A Framework to Assist Chat Operators of Mental Healthcare Services

Thiago de Oliveira Madeira, Heder Soares Bernardino, Jairo Francisco de Souza,
Nathália Munck Machado, Bruno Marcos Pinheiro da Silva, Alexandre Vieira Pereira Pacelli
Universidade Federal de Juiz de Fora, Brazil
{madeira, heder, jairo.souza}@ice.ufjf.br
{muncknathalia, brunomarcospsilva, alexandrevpp}@gmail.com

Henrique Gomide
Universidade Federal de Viçosa, Brazil
henriquepgomide@gmail.com

Abstract

Conversational agents can be used to make diagnoses, classify mental states, promote health education, and provide emotional support. The benefits of adopting conversational agents include widespread access, increased treatment engagement, and improved patient relationships with the intervention. We propose here a framework to assist chat operators of mental healthcare services, instead of a fully automated conversational agent. This design eases the adverse effects of applying chatbots in mental healthcare. The proposed framework is capable of improving the quality and reducing the time of interactions via chat between a user and a chat operator. We also present a case study in the context of health promotion on reducing tobacco use. The proposed framework uses artificial intelligence, specifically natural language processing (NLP) techniques, to classify messages from chat users. A list of suggestions is offered to the chat operator, with topics to be discussed in the session. These suggestions were created based on service protocols and the classification of previous chat sessions. The operator can also edit the suggested messages. Data collected can be used in the future to improve the quality of the suggestions offered.

1 Introduction

Due to recent advances in Natural Language Processing (NLP), chatbots are being developed and used in different domains, such as customer support, voice assistant, and medicine. Particularly, chatbots are used in medicine to diagnose medical conditions based on patients' symptoms (Srivastava and Singh, 2020), to classify mental states (Patel et al., 2019), and to promote health education (Brixey et al., 2017).

Besides, chatbots have been developed for providing advice and education on mental health conditions. Pereira and Díaz (2019) reviewed the academic literature and found applications that targeted neurological disorders (e.g., insomnia, dementia, depression), well-being, addictions, sexually-transmitted-diseases, among others. The authors pointed out that the field is still in its youth and is more focused on developing rather than testing and assessing efficacy. They also suggested that chatbots are more likely to promote health and behavior change (e.g., Andersson and Cuijpers (2009)) if integrated with human support, which had been overlooked by studies included in their review.

However, physicians believe that actual chatbots cannot effectively care for all of the patients' needs (Palanica et al., 2019). Standard chatbots cannot display human emotion, and cannot provide detailed diagnosis and treatment due to their limitation in consider all of the factors of the patient (Palanica et al., 2019). Palanica et al. (2019) also stated that healthcare chatbots can be a risk to patients when the patients do not fully understand a diagnosis.

In this work, we describe an open-source framework for developing a chatbot with human support using well-known NLP libraries. Then, we showcase an application for promoting smoking cessation using the framework.

2 Related Work

In this section, we discuss the ethical implications of developing chatbots for mental health. We also present examples of applications of healthcare assistance, emphasizing their benefits and their distinct architectural designs. Lastly, we distinguish our proposed framework from the current literature showing why it can potentially be an improvement.

People diagnosed with a mental disorder are not always inclined to seek treatment (Corrigan

et al., 2014). Bendig, Erb, Schulze-Thuesing, and Baumeister (2019) discuss the causes of this behavior, such as concerns about social opinion and negative attitudes towards drug-based treatment options, negative experiences with professional caregivers, lack of insight into their illness, and accessibility barriers like time or location (shift workers and rural communities). Chatbots are a potential solution as they can be available 24/7 over an internet connection (Cameron et al., 2018; Abd-Alrazaq et al., 2020).

It is indispensable to address ethical and social implications when applying Artificial Intelligence (AI) in healthcare (Fiske et al., 2019; Kretzschmar et al., 2019). The first thing to note is that regulations are often general and are one step behind AI's advances. Fiske et al. (2019) provided an analysis of the risks and benefits of implementing AI solutions to mental health from an ethical perspective. According to the authors, conversational agents can potentially stop working and incorporate human biases. Also, security is a high priority due to the nature of the information. Conversational agents users must be aware that they are not interacting with a human, but with an AI. Benefits include new opportunities for reaching patients (e.g., fear of stigmatization), increase treatment engagement, and improve patients' response (Fiske et al., 2019).

In addition to the ethical perspective, most healthcare conversational agents still have to be tested in randomized controlled trials. It can better determine how well the agents can assist a patient in the long run (Abd-Alrazaq et al., 2020). Research should also focus on providing better guidelines for chatbots development (Fiske et al., 2019).

Conversational agents are designed accordingly with their goals. They can have a specific task to accomplish. Symptoma (Martin et al., 2020) and Aquabot (Mujeeb et al., 2017) are examples of conversational agents for specific tasks. Symptoma differentiates more than 20,000 diseases, whereas Aquabot diagnoses Autism and Achluophobia (the fear of darkness).

On the other hand, there are non-specific task agents, such as Vik (Chaix et al., 2019) and Clara (Miner et al., 2020). The former helps patients diagnosed with breast cancer, their relatives, and friends with advice and reminders. The latter is used to share information, suggest behavior, and offer emotional support during the COVID-19 pandemic.

As this work aims to implement a conversational assistant and apply it to promote smoking cessation, it is suitable to examine the impacts of a chatbot acting in this environment. In a two-arm controlled trial, Perski et al. (2019) compared the standard version of the pro version of the Smokefree app against the standard version plus a chatbot. A total of 54,214 smokers participated in the study. After one month, researchers compared the groups for engagement (number of login sessions) and self-reported quit rates. They found that chatbot plus the standard version led to a 101% increase of engagement. However, quit rates did not differ statistically (Perski et al., 2019).

There are some examples of real-time messaging recommendations. In (TouchPal, 2008; Microsoft, 2010), keyboard applications recommend emojis based on the typed words. Gmail has a smart reply feature to suggest short responses to emails (Henderson et al., 2017). An example of a tool that supports humans in a conversation is SolutionChat (Lee et al., 2020). The framework proposed by the authors of SolutionChat can assist the moderators of a discussion group. It provides an environment where multiple users can discuss matters, express their thoughts, and vote for potential solutions. There is the presence of a human moderator to guide the debate. The moderators in charge of managing discussions are often overloaded. SolutionChat can offer suggestions to moderators and ends up promoting time-saving and even quality improvement to the discussion. According to the authors, their work is the first moderator assistance system for online chat conversation to combine summarization and real-time messaging suggestions.

Moreover, to the best of our knowledge, SolutionChat is the framework that comes closest to our proposal. Both can read messages and assist a human operator as suggesting intents. The difference between the frameworks lies in their objectives. SolutionChat's authors designed it for management purposes. It aims to identify discussion stages and featured opinions in a structured discussion. Our proposed framework aims to answer questions and offer information in the form of a *Question and Answer* approach. Also, a human is present here to confirm the suggestions. Consequently, this supervision can be used to provide implicit feedback. Healthcare is an area where the adoption of a fully automated chatbot is delicate. The presence of a

human guiding the conversation is desirable.

To overcome the risks of deploying a fully automated chatbot and benefit from the NLP techniques' advantages, we propose a conversational agent that assists human operators. The framework can classify users' utterances and provides content suggestions.

3 General Architecture

A framework to provide support to an operator of an assistance chat in the healthcare field is proposed here. Instead of being a fully automated chatbot, the proposal provides support for the chat operators. The main benefits of using the proposal are improving the quality of the conversation and reducing its time. When chat operators are supplanted (which is a latent tendency for our case study), the framework assists their training, giving them suggestions for the conversation in practice. More details are in Section 4. This section describes the general architecture, which considers the benefits of advances in NLP and overcomes the opposing sides of applying chatbots in healthcare.

In the proposed framework, the user's intent is identified when a message is sent to the chat operator. When a user sends multiple messages, they are gathered together into a single one. As a result, one intent is predicted for this new merged message. Next, a set of potential answers related to that predicted intent is selected and displayed to a chat operator. The operator then selects and edits the message. Subsequently, the operator sends the answer to the user. Data generated during the framework's execution is collected to improve the quality of the classification and suggestions. Figure 1 presents the proposed framework.

3.1 Intents and Suggestions

Users of chatting services may have many different intents. Chatbots must be aware of the users' intent to trigger a proper response. An architectural strategy to handle it is via a set of predefined intents defined by domain specialists. This set covers as much as possible queries that a chatbot may encounter (Srivastava and Prabhakar, 2020). It is worth mentioning that in addition to domain-specific queries, conversational agents are also susceptible to receive unexpected or unprompted messages [1]. An intent set covering greetings, ac-

---

[1] https://rasa.com/docs/rasa/
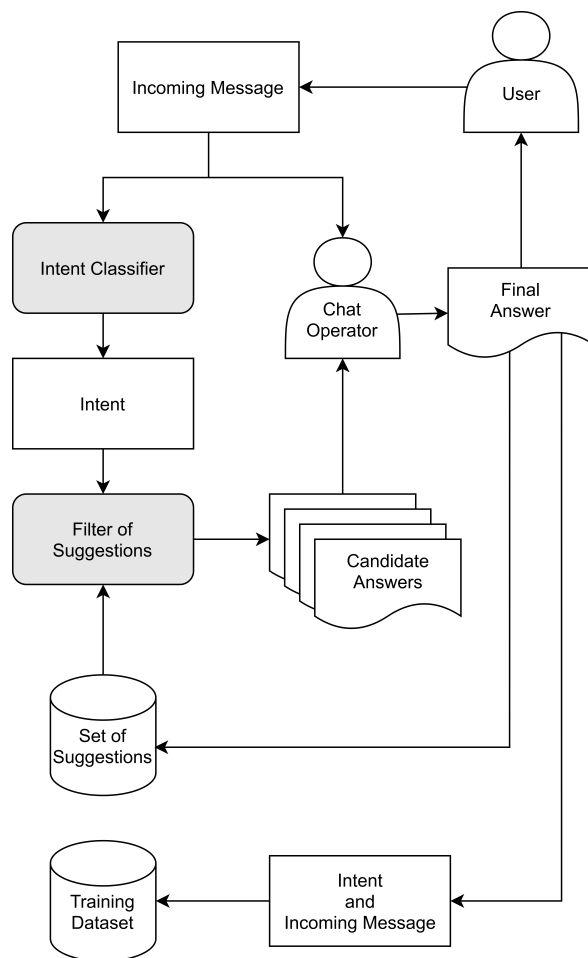dialogue-elements/small-talk/

Figure 1: General Architecture

knowledgments, reactions, and off-topic chitchat can manage surprising situations.

Domain specialists should also anticipatedly design a set of suggestions for each intent previously established. The proposed framework has a filter component responsible for displaying suggestions related to the intent predicted by the classifier to the chat operator. However, the predicted intent may not adequately address the user's utterance. In this case, the framework has to handle a misclassification. The next section discusses the classification problems and also the training set used to fit the model.

3.2 Classification and Training Set

To classify a user's utterance's intent is the same to identify what the user is trying to accomplish with its interaction. An intent classifier is trained with dialogue utterances labeled with its intents. The classification consists of predicting the intent of a given user's utterance. It is a one-off problem, where each user's utterance is associated with a

single intent (Schuurmans and Frasincar, 2019).

A chatbot may misclassify the intent of a user's utterance (Joigneau, 2018). To overcome this obstacle and enhance the accuracy of intent classifiers, Joigneau (2018) proposes methods to perform re-classifications. However, the proposed framework does not intend to substitute a human completely. Its main goal is to support a real-time conversation between a user and a chat operator. Therefore, each misclassified intent can be correctly labeled by the human operating the chat. We define it as a mis-classification when the framework can not predict at least one intent with a probability of σ higher than a threshold σ_l. In Section 4, we describe a case study of our framework and how we handled the misclassification issue.

The data used for training the classifier consists of a corpus labeled with intents from the fixed set of predefined intents. Corpus's text can either be extracted from real conversations or manually crafted. Data augmentation can also be used to increase a corpus' size (Wei and Zou, 2019). The quality and size of the dataset can impact on the classifier's accuracy (Srivastava and Prabhakar, 2020).

3.3 Feedback Module

The framework includes a feedback module. The feedback occurs in two situations. In the first situation, there is implicit feedback. When the chat operator uses a suggestion from the framework, it means that the classifier correctly predicted the intent of the user's utterance. In this case, the framework uses the input data to improve the training dataset. The training dataset incorporates the pair of the user's utterance and the predicted intent.

In the second situation, the feedback is slighter more explicit. It occurs when the chat operator edits a suggested message. The proposed framework stores the new message in the set of suggestions and can recommend it in the future.

Each time the chat operator uses a suggestion without editing, the framework adds up an internal score for that suggestion. The higher the suggestion's score, the higher is the priority for the framework to display it to the chat operator.

4   Case Study: Viva sem Tabaco

Viva sem Tabaco (VST) is a web-based intervention for smoking cessation. The website's content was adapted from evidence-based guidelines for treating tobacco addiction (Gomide et al., 2016).

VST provides information, quizzes, personalized quit smoking plans, and a chat. In this platform's chat, the chat operator is a counselor. A counselor is responsible for identifying a user's concerns and answering appropriately. The counselors are under-graduate students of health courses, trained by the psychologists' team. Due to the healthcare area's delicateness, it is desirable that counselors are well trained and adequately follow the intervention's guidelines.

A team of psychologists guarantees the quality of the VST platform's content. They create the web site's content and train new counselors. The psychologists also composed a document containing instructions for the chat interactions, specifying how the counselors should assist a user looking for help on VST. According to this document, the counselor must identify the user's need and answer with appropriate content from the VST's website.

The current version of the open-source implementation of the proposed framework is available in the Python programming language[2]. We are incorporating the framework into VST to enhance the counselor's performance by (i) keeping the conversation focused and avoid ambiguities, and (ii) reducing the response time. We used spaCy and Rasa. The former is a free, open-source library for Natural Language Processing. We imported its pre-trained word embedding for Portuguese. The latter is an open-source machine learning framework, developed to implement contextual AI assistants and chatbots. Rasa is used for generating models to classify intents of an utterance. Rasa uses SpaCy's word embeddings models to represent texts numerically.

4.1   Intents and Suggestions

Through the analysis of the document of chat assistance's instructions, we created a set of potential intents, which are listed in Table 1. The proposed framework classifies the user's utterances into one of these intents.

The team of psychologists also conceived a set of predetermined suggestions for each intent of Table 1. For each intent, the framework exhibits the set of suggestions associated with it to the counselor. So, the counselor can choose one of the suggestions and reply to the user's message. Alternatively, the message can be edited to fit the conversation bet-

---

[2] https://github.com/t-madeira/Chatbot_VsT-AeS

4

Greetings
Technical Issues
Chitchat
Acknowledgments
Farewell
TBD (To Be Determined)
Benefits of stopping and risks of smoking
Financial costs
Withdrawal and craving
Weight Gain
Anxiety and Depression
How to deal with withdrawal
Overcoming cravings
Medication
Ways to quit: gradual or cold-turkey
How to avoid relapses
Learning from previous quit attempts
Identifying slips and relapses
Most common relapses causes

Table 1: Intents defined to classify the user's utterances.

ter. Figure 2 is a representation of the counselor's perspective using the application.

4.2 Classification and Training

For the classification of the user's utterances, Rasa uses the Sklearn Intent Classifier. This classifier consists of an SVM optimized via grid search. The classifier returns probabilities σ associated with each intent, making it possible to rank the predicted intents. Due to the small size of the training data used in this case study, the classification process still has a place for improvement via feedback by gathering usage data.

In the ranking of classified intents, there can be at least one with a probability σ higher than a threshold σ_h. In this case, our implementation displays to the counselor the intents associated with the intent with the highest probability. We defined a misclassification when any predicted intent has a probability σ higher than a threshold σ_l. In this case, there is a fallback intent name TBD (to be determined). The counselor still receives suggestions but most likely has to handle the conversation by itself. When this situation occurs, the framework records the message sent by the user for future analysis.

Lastly, the highest intent's probability σ can be higher than σ_l but still lower than σ_h. That be-
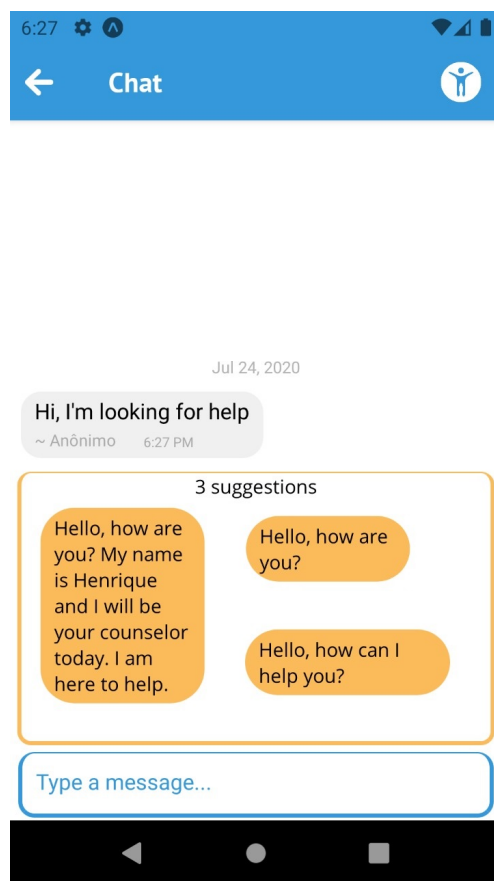

Figure 2: Counselor's view in the implemented application with examples of suggestions provided. In this illustration, the messages are answers for the intent "Greetings".

ing so, our implementation of the framework asks the counselor to solve the uncertainty and manually choose an intent from the top 3 intents of the probabilities ranking.

We used real-world conversations from the VST platform to train the classifying model. The training set consists of 373 interactions between users and counselors. We manually labeled each message presented in an interaction between a user and a counselor with a potential intent from Table 1. If it was a user's message, we labeled as trying to figure out what he or she intended to say or ask. If it was a counselor's message, we labeled as trying to figure out the user's intention that the counselor was trying to answer. We labeled 1100 messages from the 373 interactions.

4.3 Feedback

The framework can improve itself with data gathered from its usage. Each time a counselor sends a message to a user, the framework adds the intent predicted and the user's utterance to the training

5

dataset. If the highest probability σ of an intent classification in the ranking is lower than the threshold σ_h, the counselor solves the uncertainty. In this case, the intent added to the training dataset is the one chosen by the counselor. Another exception is if the framework classifies the user's utterance as TBD. In this case, nothing is added to the training dataset, and the user's utterance is recorded for future analysis.

When the counselor edits a suggestion offered by the framework, the edited suggestion is added to the set of suggestions. As previously explained in Section 3.3, each time the chat operator uses a suggestion without editing, the framework adds up an internal score for it. The higher the suggestion's score, the higher is the priority for the framework to display it to the chat operator.

5 Concluding Remarks and Future Works

In this work, we propose a framework to assist chat operators of healthcare systems. The framework classifies the user's utterances into intents. It provides real-time suggestions to the chat operators of mental healthcare services.

The advantages of adopting the proposed framework include improving the quality and reducing the time of conversations between users and counselors. The conversation's quality is increased due to the assumption that the framework's suggestions reduce ambiguity and rambling in the chat operator's discourse. The conversation's time is reduced due to the real-time suggestions offered by the framework. The chat operator does not waste time overthinking or searching for appropriate content to answer the users. A fully automated approach would be faster, but removing the human from the framework would lead to the negative characteristics of conversational agents present in the literature. Users looking for mental assistance are often mentally weakened, and a human can handle unusual situations. However, through the feedback module and further evaluation, the framework may become fully automated in the future.

Future works include adopting the framework in other health-related domains, gathering and analyzing data of its usage. The framework can be easily adapted to be used in other domains, by adding new training data, sets of intents, and sets of suggestions. Reports from users and counselors can be collected in order to evaluate the framework's efficacy.

References

Alaa Ali Abd-Alrazaq, Asma Rababeh, Mohannad Alajlani, Bridgette M Bewick, and Mowafa Househ. 2020. Effectiveness and safety of using chatbots to improve mental health: Systematic review and meta-analysis. *Journal of Medical Internet Research*, 22(7):e16021.

Gerhard Andersson and Pim Cuijpers. 2009. Internet-based and other computerized psychological treatments for adult depression: A meta-analysis. *Cognitive Behaviour Therapy*, 38(4):196–205.

Eileen Bendig, Benjamin Erb, Lea Schulze-Thuesing, and Harald Baumeister. 2019. The next generation: chatbots in clinical psychology and psychotherapy to foster mental health–a scoping review. *Verhaltenstherapie*, pages 1–13.

Jacqueline Brixey, Rens Hoegen, Wei Lan, Joshua Rusow, Karan Singla, Xusen Yin, Ron Artstein, and Anton Leuski. 2017. Shihbot: A facebook chatbot for sexual health information on hiv/aids. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, page 370–373. Association for Computational Linguistics.

Gillian Cameron, David Cameron, Gavin Megaw, Raymond Bond, Maurice Mulvenna, Siobhan O'Neill, Cherie Armour, and Michael McTear. 2018. Assessing the usability of a chatbot for mental health care. In *International Conference on Internet Science*, pages 121–132. Springer.

Benjamin Chaix, Jean-Emmanuel Bibault, Arthur Pienkowski, Guillaume Delamon, Arthur Guillemassé, Pierre Nectoux, and Benoît Brouard. 2019. When chatbots meet patients: one-year prospective study of conversations between patients with breast cancer and a chatbot. *JMIR cancer*, 5(1):e12856.

Patrick W Corrigan, Benjamin G Druss, and Deborah A Perlick. 2014. The impact of mental illness stigma on seeking and participating in mental health care. *Psychological Science in the Public Interest*, 15(2):37–70.

Amelia Fiske, Peter Henningsen, and Alena Buyx. 2019. Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *Journal of medical Internet research*, 21(5):e13216.

H. P. Gomide, H. S. Bernardino, K. Richter, L. F. Martins, and T. M. Ronzani. 2016. Development of an open-source web-based intervention for brazilian smokers – viva sem tabaco. *BMC Medical Informatics and Decision Making*, 16:103.

Matthew Henderson, Rami Al-Rfou, Brian Strope, Yun-Hsuan Sung, László Lukács, Ruiqi Guo, Sanjiv Kumar, Balint Miklos, and Ray Kurzweil. 2017. Efficient natural language response suggestion for smart reply. *arXiv preprint arXiv:1705.00652*.

Axel Joigneau. 2018. Utterances classifier for chatbots' intents.

Kira Kretzschmar, Holly Tyroll, Gabriela Pavarini, Arianna Manzini, Ilina Singh, and NeurOx Young People's Advisory Group. 2019. Can your phone be your therapist? young people's ethical perspectives on the use of fully automated conversational agents (chatbots) in mental health support. *Biomedical Informatics Insights*, 11:1178222619829083.

Sung-Chul Lee, Jaeyoon Song, Eun-Young Ko, Seongho Park, Jihee Kim, and Juho Kim. 2020. Solutionchat: Real-time moderator support for chat-based structured discussion. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–12.

Alistair Martin, Jama Nateqi, Stefanie Gruarin, Nicolas Munsch, Isselmou Abdarahmane, and Bernhard Knapp. 2020. An artificial intelligence-based first-line defence against covid-19: digitally screening citizens for risks via a chatbot. *bioRxiv*.

Microsoft. 2010. Microsoft swiftkey keyboard.

Adam S Miner, Liliana Laranjo, and A Baki Kocaballi. 2020. Chatbots in the fight against the covid-19 pandemic. *npj Digital Medicine*, 3(1):1–4.

Sana Mujeeb, Muhammad Hafeez Javed, and Tayyaba Arshad. 2017. Aquabot: a diagnostic chatbot for achluophobia and autism. *Int J Adv Comput Sci Appl*, 8.

Adam Palanica, Peter Flaschner, Anirudh Thommandram, Michael Li, and Yan Fossat. 2019. Physicians' perceptions of chatbots in health care: cross-sectional web-based survey. *Journal of medical Internet research*, 21(4):e12887.

Falguni Patel, Riya Thakore, Ishita Nandwani, and Santosh Kumar Bharti. 2019. Combating depression in students using an intelligent chatbot: A cognitive behavioral therapy. In *2019 IEEE 16th India Council International Conference (INDICON)*, page 1–4. IEEE.

Juanan Pereira and Óscar Díaz. 2019. Using health chatbots for behavior change: A mapping study. *Journal of Medical Systems*, 43(5):135.

Olga Perski, David Crane, Emma Beard, and Jamie Brown. 2019. Does the addition of a supportive chatbot promote user engagement with a smoking cessation app? an experimental study. *Digital health*, 5:2055207619880676.

Jetze Schuurmans and Flavius Frasincar. 2019. Intent classification for dialogue utterances. *IEEE Intelligent Systems*, 35(1):82–88.

Prakhar Srivastava and Nishant Singh. 2020. Automatized Medical Chatbot (Medibot). In *2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC)*, pages 351–354, Mathura, Uttar Pradesh, India. IEEE.

Saurabh Srivastava and TV Prabhakar. 2020. Intent sets: Architectural choices for building practical chatbots. In *Proceedings of the 2020 12th International Conference on Computer and Automation Engineering*, pages 194–199.

TouchPal. 2008. Touchpal keyboard.

Jason Wei and Kai Zou. 2019. Eda: Easy data augmentation techniques for boosting performance on text classification tasks. *arXiv preprint arXiv:1901.11196*.