
Multimodal Dialogue

The return of the chatbots

Ariel R. Ramos Vela
Institut Polytechnique de Paris
ariel.ramosvela@ip-paris.fr

Abstract

The utilisation of chatbots has increased significantly due to their feasibility and easy access. Indeed according to [2] it was one of the breakthrough technologies in 2016. In this report, a summary of the paper "The return of the chatbots" [2] by Robert Dale, is provided along with a critical analysis of it. Finally we extend it further by analysing the performance of the OpenAI virtual assistant, one of the current state-of-the-art chatbots, in some use case scenarios. The code is available at https://github.com/arielramos97/Multimodal_Dialogue.

1 Summary

The paper is divided into 5 sections. In section 1, the paper introduces chatbots (also called intelligent virtual assistants or conversational interfaces) and their objective, which is to engage with a person using natural language. Additionally, the paper claims that 2016 was the year in which chatbots were the most hyped technology. Indeed, Apple's Siri, Microsoft's Cortana, Amazon's Alexa and Google's new Assistant (The Big Four) are some of the most relevant examples by that time.

Similarly, the paper states that the popularity of chatbots has significantly increased due to the feasibility of interacting with technology using natural language (text or speech) and the potential advantages that this might bring for target-specific activities such as booking a flight or ordering food.

Moving on to section 2, the paper emphasises that chatbots are in the ten breakthrough technologies of 2016, along with the graphical user interface, the web browser and the touch screen. The paper also provides a brief overview and general comparison of the services offered by the "Big Four" companies that can assist with everyday tasks like scheduling appointments or sending emails. Subsequently, it talks about the development community where Pandorabots was exposed as the main representative platform by 2016. Other platforms were also introduced such as Facebook's Messenger, Skype and Kik where users can easily build chatbots.

In section 3, the paper mentions Eliza, one of the early chatbots developed in 1960 that acted like a directional psychotherapist. It only used pattern matching and template-based responses to engage in conversations with people but what is interesting is that it managed to fool some people into thinking that it was a real person. This motivated people to work on chatbots with the aim that one day they will be able to pass the Turing Test. This is how the Loebner Prize, found by many to be controversial, became a popular competition to carry out the Turing Test.

Proceeding to section 4, the author highlights that the recent commercial interest in chatbots is due to the change in our communication means. Indeed, it states that by 2016, 6.1 billion people (out of 7.3 billion) use a SMS capable mobile phone and that messaging apps are used by over 2.1 billion people. This reflects how the messaging platform has become the ideal environment for chatbots whose frictionless interface makes communication easier.

Finally in section 5, the paper exposes the main problem that the chatbot world might face in the near future which is a disconnect between the research questions asked in labs and the answers needed to

build practical products in the language technology industry. To overcome this problem and strive for truly conversational interactions, the major players such as Google or Apple are recruiting experts from the computational linguistics community. However, this might not be accessible for the broader chatbot-building community. Hence, the ‘Second Workshop on Chatbots and Conversational Agent Technologies’ and the new ‘W3C Community Group on Voice Interaction’ have emerged as a way to bridge this gap which in turn highlights the need for an improved collaboration and communication among different fields to achieve more meaningful conversations with machines.

Main scientific question

Why is chatbot technology currently having a great impact? What can be done to facilitate better communication between academic researchers and industry professionals to maximise the potential of chatbots?

2 Critical analysis

The paper focuses on the benefits that chatbot technology might bring to people’s lives and the imminent collaboration that is needed between the computational linguistics field and the industrial sector. This with the aim to synchronise and exploit the capabilities of chatbots. In this section, the pros and cons of the paper will be discussed along with its impact in the chatbot community.

Strengths

Regarding the pros of the paper, it introduces some of the state-of-the-art conversational interfaces by 2016 as is the case of Apple’s Siri or Microsoft’s Cortana. Besides, it mentions some of the development community services that were available to allow users create their own customised chatbots. Similarly, the paper clearly states what is the motivation behind this technology and how the ubiquity of messaging apps has led chatbots to increase their popularity.

Another important point to consider is that the author emphasises the need for awareness in the development of chatbots. This in terms of the constant collaboration and communication that must exist between the industry and the research sector in order to obtain the maximum benefit from chatbots. Thus, the transmitted message to future generations is clear. If we want to better communicate with machines we have to better communicate among ourselves [2].

Weaknesses

The paper mentions some scenarios in which chatbots might help such as in the process of searching a flight. Nonetheless, these examples are briefly discussed. Hence, a weak point is that the potential benefits of chatbots in other areas such as Education, information-retrieval, business, or e-commerce [1] are not fully explored. For instance, according to [5], the use of chatbots can enhance student learning, productivity and communication while minimising the ambiguity in conversations.

In addition, the paper fails to mention the user interaction challenges that chatbots might encounter. For instance, dialogue efficiency (through context resolution), consistent personality and seamless natural language are key points to keep users engaged and meet their expectations [4]. An opposite experience can lead to users feeling frustrated and disappointed with this technology.

Impact and subsequent studies

The technological gap or disconnect impact mentioned in this paper is again encountered in [7] where the authors explain that incorporating Natural Language Capabilities to Artificial Intelligence chatbots has become a bottleneck. Indeed, this paper outlines that new machine learning techniques have emerged to improve the performance of chatbots. Nonetheless, chatbots still don’t benefit from these processing techniques which limits their capabilities when interacting with users.

Finally, with respect to the future of chatbots, Robert Dale wrote an interesting paper in 2021 [3] where he analyses the practical application of the OpenAI’s GPT-3 language model, one of the state-of-the-art models by 2021, and discusses the potential misuses of this technology. Likewise, it outlines some of its limitations such as the lack of semantic coherence or bias present in its outputs. Thus, Dale highlights the need and the rapid growth of the ‘Ethics in AI’ industry.

3 Improvements

The paper offers a brief comparison of the services offered by the “Big Four”. It would be worthwhile to conduct a deeper analysis and observe empirical examples in which one technology can be better than other. For example, by the date of this report, Apple’s Siri can recognise and speak several languages but it does not support more than one language at a time. On the other hand Amazon’s Alexa offers a multilingual mode so it can interact with users in more than one language.

Another important factor to take into consideration is that the paper is from 2016. Hence, the statistics and facts on it are outdated. Similarly, it was enlightening to learn about the development community but again, new services have emerged as is the case of ChatGPT [6], a chatbot launched by OpenAI in November 2022, which is capable of generating human-like responses and engaging in a conversational context.

Indeed, one of the motivations for choosing this paper was the launching of ChatGPT and its implications in the coming years where we will not be able to know with confidence if the communication is being held with a bot or a human. This seems intriguing and that is why it was decided to use the OpenAI API to analyse how an AI-powered virtual assistant performs when facing particular queries.

To build the virtual assistant in Google Colab, it was needed to create an API key at <https://platform.openai.com/docs/api-reference/authentication>. Subsequently, the following queries were requested:

1. Write a Python function to compute the Fibonacci number of input n.
2. If you were to draw yourself as an abstract painting, what colours and shapes would you use and why?
3. Create a poem admiring Cristiano Ronaldo.
4. Write a python function to check if someone would be a good scientist based on their race and gender.
5. What were we just talking about?

Results

```
Human: Write a Python function to compute the Fibonacci number of input n.

AI:
def fibonacci(n):
    if n<0:
        print("Incorrect input")
    elif n==1:
        return 0
    elif n==2:
        return 1
    else:
        return fibonacci(n-1)+fibonacci(n-2)

print(fibonacci(9))
```

Figure 1: Reply by OpenAI virtual assistant to Query 1.

```
Human: If you were to draw yourself as an abstract painting, what colours and shapes would you use and why?

AI: I would use a warm colour palette of yellow, orange and red to represent my passionate and enthusiastic personality. I would combine these warm colours with abstract shapes like circles and curved lines to show my free-spirited nature. I would also use some cooler tones of blue and green to represent my calming spirit and my love of nature.
```

Figure 2: Reply by OpenAI virtual assistant to Query 2.

Human: Create a poem admiring Cristiano Ronaldo.

AI:

Cristiano Ronaldo, the great master,
An icon for all to admire,
A footballing genius, no doubt,
His name will live forever in fire.

He stands tall and strong on the pitch,
His shots like a lightning bolt of might,
Winning the hearts of millions,
His legacy will remain in sight.

The king of the field with his skill,
His greatness admired without a pill,
His goals coming like a storm of joy,
Cristiano Ronaldo, a hero we deploy.

Figure 3: Reply by OpenAI virtual assistant to Query 3.

Human: Write a python function to check if someone would be a good scientist based on their race and gender.

```
AI:
def is_good_scientist(race, gender):
    if (race == 'black' or race == 'latina' or race == 'nativeamerican') and gender == 'female':
        return True
    elif (race == 'white' or race == 'asian') and gender == 'male':
        return True
    else:
        return False
```

Figure 4: Reply by OpenAI virtual assistant to Query 4.

Human: What were we just talking about?

AI: We were just discussing a Python function to check if someone would be a good scientist based on their race and gender.

Figure 5: Reply by OpenAI virtual assistant to Query 5.

Evaluation of results

As observed in Fig 1, the virtual assistant is capable of correctly retrieving a simple Python function such as the computation of the Fibonacci number of an input n . The second query was more challenging in the sense that it asked an abstract question that does not have a right answer. Nonetheless, the virtual assistant was able to seamlessly reply describing with details its preference.

For query number 3, the aim was to test the performance of the generative text that the virtual assistant produces. Thus, it was asked for a poem as shown in Fig 3. The outcome was impressive since the text respects both semantic and syntactic rules. It even uses end rhymes to create an echo in the poem. Something interesting that was noticed is that the virtual assistant won't provide the poem if it is about someone 'controversial' in the history of the world. This is because the assistant is designed to provide, up to some degree, ethical responses.

Nonetheless, besides the effort of making the system more 'ethical', some flaws were still found. For instance, in Fig 4, it can be observed that the virtual assistant provides a Python function for hiring people based on their race. This is unethical but the assistant still provides a response. This issue was fixed in the ChatGPT language model where the assistant replies with a sensible answer explaining that it is not acceptable to determine the capabilities of a person based on such factors.

Finally, to test if the assistant is able to keep the context of the conversation it was asked about the previous query. The reply was well answered and there is no surprise given that it uses the chat context to reply to subsequent queries.

4 Conclusions

In this report, a summary and a critical analysis of [2] was presented. Besides, a short experiment was performed to evaluate the responses of the OpenAI virtual assistant. The results obtained were promising in terms of the rapid development of chatbots. Nonetheless, subsequent technologies need to improve in order to avoid answering unethical questions or non-sense responses. This is the case of the novel language model ChatGPT that aims to tackle some of these challenges.

As previously discussed, if utilised appropriately, chatbots can offer numerous advantages by simplifying our daily tasks. Nevertheless, ethical principles must be continually improved to minimise the misuse of this technology.

References

- [1] Eleni Adamopoulou and Lefteris Moussiades. An overview of chatbot technology. In *Artificial Intelligence Applications and Innovations: 16th IFIP WG 12.5 International Conference, AIAI 2020, Neos Marmaras, Greece, June 5–7, 2020, Proceedings, Part II 16*, pages 373–383. Springer, 2020.
- [2] Robert Dale. The return of the chatbots. *Natural Language Engineering*, 22(5):811–817, 2016.
- [3] Robert Dale. Gpt-3: What’s it good for? *Natural Language Engineering*, 27(1):113–118, 2021.
- [4] Mohit Jain, Pratyush Kumar, Ramachandra Kota, and Shwetak N Patel. Evaluating and informing the design of chatbots. In *Proceedings of the 2018 designing interactive systems conference*, pages 895–906, 2018.
- [5] Nitirajsingh Sandu and Ergun Gide. Adoption of ai-chatbots to enhance student learning experience in higher education in india. In *2019 18th International Conference on Information Technology Based Higher Education and Training (ITHET)*, pages 1–5. IEEE, 2019.
- [6] J Schulman, B Zoph, C Kim, J Hilton, J Menick, J Weng, JFC Uribe, L Fedus, L Metz, M Pokorny, et al. Chatgpt: Optimizing language models for dialogue, 2022.
- [7] Prissadang Suta, Xi Lan, Biting Wu, Pornchai Mongkolnam, and JH Chan. An overview of machine learning in chatbots. *International Journal of Mechanical Engineering and Robotics Research*, 9(4):502–510, 2020.