**Steps To Follow**

### **Project Steps**

1. **Task Selection**
   * Choose reasoning tasks (arithmetic, logic puzzles, commonsense reasoning, multi-step QA, math word problems).
2. **Prompt Design**
   * Apply different prompting methods:  
     + Chain-of-Thought (step-by-step reasoning)
     + Self-Consistency (multiple reasoning paths)
     + Reflection (double-checking answers)
3. **Model Execution**
   * Run Qwen3-4B on the selected tasks with the designed prompts.
   * Collect both final answers and intermediate reasoning steps.
4. **Evaluation**
   * Measure **correctness** (final answer accuracy).
   * Assess **reasoning quality** (logical coherence of steps).
   * Identify **error types** (hallucination, skipped steps, arithmetic mistakes, logical jumps).
5. **Comparative Analysis**
   * Optionally test a stronger LLM (e.g., GPT-4, Llama-3 70B) on the same tasks.
   * Compare reasoning depth, correctness, and error patterns.
6. **Insights & Conclusion**
   * Summarize findings about Qwen3-4B’s reasoning capabilities.
   * Highlight strengths, weaknesses, and implications for interpretability and reliability of mid-sized LLMs.