Assignment 4 Report

4.1: Prompt Design and Comparison

Test Query

During the 2025 Monaco Grand Prix, 19-year-old rookie driver Luca Martelli stunned the Formula 1 world by securing a surprise podium finish for the mid-tier team, Arden Racing. Starting from 12th on the grid, Martelli skillfully navigated the tight street circuit and capitalized on a late safety car period to move up the field. His performance has drawn praise from fans and analysts alike, marking the best result in Arden's history and positioning Martelli as a rising star in the sport.

Direct Prompt

Prompt: "Summarize the following article in 2–3 sentences."

Output:
Luca Martelli, a 19-year-old rookie, secured a surprise podium finish at the 2025 Monaco Grand Prix for Arden Racing after starting 12th and taking advantage of a late safety car. The performance marks the team's best-ever result and has earned Martelli widespread recognition. Analysts view him as a promising new talent in Formula 1.

Few-Shot Prompt

"Here are some examples of news articles and their one-sentence summaries:

Example 1:
Article: 'The local city council has approved a new zoning plan...'
Summary: 'A new shopping mall plan was approved by the city council, sparking debate over increased traffic versus economic benefits.'

Example 2:
Article: 'A rare celestial event, a total lunar eclipse, will be visible...'
Summary: 'North America will witness a rare total lunar eclipse tonight, expected to be the clearest in decades for optimal viewing.'

Task:
Summarize the following news article into one concise sentence:

"During the 2025 Monaco Grand Prix, 19-year-old rookie driver Luca Martelli stunned the Formula 1 world by securing a surprise podium finish for the mid-tier team, Arden Racing. Starting from 12th on the grid, Martelli skillfully navigated the tight street circuit and capitalized on a late safety car period to move up the field. His performance has drawn praise from fans and analysts alike, marking the best result in Arden's history and positioning Martelli as a rising star in the sport.""

Output:
Rookie driver Luca Martelli earned a surprise podium at the 2025 Monaco Grand Prix, giving Arden Racing its best-ever result and highlighting his rising potential in Formula 1.

Chain-of-Thought Prompt

Prompt:
"Summarize the following news article into one concise sentence. First, identify the main subject, then the key achievement, and finally, the potential impact."

Output:

- **Main subject:** Rookie F1 driver Luca Martelli
- **Key achievement:** Secured a surprise podium at the 2025 Monaco Grand Prix for Arden Racing
- **Potential impact:** Boosts team morale and establishes Martelli as a rising star in the sport

Concise summary:
Rookie Luca Martelli's surprise podium at Monaco elevated Arden Racing to its best-ever result and established him as a future star in Formula 1.

4.2: Prompt Tuning Experiment

Task Chosen: Sentiment Analysis

Domain: Formula 1 News & Fan Reactions

Objective: Improve the accuracy and consistency of sentiment classification in texts related to Formula 1 (e.g., driver/team performance reviews, fan comments, race reports).

Initial Prompt (Baseline)

Prompt:

"Analyze the sentiment of the following text: '{text…}'. Is it Positive, Negative, or Neutral?"

Limitations Observed:

- The model often gave vague or inconsistent outputs.
- Struggled with domain-specific tone (e.g., sarcasm, technical feedback).
- Ambiguity arose from lack of clear sentiment definitions or output constraints.

Prompt Tuning Method 1: Adding Constraints and Specific Output Format

Improved Prompt:

"Analyze the sentiment of the following Formula 1-related text and classify it strictly as 'Positive', 'Negative', or 'Neutral'.

- Use **'Positive'** for clearly favorable reactions, praise, or strong approval (e.g., great performance, impressive strategy).

- Use **'Negative'** for clearly unfavorable reactions, criticism, or disappointment (e.g., crashes, poor pit stops, team failures).
- Use **'Neutral'** for factual commentary, mixed opinions, or objective statements without clear emotional tone.

Improvements Observed:

- Reduced ambiguity by anchoring sentiment to specific emotional cues.
- Improved handling of objective race summaries and factual data.
- Still struggled with nuanced or sarcastic F1 fan language.

Prompt Tuning Method 2: Iterative Refinement with Domain-Specific Few-Shot Examples

Final Improved Prompt with F1 Contextual Examples:

"Analyze the sentiment of the following Formula 1-related text and classify it strictly as 'Positive', 'Negative', or 'Neutral'.

- **Positive:** Clearly favorable sentiment, such as praise for drivers, teams, strategies, or race results.

- **Negative:** Clearly unfavorable sentiment, including criticism, disappointment, or poor performance.

- **Neutral:** Objective, factual, or mixed statements without a strong emotional leaning.

Examples:

- Text: 'Verstappen absolutely dominated today's race. Flawless performance!'
  Sentiment: Positive

- Text: 'Ferrari messed up the strategy again. How many times will they do this?'
  Sentiment: Negative

- Text: 'Lewis Hamilton qualified 4th and will start behind Norris and Sainz.'
  Sentiment: Neutral

- Text: 'McLaren’s pace was promising, though tire wear was a concern.'
  Sentiment: Neutral

- Text: 'Terrible pit stop from Red Bull. That cost them the race win.'
  Sentiment: Negative

- Text: 'Alonso is proving why he's still one of the best — amazing drive!'
  Sentiment: Positive

Now analyze the following text:
Text: '{text…}'"

Benefits of F1-Specific Tuning

- Significantly improved classification in edge cases like technical analysis, sarcastic tweets, or neutral commentary.

- Domain knowledge (e.g., team history, race dynamics) adds contextual clarity to sentiment detection.

- Better alignment with how F1 fans, commentators, and journalists express sentiment.

4.3: Ethical Issues in Large Language Model (LLM) Use

The rapid advancement and widespread adoption of Large Language Models LLMs) have brought forth significant ethical challenges that demand careful consideration. While LLMs offer immense potential for innovation and efficiency, their deployment raises concerns across several critical domains, notably bias, fairness, and privacy. Addressing these issues is paramount to ensuring responsible and equitable development and use of AI

1. Bias in LLM Outputs

Issue:

LLMs often reflect societal, cultural, or historical biases present in the data they are trained on. This can result in:

- Gender or racial stereotypes in responses
- Unequal representation or offensive content
- Reinforcement of harmful social norms

Example:

A model may associate "nurse" with women and "engineer" with men due to biased training data.

How to Address It:

- **Diverse and representative training data:** Include data from varied cultures, backgrounds, and perspectives.

- **Bias detection tools:** Implement algorithms to audit and flag biased outputs.

- **Human oversight:** Involve domain experts to review and refine sensitive responses.

- **Post-training techniques:** Use fine-tuning or reinforcement learning to reduce bias in outputs.

2. Fairness in Access and Outcomes

Issue:

LLMs may unintentionally favor certain groups over others, leading to unfair treatment or access barriers. This can occur in:

- Automated hiring systems
- Education or healthcare assistance
- Legal or financial recommendation engines

Example:

An LLM-based resume screener might favor applicants from prestigious schools, disadvantaging those from underrepresented communities.

How to Address It:

- **Transparent algorithms:** Ensure users understand how decisions are made.
- **Inclusive design practices:** Involve stakeholders from diverse communities during development.
- **Regular fairness audits:** Continuously evaluate system performance across demographic groups.

3. Privacy Concerns

Issue:

LLMs trained on vast internet data can inadvertently memorize and regurgitate sensitive or personally identifiable information (PII).

Example:

A model might leak a user's email address or credit card number if such data was present in the training set.

How to Address It:

- **Data filtering:** Remove PII and sensitive content from training datasets.
- **Differential privacy:** Apply privacy-preserving techniques during model training.
- **Redaction tools:** Develop mechanisms to detect and prevent disclosure of private data during inference.