1. **Problem Statement:** The project aims to develop a question-answering system for artificial intelligence applications using a dataset of articles and research papers. The system will provide accurate and informative answers to user queries based on the information extracted from the dataset.

2. Data Preprocessing: I’m working with multiple articles collected from various URLs. I preprocessed the data by downloading, parsing, and tokenizing the text. Additionally, I

I applied TF-IDF vectorization to compute similarity scores between the user's query and the chunks of text from the articles.

I implemented the question-answering model using the DistilBERT model fine-tuned on the SQuAD dataset. Along with this model I incorporated a keyword matching model as a fallback option when the BERT model doesn’t produce results. I used the libraries: nltk, numpy, newspaper, sci-kit learn and transformers to implement the model. My validation method involves testing the model with user queries and evaluating the responses.