CockTail AI

Hamza Aziz Khan Vinay Kumar Mehra Sahil Sandal Shivam Srivastava
Sarthak Vijayvergiya

December 2023

Abstract

The project involves the development of a system where the Furhat virtual robot, acting as a bartender, automatically analyzes human behavior and affective states to generate socially adaptive responses during interactions with one or multiple users. The project comprises two interconnected subsystems: a User Perception sub-system and an Interaction sub-system. The User Perception sub-system uses real-time webcam input to automatically detect users' affective states by extracting facial features and applying Machine Learning techniques. The choice of facial features, ML techniques, and the decision to analyze individual users or a group is left to the project team. The Interaction sub-system integrates the affective states and user speech to generate context-aware, socially adaptive behaviors by Furhat.

1 Introduction

Welcome to the future of hospitality! Meet our Al-Powered Bartender Bot, a revolutionary system blending facial recognition technology with interactive bartending.
The project aims to develop an interactive system using Furhat as a virtual bartender in scenarios where it interacts users, engaging in simple conversations.
It's designed to enhance customer experiences by understanding and adapting to their emotions, offering personalized drink suggestions and engaging conversations. This innovative approach promises a new era of personalized hospitality.

2 Methodology

Bartender AI: The Bartender AI subsystem is designed to suggest personalized cocktail recommendations based on user preferences. It utilizes facial emotion detection to understand the user's current emotional state and tailors cocktail suggestions accordingly.
Furhat Interaction: The Furhat Interaction subsystem focuses on the interaction between the Furhat robot and the user. It receives emotion information from the Bartender AI and generates socially adaptive behaviors for Furhat based on the detected emotions and the ongoing interaction context.

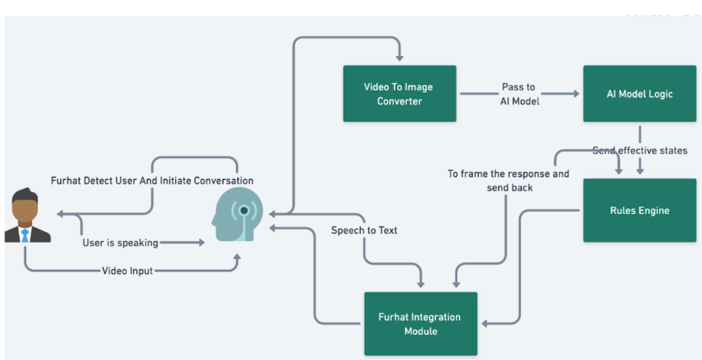

Figure 1: Technical Flow

2.1 Overall System Design

- **Scenario Choice:** We chose the bartender scenario, where Furhat[2] interacts with human users, aiming for realistic and engaging conversations.
- **High-Level Goals**: Develop an AI-powered system capable of perceiving user affective states and responding adaptively[1].
- **Specific Choices:** We employed a diverse set of models, including Decision Tree Classifier, K-Nearest Neighbors, and a tuned Support Vector Machine (SVM). Each model brings its unique strengths to the table. After careful evaluation, we chose SVM model as it gave the best results.
  Decision Tree Classifier Accuracy: Achieved accuracy of 48% on the validation set.
  K-Nearest Neighbors Accuracy: Demonstrated accuracy of 59% on the validation set.
  Tuned SVM Accuracy: The tuned Support Vector Machine showcased remarkable accuracy, reaching 62% on the validation set.
  Also, we worked with 3 emotions for furhat to respond to, i.e. happy, sad and neutral. We had achieved a total of 25 unique responses from furhat and a total of 6 gestures were used by furhat.

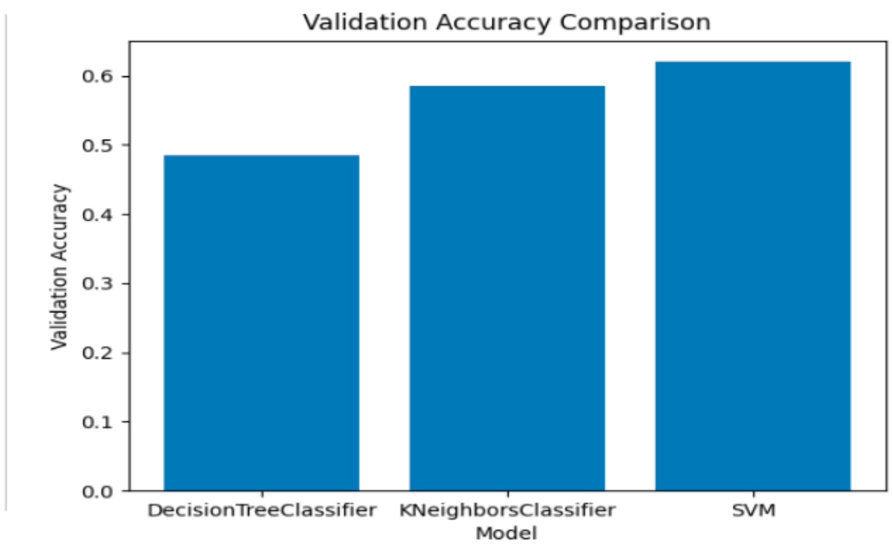

Figure 2: Validation Accuracy Comparison

2.2 User Perception sub-system:

Design: The User Perception sub-system is designed to automatically analyze human behavior and affective states by processing real-time webcam input of users interacting with Furhat. The primary goal is to extract facial features, specifically Action Units, and employ Machine Learning techniques to translate these features into a high-level representation of users' behavior, focusing on affective states such as valence or arousal.
Implementation: To achieve the intended design, we employed Py-Feat for the automatic extraction of facial features, focusing on Action Units. These features were then fed into a Machine Learning pipeline that included feature selection, model training, and validation.
Results: The User Perception sub-system demonstrated effective extraction of facial features and robust performance in detecting affective states. The choice of dataset for ML model training ensured that the system's performance aligned with the specific requirements of individual user's affective state detection.

2.3 Interactive sub-system:

Design: The Interaction Sub-system serves as the responsive component of our AI-Powered Bartender, leveraging the affective states detected by the User Perception Sub-system and the user's speech input.

The primary goal is to generate socially adaptive behaviors in Furhat. This includes dynamically adjusting Furhat's facial expressions, tone of voice, and conversational style based on the user's emotions and the context of the interaction.
The design incorporates a rule-based system that links specific affective states to predefined Furhat behaviors. For instance, if a user expresses happiness, Furhat responds with an upbeat and cheerful demeanor. Similarly, if the user appears sad, Furhat adapts by displaying empathy and providing comforting responses. This high-level design ensures a nuanced and contextually appropriate interaction between Furhat and the user.
Implementation: To implement the Interaction Sub-system, we created a set of rules mapping affective states to Furhat's behaviors. Each rule is defined based on empirical observations of human social interactions, taking into account cultural norms and the specific context of a bartending scenario. Python scripting was employed to facilitate seamless integration with Furhat's existing functionalities. The rule-based system is dynamic, allowing for adjustments and expansions as needed. For example, the system can be extended to include more nuanced responses based on the combination of affective states or the history of the conversation.
Results: The Interaction Sub-system has demonstrated effective adaptability in response to user affective states. In our evaluations, Furhat's behaviors aligned well with the detected emotions, creating a more engaging and natural interaction.

3 General Discussion

3.1 Overall Pipeline

The overall pipeline of our AI-Powered Bartender project comprises the User Perception Sub-system and the Interaction Sub-system, working collaboratively to deliver a personalized and emotionally aware bartending experience.
The User Perception Sub-system effectively captures real-time facial features and utilizes machine learning techniques to discern affective states such as valence and arousal. This information is then seamlessly passed to the Interaction Sub-system, where Furhat dynamically adjusts its behaviors based on the user's emotional cues.
The pipeline successfully creates a cohesive and adaptive interaction, aligning with our project objectives of enhancing customer experiences through personalized, emotionally intelligent bartending.

3.2 Successes in System Development

1. **Emotion Detection Accuracy:** Our facial emotion detection model demonstrated commendable accuracy in real-time webcam data. By implementing advanced machine learning techniques and training on the DiffusionFER dataset, we achieved precise recognition of happy, sad facial expressions.

2. **Integration Success:** The integration of the Bartender AI and Furhat Interaction subsystems was executed successfully. Both subsystems synchronized effectively which allows a cohesive user experience.

3. **Adaptive Behaviors:** During testing, the system exhibited user-friendly and socially adaptive behaviors. Furhat's responses were tailored accordling to the detected emotions that creates a dynamic and engaging interaction.

3.3 Challenges Faced

1. **Subsystem Integration:** Streamlining data flow and communication between the Bartender AI and Furhat Interaction subsystems required addressing differences in inputs and ensuring real-time synchronization.

2. **Facial features Dynamics:** Rapidly changing user facial features cause model to detect different emotions within small time. Which also impacted the furhat's response.

3.4 Tools Used

- In the pursuit of developing a seamless interaction between the Bartender AI and Furhat subsystems, we encountered distinct challenges necessitating a refined approach. The first challenge involved Subsystem Integration, demanding the harmonization of data flow and communication channels. To overcome this hurdle, insights and references from ChatGPT played a pivotal role in addressing disparities in inputs and achieving real-time synchronization between the subsystems.
- The second challenge centered around the dynamic nature of Facial Features, wherein rapidly changing user expressions posed difficulties for the model in accurately detecting emotions within brief intervals. This dynamism subsequently influenced Furhat's response, prompting the need for an adaptive solution. By consulting ChatGPT, we garnered valuable perspectives to mitigate the impact of swiftly changing facial features on the model's emotion recognition capabilities and, consequently, enhance Furhat's responsive behavior.

3.5 Ethical Issues

1. **Facial Recognition Data:** The User Perception Sub-system involves capturing facial features for emotion recognition. Ensuring user consent, anonymizing data, and protecting individuals' privacy rights are essential to address potential privacy concerns.
2. **User Awareness:** Users interacting with Furhat should be fully informed about the capabilities of the system, including facial recognition and emotion detection. Obtaining explicit consent for data collection and system interaction is crucial.
3. **Secure Handling of Data:** Safeguarding the collected data, including facial images and emotional states, against unauthorized access, breaches, or misuse is imperative. Employing robust encryption and secure storage practices is essential.
4. **Cross-Cultural Understanding:** Emotions and expressions vary across cultures. Ensuring that the system recognizes and respects diverse cultural norms and expressions is essential to prevent misunderstandings or misinterpretations.

4 Conclusion

The AI-Powered Bartender Bot project marks a groundbreaking stride in personalized hospitality, seamlessly integrating facial recognition and interactive bartending. Through the adeptly designed User Perception and Interaction subsystems, the system excels in real-time emotion detection and socially adaptive responses. Leveraging the DiffusionFER dataset for ML-driven emotion analysis, the User Perception sub-system lays a robust foundation. The Interaction sub-system, driven by a rule-based approach, effectively translates affective states into engaging Furhat responses. Looking ahead, the project envisions refinements, increased transparency, and considerations for user privacy, fostering a commitment to responsible AI innovation.
The AI-Powered Bartender Bot signifies a paradigm shift in personalized hospitality, showcasing technology's seamless adaptation to human emotions for memorable interactive bartending experiences. The successful implementation underscores our commitment to advancing responsible and innovative AI use. The AI-Powered Bartender Bot embodies a transformative leap in personalized customer experiences, exemplifying the harmonious convergence of technology and human emotions.

References

[1] Personalized human-robot interaction with a robot bartender, October 2022. URL: file:///C:/Users/vinay/Downloads/Personalized_Human_Robot_Interaction_wit.pdf.

[2] Furhat, October 2023. URL: https://docs.furhat.io/.

A Appendix

A Project Overview

A.1 Context

A.1.1 Elevator pitch:

A real-time interaction with the Furhat robot that responds in socially adaptive ways.

A.1.2 Objective:

The objective of the project is to develop a system that enables the Furhat virtual robot to autonomously analyze human behavior and affective states during interactions and respond with socially adaptive behaviors. The project involves two interconnected subsystems. The User Perception subsystem utilizes real-time webcam input to automatically extract facial features from human users, employing machine learning techniques to process these features into a high-level representation of affective states. The Interaction sub-system takes inputs from the User Perception sub-system, incorporating both the detected affective states and user speech. This sub-system generates socially adaptive behaviors for Furhat, responding appropriately to users' affective states and the context of the interaction.

A.1.3 Deliverables:

The project deliverables encompass a detailed project report outlining the chosen scenario, design decisions, and methodology. This report includes comprehensive documentation on the User Perception sub-system, covering facial feature extraction, machine learning techniques employed, and details regarding the training data. Similarly, the Interaction sub-system is documented, highlighting the rule-based approach for generating Furhat's adaptive behaviors in response to affective states and user speech. The project also entails a well-organized codebase for both sub-systems, accompanied by setup instructions. A live demonstration or presentation showcases the system's functionality, supported by presentation slides summarizing key aspects.

A.1.4 Success Metrics:

We'll keep checking how fast the system responds in real-time and how accurate it is. We use measures like accuracy to see how well the Machine Learning model is doing at figuring out people's feelings. To make sure the part that picks up facial features is working right, we compare its accuracy to what we know is correct. We'll also keep an eye on how easily people can change the system and what they say about it. When we're adapting the dataset, we check how well it matches what we know about people's emotions. We regularly look at our progress, discuss it together, and make any needed changes to make sure we're meeting our goals.

A.1.5 Potential Issues:

One big challenge might be finding the right balance between making the system work quickly and making sure it's really good at figuring out people's feelings. It's tough because making it work fast might clash with the need for more complex and accurate Machine Learning modules. Also, training the module to understand lots of different emotions and how people interact can take a long time. We'll need to be careful in choosing and adjusting how the module learns. Making the system flexible for users, especially in real-life situations, might be tricky too because people can have unexpected preferences and situations. So, to overcome these challenges and make sure the quick and accurate system works well, we need to plan carefully, keep testing things, and make improvements as we go.

A.2 Project Breakdown

A.2.1 Project Kickoff and Planning

- **Deadline:** End of Week 47

- **Responsible Members:** Entire Group
- Define project goals, milestones, and roles. Develop a detailed project plan.

A.2.2 Implement User Perception Sub-system

- **Deadline:** End of Week 48
- **Responsible Members:** Sarthak Vijayvergiya and Vinay Mehra
- Develop the code for acquiring real-time video input from the webcam and integrate Py-Feat or similar tools for facial feature extraction.

A.2.3 Facial Feature Extraction Module

- **Deadline:** End of Week 48
- **Responsible Members:** Sahil Sandal, Hamza Aziz Khan, and Shivam Srivastava
- Implement and test facial feature extraction module.

A.2.4 Machine Learning Model Training

- **Deadline:** End of Week 49
- **Responsible Members:** Sarthak Vijayvergiya, Vinay Mehra, and Hamza Aziz Khan
- Train Machine Learning model.

A.2.5 Integration and Testing of User Perception Sub-system

- **Deadline:** End of Week 50
- **Responsible Members:** Entire Group
- Integrate the components of the User Perception sub-system.

A.2.6 Initial Rule-Based System for Interaction Sub-system

- **Deadline:** End of Week 51
- **Responsible Members:** Entire Group
- Design and implement basic rules for Furhat's adaptive behaviors based on affective states. Set up communication between the two subsystems.

A.2.7 Documentation and Training

- **Deadline:** End of Week 52
- **Responsible Members:** Sahil Sandal and Shivam Srivastava
- Prepare comprehensive documentation and conduct effective training sessions.

A.2.8 Final Testing and Refinement

- **Deadline:** End of Week 1
- **Responsible Members:** Entire Group
- Conduct final testing and refine system components based on feedback sessions.

A.2.9 Presentation Preparation

- **Deadline:** End of Week 1
- **Responsible Members:** Entire Group
- Prepare presentation materials showcasing project capabilities and results.

A.2.10 Final Report Submission

- **Deadline:** End of Week 2
- **Responsible Members:** Entire Group
- Compile and submit the final project report, including documentation, code, and additional materials.