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

....
Aisha Hassan Abdalla Hashim
Supervisor

....
Rashidah Funke Olanrewaju
Co-Supervisor

ASST. PROF. DR. AKEEM OLOWOLAYEMO
Assistant Professor
Department of Computer Science
Kulliyyah of Information and Communication Technology
International Islamic University Malaysia

....
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.

No part of this unpublished research may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without prior written permission of the copyright holder except as provided below

1. Any material contained in or derived from this unpublished research may only be used by others in their writing with due acknowledgement.

2. IIUM or its library will have the right to make and transmit copies (print or electronic) for institutional and academic purpose.

3. The IIUM library will have the right to make, store in a retrieval system and supply copies of this unpublished research if requested by other universities and research libraries.

By signing this form, I acknowledged that I have read and understand the IIUM Intellectual Property Right and Commercialization policy.

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

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.

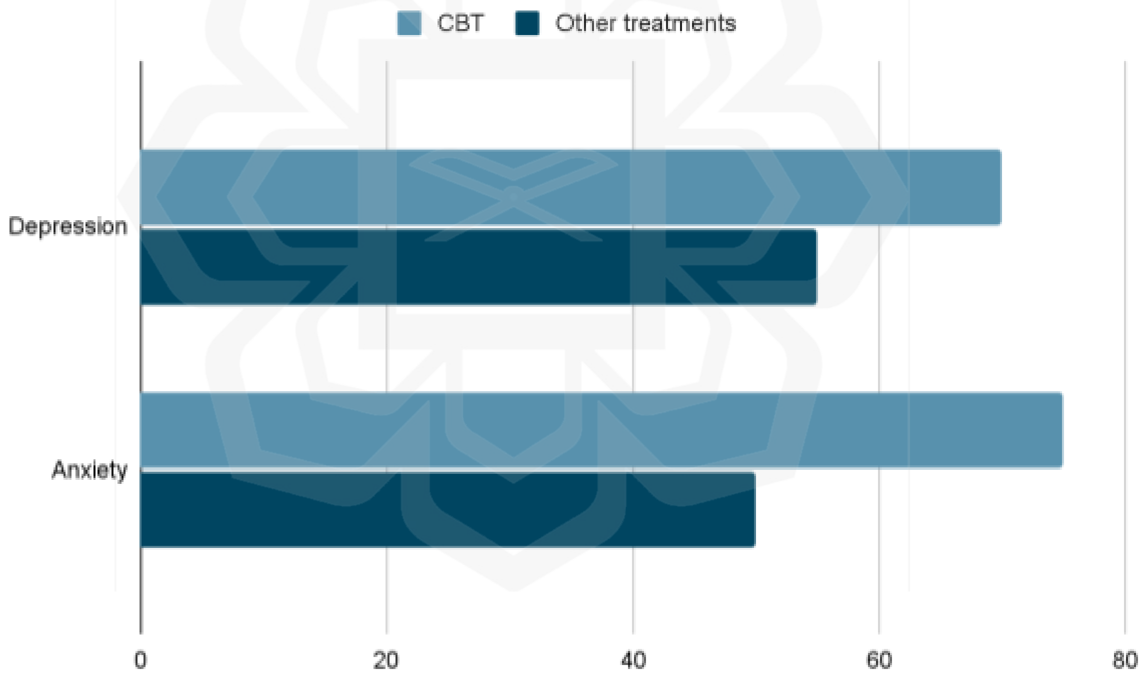

Figure 1.2 The effectiveness of CBT compared to other treatments

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.

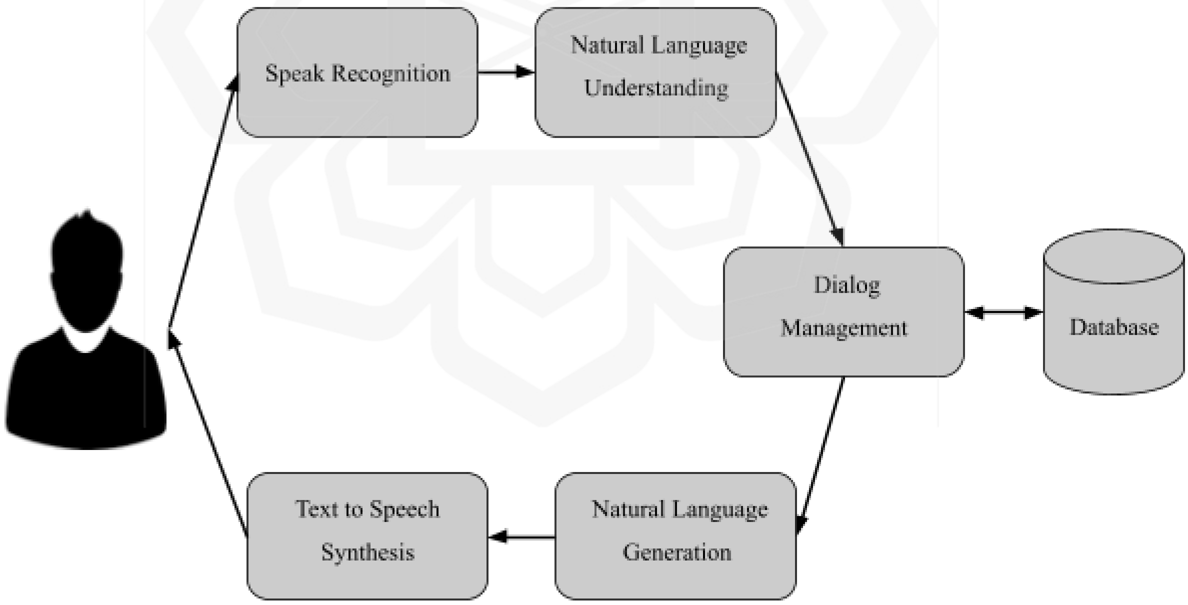

Figure 1.3 The Structure of Dialogue System

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).

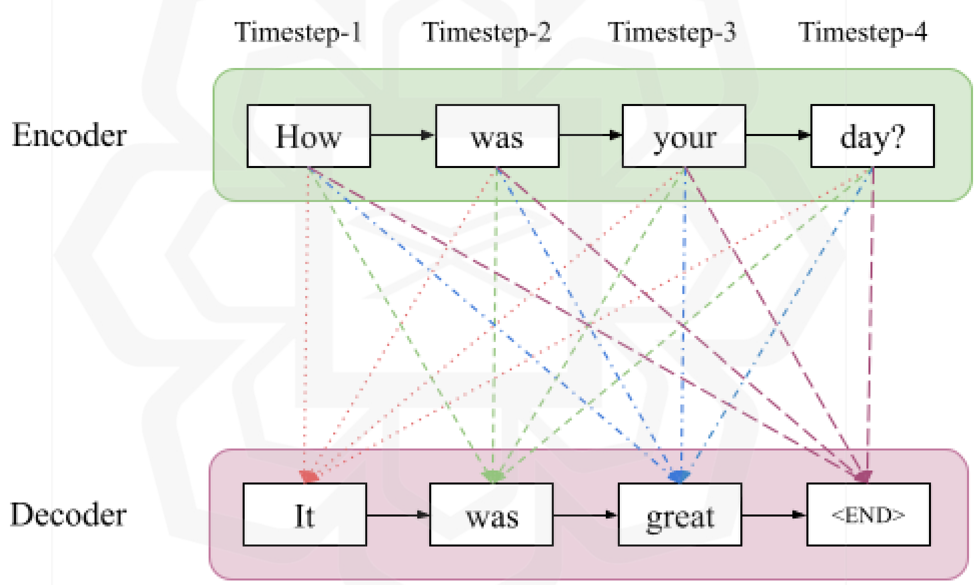

Figure 1.4 SEQ2SEQ model

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 remember and relate the previous conversations helps patients feel like they are talking to a human. This system can build more trust, resulting in a comfortable conversation flow. This thesis focuses on designing a set of algorithms that allows patients to have meaningful human-like conversations with an intelligent therapist bot that provides cost-effective approaches to improve their depression using CBT techniques.

1.3 RESEARCH OBJECTIVES

This thesis aims to design a set of algorithms that allows users to access an affordable cognitive therapist chatbot that provides expert advice on mental health.

1. To design a set of algorithms that provides vocal generative conversations
2. To develop a user-friendly software application and make it accessible to psychotherapy patients
3. To evaluate and benchmark the performance of the proposed set of algorithms.

1.4 RESEARCH METHODOLOGY

This section discusses the approaches and tools selected to design the Generative Interactive Psychotherapy Expert (GIPE) bot. Figure 1.5 presents the architect of this design. The design is divided into five sections, Bot User Experience (UX), Bot cognitions and intelligence, Data Extract, Transform, and Load (ETL), Raw Data, and finally Logging. The user's response would intuition a request via text or voice. The request is then processed by Bot UX and sent to Bot Cognition as a quarry. Bot's cognitions and intelligence are trained and updated based on the raw and processed data. After the request is processed by the Bot's intelligence, the result is sent to the Bot's UX to be processed as a reposed to the user via text and voice. This whole conversion is recorded in the conversions log to be used for improving the Bot's intelligence and intent.

Algorithm 1 presents the steps of the design of the GIPE bot. The algorithm is divided into five sections and each section use and implements different technology and techniques as presented below.

```
Algorithm 1: Design of GIPE bot
-----------------------------------------------------------------
1: Create a patient-therapist (PT) knowledge base
   ➢ Explore and gather open-sourced reliable data.
   ➢ Process the collected data based on the format.

2: Build a dialogue generation model (DGM).
   ➢ Implementing transfer learning model architecture.
   ➢ Implying HuggingFace, ParlAI, PyTorch, and GPT-2
     models.

3: Evaluated DGM.
   ➢ Automatic evaluation: Hits@1, Perplexity (ppl), and F1
     metrics.
   ➢ Human evaluation: IIUM students.

4: Train DGM with the PT knowledge base (GIPE)

5: Developing GIPE bot.
   ➢ Connect GIPE to Dialog Flow
```

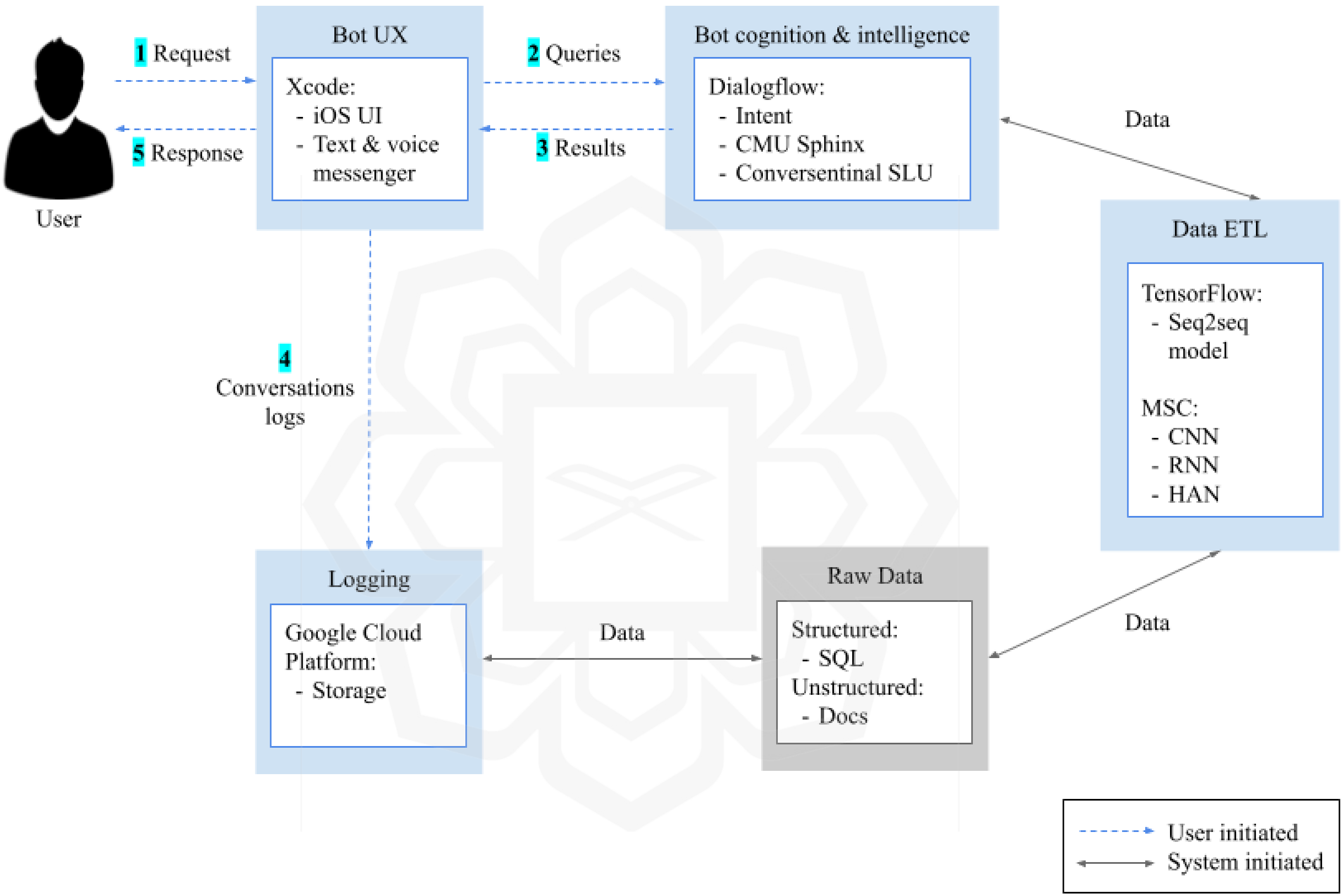

Figure 1.5 The GIPE bot’s architecture

Figure 1.6 shows the flowchart of the method process of this thesis. The steps of a certain technique or process are graphically represented in a methodology flowchart. It's a kind of flowchart that shows the many steps, jobs, and activities performed in the technique or process, in addition to the connections and dependencies among them. A methodology flowchart's goal is to give stakeholders a visual picture of the approach or process so they can comprehend and evaluate the many steps and duties involved.

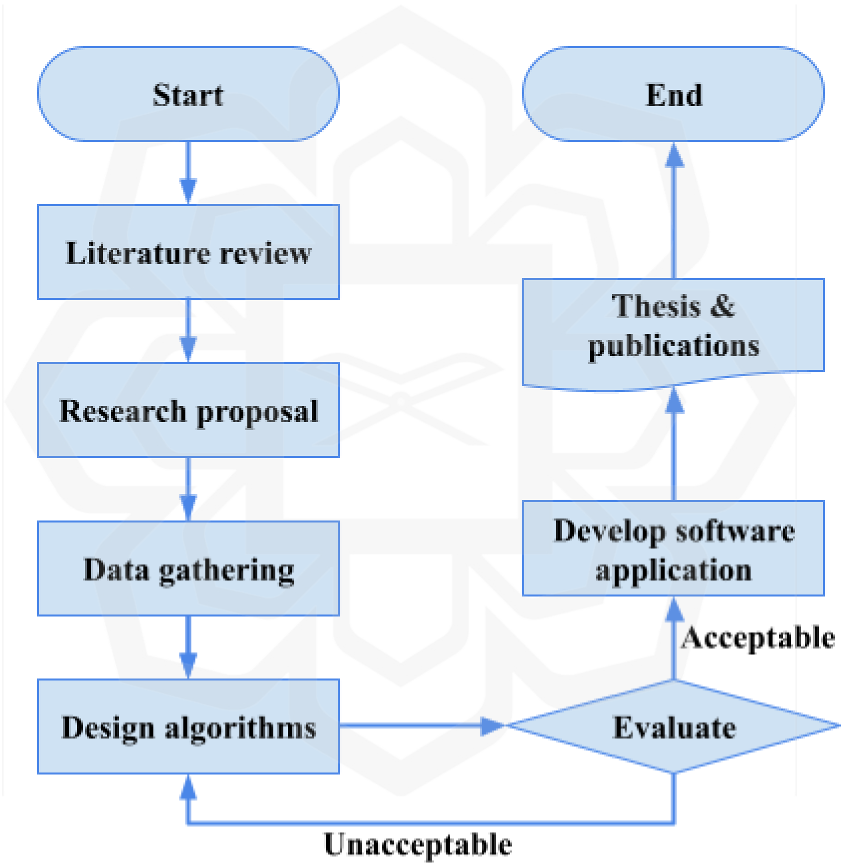

Figure 1.6 Methodology Flowchart

1.5 RESEARCH SCOPE

This research focuses on the English language dialogs only. Text and voice message conversations are supported, while video conversation is out of the scope of this research. IIUM students participated in a human evaluation of dialogue generation models.

1.6 THESIS ORGANIZATION

Chapter one introduces and provides an overview of the importance of mental health and describes the current situation of mental health facilities. After that problem statement, research objectives, and research scope are provided.

The second chapter explores a wide variety of research and applications. One section is dedicated to the most popular consultation chatbots and applications. The next section explores the technologies, their backgrounds, and their understanding of their impacts and implementations.

The third chapter is about the methodologies implemented in this study:

1. The model design and implementations of mutual dialogue generation are explored in detail.
2. Data acquisition and selections are referenced.
3. The validation techniques used are explained and explored.

Chapter four concentrates on the results and the findings. First, the data gathered is analyzed. Then the model designed is evaluated against the benchmarks. Finally, the chatbot prototype is developed.

Finally, the study's fifth chapter finishes the investigation and presents a summary of its successes, contributions, and effects. Overall, the study's conclusion chapter provides an essential review and opportunity for reflection on the work that has been done. The key conclusions, contributions, and implications of the study are succinctly summarised, along with any limitations or potential topics for further investigation. To meet the academic rigor of the study, this researcher tries to convey its outcomes in a manner that is both intelligible and accessible to a wide audience.

CHAPTER TWO

LITERATURE REVIEW

2.1 OVERVIEW

To help one's mental health, a variety of techniques are available. This chapter compares the background technologies and techniques and the prices of available treatment and consulting sessions, both in-person and online. After that, we will discuss about some of the current robots and chatbots on the market.

According to a personal website article from October 2018, even though the cost of medical care in a Malaysian government hospital could be as low as RM15 per session, patients must have a recommendation letter from a general practitioner. They will instead receive advice from a medical doctor rather than a psychotherapist. This method is not a long-term fix for a significant issue. The cost of seeing a private therapist ranges from RM200 to RM500 for each session. Even though this is a good option, not everyone can afford it.

Online apps make psychotherapy sessions more accessible and cheap. Patients would communicate with a licensed therapist throughout the sessions via text or phone. These programs include Betterhelp, Figure 2.1, and Talkspace for instance. According to their websites, in January 2021, Betterhelp and Talkspace charged between \$60 and \$80 weekly (Fader, 2017) and \$35 and \$80, respectively. Although somewhat cheaper, therapy is still unaffordable for some people and depends on their ability to get it.

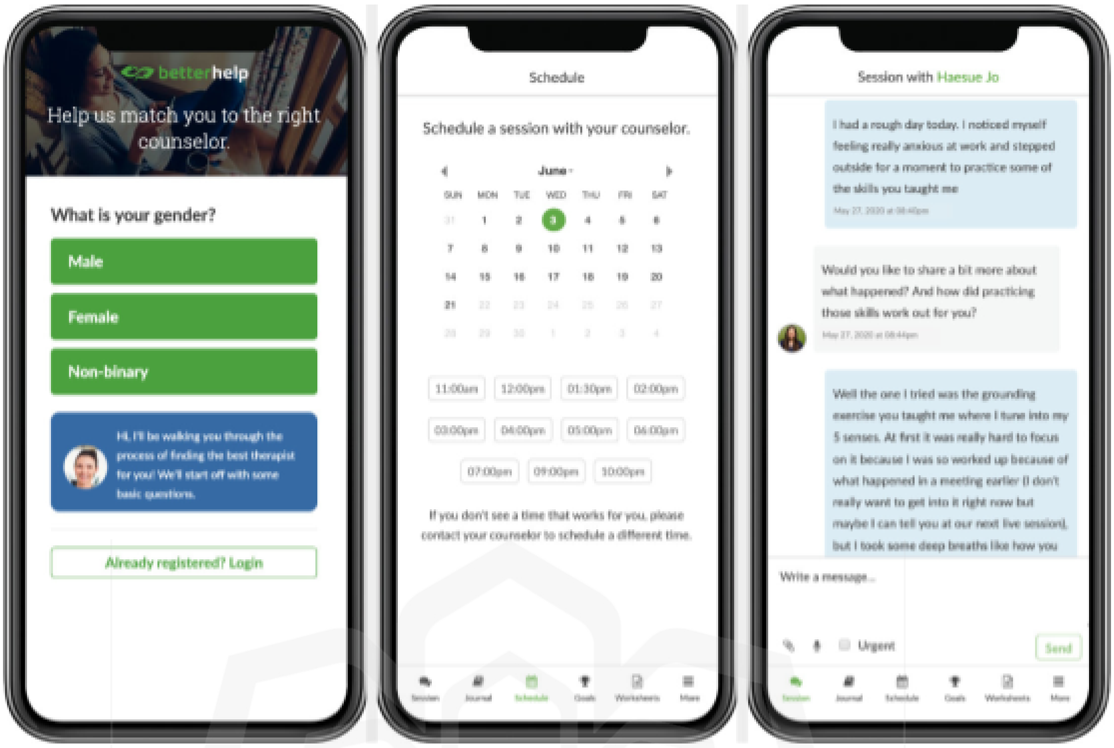

Figure 2.1 Captured from Betterhelp mobile application

2.2 BACKGROUND

Creating data-driven models for natural language conversation production has recently received extensive attention from the academic and business areas. Traditional approaches frequently rely on manually created rules and templates to find appropriate replies (Schatzmann et al., 2006). However, this process is time-consuming, and domain transfer cannot be accomplished effectively (Shang, Lu, & Li, 2015). Furthermore, the retrieval technique-based conversation system also needs a lot of pre-prepared responses to the queries (Hu et al., 2014), and such systems cannot produce new utterances in response to unknowable inquiries (Song et al., 2016).

As forums and microblogs have grown in popularity online, many discussion corpses have been created. The translation model produces the proper reaction to use these enormous bodies fully. For instance, a statistically phrase-based machine translation model was extinguished by Ritter et al. for this job (Ritter, Cherry, & Dolan, 2011). Later, the SEQ2SEQ model is presented and used in the domain of dialogue modeling and may produce a dialogue answer in a person-to-person discussion thanks to the growth of RNN in translation software (Vinyals & Le, 2015). However, the discussion information at the start of a dialogue is frequently disregarded when producing the answer because of the training signal's lengthy propagation difficulty (Serban et al., 2016). As a result, the SEQ2SEQ model has a propensity to provide general, uninformative replies like "I do not know." To properly model discourse creation, (Serban et al., 2016) created a Hierarchical Recurrent Encoder-Decoder (HRED). Each syllable is encoded into an utterance matrix by the word-level RNN. The context-level encoders retain track of previous utterances by iteratively processing the utterance vector.

Open-source informal conversation systems have made significant strides toward simulating human-like responses due to advancements in neural models and the availability of large datasets. Nevertheless, creating customized chatbots that can offer meaningful discussions and earn user confidence remains a significant challenge (Song et al., 2019). In addition, because training dialogues originate from diverse speakers, they frequently lack uniform personality features (Zhang et al., 2018). Furthermore, present chit-chat systems provide ambiguous replies (Li et al., 2016-a).

Methods such as customized reward shaping are implemented to increase the engagement of informal systems, limit generic replies (Li et al., 2016-a), and model the participants with independent factors were proposed (Li et al., 2016-b). Zhang et al., 2018, provided a more efficient technique that presets personas to chit-chat structure with the new PERSONA-CHAT dataset, figure 2.2. Two interlocutors meet for the first time in this model and have a chat to get to know one another. The personas of both interlocutors are precisely stated using multiple profile phrases in

PERSONACHAT, making it easier to teach bots with adjustable and durable personalities. For this study, the model was tested and the presented conversation in Figure 2.2 was generated.

Mazaré et al., 2018, used new Reddit data to train the PERSONA-CHAT model, while Wolf et al., 2019, fine-tuned the pre-trained language model. Although both studies show impressive outcomes, they concentrate on copying the tone of human-like responses, leaving the features of directly modeling interlocutors' awareness unexplored. Instead, this study approaches modeling from the standpoint of awareness.

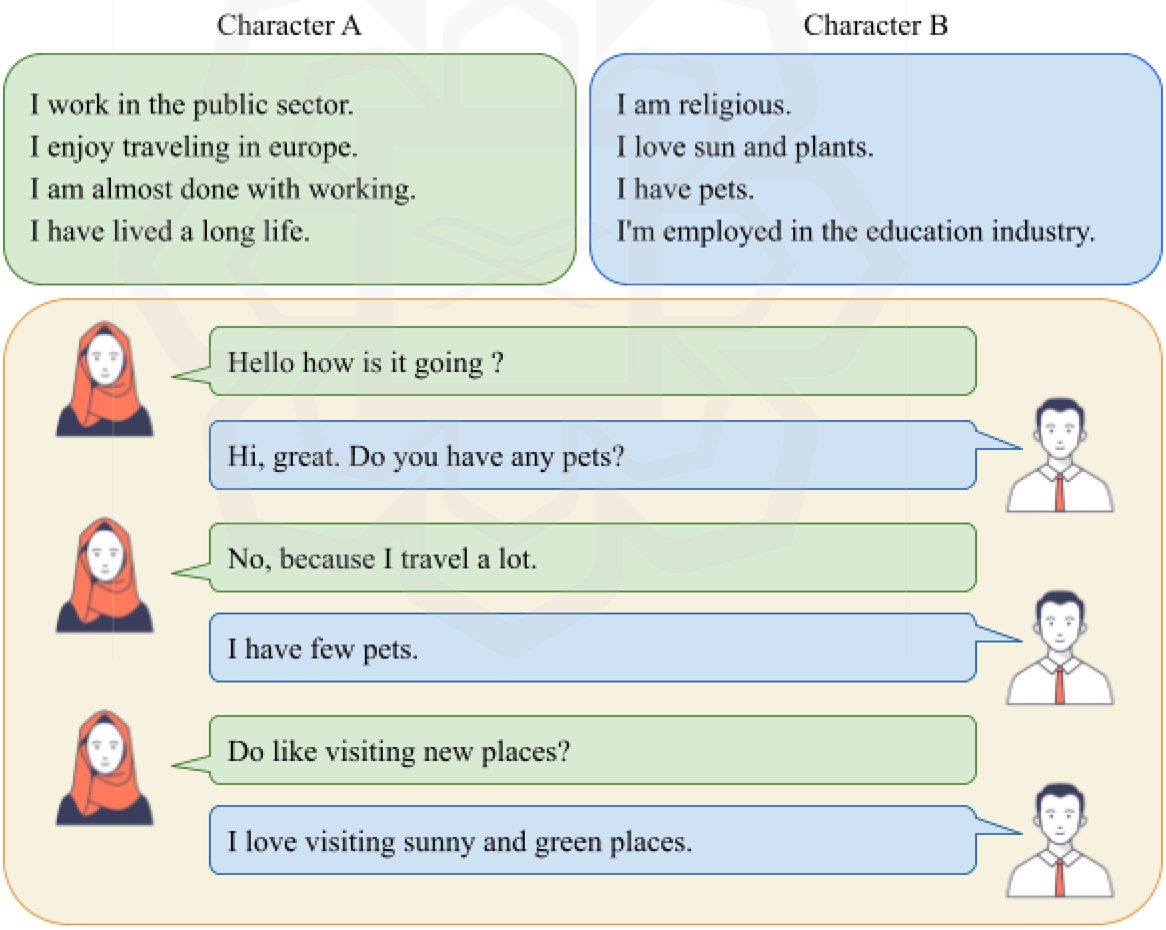

Figure 2.2 Conversation between two bots

According to a cognitive science study by Hasson et al., 2012, effective communication develops identical activation maps in both interlocutors' brains. This outcome implies that interlocutors' familiarity is an essential indicator for high-quality informal chitchat. Therefore, P^2 Bot is developed using a transmitter-receiver structure to specifically model interlocutors' awareness emphasizing mutual persona perception. That is more appropriate to define the events of exchanging information that allows interlocutors to get to know one another (Liu et al., 2020).

Supervised learning on task-specific datasets is commonly used to tackle natural language processing problems like question answering and language translation. When trained on a new dataset of millions of online articles called WebText, GPT-2 revealed that language models began to learn these tasks without direct supervision (Radford et al., 2019). The language model's answers on the CoQA dataset achieve 55 F1 when trained on a document and questions, equal to or outperforming the accuracy of three out of four baseline systems without needing the 127,000+ training samples. The model's samples exhibit these advances and include well-organized paragraphs of text. This evidence pointed to a possible direction for developing language processing algorithms that learn to accomplish tasks from natural examples.

This research uses semi-supervised training using GPT-2 and self-play fine-tuning by rewards signals indicating mutual persona perception to train the model for a customized conversational generation. Analyses using the PERSONACHAT dataset indicate the method's effectiveness in both automatic measures and human rating terms.

2.2.1 Deep Learning (DL)

Intelligent organisms can digest information to be used for future decision-making. Any approach that enables computers to replicate people's behavior is known as AI. AI is a branch of science. The capacity for learning or teaching without explicit programming is known as Machine Learning (ML). A division in the field of AI is ML. For example, DL uses Neural Networks (NN) to learn how to complete a job by extracting patterns from unprocessed data. An ML approach is called DL.

The universal approximator feature of a single-layer NN is one of the theoretical conclusions supporting NN usage. As its function approximators that are most frequently referenced. Due to the restricted processing resources, there are several practical difficulties when training a single-layer NN with a substantial number of neurons. A NN with many parameters cannot be trained due to a shortage of training data. Although there are just a few theoretical results for deep NN, they are claimed to be a more realistic approach to function approximators (Montufar et al., 2014).

Hand-engineered features are time-consuming, fragile, and not scalable in practice. Therefore DL is valuable. Three following factors account for the resurgence of DL.

1. Big data: More extensive datasets are simpler to obtain and store.
2. Hardware: Graphics Processing Units (GPUs) have a high degree of parallelizability, which enables very high processing powers.
3. Software: Significant technological advances, new models, and widely accessible toolboxes like Tensorflow.

2.2.2 Transfer Learning

The widely used sequence transduction models are based on elaborate convolutional or recurrent neural networks with both an encoder and a decoder. The top-performing models additionally use an attention mechanism to link the encoder and decoder.

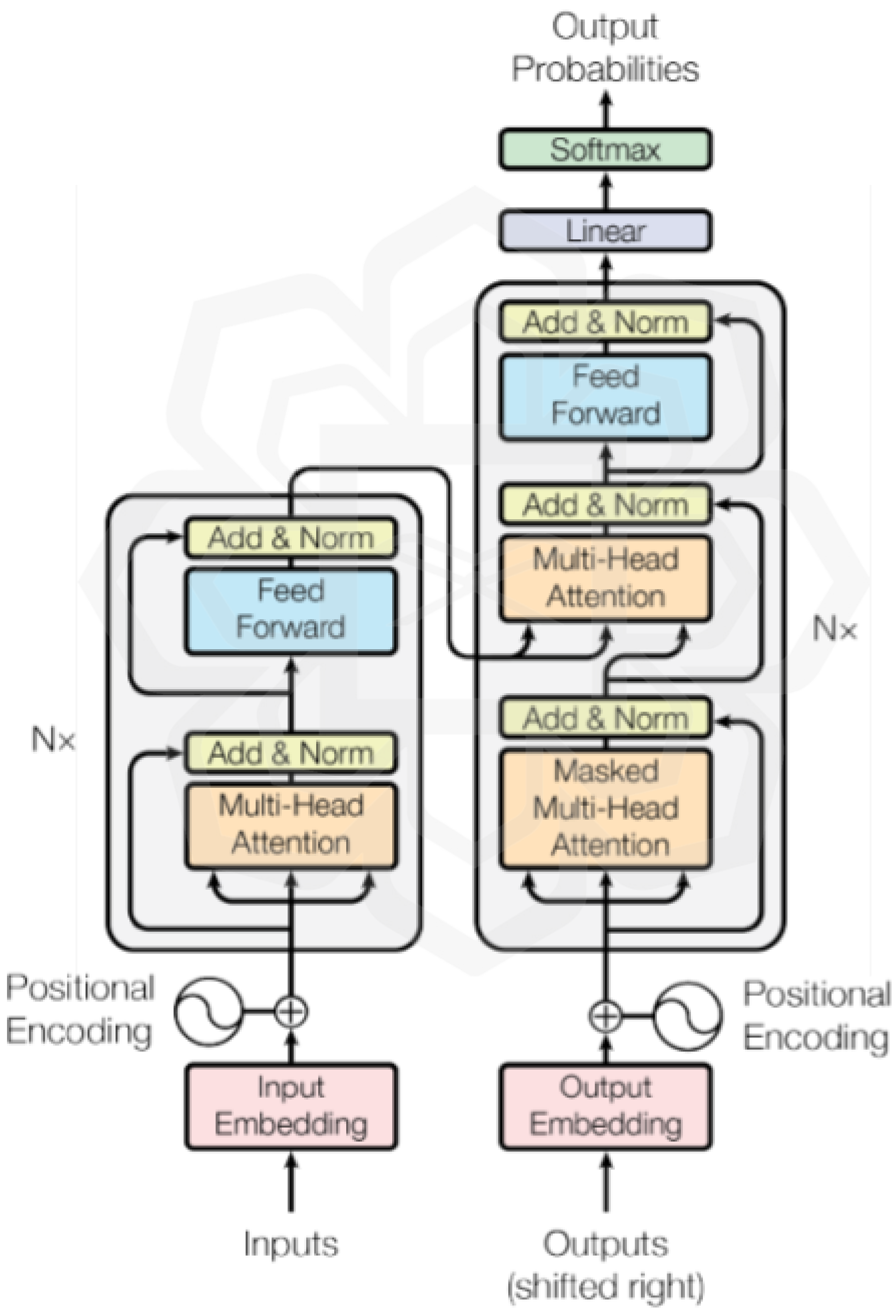

Figure 2.3 The transformer model architecture (Vaswani et al., 2017)

However, Vaswani et al., 2017, suggested a brand-new, straightforward network design called the Transformer that ultimately does away without recurrence and convolutions in favor of attention methods, figure 2.3.

A query is a collection of key-value pairs in terms of vectors. An attention function maps a query and an output vector to one another. The result is calculated as

🤖**character**:

I help around with bookkeeping and tours.

Sometimes I volunteer at an urban farm.

I am vegan.

I work at the grocery store.

🤖 PERSONA-CHAT bot
👨 User

👨 Hello there how are you?

🤖 I am doing well. How are you?

👨 I am great. Thanks. Do you like boating?

🤖 I like it when I can get away from my job at the grocery store

👨 It's my favorite activity outside of being a doctor. Do you like beaches?

🤖 I give deep sea fishing tours at the beach sometimes

👨 I volunteer at a farm, do you like animals?

🤖 I am vegan so I love animals. same here with the farm volunteering

👨 I too am a vegan, how long for you?

🤖 Just 2 years. my boss at my bookkeeping job got me started

👨 That is nice. My boss at the grocery store is nice too

🤖 Sounds like we have a lot in common. How do you feel about urban farms?

👨 I work at an urban farm, I love them. Do you?

🤖 Urban farms are the best. There is one on top of my apartment building.

👨 How lovely, mine is 2 miles up the road.

🤖 Is it all organic? I found that sometimes chemicals are necessary.

Figure 2.4 Conversation between PERSONA-CHAT bot and user a weighted sum of the values, with each value's weight determined by the query's compatibility function with its associated key.

Wolf et al., 2019 introduced a new approach to generative data-driven dialogue systems called TransferTransfo, a combination of Transfer learning based on a training scheme and a high-capacity Transformer model. A multi-task goal that includes some unsupervised prediction tasks is used for fine-tuning. The final, fine-tuned model outperformed some of the most advanced end-to-end conversational models available, including memory-augmented seq2seq and information-retrieval models.

Examining the effectiveness of open-domain conversation agents is a fascinating task in the Conversational Intelligence Challenge 21. (ConvAI2). ConvAI2 is built on the PERSONA-CHAT dataset (Zhang et al., 2018). In this crowd-sourced conversation dataset, each speaker was asked to conditional their utterances on a preset profile composed of a few sentences describing a personality. In addition, paired workers requested that they converse freely and get to know one another during the talk. For example, figure 2.4 is a conversion between chatbot with character and user. The conversation was generated by running the ConvAI2 model for this research.

2.2.3 Pre-trained Transformer-2

Generative OpenAI developed the open-source AI known as GPT-2 in February 2019 (Radford et al., 2019). GPT-2 is capable of translating text, providing replies to inquiries, summarizing sections, and producing text outcomes on a level that, while occasionally unrecognizable from humans, may occasionally become monotonous or absurd when producing lengthy passages. It is a general-purpose learner; none of these activities were mainly taught to it, and its capacity to carry them out is an extension of its general capacity to precisely synthesize the following item in any given sequence. As a "direct scale-up" of OpenAI's 2018 GPT model, GPT-2 was developed with a ten-fold increase in both the number of parameters and the size of the training dataset (Radford et al., 2018).

The GPT design leverages attention in place of earlier recurrence- and convolution-based architectures to create a DNN, especially a transformer model. The model may focus on input text sections that it thinks to be the most pertinent, thanks to attention processes (Radford et al., 2019). This model beats earlier benchmarks for RNN and LSTM-based models and enables far more parallelization.

They were curious to understand how WebText LM handled zero-shot domain transfers for language modeling, which is their core training area. They could test their model on any language model benchmark since it operates at the byte level and does not call for lossy pre-processing or tokenization. Results from language modeling datasets are frequently expressed as a number that is a scaled or exponentiated representation of the average negative log probability per canonical prediction unit, typically a character, byte, or word.

The Children's Book Test (CBT test) was developed by Hill et al. (2015) to assess how well LMs perform in various word categories. The CBT test reports accuracy on a cloze exam that was created automatically. The GPT-2 scores 93.3 percent for common nouns and 89.1 percent on named entities, which are new records. The capacity of systems to represent long-range relationships in the text is tested using the LAMBADA dataset (Paperno et al., 2016). Predicting the last word of phrases that need at least 50 context tokens for a person to anticipate correctly is the challenge at hand. Examining GPT-2's mistakes revealed that while most predictions were correct for the majority of the phrase, many were incorrect for the concluding words.

Evaluating how frequently a language model produces the proper response to factoid-style queries is one possible technique to test the information inside it. However, due to the dearth of high-quality assessment datasets, earlier demonstrations of similar behavior in neural systems, where all information is stored in parameters, including a Neural Conversational Model (Vinyals & Le, 2015), revealed qualitative results. The recently released Natural Questions dataset (Kwiatkowski et al., 2019) is an exciting resource to test this more statistically. Comparable to translation, the language model's context is seeded with illustrative question-answer pairs to infer the dataset's short response style.

When measured using the exact match metric frequently employed on reading and comprehension datasets like SQUAD, GPT-2 appropriately responds to 4.1 percent of questions. As a benchmark, the most miniature model gives the most frequent response for each question type, with an accuracy of 1.0 percent, below the most miniature model. GPT-2 provides 5.3 times more accurate responses, indicating that model ability has played a significant role in the neural systems' subpar performance on this type of task up to this point. The probability GPT-2 provides to the answers it generates is accurately calibrated, and on the 1% of inquiries it is most certainly in, GPT-2 has had an accuracy of 63.1 percent. The 30 most certain responses produced by GPT-2 for questions from the development set. The GPT-2's performance is still significantly below that of open-domain question-answering systems that combine information gathering with extractive document question-answering in a 30 to 50% range (Alberti et al., 2019).

2.2.4 Maximum Likelihood Estimation

Maximum Likelihood Estimation (MLE) is a statistical technique for estimating the parameters of a probability distribution that has been assumed given some observed data. This result is accomplished by maximizing a likelihood function to make the observed data as probable as possible given the assumed statistical model. The MLE is the location in the parameter space where the likelihood function is maximized. Maximum likelihood is a popular approach for making statistical inferences since its rationale is clear and adaptable.

The derivative test for determining maxima can be used if the probability function is differentiable. The ordinary least squares estimator, for example, maximizes the likelihood of the linear regression model, allowing the first-order requirements of the likelihood function to be explicitly solved in some circumstances. However, in the majority of cases, it will be essential to use numerical techniques to determine the probability function's maximum.

2.2.5 Reinforcement Learning

The field of ML, known as reinforcement learning (RL), studies how intelligent agents should behave in a given environment to maximize the concept of cumulative reward (Sutton et al., 1999). Besides supervised and unsupervised learning, reinforcement learning is one of the three fundamental ML paradigms.

RL differs from supervised learning in that it does not need the presence of labeled input and output pairings or the explicit correction of suboptimal behaviors. Instead, the emphasis is on striking a balance between exploitation and exploration. The benefits of supervised and RL algorithms may be combined with partially supervised RL algorithms.

2.2.6 Word Embedding

Word embedding is a term used in NLP to describe how words are represented for text analysis. Typically, this representation takes the shape of a real-valued vector that captures the word's definition, with the expectation that words close to one another in the vector space will have similar meanings. Word embeddings may be created by mapping vocabulary words or phrases to vectors of actual figures using a variety of language modeling and feature-learning approaches. NN dimensionality drops on the word co-occurrence matrix, probabilistic models, and the resolvable knowledge base method. Some techniques used to create this mapping are also explicit representations of word context.

Early methods of semantic role labeling (Toutanova, Haghighi, & Manning, 2008) concentrated on building extensive collections of linguistic characteristics as input to a linear model, frequently paired with complicated restricted inference, for example, with an ILP (Punyakanok Roth & Yih, 2008). However, a sophisticated dynamic program for accurate inference may be used to impose restrictions more successfully, as demonstrated by Täckström, Ganchev, and Das (2015). Sutton and McCallum (2005) modeled semantic role labeling and syntactic parsing, and Lewis, He, and Zettlemoyerl (2015) modeled semantic role labeling (SRL) and CCG parsing.

Syntax has been included in several studies into neural models for SRL. Compared to models without syntax on CoNLL-2009, Roth, and Lapata (2016) add syntax by embedding dependency routes, while Marcheggiani and Titov (2017) encode syntax using a graph CNN over a projected syntax tree. While our method involves the whole parsing, these efforts can only include partial dependence routes between tokens. Furthermore, Marcheggiani and Titov, 2017, indicate that on non-domain data, where our method performs well, their model is not superior to syntax-free models.

A multi-task neural network model called Linguistically-Informed Self-Attention (LISA), which was introduced by Strubell et al. (2018), successfully includes extensive language data for semantic role categorization. LISA outperforms the state-of-the-art on two benchmark SRL information, including out-of-domain datasets. Furthermore, raw tokens may be the only input for LISA's syntax. It only needs to encode the sequence once to conduct parsing, predicate detection, and role labeling for all predicates.

One attention head is trained to focus on the syntactic parents of each token to include syntax. A strong syntactic parse may also be advantageously supplied at test time without having to retrain our SRL model.

2.3 RELATED RESEARCH

The apps that are primarily free and do not need an involved therapist to work are covered in the following sections. Table 2.1 contrasts their approaches, results, and gaps.

2.3.1 Catch It

Kohonen's SOM (Self Organizing Maps) algorithm is being used in an Android app developed in partnership with the Universities of Liverpool and Manchester to treat borderline personality disorder (Jayachandran & Shyamala, 2017). However, users

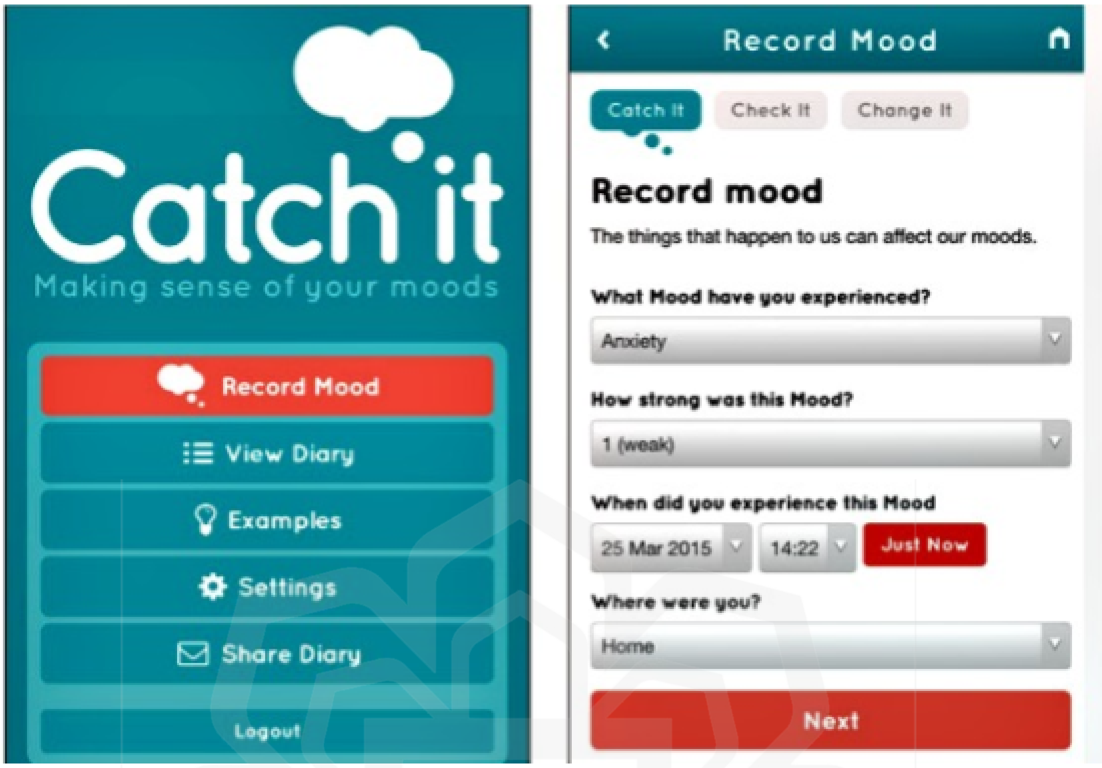

Figure 2.5 Captured from Catch It Mobile Application

find this software restrictive because it merely records their emotions and advises them to identify the triggers on their own, even though it is a first-of-its-kind application in the field of utilizing Cognitive Neuropsychology and Artificial Neural Networks. Additionally, the application's features primarily resemble forms that users must complete, which the developers refer to as "close-ended inquiries," as seen in Figure 2.5. The figure was captured from the mobile application on December 2020. As a result, the application is now ranked 3.8/5 by 141 people as of December 2020 on the Google Play Store.

2.3.2 Ryan the Companion Bot

A group of engineers at the University of Denver created Ryan, a socially-assistive robot, utilizing Program-R, a conversation management system that uses CBT to treat depression (Dino et al., 2019). Additionally, it examines the user's facial expressions and verbal answers. However, according to Ryan's therapy outcomes, it is currently not a practical option for therapists. Additionally, because it is a physical robot, a broad spectrum of consumers would be unable to operate it.

2.3.3 Youper

This software employs meditation, CBT, and commitment therapy to reduce stress and anxiety. Although the software has received positive ratings (4.9/5, particularly in the Apple App Store as of December 2021), the team has not published any studies or research to back up its effectiveness. This program is a chatbot that mostly lets users answer with a small number of choices. Therefore, the user is completing a multiple-choice quiz rather than having a conversation with the chatbot employing their own words. Technically speaking, Youper illustrates an expert system developed with therapist assistance, figure 2.6. The figure was capture from the mobile application.

2.3.4 Appsiety

This study could not test this software since it has not been released. According to the released study, it is primarily a platform that patients utilize while already seeing a therapist (Alves et al., 2019). They may then communicate with therapists more efficiently and offer more frequent updates. Although it is a helpful software for more consistent recording and communication between both patients and therapists, it is not a helpful tool in circumstances where there is no therapist accessible.

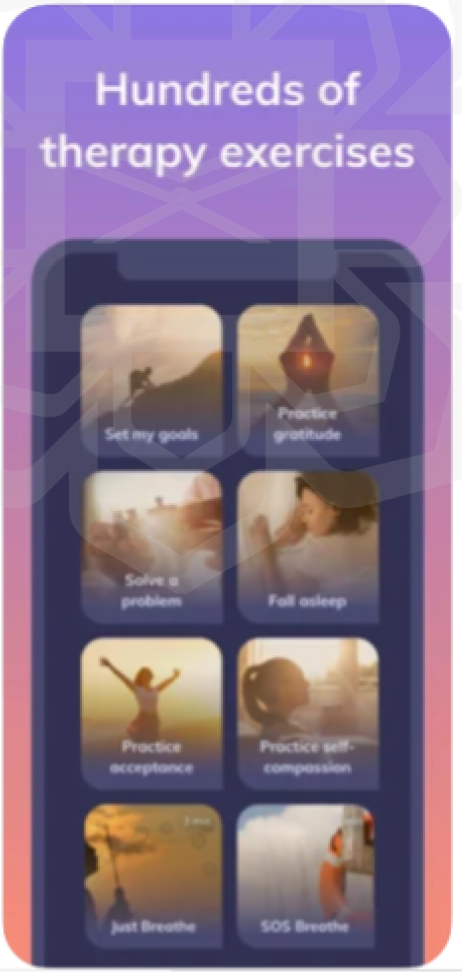

Figure 2.6 Captured from Youper Application

2.3.5 T-Bot

It is a therapeutic chatbot that aids in determining the intensity of depression and offers suggestions for ways to improve oneself (Sharma, Puri, & Rawat, 2018). Regardless, although it offers multiple tactics and aids users in determining their state of depression, it cannot resume a discussion as much as it would want by offering techniques. Essentially, it is the user's responsibility to select and employ their favorite strategies.

2.3.6 Replika

This well-known generative bot can hold a deep dialogue with users while recalling and remembering the conversions afterward (Mensio, Rizzo, & Morisio, 2018). It mimics the user's message and conversion patterns and gradually develops an equivalent speaking style. Although it technically works quite well as a conversational system and has no experience in mental health issues, it is not advised to use it in place of consulting specialists. Furthermore, relying on its replies might be risky in specific ways since it will figure out what the user loves to hear, which might even urge them to make a wrong choice because they like encouragement. In other words, the user would hear what they want to hear rather than what they need to hear.

2.3.7 COVID Coach

COVID Coach mobile application is developed by the United States Department of Veterans Affairs (VA) in 2020 to educate and update people on matters related to the COVID-19 virus (Allen, 2020). Figure 2.7 is captured from the Apple store. As of

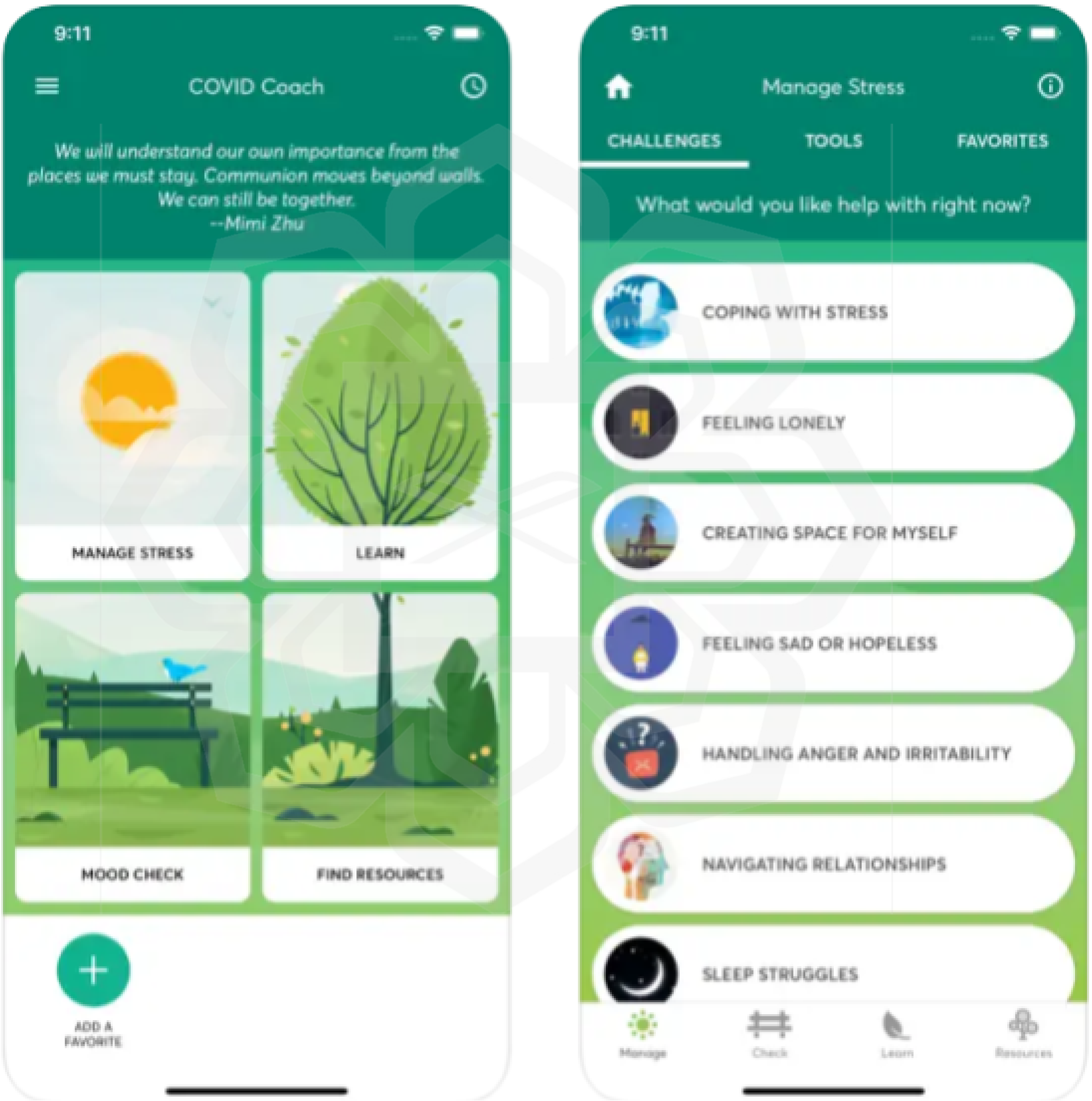

Figure 2.7 Captured from COVID Coach Application

April 2023, the application has 4.8 out of 5 stars in the Apple store. While the application has proven to be a valuable tool for self-education and awareness, there is no direct interaction between the patients and experts. Hence it cannot be much of a help for individual cases.

2.3.8 Naluri

Naluri is a Malaysian-based digital therapeutics solution published in 2017 (Lee, et al., 2019). The application provides multiple services covering both mental and physical health. Figure 2.8 is from their promotion in the Apple store with 3.7 out of 5

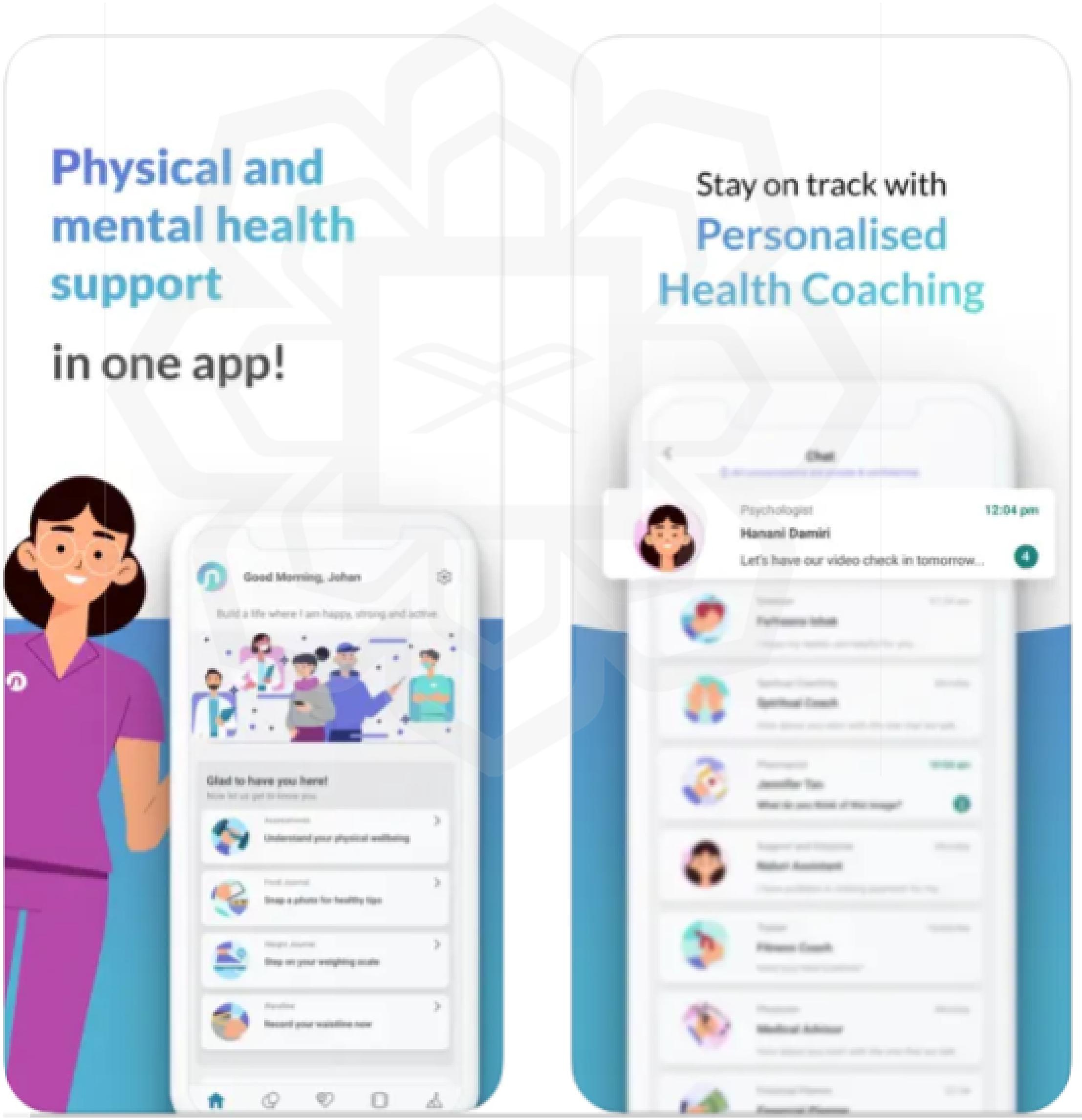

Figure 2.8 Captured from Naluri Application

stars as of March 2023. Patients are connected to psychologies and can have personalized coaching. While this application is achieving good results, still is relying on an actual therapist and their availability. There are as also some comments about how the application doesn't run smoothly which is impacting user experience.

Table 2.1 compares the available research works, implemented methods, findings, and gaps.

Table 2.1 Table of comparison

No	Research work	Methods	Findings	Gaps
1	Catch It, 2014 (Jayachan Dran & Shyamala , 2017)	- Application of the Kohonen SOM algorithm in CBT for the treatment of personality disorders - Artificial NN and cognitive neuropsychology	- Maintain a diary and preserve records. - assist users in locating triggers - One hundred forty-one people gave the app a rating of 3.8/5 on Google Play.	- Without flexibility, UI is merely a form to be filled out. - Based on user feedback, the bot's replies are nebulous and pointless.
2	Ryan the Compani on Bot, 2017 (Dino et al., 2019)	- Utilizes CBT and the Program-R conversation management system to treat depression. - Examines utterances and expressions on the face.	- Aged consumers' despair and dementia have improved. - Be interactive with users.	- Being corporeal, it is not generally available. - Ponders as a friend rather than a therapist. - At this point, primarily for entertainment.
3	Youper, 2016	- Utilizes mediation,	- Users' anxiety and tension levels	- No research was published to

		commitment therapy, and CBT. - An illustration of an expert system created with the assistance of therapists.	reportedly improved. - Receiving a 4.9/5 rating from 14.2K users in the Apple app store.	support its effectiveness. - Quiz-like. - Employs documentation and self-help primarily for many situations.
4	Appsiety, 2019 (Alves et al., 2019)	- It uses built-in sensors to gather user activity data that therapists may utilize - Users are evaluated using the System Usability Scale (SUS).	- Keeps track of patients who are already receiving treatment	- Unpublished means it cannot be tested. - Avoid helping those who lack access to therapists.
5	T-Bot, 2018 (Sharma, Puri, & Rawat, 2018)	- Based on Program E - A PHP module utilizing the A.L.I.C.E. technology	- Presents therapy options - An interactive approach to tutoring and self-help	- Absence of a verbal system - Users are unable to engage the bot in meaningful discourse. - Not open to the general public
6	Replika - 2017 (Mensio, Rizzo, & Morisio, 2018)	- NN - Generative chatbot implementing natural language interfaces	- Good at having meaningful interactions - Recollects chats and copies user behavior. - Apple app store rating: 4.7/5 from 1.9k users.	- No knowledge of mental health and no particular ethical principles - There is no documented evidence that mental health has improved.
7	Woebot:	- By psychologists	- A substantial	- According to user

	your self-care expert, 2017 (Fitzpatr-ick, Darcy, & Vierhile, 2017)	and AI specialists at Stanford - CBT on the web - Conversation Management System - NLP	decrease in depression, PHQ-9 (F=6.47; P=.01), and anxiety, GAD-7 (F1,54=9.24; P=.004) was seen in a sample of university students. - Twenty-eight users have given it 4.9/5 stars in the Apple app store.	comments, most replies are generic and script-like. - Not generative
8	COVID Coach, 2020 (Allen, 2020)	- VA, the US Department of Veterans Affairs - Safeguards for privacy	- Raising awareness of the epidemic - Excellent self-help tool - Mood and goal monitoring - 4.8/5 rating in the Apple app store from 524 users.	- Without mentioning bots - Not more than a self-management tool - There is no research on its efficacy.
9	Wysa: Mental Health Support, 2015 (Inkster, Sarda, & Subrama-nian, 2018)	- Conversational AI Agent - CBT - NLP an emotional balance	- Have a virtual coaching session with a qualified coach. - According to creators, 93 percent of users found it helpful. - It was named Apple's Best Stress App of 2019 - One hundred thirty-one people gave the app a 4.9/5 rating in the Apple App Store.	- Potential difficulties with the coaches' prejudice and availability for the expert sessions - Only around 20% less expensive than in-person counseling - Not included are spoken discussions with the chatbot.

10	Naluri, 2017	- Malaysian origin - A channel or message for healthcare professionals and patients	- Accessible to everyone - Various coaches, including therapists and nutritionists - Rating of 4/5 stars from 16 customers in the Apple app store.	- Based on complaints, the application is currently buggy. - They may have difficulties with coaches' availability and impartiality.

2.4 SUMMARY

In this chapter, all the newest technologies and implementations related to the objectives have been analyzed. Later a deep dive into the background of the science of ML and DL was shared. Finally, with the gathered knowledge and understanding in the following chapters, proper methods and implications are used to achieve the goals of this research.

Data-driven models for natural language conversation production were explored and it was realized traditional approaches are time-consuming and domain transfer is not effective. Hierarchical Recurrent Encoder-Decoder (HRED) is shown to be the best to create customized chatbots that can offer meaningful discussions and earn user confidence. P^2 Bot was introduced, which is developed using a transmitter-receiver structure to model interlocutors' awareness, emphasizing mutual persona perception. Supervised learning on task-specific datasets was explored that are used to tackle high-quality informal chitchat.

In the research background section, many applications were discussed. Kohonen's SOM algorithm is used in an Android app to treat borderline personality disorder, Ryan the Companion Bot is a socially-assistive robot, and Youper software employs meditation, CBT, and commitment therapy to reduce stress and anxiety. Youper is an expert system developed with therapist assistance, Appsiety is a platform for patients to communicate with therapists, T-Bot is a therapeutic chatbot, and Replika is a generative bot that mimics the user's message and conversion patterns. COVID Coach and NaluriNaluri are valuable tools for self-education and awareness but lack direct interaction between patients and experts.

CHAPTER THREE

METHODOLOGY

3.1 OVERVIEW

The first step of building any chatbot is designing an accurate and human-like dialogue generation model. The following sections present the details of achieving a mutual dialogue generation model. The dialogue generation selected for this study has five parts: transmitter model, supervised dialogue generation, fin-tuning the mode, reward shaping, and finally, receiver model.

After the model is designed, the gathered data are processed and analyzed. Then, the model will be trained on the prepared dataset. Finally, the proposed design and validation techniques are introduced.

3.2 RESEARCH APPROACH

Figure 3.1 represents different stages of this study's research approach. Various components of design, development, and exaltation are presented. In addition, tools and software used for each block are included. Yellow blocks represent this study's contribution blocks in each stage. Green blocks are tools, software, and techniques used for each block.

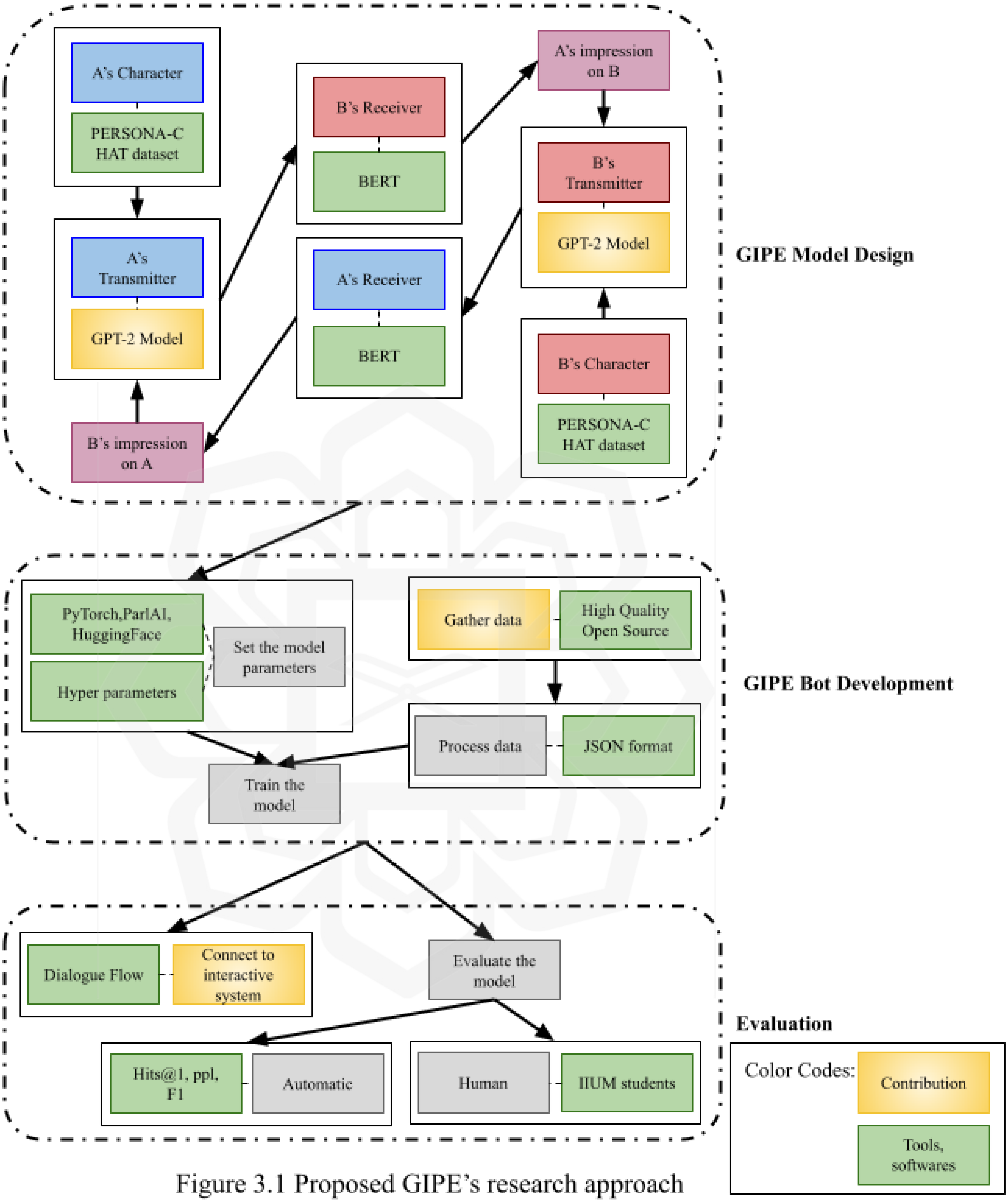

Figure 3.1 Proposed GIPE's research approach

3.3 GIPE BOT'S DESIGN

After understanding the research approach, this section goes through the GIPE bot's design. This section started with a detailed design flowchart. This flowchart is an important figure presenting this study's extensive yet detailed image. After that, the process of data acquisition is described. Gathering reliable yet accessible data was significant hence its direct impact on how the GIPE bot would respond to patients. Finally, details of the formulations behind each machine learning model and reward calculations are discussed. Understanding mathematics gives the options for modification and improvement in the future.

3.3.1 Detailed Design Flowchart

Figure 3.2 is a detailed flowchart of the GIPE design. This figure lays out a breakdown of each block and delivers a broad image of the whole research. This figure is divided into three sections: model training, dialogue generation, and user interface. The model training section presents an illustration and example of how raw data is processed, impeded, and finally used for training the model. Transmitter-Receiver machine learning algorithms are illustrated. With the help of the evaluation model, these algorithms are used for training the GIPE model. After the training phases, patients will interact with the model through the user interface. The dialogue generation section is the medium that helps process the user's input to prepare for interacting with the GIPE model and then returns the response. Users can interact with the GIPE bot via both voice and text input.

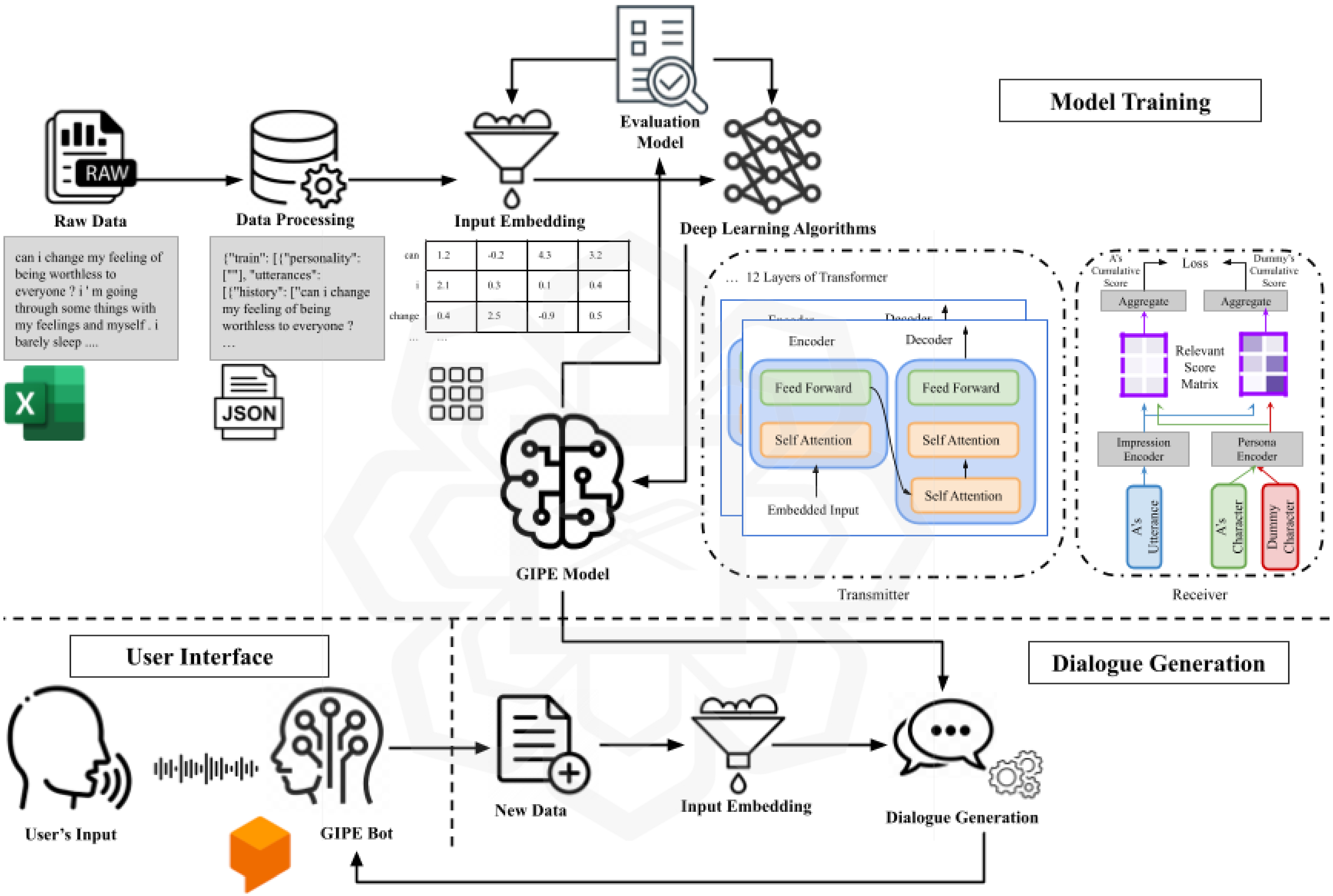

Figure 3.2 Proposed GIPE Bot's detailed architecture

3.3.2 Design Formulation

A. DATA ACQUISITION

It is particularly challenging to locate reliable statistics on mental health treatment. Moreover, the available data are a poor approximation of genuine contact between a patient and a therapist. Indeed, one can scrape Reddit, social news aggregation, and discussion websites and find intriguing therapeutic conversations between people. Still, it is almost impossible to verify the qualification of self-claimed professionals. Moreover, most of the other datasets that are accessible are either costly or proprietary.

After some investigation, a collection of therapist replies to genuine patients was discovered on the Counsel Chat website in December 2020 (*Mental health answers from Counselors I*, 25 March 20220), an open-source, relatively high-quality mental health query. Counsel Chat website is a form of an expert community. It serves as a platform for therapists to establish their credibility and connect deeply with prospective clients. On the website, therapists answer client inquiries, and visitors can "like" the most valuable answers. It is an intriguing concept that can produce some informative data.

The fact that verified therapists post these replies is an outstanding aspect of this information. Although they might not always be the best responses, they come from a subject matter expert. If Reddit data is utilized, anyone can be the one giving advice. Here, the people providing the guidance are licensed counselors. It's crucial to remember that encounters with therapists in person can differ significantly from those that are made public online. Additionally, this is not a conversation between a therapist and a client. There is only one discussion involved at a time.

Bertagnolli et al., 2020, were able to get in touch with the website's founders and access some actual data from the website. There is data analysis on these data in Chapter 4, and the dialogue generation model will be trained on these as well.

Since it is generally accepted that objective measurements and human assessment outcomes have poor correlations, human evaluation was further used in this study. Though human assessment on an extensive test dataset is costly and difficult to equate with other models in the literature, it is nevertheless necessary for dialogue development. Semantic and topical similarities are measured using the average embedding metric. This metric indicates that a high score will be obtained if the semantic content of the model-generated answer and the regression coefficients response is similar. Following prior research by Serban et al., 2017, and using the embedding measurement to some extent indicate the response quality (Vlad Serban et al., 2016).

A.A Data Conversion for Training

The required signature for the JSON file containing the training data is shown in Figure 3.3. There are two primary keys in the more significant JSON object. "valid" and "train." A collection of characteristics and utterance pairings makes up the training data called the train. Except for the validation set, valid is the same. The speaker's characteristic is described in a series of phrases called the character.

The character field was left empty for this model to denote a lack of characteristics data. Next, a list of candidate replies may be found in the candidate's section. This list includes several less-than-ideal replies to the conversation's history, where the last statement is the actual answer. Finally, it is necessary to define history. The history is just a list of characters with a new conversation turn at each place.

```
{
 "train": [
   {
     "personality": [
       "sentence",
       "sentence"
     ],
     "utterances": [
       {
         "candidates": [
           "candidate 1",
           "candidate 2",
           "true response"
         ],
         "history": [
           "response 1",
           "response 2",
           "etc..."
         ]
       }
     ]
   }
 ],
 "valid": ...
}
```

Figure 3.3 JSON file template for training data

This bot has two conversation turns one of them is from the individual imposing the inquiry and one from the therapist in reply. Thus, the question has to be answered. This format is carried out because the model does not want to train a general therapist bot but wants to find appropriate singular answers to situations. The model can be used immediately if the data is correctly prepared.

B. PROPOSED DIALOGUE GENERATION DESIGN

The Persona Perception, P^2, Bot's idea was proposed by Radford et al., 2019, to prototype an understanding between interlocutors and improve dialog generation using mutual character observation. The model has two components for achieving this task, a transmitter and a receiver, figure 3.4. The transmitter generates the conversation, while the receiver is responsible for mutual character observation. In P^2 Bot, each interlocutor A and B has a character c^A and c^B. Each of these characters is represented by P sentences, a.k.a. personalities, $\{c_1^A, \ldots, c_P^A\}$. After meeting for the first time, the two interlocutors will get to know each other via T time dialogue exchange of $(d_1^A, d_1^B, \ldots, d_T^A, d_T^B)$. A's utterance in t-th turn denoted as d_t^A. The entire of conversation archive up to t-th turn is $r_t^A = (d_1^A, \ldots, d_{t-1}^B)$. d_t^A is produced based

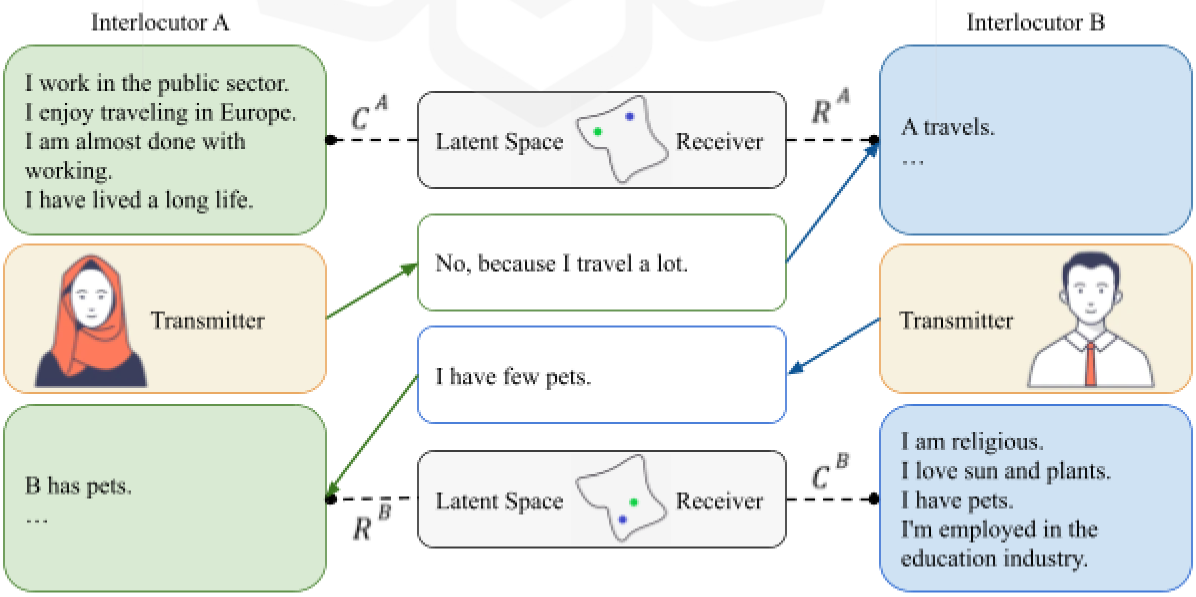

Figure 3.4 The proposed overview of dialogue generation model

on the $p(d_t^A|c^A, r_t^A)$ probability distribution and is transmitted to B so the same probes will be implemented to generate an ongoing conversation (Liu et al., 2020).

As the dialogue progresses, impressions will be formed by utterances. Each of their impressions should correspond to the other character. The receiver's goal is to determine how accurate the constructed impressions are to the real personalities. Firstly the receiver reflects impressions and characters into a latent space, then uses impression and character encoding to determine their accuracy. The relevant marks are used as mutual character perception incentives and in the transmitter's training. For example, R^A is the impression of B on A and vice versa.

B.A Transmitter Model

According to Li et al. (2016-a), dialogue production is regarded as a sequence generation difficulty in this work. The pretraining transformer language model (GPT-2) published by Radford et al. (2019) is utilized to initialize the transmitter. The two stages of training are supervised conversation production and fine-tuning of the self-play model. The supervised dialogue generation issue is optimized via the estimation of maximum likelihood (MLE). In the self-play model fine-tuning, reinforcement learning (RL) is used to urge the transmitter to develop a strategy that maximizes reward signals by simulating the discussions between two randomly assigned interlocutors. Language modeling and shared character perception are used in the reward function design.

B.B Supervised Dialogue Generation

The transmitter uses an overall design of 12 layered transformers to encode context and provide the output, figure 3.5. "Transformer Block" is abbreviated as "Block." Arrows connect the present block to the following layer's blocks. By allocating an embedding to each absolute location in the sequence, position information is included in the block through position encoding. In this case, the interior architecture of the block is absent—the training objective disregards [MASK] tokens. [CS] token is the indication that the new Character Started. , and refer the readers to Vaswani et al. (2017) for more details. [EOS] token means the End Of the Sentence, [SOS] token means the Start of the Sentence.

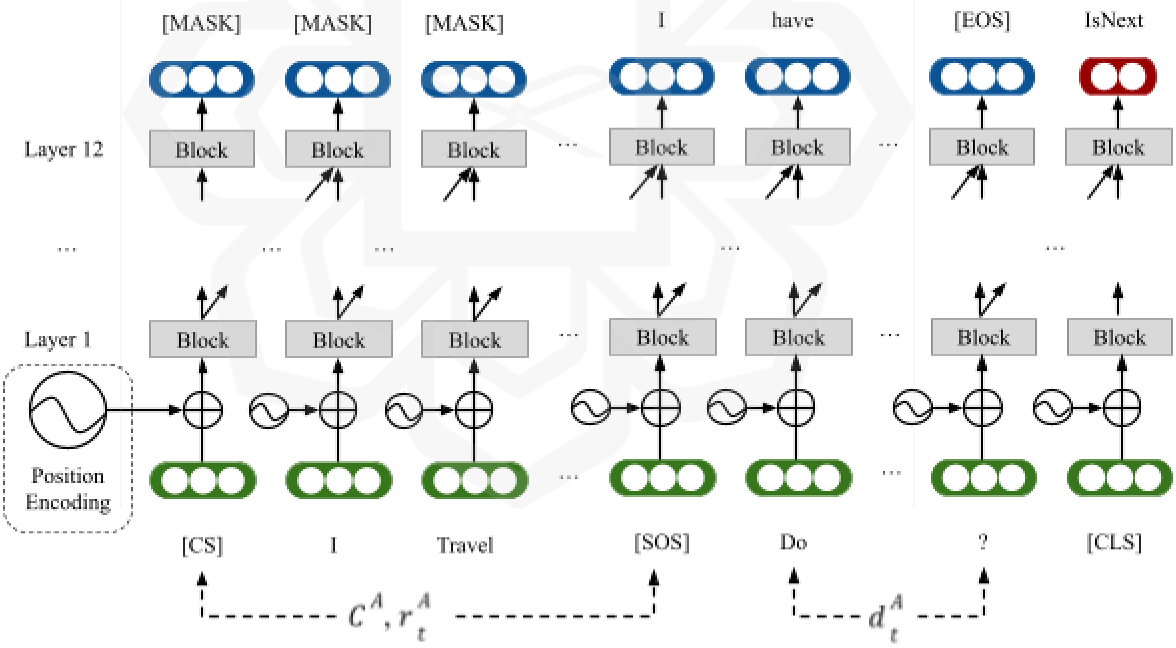

Figure 3.5 The proposed transmitter's architecture

Maximizing the conditional log-likelihood provided by the training instance is the development objective of MLE (Zhang et al., 2018).

$$\mathcal{L}_{MLE} = \sum_{n} log\, p\theta\, (d^{A}_{t,n} |\, c^{A},\, r^{A}_{t},\, d^{A}_{t<n}) \qquad (3.1)$$

θ is the transmitter's parameter. $d^{A}_{t,n}$ indicates the token of t[th] in d^{A}_{t} and $d^{A}_{t<n}$ means the sequence of tokens before the n[th] token.

The transmitter selects as a prediction the response candidate that maximizes the length-normalized score after applying beam search to store the highest response candidate $\widehat{d}^{A}_{t}$ during inference.

$$d^{A*}_{t} = arg\, max \frac{log\, p\theta\, (\widehat{d}^{A}_{t} |\, c^{A}, r^{A}_{t})}{|\widehat{d}^{A}_{t}\,|} \qquad (3.2)$$

In addition to teaching the transmitter to produce replies, they also train it to recognize if a response is a subsequent speech in the context (Wolf et al., 2019). Concretely, the created tokens have a unique token [CLS] added to each end. In the final transformer layer, a classifier is constructed over the token's hidden state, as shown by the red rectangle in Figure 3.5. We randomly choose a distractor during training for each answer, and we teach the classifier to provide the responder with a more outstanding score than the distractor. Equation 2 is expanded as follows, denoting as $x_{t} = 1$ the indicator indicating the produced answer $\widehat{d}^{A}_{t}$ is projected to be the following utterance and a hyper-parameter.

$$d_t^{A*} = arg\,max(\, \alpha \frac{log\, p\theta\, (\widehat{d}_n^A |\, c^A, r_t^A)}{|\widehat{d}_t^A\,|} \; + \; (1 - \; \alpha)\, log\, p\theta\, (\, x_t = \; 1 \,|\, c^A,\, r_t^A, \widehat{d}_t^A)\,)$$

(3.3)

B.C Self-play Model Fine-Tuning

While supervised conversation generation replicates human-like replies, comprehension is not its primary goal. To maximize mutual character perception, Transmitter is hence tuned using reinforcement learning (Lewis et al, 2017).

In particular, the two transmitters converse with one another for several rounds. While the other Transmitter acts as a learnable agent, the first one acts as a user with its settings locked. During the self-play, the learnable agent's parameter is adjusted. In the experiments, interlocutor A initiates a dialogue and can serve as the user without compromising generality. And B can then serve as the learnable agent.

Formulating the RL in the model requires a state composed of both character and dialogue history (Sutton et al., 1999). For example, A's status at turn n is $s_t^A \;=\; \{c^A,\, r_t^A\}$. The response generates action, a_t^A. Taking the status as the input, θ defined the policy as $p\theta(a_t^A|s_t^A)$, through which the agent learns to generate responses.

A and B proceed to generate responses until reaching the given limit. But in this stage, dialogue is composed, and the collected reward is used to optimize θ implementing policy gradients. $R(a_t^A)$ is the generated reward for A at t position. The reward is optimized using the likelihood ratio (Williams, 1992).

B.D Reward Shaping (RS)

To achieve a high-quality conversation, three rewards systems are designed (Liu et al., 2020).

RS.1 Language

The produced replies must follow human language patterns, which a GPT-2 model can assess.

$$R_1(a_t^A) = \frac{1}{|a_t^A|} \sum_n log\, plm\, (a_{t,n}^A | a_{t<n}^A) \qquad (3.4)$$

RS.2 Coherence

Discourse coherence is not taken into account while evaluating the language score. However, a fair reaction should connect meaning and context, another crucial component of humanlike responses. The proficient Next Utterance Predictor is employed to account for speech coherence.

$$R_2(a_t^A) = log\, p\theta\, (x_t = 1 | a_t^A, s_t^A) \qquad (3.5)$$

RS.3 Mutual P^2

This reward is designed to encourage understanding between interlocutors explicitly. The entire dialogue's long-term objective is mutual character perception. Therefore the immediate consequences of the present action could not become apparent for some time.

$$R_3(a_t^A) = r(a_t^A) + \sum_{k=t+1}^{N} \beta^{2(k-t)-1} r(d_t^{B*}) + \beta^{2(k-t)} r(a_k^A)) \qquad (3.6)$$

β denotes the discount factor and $r(a_t^A)$ is P^2 the score in the n-th turn.

The final reward of P^2 bot is:

$$R = \gamma_1 R_1 + \gamma_2 R_2 + \gamma_3 R_3 \qquad (3.7)$$

γ_1, γ_2, and γ_3 are hyperparameters.

B.E Receiver Model

The receiver is designed to determine the distance between generated perceptions and underlying personalities via negative sampling. In particular, while training, arbitrarily a character distractor was sampled c^z. The receiver is trained to recognize the real c^A from $\{c^A, c^z\}$. In inference, the Receiver is accountable for delivering a suitable applicable outcome for each phrase to represent the suggested shared persona perception. As part of the incentives, the score is added to the self-play fine-tuning on Transmitter.

As explained earlier, the receiver includes two encoders, one for the impression and the other for the persona. Both encoders are initialized by Bidirectional Encoder Representations (BERT), Devlin, et al., 2019. In addition, an open-source ML framework for NLP is called BERT. BERT leverages the surrounding text to provide a context to assist computers in grasping the implications of ambiguous words in the text. The BERT framework may be adjusted using question-and-answer datasets after pre-training on Wikipedia text.

After that, the average of all the representations is taken to get a fixed v-volume vector for a single sentence. Given d_t^A and c^A, the receiver implements the following function to acquire the d_t^A's personal perception grade.

$$\text{grade}\ (d_t^A, c^A) = \frac{aggregation\ (R_t^A\ (C^A)\top)}{\sqrt{v}} \qquad (3.8)$$

R_t^A and C^A are the impression encoding and persona encoding of d_t^A and c^A respectively. ⊤, transpose, is an operator that flips a matrix across diagonally.

C. EVALUATION TECHNIQUES

The efficiency of response creation is evaluated in this study using two assessment metrics (Serban et al., 2017). Each of these is the Embedding Average metric (EACosine), which evaluates how much the ground truth response and model reaction's embedding vectors resemble one another in terms of cosine. The utterance embedding vector represents the average of an utterance's phrase embeddings. It acts as a tool to encourage semantic similarity. Another tool metric is the Bilingual Evaluation Understudy (BLEU) (Papineni et al., 2002), which tracks the number of n-gram matches between the reference responses and the generated responses. There are no position-specific matches. The two methods mentioned above are widely used to determine how similar two texts are.

Following other research (Li et al., 2015-c), (Serban et al., 2017), also assessments were conducted utilizing distinct-1, distinct-2, average entropy, and several entities per answer to gauge the potential of the suggested models to foster variety and offer additional details during conversation creation.

Retrieval-based, generative-based, and pre-trained finetune-based algorithms make up the baselines. An analogous dual architecture was presented within the Dually Interactive Matching Network baseline (Gu et al., 2019) to match the reactions and their related circumstances. Furthermore, across retrieval-based benchmarks that made use of the memory network and profile data, the KV Profile Memory benchmarking (Zhang et al., 2018) served as the recognized standard. As generative baselines for conversation development, Language Model, Generative Profile Memory (Zhang et al., 2018), and SEQ2SEQ using a learning algorithm (Bahdanau et al., 2015) have been used.

3.4. DEVELOPING AN INTERACTIVE VOICE RESPONSE SYSTEM

After the consultation ML model is trained on the acquired data, and the model is evaluated, the final stage of this research is to develop a dialogue-spoken model application that can interact with patients via both text and voice.

To develop the application, DialogFlow was employed. It uses Dialogflow. Dialog Flow is an NLU platform that uses conversational SLU techniques and CMU Sphinx (speech recognition systems). A conversational user interface can be designed and integrated into mobile apps, online applications, gadgets, chatbots, interactive voice response systems, and other applications. Dialogflow may examine a variety of user inputs, such as text or audio sources (like a voice recording). Additionally, it can respond to consumers through text or artificial speech.

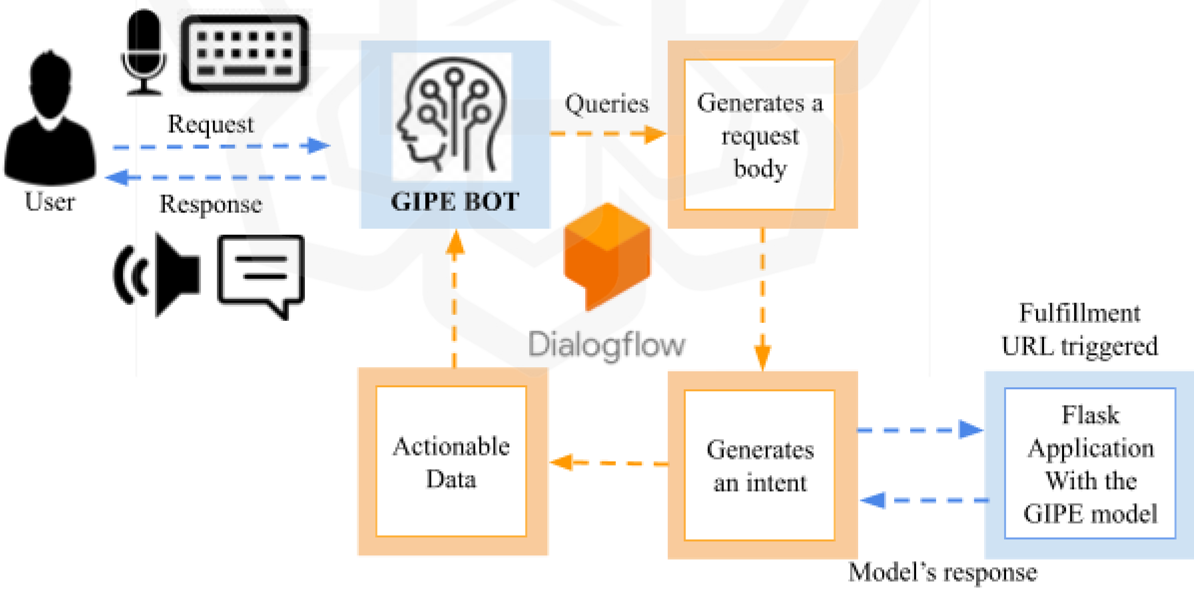

Figure 3.6 The proposed interactive voice response chatbot architecture

The trained ML model is uploaded to a public HTTPS URL to achieve this. This application flask allows direct interaction with the model. Then after creating a project in Dialogflow, intents are linked to the Flask application. Figure 3.6 presents this system's architect for a better visual understanding.

3.5 SUMMARY

In this chapter, the research approach, the model design, and the formulation were completed, and the data was gathered. Each section provided detailed figures and calculations. Each block of the design chart is explored and explained. The derailed design flowchart branched into three sections: model training, dialogue generation, as well as the user interface.

As an illustration of how raw data is processed, hindered, and ultimately utilized for training the model, transmitter-receiver machine learning methods are employed. Following training, patients will use voice and text input to communicate with the model via the user interface. A type of expert community is the Counsel Chat website (Mental health answers from counselors, 25 March 20220). It acts as an avenue to help therapists build their reputations. Therapists respond to client questions on the website, and users can "like" the responses. The last step in this research is to create a dialogue-spoken system application once the consultation machine learning (ML) framework has been trained on the gathered data and assessed. With all of this, model implementation and validation are the next steps explored in the following chapter.

CHAPTER FOUR

RESULTS AND ANALYSIS

4.1 OVERVIEW

This chapter presents the results and the comprehensive findings of this study in different stages and their analysis. Firstly several informative charts are introduced, offering visualize representative of different aspects and frequencies of the data. After that, the model implementations and different parameters implemented are explained. This section is followed by results of the dialogue generation model, automatic evaluation and human evaluation, some raw examples of users' conversations with the consultation model, and the mobile application prototype. Alternatively, the results are inspected and apprehended truly, bringing this chapgbert to a fulfilling conculsion.

4.2. DATASET ANALYSIS

This data analysis is based on rich dataset provided by the Counsel Chat website,(*Mental health answers from Counselors I*, 25 March 2020), to Bertagnolli et al., 2020. The Counsel Chat website serves as a platform where individuals can seek professional advice and guidance on a wide range of topics. The dataset used for this study comprises anonymized records of user interactions with the consultation model deployed on the website. These interactions encompassed various domains, including mental health, relationship advice, career guidance, and more. The dataset offers a diverse and representative sample of real-world conversations, enabling a comprehensive examination of the effectiveness and performance of the dialogue generation model. By leveraging this valuable resource, the study gains valuable insights into the users' needs, preferences, and the efficacy of the consultation model in addressing their concerns. The dataset is shown as a CSV file including eleven lines. Table 4.1 presents the columns’ headers and their descriptions.

Table 4.1 Data Analysis Columns

No	Column header	Description
1	questionID	A particular question identifier that is different for every question.
2	questionTitle	The question's title on the counsel chat page.
3	questionText	The bulk of the client's inquiry to the counselor.
4	questionLink	The most recent URL for the query.
5	topic	The heading of the question was listed under it.
6	therapistInfo	A description of each therapist, typically including their name and areas of expertise.
7	therapistURL	A connection to the counselor's profile on counsel chat answer.
8	answerText	The therapist's answer to the query.
9	upvotes	The number of upvotes for the answer.
10	views	How many times has the question been viewed.
11	split	Divided data into training, validation, and testing groups.

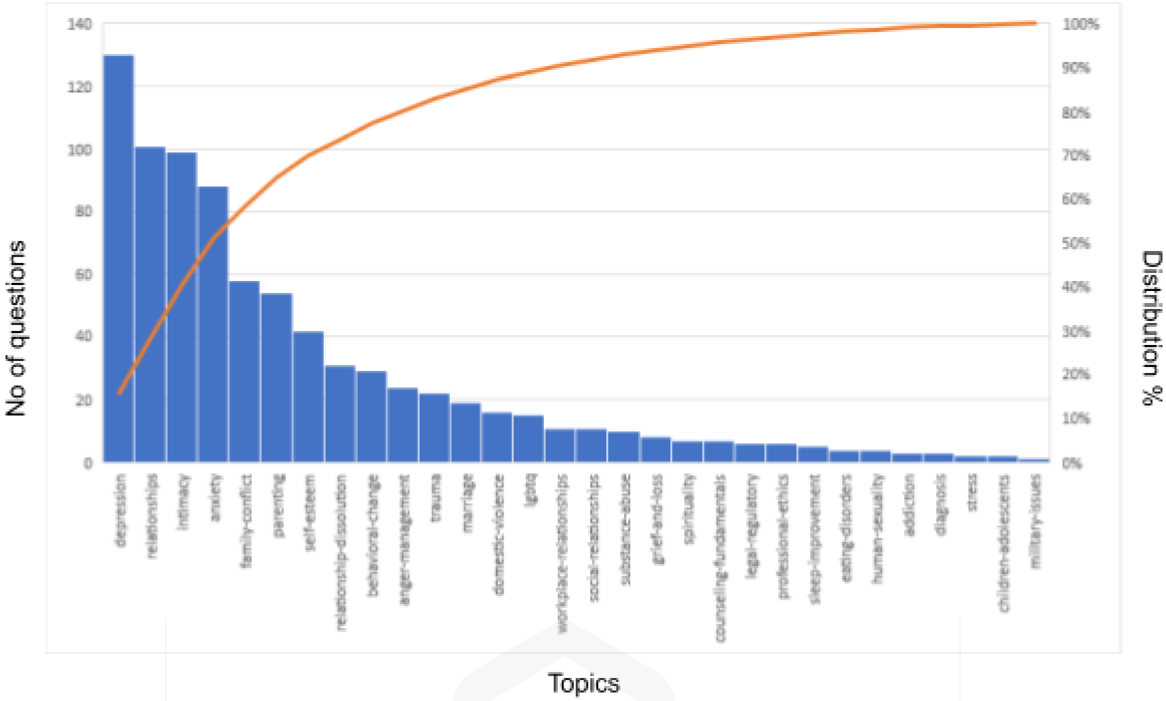

Figure 4.1 Number of questions per topics

There are 818 unique questions in the dataset, encompassing a wide range of topics and concerce. Overall, 2,129 responses are provided to these questions, reqfelcting engagement and depth of the converstions. The questions are categorized into 30 distinct topics, including depression, anxiety, and self-esteem, figure 4.1. Depression is the topic with the most questions, with about 130. After that relationships and intimacy cover approximately 100 questions each. Anxiety is in fourth place with about 90 questions. Given the substantial coverage of disorders such as depression and anxiety, this dataset holds immense value and reliability for the study's focus. It offers a robust foundation for examining and understanding the dynamics of these disorders within the context of the consultation model, allowing for meaningful insights and informed analysis.

Figure 4.2 shows the distribution of the number of responses per question, offering valuable insights into the engagement level of therapists. About 75 percent of questions in the dataset have one or two answers by there are questions that are highly engaged therapists. However, there are certain questions that have attracted a considerably higher level of engagement from the therapist. The most responded question is, “Do I have too many issues for counseling?”. This particular question

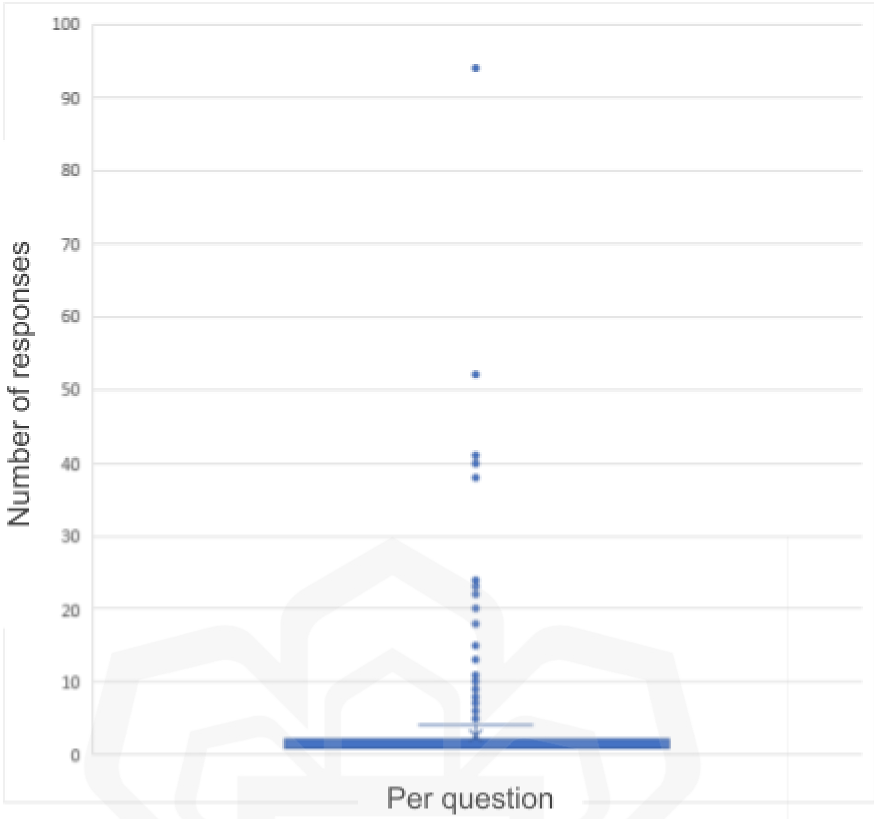

Figure 4.2 Number of responses per question

seems to have struck a chord with therapists, eliciting a wide range of perspectives and approaches. The high response rate suggests that therapists have differing opinions and insights on this matter, possibly reflecting the complexity and individuality of each client's circumstances. Having multiple answers to the same questions provides a more diverse training dataset as well as covering different perspectives. By encompassing various viewpoints and approaches, this dataset enhances the training process for counseling models. It allows for a broader understanding of the therapeutic landscape, accommodating the multitude of perspectives that therapists may hold. This diversity is essential for ensuring the model's ability to address the unique needs and concerns of individuals seeking counseling, fostering a more inclusive and comprehensive approach to the practice.

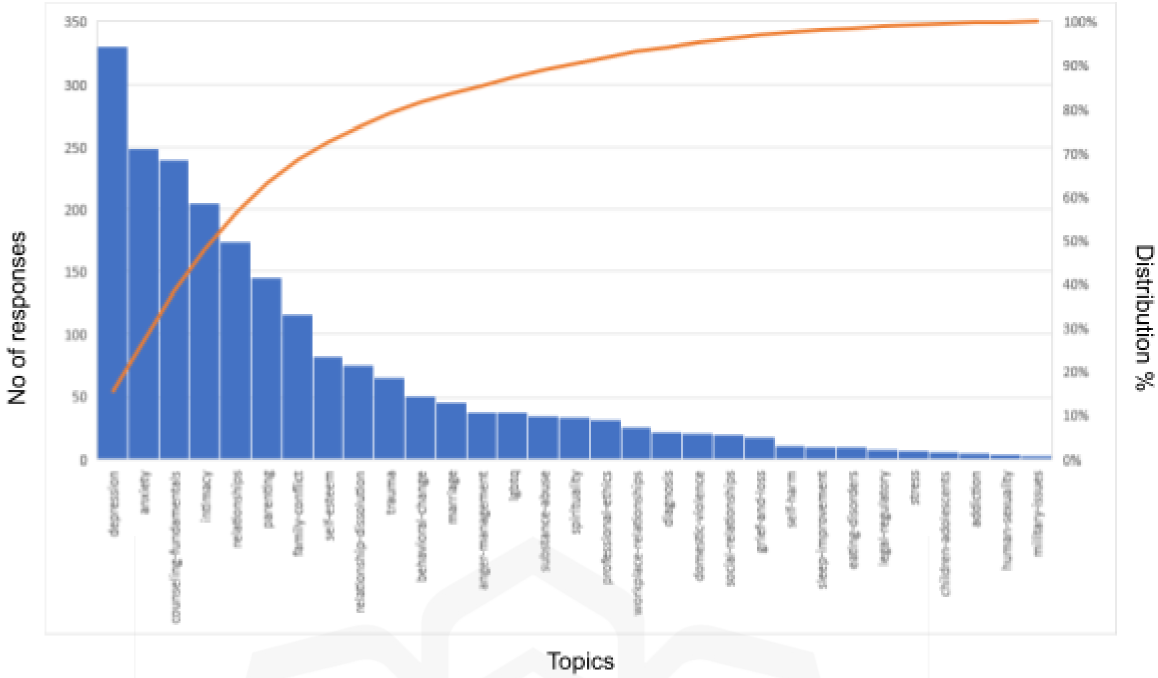

Figure 4.3. Responses counts for each topic

Figure 4.3 represents the number of answers per topic. Topics related to depression and anxiety received the most response with approximately 330 and 250 respectively. Meaning about 580 out of 2000 questions are directly related to the topics of interest of this study. Depression, relationships, intimacy, and anxieties are the top questioned topics in order, but interestingly consulting fundamentals found its place as the top 3 most answered topics.

The dataset analysis further reveals an interesting pattern regarding the length of questions and corresponding therapist answers. While the majority of questions are relatively brief, therapists tend to provide more extended and in-depth answers. This suggests that therapists prioritize providing comprehensive responses to address the concerns and needs of individuals seeking counseling.

Specifically, the average length of the questions in the dataset was approximately 64 words. In contrast, therapists' answers exhibited a higher average length, averaging around 170 words. This stark contrast indicates that therapists are

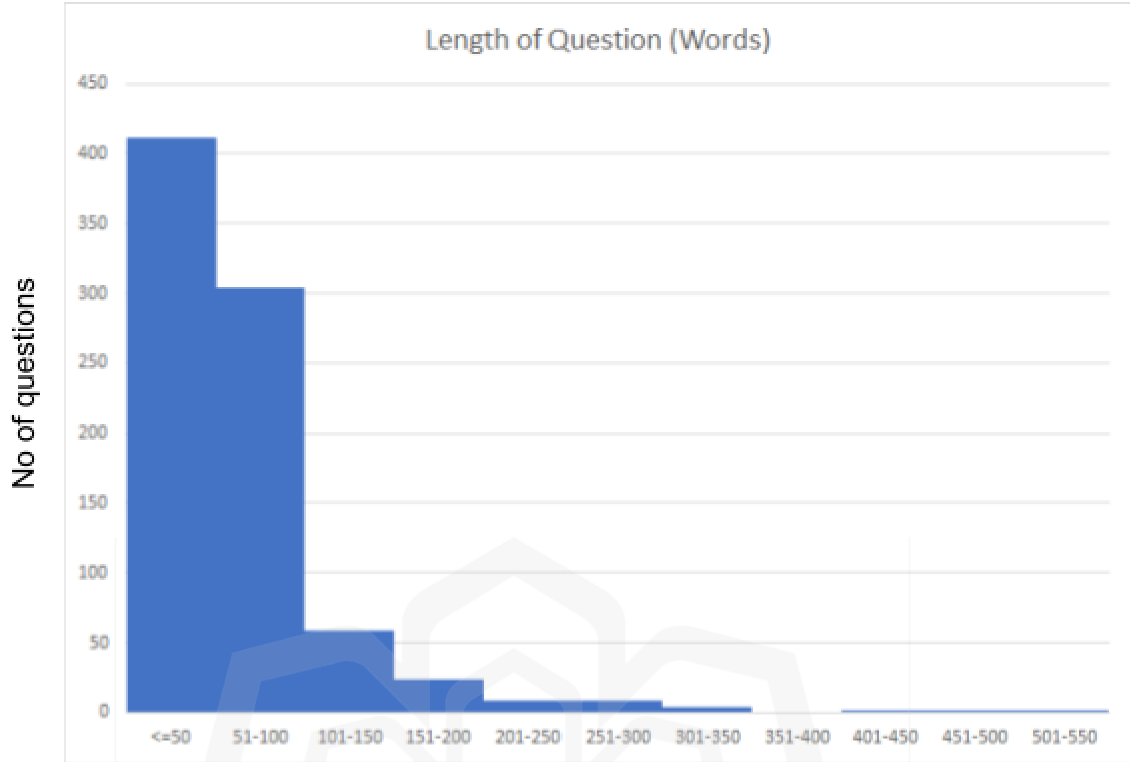

Figure 4.4 Distribution of length of questions

not only addressing the immediate queries but also taking the opportunity to offer more detailed explanations, insights, and guidance to individuals seeking their expertise.

Figure 4.4 provides additional insights into the distribution of question lengths. It reveals that approximately 410 questions in the dataset consist of fewer than 50 words, indicating a substantial portion of concise inquiries. Furthermore, Figure 4.5 demonstrates that more than 1000 questions fall within the range of 50 to 100 words, signifying a significant number of moderately sized questions.

The presence of these longer answers from therapists has significant implications for training models. By responding with more extensive and comprehensive answers, therapists provide valuable context, analysis, and support to individuals seeking guidance. Consequently, when the trained model is exposed to these longer answers during the training process, it is prompted to generate responses

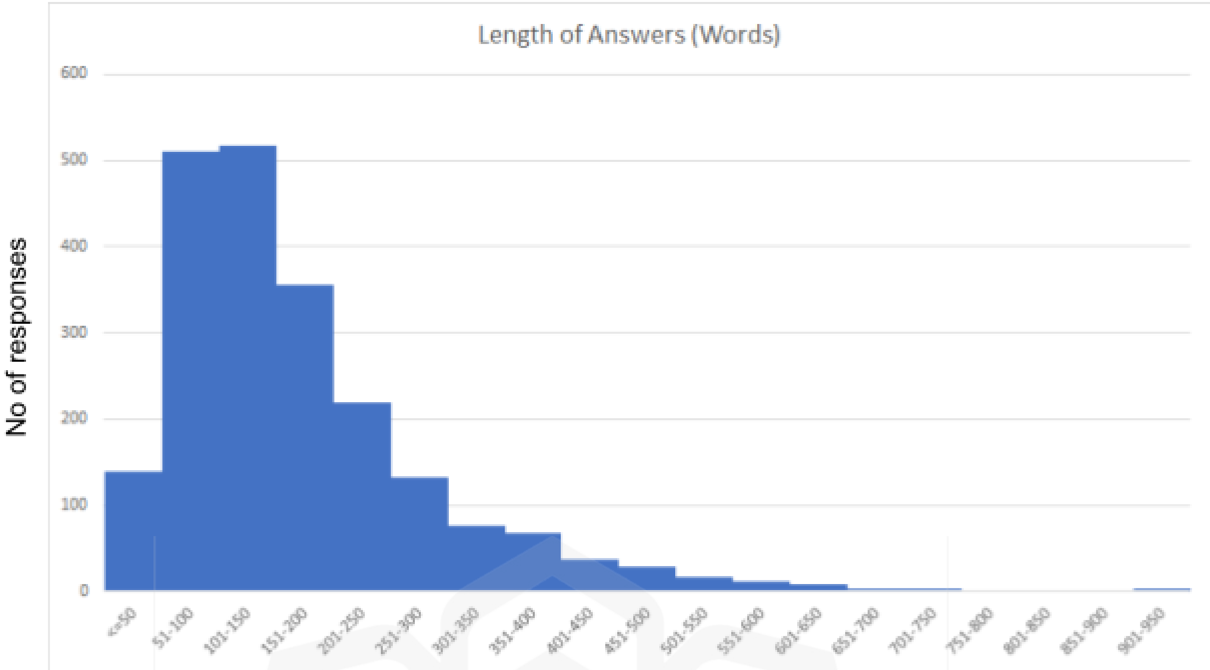

Figure 4.5 Distribution of length of responses

that are more elaborate, detailed, and encompassing. This encourages the model to develop a capacity for providing in-depth and comprehensive answers, thereby reflecting the expertise and depth of knowledge possessed by experienced therapists.

4.3 RESULTS AND ANALYSIS

The implementation was built on the transformers library from HuggingFace, ParlAI, and PyTorch (Paszke et al., 2019). (Wolf et al., 2019). The receiver as well as the transmitter in supervised learning are powered by Adam's optimizer (Kingma & Ba, 2014), which has a learning rate of 6.25e-5.

The learning rate of the self-play section of the transmitter was defined as 1e-6. The hyper-parameters γ_1, γ_2, γ_3, and β were defined as 0.5, 0.5, 0.4, 0.1, and 0.1 respectively. The transmitter underwent two epochs of supervised training, and 2000 dialogues of self-play fine-tuning with three turns each were used. The size of the beam search was set at 2 (Liu et al., 2020).

This headline presents different results collected during this study. First, the results of an automatic and human assessment of the dialogue generation model are presented. After that, some of the results of raw conversions of the consultation model are displayed. Finally, the prototype of the model applications is shown.

4.3.1 Results of Dialogue Generation

Two methods were used to evaluate the dialogue generation model: automatic evaluation and human evolution. Hits@1(%), ppl, and F1 are the automatic evaluation techniques used, and human evolution concentrates on human-like and informative model responses sounded to humans.

4.3.1.1 Automatic Evaluation

This study's benchmarks are divided into three types: retrieval-based, generative-based, and pre-train-fine tune-based models.

The Liu, & et al., 2020 dataset was used as a source. The formal bassline that employed the memory network in conjunction with profile details was KV Profile Memory (Zhang et al., 2018). A dual matching architecture was offered by a dual interactive matching network (Papineni et al., 2002) to match the responses and their relevant contexts among the retrieval-based baselines. The development set was used to evaluate all findings since the test set wasn't made available to the general public.

Table 4.1 Results of automated comparisons of various techniques on the PERSONA-CHAT dataset (Liu, & et al., 2020).

Category	Model	Original			Revised		
		Hits@1 (%)	ppl	F1(%)	Hits@1 (%)	ppl	F1(%)
Retrieval	KV Profile Memory	54.8	-	14.25	38.1	-	13.65
	Dually Interactive Matching	78.8	-	-	70.7	-	-
Generative	Generative Profile Memory	10.2	35.01	16.29	9.9	34.94	15.71
	Language Model	-	50.67	16.30	-	51.61	13.59
	SEQ2SEQ-ATTN	12.5	35.07	16.82	9.8	39.54	15.52
Pretrain Fine-tune	Lost In Conversation	17.8	-	17.79	16.2	-	16.83
	Transfertransfo	82.1	17.51	19.09	-	-	-
	P^2 Bot	81.9	15.12	19.77	68.6	18.89	19.08
	GIPE Bot	82.2	15.04	19.67	68.8	19.02	18.67

The experimental outcomes for automated measurements are presented in Table 4.1 Hits@1, Perplexity (ppl), and F1 are reported as the official automated measures to assess the approaches, adopting Zhang et al., 2018. Hits@1 is the likelihood that the actual response will score the top out of 20 possible responses. Lower values denote greater performance. Perplexity quantifies the negative log probability of the proper order produced by the model. The symmetrical mean of word-level recall and accuracy is F1. This research performs better than virtually all baselines and delivers solid Hits@1 performance along with highly competitive ppl and F1 performance. This method still performs best in the updated mode, with a 13% relative improvement in people's accomplishments in contact with the most robust baseline. It is significant to mention that we also attempted to use Hits@1 as the prize, but the outcome was not good.

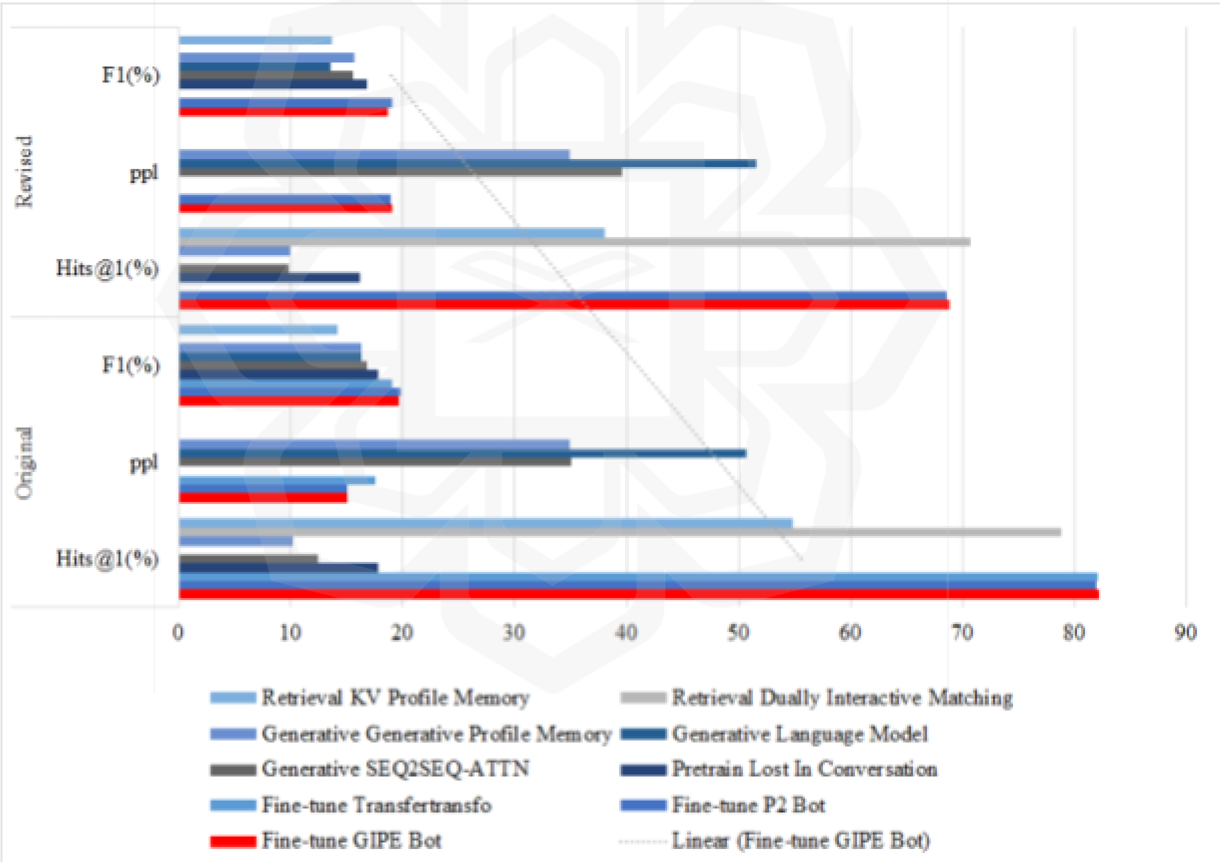

Figure 4.6 Automatic evaluation results graph

Figure 4.6 is a presentation of the automatic evaluation results in graph form providing a visual representation of the performance of different approaches. In the red bar, the GIPE bot's results are superior in some metrics, such as Hits@1, both in original and revised data, while presenting reliable results based on other metrics. These superior results in Hits@1 suggest that the GIPE bot is particularly effective at generating responses that align closely with the expected or desired answer.

However, it is important to note that the GIPE bot also demonstrates reliable results based on other evaluation metrics. While the specific metrics are not mentioned in the given statement, it can be inferred that the GIPE bot performs consistently and reliably across multiple evaluation criteria. These additional metrics likely assess various aspects of the model's performance, such as overall response quality, coherence, fluency, and relevance.

On the other hand, the experimental outcomes for automated measurements are presented in Table 4.1. y adopting these established evaluation metrics, the study ensures a comprehensive assessment of the different approaches, allowing for a fair and objective comparison of their performance. The results presented in Table 4.1 provide valuable insights into how well each approach performs in terms of accuracy, language understanding, and text generation quality.

4.3.1.2 Human Evaluation

No automated metric is ideal for assessing work with such an open domain, as stated by (Liu et al., 2020). Therefore this study used their human evaluation of state-of-art benchmarks and compared it with its results. We randomly selected 200 responses generated by different approaches from the original development set and asked students to rate them. The data was collected in the form of an anonymous electronic questionnaire. The scale goes from 1 to 4. Rating 1 indicates that the response is only correct in terms of sentences and phrases without much consideration for context. A rating of 2 suggested that the response is also comprehensible with the context. a rating of 3 indicates that the response is reasonable as well as compelling and informative rather than just a one-word answer such as "Yes." Lastly, a rating of 4 indicates that the answer is consistent with the interlocutor's character, which is crucial for determining whether the model can use the person. The results of this human evaluation were then compared to the automated evaluation metrics discussed earlier and they are similar to those of the automated evaluation, as shown in table 4.2, proving the superiority of this research over the baselines. This similarity in results further supports the superiority of the research approach utilized in the study over the baseline approaches. It suggests that the model developed in the research outperformed other approaches not only according to automated metrics but also based on human judgment, which is crucial for assessing the quality and relevance of the generated responses.

Table 4.3 Human evaluation results.

Model	1(%)	2(%)	3(%)	4(%)	Avg
Lost In Conversation	26.3	48.7	22.0	3.0	2.017
Transfertransfo	41.7	25.3	28.7	4.3	1.956
P^2 Bot	18.9	26.3	28.6	26.2	2.621
GIPE Bot	17.7	26.2	28.5	27.6	2.66

Figure 4.7 is a virtual illustration of human evaluation responses, providing insights into the performance of the GIPE bot. As presented in the graph, the GIPE bot received the highest number of 4 responses indicating that the responses represented the characters' personalities. This suggests that the GIPE bot exhibited a strong ability to capture and emulate the unique traits, behaviors, and speech patterns of different individuals, enhancing the overall conversational experience and authenticity. This graph is an excellent presentation of awareness by the GIPE bot.

Also, the GIPE bot received the lowest number of 1 response meaning the majority of its responses were coherent, meaningful, and grammatically correct. These results have successfully fulfilled one of this study's goals of a coherent generative chatbot. Hopefully, with a higher number of relevant data, the coherency has a promise of improving even higher.

It is important to note that these positive results are a promising indication of the GIPE bot's performance. However, the study acknowledges that with a higher volume of relevant data, the coherency and quality of the responses have the potential to improve even further. This suggests that as the chatbot is exposed to a larger and

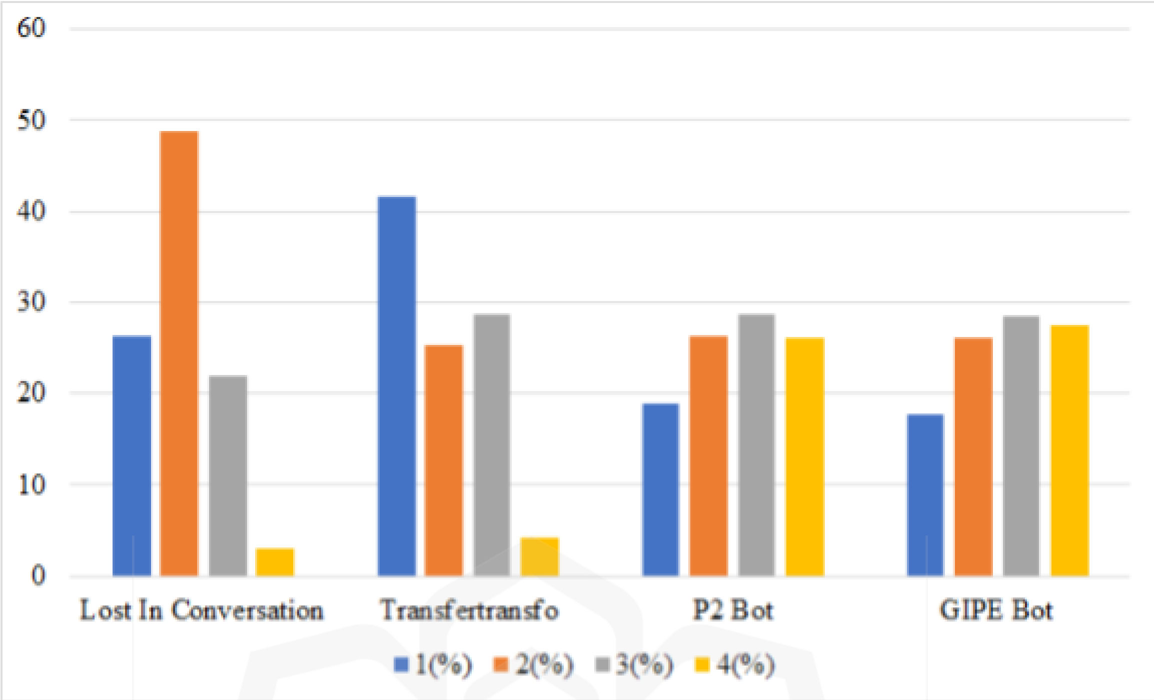

Figure 4.7 Human evaluation results graph

more diverse dataset, it can benefit from a broader range of examples and patterns, potentially leading to enhanced coherency and accuracy in generating responses.

Overall, the insights provided by Figure 4.7 highlight the GIPE bot's strengths in capturing character personalities and producing coherent responses. These findings contribute to the study's objective of developing a generative chatbot that can engage in meaningful and contextually appropriate conversations, with the potential for further improvement as more relevant data is incorporated into its training.

4.3.2 Consultation Model

Following the training of the dialogue generation model using the dataset acquired, the consultation chatbot was tested with some simulated conversations. Figure 4.8

shows the consultation chatbot's conversation with the user. As observed, the chatbot is empathetic and provides relevant information and techniques to help the user. As presented in the example, the chatbot shows awareness; hence the responses sound human-like and relevant. Also, it can remember and reuse the information provided earlier. This promising feature illustrates the generative ability of the GIPE bot. As expected of an expert chatbot, the responses are informative and emphatic.

In the presented example, the chatbot's responses reflect an understanding of the user's concerns and needs, effectively providing guidance and support. This empathetic approach enhances the user experience and contributes to the chatbot's effectiveness as a counseling tool. By demonstrating awareness, the chatbot is able to deliver responses that are not only contextually appropriate but also emotionally sensitive, thus mimicking the qualities one might expect from a human counselor.

Additionally, the chatbot showcases the ability to remember and reuse information provided earlier in the conversation. This feature highlights the generative ability of the GIPE bot, allowing it to maintain continuity and build upon previous

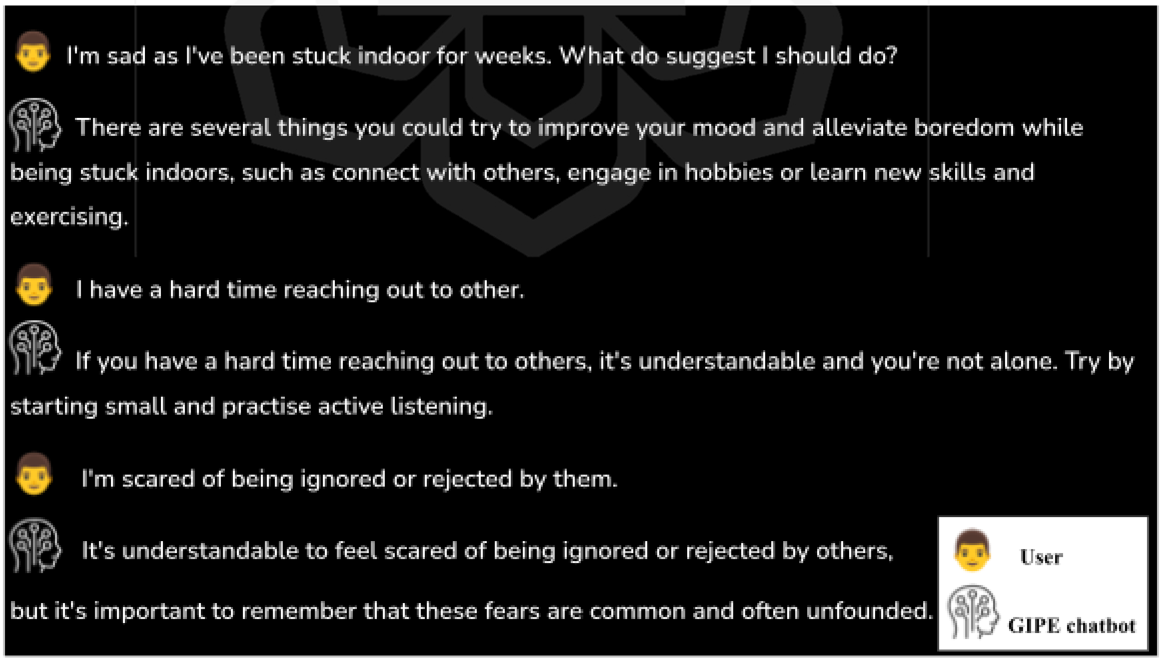

Figure 4.8 Example of the GIPE chatbot conversation with patient

interactions with the user. By recalling and leveraging previously shared information, the chatbot can offer personalized and tailored responses, enhancing the conversational experience and providing more targeted support.

Overall, the consultation chatbot's performance in the simulated conversations exemplifies the characteristics expected of an expert chatbot. The responses are informative, providing valuable insights and information to address the user's concerns. Furthermore, the chatbot's empathetic and human-like approach, combined with its generative ability and memory retention, contributes to a promising and effective counseling experience for users seeking guidance and support.

While these results seem highly promising, the true potential of the GIPE chatbot can only be tested in a clinical environment and with actual patients. In the presence of psychologists and consultants, patients' current mental health should be established. Then they will have regular interaction with the GIPE chatbot. After a fixed set of time, their mental health should be examined. This approach is the only way that the true impact of the GIPE bot can be fully understood and measured.

4.3.3 Mobile application

The final stage of this study was developing a user-friendly mobile application. This application can keep the flow of natural conversations with patients diagnosed with depression and recommend techniques to overcome it. We are also hoping to reach youth around the globe without access to mental healthcare. The mobile application's prototype is presented in figure 4.9.

As shown, users can both type their messages to the GIPE bot or send a voice note which then will be converted to text. The design is simple and direct to not cause any confusion for the users as well as to bring the learning curve to the lowest possible. A record of conversions is kept on the user's device for future reference.

The design of the mobile application emphasizes simplicity and directness. The aim is to ensure that users can navigate the application without confusion or difficulty, minimizing the learning curve. By keeping the design straightforward and intuitive, the developers strive to create a user-friendly experience that encourages engagement and openness.

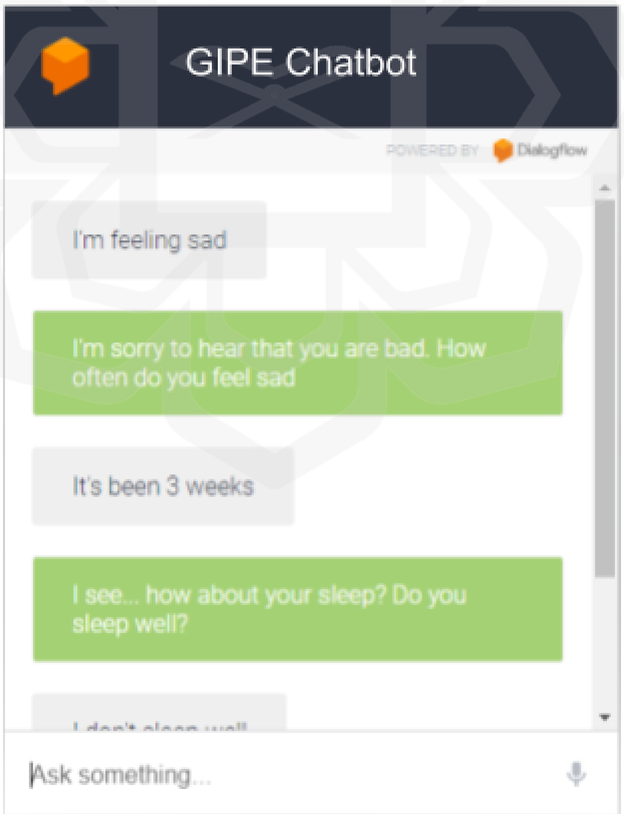

Figure 4.9 The mobile application prototype

To provide continuity and enable users to refer back to their previous conversations, a record of conversions is stored on the user's device. This feature allows users to review their past discussions, maintain a sense of progress, and access any valuable insights or guidance shared by the chatbot. By having a record of conversations readily available, the mobile application enhances the user's ability to track their journey and refer to relevant information when needed.

Overall, the development of this user-friendly mobile application serves as a valuable tool to facilitate conversations, offer recommendations, and provide support for individuals dealing with depression. By leveraging technology and creating an accessible platform, the study aims to extend mental healthcare services to a wider audience, particularly youth worldwide who may lack access to traditional forms of therapy or counseling.

4.4 DISCUSSION

This study provided data analysis based on information provided by the Counsel Chat website (*Mental health answers from Counselors*, 25 March 20220), to Bertagnolli et al., 2020. There are 818 unique questions, and overall, 2,129 responses to them. About 75 percnet of questions have one or two responses by there are questions that are highly engaged therapists. The most responded question is, "Do I have too many issues for counseling?". Then the data conversion was achieved by the two primary keys in the more significant JSON object, "valid" and "train." A collection of characteristics and utterance pairings makes up the training data called the train. This collection is just a list of characters with a new conversation turn at each place. The character field was left empty for this model to denote a lack of characteristics data.

After that, the dialogue-generation machine-learning model was trained on the processed dataset. The implementation was based on the transformers library from HuggingFace, ParlAI, and PyTorch. The model achieved 82.2% accuracy on Hits@1(%) in the original data and 68.8% on the revised data. This result shows an improvement compared to the benchmark. Humans also evaluated the model based on relevance, coherency, grammar, and awareness. The GIPE model received 2.66 on human evaluation representing the highest coherency and awareness. Based on the evaluation, the majority of responses were relevant and human-like.

Finally, the model was connected to the Dialogflow system. This stage provides the interactive voice feature with a user-friendly user interface. Users have the option of interacting with the GIPE bot via both voice and text messages. The mobile application's design is simple and to the point, eliminating any confusion. Any user with limited knowledge of technology can use it within a few minutes of tutoring.

4.5 SUMMARY

This chapter serves as a comprehensive overview of the achievements obtained throughout the study, summarizing the key findings and outcomes. It begins by presenting various representations of the gathered data and discussing the implemented processing methods. This initial step ensures that the data is organized and prepared for further analysis and model training.

The chapter then progresses to highlight the results of the dialogue-generation machine learning phase. This stage involves training models to generate responses based on the acquired data. The outcomes of this training process are discussed, shedding light on the model's performance in generating contextually appropriate and coherent responses.

To assess the effectiveness of the developed models, both automatic and human evaluation results are presented and explained. Automatic evaluation metrics provide quantitative measures of the model's performance, while human evaluation involves subjective judgments from human assessors. By considering both approaches, the study gains a comprehensive understanding of how well the models perform and how they compare to human-generated responses. The promising results obtained from these evaluations validate the study's objectives and demonstrate the efficacy of the developed approach.

Finally, the chapter concludes by unveiling the consultations model, providing examples of simulated conversations between the chatbot and users. These examples showcase the chatbot's ability to engage in empathetic and informative exchanges, demonstrating its potential as a valuable tool in assisting individuals dealing with depression. Additionally, the chapter introduces the user-friendly mobile application developed as a practical implementation of the research. The mobile application aims to provide a platform for natural conversations and offer techniques to overcome depression, with the goal of reaching global youth who may lack access to mental healthcare.

Overall, this chapter effectively summarizes the various stages and achievements of the study, from data processing and dialogue-generation results to evaluation findings and the practical implementation of a mobile application. These accomplishments collectively contribute to advancing the understanding and application of generative chatbot technology in the context of mental health support.

CHAPTER FIVE

CONCLUSION AND FUTURE WORK

5.1 CONCLUSION

The thesis aimed primary to design a set of algorithms that allows users access to an affordable cognitive therapist chatbot capable of proving expert advice on mental health. The study aimed to address the increasing need for accessible and cost-effective mental health support by leveraging the capabilities of artificial intelligence and natural language processing. The following presents the achievements of this study concerning the objectives selected.

The study successfully designed and implemented algorithms that enabled the development of a cognitive therapist chatbot. Firstly a dialogue generation model with Mutual Character and Semi-Supervised Multi Task Learners and Awareness was designed. This design was based on a transmitter-receiver model. The model was implemented using different machine-learning techniques and open-source libraries such as PyTorch, ParlAI, and HuggingFace. The model was trained on the reliable processed open-source data gathered from a consultation experts' community website.

A crucial objective of this study was to make mental health support more accessible and affordable. By developing an AI-powered chatbot, the study aimed to provide a cost-effective solution that could be accessed by users at their convenience, without the need for expensive therapy sessions or appointments. This achievement aligns with the goal of reducing barriers to mental healthcare and reaching individuals who may have limited resources or face financial constraints.

The chatbot was trained to provide expert advice on mental health concerns. Through the utilization of relevant datasets and expert knowledge, the study aimed to ensure that the chatbot's responses were informed, accurate, and aligned with established therapeutic principles. By offering expert advice, the chatbot aimed to provide users with valuable insights and guidance to support their mental well-being.

The study also focused on enhancing the user experience to foster engagement and usability. After the model was designed, it was connected to a user-friendly mobile application with a simple yet effective user interface. The design of the user interface and the mobile application prototype aimed to provide a seamless and user-friendly experience, minimizing confusion and maximizing ease of use. The mobile application design allows users of different technological backgrounds to use it with only a few minutes of training.

Finally, the model was tested on automatic elevation and human assasment. The purpose of these evaluations was to assess the model's performance and compare it to existing benchmarks. The GIPE model performed exceptionally on all the evaluations, breaking accuracy in a number of them compared to the benchmarks. t demonstrates that the model not only surpasses existing benchmarks but also generates responses that are highly accurate, relevant, and aligned with human expectations. These findings validate the effectiveness and superiority of the GIPE model in the context of providing expert advice on mental health.

In summary, this study successfully achieved its objectives by designing algorithms that enabled the development of an affordable cognitive therapist chatbot capable of providing expert advice on mental health. The accomplishments in algorithm design, affordability, expert advice provision, and user experience contribute to addressing the pressing need for accessible and cost-effective mental healthcare solutions.

5.2 CONTRIBUTION

The dialogue-generated model designed during this study achieved 82.2% accuracy on Hits@1(%) in the original data and 68.8% on the revised data. Also, 2.66 on human evaluation is an impressive accuracy. The impression of this model in the human evaluation shows excellent promise; hence, its responses are grammatically correct. It also illustrates the chatbot's awareness, coherency, and generative abilities.

A conclusion chatbot developed during this study with high-quality, real-life consultation responses from psychology professionals resulted in a chatbot with traces of empathetic and human-like repose. Furthermore, the replies are informative, which is an excellent promise in case of support for testing in a clinical trial setting. Finally, a simple and user-friendly vocal interactive mobile application is designed for people of different technological backgrounds.

5.3 FUTURE WORK

Its impact can only be tested after a clinical trial for a psychology-related work like this. In the future, involving psychologists to evaluate the model's response similar to the human evaluation technique but considering the responses of the model as psychologically correct and informative can help improve the model tremendously.

Also, involving patients can provide the most critical feedback for improving the design.

In training, an expert dialogue generation model such as GIPE, a more significant and diverse number of datasets, can help improve the chatbot's responses to be more informative and expert sounding. Also, data from professionals with different techniques can diversify the chatbot's replies and reactions.

Another implementation to be considered is connecting this chatbot to a physical device to smaller devices such as smartwatches or voice-activating devices that can help patients reach out anytime needed without carrying their phones.

REFERENCES

Abas, A. S., & Sukaimi, S. A. (2018). Dr Wan Azizah: Time to review Malaysia's mental health policy. In *New Straits Times*.

Alberti, C., Lee, K., & Collins, M. (2019). A bert baseline for the natural questions. *arXiv preprint arXiv:1901.08634*.

Allen, S. (2020). COVID-19 Is Straining Mental Health—Could Technology Be the Answer?. *IEEE pulse*, *11*(4), 8-13.

Alves, B., Oliveira, I., Pratas, C., & Pereira, A. (2019, February). A mobile application to complement face-to-face interactions in psychological intervention for social anxiety management. In *2019 IEEE 6th Portuguese Meeting on Bioengineering (ENBENG)* (pp. 1-4). IEEE.

Lamberterie, C. de. (1990). *Les adjectifs grecs en -us: Sémantique et comparaison*. Amazon. Retrieved April 17, 2020, from https://developer.amazon.com/en-US/alexa

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Bernama. (2020, October 10). *Survey finds almost half a million Malaysians experience symptoms of depression.* Retrieved December 22, 2020,from https://www.theedgemarkets.com/article/survey-finds-almost-half-million-malaysians-experience-symptoms-depression

Bertagnolli, N., Lord, G., Ström, E., & Lee, P. (2020, October 7). *Counsel chat: Bootstrapping high-quality therapy data*. Medium. Retrieved December 2020, from https://towardsdatascience.com/counsel-chat-bootstrapping-high-quality-therapy-data-971b419f33da

Britz, D. (2015). Understanding convolutional neural networks for NLP. *URL: http://www. wildml. com/2015/11/understanding-convolutional-neuralnetworks-for-nlp/(visited on 11/07/2015)*.

Brunier, A., & Drysdale, C. (2020, October 5). *COVID-19 disrupting mental health services in most countries, WHO survey*. Retrieved December 19, 2020, from https://www.who.int/news/item/05-10-2020-covid-19-disrupting-mental-health-services-in-most-countries-who-survey

Butler, A. C., Chapman, J. E., Forman, E. M., & Beck, A. T. (2006). The empirical status of cognitive-behavioral therapy: a review of meta-analyses. *Clinical psychology review*, 26(1), 17-31.

Carlo, A. D., Hosseini Ghomi, R., Renn, B. N., & Areán, P. A. (2019). By the numbers: ratings and utilization of behavioral health mobile applications. *NPJ digital medicine*, *2*(1), 54.

CounselChat. (n.d.). *Mental health answers from counselors*. CounselChat. Retrieved December 2020, from https://counselchat.com/

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Dahl, G. E., Sainath, T. N., & Hinton, G. E. (2013, May). Improving deep neural networks for LVCSR using rectified linear units and dropout. In *2013 IEEE international conference on acoustics, speech and signal processing* (pp. 8609-8613). IEEE.

Dino, F., Zandie, R., Abdollahi, H., Schoeder, S., & Mahoor, M. H. (2019, November). Delivering cognitive behavioral therapy using a conversational social robot. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 2089-2095). IEEE.

Fader, S. (2017, June 05). *How Much Does Therapy Cost? Can I Afford To See A Counselor?* Retrieved January 02, 2021, from https://www.betterhelp.com/advice/therapy/can-i-afford-to-see-a-counselor-how-much-does-therapy-cost/

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR mental health*, *4*(2), e7785.

Get healthier. One step at a time. (n.d.). Retrieved January 03, 2021, from https://www.naluri.life/

Google. (n.d.). *Dialogflow documentation | google cloud*. Google. Retrieved September 6, 2021, from https://cloud.google.com/dialogflow/docs

Gu, J. C., Ling, Z. H., Zhu, X., & Liu, Q. (2019). Dually interactive matching network for personalized response selection in retrieval-based chatbots. *arXiv preprint arXiv:1908.05859*.

Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., & Keysers, C. (2012). Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends in cognitive sciences*, *16*(2), 114-121.

Hill, F., Bordes, A., Chopra, S., & Weston, J. (2015). The goldilocks principle: Reading children's books with explicit memory representations. *arXiv preprint arXiv:1511.02301*.

Hofmann, S. G., Asmundson, G. J., & Beck, A. T. (2013). The science of cognitive therapy. *Behavior therapy*, 44(2), 199-212.

Hu, B., Lu, Z., Li, H., & Chen, Q. (2014). Convolutional neural network architectures for matching natural language sentences. *Advances in neural information processing systems*, *27*.

Industry market research, reports, and Statistics. IBISWorld. (2016, March). Retrieved May 20, 2020, from http://www.ibisworld.com/industry/behavioral-therapists.html

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.

Jayachandran, C. M., & Shyamala, K. (2017, February). Kohonen SOM deployment in android app-based cognitive behavioral therapy for personality disorders. In *2017 World Congress on Computing and Communication Technologies (WCCCT)* (pp. 255-257). IEEE.

Jia-Chen Gu, Zhen-Hua Ling, Xiaodan Zhu, and Quan Liu. 2019. Dually interactive matching network for personalized response selection in retrieval-based chatbots. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019*, Hong Kong, China. Association for Computational Linguistics.

Jurinec, N., & Schienle, A. (2020). Utilizing placebos to leverage effects of cognitive-behavioral therapy in patients with depression. *Journal of affective disorders*, *277*, 779-784.

Kaur, D., & Whalley, D. (2020, July 05). *What is Cognitive Behavioral Therapy* (CBT)? Retrieved January 02, 2021, from https://www.psychologytools.com/self-help/what-is-cbt/

Këpuska, V., & Bohouta, G. (2017, November). Improving wake-up-word and general speech recognition systems. In *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)* (pp. 318-321). IEEE.

Khalaf, A. A., Hashim, A. H. A., Olowolayemo, A., & Funke, R. (2021). Artificial Intelligent Applications for Mental Health Support: A Review paper. *Engineering Professional Ethics and Education 2021 (ICEPEE'21)*, 22.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kwiatkowski, T., Palomaki, J., Redfield, O., Collins, M., Parikh, A., Alberti, C., ... & Petrov, S. (2019). Natural questions: a benchmark for question answering research. *Transactions of the Association for Computational Linguistics*, *7*, 453-466.

Lee, J., Sabrina, S., Khaw, C., Salim, Z., & Sivakumar, Y. (2019, March 27). This *founder once made AirAsia x financially fit. now he makes m'sians physically fit.* Vulcan Post. Retrieved March 5, 2023, from https://vulcanpost.com/659717/naluri-health-life-coaching-service-app-malaysia/

Lewis, M., He, L., & Zettlemoyer, L. (2015, September). Joint A* CCG parsing and semantic role labelling. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (pp. 1444-1454).

Lewis, M., Yarats, D., Dauphin, Y. N., Parikh, D., & Batra, D. (2017). Deal or no deal? end-to-end learning for negotiation dialogues. *arXiv preprint arXiv:1706.05125*.

Li, J., Galley, M., Brockett, C., Gao, J., & Dolan, B. (2015-c). A diversity-promoting objective function for neural conversation models. *arXiv preprint arXiv:1510.03055*.

Li, J., Galley, M., Brockett, C., Spithourakis, G. P., Gao, J., & Dolan, B. (2016-b). A persona-based neural conversation model. *arXiv preprint arXiv:1603.06155*.

Li, J., Monroe, W., Ritter, A., Galley, M., Gao, J., & Jurafsky, D. (2016-a). Deep reinforcement learning for dialogue generation. *arXiv preprint arXiv:1606.01541*.

Liu, Q., Chen, Y., Chen, B., Lou, J. G., Chen, Z., Zhou, B., & Zhang, D. (2020). You impress me: Dialogue generation via mutual persona perception. *arXiv preprint arXiv:2004.05388*.

Marcheggiani, D., Frolov, A., & Titov, I. (2017). A simple and accurate syntax-agnostic neural model for dependency-based semantic role labeling. *arXiv preprint arXiv:1701.02593*.

Mazaré, P. E., Humeau, S., Raison, M., & Bordes, A. (2018). Training millions of personalized dialogue agents. *arXiv preprint arXiv:1809.01984*.

Mensio, M., Rizzo, G., & Morisio, M. (2018, April). The rise of emotion-aware conversational agents: threats in digital emotions. In *Companion Proceedings of The Web Conference 2018* (pp. 1541-1544).

Mental health is a human right. (2018, May 24). Retrieved January 01, 2021, from https://www.ohchr.org/EN/NewsEvents/Pages/MentalHealthIsAhumanright.aspx

Montufar, G. F., Pascanu, R., Cho, K., & Bengio, Y. (2014). On the number of linear regions of deep neural networks. *Advances in neural information processing systems*, *27*.

N. (2018, October 22). *Costs of Mental Health Treatment in Malaysia*. Retrieved January 02, 2021, from https://imfunemployed.com/2018/10/17/costs-of-mental-health-treatment-in-malaysia/

N. Panchal et al., (2020) *The implications of COVID-19 for mental health and substance use, Kaiser Family Foundation*, Apr. 2020, [online] Available: https://www.kff.org/coronavirus-covid-19/issue-brief/the-implications-of-covid-19-for-mental-health-and-substance-use/.

Online therapy and medication for depression and anxiety. Youper. (n.d.). Retrieved November 10, 2020, from https://www.youper.ai/

Online therapy with a licensed therapist. (n.d.). Retrieved January 02, 2021, from https://try.talkspace.com/affiliate?utm_content=93052422

Paperno, D., Kruszewski, G., Lazaridou, A., Pham, Q. N., Bernardi, R., Pezzelle, S., ... & Fernández, R. (2016). The LAMBADA dataset: Word prediction requiring a broad discourse context. *arXiv preprint arXiv:1606.06031*.

Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002, July). Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics* (pp. 311-318).

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... & Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, *32*.

Patel, F., Thakore, R., Nandwani, I., & Bharti, S. K. (2019, December). Combating depression in students using an intelligent chatBot: a cognitive behavioral therapy. In *2019 IEEE 16th India Council International Conference (INDICON)* (pp. 1-4). IEEE.

Punyakanok, V., Roth, D., & Yih, W. T. (2008). The importance of syntactic parsing and inference in semantic role labeling. *Computational Linguistics*, *34*(2), 257-287.

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI blog, 1*(8), 9.

Ritter, A., Cherry, C., & Dolan, B. (2011, January). Data-driven response generation in social media. In *Empirical Methods in Natural Language Processing (EMNLP)*.

Roth, M., & Lapata, M. (2016). Neural semantic role labeling with dependency path embeddings. *arXiv preprint arXiv:1605.07515*.

Sainath, T. N., Vinyals, O., Senior, A., & Sak, H. (2015, April). Convolutional, long short-term memory, fully connected deep neural networks. In *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 4580-4584). IEEE.

Schatzmann, J., Weilhammer, K., Stuttle, M., & Young, S. (2006). A survey of statistical user simulation techniques for reinforcement-learning of dialogue management strategies. *The knowledge engineering review*, *21*(2), 97-126.

Serban, I., Sordoni, A., Bengio, Y., Courville, A., & Pineau, J. (2016, March). Building end-to-end dialogue systems using generative hierarchical neural network models. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 30, No. 1).

Serban, I., Sordoni, A., Lowe, R., Charlin, L., Pineau, J., Courville, A., & Bengio, Y. (2017, February). A hierarchical latent variable encoder-decoder model for generating dialogues. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 31, No. 1).

Shah, D. (2020, July 28). *Generative chatbots using the seq2seq model*! Retrieved November 15, 2020, from https://towardsdatascience.com/generative-chatbots-using-the-seq2seq-model-d411c8738ab5

Shang, L., Lu, Z., & Li, H. (2015). Neural responding machine for short-text conversation. *arXiv preprint arXiv:1503.02364*.

Sharma, B., Puri, H., & Rawat, D. (2018, April). Digital psychiatry-curbing depression using therapy chatbot and depression analysis. In *2018 Second International Conference on Inventive Communication and Computational*

Technologies (ICICCT) (pp. 627-631). IEEE.

Sheikh, S. A., Tiwari, V., & Singhal, S. (2019, September). Generative model chatbot for human resource using deep learning. In *2019 International Conference on Data Science and Engineering (ICDSE)* (pp. 126-132). IEEE.

Song, H., Zhang, W. N., Cui, Y., Wang, D., & Liu, T. (2019). Exploiting persona information for diverse generation of conversational responses. *arXiv preprint arXiv:1905.12188*.

Song, Y., Yan, R., Li, X., Zhao, D., & Zhang, M. (2016). Two are better than one: An ensemble of retrieval-and generation-based dialog systems. *arXiv preprint arXiv:1610.07149*.

Strubell, E., Verga, P., Andor, D., Weiss, D., & McCallum, A. (2018). Linguistically-informed self-attention for semantic role labeling. *arXiv preprint arXiv:1804.08199*.

Sutton, R. S., McAllester, D., Singh, S., & Mansour, Y. (1999). Policy gradient methods for reinforcement learning with function approximation. *Advances in neural information processing systems*, *12*.

Täckström, O., Ganchev, K., & Das, D. (2015). Efficient inference and structured learning for semantic role labeling. *Transactions of the Association for Computational Linguistics*, *3*, 29-41.

Toutanova, K., Haghighi, A., & Manning, C. D. (2008). A global joint model for semantic role labeling. *Computational Linguistics*, *34*(2), 161-191.

UN High Commissioner for Refugees (UNHCR), *The Sustainable Development Goals and Addressing Statelessness*, March 2017, available at: https://www.refworld.org/docid/58b6e3364.html [accessed 6 September 2020]

Varghese, E., & Pillai, M. R. (2018, April). A standalone generative conversational interface using deep learning. In *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)* (pp. 1915-1920). IEEE.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural* *information processing systems, 30*.

Vinyals, O., & Le, Q. (2015). A neural conversational model. *arXiv preprint arXiv:1506.05869*.

Vlad Serban, I., Klinger, T., Tesauro, G., Talamadupula, K., Zhou, B., Bengio, Y., & Courville, A. (2016). Multiresolution Recurrent Neural Networks: An Application to Dialogue Response Generation. *arXiv e-prints*, arXiv-1606.

Williams, R. J. (1992). Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Reinforcement learning*, 5-32.

Wolf, T., Sanh, V., Chaumond, J., & Delangue, C. (2019). Transfertransfo: A transfer learning approach for neural network based conversational agents. *arXiv preprint arXiv:1901.08149*.

World Health Organization. (n.d.). *Depression*. World Health Organization. Retrieved December 19, 2020, from https://www.who.int/news-room/fact-sheets/detail/depression

World Health Organization. (2022, March 2). *Covid-19 pandemic triggers 25% increase in prevalence of anxiety and depression worldwide.* World Health Organization. Retrieved April 6, 2022, from https://www.who.int/news/item/02-03-2022-covid-19-pandemic-triggers-25-increase-in-prevalence-of-anxiety-and-depression-worldwide

Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., & Hovy, E. (2016, June). Hierarchical attention networks for document classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies* (pp. 1480-1489).

Zhang, S., Dinan, E., Urbanek, J., Szlam, A., Kiela, D., & Weston, J. (2018). Personalizing dialogue agents: I have a dog, do you have pets too?. *arXiv preprint arXiv:1801.07243*.

APPENDIX I: Publications

INTERNATIONAL CONFERENCE ON ENGINEERING PROFESSIONAL ETHICS AND EDUCATION(ICEPEE'21)

Artificial Intelligent Applications for Mental Health Support: A Review paper

Ayesheh Ahrari Khalaf[1], Aisha Hassan Abdalla Hashim[1], Akeem Olowolayemo[2], and Rashidah Funke Olanrewaju[1].

[1] *Department of Electrical and Computer Engineering, Faculty of Engineering, IIUM, Kuala Lumpur, Malaysia*
[2] *Department of Computer Science, Faculty of Information and Communication Technology, IIUM, Kuala Lumpur, Malaysia*

Abstract— Mental health is every human's right. Physical health cannot be achieved without a healthy mind. Good mental health helps dealing with the stress of life, affects physical health, and allows building and maintaining strong human relationships. Individuals with a healthy mind can be a valuable part of society and contribute to their community while feeling fulfilled and satisfied. While its importance is obvious, almost everyone at some points in their lives faces some life-altering experiences that could challenge their mental health, such as trauma, abuse and family problems, genetics, and lifestyles. At the moment, the treatment options are majorly limited to attending psychotherapy, medical therapy, and self-help.

Unfortunately at the moment, the option of attending a psychotherapy session is not available or affordable for everyone around the world. Also, accessibility to professional therapists is limited. Besides that even though therapists are trained to be unbiased and fair, it is always harder for humans to overcome their personal preferences and biases. Sometimes that is way easier to control and manipulate with machines. This paper will review the latest advancements in technology, from mobile applications to robots, that are designed to help with mental health matters.

Keywords: *Artificial Intelligence, Machine Learning, Mental Health, Therapeutic Chatbot, Deep Learning Approaches.*

1. INTRODUCTION

The World Health Organization (WHO) reported in January 2020 that more than 264 million are suffering from depression. It also mentioned depression as the leading cause of disability, and it may lead to suicide [1]. After the COVID-19 pandemic, WHO reported in October 2020, that based on their recent survey, less than 1% of international budgets for health is dedicated to mental health. While during the pandemic mental health services were disrupted in 93% of countries around the world [2].

The National Institute of Mental Health reported that more than a quarter of American adults face some sort of depression or anxiety every year [3]. Based on the study done in 2017 on 273,203 individuals in Malaysia, more than 6.7% of them were suffering from some level of depression [4]. The National Health and Morbidity Survey (NHMS 2019) reported, almost half a million Malaysians are experiencing some signs of depression. Also, 424,000 children are facing mental health issues [5]. Based on a Kaiser Family Foundation poll, about half of American adults are concerned about the effects of the COVID-19 pandemic on their mental health [6]. It can only conclude, the increase of depression and mental health in Malaysia as well.

More than 165,000 healthcare apps are available in smartphones app stores at the moment, of which only 6% are related to mental health matters [7]. As mentioned in the UNHCR's Sustainable Development Goals (SDG), Fig 1, SDG1: no poverty, SDG2: no hunger, SDG3: good health and well-being and SDG16: peace and justice are every human being's right. UNHCR published an article in May 2018, "Mental health is a human right". The complication of participating in daily activities such as attending school or work, personal and communal responsibilities with an ill mind is mentioned. Mr. Dainius Pūras states while it is evident "there cannot be health without mental health", there still is not merely as much attention and budget allocated to it compared to physical health anywhere in the world [8].

Generative Interactive Psychotherapy Expert (GIPE) Bot

Ayesheh Ahrari Khalaf [1†], Prof. Dr. Aisha Hassan Abdalla Hashim[2††], Asst. Prof. Dr. Akeem Olowolayemo, and Assoc. Prof. Dr. Rashidah Funke Olanrewaju,
aahraryk1992@gmail.com aisha@iium.edu.my
International Islamic University Malaysia, Gombak, Kuala Lumpur, Malaysia

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. 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. A smart generative cognitive behavioral therapy with spoken dialogue systems support was then developed using a model Persona Perception (P2) bot with Generative Pre-trained Transformer-2 (GPT-2). The model 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.

Keywords:
Machine Learning, Mental Health, Therapeutic Chatbot, Deep Learning Approaches, GPT-2.

1. Introduction

In January 2020, the World Health Organization (WHO) estimated that more than 264 million people worldwide experience depression [1]. 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 fewer than 2% of global health funds are allocated to mental health [2].

Currently, just 6% of 165,000 healthcare applications accessible in smartphone application stores are focused on mental health issues [3].

In Cognitive Behavioral Therapy (CBT), several cognitive and behavioral interactions are used as a well-known and scientifically validated treatment [4]. 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 [5]. Cognitive-behavioral approaches are therefore presented and practiced during treatment sessions with classwork to internalize the new behavior [6]. 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 depicts the efficacy of CBT according to Kaur and Whalley, 2020 [6].

Spoken dialog systems are recently finding their way into all intelligent 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

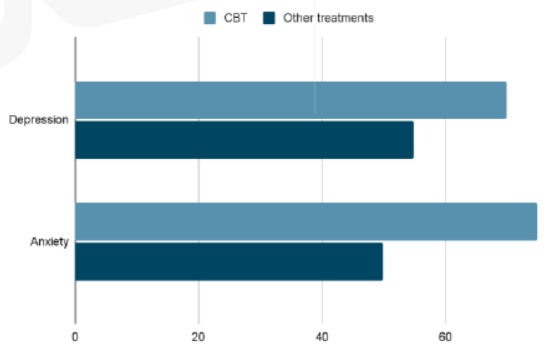

Figure 1: The effectiveness of CBT compared to other treatments

Manuscript received April 13, 2023
Manuscript revised
https://doi.org/

APPENDIX II: Codes

training_transmitter.py

```
import os
import random
import torch
from agents.transmitter.transmitter import ARCH_CHOICE
from parlai.scripts.train_model import setup_args as setup_dict_args, TrainLoop

# if is original, train model on original data; otherwise on revised data.
IS_ORIGINAL = False

TRANSMITTER_DIR = './tmp/transmitter'
VERSION = "transmitter_revised"


def setup_task():
    if IS_ORIGINAL:
        task_name = 'tasks.convai2transmitter.agents:SelfOriginalTeacher'
    else:
        task_name = 'tasks.convai2transmitter.agents:SelfRevisedTeacher'
    return task_name


def setup_seed(seed=706123):#1706123
    # random seed, to evaluate the performance
    torch.random.manual_seed(seed)
    torch.cuda.manual_seed(seed)
    random.seed(seed)


def gpt_setting():
    return 2, 1e-4, 'gpt_custom', 1.0  #10


def lstm_setting():
    return 8, 3, 'sgd', 0.1 #64


def setup_args():
    """
    Use create test env setting
    :return: opt
    """
```

```
    parser = setup_dict_args()
    exp_name = VERSION
    n_epoches = 100
    beam_size = 2
    encode_layers = 2
    decode_layers = 2
    embedding_size = 256
    turn_emed_size = 50
    encoder_turn_use = False
    encoder_dis_use = False
    encoder_hidden_size = 1024
    decoder_hidden_size = 1024
    encode_max_seq_len = 256
    decode_max_seq_len = 32
    smoothing = 0.05
    dropout = 0.1
    embedding_type = 'glove'
    momentum = 0.9
    persona_append_strategy = 'concat'
    history_append_strategy = -1
    select_persona = False
    shuffle_persona = True
    share_decoder_input_output_embed = False
    num_train_epochs = 4

    if ARCH_CHOICE == 'gpt2':
        batchsize, lr, optimizer, gradient_clip = gpt_setting()
    else:
        batchsize, lr, optimizer, gradient_clip = lstm_setting()

    task_name = setup_task()
    parser.set_defaults(
        task=task_name,
        rank_candidates=False,
        # task='tasks.convai2transmitter.agents:SelfRevisedTeacher:no_cands',
        model='agents.transmitter.transmitter:TransformerAgent',
        model_file='./tmp/transmitter/{}.model'.format(exp_name),
        dict_tokenizer='split',
        datatype='train',
        gpt_lr=6.25e-5,
        n_epoches=n_epoches,
        num_epochs=num_train_epochs,
        batchsize=batchsize,
        beam_size=beam_size,
        encoder_layers=encode_layers,
        decoder_layers=decode_layers,
        encoder_embed_dim=embedding_size,
        encoder_turn_dim=turn_emed_size,
        encoder_turn_use=encoder_turn_use,
        encoder_dis_use=encoder_dis_use,
```

```
        decoder_embed_dim=embedding_size,
        encode_max_seq_len=encode_max_seq_len,
        decode_max_seq_len=decode_max_seq_len,
        select_persona=select_persona,
        shuffle_persona=shuffle_persona,
        persona_append_strategy=persona_append_strategy,
        history_append_strategy=history_append_strategy,
        encoder_bidirectional=False,
        encoder_hidden_size=encoder_hidden_size,
        decoder_hidden_size=decoder_hidden_size,
        smoothing=smoothing,
        lr=lr,
        dropout=dropout,
        encoder_dropout_in=dropout,
        encoder_dropout_out=0,
        decoder_dropout_in=dropout,
        decoder_dropout_out=0,
        share_decoder_input_output_embed=share_decoder_input_output_embed,
        gradient_clip=gradient_clip,
        lookuptable='enc_dec',
        optimizer=optimizer,
        embedding_type=embedding_type,
        momentum=momentum,
        # rough enough
        validation_max_exs=-1,
        validation_every_n_secs=3600,
        validation_metric='ppl',
        validation_metric_mode='min',
        validation_patience=10,
        log_every_n_secs=30,
        gpu=0,
        # logging configuration
        exp=exp_name,
        tensorboard_log=True,
        tensorboard_tag='exp',
        train_report_metrics='ppl,f1,hits@1',
        tensorboard_metrics='ppl,f1,hits@1',
    )
    return parser


if __name__ == '__main__':
    opt = setup_args()
    setup_seed()
    TrainLoop(opt).train()
```

AYESHEH AHRARI K.   MSCIE   2023   IIUM