Prompt Engineering for Thematic Analysis

TIM 175 WEEK 2 LAB

In our first project (Weeks 2-5), we will use Generative AI to extract themes and create rich narratives around career insights grounded in the stories of professionals in the community shared on YFIOB's What-To-Be podcast. In Weeks 2-3, you will be focused on identifying relevant passages from the transcripts. Then in Weeks 4-5, you will focus on developing high-quality themes and rich narratives. This week, you will apply prompt engineering techniques to identifying relevant passages, including an **individual deliverable (due Saturday at 11:59pm)** and a **team deliverable (due Monday 11:59pm)**

Readings: see HW 1 and HW 2 Prelab readings on thematic analysis and prompt engineering.

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

Brief Task Overview
Individual deliverable
- ☐ Create a workbook on [LastMileAI](#)
- ☐ Determine the transcripts you will be using for your task
- ☐ Create a spreadsheet to document your results, use this [template](#)
- ☐ Break down your transcripts into small excerpts
- ☐ Design a simple prompt to create codes and extract quotes from your excerpts
- ☐ Apply and refine your prompts
- ☐ Reflect and submit your individual work using this [Google form](#)

Team deliverable
- ☐ In-section peer review, provide constructive feedback to your teammate's deliverable
- ☐ Compile your reviews, create a copy of this [template](#)
- ☐ Iterate as a team to create a final prompt
- ☐ Reflect and submit your group work using this [Google form](#)

You can use ChatGPT or other GenAI tools to inform any part of the assignment but: (1) you need to first form your own independent thoughts, (2) every word included in the submission needs to be something you've read, thought about, and decided to include, and (3) you should strive towards submitting the highest quality work you can rather than mediocre work that meets the requirements.

Individual Assignment Instructions

Create a workbook on LastMileAI. Go to https://lastmileai.dev, create an account with your UCSC email, and create a new workbook titled "Project 1 - Week 2 Lab". This is where you'll be constructing your prompts. Remember to use the GPT-4 model.

Warning! Make sure you don't exceed the weekly token limit in LastMileAI. This can happen if you change the models to a heavy model like TTS. So please stick to using GPT-4 or ChatGPT models and do not use LastMileAI for anything irrelevant from the Lab.

If you cannot use LastMileAI for any technical reason and you receive an error when you run a cell, try restarting LastMileAI. If it does not work, please inform a TA/tutor. You may then use this Google Colab as a template to generate the rest of your outputs.

Determine the transcripts you will be using for your task. Each team will be designing prompts and applying them to conduct a thematic analysis on **four new** interview transcripts. In the individual portion of the assignment, you will be responsible for **two** of the four new transcripts (check the assignment in the 'Teams' tab of the course spreadsheet for which two).

Teams 1-4
- 003_SukhSingh_CEOofCodeNaturally.txt
- 008_IrvinLemus_EthicalHACKERComputerInformationSystemsInstructor.txt
- 011_Brook Ewoldsen_Fashion Institute Of Design (FIDM) .txt
- 012 _Rob Gaukel_Chief Operating Officer of Verve Coffee.txt

Teams 5-8
- 0019_Jose Ibarra_Resident Services Manager.txt
- 0020_Lauren Donnelley Crocker_Development Director Seymour Marine Discovery Center.txt
- 025_AkiWilliams_ChiefOperationsOfficerDefibThisFlightNurse.txt
- 039_RachelRassmussen_HRDirector&BaristaatCatandCloudCoffee.txt

Teams 9-12
- 077_Burr Guthrie Coordinator of Adult Education at PVUSD.txt
- 081_JessicaRidgeway_ExecutiveDirectorOfFarmDiscovery.txt
- 091_SusanO'ConnorFraser_TamCommunications.txt
- 096_Lisa Uttal _MarineBiologistNOAA.txt

Teams 13-16
- 097_Rachel Anne Goodman _ KSQD, Journalist and Teacher_otter_ai.txt
- 098_Andrea Mollenauer _ Chef Instructor Cabrillo College Culinary_otter_ai (1).txt
- 121_ErinHubach_MontereyBayAquariumInterpreter&VolunteerSupervisor.txt
- 136_DanielleGlagola_MarketingCommunicationsCustomerServiceDirector.txt

Team 17-20
- [151_NadaMiljkovic_ArtistUCSCProfessorRadioPersonalityandEntrepreneur.txt](#)
- [152_ReneGolder_PrincipalBayViewElementarySchool.txt](#)
- [154_PatriciaGreenway_SantaCruzCitySchoolsCareerDevelopmentSpecialist.txt](#)
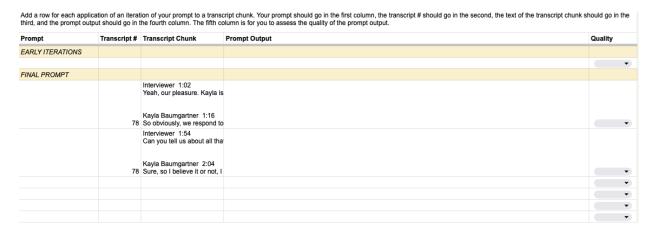- [165_ChristinaOrtega_AcademicCounselor.txt](#)

Team 21-24
- [166_DanteSearcy_HRGeneralist.txt](#)
- [169_KyleVasquez_SevenDirectionsInstitute.txt](#)
- [170_JohnLaird_CaliforniaStateSenator.txt](#)
- [184_DeanBussiere_Chef.txt](#)

Create a spreadsheet to document your results. Copy this [prompt engineering spreadsheet](#), which you will eventually use to document the results from your prompt iterations.

Break down your transcripts into small excerpts. For each transcript, you need to break it up into chunks since you will be applying the prompt you define to one chunk at a time. A simple way to chunk your transcript, which we will use in this assignment, is to use question-answer pairs (i.e a question is asked by the interviewer and an answer is given by the interviewee would be one chunk).

You should add each chunk underneath the "FINAL PROMPT" section of the given spreadsheet, with each chunk given a separate row, filling in columns B and C like below:

Add a row for each application of an iteration of your prompt to a transcript chunk. Your prompt should go in the first column, the transcript # should go in the second, the text of the transcript chunk should go in the third, and the prompt output should go in the fourth column. The fifth column is for you to assess the quality of the prompt output.

Prompt	Transcript #	Transcript Chunk	Prompt Output	Quality
EARLY ITERATIONS				
				▼
FINAL PROMPT				
	78	Interviewer 1:02 Yeah, our pleasure. Kayla is Kayla Baumgartner 1:16 So obviously, we respond to		▼
	78	Interviewer 1:54 Can you tell us about all that Kayla Baumgartner 2:04 Sure, so I believe it or not, I		▼
				▼
				▼
				▼
				▼

Design a simple prompt. Craft a prompt in your LastMileAI workbook that takes a single transcript chunk, identifies whether there are one or more passages relevant to the research question, and then for each relevant passage, extracts the relevant quote and defines an annotation (a code / theme) that captures how the quote might be relevant to the research question.

As a reminder, the research question is the same as in Homework 1:

> *"What career insights, perspectives or tips do professionals have that might be helpful for a young person who is just beginning their career journey?"*

You can even have two separate prompts (one to generate relevant codes/themes and one to extract the quotes, or vice versa).

Apply and refine your prompts: This is the most important part. As you apply your prompt to Q&A transcript chunks, you need to refine your prompt using prompt engineering techniques to iteratively improve the quality of your output. You should be looking at whether all relevant passages were assigned a code and whether the code captures the essence of how it relates to the research question.

We recommend you use one Q&A chunk at a time to refine your iterations. **You also need to record the results of your iterations in the prompt engineering spreadsheet you created.** Your initial iterations should go in the "EARLY ITERATIONS" section. When starting with your first Q&A chunk, copy it from the "FINAL PROMPT" section to the "EARLY ITERATIONS" section to record your initial iterations. Once your prompt works well on the first Q&A chunk, apply it to a second and third and so on, continuing to copy these chunks and record the results under the "EARLY ITERATIONS" section.

Continue until you: 1) have confidence that your prompt will produce high-quality results, and 2) have **at least** made 5 iterations of your prompt with the latest applied to and producing high-quality results on at least 10 diverse chunks. An iteration also **does not** mean changing 1 or 2 words - you should be applying different techniques / noticeable changes to the structure / detail of your prompt in each iteration. If you feel like you are already getting high-quality results before your 5th iteration, push for even higher-quality results or try more challenging Q&A pairs!

Assess your final prompt. Once you are satisfied, apply your best prompt to all the chunks under the "FINAL PROMPT" section. Then assess how well it did on each chunk by labeling the quality of the outcome as either "High", "Ok", or "Poor" quality in column F of your spreadsheet.

Reflect and submit. Finally, reflect on the individual activity. Submit your responses to the below questions directly on the Google Form along with your **spreadsheet** of prompt iterations and results. **IMPORTANT: MAKE SURE TO GIVE US PERMISSIONS TO ACCESS DOCUMENTS**

1. Please provide a one paragraph reflection of the prompting techniques you used and your iterative process. What aspects of your prompt(s) did you change in each iteration? What was the result?

2. Write a one paragraph reflection on what you thought was particularly effective in improving the quality of your outputs.

3. Write one paragraph reflection on what your final outputs are still failing in.

Team Assignment Instructions

In-section peer review of prompts. 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 prompts and outcomes (in their LastMile AI and results spreadsheet), take note of specific techniques they used and how they could have improved. Then:
- Write a two sentence reflection on what your team member did well in. For example, did they use various high techniques that resulted in a high-quality output?
- Write a two sentence reflection on a concrete, specific way in which your partner could have improved their prompt. 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 another copy of the prompt engineering 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.

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.

Iterate as a team. Work together to create a single final team version of the prompt, 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. specific structure of the prompt, different techniques used, etc). Note: you do not have to use something from every individual submission. You should be trying to create a final prompt that produces the highest quality outcomes you can achieve.

In the same copy of the prompt engineering spreadsheet where you compiled your peer reviews, fill in the 'FINAL PROMPT' section under the 'Prompt Results' tab using the final prompt you came up with applied to all **four** of your assigned transcripts (You can get rid of the "EARLY ITERATIONS" section). **Additionally, for the team deliverable, you also should run your prompt on the two transcripts from homework 1.** So your final submission will include Prompt Results of 6 transcripts (4 assigned transcripts + 2 transcripts from HW1).

Reflect and submit. Submit your responses to the below questions directly on the [Google Form](#) along with your spreadsheet of your final prompt and results. **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 prompting techniques to use? What influenced your prompt's structure? 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:
https://docs.google.com/spreadsheets/d/1-4O8NBYEi2cyfCHmPspuptSPHP5bJ_qnZcQ-mVKoZM4/edit?gid=591800225#gid=591800225