AI Based News Authentication System



Session: 2022 – 2026

**Submitted by:**

Aleena Sheikh 2022-CS-130

Ayesha Khalid 2022-CS-153

**Submitted to:**

Mr. Samyan Qayyum Wahla

Department of Computer Science

**University of Engineering and Technology Lahore Pakistan**

Contents

[**Software Requirement Specification (SRS) for AI-Based News Authentication System** 3](#_Toc185193545)

[**1. Introduction** 3](#_Toc185193546)

[1.1 Purpose 3](#_Toc185193547)

[1.2 Scope 3](#_Toc185193548)

[1.3 Definitions, Acronyms, and Abbreviations 3](#_Toc185193549)

[1.4 References 4](#_Toc185193550)

[1.5 Conclusions (Change to Overview) 4](#_Toc185193551)

[**2. General Description** 4](#_Toc185193552)

[2.1 Process Concept 4](#_Toc185193553)

[2.1.1 System Interfaces 4](#_Toc185193554)

[2.1.2 Connections 4](#_Toc185193555)

[2.1.3 Hardware Interface 4](#_Toc185193556)

[2.1.4 Software Interfaces 5](#_Toc185193557)

[2.1.5 Communication Interfaces 5](#_Toc185193558)

[2.1.6 Memory Constraints 5](#_Toc185193559)

[2.1.7 Operations 5](#_Toc185193560)

[2.1.8 Site Adaptation Requirements 5](#_Toc185193561)

[2.2 Product Functions 5](#_Toc185193562)

[2.3 User Characteristics 5](#_Toc185193563)

[2.4 Constraints 6](#_Toc185193564)

[2.5 Assumptions and Dependencies 6](#_Toc185193565)

[**3. System Requirements** 6](#_Toc185193566)

[3.1 Functional Requirements 6](#_Toc185193567)

[3.1.1 API Data Retrieval 6](#_Toc185193568)

[Use Case: API Data Retrieval 6](#_Toc185193569)

[3.1.2 Data Processing and Cleaning 6](#_Toc185193570)

[Use Case: Data Processing and Cleaning 7](#_Toc185193571)

[3.1.3 Vector Database Management 7](#_Toc185193572)

[Use Case: Vector Database Management 7](#_Toc185193573)

[3.1.4 News Authentication and Comparison 7](#_Toc185193574)

[Use Case: News Authentication and Comparison 7](#_Toc185193575)

[**4.** **Wireframe** 8](#_Toc185193576)

[**5.** **Glossary** 8](#_Toc185193577)

[**6.** **Acronyms** 8](#_Toc185193578)

**Software Requirement Specification (SRS) for AI-Based News Authentication System**

# **1. Introduction**

This Software Requirement Specification (SRS) document details the intended and non-functional requirements for the AI-Based News Authentication System. It seeks to ensure clarity throughout the process, including design, implementation, and testing of the project.

## 1.1 Purpose

## This section will cover the requirements for the AI-Based News Authentication System. By using semantic evaluation and gadget mastering to evaluate information articles, the gadget can assist reduce the unfold of misinformation.

## 1.2 Scope

## This AI-Based News Validation System will:

## You can fetch news data from various APIs.

## Store the data in a vector database for efficient processing

## Use semantic analysis to detect article authenticity

## This flow helps the media organizations journalists and general public to know the authenticity of the news.

## 1.3 Definitions, Acronyms, and Abbreviations

* **SRS:** Software Requirements Specification
* **AI:** Artificial Intelligence
* **API:** Application Program Interface
* **Vector Database:** Database specialized for easy (and fast) retrieval of large amounts of scale vector data.
* **Semantic Analysis:** The capability to perform natural language processing tasks such as text classification and named entity recognition.
* **Authenticity Score:** A metric that calculates the authenticity of the news

## 1.4 References

* <https://www.researchgate.net/publication/382296709_application_about_conceptual_systems_used_in_information_acquisition>

## 1.5 Conclusions (Change to Overview)

The structure of this sentence is:

* Section 1 gives the overview of the system and tells the purpose.
* Section 2 tells the description of the system , target users and the constraints.
* Section 3 explains the system’s functional and non-functional requirements.

# **2. General Description**

## 2.1 Process Concept

This is the system which is designed to remove the misinformation. And with the existing information it will analyze whether the news is correct or not. The advantages of our system are:

* **We are retrieving real time data from different APIs.**
* **Efficient binding of data with the help of vector database.**
* **For semantic analysis , advance techniques are used.**

### 2.1.1 System Interfaces

* **API Interface:** For media data it facilitates the integration.
* **User Interface (UI):** Allows the user to input their news and see the results.
* **Admin Interface:** It allows to update and view the parameters of the system.

### 2.1.2 Connections

* **User Interface (UI):** Users can input their news and see the results.
* **API Interface:** Accepts the data from media APIs.
* **Database Interface:** From vector database it retrieves and stores the data.

### 2.1.3 Hardware Interface

* **Standard Server:** The Standard Server acts as a host for the content and processes AI procedures.
* **End-user Devices (PC, Tablet, Smartphone):** PC, Tablet and Smartphone serves as interfaces through which users can interact with and utilize the system.

### 2.1.4 Software Interfaces

* **Operating System:** Windows, Linux, or macOS.
* **Database:** A vector database (e.g., Pinecone, Milvus).
* **Programming Languages:** Python, JavaScript.
* **AI Libraries:** TensorFlow, PyTorch, SpaCy.

### 2.1.5 Communication Interfaces

* **Internet Connection:** Gives the access to the News API.
* **Secure Communication Protocol (HTTPS):** Ensures the secure transmission of data.

### 2.1.6 Memory Constraints

This system handles a large amount of data in efficient way. It is best to work with a server that has at least 16 GB of RAM storage in order to handle the data and perform the AI calculations.

### 2.1.7 Operations

* Data retrieving via APIs is continuous.
* Analyzing the content of the media as the media is being processed.
* Persistence and query of data contained for vector databases.
* Authentication result is presented back to the users

### 2.1.8 Site Adaptation Requirements

This system will be web-based and there is no need for the specific physical solution. It would be accessible.

## 2.2 Product Functions

The system is supposed to perform the following task:

* **Retrieve Media Data:** Retrieves the data from different APIs.
* **Store Data:** Saves the data in the vector database.
* **Analyze Media:** Utilize AI-based semantic analysis to verify authenticity.
* **Display Results:** Present an authenticity score or status to the user.
* **User Feedback:** Allow users to provide feedback on the authenticity result.

## 2.3 User Characteristics

* **General Users:** Individuals seeking to verify the authenticity of news articles (no technical expertise required).
* **Media Professionals:** Journalists and editors who use the system for quick news verification (moderate technical knowledge).
* **Administrators:** Personnel responsible for system maintenance, API connection monitoring, and AI model updates (high technical expertise).

## 2.4 Constraints

* **Regulatory Compliance:** The system must adhere to data privacy laws, particularly concerning sensitive user data and privacy.
* **API Rate Limits:** Some APIs may limit the number of requests per minute or hour, impacting data collection.
* **Processing Time:** Semantic analysis should not take more than a few seconds per article.

## 2.5 Assumptions and Dependencies

* The APIs used to fetch news data are reliable and provide accurate information.
* Users have access to a stable internet connection.
* The vector database and AI models are scalable to handle increased user traffic.

Here is an expanded section with additional functional requirements and use cases based on your project's goals.

# **3. System Requirements**

## 3.1 Functional Requirements

### 3.1.1 API Data Retrieval

* **Objective**: Retrieve news data from some free news channels' APIs.
* **Scope**: System should connect with API sources to get news data after some delay.
* **Process**:
  + The system connects to APIs of news channels.
  + It analyzes the incoming data and stores it in a structured format.
* **Output**: The retrieved data is stored in a raw data form for further processing.

### Use Case: API Data Retrieval

* **Title**: Retrieve news data from APIs
* **Actors**: System
* **Main Flow**:
  1. System initiates a scheduled call to each configured news API.
  2. Retrieves the data from each source.
  3. Converts the data into the defined format for storage.
  4. Saves the data in a raw storage location.

### 3.1.2 Data Processing and Cleaning

* **Objective**: Clean and process the retrieved news data.
* **Scope**: The system should be able to normalize and delete completely or partially any fields that are not relevant.
* **Process**:
  + The objective of data preprocessing is to eliminate duplication, irrelevant information, and different data formats.
* **Output**: Cleaned data is stored in the database for vectorization.

### Use Case: Data Processing and Cleaning

* **Title**: Process and clean retrieved data
* **Actors**: System
* **Main Flow**:
  1. System loads the raw news data from storage.
  2. Identifies and removes duplicate records.
  3. Removes any irrelevant data fields.
  4. Normalize the data format.
  5. Saves the processed data in the database.

### 3.1.3 Vector Database Management

* **Objective**: Store processed data in a vector format to enable efficient searching and comparison.
* **Scope**: System should convert news data into vectors and manage these within a vector database.
* **Process**:
  + Data is converted into vector format for efficient comparison.
  + Vector database handles indexing and storage of the vectored data.
* **Output**: News data is stored as vectors, ready for querying and analysis.

### Use Case: Vector Database Management

* **Title**: Store processed data in vector format
* **Actors**: System
* **Main Flow**:
  1. System reads cleaned data from the main database.
  2. Converts each news entry into a vector.
  3. Stores the vectored data in the vector database with relevant indexing.

### 3.1.4 News Authentication and Comparison

* **Objective**: Analyze and authenticate news articles based on similarity checks and advanced search.
* **Scope**: System should use vector comparisons and search queries to detect potential misinformation.
* **Process**:
  + System runs similarity checks between new and existing news entries.
  + Advanced queries help identify patterns indicating misinformation.
* **Output**: Authentication results flag articles as reliable or potentially fake.

### Use Case: News Authentication and Comparison

* **Title**: Authenticate news articles
* **Actors**: System
* **Main Flow**:
  1. System receives a news article to authenticate.
  2. Runs a similarity check against trusted sources in the vector database.
  3. Executes advanced queries to detect mismatches or suspicious patterns.
  4. Flags the news as authentic or potentially fake.

# **Wireframe**

# **Glossary**

 **API (Application Programming Interface)**: A set of protocols and tools for building and interacting with software applications, enabling data sharing between systems.

 **Vector Database**: A database optimized for storing data in vector form, allowing efficient similarity searches, often used for tasks like recommendation and fraud detection.

 **Similarity Search**: A method of finding similar items within a database, commonly based on vector distance calculations.

 **Data Preprocessing**: Cleaning and transforming raw data to prepare it for analysis.

 **False News/Misinformation**: Content that is intentionally or unintentionally misleading or factually incorrect.

 **Authentication**: Verification of the authenticity or accuracy of a piece of information.

# **Acronyms**

 **API**: Application Programming Interface

 **SRS**: Software Requirements Specification

 **DB**: Database