Weekly Project Report

Name: Anbhi Thakur
Date: 26 July 2025
Week: July 20 – July 26, 2025
Project Focus: Text to Prompt Compression – Research, Documentation, and Implementation

---

Summary

This week was dedicated to deep-diving into the concept, methodology, and practical implications of Text to Prompt Compression (TPC), a technique aimed at optimizing input prompts for large language models. The work involved literature research, experimentation with prompt engineering tools, and the drafting of a comprehensive research report.

Objectives for the Week

1. Understand the theoretical foundation of TPC
2. Analyze token economy and efficiency concerns
3. Explore and test TPC techniques (manual and automated)
4. Document findings in a well-structured and human-centric format
5. Create a plagiarism-free final report

Tasks Completed

1. Literature Review

- Studied foundational papers on prompt engineering and tokenization.
- Explored works by OpenAI, Hugging Face, and academic resources related to NLP compression and summarization.
- Compared classical summarization with prompt compression.

2. Tool Testing

- Used OpenAI's API to test compression accuracy and token count reduction.
- Explored spaCy for keyword extraction and phrase simplification.
- Evaluated LangChain's prompt optimization workflows.

3. Methodology Design

- Designed a hybrid methodology combining:
    - Abstractive summarization
    - Keyword filtering
    - Template reengineering
    - Controlled human rewriting

4. Case Studies & Applications

- Developed sample scenarios for:
    - Chatbot improvement
    - Legal text summarization
    - Academic note condensation
    - Customer service automation

5. Final Report Drafting

- Composed the complete report titled **"Text to Prompt Compression Report"**, formatted professionally with:
    - Abstract
    - TOC
    - 12 Sections (Intro to References)
    - Humanized, unique, and plagiarism-free content

Insights Gained

- Effective prompt compression significantly reduces inference costs without compromising performance.
- Maintaining semantic integrity requires balancing brevity with contextual depth.
- Human-in-the-loop models enhance compression quality in high-risk domains.
- Language models respond better to structured, specific prompts than verbose or vague ones.

Challenges Faced

- Avoiding over-compression that may lead to loss of intent
- Ensuring unbiased output after rewriting compressed prompts
- Achieving 100% originality in content while maintaining technical accuracy

---

Tools & Technologies Used

- OpenAI GPT-4
- spaCy
- LangChain Prompt Optimization Toolkit
- Grammarly + LanguageTool (for tone, grammar check)
- Microsoft Word & Markdown Editors

Upcoming Week Plan

1. Develop a small Streamlit app for Text to Prompt Compression demo
2. Integrate summarization and prompt testing APIs
3. Explore multilingual TPC frameworks
4. Prepare a 5-minute video explanation of TPC
5. Collect user feedback on prompt clarity and compression quality

---

Overall Reflection

This week was a breakthrough in understanding the synergy between human communication and AI interpretation. Compressing text into prompts is more than just a technical task—it's about translating intent into action, efficiently. The exercise brought forward the importance of clarity, minimalism, and linguistic intuition when dealing with powerful language models.