# Large Language Models for Virtual Human Gesture Selection

Parisa Ghanad Torshizi, Laura Hensel, Ari Shapiro, Stacy Marsella

Northeastern University, University of Glasgow

## Problem Statement

- **Co-speech gestures are integral to interactions.**
  Gestures impact comprehension, engagement, recall, and perception in human-human and human-agent interactions.
- **Recent work on gesture generation: Challenges**
  Data-driven approaches:
  Produce a statistical average over gestures
  Have difficulty conveying meaning
  Ignore role and context differences
- **Gesture generation should be context- and role-specific**
  Clinicians, politicians, and comedians use different gestures because they seek to achieve different goals in different contexts
- **Selecting when to gesture, as well as appropriate gestures**
  Identifying the emphasis in speech (rheme-theme and prosody cues) is critical

## Contributions

1. Exploring GPT-4's encoded information on gestures
2. Exploring GPT-4's ability to select appropriate gestures
3. Exploring GPT-4's capabilities to determine when to gesture
4. Evaluating different prompting approaches
5. Automating gesture selection from GPT-4 into a behavior specification
6. Implementing this gesture selection system within the virtual human framework SIMA(Socially Intelligent Multimodal Agent)

## Research Questions

**RQ1:** Do LLMs encode gesture information?
**RQ2:** How do we prompt LLM to select appropriate gestures?
**RQ3:** Do LLMs encode information to determine when a virtual human should gesture?

## Methodology

**To address RQ2:**
Assessment of alternative prompting approaches on the appropriateness of gestures

We annotated the speech of an actual speaker to derive gesture classes and annotated examples

| | |
|---|---|
| **Approach 0:**<br>Gesture Classes NOT provided,<br>Annotated examples NOT provided | **Approach 1:**<br>Gesture Classes provided,<br>Annotated examples NOT provided |
| **Approach 2:**<br>Gesture Classes NOT provided,<br>Annotated examples provided | **Approach 3:**<br>Gesture Classes provided,<br>Annotated examples provided |

**To address RQ3:**
GPT-4 enabled rheme and theme analysis
**Gestures co-occur with the rheme part of an utterance**

## Results

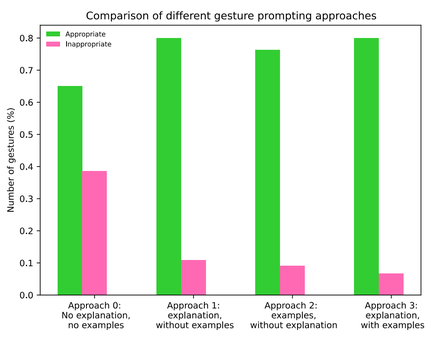
**Results of RQ1:**

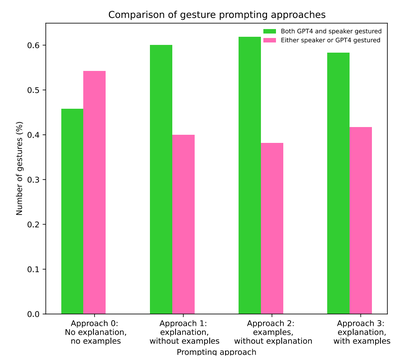
**GPT4: Image schemas** are recurring, dynamic patterns of our perceptual interactions and motor programs … basic cognitive structures that arise from bodily interactions with the world … are understood to be one of the fundamental building blocks of human thought.

**GPT4:** Here's a repertoire of hand gestures that **clinicians** often use:

**Palms Open and Upward:**
Indicates openness
**Hand to Chest:**
Expresses sincerity
**Enclosed Hands or Fingers Interlocked:** attentive listening.

**Results of RQ2:**

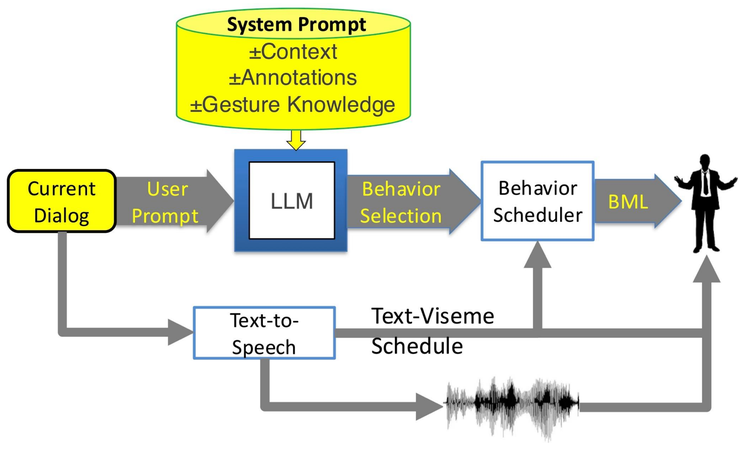



**Results of RQ3:**

43 out of 49 gestures in the annotations occurred within rheme identified by GPT-4

## Implementation




## Conclusions and Future Work

**Overall, the semantic appropriateness of gestures was impressive.**
**Alternative prompting approaches are suitable for different uses.**
- **Approach 0** → creative gesture → Suggesting gestures to a designer
- **Approaches 1,2,3** → Constrained gestures → Automating gesture selection

**Rheme-theme analysis:** Narrow down gesture selection
**Future Work**: Fine-tune smaller models for real-time performance.

## References


- Calbris, Geneviève. "Elements of meaning in gesture." (2011): 1-398.
- McNeill, David. "Hand and mind: What gestures reveal about thought." (1992).
- Nyatsanga, Simbarashe, et al. "A comprehensive review of data-driven co-speech gesture generation." *Computer Graphics Forum*. Vol. 42. No. 2. 2023.
- Torshizi, Parisa Ghanad, et al. "Large Language Models for Virtual Human Gesture Selection." *arXiv preprint arXiv:2503.14408* (2025).


This research is based upon work supported by the NSF Grant #2128743.