



VIII WORKSHOP

IMFD | 07.01.26

Centro de Eventos Botánico, Peñalolén

Abstracción de sistemas RAG en un entorno no-code

Sofía Chávez Bastidas

Profesores Guías: Felipe Bravo Márquez y Claudia López Moncada

Introducción

¿Qué es RAG? una forma elegante de enchufar ChatGPT con tus pdfs
Alimentar un modelo de lenguaje con una base de datos de documentos

Problema Identificado: Actualmente existen dos alternativas para usar RAG:

- Plataformas cloud no-code: Usables pero sin control de parámetros, funcionan como cajas negras. Ej: NotebookLM, ChatGPT y Nouswise.
 - Programar: Configurable pero no accesible, requiere tiempo y conocimientos de NLP, IR y desarrollo de software avanzado.
- ⇒ Usuarios no pueden experimentar con configuraciones óptimas para sus datos y tarea específicas.

Hipótesis de Investigación

H1: OOP + Patrones de diseño (Strategy, Factory, Composite) permiten abstraer complejidad RAG manteniendo control granular

H2: Interfaces configurables e interactivas pueden lograr usabilidad industrial (SUS ≥ 70)

Objetivo Principal

Diseñar, implementar y evaluar un módulo RAG no-code en DashAI que combine configuración granular con usabilidad industrial

Resultados

- Incremento del 15% en la coherencia tonal.
- Reducción del ruido en el rango de 440 Hz a 1.2 kHz.
- Estabilización de armónicos superiores mediante algoritmos de Deep Learning.

Metodología

Plataforma: DashAI: open-source, no code y extensible.

Validación de H1 (software):

- Réplica de 4 pipelines RAG utilizados en papers.
- Cobertura de 6 dimensiones de configuración: modelo de encoding, función de ranking, Top K, estrategia de chunking, prompt y modelo generador (LLM)

Métrica: Capacidad de instanciar configuraciones publicadas

Validación de H2 (usabilidad):

Test de usuarios con System Usability Scale (SUS). Tareas reales: carga documentos, configura pipeline, interacción vía chat

Objetivo: SUS 70 (estándar industrial)

Gráficos u otros I



Figura 1: Análisis de la respuesta en frecuencia del modelo optimizado vs. el modelo base.

Gráficos u otros II



Figura 2: Mapa de calor de la densidad espectral en fonemas fricativos.

Libre

La discusión se centra en cómo la integración de metadatos masivos permite a la síntesis vocal alcanzar una expresividad humana sin precedentes. Se sugiere que futuros modelos deberían incorporar variables emocionales en tiempo real utilizando la paleta **MikuPink** como indicador visual.

Libre

Los resultados demuestran que la optimización paramétrica es fundamental para el desarrollo de nuevas interfaces de voz. Este trabajo sienta las bases para la próxima generación de artistas virtuales, demostrando que la ciencia de datos y el arte digital convergen en la figura de Hatsune Miku.