Developing a Novel Approach to Video-Based Fall Risk Assessment in Home Healthcare Using Multimodal Large Language Models: A Pilot Study

Pallavi Gupta, PhD, MS
Postdoctoral Research Scientist
Columbia University School of Nursing

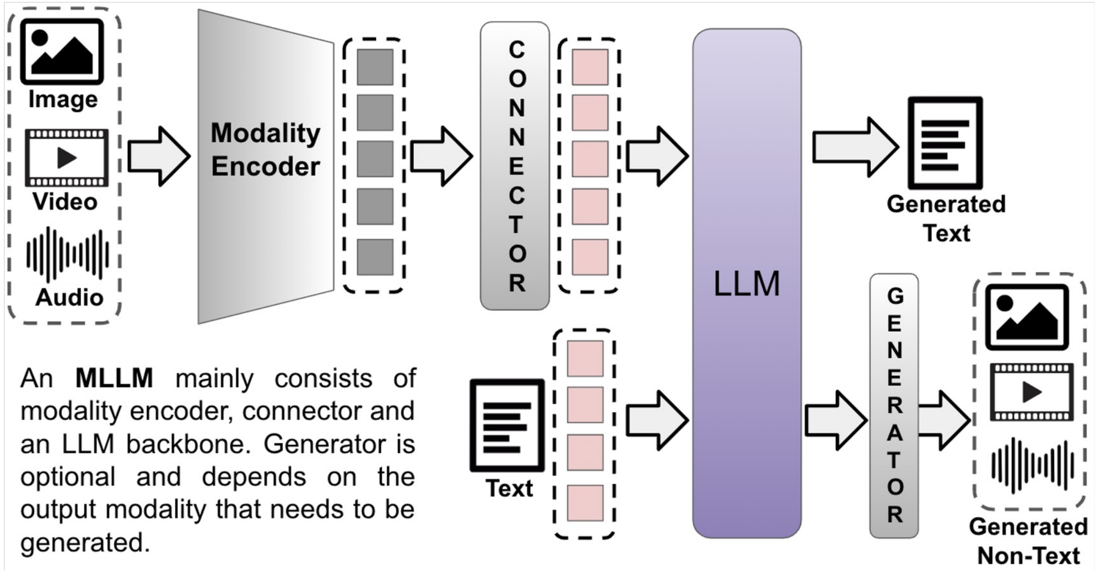
H3IT Conference
New Orleans, Louisiana
November 1, 2025

COLUMBIA
SCHOOL OF NURSING

Evolution: From LLMs to Multimodal LLMs

Multimodal Large Language Models (MLLMs)

- MLLMs are extensions of LLMs.
- Incorporate LLM as a core component.
- LLM backbone acts as the "brain" or central reasoning engine of the MLLM, processes embeddings and generates text-based representations of the response[1].



An **MLLM** mainly consists of modality encoder, connector and an LLM backbone. Generator is optional and depends on the output modality that needs to be generated.

COLUMBIA
SCHOOL OF NURSING

Multimodals for audio-visual understanding.

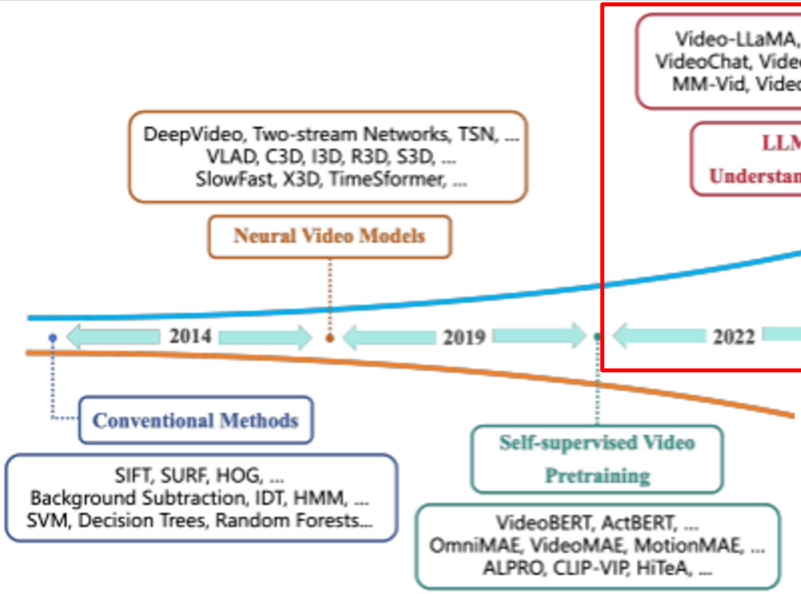


Figure 1: The development of video understanding methods can be sum
(1) Conventional Methods, (2) Neural Video Models, (3) Self-supervised V
Large Language Models for Video Understanding, i.e., Vid-LLMs. Their i
continuously improving, and they possess the potential for further enhance

Selection Criteria:

- What is our input and output modalities?
- What suits best for our success metric and our data?

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE

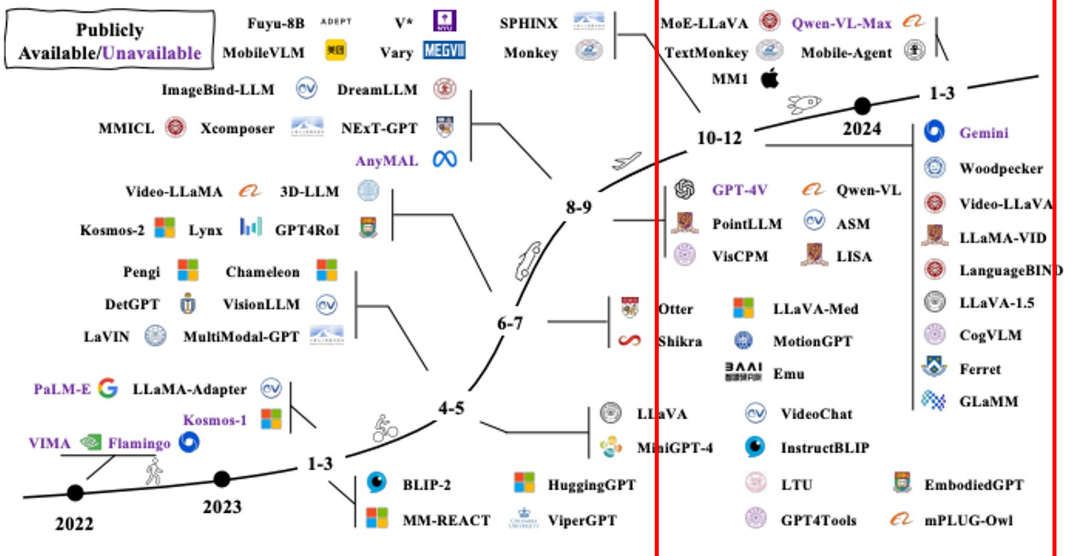


Fig. 1: A timeline of representative MLLMs. We are witnessing rapid growth in this field. More works can be found in our released GitHub page, which is updated daily.

Multimodal Large Language Models

From Data Bottlenecks to Plug-and-Play AI

COLUMBIA
SCHOOL OF NURSING

- **Web-scale training corpora** → massive, diverse datasets
- **Resilience** → robust to noise; handles messy, real-world inputs
- **Speed** → no data-scarcity bottlenecks; skip bespoke collection and long training
- **Access** → off-the-shelf models for immediate inference

Ideation happened

We can use MLLMs to pull useful details from home-care videos. More Useful Information!!!!

Initiated feasibility testing of multimodal LLMs for fall-risk detection in home healthcare.

Video Based Fall Risk Detection

- **Falls** - major driver of injury, hospitalization, mortality.
- **Magnitude - ~37.3M** medically attended falls/year globally (2021); **$50B+** U.S. costs (2015 est); ~**646k** deaths/year (2021)[4,5,6,7].
- **Home health gap -** Traditional tools & wearables don't capture the **dynamic, multifactorial** mix of:
  - **Intrinsic** (patient factors)
  - **Extrinsic** (home environment)
  - **Behavioral** (daily activities)
- Emerging signals (smart-home ambient sensors) are promising but **costly** and hard to **integrate** into workflow.

Method

Simulated Video Data Mimicking Home Healthcare Nurse Visits

Processing Video Data (Visual Frame)

Identify Fall Risk Factors

Standardized Prompts

LLaVA-NeXT-Video-7B-hf (Multimodal LLM)

Fall Risk Factors Detected

Method

COLUMBIA
SCHOOL OF NURSING

Simulated Video Data
Mimicking Home Healthcare
Nurse Visits

COLUMBIA
SCHOOL OF NURSING

Simulated Video Data Generation

Collaborators at University of Minnesota

Dr. Martin Michalowski, PhD, FAMIA, FIAHSI
School of Nursing Foundation Research Professor
Medical Informatics | School of Nursing
University of Minnesota

Meijia Song
PhD Student
School of Nursing
University of Minnesota

COLUMBIA
SCHOOL OF NURSING

Simulated Video Data Generation

Development of Simulation Transcripts

- **Replicate realistic patient**–provider interactions.
- Follow clinical guidelines, safety protocols, and documentation standards.

Scenario Preparation & Environmental Setup

- **Simulated home healthcare environment** created for realism.
- **Props included:** food items, charging cables, trash cans, mobility aids.

Quality Assurance Procedures

- **Rehearsals** and **supply checklists** ensured scenario readiness.
- **Equipment tests** performed before recording.
- **Real-time monitoring** and **post-recording reviews** validated data quality.

COLUMBIA
SCHOOL OF NURSING

Examples of Video Screenshots

Figure: Sequential Frames Extracted from Video to Showcase Scenarios. The patient shows the following: (a) sitting in a wheelchair; (b) woman wearing glasses; (c and d) exercising; (e) sits back again in the wheelchair.

Method

Simulated Video Data Mimicking Home Healthcare Nurse Visits

Identify Fall Risk Factors

Risk Factors Extraction From Literature

COLUMBIA
SCHOOL OF NURSING

Goal: Capture the interplay among **intrinsic**, **extrinsic (environmental)**, and **behavioral** factors to estimate fall risk.

Intrinsic (Personal Factors)	Extrinsic (Environmental Factors)	Behavioral Factors
- Age - Gender - Poor vision - Impaired mobility - Difficulty walking on uneven surfaces - Difficulty negotiating obstacles	- Slippery floor - Unsafe floor mat - Cluttered living space - Wearing inappropriate footwear - At increased risk for environmental injury	- Doing activities like walking or exercising on a cluttered/uncluttered floor

Method

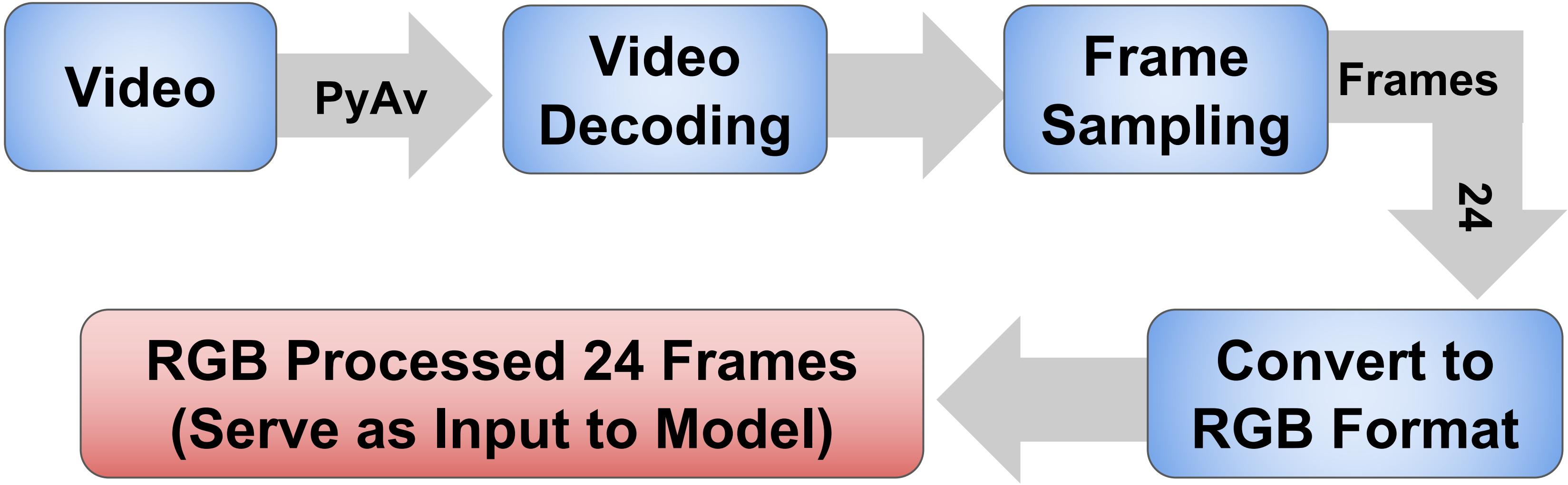
COLUMBIA
SCHOOL OF NURSING

Simulated Video Data Mimicking Home Healthcare Nurse Visits

Processing Video Data (Visual Frame)

Identify Fall Risk Factors

Video Preprocessing



Method

Simulated Video Data Mimicking Home Healthcare Nurse Visits

Processing Video Data (Visual Frame)

Identify Fall Risk Factors

Standardized Prompts

Iterative Prompting

- Prompts were developed to ask for the presence of each risk factor.
- Two-tier prompting:
  - Initial Evaluation with Direct Prompts
    - Concise prompts for simple cues
      - e.g. cluttered floor, uneven surfaces etc
  - Refinement with Elaborated Prompts
    - Elaborated prompts for complex reasoning
      - e.g. Shoes are appropriate (how do we define appropriate?)
- Benchmark prompt set finalized.

Method

Simulated Video Data Mimicking Home Healthcare Nurse Visits

Processing Video Data (Visual Frame)

Identify Fall Risk Factors

Standardized Prompts

LLaVA-NeXT-Video-7B-hf (Multimodal LLM)

Multimodal LLM Used

Model Used: LLaVA-NeXT-Video-7B-hf (Multimodal LLM)

Why did we choose this model?

- **Local, privacy-first:** Chose **LLaVA-NeXT-Video-7B-hf**, run **on-prem from Hugging Face** (no API/data egress), supporting HIPAA-aligned workflows.
- **Open & reliable: Open-source**, actively maintained, with clear docs—enabling inspection, reproducibility, and stable deployment.
- **Right-sized compute: 7B parameters** model require optimal memory limits while delivering competitive accuracy; larger models raised computational cost with limited benefit.

Input to MLLM

- **Multimodal setup: 24 sampled frames + structured prompts** to capture intrinsic/extrinsic/behavioral risks; outputs checked against HHC expert ratings. Performed well.
- **Robust inference: 3 runs per prompt + majority vote** and **iterative prompt engineering** → standardized benchmark prompts for consistent, reproducible assessments.

Results: Evaluation vs. Human Experts

Fall Risk Factors	Standardized Prompts	MLLM Results	Expert Evaluation
Category 1: Direct Prompts for Easy Inferences			
1. Age	Based on visible characteristics in the video, does the patient appear to be an older adult? Respond 'Yes, 'No,' or 'Insufficient information.'	Yes	No
2. Gender	Based on visible features in the video, does the patient appear to be a female or a male? Answer 'male' or 'female.'	Female	Female
3. Difficulty walking	Based on the video, does the patient display any visible signs of difficulty walking that could increase their risk of falling? Respond 'Yes, 'No,' or 'Insufficient information.'	Yes	Yes
4. Impaired Mobility	Based on the video, does the patient show any visible signs of impaired mobility that could increase their risk of falling? Respond 'Yes, 'No,' or 'Insufficient information.'	Yes	Yes
5. Uneven Surfaces	Based on the video, is there any visible indication that the floor is uneven in a way that could increase the risk of falling? Respond 'Yes, 'No,' or 'Insufficient information.'	Yes	Yes
6. At increased risk for environmental injury	Based on the video, is the patient at increased risk of environmental injury that could raise their risk of falling? Respond 'Yes, 'No,' or 'Insufficient information.'	Yes	Yes

Context lifts performance: precise criteria (e.g., shoe traction/support/stability) → correct inference vs vague “inappropriate shoes.”

COLUMBIA
SCHOOL OF NURSING

7. Cluttered Living Space	Based on the video, does the space appear visually cluttered in a way that could increase the patient's risk of falling? Respond ’Yes, ’No,’ or ’Insufficient information.’	Yes	Yes
Category 2: Elaborated prompts requiring deeper reasoning			
8. Wearing appropriate footwear	Analyze the video and determine what kind of shoes the patient is wearing. Assess whether the shoes provide adequate traction, support, or stability. Respond ’Yes, ’No,’ or ’Insufficient information.’	Yes	Yes
9. Exercise (Indoor home)	Analyze the video and determine if the person is exercising on the floor. For floor-based exercises, look for actions such as lying, sitting, kneeling, or placing hands/feet on the floor, and identify movements like push-ups, planks, yoga poses, or stretches. Observe upright movements such as squats, lunges, or balance drills for standing exercises. Note any use of mats, equipment, or visible effort to maintain stability during the exercise. Respond ’Yes, ’No,’ or ’Insufficient information.’	Yes	Yes

Fails without evidence or specialized knowledge

Category 3: Insufficient Information			
10. Poor Vision(With Direct Prompt)	Based on observable signs in the video, does the patient appear to have impaired vision? Respond ’Yes, ’No,’ or ’Insufficient information.’	Yes	Insufficient information/ Require Clinical Expertise
11. Poor Vision (With Elaborated Prompt)	Observe the person in the video and assess their visual ability based on their movements. Look for signs of impaired vision, such as hesitation while walking, difficulty detecting obstacles, excessive squinting, or frequent misjudgment of distances. Describe signs and estimate the likelihood of visual impairment if signs are present. If no such behavior is observed, state that the person appears to have normal vision. Respond ’Yes, ’No,’ or ’Insufficient information.’	Yes	Insufficient Information/ Require Clinical Expertise
12. Slippery Floor (With Direct Prompt)	Based on the video, does the floor appear slippery in a way that could increase the risk of falling? Respond ’Yes, ’No,’ or ’Insufficient information.’	Yes	Insufficient information
13. Slippery Floor (With Elaborated Prompt)	Observe the person’s walking pattern and body movements to determine if the floor is slippery. Look for signs such as sudden slips, loss of balance, hesitant or cautious walking, quick posture adjustments, or near-falls. If any of these behaviors are present, describe them and assess whether they indicate a slippery surface. If no such signs are observed, state that the floor appears stable. Respond ’Yes, ’No,’ or ’Insufficient information.’	Yes	Insufficient information

Results: Key Takeaways

- **Works well when cues are obvious**: wheelchair/crutches → *mobility impairment*; clutter/rugs → *environmental hazard*.
- **Context lifts performance**: precise criteria (e.g., shoe traction/support/stability) → correct inference vs vague “inappropriate shoes.”
- **Fails without evidence or specialized knowledge**: *age* mis-ID; *poor vision* inferred from glasses; *slippery floor* with no visual proof.

Prompt specificity + visible cues ⇒ high accuracy;

Absent cues/clinical nuance ⇒ errors.

Conclusion

- **Feasibility shown:** A compact, locally run MLLM (LLaVA-NeXT-Video-7B-hf) can assess fall risk from in-home video frames in HHC settings.
- **Privacy-preserving and practical:** Open-source, on-prem inference supports HIPAA-aligned workflows and low-resource deployment.
- **Limitations:** Fails on clinically nuanced or evidence-sparse items (e.g., age, poor vision, “slippery floor”); tends to answer even with insufficient evidence (hallucination).
- **Viable and cost-effective:** Compact MLLMs can scale video-based fall-risk assessment in HHC.

COLUMBIA
SCHOOL OF NURSING

Future Work

- **Path forward:** Validate on diverse real-world data; expand to temporal reasoning across days and add multimodal inputs (e.g., audio); consider rule-based layers + unified risk score.
- **Clinical impact:** With refinement, this approach can enhance remote monitoring, personalize fall prevention, and improve outcomes for aging populations.
- **Patient context integration:** Pull EHR facts (age, comorbidities, meds) via privacy-preserving pipelines to ground in clinical reality.

Nurse Assist – AI

(Recently Funded By American Nurse Foundation)

- An AI system to support **home-healthcare nurses**.
- Uses **video, audio, and EHR data** to cut documentation time and provide decision support so nurses can make more informed care recommendations.
- **Data collection -** In partnership with **VNS Health**.

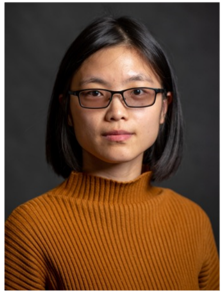
Funding: Selected in **May 2025** as one of three projects in the American Nurses Foundation's **Reimagining Nursing Initiative.**

Acknowledgments

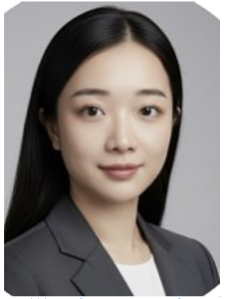
Dr. Maxim Topaz
PhD, MA, RN

Dr. Martin Michalowski PhD, MS
(Collaborator at University of Minnesota)

Dr. Xiao Hu PhD, MS
(Collaborator
Emory University)



Dr. Zhihong Zhang PhD, RN



Meijia Song, BSN

National Institutes of Health

COLUMBIA
SCHOOL OF NURSING

References

1. Gupta P, Zhang Z, Song M, Michalowski M, Hu X, Stiglic G, Topaz M. Rapid review: Growing usage of Multimodal Large Language Models in healthcare. J Biomed Inform. 2025 Sep;169:104875. doi: 10.1016/j.jbi.2025.104875. Epub 2025 Aug 5. PMID: 40754135.
2. Yin, Shukang, Chaoyou Fu, Sirui Zhao, Ke Li, Xing Sun, Tong Xu, and Enhong Chen. "A survey on multimodal large language models." *National Science Review* 11, no. 12 (2024): nwae403.
3. Jin, Yizhang, Jian Li, Yexin Liu, Tianjun Gu, Kai Wu, Zhengkai Jiang, Muyang He et al. "Efficient multimodal large language models: A survey." *arXiv preprint arXiv:2405.10739* (2024).
4. World Health Organization. Falls;. Accessed: 2019-08-02. https://www.who.int/news-room/fact-sheets/ detail/falls.
5. Florence CS, Bergen G, Atherly A, Burns E, Stevens J, Drake C. Medical costs of fatal and nonfatal falls in older adults. Journal of the American Geriatrics Society. 2018;66(4):693-8.
6. Delbaere K, Close JC, Brodaty H, Sachdev P, Lord SR. Determinants of disparities between perceived and physiological risk of falling among elderly people: cohort study. Bmj. 2010;341.
7. Haines TP, Hill AM, Hill KD, McPhail S, Oliver D, Brauer S, et al. Patient education to prevent falls among older hospital inpatients: a randomized controlled trial. Archives of internal medicine. 2011;171(6):516-24.
8. Jehu D, Davis J, Falck R, Bennett K, Tai D, Souza M, et al. Risk factors for recurrent falls in older adults: A systematic review with meta-analysis. Maturitas. 2021;144:23-8.

Thank You

Questions

Extra Slides

Leaderboard for MLLM

https://video-mme.github.io/home_page.html#leaderboard