



# Reto CicerónIA– Sistema de procesamiento de sonido y texto basado en IA para asistente al evaluador y juez en debates universitarios

Pablo Pavón Mariño

3 noviembre 2025

## 1 INTRODUCCIÓN

El presente reto tiene como objetivo el desarrollo de un sistema de inteligencia artificial capaz de emular la labor de un juez en un debate académico, analizando tanto el contenido del discurso como la forma en la que los participantes se expresan. Se trata de una iniciativa conjunta entre la Asociación Universitaria de Debate de la UPCT y la Asociación de Estudiantes de Inteligencia Artificial, quienes unirán sus conocimientos y capacidades para abordar de manera integral un problema con un alto componente técnico y humano. Se ha puesto como nombre tentativo del sistema *CicerónIA*, en honor a Marco Tulio Cicerón (106 a. C. – 43 a. C), famoso político, jurista y orador de tiempos de la República romana.

Por un lado, la Asociación de Debate aportará su experiencia en oratoria, formatos de competición, rúbricas de evaluación y criterios de puntuación, esenciales para que el sistema pueda aprender qué aspectos son valorados por un juez humano: claridad argumental, capacidad de refutación, coherencia, dominio del tiempo, elocuencia o seguridad, entre otros. Además, esta asociación facilitará grabaciones de debates reales, servirá como entorno de validación del sistema y participará en el diseño iterativo de los criterios de evaluación automatizada.

Nos centraremos en el formato de debate académico, que cuenta con dos equipos que debaten un tema, uno a favor otro en contra, con 4 turnos de intervención cada uno de 3-4 minutos (introducción, refutación 1, refutación 2, conclusión), y unas reglas de intervención pautadas (p.e. creo que <https://ledu.es/> se pueden ver estas normas).

Por otro lado, la Asociación de IA se encargará del desarrollo tecnológico del sistema, combinando técnicas de procesamiento de audio, lenguaje natural y análisis multimodal. En la primera fase, se integrarán herramientas de reconocimiento automático de voz con diarización (identificación de los distintos oradores) y de análisis de prosodia (entonación, ritmo, pausas, tono, variabilidad, etc.), con el fin de extraer información tanto objetiva como subjetiva del desempeño de cada participante. Posteriormente, mediante el uso de modelos de lenguaje avanzados y un sistema de *prompting* supervisado, el juez IA generará puntuaciones y comentarios razonados siguiendo las rúbricas empleadas habitualmente en competiciones de debate.

El objetivo inicial del reto es lograr un prototipo funcional que, a partir de un audio de debate, sea capaz de:

1. Transcribir automáticamente las intervenciones, identificando a los distintos oradores.
2. Analizar métricas de oratoria y expresividad basadas en el sonido y la prosodia.
3. Evaluar el contenido argumental de cada intervención según los criterios del formato de debate.
4. Emitir un informe de puntuación y retroalimentación similar al que elaboraría un juez humano.



Este reto supone una oportunidad única para fomentar la colaboración entre áreas tradicionalmente separadas —tecnología y oratoria— y para avanzar en el desarrollo de sistemas de IA explicables y aplicados a contextos educativos. La iniciativa no sólo permitirá a los estudiantes trabajar en un proyecto de frontera, sino que también contribuirá a modernizar la práctica del debate académico, dotándola de herramientas de análisis objetivo y de apoyo al aprendizaje.

## 2 ARQUITECTURA DEL SISTEMA

El sistema a desarrollar constará de una arquitectura clásica web cliente-servidor, diseñada para permitir la gestión, registro y evaluación automatizada de debates académicos. Su finalidad es asistir a los jueces humanos mediante el análisis de audio y texto, ofreciendo puntuaciones y observaciones basadas en criterios objetivos y reproducibles.

El proyecto combinará tecnologías de procesamiento de voz, análisis de prosodia, modelos de lenguaje (LLM) y visualización web, integradas en una plataforma alojada en un servidor propio en la nube de la universidad.

### 2.1 Estructura general del sistema

El sistema se organiza en dos componentes principales:

#### 2.1.1 Interfaz web (front-end)

Será la puerta de entrada para los usuarios (jueces, organizadores o miembros del club de debate). A través de esta interfaz, el usuario podrá:

- **Crear un debate** e introducir la información general: título, tema, fecha, los nombres de los participantes, asignando sus roles (Introducción, Refutador 1, Refutador 2 y Conclusor) en los equipos A y B.
- **Controlar el flujo del debate**: mediante botones o marcadores visuales podrá iniciar y detener la grabación de cada una de las 8 intervenciones (4 por equipo).
- **Visualizar resultados parciales y finales**, incluyendo:
  - Estadísticos básicos (tiempos de intervención, pausas, ritmo, palabras por minuto, etc.).
  - Valoraciones automáticas preliminares tras cada intervención.
  - Informe final con las puntuaciones agregadas y observaciones del sistema.

El front-end se desarrollará como una aplicación web, empleando un framework a elegir, buscando funcionalidad y facilidad de uso, antes de espectacularidad. La conexión con el backend será mediante API REST, siguiendo lo habitual en estos casos.

Los datos del debate (audio, transcripción, resultados) se almacenarán temporalmente en el servidor para su posterior revisión y validación.

#### 2.1.2 Back-end de procesamiento y análisis

El back-end implementará la lógica de negocio y las funciones de análisis. Exportará una REST-API. Puede desarrollarse en cualquier plataforma en la que los alumnos se sientan cómodos. Se proponen los siguientes módulos:

##### a) Módulo de procesamiento de audio y transcripción (WhisperX)

- Se empleará el sistema WhisperX (MIT License, uso gratuito), basado en Whisper de OpenAI, con mejoras en alineación temporal y diarización.
- Este módulo realizará:
  - Transcripción del audio a texto en español.
  - Alineación palabra a palabra (timestamps).
  - Identificación automática de los distintos hablantes (*diarization*), asignando etiquetas como SPEAKER\_00, SPEAKER\_01, etc.

La documentación y código fuente están disponibles en: <https://github.com/m-bain/whisperX>. Ejemplo abreviado de salida JSON que genera WhisperX:

```
{
  "segments": [
    {
      "start": 0.0,
      "end": 7.6,
      "speaker": "SPEAKER_00",
      "text": "Buenas tardes a todos...",
      "words": [
        {"word": "Buenas", "start": 0.1, "end": 0.5, "score": 0.99},
        {"word": "tardes", "start": 0.5, "end": 0.9, "score": 0.98}
      ]
    }
  ],
  "speakers": [
    {"speaker": "SPEAKER_00", "total_time": 120.5},
    {"speaker": "SPEAKER_01", "total_time": 118.3}
  ]
}
```

A partir de esta información, el sistema calculará métricas como:

- Tiempo total de intervención.
- Ritmo de habla (palabras por minuto).
- Pausas, interrupciones o solapamientos.
- Reparto del tiempo entre los oradores.
- ...

#### **b) Módulo de análisis prosódico (openSMILE 2.0)**

- Se usará openSMILE 2.0 (Apache License 2.0, uso gratuito), desarrollado por el audEERING Research Group y la Universidad Técnica de Múnich.
- Este software extrae automáticamente características acústicas y prosódicas del audio (sin requerir GPU), entre ellas:
  - F0 (pitch): media, rango, variabilidad.

- Energía (RMS/loudness): media, desviación, picos.
- Jitter y shimmer: estabilidad de la voz.
- HNR (Harmonics-to-Noise Ratio): claridad y control vocal.
- Spectral centroid / flux / slope: brillo o timbre del sonido.
- Tasa de voz/silencio y duración de pausas: fluidez.

OpenSmile proporciona ya “sin esfuerzo” el conjunto existente de parámetros eGeMAPS (88 features), reconocido internacionalmente como estándar para el análisis de prosodia y calidad de la voz (Eyben et al., *IEEE Trans. Affective Computing*, 2016). Usando procedimientos ya conocidos, estos parámetros se transformarán en estadísticos agregados (medias, desviaciones, percentiles, IQRs) y se normalizarán, permitiendo comparaciones entre oradores y turnos.

Por ejemplo, de acuerdo con la literatura, una mayor variabilidad del F0 y de la energía se asocia con mayor elocuencia y expresividad (Müller et al., *Interspeech*, 2013; Niebuhr, *Frontiers in Communication*, 2020), mientras que un bajo jitter/shimmer y un HNR alto se relacionan con seguridad y control vocal (Aronson & Pearson, *Journal of Voice*, 2018). Existen ya resultados al respecto, por lo que debemos, en un primer paso, únicamente aplicarlos.

### c) Módulo de evaluación semántica (LLM)

Los resultados de WhisperX y openSMILE se integrarán en un prompt estructurado para un modelo de lenguaje (LLM), que interpretará tanto la argumentación (coherencia, refutación, estructura) como la expresión oral. Por comodidad, el modelo de lenguaje se utilizará un modelo en la nube accedido vía API (p.e. de la familia ChatGPT-5), a través de Azure.

El LLM generará un informe textual y puntuaciones por criterios, siguiendo las rúbricas proporcionadas por la Asociación de Debate, que pueden ser en un primer momento una transcripción de las reglas de debate universitario, y del manual de jueces, adaptadas a este caso.

### d) Módulo de almacenamiento y resultados

- Almacena los audios, transcripciones, métricas y resultados en una base de datos (puede ser cualquiera, con la que los desarrolladores estén más cómodos).
- Expone los resultados al front-end a través de una API REST.
- Permite la exportación de informes (PDF/JSON) y su uso para análisis posteriores.

## 2.2 Despliegue y alojamiento

El sistema se alojará en un **servidor propio en la nube**, proporcionado por la UPCT o patrocinador, configurado con GPU (para acelerar WhisperX) y CPU optimizada (para openSMILE). Esto garantiza el control de datos y cumplimiento con las políticas de la UPCT sobre protección y procesamiento local de información.

Se proporcionará acceso a API para uso de modelo LLM, con una cuota de utilización. Deberá emplearse únicamente para este proyecto.

Se estudiarán vías de **financiación institucional o externa** (fondos de innovación docente, retos de estudiantes, patrocinio o colaboración con empresas tecnológicas), con vistas a mantener el servicio de forma sostenible.

### 3 ROL DE LA ASOCIACIÓN DE IA Y LA ASOCIACIÓN DE DEBATE

La Asociación de IA actuará como el proveedor tecnológico, relacionado con los aspectos de desarrollo tal y como se ha descrito más arriba. Por otro lado, la Asociación de Debate actuará como experto de dominio (domain expert), con responsabilidades clave en la definición y validación del sistema:

- Proporcionar audios reales de debates y autorizar su uso para entrenamiento y validación.
- Definir y consensuar las rúbricas de evaluación (criterios, escalas, ejemplos de buen desempeño).
- Colaborar en el diseño de los *prompts* empleados por el modelo de lenguaje para asegurar que las valoraciones sean coherentes con los criterios de los jueces humanos.
- Participar en la especificación del interfaz gráfico, indicando qué información resulta más útil durante una competición (tiempos, métricas, comparativas entre equipos, resúmenes narrativos, etc.).
- Evaluar el sistema en entornos reales, comparando sus juicios con los de jueces humanos y proponiendo mejoras.

Esta colaboración es esencial para garantizar que el resultado no sea solo técnicamente sólido, sino también relevante y aceptado por la comunidad de debate.

En todo caso, los roles pueden fluctuar según los intereses y disponibilidad de cada uno. Se trata de un reto abierto, la colaboración la marcan los mismos alumnos que se autoorganizan en gran medida.

### 4 PLANIFICACIÓN TEMPORAL

El objetivo final es tener un sistema prototipo, con todas las funcionalidades posibles, para Mayo de 2026, donde se presentaría en un foro que se espera multitudinario, que incluirá a toda la comunidad universitaria, empresas, y el público en general. Este foro reuniría no solo este reto, si no otros retos con otras asociaciones, que esperamos poder presentar pronto.

Para llegar a este punto, el primer paso es conformar un equipo de personas interesadas, y comenzar a dar los primeros pasos con ellos. Con ese equipo de personas, planteamos un plan de trabajo flexible, que se adapte a la realidad de los alumnos (tiempos, dedicación).