

Post-Survey

- **[Short answer 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100]** Of the questions you have just labeled, what percentage of them have you seen verbatim beforehand (e.g. in your classes, or when preparing for USMLE)?
- **[Short answer 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100]** Of the questions you have just labeled, what percentage of them have you seen verbatim?
- **[Short answer 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100]** Of the questions you have just labeled, what do you think your accuracy (percentage) was in answering the multiple choice questions?
- **[Short answer 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100]** Suppose we were to give the same clinical questions to the current best LLM (e.g. GPT o3, Claude 3.7 Sonnet), what do you think its accuracy would be on the multiple choice questions?
- **[1: Not Confident, 2:Slightly confident, 3: Moderately confident, 4: Very confident, 5: Extremely confident]** How confident were you, on average, in your answers for the relevance scores?
- **[1: Extreme ambiguity, 2: High ambiguity, 3: Moderate ambiguity, 4: Slight ambiguity, 5: No ambiguity at all]** How much ambiguity do you think there is in the true relevance labels?
- **[1: Not at all aligned, 2: Slightly aligned, 3: Moderately aligned, 4: Strongly aligned, 5: Very closely aligned]** To what degree do you think the process of answering multiple choice clinical QAs correlates with real-world day-to-day medical practice?