**🟥 Slide 1 – Title**

**[20 sec]**

"Hi everyone, we’re Bhaskar and Pallavi, and today we’ll be presenting our work on how users develop trust in AI-generated medical diagnoses — specifically when those diagnoses come from very different kinds of models: a general-purpose LLM like Gemini, and a domain-specific model like BioGPT."

**🟥 Slide 2 – Motivation + RQs**

**[20 sec]**

"Our motivation came from a simple but important question: when AI models disagree, who do users trust — and why?

Generalist models like Gemini explain fluently but may lack precision. Domain-specific models like BioGPT are accurate, but less human-readable. In medicine, where users are often non-experts, this tradeoff really matters.

We focused on three questions:

1. How do users resolve conflicting AI outputs?
2. How do explanation style and confidence influence trust?
3. How do individual biases and background shape decisions?"

**🟥 Slide 3 – Research Design**

**[30 sec]**

"We designed a between-subjects study with a 2×2 structure: participants received either text or image-based inputs, and either consistent or conflicting diagnoses from the two models.

Each participant saw both outputs, rated trust and confidence, and told us which model they preferred and why.

We collected both quantitative data — like Likert ratings and bias scores — and qualitative feedback to understand the reasoning behind trust."

**🟥 Slide 4 – System Flow**

**[20 sec]**

"This diagram shows how the system worked: from the prompt, to diagnosis generation by both models, to user interaction and rating.

We also randomized model order to test anchoring bias, and included trials where AI confidence scores were explicitly shown to see how certainty affected perception."

**🟥 Slide 5 – Bias Behavior Analysis**

**[30 sec]**

"Let’s start with cognitive bias. We measured self-reported scores for anchoring, automation, and confirmation bias, and compared them with behavioral indicators.

Surprisingly, there was almost no alignment. For example, anchoring bias had a correlation near zero. Users believed they weren’t biased — and they were right, but not for the reasons they thought. Their decisions weren’t driven by bias awareness, but rather by model clarity and perceived explanation quality.

This disconnect shows that self-perceived bias doesn't predict how users actually interact with AI."

**🟥 Slide 6 – Hypothesis Testing**

**[45 sec]**

"We tested four core hypotheses.  
First: users would trust BioGPT more — not supported. Gemini actually had slightly higher trust ratings.  
Second: trust drops during conflict — supported, and significantly so.  
Third: model shown first would be preferred — not supported, no anchoring effect.  
And fourth: confidence is lower in conflicting cases — supported again.

Together, these results show that **model consistency and clarity** influence trust more than domain training or presentation order."

**🟥 Slide 7 – Expertise & Experience**

**[30 sec]**

"We then looked at how background traits influenced trust.  
Medical knowledge consistently increased trust, especially in conflict — users could evaluate both models critically.  
But AI experience had the opposite effect: experienced users trusted **less** when the models disagreed — possibly because they were more attuned to inconsistency."

**🟥 Slide 8 – Sentiment & Thematic Analysis**

**[35 sec]**

"Our open-ended feedback revealed more nuance.  
Using RoBERTa, we found that Gemini's responses were associated with more **positive sentiment** — users described them as 'clear' and 'easy to understand'.  
BioGPT was often described as 'technical' or 'detailed', but harder to follow.  
Thematic coding confirmed this — Gemini’s clarity earned it trust, even when it wasn’t as medically dense."

**🟥 Slide 9 – Takeaways**

**[30 sec]**

We found that users trusted clearer explanations over clinically accurate ones — Gemini often won over BioGPT.

Trust dropped sharply during conflicts, and users weren’t aware of their own biases in how they judged AI.

Interestingly, medical expertise increased trust, but AI experience made users more skeptical.

And tone mattered — human-like, coherent language led to more positive reactions.

Implication? In health AI, how you explain matters just as much as what you explain.

I’ll summarize: BioGPT was trusted more by experts, Gemini was trusted more by everyone else — and clarity consistently won over correctness.

**🟥 Slide 10 – Future Work**

**[15 sec]**

"Next, we’d like to test this with medical professionals, explore interactive explainability tools, and even run debates between models to see how that shapes trust."

**🟥 Slide 11 – Closing**

**[10 sec]**

"Thanks for listening. We’d love to hear your questions — especially about how we might design AI explanations that feel both trustworthy and understandable."