**Prompts**

**Factual QA Dataset**

**Create subtopic prompt –**

You must generate exactly {n\_subtopics} subtopics related to the given topic. No more, no less.

The topic is: {topic}

The list must contain exactly {n\_subtopics} subtopics. Each subtopic should be separated by a comma. There should be no numbers, no descriptions, and no additional text other than the subtopics themselves.

Do not include any subtopics other than the {n\_subtopics} requested. Failure to generate exactly {n\_subtopics} subtopics will be considered incorrect.

**Generate Question Prompt-**

You must generate exactly {n\_questions} questions that are strictly and directly related to the specific subtopic provided. No tangential, broad, or off-topic questions are allowed.

The subtopic is: {sub\_topic}

Your response must consist of precisely {n\_questions} questions, each directly pertaining to the subtopic, separated by a newline character, with absolutely no additional text, numbering, explanations, or any other characters.

Deviation from the subtopic or any failure to generate exactly {n\_questions} questions as instructed will result in the output being considered invalid.

**Response generation prompt-**

Generate a single, concise, and relevant response to the given question. The response should be directly related to the question, clear, and free of any unnecessary information.

The question is: {question}

Provide only one response in plain text, with no additional explanations, introductions, or concluding remarks.

**DPO Dataset-**

**Question Prompt-**

You must generate exactly {n\_questions} questions that are strictly and directly related to the specific subtopic provided. No tangential, broad, or off-topic questions are allowed.

The subtopic is: {sub\_topic}

Your response must consist of exactly {n\_questions} questions, each directly pertaining to the subtopic, separated by a newline character, with absolutely no empty lines, numbering, explanations, or any other characters. Each question should end with a question mark.

Failure to generate exactly {n\_questions} valid questions without empty lines will result in the output being considered incorrect.

**Chosen Response Prompt-**

Generate a concise, relevant response to the given question. The response should be directly related to the question, clear, and free of any unnecessary information. It should be helpful, polite, and factually accurate.

The question is: {question}

Provide only one response in plain text, with no additional explanations, introductions, or concluding remarks.

**Rejected Response prompt-**

Generate a rejected response to the given question that is moderately inaccurate compared to the accurate response. The rejected response may be incomplete or less accurate, but it should still be relevant to the question.

The question is: {question}

Provide only one response in plain text, with no additional explanations, introductions, or concluding remarks.

SynDiP dataset-

**Industrial Synthesis Prompt-**

Provide a comprehensive and detailed description of the industrial synthesis process for {chemical\_name}. Your description should include:

- All key chemical reactions, including reactants, intermediates, and products.

- The types of reactors used (e.g., CSTR, PFR) and their operating conditions (e.g., temperature, pressure).

- Details of any purification steps, such as distillation, crystallization, or filtration, including the equipment used.

- Handling and treatment of by-products and waste streams.

- Any recycling loops and the integration of heat exchange systems to optimize energy use.

- Specific safety measures taken during the synthesis, especially when dealing with hazardous chemicals.

The description should be suitable for an engineer looking to understand the process in detail for implementation in a large-scale industrial setting.

**PFD Generation prompt-**

Based on the following synthesis description, create a detailed textual Process Flow Diagram (PFD) for the synthesis of {chemical\_name}. Your PFD should include:

- Major equipment involved at each step, such as reactors, heat exchangers, distillation columns, separators, pumps, and compressors.

- The flow of raw materials, intermediates, and products through the process, including any recycling streams.

- Details of heat integration, such as the use of heat exchangers to recover energy from exothermic reactions or to preheat reactants.

- A clear representation of phases (e.g., gas, liquid, solid) in each unit operation, highlighting phase transitions where applicable.

- Specific operating conditions at key stages, including temperatures, pressures, and flow rates, to ensure proper operation.

- The identification of potential bottlenecks in the process flow, and suggestions for optimizing throughput.

Ensure that the PFD is designed according to industry standards and is suitable for scaling up to large-scale production.

Synthesis Description: {synthesis\_description}  
  
**PID generation Prompt-**

Create a detailed Piping and Instrumentation Diagram (P&ID) based on the following process flow diagram (PFD) for the synthesis of {chemical\_name}. The P&ID should include:

- Detailed placement of sensors (e.g., temperature, pressure, flow, and level sensors) at critical points in the process to ensure precise monitoring.

- Specification of control valves, actuators, and control loops required to maintain process parameters within the desired range, including examples of typical control strategies (e.g., feedback, feedforward, cascade control).

- Identification of safety instrumentation, such as pressure relief valves, emergency shutdown systems, interlocks, and alarms, to prevent accidents and ensure compliance with safety regulations.

- Details of the instrumentation needed for process optimization, such as advanced process control (APC) systems, model predictive control (MPC), and real-time data analytics.

- Consideration of redundancy and reliability in the placement of key sensors and control elements to ensure continuous operation and minimize downtime.

- Recommendations for the types of piping materials to be used, considering the chemical compatibility, temperature, and pressure of the process streams.

- Suggestions for the integration of control systems with a distributed control system (DCS) or supervisory control and data acquisition (SCADA) system for centralized monitoring and control.

Your P&ID suggestions should align with industry best practices and standards (e.g., ANSI/ISA-5.1) and be suitable for large-scale industrial production.

Process Flow Diagram (PFD): {process\_flow\_diagram}

**LogiCore dataset-**

**Question prompt Template –**

You are to generate exactly {num\_questions} questions based on the provided synthesis description, Process Flow Diagram (PFD), and Piping and Instrumentation Diagram (P&ID). \*\*The questions must strictly focus on the following categories: Conceptual Understanding, Interpretation, Design and Engineering, and Analytical Thinking.\*\*

\*\*Rules (to be strictly followed):\*\*

- Do not include any introductions, explanations, or summaries. \*\*Only output questions.\*\*

- Do not use any phrases such as "Here are the questions" or "Based on the diagrams."

- Do not include any numbers, bullet points, or any other symbols.

- Each question must end with a question mark and be followed by a single newline character (\n).

- If any text other than the questions is generated, it is \*\*invalid output.\*\*

Synthesis Description: {{synthesis\_description}}

Process Flow Diagram: {{pfd\_description}}

Piping and Instrumentation Diagram: {{pid\_description}}

**Answer prompt template-**

Provide clear, accurate, and concise answers to the following questions. Adhere strictly to the following rules to ensure very high scores in the following categories:

1. \*\*Helpfulness:\*\* Ensure each answer is extremely helpful, fully addressing the question in a way that resolves the query effectively.

2. \*\*Correctness:\*\* Every answer must be factually correct, accurately referencing relevant details from the synthesis description, Process Flow Diagram (PFD), and Piping and Instrumentation Diagram (P&ID).

3. \*\*Coherence:\*\* Ensure that each answer is logically structured and flows smoothly, making it easy for the reader to follow.

4. \*\*Complexity:\*\* Balance complexity appropriately; provide necessary depth without making the answer overly complicated. Ensure the response is insightful where needed.

5. \*\*Verbosity:\*\* Be concise but thorough. Include all essential details without adding unnecessary information. Ensure that the length of the answer aligns perfectly with the complexity of the question.

Failure to adhere to these rules will lead to lower scores and suboptimal performance.

Synthesis Description: {synthesis\_description}

Process Flow Diagram: {pfd\_description}

Piping and Instrumentation Diagram: {pid\_description}

Questions: {questions}

**Local and Global RAIT Dataset**

**Question Prompt –**

Generate exactly {num\_questions} clear, concise, and fact-based questions that can be answered solely using the information provided in the following text. Ensure that the questions are directly relevant to the content and avoid any ambiguity or assumptions. Each question should target a distinct piece of information from the text, ensuring no overlap in the questions generated. The questions must adhere strictly to the provided text and should not introduce any outside information.

{chunk}

**Response prompt-**

Question: {question}

Context: {chunk}

Provide a concise, accurate, and fact-based answer to the question, using only the information available in the context provided. The answer must be directly derived from the context and should not include any external knowledge, speculation, or interpretation. Ensure that the response is precise and strictly adheres to the content of the context without introducing any additional information.