Prompt Workflows for Theme Development

TIM 175 WEEK 4 LAB

In the past two weeks, you started using GenAI for the initial coding process for extracting relevant passages and for defining codes/themes from transcript excerpts. This week you will be extending your prompts to cover the entire process of thematic analysis from initial coding to theme development. You will define a prompt workflow that includes the initial coding prompts and evaluations from Weeks 1 and 2 and utilizes the produced initial codes and quotes to define a final set of themes and subthemes. This will include an **individual deliverable (due Saturday at 11:59pm)** and a **team deliverable (due Monday 11:59pm)**

Readings:
See Week 4 PreLab or Course Spreadsheet.

Submission Link
[Week 4 TIM 175 Submission Form (Spring 2025)](#)

Brief Task Overview
Individual deliverable
- Understand Prompt Workflow Diagrams and Create Your Own Diagram
    1. Read the Concept Induction Paper
    2. Write a short reflection on the Paper
    3. Design your own workflow diagram.
- Implement Your Initial Coding Prompts in a Google Colab
    4. Make a copy of the [Google Colab workbook for thematic analysis](#) and ensure you have the API key setup
    5. Upload a transcript and run the provided code for splitting it into an array of question-answer excerpts
    6. Implement your past prompts for initial coding in the Google Colab
- Complete the Remainder of the Prompt Workflow for Thematic Analysis
    7. Import your generated HW 2 or HW 3 outputs to create a list of all your code/quote pairs using this [Google Spreadsheet](#)
    8. Implement the remaining steps in your workflow for creating overall themes and subthemes.
    9. Refine your prompts
    10. Document the best version of your final output on the [Google Form](#)

Team deliverable
1. In-section peer review of evaluation
2. Compile your reviews, make another copy of this spreadsheet
3. Iterate as a team
4. Reflect and submit using [the Google Form](#)

Individual Assignment Instructions

Task 1: Understand Prompt Workflow Diagrams and Create Your Own Diagram

Read the Concept Induction Paper. For this assignment, it is important that you read the [Concept Induction Paper](). Pay attention to:
- How they design their workflow, i.e. how they break down the problem into multiple smaller subtasks. You want to build intuition on how complex tasks can be decomposed into smaller ones (the lecture and other readings also covered many other examples for building intuition),
- How they visualize the workflow in a workflow diagram (Figure 3).

Look at the appendix to see their detailed prompts. You can also see these in their [open-source code]().

Write a short reflection on the Paper. Write a one-paragraph reflection of what you learned and understood from the paper, their workflow design, diagram, their prompts and evaluation. You will submit this one-paragraph reflection directly to the Submission Form.

Design your own workflow diagram. Read through the rest of this Google Document to fully understand the task of thematic analysis (Task 2) and then create a workflow diagram, similar to the one in the paper (Fig 3), for the task. Your figure should show how you designed your workflow, i.e. how you break down the task of thematic analysis into a series of smaller steps starting from a set of interview transcripts to a final set of themes/subthemes and associated passages. You will submit a link to a document containing your workflow diagram to the Submission Form.

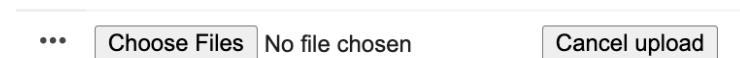## Task 2: Implement Your Initial Coding Prompts in a Google Colab

The first part of your thematic analysis workflow will be the initial coding portion which produces a set of codes and quotes from a set of transcripts. You have already defined prompts that extract codes and quotes from transcript chunks and evaluated them using evaluation prompts. Previously, you did this in LastMileAI. Your task is to now implement this in a Google Colab.

Make a copy of the Google Colab workbook for thematic analysis and ensure you have the API key setup. Make a copy of [this Google Colab Template](#) to create a new prompt workflow for conducting the thematic analysis. Make sure you have an API key added with the name "GEMINI_TOKEN" and that you toggle "Notebook Access" on following the same steps as the Prelab.

Change runtime type to T4 GPU. On the top right corner you will see your connection status and a dropdown menu for "Additional connection options". From there change your runtime type to T4 GPU and save. GPU connection is preferred for AI tasks over CPUs primarily due to their parallel processing capabilities. This will make your notebook run more efficiently.

Upload a transcript and run the provided code for splitting it into an array of question-answer excerpts. The first step of the workflow is to take a set of interview transcripts and process them into question-answer pair excerpts to feed to the next part of your workflow. We have already provided you with a code for processing one transcript. You need not to make any changes for this step. Simply run the Google Colab cells and upload a transcript as a txt file. You will see the below UI for uploading a file when you run the cell.

Implement your past prompts for initial coding in the Google Colab. The next step of the workflow is to extract initial codes and quotes from the transcript excerpts. Since you have already defined prompts for this in the previous two weeks, you just need to implement these within the Colab. You should apply the main prompt to the excerpt and then apply your evaluation prompts to the produced JSON array (including looping over the array to apply evaluations that are designed for a single code/quote pair). For this assignment, you do not need to then use the evaluation results to further improve the generated output (though if you do, that could definitely make your submission an outstanding one). Test your code on a few excerpts to make sure it returns the expected JSON array, but you do not need to apply it to all excerpts or record your results.
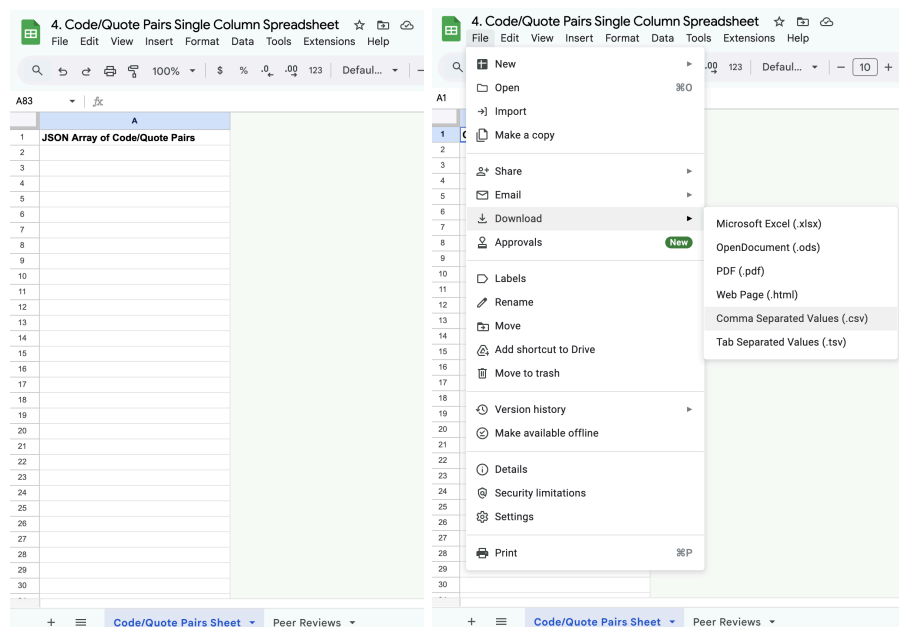
Task 3: Complete the Remainder of the Prompt Workflow for Thematic Analysis

After obtaining a list of relevant passages and initial codes, we would like to utilize the extracted codes/quotes to create a set of high-quality themes and subthemes that capture the entire set of career insights and organize them into higher-level types of career insights. An ideal set of themes and subthemes is coherent, distinct, and grounded, and helps to communicate the "shape of the data". In these final steps, you will implement prompts for determining the set of themes and subthemes.

Import your generated HW 2 or HW 3 outputs to create a list of all your code/quote pairs. In the normal workflow, we would apply the code you wrote above to all the excerpts for all six transcripts to obtain a list of all extracted code/quote pairs. However, we will not ask you to do that to save your API call quota and because you have already extracted code/quote pairs for all the excerpts in your previous six transcripts.

Instead, you are going to upload your generated code-quote pairs from previous weeks as a CSV file so that it can be processed into a list that you can use for subsequent steps in your workflow. For this, you will first need to manually put together your full set of code/quote pairs (the JSON arrays produced from your main prompt)into a Google Spreadsheet with only **ONE** column and download that as a CSV file. The Google Collab will ask you to upload this CSV and will process it into a list for you itself.

Note: we understand that the outputs from HW 3 were only the raw outputs from your main prompt and did not utilize your evaluation prompts to further improve the quality. However, we will just work with these for the purpose of this assignment.

Implement the remaining steps in your workflow for creating overall themes and subthemes. After defining initial codes and extracting relevant passages, we would like to find patterns or common themes across those codes/passages to determine a single set of themes and subthemes that describe the space of career insights.

Your workflow should take the set of code/quote pairs that you generated and produce a set of overall themes and subthemes. Your final output of the prompt workflow should be a hierarchical outline of your themes, subthemes, and 1-2 quotes per subtheme chosen to be the most salient, essence-capturing, or evocative quotes for illustrating that theme/subtheme. This should not be done in a single prompt, but should involve a series of prompts. For example, you might first extract clusters before developing high-level themes or determining whether themes need to be split or combined or determining the best representative quotes per theme.

Here is an example of a format that would show the hierarchical set of themes and subthemes along with an illustrative quote per subtheme:

<Theme 1>

* <Subtheme 1.1>: <quote>
* <Subtheme 1.2>: <quote>
* ...

<Theme 2>
* <Subtheme 2.1>: <quote>
* ...

...

Refine your prompts. Refine your prompts by using prompt engineering techniques, evaluations, examples or guidelines that will help you create high-quality themes and subthemes.

Document the best version of your final output. You will submit your best attempt at the final set of themes, subthemes and salient quotes in the Google Form.

Here are some tips / reminders to keep in mind about defining themes and subthemes:

- Themes are clusters of passages that are similar to each other in some way that relates to the research question. The theme should clearly and concisely capture the main essence of how the code/quote pairs in a cluster relate to the research question.

- Themes can be combined or split to create a hierarchy of higher-level themes and lower-level subthemes. A good organization of themes into higher-level themes and subthemes helps to communicate the shape of the data (i.e. the space of career insights).

- You want this hierarchy of themes and subthemes to be distinct, coherent, and grounded. Distinct themes help to map the space around career insights. Grounded means that a theme should reflect the data categorized to that theme. Coherent means that the data within a theme cohere together.

- If there are too many themes that it is hard to wrap your mind around them, they should probably be grouped together into higher-level themes. If the data within a single theme seems to have distinct things to talk about, it might be worth splitting it into smaller themes. Note: it doesn't make sense to have a theme with a single theme underneath it as that defeats the purpose.

Reflect and submit. Finally, reflect on the individual activity. Submit your responses to the below questions directly on the Google Form along with your workflow diagram, prompt workflow link and final set of themes. **IMPORTANT: MAKE SURE TO GIVE US PERMISSIONS TO ACCESS DOCUMENTS**

1. Please provide a one paragraph reflection of what you learned and understood from the Concept Induction Paper including their workflow design, diagram, prompts or evaluation.

2. Please provide a one paragraph reflection of how you designed your initial prompt workflow (i.e how you broke it down into multiple steps). What did you learn from creating the workflow and the workflow diagram? What advantages/improvements were you able to make by having a series of prompts as opposed to a single prompt?

3. Please provide a one paragraph reflection on your process of refining your prompt workflow? What changes did you incorporate in your prompt workflow to improve it?

4. What do you think is lacking in your workflow and how would you want to improve it?

Team Assignment Instructions

In-section peer review of evaluation. Among the team members who are present at the section, conduct a peer review of each other's work. We don't have a strong preference for how you assign peer reviewers, but one way to do it is to use a cycle A → B → C → D → A. In other words, if you have 4 team members present (A, B, C, D), have A peer review B, B peer review C, C peer review D, and D peer review A.

Examining your team member's prompt workflow google collab, their workflow diagram and their final set of themes/subthemes, take note of specific ways they could have improved in the diagram or prompts. Then:

- Write a two sentence reflection on what your team member did well in. For example, how did they break down the task into a series of steps? Were there any prompt engineering techniques they used well?
- Write a two sentence reflection on a concrete, specific way in which your partner could have improved their diagram, workflow or their prompts. You should focus on what is most important / would make the biggest impact or improvement.

Note: The course staff will share some initial observations and pointers, but this is only based on a quick review of the individual submissions. You should take what they say into account, but make sure to be detail-oriented in thinking about the submission you are reviewing and how it should improve.

Compile your reviews. Create a copy of [the Peer Review spreadsheet](#) and compile the reviews from all of the participating team members in the 'Peer Reviews' tab. Specify for each review, who was the reviewer, who was being reviewed, the two sentence reflection on what your team member did well in, and the two sentence reflection on the most important way they could have improved.

<mark>We are looking at the quality of your reviews, so the way to maximize your points is to write the most helpful critique that points out the biggest way each team member can improve.</mark>

Iterate as a team. Work together to create a single final team version of the prompt workflow, building on the work you have already done in the individual submission and considering what you might want to draw from each submission (e.g. workflow design, diagram, specific prompts.) Note: you do not have to use something from every individual submission. You should be trying to create a final prompt workflow that produces the highest quality outcomes you can achieve.

Reflect and submit. Submit your responses to the below questions directly on the [Google Form](#) along with your final prompt workflow, diagram and final set of themes/subthemes. **IMPORTANT: MAKE SURE TO GIVE US PERMISSIONS TO ACCESS DOCUMENTS**

1. Write 3-4 sentences on your team's thought process. How did you choose what prompt workflow design to use? What influenced your changes to the final prompt workflow? What

further improvements can be made? What feedback did you incorporate (from peer review or your tutor)?
2. Write a sentence or two describing the team dynamics. Were there any challenges you faced working in your team and how did you overcome them?
3. Please list each member of your team, whether they attended and engaged in section discussions, and their specific contributions.

Evaluation Rubrics

Individual submissions will be graded on a Check+, Check, Check-, Minus+, Minus scale according to the below rubric. The purpose of individual submissions is primarily to ensure that all members are contributing to their team, so graders will not be providing feedback on these.

Check+   Outstanding, one of the best in the class (102%),
Check   High quality, though not one of the best in minor ways (95%),
Check-   Completed the work, but needs significant improvement (80%),
Minus+   Low quality or missing significant portions (40%)
Minus   Did not do the work or barely did any work (0%)

Team submissions will be graded according to detailed rubrics to be posted on the rubric spreadsheet