

Empathetic AI: Real-Time Multimodal Emotion Detection in CHATBOT

Presented By:

Syed Shakil Mahmud

Id: 22208019

Session: 2021-22

**Department of CSE
Comilla University**

Supervised By:

Dr. Mahamudul Hasan, SMIEEE

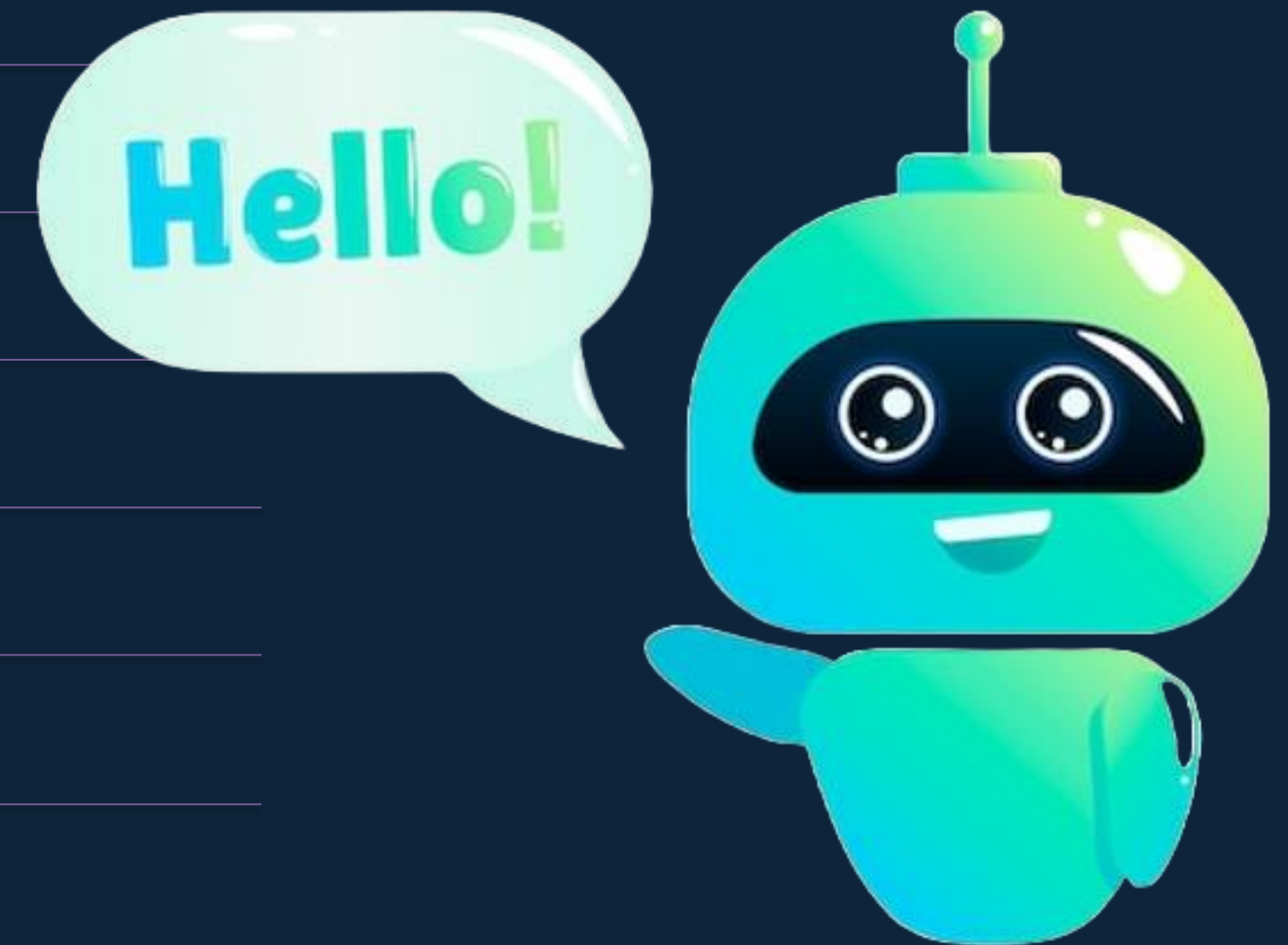
Associate Professor

Department of CSE

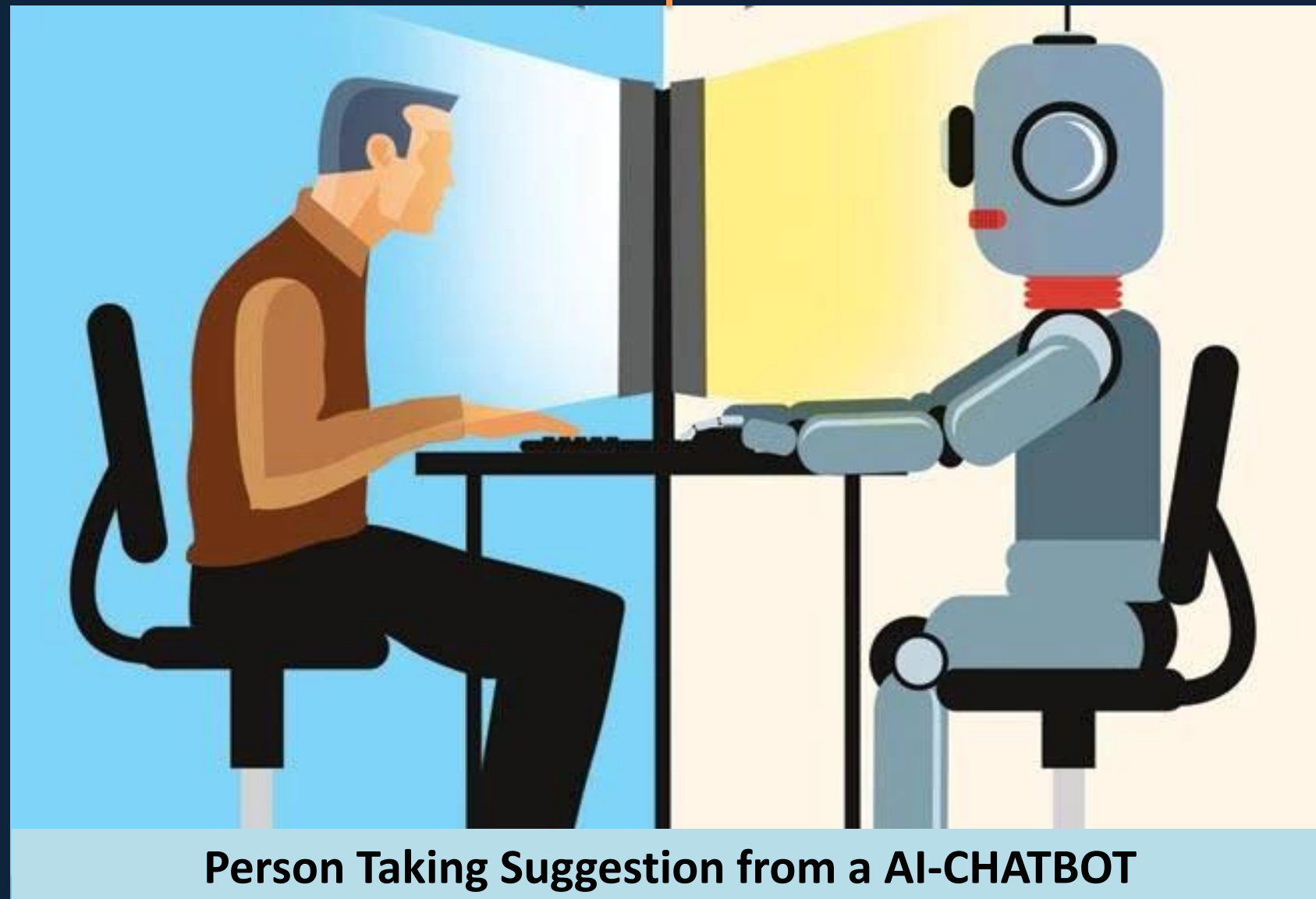
Comilla University

Content Outlines

- 🎯 **Introduction**
- 🎯 **Motivation**
- 🎯 **Objectives:** The novelty of our developed system.
- 🎯 **Literature Review:** Existing system limitations
- 🎯 **Contribution:** What Made Our Model Extraordinary?
- 🎯 **Methodology**
- 🎯 **Experimental Results**
- 🎯 **Conclusion and Future Work**
- 🎯 **References**



Let's Introduce with Our System



Person Taking Suggestion from a AI-CHATBOT

Emotions play a critical role in human interaction.

Multimodal data (facial expressions, speech, gestures) conveys emotions.

Goal: Build an emotion recognition system using audio-visual data.

Integrate this system into a CHATBOT for real-time emotion detection.

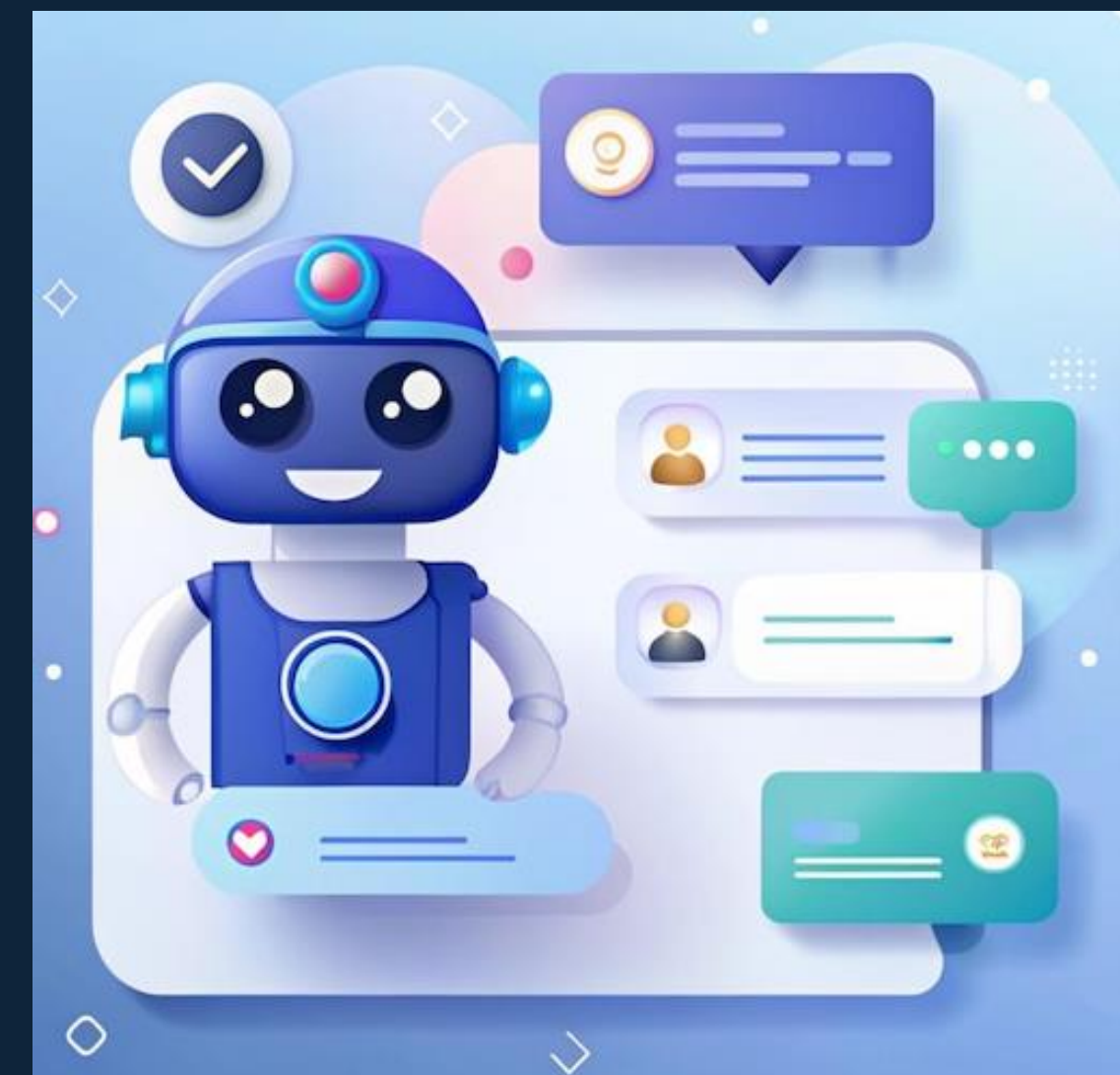
Applications: Call centers, healthcare, marketing, education.

Motivation behind the Idea!



- Enhance healthcare by early detection of mental health conditions.
- Provide more empathetic and responsive human-computer interactions.
- Real-time emotion recognition for better decision-making.

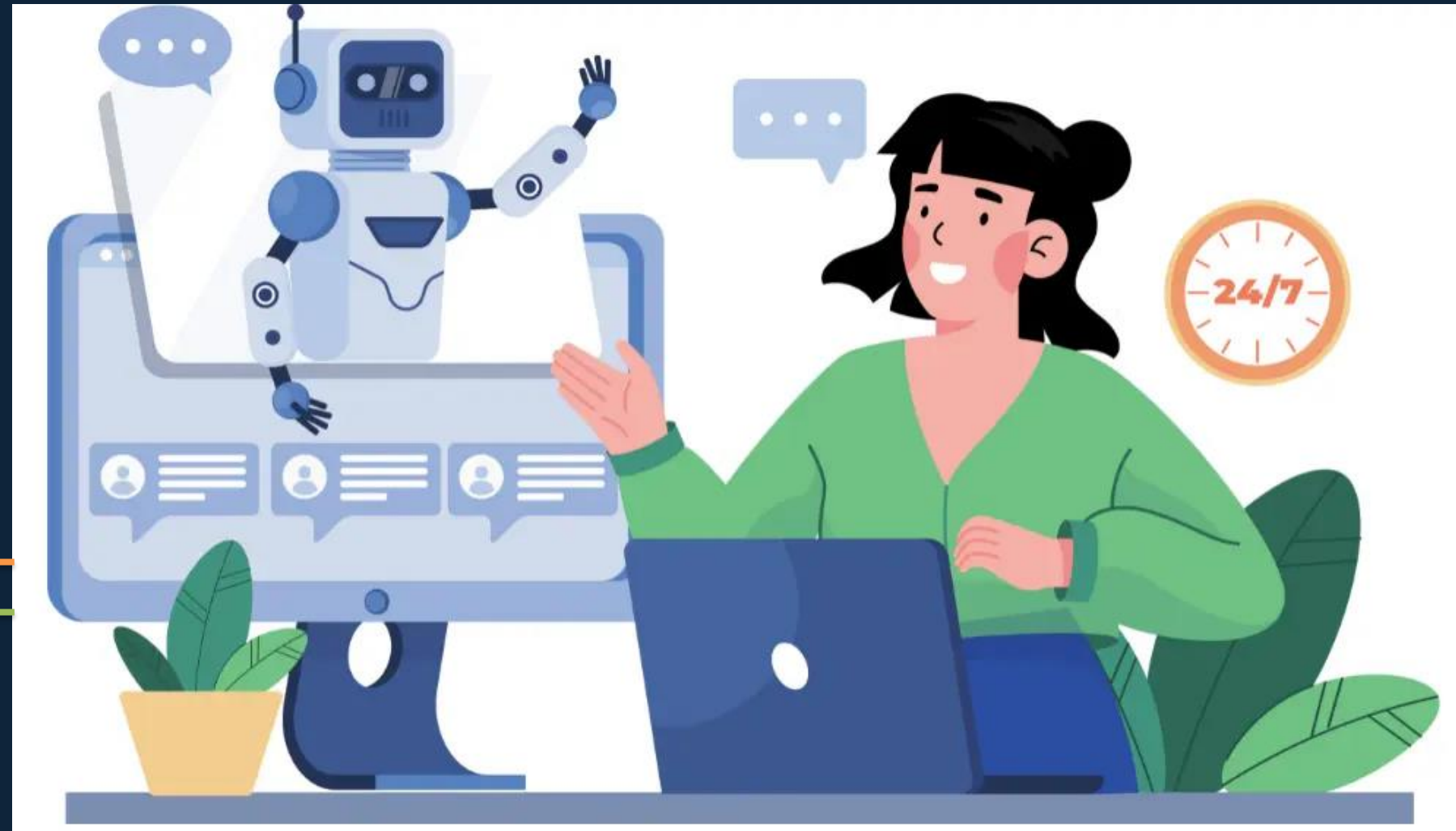
- Traditional systems do not capture the complexity of emotions.
- Improving customer service through emotion detection.
- Assist teachers in understanding students' engagement and feelings.



The novelty of our developed system.

Necessary objectives of our System

Integrate the model into a CHATBOT system for real-time interaction.



Replace human agents and can perform general tasks by recognizing emotions.

Develop a multimodal emotion recognition model using audio-visual data.

Person Taking Suggestion from a AI-CHATBOT

Deploy in practical settings: call centers, education, healthcare.

Utilize the IEMOCAP dataset for training and feature extraction.

Provide Instant Support

Automate Your Social Media Support

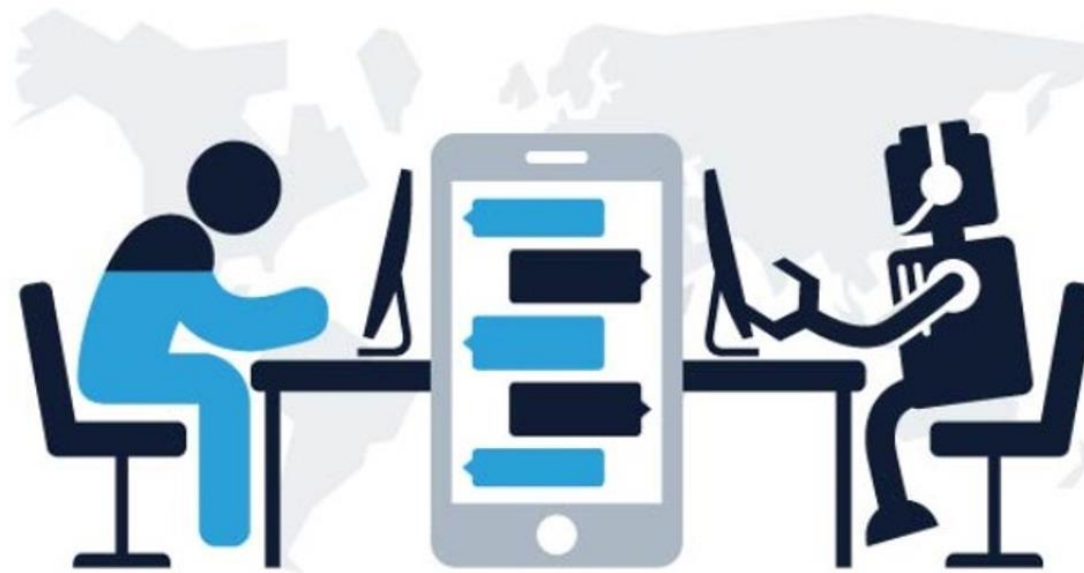
Improve Productivity of Support Agents

Deliver a Better Experience

Collect Real-time Feedback

Stay Available
24×7

67% of consumers worldwide used a chatbot for customer support in the past year



source: <https://www.revechat.com>

Ensure a Seamless Hybrid Support

Achieve Easy Scalability

Minimize Customer Support Costs

Literature Review

Existing system limitations.

Author	Domain Name	Solved Problem	Method	Limitations
[1] Almulla, M. A. [2024]	A MER system using deep convolution neural networks	Recognizing emotions from audio, video and text data	Deep convolution, Decision level fusion	<ul style="list-style-type: none"> - Can recognize only 7 emotions - No deployment in real world scenarios
[2] Zhang, S., Yang, Y., Chen, C., Liu, R., Tao, X., Guo, W., ... & Zhao, X. [2023]	MER based on audio and text by using hybrid attention networks	Detect emotion through Audio and text	Hybrid attention networks	<ul style="list-style-type: none"> - Only used audio and text modalities. - Less modalities - No real time deployment
[3] Bhat, A. A., Kavitha, S., Satapathy, S. M., & Kavipriya, J. [2024]	Real Time Bimodal Emotion Recognition using Hybridized Deep Learning Techniques	This project tried to recognizing emotion using audio and video	PCA before training, Uses Deep Learning, Techniques. Hybrid CNN-LSTM is used	<ul style="list-style-type: none"> - Audio is not used - Low accuracy - No updated technology used - No real time deployment
[4] Tan, Y., Sun, Z., Duan, F., Solé-Casals, J., & Caiafa, C. F. [2021]	A MER method based on facial expressions & electroencephalography	They focused on recognize emotion through image and electroencephalography	Fusion by HRI (Human Robot Interaction) system	<ul style="list-style-type: none"> - System applied in lab - Need to adjust the system in daily life
[5] Praveen, R. G., Cardinal, P., & Granger, E. [2023]	Audio-Visual Fusion for ER in the Valence-Arousal Space Using Joint Cross-Attention	They focused on Emotion recognition in the valence arousal space	Used Joint Cross-Attention model for A-V fusion	<ul style="list-style-type: none"> - No real time integration with CHATBOT

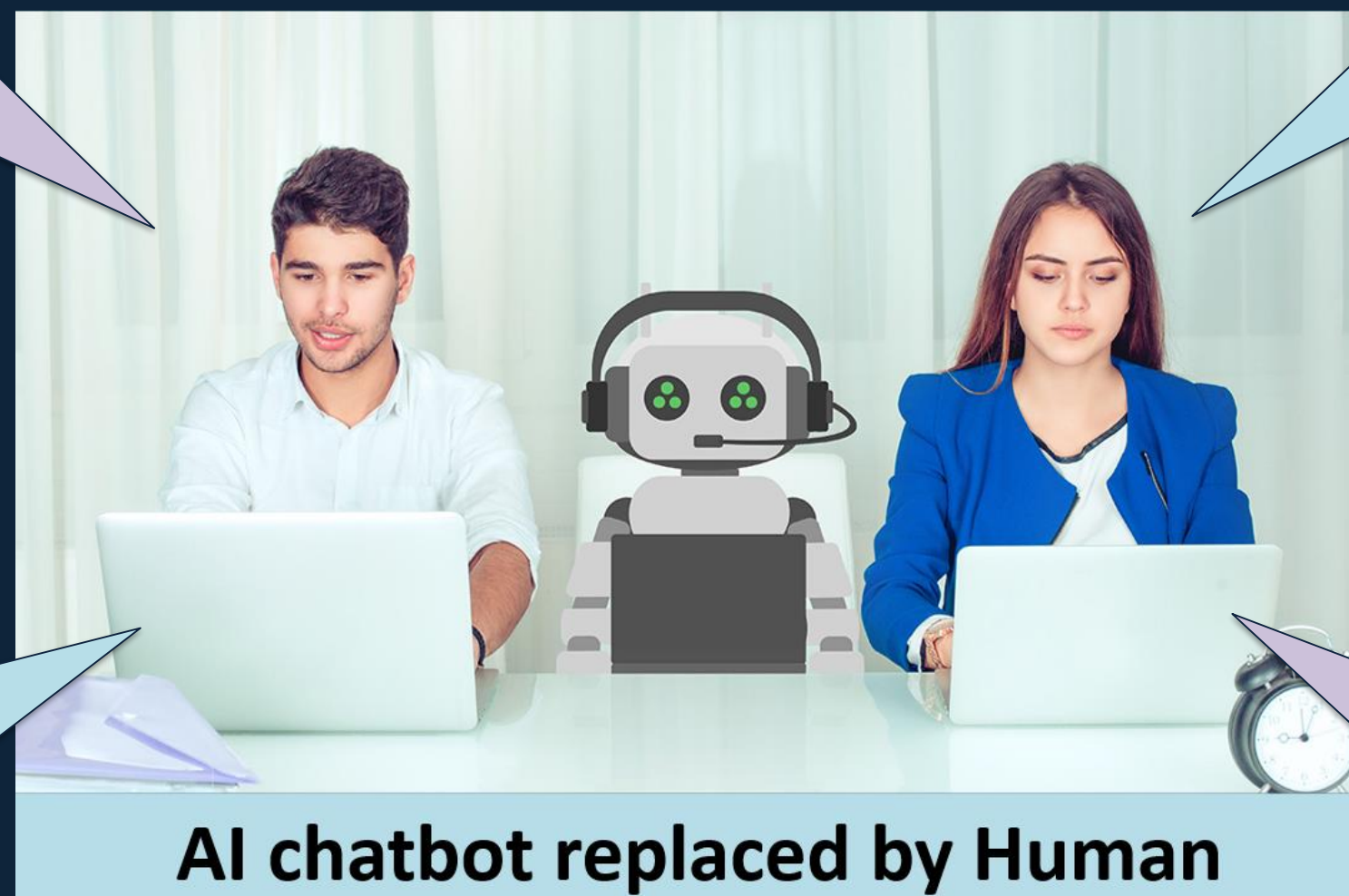
What Made Our Model Extraordinary?

The Contribution of us that made the model unique!

Develop a reliable multimodal emotion recognition system using deep learning.

Use of the IEMOCAP dataset with emotion annotations across modalities.

Real-time integration with CHATBOTS for real-world applications.



AI chatbot replaced by Human

Enhanced model accuracy through the fusion of separate audio and video models.

Methodology(Cont.)

A. Data Collection

🎯 The Interactive Emotional Dyadic Motion Capture (IEMOCAP) database is an acted, multimodal and multispeaker database

🎯 We have Collected the IEMOCAP dataset from SAIL lab at USC

🎯 It contains approximately 12 hours of audiovisual data, including video, speech, motion capture of face, text transcriptions.

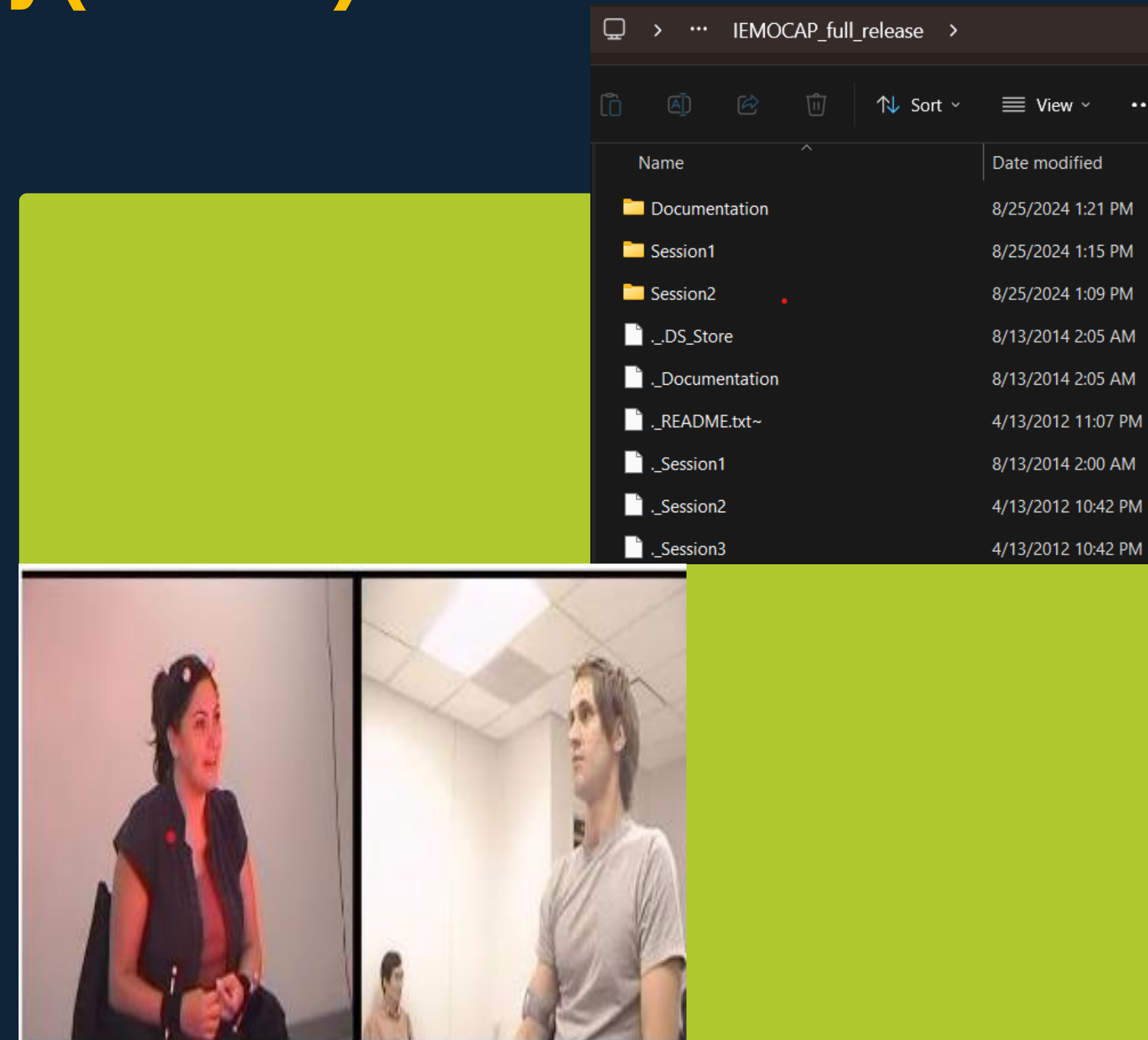
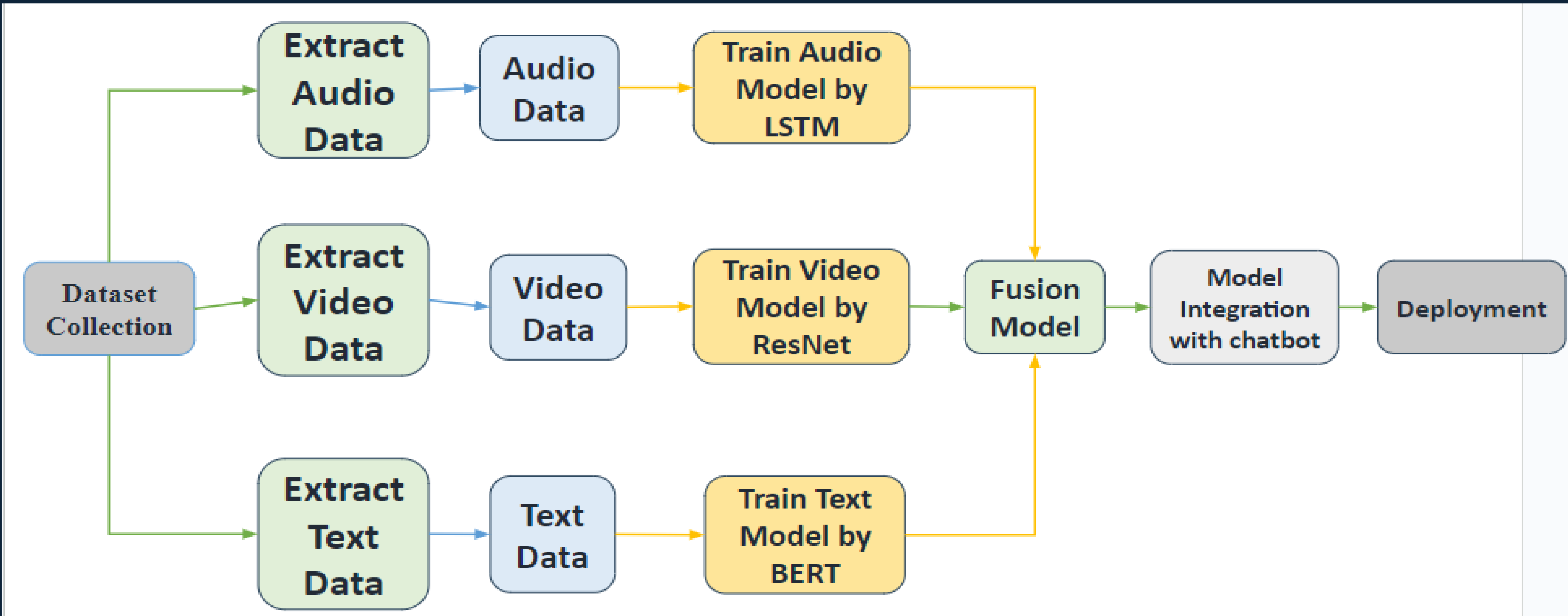


Figure 1. Sample Dataset

Methodology(Cont.)

B. Overall Processing



What Are We Expecting In The Long Run?

Accurate real-time emotion recognition in CHATBOTS.

Deployment in diverse fields: customer service, healthcare, education.

Enhanced customer interaction and user experience.

Establishment of a foundation for future research in multimodal emotion recognition.







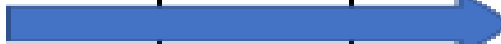


Conclusion and Future Work

- **This project aims to develop an emotion recognition model integrated into a CHATBOT.**
- **Real-time detection of emotions in practical settings (call centers, healthcare, etc.).**
- **The CHATBOT will improve user interactions through empathy and responsiveness.**

OUR NEXT STEP

- 🎯 **We will make a dataset consisting of Bengali voices and shoot video faces with Bangla transcripts**

Timeline

No	Activity / Task	1 st & 2 nd Semester											
		Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	June
01	Meetings with supervisor												
02	Topic search												
03	An in-depth review of the literature on relevant existing algorithms												
04	Evaluation of the designed algorithms and verifications												
05	Find an optimum ML algorithm and model to improve consensus components												
06	Refine the designed algorithms for the best fit if necessary												
07	Implementation												
08	Result analysis & Evaluation												
09	Final thesis writing												

References

- [1] **Almulla, M. A. (2024). A multimodal emotion recognition system using deep convolution neural networks. Journal of Engineering Research.**
- [2] **Zhang, S., Yang, Y., Chen, C., Liu, R., Tao, X., Guo, W., ... & Zhao, X. (2023). Multimodal emotion recognition based on audio and text by using hybrid attention networks. Biomedical Signal Processing and Control, 85, 105052.**
- [3] **Bhat, A. A., Kavitha, S., Satapathy, S. M., & Kavipriya, J. (2024). Real Time Bimodal Emotion Recognition using Hybridized Deep Learning Techniques. Procedia Computer Science, 235, 1772-1781.**
- [4] **Tan, Y., Sun, Z., Duan, F., Solé-Casals, J., & Caiafa, C. F. (2021). A multimodal emotion recognition method based on facial expressions and electroencephalography. Biomedical Signal Processing and Control, 70, 103029.**
- [5] **Praveen, R. G., Cardinal, P., & Granger, E. (2023). Audio-visual fusion for emotion recognition in the valence-arousal space using joint cross-attention. IEEE Transactions on Biometrics, Behavior, and Identity Science.**
- [6] **Dresvyanskiy, D., Markitantov, M., Yu, J., Kaya, H., & Karpov, A. (2024). Multi-modal Arousal and Valence Estimation under Noisy Conditions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4773-4783).**

Thank You!

