**Functional Document**

1. Introduction
2. Product Goal
3. Demography (Users, Location)
4. Business Processes
5. Features
   1. Feature #1
      1. Description
      2. User Story
6. Authorization Matrix
7. Assumptions

**Fake Profile Identification using Machine Learning**

1. **Introduction**

The rapid rise of social media has created a massive platform for communication, connection, and content sharing. However, with this growth has come the problem of fake profiles, which are used for various malicious activities such as phishing, spreading misinformation, and identity theft. Fake profiles erode the trustworthiness of social media platforms, leading to security concerns and deteriorating user experiences.

This document outlines the functional requirements for a machine learning and NLP-based system to identify and manage fake profiles in social media environments. The system aims to provide an automated, scalable solution capable of accurately detecting fake profiles through the analysis of user behavior, content, and interactions.

2. **Product Goal**

The primary goal of this product is to develop a robust system that can accurately detect and classify fake profiles on social media platforms using machine learning (ML) and natural language processing (NLP) techniques. The system should:

* Enhance security by reducing the number of fake profiles.
* Protect users from scams, fraud, and misinformation.
* Increase trust and user engagement on social platforms by ensuring authenticity.

3. **Demography (Users, Location)**

* **Users**:
  + **Social Media Users**: Individuals who use social media platforms and may interact with or be targeted by fake profiles.
  + **Social Media Platforms/Administrators**: Social media companies and their administrative teams that monitor platform activity and ensure policy compliance.
  + **Researchers/Developers**: Individuals or organizations working on enhancing machine learning models to identify fake profiles.
* **Location**:
  + Global reach, as social media platforms operate worldwide. The solution must be adaptable to various geographic regions and capable of handling user-generated content in multiple languages, depending on the platform’s reach.

4. **Business Processes**

 **Data Collection**:

* Collect user profile information, content shared, interaction patterns, and user activity data from the social media platform.

 **Feature Extraction**:

* Extract features from the collected data, such as profile metadata (e.g., age, profile picture), behavioral patterns (e.g., posting frequency), and content attributes (e.g., linguistic cues from posts using NLP).

 **Model Training**:

* Train machine learning models using labeled datasets (both fake and genuine profiles). Supervised learning and NLP techniques will be used to identify patterns indicating fake profiles.

 **Profile Classification**:

* Run the trained model on real-time data from user profiles. The system will classify each profile as either fake or genuine based on its features.

 **Alert Generation & Action**:

* If a profile is identified as fake, the system will flag it for review by the platform administrator. Based on the platform’s policy, appropriate action (e.g., suspension, ban, or warning) will be taken.

 **Model Evaluation and Updates**:

* Continuously evaluate the model’s accuracy and update it with new data to improve performance.

5. **Features**

**Description**

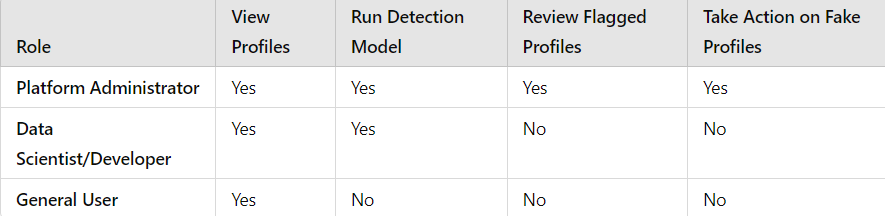
This feature focuses on analyzing user behavior on the platform to identify fake profiles. It examines:

* Frequency and timing of posts (e.g., large volumes in short periods).
* Interactions with other users (e.g., spamming messages or friend requests).
* Network connections (e.g., lack of mutual friends, connections to other flagged accounts).

**2. User Story**

As a **platform user**, I want the platform to automatically block fake profiles so that I can have a safer and more trustworthy experience when interacting with others.

6. **Authorization Matrix**



7. **Assumptions**

 The social media platform will provide access to the necessary user data and activity logs for analysis.

 The dataset used for model training contains a balanced distribution of fake and genuine profiles, and the ground truth labels are accurate.

 The system will work with both English and non-English languages, assuming language processing models are available for supported languages.

 The detection model will be updated periodically with new data to account for evolving tactics used by fake profiles.

 The system will operate in compliance with data privacy and user consent regulations, such as GDPR or CCPA.

 The system will be able to handle high-volume traffic and data in real-time without significant latency issues.

 Users and administrators will have access to detailed reports of flagged profiles and model performance.