UE22CS320A- Project Phase - 2

Capstone Project –Phase 2 –review 1

Project Title : Real Time Context-Aware Communication Assistance for Autistic Individuals

Project ID : 166

Project Guide : Dr. Suja C.M

Project Team : Nishruth Chowdary (PES2UG22CS194)
Gowri M.Budhial (PES2UG22CS202)
Joshua John (PES2UG22CS238)
Jumanah Maheen (PES2UG22CS241)

Agenda

- ➢ Problem Statement
- ➢ Abstract and Scope
- ➢ Suggestions from Phase-1
- ➢ Requirements Specification
- ➢ Design Approach
- ➢ Design Constraints, Assumptions & Dependencies
- ➢ Proposed System / Approach
- ➢ Architecture
- ➢ Design Description
- ➢ Project Progress
- ➢ References

Problem Statement

Real –Time Communication assistance intent for Autistic Individuals.

Developing a machine learning model to predict and assist in generating communication intent for autistic individuals.

- Analyzing real-time facial expressions, behavioral cues, and environmental factors.
- Interpret emotions and suggest responses tailored to individual needs.
- Enhance communication and provide supportive interactions.
- Improve daily interactions by predicting communication intent.

Abstract & Scope

Abstract

- Effective communication is a persistent challenge for many autistic individuals. This often leads to frustration, social isolation, and misunderstandings in daily interactions. Traditional systems, relying on static templates or predefined vocabularies, fail to adapt to the nuanced and dynamic communication needs of autistic individuals.

- This project aims to develop a **machine learning-based system** capable of real-time prediction and generation of communication intent by analysing multimodal data, including facial expressions, behavioural cues, and environmental context. The system will detect emotional states such as frustration, happiness, or anxiety and provide personalized, contextually relevant responses tailored to the individual's situation and emotions.

Abstract & Scope

Scope

- **Facial Expression and Behaviour Analysis**:
  - Utilising video data to detect emotional states by analysing facial expressions and body language patterns.
- **Emotion Interpretation**:
  - Employing real-time ML techniques to interpret emotional states, enabling the system to understand nuanced emotional changes.
- **Multimodal Data Integration**:
  - Combine inputs from facial expressions, behavioural cues, and environmental context to enhance sentiment classification.
- **Response Generation**:
  - Provide tailored, context-aware responses that align with the user's emotions and conversational needs.
- **Impact on Daily Communication**:
  - Enhance day-to-day interactions by predicting intent, generating empathetic responses, and reducing frustration for both the individual and their communication partners.

Suggestions from Phase - 1

Feasibility of Real-Time Processing

Optimized Model Design

- **Lightweight CNNs & LSTMs** ensure low-latency inference.
- Real -Time Video Processing using OpenCV

Feasibility of Data Collection & Processing

Data Availability

- Publicly available datasets (**AffectNet, FADC, Emoreact**) provide **facial and behavioral data**.

Requirement Specifications

Functional Requirements

1. Data Collection & Processing
    - The system should capture facial expressions from images or videos.
    - The system should detect behavioral gestures (e.g., hand flapping, gaze aversion).
    - The system should record environmental context (e.g., noise levels, lighting conditions).
    - The system should preprocess and normalize multimodal data for uniform feature extraction.

2. Multimodal Analysis & Intent Prediction
    - The system should extract features from each modality (facial, behavioral, environmental).
    - The system should use fusion to merge features before classification.
    - The system should predict communication intent (e.g., request help, express joy).
    - The system should generate adaptive responses based on detected intent.

Requirement Specifications

Functional Requirements

3. Personalization & Learning
    - The system should allow user-specific fine-tuning for personalized predictions.
    - The system should integrate feedback mechanisms for model improvement.

4. User Interface & Deployment
    - The system should provide a simple, interactive UI for caregivers and therapists.
    - The system should work on mobile, cloud, and edge devices for real-time deployment.

Requirement Specifications

Non - Functional Requirements

1. High Accuracy & Scalability
    - The system should process real-time video and sensor data with low latency.
    - The system should support scalability to accommodate additional users and datasets.

2. Security & Privacy
    - The system must be secure for handling sensitive personal data.

3. Maintainability & Compatibility
    - The system should be modular to support future enhancements.
    - It should be compatible with TensorFlow, PyTorch, and cloud services.

Requirement Specifications

Technologies Used & Why

1. TensorFlow / PyTorch

Powerful frameworks for developing & training deep learning models for emotion recognition, behavior analysis, & multimodal fusion.

2. OpenCV & Mediapipe

Essential for facial expression recognition, video processing, and landmark detection in real-time.

3. Scikit-learn

Used for training simpler models like classifiers for behavior analysis and environmental input prediction.

4. React/NodeJs

To develop the web or mobile interfaces for caregivers and users to interact with the system.

5. MySQL/MongoDb

For storing user data, logs, training data, and predictions for later analysis and customization.

6. Cloud Computing (AWS)

Scalable resources for model training and real-time data processing, ensuring system availability for multiple users.

7. Multimodal Fusion

To combine facial, behavioral, and environmental data effectively, leading to more accurate communication intent predictions.

Design Approach Followed & Justification

Approach

➔ **Multimodal Machine Learning with Late Fusion**

The system follows a multimodal deep learning approach with late fusion, meaning that different modalities (facial expressions, behavioral gestures, environmental context) are processed separately before being combined at a later stage for final decision-making.

Design Constraints, Assumptions & Dependency

Design Constraints

1. **Real-Time Processing** → Requires low-latency models (optimized CNNs, LSTMs) for quick inference.
2. **Data Availability & Annotation** → Limited autism-specific datasets, requiring transfer learning.
3. **User Variability** → Different individuals exhibit unique gestures and expressions, necessitating personalization.
4. **Multimodal Synchronization** → Aligning facial, behavioral, and environmental data is complex, requiring time synchronization techniques.
5. **Ethical & Privacy Concerns** → Needs secure data storage and user consent mechanisms to ensure privacy.

Design Constraints, Assumptions & Dependency

Assumptions in the Design Approach

1. **Multimodal Data Enhances Accuracy** → Combining **facial, behavioral, and environmental data** improves intent prediction.
2. **Optimized Models Enable Real-Time Inference** → **Edge computing** and **lightweight architectures** will allow low-latency predictions.
3. **Personalization is Achievable** → Fine-tuning with **small-scale user data** will adapt the system to individuals.

Design Constraints, Assumptions & Dependency

Dependencies & Impact

1. **Dataset Dependencies** → Requires quality labeled data (AffectNet, FADC, EmoReact, DEAP).
2. **Hardware Dependencies** → Needs GPUs (NVIDIA A100, RTX 3090) for training & edge processing for real-time use.
3. **Model Dependencies** → Uses CNNs (Facial), LSTMs (Behavioral), Transformer-based Fusion.
4. **Privacy & Ethics** → Must comply with GDPR, HIPAA for facial & behavioral data security.
5. **Personalization & Feedback** → Requires users feedback for fine-tuning predictions.

Design Details

Project Dependencies & Required Changes

1. **Novelty & Innovativeness** – Uses multimodal AI (Facial, Behavioral, Environmental) for autism communication.
    - **Change**: Improve autism-specific datasets for better accuracy.
2. **Interoperability** – Integrates TensorFlow, PyTorch, and Edge Computing for real-time use.
    - **Change**: Standardize APIs for multimodal data fusion.
3. **Performance & Reliability** – Requires GPU acceleration (CUDA, TensorRT) for real-time inference.
    - **Change**: Optimize using pruning.
4. **Security & Privacy** – Handles sensitive facial and behavioral data (GDPR, HIPAA compliance).
    - **Change**: Implement data encryption and secure storage.
5. **Legacy to Modernization** – Shifts from static rule-based systems to adaptive AI.
    - **Change**: Enable real-time learning and personalization.
6. **Reusability & Compatibility** – Can be extended to healthcare, education, and smart assistants.
    - **Change**: Follow a modular design for broader integration.

Proposed System / Approach

Proposed System: Late Fusion

Reasons for Choosing Late Fusion

1. **Independent Processing** – Each modality (facial, behavioral, environmental) is processed separately, allowing independent model improvements.
2. **Real-Time Efficiency** – Since models run separately before merging, inference is faster.
3. **Flexibility** – Each model can be updated without affecting the others.
4. **Handling Missing Data** – If one modality is unavailable, the system can still make predictions using the others.
5. **Lower Computational Cost** – Requires less processing power as individual models can be optimized separately.

Proposed System / Approach

Approach 2: Mid-Level Fusion (Alternative Approach)

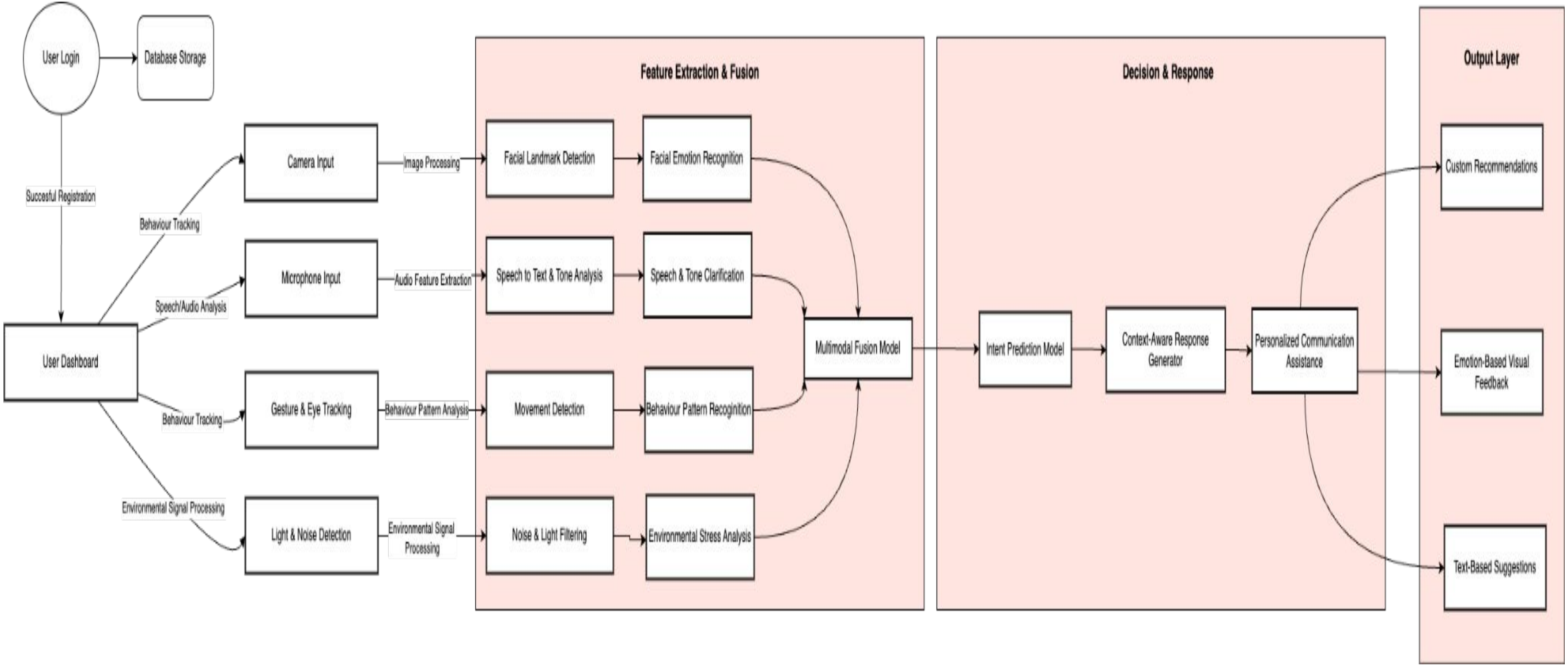
Reasons for Considering Mid-Level Fusion

1. **Better Accuracy** – Combining features at an intermediate stage allows deeper interaction between modalities.
2. **Improved Synchronization** – Data from different modalities is merged before classification, reducing alignment issues.
3. **Context-Aware Decision Making** – Environmental and behavioral data influence facial expression analysis for better predictions.

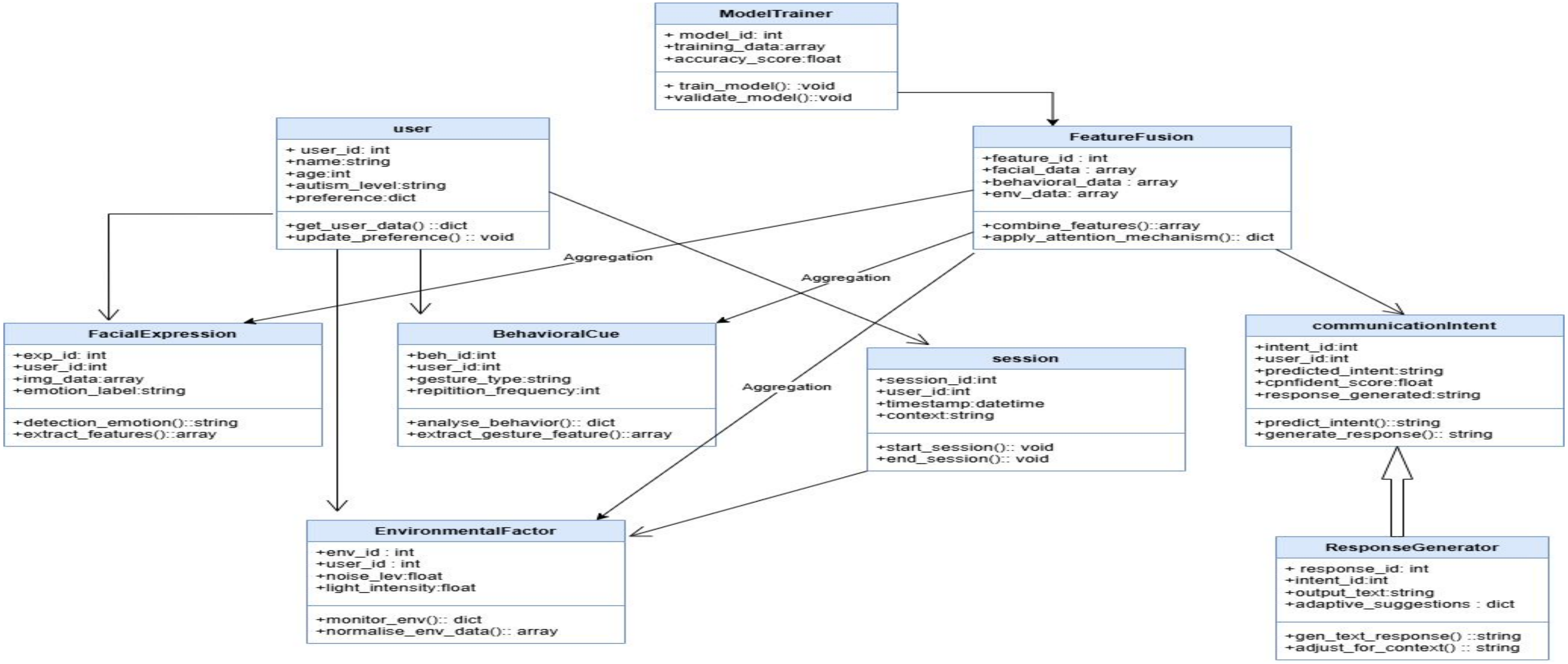
Drawbacks of Mid-Level Fusion

1. **Higher Computational Cost** – Processing and merging features before classification requires more resources.
2. **Complex Training** – Tuning the fusion layers correctly requires additional effort.
3. **Dependency on All Modalities** – If one modality is missing, the feature combination may be incomplete.

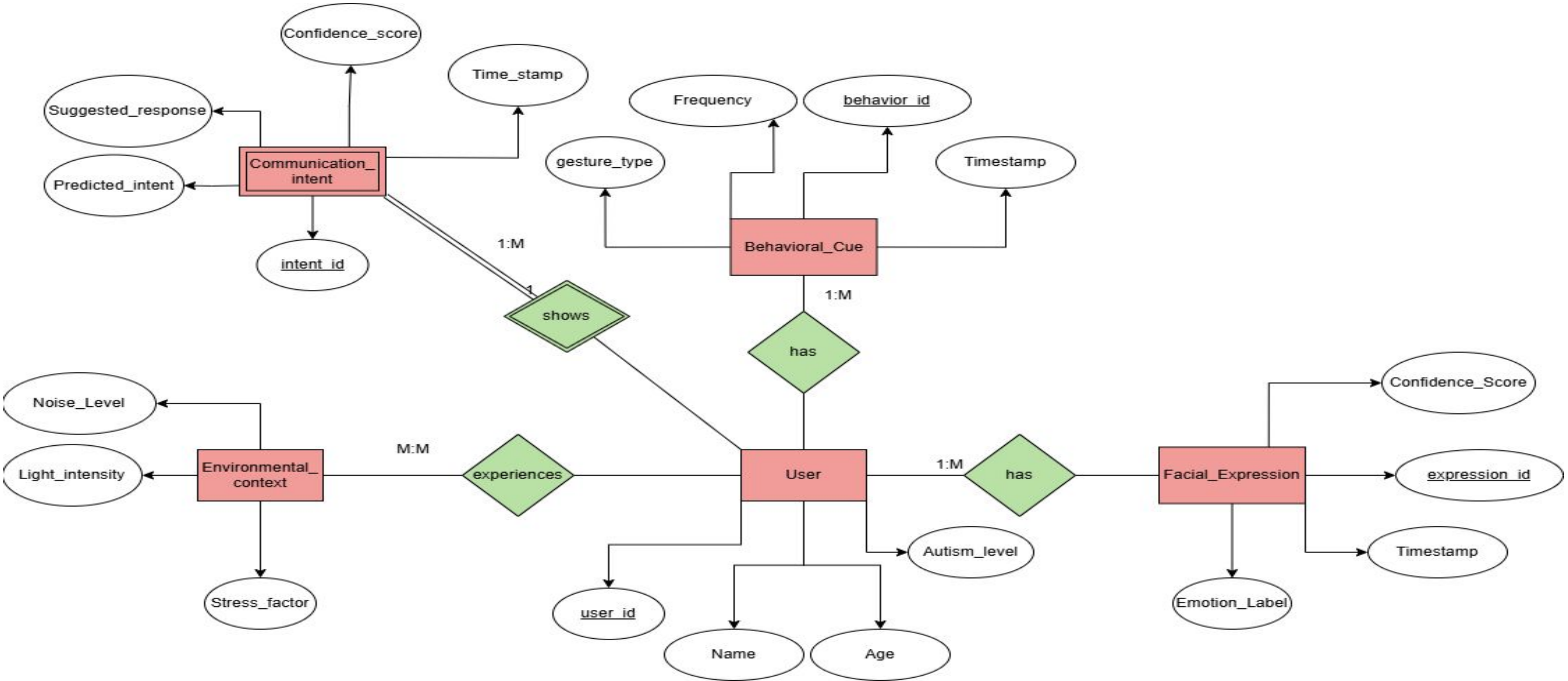
Architecture



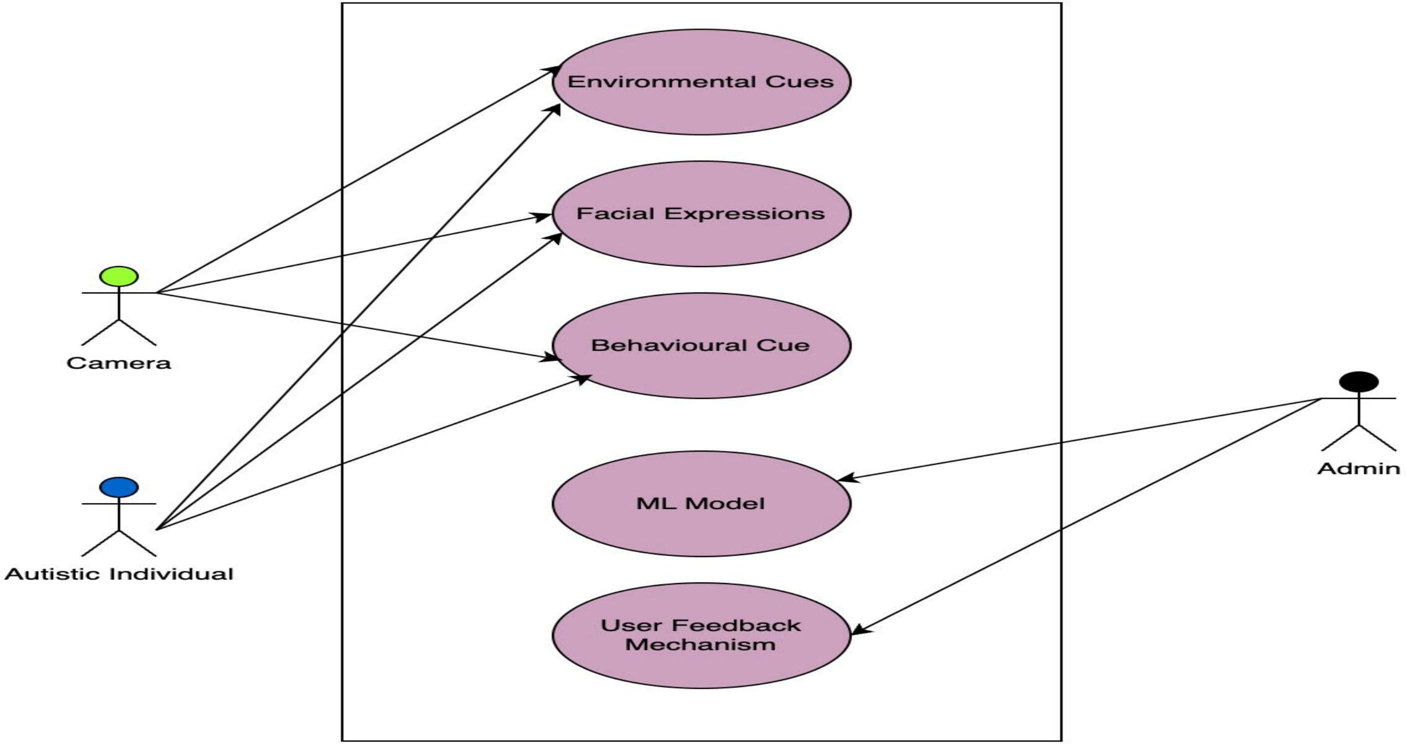
Master Class



ER Diagram


Confidence_score
Time_stamp
Suggested_response
Communication_intent
Predicted_intent
intent_id
1:M
1
shows
Frequency
behavior_id
gesture_type
Timestamp
Behavioral_Cue
1:M
has
Noise_Level
Light_intensity
Environmental_context
Stress_factor
M:M
experiences
User
user_id
Name
Age
Autism_level
1:M
has
Facial_Expression
Confidence_Score
expression_id
Timestamp
Emotion_Label

Use Case Diagram



External Interfaces

1. User Interfaces (UI/UX)

- **Web-Based Interface (React.js)** – Enables users to interact with the system, view predictions, and adjust responses.
- **Dashboard for Communication Intent Prediction** – Displays model outputs, predicted intent, confidence scores, and real-time analysis of facial, behavioral, and environmental data.

2. Hardware Interfaces

- **Microphone & Camera** – Captures speech, facial expressions, and behavioral cues for multimodal communication intent detection.

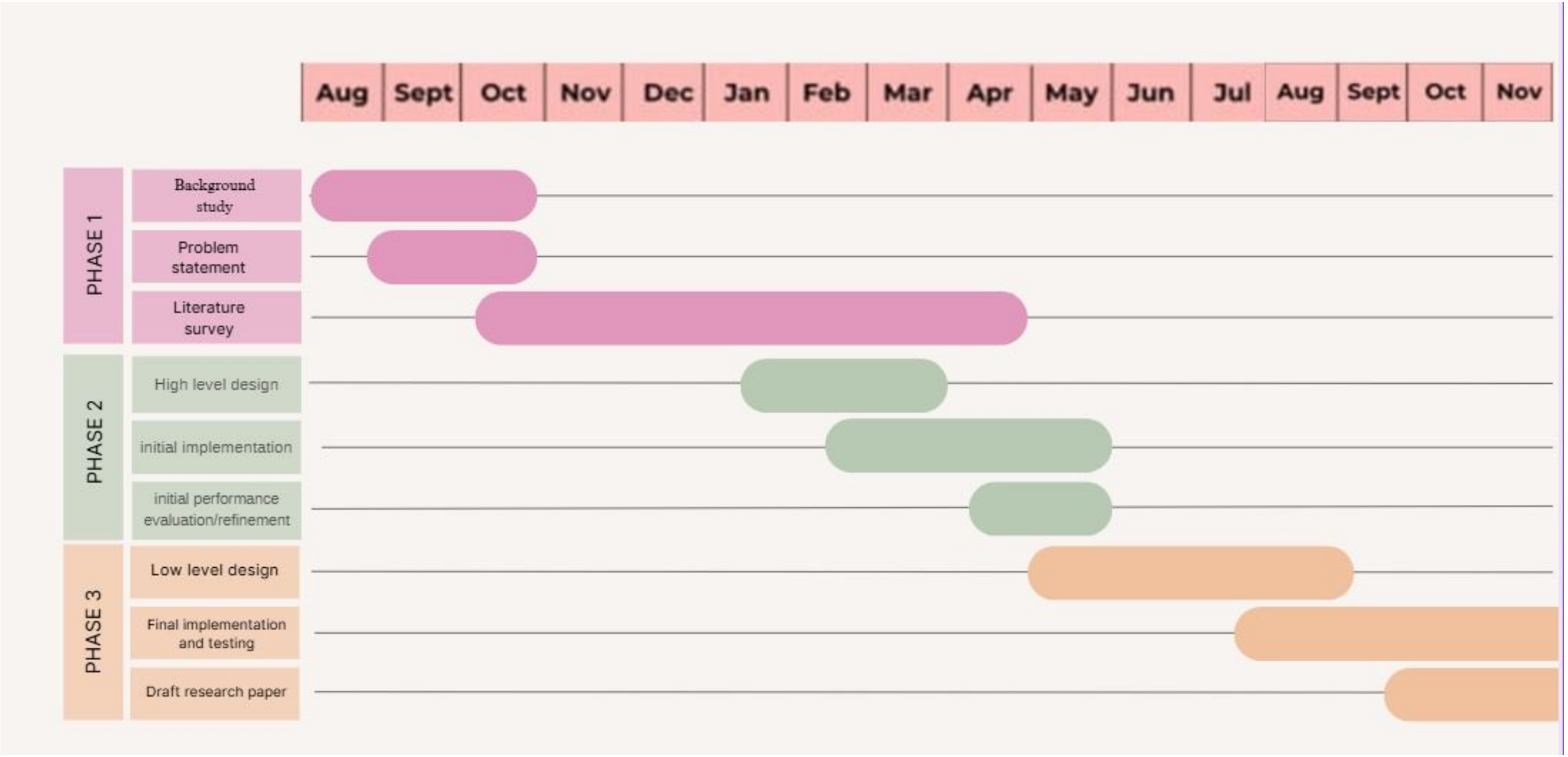
3. Software Interfaces

- **TensorFlow/PyTorch/Keras** – Used for training and inference of deep learning models for facial, behavioral, and environmental analysis.
- **OpenCV & ResNet50** – Extracts facial expression features from video frames for emotion recognition.

4. Database Interfaces

- **MySQL (SQL Database)** – Stores user profiles, interaction logs, model predictions, and caregiver feedback.
- **CSV Files (train.csv, test.csv)** – Holds labeled training and testing datasets for multimodal analysis.
- **MongoDB (if required for NoSQL storage)** – Can store real-time logs and multimodal metadata for continuous learning.

Capstone (Phase-1, Phase-2, Phase-3) Project Timeline



Conclusion

1. The development of a context-aware communication assistance system for autistic individuals addresses the critical challenge of expressing emotions, needs, and intentions in real-time.
2. By integrating facial expressions, behavioral gestures, and environmental factors, the system enhances the ability of autistic individuals to communicate effectively.
3. The use of multimodal machine learning with Late Fusion ensures that each modality is processed independently before combining results, allowing for real-time decision-making, flexibility, and modular improvements.
4. Despite challenges such as data availability, synchronization issues, and computational constraints, the system remains a promising step toward personalized assistive communication technologies.
5. This project contributes to the broader goal of inclusive AI-driven assistive technology, ensuring greater independence, reduced frustration, and improved social interactions for autistic individuals.

References

AACessTalk: Fostering Communication between Minimally Verbal Autistic Children and Parents with Contextual Guidance and Card Recommendation:
https://arxiv.org/abs/2409.09641v2

Dasom Choi, SoHyun Park, et al. AACessTalk: Fostering Communication between Minimally Verbal Autistic Children and Parents with Contextual Guidance and Card Recommendation.
https://arxiv.org/abs/2409.09641v2

Robin Cromjongh, et al. Adaptive Context-Aware Planning Support for Students with Autism.
https://dl.acm.org/doi/abs/10.1145/3565472.3595610

Ziying Tang, Jin Guo, Sheng Miao, Subrata Acharya, Jinjuan Heidi Feng, et al. Ambient Intelligence Based Context-Aware Assistive System to
Improve Independence for People with Autism Spectrum Disorder
https://ieeexplore.ieee.org/document/7427601

Antonio Iannone 1 and Daniele Giansanti, et al. Breaking Barriers—The Intersection of AI and Assistive Technology in Autism Care: A Narrative Review
https://pubmed.ncbi.nlm.nih.gov/38248742/

Thank You