TASK-2

**2.1 Prompt used**

Generate CSV: Class (6), Subject (infer), Ch Num (infer), Ch Name (infer), Topic Num, Topic Name (sec heading or quote text), Other content.  
"Other content" (quote if has comma): "Figure it Out", "Table X:[Title]", "Example:[summary]", "Activity X.Y:[Title]", "Quote:'[text]'", or "Image:[desc]".  
Start each Topic Name row with a blank "Other content", then new rows for each allowed item under it. Strictly mention the sub-headings mentioned in other contents only, also if there's a “comma” in the sentence, then enclose it with quotation marks.

**2.2 What technique is used to extract data**

The main way the information was pulled out was by guiding an AI (a smart computer program) with step-by-step instructions given in plain language.

Here’s how it worked:

1. **The User Provided the Text:** The user first got the raw text from the PDF pages using a separate tool (OCR). This text was then given to the AI. The AI didn't read the PDF images directly.
2. **The AI Followed Instructions:** The AI then "read" the instructions (prompts) provided by the user. These instructions told the AI what to look for in the text, such as:
   * Finding main sections and headings.
   * Spotting specific items like "Figure it Out" boxes, tables, examples, activities, quotes, or image descriptions.
   * Organizing the found information into specific columns for a CSV file.
   * Applying formatting rules, like putting quotation marks around text that included commas, or creating short summaries.
   * Ignoring any text that wasn't specifically asked for.
3. **Creating the Structured List:** Finally, after understanding all the instructions correctly, the AI generated the information in the requested CSV format, ready for use in a spreadsheet program.

2.3 Your step-by-step process plan: how the document was fed to the AI, cleaned, and converted

It all started with the user having a specific idea for organizing information from PDF textbook pages. Here's how they worked with a smart AI helper to make it happen:

1. **Initial Prompt :** The process began by giving the AI helper a set of instructions (a prompt). This first set of directions outlined the basic idea: "I want a list (like a spreadsheet) with columns for things like Class, Subject, Chapter, Topic, and some other specific content I'm interested in.
2. **The AI Helper's First Attempt:** The AI read through the text, trying its best to follow those initial instructions. It looked for chapter details and started to identify what seemed like the main topics. However, since every book is a bit different (some have numbered sections, others use big questions to guide you), the AI sometimes got a bit confused or missed things it was supposed to find.
3. **Review and Refine:** The user looked carefully at what the AI helper produced. They’d notice if:
   * Some text was split into the wrong columns (especially if there were commas in the middle of something).
   * The AI included extra bits of text that weren't needed.
   * The "Topic Name" or "Topic Number" wasn't quite right for how the book was laid out.
   * Some specific items they wanted weren't picked up.

Based on this, much clearer instructions were given. For example:

* + "Okay, for the 'Other content' column, I *only* want you to list it if it's exactly a 'Figure it Out' box, a 'Table', or an 'Example'. Nothing else."
  + "When you make the list, if any text you put in a cell has a comma, please put quotation marks around that whole piece of text."
  + "For the math book, use the section numbers for 'Topic Number.' For the science book, use the big questions in the blue boxes as 'Topic Names' and just number them 1, 2, 3..."
  + "And make sure each new topic starts with a blank spot in that 'Other content' column before you list the items under it."

1. **The AI Tries Again:** With these more detailed instructions, the AI would process the text again, getting closer to what we wanted. This back-and-forth of checking and giving more precise directions, and the AI trying again, happened several times.
2. **The Final List:** After these rounds of refinement, the AI was able to produce the final list (the CSV file) exactly as envisioned, neatly organized and following all the specific rules for both types of textbook chapters.

2.4 Any challenges faced and how they were solved

**The Comma Problem:**

One tricky bit was when the information itself had commas in it.

* **The Challenge:** Imagine the AI helper was writing down a summary for an "Example" like this: "The pattern is, first you add 2, then you add 3." Or an "Image" description like: "A desert, a coast, and an ocean."  
  When you make a simple list (like a CSV file) for a spreadsheet, the comma is usually the special signal that says, "Okay, start a new column here!" So, if the AI just wrote that summary with its commas, the spreadsheet would get confused and split it up into tiny pieces across several columns, which wasn't what we wanted.
* **The Fix:** Using special rule: "If any piece of text you're putting into a cell for 'Other content' has its own commas inside it, please be sure to put quotation marks (") around that *entire* piece of text."  
  So, instead of just writing: Example: The pattern is, first you add 2, then you add 3.  
  The AI learned to write: "Example: The pattern is, first you add 2, then you add 3."

2.5 A brief summary of what you learned

**The AI is Smart, But It Needs a Good Map (Your Instructions!):**

The AI is powerful and can do a lot, but it doesn't automatically know exactly what you want it to do with something like a textbook chapter.

It needed very precise directions (guidance and corrections) to ensure that the final list was exactly how we wanted it, especially since we had very specific rules for organizing the information.

The user played a big part by checking the AI's work and telling it exactly what to fix or change. It was like the user was the guide, making sure the AI stayed on the right path.