Utilizing Generative AI for Patient Behavioral Assessment

Mostafa Habibi, Mehrdad Nourani
Department of Electrical & Computer Engineering
The University of Texas at Dallas
Richardson, Texas 75080, USA

Outline

- Background & Motivation
  - AI-Powered Solution
- Prior Works
- Proposed Methodology
  - Interaction Modeling and Detection
  - Prompt Engineering
  - Summarization using LLM
- Experimental Results
- Conclusion and Future Work

Motivation & Background

- Patient behavior (e.g. ambulation, physical/social activity and interactions) monitoring has a significant impact on their treatment plan and recovery.

- Traditional Monitoring Methods
  - Wearable devices
  - Visual observation

- Current Monitoring Practices
  - Visual observation by nurses/caregivers
  - Manual event recording

- These methods are **inaccurate, incomplete, and labor-intensive.**

AI-Powered Solution

- AI offers potential to develop **non-intrusive**, **real-time**, and **scalable** monitoring systems.

- **Computer Vision** is a field of Artificial Intelligence (AI) that enables machines to **interpret, understand, and extract meaningful information from visual inputs** such as images and videos.

- **Generative AI** is a field of AI that can **create new content**, such as text, images, audio, or video, based on patterns learned from existing data.
  - **Large Language Model (LLM)** techniques are a subset of Generative AI focused **on natural language**.

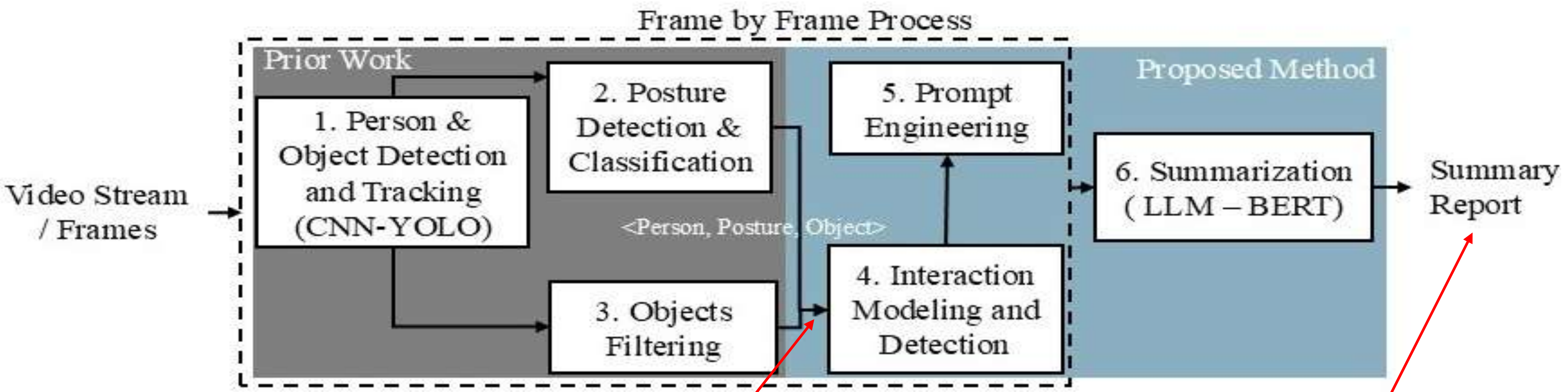
- Computer Vision for object & behavior detection
- LLMs for automatic report generation

Prior Works

- AI-Based Patient Monitoring
  - Computer vision & machine learning to automate monitoring [Z. You, et al., BHI 2021]
  - Reporting general ambulation data (no clinical report) [A. Steele et al., EMBC 2023]

- Our prior work [M. Habibi et al., EMBC 2024]
  - YOLO-based Object Detection
    - Retrained YOLO models for hospital-specific object recognition using 23,704 annotated clinical environment images.
  - Posture Classification with Random Forest
    - Mediapipe landmarks → 33 feature vectors → classification
    - Three main postures: Standing, Sitting, Lying down

Proposed Methodology



- **Inputs** to the proposed method
  - **Patient and object information**
  - **Patient posture**

- **Outputs**
  - **Generate automated patient behavior reports**

Interaction Modeling and Detection

- We create a relational table to ensure **meaningful associations** between the **patient's posture** and **clinically relevant objects**.

Posture	Objects	Posture	Objects	Posture	Objects
Sit	Bed	Stand	Cane	Lie	Bed
	Chair		Walker		Sofa
	Sofa		Crutch		Table
	Table		IV Pole		
	Wheelchair				

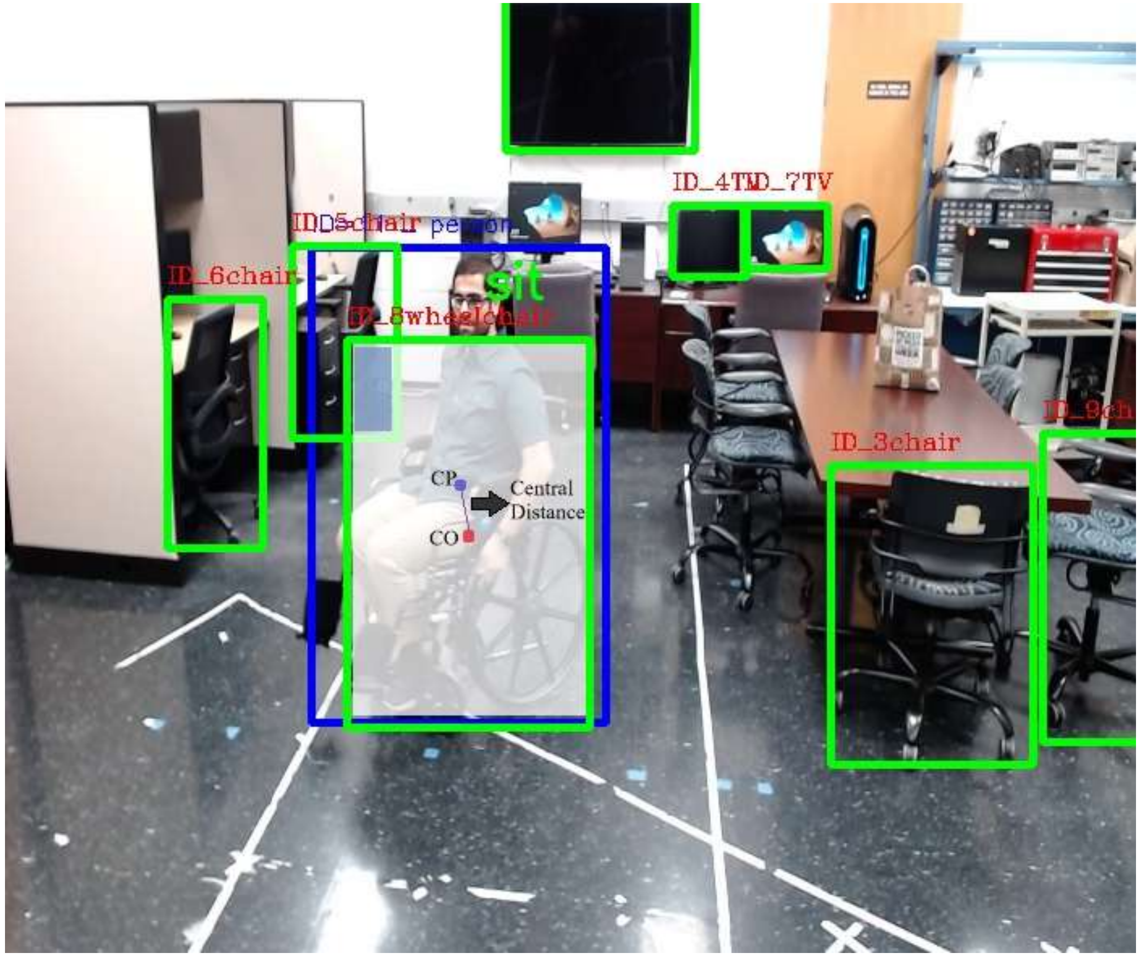
Interaction Modeling and Detection

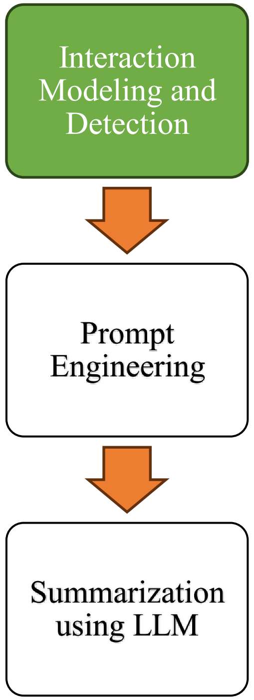
Prompt Engineering

Summarization using LLM

Interaction Modeling and Detection (Cont.)

- **Metrics** used for interaction modeling
  - **Intersection over Union (IoU):** Measures overlap (of bounding boxes) between objects.
  - **Central Distance**: Determines spatial proximity.
  - **Depth Estimation (AdaBins):** Handles 2D ambiguity in object size & distance.

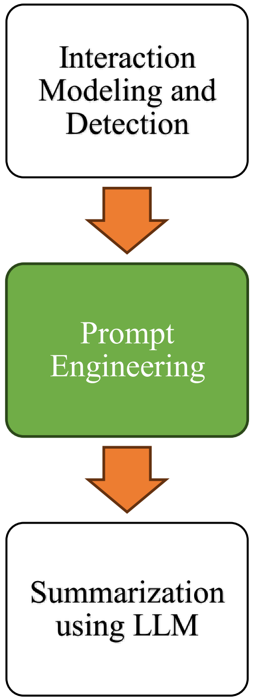




Prompt Engineering

- **Convert** patient-object interactions into **structured text** prompts (sentence or phrase)

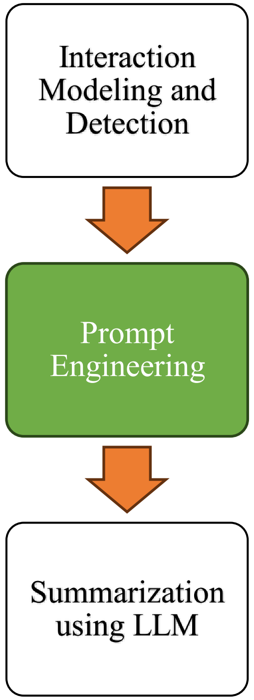
- Apply **real-world constraints** to recognize interactions (based on interaction / distances)
  - **[Subject] + [Posture Verb] + [Preposition] + [Object] + [Duration]**
    - Example: The patient sits on the chair for 10 seconds.
  - **[Subject] + [Posture Verb] + [Location Phrase]**
    - Example: The patient stands near the bed.



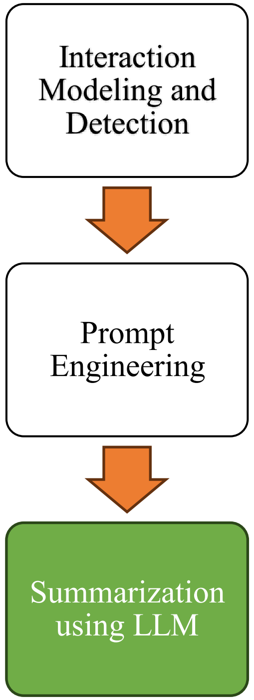
Prompt Engineering (Cont.)

- These **per-frame prompts** act as **building blocks** for summarization and reporting.
- Blocks help to track **Behavioral trends, Unusual patterns,** and **Response to the environment**
- Example of five generated prompts that describe the detected interactions and spatial relationships within the scene.

Frame #	Generated Prompt
13	Person stands close to a table and two chairs and far from TVs.
14	Person stands close to a chair and far from other chair and table and TVs.
15	Person sits in wheelchair and far from table and chairs and TVs.
16	Person stands near wheelchair and far from table and chairs and TVs.
17	Person stands between chair and wheelchair.



Summarization Using LLM



- **BERT model** performs summarization by **understanding the context and meaning** of the text, then identifying and **extracting the most important sentences** or phrases to form a **concise summary**.
  - Leverages **pre-trained** language knowledge to identify important information
  - Works **without any additional training** (zero-shot setting)

- The model **combines** multiple **frame-level prompts** into a **structured summary** that captures the overall activity within a specific time window or full session.

Summarization Using LLM (Cont.)

- Generated Summary for a Sample Video

> *A person is present in a room containing several chairs, TVs, a table, and a wheelchair. The person stands near the table for* ***three seconds****. Then, the person sits in the wheelchair for* ***ten seconds*** *and moves it. Afterward, the person stands between a chair and the wheelchair, positioned far from the table, for* ***five seconds****.*

- Transformed informative **summary report assists** clinician's and caregivers to
  - **Quickly understand** patient's **key activities & ambulation**
  - **Notice** the important **behaviors, anomalies, or routine interactions**
  - **See clear messages** as opposed to long sequences of raw prompts

Interaction Modeling and Detection

Prompt Engineering

Summarization using LLM

Experimental Results – Setup & Data Collection

- **Computational Platform**
  - Laptop with 32GB of RAM, an Intel i9-12900HK CPU, and an NVIDIA GeForce RTX 3050 GPU
  - Customized MediaPipe using Python, and TensorFlow Lite for efficient model execution
  - A regular webcam Logitech C920S camera

- **Eight Subjects**
  - Institutional Review Board IRB-24-776 was approved by University of Texas at Dallas.
  - Each volunteer performing three actions (Sitting, Standing, Lying down) in controlled setting.

- **Videos Collected**
  - Eight recorded video segments (each approximately 20 minutes) covering 24 scenarios.

Experimental Results – HOI Detection Accuracy

- Accuracy for different postures

Postures	# of Test Frames	# of Correct Detections	Accuracy(%)
Siting	2174	1783	85
Laying	1947	1938	89
Standing	1870	1851	94

- Findings
  - Standing detection (highest) – clear posture distinction
  - Lying detection – slightly lower due to possible occlusions
  - Sitting detection (lowest) – more challenging due to environmental variations and occlusions

Conclusions & Future Work

- Our AI-based patient monitoring system enables
  - **Efficiency**: Non-intrusive for patients and less workload for caregivers
  - **Accuracy**: Automated and with continuous patient/objective tracking
  - **Scalability**: Deployable across hospitals, rehabilitation, and elderly care centers

- Future Works
  - Address occlusion & lighting challenges
  - Fine-tune BERT for medical-specific summarization
  - Validate the output summaries in real-world clinical environment

Thanks for Your Attention!

Mostafa Habibi, Mehrdad Nourani
Department of ECE
The University of Texas at Dallas
{mostafa.habibidehsheikhi, nourani}@utdallas.edu