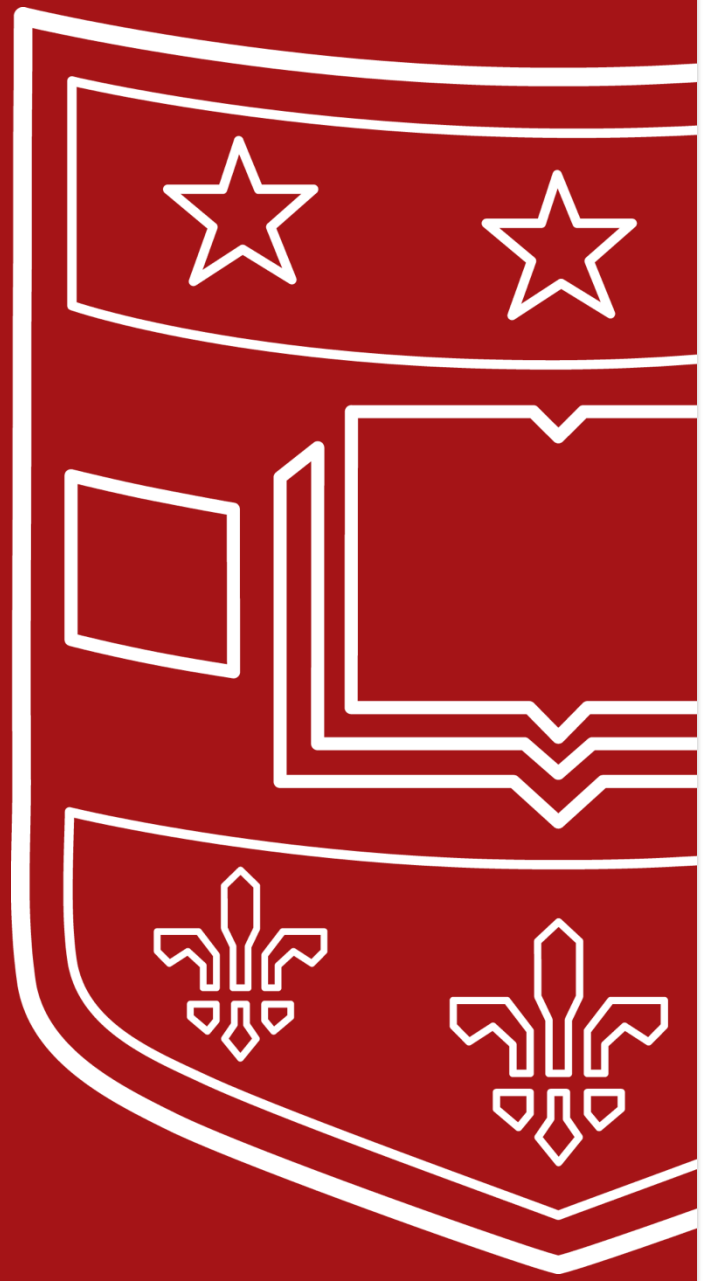
Foundation Models Across Health Data Modalities

Chenyang Lu
Director, AI for Health Institute
Fullgraf Professor
Computer Science · Anesthesiology · Medicine · Neurosurgery · Psychiatry · Public Health

Washington University in St. Louis



Foundation Models Across Health Data Modalities

CLINICAL TEXT
Clinical notes → surgical complications

LONGITUDINAL EHR
Patient histories → early detect CSM

VISION + LANGUAGE
Cooking videos → cognitive errors

Foundation models

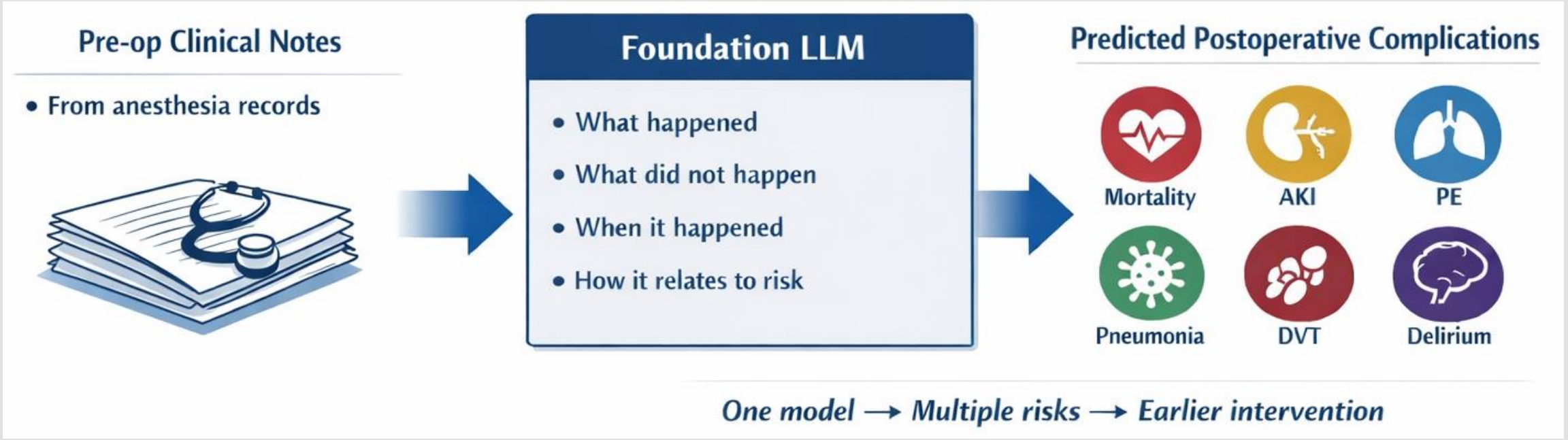
Pretrain on broad data, then adapt representations to a specific health task.

Multimodal health impact

Extract signals that conventional, task-specific models struggle to capture.

Predict Postoperative Risks from Clinical Notes

- Clinical notes capture nuanced clinical judgment that structured data cannot.
- **Large Language Models (LLMs)** translate narrative documentation into risk signals.

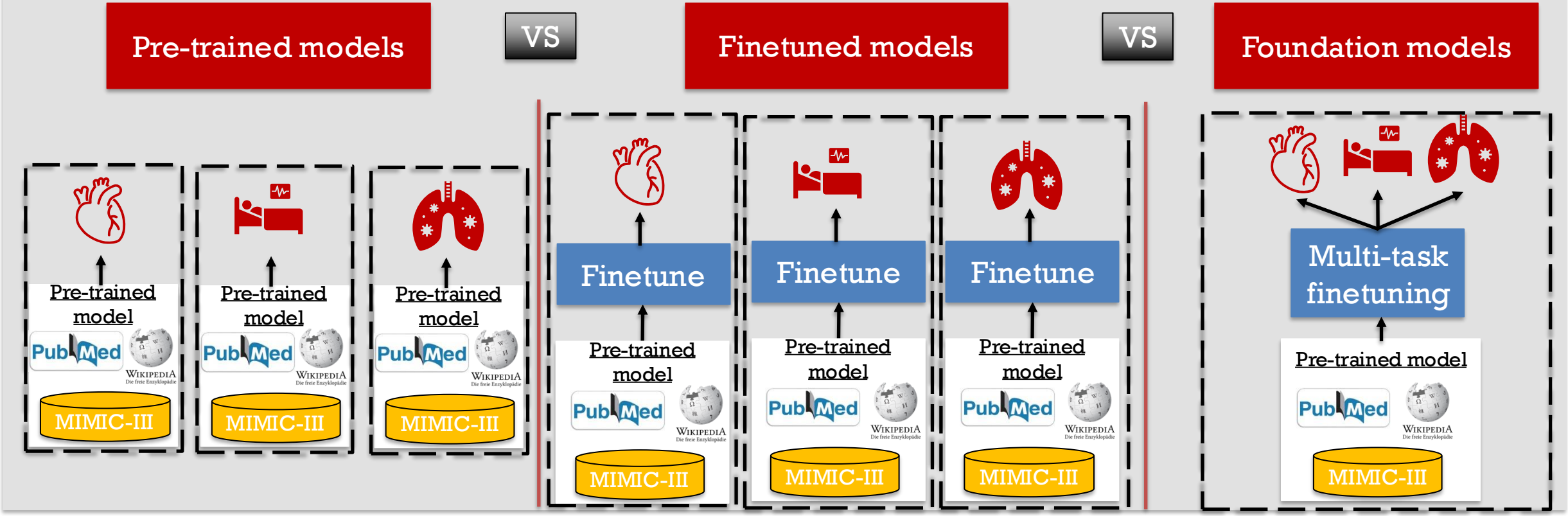


Dataset: 84,875 pre-operative anesthesia notes from Barnes-Jewish Hospital (BJH)

Work with Joanna Abraham, Thomas Kannampallil (Anesthesiology), Charles Alba (DCDS)

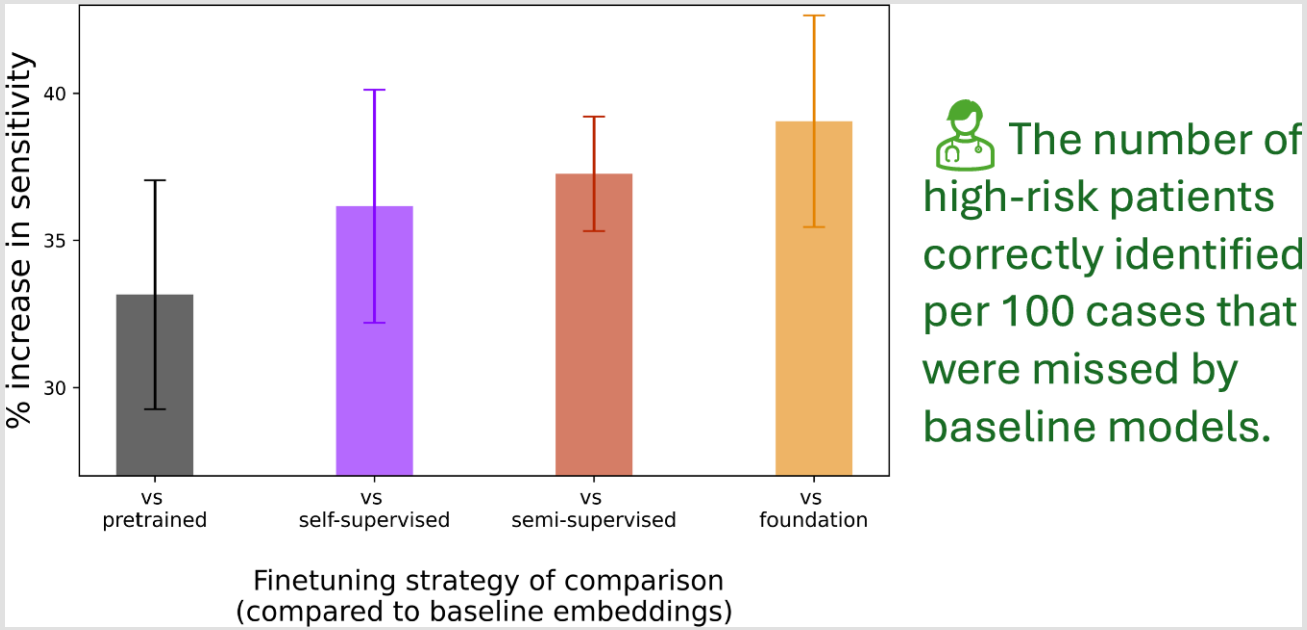
LLMs for Predicting Postoperative Risks

1. Do **pre-trained** LLMs outperform traditional NLP?
2. Does **fine-tuning** to surgical notes improve performance?
3. Can a single **foundation** model predict multiple risks?



Language Models Improve Surgical Risk Prediction

- Pretrained: **+33** percentage point increase over NLP in sensitivity (at 95% specificity)
- **One** foundation model predicts **multiple** complications
  - **Outperforms** individually fine-tuned models
  - **Simplifies** clinical deployment



The number of high-risk patients correctly identified per 100 cases that were missed by baseline models.

SurgicalPLAN

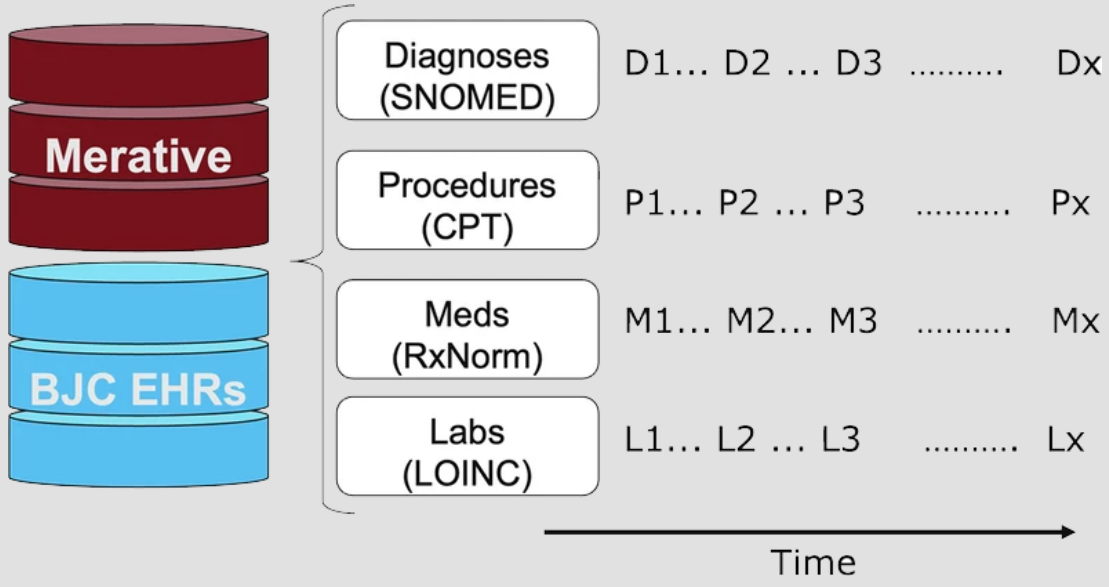
ACS AI Course

Alba C, Xue B, Abraham J, Kannampallil T, Lu C. The foundational capabilities of large language models in predicting postoperative risks using clinical notes. *npj Digital Medicine*. 8, 95 (2025).

Predict Cervical Spondylotic Myelopathy (CSM)

- CSM is the **leading cause** of spinal cord dysfunction in older adults
- Diagnosis is frequently delayed by **2–6 years** after symptom onset
- Goal: Predict CSM using **longitudinal EHR data** before clinical diagnosis

Data
- Merative MarketScan: ~1.4M patients' longitudinal claims data
- BJC Health: ~500K patients' institutional EHRs



Work with Jacob Greenberg, Salim Yakdan (Neurosurgery), Ben Warner (CSE)

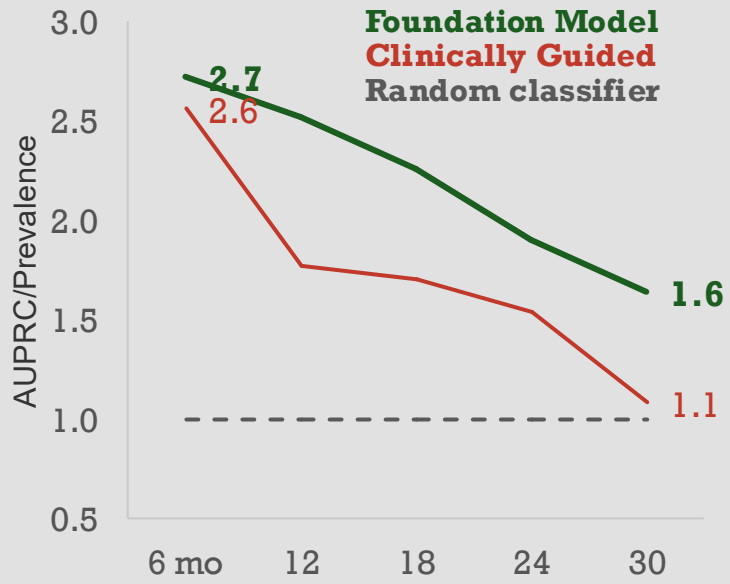
AI-Driven Early Detection of CSM

INTERNAL VALIDATION:
Predictive performance up to 30 months in advance

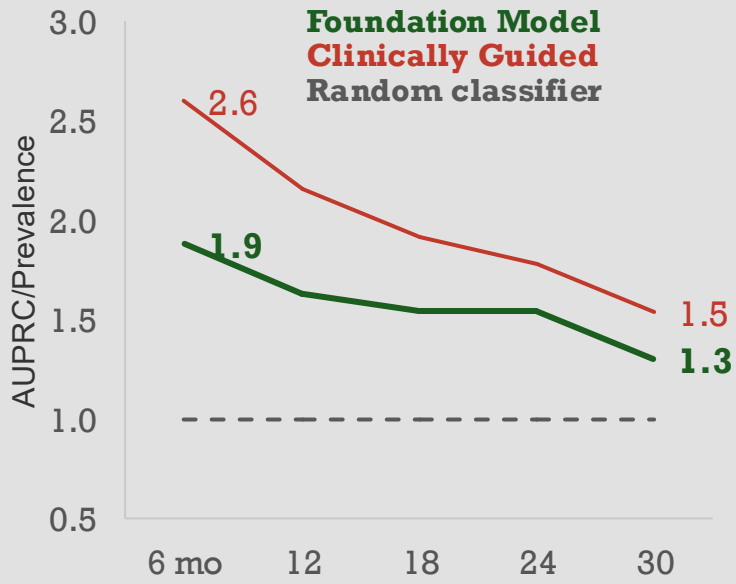
EXTERNAL VALIDATION:
Clinically guided model retains its performance

KEY TAKEAWAY:
Clinical knowledge drives better generalization

Internal Validation (BJC)



External Validation (Merative)



Foundation Model
145M parameters; pretrained on 2.57M Stanford patients; enriched with 5K CSM-associated codes.

Clinically Guided
391K-parameter, 3-layer neural network using counts of 497 clinician-selected codes.

Yakdan S, Warner BC, et al. Clinically-guided models or foundation models? Predicting cervical spondylotic myelopathy from electronic health records. *npj Digital Medicine*. 2026;9:153.

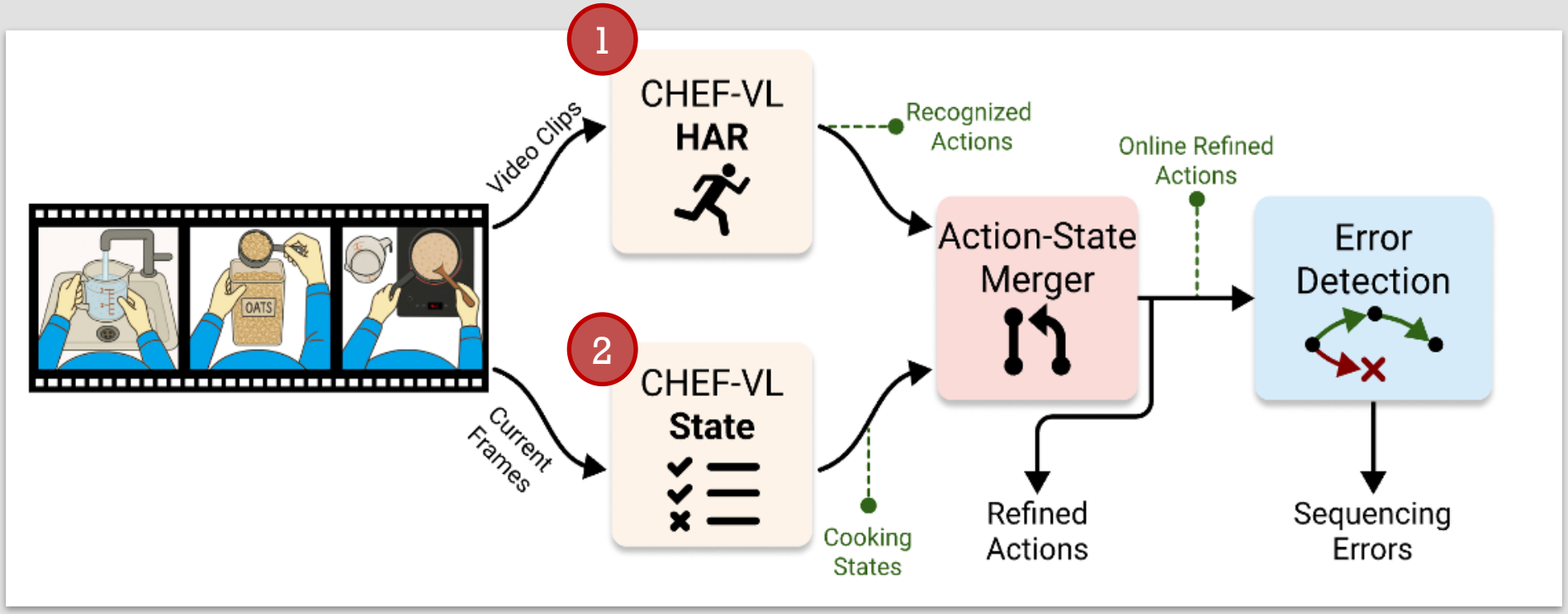
Detecting Cognitive Errors

- Individuals with **Subjective Cognitive Decline (SCD)**
  - Struggle with everyday tasks
  - **1 in 9** people over 45 experience SCD
- **Smart Kitchen**: Detect cognitive errors and support cooking tasks
  - Collected videos of participants cooking oatmeal
  - 107 participants (27: self-reported SCD, 22: mild cognitive impairment).



Joint work with Lisa Connor, Carolyn Baum (Occupational Therapy), Ruiqi Wang (CSE)

Activity Recognition with Vison Language Models



1. **CHEF-VL-HAR** for action recognition
2. **CHEF-VL-State** for state understanding



Performance: Detecting Sequencing Errors

- **Pretrained** VLM provides a strong zero-shot baseline
- **Fine-tuning** drives large gains in error detection accuracy
- Adding **state** information delivers the highest overall performance

Method	Precision (%)	Recall (%)	F1-Score (%)
LSTR	20.42	60.67	19.07
Qwen2.5-VL	33.00	35.00	17.00
+ Fine-tune → Qwen2.5-VL-Finetune	49.83	70.67	44.50
+ State → **CHEF-VL (Ours)**	**53.00**	**71.67**	**48.43**

Wang R, Gao P, Lynch P, Liu T, Lee Y, Baum CM, Connor LT, Lu C. CHEF-VL: Detecting cognitive sequencing errors in cooking with vision-language models. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2025;9(4):220. doi:10.1145/3770714.

Foundation Models Unlock Multimodal Insights

SURGICAL CARE
Anticipate complications from clinical notes

NEUROLOGIC CARE
Detect CSM earlier from longitudinal EHR

COGNITIVE HEALTH
Recognize errors in everyday activities

Why foundation models matter

They learn reusable representations from rich, longitudinal, and contextual health data.

What makes them useful

Task adaptation, clinical grounding, and rigorous validation turn scale into health impact.