Exempt Research
SECTION I

1. Title of Protocol:

Leveraging AI for Clinical ReasoningAssessment and Quantitative Approaches to Measurement: Multi-Institutional Implementation of a Novel Large Language Model

2. Responsible Personnel:

A. Principal Investigator (PI):

Rohlfsen, Cory J - Dept Of IM Gen Hospitalist Home - 402-559-7591 - cory.rohlfsen@unmc.edu - alt #: 402-397-0445 - degree: BS - address: MSB 5554 UNMC Midtown (Zip 3331) - phone: 402-836-9598

B. Secondary Investigator (SI):

Sorrick, Alex G - Int Med General Medicine - 402-559-7299 - alexandra.sorrick@unmc.edu - alt #: 402-559-9605 - degree: BS - address: WHM 3042B UNMC Midtown (Zip 6430) - phone: 9-7299

C. Participating Personnel:

D. Lead Coordinator:

E. Coordinator(s):

Are you adding a clinical trial management group?
No

F. Data/Administrative Personnel:

G. Are you a student or house officer?
No

3. Funding Source:
Check all that apply and provide the source of the funding.
♦ Cooperative Group: multi-institutional team (David Liebovitz from Northwestern, Andrew Parsons from UVA, Sara Vick from UKentucky, Brian Locke from InterMountain Health, and

Shuhan He from MGH)
Center for Clinical and Translational Research (CCTR)
Federal (ie., NIH) Grant - Provide source:
Other Grant:
Departmental funding
Commercial - Provide company name:
Department of Defense
Other (e.g. personal funding) - Provide source:

4. Study Sites

A. Provide the names and locations of all study sites where this research will be conducted under the oversight of the UNMC IRB or Joint Pediatric IRB.

Consent will be obtained on-line in a wrap agreement. Data will be collected via surveys and transcripts of ChatBot interview simulations. No patient data will be collected.

B. Will the research be conducted external sites under the oversight of an external IRB?

Yes

Will UNMC, NM, CHMC, and UNO investigators or staff interact with subjects, collect data or solicit consent at those external sites?

Yes

List the sites

UNMC
Northwestern University
University of Virginia
University of Kentucky
University of Utah, Intermountain Health
Harvard University

C. Does UNMC, CHMC or UNO serve as the lead site with responsibility for data and/or safety monitoring?

Yes

List the sites

While UNMC is taking the lead for deployment of surveys (first group of students to participate), data collection, management, and safety monitoring will be performed by David Liebovitz, the Associate Vice Chair for Clinical Informatics at Northwestern University

D. Does this study involve any international sites where the PI will either; 1) conduct 2) supervise or 3) receive / ship HBM or data to / from UNMC?
No

E. Does this study involve face to face contact with subjects?
No

5. Principal Investigator's Assurance
The PI understands and accepts the following obligations to protect the rights and welfare of research subjects in this study:

- **All information in this application is complete and accurate.**
- **I will conduct the research as described in the application and the protocol.**
- **I will ensure that all research personnel are qualified and properly trained.**
- **I will fulfill my responsibilities as PI, described in** https://guides.unmc.edu/books/hrpp-policies-and-procedures/page/126-pi-qualifications-and-responsibilities
- **I will follow all applicable HRPP and institutional polices, and all applicable laws, statutes and regulations.**

Rohlfsen, Cory J - 2025-04-17 23:11:56.560

SECTION II
PURPOSE OF THE STUDY AND BACKGROUND

1. Purpose of the Study

A. What are the specific scientific objectives of the research?

1) To determine whether clinical reasoning excellence can be measured.
2) To explore signatures of clinical reasoning excellence through applications of Shannon's entropy (measured as information gain across a patient interview)
3) To understand and evaluate the learner experience in a standardized ChatBot Observed Structured Clinical Exercise (OSCE)

B. What will be done with the information generated from this research once it is collected and analyzed?

- ♦ Prepare for publication
- ♦ Present within the institution
- ♦ Present outside the institution

Report back to the sponsoring agency for internal use
Other

2. Background and Rationale

Provide background information regarding the proposed research, including a critical evaluation of existing knowledge, and specifically identify the information gaps that the project is intended to fill. This section should clearly support the purpose of the study, contain appropriate key literature citations.

Clinical reasoning, defined as both the cognitive process and outcome of medical decision-making, is a critical ability for clinicians. Despite its importance, available tools for assessing learners' clinical reasoning remain limited and reliant on direct observation, which is resource-intensive, dependent on the expertise of individual faculty, and difficult to scale. Artificial intelligence (AI) presents a unique opportunity to fill this gap by automating standardized tools and rubrics for assessing clinical reasoning (Zhou et al.,*JMIR Medical Education 2025). Our multi-institutional team developed a curated case and rubric specifically designed to assess clinical reasoning, powered by a large language model (LLM) to simulate a patient encounter. The learner directly interacts with the chatbot, taking a patient history, synthesizing key findings, and pivoting their line of questioning as new information emerges. As opposed to traditional OSCEs which rely on binary checklists – whether the learner asked particular questions and ultimately identified the final diagnosis correctly – this chatbot focuses on the entire clinical reasoning process to provide automated comprehensive assessment data to learner and coach. Using layered assessments, the LLM provides personalized feedback on not only the accuracy, but also*

the efficiency, thoroughness, and degree of uncertainty that remains after the encounter concludes. This feedback includes qualitative as well as novel quantitative metrics.

CHARACTERISTICS OF THE SUBJECT POPULATION

3. Accrual
A. What is the number of subjects that will be participating in this research at all sites under the oversight of the UNMC IRB or Joint Pediatric IRB?
100

B. What is the statistical or other justification for this number?
The study is largely exploratory. Based on our preliminary "pilot testing" of the platform, we think generating 25 transcripts across 4 institutions will allow us to reach saturation in the qualitative learning experience (survey-based feedback from participants). As for entropy reduction, we are more interested in feasibility and proof of concept than targeting a specific amount of information gained per question asked. The rationale for this is we don't know if entropy reduction is even a marker of excellence in clinical reasoning. Our enrollment will be diluted by level of training: 1/3rd medical student, 1/3rd resident, and 1/3rd clinical attending. We don't have the bandwidth to validate more than 100 transcripts before our September presentation in Chicago, so we want to keep it small.

4. Gender of the Subjects
Are there any enrollment restrictions based on gender?
No

5. Age Range of the Subjects
A. What is the age range of the subjects?
> 21 years old

B. What is the rationale for selecting this age range?
Adult learners in medical school, residency, or independent medical practice

C. Will children (18 years of age or younger) be included in this research?
No

6. Race and Ethnicity
Are there any subject enrollment restrictions based upon race or ethnic origin?
No

7. Inclusion Criteria

What are the specific inclusion criteria?
Medical student or actively practicing medicine as a post-graduate resident or faculty member

METHODS AND PROCEDURES

8. Description of Procedures

A. Describe in a brief and concise way the procedures, evaluations and/or tests that will be done.
Learners from various institutions will receive an invite to participate in the ChatBot study. Those who accept the invitation will receive a link to a closed-loop streamlit website that is password protected. Once logged in as a guest, they will review the consent agreement and instructions for interacting with the ChatBot. They will be instructed to say they are "Dr. C" using the first letter of their first name to avoid identifying themselves. They can opt to interview the patient by using their microphone (voice) or typing into the computer (digital interaction). No audio recordings will be saved in the cloud. No video recordings will be conducted. Non-identifying demographic data will be collected in a survey separate from the streamlit site after the interview. The feedback for the learner will be immediate and self-contained within the streamlit site. Once the participant has completed their clinical interview and evaluation, they will click "formulate assessment" to receive their personalized assessment and feedback on clinical reasoning.

B. Does the research involve review of identifiable private information?
No

9. Confidentiality

Where will the research data be stored during the study and how will it be secured?
Participant interview transcripts and surveys will be stored in a secure cloud environment separate from UNMC and managed by Dr. David Liebovitz.

10. Privacy

How will the subject's privacy be protected?
All audio recordings will be immediately transcribed into text and subsequently deleted. Participants will be encouraged to introduce themselves as "Dr. C" (first letter of first name). That ChatBot has been tested over 100 times and does not nudge or inquire the learner to disclose additional personal information.

RISK/BENEFIT ASSESSMENT

11. Potential Risks

A. Are there any potential risks associated with the research procedures, interventions or tests?
No

B. Will the research involve the collection of identifiable private information?
No

12. Potential Benefits to the Subject
Are there potential benefits to the subjects that may reasonably be expected from participation in the research?
Yes

Describe.
They'll get to explore a new technology and receive non-evaluative, formative feedback on their clinical reasoning skills in a low stakes environment.

13. Potential Benefits to Society
Describe the potential benefits to society that may reasonably be expected to result from this research.
OSCEs currently cost ~$50,000 at each institution to recruit standardized patients, administer across the student body, and grade. Standardized assessments of clinical reasoning skills using ChatBot will significantly decrease this cost and make feedback of patient interviewing and clinical reasoning skills accessible to student in low resource training contexts.

FINANCIAL COMPENSATION

14. Compensation to the Subject for Participation
Will the subject receive any compensation for participation?
No

SUBJECT IDENTIFICATION, RECRUITMENT AND CONSENT/ASSENT

15.Method of Subject Identification and Recruitment
A. How will prospective subjects be identified?
Each university site champion will identify participants at 3 different phases of training (e.g. pre-clinical, Internal Medicine rotation, Phase 3 OSCE preparation, residency, or faculty division meeting). UNMC site will leverage the M3 Internal Medicine clerkship (Phase 2 of UME curriculum) and residency prep courses (Phase 3 of UME curriculum)

B. How do investigators have ethical access to the names of potential subjects?
We regularly teach lectures and small groups in these curricula.

C. How will potential subjects be contacted for recruitment into the study? Note: If recruitment material (for example fliers or emails) will be used, they must be uploaded as documents for review and approval.
Word of mouth and e-mail invites

16. Informed Consent
Select the appropriate options for how informed consent will be obtained. Note: All consent forms must be uploaded as documents for review and approval.
Direct (in person) contact with investigator
Phone contact with investigator
♦ Signed narrative consent form
Unsigned consent form, email or cover letter
Introductory paragraph (survey research)
No consent (retrospective data collection ONLY)
Other

LITERATURE REVIEW

17. References
Provide a full listing of the references cited in the Background (section II.2) above.
Zhou, Y., et al. (2025). *Large Language Model–Based Assessment of Clinical Reasoning Documentation in the Electronic Health Record: Development and Validation Across Two Institutions. JMIR Medical Education, 11(1), e67967*