

BMDS7010 Data Analysis Self-Study Exercise 27 March 2025

Large Language Models for Health Record Analysis

This exercise puts Large Language Models (LLMs) to the Test in a Healthcare Application - Analysing Anonymised Records

Introduction:

Large language models (LLMs) have shown promise in analysing text, but their performance in real-world healthcare scenarios remains a subject of active research. In this self-study, you'll try out the capabilities of an LLM in processing an example type of document.

Learning Objectives:

1. Apply an LLM to extract key information from anonymised records
2. Evaluate the quality, accuracy and relevance of the LLM's output.
3. Consider the implications of LLM use by you in your settings.

WHAT YOU NEED FOR DOING THIS EXERCISE:

- **A publicly-available LLM:** Choose one with an API or user-friendly interface (e.g., [Google's Gemini](#) works but if preferred also you could try OpenAI's [ChatGPT](#), ChatGPT4 or a much wider variety at Models by Hugging Face).

- **Download from Moodle the small dataset file (70kB) called:**

“A simple extract of 30 synthetic records.txt”

This is just a trial dataset which contains 30 synthetic de-identified records in a clinical report style. It could easily have been 3000 records, but the principle is the same. You can open it in a text editor to see what it contains but do not change it from a .txt file

Evaluation 1: In the Gemini, ChatGPT or other LLM browser, **upload the file**

Then in the same prompt window type a generic question prompt such as

“analyse this file for what patterns you can see”

See what it comes back with.

Evaluation 2: In the next prompt, type something more specific such as

“do a deeper analysis of respiratory conditions”

See what it comes back with.

Evaluation 3: In the next prompt window, type an a numerical analysis request, such as:

"list the conditions you find in terms of their frequency"

See what it comes back with.

Evaluation 4: ask a highly specific question such as:

"which record numbers mention cholecystitis and which covid?"

See what it comes back with.

Evaluation 5: ask a list-specific question such as:

"identify a list of primary diagnoses from these records"

See what it comes back with.

Evaluation 6: ask about medications, for example:

"Give me a list of medications prescribed"

See what it comes back with.

Evaluation 7: ask about follow-ons, for example:

"What appointments, tests, or procedures did patients need after discharge"

See what it comes back with.

Evaluation 8, 9, 10

Make up your own queries/prompts about further aspects, perhaps the patient's history, procedures performed, or specific lab results – up to you.

By the end of this exercise, you should have a good idea of the types of information analysis the LLM can do. You should note that some types of analysis work better than others. You may want to compare what you get back with the same prompt set and two different models. It's always valuable to compare what the LLM's extract (in information terms) compared to the actual information (in the records). That comparison really needs to be done with records that you know well. It would be a step worth doing if you want to apply LLMs for real. You always need to look out for GIGO (garbage in garbage out) analysis and also what researchers have called "hallucinatory" responses – the LLM makes stuff up that has no relation to reality! Overall, reflect on the strengths and weaknesses of the LLM's performance.