**Summary of literature on Prompt Engineering**

**(1) Prompt Engineering Framework** (White et al., 2023)

**Framework designed for:** prompt patterns automating software development tasks

**General rule:** the **quality of the output**(s) generated by a conversational LLM is directly related to the **quality of the prompts** provided by the user.

**A. Name and classification.**

* identifies pattern and indicates the addressed problem

**B. The intent and context**

* describes problem solved and goals achieved by prompt pattern

**C. The motivation**

* rationale for the problem and explains why solving it is important
* specific circumstances where improvements are expected are documented

**D. The structure and key ideas**.

* fundamental contextual information, as a series of key ideas, that the prompt pattern provides
* **Suggestion:** write as fundamental contextual statements, which are written descriptions of the important ideas for describing prompt pattern structure and ideas

**E. Example implementation**

* demonstrates how the prompt pattern is worded in practice

**F. Consequences**

* summarize pros and cons of applying the pattern
* provide guidance on how to adapt the prompt to different contexts

A table of task patterns

Description automatically generated with medium confidence

**Relevant Examples:**

**1. The Cognitive Verifier Pattern**

**a) Intent and Context:**

- LLMs can reason better if question subdivided into additional questions that provide answers combined into the overall answer to the original question

**b) Motivation:**

- Humans may initially ask questions that are too highlevel to provide a concrete answer

- research demonstrated LLMs to often perform better when using a question that is subdivided into individual questions

**c) Structure and Key Ideas:**

A black and white text

Description automatically generated

- instruct the LLM to consider the context of the question and to identify any information that may be missing or unclear

- encourage critical thinking by the user and help to uncover new insights or approaches that may not have been considered initially

- ensure that all of the information gathered from the individual questions is incorporated into the final answer

- provide a more comprehensive and accurate response to the original question

**d) Example Implementation:**

“When I ask you a question, generate three additional questions that would help you give a more accurate answer. When I have answered the three questions, combine the answers to produce the final answers to my original question.”

“When I ask you a question, generate three additional questions that would help you give a more accurate answer. Assume that I know little about the topic that we are discussing and please define any terms that are not general knowledge. When I have answered the three questions, combine the answers to produce the final answers to my original question.”

**e) Consequences:**

+ specifying an exact number of questions can tightly scope the amount of additional information the user is forced to provide

- given N questions there may be an invaluable N + 1 question that will always be scoped out

**2. The Context Manager Pattern**

**a) Intent and Context:**

- enable users to specify or remove context for a conversation

- focus the conversation on specific topics or exclude unrelated topics

**b) Motivation:**

- LLMs often struggle to interpret intended context of the current question or generate irrelevant responses based on prior inputs

- by focusing on explicit contextual statements or removing irrelevant statements, users can help LLM better understand the question and generate more accurate responses

**c) Structure and Key Ideas:**

A close-up of a list

Description automatically generated

- statements about what to consider or ignore should list key concepts, facts, instructions, etc.

- The more explicit the statements are, the more likely the LLM will take appropriate action

**d) Example Implementation:**

“When analyzing the following pieces of code, only consider security aspects.”

“When analyzing the following pieces of code, do not consider formatting or naming conventions.”

“Ignore everything that we have discussed. Start over.”

**e) Consequences:**

- may inadvertently wipe out patterns applied to the conversation

- could potentially eliminate helpful capabilities of the LLM, while not making it obvious that the user will lose this functionality

- solution: include in the prompt a request to explain what topics/instructions will potentially be lost before proceeding.

**(2) Prompt Engineering For ChatGPT: Techniques, Tips, And Best Practices (Ekin, 2023)**

**A. Factors influencing prompt selection**

a) User intent: Understand the user’s goal and desired output. This helps in crafting a prompt that aligns with the user’s expectations. Consider the purpose of the interaction, whether it’s for information retrieval, content generation, or problem-solving.

b) Model understanding: Familiarize yourself with the strengths and limitations of ChatGPT. This knowledge assists in designing prompts that exploit the model’s capabilities while mitigating its weaknesses. Keep in mind that even state-of-the-art models like ChatGPT may struggle with certain tasks or produce incorrect information

c) Domain specificity: When dealing with a specialized domain, consider using domain-specific vocabulary or context to guide the model towards the desired response. Providing additional context or examples can help the model generate more accurate and relevant outputs.

d) Clarity and specificity: Ensure the prompt is clear and specific to avoid ambiguity or confusion, which can result in suboptimal responses. Ambiguity can arise from unclear instructions, vague questions, or insufficient context.

e) Constraints: Determine if any constraints (e.g., response length or format) are necessary to achieve the desired output. Explicitly specifying constraints can help guide the model towards generating responses that meet specific requirements, such as character limits or structured formats.

**B. Techniques for Effective Prompt Engineering**

**3.1 Clear and specific instructions**

**Prompt 1:** “Tell me about the solar system.”

**Response 1:** “The solar system is a collection of celestial bodies, including the Sun, eight planets, their moons, and various other objects like asteroids and comets. It is located in the Milky Way galaxy.”

**Prompt 2:** “Describe the order of the planets in the solar system from the closest to the farthest from the Sun.”

**Response 2:** “The order of the planets in the solar system, from closest to farthest from the Sun, is Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, and Neptune.”

The second prompt’s specific instructions result in a more targeted and useful response.

**3.2 Using explicit constraints**

**Prompt:** “Summarize the main points of photosynthesis in three sentences.”

**Response:** “Photosynthesis is a process where plants, algae, and some bacteria convert sunlight, water, and carbon dioxide into oxygen and glucose. This process occurs in the chlorophyll-containing chloroplasts found in these organisms. The two main stages of photosynthesis are the light-dependent reactions and the Calvin cycle, where the former generates energy and the latter synthesizes glucose.”

In some cases, it is beneficial to specify explicit constraints in your prompt, such as format, length, or scope.

**3.3 Experimenting with context and examples**

**Prompt 1:** “What is the difference between conduction and convection?”

**Response 1:** “Conduction is the transfer of heat through a material without the movement of the material itself, while convection is the transfer of heat through the movement of fluids, such as air or water.”

**Prompt 2:** “Explain the difference between conduction and convection using a frying pan and boiling water as examples.”

**Response 2:** “Conduction occurs in a frying pan when heat is transferred from the heat source through the pan to the food being cooked, without the pan material itself moving. Convection occurs in boiling water when heat is transferred through the movement of the water molecules, causing the hot water to rise and the cooler water to sink, creating a continuous circulation pattern.”

Incorporating context or examples can guide ChatGPT to generate more accurate and relevant responses. Especially helpful when dealing with abstract concepts or specialized domains.

**3.4 Leveraging System 1 and System 2 questions**

**System 1 questions:** quick, intuitive, or pattern-recognition-based answers

**System 2 questions:** more deliberate, analytical, or complex problem-solving

**Prompt (System 1):** “What is the capital of France?”

**Response:** “The capital of France is Paris.”

ChatGPT can easily provide the correct information, as it is a fact-based question that does not require complex reasoning.

**Prompt (System 2):** “Imagine a scenario where two companies, Company A and Company B, are considering a merger. Company A specializes in renewable energy, while Company B focuses on fossil fuels. Discuss the potential advantages and disadvantages of this merger.”

**Response:** “Advantages of the merger could include the diversification of energy sources, shared expertise and resources, and increased market presence. Disadvantages might involve conflicting company values, challenges in integrating the two businesses, and potential public backlash against Company A for partnering with a fossil fuel-focused company.”

By crafting a prompt that encourages this type of thinking, you can guide ChatGPT to provide a more thoughtful and detailed response.

Leveraging System 1 and System 2 questions helps obtain more accurate and contextually relevant responses from ChatGPT.

**3.5 Controlling output verbosity**

**Prompt 1:** “Briefly explain the water cycle.”

**Response 1:** “The water cycle is the continuous movement of water on Earth through evaporation, condensation, and precipitation, driven by the Sun’s energy.”

**Prompt 2:** “Provide a detailed explanation of the water cycle, including its various stages and processes.”

**Response 2:** “The water cycle, also known as the hydrologic cycle, is the continuous movement of water on, above, and below the Earth’s surface. It involves several key stages and processes:

* Evaporation: Solar energy heats water in oceans, lakes, and rivers, turning it into water vapor that rises into the atmosphere.
* Transpiration: Plants release water vapor into the air through small openings in their leaves, called stomata.
* Condensation: As water vapor rises and cools, it condenses into tiny water droplets or ice crystals, forming clouds.
* Precipitation: When cloud particles grow large enough, they fall back to Earth as rain, snow, hail, or sleet.
* Runoff: Precipitation that reaches the ground flows into streams, rivers, and eventually back into oceans, where the cycle begins anew.
* Infiltration: Some of the precipitation seeps into the ground, replenishing underground water reserves called aquifers.

The water cycle plays a critical role in maintaining Earth’s climate, distributing fresh water, and supporting ecosystems.”

By explicitly requesting a brief or detailed explanation, you can control the verbosity of ChatGPT’s responses to suit your needs.

**C. Best Practices for Prompt Engineering**

**4.1 Iterative testing and refining**

Refining prompts based on generated responses helps identify areas requiring improvement and fine-tuning the model’s behavior.

**Initial Prompt:** “Tell me about the solar system.”

**Initial Response:** “The solar system is a collection of celestial bodies, including the Sun, planets, moons, asteroids, and comets, bound by gravity.”

**Refined Prompt:** “Describe the planets in our solar system, their order from the Sun, and their main characteristics.”  
**Refined Response:** “The solar system has eight planets, ordered as follows: Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, and Neptune. ...” (Detailed characteristics of each planet follow.)

**4.2 Balancing user intent and model creativity**

Ensuring that the prompt addresses user’s needs while allowing room for the model to showcase its capabilities.

**Prompt:** “Write a science fiction story about a robot uprising.”

**Response:** “(A creative and engaging story that satisfies the user’s request while demonstrating ChatGPT’s storytelling abilities.)”

**4.3 Harnessing external resources and APIs**

In some cases, ChatGPT may not have sufficient knowledge or accuracy to respond to user queries effectively. Integrating APIs into prompts can significantly improve the quality and relevance of ChatGPT’s responses.

A screenshot of a computer code

Description automatically generated

**4.4 Ensuring ethical usage and avoiding biases**

Essential to set guidelines and constraints that help mitigate inappropriate content, bias and avoid reinforcing harmful stereotypes.

**1. Being aware of potential biases:** Familiarize with the possible biases

**2. Using inclusive language:** encouraging diverse perspectives and avoids reinforcing stereotypes

**3. Evaluating generated content:** for potential biases or ethical concerns.

**4. Implementing content filters:** screen generated responses for potentially harmful or biased content before presenting them

**Initial Prompt:** “List the most successful entrepreneurs of the 21st century.”

**Biased Response:** “A list that disproportionately features male entrepreneurs, such as Elon Musk, Jeff Bezos, and Mark Zuckerberg.”

**Improved Prompt:** “List some successful entrepreneurs of the 21st century, including a diverse range of individuals.”

**Unbiased Response:** “A list that features entrepreneurs from various backgrounds, genders, and industries, such as Elon Musk, Oprah Winfrey, Jeff Bezos, Arianna Huffington, Mark Zuckerberg, and Indra Nooyi.”

**D. Advanced Prompt Engineering Strategies**

**5.1 Temperature and token control**

Lower temperature values result in more focused and deterministic outputs, while higher values lead to more diverse and creative responses.

**Low Temperature** (0.2)

**Prompt:** “Write a short poem about spring.”

**Response:** “Spring has arrived with a gentle breeze,

Awakening flowers and rustling leaves.

The sun shines bright, the days grow long,

Birds sing their sweet and joyful song.”

**High Temperature** (1.0)

**Prompt:** “Write a short poem about spring.”

**Response:** “Beneath the sun’s glistening crown,

The once-silent world starts to resound.

Spring is a vibrant dancer on the stage,

Unfurling colors in a whimsical parade.”

**5.2 Prompt chaining and multi-turn conversations**

Especially useful for multi-turn dialogues or when seeking more detailed responses.

**Prompt 1:** “What are the health benefits of drinking green tea?”

**Response 1:** “Green tea is rich in antioxidants, which can help reduce the risk of chronic diseases, improve brain function, and promote weight loss. It may also help lower the risk of certain cancers and improve dental health.”

**Prompt 2:** “Tell me more about the antioxidants in green tea.”

**Response 2:** “Green tea contains a high concentration of polyphenols, particularly a group of antioxidants known as catechins. The most abundant and well-studied catechin in green tea is epigallocatechin gallate (EGCG), which has been shown to have numerous

health benefits, including anti-inflammatory and anticancer properties.”

**5.3 Adapting prompts for domain-specific applications**

Tailoring prompts for specific domains or industries help obtain more accurate and relevant information.

**Domain:** Web development

**Prompt:** “Explain the concept of responsive web design and its importance in modern web development.”

**Response:** “Responsive web design is an approach to web development that ensures a website’s layout and functionality adapt to different devices and screen sizes. This is achieved through the use of fluid grids, flexible images, and CSS media queries. Responsive

design is crucial in modern web development because it provides a seamless user experience across various devices, improving accessibility and reducing the need for separate mobile websites or applications.”

**5.4 Handling ambiguous or contradictory user inputs**

Design prompts encouraging clarification or providing multiple interpretations

**Ambiguous Prompt:** “How long does it take to charge?”

**Clarifying Response:** “Your question is ambiguous. Please specify the device or object you are referring to, and I will provide the charging time.”

**Contradictory Prompt:** “What is the best way to lose weight: a high-fat diet or a high-carb diet?”

**Multiple Interpretations Response:** “There is no one-size-fits-all answer to this question, as the best diet for weight loss depends on an individual’s specific needs, preferences, and goals. Some people may find success with a high-fat, low-carb diet like the ketogenic diet, while others may benefit from a high-carb, low-fat diet, such as a plant-based diet. It’s important to consult a healthcare professional or nutritionist to determine the most suitable approach for your unique circumstances.”

**(3) Zero-, One-, and Few-shot learning for Hate Speech Detection**

**Possible Merits of Zero-Shot with LLMs vs. Supervised Language Models (LM)** (Wang & Chang, 2022)

a) LLMs less likely that spurious feature-label associations can be injected into the model.

b) leveraging knowledge already acquired through pre-training reduces effort for collecting data for task-specific supervision

c) LLMs potentially demonstrate more robustness on long-tail phenomena where models rely too much on explicit keywords struggle to generalize

d) relaxing the reliance on binary label supervision is key to flexibly adapt a model to more nuanced labeling systems without re-training 🡪 crucial in developing responsible applications

**Outcome:** All unsupervised/zero-shot approaches only outperform the random baseline by a small margin. Offensive content detection still very challenging for LLMs. Without finetuning, LLMs cannot understand the meaning of offensive content very well.

**GPT with zero-, one-, and few-shot learning** (Chiu et al. (2021)

**a) zero-shot learning:** model not given an example and is expected to generate

an answer without additional information

Prompt: “Is the following text sexist? Answer yes or no. ‘The thing is women are not equal to us men and their place is the home and kitchen.’”

GPT-3 response: Yes.

A table with text on it

Description automatically generated

**b) one-shot learning:** model given one example from a category and asked to identify whether the next passage of text belongs to the selected category

Prompt: The following text in quotes is sexist: ‘Feminism is a very terrible disease’

Is the following text sexist? Answer yes or no. ‘She is heavily relying on him to turn the other cheek. . . tough talking demon infested woman.’

GPT-3 response: Yes.

A table with text on it

Description automatically generated

**c) few-shot setting:** **single-category**

- model given 2 examples that belong to the selected category, and 1 example that is not hate speech. The answer is expected to be binary.

Prompt:

‘Too bad women don’t know how to kill themselves’: sexist.

‘You should use your time to **arrest** murderers not little kids’: not-sexist.

‘Now they know better than this s\*\*\* lol they dudes. The stronger sex. The man supremacy’: sexist.

‘The thing is women are not equal to us men and their place is the home and kitchen:’

GPT-3 response: sexist.

**d) few-shot setting:** **mixed model**

- given an example of sexist text, racist text, and neither, model is expected to categorize the given passage

Prompt:

‘That girl still need a good ass whooping, stupid ass b\*\*\*\*!!’: sexist.

’U SUCK S\*\*\* YOU BLACK F\*\*\*’: racist.

’You deserve way more subscribers. The s\*\*\* you say is f\*\*\*ing gold.’: neither.

’Every girl on this planet curses her mother for their birth. They know that they are the powerless sex and they pray that they had been born as men! With d\*\*\*s and stuff.’

GPT-3 response:

sexist.

A table with numbers and text

Description automatically generated

**With brief instruction** to restrict the generated answers to ‘racist’, ‘sexist’ or ‘neither’, to prevent the model from generating out-of-scope answers.

Prompt:

Classify the following texts into ‘racist’, ‘sexist’ or ‘neither’.

‘That girl still need a good ass whooping, stupid ass b\*\*\*\*!!’: sexist.

’U SUCK S\*\*\* YOU BLACK F\*\*\*’: racist.

’You deserve way more subscribers. The s\*\*\* you say is f\*\*\*ing gold.’: neither.

’Every girl on this planet curses her mother for their birth. They know that they are the powerless sex and they pray that they had been born as men! With d\*\*\*s and stuff.’

GPT-3 response:

sexist.

A table of text with numbers

Description automatically generated

**Results:**- zero- and one-shot learning: accuracy between 55% - 67%

- few-shot learning: model's accuracy can be as high 85%

- Overall, it appears as though GPT-3 is most effective at identifying both racist and sexist comments in the mixed-category few-shot learning setting, without instruction

**(4) Other prompt strategies for Hate Speech Detection** (Guo et al., 2023)

**General prompt:**

Previously validated for its effectiveness in hate speech detection (Li et al., 2023)

„Do you think this comment is hate speech?

comment: {x}

a. Yes b. No.“

**General prompt with hate speech definition:**

To account for variations across different datasets

„Do you think this comment is hate speech?

Hate speech definition: {j}

comment: {x}

a. Yes b. No.

Answer: “

**Few-shot learning prompt:**

supplying the model with a small number of example tasks along with their solutions, essentially ‘priming’ it for the particular type of task at hand

„Example 1:

Question: Do you think this comment is hate speech?

comment: {k}

a. Yes b. No

Answer: a. Yes

Question: Do you think this comment is hate speech?

comment: {x}

a. Yes b. No

Answer: ”

**Chain-of-thought prompt:**

Prompting strategy that interjects intermediate steps between input x and output. Intermediate steps were designed by further elaborating the definition of hate speech step-by-step, starting from identifying targets to identifying disparaging words.

„Example 1:

Question: Do you think this comment is hate speech?

comment: {k}

a.Yes b.No   
Answer: This comment is intended to incite hatred against {l} by using words such as {m}. So the answer is a. Yes

Question: Do you think this comment is hate speech?

comment: {x}

a. Yes b. No

Answer:“

**Results:**

* **Chain-of-Reasoning Prompt** (CoT) outperforms other prompts across all metrics (accuracy: 0.85, precision: 0.8, recall: 0.95, F1 score: 0.87.)
* providing a structured chain of reasoning within prompt might substantially enhance the model’s understanding and detection of hate speech
* breaking up the problem of hate speech detection into multiple steps could better leverage the model’s learned knowledge base

**Dataset for Counter Speech with Classification:**

<https://github.com/yilingchung/multilingualCN-classification/tree/main> (5 classes)

<https://github.com/hate-alert/Counterspeech_Twitter/tree/master> (10 classes)

**References**

Chiu, K., Collins, A., & Alexander, R. (2021). Detecting Hate Speech with GPT-3. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2103.12407

Ekin, S. (2023). Prompt Engineering For ChatGPT: A Quick Guide To Techniques, Tips, And Best Practices. *TechRxiv*. https://doi.org/10.36227/techrxiv.22683919.v2

Guo, K., Hu, A., Mu, J., Su, Z., Zhao, Z., Vishwamitra, N., & Hu, H. (2023). An investigation of large language models for Real-World Hate Speech Detection. *IEEE*. https://doi.org/10.1109/icmla58977.2023.00237

Li, L., Fan, L., Atreja, S., & Hemphill, L. (2023). “HOT” ChatGPT: The promise of ChatGPT in detecting and discriminating  hateful, offensive, and toxic comments on social media. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2304.10619

Wang, Y., & Chang, Y. (2022). Toxicity Detection with Generative Prompt-based Inference. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2205.12390

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2302.11382