

API Trade-off Analysis for Meeting Analysis App

This document compares the key APIs used in the Meeting Analysis App in terms of performance, security, privacy, and scalability, incorporating actual test results from the application

1. Assembly AI (Audio Transcription)

Performance Metrics

Latency: Moderate; transcription can take several seconds to minutes depending on audio length.

Accuracy: High accuracy for clear audio; supports multiple languages and accents.

Cost: Usage-based pricing; can become costly with large volumes of audio.

Test Results: Basic unit tests passed successfully, confirming API integration and transcription flow works as expected.

Security and Privacy

- Uses HTTPS for data transmission.
- Data retention policies vary; sensitive audio data is uploaded to their servers.
- Requires secure API key management.

Scaling Challenges and Solutions

- Handles large volumes via cloud infrastructure.
- Rate limits apply; batching and asynchronous processing recommended.
- Scaling is managed by the provider, reducing client-side complexity.

2. Cohere AI (Text Summarization)

Performance Metrics

Latency: Low to moderate; typically seconds per request.

Accuracy: High-quality summarization with contextual understanding.

Cost: Subscription or pay-per-use; costs increase with usage.

Test Results: Summarization function tested indirectly via integration tests; results returned as expected.

Security and Privacy

- Data sent to Cohere's servers; encrypted in transit.
- API keys must be securely stored.
- Privacy depends on provider's data policies.

Scaling Challenges and Solutions

- Cloud-based scaling handled by provider.
- Rate limits require request management.
- Suitable for real-time and batch processing.

3. Hugging Face (Sentiment Analysis)

Performance Metrics

Latency: Low; inference typically under a second for short texts.

Accuracy: Good accuracy with state-of-the-art models.

Cost: Free tier available; paid plans for higher usage.

Test Results: Sentiment analysis function tested via unit tests; classification labels returned correctly.

Security and Privacy

- Data sent to Hugging Face servers; encrypted transmission.
- API keys required and must be protected.
- Open-source models can be self-hosted for enhanced privacy.

Scaling Challenges and Solutions

- Cloud-hosted models scale automatically.
- Self-hosting requires infrastructure and maintenance.
- Rate limits apply on hosted API.

Recommendations

Performance: Assembly AI is the bottleneck due to audio processing time; asynchronous handling recommended.

Security: Use environment variables and secure storage for API keys; consider self-hosting Hugging Face models for sensitive data.

Cost: Monitor usage to control costs; optimize audio length and summarization requests.

Scaling: Leverage provider-managed scaling; implement retry and backoff strategies for rate limits.

This trade-off analysis helps guide decisions on API usage balancing performance, cost, security, and scalability for the Meeting Analysis App.