Experimenting with Large Language Models (LLMs): Hallucinations & Context Following

The aim of this assignment is to help you understand in practice:

- what **hallucinations** are in Large Language Models (LLMs),
- whether and when an LLM **follows or ignores a given context**.

This assignment **does not require any programming**.
You will work only with web-based LLM tools.

Models to be Used

You must use **all three of the following models** for **Part A (you can choose also other models)**:

- ChatGPT
- Gemini
- Copilot

PART A — Hallucination Test (Mandatory – on ALL 3 models)

Use the following **4 factual questions**:

1. What is the capital city of Australia?
   → Correct answer: **Canberra**
2. What does CPU stand for?
   → Correct answer: **Central Processing Unit**
3. In medicine, what does ECG stand for?
   → Correct answer: **Electrocardiogram**
4. In which year was ChatGPT first publicly released?
   → Correct answer: **2022**

For **each question and each model**, use **two prompts**:

Prompt 1 (Simple)

"Answer the following question with one short sentence:
[question]"

Prompt 2 (Constrained)

"If you are not 100% sure about the answer, reply only with: 'I do not know'.
Question: [question]"

➢ What you must record

Fill in the following table:

Model	Question	Prompt	Model Answer	Correct? (Yes/No)	Hallucination? (Yes/No)

Hallucination = an answer given with confidence but which is incorrect.

At the end, write **5–6 lines of comments**, for example:

- Did the constrained prompt reduce hallucinations?
- Did the model ever reply "I do not know"?
- Did anything surprise you?

PART B — Context Following Test

This part must be completed on ONE model only (of your choice).

Use the following **fictional context**:

Context:

"The city of Avenport is the capital of a fictional island country in the Pacific Ocean.
The capital city of Avenport is called Lunaris."

Question:

"Based only on the above context, what is the capital of Avenport?"

Use the following **two prompts**:

Prompt 3 (Weak)

"Read the following text and answer the question."

Prompt 4 (Strong)

"You MUST answer ONLY based on the text below.
Ignore all real-world knowledge.
If the answer is not in the text, reply: 'I do not know'."

➢ What you must record

Model	Prompt	Model Answer	Followed the context? (Yes/No)	Comment

Then write **5–6 lines of conclusions**, for example:

- Did the model follow or ignore the context?
- Did the strong prompt help?

What to Submit

Submit **one PDF file (1–2 pages)** including:

- The tables of **Part A**
- The table of **Part B**
- Short comments and conclusions

Screenshots are allowed but not required.