UNDERGRADUATE SUMMER RESEARCH PROGRAM PROPOSAL

- Project title

  Large language models for health data summarization

- Faculty advisor

  Scott Wallace

- Project description
  This project will examine using a Large Language Model (e.g., ChatGPT) to summarize health data from the college of nursing. We will explore various methods for prompt engineering and various LLM models to generate summaries of the text. Findings will be compared to hand coded results and to traditional methods such as term frequency counting.

- Deliverables
  - State what results, outcomes, or products are expected at the end of the project.

- Time requirements
  160 hours, Flexible Mon-Fri, 8am-5pm May 16 – July 15

- Constraints
  - None

- Required skills and knowledge
  - Strong knowledge of python

- Preferred qualifications
  - Ideally the student has some experience with machine learning and preferably has taken a course on machine learning (e.g., CS 483)