**Phase 1: Problem Definition and Design Thinking**

The scope of this document is to identify the problem and evolving nature of fake news, the need for large and diverse datasets, and the limitations of NLP models. It's important to continually update and fine-tune your model to stay effective in identifying new forms of misinformation.

**Problem Definition**

The problem is to develop a fake news detection model using a Kaggle dataset. The goal is to distinguish between genuine and fake news articles based on their titles and text. This project involves using natural language processing (NLP) techniques to preprocess the text data, building a machine learning model for classification, and evaluating the model's performance.

**Design Thinking:**

⚫ Choose the fake news dataset available on Kaggle, containing articles titles and text, along with their labels (genuine or fake).

⚫Clean and preprocess the textual data to prepare it for analysis.

⚫Utilize techniques like TF-IDF (Term Frequency-Inverse Document Frequency) or word embeddings to convert text into numerical features.

⚫Select a suitable classification algorithm (e.g., Logistic Regression, Random Forest, or Neural Networks) for the fake news detection task.

⚫Train the selected model using the preprocessed data.

⚫Evaluate the model's performance using metrics like accuracy, precision, recall, F1-score, and ROC-AUC.

Note:file naming convention:AI-phase-1