GenAI Assignment Q3  
  
**Proprietary LLMs** : Pick any one the proprietary LLMs, and compare the Accuracy, Speed, Cost for a same task. To make more challenging, increase the complexity of the task and measure below KPIs multiple times.

**KPI**

* Accuracy
* Speed
* Cost

## Iteration 1:

Common Initial Prompt:

I want to build a quiz app using python. Ask the full name and date of birth of the user first. We would need to store the user details somewhere and if he tries to login again, then he can login again with his full name and date of birth. the user will be given four options. The user will click on the four options and then the colour of the right option should be green if he selects it and red if he gets it wrong. Take ten random questions from your world general knowledge and their four potential answers with one option as correct and rest of them wrong. Keep increasing the difficulty level if he starts getting the answers right. Show the total score finally and also for each user, the score must be stored with the timestamp as well. Make the UI and UX visually appealing.

### ChatGPT:

* A few prompts:
  + give me more question and answers so that they can easily integrate with the above code. Keep incrementing the difficulty of questions.
* It took me 24 prompts to get the final app with most prompts were about fine tuning the app.

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Bard:

* A few prompts:
  + build a json file for questions along with their answers to get along with the code you have given me. gradually increase the difficulty of questions.
  + only one option changes every time in the game window, I see Paris in most of the other options
* Initially, Bard gave me only text responses, no code, I had to explicitly ask it to generate the python code.
* This was the best Bard could come up with 33 prompts and manual intervention in the code. And that too with only one option loading, rest of the options are repeated ones.  
    
  A screenshot of a computer

  Description automatically generated

Iteration 2:  
  
Common Initial Prompt:  
I want to build an interactive python quiz app. initially, the app asks about the full name, email and date of birth. and assign a unique id to that user based on his credentials. everytime a user logs in, the score and start and end timestamp of that quiz should be stored in a json file. the quiz consists of GK questions and are retrieved from another json file. Create 10 questions for the app. at the end of the quiz, show summary of the session and the score. the summary includes whether the user chose the correct option or incorrect Give UI to the app.

### ChatGPT:

* A few prompts:
  + just a simple json format which can be integrated with the quiz code. remove any indexing in the json file
  + now i want the summary to include the question and all the options with the correct answer and the option which the user chose. use red green color-coding scheme also include the total time taken by the user to finish the quiz so that we can rank the users in future
* It took me 5 prompts to get a working app.
* It took me 16 prompts to get the final app with some UI with ChatGPT
* This time, even after multiple trials, ChatGPT could not give me the correct final score.  
    
  A screenshot of a computer

  Description automatically generated A screenshot of a computer

  Description automatically generated

A screenshot of a computer

Description automatically generated

Bard:

* A few prompts:
  + give me the python code, not the text.
  + generate the code from where you left off.
  + again, the user details variables are uninitialized. rectify it.
  + initialise the root variable also.
  + just give me the summary function
* Again, Bard gave me only text responses, no code.
* This time, it gave me modular code, not everything in just one function.
* This was the best Bard could come up after 17 prompts, the app could not even load the quiz on the UI. It kept making the previous mistakes which I explicitly asked it to rectify.  
    
  A screenshot of a computer

  Description automatically generated

Iteration 3:  
  
Common Initial Prompt:  
I want to build a python-based quiz app It should ask about the user details and the details will be stored in a json file. the timestamp of start and end of quiz and total time taken for quiz completion must be noted The questions must come from a json file There must be a quiz summary after the quiz ends where each question must be displayed with the correct answer as well as what the user chose The app must have a decent UI.

### ChatGPT:

* A few prompts:
  + now give me a json file of 10 quiz questions of increasing difficulty
  + missing things in the app:  
    not asking for user details like name, email and storing them in the json file afterwards not showing all questions with their respective answers and chosen answers in a red green coding scheme
* This time, ChatGPT gave me the full app in 3 prompts.
* Generated the correct code for everything but the user json file does not make sense.
* ChatGPT gave me the directory structure for the codebase this time.

A screenshot of a quiz

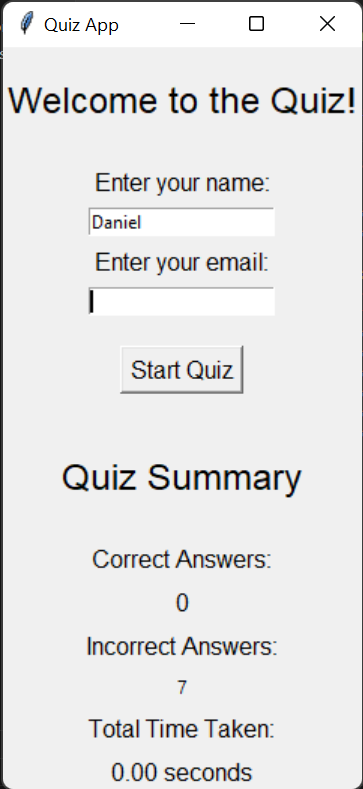
Description automatically generated A screenshot of a quiz

Description automatically generated  
  
A screenshot of a cell phone

Description automatically generated

Bard:

* A few prompts:
  + give me the python code, not the text.
  + generate the code from where you left off.
  + I do not want user to resume quiz so do not save and load progress
* This time, without asking, Bard generated some code to save the user sessions so that he can resume from wherever it left off.
* Bard does not generate long code scripts like ChatGPT does. In other words, it does not output large number of tokens. It stops midway.
* Generated the correct code for everything but the user json file does not make sense.
* After 20 prompts,



Iteration 4:  
  
Common Initial Prompt:  
Build a working python quiz app with basic UI ask the user his details and store it in a json file, so that if he logins again, make changes in that file only. also store the quiz details in that file for that session take the questions from another json file. generate a few questions for me with increasing difficulty show quiz summary after the quiz ends.

### ChatGPT:

* A few prompts:
  + Error: No binding for nonlocal "score" found
  + I cannot submit my option after the first submission and i want buttons instead of radio buttons also the user details window should disappear after saving his details
  + The quiz does not start and script ends, correct it.
* After 12 or so prompts, it did not return the desired result.
* The quiz did not even start this time, and the script ended.
* Gets stuck in a loop of being prompted and returning the same output.
* ChatGPT could not find out the reason this time and when prompted to regenerate the code, it returned the same code each time I asked.

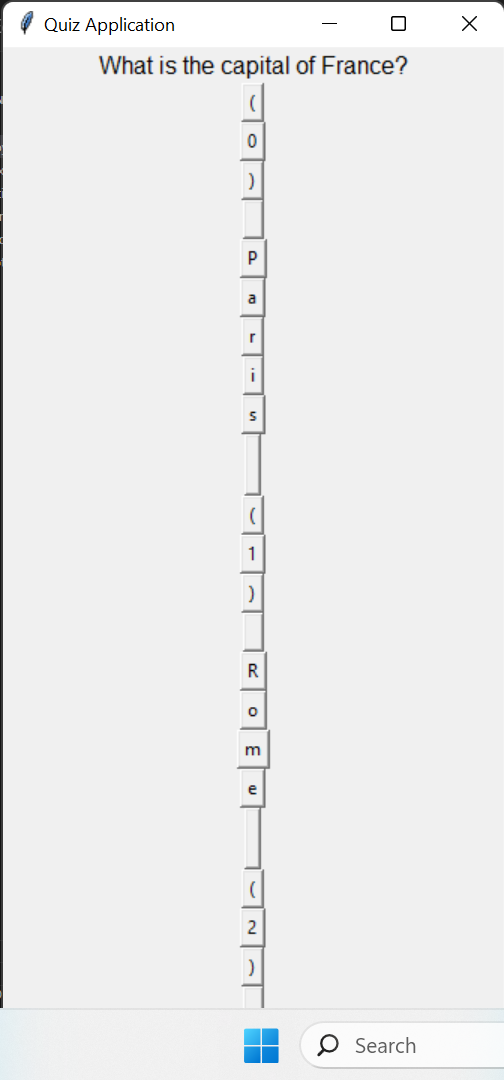
Bard:

* A few prompts:
  + the GUI should contain the clickable buttons as the options.

and the following is printed each time user submits:

"Incorrect. The correct answer is Russia".

* + Maybe you are iterating over each character of each option, take this into account and rectify.
* It started by generating a good working script, but the output was on command line.
* When prompted to generate UI using *tkinter* library, each character of every option became a clickable option.
* Upon asking to rectify it and some manual intervention as well, it was unable to give me the correct code and was generating the output which had no effect.



##### Key Takeaways:

* + These AI tools are there for our assistance mostly, end to end working code generation is still a remote concept but we will be there very soon.
  + Naïve errors they produce in the code mostly include uninitialized variables, no reference to variables, or accessing and operating on wrong data types, etc.
  + They generate ready to use boilerplate code. Useful in the sense of integrating some functionality with the existing codebase.
  + When assigned with multiple tasks in a single prompt, any GenAI model is not able to focus completely. Some of the functional requirements will be lost and the user has to ask the model to develop that functionality.
  + **ChatGPT**:
    - It takes very less prompts to generate the boilerplate code or decent working code as well.
    - Generates modular code mostly without explicitly saying.
    - Longer context remembering.
    - Input and output token size is quite large unlike Bard.
    - Takes an extra step while clarifying things.
    - Max token size is 4096 tokens for the free version and 8192 for the paid version.
  + **Bard**:
    - Initially, it generates text only unless explicitly asked about generating code.
    - It tends to repeat the past mistakes.  
      For example, if I ask it to import some package, it will do the desired once, but it will forget the context in future prompts. So, five or six extra prompts are required to get a minor functionality again. Maybe the attention module for the new prompts is weak.
    - Less number of output tokens unlike ChatGPT. May generate full code initially, but after a few prompts while generating the code, it stops midway and even after asking it to continue from where it left, it generates the whole code again and gets stuck again near the previous checkpoint. So, it can create a kind of deadlock situation.
    - Upon asking to modify something, it just changes function(s) but the effect of those changes is not incorporated in the rest of the code.
    - Max input token size is upto 1000 tokens and the output token size is unspecified.