### **Project Title:** Scalable & Secure Interactive AI-Powered News Podcast Platform

**Abstract**

This project focuses on the development of a highly available, secure, and interactive cloud-native application that redefines news consumption. Unlike traditional static news feeds, this platform empowers users to discover trending articles across various categories and instantly convert them into high-quality, AI-generated audio podcasts on demand. The solution is architected for enterprise-grade scalability, utilizing a **"pull" model** where content is synthesized in real-time using Generative AI. Built on **AWS**, the infrastructure leverages a secure **Virtual Private Cloud (VPC)** network topology, **Auto Scaling**, and **Load Balancing** to ensure fault tolerance, security, and low latency for a seamless user experience.

**Objective**

The primary objective is to build a production-ready, full-stack web application that solves information overload through personalization and automation. Key goals include:

* **On-Demand Personalization:** Enabling users to select specific articles for instant audio conversion using **Amazon Bedrock** for context-aware summarization and **Amazon Polly** for lifelike neural speech synthesis.
* **High Availability & Scalability:** Implementing an **Auto Scaling Group (ASG)** and **Application Load Balancer (ALB)** to dynamically handle traffic spikes and ensure zero downtime.
* **Robust Security:** Enforcing strict network isolation via a custom **VPC** with private subnets, and implementing Role-Based Access Control (RBAC) using **Amazon Cognito** to distinguish between General and Admin users.
* **Cost Efficiency:** utilizing **Amazon DynamoDB** as a caching layer to prevent redundant AI processing for frequently requested articles.

**Architecture Overview**

The system adopts a modern 3-tier web architecture designed for security and performance:

* **Frontend & Backend:** A **React.js** Single Page Application (SPA) serves as the user interface, while **FastAPI (Python)** acts as the backend orchestrator. Both are deployed on **Amazon EC2** instances within an **Auto Scaling Group** to ensure responsiveness.
* **Network & Security:** The infrastructure is isolated within a custom **VPC**. An **Application Load Balancer (ALB)** manages incoming traffic and handles SSL/TLS termination (HTTPS) via **AWS Certificate Manager**. EC2 instances reside in private subnets, accessing external services securely via **NAT Gateways**.
* **AI & Data Services:** **AWS Boto3** is used to integrate **Amazon Bedrock** (LLM-based script generation) and **Amazon Polly** (Text-to-Speech). **Amazon S3** stores the generated audio files, while **Amazon DynamoDB** manages user profiles, article metadata, and audio caching state.
* **Authentication:** **Amazon Cognito** secures the application with industry-standard authentication and user group management.

**Workflow and Key Interactions**

1. **User Access:** A user navigates to the custom domain (managed by **Route 53**). Traffic is routed through the **ALB** to an available EC2 instance. The user logs in securely via **Cognito**.
2. **Content Discovery:** The user selects a news category (e.g., Technology) on the **React** dashboard. The **FastAPI** backend fetches real-time trending headlines from an external **News API**.
3. **On-Demand Processing:** When the user clicks "Play" on an article:
   * The backend queries **DynamoDB** to check if an audio file already exists (Cache Check).
   * *If New:* **Boto3** sends the article text to **Amazon Bedrock** to generate a podcast script, then sends the script to **Amazon Polly** to synthesize speech.
   * *Storage:* The resulting MP3 is uploaded to **S3**, and the metadata is saved to **DynamoDB**.
4. **Playback:** The secure S3 URL is returned to the frontend, allowing the user to stream the audio podcast immediately.
5. **Admin Management:** Admin users can access restricted endpoints to clear the cache or remove specific articles from the database.

**Conclusion**

The Interactive AI-Powered News Podcast Platform demonstrates a comprehensive mastery of modern cloud engineering. By combining the reactive capabilities of a full-stack web application with the immense potential of Serverless AI and the robustness of a secure, load-balanced AWS infrastructure, it delivers a scalable, production-ready platform. This project successfully bridges the gap between raw information and personalized, accessible media, offering a sophisticated solution to the modern problem of digital content consumption.