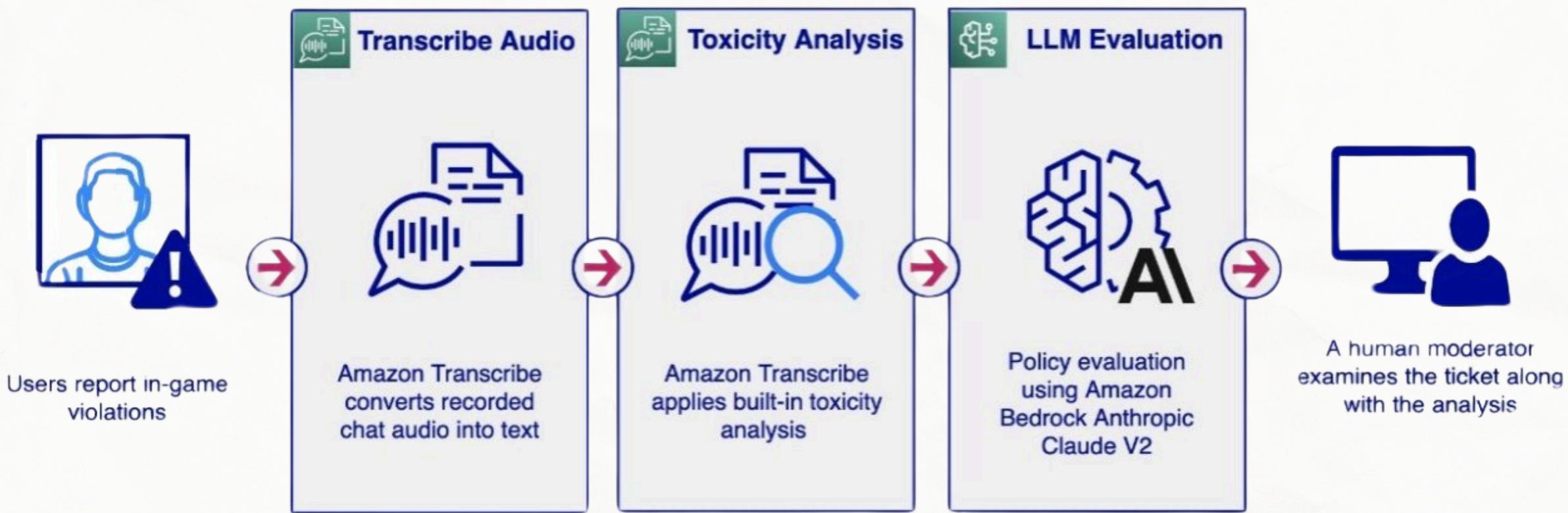


10

Use Cases of **RAG WITH AUDIO SIMILARITY**



1. Content Moderation

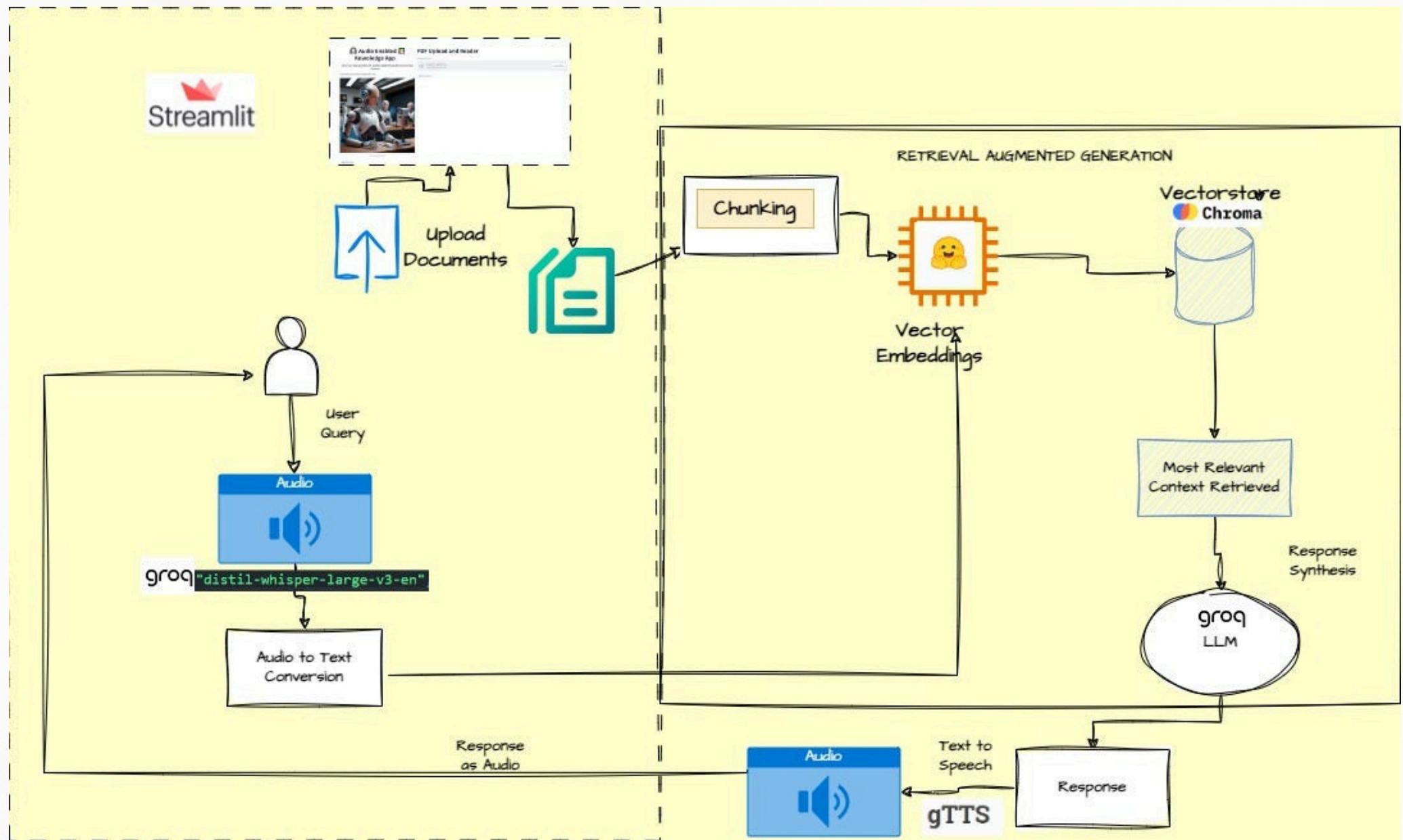


*credit: AWS

Content moderation ensures user-generated content follows platform guidelines and legal rules, fostering a safe online space

- **Detecting Harmful Content:** Identifying hate speech, harassment, or inappropriate content in audio, even if nuanced or disguised.
- **Analyzing Music:** Comparing tracks to detect copyright infringement, protecting creators and ensuring platform compliance.
- **Identifying Misinformation:** Analyzing audio for misinformation or propaganda, combating its spread.

2. Voice Search Enhancement

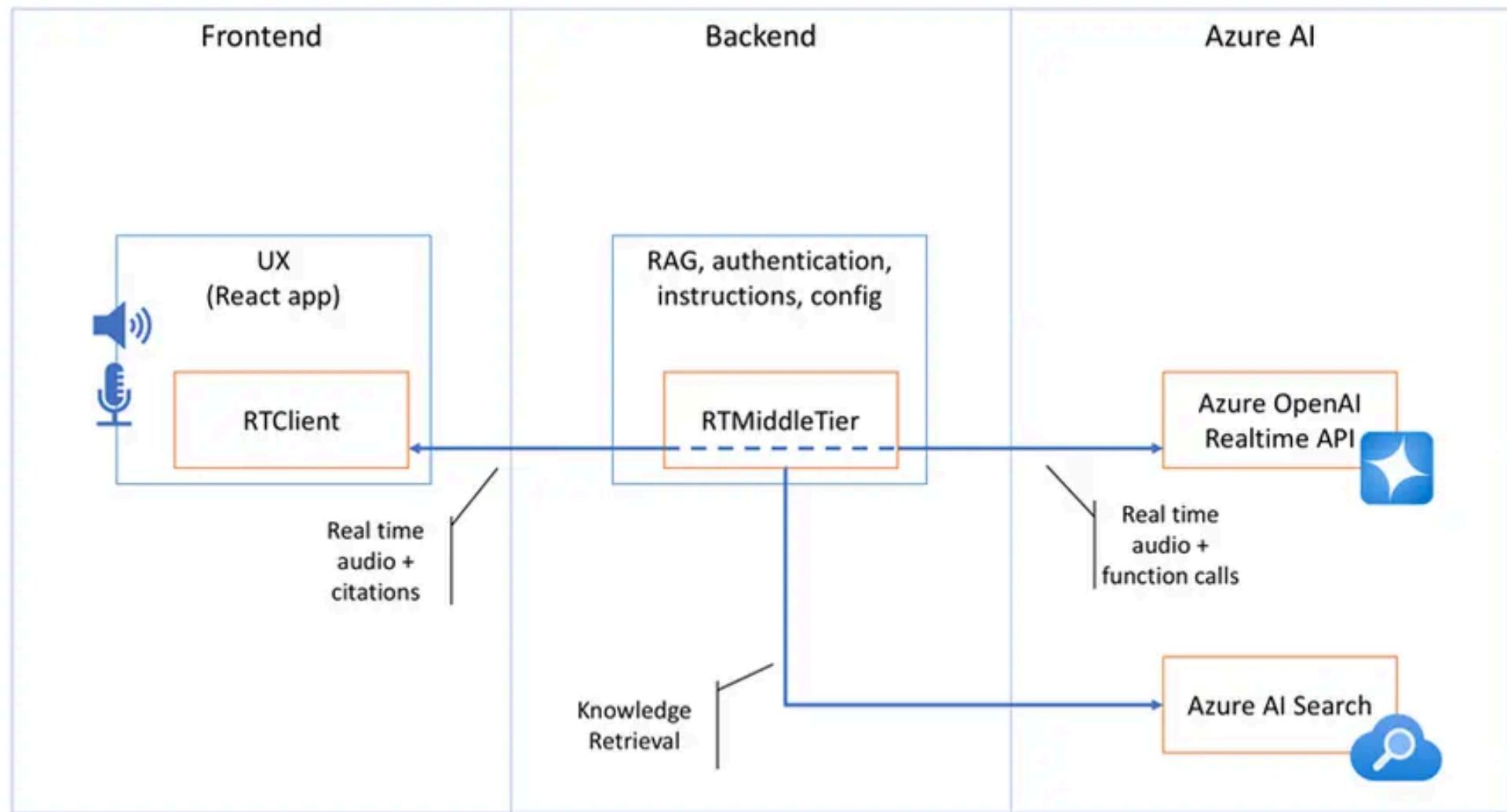


Voice search enhancement improves the accuracy and effectiveness of voice queries, making information retrieval more natural and seamless

- **Understanding Spoken Language:** It can accurately interpret spoken queries, even with accents, background noise, or imperfect pronunciation.
- **Finding Relevant Audio:** It can search for information within audio content, such as podcasts or lectures, to provide precise answers to voice queries.
- **Providing Contextual Responses:** It can understand the intent behind a voice query and provide more relevant and personalized results.

3. Customer Service Analytics

VoiceRAG: app pattern for RAG with real-time audio-enabled models

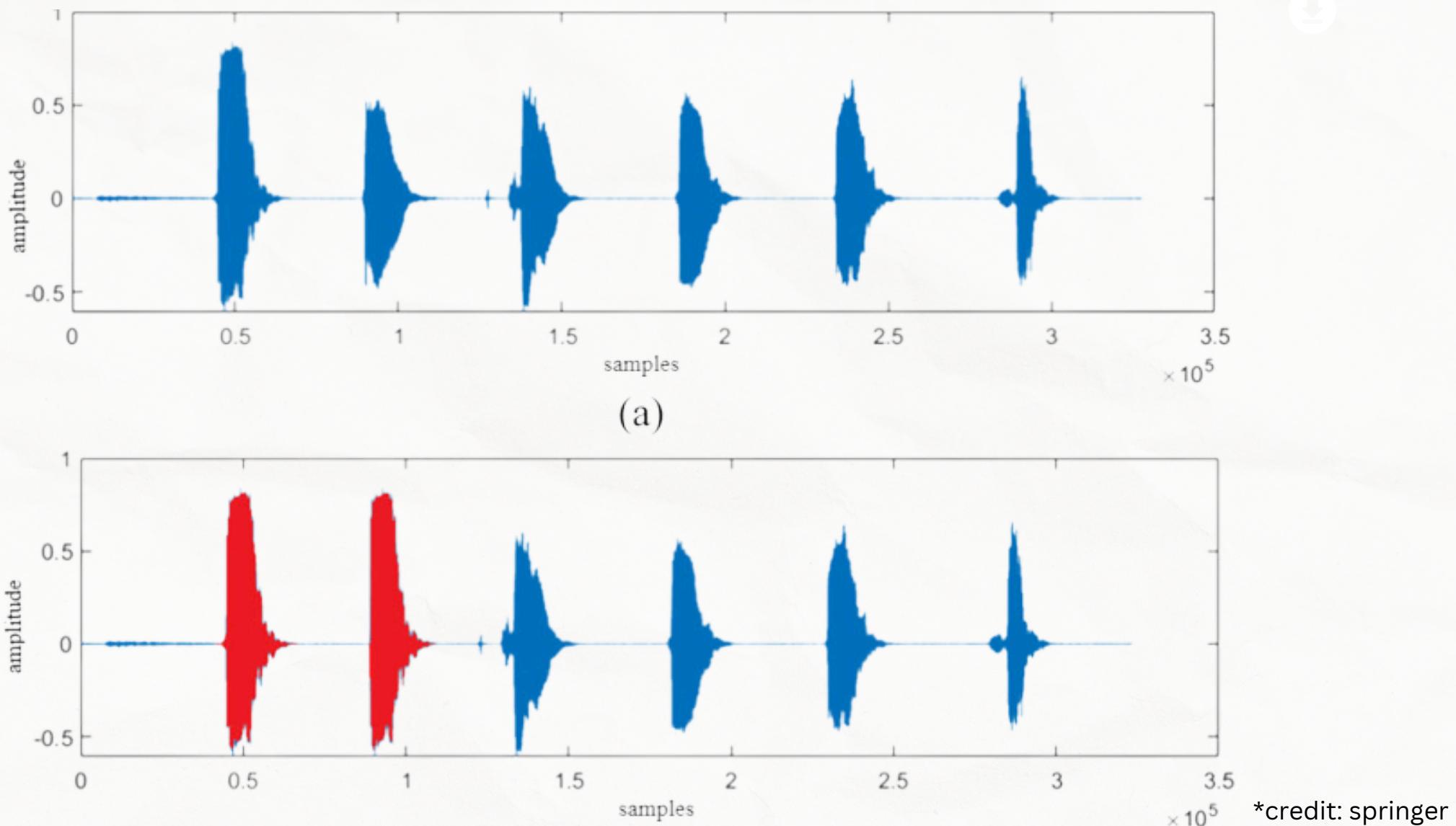


*credit: hyperight

Customer service analytics analyzes data from interactions (calls, chats, emails) to understand behaviour, enhance service, and optimize processes

- **Sentiment Analysis:** Detects customer emotions (frustration, satisfaction, etc.) for better understanding.
- **Issue Identification:** Pinpoints recurring customer problems to improve products/services.
- **Agent Evaluation:** Assesses agent performance to identify strengths and training needs.

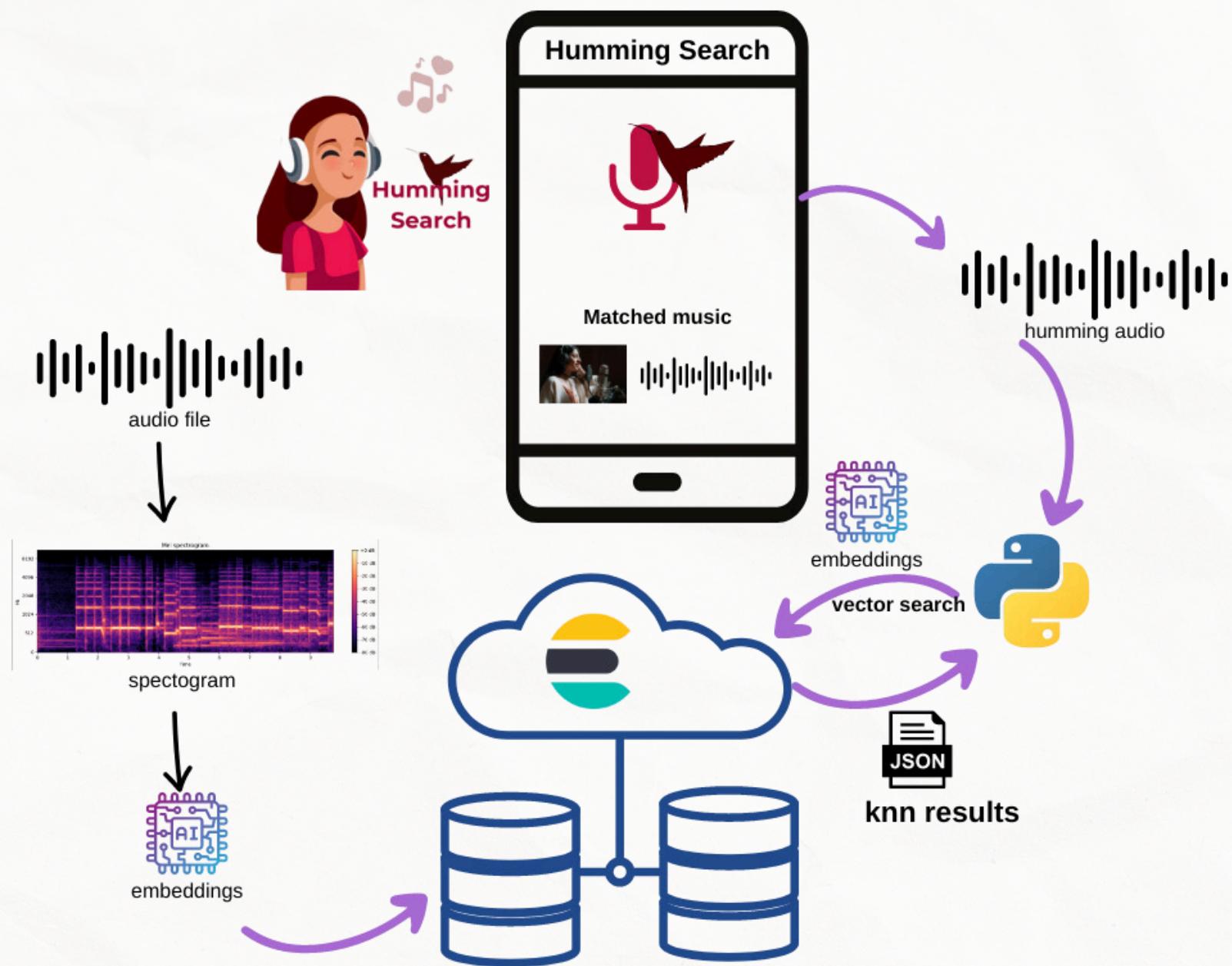
4. Forensic Audio Analysis



Forensic audio analysis involves the scientific examination and interpretation of audio recordings for use as evidence in legal proceedings or investigations.

- **Automating Matching:** Quickly finds similar sounds in large databases, even if degraded or altered.
- **Retrieving Context:** Provides associated information like time and location for matched audio.
- **Improving Accuracy:** Combines audio matching with related text (transcripts, etc.) for better results.
- **Handling Poor Quality:** Works with noisy or degraded audio.

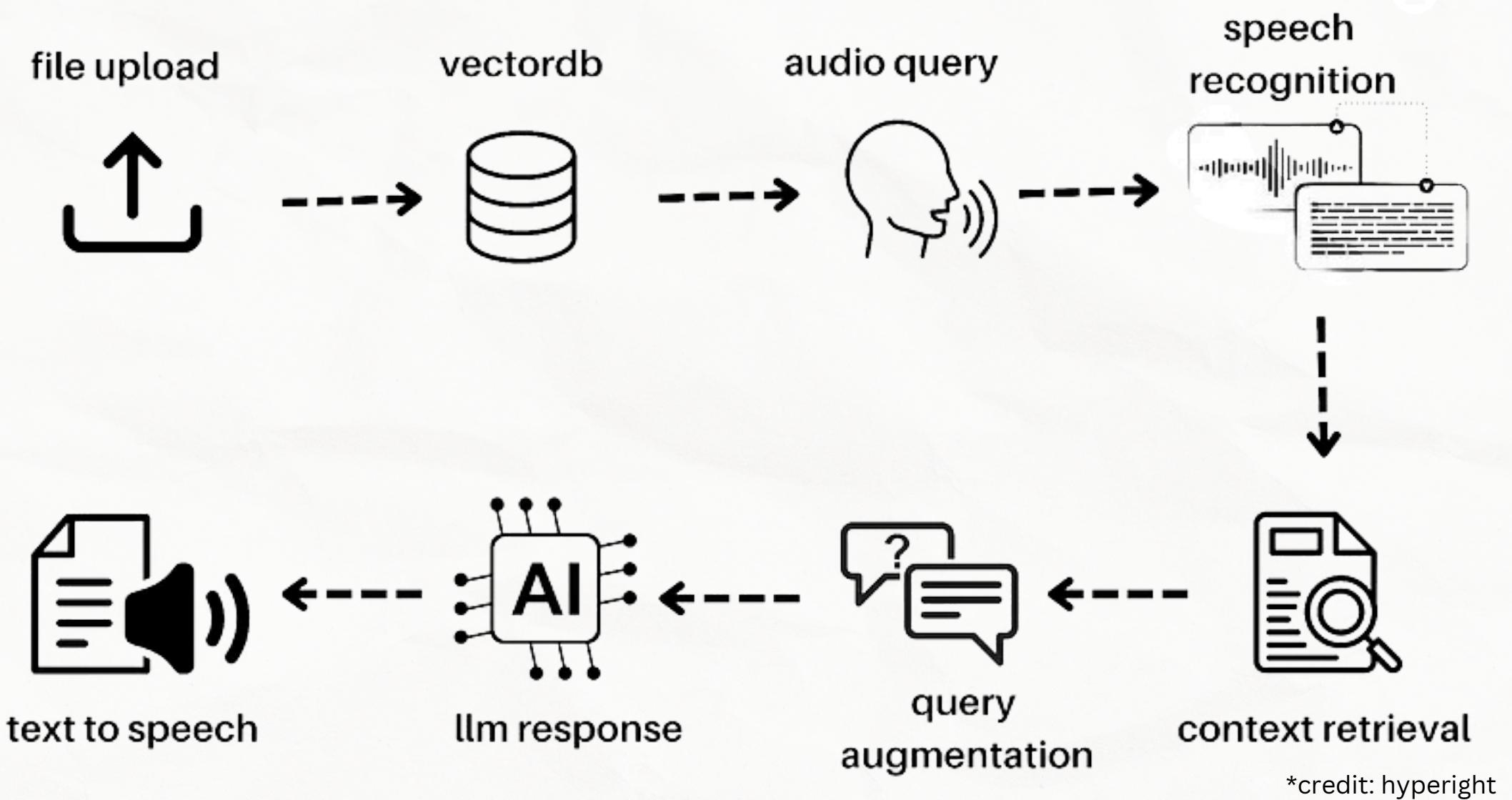
5. Music Recommendation



Music recommendation systems suggest songs based on user preferences using various techniques to enhance personalization.

- **Finding Acoustically Similar Music:** Find songs that sound similar, leading to more personalised recommendations
- **Understanding Nuances in Music:** It analyzes musical elements like melody, harmony, rhythm, to find hidden similarities beyond traditional methods
- **Handling Diverse Audio:** It can analyze audio inputs like humming, describing a melody, or providing a short song clip.

6. Podcast Analysis

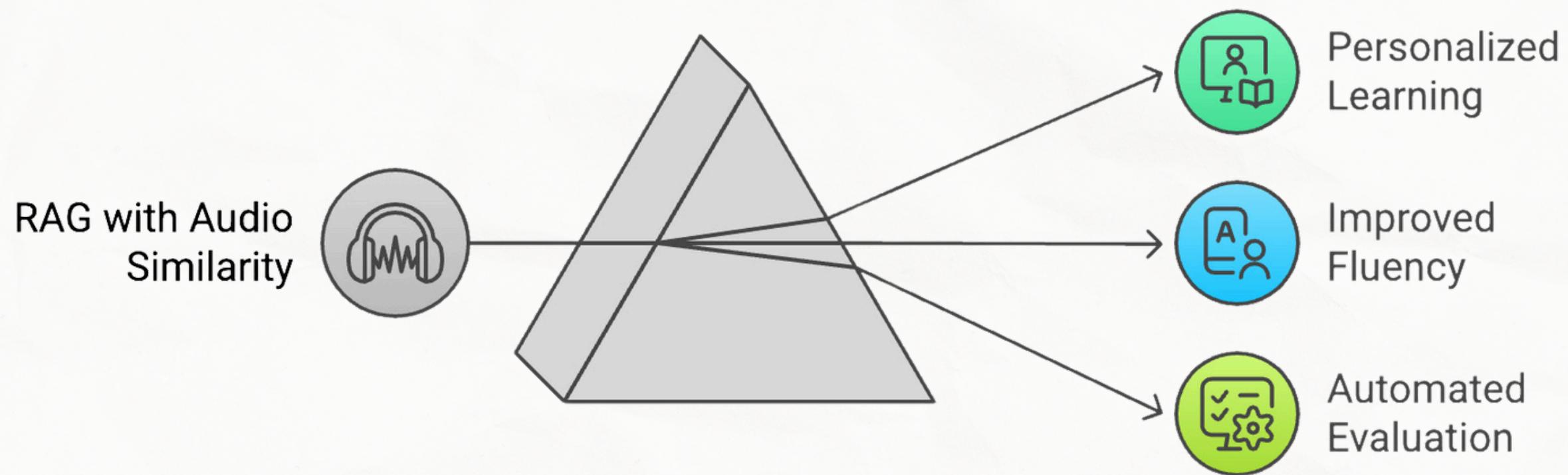


*credit: hyperight

Podcast analysis examines audio and transcripts to understand audience behavior, identify trends, and extract key insights for better content and marketing

- **Precise Retrieval:** Finds specific podcast moments based on sounds, not just text transcripts.
- **Topic ID:** Identifies and groups segments by topic, even with varied wording.
- **Speaker Analysis:** Distinguishes and potentially identifies speakers for conversation analysis.
- **Sentiment Analysis:** Detects the emotional tone of podcast segments.

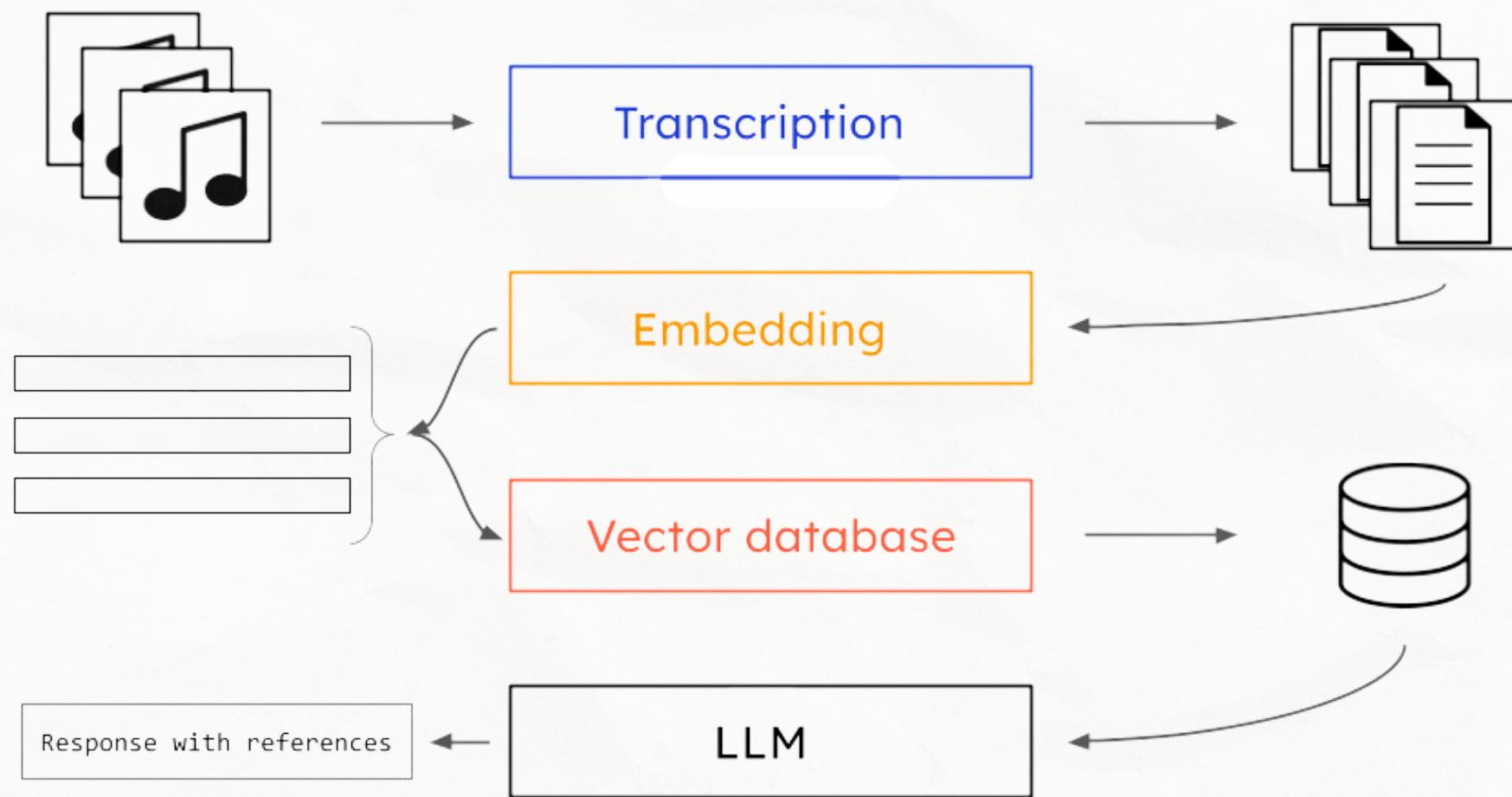
7. Language Learning



Language learning involves mastering pronunciation, which RAG with Audio Similarity helps improve through targeted feedback.

- **Personalized Learning:** RAG customizes content to address learners' specific challenges, enhancing learning outcomes.
- **Improved Fluency:** Focusing on pronunciation, RAG boosts fluency and confidence in speaking.
- **Automated Evaluation:** RAG automates pronunciation evaluation, providing instant feedback without a human instructor.

8. Accessibility for the Visually Impaired



*credit: AssemblyAI

Accessing audio content can be challenging for the visually impaired. RAG with Audio Similarity enhances accessibility, offering richer experiences.

- **Deeper Understanding:** Detailed audio descriptions help visually impaired individuals grasp nuances missed in transcripts.
- **Personalized Descriptions:** RAG can tailor descriptions to individual preferences (e.g., environment vs. speakers).
- **Enhanced Inclusion:** Makes audio content more accessible for full participation in various settings.

9. Automated Audio Tagging

Acoustic Analysis

Analyzes audio to identify sounds.

Keyword Generation

Automatically creates relevant tags for audio.

Organization & Search

Facilitates easy organization and searching of audio.

Organizing large audio collections can be a time-consuming task. RAG with Audio Similarity automates this process by automatically tagging audio files with relevant keywords.

- **Acoustic Analysis:** RAG analyzes audio, identifying sounds like speech, music, or events. It understands acoustic properties, not just sound detection.
- **Keyword Generation:** RAG automatically creates relevant tags based on the audio analysis. A conversation might be tagged "meeting," "discussion," etc.
- **Organization & Search:** Auto-tags make audio libraries easier to organize and search. Users quickly find audio by content, not just filenames.

10. Interactive storytelling



Enhanced Immersion

The story becomes more engaging as it responds to the user's voice.



Personalized Experience

Each user's unique voice input creates a memorable storytelling experience.



Voice Input Analysis

Analyzes spoken input, understanding both content and emotional tone.

Interactive storytelling becomes more immersive when it adapts to user input. RAG with Audio Similarity enables dynamic, voice-responsive narratives.

- **Enhanced Immersion:** Responding to the user's voice makes the story more immersive, making choices and emotions feel impactful.
- **Personalized Experience:** Each user's experience is unique, shaped by their voice input. This creates a more personal and memorable storytelling experience.
- **Voice Input Analysis:** RAG analyzes both the words and emotional tone of the user's spoken input. It understands not just what is said, but how it's said.

Chunking Strategy

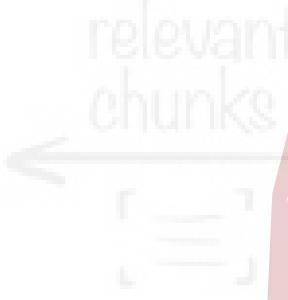
- Chunk Size
- Overlap

Embedding Strategy

E5, BERT



relevant
chunks



MINISTRY OF
HOME AFFAIRS



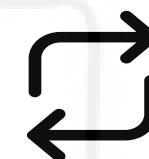
**Follow to stay updated on
Generative AI**



LIKE



COMMENT



REPOST