Natural Language Processing
CS - 429
Assignment Report

Group 3
Parth Jha - 2021B3A71042P
Ojas Jain - 2021B3A70976P

Introduction

This assignment explores the use of **natural language processing (NLP) techniques** to improve access to welfare schemes by converting **textual descriptions into visually descriptive image prompts**. The objective is to make **welfare scheme information more accessible through images**, bridging language barriers that may hinder understanding.

The project involves several key steps:

1. Extracting key information from welfare scheme texts
2. Generating descriptive image prompts based on the extracted information
3. Ensuring consistency across the generated images by maintaining entity consistency
4. Evaluating the quality and accuracy of the generated images based on coherence and consistency

For the initial text summarization task, two approaches were explored - using a **large language model (LLM)** to generate written summaries, and leveraging a **transformer-based T5 model**. The performance of these two summarization methods was compared using **BLEU and ROUGE scores**. The results indicate that the **LLM-generated summaries outperformed the transformer model**, showing better alignment with the reference texts.

To further improve the **text-to-image generation capabilities**, the project will draw insights from the blog post "https://ar5iv.labs.arxiv.org/html/2109.06977" and the research paper "Dynamic Prompt Optimizing for Text-to-Image Generation". These resources provide guidance on **enhancing prompt engineering and generation techniques**, which will be incorporated in the final submission to **generate higher-quality and more consistent image prompts** from the welfare scheme texts.

The goal is to create an end-to-end system that can effectively **extract key information from textual descriptions**, **generate corresponding image prompts**, and **produce visually descriptive outputs** that make **welfare scheme details more accessible** to a wide range of users, regardless of their language proficiency.

Methodology

Information Extraction

Collection of Data

In the **Information Extraction** section, we collected information on **ten government schemes**, using reliable sources such as **official government websites, policy documents**, and **reputable news publications**. By focusing on aspects like **eligibility, benefits, target population**, and **financial provisions**, we ensured the **completeness** and **relevance** of the collected data. Key sources included the Government of India's **Ministry websites** (e.g., Finance, Rural Development, Health and Family Welfare) for official details, along with **policy brochures and FAQs** for further insights. Additionally, **news outlets** such as The Economic Times and the Press Information Bureau provided supplementary information and helped confirm the broader public context and reception of these schemes.

Each scheme's details were stored in individual text files named **scheme1.txt** through **scheme10.txt**, with each file containing information specific to one scheme. This structured approach facilitated a systematic analysis in later stages of the assignment, enabling efficient **sentiment analysis, frequency comparisons**, and **information retention checks** through code-based methods.

Generating Write-ups with LLMs

The process of generating write-ups for each scheme was automated using the **Google Gemini 1.5 Flash API**. For each scheme, three structured write-ups were generated: one detailing the **beneficiary and problem statement**, another outlining the **application process and benefits**, and the third explaining the **outcome and impact**. The code reads scheme details from text files (e.g., scheme1.txt, scheme2.txt) and sends prompts to the API to generate relevant content based on the provided scheme information. The resulting write-ups are then saved in new files with the suffix _llm_writeup.txt.

This approach streamlines the extraction of critical information from the schemes, ensuring consistent and concise summaries without manual intervention. The use of the **Google Gemini 1.5 Flash model** helps in crafting informative and structured content, making it easier to analyze and present the schemes' details in a standardized format.

Generating Write-ups with Transformers

For generating summaries of welfare scheme texts, the **T5-base transformer model** from Hugging Face's **Transformers library** was utilized. This approach processes files named like scheme1.txt, scheme2.txt, and so on. For each scheme, three specific write-ups are generated: **Beneficiary and Problem Statement**, **Application Process and Benefits**, and **Outcome and Impact**. The model is initialized using the **T5 tokenizer** and **T5ForConditionalGeneration** to prepare the data and generate concise summaries.

The summaries are generated by feeding the scheme content into the T5 model with appropriate prompts tailored to each section of the write-up. The output summaries are then saved in files with the suffix _t5_writeup.txt. This method ensures that the generated content is well-structured and relevant to the specific aspects of the welfare scheme, providing a standardized summary format for further analysis.

Prompt Generation

In our project, we employed **Large Language Models (LLMs)** to create prompts tailored for different presentation styles. Specifically, we generated **two types of prompts** for each scheme description:

1. **Story-Based Prompts**: These prompts are crafted in a narrative style to capture attention by depicting scenarios in a storytelling format. They are designed to engage readers by describing characters, settings, and situations in a relatable manner. The story-based prompts are stored in **results.txt**.

2. **Descriptive-Based Prompts**: These prompts focus on clarity and are structured for use in **informative posters**. They emphasize key details and relevant quantitative data, making it easy to understand important aspects such as scheme benefits, application processes, and expected outcomes. The descriptive-based prompts are stored in **resultsllm.txt**.

Using LLMs allowed us to generate prompts that are both engaging and informative, ensuring they effectively meet the needs of diverse presentation contexts. The two prompt types offer flexibility, with story-based prompts designed for a compelling narrative appeal and descriptive prompts optimized for clear, factual communication.

Ensuring Entity Consistency

To ensure this consistency, we developed a Python function that replaced each entity in the prompts with its designated identifier. We designed three prompt templates to describe different aspects of each scheme: the beneficiary's problem, the benefits of applying to the scheme, and the positive outcomes achieved through it. Each prompt referenced the identified entities consistently, ensuring that elements such as the farmer's struggles, the application process, and the final benefits aligned seamlessly across the descriptions. This approach, supported by LLMs for more flexible prompt generation, resulted in a set of coherent, entity-consistent prompts that provided a clear and unified narrative for each welfare scheme.

Image Generation

For generating images, we used the DALL-E model, manually entering each prompt to create visuals aligned with our objectives. We experimented with two types of prompts: story-based prompts from `results.txt` and descriptive, informative prompts from `resultsllm.txt`. By testing both prompt types, we aimed to assess their effectiveness in producing images that effectively communicated the themes of each welfare scheme.

Our findings indicated that the descriptive, informative prompts in `resultsllm.txt` were more successful in generating clear and informative posters. These prompts provided structured information and context, allowing DALL-E to focus on key elements of the scheme, resulting in visuals that were more aligned with our goal of informative representation. The story-based prompts, while engaging, did not achieve the same level of clarity in conveying specific details. Thus, we opted to prioritize the informative prompts for our final images, which better captured the essence of each scheme in a visually accessible way.

The generated images are presented later in this report, where they illustrate the results of our prompt design approach and the DALL-E model's interpretation of the welfare schemes.

Self - Evaluation

Write-up Evaluation

Manual Evaluation

Manual evaluation was performed to ensure the accuracy of the extracted information. We cross-checked key details such as **objectives**, **name of scheme**, **eligibility criteria**, and **expected outcomes** by manually reading the content of the schemes. This evaluation helped in confirming that the automated summaries aligned with the original scheme details and included all relevant information.

For example, in **Scheme 1** and **Scheme 2**, we carefully reviewed the following:

Objectives: ✔️**Name of Scheme**: ✔️ **Eligibility Criteria**: ✔️**Expected Outcomes**: ✔️

By verifying these aspects through manual inspection, we ensured that the content provided in the summaries was accurate and comprehensive. This manual review served as an additional layer of validation to maintain the quality and reliability of the generated write-ups.

Evaluation of Sentiment and Word Frequency Consistency

Sentiment Analysis

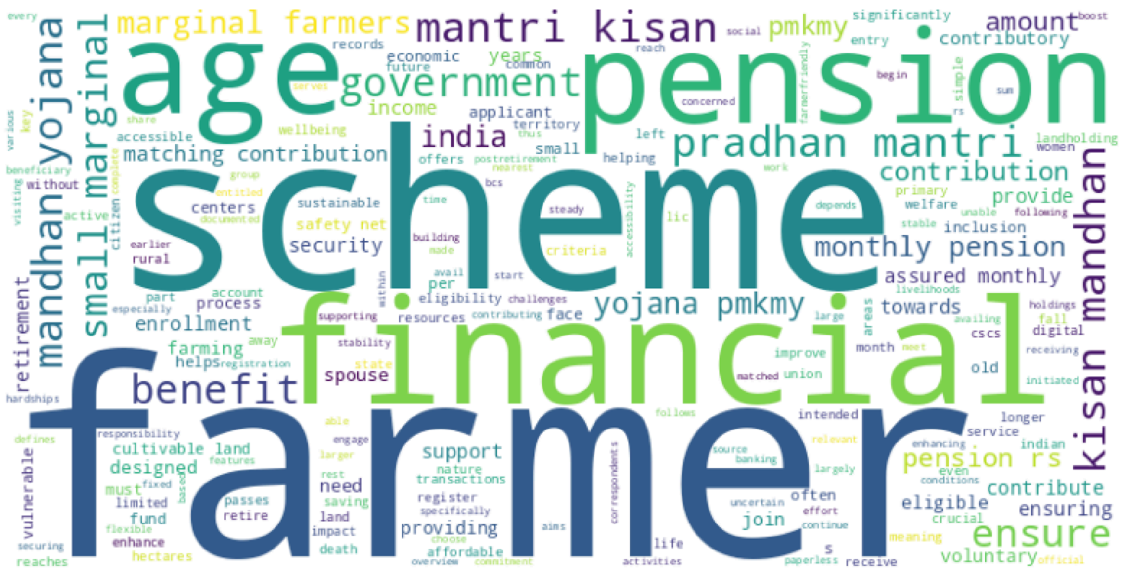
Sentiment analysis showed that the polarity (positive/negative tone) of the writeups was generally lower than the originals, indicating a slight decrease in positivity. Subjectivity values were also lower, suggesting a more neutral, objective tone in the writeups.

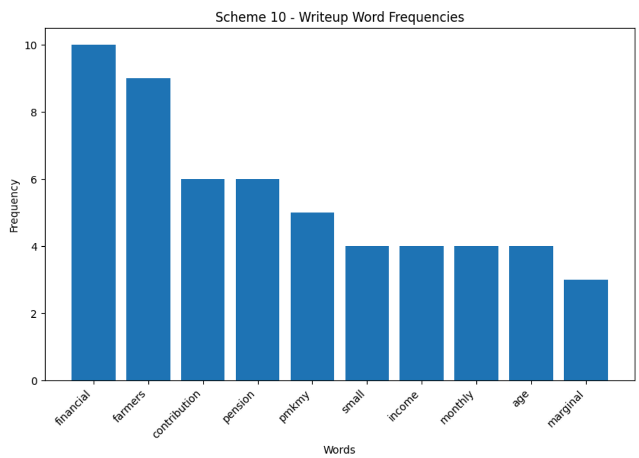
Word Frequency Analysis

Key themes such as "financial," "farmers," and "scheme" were consistently highlighted in both the original texts and the writeups. While the word "farmers" appeared slightly more in the originals, the writeups maintained this emphasis with a more formal structure.

Results and Visualizations

Results and visualizations, including bar charts and word clouds, comparing the original texts with their writeups, are available in the Python notebook. These confirm that the writeups preserve the core message with a refined, formal tone. Examples of these visualizations are shown below.





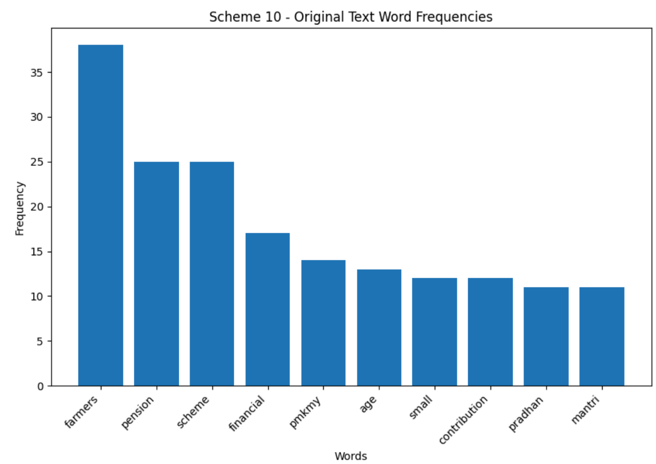


Image Evaluation

In evaluating the generated images from prompts, we manually assessed the coherence and consistency of the outputs to ensure their quality and alignment with the input instructions. Coherence was measured by how well the elements within each image fit together logically, ensuring that all aspects of the prompt were represented accurately without contradictions. For example, if the prompt specified a scene with specific objects or actions, the evaluation considered whether these elements were present and interacted in a believable way.

Consistency, on the other hand, was evaluated by examining whether the generated images adhered to the same visual style, structure, and interpretation across multiple variations of the same prompt. This involved checking if similar prompts produced consistent visual outcomes, with consistent attributes such as color schemes, object placement, and overall theme. By combining these two metrics, we were able to qualitatively assess the quality of the generated images, ensuring that they were both visually coherent and consistent across different prompts.

Phase 2

Automatic Prompt Engineering

This phase 2 improvement has been inspired from “https://arxiv.org/abs/2211.01910”.

In our project, we implemented **Automatic Prompt Engineering (APE)** using **Google's Gemini model** to enhance the precision and effectiveness of image generation prompts. APE serves as an **AI-driven refinement system** that transforms basic prompts into structured, detailed instructions that better communicate creative intent to image generation models.

The code employs **Google's Gemini API** to perform a comprehensive analysis, breaking down the prompt into **distinct categories** and enhancing each with specific details. This structured approach ensures that all critical elements - from **technical specifications** to **artistic intentions** - are clearly defined.

Example (data/llm_data/scheme2_llm_writeup)

Title: The Plight of Informal Traders and Shopkeepers: The Promise of the National Pension Scheme**

Category	APE-Enhanced Details
Main Subject	- Elderly shopkeeper (60s) ✦
- Weary expression ✦
- Sitting position specified ✦
Style Elements	- Photorealistic approach ✧
- Documentary photography style ✧
- Real-life struggle emphasis ✧
Color Treatment	- Earthy tone palette ★
- Strategic accent colors ★
- Logo highlight consideration ★
Composition	- Close-up framing ◆
- Store context included ◆
- Background depth elements ◆
Data Visualization	- 92% coverage statistic ▲
- Rs. 3,000 pension amount ▲
- Government contribution note ▲
Technical Specs	- Precise dimensions (1280x720) ■
- DPI specification (72) ■
- Color space definition (RGB) ■

Prompt Designing

This section of phase 2 is inspired from “https://ar5iv.labs.arxiv.org/html/2109.06977”. However, this portion was applied during phase 1 itself and the results are available in **data/llm_data** folder.

Refer to Appendix A of the given paper for more information.

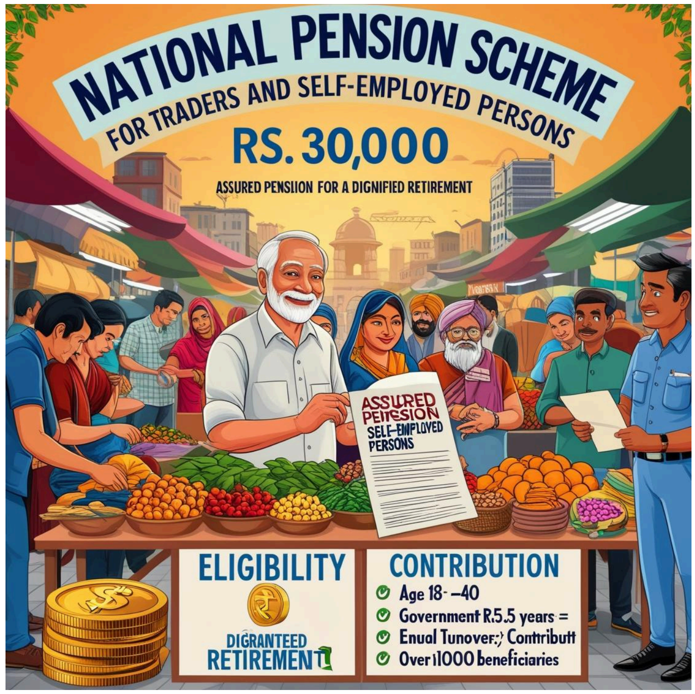
<u>Example Improvements in Image Generation</u>

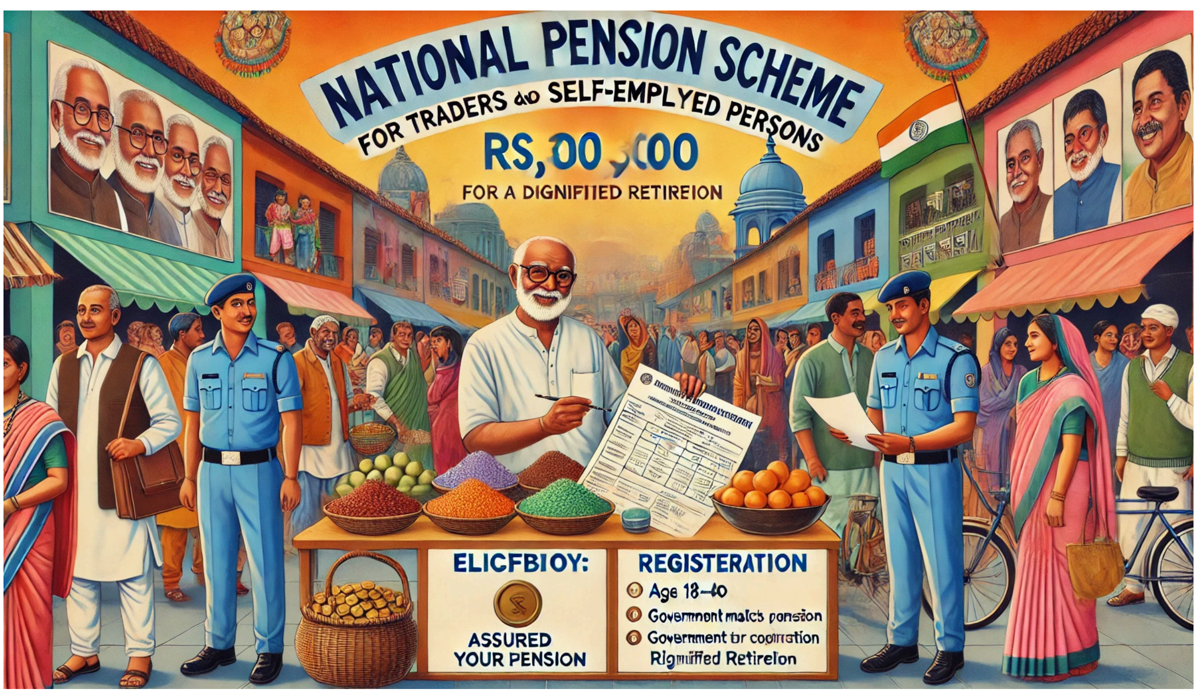
Phase II

Phase I

Improvements->Images generated in the phase 2 depicts clear text in comparison to the image generated in phase 1.The image generated in the phase 2 has is more descriptive , just like an informative poster(containing pension amount etc).Also the first image gives a good street view.

Phase II





Phase I

Improvements->Images generated in the phase 2 depicts clear text in comparison to the image generated in phase 1.The image generated in the phase 2 is showing 30000 value clearly (not in phase

1 image).Also the first image has all the information in alignment with the prompt,nothing fake deceptive or fake.

Phase II

Phase I

Improvements->Images generated in the phase 2 depicts clear text in comparison to the image generated in phase 1.The image generated in the phase 2 has is more descriptive , just like an informative poster(containing scheme name and benefits etc).Also the first image gives a good representation of the family with human look alike characters instead of animated characters as in phase 1 image.

Challenges and Solutions

Challenge: Scene Specification for Visual Accuracy
Solution: The prompt generation template was modified to include explicit scene descriptions, such as the setting and beneficiary expressions. This enhancement ensured that the image generation model had precise guidance for each visual element.

Challenge: Model Limitations in Image Generation
Solution: Where the model struggled to generate precise imagery, we iteratively refined prompts and adjusted parameters to achieve closer visual alignment with the intended scene.

Challenge: Effectiveness of Prompt Types
Solution: Through experimentation, we found that informative prompts, rather than narrative-style story prompts, produced better results. We refined our prompt structure by explicitly dividing the information into clear categories, such as scene, background, and key details. This approach provided the model with more focused and structured guidance, improving the accuracy and clarity of the generated images. By breaking down the prompt into specific components, we enabled the model to better understand the visual elements and their relationships, resulting in more precise and consistent image generation.

Conclusion

This project successfully explored the use of **NLP and image generation models** to make **welfare schemes more accessible** through **visual prompts**. By extracting key information from scheme texts and generating **structured, informative prompts**, we found that these prompts led to **better image quality and clarity** compared to narrative-based ones. **Manual evaluation and sentiment analysis** ensured the generated content was **accurate and consistent** with the original schemes.

In **Phase 2**, we introduced **Automatic Prompt Engineering (APE)** using **Google's Gemini model** to refine image generation prompts. APE enhanced the **precision of prompts** by breaking them down into **structured categories**, adding detail to each element, and ensuring better communication of **creative intent**. This resulted in images that were more **descriptive**, **clear**, and **aligned** with the scheme's objectives.

Key findings include that **informative, APE-enhanced prompts** produced more **precise and consistent images**, effectively addressing challenges like **scene specification** and **model limitations**. The improved images clearly depicted elements such as **pension amounts** and **scheme benefits**, ensuring better representation of the intended message. This approach demonstrated that combining **NLP techniques with image generation** can significantly improve the **accessibility and clarity** of welfare scheme information, making it easier for a wider audience to understand and engage with the material.