# Problem Statement:

**Mission Brief: Advanced Natural Language Query System for Naval Intelligence**

**Role:** AI Engineer, Central Intelligence Team, Navy

**Objective:**

Harness your expertise to engineer an advanced Question Answering (QA) system. This system is designed to interpret succinct queries in English and derive specific data from provided datasets, crucial for strategic naval operations.

**Assignment Details:**

Your mission involves two primary datasets and corresponding explanatory documents. Your challenge is to design a QA system that understands natural language inputs and outputs structured data responses. This system will be a cornerstone for internal operations, enhancing our data retrieval capabilities with precision.

**Scenario Example:**

Consider the datasets: ship\_data and grid\_details. The ship\_data includes columns such as ship\_name and ship\_priority.

For a query like, "What is the priority of the ship Nebula?" your system should output:

**{**

**'select\_column' : 'ship\_priority',**

**'from\_table' : 'ship\_data',**

**'where': [{**

**'column': 'ship\_name',**

**'relation': '=',**

**'value' : ‘nebula’}]**

**}**

**Output Specification:**

Your system should generate a Python dictionary with:

* select\_column: Data to be retrieved.
* from\_table: Dataset containing the necessary data.
* where: List of conditions to refine the search.

Each condition within where is a dictionary specifying the column, relation, and value.

**Development Resources:**

You will be provided with dummy datasets for development purposes. These include:

* Two main datasets.
* Two documents detailing the datasets.
* Sample question-answer pairs for familiarising with the response format.

**System Requirements:**

* The system must run locally. External cloud APIs are strictly prohibited.
* Design the system to be scalable. Avoid hardcoding and reliance solely on regex-based solutions.
* The system should be executable from the terminal, with no UI components.

**Evaluation Protocol:**

Your system will be assessed using an undisclosed dataset to ensure its effectiveness in real-world scenarios.