



# HyLECA: A Framework for Developing Hybrid Long-term Engaging Controlled Conversational Agents

**Erkan Başar**

Communication and Media Group,  
Behavioural Science Institute,  
Radboud University  
Nijmegen, The Netherlands  
erkan.basar@ru.nl

**Divyaa Balaji**

Amsterdam School of Communication  
Research, University of Amsterdam  
Amsterdam, The Netherlands  
d.balaji@uva.nl

**Linwei He**

Department of Communication and  
Cognition, TSHD, Tilburg University  
Tilburg, The Netherlands  
l.he\_1@tilburguniversity.edu

**Iris Hendrickx**

Centre for Language Studies, Centre  
for Language and Speech Technology,  
Radboud University  
Nijmegen, The Netherlands  
iris.hendrickx@ru.nl

**Emiel Kraahmer**

Department of Communication and  
Cognition, TSHD, Tilburg University  
Tilburg, The Netherlands  
e.j.kraahmer@tilburguniversity.edu

**Gert-Jan de Bruijn**

Department of Communication  
Studies, University of Antwerp  
Antwerp, Belgium  
gert-jan.debruijn@uantwerpen.be

**Tibor Bosse**

Communication and Media Group,  
Behavioural Science Institute,  
Radboud University  
Nijmegen, The Netherlands  
tibor.bosse@ru.nl

## ABSTRACT

We present HyLECA, an open-source framework designed for the development of long-term engaging controlled conversational agents. HyLECA's dialogue manager employs a hybrid architecture, combining rule-based methods for controlled dialogue flows with retrieval-based and generation-based approaches to enhance the utterance variability and flexibility. The motivation behind HyLECA lies in enhancing user engagement and enjoyment in task-oriented chatbots by leveraging the natural language generation capabilities of open-domain large language models within the confines of predetermined dialogue flows. Moreover, we discuss the technical capabilities, potential applications, relevance, and adaptability of the system. Lastly, we report preliminary findings from integrating state-of-the-art large language models in simulating a conversation centred on smoking cessation.

## CCS CONCEPTS

• **Human-centered computing** → **Natural language interfaces.**

## KEYWORDS

hybrid conversational agents, task-oriented dialogue systems, natural language generation

## ACM Reference Format:

Erkan Başar, Divyaa Balaji, Linwei He, Iris Hendrickx, Emiel Kraahmer, Gert-Jan de Bruijn, and Tibor Bosse. 2023. HyLECA: A Framework for Developing Hybrid Long-term Engaging Controlled Conversational Agents. In *ACM conference on Conversational User Interfaces (CUI '23)*, July 19–21, 2023, Eindhoven, Netherlands. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3571884.3604404>

## 1 INTRODUCTION

Nowadays, large language models (LLMs) demonstrate daily advancements in their natural language generation (NLG) capabilities. However, their unpredictable nature poses risks for their implementation in chatbots within sensitive domains, such as health communication [3]. Furthermore, their lack of understanding of the world constrains their capacity to engage in complex, nuanced conversations in a long-term setting. Consequently, there is a preference for pre-scripted dialogue management approaches in the application domain, as they allow designers to maintain full control over the dialogue. While preserving conversational integrity is essential, it is equally important to address the primary drawbacks of pre-scripted conversational agents: repetitiveness and lack of flexibility. Designers of pre-scripted chatbots can only anticipate a limited number of user replies, making it labor-intensive to expand the coverage of responses. Consequently, it is reasonable for designers to prioritize relevance over variety, resulting in the creation of generic, catch-all chatbot utterances. While generic utterances frequently adapt well to their context, they may not adequately address users' specific concerns and requests. Issues like inadequate specificity in conversations can be disliked by humans, potentially weakening user engagement [12].

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).  
CUI '23, July 19–21, 2023, Eindhoven, Netherlands  
© 2023 Copyright held by the owner/author(s).  
ACM ISBN 979-8-4007-0014-9/23/07.  
<https://doi.org/10.1145/3571884.3604404>

The application area of chatbots are seemingly endless, with numerous frameworks (e.g., Rasa, DialogueFlow, IBM Watson) available to facilitate the development of chatbots that effectively address short-term, focused tasks, such as customer service. Nevertheless, certain application domains (such as counselling) benefit from chatbots that can maintain multiple sessions to foster meaningful engagement and progress with users. Smoking cessation counselling serves as a prime example, where smokers require long-term support from their counsellors during their quitting process [8]. Despite this need, the amount of smoking cessation-focused conversational agents equipped with long-term capabilities remains insufficient to meet the demand [6]. Furthermore, when considering the aforementioned repetitiveness issue associated with pre-scripted chatbots, the development of an engaging long-term conversational agent with controlled dialogue flows becomes a laborious task, requiring substantial maintenance.

In this paper, we present the **Hybrid Long-term Engaging Controlled Conversational Agents (HyLECA)**, an open-source chatbot development framework designed to create controlled long-term conversational agents with integrated NLG capabilities for enhanced engagement. HyLECA is an implementation of a hybrid architecture that combines rule-based, retrieval-based, and generation-based approaches, aiming to augment the engagement potential of task-oriented chatbots, particularly in highly sensitive domains that necessitate long-term support [2]. Our hybrid approach strives to integrate the strengths of each approach, addressing their respective limitations and complementing one another for a more effective solution.

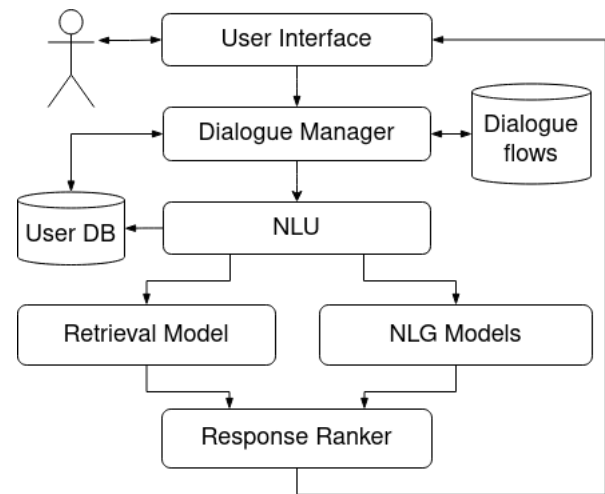
HyLECA is an open-source framework developed in Python, which can be found on a publicly accessible Git repository<sup>1</sup>. The platform serves as a research prototype, designed to investigate the capabilities and applicabilities of the emerging natural language generation technologies. It is not intended for utilizing the outputs of large language models in direct, unsupervised application in sensitive domains.

## 2 RELATED WORK

In recent years, we see an abundance of emerging chatbot applications, showing that a successful chatbot design is not a trivial task [11]. Pre-scripted chatbot design relies on creating dialogue flows where chatbot utterances follow user input. Rasa Core's *stories* allows coding the potential chatbot utterances and actions for a task-oriented dialogue state [4]. These utterances and actions are executed based on the results of intent detection and information extraction performed on the user input. Zamanirad et al. [14] used dialogue acts and intent detection to execute dialogue states to perform the predefined tasks. Similar to these approaches, we implement dialogue states to which specific tasks can be assigned. However, rather than employing intent detection, we adopt a rule-based approach in which the next state is determined based on predefined conditions, ensuring full control over the dialogue flows. Moreover, we emphasize the use of retrieval-based approaches to enhance the specificity and variety of pre-scripted chatbot responses. We also introduce the integration of natural language generation (NLG)

models, which contribute flexibility to the conversation. Papaioanou et al. [10] developed a platform for an ensemble of chatbots, including custom-trained sequence-to-sequence neural generation models, while we leverage the generative capabilities of relatively recently developed open-domain large language models. Numerous other studies have proposed hybrid methods, combining generative and retrieval approaches. For instance, Shalymov et al. [13] introduced GRTr, a platform featuring hybrid generative-retrieval models based on GPT-2 for rapid domain adaptation. In contrast, our proposed system primarily employs rule-based approaches for dialogue management and uses retrieval and generative approaches on top of that.

## 3 SYSTEM ARCHITECTURE



**Figure 1: An overview of the system architecture, showing the data flow in the system.**

The HyLECA system is fundamentally designed to follow pre-determined dialogue flows by implementing a rule-based dialogue management approach. It incorporates all standard components and features typically found in chatbot frameworks, such as a dialogue manager, natural language understanding (NLU) module, response generation, action handling, and slot filling. Additionally, HyLECA records user data and conversation history, enabling the creation of chatbots with long-term interaction capabilities. At its core, the HyLECA dialogue manager serves as the central hub, loading relevant dialogue flows and user data for use as inputs throughout the system. Consequently, the entire system considers specific user information (e.g. name, age) and past conversations during its processing (shown as ‘User DB’ in Figure 1).

The response generation consists of two main components: a retrieval module and a generation module. At each turn, candidate pre-scripted chatbot utterances are chosen according to the corresponding dialogue state and filtered based on the fulfilment of their predefined conditions. These candidate utterances are then processed by the built-in retrieval model for ranking, which is determined by their similarity to the conversation context. Simultaneously, each NLG model within the generation module produces

<sup>1</sup><https://github.com/look-whos-talking-chatbots/hyleca>

response candidates. The generated and retrieved candidates are given to the response ranker, which returns the highest ranking candidate as the selected utterance <sup>2</sup>.

### 3.1 Dialogue States

In our design, the building blocks of dialogue flows are dialogue states. A dialogue state holds the information necessary to execute the corresponding state as designed by the developers of the chatbot. Hereunder, we introduce the utilities of a dialogue state and discuss the possibilities.

**3.1.1 Utterances:** A collection of pre-scripted chatbot utterances can be assigned to a dialogue state, serving as potential responses. These utterances are categorized into two groups: *initiators* and *responses*. Initiator utterances function at the start of a dialogue state, facilitating the discussion topic. Conversely, response utterances are employed after the initial user input is obtained, acting as potential replies and subsequent interactions. HyLECA conveys these pre-scripted utterances to the retrieval model, which identifies the most contextually relevant option. To enhance the chatbot’s coverage and coherence within a dialogue state, it is advisable to provide multiple initiator and response utterances.

**3.1.2 NLU:** The natural language understanding component of the system comprises two primary components: text classification and entity identification. Upon receiving user input, the system employs designated NLU models for the corresponding dialogue state. The detected information is stored in user data for further utilization. While the system employs integrated general-purpose NLU tools (e.g., name recognition, sentiment analysis), developers have the option to incorporate custom functions for each chatbot.

**3.1.3 NLG:** The HyLECA system is engineered to operate with multiple generation models simultaneously. It comes equipped with a few pre-trained open-domain natural language generation models, such as GPT-3.5, GPT-4[9], and DialoGPT[15]. Moreover, it is possible to integrate custom NLG models or other LLMs. This enables the use of NLG models specifically fine-tuned to the context or the topic of specific dialogue states. Alternatively, we can specify that no generation model should be involved in the response generation process for certain dialogue states, thereby enforcing adherence to only rule-based and retrieval-based approaches in those instances.

**3.1.4 Conditions:** The system utilizes a rule-based approach to determine the triggering of each dialogue state and chatbot utterance. This is achieved by binding them to conditions that compare specific user data to a value, utilizing operators such as *equal*, *not-equal*, *greater-than*, *smaller-than*, *includes*, and *not-includes*. All conditions must be satisfied for a state to be triggered or a chatbot utterance to be used in response.

**3.1.5 Actions:** The system supports the implementation and execution of custom actions within the dialogue states. Actions can be any custom functionality required by the chatbot developer, varying from a basic value verification in user data to executing an external API call. Custom actions can be designated to execute at three specific positions: (1) immediately when the dialogue state is

activated, (2) repeatedly after each user input is received within the dialogue state, and (3) just before the dialogue state is terminated.

### 3.2 Processing a Dialogue State

Initially, when a new dialogue state is triggered, the dialogue manager selects an initiator chatbot utterance from a predefined set, as visualized in Figure 2. The dialogue manager is designed to maintain the user in the current dialogue state until at least one of the state’s ending conditions is satisfied. During this time, the dialogue manager may need to repeat the dialogue state. Until then, the state conditions and response utterances can be configured to request the user to rephrase their utterance in the event that the NLU component fails to detect the required information. Additionally, the complexity of the state conditions can be adjusted to generate various scenarios. For example, a condition could be set to obtain a certain number of valid responses to a question. In such a scenario, the dialogue manager can continue to provide chatbot responses to the user input and request further elaboration or alternative ideas.

Upon meeting the termination conditions of the dialogue state, the dialogue manager initiates a counter to keep track of the number of exchanges between the NLG models and the user. The dialogue state is then terminated, and the next state is triggered once the maximum exchange limit is reached. Currently, the dialogue manager is provided with a constant maximum exchange limit. In HyLECA, there is a sequential connection between the dialogue states, allowing for explicit definition of the subsequent state to be initiated after the current state concludes. The state triggers can also incorporate conditions based on the information identified by the NLU module during the conversation, enabling the call of different states.

## 4 LONG-TERM INTERACTIONS

The storage of the conversation history enables chatbots to resume a conversation from the point at which it was left. Although this approach is effective in practice, it can result in an unnatural communication style. For instance, a user may disengage from the conversation and later return to it. It is typical for humans to greet someone after an absence, whereas a chatbot may simply continue the dialogue without acknowledging any elapsed time.

To mitigate this issue, we propose a solution wherein the chatbot initiates a new conversation session if the user remains unresponsive for a predefined duration. This is achieved by the dialogue manager placing the user in a designated dialogue state, referred to as the *on-hold* state. Upon their return, users in the *on-hold* state are greeted with a new message and reminded of the previous dialogue topic. Therefore, after prolonged periods of inactivity, users are not expected to immediately respond with an utterance relevant to the previous dialogue state, but instead simply acknowledge the chatbot’s greeting. This approach ensures that the conversation remains coherent and minimizes the potential for misunderstandings.

## 5 PRELIMINARY EVALUATION

HyLECA has been employed in experimental settings to assess the effectiveness of using chatbots for interventions in two health-related domains: smoking cessation [7] and sexual health promotion

<sup>2</sup>More detail on the response generation process can be found in our previous work [2].

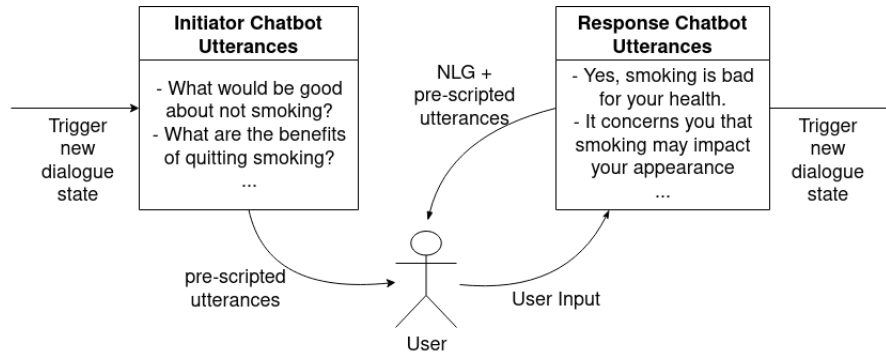


Figure 2: A visualization of the chatbot utterance handling within a dialogue state.

[1]. While these experiments did not involve natural language generation and primarily utilized the system’s pre-scripted features, they supplied essential conversational data for evaluating the potential of open-domain, large language models. By leveraging the collected human-bot conversational dataset, we have simulated scenarios in which the GPT-3.5-turbo and GPT-4 models generated alternative utterances to replace the retrieved human-authored responses from the original chatbot configuration. The GPT models were provided with the most recent five exchanges between the user and chatbot as conversation history.

Although we have not conducted a comprehensive evaluation of the generated content yet, it has been observed that GPT models can produce coherent utterances addressing users’ specific issues. Table 1 shows examples where users mention specific negative consequences of smoking. By design, pre-scripted utterances are composed in a generic tone, with a limited number of utterances referring to specific issues. In the first example, there was no pre-scripted utterance explicitly addressing cancer; therefore, the dialogue manager retrieved a generic utterance pertaining to the user’s health concerns. Conversely, both GPT-4 and GPT-3.5 produced utterances explicitly discussing the connection between smoking and cancer. The second example presents a scenario in which the user mentioned three negative consequences. The chatbot dialogue flow was initially designed to accommodate a single topic at a time, resulting in the retrieved pre-scripted utterance addressing only one of the topics. In this example, GPT-3.5 acknowledged the user’s multiple inputs, but still proposed to concentrate on a single topic. GPT-4, on the other hand, generated an utterance addressing all three topics within a single utterance.

## 6 RELEVANCE & ADAPTABILITY

Although HyLECA has initially been developed for highly sensitive domains, and tested in two specific topics (smoking cessation and sexual health promotion), chatbots for any other purposes can also be seamlessly integrated. Chatbot designers can create their own dialogue flows and without having to make any changes to HyLECA’s core dialogue management code. Moreover, HyLECA features a modular design, allowing more tech-savvy developers to add or replace components (such as custom actions, NLU functions, and NLG models) specific to their chatbots without disrupting the overall system. The system primarily focuses on developing and

maintaining text-based conversational agents, but it can also be integrated into an embodied virtual agent as a dialogue engine.

HyLECA’s modular design makes it relevant for any chatbot designer seeking to integrate NLG capabilities into their pre-scripted dialogue flows. Although the suitability of end-to-end NLG approaches for specific domains remains uncertain due to their unpredictable nature, HyLECA can serve as a platform to conduct real-time tests or simulations of these models in specific situations.

The HyLECA software presented in this paper does not include a visual platform. As the software’s authors, we have currently paired the system with a self-maintained version of the open-source Rocket.Chat<sup>3</sup> platform to make our chatbots publicly accessible<sup>4</sup>. Furthermore, as part of our future work, we plan to integrate the Tilbot software [5] as a visual platform to design the pre-scripted dialogue flows which then can be used in the HyLECA system.

We aim to demonstrate HyLECA by engaging in (and encouraging the audience to participate in) conversations with the GPT-3.5 integrated versions of our chatbots, specifically designed for smoking cessation and sexual health promotion.

## 7 CONCLUSIONS

We introduced our open-source chatbot development framework, HyLECA, and described its features. The framework aims to help chatbot designers to create controlled long-term conversational agents with integrated NLG capabilities for enhanced engagement. Its hybrid architecture combines the controllability of pre-scripted dialogue flows and retrieval-based approaches with the flexibility of natural language generation models. We showed the potential utility of our system for constructing specific types of conversational agents, and its relevance for chatbot designers interested in incorporating NLG capabilities into their pre-scripted dialogue flows.

We can observe a plethora of related studies on systematic evaluations of LLMs and GPT-3.5 in particular for a variety of tasks. We have also presented preliminary results on utilizing current state-of-the-art large language models, GPT-3.5 and GPT-4. These results and examples only showcase the potential of LLMs but do not yet appraise the suitability and effectiveness of these LLMs for practical implementations. As a next step, we plan to conduct extensive

<sup>3</sup><https://www.rocket.chat/>

<sup>4</sup><https://lwt.cls.ru.nl/chat>



Agent	Utterance
BOT	What do you see as a not-so-good thing if you continue smoking as you are?
USER	cancer
RETRIEVED	Sounds like you would like to be healthier overall, and stopping smoking might be a big step towards that.
GPT-3.5	Yes, cancer is definitely one of the significant risks in smoking, and it can be life-threatening if not caught early. Almost 30% of cancer deaths in the US are linked to smoking.
GPT-4	Cancer is a significant concern for many people, and it's understandable that you would be worried about the risk of developing it due to smoking. It's essential to be aware of the potential consequences and make informed decisions about your health and well-being.
BOT	What do you see as a not-so-good thing if you continue smoking as you are?
USER	bad lungs, a lot of money spent, the smell
RETRIEVED	So you want to take care of your finance, and stopping smoking might be an important step you can take.
GPT-3.5	Those are great points. Let's focus on the money part. Have you ever calculated how much you spend on smoking in a week or in a month?
GPT-4	Those are valid concerns. Bad lungs can lead to breathing difficulties and various illnesses, spending money on cigarettes can affect your financial stability, and the lingering smell can be unpleasant for others around you, as well as affecting your own sense of freshness.

**Table 1: Example conversations from smoking cessation chatbot with the retrieved and generated utterances.**

human evaluations of the generated content within chatbot-user conversations to assess its contextual appropriateness, specificity, naturalness, and engagement. Following these evaluations, we plan to implement a response ranker that can assess the quality of the utterances based on these four criteria to add control over the NLG models.

## ACKNOWLEDGMENTS

This project is partly financed by the Dutch Research Council (NWO) with project number 406.DI.19.054.

## REFERENCES

- [1] Divyaa Balaji, Margot van der Goot, Gert-Jan de Bruijn, Tibor Bosse, and Reinout Wiers. 2023. Using counselling-inspired relational strategies to facilitate self-disclosure with a chatbot in a sensitive domain: A qualitative study. (2023). Manuscript submitted for publication.
- [2] Erkan Başar, Iris Hendrickx, Emiel Kraemer, Gert-Jan de Bruijn, and Tibor Bosse. 2022. Hints of Independence in a Pre-scripted World: On Controlled Usage of Open-domain Language Models for Chatbots in Highly Sensitive Domains. In *Proceedings of the 14th International Conference on Agents and Artificial Intelligence (ICAART)*. INSTICC, SciTePress. <https://doi.org/10.5220/0010914300003116>
- [3] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922>
- [4] Tom Bocklisch, Joey Faulkner, Nick Pawlowski, and Alan Nichol. 2017. Rasa: Open Source Language Understanding and Dialogue Management. In *Proceedings of the Conversational AI Workshop at the 31st Conference on Neural Information Processing Systems (NeurIPS 2017)*. <https://doi.org/10.48550/arXiv.1712.05181>
- [5] Jan de Wit and Anouck Braggaar. 2023. Tilbot: A Visual Design Platform to Facilitate Open Science Research into Conversational User Interfaces. In *Proceedings of the 5th Conference on Conversational User Interfaces (CUI '23)*. Association for Computing Machinery.
- [6] Linwei He, Divyaa Balaji, Reinout Wiers, Marjolijn L. Antheunis, and Emiel Kraemer. 2022. Effectiveness and Acceptability of Conversational Agents for Smoking Cessation: A Systematic Review and Meta-analysis. *Nicotine & Tobacco Research* (2022). <https://doi.org/10.1093/ntr/ntac281>
- [7] Linwei He, Erkan Başar, Emiel Kraemer, Reinout Wiers, and Marjolijn Antheunis. 2023. Effectiveness and user experience of a smoking cessation chatbot: A mixed-methods study comparing motivational interviewing and confrontational counselling. (2023). Manuscript submitted for publication.
- [8] Robin Mermelstein, Sheldon Cohen, Edward Lichtenstein, John S Baer, and Tom Kamarck. 1986. Social support and smoking cessation and maintenance. *Journal of consulting and clinical psychology* 54, 4 (1986).
- [9] OpenAI. 2023. GPT-4 Technical Report. arXiv:2303.08774
- [10] Ioannis Papaioannou, Amanda Cercas Curry, Jose L Part, Igor Shalyminov, Xinnuo Xu, Yanchao Yu, Ondrej Dušek, Verena Rieser, and Oliver Lemon. 2017. Alana: Social dialogue using an ensemble model and a ranker trained on user feedback. In *Alexa Prize SocialBot Grand Challenge 1 Proceedings*.
- [11] Amon Rapp, Lorenzo Curti, and Arianna Boldi. 2021. The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies* 151 (2021). <https://doi.org/10.1016/j.ijhcs.2021.102630>
- [12] Abigail See, Stephen Roller, Douwe Kiela, and Jason Weston. 2019. What makes a good conversation? How controllable attributes affect human judgments. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. ACL. <https://doi.org/10.18653/v1/N19-1170>
- [13] Igor Shalyminov, Alessandro Sordani, Adam Atkinson, and Hannes Schulz. 2021. GRTr: Generative-Retrieval Transformers for Data-Efficient Dialogue Domain Adaptation. *IEEE/ACM Trans. Audio, Speech and Lang. Proc.* (2021). <https://doi.org/10.1109/TASLP.2021.3074779>
- [14] Shayan Zamanirad, Boualem Benatallah, Carlos Rodriguez, Mohammadali Yaghoobzadehfard, Sara Bouguelia, and Hayet Brabra. 2020. State machine based human-bot conversation model and services. In *International Conference on Advanced Information Systems Engineering*. Springer. [https://doi.org/10.1007/978-3-030-49435-3\\_13](https://doi.org/10.1007/978-3-030-49435-3_13)
- [15] Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and Bill Dolan. 2020. DIALOGPT: Large-Scale Generative Pre-training for Conversational Response Generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. ACL. <https://doi.org/10.18653/v1/2020.acl-demos.30>