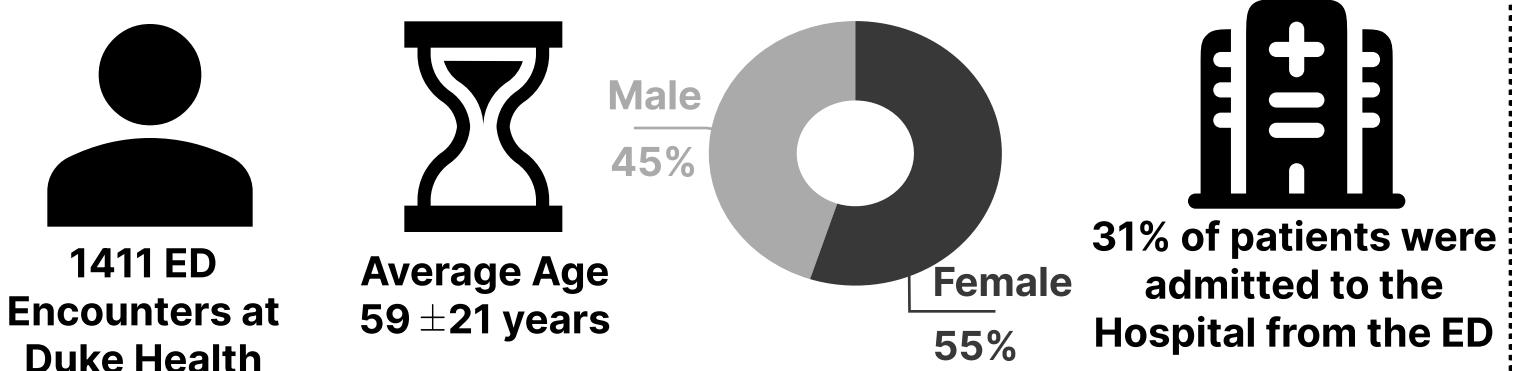


Pranav Manjunath<sup>1</sup>, Syed M. Adil<sup>2</sup>, Benjamin Wissel<sup>2</sup>, Daniel Sexton<sup>2</sup>, Brian Lerner<sup>3</sup>, Timothy W. Dunn<sup>1,3</sup>**Background**

Traumatic Brain Injury (TBI) is a highly heterogeneous condition, with patients presenting with diverse injury mechanisms, severities, and clinical trajectories—most of which begin in the Emergency Department (ED). Given this variability, providing **individualized context is essential** for supporting accurate and timely clinical decision-making. However, clinical guidelines alone are insufficient for predicting personalized, real-world outcomes. We propose augmenting small language models (SLM) with real-world evidence (retrieved from historically similar patients) alongside established clinical guidelines to generate more individualized and clinically grounded predictions of **ED Disposition in TBI care<sup>b</sup>**.

**Data Cohort****Results**  
**ED Disposition Prediction**

Clinical Guidelines alone make the model conservative and over admits patients (Table 1)

| Prompting Strategies   | Phi   |       |       | Qwen  |       |       |
|------------------------|-------|-------|-------|-------|-------|-------|
|                        | F1    | SN    | SP    | F1    | SN    | SP    |
| No Context - Baseline  | 0.586 | 0.728 | 0.656 | 0.386 | 0.311 | 0.865 |
| Guidelines (Top 1)     | 0.544 | 0.821 | 0.455 | 0.527 | 0.957 | 0.238 |
| Guidelines (Top 2)     | 0.525 | 0.901 | 0.293 | 0.539 | 0.930 | 0.310 |
| Guidelines (Top 3)     | 0.512 | 0.948 | 0.202 | 0.547 | 0.923 | 0.340 |
| Similar Patients (n=3) | 0.585 | 0.782 | 0.594 | 0.600 | 0.553 | 0.868 |
| Similar Patients (n=5) | 0.625 | 0.853 | 0.601 | 0.653 | 0.707 | 0.792 |
| Similar Patients (n=7) | 0.634 | 0.884 | 0.589 | 0.626 | 0.683 | 0.774 |
| Random Patients (n=7)  | 0.540 | 0.887 | 0.365 | 0.565 | 0.644 | 0.711 |
| Dual Retrieval         | 0.626 | 0.698 | 0.758 | 0.563 | 0.912 | 0.399 |

Table 1: ED Disposition Prediction Results (SN: Sensitivity, SP: Specificity)

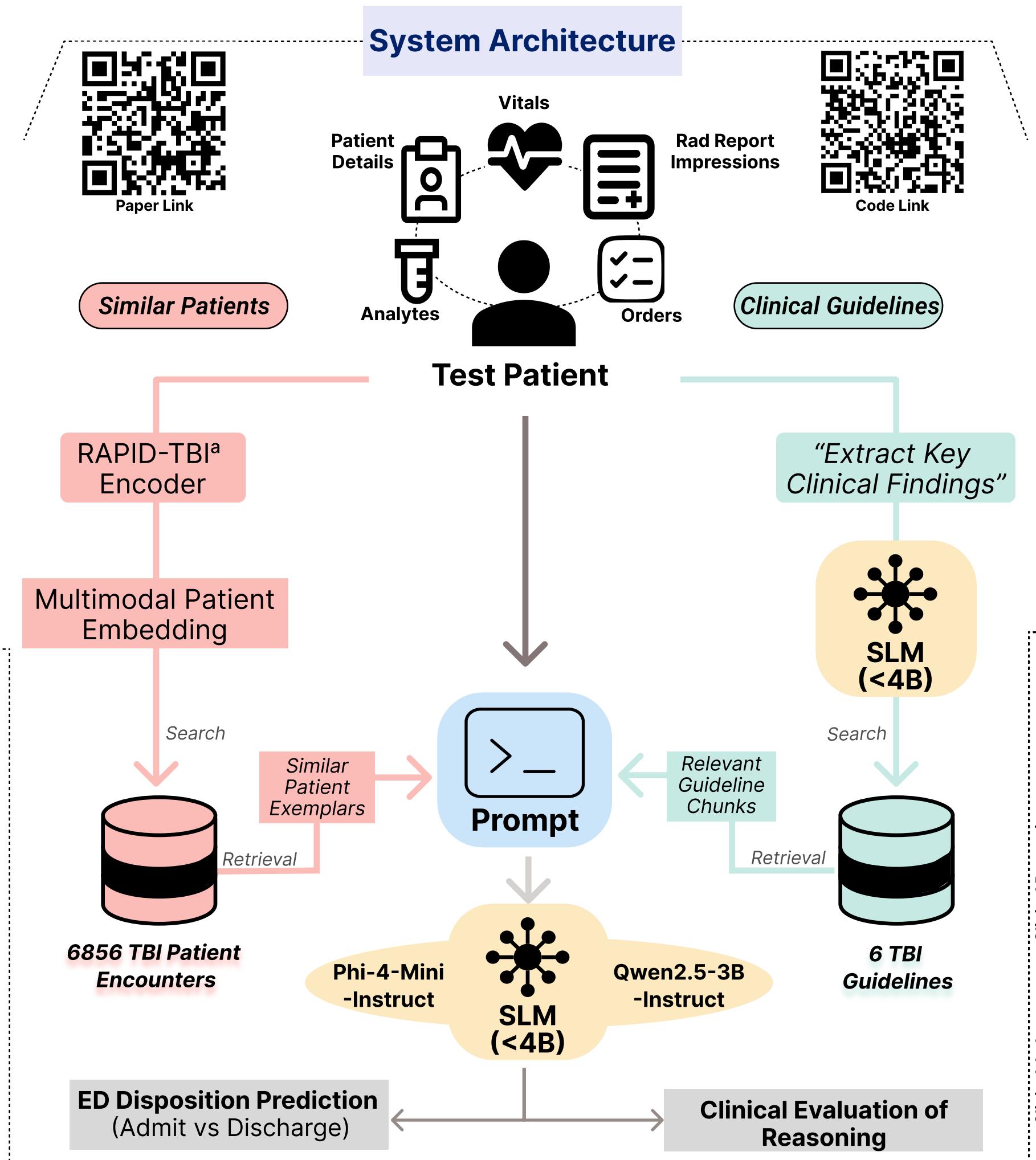
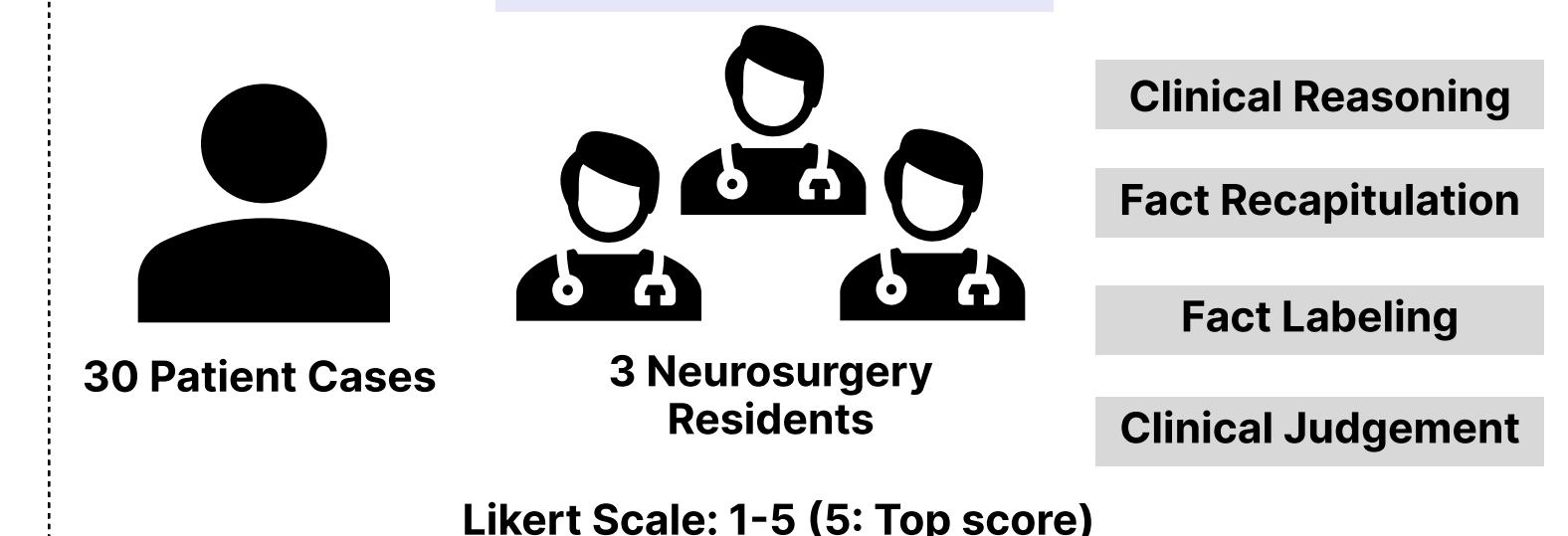
**How does Personalized Context impact the prediction of language models on ED Disposition of patients with TBI?**

Fig 1: System Architecture: Dual Retrieval-based Prediction and Reasoning for Personalized TBI Prognosis

**Key Takeaways**

- Prediction: Language models benefit from real-world evidence (historical similar patients) to predict more personalized and accurate clinical real world outcomes (Table 1).
- Reasoning: Tends to improve when combined with both guidelines and similar patients as context (Fig. 3).
- However, in clinical settings, be cautious in over-relying on these models due to inherent training-data biases.

**Reasoning Evaluation****Results**  
**Reasoning Evaluation**

Reasoning tends to overfit to similar patients, introducing medical hallucinations (Fig. 2)

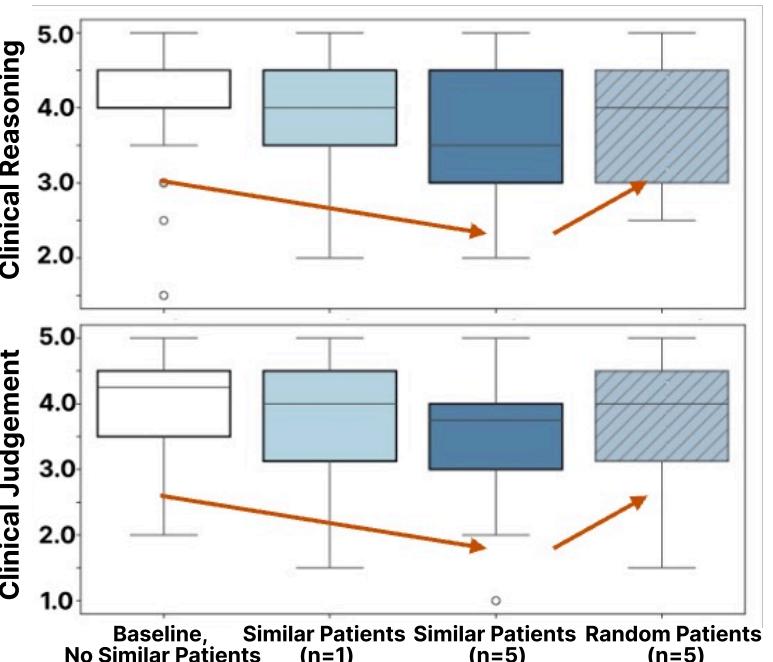


Fig 2: The effect of similar patients on Phi model reasoning

Adding guidelines with similar patients tends to improve reasoning quality\* (Fig. 3)

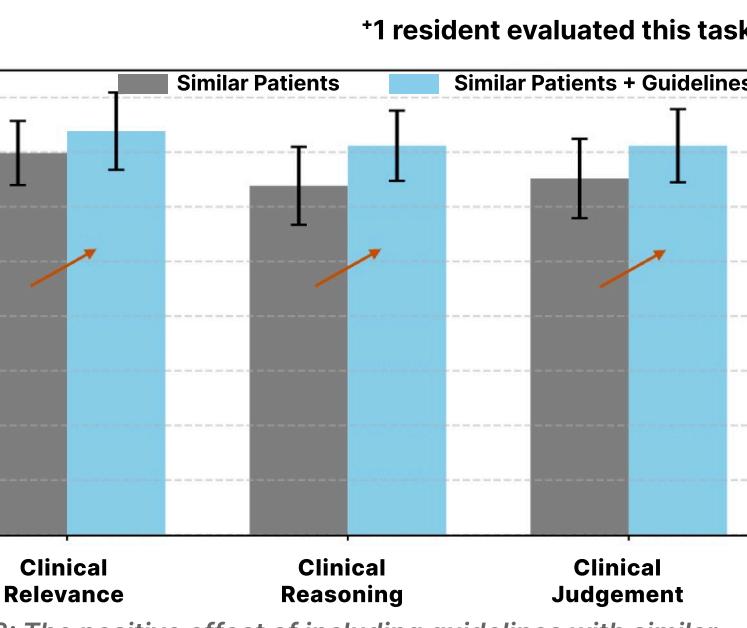


Fig 3: The positive effect of including guidelines with similar patients on Phi model reasoning