GOLDEN-BOY: AN INTELLIGENT CHATBOT TO INFORM AND MITIGATE RACIAL BIAS IN FACIAL RECOGNITION TECHNOLOGY

Abstract

Facial recognition has become a widely adopted technology that has integrated itself in various sectors around the world from law enforcement, phone security, and even grocery shopping. According to Harvard University's *Science In The News*, it is estimated that almost half of American adults – over 117 million people, as of 2016 – have photos within a facial recognition network used by law enforcement. This participation occurs without consent, or even awareness, and is bolstered by a lack of legislative oversight.The increasing rise of this technology has lead to an equal rise of anxiety towards the widespread adoption of facial recognition, due to its historical racial bias affecting marginalized groups who are already most at risk for racial profiling. This paper aims to provide a solution for the lack of diversity behind the training of facial recognition AI by creating a chatbot architecture that can be integrated into software such as Slack for the purpose of informing the public about the current state of this technology and its wide adoption. The benefits of this chatbot would be the spread of information regarding the current weakness of artificial intelligence and how it can be corrected to lessen the risk for marginalized groups.

Introduction

Face recognition has become a powerful technology with significant implications in both criminal justice and everyday life. These results have prompted immediate responses, shaping an ongoing discourse around equity in face recognition. Facial recognition and artificial intelligence technology has exploded in recent years, giving way for the flaws of this technology to come to the surface. It should be noted that this chatbot is merely for the purpose of informing those who are exposed to artificial intelligence, and further research into the training diversity of these machine learning programs must be followed. The increasing rise of this technology has lead to an equal rise of anxiety towards the widespread adoption of facial recognition, due to its historical racial bias affecting marginalized groups who are already most at risk for racial profiling. This paper aims to provide a solution for the lack of diversity behind the training of facial recognition AI by creating a chatbot architecture that can be integrated into software such as Slack for the purpose of informing the public about the current state of this technology and its wide adoption. The benefits of this chatbot would be the spread of information regarding the current weakness of artificial intelligence and how it can be corrected to lessen the risk for marginalized groups.

2 Golden-Boy Design

Golden-Boy was designed and developed as a conversational agent. It is a social service powered by rule and artificial intelligence that helps users interact via a chat interface (Sanjeevi, 2018) to provide users with more about the AI facial recognition bias and mitigation exploration.

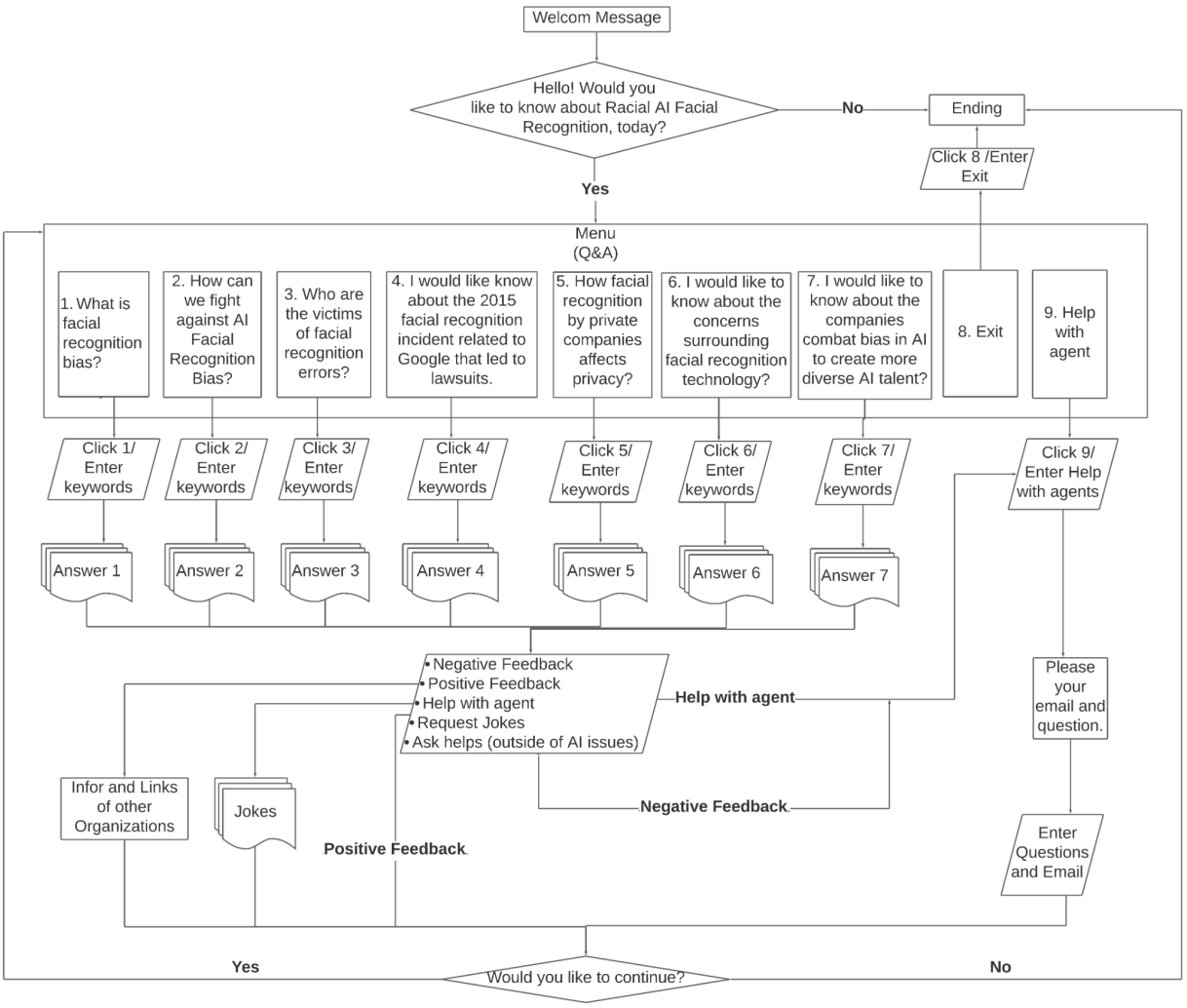
We implemented Golden-Boy on the IBM Waston Assistant Lite plan and integrated chatbots integrated into chat apps (e.g., Slack). Users can communicate with Golden-Boy on web browsers and mobile apps. Based on the conversation flow of Golden-Boy, as illustrated in Figure 1, IBM Watson Assistant provides GB capabilities to recognize users' intents and entities, and then, they will interact with users by responding and requesting. Using machine learning, chatbots are able to conduct text based on conversations with users to answer basic questions, provide information, and offer interactive support in various domains. (Sanjeevi, 2018). We also trained Golden-Boy with an open domain that helps users communicate with Chatbots anywhere, and are able to type or ask anything (Sanjeevi, 2018). In particular, we build the RQAs as FQAs with seven main questions and answers as social-exploration and mitigation exploration:

- RQ1: What is facial recognition bias?
- RQ2: How can we fight against AI Facial Recognition Bias?
- RQ3: Who are the victims of facial recognition errors?
- RQ4: I would like to know about the 2015 facial recognition incident related to Google that led to lawsuits.
- RQ5: How facial recognition by private companies affect privacy?
- RQ6: I would like to know about the concerns surrounding facial recognition

technology?

- RQ7: I would like to know about the companies combates bias in AI to create more diverse AI talent.

Also, IBM Watson Assistant assisted GoldenBoy in training chatbot applications more human (Adamopoulou and Moussiades, 2020) such as creating jokes and providing information of 10 organizations that support racial equality.



2.1 Scenario

Artificial Intelligence facial recognition brings many benefits to many industries nowadays, such as healthcare, security, airport boarding, and proctoring (RecFace,. 2021). However, AI facial recognition has negative social interaction from users because it contains the bias the AI to exhibit racism, sexism, or other types of discrimination. In June 2020, IBM announced to cancel the facial recognition programs to advance racial equity in law enforcement as bias (Ziady, 2020).

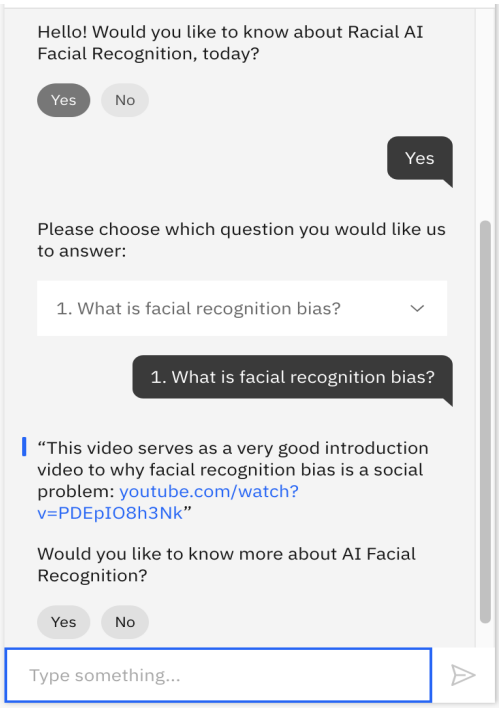
With Golden-Boy, users are prompted with two different features such as social exploration and mitigation exploration. Golden-Boy provides educational research to help users explore accurate information about the issue of AI facial recognition as social issues. We implemented IBM Watson Assistant, which provides users with fast, consistent, and accurate answers across any application and device. Also, our AI assistant learns from user conversations and improves its ability to explain racial AI recognition. It helps reduce the frustration of long-time wait times, tedious searches, and unhelpful chatbots.

2.2 Features

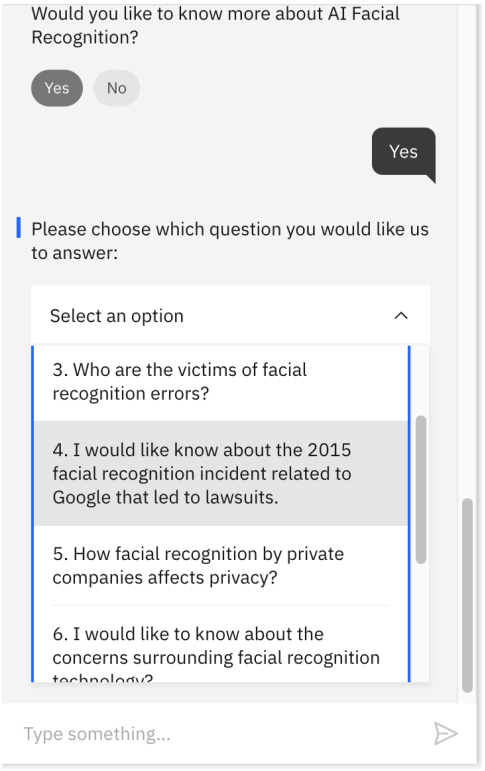
In this section, we present the main features of GOLDENBOY mentioned above. The chatbot's main characteristic is informative. It shows all the possible menus to the users, and waits for the user's response to decide what actions to take next. The chatbot menus can be categorized into 3 sections. Users can find these options when they launch the chatbot after the greeting conversation.

2.2.1 Background knowledge

Users can access the overall knowledge and background information about the topic in this section. If the user indicates that he or she wants to learn about Artificial Intelligence Facial Recognition Biases, the first menu of the chatbot provides a Youtube video link to the user. The video explains; 1) how facial recognition technology is actually used in many different places, 2) what are the benefits of using the technology, 3) what are the concerns and problems about facial recognition.
This video functions as a basic introduction to the topic that chatbot is going to provide to users. It raises the issue and problem to the surface that can be brought to society, when there are inappropriate facial recognition biases. This part boosts users' interest in the topic.



2.2.2 Actual Cases



After showcasing the introduction, chatbot provides more detailed information and actual examples to users. Through these questions and answers, users are able to get more detailed information about the issue. For example, users can be informed of various accidents and cases that have actually happened previously. It helps users to put themselves in the same situation as victims and feel the same way to them. This is the main section of the chatbot to achieve the main purpose of this project to raise awareness of society and find the solution for the issue.

For example, It provides a video of Joy Buolamwini, a MIT graduate student, whose face was not detected from various occasions by AI facial recognition devices, even from her own project, because all of those facial recognition uses same training sets on the internet, which is not enough to totally indicate the entire various ethnicities and face shape. It becomes a huge problem when police start to use facial recognition software to their “crime fighting arsenal”. It can cause huge misunderstanding to them and bring unnecessary conflict and danger to both police and civilians.

Another example is that Google Images recognized and categorized some pictures of black people as ‘gorillas’. It can be really stressful and anxious to face this discrimination by AI technology which is widely used in the world. It is a big social issue, and problematic to use this technology until they polish the software much more precisely in order to prevent any kind of uncomfortable and upset moment to users. It can also bring huge damage to the enterprises who use it as well. One mistake from the AI technology can be spread to the entire world through the internet within a day, and it can cause tremendous monetary penalties and damage to their reputation.

2.2.3 Solutions for the problem

In previous sections, chatbot gathered users' attention and interest by providing introductory information, and raised the severity and truly put users in the same shoes as victims by providing actual cases and what problems can arise from it. In this section chatbot provides some examples of how this AI technology can be fixed. It might be easily thought by people that there is nothing that normal people actually do or it is AI scientists or software developers’ job to fix the problem. However, changes do not come automatically if no one knows the seriousness of the issue or recognizes it as a problem. The more people who are educated will bring the more voices about correctness to the surface, and it will make a pressure for the change to those producers who make the AI technology and organizations who use the AI technology.

3 Methodology

We conducted five interview studies with five different candidates based on a semi-structured interview template that our team created. The interviews' purpose was to get deeper insights from potential users of our chatbot. The interviews were done to collect information from users based on three research questions: What did they like, what did they not like, and what did they want to see in the future of our chatbot.

3.1 Demographics

The study was done on five different individuals with their age ranging from 18 to 28. Two of the five interviewees are university students, one is a childcare worker, one is a marketing & sales manager; and last but not least, an account representative. Two of our candidates are male, and the other three are females.

3.2 Questions

Our thematic analysis of the responses that we acquired from our candidates were conducted around the questions below:

Q1: What are the strengths of our GoldenBoy chatbot?

Our interview participants had similar input to our chatbot that its strength relies on being informative, convenient to use, and credibility.

Informative: The purpose of our chatbot was to mainly spread awareness to both beginners and intermediate level learners of facial recognition. Whether you have never heard of facial recognition biases, or have heard little about it, our chatbot can still teach and inform users about the danger of facial recognition biases to our society, and its target victims; while introducing users to what facial recognition bias is.

Convenient: 100% of our participants stated that the GoldenBoy chatbot is very convenient to use, and easy to navigate. Whether a person is tech-savvy, or a beginner in technology usage, you can still put our chatbot to use with ease. The chatbot is also accessible by anyone.

Credibility: Besides its convenient quality, three of our participants noted that the

responses that our chatbot gave are credible due to its functional advantage in providing resources to every answer. Moreover, the references were sourced from sites and videos that have back-up credentials including research companies and studies from M.I.T.

Q2. What are the weaknesses of the GoldenBoy chatbot?

According to our research interviews, most participants mentioned that our chatbot lacks interactive functions and features. Interactive functionalities can keep the chatbot feels fresh, and fun to use. To add, participants also noted that our sources included too many articles that require intense and long reading time.

Q3. What can we do to improve the GoldenBoy chatbot?

In accordance with our team analysis of the responses from the interviews, we found that the improvement of GoldenBoy will rely heavily on adding interactive elements such as pictures and videos to our chatbot to keep the chatbot feeling fresh with aesthetics and be less boring. Moreover, summarizing the answers to the questions in the chatbot will make the chatbot look more appealing with concise and straightforward answers.

4 Conclusion

In conclusion, the use of facial recognition is seen throughout various sectors in the economy and its impact is bigger than ever. Despite the good intentions of this technology, it has led people to question if the good outweighs the flaws and social problems that come with it. More specifically, racial bias affecting marginalized groups and further contributing to the social issue of racial profiling. Goldenboy looks to provide a solution for the lack of diversity behind the training of facial recognition. It is a chatbot program powered by IBM Watson that can be integrated into software such as Slack and was developed to act as a conversational agent. The goal of Goldenboy is to share information regarding the current weakness of artificial intelligence and how it can be corrected to lessen the risk for marginalized groups. During our time working on Goldenboy, we conducted five interview studies with five different candidates. Our findings showed that this chatbot is informative, convenient to use, and has good credibility. It also showed Goldenboy lacked interactive functions and features. In the future, our team looks forward to working on Goldenboy and implementing interactive elements such as pictures, videos, and other elements to enhance our chatbot even more.

References

Adamopoulou, E,. and Moussiades, L,. 2020. “An Overview of Chatbot Technology.” Department of Computer Science, International Hellenic University, Agios Loukas, 65404 Kavala, Greece. (doi: 10.1007/978-3-030-49186-4_31)

Anna, Roger Peterson, Pavitra Goswami, Dumbfounded, Don Quijote, Roshni Kargi, & Ramona Herner. (2020, October 26). *Racial Discrimination in Face Recognition Technology*. Science in the News. https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/.

RecFaces. 2021. “What is AI Facial Recognition Tech and How does It Work?” RecFaces, RecFaces (https://recfaces.com/articles/ai-facial-recognition, accessed May 14, 2021).

Sanjeevi, M. 2018. “Chapter 11: ChatBots to Question & Answer systems.” Medium. Medium. (https://medium.com/deep-math-machine-learning-ai/chapter-11-chatbots-to-question-answer-systems-e06c648ac22a, accessed May 14, 2021).

Ziady, H. 2020. “IBM is canceling its facial recognition program,” CNN Business, CNN Business, June 9. (https://www.cnn.com/2020/06/09/tech/ibm-facial-recognition-blm/index.html, accessed May 14, 2021).