



# Large Language Model–Assisted Generation of Patient-Facing After-Visit Summaries at Hospital Discharge

UC San Diego Health  
Hospital Medicine

Milla Kviatkovsky, DO, MPH, Caden Stewart, BS; Annalise McDonald, MD; Yan Xing, MD; Sara Gianella, MD; Anna Gelberg, MD; Bryan Huang, MD; Shadi Mayasy, MD; Kristina Djekic, MD; Maryann Ally, MD; Ian Jenkins, MD, Benjamin Smarr, PhD; Karandeep Singh, MD; Robert El-Kareh, MD

## Background and Problem

**1/2**  
Clinician time spend on EHR tasks

**50%**  
Patients misunderstand critical hospitalization info



- 1) After-visit summaries (AVS) are frequently incomplete, overly technical, or written well beyond the health literacy level of the average patient.
- 2) Clinicians face a dual burden: accurate documentation for providers and comprehensible instructions for patients.

## Solution

Standard Workflow

LLM Supported Workflow

A structured human-oversight model ensures LLM-generated AVSs are clinically reviewed before reaching the patient

Complex Discharge

Complex Discharge

Write Hospital Course

Write Hospital Course

Prompt:  
patient-centered language  
· 6th-grade reading level  
· avoid jargon  
· individualized recommendations

Manually Translate to AVS  
• Average time: 10-15min  
• AVS often above reading level

Prompt LLM To draft into AVS  
• Average time: 2-3 min  
• Plain-language AVS at 6<sup>th</sup> grade reading level

## Methods

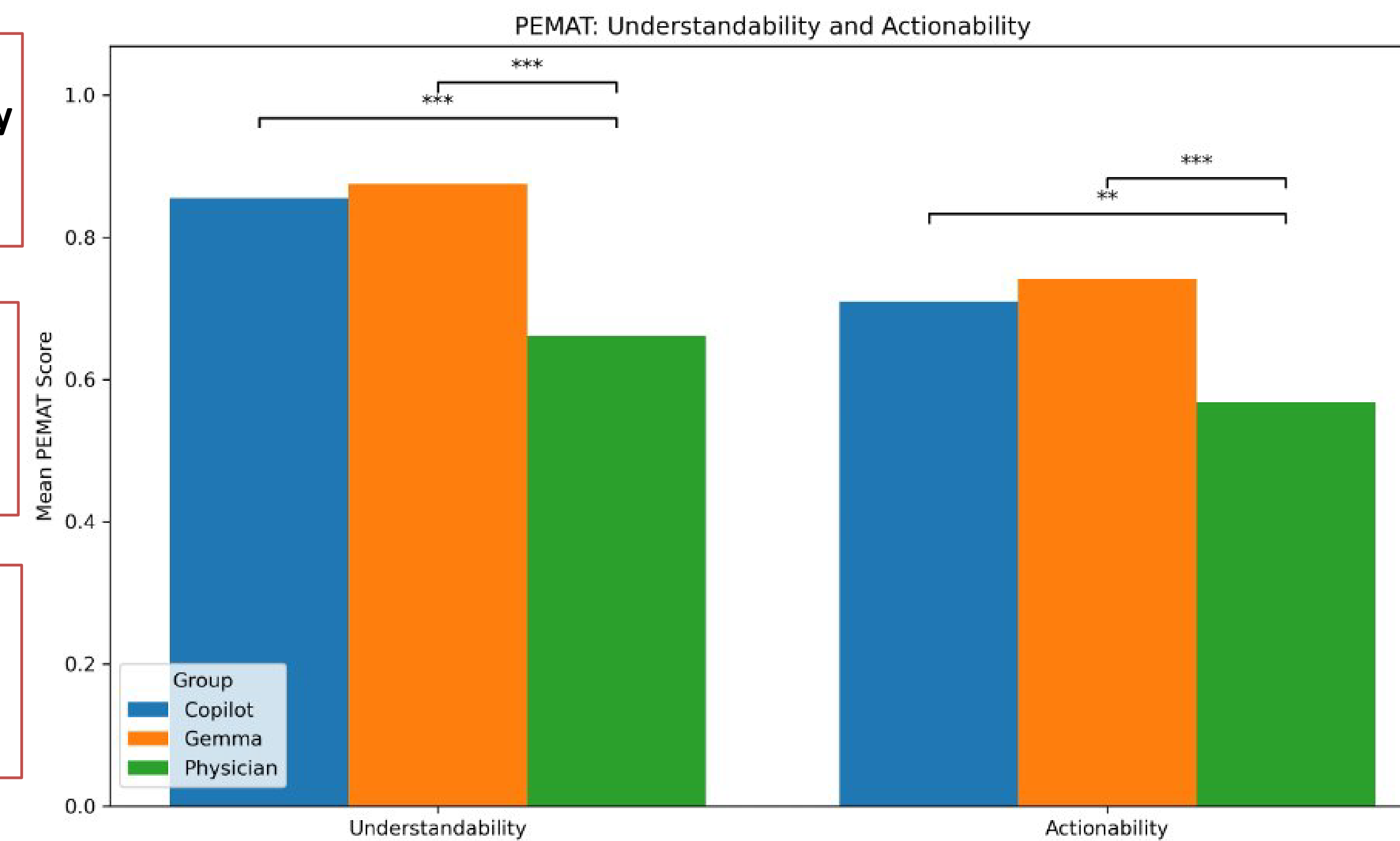
<b>Design</b>	Retrospective, blinded, cross-sectional · IRB exempt
<b>Population</b>	50 adults discharged home (Jan–Dec 2023)
<b>AVS Types</b>	3 per patient: Clinician-authored · MS Copilot (GPT-4) · Gemma 3n 2B
<b>Reviewers</b>	5 blinded attending physicians · 10 charts each · 3 AVSs/chart
<b>Instruments</b>	PEMAT v2.0 (understandability & actionability) + AVS Rubric (5 domains + harm)

## Results PEMAT

**+21%:**  
Understandability  
Gemma vs.  
Physician

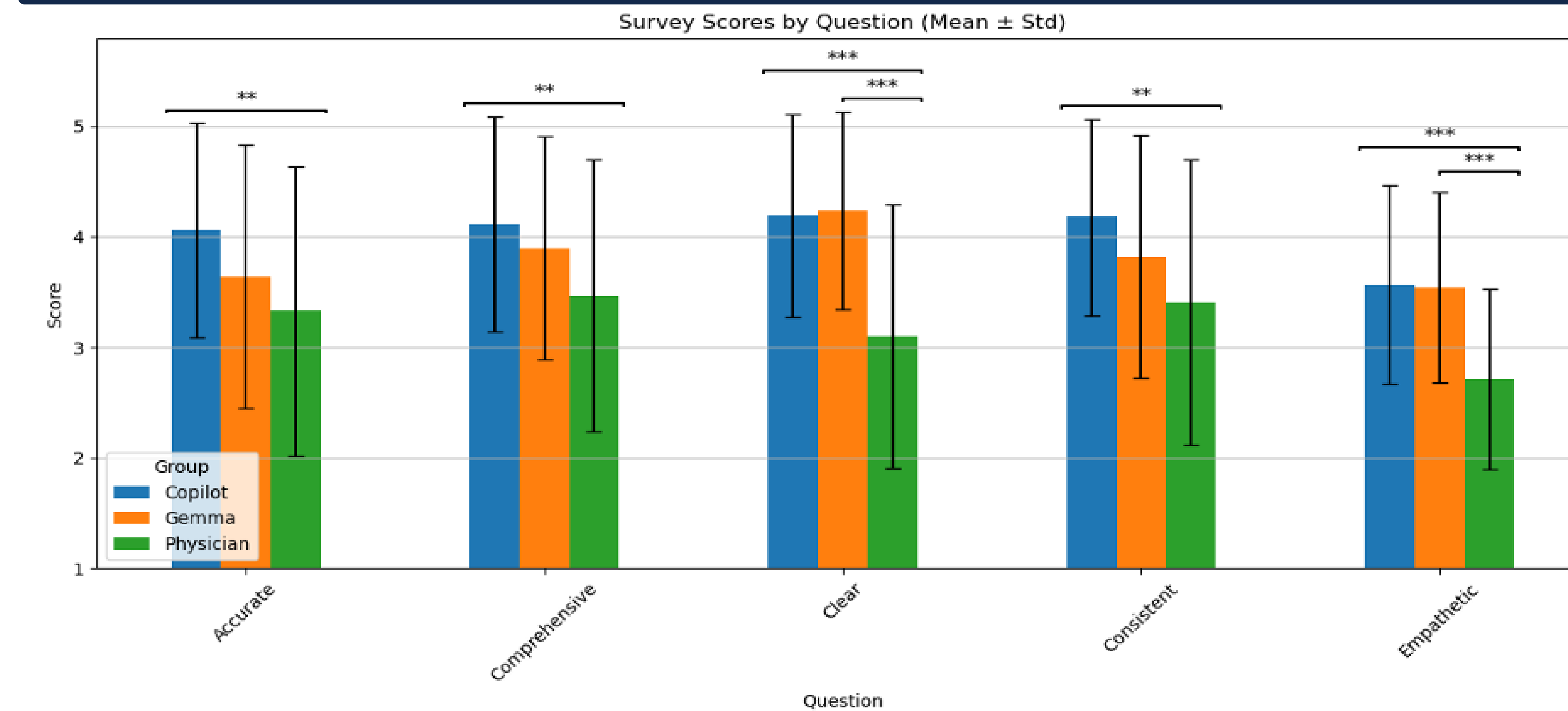
**+18%:**  
Actionability  
Gemma vs.  
Physician

**P<0.001 both models**



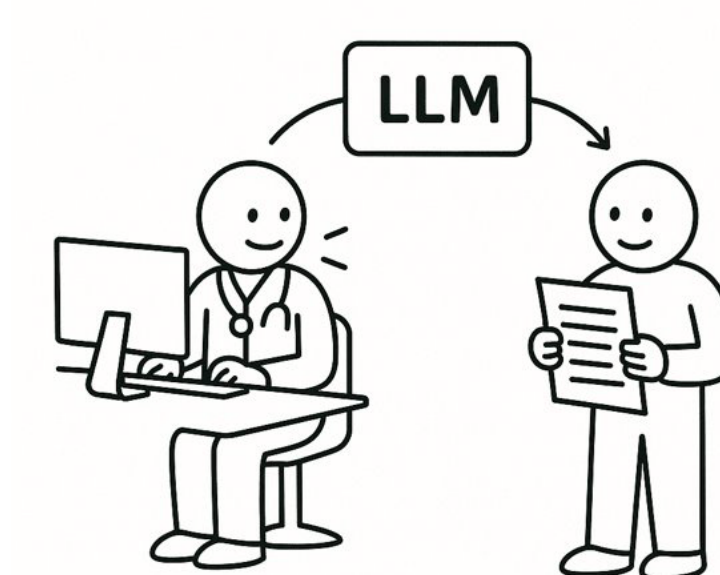
Item Level Analysis: Largest gains seen in Clarity of Purpose, Simplicity of Language, Proper Definition of Medical Terms. Scores exceeded physicians by 25-35%

## Results AVS Rubric

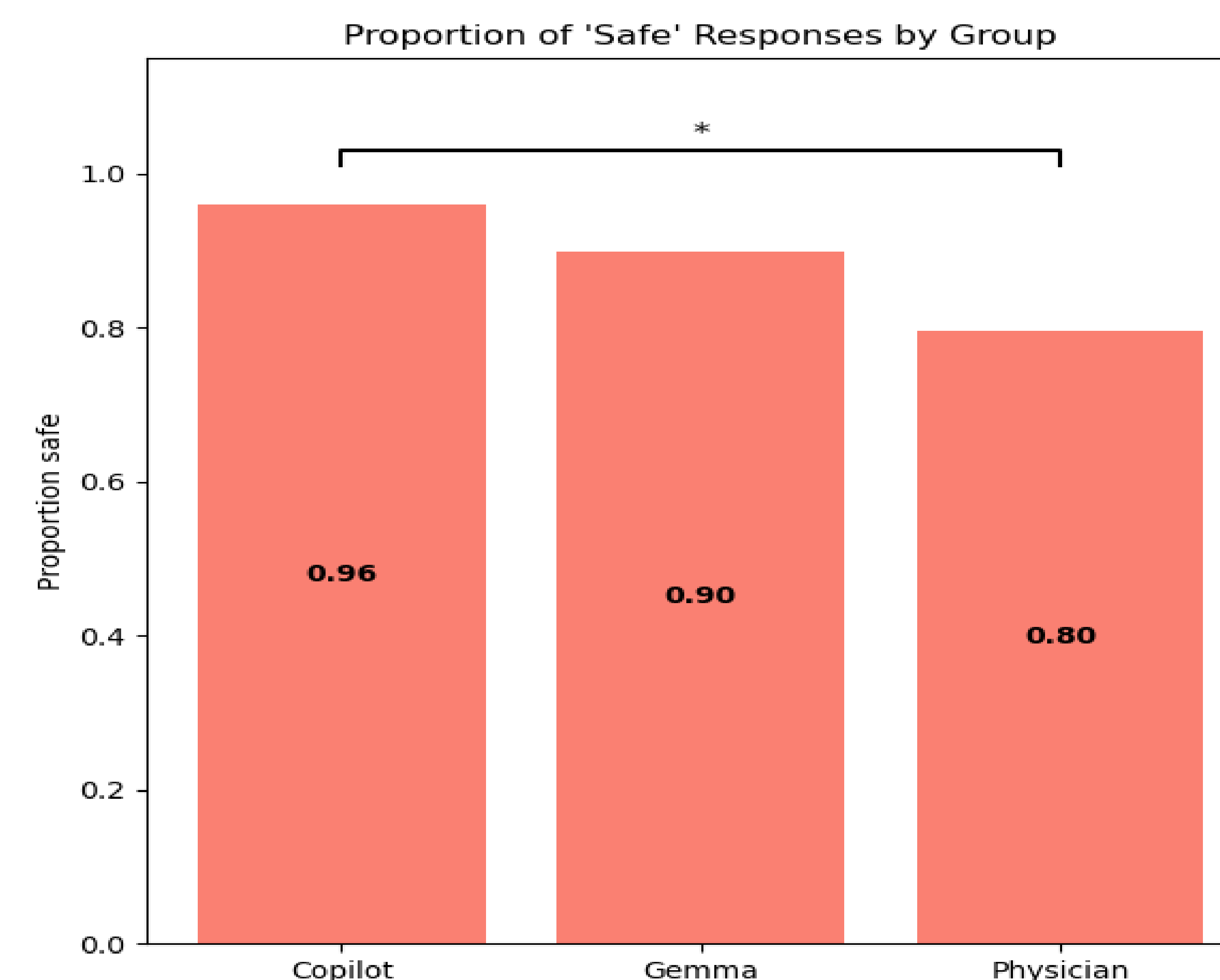


Both LLMs significantly outperformed physician-authored AVSs across all 5 domains (p < 0.001 for clarity & empathy).  
No significant difference between Copilot and Gemma.

## Results AVS Rubric for Safety: No Increased Risk of Harm



Copilot: 96% safe  
Gemma: 90% safe  
Physician: 80% safe  
(p = 0.02, Copilot vs. Physician)



## Conclusions and Next Steps

- 1) LLM-generated AVSs are significantly more readable, actionable, and patient-centered than physician-authored AVSs — across both PEMAT and AVS Rubric, as graded by blinded attending physicians.
- 2) Blinded physician graders did not find LLM-generated discharge instructions to carry increased risk of harm. Safety concerns are not a barrier to implementation in a physician-in-the-loop model.
- 3) GPT-4 and open-source Gemma perform comparably — offering flexible, accessible implementation paths without requiring proprietary infrastructure.

### Roadmap & Next Steps

- Phase 1 — Retrospective Validation:** Complete. Results presented today.
- Phase 2 — Prospective Pilot:** Select physicians · Train on workflow · Evaluate quality & time savings · Patient perspectives
- Phase 3 — Scale and Expansion:** Model refinement · EHR integration · Real-world outcomes · Governance & bias auditing

Human oversight, governance, auditability, and bias monitoring remain essential across all phases.