IOTA CLUSTER

GC-25

MACHINE LEARNING HACKATHON

Introduction

Welcome to the GC-25 ML Hackathon! This two-day competition challenges hostel teams to push the boundaries of AI through two cutting-edge Problem Statements:

- **NLP Track:** Build an advanced voice controlled AI desktop assistant similar to JARVIS.
- **CV Track:** Develop an intelligent video understanding system that answers natural language questions.

Participants will have 24 hours to build their solution, followed by presentations, where teams will showcase their solutions.

EVENT TIMELINE

DEVELOPMENT PHASE: 11th Jan'25 11:59am → 12th Jan'25 11:59am

PRESENTATIONS: 12th Jan'25 9-11pm (Time slots will be communicated to the teams later)

TEAM FORMATION

Two teams can participate per hall (one per PS) with a maximum of 4 members.

POINTS DISTRIBUTION

Per Problem Statement:

- 1st place : 500 points
- 2nd place : 400 points
- 3rd place : 250 points
- Participation points : 50 points

If there are entries for both the Problem Statements from a hostel their participtotal points will be added up.

-> ## NLP TRACK

Problem Statement: Advanced AI Desktop Assistant

Create a comprehensive voice-enabled AI assistant that serves as a powerful interface for computer interaction and task automation. The assistant should

demonstrate advanced contextual understanding and multimodal capabilities.

Core Requirements:

1. Voice Interaction

   - Natural and real-time speech-to-text and text-to-speech capabilities

   - Support for multiple voice accents and speaking styles

2. Computer Control Capabilities

   - File system operations (create, read, update, delete)

   - Application control (launch, close, interact)

   - System settings management

   - Basic automation capabilities

3. Contextual Understanding

   - Maintain persistent chat history per user

   - Context-aware responses across conversation sessions

   - Understanding and executing multi-step commands

   - Memory of user preferences and past interactions

Additional Features:

1. Multi user support with advanced authentication

   - Face recognition and/or Voice biometric verification for user authentication

   - Multi-user support with personalized profiles

2. Multimodal Interaction

   - Webcam-based visual question answering

- Screen context understanding
- Basic Gesture Control
- GUI element interaction and understanding

3. Task Automation
   - Custom macro creation and execution
   - Workflow automation
   - Task scheduling and reminders
   - Script generation and execution

Evaluation Criteria:

- **Core Functionalities (60%)**

1. **Voice Interaction (20%)**: Accuracy, naturalness, and real-time performance.
2. **Computer Control (20%)**: Reliability in file operations, app control, and system management.
3. **Contextual Understanding (20%)**: Coherence, memory of preferences, and multi-step command execution.

- **Additional Features (40%)**

1. **Multi-User Support (10%)**: Secure authentication and personalized profiles.
2. **Multimodal Interaction (10%)**: Visual Q&A, screen context analysis, and gesture control.
3. **Task Automation (10%)**: Macro creation, scheduling, and script execution.
4. **User Experience and Innovation (10%)**: Usability, error handling, accessibility, and novel features.

-> **CV TRACK**

Problem Statement: Video Question Answering System

Develop a Video Question Answering (VideoQA) system that can answer natural language questions based on video clips. The system should demonstrate sophisticated understanding of both visual and audio content while providing accurate, contextually relevant responses.

Core Requirements:

1. Question Types

   - Handle diverse question categories (what, when, where, who, how)

   - Process complex queries about video content

   - Understanding and executing multi-step commands

   - Generate natural language responses

2. Multimodal Processing

   - Integrate visual frame analysis

   - Process audio information

   - Combine multiple modalities for comprehensive understanding

3. Technical Specifications

   - Process video clips up to 30 seconds

   - Support minimum resolution of 720p (1280 x 720)

   - Response delay as close to real-time as possible

Additional Features:

1. Contextual Understanding

    - Handle multi-step questions

    - Remember context from previous questions

    - Cross-reference information across video segments

2. Domain Coverage

    - Support various domains (entertainment, sports, news, daily activities)

    - Handle both audio-visual and visual-only content

    - Adapt to different video styles and formats

Testing Dataset:

The system will be evaluated using a custom dataset incorporating elements similar to:

MSVD, ActivityNet and NextQA.

Test videos will include:

- Both audio-visual and visual-only content
- Multiple domains and scenarios
- Various video quality levels and formats

Evaluation Criteria:

- **Accuracy (40%)**: Precision, contextual relevance, and consistency of answers across diverse question types and scenarios for the test set.
- **Technical Implementation (20%)**: Efficiency, scalability, and robustness in handling video clips with variable resolution and minimal response delay.
- **Multimodal Processing (20%)**: Effective integration of visual, audio, and other modalities for comprehensive understanding of video content.

- **Novelty (20%):** Introduction of innovative features or approaches that enhance the system's capabilities and user experience.

Submission Requirements

1. GitHub repository containing:

- Complete source code
- Documentation
- Setup instructions
- Requirements.txt file
- Demo video (max 5 minutes)

2. README file with:

- Project description
- Architecture diagram
- Installation steps
- Usage instructions

Note - For the CV Track we expect you to clearly define a way for testing your system (can be a simple function that takes in video and questions regarding it or a GUI)

Presentation Guidelines

1. 10 Minute slots will be allotted to each team for their presentations with a buffer of 2 minutes for setup.

2. There are no restrictions on the number of slides in the presentations.

3. The presentations need to be submitted through a separate link 30 minutes before the presentations. The link will be active for the period 8:15pm - 8:30pm 12’th Jan.

General Guidelines

1. Deadlines: Late submissions will incur penalties of 10% for up to 30 minutes. Submissions beyond this will not be accepted.

2. Integrity: Plagiarism or tampering with the components will lead to disqualification.

3. Judging: The organizers reserve the right to modify rules and judging criteria at any time.

Contact- Yugam Bhatt

7082167159