Product Requirement Document (PRD) for AI Presentation Coach

# 1. Problem Statement

Presentations are an essential skill in academic, professional, and personal spheres, yet many people struggle with creating impactful presentations that communicate effectively. The challenge includes poor slide design, unclear messaging, ineffective voice modulation, and non-verbal cues. Traditional presentation training often requires in-person coaching, which can be time-consuming, expensive, and inconsistent.  
  
An AI-driven coach can provide real-time, actionable feedback on multiple aspects of a presentation—content, delivery, and visual appeal—helping individuals enhance their presentation skills more effectively and affordably.

# 2. Solution Overview

The AI Presentation Coach aims to provide users with tailored, comprehensive feedback to improve their presentation skills. The system will analyze slide content, vocal tone, and body language through a combination of Natural Language Processing (NLP), speech analysis, and video analysis. This feedback will help users refine their presentation structure, delivery style, and overall impact.

# 3. Key Features

## 3.1 Content Feedback

- Slide Structure and Readability: The AI analyzes slide content to ensure that text is concise, coherent, and organized. It flags overloading of slides with too much information.  
- Visual Design Feedback: The system provides suggestions on visual elements (color contrast, font size, spacing) based on best design practices.  
- Messaging Alignment: Checks if the presentation's key messages are clear and align with the intended purpose or audience expectations.  
- Data Insights: Identifies whether charts, graphs, and data are presented in a clear and impactful manner.

## 3.2 Voice and Tone Feedback

- Voice Clarity: The AI assesses the speaker's clarity, pace, and diction, offering feedback on improving vocal delivery.  
- Tone Analysis: Analyzes whether the speaker's tone is appropriate for different sections (informative, persuasive, engaging, etc.) and offers suggestions for tone modulation.  
- Pacing Recommendations: Tracks the speed of speech and suggests areas to slow down or speed up for maximum impact.

## 3.3 Non-Verbal Feedback (Video Analysis)

- Body Language: Analyzes facial expressions, posture, and gestures to evaluate whether the speaker is engaging, confident, and approachable.  
- Eye Contact: Tracks eye contact frequency and provides suggestions to improve engagement.  
- Filler Words Detection: Identifies and counts filler words (e.g., “um,” “uh”) and suggests areas for improvement.  
- Hand Gestures: Assesses whether hand gestures are too few, too many, or mismatched with the spoken message.

## 3.4 Presentation Coaching Mode

- Practice Mode: Allows users to practice their presentations and receive real-time feedback on each slide.  
- Summary Feedback: After the practice, the coach provides a summary, highlighting areas of improvement, such as body language, slide flow, and voice modulation.  
- Actionable Suggestions: Specific, tailored suggestions that can be implemented immediately, such as tweaking slide content or practicing certain vocal tones.  
- Comparative Analytics: Tracks user performance over time, helping them identify areas of consistent improvement or persistent challenges.

## 3.5 Integration Capabilities

- Integration with Popular Presentation Tools: Seamless integration with Microsoft PowerPoint, Google Slides, or Prezi to directly analyze slides.  
- Video Upload Option: Users can upload their recorded presentations to receive feedback, making the tool more flexible.  
- Multilingual Support: Feedback in multiple languages to serve a global user base.

# 4. User Personas

1. Professionals: Employees looking to improve their presentation skills for meetings, conferences, or pitches.  
2. Students/Academics: Individuals preparing for academic presentations, including dissertations and class presentations.  
3. Public Speakers: People who regularly speak in public forums, seeking feedback on delivery and engagement.  
4. Entrepreneurs: Startups and business founders practicing their pitches for investors.

# 5. Technological Overview

## 5.1 Frontend

- Framework: React.js or Vue.js for building a responsive, interactive user interface.  
- UI Design: Tailored for ease of use with a simple workflow—upload slides, video, or record directly within the app.  
- Real-Time Feedback Visualization: Use of WebSocket or REST APIs to receive real-time feedback from the backend and display it to users in an intuitive manner.  
- Device Compatibility: Design for cross-platform usability (web, mobile, tablet).

## 5.2 Backend

- Framework: Django (Python) or Node.js (JavaScript) for handling REST APIs and server-side processing.  
- Speech Analysis: Python’s SpeechRecognition or Google Cloud Speech-to-Text API for extracting speech features and performing analysis.  
- NLP & Content Analysis: Use of Hugging Face transformers for NLP tasks to analyze slide content, provide text summaries, and feedback on message clarity.  
- Video Analysis: OpenCV and MediaPipe for tracking body posture, hand gestures, and eye movement in the video.  
- Voice Modulation and Tone Analysis: Use PyDub for voice modulation analysis, or integrate with deep-learning models for advanced speech sentiment analysis.

## 5.3 AI and Machine Learning

- Natural Language Processing (NLP): For understanding slide content and ensuring alignment of messaging.  
- Computer Vision: For analyzing body language, facial expressions, and gestures from the video feed.  
- Speech Recognition and Analysis: For feedback on vocal delivery, tone, and filler words.  
- Feedback Engine: Use machine learning models for continuous improvement, collecting data on user feedback and adjusting recommendations.

## 5.4 Data Storage

- Database: PostgreSQL or MongoDB to store user presentations, feedback history, and user performance analytics.  
- Cloud Storage: Use Amazon S3 or Google Cloud Storage for storing video uploads and presentation materials.

## 5.5 Scalability

- Cloud Deployment: Leverage AWS, Google Cloud, or Microsoft Azure for scalable server infrastructure, supporting real-time processing for a large number of users.  
- Microservices Architecture: To decouple the various services (speech analysis, NLP, computer vision) and allow independent scaling.

# 6. Potential Challenges

- Latency: Real-time feedback may introduce latency, especially during video analysis. Optimizing for speed without compromising accuracy will be critical.  
- Data Privacy: Storing and processing video data must comply with privacy regulations like GDPR.  
- Accuracy of Non-verbal Feedback: Complex human gestures and expressions can be difficult for AI to interpret accurately across different cultural contexts.

# 7. Success Metrics

- User Engagement: Number of users practicing their presentations and using the feedback system regularly.  
- Improvement Rate: Measurable improvement in user performance over time based on feedback analytics.  
- Feedback Quality: High user satisfaction with the actionable feedback provided by the system.  
- System Performance: Low latency, accurate speech and video analysis, and robust handling of large volumes of data.