تقديمها على مر السنين لتشخيص الاضطرابات النفسية أو علاجها ، إلا أنه لا يوجد روبوت محادثة سهل الاستخدام. يهدف هذا البحث إلى تحسين جودة توليد المحادثة من خلال تطبيق نموذج GPT-2 على P^2 Bot. نموذج GPT-2 هو نموذج محول معامل 1.5B ينتج دقة متطورة حسب التقييم للتعلم من الصفر في 7 من أصل 8 مجموعة بيانات لنمذجة اللغة. تُظهر الملاحظات على مجموعة بيانات كبيرة مفتوحة المصدر PERSONA-CHAT أن التقنية ناجحة، مع بعض التحسينات مقارنة بأحدث طراز للخطوط الأساسية في كل من المقاييس التلقائية والتقييمات البشرية. حقق النموذج دقة 82.2٪ بزيادة 1% عن البيانات الأصلية و 68.8٪ على البيانات المنقحة. وفي التقييم البشري، سجل النموذج متوسط 2.66 مما يعني أن الردود المقدمة كانت متماسكة وغنية بالمعلومات. بعد ذلك، تم تنفيذ النموذج باستخدام تقنيات حديثة في VPAs، مثل التعرف على الصوت وفهم اللغة الطبيعية

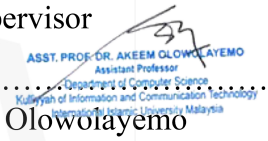
(NLU) و خوارزمية تحويل النص إلى كلام لتطوير معالجة السلوك الإدراكي الذكي الذي يدعم أنظمة الحوار المنطوقة. يمكن لهذا النموذج إجراء محادثات علاجية مع المستخدمين باستخدام واجهة مستخدم للمحادثة النصية والصوتية، مما يجعله نظامًا علاجيًا مميزاً مدعوم بأنظمة استجابة صوتية تفاعلية.

APPROVAL PAGE

I certify that I have supervised and read this study and that in my opinion, it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a thesis for the degree of Master of Science in Computer and Information Engineering.

……
Aisha Hassan Abdalla Hashim
Supervisor

……
Rashidah Funke Olanrewaju
Co-Supervisor



……
Akeem Olowolayemo
Co-Supervisor

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a thesis for the degree of Master of Science in Computer and Information Engineering.

……
Rafiqul Islam
Internal Examiner

……
Rashid Abdelhaleem Saeed
External Examiner

This thesis was submitted to the Department of Electrical and Computer Engineering and is accepted as a fulfillment of the requirement for the degree of Master of Science in Computer and Information Engineering.

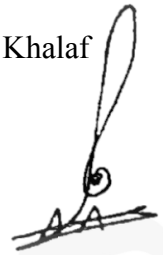
……
Rafiqul Islam
Head, Department of Electrical and Computer Engineering

This dissertation was submitted to the Kulliyyah of Engineering and is accepted as a fulfillment of the requirement for the degree of Master of Science in Computer and Information Engineering.

……
Sany Izan Ihsan
Dean, Kulliyyah of Engineering

DECLARATION

I hereby declare that this dissertation is the result of my own investigations, except where otherwise stated. I also declare that it has not been previously or concurrently submitted as a whole for any other degrees at IIUM or other institutions.

Ayesheh Ahrari Khalaf 

Date: 08 May 2023

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

DECLARATION OF COPYRIGHT AND AFFIRMATION OF FAIR USE OF UNPUBLISHED RESEARCH

GENERATIVE COGNITIVE BEHAVIORAL THERAPY WITH SPOKEN DIALOG SYSTEMS' SUPPORT

I declare that the copyright holder of this thesis are jointly owned by Ayesheh Ahrari Khalaf and IIUM.

Affirmed by Ayesheh Ahrari Khalaf

08/05/2023

This thesis is dedicated to my parents for going through all the challenges of this life to ensure I get a chance to fulfill my dreams.

ACKNOWLEDGEMENTS

Praise be to Allah (SWT), the Lord of both worlds. Prayers and peace be upon our Prophet, Muhammad, his family, and his companions.

I praise Allah (SWT) first and foremost for allowing me to complete my research and for all of these ongoing, unfathomable blessings. His Mercies and Blessings on me have made it possible for me to finish my thesis, even if it has been arduous.

My supervisor, Prof. Dr. Aisha Hassan Abdalla Hashim, whose devoted nature, compassion, promptitude, thoroughness, and motherhood have made it possible for me to complete my work successfully, deserves the highest amount of gratitude. She provided thorough remarks, helpful recommendations, and thought-provoking questions that significantly improved this thesis, and I want to express my gratitude for them. Her intelligent remarks, ideas, and questions were the result of her outstanding understanding of the purpose and content of this book, and they greatly aided my learning. Despite her obligations, she found time to discuss and counsel me whenever I needed it. She provided me with moral support, which without a doubt gave me a lift while I built and wrote the draft of this research report. I also like to thank my co-supervisors, Asst. Prof. Dr. Akeem Olowolayemo and Assoc. Prof. Dr. Rashidah Funke Olanrewajue, whose help and collaboration made this research possible.

My appreciation also extends to my dear parents for their prayers, patience, and compassion.

We praise Allah once more for His unending kindness toward us, one of which is allowing us to complete the composition of this thesis. Alhamdulillah.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

c^A	Interlocutor A's character
c^B	Interlocutor B's character
d_t^A	A's utterance in t[th] trun
P	Personalities
r_t^A	The entire of conversation archive up to t[th] turn
θ	The transmitter's parameter
$d_{t,n}^A$	The token of t[th] in d_t^A
$d_{t<n}^A$	The sequence of tokens before the n-th token
α	A hyper-parameter
p(x)	Probability distribution function of x
λ_{MLE}	Maximum log-likelihood function
$\widehat{d}_t^A$	Maximum length-normalized scored after beam search
$R(a_t^A)$	Generated reward for A at t position
R_1	Language reward shape function
R_2	Coherence reward shape function

R_3 Mutual P^2 reward shape function

β The discount factor

$r(a_t^A)$ P^2 the score in the n-th turn

γ_n hyper parameters

[CS] Character Started token

[EOS] End Of Sentence token

[SOS] Start Of Sentence token

[CLS] Classification token

[MASK] Masking token

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
BLEU	Bilingual Evaluation Understudy
VPAs	Virtual Personal Assistants
P^2	Persona Perception
NLU	Natural Language Understanding
WHO	World Health Organization
UNHCR	United Nations High Commissioner for Refugees
OHCHR	United Nations High Commissioner for Human Rights
CBT	Cognitive Behavioral Therapy
CBT test	Children's Book Test
DNN	Deep Neural Networks
ASR	Automatic Speech Recognition
CNN	Convolutional Neural Networks
LSTM	Long Short-Term Memory
IIUM	International Islamic University Malaysia
RNN	Recurrent Neural Network
HAN	Hierarchical Attention Network
SOM	Self Organizing Maps
SUS	System Usability Scale
GPT-2	Generative Pre-trained Transformer-2
ML	Machine Learning
DL	Deep Learning
NN	Neural Network
GPUs	Graphics Processing Units
HRED	Hierarchical Recurrent Encoder-Decoder
SEQ2SEQ	Sequence to Sequence

app.	appendix	n.p.	no place: no publisher
art./arts.	article/articles	no./no.s	number/numbers
b.	born	n. s.	new series
bk./bks.	book/books	o. s.	old series
C. P. C.	Criminal Procedure Code	P. B. U. H.	Peace Be Upon Him
c.	copyright	P. L. D.	All Pakistan Legal Decisions
ca.	(circa): about, approximately	P. P. C.	Pakistan Penal Code
cf.	compare	p./pars.	paragraph/paragraphs
ch.	chapter (in legal *fi*rms)	passim	here and there
chap./chaps.	chapter/chapters	pt./pts.	part/parts
col./cols.	column/columns	q. v.	*(quode vide)*: which see
comp./comps.	compiler/compilers; compiled by	Q. Sh	Qanun – E Shahadat
dept./depts.	department/departments	S. L. J.	The Sudan, Law, Journal
d	died	S. W. T.	Subhanahu Wa Ta'ala (Praise be to Allah and the Most High)
div./divs.	division/divisions		
e. g	*(exempligratia)*; for example	sc.	scene
ed./eds.	edition/editions; editor, edited by	sec./secs.	section/sections
et al.	*(et alia)*: and others	sic.	so, thus
et seq	*(et sequers)*: and the following	s. l.	*(sinoloco)*: no place of publication
etc	*(et cetera)*: and so forth pages that follow	s. n.	*(sine nomine)*: no publisher
*fi*g./*fi*gs.	*fi*gure/*fi*gures	s. v.	*(sub-verbo, sub-voce) under the word of heading*
ibid.	*(ibidem)*: in the same place	trans.	translator/translated by
id	*(idem)*: the same below	v./vv.	verse/verses
L. E.	Law of Evidence	viz.	*(videlicet)*: namely
l. v.	*(locus variis)*: various places (of publication)	vol./vols.	volume/volumes
ms./mss.	manuscript/manuscripts		
n. d.	no date		

CHAPTER ONE

INTRODUCTION

1.1 OVERVIEW

In January 2020, the World Health Organization (WHO) estimated that more than 264 million people worldwide experience depression. Additionally, it stated that depression is the main contributor to disability and might result in suicide. Following the Coronavirus Disease 2019 (COVID-19) pandemic, WHO remarked in March 2022 that less than 2% of global health funds are allocated to mental health (*Covid-19 pandemic triggers 25% increase in the prevalence of anxiety and depression worldwide*, 2022).

According to industry market research, surveys, and statistics published in March 2016 and cited by the National Institute of Mental Health, more than a quarter of Americans experience depression or anxiety yearly. According to a survey by the Kaiser Family Foundation, approximately half of the American citizens are worried about how the COVID-19 epidemic will affect their mental well-being (N. Panchal et al., 2020). According to a 2017 survey of 273,203 Malaysian citizens, upwards of 6.7% had some degree of depressive episodes (Abas & Sukaimi, 2018). Nearly 500,000 Malaysians, according to the National Health and Morbidity Survey (NHMS 2019), are showing some symptoms of depression. In addition, 424,000 kids are dealing with mental health concerns (Bernama, 2020). The two surveys also conclude the upsurge in depression and mental health in Malaysia.

Currently, just 6% of 165,000 healthcare applications accessible in smartphone application stores are focused on mental health issues (Carlo et al., 2019). As stated in

Figure 1.1, number 3 of the sustainable development objectives of the United Nations High Commissioner for Refugees (UNHCR), every human being has a right for good health and well-being. In May 2018, the UNHCR issued an essay titled "Mental health is a human right." The difficulty of carrying out everyday commitments, including going to work or school, as well as one's and other people's obligations, is brought up. While it is clear that "there cannot be health without mental health," according to Mr. Dainius Pras, there is still not nearly as much focus and funding given to it as there is to physical health anywhere in the world. The Office of the United Nations High Commissioner for Human Rights (OHCHR) provided this report in May 2018. With this research, we hope to help achieve a small portion of this goal worldwide.

Figure 1.1 UNHCR's Sustainable Development Goals (*The Sustainable Development Goals and Addressing Statelessness*, 2017)

Cognitive Behavioral Therapy (CBT) implements various cognitive and behavioral interactions as a well-known and scientifically validated treatment (Hofmann, Asmundson, & Beck, 2013). CBT's concept is based on the significance of false ideas and mindsets, improper information processing, and unhelpful behavior as the risk factors for depression (Butler et al., 2006). Cognitive-behavioral approaches

are therefore presented and practiced during treatment sessions with classwork to internalize the new behavior (Jurinec & Schienle, 2020). Because most of these sessions are conversation based, the CBT approach is suitable for this research. While keeping track of the patient's assignments and progress, the chatbot will record their dialogues. The assignments and progress reports help them modify their views and thinking through time. Figure 1.2 depicts the efficacy of CBT according to Kaur and Whalley, 2020. This study compared CBT with more common treatment methods such as medications or psychotherapy. As the figure presents, CBT shows better responses compare to other treatments when it comes to disorders such as depression and anxiety.

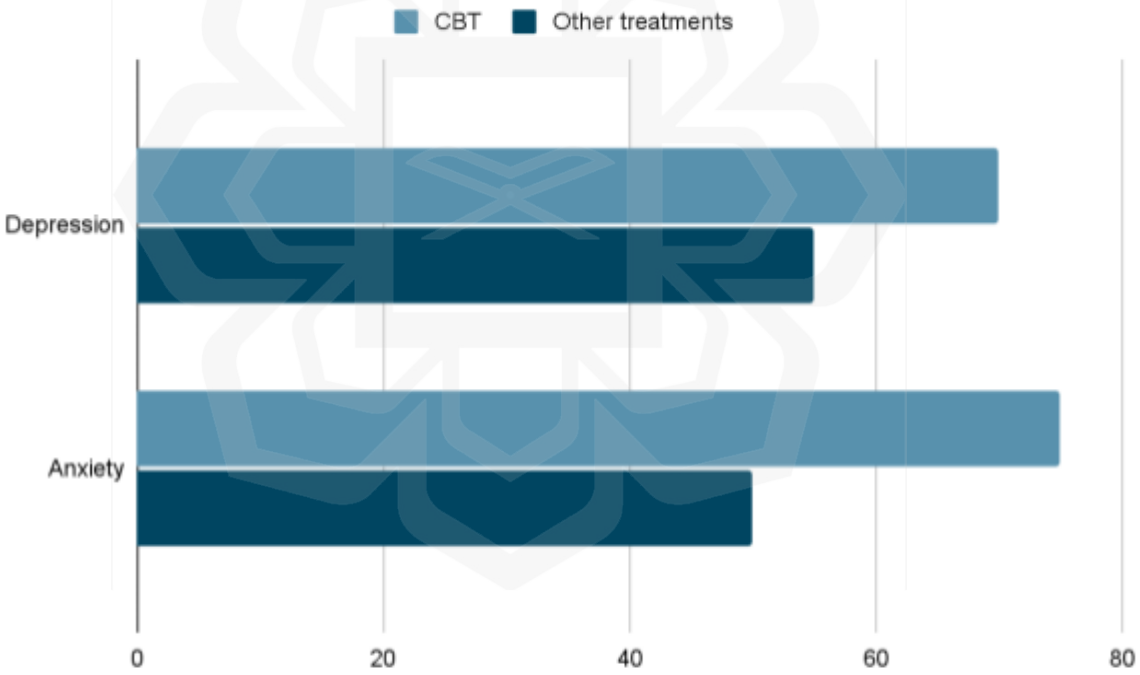


Figure 1.2 The effectiveness of CBT compared to other treatments

Spoken dialog systems are recently finding their way into all smart devices and gadgets. It provided user-friendly, efficient, and human-like communication for the

users. These technologies are implemented in education, government, business, and entertainment industries. Nevertheless, they are yet to prove their benefits, particularly in the mental health sector. In the world of Virtual Personal Assistants (VPAs), there are different methods. Every company has its preferred method and implementation. For example, Google Assistant uses Deep Neural Networks (DNN) that generally focus on the main components of dialog systems (Këpuska & Bohouta, 2017). On the other hand, Amazon is benefiting from Automatic Speech Recognition (ASR) methods and Natural Language Understanding (NLU), as mentioned on their website.

It is expected that during a typical conversation, each participant takes a turn to talk. The same is anticipated of a spoken dialog system, a natural and efficient interaction. There are six main components in each spoken dialog system, as presented in Figure 1.3.

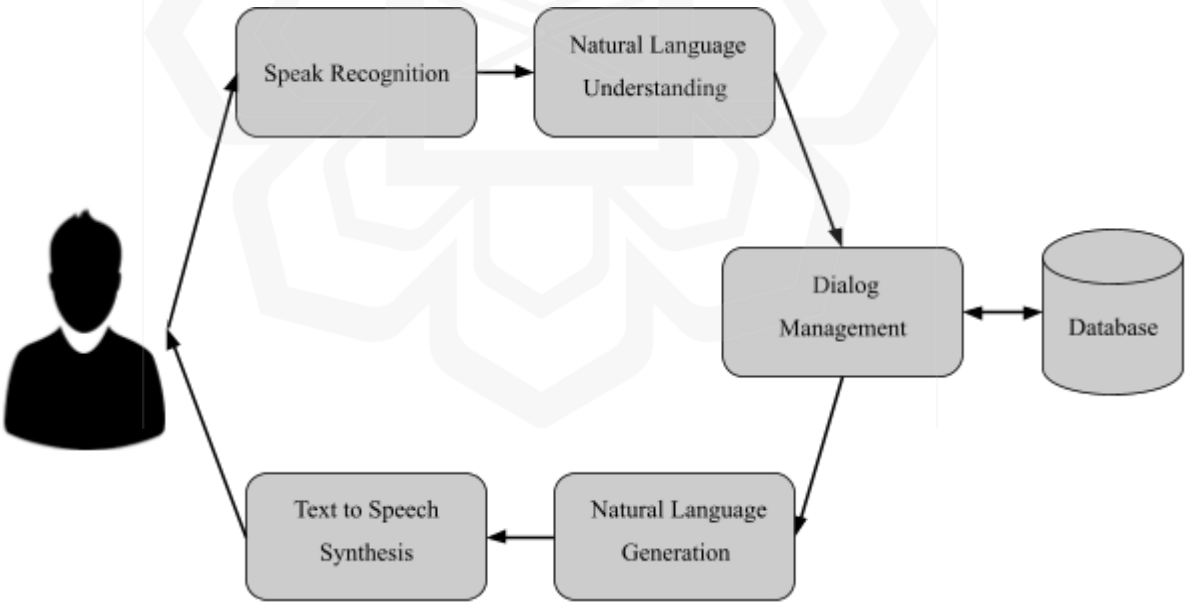


Figure 1.3 The Structure of Dialogue System

Generative conversational chatbots are designed to have a conversation with users. Generative chatbots' most attractive and unique feature is that they improve over time by obtaining past interactions. There are two approaches to designing conversational models, rule-based and machine-based. Rule-based interactions are based on predefined rules, while machine-based is learning and improving over time by utilizing deep-learning techniques. Generative models fall in the machine-based category. While improving based on the question and past interactions makes them more innovative, it is also more prone to error. Training with larger datasets can help improve their accuracy (Varghese & Pillai, 2018).

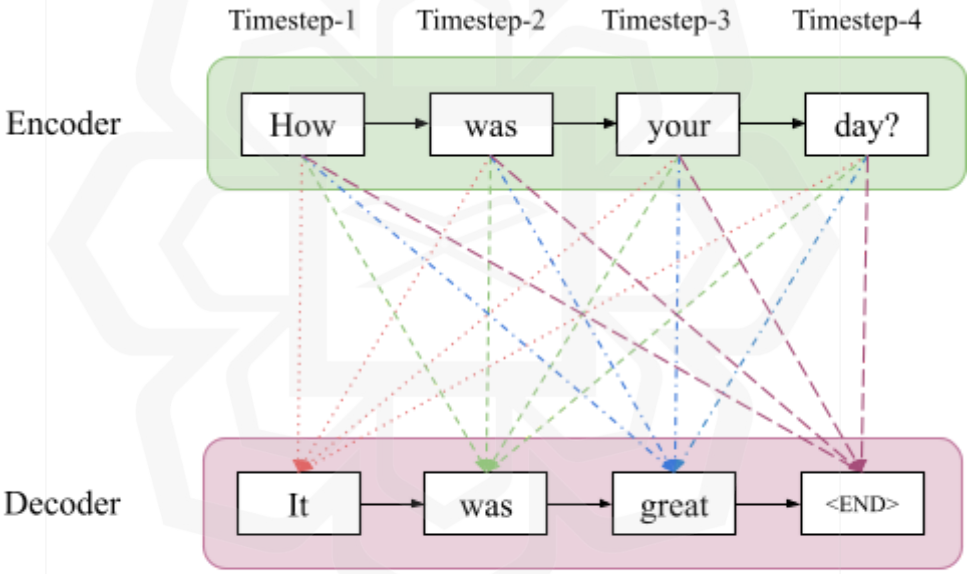


Figure 1.4 SEQ2SEQ model

Following are some of the classifications of generative-based models.

- SEQ2SEQ model learning strategies are Intermittent Neural Networks, DNN, and Convolutional Neural Networks (CNN) (Dahl, Sainath, & Hinton, 2013), figure 1.4.

- Long Short-Term Memory (LSTM) classifier uses feedback connections and processes entire sequences of data (Sheikh, Tiwari, & Singhal, 2019).

1.2 PROBLEM STATEMENT

Everyone has a right to posses good mental health. Without mental health, one cannot be considered physically well. Good mental health helps deal with the stress of life, affects physical health, and allows for building and maintaining strong human relationships. It also has an impact on physical health. People with mental health issues may make meaningful contributions to society and their local communities while feeling content and happy. Nearly everyone encounters traumatic events, abusive situations, family issues, hereditary issues, and lifestyle choices at some time in their lives. That is why it is necessary to have access to psychological tools and techniques to deal with these issues. The only significant therapeutic choices are currently self-help groups, medical treatments, and psychotherapy.

Unfortunately, not everyone around the globe has access to or can afford to attend a psychotherapy session at this time. Therefore, the number of professional therapists is limited. Additionally, although therapists are taught to be objective and fair, it is not easy for people to overcome their prejudices and preferences. Sometimes, machines are much better from this perspective. Even though there are several chatbots available, neither of them can have a vocal conversation with the patients and mostly are quiz-like. This observation means they ask the user to select from the available answers rather than share their opinions. These bots are also limited because they do not consider the patient's history, making the conversation sound robotic and repetitive. A meaningful vocal conversation with a generative chatbot that can