**Proposed Title:** Language Guided Navigation: A VLM Approach

**Research Question:**

How can vision-language models (VLMs) be integrated into robotic navigation systems to enhance their understanding and navigation of complex environments based on human-provided natural language instructions?

**Goal**:

* Develop agent’s navigation system that can effectively interpret and follow natural language instructions.
* Improve the agent’s ability to understand and reason about spatial relationships between objects.
* Enhance the agent's ability to understand tasks based on human-provided natural language descriptions and visual cues, ensuring that humans can maintain control over the robots while increasing the robots' independence.

**Approach**:

* Utilize a pre-trained VLM: Leverage a SOTA VLM (e.g., CLIP, ViT) to provide strong visual-language understanding capabilities.
* Integrate VLM with robotic navigation framework: Connect the VLM to a robotic navigation system, such as ROS or ROS2.
* Develop a task-oriented dialogue system: Create a dialogue system that can understand and respond to natural language instructions related to navigation tasks.
* Implement spatial reasoning: Use the VLM to reason about spatial relationships between objects and landmarks, enabling the robot to plan efficient paths.
* Test and evaluate: Evaluate the performance of the system on various navigation tasks.

**Novel Contributions:**

1. Development of a dialogue system specifically tailored for robotic navigation tasks, allowing the robot to interpret and follow complex, multi-step instructions in natural language.

2. Enhanced spatial reasoning capabilities within a vision-language model for improved path planning and object recognition in dynamic environments.

**Experiment Setup:**

1. Datasets: Utilize synthetic datasets (e.g., AI2-THOR for 3D environments) for initial training then fine-tune on real-world datasets such as Google’s Robotics Object Search, which involves searching for and identifying objects in real-world settings.

2. Metrics:

- Success Rate: Measure the percentage of successfully completed tasks.

- Task Completion Time: Evaluate the efficiency of task completion, specifically how quickly and effectively the robot can navigate to a target or perform a task.

- Spatial Reasoning Accuracy: Assess the system's ability to understand and navigate spatial relationships between objects in the environment.

3. Baseline Model:

Use a SOTA vision-language model such as CLIP or ViLBERT, in combination with well-established navigation algorithms like RRT (Rapidly-exploring Random Trees) or A\* Pathfinding, to provide a foundation for comparison.

**Steps to Evaluate the Idea:**

1. Sim2Real Transfer: Train the system in simulated environments (e.g., AI2-THOR) to handle diverse navigation and object search tasks, then transfer it to real-world scenarios with varying lighting, object arrangements, and environmental conditions to evaluate generalization and adaptability.

2. Human Control & Input: Test the robot’s ability to balance autonomy with human control, ensuring that humans can intervene or guide the robot when necessary. This will assess how well the robot adapts to real-time human input and oversight.