**Project Proposal: AI-Powered Tech News Summarizer**

**Statement of Problem**

In today's digital era, information overload has become a challenge for many users. People struggle to keep up with the latest news across multiple sources, making it difficult to stay informed without spending excessive time reading. Additionally, news articles often contain bias or unnecessary details, making it harder for users to extract the essential information they need.

This project will focus on scraping news articles from reputable sources or sources of choice and utilizing AI to generate concise, unbiased summaries. The collected data will include article headlines, publication dates, source URLs, and full article content. The AI model will process these articles to generate key insights and summaries while ensuring fairness and accuracy.

**Proposed Solution**

This project aims to develop an AI-powered news summarizer that scrapes articles from multiple sources and generates summaries tailored to users' interests.

**Data Processing and Transformation**

* **Source of Data**: Tech News websites: Tech crunch, Engadget, Wired, and APIs
* **Data Transformations**:
  + Scraping: Collecting news articles from selected sources
  + Cleaning: Removing advertisements, unnecessary HTML tags, and duplicate content
  + Summarization: Using AI models to generate concise summaries
* **Storage**: The processed data will be stored in a structured database for efficient retrieval

**Tools and Techniques**

* **Programming Languages**: Python (BeautifulSoup, pandas, spaCy, transformers)
* **Database**: PostgreSQL or MongoDB
* **AI Model**: Pre-trained NLP models such as BART, T5, or GPT-based models for summarization
* **Deployment**: Web-based dashboard for users to access personalized summaries

**Data Output**

* A webpage displaying summarized news articles categorized by user interest
* (Optional) An API that allows integration with other applications for customized news summaries

**Value Proposition**

**Efficiency**

This project will significantly reduce the time users spend reading lengthy articles by presenting key takeaways concisely.

**Insights**

* Users will receive a curated list of summarized articles based on their interests
* Reduction in exposure to misleading headlines or biased information
* Enhanced accessibility to news for people with limited time

**Strategic Value**

* Promotes media literacy by offering summaries
* Supports efficient decision-making by providing essential news highlights

**Timeline and Resources**

**Timeline**

* **Week 1-2**: Data source identification and scraping setup
* **Week 3-4**: Data processing and AI model selection
* **Week 5-6**: Summarization model training and testing
* **Week 7**: Web application and API development
* **Week 8**: Final testing and deployment

**Resources**

* **Software & Hardware**:
  + Python environment with necessary libraries
  + Database management tools (PostgreSQL, MongoDB)
  + Cloud hosting for deployment (AWS, Heroku)
* **Data Access**: News websites, APIs