Project Workflow: Faculty Feedback Categorization & Analysis

1. Data Collection and Preparation

Step 1.1 – Organize Feedback by Professor

- Gathered all written responses from faculty.
- Created separate files for each **professor**, keeping data modular and manageable.

Step 1.2 – Divide Responses by Student

- Within each professor file, split responses for each **student**.
- Used simple numeric labels (e.g. 1, 2, 3) to distinguish students' feedback sections.

2. Structuring the Data for Processing

Step 2.1 – Convert to JSON Format

- Transformed each file into structured JSON objects, adding:
    - `professor_id`
    - `student_id`
    - `sentence_id`

Step 2.2 – Sentence Segmentation

- Split each student's feedback into individual **sentences**.
- Assigned each a unique `sentence_id`, tied to `student_id`, to prepare for granular classification.

3. Iterative Workflow Using Gemini + GPT-4o

Throughout this process, **multiple sessions with ChatGPT-4o** were used to:

- Plan, test, and refine each stage
- Troubleshoot edge cases
- Script transformations, merges, and category alignment

---

4. Concern vs Benefit Classification (Gemini 2.5)

Step 4.1 – Determine if Sentence = Concern or Benefit

- Used Gemini 2.5 to classify each sentence into one of two high-level sentiments:
  - **Concern**
  - **Benefit**

Step 4.2 – Split Files Accordingly

- Created two distinct files: one for **concern-related** sentences, one for **benefit-related** sentences.
- This reduced token load and helped focus categorization prompts.

---

5. Category Assignment Using Survey-Derived Labels

Step 5.1 – Use Separate Prompts for Each Category Set

- **Prompt for Concerns:**

  "With this file, it is only sentences that have been marked 'concern'. Assign the most relevant category from the list of concerns to each `sentence_id` please. If you cannot make a determination, return a blank array."

  Concern Categories:

  - Promotes Academic Misconduct

    - Prevalence of AI Errors or “Hallucinations”
    - Increased Bias and Stereotypes
    - Negative Environmental Impacts
    - Ignores Intellectual Property Rights
    - Limits my Voice, Creativity and/or Critical Thinking

- **Prompt for Benefits:**

  “With this file, it is only sentences that have been marked ‘benefit’. Assign the most relevant category from the list of benefits to each `sentence_id` please. If you cannot make a determination, return a blank array.”

  Benefit Categories:

    - Saves Time and Effort
    - Improves Accuracy and Reduces Human Error
    - Promotes Diverse Perspectives and Reduces Bias
    - Supports Environmental Sustainability
    - Promotes Citing Sources
    - Boosts my Voice, Creativity, and/or Critical Thinking

Step 5.2 – **Assign Categories to Each Sentence**

- Gemini returned one category per sentence when confident, or a blank array when not.
- This enabled clean downstream filtering and analysis.

---

6. Emotional Intensity Tagging

Step 6.1 – **Evaluate Affective Tone per Sentence**

- Used Gemini to label each sentence with **Emotional Intensity (EI)**:
    - `Neutral`

- `Mild`
- `Strong`

This provided a qualitative lens to complement categorical classification, showing how strongly professors felt about particular issues or values.

---

7. Filtering, Formatting, and Cleanup

Step 7.1 – **Remove Unlabeled Sentences**

- Any sentence that lacked a valid category label was filtered out to maintain a focused dataset.

Step 7.2 – **Convert Final Data to CSV**

- JSON files (with metadata, categories, and EI labels) were converted into CSV format for analysis and visualization.

---

8. Merging and Visualization

Step 8.1 – **Re-Merge by Professor Using GPT-4.1**

- Once categorized and enriched, JSON files were recombined per professor for context preservation.

Step 8.2 – **Create a Master JSON File**

- All individual professor JSONs were merged into a single master dataset for cohort-level insights.

Step 8.3 – **Upload to ChatGPT-4o for Charts**

- Uploaded the final master JSON to GPT-4o for:
  - CSV transformation (if needed)
  - Chart generation (e.g. category frequency, EI distributions)

---

9. Sharing and Collaboration

Step 9.1 – Distribute Final Outputs

- Shared the CSVs and charts with coworkers and collaborators.
- Files were cleaned, labeled, and ready for further synthesis or reporting.
