Background

Organizations increasingly use AI-powered language models, such as GPT, to assist with information retrieval and process automation. However, unrestricted AI responses can expose sensitive information, creating security risks. A need exists for a system that allows users to query an AI model, such as GPT, but restricts the answers based on user-specific access levels. Biometric authentication, especially facial recognition, offers a reliable method for verifying a user's identity and access level, ensuring that responses from the AI are limited to information within their authorization.

Summary of Invention

The proposed system integrates a language model API with facial recognition technology to:

1. Authenticate users and determine their access level.
2. Retrieve and deliver AI responses based on restricted content permissions.
3. Store and track access attempts and responses for auditing.

The system begins with facial detection and recognition to confirm the user's identity. Once verified, the user's access level is checked against a database of content permissions, determining which datasets or information the AI model can retrieve. When the user submits a query, the AI model filters its response to only include data accessible to that specific user, ensuring data privacy and regulatory compliance.

Detailed Description

1. System Components

- **Biometric Authentication Module**: A facial recognition module that verifies a user's identity. This module interfaces with a database containing authorized user profiles and access levels.
- **Natural Language Processing Module**: An API interface with a language model (e.g., GPT) for handling user queries and generating responses.
- **Access Control Module**: A system that restricts responses based on the user's access level, filtering out sensitive information that is beyond their clearance.
- **Audit and Logging Module**: Tracks user access attempts, queries, and AI responses for compliance auditing.

2. Method for Controlled Information Retrieval

- **Step 1: User Authentication via Facial Recognition**

    - When a user initiates a session, the system activates facial detection and recognition.
    - The user's facial features are captured, encoded, and compared with a database of stored facial profiles.
    - If a match is found, the system retrieves the user's access level.
- **Step 2: Query Submission and Access Verification**
    - After successful authentication, the user can submit a query to the language model.
    - The Access Control Module verifies the user's access level, filtering content permissions associated with their profile.
- **Step 3: AI Response Generation with Restricted Output**
    - The language model generates a response to the user's query.
    - The response is dynamically filtered by the Access Control Module to exclude any information not authorized for the user's level.
    - Only authorized information is returned to the user, ensuring compliance with data privacy and security policies.
- **Step 4: Logging and Monitoring**
    - The system logs each access attempt, query, and AI response for auditing purposes.
    - This data is stored securely for periodic review, ensuring that access remains compliant with security standards.

3. Technical Specifications

- **Facial Recognition Accuracy**: The system employs a facial recognition module with high accuracy to prevent unauthorized access. It may use multi-factor authentication (e.g., biometric data and secondary PIN verification) for sensitive information access.
- **Natural Language Processing API**: The language model is trained to handle queries in a variety of contexts and can respond with detailed, context-aware answers.
- **Dynamic Filtering Mechanism**: The Access Control Module includes a filter layer that parses AI responses to remove unauthorized data, ensuring output is aligned with the user's access level.

4. Use Case Scenarios

- **Medical Facilities**: Authorized healthcare providers can access patient records via ChatGPT, but only for patients they are authorized to view. Facial recognition ensures that only verified providers receive relevant information.
- **Corporate Knowledge Management**: Employees can use ChatGPT to query internal documents, but responses are restricted based on their role. Facial authentication ensures that only authorized personnel can retrieve specific documents.
- **Legal Firms**: Lawyers can query case records or legal documents, but only within the bounds of their authorized cases, verified through facial recognition.