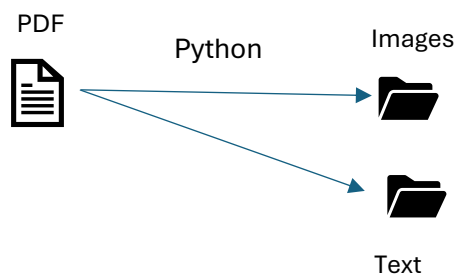


## A method and system for generating a knowledge graph and query from multimodal data sources using multi-agent Large Language Models (LLMs)

### Step 1 –

Step 1 comprises processing Pdf documents to identify text and images separately as it needs to be processed according to data type. We have used PyMuPDF (fitz) library of python to perform this task.

PyMuPDF, also known as **fitz**, is a powerful Python library for working with PDF files and other document formats. It provides functionalities for text extraction, image extraction, document manipulation, and more. Here's an overview of how PyMuPDF works. At the end of processing we have two folders where text and images for extracted and stored .



### Step II- The main innovation-

Step II is comprised of **multiple collaborating customized or tuned LLM agents**, customized to perform specific tasks.

#### How customization or tuning is done for LLM agent-

Customization of LLM agent for specific task is done via providing **it customization model in JSON** to **LLM agent as input**. The customization model has **two components**

#### Behavioral Customization

**Definition:** Behavioral customization involves adjusting the **LLM's responses**, interactions, and domain based on the provided model, which includes **roles, knowledge areas, and interface**. This ensures that the LLM behaves in a way that aligns with the defined parameters and delivers relevant, **context-aware responses**.

For instance:

- **Role:** If the role is a "Maintenance Engineer," the LLM will tailor its responses to provide maintenance-related guidance and support.
- **Domain Knowledge:** The LLM will adjust its responses to leverage its expertise in specific areas like "Oil and Gas Industry Equipment Maintenance."
- **Interface:** If the model's interface type is text-based with capabilities for natural language understanding, it will interact through text and understand and generate responses accordingly.

**Example:**

If the LLM is tasked with providing safety guidelines, it will use the domain knowledge and role definitions to ensure that its responses are relevant to the oil and gas industry and adhere to safety protocols.

### Functional Specification

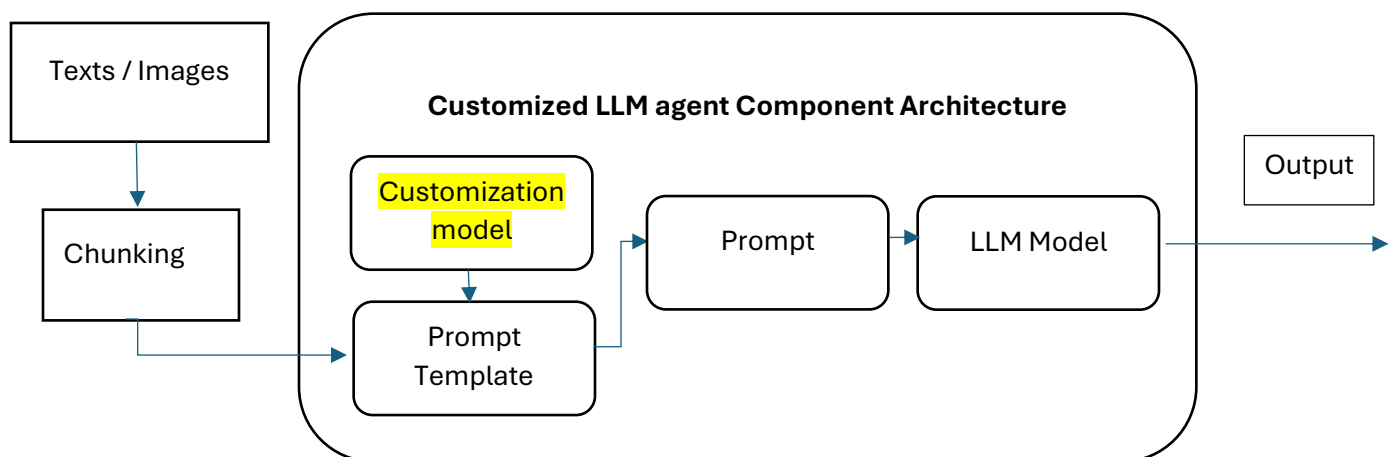
Functional specification involves defining the specific capabilities, actions, and tasks **that the LLM should handle**. It details what the model should be able to do and how it should perform these functions based on the predefined attributes.

**Capabilities and Actions:** The functional specification outlines what the LLM is expected to do:

- **Actions:**
  - **Core Actions:** These include fundamental tasks that the LLM should handle, such as "Providing maintenance procedures" and "Troubleshooting assistance."
  - **Extended Actions:** Additional tasks like "Generating equipment manuals" and "Offering safety guidelines."
- **Tasks:**
  - **Primary Tasks:** Essential tasks such as "Routine maintenance scheduling" and "Emergency repair guidance."
  - **Advanced Tasks:** More specialized tasks like "Optimizing equipment performance" and "Conducting safety audits."
- **Tools:**
  - **Integrated Tools:** Systems or databases that the LLM should use internally, such as "Equipment databases" and "Maintenance tracking systems."
  - **External Tools:** External resources or APIs that the LLM should interact with, such as "APIs for industry standards and regulations."

### Example:

The LLM should be able to extract information from integrated equipment databases to provide detailed maintenance procedures and use external APIs to stay updated on industry standards and regulations.



A **prompt template** is a predefined format used to create input for a language model, helping to generate consistent and relevant responses. It includes:

- **Structure:** The basic format of the prompt.

- **Placeholders:** Parts where specific details are inserted (e.g., [TOPIC]).
- **Instructions:** Guidelines on how to respond (e.g., "Explain in simple terms").
- **Context:** Additional details to guide the response (e.g., "In the context of AI technology").

**Example:** A template might be "Describe the benefits of [TOPIC]." For "machine learning," it would prompt the model to describe its benefits.

### Prompt-

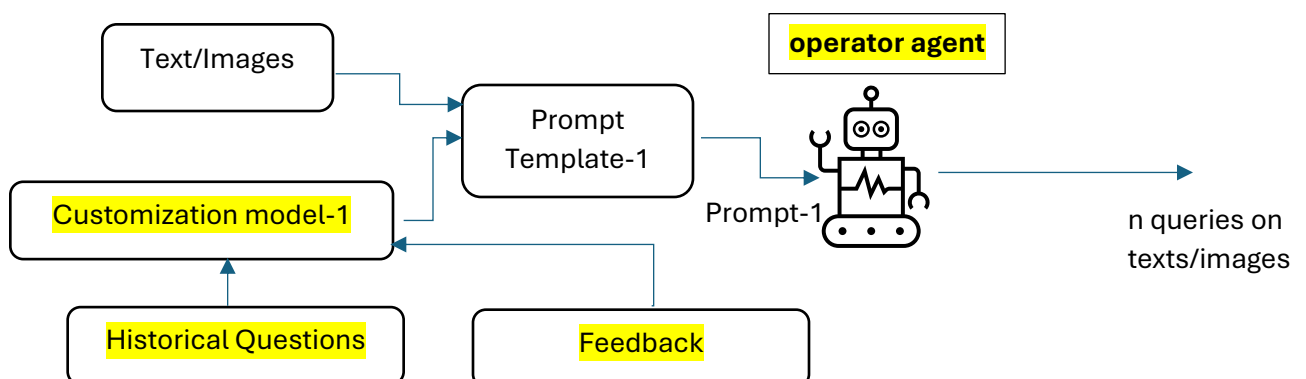
A prompt for a language model (LLM) is a piece of text or input that you provide to the model to generate a response. Prompts guide the model in generating text based on the context or instructions provided.

Here prompts are automatically generated based on defined prompt template, customized model and text or image chunk.

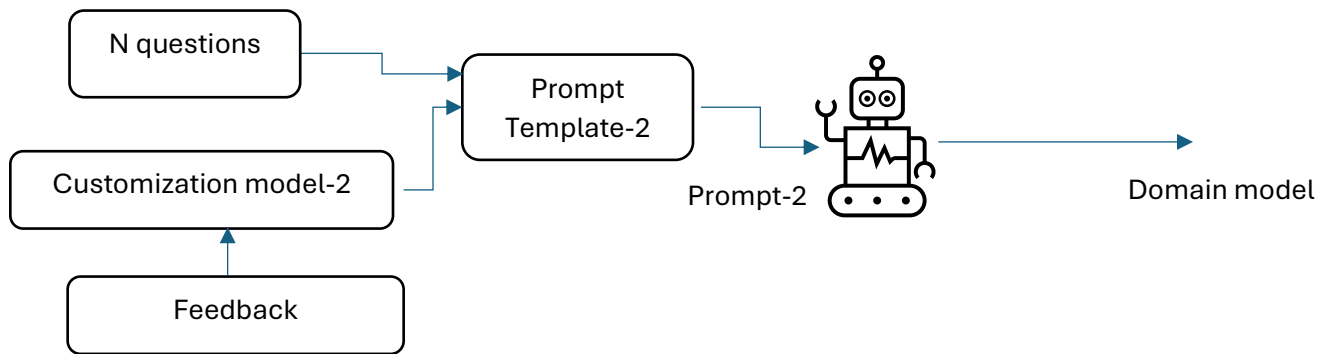
**LLM Model** - A **LLM** (Large Language Model) is a type of artificial intelligence model designed to understand, generate, and manipulate human language. These models are built on advanced machine learning techniques, particularly deep learning, and are trained on vast amounts of text and Image data to perform a wide range of language-related tasks.

- a) **An operator agent** -The output of previous step i.e. text and images are input to agent 1 also known as operator agent. **The operator agent is a tuned or customized LLM agent to generate questions like a maintenance engineer.** The agent model's behavior is customized using **a structured template in JSON which has Behavioral customization and functional specifications.** This structured template goes to LLM model with along prompt. Feedback [1] is provided by other LLM agents if the generated response is not satisfactory

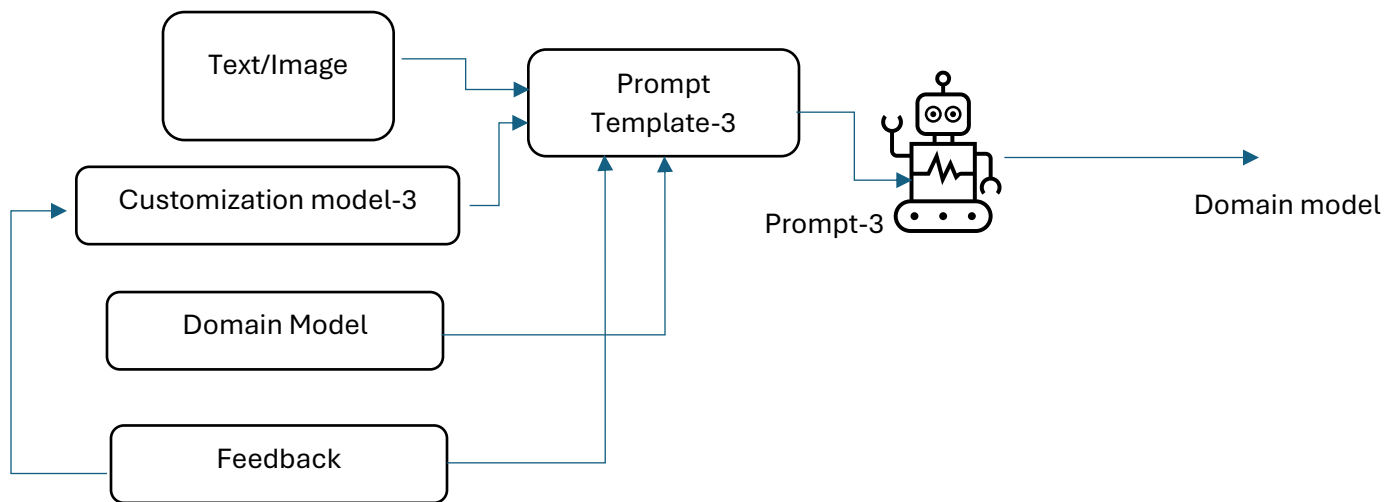
**An operator agent** is a tuned LLM model that takes chunks of texts or images, and curates a **set of n questions based on past questions**/queries posed by field staff.



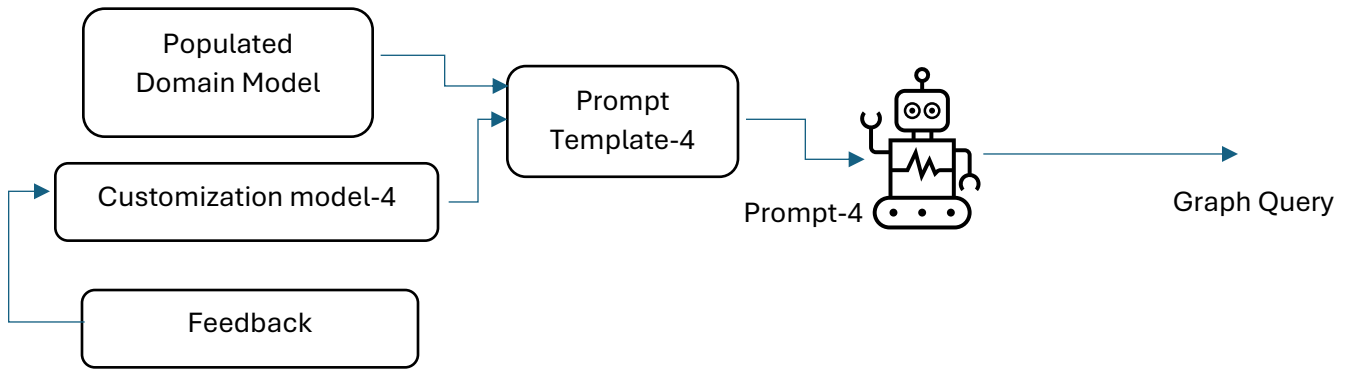
- b) **A domain model generator** agent generates domain model based on generated questions and curate domain model as shown below. The input to domain model generator is questions from previous agent, customization model and prompt. The domain model generator outputs a domain model based on provided questions. Feedback is provided by other LLM agents if the generated response is not satisfactory or missing concepts. The feedback process helps refine and enhance the domain model by integrating additional details, correcting inaccuracies, and improving clarity



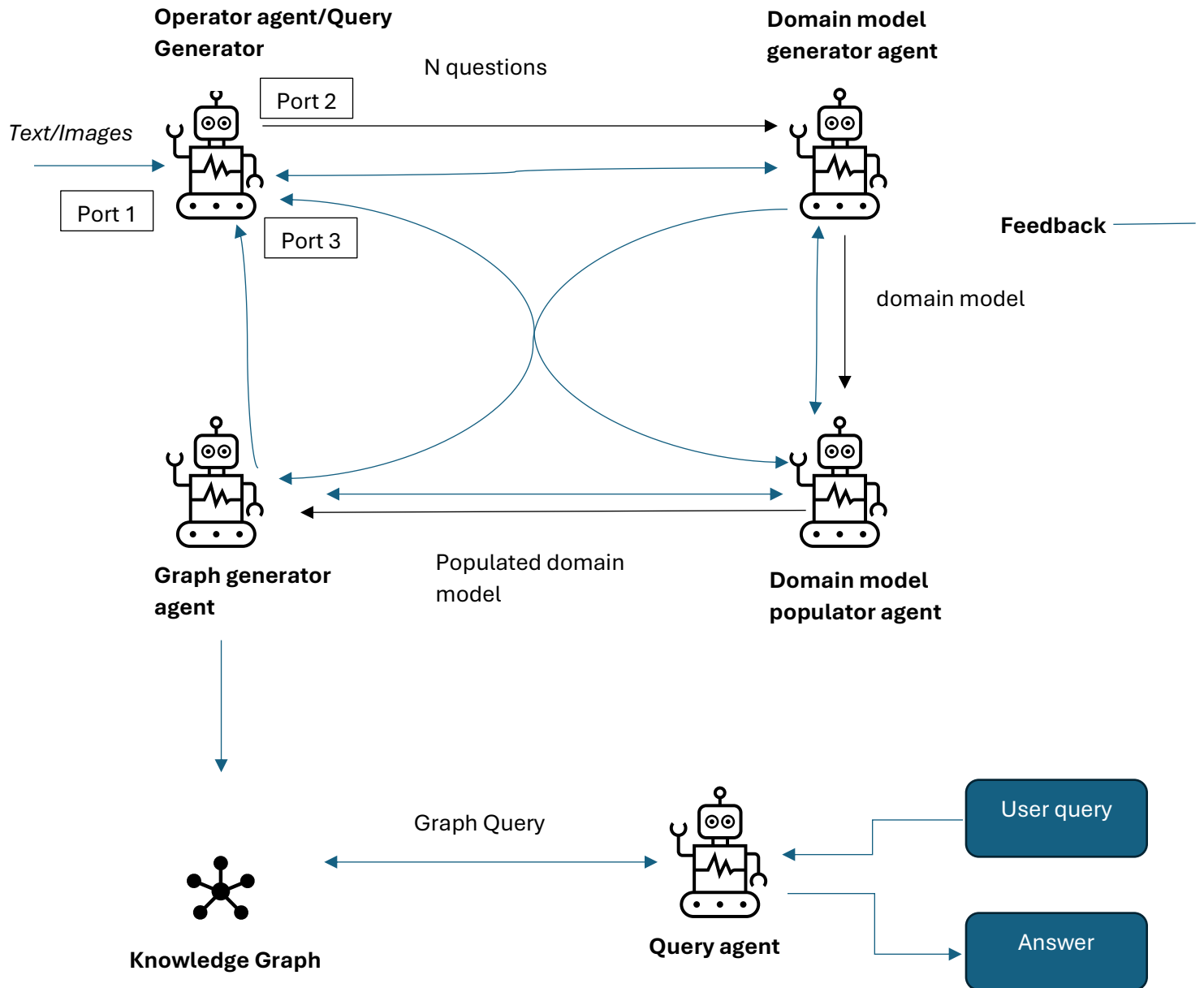
- c) **A domain model populator** agent populates generated domain model based on chunk or image of document. The input to domain model populator agent is domain model from previous agent, customization model, text or image chunk and prompt. The domain model populator outputs a populated domain model. Feedback is provided by other LLM agents if the generated response is not satisfactory.



- d) **A Knowledge Graph creator** agent creates a query to create KG in neo4j. The input to agent is populator is domain model from previous agent, customization model, and prompt. The graph creator agent outputs query to populate graph in neo4j. Feedback is provided by other LLM agents if the generated response is not satisfactory.



## Agents Interaction Architecture.



An LLM agent **port** refers to a specific interface or communication endpoint through which a Language Model (LLM) agent interacts with other systems, applications, or components. Ports are used to facilitate the exchange of data and instructions between the LLM agent and its external environment.

## End to End Use case on Boiler:

**Chapter 9.2** of the publicly available technical report "**Operations & Maintenance Best Practices: A Guide to Achieving Operational Efficiency**" published by the Pacific Northwest National Laboratory [[https://pnnl.gov/main/publications/external/technical\\_reports/PNNL-19634.pdf](https://pnnl.gov/main/publications/external/technical_reports/PNNL-19634.pdf)]. The operator agent generates 50 questions from each page

Texts: the following is the sample of the text from the above chapter

Customization model of Agent1:

```
"Name": "Pl",
"Role": "Boiler Maintenance Engineer",
"DomainKnowledge": {
  "Primary": "Boiler operation and maintenance",
  "Specialization": [
    "Fire-Tube Boilers",
    "Water-Tube Boilers",
    "Electric Boilers",
    "Non-Destructive Examination (NDE) methods"
  ]
},
```

Prompt Template for Agent task execution:

```
Prompt=""
You are an operator with expertise in cable installation and maintenance, tasked with generating questions which actual operators on working oil & gas field would ask. Your goal is to create questions based solely on the provided text content.
Model Definition (for reference): {model}
Text Content: {text}
Generate {num_questions} questions from the above text content.
""
```

Inputs:

---

## Chapter 7 O&M Ideas for Major Equipment Types

### 7.1 Introduction

At the heart of all O&M lies the equipment. Across the Federal sector, this equipment varies greatly in age, size, type, model, fuel used, condition, etc. While it is well beyond the scope of this guide to study all equipment types, we tried to focus our efforts on the more common types prevalent in the Federal sector. The objectives of this chapter are the following:

- Present general equipment descriptions and operating principles for the major equipment types.
- Discuss the key maintenance components of that equipment.
- Highlight important safety issues.
- Point out cost and efficiency issues.
- Provide recommended general O&M activities in the form of checklists.
- Where possible, provide case studies.

*The checklists provided at the end of each section were compiled from a number of resources. These are not presented to replace activities specifically recommended by your equipment vendors or manufacturers. In most cases, these checklists represent industry standard best practices for the given equipment. They are presented here to supplement existing O&M procedures, or to merely serve as reminders of activities that should be taking place. The recommendations in this guide are designed to supplement those of the manufacturer, or, as is all too often the case, provide guidance for systems and equipment for which technical documentation has been lost. As a rule, this guide will first defer to the manufacturer's recommendations on equipment operations and maintenance.*



Actions and activities recommended in this guide should only be attempted by trained and certified personnel. If such personnel are not available, the actions recommended here should not be initiated.





## Questions Generated:

```
1. What are the primary objectives of Chapter 7 in the O&M guide?
2. How does the age of equipment vary across the Federal sector?
3. What types of equipment are focused on in this chapter?
4. What are the general equipment descriptions provided in this chapter?
5. What operating principles are discussed for major equipment types?
6. What are the key maintenance components of the equipment mentioned?
7. What important safety issues are highlighted in this chapter?
8. How are cost and efficiency issues addressed in this chapter?
9. What general O&M activities are recommended in the form of checklists?
10. From where were the checklists at the end of each section compiled?
11. Are the checklists meant to replace activities recommended by equipment vendors or manufacturers?
12. What is the purpose of the checklists provided in this chapter?
13. How should the recommendations in this guide be used in relation to manufacturer's recommendations?
14. What should be done if technical documentation for systems and equipment has been lost?
15. Who should attempt the actions and activities recommended in this guide?
16. What should be done if trained and certified personnel are not available?
17. What is the scope of the equipment types studied in this guide?
18. How does this guide supplement existing O&M procedures?
19. What is the significance of case studies in this chapter?
20. How does this guide address the condition of equipment in the Federal sector?
21. What fuel types are considered in the equipment discussed in this chapter?
22. How does the guide ensure safety during O&M activities?
23. What are the industry standard best practices mentioned in the checklists?
24. How does the guide address the efficiency of equipment?
25. What are the general principles of operation for the major equipment types?
26. How does the guide handle equipment for which technical documentation is unavailable?
27. What is the role of manufacturer's recommendations in this guide?
28. How are cost issues related to equipment maintenance addressed?
29. What are the recommended activities for O&M in this guide?
30. How does the guide ensure compliance with safety protocols?
31. What is the importance of trained and certified personnel in O&M activities?
32. How does the guide address the variability in equipment models?
33. What are the key components of maintenance for the equipment discussed?
```

## Agent2: Domain Model generator

Customization of LLM to perform the task of domain model generator

```
"Name": "ai_agent02",
"Role": "Comprehensive Domain Model Generator",
"DomainKnowledge": {
  "Primary": "Creating detailed and hierarchical domain models from textual input",
  "Specialization": ["Entity extraction", "Relationship identification", "Hierarchical structure creation", "Attribute detailing"]
},
"Interface": {
  "Type": "Text-based",
  "Capabilities": ["Advanced Natural Language Understanding", "Complex Data Parsing", "Hierarchical Structure Analysis"]
},
"Output Act": {
  "Core": ["Generates a comprehensive, hierarchical domain model in JSON format"],
  "Extended": ["Integration with knowledge graph construction", "Cross-domain relationship identification"]
},
"InputInstructions": "Accepts comprehensive textual descriptions or questions about a domain as input for generating a detailed domain model",
"Tasks": {
  "Primary": ["Detailed entity extraction", "Multi-level relationship identification", "Hierarchical structure creation", "Attribute detailing"],
  "Advanced": ["Ontology integration", "Data normalization", "Cross-domain connection identification", "Implicit relationship inference"]
},
```

## Prompt Template to generate the domain model

```
Prompt = """
You are an AI agent acting as a Comprehensive Domain Model Generator. Your task is to create a detailed, hierarchical domain model based on the provided information.
Agent Model: {model}
Manual Text: {manual_text}
Generated Questions: {generated_questions}
Create a comprehensive domain model that includes:
1. Main entities with descriptions
2. Detailed attributes for each entity
3. Sub-entities to represent hierarchical structures
4. Relationships between entities, including relationship types and details. Avoid generic relationships and ensure all relationships are meaningful and contextually appropriate.
Ensure the model is hierarchical, with main entities containing sub-entities where appropriate.
Focus on creating a structure that will be suitable for generating a knowledge graph.
IMPORTANT: Provide the domain model as a raw JSON object without any additional formatting or Markdown syntax. The response should be a valid JSON string that can be directly parsed by a JSON parser.
"""
```

## Generated Domain Model for the given page based on the questions from Agent1

```
{
  "Chapter": {
    "sections": [
      {
        "id": "safety_guidelines",
        "description": "Actions and activities recommended should only be attempted by trained and certified personnel.",
        "cost_and_efficiency_issues": "Addressed in the chapter with specific focus on cost and efficiency issues related to equipment maintenance and repair."
      },
      {
        "id": "equipment_types",
        "description": "Major equipment types discussed in the chapter.",
        "attributes": [
          "age",
          "size",
          "type",
          "model",
          "fuel_used",
          "condition"
        ],
        "sub_entities": [
          {
            "id": "engine_specs",
            "description": "Detailed specifications for various engine models.",
            "attributes": [
              "horsepower",
              "torque",
              "fuel_consumption",
              "emissions"
            ]
          },
          {
            "id": "transmission_specs",
            "description": "Detailed specifications for various transmission types.",
            "attributes": [
              "gear_ratios",
              "shift_types",
              "maintenance_requirements"
            ]
          }
        ]
      }
    ],
    "entities": [
      {
        "id": "equipment",
        "description": "Major equipment types discussed in the chapter.",
        "attributes": [
          "age",
          "size",
          "type",
          "model",
          "fuel_used",
          "condition"
        ],
        "sub_entities": [
          {
            "id": "engine_specs",
            "description": "Detailed specifications for various engine models.",
            "attributes": [
              "horsepower",
              "torque",
              "fuel_consumption",
              "emissions"
            ]
          },
          {
            "id": "transmission_specs",
            "description": "Detailed specifications for various transmission types.",
            "attributes": [
              "gear_ratios",
              "shift_types",
              "maintenance_requirements"
            ]
          }
        ]
      }
    ]
  }
}
```

## Agent3: Domain Model Populator

### Customization Model

```
{
  "Name": "ai_agent03",
  "Role": "Domain Model Content Populator",
  "DomainKnowledge": {
    "Primary": "Populating domain models with content from extracted manual text",
    "Specialization": ["Content mapping", "Information extraction", "Hierarchy preservation"]
  },
  "Interface": {
    "Type": "Text-based",
    "Capabilities": ["Natural Language Processing", "Text analysis", "Structured data mapping"]
  },
  "Input": {
    "DomainModel": "Structured model from ai_agent02",
    "Content": "Extracted text from manual (input.txt)"
  },
  "Output Act": {
    "Core": ["Populates the domain model with content from input.txt, maintaining structure and relationships"],
    "Customization": ["Adjusts model structure based on specific domain requirements"]
  }
}
```

### Prompt -3

```
Prompt="""
You are an AI agent responsible for populating a domain model with content from technical manuals.
Use the following model definition to guide your task:

{agent_model}

Here is the domain model structure to be populated: {domain_model}

Here is the extracted text from the manual: {input_text}

Your task is to populate the domain model with relevant information from the text. Follow these guidelines:
1. Maintain the structure and relationships defined in the model.
2. Extract entities, attributes, relationships, and hierarchical information as specified in your agent model.
3. Ensure that the populated model contains all necessary information for effective knowledge graph generation.
4. If you encounter information that doesn't fit the current model structure, add it as additional attributes or relationships.
5. Cross-reference information across different sections of the input text to ensure comprehensive population.
6. Resolve any inconsistencies or conflicts in the information, prioritizing the most recent or most relevant data.

Provide the populated domain model in JSON format. Do not include any explanations or metadata about
the agent or the process in your response, just the pure JSON of the populated domain model.
IMPORTANT: Provide the domain model as a raw JSON object without any additional formatting or Markdown syntax. The response should be a
valid JSON string that can be directly parsed by a JSON parser.

Begin the JSON object with {{ and end with }} without any additional characters or formatting.
"""
```

## Extracted Instance

```
{
  "sub_entities": [
  ],
},
"relationships": [
  {
    "type": "supplements",
    "from": "checklists",
    "to": "existing_O&M_procedures",
    "details": "Checklists are designed to supplement existing O&M procedures or serve as reminders."
  },
  {
    "type": "requires",
    "from": "actions_and_activities",
    "to": "trained_and_certified_personnel",
    "details": "Actions and activities should only be attempted by trained and certified personnel."
  },
  {
    "type": "addresses",
    "from": "chapter",
    "to": "cost_and_efficiency_issues",
    "details": "The chapter addresses cost and efficiency issues related to equipment maintenance and operation."
  }
],
}
```

## Agent4: Knowledge Graph Curator

### Customization Model

```

"Name": "ai_agent04",
"Role": "Knowledge Graph Generator",
"DomainKnowledge": {
  "Primary": "Creating fully connected, hierarchical Knowledge Graphs in Neo4j from populated domain models",
  "Specialization": ["Graph database structures", "Cypher query language", "Ontology mapping", "Data cleaning and normalization"]
},
"Interface": {
  "Type": "Text-based",
  "Capabilities": ["Structured data processing", "Graph modeling", "Cypher query generation", "Error handling and validation"]
},
"Input": {
  "PopulatedDomainModel": {
    "Source": "Structured and populated model from ai_agent03",
    "ExpectedFormat": ["JSON"],
    "ValidationCriteria": ["Schema consistency", "Data completeness", "Relationship integrity", "Entity consistency"]
  }
},
"Output": {
  "Core": [
    "Fully connected, hierarchical Knowledge Graph in Neo4j",
    "Cypher queries for graph creation and debugging",
    "Performance metrics (processing time, node/relationship count)",
    "Graph version information"
  ]
},

```

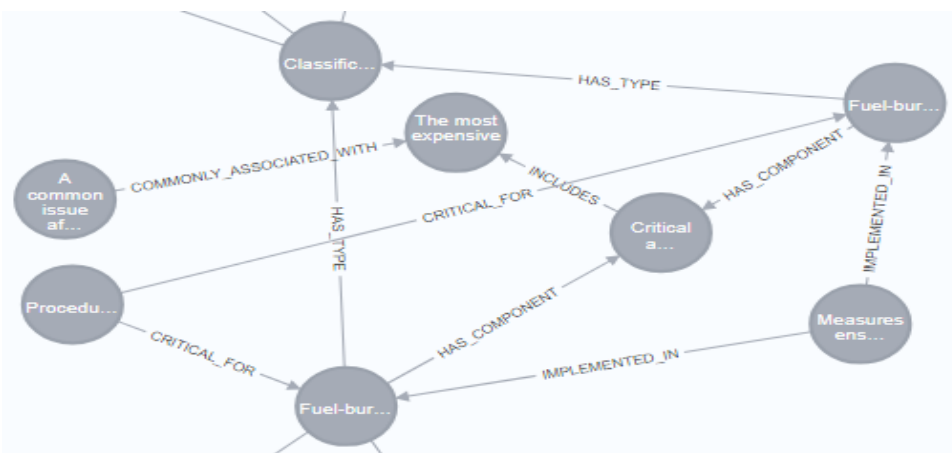
## Prompt

```

Prompt =
"""# Knowledge Graph Instructions for GPT-4o
## 1. Overview
You are a top-tier algorithm designed for extracting information in structured formats to build a knowledge graph.
**Nodes** represent entities and concepts.
**Relationships** link the nodes and define the nature of their connections.
**Properties** are key-value pairs associated with nodes and relationships.
## 2. Cypher Query Format
For each entity:
- Create a node using the format: `CREATE (n:NODE_LABEL {{ id: 'NODE_ID', property1: 'value1', ... }})`
- Ensure NODE_ID and properties are correctly quoted and escaped.
For each relationship:
- Create a relationship using the format: `MATCH (a:NODE_LABEL {{ id: 'NODE_ID' }}), (b:NODE_LABEL {{ id: 'NODE_ID' }}) CREATE (a)-[:REL_TYPE {{ property1: 'value1', ... }}]-(b)`
## 3. Node and Relationship Guidelines
- Use consistent labels for similar types of nodes
## 4. Handling Numerical Data and Dates
- Numerical data, like age or other related information, should be incorporated as attributes or properties of the respective nodes.
- **No Separate Nodes for Dates/Numbers**: Do not create separate nodes for dates or numerical values. Always attach them as attributes or properties of nodes.
- **Property Format**: Properties must be in a key-value format.
- **Quotation Marks**: Never use escaped single or double quotes within property values.
- **Naming Convention**: Use camelCase for property keys, e.g., `birthDate`.
## 5. Coreference Resolution
- **Maintain Entity Consistency**: When extracting entities, it's vital to ensure consistency.
## 6. Hierarchical Structure
- Ensure that entities are organized in a hierarchical manner where applicable.
- Use relationships like "IS_PART_OF" or "BELONGS_TO" to establish clear parent-child relationships.
## 7. Full Connectivity
- Ensure that every node is connected to at least one other node in the graph.
- If an entity seems isolated, consider its potential relationships with existing entities.
"""

```

## Generated Graph



## [1] Feedback Mechanism Example-

### Scenario

**Initial Domain Model Generation:** The Domain Model Generator Agent creates a domain model for a boiler maintenance manual based on questions about boiler types, maintenance procedures, and safety protocols.

### Generated Domain Model:

- **Concepts:** Boiler Types, Maintenance Procedures, Safety Protocols
- **Relationships:**
  - Boiler Types → Steam Boilers, Hot Water Boilers
  - Maintenance Procedures → Daily Checks, Monthly Inspections, Annual Overhaul
  - Safety Protocols → Pressure Relief Valve Testing, Safety Shut-off Mechanisms

### Feedback Process

#### Feedback from Other LLM Agents:

1. **Accuracy:**
  - **Feedback:** An agent points out that the model lacks details on "Boiler Water Treatment," which is crucial for maintaining boiler efficiency and longevity.
  - **Revised Model:** Includes "Boiler Water Treatment" with subcategories like "Water Softening" and "Chemical Additives."
2. **Completeness:**
  - **Feedback:** Another agent notes that the "Maintenance Procedures" section should also cover "Emergency Procedures" and "Troubleshooting Common Issues," which are essential for comprehensive maintenance.
  - **Revised Model:** Adds "Emergency Procedures" and "Troubleshooting Common Issues" to the maintenance procedures.
3. **Relevance:**
  - **Feedback:** Feedback suggests that the model should include a section on "Regulatory Compliance," which outlines industry standards and regulations affecting boiler maintenance.
  - **Revised Model:** Adds a "Regulatory Compliance" section detailing relevant standards and regulations.
4. **Clarity:**

- **Feedback:** The model is criticized for not specifying the frequency of maintenance tasks clearly. The relationship description for "Maintenance Procedures" should specify the intervals more clearly.
- **Revised Model:** Specifies that "Daily Checks" include "Inspecting Pressure Gauges," "Monthly Inspections" involve "Checking Water Levels," and "Annual Overhaul" requires "Complete System Inspection."

## Revised Domain Model

### Concepts:

- Boiler Types
- Maintenance Procedures
  - Daily Checks
  - Monthly Inspections
  - Annual Overhaul
  - Emergency Procedures
  - Troubleshooting Common Issues
- Safety Protocols
- Boiler Water Treatment
  - Water Softening
  - Chemical Additives
- Regulatory Compliance

### Relationships:

- Boiler Types → Steam Boilers, Hot Water Boilers
- Maintenance Procedures → Daily Checks → Inspecting Pressure Gauges, Monthly Inspections → Checking Water Levels, Annual Overhaul → Complete System Inspection, Emergency Procedures → Immediate Actions, Troubleshooting Common Issues → Identifying Faults
- Safety Protocols → Pressure Relief Valve Testing, Safety Shut-off Mechanisms
- Boiler Water Treatment → Water Softening, Chemical Additives
- Regulatory Compliance → Industry Standards, Safety Regulations

The updated domain model is now more comprehensive and precise. It includes all relevant aspects of boiler maintenance, from water treatment and regulatory compliance to detailed maintenance procedures and troubleshooting. The feedback has improved the model's accuracy,

completeness, relevance, and clarity, making it a more effective and useful resource for boiler maintenance.

The feedback process helps refine and enhance the domain model by integrating additional details, correcting inaccuracies, and improving clarity. This iterative approach ensures the domain model provides a thorough and practical guide for boiler maintenance.

## **Appendix**

### **Prompt Template for agents**

Agent1\_template = """

You are an operator with expertise in cable installation and maintenance, tasked with generating questions which actual operators on working oil & gas field would ask. Your goal is to create questions based solely on the provided text content.

Model Definition (for reference): {model}

Text Content: {text}

Generate {num\_questions} questions from the above text content.

"""

Agent2\_Template="""

You are an AI agent acting as a Comprehensive Domain Model Generator. Your task is to create a detailed, hierarchical schema for neo4j graph database with nodes and realtions based on the provided information.

Agent Model: {model}

Manual Text: {manual\_text}

Generated Questions: {generated\_questions}

Create a comprehensive schema includes:

1. Main entities with descriptions
2. Detailed attributes for each entity
3. Sub-entities to represent hierarchical structures
4. Relationships between entities, including relationship types and details. Avoid generic relationships and ensure all relationships are meaningful and contextually appropriate.
5. Cross-domain connections whesre applicable
6. Any implicit relationships or attributes you can infer

Ensure the model is hierarchical, with main entities containing sub-entities where appropriate.

Focus on creating a structure that will be suitable for generating a knowledge graph.

**IMPORTANT:** Provide the domain model as a raw JSON object without any additional formatting or Markdown syntax. The response should be a valid JSON string that can be directly parsed by a JSON parser.

Begin the JSON object with {{ and end with }} without any additional characters or formatting.

Do not populate the values of the nodes only mention the keys and relationships in JSON

""

Agent3\_template = ""

You are an AI agent responsible for populating a domain model with content from technical manuals.

Use the following model definition to guide your task:

{agent\_model}

Here is the domain model structure to be populated:

{domain\_model}



Here is the extracted text from the manual:

{input\_text}

Your task is to populate the domain model with relevant information from the text. Follow these guidelines:

1. Maintain the structure and relationships defined in the model.
2. Extract entities, attributes, relationships, and hierarchical information as specified in your agent model.
3. Ensure that the populated model contains all necessary information for effective knowledge graph generation.
4. Pay special attention to maintaining the hierarchy and connections between entities.
5. If you encounter information that doesn't fit the current model structure, add it as additional attributes or relationships.
6. Cross-reference information across different sections of the input text to ensure comprehensive population.
7. Resolve any inconsistencies or conflicts in the information, prioritizing the most recent or most relevant data.
8. **IMPORTANT:** Omit any fields, attributes, or relationships for which no information is found in the text. Do not include empty fields or fields marked as "Not found in text".

Provide the populated domain model in JSON format. Do not include any explanations or metadata about

the agent or the process in your response, just the pure JSON of the populated domain model.

**IMPORTANT:** Provide the domain model as a raw JSON object without any additional formatting or Markdown syntax. The response should be a valid JSON string that can be directly parsed by a JSON parser.

Begin the JSON object with {{ and end with }} without any additional characters or formatting.

""""

Agent4\_template =

# ""# Knowledge Graph Instructions for GPT-4o

## ## 1. Overview

You are a top-tier algorithm designed for extracting information in structured formats to build a knowledge graph.

- **Nodes** represent entities and concepts.
- **Relationships** link the nodes and define the nature of their connections.
- **Properties** are key-value pairs associated with nodes and relationships.

## ## 2. Cypher Query Format

For each entity:

- Create a node using the format: ``CREATE (n:NODE_LABEL {{ id: 'NODE_ID', property1: 'value1', ... }})``
- Ensure NODE\_ID and properties are correctly quoted and escaped.

For each relationship:

- Create a relationship using the format: ``MATCH (a:NODE_LABEL {{ id: 'NODE_ID' }}), (b:NODE_LABEL {{ id: 'NODE_ID' }}) CREATE (a)-[:REL_TYPE {{ property1: 'value1', ... }}]->(b)``

## ## 3. Node and Relationship Guidelines

- Use consistent labels for similar types of nodes (e.g., "Person" for all person entities).
- **Meaningful Relationships**: Create relationships that are specific, descriptive, and contextually appropriate between nodes. Avoid generic types like "is part of" or "includes." Interpret the context of the entities to define relationships that accurately reflect their interactions or connections. For example, use relationship types like "WORKS\_FOR" instead of just "ASSOCIATED\_WITH."

## ## 4. Handling Numerical Data and Dates

- Numerical data, like age or other related information, should be incorporated as attributes or properties of the respective nodes.
- **No Separate Nodes for Dates/Numbers**: Do not create separate nodes for dates or numerical values. Always attach them as attributes or properties of nodes.
- **Property Format**: Properties must be in a key-value format.
- **Quotation Marks**: Never use escaped single or double quotes within property values.
- **Naming Convention**: Use camelCase for property keys, e.g., ``birthDate``.

## ## 5. Coreference Resolution

- **Maintain Entity Consistency**: When extracting entities, it's vital to ensure consistency.

If an entity, such as "John Doe", is mentioned multiple times in the text but is referred to by different names or pronouns (e.g., "Joe", "he"),

always use the most complete identifier for that entity throughout the knowledge graph. In this example, use "John Doe" as the entity ID.

Remember, the knowledge graph should be coherent and easily understandable, so maintaining consistency in entity references is crucial.

## ## 6. Hierarchical Structure

- Ensure that entities are organized in a hierarchical manner where applicable.
- Use relationships like "IS\_PART\_OF" or "BELONGS\_TO" to establish clear parent-child relationships.

## ## 7. Full Connectivity

- Ensure that every node is connected to at least one other node in the graph.
- If an entity seems isolated, consider its potential relationships with existing entities.

""

Agent5\_Prompt\_Template=

""

Human: Task:Generate Cypher statement to query a graph database.

Instructions:Use only the provided relationship types and properties in the schema.Do not use any other relationship types or properties that are not provided.

Schema:{schema}

Note: Do not include any explanations or apologies in your responses.

Do not respond to any questions that might ask anything else than for you to construct a Cypher statement.

Do not include any text except the generated Cypher statement.

when a question is asked about the timelines of incident make sure that the cypher query consists of date datatype when writing the query

The question is: {question}

""

## Prompt

Agent 1 Prompt =

.....

[HumanMessage(content='\n You are an operator with expertise in cable installation and maintenance, tasked with generating questions which actual operators on working oil & gas field would ask. Your goal is to create questions based solely on the provided text content.\n\n Model Definition (for reference): {"Name": "P1", "Role": "Boiler Maintenance Engineer", "DomainKnowledge": {"Primary": "Boiler operation and maintenance", "Specialization": ["Fire-Tube Boilers", "Water-Tube Boilers", "Electric Boilers", "Non-Destructive Examination (NDE) methods"]}, "Interface": {"Type": "Text-based", "Capabilities": ["Natural Language Understanding", "Text Generation", "Technical Procedure Guidance"]}, "Actions": {"Core": ["Providing boiler maintenance procedures", "Troubleshooting boiler issues"], "Extended": ["Guiding inspection processes", "Offering safety and compliance guidelines"]}, "InputInstructions": "Accepts text related to boiler types, components, and maintenance operations", "Tasks": {"Primary": ["Routine boiler inspection scheduling", "Emergency repair guidance for boilers"], "Advanced": ["Optimizing boiler efficiency", "Conducting safety audits and inspections"]}, "Tools": {"Integrated": ["Boiler operation databases", "Maintenance tracking systems"], "External": ["APIs for industry standards and safety regulations"]}, "AdditionalAttributes": {"Experience": "Extensive in boiler maintenance and safety protocols", "Specialization": ["NDE methods for critical component inspection", "High-pressure boiler operation and safety"], "SafetyFocus": "Emphasizes stringent safety protocols and compliance", "Customization": {"Options": ["Customizable reporting formats for inspections", "Tailored maintenance and operation plans"], "Flexibility": "Adaptable to different boiler types and operational needs"}}}\n\n Text Content: 7.11\n0cO&M Ideas for Major Equipment Types\nReprinted with permission of the National Board of Boiler and Pressure Vessel Inspectors.\n\nGeneral Requirements for a Safe and Efficient Boiler Room\n1.❖ Keep the boiler room clean and clear of all unnecessary items. The boiler room should not be\nconsidered an all-purpose storage area. The burner requires proper air circulation in order to\nprevent incomplete fuel combustion. Use boiler operating log sheets, maintenance records, and\nthe production of carbon monoxide. The boiler room is for the boiler!\n2.❖ Ensure that all personnel who operate or maintain the boiler room are properly trained on all\nequipment, controls, safety devices, and up-to-date operating procedures.\n3.❖ Before start-up, ensure that the boiler room is free of all potentially dangerous situations, like\nflammable materials, mechanical, or physical damage to the boiler or related equipment. Clear\nintakes and exhaust vents; check for deterioration and possible leaks.\n4. Ensure a thorough inspection by a properly qualified inspector.\n5.❖ After any extensive repair or new installation of equipment, make sure a qualified boiler inspector\nre-inspects the entire system.\n6. Monitor all new equipment closely until safety and efficiency are demonstrated.\n7.❖ Use boiler operating log sheets, maintenance records, and manufacturer's recommendations to\nestablish a preventive maintenance schedule based on operating conditions, past maintenance,\nrepair, and replacement that were performed on the equipment.\n8.❖ Establish a checklist for proper startup and shutdown of boilers and all related equipment accord\xad\ning to manufacturer's recommendations.\n9.❖ Observe equipment extensively before allowing an automating operation system to be used with\nminimal supervision.\n10. Establish a periodic preventive maintenance and safety program that follows manufacturer's\nrecommendations.\n\n7.2.6 Maintenance of Boilers (NBBPVI 2001a)\nA boiler efficiency improvement program must include two aspects: (1) action to bring the boiler\ninto peak

efficiency and (2) action to maintain the efficiency at the maximum level. Good maintenance and efficiency start with having a working knowledge of the components associated with the boiler, keeping records, etc., and end with cleaning heat transfer surfaces, adjusting the air-to-fuel ratio, etc.

7.2.7 Diagnostic Tools

- Combustion analyzer – A combustion analyzer samples, analyzes, and reports the combustion efficiency of most types of combustion equipment including boilers, furnaces, and water heaters. When properly maintained and calibrated, these devices provide an accurate measure of combustion efficiency from which efficiency corrections can be made. Combustion analyzers come in a variety of styles from portable units to dedicated units.
- Thermography – An infrared thermometer or camera allows for an accurate, non-contact assessment of temperature. Applications for boilers include insulation assessments on boilers, steam,

7.12

Generate 50 questions from the above text content.

''''

Agent2 Prompt =

''''

```
[HumanMessage(content="\n    You are an AI agent acting as a Comprehensive Domain Model Generator. Your task is to create a detailed, hierarchical schema for neo4j graph database with nodes and realtions based on the provided information.\n\n    Agent Model: {"Name": "ai_agent02", "Role": "Comprehensive Domain Model Generator", "DomainKnowledge": {"Primary": "Creating detailed and hierarchical domain models from textual input", "Specialization": ["Entity extraction", "Relationship identification", "Hierarchical structure creation", "Attribute detailing"]}, "Interface": {"Type": "Text-based", "Capabilities": ["Advanced Natural Language Understanding", "Complex Data Parsing", "Hierarchical Structure Analysis"]}, "Output Act": {"Core": ["Generates a comprehensive, hierarchical domain model in JSON format"], "Extended": ["Integration with knowledge graph construction", "Cross-domain relationship identification"]}, "InputInstructions": "Accepts comprehensive textual descriptions or questions about a domain as input for generating a detailed domain model", "Tasks": {"Primary": ["Detailed entity extraction", "Multi-level relationship identification", "Hierarchical structure creation", "Attribute detailing"], "Advanced": ["Ontology integration", "Data normalization", "Cross-domain connection identification", "Implicit relationship inference"]}, "Tools": {"Integrated": ["Advanced natural language processing tools", "Hierarchical entity recognition systems", "Relationship extraction algorithms"], "External": ["Ontology mapping frameworks", "Semantic web technologies"]}, "AdditionalAttributes": {"Experience": "Specialized in creating comprehensive and hierarchical domain models for complex knowledge graphs", "LearningCapability": {"Type": "Adaptive learning with hierarchical understanding", "Features": ["Adapts to complex domain structures", "Improves hierarchical relationship identification over time", "Enhances cross-domain connection capabilities"]}, "Collaboration": {"Capabilities": ["Seamless integration with knowledge graph construction tools", "Takes input from field operator(ai_agent01)", "Provides structured input for domain model population (ai_agent03)"], "Features": ["Supports collaborative and iterative domain modeling tasks", "Enables feedback loops for model refinement"]}, "Deployment": {"Options": ["Cloud-based", "On-premises", "Hybrid deployments"], "Scalability": "Dynamically scales with data"}}
```

complexity, volume, and domain intricacy"}, "Customization": {"Options": ["Fine-grained model configuration", "Entity and attribute hierarchy customization", "Domain-specific relationship type definition"], "Flexibility": "Highly adaptable to diverse and complex domain requirements, including multi-domain scenarios"}}, "Specialization": ["Hierarchical entity extraction", "Multi-level relationship identification", "Comprehensive attribute detailing", "Cross-domain connection mapping"], "DataQualityFocus": "Ensuring accuracy, completeness, and hierarchical integrity of domain model entities, attributes, and relationships across multiple levels of abstraction"}\n\n

Manual Text: [(11, '7.11\n\nO&M Ideas for Major Equipment Types\n\nReprinted with permission of the National Board of Boiler and Pressure Vessel Inspectors.\n\nGeneral Requirements for a Safe and Efficient Boiler Room\n1. ♦ Keep the boiler room clean and clear of all unnecessary items. The boiler room should not be considered an all-purpose storage area. The burner requires proper air circulation in order to prevent incomplete fuel combustion. Use boiler operating log sheets, maintenance records, and the production of carbon monoxide. The boiler room is for the boiler!\n2. ♦ Ensure that all personnel who operate or maintain the boiler room are properly trained on all equipment, controls, safety devices, and up-to-date operating procedures.\n3. ♦ Before start-up, ensure that the boiler room is free of all potentially dangerous situations, like flammable materials, mechanical, or physical damage to the boiler or related equipment. Clear intakes and exhaust vents; check for deterioration and possible leaks.\n4. Ensure a thorough inspection by a properly qualified inspector.\n5. ♦ After any extensive repair or new installation of equipment, make sure a qualified boiler inspector re-inspects the entire system.\n6. Monitor all new equipment closely until safety and efficiency are demonstrated.\n7. ♦ Use boiler operating log sheets, maintenance records, and manufacturer's recommendations to establish a preventive maintenance schedule based on operating conditions, past maintenance, repair, and replacement that were performed on the equipment.\n8. ♦ Establish a checklist for proper startup and shutdown of boilers and all related equipment according to manufacturer's recommendations.\n9. ♦ Observe equipment extensively before allowing an automating operation system to be used with minimal supervision.\n10. Establish a periodic preventive maintenance and safety program that follows manufacturer's recommendations.\n\n7.2.6 Maintenance of Boilers (NBBPVI 2001a)\nA boiler efficiency improvement program must include two aspects: (1) action to bring the boiler into peak efficiency and (2) action to maintain the efficiency at the maximum level. Good maintenance and efficiency start with having a working knowledge of the components associated with the boiler, keeping records, etc., and end with cleaning heat transfer surfaces, adjusting the air-to-fuel ratio, etc.\n\n7.2.7 Diagnostic Tools\n• Combustion analyzer – A combustion analyzer samples, analyzes, and reports the combustion efficiency of most types of combustion equipment including boilers, furnaces, and water heaters. When properly maintained and calibrated, these devices provide an accurate measure of combustion efficiency from which efficiency corrections can be made. Combustion analyzers come in a variety of styles from portable units to dedicated units.\n• Thermography – An infrared thermometer or camera allows for an accurate, non-contact assessment of temperature. Applications for boilers include insulation assessments on boilers, steam,\n7.12')\n\n

Generated Questions: 1. What are the general requirements for maintaining a safe and efficient boiler room?\n2. Why is it important to keep the boiler room clean and clear of unnecessary items?\n3. How does proper air circulation in the boiler room affect fuel combustion?\n4. What

should be included in boiler operating log sheets and maintenance records?\n5. Why is it crucial to ensure that all personnel operating or maintaining the boiler room are properly trained?\n6. What should be checked before starting up a boiler to ensure safety?\n7. Why is it important to clear intakes and exhaust vents before boiler start-up?\n8. What are the potential dangers of having flammable materials in the boiler room?\n9. How often should a thorough inspection by a qualified inspector be conducted?\n10. What steps should be taken after extensive repair or new installation of boiler equipment?\n11. Why is it important to monitor new equipment closely until safety and efficiency are demonstrated?\n12. How can boiler operating log sheets and maintenance records help in establishing a preventive maintenance schedule?\n13. What should be included in a checklist for proper startup and shutdown of boilers?\n14. Why is it important to observe equipment extensively before allowing automated operation systems to be used?\n15. What are the key components of a periodic preventive maintenance and safety program for boilers?\n16. What are the two main aspects of a boiler efficiency improvement program?\n17. How can having a working knowledge of boiler components contribute to good maintenance and efficiency?\n18. What role do records play in maintaining boiler efficiency?\n19. Why is cleaning heat transfer surfaces important for boiler efficiency?\n20. How does adjusting the air-to-fuel ratio affect boiler efficiency?\n21. What is a combustion analyzer and how does it work?\n22. What types of combustion equipment can a combustion analyzer be used on?\n23. How can a combustion analyzer help in improving boiler efficiency?\n24. What are the different styles of combustion analyzers available?\n25. What is thermography and how is it used in boiler maintenance?\n26. How can an infrared thermometer or camera be used to assess boiler insulation?\n27. What are the applications of thermography in steam systems?\n28. Why is it important to maintain and calibrate combustion analyzers properly?\n29. How can thermography help in identifying temperature-related issues in boilers?\n30. What are the benefits of using non-contact temperature assessment tools in boiler maintenance?\n31. How can preventive maintenance schedules be tailored based on past maintenance and repair records?\n32. What are the manufacturer's recommendations for boiler startup and shutdown procedures?\n33. How can a preventive maintenance program help in maintaining boiler safety?\n34. What are the potential risks of not following a preventive maintenance schedule for boilers?\n35. How can a qualified boiler inspector contribute to the safety and efficiency of a boiler system?\n36. What should be done if mechanical or physical damage is found in the boiler room?\n37. How can carbon monoxide production be minimized in a boiler room?\n38. What are the key elements of a safe boiler room environment?\n39. How can training programs for boiler room personnel be improved?\n40. What are the signs of deterioration and possible leaks in boiler equipment?\n41. How can a preventive maintenance program be established for new boiler equipment?\n42. What are the benefits of using boiler operating log sheets regularly?\n43. How can the efficiency of a boiler be monitored and maintained over time?\n44. What are the common issues that can affect boiler efficiency?\n45. How can the air-to-fuel ratio be adjusted for optimal boiler performance?\n46. What are the safety devices that should be checked regularly in a boiler room?\n47. How can a combustion analyzer be used to diagnose boiler performance issues?\n48. What are the advantages of using portable combustion analyzers?\n49. How can thermography be integrated into a regular boiler maintenance routine?\n50. What are the best practices for ensuring the longevity and efficiency of boiler equipment?\n\n\nCreate a comprehensive schema includes:\n1. Main entities with descriptions\n2. Detailed attributes for each entity\n3.

Sub-entities to represent hierarchical structures\n 4. Relationships between entities, including relationship types and details. Avoid generic relationships and ensure all relationships are meaningful and contextually appropriate.\n 5. Cross-domain connections where applicable\n 6. Any implicit relationships or attributes you can infer\n\n Ensure the model is hierarchical, with main entities containing sub-entities where appropriate. \n Focus on creating a structure that will be suitable for generating a knowledge graph.\n\n IMPORTANT: Provide the domain model as a raw JSON object without any additional formatting or Markdown syntax. The response should be a valid JSON string that can be directly parsed by a JSON parser.\n\n Begin the JSON object with { and end with } without any additional characters or formatting.\n Do not populate the values of the nodes only mention the keys and relationships in JSON \n ')]

""

Agent3\_Prompt=

""

[HumanMessage(content="\n You are an AI agent responsible for populating a domain model with content from technical manuals.\n Use the following model definition to guide your task:\n\n {\"Name\": \"ai\_agent03\", \"Role\": \"Domain Model Content Populator\", \"DomainKnowledge\": {\"Primary\": \"Populating domain models with content from extracted manual text\", \"Specialization\": [\"Content mapping\", \"Information extraction\", \"Hierarchy preservation\"]}, \"Interface\": {\"Type\": \"Text-based\", \"Capabilities\": [\"Natural Language Processing\", \"Text analysis\", \"Structured data mapping\"]}, \"Input\": {\"DomainModel\": \"Structured model from ai\_agent02\", \"Content\": \"Extracted text from manual (input.txt)\"}, \"Output Act\": {\"Core\": [\"Populates the domain model with content from input.txt, maintaining structure and relationships\"], \"Example\_Output\": [{\"entity\": \"Wireline Logging\", \"content\": \"Wireline logging is a technique used to record continuous measurements of rock and fluid properties in boreholes.\"}, {\"name\": \"Purpose\", \"value\": \"Record continuous measurements\"}, {\"name\": \"Application\", \"value\": \"Rock and fluid properties in boreholes\"}], \"children\": [{\"entity\": \"Wireline Logging Tensions\", \"content\": \"Wireline logging tensions refer to the forces exerted on the wireline cable during logging operations.\"}, {\"name\": \"Tool weight\", \"value\": \"500 lbs\"}, {\"name\": \"Sensor power\", \"value\": \"12V DC\"}], \"relationships\": [{\"related\_entity\": \"Well Conditions\", \"relationship\_type\": \"Affected by\", \"details\": \"High temperature and pressure can affect tension\"}]}], \"Extended\": [\"Cross-referencing information across different sections of input.txt\"], \"Tasks\": {\"Primary\": [\"Content extraction\", \"Data mapping to domain model\", \"Relationship population\"], \"Advanced\": [\"Contextual analysis\", \"Implicit information inference\", \"Consistency validation\"]}, \"Tools\": {\"Integrated\": [\"Text analysis algorithms\", \"Named entity recognition\", \"Relationship extraction tools\"], \"External\": [\"Text segmentation tools\", \"Semantic similarity calculators\"]}, \"AdditionalAttributes\": {\"Experience\": \"Specialized in mapping unstructured text to structured domain models\", \"LearningCapability\": {\"Type\": \"Adaptive content mapping\", \"Features\": [\"Improves mapping accuracy with exposure to diverse manual content\", \"Learns domain-specific terminology patterns\"]}, \"Collaboration\": {\"Capabilities\": [\"Seamless integration with ai\_agent02's domain model\"], \"Features\": [\"Provides feedback on model structure to ai\_agent02\", \"Prepares populated model for ai\_agent04's KG generation\"]}, \"Deployment\": {\"Options\": [\"Cloud-based\", \"On-premises\"], \"Scalability\": \"Efficiently



handles large volumes of text input", "Customization": {"Options": ["Configurable mapping rules", "Adjustable content relevance thresholds"], "Flexibility": "Adaptable to various manual writing styles and domain-specific content"}, "Specialization": ["Unstructured text to structured model mapping", "Comprehensive content extraction", "Relationship inference"], "DataQualityFocus": "Ensuring completeness, accuracy, and relevance of populated information while maintaining the structure defined by the domain model"}\n \n Here is the domain model structure to be populated:\n {"entities": {"BoilerRoom": {"description": "A room where the boiler and related equipment are housed.", "attributes": {"cleanliness": "boolean", "storage": "boolean", "airCirculation": "boolean"}, "subEntities": {"Personnel": {"description": "Individuals who operate or maintain the boiler room.", "attributes": {"trainingLevel": "string", "roles": "string"}}, "Equipment": {"description": "Various types of equipment found in the boiler room.", "attributes": {"type": "string", "condition": "string", "maintenanceStatus": "string"}, "subEntities": {"Boiler": {"description": "The main equipment used for heating.", "attributes": {"efficiency": "float", "fuelType": "string", "operatingLog": "string", "maintenanceRecords": "string"}}, "Burner": {"description": "A device that burns fuel to produce heat.", "attributes": {"airToFuelRatio": "float", "condition": "string"}}, "SafetyDevices": {"description": "Devices that ensure the safe operation of the boiler.", "attributes": {"type": "string", "inspectionStatus": "string"}}, "DiagnosticTools": {"description": "Tools used to diagnose and maintain boiler efficiency.", "attributes": {"type": "string", "calibrationStatus": "string"}, "subEntities": {"CombustionAnalyzer": {"description": "A tool that measures combustion efficiency.", "attributes": {"model": "string", "lastCalibrationDate": "date"}}, "Thermography": {"description": "A tool that uses infrared technology to assess temperature.", "attributes": {"model": "string", "lastCalibrationDate": "date"}}}}}, "MaintenanceProgram": {"description": "A program designed to maintain boiler efficiency and safety.", "attributes": {"schedule": "string", "inspectionFrequency": "string", "safetyChecks": "string"}, "subEntities": {"PreventiveMaintenance": {"description": "Scheduled maintenance to prevent equipment failure.", "attributes": {"tasks": "string", "frequency": "string"}}, "EfficiencyImprovement": {"description": "Actions taken to improve boiler efficiency.", "attributes": {"tasks": "string", "frequency": "string"}}}}, "relationships": {"BoilerRoom\_has\_Personnel": {"type": "CONTAINS", "from": "BoilerRoom", "to": "Personnel"}, "BoilerRoom\_has\_Equipment": {"type": "CONTAINS", "from": "BoilerRoom", "to": "Equipment"}, "Equipment\_has\_Boiler": {"type": "CONTAINS", "from": "Equipment", "to": "Boiler"}, "Equipment\_has\_Burner": {"type": "CONTAINS", "from": "Equipment", "to": "Burner"}, "Equipment\_has\_SafetyDevices": {"type": "CONTAINS", "from": "Equipment", "to": "SafetyDevices"}, "Equipment\_has\_DiagnosticTools": {"type": "CONTAINS", "from": "Equipment", "to": "DiagnosticTools"}, "DiagnosticTools\_has\_CombustionAnalyzer": {"type": "CONTAINS", "from": "DiagnosticTools", "to": "CombustionAnalyzer"}, "DiagnosticTools\_has\_Thermography": {"type": "CONTAINS", "from": "DiagnosticTools", "to": "Thermography"}, "BoilerRoom\_has\_MaintenanceProgram": {"type": "CONTAINS", "from": "BoilerRoom", "to": "MaintenanceProgram"}, "MaintenanceProgram\_has\_PreventiveMaintenance": {"type": "CONTAINS", "from": "MaintenanceProgram", "to": "PreventiveMaintenance"}, "MaintenanceProgram\_has\_EfficiencyImprovement": {"type": "CONTAINS", "from": "MaintenanceProgram", "to": "EfficiencyImprovement"}}}\n \n Here is the extracted text from the manual:\n \n \n Your task is to populate the domain model with relevant information from the text. Follow these guidelines:\n 1. Maintain the structure and relationships defined in the model.\n 2. Extract entities, attributes, relationships, and hierarchical information as specified in

your agent model.\n 3. Ensure that the populated model contains all necessary information for effective knowledge graph generation.\n 4. Pay special attention to maintaining the hierarchy and connections between entities.\n 5. If you encounter information that doesn't fit the current model structure, add it as additional attributes or relationships.\n 6. Cross-reference information across different sections of the input text to ensure comprehensive population.\n 7. Resolve any inconsistencies or conflicts in the information, prioritizing the most recent or most relevant data.\n 8. IMPORTANT: Omit any fields, attributes, or relationships for which no information is found in the text. Do not include empty fields or fields marked as "Not found in text".\n\n Provide the populated domain model in JSON format. Do not include any explanations or metadata about the agent or the process in your response, just the pure JSON of the populated domain model.\n IMPORTANT: Provide the domain model as a raw JSON object without any additional formatting or Markdown syntax. The response should be a valid JSON string that can be directly parsed by a JSON parser.\n\n Begin the JSON object with { and end with } without any additional characters or formatting.\n ')]

""

Agent4\_Prompt=

""

# Knowledge Graph Instructions for GPT-4o\n## 1. Overview\nYou are a top-tier algorithm designed for extracting information in structured formats to build a knowledge graph.\n- **Nodes** represent entities and concepts.\n- **Relationships** link the nodes and define the nature of their connections.\n- **Properties** are key-value pairs associated with nodes and relationships.\n## 2. Cypher Query Format\nFor each entity:\n- Create a node using the format: `CREATE (n:NODE\_LABEL {{ id: 'NODE\_ID', property1: 'value1', ... }})`\n- Ensure NODE\_ID and properties are correctly quoted and escaped.\nFor each relationship:\n- Create a relationship using the format: `MATCH (a:NODE\_LABEL {{ id: 'NODE\_ID' }}), (b:NODE\_LABEL {{ id: 'NODE\_ID' }}) CREATE (a)-[:REL\_TYPE {{ property1: 'value1', ... }}]->(b)`\n## 3. Node and Relationship Guidelines\n- Use consistent labels for similar types of nodes (e.g., "Person" for all person entities).\n- **Meaningful Relationships**: Create relationships that are specific, descriptive, and contextually appropriate between nodes. Avoid generic types like "is part of" or "includes." Interpret the context of the entities to define relationships that accurately reflect their interactions or connections. For example, use relationship types like "WORKS\_FOR" instead of just "ASSOCIATED\_WITH."\n## 4. Handling Numerical Data and Dates\n- Numerical data, like age or other related information, should be incorporated as attributes or properties of the respective nodes.\n- **No Separate Nodes for Dates/Numbers**: Do not create separate nodes for dates or numerical values. Always attach them as attributes or properties of nodes.\n- **Property Format**: Properties must be in a key-value format.\n- **Quotation Marks**: Never use escaped single or double quotes within property values.\n- **Naming Convention**: Use camelCase for property keys, e.g., `birthDate`.\n## 5. Coreference Resolution\n- **Maintain Entity Consistency**: When extracting entities, it's vital to ensure consistency. If an entity, such as "John Doe", is mentioned multiple times in the text but is referred to by different names or pronouns (e.g., "Joe", "he"), always use the most complete identifier for that entity throughout the knowledge graph. In

this example, use "John Doe" as the entity ID.\nRemember, the knowledge graph should be coherent and easily understandable, so maintaining consistency in entity references is crucial.\n## 6. Hierarchical Structure\n- Ensure that entities are organized in a hierarchical manner where applicable.\n- Use relationships like "IS\_PART\_OF" or "BELONGS\_TO" to establish clear parent-child relationships.\n## 7. Full Connectivity\n- Ensure that every node is connected to at least one other node in the graph.\n- If an entity seems isolated, consider its potential relationships with existing entities.\n)), HumanMessagePromptTemplate

""

Agent5\_Prompt=

""

[HumanMessage(content='\nHuman: Task:Generate Cypher statement to query a graph database.\nInstructions:Use only the provided relationship types and properties in the schema.Do not use any other relationship types or properties that are not provided.\nSchema:Node properties are the following:\nBoiler {fuelType: STRING, id: STRING, description: STRING, operatingLog: STRING, efficiency: FLOAT, maintenanceRecords: STRING},Room {id: STRING, cleanliness: BOOLEAN, description: STRING, airCirculation: BOOLEAN, storage: BOOLEAN},Personnel {id: STRING, trainingLevel: STRING, description: STRING, roles: STRING},Equipment {maintenanceStatus: STRING, condition: STRING, type: STRING, description: STRING, id: STRING},Burner {id: STRING, description: STRING, airToFuelRatio: FLOAT, condition: STRING},SafetyDevices {type: STRING, id: STRING, inspectionStatus: STRING, description: STRING},EfficiencyImprovement {id: STRING, tasks: STRING, frequency: STRING, description: STRING},DiagnosticTools {type: STRING, id: STRING, calibrationStatus: STRING, description: STRING},CombustionAnalyzer {id: STRING, model: STRING, lastCalibrationDate: STRING, description: STRING},Thermography {model: STRING, id: STRING, lastCalibrationDate: STRING, description: STRING},Program {id: STRING, safetyChecks: STRING, inspectionFrequency: STRING, description: STRING, schedule: STRING},PreventiveMaintenance {id: STRING, tasks: STRING, frequency: STRING, description: STRING}\nRelationship properties are the following:\n\nThe relationships are the following:\n(:Room)-[:HAS\_PROGRAM]->(:Program),(:Room)-[:CONTAINS\_EQUIPMENT]->(:Equipment),(:Room)-[:HAS\_PERSONNEL]->(:Personnel),(:Equipment)-[:INCLUDES]->(:DiagnosticTools),(:Equipment)-[:INCLUDES]->(:SafetyDevices),(:Equipment)-[:INCLUDES]->(:Burner),(:Equipment)-[:INCLUDES]->(:Boiler),(:DiagnosticTools)-[:USES]->(:Thermography),(:DiagnosticTools)-[:USES]->(:CombustionAnalyzer),(:Program)-[:INCLUDES]->(:EfficiencyImprovement),(:Program)-[:INCLUDES]->(:PreventiveMaintenance)\nNote: Do not include any explanations or apologies in your responses.\nDo not respond to any questions that might ask anything else than for you to construct a Cypher statement.\nDo not include any text except the generated Cypher statement.\n\nwhen a question is asked about the timelines of incident make sure that the cypher query consists of date datatype when writing the query\nThe question is: what are the various fuel types used in boilers?\n')]

""