

Phase 1: Problem Definition and Design Thinking

PROJECT-7 FAKE NEWS DETECTION USING NLP

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:

Functionality:

Fake news detection using NLP involves analyzing the content of a news piece against reliable sources or databases, sentimental analysis, semantic analysis, fact-checking, and recognizing sensationalism and clickbait.

Data Collection:

To train our model, we are going to use a large dataset of news articles, including both real and fake examples, from the Kaggle Dataset to train the AI model.

Data Preprocessing:

Since the data may involve tasks like tokenization, stop word removal, and text normalization, we are going to clean and preprocess the data.

Feature Extraction:

We are going to extract relevant features from the text data, such as word embeddings or TF-IDF vectors, to represent the articles.

Natural Language Processing (NLP):

To train our chatbot understand/recognize various user inputs it is necessary to implement NLP&NLU techniques. As the chatbot is created using python, python libraries like NLTK (natural language tool kit) and RASA-NLU can be used to analyse the user inputs. It will help the chatbot to process the output in a conversational manner.

Training:

We are going to train the selected model on the labeled dataset, adjusting hyperparameters and using techniques like cross-validation to ensure good performance.

Evaluation:

We are going to assess the model's performance using metrics like accuracy, precision, recall, and F1 score on a separate test dataset.

Fine-Tuning:

Since it is a rejection type mode, we have to refine the model and its features to improve accuracy and reduce false positives/negatives.

Deployment:

We are going to integrate the trained model into an application and develop a user-friendly interface for users to input news articles and receive detection results.

Continuous Updation:

We are going to continuously update the model with new data to adapt to evolving fake news tactics.

Finally it's our responsibility to announce that eventhough AI can assist in fake news detection, it's not foolproof, and We, human fact-checkers should remain part of the process to verify results.