

# # PROJECT NOTES

## ## TIMELINE

1. Gesture recognition & embedding mapping
  - a. Set up gesture recognition
    - [ ] find psychology-based library of gestures
      - ekman & friesen's five types of gestures:
        - emblems
        - illustrators
        - affect displays
        - regulators
        - adaptors
      - kendon's gesture studies
        - phases: preparation, stroke (main motion), retraction.
    - [ ] define core set of gestures
      - common across cultures
      - relevant to conversational AI
      - detectable via existing pose models
    - [ ] collect example data & create labels
      - CMU panoptic or MPII human pose for general body language
    - [ ] mediapipe or openpose
    - [ ] extract features like joint angles
  - b. Convert gestures to embeddings
    - [ ] map keypoints into vectors (PCS or t-SNE)
    - [ ] normalize vectors (L2 normalization)
  - c. Define gesture categories & tokens
    - [ ] manually classify key gestures (e.g., "engaged", "defensive", "confused", etc.)
    - [ ] assign a discrete token for each category
2. LLM integration & testing
  - a. Integrate with LLM (ollama)
    - [ ] concatenate gesture tokens with text input
    - [ ] adjust prompts to include nonverbal context
  - b. Fine-tune responses & test interaction
    - [ ] conduct real-time testing with video input

- [ ] adjust mapping if gestures aren't recognized well
- [ ] build simple UI for demonstration (optional)

## for next week

- [ ] get enough reference literature (and make sure we are doing something more unique) and start implementing mediapipe with python code -> get livestream demo where it detects a gesture from camera
- [ ] embed body language into conversation to add context for LLM conversation,

## REFERENCES

<https://github.com/CMU-Perceptual-Computing-Lab/openpose>