

Theano: A Greek-speaking conversational agent for COVID-19

Nikoletta Ventoura¹, Kosmas Palios¹, Ioannis Vasilakis¹, Georgios Paraskevopoulos^{1,2}, Athanasios Katsamanis^{1,3}, Vassilis Katsouros¹

¹ Institute for Language and Speech Processing, Athena Research Center, Athens, Greece

² School of ECE, National Technical University of Athens, Athens, Greece

³ Behavioral Signals Technologies, Los Angeles, CA, USA

nikoletta.ventoura@athenarc.gr, kosmas.palios@athenarc.gr
yannis.vasilakis@athenarc.gr, g.paraskevopoulos@athenarc.gr
nkatsam@athenarc.gr, vsk@athenarc.gr

Abstract

Conversational Agents (CAs) can be a proxy for disseminating information and providing support to the public, especially in times of crisis. CAs can scale to reach larger numbers of end-users than human operators, while they can offer information interactively and engagingly. In this work, we present Theano, a Greek-speaking virtual assistant for COVID-19. Theano presents users with COVID-19 statistics and facts and informs users about the best health practices as well as the latest COVID-19 related guidelines. Additionally, Theano provides support to end-users by helping them self-assess their symptoms and redirecting them to first-line health workers. The relevant, localized information that Theano provides, makes it a valuable tool for combating COVID-19 in Greece. Theano has already conversed with different users in more than 170 different conversations through a web interface as a chatbot and over the phone as a voice bot.

1 Introduction

The current COVID-19 pandemic has presented a global challenge for citizens, health and government structures, and the global economy. One key characteristic of the current crisis is, that it cannot be combated with centralized handling alone, but requires a significant degree of cooperation from the public. Specifically, citizens adapt their lifestyle by adhering to measures limiting their daily contacts through physical distancing guidelines, stay-at-home orders, and curfews. Additionally, citizens are urged to avoid unnecessary visits to the hospitals. This transformation in everyday life has been a cause of stress for all areas of society. The best way to help people handle this crisis responsibly is to disseminate accurate information about the current state of affairs so that citizens are

well-informed of expert recommendations and the status of the pandemic.

Intelligent conversational agents (CAs) can be a powerful tool for information dissemination and user support. Specifically, CAs have been extensively used for health applications, e.g., to help users quit smoking¹, for mental health support (Cameron et al., 2017, 2018; Grové, 2020), and HIV counselling (van Heerden et al., 2017). CAs have also been used for other social applications, e.g., fact checking (Gupta et al., 2021) and encouraging altruistic behavior (Wang et al., 2019). In the context of the COVID-19 crisis, CAs have been proposed for symptom self-checking, i.e., Clara², information dissemination, i.e., HealthyBuddy³, and combating misinformation, i.e., Jennifer⁴ (Li et al., 2020).

One issue that arises is that most CAs are created for English-speaking demographics, but in many cases the policies and information about the pandemic are region-dependent. Developing applications for under-resourced languages is challenging, due to the lack of data and the limited user base (Hovy and Spruit, 2016). Nevertheless, we need to overcome language barriers and to provide contextualized information with respect to local policies, circumstances and statistics. Theano is debunking myths and conspiracy theories that are prevalent in Greece, such as the negative effects that masks have in children or the involvement of the Greek church in COVID-19 policies. We have not added information about fake news that have spread abroad and do not interest the Greek public (e.g., drinking disinfectant to stop COVID-19). Therefore, it is essential to develop localized, native CAs.

In this work, we present Theano, a Greek speak-

¹WHO Florence

²CDC Clara

³UNICEF, WHO/Europe HealthBuddy+

⁴Jennifer

ing CA for COVID-19. Theano is a multipurpose CA, aiming to inform and support the Greek population. Specifically, Theano provides up-to-date localized statistics about the pandemic and expert guidelines. Furthermore, Theano debunks common myths about the use of masks, the need for vaccination, etc. Besides informing users, Theano can also support them, by helping them to self-assess the severity of their symptoms and direct them to the best avenues for receiving expert help. We make Theano available through a user-friendly web interface as a chatbot and through a Twilio telephone number as a voice assistant. The virtual assistant has already conversed with different users in more than 170 different conversations. Theano is continuously improved based on user feedback, by retraining the model using user conversations and including frequently requested features and improvements. Code is available as open-source⁵

2 Background Work

From a technical standpoint, CAs are systems that integrate multiple components in order to understand the user’s query (Natural Language Understanding – NLU), isolate spans of interest in the input phrases (Entity Extraction) and select an appropriate action or response based on the input (Dialogue Management). Initial NLU systems were based on rule-based systems that use elaborate grammars (Allen, 1988; Dowding et al., 1993) and statistical language modeling (Suen, 1979), while recent research focuses on neural approaches, i.e., deep belief networks (Sarikaya et al., 2014), Recurrent Neural Networks (RNNs) (Yao et al., 2013) and Transformer (Vaswani et al., 2017) models (Liu et al., 2019; Dong et al., 2019; Bunk et al., 2020). Similarly, research for Entity Extraction (or Named Entity Recognition) systems, has moved from rule-based approaches and elaborate feature engineering (Mikheev et al., 1999; Collins and Singer, 1999) to approaches based on Conditional Random Fields (Lafferty et al., 2001) and neural networks (Chiu and Nichols, 2016; Devlin et al., 2019). Dialogue Management approaches include hierarchical plan-based systems (Bohus and Rudnicky, 2009), Hierarchical RNNs (Serban et al., 2016) and Transformers (Vlasov et al., 2020). These components have been integrated in dialogue frameworks that provide a streamlined developer experience. Popular open-source dialogue frameworks include Deep-

⁵GitLab Repository

pavlov (Burtsev et al., 2018) and Rasa (Bocklisch et al., 2017)

3 Conversation Design

Theano aims to provide a first line of support for end users during the pandemic. Our design choices are based on two key goals: a) keeping users well-informed about the current status of the pandemic in order to enable them to make good health choices and b) reducing panic by helping users self-assess their symptoms and directing them to first-line health-care providers. This objective has also been suggested by McKillop et al. (2020) and Følstad et al. (2018).

On the one hand, Theano provides general information about COVID-19. There is access to the number of COVID-19 new cases and deaths in Greece, other countries, and worldwide. Moreover, she knows the availability of Intensive Care Units, as well as the number of people that have been vaccinated on a specific date and in total in Greece. We have also included myth busters concerning the usage of masks and other conspiracy theories, e.g., that masks delay the development of children’s lungs. There is a wide range of Frequently Asked Questions about COVID-19 supported, from how it started, ways of transmission and protection to how to wash hands or whether one should be wearing gloves in public.⁶

On the other hand, Theano can list the common COVID-19 symptoms and in case the user experiences symptoms, through a mixed initiative conversation, she starts a diagnostic survey. In all stages of the dialogue, she clearly states that she is not a doctor and the user should speak with the local authorities if they continue having symptoms. She can also find pharmacies that are open in the area that the user requests. In addition, Theano can advise users who have been in contact with a COVID-19 patient, i.e., she suggests that they get tested and informs them where they can obtain free tests provided by local authorities.

To better facilitate the fulfilment of the two goals mentioned earlier, namely, being informative, and also keeping users calm, we have created a relevant Persona for Theano. The importance of analytical design of the Persona is analyzed by Pearl (2016). Theano has the persona of a support agent that is honest, polite, and direct, but at the same time friendly. The core design principles we followed

⁶See Section 5 for our data sources.

are:

Engagement. We want Theano to be part of the daily life of users, so she has to be interesting and not only respond to questions. [Ruane et al. \(2019\)](#) highlight the importance of engagement when dealing with precarious topics. Theano tries to keep the users interested in the conversation at all dialogue stages, by suggesting new and relevant conversation topics. Theano is also ready to chitchat, which helps keeping certain users better engaged. Moreover, chitchatting includes general questions about Theano, e.g. age, favourite color and music, holidays, hobbies, weather, languages, so that the users can get to know the Persona in-depth and have an alternative from the COVID-related topics. As shown in [Fig. 3](#), and will be discussed in the following, chitchat has actually been found to be one of Theano’s most popular features.

Consistency. We want users to get the information they ask for, fast. Theano has short responses and doesn’t contradict herself by giving conflicting information. We want the user to trust Theano. Short responses are also preferable for allowing spoken communication with Theano (on the phone). Longer responses on the phone tend to quickly tire users. This objective has been proposed by [Bickmore and Giorgino \(2004\)](#).

Clarity. The topic that Theano is handling is serious, that is the reason why we try to present pieces of information in a clear and concise manner, that is easily understood by everyone. By being explicit in every stage of the dialogue, we also ensure that the user acknowledges the capabilities of the CA and does not expect anything more than what is offered about a sensitive medical topic like COVID-19 ([Laranjo et al., 2018](#)).

Empathy. Theano is handling questions that are serious and have an impact on people’s lives, in a period of profound crisis. She tries to mitigate user fear, comprehends the worries that the users express and offers emotional support with her responses, which is very valuable during the pandemic, as supported by [Miner et al. \(2020\)](#), [De Genaro et al. \(2020\)](#) and [Rashkin et al. \(2019\)](#). For example:

USER : I am positive. What can I do?

THEANO : Since you know you are positive, it is important to stay calm and not panic. You should stay in contact with your doctor and the local authorities. In case your symptoms worsen you should go to the hospital. I think

that it’s all going to be okay. Try to stay calm.

4 System Description

[Fig. 1](#) shows the architecture of Theano. The user’s query is passed through a preprocessing pipeline that normalizes the raw text. Then, textual features are extracted from the raw text and are fed to Entity Extraction and the Intent Classification modules. Extracted entities are augmented with the output of a rule-based module based on External Lookup Tables. The state tracker saves the recognized intent and the extracted entities and passes them to the Dialogue Management module, which determines an appropriate response based on the user’s query. All intermediate states are logged into an external database. Theano is implemented using the Rasa open-source dialogue framework, which provides a clean way to implement our dialogue pipeline.

In detail, for pre-processing, we use an in-house spell-checker based on a convolutional sequence to sequence architecture. The spell checker is trained on various synthetically corrupted Greek texts and can correct spelling and accenting errors for small Greek phrases. Online conversations in Greek often contain Greek words written in Latin characters (also known as Greeklisch). Therefore, we also include as a preprocessing step a state-machine based Greeklisch to Greek translator. The clean text is finally tokenized on whitespaces. During the Featurization step, we extract Bag of Words and N-gram features for the input tokens.

Entity extraction is performed using a Conditional Random Field (CRF) Named Entity Recognition (NER) module ([Lafferty et al., 2001](#); [Bocklisch et al., 2017](#)). Additionally, we use Duckling⁷, which provides a regex Lookup Table for time and date entities and aggregate the extracted entities with the CRF outputs. For example, for the message “What pharmacies will be open in Athens tomorrow morning at 8?”, the CRF will extract “Athens” for the entity “city” and Duckling will extract the correct date and time string.

Intents are organized in a hierarchy, where high-level intents are recognized first (e.g., FAQ), and then fine-grained intents are extracted. This decision allows easily scaling up the dialogue system to a large number of intents with limited training data. Therefore, intent recognition is performed using a hierarchy of Dual Intent Entity Transformer (DIET) ([Bunk et al., 2020](#)) classifiers. DIET is

⁷[Duckling Github repository](#)

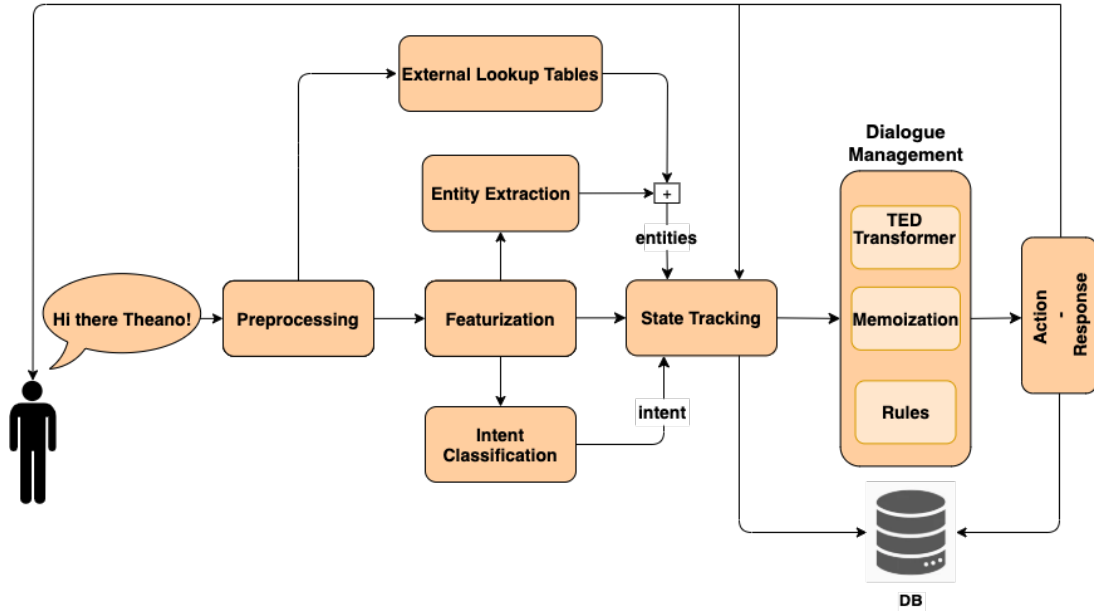


Figure 1: Architecture overview of Theano. The key components are the Preprocessing and Featurization steps, the Entity Extraction, the Intent Classification, and the Dialogue Management.

a Transformer based architecture that operates on Bag of Words features. The sparse features are embedded using a feedforward layer and then encoded using a 2-layer Transformer encoder with relative positional embeddings in the attention layers (Shaw et al., 2018). The first DIET classifier classifies coarse-grained intent classes, and then the input is routed to the second level of classifiers (i.e. response selectors). The output is the fine-grained recognized intent and the respective classifier confidence. An intent is considered to be recognized if the confidence level exceeds a specified threshold, otherwise a fallback strategy is adopted (e.g. “I did not understand your question. Can you rephrase?”).

For dialogue management, we use three policies with different levels of complexity and priority. First, we use a Rule-based policy that explicitly maps a small, selected set of important intents to the desired actions or responses. Then, we fall back to a memoization policy that retrieves identical user inputs from past conversations. If the rule-based policy and the memoization policy do not produce a response, we use TED policy (Vlasov et al., 2020), a Transformer-based classifier for dialogue act classification. TED receives the user inputs, recognized intents, and extracted entities across a fixed history window, and outputs the appropriate response with respect to the user input.

It is important that detailed statistics and logs of past conversations are kept, in order to continu-

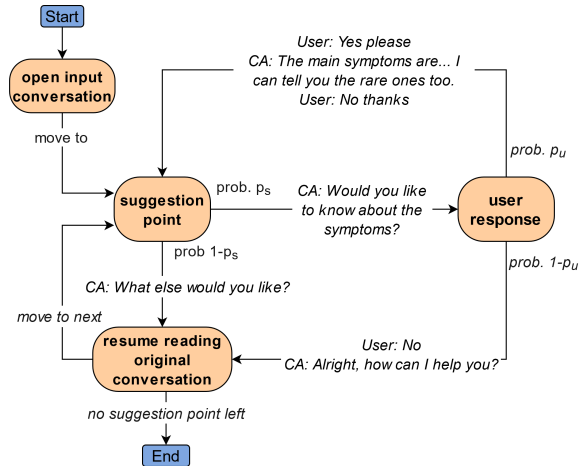


Figure 2: Overview of the synthetic data generator

ously improve Theano’s performance. To this end, we track all conversations, and anonymized versions of the user inputs, recognized intents, and extracted entities are saved into a database. Thus, we can evaluate past system performance and improve the dialogue stories using expert knowledge from linguists.

4.1 Smart suggestions

An early test release of Theano showed us that users are typically unaware of the themes that the CA covers; consequently, they used to ask questions that Theano could not respond to while ignoring many of her functionalities. The dialogue flow did not allow transitioning between themes because the

dialogue was limited to start with a question from the user, an exchange of messages on the topic, and, in the end, the CA would “What else can I do for you?”. At that point, we realized that this question, which is at the end of every sub-conversation, is also what we call a suggestion point: a point where the bot could possibly say “I know about Y too, would you like me to tell you?”. Suggestion Y should not be a random topic but something new to the conversation. Another question is how often the CA should suggest something; there is not an easy answer to this. We have experimented with the value of p_s , the probability of the CA making a suggestion upon reaching a suggestion point.

To implement this feature, we have to augment the training data of the TED policy with examples of conversations driven - at least partially - by the CA’s suggestions. Due to the large amount of possible suggestions, instead of manually generating example conversations, we opt for a synthetic data generation process. During this process, we augment existing conversations in our training set with one or more additional turns that are driven by Theano’s suggestions.

Fig. 2 shows the process of generating the augmented conversation samples. The input is an existing conversation with one or more suggestion points as well as p_s and p_u . The latter is used to model how willing a user is to accept a suggestion.

When the generator reaches a suggestion point, a suggestion loop begins. Firstly, it has to decide whether to suggest something from a topic previously not discussed (with probability p_s) or to leave the suggestion point unaltered, i.e., a generic prompt. If the generator makes a suggestion, then the proposed topic is either rejected or accepted by the simulated user with probability p_u . If the user rejects the suggestion, then the script adds the user’s refusal and proceeds to the rest of the conversation. Otherwise, if the user accepts the suggestion, the generator appends a sub-conversation on the selected topic, which is followed by another suggestion point. The aforementioned procedure keeps repeating until either no suggestion is made or the user has chosen to disagree.

5 Integrations

Data Sources: Table 1 shows the data sources we have integrated into our backend to obtain the latest COVID-19 statistics, suggestions and facts. We prefer to use official or trusted sources, e.g., CDC,

Data Source	Extracted Information
Covid API ⁸	Foreign #cases , #deaths statistics
Nyrro’s public domain ⁹	Greek #cases, #deaths, ICU availability statistics
European Centre of Disease Control (ECDC) ¹⁰	FAQs, diagnostic form
Centre of Disease Control (CDC) ¹¹	FAQs
Our World in Data (OWiD) ¹²	Vaccine progress
Vrisko.gr ¹³	Pharmacy locations, open hours
World Health Organization (WHO) ¹⁴	FAQs, mask facts
National Public Health Organization (NPHO) ¹⁵	Greek FAQs, mask facts

Table 1: Data sources for COVID-19 stats and facts.

WHO, NPHO etc. Our main sources for statistics are the Covid API, OWiD and Nyrro’s public data. Covid API and OWiD are trusted sources for global COVID-19 related statistics, while Nyrro’s data is an individual’s effort to aggregate latest COVID-19 statistics about Greece. Nyrro’s is not an official source, but we opt to use it, because it is more up to date than OWiD and easier to access programmatically than statistics provided by Greek official sources (i.e. NPHO). Nevertheless we continuously cross-check the accuracy of Nyrro’s statistics with the official sources, both manually and through regression tests. For COVID-19 related facts and expert suggestions, our main sources are the FAQs provided by ECDC, CDC, WHO and NPHO. Finally, in order to locate open pharmacies near the user’s location, we utilize vrisko.gr, a site that provides information about the location and open hours of businesses through an API.

Voice and Chat bot: In addition to the described chat functionality, Theano is also available as a voice bot. We integrate Google Automatic Speech Recognition (ASR) and Text to Speech (TTS) services for the Greek language. We use a master service, which receives voice input through websockets, routes the intermediate outputs to the appropriate microservices and then streams the voice response to the user. Overall, Theano is available through the following channels:

Web chat: Sends textual input to the master service

⁷Covid API

⁸Nyrros’ Spreadsheet

⁹ECDC FAQ on Covid-19

¹⁰CDC FAQ on Covid-19

¹¹OWiD Github repository

¹²Pharmacies on duty through Vrisko

¹³WHO FAQ on Covid-19

¹⁴NPHO FAQ on Covid-19

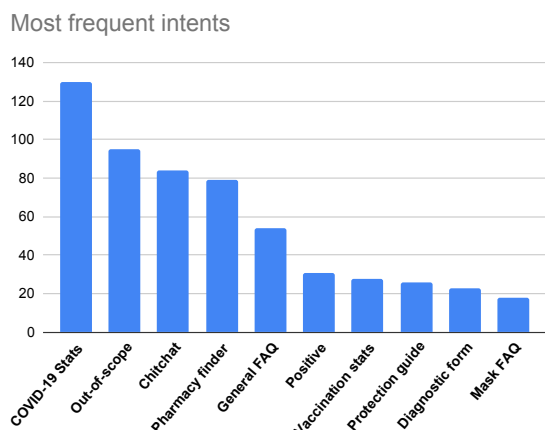


Figure 3: 10 most frequent intents asked by users in April 2021

text endpoint.

Twilio: Allows consuming voice streams from the telephone network. It receives the phone traffic, and forwards it to the master service websocket.

Web-based Voice bot: Connects directly to the master service, using websockets.

6 System Evaluation

We have constructed a training set that consists of 3719 real and synthetic conversations, which is constantly evolving based on real-world usage. Our dataset includes 37 core intents at the first level of the intent hierarchy. Four of these intents, i.e. Chitchat, General FAQ, Mask FAQ and Out-of-scope, extend to a second level in the hierarchy with 13, 31, 13 and 10 sub-intents respectively.

Fig. 3 shows the ten most frequent questions asked by users during a release in April 2021. There is a preference in the COVID-19 statistics and the pharmacy finder features, with the out-of-scope response selector as a close second. The latter includes topics that Theano does not cover, for example, lockdown updates or hospital operating hours; instead of resulting in a fallback action, in this case, Theano redirects the user to the appropriate website. An interesting fact is the tendency of users to chit-chatting.

We train DIET classifier for the 37 core intents for 10 epochs with batch size 32. We use the default values provided by Rasa for all other hyperparameters. Similarly we train separate classifiers for the sub-intent categories of Chitchat, Mask FAQ, General FAQ and Out of Scope for 25, 60, 50 and 50 epochs respectively. This yields an overall intent

recognition accuracy of $92.2\% \pm 0.002$ for the core 37 intents and $92.0\% \pm 0.002$ for the 67 sub-intents during 5-fold cross-validation. The respective F1 scores are $92.1\% \pm 0.008$ and $91.8\% \pm 0.016$.

However, these proportions fall down to 86% when the system interacts with real users. During inference, if classifier confidence is small (i.e. less than 0.5), or the classifier is unsure (i.e. confidence for top two intents close – less than 0.1 difference) we choose the intent Fallback action.

For the dialogue manager, we configure memoization policy to memorize a window of 5 dialogue turns. TED classifier is configured with a history window of 10 turns and we set the output dimension to 64. We use default values for all other hyperparameters. TED is trained for 8 epochs with batch size 16, yielding 95.3% test accuracy. During inference we choose the fallback response if the output confidence of TED is less than 0.4.

Additionally to the test accuracy, we measure intent classification accuracy when the system is used by real users in the wild. Fig. 4 shows the confusion matrices over two user evaluation periods. The first user evaluation was performed in December 2020 in tandem with the first release of Theano, when we received 101 conversations. The second evaluation was completed in April 2021 with the second release of Theano, when we received 73 conversations. All the incoming data have been annotated by linguists. One thing to notice is that Theano was updated with more intents during the second release. Included intents are selected based on user feedback and the current developments. While more intents are introduced, the overall NLU performance is improved (79.5% accuracy in December 2020 versus 86.7% recognition accuracy in April). Finally we observe in both evaluation periods, most misclassified intents are successfully captured by the fallback strategy.

6.1 Qualitative comparison with other CAs

Table 2 shows the main features of Clara, Jennifer, and Theano. We examine five main functionalities that at least one of the CAs supports, or partially supports, which means that even if it does not provide an interactive response, it redirects the user to an appropriate website.

The main purpose of Clara is to provide self checking. The chatbot has a specific purpose, i.e. symptom self-checking. Clara does not allow for arbitrary textual input, instead allows users to se-

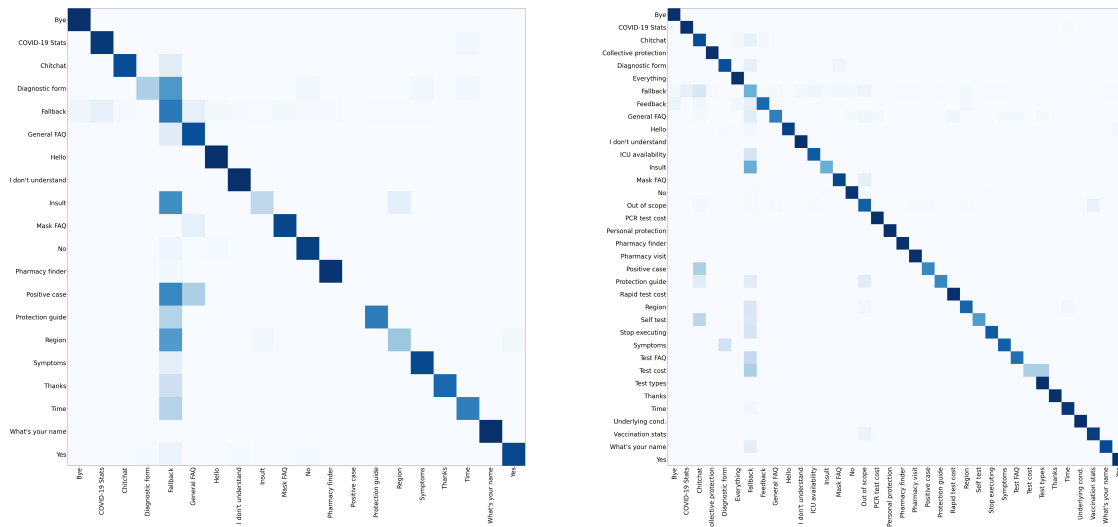


Figure 4: Confusion matrices for intent recognition in real conversations over two periods of user evaluation, i.e. December 2020 (left) and April 2021 (right).

	Theano	Jennifer	Clara
Language	GR	EN	EN
Voice interface	✓	✗	✗
COVID-19 statistics	✓	✓	✗
Diagnostic Form	✓	✗	✓
Treatment details	✗	✓	✗
Ways of protection	✓	✓	✓
Vaccine information	✓*	✓	✗
Chitchat	✓	✓	✗

Table 2: Comparison of supported features of Clara, Jennifer and Theano. ✗: not supported, ✓: supported, ✓*: partially supported

lect one query among a set of fixed choices. The purpose of this decision is to have a clear understanding of the symptoms, assess them, consult the user according to the responses, and give instructions about the ways of protection. The diagnostic forms have similar elements both in Theano and Clara, as both of them follow the same methodology. In particular, Theano and Clara present the main symptoms and ask the user which symptoms they experience. They then proceed with the more rare ones and finally advise the user per the best avenues of medical care they should seek.

Jennifer and Theano have similar functionalities, but the responses have different goals, apart from the common purpose of accurately informing the users. Jennifer’s responses are longer, and at the end, she redirects the user to her sources, whereas Theano focuses more on user engagement by giving

smaller, condensed answers, unless the user asks for details. Furthermore, Jennifer handles questions that she cannot reply with a diplomatic way by apologizing and suggesting that her developers will add this feature in the future. In comparison, Theano tries to overcome the fallback, by suggesting the topics that she is aware of, without explaining why she could not respond to the specific question. Jennifer offers general information about vaccines, for example which antiviral drugs are used, something that Theano is not capable of. Theano’s goal is the dissemination of information that the users understand, without using specialized medical vocabulary. Both Jennifer and Theano cover similar chitchat topics; they give information about themselves (e.g., age, location, functionalities) and their interests (e.g., hobbies). The only difference is the content of their responses because Jennifer has a robotic persona, whereas Theano is anthropomorphic. To be more precise, a robotic persona means that the CA does not incorporate human-like elements, while a more anthropomorphic one imitates human characteristics (e.g., Theano has a favourite holiday island). From our evaluation, we present the responses that the two CAs have in the same question.

Conversation example with Theano:

USER : Do you have any hobbies?
 THEANO : Sure. I love listening to music.

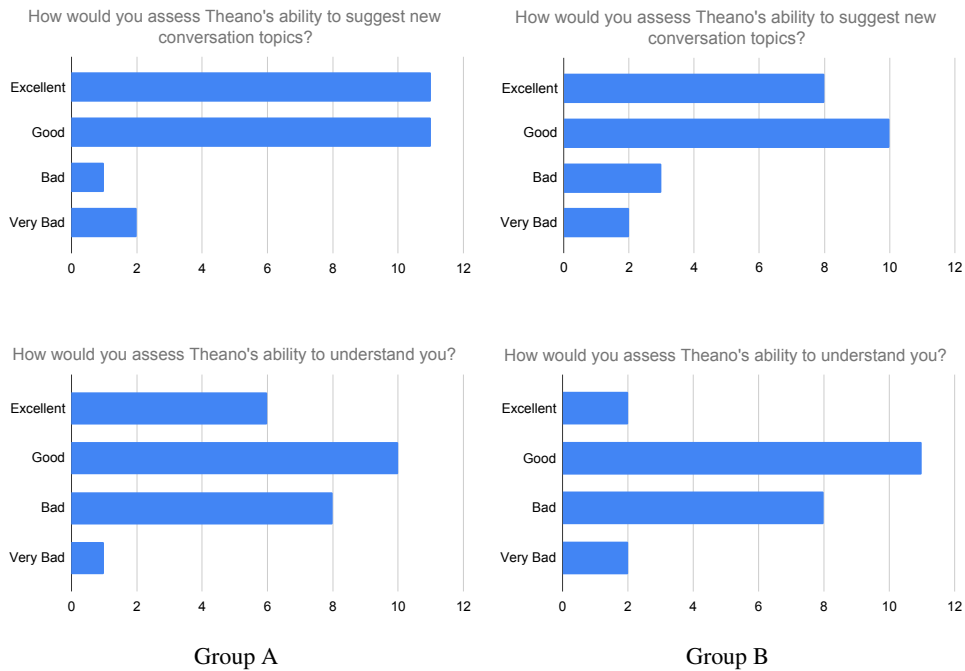


Figure 5: User assessment during A/B testing of the smart suggestion feature. We ask users to provide an assessment of Theano’s comprehension ability and about the new conversation suggestion feature. For Group A we suggest new topics with $p_s = 0.2$, while for Group B with $p_s = 0.8$.

USER : What kind of music do you like?
 THEANO : I prefer alternative Greek music.
 USER : Where are you?
 THEANO : I live in Athens.

Conversation example with Jennifer:

USER : Do you have any hobbies?
 JENNIFER : understanding people - what they think and how they feel
 USER : Where are you?
 JENNIFER : I live in the cloud - my address would be something like: 128.0.0.0

While Jennifer and Clara are specialized chatbots for information dissemination and symptom self-checking respectively, Theano includes both functionalities. The role of Theano is not limited to the dissemination of accurate information concerning COVID-19, but also to provide a holistic support to the user by being empathetic, engaging, clear, and consistent. Jennifer and Clara are available for English-speaking users, while Theano is available in Greek. Finally, Theano includes a voice interface, whereas users can only interact with Jennifer and Clara via chat.

6.2 Smart Suggestion - A/B testing

An A/B test with real users was performed to evaluate the new smart suggestion feature, and specifically the optimal suggestion probability p_s . During the A/B test, users interacted with two versions of Theano. The first version, presented to Group A, was configured with $p_s = 0.2$ (few suggestions), while the second version, presented to Group B, was configured with $p_s = 0.8$ (lots of suggestions). Seventy three individuals aged 15-65 participated in the A/B test, and the system divided them into groups randomly (36 in Group A and 37 in Group B). At the end of the interaction, the users completed a survey. In Fig. 5 we present user assessments for two survey questions regarding the new suggestion feature and Theano’s overall comprehension ability.

Users prefer the CA suggestions to be less intrusive, as, in Group A, the Excellent to Bad and Very Bad ratio is greater than the one from Group B. Also, the normalized sum of Bad and Very Bad counts is less than the respective sum from Group B. A remarkable observation is that the assessment of Theano’s comprehension ability is worse in Group B than in Group A. Although the similarity between the two groups is noteworthy, Group A has overall better reviews. This fact indicates that deeper, CA-

driven conversations may be more prone to intent recognition errors. With a higher p_s we increase the engagement, but it appears to negatively affect how much the user feels understood by the CA. This suggests that the optimal value for p_s is somewhere between 0.2 and 0.8, with the purpose of engaging the user without losing the mutual understanding between user and CA.

Furthermore, the average number of conversation turns was measured as a proxy to assess user engagement, which is 10.74 and 13.06 average turns for Group A and B respectively. These figures indicate that conversations last longer when the CA directs the user by making suggestions frequently, even though this may upset users that want to have better control of the conversation.

7 Conclusions and Future Work

We present Theano, a CA for COVID-19 information dissemination and symptom self-checking in Greek. We believe, Theano is a valuable contribution to the list of digital tools for battling COVID-19, as it provides a feature-rich and scalable support system tailored to the needs of our user-base. Theano needs to be constantly updated with new intents and conversation topics without sacrificing NLU performance, in order to keep up with the latest developments. We show that this is possible; in the second iteration, we double the number of supported intents and even improve the intent recognition performance in the real-world user evaluation. We also propose a simple smart suggestion feature, in order to improve user engagement and interactively show Theano's capabilities to users. During our A/B testing, we receive encouraging -though not conclusive- results, that this can improve user engagement, as the use of this feature leads to higher number of conversation turns.

In the future, we want to keep improving Theano's abilities by continuously adding intents based on our users needs. We also want to introduce a reinforcement learning based smart suggestion module, for open domain smart suggestions. Finally, we want to develop a large scale deployment of Theano and reach a wider user base.

Acknowledgements

- We want to thank the reviewers for their constructive feedback and our user-base for helping us improve Theano.

- This research has been co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project safety4all with code:T1EDK-04248).

References

- James Allen. 1988. *Natural language understanding*. Benjamin-Cummings Publishing Co., Inc.
- Timothy Bickmore and Toni Giorgino. 2004. Some novel aspects of health communication from a dialogue systems perspective. In *AAAI Fall Symposium on Dialogue Systems for Health Communication*, pages 275–291.
- Tom Bocklisch, Joey Faulkner, Nick Pawlowski, and Alan Nichol. 2017. [Rasa: Open source language understanding and dialogue management](#). *CoRR*, abs/1712.05181.
- Dan Bohus and Alexander I Rudnicky. 2009. The ravenclaw dialog management framework: Architecture and systems. *Computer Speech & Language*, 23(3):332–361.
- Tanja Bunk, Daksh Varshneya, Vladimir Vlasov, and Alan Nichol. 2020. [Diet: Lightweight language understanding for dialogue systems](#).
- Mikhail Burtsev, Alexander Seliverstov, Rafael Airapetyan, Mikhail Arkhipov, Dilyara Baymurzina, Nickolay Bushkov, Olga Gureenkova, Taras Khakhulin, Yuri Kuratov, Denis Kuznetsov, et al. 2018. Deppavlov: Open-source library for dialogue systems. In *Proceedings of ACL 2018, System Demonstrations*, pages 122–127.
- Gillian Cameron, David Cameron, Gavin Megaw, Raymond Bond, Maurice Mulvenna, Siobhan O'Neill, Cherie Armour, and Michael McTear. 2017. Towards a chatbot for digital counselling. In *Proceedings of the 31st International BCS Human Computer Interaction Conference (HCI 2017) 31*, pages 1–7.
- 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.
- Jason PC Chiu and Eric Nichols. 2016. Named entity recognition with bidirectional lstm-cnns. *Transactions of the Association for Computational Linguistics*, 4:357–370.

- Michael Collins and Yoram Singer. 1999. Unsupervised models for named entity classification. In *1999 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*.
- Mauro De Gennaro, Eva G Krumhuber, and Gale Lucas. 2020. Effectiveness of an empathic chatbot in combating adverse effects of social exclusion on mood. *Frontiers in psychology*, 10:3061.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186.
- Li Dong, Nan Yang, Wenhui Wang, Furu Wei, Xiaodong Liu, Yu Wang, Jianfeng Gao, Ming Zhou, and Hsiao-Wuen Hon. 2019. [Unified Language Model Pre-training for Natural Language Understanding and Generation](#). In *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*.
- John Dowding, Jean Mark Gawron, Douglas Appelt, John Bear, Lynn Cherny, Robert C Moore, and Douglas B Moran. 1993. Gemini: A natural language system for spoken-language understanding. In *31st Annual Meeting of the Association for Computational Linguistics*, pages 54–61.
- Asbjørn Følstad, Petter Bae Brandtzæg, Tom Feltwell, Effie LC Law, Manfred Tscheligi, and Ewa A Luger. 2018. Sig: chatbots for social good. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–4.
- Christine Grové. 2020. Co-developing a mental health and wellbeing chatbot with and for young people. *Frontiers in Psychiatry*, 11:1664.
- Ankur Gupta, Yash Varun, Prarthana Das, Nithya Muttineni, Parth Srivastava, Hamim Zafar, Tanmoy Chakraborty, and Swaprava Nath. 2021. Truthbot: An automated conversational tool for intent learning, curated information presenting, and fake news alerting. *arXiv preprint arXiv:2102.00509*.
- Alastair van Heerden, Xolani Ntinga, and Khanya Vilakazi. 2017. The potential of conversational agents to provide a rapid hiv counseling and testing services. In *2017 international conference on the frontiers and advances in data science (FADS)*, pages 80–85. IEEE.
- Dirk Hovy and Shannon L Spruit. 2016. The social impact of natural language processing. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 591–598.
- John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning, ICML '01*, page 282–289, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- Liliana Laranjo, Adam G Dunn, Huong Ly Tong, Ahmet Baki Kocaballi, Jessica Chen, Rabia Bashir, Didi Surian, Blanca Gallego, Farah Magrabi, Annie YS Lau, et al. 2018. Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association*, 25(9):1248–1258.
- Yunyao Li, Tyrone Grandison, Patricia Silveyra, Ali Douraghy, Xinyu Guan, Thomas Kieselbach, Chengkai Li, and Haiqi Zhang. 2020. [Jennifer for COVID-19: An NLP-powered chatbot built for the people and by the people to combat misinformation](#). In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*, Online. Association for Computational Linguistics.
- Xiaodong Liu, Pengcheng He, Weizhu Chen, and Jianfeng Gao. 2019. Multi-task deep neural networks for natural language understanding. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4487–4496.
- Mollie McKillop, Brett R South, Anita Preininger, Mitch Mason, and Gretchen Purcell Jackson. 2020. Leveraging conversational technology to answer common covid-19 questions. *Journal of the American Medical Informatics Association*.
- Andrei Mikheev, Marc Moens, and Claire Grover. 1999. Named entity recognition without gazetteers. In *Ninth Conference of the European Chapter of the Association for Computational Linguistics*.
- 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.
- Cathy Pearl. 2016. *Designing voice user interfaces: principles of conversational experiences*. ” O’Reilly Media, Inc.”.
- Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. 2019. [Towards empathetic open-domain conversation models: A new benchmark and dataset](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5370–5381, Florence, Italy. Association for Computational Linguistics.
- Elayne Ruane, Abeba Birhane, and Anthony Ventresque. 2019. Conversational ai: Social and ethical considerations. In *AICS*, pages 104–115.
- Ruhi Sarikaya, Geoffrey E Hinton, and Anoop Deoras. 2014. Application of deep belief networks for natural language understanding. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22(4):778–784.

- Iulian Serban, Alessandro Sordoni, Yoshua Bengio, Aaron Courville, and Joelle Pineau. 2016. Building end-to-end dialogue systems using generative hierarchical neural network models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30.
- Peter Shaw, Jakob Uszkoreit, and Ashish Vaswani. 2018. [Self-attention with relative position representations](#). In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 464–468, New Orleans, Louisiana. Association for Computational Linguistics.
- Ching Y Suen. 1979. N-gram statistics for natural language understanding and text processing. *IEEE transactions on pattern analysis and machine intelligence*, pages 164–172.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. [Attention is all you need](#). In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.
- Vladimir Vlasov, Johannes E. M. Mosig, and Alan Nichol. 2020. [Dialogue transformers](#).
- Xuwei Wang, Weiyan Shi, Richard Kim, Yoojung Oh, Sijia Yang, Jingwen Zhang, and Zhou Yu. 2019. Persuasion for good: Towards a personalized persuasive dialogue system for social good. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5635–5649.
- Kaisheng Yao, Geoffrey Zweig, Mei-Yuh Hwang, Yangyang Shi, and Dong Yu. 2013. Recurrent neural networks for language understanding. In *Inter-speech*, pages 2524–2528.