Of course! Let’s build a **clear roadmap** by **dividing your project into blocks** (almost like stages or modules).  
This way, you'll know exactly **what steps to follow**, **what you need to learn or build at each stage**, and **how they connect**.

**📚 Roadmap for Your Bachelor's Project**

**Title: Semantic Analysis and Conceptual Title Generation from Long Texts Using LLMs**

**🧩 Block 1: Foundations (Learning & Research Phase)**

**Goals**:

* Understand LLMs (GPT, BERT, etc.)
* Study text summarization techniques (extractive and abstractive)
* Learn semantic analysis (embeddings, topic modeling, semantic similarity)

**Tasks**:

* Read about Transformers, LLMs (e.g., BERT, GPT, T5, PEGASUS).
* Study summarization techniques (Pointer-Generator, BERTSum, PEGASUS, Longformer, etc.).
* Learn about semantic embeddings (Sentence-BERT, Word2Vec) and topic models (LDA, BERTopic).
* Try small experiments (e.g., summarizing short texts with HuggingFace models).

**🧩 Block 2: Dataset Preparation**

**Goals**:

* Find or create a dataset of long documents.
* Preprocess the dataset: cleaning, formatting.

**Tasks**:

* Download large text datasets (Wikipedia dumps, arXiv papers, books, etc.)
* Clean texts: remove noise (ads, footers, irrelevant info).
* Decide on annotation:
  + If available: use existing summaries/titles.
  + If not: prepare your own manually for small samples.

**🧩 Block 3: Chunking Long Documents**

**Goals**:

* Handle documents longer than model input limits (e.g., 4K-32K tokens).

**Tasks**:

* Implement **Semantic Chunking**:
  + Split texts based on meaning, not fixed size.
  + Use Sentence-BERT or embeddings to find semantic boundaries.
* Overlap chunks slightly to maintain context between parts.

**🧩 Block 4: Summarization of Chunks**

**Goals**:

* Summarize each chunk independently to get local summaries.

**Tasks**:

* Fine-tune or prompt an LLM (e.g., GPT-3.5, T5, BART, Pegasus).
* For each chunk:
  + Perform summarization.
  + Focus on preserving key concepts, avoiding redundancy.

**🧩 Block 5: Global Merging and Semantic Analysis**

**Goals**:

* Merge chunk summaries into one unified overview.
* Extract key topics, viewpoints, beliefs from the global summary.

**Tasks**:

* Apply:
  + **Topic modeling** (e.g., BERTopic) on combined summaries.
  + **Semantic clustering** (Sentence-BERT + KMeans) for theme extraction.
* Analyze author's viewpoint: polarity, opinionated statements.

**🧩 Block 6: Title (Conceptual Headline) Generation**

**Goals**:

* Generate a smart, semantic-rich title based on the global summary.

**Tasks**:

* Fine-tune a headline-generation model OR use prompting.
* Examples:
  + "Based on this global summary, generate a 5-word conceptual title."
* Optionally: Experiment with coarse-to-fine models (extract important sentences first, then generate title).

**🧩 Block 7: Evaluation**

**Goals**:

* Measure how good your summaries and titles are.

**Tasks**:

* Use **ROUGE** scores for summaries.
* Use **Semantic similarity** (BERTScore) to compare title vs. full text.
* Human evaluation: Are the titles accurate, attractive, and informative?

**🧩 Block 8: Deployment or Demo System (Optional but recommended)**

**Goals**:

* Build a simple app or system demo.

**Tasks**:

* Create a simple UI:
  + Upload a long text.
  + See semantic chunks, summaries, extracted topics, and generated title.
* Use libraries like **Streamlit**, **Gradio**, or a simple web app.

**📈 Visual Structure (Simple Flow)**

Learning ➡ Dataset ➡ Chunking ➡ Summarizing ➡ Merging + Analyzing ➡ Title Generation ➡ Evaluation ➡ (Optional: Deployment)

**⚡ Bonus Tips for Success**

* Focus your early energy on **Blocks 1-3**. Without clean data and chunking, the rest will struggle.
* Modularize everything: write code for each block separately (chunker, summarizer, title generator).
* Keep evaluation simple but meaningful: 2–3 automatic metrics + small manual review.
* Always test your models **step-by-step** before combining everything.

Would you also like me to make a **visual roadmap chart or Gantt chart** 📅 for you to track it even easier? (It would look really cool and super helpful for your project plan!) 🚀  
Let me know!