

Integración de herramientas de inteligencia artificial para el diseño e implementación de un curso virtual interactivo de mezcla y masterización utilizando Moodle y Ollama, aplicado a la Carrera de Software de la Universidad de las Fuerzas Armadas ESPE

Cortez Castro Kenneth Alexander

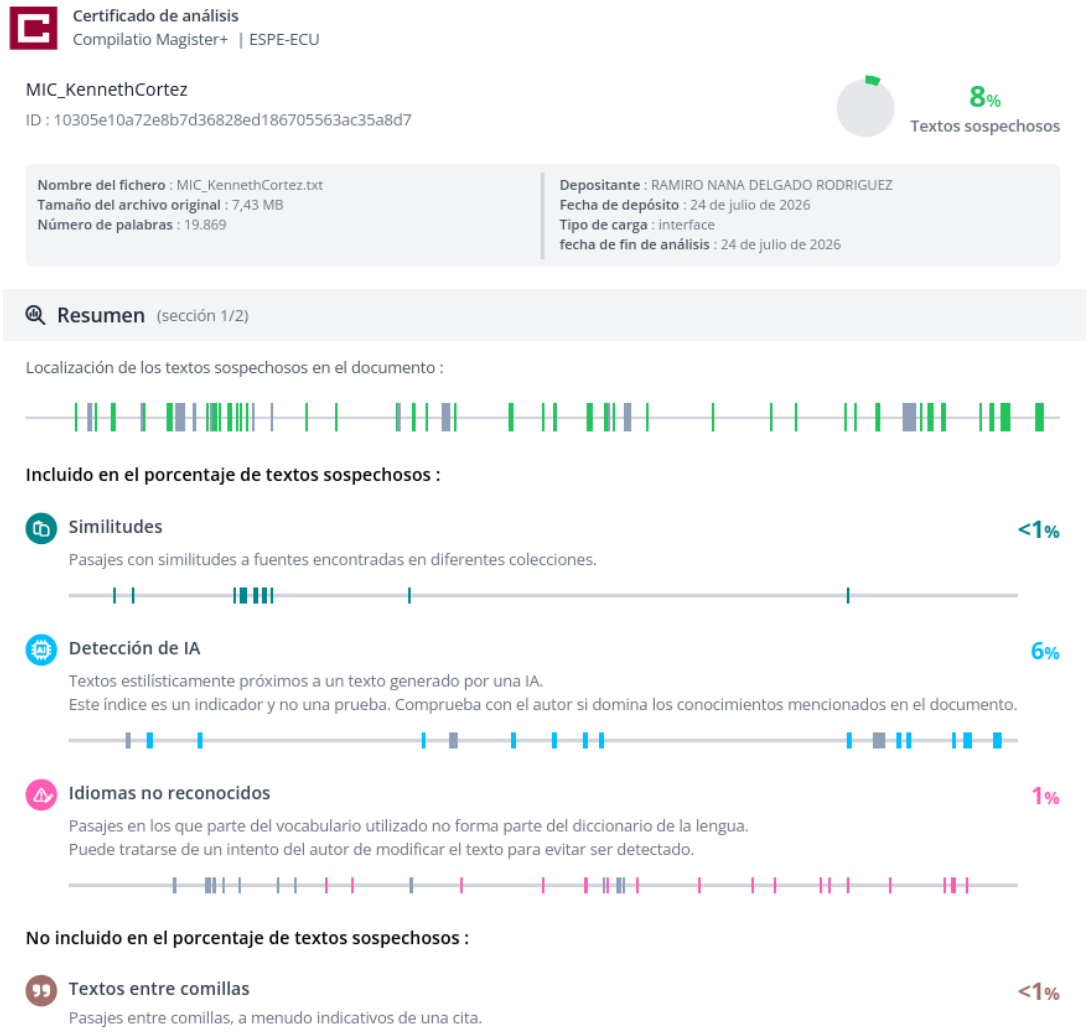
Departamento de Ciencias de la Computación

Carrera de Tecnologías de la Información

Trabajo de integración curricular, previo a la obtención del título de Ingeniero en Tecnologías de la Información

Ing. Delgado Rodríguez, Ramiro Nanac, PhD

31 de julio de 2026



Certificado de análisis
Compilatio Magister+ | ESPE-ECU

MIC_KennethCortez
ID : 10305e10a72e8b7d36828ed186705563ac35a8d7

8%
Textos sospechosos

Nombre del fichero : MIC_KennethCortez.txt
Tamaño del archivo original : 7,43 MB
Número de palabras : 19.869

Depositante : RAMIRO NANA DELGADO RODRIGUEZ
Fecha de depósito : 24 de julio de 2026
Tipo de carga : interface
fecha de fin de análisis : 24 de julio de 2026

Resumen (sección 1/2)

Localización de los textos sospechosos en el documento :

Incluido en el porcentaje de textos sospechosos :

Similitudes <1%
Pasajes con similitudes a fuentes encontradas en diferentes colecciones.

Detección de IA 6%
Textos estilísticamente próximos a un texto generado por una IA.
Este índice es un indicador y no una prueba. Comprueba con el autor si domina los conocimientos mencionados en el documento.

Idiomas no reconocidos 1%
Pasajes en los que parte del vocabulario utilizado no forma parte del diccionario de la lengua.
Puede tratarse de un intento del autor de modificar el texto para evitar ser detectado.

No incluido en el porcentaje de textos sospechosos :

Textos entre comillas <1%
Pasajes entre comillas, a menudo indicativos de una cita.

Validar únicamente en FirmaEC.
Firmado electrónicamente por:
RAMIRO NANAC
DELGADO RODRIGUEZ

Ing. Delgado Rodríguez, Ramiro Nanac, PhD

Director

Departamento de Ciencias de la Computación

Carrera de Tecnologías de la Información

Certificación

Certifico que el trabajo de integración curricular: "**Integración de herramientas de inteligencia artificial para el diseño e implementación de un curso virtual interactivo de mezcla y masterización utilizando Moodle y Ollama, aplicado a la Carrera de Software de la Universidad de las Fuerzas Armadas ESPE**" fue realizado por el señor **Cortez Castro, Kenneth Alexander,** el mismo que cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE; habiéndose efectuado una revisión integral de forma y de fondo conforme a los lineamientos y formatos institucionales vigentes. Además, fue revisado y analizada en su totalidad por la herramienta de prevención y/o verificación de similitud de contenidos; razón por la cual me permito acreditar y autorizar para que se lo sustente públicamente.

Sangolquí, 31 de julio de 2026

Validar únicamente en FirmaEC.
Firmado electrónicamente por:
RAMIRO NANAC
DELGADO RODRIGUEZ

Ing. Delgado Rodríguez, Ramiro Nanac, PhD

Director

C.C.: 1707019178

Departamento de Ciencias de la Computación

Carrera de Tecnologías de la Información

Responsabilidad de Autoría

Yo, **Cortez Castro Kenneth Alexander**, con cédula de ciudadanía n°1721613998, declaro que el contenido, ideas y criterios del trabajo de integración curricular**: Integración de herramientas de inteligencia artificial para el diseño e implementación de un curso virtual interactivo de mezcla y masterización utilizando Moodle y Ollama, aplicado a la Carrera de Software de la Universidad de las Fuerzas Armadas ESPE** es de mi autoría y responsabilidad, cumpliendo con los requisitos legales, teóricos, científicos, técnicos, y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí, 31 de julio de 2026

Validar únicamente en FirmaEC.
Firmado electrónicamente por:
KENNETH ALEXANDER
CORTEZ CASTRO

Cortez Castro, Kenneth Alexander

C.C.: 1721613998

Departamento de Ciencias de la Computación

Carrera de Tecnologías de la Información

Autorización de Publicación

Yo **Cortez Castro Kenneth Alexander**, con cédula de ciudadanía n°1721613998, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de integración curricular: **Título: Integración de herramientas de inteligencia artificial para el diseño e implementación de un curso virtual interactivo de mezcla y masterización utilizando Moodle y Ollama, aplicado a la Carrera de Software de la Universidad de las Fuerzas Armadas ESPE** en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi/nuestra responsabilidad.

Sangolquí, 31 de julio de 2026

Validar únicamente en FirmaEC.
Firmado electrónicamente por:
KENNETH ALEXANDER
CORTEZ CASTRO

Cortez Castro, Kenneth Alexander

C.C.: 1721613998

Dedicatoria

Este trabajo se lo dedico a mis padres Narda y Leonardo, a mis hermanos Arelys y Ronny, a mi abuelita Norma y a mi tía Verónica a quienes les debo mi vida entera, sin ellos no habría sido posible todo esto. A mis tíos Myriam, Jonh, Glenda, Fausto por siempre haberme apoyado en todo este proceso. Por último, a mis mascotitas Pelucha, Pelucas y Keiko †.

Agradecimiento

Quiero agradecer a mis primos Kively, Ana Paula, Cynthia, Gabriel por los momentos que me han regalado y su apoyo durante todo este tiempo.

A Ronny G. por su apoyo incondicional y también por ser como un hermano mayor para mí.

A todos mis amigos, compañeros y excompañeros de la carrera, y en especial a Michael, Saúl, Christian, Mateo, Sebastián, Nicole, Naomi, Stephani, Santiago, Irving, Christopher y Gabriel por brindarme tantos buenos momentos y hacer que mi estancia en la universidad no fuera tan difícil.

A mi tutor Ramiro Delgado por otorgarme su apoyo en la culminación de este trabajo.

Índice de contenidos

Certificación 3
Responsabilidad de Autoría 4
Autorización de Publicación 5
Dedicatoria 6
Agradecimiento 7
Resumen 15
Abstract 16
Capítulo I 17
Introducción 17
Planteamiento del problema 19
Formulación del problema o pregunta de investigación 20
Objetivos 21
Objetivo general 21
Objetivos específicos 21
Justificación 22
Alcance y delimitación 24
Capítulo II 26
Estado del arte 26
Revisión sistemática PRISMA 2020 26
Protocolo de búsqueda y criterios de selección 26
Diagrama PRISMA y matriz de estudios incluidos 29
Hallazgos por eje temático 35
Brechas identificadas 38
Marco teórico 40

Diseño instruccional y evaluación formativa en entornos virtuales 40
Plataformas LMS y Moodle en educación superior 41
Inteligencia artificial generativa aplicada a la educación 41
H5P, video interactivo y señales de aprendizaje 42
Analítica del aprendizaje y adaptación formativa 42
Modelos locales de lenguaje, Ollama y privacidad institucional 43
Recuperación aumentada por generación y tutoría contextual 44
Evaluación de calidad de un tutor basado en RAG 44
Calidad de software y validación funcional 45
Fundamentos mínimos de mezcla y masterización como dominio de aplicación 46
Escalas de percepción, consistencia interna y concordancia 46
Marco conceptual 47
Conceptos operacionales del proyecto 47
Relación entre Moodle, Ollama, RAG, tutor IA y curso virtual interactivo 47
Capítulo III 48
Metodología 48
Enfoque de investigación 48
Tipo y diseño de investigación 48
Unidad de análisis y contexto de aplicación 49
Fases metodológicas 49
Metodología de desarrollo del aplicativo 49
Herramientas y entorno de desarrollo 51
Técnicas e instrumentos 52
Validación por percepción de usuarios 54
Criterios de validación 55
Consideraciones éticas 55

Limitaciones metodológicas 56
Capítulo IV 57
Análisis y discusión de los resultados 57
Análisis de requerimientos 57
Requerimientos funcionales 57
Requerimientos no funcionales 58
Requerimientos pedagógicos 59
Diseño de la solución 59
Arquitectura general en capas 60
Diseño instruccional del curso virtual 62
Diseño del tutor contextual 63
Estrategia RAG y base de conocimiento 64
Implementación de la solución 65
Configuración del curso y despliegue en producción 65
Especificaciones técnicas del servidor y modelos de lenguaje empleados 66
Integración del backend con el modelo local y flujo teacher-driven RAG 69
Videos interactivos H5P y señales de aprendizaje 70
Tutor adaptativo, interfaz del estudiante y del docente 71
Resultados de implementación 73
Validación funcional del sistema 74
Pruebas automatizadas de backend y contratos de frontend 74
Validación de recuperación y de preguntas sobre recursos 74
Verificación en el entorno de producción 75
Resultados de la prueba piloto con estudiantes 75
Actividad H5P y desempeño observado 76
Conceptos con mayor necesidad de refuerzo 78

Interacción con el tutor durante el piloto 80
Calidad rubricada de las respuestas del tutor: matriz de evaluación docente 81
Diseño de la rúbrica y perfiles evaluadores 81
Resultados globales y por criterio 81
Puntaje por lección y casos observados 83
Resultados de la validación por percepción de usuarios 85
Resultados del grupo de estudiantes 86
Resultados del grupo docente 88
Resultado de la validación técnica del perfil administrativo 89
Análisis de las respuestas abiertas 91
Discusión de los resultados de percepción 93
Limitaciones metodológicas de la validación con usuarios 93
Áreas de mejora identificadas 96
Discusión de resultados 98
Lectura conjunta de la evidencia 98
Cumplimiento de requerimientos y de objetivos 98
Limitaciones del sistema y deuda técnica 101
Capítulo V 103
Conclusiones 103
Recomendaciones 104
Trabajos futuros 105

Índice de tablas

Tabla 1. Estrategia de búsqueda aplicada en cada base de datos 28

Tabla 2. Criterios de inclusión y exclusión de la revisión sistemática 29

Tabla 3. Matriz de estudios incluidos en la revisión sistemática 32

Tabla 4. Distribución de los estudios incluidos por eje temático 37

Tabla 5. Brechas identificadas en la literatura y aporte del presente trabajo 40

Tabla 6. Incrementos del desarrollo del aplicativo y su criterio de aceptación 51

Tabla 7. Herramientas y entorno empleados en el desarrollo del aplicativo 52

Tabla 8. Requerimientos funcionales del sistema 58

Tabla 9. Requerimientos no funcionales del sistema 59

Tabla 10. Requerimientos pedagógicos de la solución 60

Tabla 11. Componentes de la arquitectura de la plataforma 63

Tabla 12. Especificaciones técnicas del servidor de producción 69

Tabla 13. Modelos de lenguaje empleados por el sistema y su función 70

Tabla 14. Indicadores de contenido e índice de la Sección 0 del curso piloto 75

Tabla 15. Resultados de las pruebas automatizadas de backend y frontend 76

Tabla 16. Resultados de la validación funcional del sistema 77

Tabla 17. Participación y desempeño en las evaluaciones interactivas por lección 78

Tabla 18. Distribución global de los niveles de logro 80

Tabla 19. Conceptos con mayor proporción de fallo en la Sección 0 81

Tabla 20. Indicadores de la interacción con el tutor durante la prueba piloto 82

Tabla 21. Puntaje global del tutor según cada perfil evaluador 83

Tabla 22. Puntuación media por criterio de la rúbrica según perfil evaluador 84

Tabla 23. Puntaje agregado de las respuestas del tutor por lección 85

Tabla 24. Participantes de la validación por percepción según perfil 87

Tabla 25. Dimensiones evaluadas en cada instrumento de percepción 88

Tabla 26. Resultados generales del grupo de estudiantes por dimensión (n = 20) 88

Tabla 27. Resultados descriptivos del grupo docente por dimensión (n = 3) 90

Tabla 28. Valoración técnica del administrador de sistemas (n = 1) 91

Tabla 29. Síntesis temática de las respuestas abiertas por perfil 93

Tabla 30. Limitaciones metodológicas de la validación con usuarios 96

Tabla 31. Áreas de mejora priorizadas y evidencia que las sustenta 98

Tabla 32. Trazabilidad de los objetivos específicos: actividad, evidencia, resultado y limitación 101

Índice de figuras

Figura 1. Diagrama de flujo PRISMA 2020 del proceso de selección de estudios 31

Figura 2. Diagrama de procesos de la plataforma en notación BPMN 62

Figura 3. Arquitectura general de la plataforma 63

Figura 4. Secuencia de uso de la plataforma por el estudiante 65

Figura 5. Flujo de construcción del contexto del tutor por lección 66

Figura 6. Flujo teacher-driven RAG de la plataforma 67

Figura 7. Pantalla de inicio de sesión de la plataforma 68

Figura 8. Vista docente para gestionar el conocimiento del tutor 72

Figura 9. Configuración de un video interactivo H5P 73

Figura 10. Interfaz del estudiante con video, chat del tutor y recursos 74

Figura 11. Orientación adaptativa del tutor con minuto, recurso y micropráctica 74

Figura 12. Acierto promedio en las evaluaciones interactivas por lección 79

Figura 13. Conceptos con mayor proporción de fallo en la Sección 0 81

Figura 14. Puntuación media del tutor por criterio y perfil evaluador 84

Figura 15. Percepción de los estudiantes por dimensión del cuestionario 89

Resumen

El trabajo aborda una dificultad presente en los cursos virtuales de carácter práctico: ofrecer retroalimentación frecuente y técnicamente sustentada mientras el estudiante realiza las actividades. La mezcla y masterización de audio se utilizó como caso de aplicación. Para ello se integraron Moodle, una interfaz desarrollada en React y servicios de inteligencia artificial instalados en la infraestructura del proyecto. FastAPI expone las funciones del backend y LangGraph organiza el flujo del tutor. MariaDB almacenó los registros académicos, mientras que ChromaDB permitió consultar los pasajes aprobados del curso. Ollama ejecutó el modelo de lenguaje y H5P se empleó para los videos y las actividades. El profesor decidió desde la interfaz cuáles eran las fuentes disponibles para el tutor. Si los resultados de H5P mostraban una dificultad, el estudiante recibía el concepto que debía repasar, el minuto del video, un recurso y una práctica breve. La Sección 0 reunió siete lecciones, siete videos y 29 actividades. De los 20 estudiantes, 15 completaron el recorrido; los registros mostraron 70 % de aciertos. La batería del tutor alcanzó 90,4 %, con un ICC de 0,92 y sin información inventada. Los resultados técnicos y las valoraciones de los participantes se detallan en el documento. Estos datos describen el piloto, no una ganancia de aprendizaje.

Palabras clave: tutoría inteligente, inteligencia artificial generativa, Moodle, recuperación aumentada por generación, retroalimentación adaptativa.

Abstract

Learning to mix and master audio requires repeated listening, correction, and practice. In an online course, however, the teacher cannot accompany every attempt. This study developed a seven-lesson Moodle section that connects H5P activities with a locally deployed tutor. The teacher chooses the material available to the tutor; Ollama generates the response and ChromaDB retrieves the approved passages. Requirements were obtained from a PRISMA 2020 review of Springer Link, IEEE Xplore, and Scopus. The implementation was examined with backend tests, retrieval checks, H5P logs, user questionnaires, and a battery of 28 tutor questions reviewed by three specialists. The backend passed 263 tests. Retrieval succeeded in 21 of 21 queries and in 6 of 6 questions about course resources. Twenty students participated in the pilot and 15 reached the final lesson. Their H5P activity produced 131 attempts and 544 answers, 70% of which were correct. The tutor obtained 911 of 1,008 available rubric points, equivalent to 90.4%. Agreement among reviewers reached an ICC of 0.92, and none of them identified fabricated information. Students reported an average perception of 4.33 out of 5, with 91.5% favorable answers. These findings show that the components can operate together and provide guidance based on teacher-approved content. They do not establish a learning gain because the pilot had neither a baseline nor a comparison group.

Keywords: artificial intelligence, retrieval-augmented generation, Moodle, intelligent tutoring, adaptive feedback.

Capítulo I

Introducción

Este capítulo delimita el problema que motivó el trabajo, formula la pregunta de investigación, enuncia los objetivos, expone la justificación académica, técnica, institucional y pedagógica, y precisa el alcance asumido.

La formación para ingenieros en Tecnologías de la Información hoy en día implica enfrentarse a un entorno digital que evoluciona rápidamente. Existe una demanda real por profesionales capaces de combinar el rigor técnico con conocimientos muy específicos en áreas concretas. Ante este escenario, la inteligencia artificial generativa se ha incorporado con rapidez a la educación superior y se explora como una herramienta para diseñar entornos de aprendizaje, siempre que medien propósitos pedagógicos definidos, supervisión humana y evaluación de la calidad de las respuestas (UNESCO, 2023) que se ajustan al ritmo del estudiante y le ofrecen acompañamiento justo durante la práctica y no solo al momento de rendir exámenes.

De hecho, investigaciones recientes confirman que los modelos masivos de lenguaje han mostrado potencial para desempeñar funciones de tutoría y generar retroalimentación oportuna, siempre que exista un propósito pedagógico claro y controles estrictos de calidad (Kasneci et al., 2023; Wu et al., 2025). En el ámbito de la producción musical, el aprendizaje profundo ya se encarga de analizar espectros acústicos, aislar fuentes sonoras o sugerir parámetros durante los procesos de mezcla y masterización. Estos avances técnicos apoyan tareas propias de la ingeniería de audio que tradicionalmente dependían del oído experto.

Sin embargo, las universidades aún no incorporan plenamente estas ventajas en sus planes académicos. El estudiante universitario sigue limitado a las horas disponibles del profesor para sentarse con él, escuchar sus pistas, señalar errores y solicitar nuevas versiones.

Para desarrollar esta solución se tomó Moodle como base operativa. Al ser el sistema oficial de la ESPE para gestionar materias open source, resulta ideal para organizar las

actividades estudiantiles. La intención es aprovechar sus capacidades para llevar a cabo sesiones guiadas de práctica más allá del uso tradicional como plataforma para asignar notas finales. Lamentablemente, es común que en educación superior esta plataforma se limite a funcionar como cartelera de recursos o para realizar cuestionarios rápidos (Gamage et al., 2022; Ladrón-de-Guevara Moreno et al., 2024).

El proyecto propone estructurar un curso sobre mezcla y masterización apoyado por un tutor inteligente contextual. Este agente corre localmente mediante Ollama, lo que significa que el modelo opera dentro de los propios servidores del proyecto. Esta arquitectura no es casual: permite mantener bajo control institucional las consultas estudiantiles, un aspecto crítico si consideramos las advertencias sobre riesgos derivados de la datificación señaladas por Marín y Tur (2023).

Es importante aclarar que el sistema no pretende reemplazar al docente humano; su función es brindar ayuda durante los ejercicios con base exclusiva en el material previamente aprobado por el profesor.

Para la parte visual se utilizó H5P; los videos diseñados con este marco interactivo capturan las respuestas del estudiante y las convierten en métricas que el tutor emplea para orientar el repaso.

Todo este trabajo abarca desde el diseño conceptual hasta las pruebas de estrés realizadas sobre la primera unidad del curso piloto desarrollado íntegramente para la Carrera de Software en ESPE.

Cabe señalar ciertas limitaciones: no se busca comparar cuál modelo lingüístico es superior ni tampoco programar un software comercial destinado a estudios profesionales de grabación.

Los archivos de audio se incorporaron con el único fin de proporcionar al tutor un escenario real para definir las prácticas y aplicar criterios técnicos.

El trabajo se divide en cinco capítulos que describen el ciclo de vida del proyecto. El primero se encarga de delimitar la problemática, establecer los objetivos y justificar el motivo y propósito del desarrollo de esta herramienta. En el segundo capítulo se detalla el rastreo bibliográfico sistematizado (PRISMA) y se establecen las bases teóricas que sustentan la plataforma. Después en el tercero se profundiza en las metodologías empleadas y también se explica la construcción de los instrumentos de medición. El cuarto bloque presenta el desarrollo técnico: listado de requerimientos, topología arquitectónica, codificación, realización de pruebas de estrés y un análisis de los resultados obtenidos. Finalmente, el quinto segmento concluye con un balance general, sugerencias operativas y las posibles vías para la mejora posterior del sistema.

Planteamiento del problema

Aprender habilidades prácticas complejas, como la técnica de mezcla y masterización de audio, requiere una supervisión constante para corregir decisiones en tiempo real. A nivel global, aunque el procesamiento computacional apoyado en aprendizaje profundo ya permite asistir o automatizar gran parte de estas tareas (Martínez-Ramírez et al., 2022; Koszewski et al., 2023), su aplicación pedagógica en entornos formales es todavía limitada. La literatura muestra que existe una relación directa entre la calidad de la retroalimentación y la autorregulación del estudiante (Morris et al., 2021; Nedrehagen et al., 2025). Sin embargo, el principal desafío radica en que, en grupos numerosos, la demanda por acompañamiento individualizado supera ampliamente el tiempo disponible para revisión por parte del docente, de modo que los estudiantes quedan sujetos a tutoriales genéricos de videos en YouTube sin una guía continua.

En el contexto institucional universitario, sistemas como Moodle son el estándar oficial para administrar cursos y registrar actividades académicas, sin embargo, la evidencia indica que estas plataformas suelen usarse mayormente como repositorios estáticos o para evaluaciones tipo opción múltiple y desaprovechan su potencial para el apoyo de prácticas

guiadas (Gamage et al., 2022; Ladrón-de-Guevara Moreno et al., 2024). A esto se suma el reto relacionado con la gestión de la información debido a que incorporar tutores basados en inteligencia artificial alojados en servidores externos comerciales expone datos y patrones de aprendizaje ax terceros. Ante este volumen de recopilación de datos, resulta indispensable fijar directrices estrictas de privacidad que resguarden la información de los usuarios dentro del entorno académico (Marín & Tur, 2023).

Al observar específicamente la realidad de la Carrera de Software en la Universidad de las Fuerzas Armadas ESPE, se evidencia una fuerte desconexión: la plataforma Moodle, el diseño curricular y las nacientes herramientas de inteligencia artificial funcionan como islas completamente independientes. Los estudiantes conocían los principios del análisis de señales, pero el curso no aprovechaba sus respuestas para indicarles qué contenido necesitaba repaso. El prototipo atendió ese problema mediante la conexión de Moodle con los materiales curriculares y con un modelo instalado localmente. De esta manera, las orientaciones respetaban lo definido por el profesor y las consultas permanecían dentro de la infraestructura del proyecto.

Formulación del problema o pregunta de investigación

A partir de esta problemática, la investigación busca responder la siguiente pregunta:

¿De qué manera la integración de un tutor de inteligencia artificial generativa ejecutado localmente, basado en recuperación aumentada por generación gobernada por el docente y en señales de aprendizaje provenientes de actividades H5P, permite diseñar e implementar un curso virtual interactivo de mezcla y masterización con retroalimentación adaptativa sobre la plataforma Moodle?

La interrogante se abordó en tres etapas. En la primera se definieron los requerimientos del curso y del tutor. En la segunda se estableció la conexión entre Moodle y el modelo local y se aseguró además la conservación de la fuente utilizada en cada respuesta. Por último, en la tercera se establecieron los criterios de revisión de la recuperación de contenidos, los eventos

que registra H5P y las orientaciones que da al estudiante. Esta organización permitió el desarrollo de un prototipo, previo a la evaluación de su funcionamiento.

Preguntas de investigación

- ¿Qué condiciones técnicas, pedagógicas y funcionales requiere el curso para conectar Moodle, las actividades H5P y un tutor local?
- ¿De qué forma pueden integrarse Moodle, H5P, Ollama y la recuperación de documentos sin retirar al docente el control de las fuentes?
- ¿Qué muestran las pruebas técnicas, los registros del piloto, la revisión del tutor y las opiniones de los participantes sobre el funcionamiento de la propuesta?

Objetivos

Objetivo general

Integrar herramientas de inteligencia artificial mediante agentes locales basados en Ollama en el diseño e implementación de un curso virtual interactivo de mezcla y masterización sobre la plataforma Moodle, con el fin de fortalecer las competencias técnicas interdisciplinarias en audio digital de los estudiantes de la Carrera de Software de la Universidad de las Fuerzas Armadas ESPE.

Objetivos específicos

- Realizar una revisión sistemática de literatura aplicando la metodología PRISMA sobre enseñanza virtual de producción musical, IA aplicada a audio y plataformas LMS.
- Definir los requerimientos técnicos, pedagógicos y funcionales del curso en Moodle aplicando las mejores prácticas de la ingeniería de requisitos.
- Implementar el curso virtual en Moodle y desarrollar los componentes interactivos basados en IA usando Ollama.

- Evaluar la efectividad del curso y la precisión del sistema de retroalimentación asistida por IA.
- Elaborar conclusiones, recomendaciones y trabajos futuros.

Justificación

La relevancia del presente proyecto de investigación se puede entender desde cuatro perspectivas que lo complementan: la académica, la técnica, la institucional y la pedagógica. En el ámbito académico, el principal aporte radica en lograr la convergencia de varias disciplinas que habitualmente se enseñan por separado, con el propósito de entrelazar la infraestructura virtual, el procesamiento computacional de señales de audio y el auge actual de la inteligencia artificial generativa dentro de las aulas, lo cual resulta fundamental en la actualidad.

Aunque existe abundante literatura sobre cada área de forma aislada, es verdaderamente poco común encontrar estudios que integren estos campos dentro de un ecosistema real para clases presenciales o virtuales, un vacío que esta investigación busca solucionar. Autores como Wu et al. (2025) y Reddig et al. (2025) enfatizan este punto crucial: un chatbot carece realmente de utilidad si no está rigurosamente alineado con un plan docente estructurado y evaluable. En respuesta directa a esta necesidad pedagógica, esta investigación documenta abiertamente toda la arquitectura tecnológica, de modo que todo el proceso resulta completamente transparente y verificable, desde su inicio hasta su finalización.

Desde el aspecto técnico, se tomó una decisión clave: instalar y ejecutar el modelo con Ollama en servidores propios. Esta decisión contribuyó a disminuir la dependencia de servicios externos al mantener las consultas fuera de las APIs comerciales. Por ello, cada orientación conserva la referencia al documento que la fundamenta y el tutor responde únicamente con el contenido aprobado previamente. La implementación local se acompañó de medidas de control de acceso, monitoreo y respaldo. Además, Moodle permaneció como la plataforma central

debido a su uso institucional previo, mientras que los microservicios se añadieron a la infraestructura existente sin alterar la administración del curso ni los registros académicos.

En el aspecto institucional, uno de los mayores retos fue proteger la privacidad del estudiantado, porque los datos escolares pueden quedar expuestos a usos indebidos o fines comerciales cuando se recurre a tecnologías de terceros y no haya marcos legales claros que regulen su tratamiento (Marín & Tur, 2023). Al desplegar la inteligencia artificial dentro del perímetro local del campus universitario se reduce la transmisión de esta información a terceros; no obstante, la protección efectiva no depende únicamente del despliegue local, sino que exige combinarlo con minimización de datos, control de accesos, políticas de conservación y seguridad, en línea con los deberes previstos en la Ley Orgánica de Protección de Datos Personales del Ecuador y su reglamento (Ley Orgánica de Protección de Datos Personales, 2021).

Por último, desde el aspecto formativo, se busca mitigar una frustración común tanto para los estudiantes como para los docentes dentro de materias técnicas, y es el no poder supervisar a veinte personas en cada pantalla al mismo tiempo. La bibliografía sobre procesos de evaluación (Morris et al., 2021; Nedrehagen et al., 2025) resalta que corregir un error en el momento preciso resulta mucho más efectivo que hacerlo semanas después, y justamente este tutor basado en inteligencia artificial se encarga de esa primera línea y apoya los ensayos prácticos sin sustitución de la autoridad del profesor titular. Se seleccionaron los procesos de mezcla y masterización de sonido como banco de pruebas porque requieren un estado de alerta total, que obliga al usuario a afinar su escucha, comprender valores físicos y justificar las decisiones de utilizar ciertos plugins específicos para la modificación del audio.

Toda la implementación aprovechó la tecnología ya disponible y aplicó estándares comprobados en ingeniería de software para someter el código a pruebas; así se logró poner en marcha la primera etapa del curso piloto dentro de los plazos del semestre vigente y sin dependencia de servicios pagos de inteligencia artificial.

La contribución de este trabajo de titulación no se agota en la redacción de un documento teórico, sino que aporta soluciones tecnológicas completamente operativas y sujetas a auditoría. Como respaldo práctico a la investigación, se adjunta el aula virtual ya configurada, el código fuente que da vida al orquestador, los manuales técnicos para levantar la infraestructura desde cero y la base de datos cruda con las métricas obtenidas durante las evaluaciones. Los componentes entregados permiten revisar el funcionamiento del prototipo y ofrecen una base para incorporar tutores semejantes en otras asignaturas. El proyecto también se relaciona con el Objetivo de Desarrollo Sostenible 4, debido a que utiliza tecnología educativa para ampliar las posibilidades de acompañamiento durante la práctica. Esta relación expresa la finalidad del desarrollo y no demuestra por sí misma un impacto institucional.

Alcance y delimitación

Desde la perspectiva operativa, el proyecto considera la implementación de un entorno educativo conformado por una interfaz web renovada, un intermediario de servicios, un backend encargado de la lógica del sistema, la base de datos relacional de Moodle, un repositorio vectorial y un motor generativo desplegado localmente. Bajo este esquema tecnológico cobró vida la “Sección 0: El sistema de decisión”, un módulo introductorio conformado por siete lecciones apoyadas con material H5P, lecturas seleccionadas, un tutor conversacional anclado a cada temática y un chat global para resolver dudas generales. El control total sobre la información recae en manos del profesor ya que a través del mismo sitio web puede colocar resúmenes de las lecciones, subir recursos para cada una y ajustar las directrices del tutor sin necesidad de tener conocimientos en programación o bases vectoriales.

A medida que el estudiante interactúa con el reproductor H5P, el sistema recopila sus respuestas y en base a sus métricas sobre aciertos o errores, el algoritmo dirige la atención del alumno hacia sus puntos más débiles, le señala con exactitud qué minuto del video debería revisar de nuevo y lo invita a completar un desafío rápido.

La fiabilidad del desarrollo fue evaluada mediante tres canales distintos: primero se realizaron pruebas automatizadas sobre el backend para certificar precisión al recuperar fragmentos textuales en ambiente productivo. 20 estudiantes iniciaron el piloto y 15 completaron las siete lecciones. Los registros permitieron reconocer las dudas repetidas y los contenidos que requerían de una orientación. Tres especialistas calificaron las 28 respuestas del tutor; después se aplicaron cuestionarios diferentes a los estudiantes, a los docentes y a el personal técnico.

Los datos muestran cómo respondió el software y qué actividad se registró durante el piloto. Sin una línea de base, un grupo de comparación y un seguimiento prolongado, no es posible el atribuir al sistema una mejora del aprendizaje. Tampoco puede suponerse que los resultados se repetirán en estudiantes con otros perfiles. La medición de cambios en el rendimiento deberá abordarse en una investigación posterior con un diseño experimental apropiado. De igual manera, el proyecto no busca determinar qué motor de red neuronal funciona mejor, ni pretende sustituir las herramientas profesionales empleadas actualmente en la industria musical. El procesamiento de las ondas de audio se utilizó únicamente como un medio técnico para estructurar las tareas y evaluar la habilidad del agente conversacional.

Capítulo II

Estado del arte

Este capítulo presenta la revisión sistemática realizada con el protocolo PRISMA 2020, los hallazgos por eje temático, las brechas detectadas y el marco teórico y conceptual que sustentan el diseño de la plataforma.

Revisión sistemática PRISMA 2020

Con la finalidad de fundamentar el diseño de la solución con la evidencia reciente, se realizó una revisión sistemática de la literatura conforme al protocolo PRISMA 2020, que establece las pautas para la identificación, la selección, la evaluación de elegibilidad y la inclusión de estudios, así como para reportar cada una de estas fases con total transparencia (Page et al., 2021). Se revisaron las experiencias educativas que combinaran modelos generativos con la tutoría, RAG, LMS, H5P o algunos de los mecanismos de adaptación.

El proceso quedó registrado con PRISMA 2020. Los resultados de los tres repositorios se reunieron y se depuraron antes de revisar los títulos y los resúmenes. Los trabajos pertinentes pasaron a la lectura completa. La matriz conserva el resultado de cada una de las evaluaciones y la causa aplicada.

Protocolo de búsqueda y criterios de selección

Las búsquedas se efectuaron en Springer Link, IEEE Xplore y Scopus debido a su cobertura en las áreas de ingeniería de software, inteligencia artificial y tecnología educativa. Las consultas generaron un total de 1865 registros, distribuidos en 796 resultados de Springer Link, 479 de IEEE Xplore y 590 de Scopus. Los cuatro grupos temáticos fueron combinados mediante el operador AND, mientras que los términos relacionados dentro de cada grupo se vincularon mediante OR. Estos grupos incluyeron modelos generativos, tutores basados en RAG, Moodle y H5P, además de analítica de aprendizaje y retroalimentación. La Tabla 1 presenta las ecuaciones, los campos consultados y los resultados de cada repositorio. La

lógica de búsqueda se mantuvo; solamente se adaptaron comillas, comodines y otros elementos exigidos por cada plataforma.

Tabla 1

Estrategia de búsqueda aplicada en cada base de datos

Base	Expresión de búsqueda	Campos	Filtros aplicados	Registros
Springer Link	("generative artificial intelligence" OR "large language model" OR LLM) AND ("intelligent tutoring" OR "AI tutor" OR "retrieval-augmented generation" OR RAG) AND ("learning management system" OR Moodle OR H5P) AND (feedback OR "formative assessment" OR "learning analytics")	Título, resumen y palabras clave	Periodo 2023–2026; artículos de revista y capítulos; inglés y español; búsqueda del 2 de mayo de 2026	796
IEEE Xplore	("All Metadata":"generative AI" OR "large language model") AND ("All Metadata":"intelligent tutoring" OR "retrieval augmented generation") AND ("All Metadata":"learning management system" OR Moodle OR H5P) AND ("All Metadata":feedback OR "learning analytics")	Todos los metadatos	Periodo 2023–2026; artículos de revista y actas de congreso; búsqueda del 2 de mayo de 2026	479

Base	Expresión de búsqueda	Campos	Filtros aplicados	Registros
Scopus	TITLE-ABS-KEY(("generative AI" OR "large language model*") AND ("intelligent tutor*" OR "retrieval-augmented generation") AND ("learning management system" OR Moodle OR H5P) AND (feedback OR "formative assessment" OR "learning analytics"))	Título, resumen y palabras clave	PUBYEAR > 2022; DOCTYPE artículo o ponencia; idioma inglés o español; búsqueda del 2 de mayo de 2026	590

La selección de los estudios se realizó con base en criterios de inclusión y exclusión establecidos previamente al proceso de cribado, con el propósito de reducir la subjetividad en las decisiones de selección.. Se consideraron únicamente publicaciones posteriores a 2022, dado que el interés de la revisión se concentra en la generación de sistemas educativos posteriores a la difusión masiva de los modelos de lenguaje de gran tamaño. La Tabla 2 resume los criterios aplicados y la razón de exclusión asociada a cada uno.

Tabla 2

Criterios de inclusión y exclusión de la revisión sistemática

Dimensión	Criterio de inclusión	Criterio de exclusión (razón asociada)
Periodo de publicación	Estudios publicados entre 2023 y 2026	Publicaciones de 2022 o anteriores (Razón 1)
Pertinencia temática	Relación directa con IA educativa, tutores inteligentes o RAG en contextos de enseñanza	Sin relación con IA, tutor o RAG educativo (Razón 2)
Ámbito de aplicación	Investigaciones situadas en el ámbito educativo o de aprendizaje	Estudios fuera del ámbito educativo (Razón 3)

Dimensión	Criterio de inclusión	Criterio de exclusión (razón asociada)
Tipo de contribución	Trabajos que se desarrollan, se implementan, se evalúan o se analizan a un sistema, a una plataforma, a un tutor, a un chatbot, LMS, H5P, a una analítica o un mecanismo de retroalimentación	Trabajos sin desarrollo, implementación ni evaluación de un sistema o mecanismo (Razón 4)
Componente evaluativo	Estudios que reportan evaluación, retroalimentación, analítica, personalización o desempeño del estudiante	Sin evaluación, feedback, analítica ni personalización (Razón 5)
Acceso y tipo documental	Documentos completos, accesibles y de tipo documental válido	Idioma no considerado, documento incompleto, duplicado residual o tipo documental no válido (Razón 6)

Diagrama PRISMA y matriz de estudios incluidos

La Figura 1 resume el recorrido de selección. Se identificaron 1865 registros y se compararon títulos, autores y DOI. Gracias a esa revisión se detectaron 292 duplicados, por lo que 1573 registros continuaron al cribado.

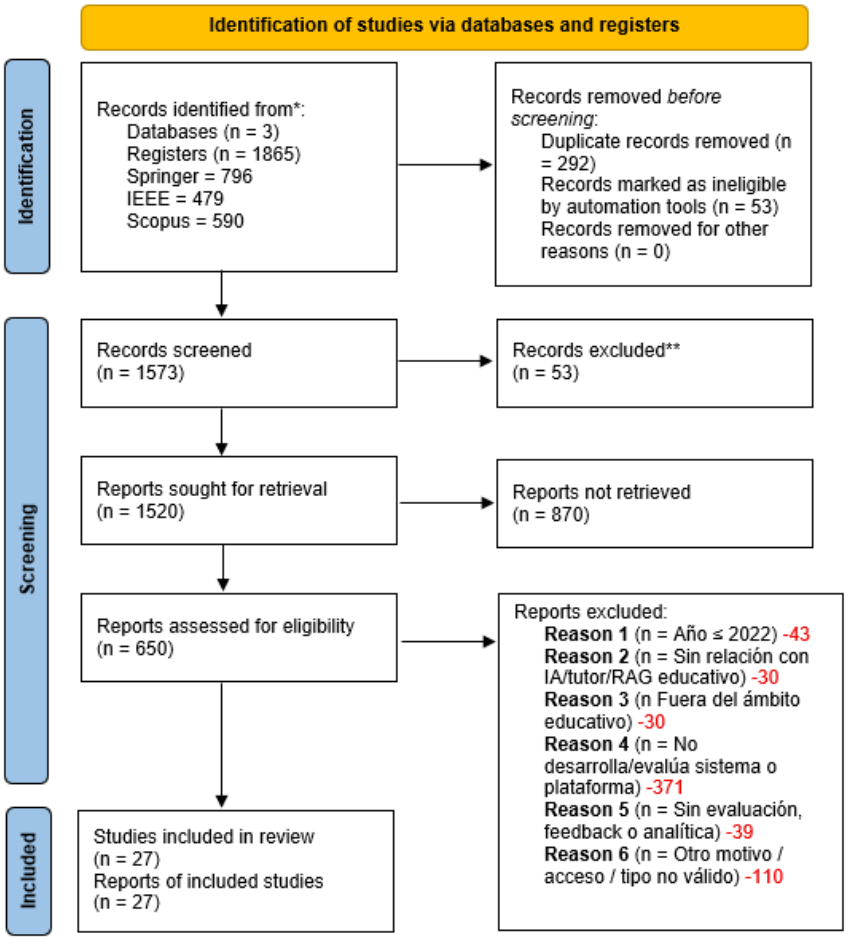
Antes de revisar títulos y resúmenes se aplicaron reglas sobre el año y el tipo de documento a los metadatos exportados. Cincuenta y tres registros no cumplieron esos criterios y se los retiraron.

Después del cribado se buscaron los textos completos de 1520 registros. No fue posible obtener 870 por falta de suscripciones, por lo que se recurrió a copias abiertas o depositadas por sus autores. De los 650 documentos revisados se excluyeron 623: 43 por el periodo, 30 por el tema, 30 por estar fuera de educación, 371 porque no desarrollaban o evaluaban herramientas, 39 por no incluir retroalimentación y 110 por problemas documentales. Al final

quedaron 27 artículos. La falta de una herramienta desarrollada o evaluada explicó el 60 % de las exclusiones.

Figura 1

Diagrama de flujo PRISMA 2020 del proceso de selección de estudios



Nota. Elaboración propia a partir de la plantilla de diagrama de flujo PRISMA 2020 (Page et al., 2021).

La revisión incluyó 27 estudios. Veinticuatro fueron publicados en 2025 o 2026; trece procedían de Springer Link, once de Scopus y tres de IEEE Xplore. La Tabla 3 presenta sus autores, títulos y fuentes de publicación.

Tabla 3

Matriz de estudios incluidos en la revisión sistemática

N.º	Autores (año)	Título	Publicación	Base
1	Semerikov et al. (2026)	Development and validation of an adaptive learning readiness assessment framework for Moodle courses	Journal of New Approaches in Educational Research	Springer
2	Villegas-Ch et al. (2025)	Explainable educational assistant integrated in Moodle: Automated semantic assessment and adaptive tutoring based on NLP and XAI	Discover Artificial Intelligence	Springer
3	Saez et al. (2026)	Evaluating large language models for AI-assisted grading: A framework and case study in higher education	Scientific Reports	Springer
4	Paavilainen et al. (2026)	Implementing learning design with learning analytics to scaffold the regulation of collaboration in primary education	Smart Learning Environments	Springer
5	Ahangama (2026)	Designing assessments in the generative AI era: A tailored assessment framework for ICT tertiary education	International Journal of Educational Technology in Higher Education	Springer

N.º	Autores (año)	Título	Publicación	Base
6	Liu, J. et al. (2025)	Design and evaluation of ChatGPT-MWPS: An AI-enhanced learning system for improving primary students' mathematical word problem solving	Smart Learning Environments	Springer
7	Alfredo et al. (2025)	Co-designing AI-powered learning analytics: Bringing students and teachers together	International Journal of Educational Technology in Higher Education	Springer
8	Uzun et al. (2025)	Engagement with analytics feedback and its relationship to self-regulated learning competence and course performance	International Journal of Educational Technology in Higher Education	Springer
9	Steinert et al. (2024)	Harnessing large language models to develop research-based learning assistants for formative feedback	Smart Learning Environments	Springer
10	Zhou et al. (2025)	Deep learning based knowledge tracing in intelligent tutoring systems	Scientific Reports	Springer
11	Scheu et al. (2026)	Designing a mobile chatbot-based learning journaling system for intrinsic motivation and engagement	International Journal of Educational Technology in Higher Education	Springer

N.º	Autores (año)	Título	Publicación	Base
12	Fu (2025)	Integrating reinforcement learning with dynamic knowledge tracing for personalized learning path optimization	Scientific Reports	Springer
13	Yin et al. (2025)	Effects of different AI-driven chatbot feedback on learning outcomes and brain activity	npj Science of Learning	Springer
14	Lee, Y. O. et al. (2026)	Personalized auto-grading and feedback system for constructive geometry tasks using large language models on an online math platform	IEEE Access	IEEE Xplore
15	Chang et al. (2026)	A dual-agent intelligent tutoring system for teaching two-step fraction word problems	IEEE Transactions on Learning Technologies	IEEE Xplore
16	Lee & Joe (2025)	A GPT-based code review system with accurate feedback for programming education	IEEE Access	IEEE Xplore
17	Tobler & Köhler (2026)	A lecture-specific AI-based tutor for higher education: Pedagogical design and empirical evaluation	Education Sciences	Scopus

N.º	Autores (año)	Título	Publicación	Base
18	Sappa (2025)	Assessing the impact of large language models on the scalability and efficiency of automated feedback mechanisms in massive open online courses	Indian Journal of Information Sources and Services	Scopus
19	Hilliger et al. (2026)	A GAI-based chatbot for scaffolding self-regulated learning: Insights from a design-based research approach	Journal of Learning Analytics	Scopus
20	Suen & Hung (2026)	An interpretable closed-loop intelligent tutoring system for multimodal affective feedback in asynchronous presentation training	IEEE Transactions on Learning Technologies	Scopus
21	Yermaganbetova et al. (2026)	Evaluating the impact of an AI-integrated learning platform on student performance: A quasi-experimental study	Frontiers in Education	Scopus
22	Valente et al. (2026)	An AI-driven conversational platform for the learning and assessment of programming logic	IMX 2026 — Proceedings of the ACM International Conference on Interactive Media Experiences	Scopus
23	Martínez-Araneda et al. (2025)	Bridging code and timely feedback: Integrating generative AI into a programming platform	PeerJ Computer Science	Scopus

N.º	Autores (año)	Título	Publicación	Base
24	Papachristou et al. (2026)	A fuzzy logic-driven system for interpretable and behavior-aware student assessment: E-Teacher Assistant case study	Electronics	Scopus
25	Gaeta et al. (2025)	Enhancing traditional ITS architectures with large language models for generating motivational feedback	Computers and Education: Artificial Intelligence	Scopus
26	Liu, S. et al. (2024)	Advancing generative intelligent tutoring systems with GPT-4: Design, evaluation, and a modular framework for future learning platforms	Electronics	Scopus
27	Yildirim-Erbasli & Bulut (2023)	Conversation-based assessment: A novel approach to boosting test-taking effort in digital formative assessment	Computers and Education: Artificial Intelligence	Scopus

Hallazgos por eje temático

Para el análisis, los estudios incluidos se agruparon en seis ejes temáticos derivados de la clasificación registrada en la matriz de revisión. Como la mayoría de los trabajos aborda varias dimensiones a la vez, los ejes no son mutuamente excluyentes; la Tabla 4 muestra la distribución y algunos estudios representativos de cada uno.

Tabla 4

Distribución de los estudios incluidos por eje temático

Eje temático	n	Estudios representativos
IA generativa, LLM y tutores conversacionales	24	Villegas-Ch et al. (2025); Saez et al. (2026); Steinert et al. (2024); Liu, S. et al. (2024)
Tutoría inteligente, adaptación y personalización	14	Chang et al. (2026); Zhou et al. (2025); Fu (2025); Gaeta et al. (2025)
LMS, Moodle y cursos virtuales	9	Semerikov et al. (2026); Yermaganbetova et al. (2026); Sappa (2025)
Evaluación digital y analítica del aprendizaje	11	Paavilainen et al. (2026); Alfredo et al. (2025); Ahangama (2026)
Retroalimentación y recomendación	11	Yin et al. (2025); Lee y Joe (2025); Suen y Hung (2026)
Diseño y evaluación de sistemas educativos con tecnología	27	Hilliger et al. (2026); Scheu et al. (2026); Tobler y Köhler (2026)

El primer eje muestra un cambio desde los chatbots generales hacia asistentes diseñados específicamente con el material de una asignatura. Villegas-Ch et al. (2025) integran en Moodle un asistente explicable que incluye evaluación semántica y tutoría adaptativa; Saez et al. (2026) y Lee, Y. O. et al. (2026) investigan la calificación asistida por modelos de lenguaje, mientras que Steinert et al. (2024) y Liu, S. et al. (2024) describen asistentes formativos y arquitecturas modulares. La capacidad técnica ya está disponible; lo que queda por resolver es garantizar la fiabilidad y fundamentar cada respuesta en evidencia verificable.

El segundo eje se refiere a la adaptación y personalización. Zhou et al. (2025) y Fu (2025) emplean el rastreo del conocimiento y el aprendizaje por refuerzo para ajustar las rutas de estudio; Chang et al. (2026) distribuyen la tutoría entre dos agentes; Gaeta et al. (2025) fusionan estímulos motivacionales con infraestructuras convencionales, mientras que Suen y Hung (2026) suman modelos de retroalimentación afectiva que resultan fáciles de interpretar. Todos estos esfuerzos logran articular a las tecnologías maduras con la generación automática

de texto; sin embargo, persiste un desafío muy clave: el dotar al sistema de la capacidad para justificar y explicar de forma transparente cada una de sus decisiones pedagógicas (Papachristou et al., 2026).

Por otro lado, la tercera línea de investigación reúne a los entornos virtuales para la gestión del aprendizaje. Aquí figuran estudios como el de Semerikov et al. (2026), quienes auditan qué tan listos están los cursos en Moodle para soportar metodologías adaptativas. A la vez, Yermaganbetova et al. (2026) validan un entorno educativo con la inteligencia artificial a través del diseño cuasiexperimental. También destaca Sappa (2025) al medir qué tan escalable resulta automatizar las correcciones, a la par que Martínez-Araneda et al. (2025) y Valente et al. (2026) consiguen acoplar motores generativos dentro de aulas de programación. Pese a estos avances, se nota un patrón común: la gran mayoría de estas arquitecturas dependen fuertemente de proveedores comerciales en la nube y apenas logran integrarse de forma superficial con el sistema de gestión académica.

La analítica de datos y las nuevas formas de evaluación digital protagonizan el cuarto bloque temático. Paavilainen et al. (2026), por ejemplo, conectan la planificación de tareas con métricas de uso para fomentar el trabajo colaborativo. Alfredo et al. (2025) subrayan lo vital que resulta involucrar tanto a profesores como alumnos al momento de configurar estos medidores, mientras que Uzun et al. (2025) vinculan directamente dichas estadísticas con el rendimiento y la autonomía del estudiante.

Con la exploración de nuevas metodologías, Ahangama (2026) reflexiona sobre cómo auditar el conocimiento en plena era de la inteligencia artificial, y su aporte se complementa con la visión de Yildirim-Erbasli y Bulut (2023) sobre las pruebas en formato de diálogo. De todo este grupo se extrae una lección clara: aunque rastrear el comportamiento del alumno genera métricas invaluables, son contadas las propuestas que logran inyectar esa información en tiempo real hacia un asistente virtual.

En cuanto al quinto enfoque, la atención recae puramente sobre la mecánica de retroalimentación. Yin et al. (2025) se dedican a contrastar el impacto real que tienen las diferentes respuestas emitidas por un agente conversacional. Desde otra óptica, Lee y Joe (2025) ponen a prueba las revisiones algorítmicas de código fuente para universitarios de informática. Simultáneamente, Hilliger et al. (2026) estructuran un bot interactivo con la misión de impulsar la autogestión del ritmo de estudio en el alumno.

Finalmente, el último eje reúne estudios dedicados a diseñar y evaluar sistemas educativos completos: Tobler y Köhler (2026) prueban un tutor construido para una asignatura específica; Scheu et al. (2026) desarrollan un sistema conversacional para registros reflexivos, mientras que Liu, J. et al. (2025) evalúan una plataforma destinada a resolver problemas matemáticos complejos.

Aunque sus métodos evaluativos son más rigurosos que los trabajos anteriores, aún son escasos los sistemas que integren LMS, contenido interactivo, analítica educativa junto con un tutor generativo ejecutado localmente.

Brechas identificadas

La revisión pone en evidencia avances significativos en tutores inteligentes, evaluación formativa, analítica educativa y plataformas digitales educativas; sin embargo pocas soluciones reúnen todos estos componentes simultáneamente.

En particular destacan como escasos los trabajos que combinan Moodle con H5P y un corpus administrado por docentes junto a modelos locales e intervención basada estrictamente en desempeño.

La Tabla 5 recoge estas brechas e indica cuáles son abordadas por la plataforma desarrollada específicamente para este estudio.

Tabla 5

Brechas identificadas en la literatura y aporte del presente trabajo

Brecha identificada	Evidencia en la literatura	Aporte del presente trabajo
Integración de extremo a extremo entre LMS, contenidos interactivos, RAG y tutor generativo	Los estudios abordan componentes de forma aislada: asistentes en Moodle, analíticas o tutores, pero rara vez el flujo completo	Plataforma que articula Moodle, H5P, ChromaDB y Ollama en una sola arquitectura orientada a servicios
Gobernanza docente del corpus de conocimiento del tutor	El contenido que alimenta a los tutores suele ser curado por desarrolladores o proviene del modelo general	Flujo teacher-driven RAG: el docente aprueba transcripciones, recursos y contexto desde la interfaz, sin manipular artefactos técnicos
Uso de modelos de lenguaje locales en entornos educativos	Predominan servicios comerciales en la nube, con implicaciones de costo y privacidad	Inferencia local con Ollama, que mantiene los datos de estudiantes y contenidos dentro de la infraestructura institucional
Conexión en tiempo real entre señales de aprendizaje y tutoría generativa	Las analíticas del aprendizaje se reportan en los tableros o en los informes, desacopladas del tutor	Señales de aprendizaje generadas por interacciones H5P que orientan la guía del tutor hacia conceptos, minutos de video, recursos y microprácticas
Dominios de aplicación creativos y técnico-artísticos	Predominan programación, matemática e idiomas como dominios de prueba	Curso de mezcla y masterización de audio como caso de aplicación del diseño instruccional y del tutor contextual

Estas brechas definen el ámbito de contribución del proyecto: no se busca desarrollar un nuevo modelo de lenguaje ni un algoritmo de adaptación innovador, sino diseñar, implementar y validar funcionalmente una integración coherente de componentes ya consolidados, guiada por un principio de gobernanza docente del conocimiento y situada en un campo de aplicación poco explorado en la literatura.

Marco teórico

Diseño instruccional y evaluación formativa en entornos virtuales

El propósito central de la evaluación formativa es capturar métricas en tiempo real mientras transcurre la actividad y ofrecer herramientas prácticas para que tanto el alumnado como el profesorado adapten sus estrategias sobre la marcha.

La eficacia de este modelo no radica en saturar al estudiante con múltiples observaciones, sino en intervenir en el instante exacto con indicaciones sumamente específicas y ancladas a las reglas de la materia (Morris et al., 2021). Bajo esta misma lógica, autores como Nedrehagen et al. (2025) argumentan que la retroalimentación debe fluir como una conversación constante durante toda la ruta de aprendizaje, sin la antigua costumbre de emitir un veredicto final o aislado. Al llevar este concepto hacia los ecosistemas digitales, la teoría se materializa a través de canales conversacionales que habilitan espacios inmediatos para la corrección (Yildirim-Erbasli & Bulut, 2023).

El diseño instruccional organiza objetivos, contenidos, actividades y evaluación dentro de una misma secuencia (Paavilainen et al., 2026). Cuando la plataforma registra lo que sucede durante esa secuencia, los datos pueden apoyar al estudiante en tiempo real durante la actividad y no solo alimentar informes posteriores (Paavilainen et al., 2026). Por ello cada lección combina video, recursos y práctica, y las respuestas del estudiante determinan qué orientación se presenta a continuación.

Plataformas LMS y Moodle en educación superior

Los sistemas de gestión del aprendizaje centralizan la administración de cursos, usuarios, contenidos y calificaciones en educación superior. Moodle es una de las plataformas open source más adoptadas y cuenta con una comunidad investigadora consolidada sobre sus usos en enseñanza y aprendizaje (Gamage et al., 2022). Sin embargo, diversos estudios señalan que su uso habitual se limita a distribuir recursos y aplicar cuestionarios, con un aprovechamiento reducido de sus potencialidades formativas (Ladrón-de-Guevara Moreno et al., 2024).

Dos motivos hacen a Moodle especialmente adecuado para este proyecto. Sus servicios web permiten conectarlo con aplicaciones externas sin perder el control centralizado sobre cursos y usuarios (Moodle, 2025). Además, la literatura ya ofrece marcos para evaluar si un curso Moodle está preparado para experiencias adaptativas, lo que confirma la relevancia de ampliar sus funciones personalizadoras (Semerikov et al., 2026).

Inteligencia artificial generativa aplicada a la educación

Los modelos de lenguaje generan explicaciones y retroalimentación sin depender de reglas escritas caso por caso. Esta flexibilidad también implica riesgos: sesgos, respuestas incorrectas expresadas con seguridad o dependencia excesiva del estudiante son algunos problemas señalados por Kasneci et al. (2023). Reddig et al. (2025) y Steinert et al. (2024) obtienen mejores resultados cuando las respuestas se condicionan tanto por la tarea como por información específica del estudiante, en lugar de basarse únicamente en el conocimiento general del modelo.

La irrupción de estas herramientas tecnológicas exige, ineludiblemente, una reconfiguración de los esquemas de evaluación. Al respecto, Ahangama (2026) plantea la necesidad de estructurar nuevos marcos de calificación pensados expresamente para la era de la IA generativa en las universidades. En paralelo, Saez et al. (2026) analizan los beneficios y las barreras que surgen al delegar las correcciones a motores lingüísticos. A partir de toda esta

literatura, la presente investigación adopta una premisa fundamental: el verdadero potencial didáctico de un modelo de lenguaje no radica en su vasto conocimiento general, sino en la destreza que tenga para manipular los textos exactos de la asignatura y en el respeto permanente a las directrices impuestas por el profesor.

H5P, video interactivo y señales de aprendizaje

Para materializar este enfoque, se recurrió a H5P, un entorno de código abierto especializado en la creación de recursos dinámicos, como videos interactivos, cuestionarios integrados y ejercicios de repaso. Su gran ventaja es que se acopla de forma nativa a sistemas como Moodle y reporta la actividad del alumnado a través de protocolos informáticos estandarizados (H5P Group, 2025).

En el video interactivo, las evaluaciones se insertan en puntos estratégicos de la línea de tiempo. Gracias a esto, cualquier respuesta ingresada por el usuario queda amarrada a un segundo exacto de la grabación y, en consecuencia, al tema puntual que se expone en ese fragmento. Es precisamente esta triangulación entre la pregunta, la marca de tiempo y el concepto lo que transforma una simple respuesta en una auténtica "señal de aprendizaje" y aporta métricas tangibles para detectar dónde tropieza el alumno. Toda esta interacción se empaqueta y se guarda con el protocolo xAPI (ADL Initiative, 2025).

Con todo, la simple acumulación de bases de datos carece de sentido práctico. Tal como afirman Alfredo et al. (2025) y Uzun et al. (2025), la analítica cobra verdadera relevancia únicamente cuando desencadena una intervención concreta; para efectos de este desarrollo, esa acción se traduce en desplegar una ruta de nivelación directamente en la pantalla de la clase.

Analítica del aprendizaje y adaptación formativa

El campo de la analítica educativa se dedica a procesar el rastro digital que dejan los alumnos, con el fin de interpretar y potenciar su trayectoria académica. Si bien las bitácoras de un aula virtual son excelentes insumos para detonar alertas, emitir sugerencias o brindar

apoyo, el simple hecho de almacenar estos registros no es prueba irrefutable de que el estudiante haya asimilado el conocimiento (Pan et al., 2024).

A partir de esta base, el actual proyecto no asume las métricas de H5P, ya sean de equivocaciones, aciertos, reintentos o tiempos de permanencia, como un veredicto definitivo sobre las habilidades adquiridas. Los registros de H5P funcionan aquí como señales de actividad. El sistema los consulta para seleccionar el contenido de repaso, pero no los considera una demostración de que se adquirió una competencia. La orientación cambia de acuerdo con los resultados observados, sin modificar el currículo ni atribuir al estudiante conocimientos que no fueron medidos.

Las bandas únicamente determinan si corresponde continuar o revisar un contenido. Sin embargo, los puntos de corte dependen del contexto y de la dificultad de las tareas, por lo que las bandas empleadas en el piloto tienen una función operativa y no constituyen una escala universal de dominio (Winget y Persky, 2022).

Modelos locales de lenguaje, Ollama y privacidad institucional

Llevar los motores generativos a las aulas obliga a resolver un dilema fundamental: establecer quién custodia realmente los apuntes de la materia y las interacciones del alumnado. Al depender de proveedores externos, todo este flujo de datos se procesa en servidores comerciales sometidos a políticas corporativas que escapan de la jurisdicción universitaria. Dado que el simple historial de preguntas de un estudiante es capaz de dejar al descubierto sus debilidades académicas, sus ritmos de estudio y hasta datos de carácter sensible, resulta imperativo que los escudos de seguridad nazcan entrelazados con la propia arquitectura del software, en lugar de añadirse después como una simple medida correctiva (Marín & Tur, 2023).

La ejecución local mantiene las conversaciones y el corpus dentro de la infraestructura institucional; este tratamiento debe alinearse igualmente con los deberes de protección y seguridad de datos previstos en la normativa ecuatoriana (Ley Orgánica de Protección de

Datos Personales, 2021). Ollama facilita este despliegue: descarga, ejecuta y expone modelos abiertos mediante una interfaz local (Ollama, 2025). Esta opción tiene un límite claro: en este proyecto, la selección del modelo estuvo condicionada por la memoria y la capacidad de procesamiento del servidor disponible. Para compensar esto, la plataforma recupera contenido del curso antes de generar la respuesta y así disminuye la dependencia del conocimiento general del modelo.

Recuperación aumentada por generación y tutoría contextual

La recuperación aumentada por generación, o RAG, antepone una búsqueda a la consulta del modelo. El sistema identifica los fragmentos más relevantes para la pregunta y los añade como contexto para generar la respuesta (Lewis et al., 2020); así, la base puede actualizarse sin necesidad de reentrenar el modelo y la conservación de los fragmentos recuperados y sus metadatos puede facilitar la trazabilidad del material utilizado. ChromaDB almacena esos fragmentos como las representaciones vectoriales y los recupera mediante la similitud semántica (Chroma, 2025).

En tutoría, RAG permite que el asistente responda con base en los materiales específicos de la asignatura. Tobler y Köhler (2026) evaluaron un tutor específico por cátedra y Villegas-Ch et al. (2025) demostraron su utilidad integrada en Moodle. La propuesta añade el control del profesor al proceso RAG. El docente selecciona las transcripciones, los recursos y el contexto que pueden entrar en la base. En este estudio, esa administración se denomina teacher-driven RAG.

Evaluación de calidad de un tutor basado en RAG

La evaluación de RAG debe comprobar tanto la pertinencia de los pasajes recuperados como el respaldo que esos pasajes ofrecen a la respuesta. En educación, una respuesta recuperada también tiene que ser comprensible y permitir que el estudiante continúe la tarea. Incorporar pasajes previos disminuye la generación de afirmaciones sin respaldo, pero no la

hace imposible. El modelo todavía puede interpretar mal una fuente o ir más allá de lo que realmente afirma el contexto (Es et al., 2024).

Para evaluar un tutor educativo, la fidelidad a las fuentes debe revisarse junto con la utilidad pedagógica de la respuesta. Una respuesta puede ser correcta y, aun así, ofrecer poca orientación al estudiante. Por ello, la evaluación de tutores basados en modelos de lenguaje también considera dimensiones como coherencia, claridad, orientación, posibilidad de actuar sobre la respuesta y adecuación a la dificultad observada en el estudiante (Maurya et al., 2025). En este trabajo, “información inventada” se refiere operacionalmente a una afirmación que no está respaldada por el corpus aprobado o que contradice sus contenidos.

Calidad de software y validación funcional

No todas las propiedades del software tienen la misma importancia en cada contexto de uso. La norma ISO/IEC 25010:2023 ofrece una referencia para examinar la adecuación funcional, la eficiencia de desempeño, la compatibilidad, la usabilidad, la fiabilidad, la seguridad, la mantenibilidad y la portabilidad. En el prototipo, los requerimientos funcionales, no funcionales y pedagógicos definieron los comportamientos y características que debían ser verificados durante la evaluación. en la Sección 0 (International Organization for Standardization, 2023).

Los casos automatizados permitieron verificar las funciones específicas del backend. Por otro lado, las pruebas de contrato evaluaron la estructura de los datos intercambiados con la interfaz. Después del despliegue se comprobó que los servicios principales siguieran disponibles. Los resultados corresponden a la versión y a las condiciones empleadas en la prueba. No prueban que el software carezca de errores ni anticipan su respuesta ante cargas diferentes. En esta investigación, sus resultados se limitan a los escenarios ejecutados durante la validación del prototipo.

Fundamentos mínimos de mezcla y masterización como dominio de aplicación

La mezcla organiza y procesa las pistas mediante ecualización, dinámica, paneo y efectos hasta construir una salida estéreo coherente. La masterización toma esa salida y ajusta su nivel y balance para distintos sistemas de reproducción. En ambas etapas se combinan la lectura de medidores, la escucha y las decisiones técnicas. Curiosamente, la automatización de la mezcla ha despertado un enorme interés dentro de la ingeniería informática e impulsa los proyectos apoyados en las redes neuronales profundas (Koszewski et al., 2023; Martínez-Ramírez et al., 2022); esto confirma que la gran parte del razonamiento acústico es susceptible de modelarse computacionalmente. Las actividades del piloto se construyeron a partir de decisiones habituales de mezcla y masterización. El tutor explicaba esos procedimientos con la consulta exclusiva de los contenidos revisados por el profesor.

Escalas de percepción, consistencia interna y concordancia

Las respuestas de la escala Likert permitieron ordenar el grado de acuerdo con cada afirmación. Para cada dimensión se calculó el alfa de Cronbach. Un coeficiente alto muestra que los reactivos están relacionados, aunque también puede aparecer cuando varias preguntas resultan demasiado parecidas. Por sí solo, este valor no demuestra validez, unidimensionalidad ni estabilidad en otras muestras (McNeish, 2018). En el presente estudio se informa únicamente como una referencia exploratoria.

La fiabilidad interevaluador permite conocer cuánto varían las puntuaciones cuando varias personas aplican la misma rúbrica. El ICC cuantifica esa concordancia, pero su variante debe elegirse de acuerdo con el diseño, el tipo de acuerdo y la interpretación individual o promedio de las calificaciones. El acuerdo absoluto exige puntuaciones semejantes, mientras que la consistencia solo exige que los evaluadores mantengan un patrón relativo parecido; asimismo, la medida individual representa la fiabilidad de un solo evaluador (Koo y Li, 2016).

Marco conceptual

Conceptos operacionales del proyecto

Dentro de este marco de investigación, se entiende como curso virtual interactivo a toda aquella sucesión de módulos que entrelaza material audiovisual, documentación de apoyo y ejercicios prácticos, con una interfaz vinculada de forma directa con Moodle.

Con teacher-driven RAG, el docente elige desde la interfaz los documentos que se guardan en ChromaDB y el tutor solo consulta ese material. H5P aporta los intentos, las respuestas y los datos de reproducción. Cuando aparece una dificultad, queda disponible una guía con el concepto, el minuto, un recurso y una actividad breve. El chat general usa una ayuda neutral; en cada lección, ese nivel puede configurarse de otra manera. Precisamente, para someter a prueba todo este entramado lógico, se construyó y desplegó en su totalidad la "Sección 0: El sistema de decisión".

Relación entre Moodle, Ollama, RAG, tutor IA y curso virtual interactivo

En este ecosistema, Moodle actúa como la bóveda principal encargada de salvaguardar el expediente académico, con el listado de matriculados, la malla de lecciones y los paquetes evaluativos de H5P. De manera paralela, la aplicación web asume el rol de visualizar dichos módulos y monitorear permanentemente las acciones del alumno. Durante este proceso, en la parte interna del sistema, los documentos previamente validados son segmentados e incorporados automáticamente en la base de datos vectorial para su posterior recuperación. Ante una pregunta, el backend recupera pasajes de la lección y añade los resultados H5P recientes del estudiante. Ollama recibe ese contexto y genera la respuesta.

El resultado de cada actividad queda registrado. Si requiere repaso, el tutor consulta las fuentes autorizadas. El pr ofesor controla los recursos y la configuración. El Capítulo IV muestra la arquitectura de este flujo.

Capítulo III

Metodología

Este capítulo describe el enfoque, el tipo y el diseño de investigación, la unidad de análisis, las fases metodológicas, la metodología de desarrollo del aplicativo, las herramientas empleadas, los instrumentos, los criterios de validación y las limitaciones asumidas.

Enfoque de investigación

El trabajo se concentró en construir la solución y observar su uso en la Sección 0. No se planteó estimar un efecto para toda la población. En primer lugar, se comprobó el funcionamiento de los servicios; después se revisaron los registros de H5P y las respuestas del tutor; por último, se recogió la opinión de los participantes. La plataforma se completó mediante las entregas sucesivas.

El análisis no trató como equivalentes todos los datos obtenidos. Las pruebas, los contratos de interfaz, las consultas de recuperación, la disponibilidad de los servicios y la latencia se revisaron como la evidencia técnica. Los intentos H5P, los aciertos, los conceptos pendientes y las consultas al tutor correspondieron al uso de la plataforma. En un tercer grupo quedaron las respuestas de estudiantes, docentes y administración. Para los dos primeros grupos se emplearon conteos, porcentajes y tiempos; los cuestionarios se resumieron por perfil y sus respuestas abiertas se clasificaron según los temas mencionados.

Tipo y diseño de investigación

El estudio partió de un problema concreto y terminó en un prototipo que podía comprobarse, por lo que se trabajó como una investigación aplicada con desarrollo experimental (OECD, 2015). No se asignaron tratamientos, no se manipularon las variables y tampoco se compararon los grupos; el diseño fue no experimental. La revisión PRISMA 2020 sirvió para establecer los requerimientos. Con esa base se construyó la plataforma por las iteraciones y se la examinó mediante las pruebas técnicas, los registros producidos durante el piloto y los cuestionarios posteriores.

Unidad de análisis y contexto de aplicación

El análisis se centró en la Sección 0, El sistema de decisión. Las siete lecciones —SEC2-R55 a SEC2-R61— contienen videos H5P, recursos aprobados, tutores por lección, un chat general y registros de actividad. El desarrollo se realizó en la Carrera de Software de la ESPE y todos los servicios funcionaron en el servidor del proyecto.

Fases metodológicas

El proyecto se desarrolló en seis fases. La revisión en Springer Link, IEEE Xplore y Scopus permitió formular los requerimientos. Luego se prepararon las lecciones, los guiones, las actividades y los recursos de la Sección 0. La siguiente fase conectó la interfaz, el backend, Moodle, MariaDB, ChromaDB y Ollama. La aprobación del material se trasladó después a la interfaz docente, de modo que el profesor no tuviera que modificar archivos técnicos. Después se configuraron los siete videos y sus eventos. La comprobación técnica precedió al piloto con 20 estudiantes, la revisión de 28 respuestas y los cuestionarios de los tres perfiles.

Metodología de desarrollo del aplicativo

La construcción del aplicativo no siguió un marco ágil de equipo como Scrum, porque el desarrollo estuvo a cargo de una sola persona y no existieron los roles ni las ceremonias que ese marco supone. Se adoptó un ciclo de vida iterativo e incremental, apropiado para proyectos cuyos requerimientos se precisan a medida que el producto se vuelve verificable (Pressman y Maxim, 2020; Sommerville, 2016). Cada incremento cerró con un entregable ejecutable, desplegado y probado, de modo que el alcance creció sobre una base ya validada.

El desarrollo se organizó en ocho incrementos, cada uno con un objetivo funcional propio y un criterio de aceptación verificable, según resume la Tabla 6. Los incrementos 1 a 4 levantaron la plataforma y el tutor; los incrementos 5 y 6 trasladaron el control del conocimiento al docente e incorporaron las señales de aprendizaje; los incrementos 7 y 8 cerraron la experiencia de uso, la documentación y el despliegue. El paso de un incremento al siguiente exigió que las pruebas del backend y los contratos del frontend se ejecutaran sin fallos.

Tabla 6

Incrementos del desarrollo del aplicativo y su criterio de aceptación

Incremento	Objetivo funcional	Criterio de aceptación verificado
1	Plataforma base y arquitectura orientada a servicios	Interfaz, puerta de enlace y backend bajo un solo origen
2	Tutor con recuperación aumentada y verificación posterior	Respuesta con fuentes y bloqueo fuera del dominio
3	Autoría docente: perfil pedagógico, asistente y momentos	Borrador editable promovido sin reindexación
4	Corpus por secciones y migración de ejes a secciones	Recuperación por curso, sección y lección
5	Flujo teacher-driven RAG en la Sección 0	Corpus canónico sustituido por recursos del docente
6	Videos H5P, señales de aprendizaje y tutor adaptativo	Guía con concepto, minuto, recurso y micropráctica
7	Cierre de la experiencia de uso, chat general y nivel de ayuda	Guía persistente y chat general con respuesta neutral
8	Documentación, auditoría y despliegue de cierre	Pruebas, contratos, salud y prueba de humo correctos

Nota. Cada incremento cerró con un entregable desplegado en el servidor del proyecto y con su evidencia registrada en los reportes de cierre.

Las prácticas de ingeniería se mantuvieron constantes: control de versiones con ramas integradas a la rama principal antes de cada despliegue, pruebas automatizadas escritas junto con el código, verificación de los contratos entre el frontend y el backend, y comprobación del entorno de producción con pruebas de salud y de humo. El modelo de calidad de la norma ISO/IEC 25010:2023 orientó los atributos observados en cada cierre (International Organization for Standardization, 2023).

Herramientas y entorno de desarrollo

El trabajo se realizó sobre un entorno local de desarrollo y un servidor de despliegue. La Tabla 7 detalla las herramientas empleadas en cada actividad. Las características del servidor y de los modelos de lenguaje se describen en el Capítulo IV.

Tabla 7

Herramientas y entorno empleados en el desarrollo del aplicativo

Actividad	Herramienta	Uso en el proyecto
Código y versiones	Visual Studio Code y Git	Escritura del código y trazabilidad de cada incremento
Backend	Python, FastAPI y LangGraph	Chat, recuperación, autoría docente y señales
Frontend	React, Vite y Tailwind CSS	Interfaz de estudiante, docente y administrador
Dependencias	pip y npm	Instalación reproducible de las bibliotecas
Pruebas y análisis	pytest, ESLint y Node.js	Pruebas del backend y contratos del frontend
Verificación en servidor	Guiones de consola y cURL	Pruebas de salud y de humo en producción
Despliegue	Docker, Docker Compose y nginx	Ejecución de los servicios y enrutamiento único
Datos	MariaDB y ChromaDB	Registros del curso e índice de pasajes aprobados
Modelos locales	Ollama y faster-whisper	Modelos de lenguaje y transcripción en el servidor
Contenido interactivo	H5P sobre Moodle	Siete videos interactivos y 29 interacciones
Observabilidad	Loki, Promtail y Grafana	Registro de los eventos de cada solicitud
Modelado y redacción	Bizagi Modeler, Mermaid y Microsoft Word	Diagramas del sistema e informe del trabajo

Técnicas e instrumentos

La revisión, el desarrollo y el piloto no se examinaron con el mismo instrumento. La matriz PRISMA conservó la decisión tomada para cada publicación. Los diagramas registraron la arquitectura y sus flujos. La recuperación se comprobó mediante 21 consultas distribuidas por lección y seis preguntas sobre los recursos aprobados; H5P aportó las señales. Los reportes de cierre documentaron ChromaDB y los demás servicios. En el piloto se añadieron la rúbrica del tutor y los cuestionarios.

El propio sistema se encargó de almacenar un registro detallado por cada estudiante y lección, esto incluyó la calificación obtenida en los videos H5P, el detalle de cada respuesta ingresada, el concepto abordado y el nivel de logro alcanzado. El tutor registró la latencia, la intención de la consulta y el uso de las señales y de las evidencias. Antes del análisis se definieron tres bandas para activar las guías: menos de 60 % le correspondía a «conviene reforzar», del 60 % al 79 % a «en progreso» y desde el 80 % a «desempeño esperado». Estas categorías se las utilizaron dentro del prototipo; no forman parte de la escala oficial de la ESPE ni certifican el dominio. El 80 % se apoyó en referencias de aprendizaje para el dominio y el 60 % se fijó como la alerta operativa del prototipo (Winget y Persky, 2022; Akinpelu et al., 2024).

La batería reunió 28 preguntas, cuatro por lección y diferentes de las consultas libres. Los tres especialistas calificaron las mismas respuestas de 0 a 2 en seis criterios: la corrección técnica, la coherencia, la claridad, el uso de recursos, la utilidad y la ausencia de la información inventada. El especialista en audio revisó las decisiones técnicas; el docente, la explicación; y el especialista en IA, el respaldo recuperado. La elección de estos aspectos se apoyó en las evaluaciones recientes de los sistemas RAG y de los tutores basados en modelos de lenguaje (Es et al., 2024; Maurya et al., 2025). Las 28 preguntas corresponden al alcance del piloto y no representan todas las consultas que el sistema podría recibir.

Cabe mencionar que las actividades interactivas en H5P no funcionaron bajo una lógica de rúbrica analítica compleja. Cada actividad H5P registró una respuesta como correcta o

incorrecta. La suma se utilizó para asignar uno de los tres niveles y decidir qué contenido debía repasarse. La rúbrica de seis criterios se aplicó únicamente a las respuestas del tutor.

Después, para obtener el porcentaje del puntaje máximo de la rúbrica (puntaje normalizado de calidad de las respuestas) se dividió el puntaje real obtenido entre el máximo posible. El puntaje máximo resultó del producto de las 28 preguntas por los seis criterios y por los dos puntos disponibles en cada criterio. Así, cada evaluador podía asignar hasta 336 puntos y el total de los tres evaluadores era de 1008. El tutor obtuvo 911 puntos, equivalentes al 90,4 %. Este valor en sí, resume el desempeño en la batería aplicada; no corresponde a una precisión de la clasificación ni demuestra que el sistema produzca el efecto causal.

$$\text{PM} = p \cdot c \cdot m \cdot j = 28 \cdot 6 \cdot 2 \cdot 3 = 1008$$

$$\text{Puntaje del tutor} = \frac{911}{1008} \times 100\% = 90{,}4\%$$

donde p es el número de las preguntas de la batería (28), c son los criterios de la rúbrica (6), m es el puntaje máximo por criterio (2) y j es el número de los evaluadores (3); 911 es el puntaje total obtenido. El porcentaje expresa la parte del puntaje máximo conseguido en la rúbrica; no es una medida de precisión de la clasificación.

El ICC se calculó con las puntuaciones que los tres especialistas asignaron a las 28 respuestas. Se utilizó un modelo bidireccional mixto, de acuerdo absoluto y medida individual, porque interesaba conocer si otorgaban valores semejantes y no únicamente si coincidían en el orden de los casos. La medida individual estima la fiabilidad de la puntuación otorgada por un solo evaluador (Koo y Li, 2016). A su vez, emplear la medida individual ayuda a interpretar qué tan confiable sería el proceso si la revisión recayera sobre un único profesor. Este análisis estadístico tomó en cuenta las notas globales de las 28 interacciones y quedó plasmado en la hoja de trabajo correspondiente.

$$\text{ICC} = \frac{\text{MS}_\text{R} - \text{MS}_\text{E}}{\text{MS}_\text{R} + (k - 1)\text{MS}_\text{E} + \left(\frac{k}{n}\right)(\text{MS}_\text{C} - \text{MS}_\text{E})}$$

donde MS_r, MSc y MS_e son los cuadrados medios entre respuestas, entre evaluadores y del error, respectivamente; n = 28 respuestas y k = 3 evaluadores. El coeficiente y su intervalo de confianza se obtuvieron con software estadístico conforme a Koo y Li (2016).

Por último, la validación no estaría completa sin considerar la perspectiva de los involucrados. Con este fin, se aplicaron tres cuestionarios estructurados y adaptados a cada tipo de perfil de usuario mediante formularios web una vez concluida la fase piloto, con la garantía de que la participación fuera totalmente voluntaria y anónima. La composición exacta de estas encuestas, sus escalas de valoración y el tratamiento de los datos se detallan a profundidad en el apartado siguiente.

Validación por percepción de usuarios

Después del uso respondieron los cuestionarios 20 estudiantes, tres docentes y un administrador, seleccionados por disponibilidad. 11 estudiantes indicaron el nivel de principiante en audio, 8 en el nivel intermedio y 1 en el nivel avanzado. Cada perfil se analizó por separado. Para los estudiantes se calculó la media, la desviación estándar, el porcentaje favorable y el alfa de Cronbach exploratorio (Ferrando et al., 2025; McNeish, 2018). Las respuestas de los docentes y las del administrador se resumieron sin los identificadores. Un investigador agrupó a los comentarios; sin una segunda codificación, el análisis cualitativo se limita al piloto (Braun y Clarke, 2024).

Los formularios no tuvieron la misma extensión. El de los estudiantes reunió 20 reactivos cerrados y tres preguntas abiertas; el de los docentes, 18 reactivos y cuatro preguntas abiertas; y el del administrador, 18 reactivos cerrados. Las opciones fueron: 1, totalmente en desacuerdo; 2, en desacuerdo; 3, ni de acuerdo ni en desacuerdo; 4, de acuerdo; y 5, totalmente de acuerdo. Para revisar los comentarios, primero se conservaron separados los tres perfiles y después se reunieron las ideas que aparecían más de una vez. La clasificación fue realizada por el investigador, de modo que se la describe en las respuestas del piloto y no pretende demostrar la saturación temática alguna.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$s = \sqrt{\frac{1}{(n-1)}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\% \text{ favorable} = \frac{n_{4,5}}{n} \times 100$$

Para interpretar las fórmulas, x_i corresponde a cada una de las respuestas, $\bar{x}$ al promedio y n al número de observaciones del cuestionario. $n_{4,5}$ reúne únicamente las respuestas con valores de 4 o de 5.

Criterios de validación

La comprobación técnica incluyó las pruebas, los contratos, las 21 consultas, las 6 preguntas sobre los recursos, las señales H5P y los controles del despliegue. La batería se aceptó con el 80 %, el 1,8 sobre 2 y ninguna información inventada. En los estudiantes se usaron el 4,00 y más del 80 % favorable. No se definieron los límites para los docentes ni para el administrador. El contraste final aparece en el Capítulo IV.

Los puntos de corte ya estaban definidos cuando comenzó el piloto. Se tomó el 80 % como el desempeño esperado y el 60 % como el aviso para iniciar el refuerzo. Su función se limita a orientar el comportamiento del prototipo; no corresponden a la escala de las calificaciones de la ESPE. También se acotó la revisión del tutor a las 28 preguntas, a los 3 evaluadores y a los 6 criterios. Por esta razón, el resultado informa sobre esa batería concreta y no permite anticipar cómo responderá el sistema ante cualquier consulta.

Consideraciones éticas

Cada participante conoció el propósito académico y respondió voluntariamente los cuestionarios sin identificadores. Los resultados se analizaron por perfil. Se utilizaron los materiales propios o los autorizados; los documentos y las consultas permanecieron en la infraestructura del proyecto y no se enviaron a los servicios externos.

En los formularios no se pidieron los nombres ni los correos. Para el análisis se conservaron únicamente las respuestas y el perfil del participante; los resultados se muestran agrupados y se utilizaron para este trabajo académico.

Una investigación que añada las mediciones del aprendizaje o del seguimiento a largo plazo deberá solicitar su propia aprobación institucional.

Limitaciones metodológicas

La evaluación reúne las pruebas, los registros de desempeño y las opiniones, pero no observa el aprendizaje en varios de los momentos. El 70 % corresponde a las respuestas obtenidas en las actividades H5P. Para atribuir un cambio al sistema harían falta mediciones equivalentes antes y después y, cuando sea posible, un grupo de referencia.

La muestra por conveniencia estuvo integrada por 20 estudiantes, 3 docentes y un administrador. El número de participantes no permite extrapolar los resultados. La participación voluntaria y el conocimiento del estudio también pudieron influir en las respuestas. Por esa razón, los datos docentes y técnicos se presentan descriptivamente.

La implementación completa cubrió solamente la Sección 0. El tutor se revisó con tres jueces, 28 preguntas y una corrida. El ICC describe ese conjunto y no todas las preguntas posibles. Como el modelo puede variar entre ejecuciones, las calificaciones corresponden a esa corrida. Los indicadores técnicos son los valores consolidados de reportes reproducibles al cierre.

Capítulo IV

Análisis y discusión de los resultados

Este capítulo desarrolla el análisis de requerimientos, el diseño de la solución, su implementación, los resultados técnicos de la construcción de la Sección 0 del curso piloto y la validación funcional del sistema, para cerrar con la discusión de los hallazgos frente a los objetivos planteados. Los datos reportados provienen de los reportes de auditoría y validación generados durante el cierre del desarrollo.

Análisis de requerimientos

Los requerimientos surgieron de la revisión del Capítulo II, de las necesidades concretas del curso de mezcla y masterización y de las restricciones institucionales sobre privacidad y control de datos. Para facilitar su seguimiento se agruparon en funcionales, no funcionales y pedagógicos, y se priorizaron según su impacto en el recorrido completo de la Sección 0.

Requerimientos funcionales

La Tabla 8 reúne las funciones que debía ofrecer la plataforma. El recorrido esperado arranca en una lección con video H5P; las respuestas generan información sobre el desempeño y el tutor recurre al material aprobado por el profesor para explicar o proponer un repaso. Cada requerimiento describe un paso de ese flujo.

Tabla 8

Requerimientos funcionales del sistema

Código	Descripción	Prioridad
RF01	Gestionar el curso piloto y sus lecciones en Moodle, con roles diferenciados de estudiante y docente.	Alta
RF02	Reproducir videos interactivos H5P con preguntas incrustadas en cada lección.	Alta
RF03	Registrar los eventos de interacción del estudiante con los videos H5P y derivar señales de aprendizaje.	Alta

Código	Descripción	Prioridad
RF04	Ofrecer un tutor de inteligencia artificial por lección que responda con base en el contexto aprobado por el docente.	Alta
RF05	Ofrecer un chat general del curso con alcance neutral, sin personalización por lección.	Media
RF06	Permitir al docente cargar, describir y aprobar recursos y transcripciones desde la interfaz, sin manipular archivos técnicos.	Alta
RF07	Indexar de forma automática en la base vectorial el material previamente aprobado por el docente.	Alta
RF08	Generar orientaciones adaptativas que remitan al concepto con mayor dificultad, al segmento correspondiente del video, al recurso de apoyo y a la micropráctica más adecuados para el estudiante.	Alta
RF09	Permitir configurar el tono y el nivel de ayuda del tutor por lección.	Media
RF10	Presentar la guía adaptativa de forma recuperable mediante los indicadores visuales y los sonoros.	Media

Requerimientos no funcionales

Los requerimientos no funcionales, recogidos en la Tabla 9, priorizan la privacidad de los datos, la operabilidad del despliegue y la posibilidad de verificar el comportamiento del sistema mediante pruebas automatizadas.

Tabla 9

Requerimientos no funcionales del sistema

Código	Descripción	Prioridad
RNF01	Ejecutar la inferencia del modelo de lenguaje en infraestructura propia, sin envío de datos de estudiantes a servicios externos.	Alta
RNF02	Adoptar una arquitectura orientada a servicios, contenedorizada y reproducible mediante Docker Compose.	Alta
RNF03	Permitir verificar la disponibilidad mediante comprobaciones de salud y pruebas de humo en el entorno de producción.	Alta
RNF04	Respaldar la mantenibilidad con una suite de pruebas automatizadas de backend y contratos de frontend.	Alta

Código	Descripción	Prioridad
RNF05	Posibilitar la extensión a nuevas secciones y cursos sin rediseñar la arquitectura.	Media
RNF06	Garantizar la trazabilidad de las respuestas del tutor hacia el contexto aprobado por el docente.	Alta
RNF07	Ofrecer interfaces utilizables por estudiantes y docentes sin conocimientos técnicos de la infraestructura.	Media

Requerimientos pedagógicos

Junto a los requerimientos técnicos, el diseño atendió requerimientos pedagógicos dirigidos a que la tecnología sostenga la evaluación formativa y la autonomía del estudiante, según resume la Tabla 10.

Tabla 10

Requerimientos pedagógicos de la solución

Código	Requerimiento pedagógico
RP01	Integrar evaluación formativa embebida en el video mediante preguntas H5P distribuidas a lo largo de cada lección.
RP02	Proporcionar retroalimentación inmediata y accionable a partir del desempeño observado del estudiante.
RP03	Mantener la gobernanza del docente sobre el contenido que sustenta las respuestas del tutor.
RP04	Alinear la orientación adaptativa con los conceptos y los objetivos de cada lección.
RP05	Graduar la ayuda ofrecida para fomentar la autonomía del estudiante en lugar de sustituir al esfuerzo.
RP06	Minimizar las respuestas ajenas al contexto del curso en las interacciones con el tutor.

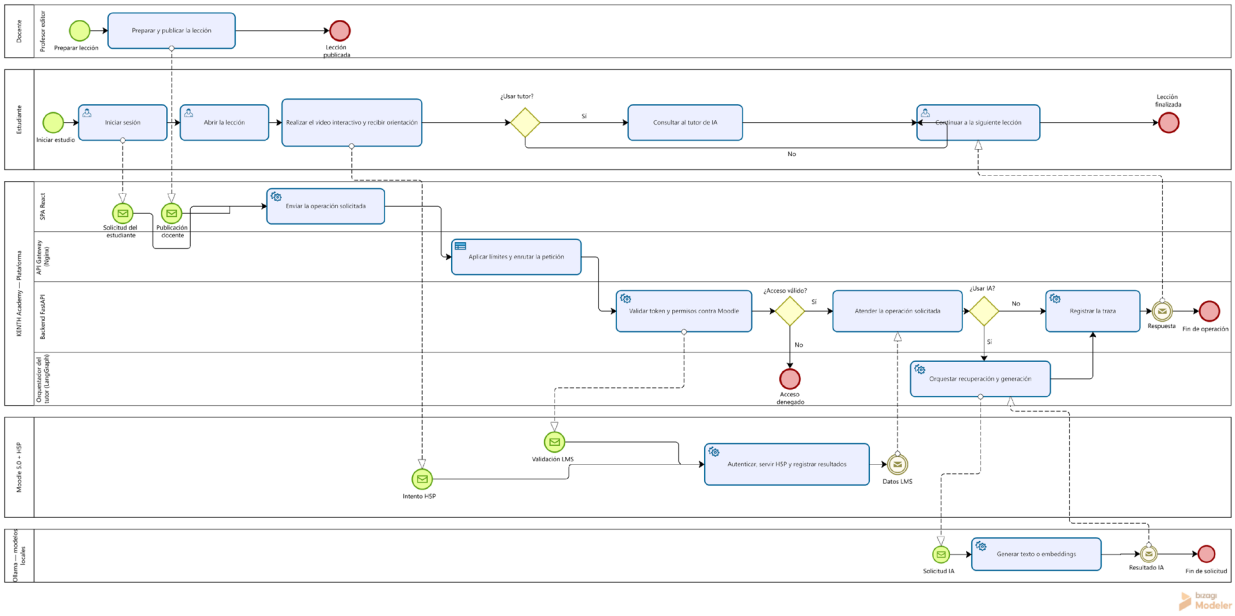
Diseño de la solución

Toda la estrategia de ingeniería se cimentó sobre un par de reglas maestras. En primer lugar, se delimitó con precisión el ámbito de cada bloque operativo y se los entrelazó mediante

una red de servicios para que cualquier módulo pudiera actualizarse o sustituirse sin necesidad de reescribir la plataforma entera. En segundo lugar, se blindó la soberanía del profesor sobre los saberes del curso: ningún documento o recurso pedagógico llega a incorporarse al cerebro del asistente virtual si antes no ha sido subido y validado de forma explícita desde el panel de control. La Figura 2 modela ese diseño como un proceso de negocio y sitúa cada actividad en el carril del responsable que la ejecuta.

Figura 2

Diagrama de procesos de la plataforma en notación BPMN



Nota. Elaboración propia con Bizagi Modeler. Los carriles corresponden al docente, al estudiante, a los servicios de la plataforma, a Moodle con H5P y a los modelos locales. Las líneas continuas representan flujo de secuencia y las discontinuas, flujo de mensaje entre participantes.

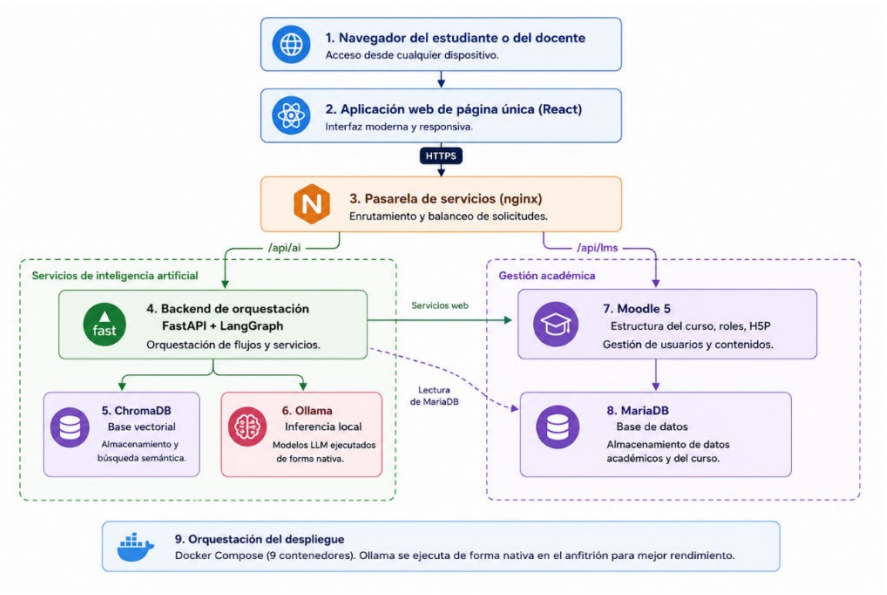
Arquitectura general en capas

La infraestructura informática se estructuró en capas bien definidas. En la capa visual interviene React —la biblioteca elegida para dar forma a la interfaz web—, encargada de

agrupar tanto las pantallas destinadas al alumnado como los paneles de gestión docente. Debajo de esta capa se encuentra Nginx, que actúa como servidor web y proxy inverso, y dirige cada solicitud entrante hacia el entorno de Moodle o hacia los microservicios de inteligencia artificial correspondientes. FastAPI expone los servicios del backend, mientras que LangGraph gestiona los estados del tutor. Moodle y MariaDB almacenan la información académica y las tablas específicas del proyecto. Por su parte, ChromaDB almacena los fragmentos autorizados empleados en la búsqueda semántica, y Ollama ejecuta el modelo dentro de la infraestructura local. Docker Compose define y coordina los contenedores necesarios para reproducir el despliegue. La topología de estas conexiones queda reflejada de manera gráfica en la Figura 3.

Figura 3

Arquitectura general de la plataforma



Nota. Elaboración propia. Los prefijos /api/ai y /api/lms corresponden a las rutas del servicio de inteligencia artificial y de Moodle, respectivamente.

La Tabla 11 detalla los componentes principales, la tecnología que los implementa y la responsabilidad que cumple cada uno dentro del sistema.

Tabla 11

Componentes de la arquitectura de la plataforma

Componente	Tecnología	Responsabilidad
Aplicación de la página única	React	Interfaz del estudiante y del docente: lecciones, videos interactivos, tutor por lección, chat general y gestión de recursos.
Pasarela de servicios	nginx	Enrutamiento de las peticiones y el punto único de entrada al sistema; límites de la tasa y los tiempos de espera.
Backend de orquestación	FastAPI y LangGraph	Construcción del contexto del tutor, flujo teacher-driven RAG, señales de aprendizaje y la coordinación de los servicios.
Gestión del aprendizaje	Moodle	Estructura del curso, matrícula, roles y actividades; contenidos H5P (mod_hvp).
Base de datos académica	MariaDB	Persistencia de la información de Moodle y de las tablas del proyecto.
Base vectorial	ChromaDB	Almacenar e indexar semánticamente el material validado por el docente.
Inferencia local	Ollama	Implementación del modelo de lenguaje en la infraestructura local.
Contenido interactivo	H5P (Interactive Video)	Videos con preguntas incrustadas que registran los eventos de interacción del estudiante.
Orquestación de despliegue	Docker Compose	Definición y configuración del entorno de producción.
Observabilidad	Loki, Promtail y Grafana	Registro estructurado de cada solicitud para garantizar la trazabilidad operativa.

Diseño instruccional del curso virtual

La implementación completa del modelo pedagógico se realizó en la Sección 0, denominada El sistema de decisión. Esta sección contiene siete lecciones, desde SEC2-R55 hasta SEC2-R61, cada una con un video H5P, recursos de apoyo y un tutor asociado. El

estudiante revisa el contenido, responde las actividades y, cuando los registros muestran una dificultad, recibe una orientación con el concepto, el minuto del video, el recurso y una práctica breve. La Figura 4 representa la secuencia.

Figura 4

Secuencia de uso de la plataforma por el estudiante



Nota. Elaboración propia.

Diseño del tutor contextual

Con el propósito de erradicar las respuestas genéricas o desconectadas de la materia en curso, el asistente virtual queda limitado al contenido de cada lección. El backend utiliza el material aprobado, los resultados H5P recientes y la configuración docente de tono y ayuda. Con esos datos, el tutor puede orientar mediante preguntas o responder de forma directa sin apartarse de la lección. El chat general atiende las consultas transversales con una configuración neutral. La organización de estas capas de contexto que alimentan al modelo se detalla gráficamente en la Figura 5.

Figura 5

Flujo de construcción del contexto del tutor por lección



Nota. Elaboración propia. El material aprobado se indexa y recupera. El contexto de la actividad y las señales de aprendizaje se incorporan al prompt, no a la base vectorial.

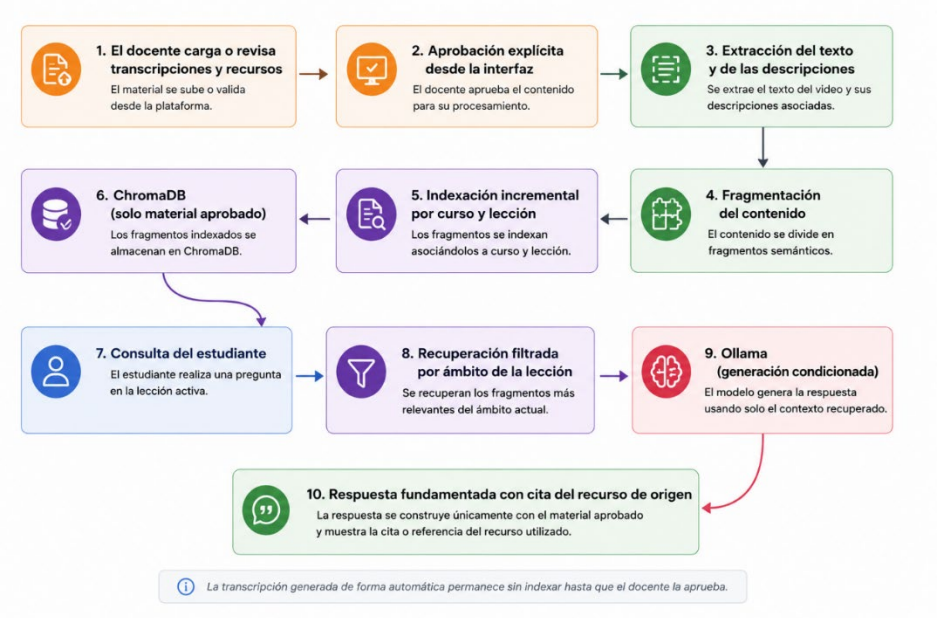
Estrategia RAG y base de conocimiento

El docente gestionó el corpus mediante la interfaz, con la carga de archivos y transcripciones, así como la validación de las fuentes que utilizaría el tutor. El backend procesó el contenido autorizado, lo segmentó y lo almacenó en ChromaDB. De esta forma, no fue necesario intervenir directamente en la base de datos ni realizar modificaciones sobre archivos técnicos.

El tutor consulta solo el material aprobado para la lección activa. Los filtros de curso, sección, lección y bloque acotan la búsqueda; los fragmentos más afines se añaden al contexto con sus recursos de respaldo. La Figura 6 muestra el recorrido desde la carga hasta la respuesta.

Figura 6

Flujo teacher-driven RAG de la plataforma



Nota. Elaboración propia. La indexación se actualiza por lección: los fragmentos anteriores de esa fuente se eliminan y se insertan de nuevo, sin reconstruir el índice completo.

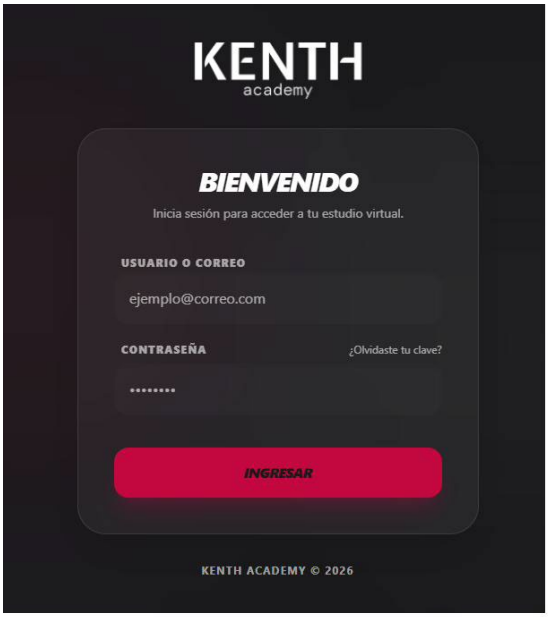
Implementación de la solución

Configuración del curso y despliegue en producción

Toda la asignatura quedó montada sobre una versión de Moodle 5 desplegada en el servidor del proyecto. Nueve componentes y los identificadores persistentes enlazan cada lección con Moodle sin que el orden modifique sus metadatos. Nueve contenedores ejecutan la interfaz, la pasarela, el backend, Moodle, sus tareas, MariaDB y la observabilidad; Ollama utiliza la GPU desde el anfitrión. Cada actualización parte de la rama principal y termina con pruebas de humo y estado. El diseño de la pantalla de bienvenida se ilustra en la Figura 7.

Figura 7

Pantalla de inicio de sesión de la plataforma



Nota. Captura del entorno de producción del proyecto.

Especificaciones técnicas del servidor y modelos de lenguaje empleados

El rendimiento del tutor depende del servidor y de los modelos disponibles, así que ambos se documentaron a partir de la instalación real. Los datos que siguen provienen de una inspección del entorno de producción y no de una configuración teórica; cuando los archivos de ejemplo del repositorio no coinciden con lo que se ejecuta, se reporta el valor observado en el servidor.

El sistema corre sobre un único servidor de propósito general con aceleración por unidad de procesamiento gráfico, condición que vuelve viable la inferencia local de modelos de lenguaje con tiempos de respuesta propios de una interacción conversacional. La Tabla 12 detalla las características de la infraestructura y las versiones de software observadas.

Tabla 12

Especificaciones técnicas del servidor de producción

Categoría	Componente	Especificación observada
Sistema	Sistema operativo	Ubuntu 26.04 LTS
	Núcleo	Linux 7.0.0-22-generic (x86_64)
Cómputo	Procesador	Intel Core Ultra 9 285
	CPU lógicas	24 (1 socket, 24 núcleos, 1 hilo por núcleo)
Memoria	RAM	60 GiB totales
	Memoria de intercambio	8 GiB
Almacenamiento	Disco raíz	1,9 TiB (LVM sobre ext4); 13 % de uso observado
Aceleración	GPU	NVIDIA GeForce RTX 5070 Ti
	Memoria de video	16 GB (16 303 MiB)
	Controlador NVIDIA	595.71.05
Plataforma	Docker / Docker Compose	29.6.1 / v5.3.0
	Ollama	0.30.10 (servicio nativo, systemd, puerto 11434)
	Python (host)	3.14.4
	Bibliotecas del backend	chromadb 1.5.9; langchain-ollama 1.1.0; faster-whisper 1.2.1
Despliegue	Contenedores en producción	9 servicios (aplicación web, pasarela, backend, Moodle y su cron, MariaDB y observabilidad)
	Puertos publicados	8090 (pasarela pública), 8091 (Moodle), 3000 (observabilidad, solo local)

Nota. Las especificaciones se comprobaron directamente. Moodle y MariaDB funcionan en contenedores, mientras que Ollama se ejecuta en el anfitrión y utiliza la aceleración disponible.

Las distintas cargas de trabajo vinculadas a la inteligencia artificial se distribuyen de manera estratégica entre varios motores locales. El chat utilizó un modelo de ocho mil millones de parámetros. Su selección respondió al equilibrio observado entre velocidad, calidad y

memoria de video disponible. En paralelo, un motor secundario se encarga de calcular las representaciones vectoriales para poblar el índice. De la configuración del perfil pedagógico se ocupa un modelo de mayor envergadura que opera de fondo, y queda abierta la posibilidad de invocar a un motor especializado en caso de requerir una auditoría complementaria. Asimismo, la interpretación de elementos gráficos recae en una red multimodal, mientras que el procesamiento de las transcripciones se encomienda a un módulo local que funciona de forma totalmente independiente de Ollama. Todas estas asignaciones operativas se encuentran desglosadas en la Tabla 13.

Tabla 13

Modelos de lenguaje empleados por el sistema y su función

Modelo	**Motor**	**Función en el sistema**	**Condición de uso**
llama3.1:8b	Ollama	Generación de las respuestas del tutor contextual en vivo (chat RAG).	Productivo, en línea
nomic-embed-text	Ollama	Generación de representaciones vectoriales para indexar y consultar la base de conocimiento.	Productivo, en línea
qwen3:14b	Ollama	Asistente de preparación del tutor: borradores de perfil pedagógico y resumen de transcripciones extensas.	Productivo, en segundo plano
qwen3-vl:4b-instruct	Ollama	Descripción y clasificación de imágenes o capturas dentro del flujo del tutor.	Productivo, multimodal
deepseek-r1:32b	Ollama	Revisión opcional de los borradores del asistente de preparación cuando se solicita calidad máxima.	Opcional
whisper:small	faster-whisper	Transcripción automática de los videos de las lecciones, sujeta a la revisión y la aprobación del docente antes de indexarse.	Productivo, en segundo plano

Nota. Los cinco primeros modelos se ejecutan mediante Ollama. La transcripción se procesa desde el backend sobre CPU con cuantización de ocho bits. La comprobación de estado revisa la disponibilidad del modelo de conversación y del modelo de representaciones vectoriales.

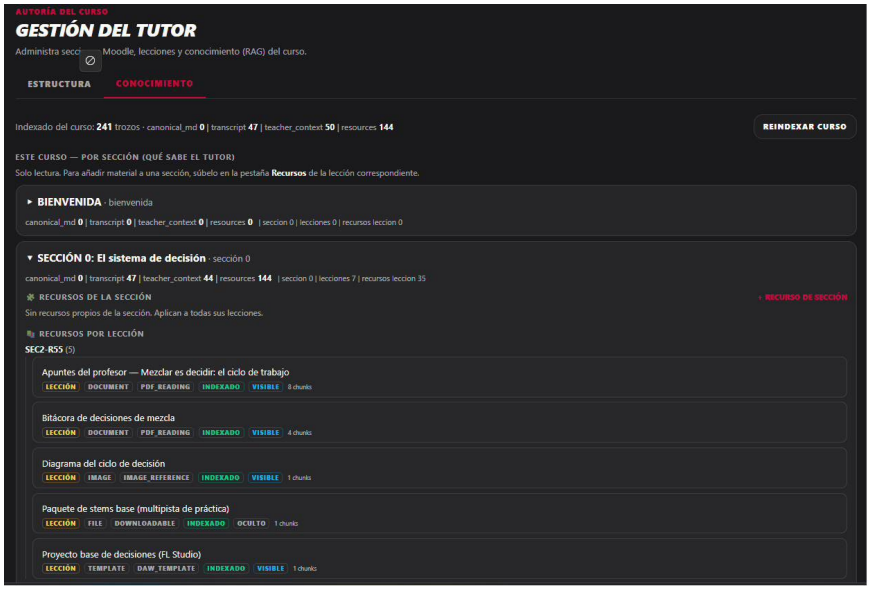
La ejecución local mantiene las consultas y el material dentro de la infraestructura del proyecto. El modelo de conversación se seleccionó de acuerdo con la memoria gráfica disponible. Por otra parte, el hecho de trabajar con un modelo de chat de escala moderada —limitado de forma directa por los gigabytes de memoria gráfica disponibles— justifica por qué la arquitectura compensa su menor naturalidad discursiva con el anclaje de las respuestas al temario autorizado por la cátedra y con la delegación de las sugerencias de minuto y material de apoyo en un itinerario determinista.

Integración del backend con el modelo local y flujo teacher-driven RAG

Para establecer la comunicación con Ollama, FastAPI opera a través de su interfaz nativa: un modelo se encarga de redactar las contestaciones, mientras otro traduce el contenido a vectores matemáticos. El flujo de control docente añade filtros previos a cualquier indexación. De este modo, una transcripción generada de manera automática permanece en estado de pausa hasta que el profesor le otorga el visto bueno; simultáneamente, el asistente redacta una propuesta de configuración que también debe ser validada, y cada recurso define por separado si tendrá visibilidad pública y si pasará a formar parte del índice vectorial. Al momento de aplicar cambios, el software actualiza únicamente la lección implicada. La Figura 8 ilustra esta consola de administración enfocada en el rol del docente.

Figura 8

Vista docente para gestionar el conocimiento del tutor



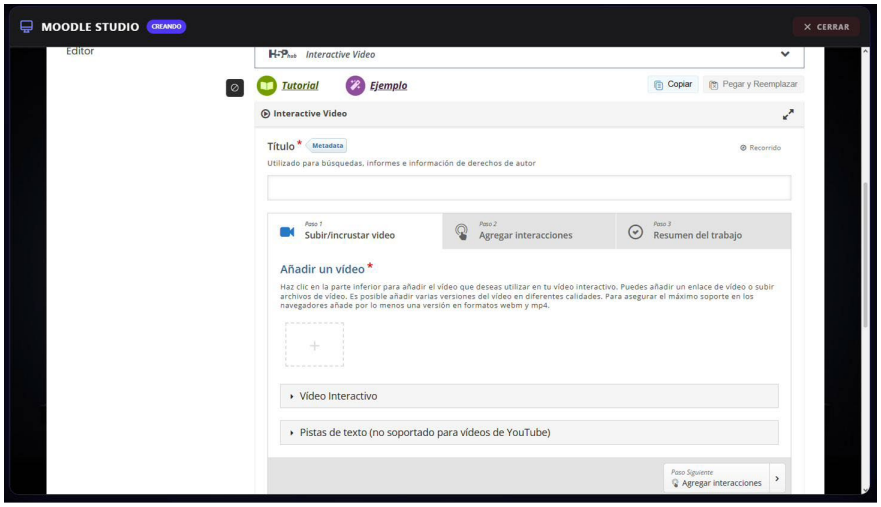
Nota. Captura del entorno de producción del proyecto.

Videos interactivos H5P y señales de aprendizaje

La totalidad del catálogo audiovisual se transformó en los módulos interactivos de tipo H5P (Interactive Video), con un total de 29 puntos de evaluación distribuidos estratégicamente a lo largo de su reproducción, desde los reactivos de opción múltiple y falso/verdadero, hasta las dinámicas de la síntesis. Las actividades H5P se añadieron sin reemplazar los recursos originales de Moodle. El manifiesto relaciona cada pregunta con el concepto evaluado y con una ruta de repaso formada por un minuto del video, un recurso y una práctica breve. También se utiliza para mostrar las actividades y procesar los resultados registrados. Mientras Moodle archiva las respuestas con el estándar xAPI y el libro de notas tradicional, el backend procesa posteriormente estas métricas para arrojar los porcentajes de acierto, los niveles de logro alcanzados y los temas con mayor índice de error. La Figura 9 muestra la pantalla de configuración asignada a una lección.

Figura 9

Configuración de un video interactivo H5P



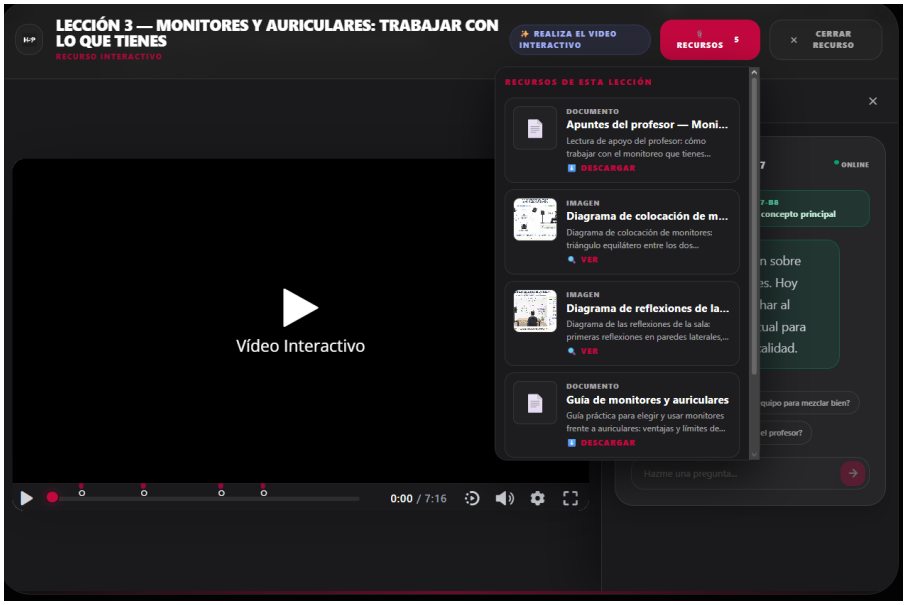
Nota. Captura del entorno de producción del proyecto.

Tutor adaptativo, interfaz del estudiante y del docente

La orientación llega al estudiante por dos vías. La guía determinística se arma sin pedir texto al modelo y puede cubrir hasta tres conceptos ordenados por prioridad; por eso el minuto y el recurso proceden siempre del manifiesto del curso. En el chat de la lección, esas mismas señales se incorporan al contexto para que la respuesta tenga en cuenta el desempeño reciente. La guía se guarda por curso y lección, persiste al recargar la página, evita repetir el sonido y caduca a los siete días. Las Figuras 10 y 11 muestran la vista del estudiante y un ejemplo de guía.

Figura 10

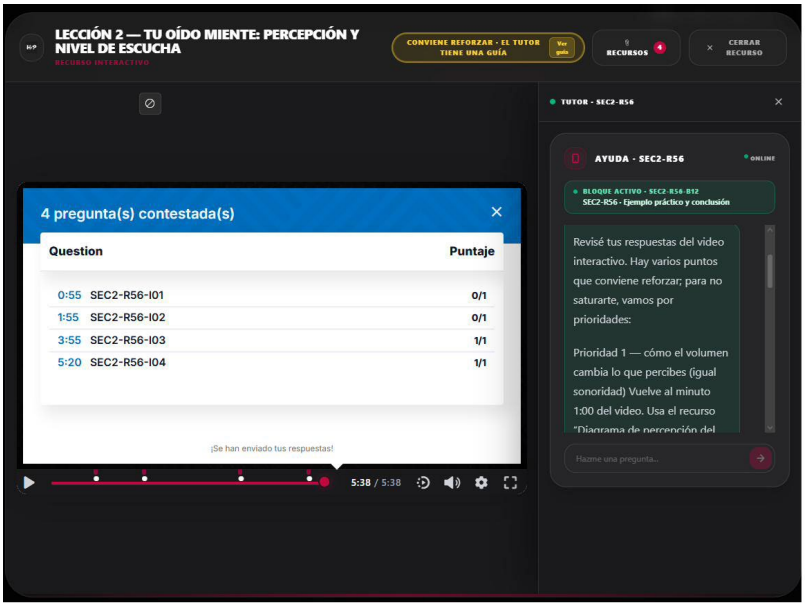
Interfaz del estudiante con video, chat del tutor y recursos



Nota. Captura del entorno de producción del proyecto.

Figura 11

Orientación adaptativa del tutor con minuto, recurso y micropráctica



Nota. Captura del entorno de producción del proyecto.

Resultados de implementación

La construcción de la Sección 0 se completó de extremo a extremo. La Tabla 14 concentra los principales indicadores de contenido e índice obtenidos al cierre del desarrollo, según los reportes de auditoría del proyecto.

Tabla 14

Indicadores de contenido e índice de la Sección 0 del curso piloto

Indicador	Valor
Lecciones implementadas (SEC2-R55 a SEC2-R61)	7
Videos interactivos H5P (Interactive Video)	7
Interacciones H5P evaluativas	29
Fragmentos en la base vectorial (curso piloto)	241
Fragmentos de texto de recursos (resource_text)	123
Fragmentos de descripción de recursos (resource_description)	21
Fragmentos canónicos activos tras la supersesión	0 (208 sustituidos)
Recursos docentes reales cargados	35
Resultados xAPI registrados en la verificación interna previa al piloto	49
Intentos de evaluación registrados en la prueba piloto (20 estudiantes)	131
Interacciones H5P respondidas en la prueba piloto	544

Nota. Los conteos proceden de ChromaDB y de los reportes de validación. Los 208 fragmentos utilizados durante el desarrollo inicial fueron desactivados y sustituidos por el material ingresado desde la interfaz docente.

El índice final contenía 241 fragmentos asociados con las siete lecciones y el contexto aprobado. Los resultados H5P se incorporan temporalmente al elaborar la petición, pero no quedan almacenados en ChromaDB. Así se mantiene la diferencia entre los datos personales y las fuentes del curso.

Validación funcional del sistema

Pruebas automatizadas de backend y contratos de frontend

Las pruebas del backend se ejecutaron en un contenedor aislado y con una base vectorial temporal, con un resultado de 263 casos aprobados, cuatro omitidos y ninguno fallido. La cobertura abarca permisos, roles de autoría, perfil pedagógico, preparación del tutor, control de dominio, ingesta segura, señales y orientación para varios conceptos. En la interfaz no aparecieron errores de análisis estático, la compilación terminó correctamente y se aprobaron los tres contratos sobre secciones de Moodle, fuentes del chat y recuperación de la guía. La Tabla 15 resume la ejecución.

Tabla 15

Resultados de las pruebas automatizadas de backend y frontend

Prueba	Resultado
Suite de pruebas del backend (pytest, contenedor aislado)	263 aprobadas, 4 omitidas, 0 fallidas
Análisis estático del frontend	0 errores (4 advertencias preexistentes)
Construcción del paquete del frontend	Correcta
Contratos de frontend (secciones, fuentes del chat, guía)	3 de 3 verificados

Validación de recuperación y de preguntas sobre recursos

La batería de recuperación de información sobre la Sección 0, formada por veintiún preguntas que abarcan las siete lecciones, respondió correctamente en los veintiún casos y recuperó fragmentos del contexto aprobado por el docente. En paralelo, un conjunto de seis preguntas de control sobre los recursos docentes cargados se respondió por completo, sin exponer material marcado como no visible para el estudiante. La Tabla 16 sintetiza estos resultados junto con las verificaciones de las interacciones H5P.

Tabla 16

Resultados de la validación funcional del sistema

Validación	Resultado
Batería de recuperación de información (Sección 0)	21 de 21
Preguntas de control sobre recursos docentes	6 de 6
Videos H5P con interacciones activas	7 de 7
Interacciones H5P evaluables	29
Auditoría de la base vectorial antes y después del cierre	241 = 241 (sin cambios)

Verificación en el entorno de producción

La comprobación de salud confirmó el backend y la conexión con Moodle, MariaDB, ChromaDB y Ollama, así como la presencia de los modelos requeridos y de los 241 fragmentos del índice. La prueba de humo aprobó once casos: acceso público, rechazo de solicitudes no autorizadas, disponibilidad de los servicios y funcionamiento de nueve contenedores.

En la primera lección, un estudiante falló dos de cinco preguntas. El sistema ordenó los conceptos pendientes y asignó a cada uno un minuto, un recurso y una práctica. La guía indicó por dónde empezar sin exponer datos internos ni sancionar, según RP02 y RP05.

Resultados de la prueba piloto con estudiantes

Los 20 estudiantes comenzaron la Sección 0 y 15 completaron la 7ma lección. 2 de ellos llegaron a la 6ta, 2 llegaron a la 5ta y 1 llegó a la 4ta. Los promedios se los calcularon únicamente con quienes cursaron cada una de las lecciones. En total se registraron 131 intentos, 544 respuestas H5P y 99 consultas en 81 conversaciones. Hubo 380 respuestas correctas, equivalentes al 70 %, mientras que el promedio de los porcentajes individuales fue del 68 %. Sin una medición inicial, una evaluación posterior equivalente y un grupo comparativo, estos resultados no demuestran una mejora causada por el sistema.

$$\text{Acierto global} = \frac{380}{544} \times 100\% = 70\%$$

donde 380 es el número de respuestas correctas y 544 el total de respuestas registradas en H5P. El promedio de las medias individuales de cada estudiante fue 68 %. Estos valores describen el desempeño durante las actividades, no una ganancia de aprendizaje.

Actividad H5P y desempeño observado

La Tabla 17 detalla la participación y el acierto en cada lección. Los veinte estudiantes cursaron las cuatro primeras; a partir de la quinta el grupo se redujo hasta los quince de la lección final. El promedio más alto fue 77 % en la 0.1 y el más bajo, 58 % en la 0.7. La Figura 12 presenta los porcentajes junto al número de participantes, dato imprescindible para leer las barras finales.

Tabla 17

Participación y desempeño en las evaluaciones interactivas por lección

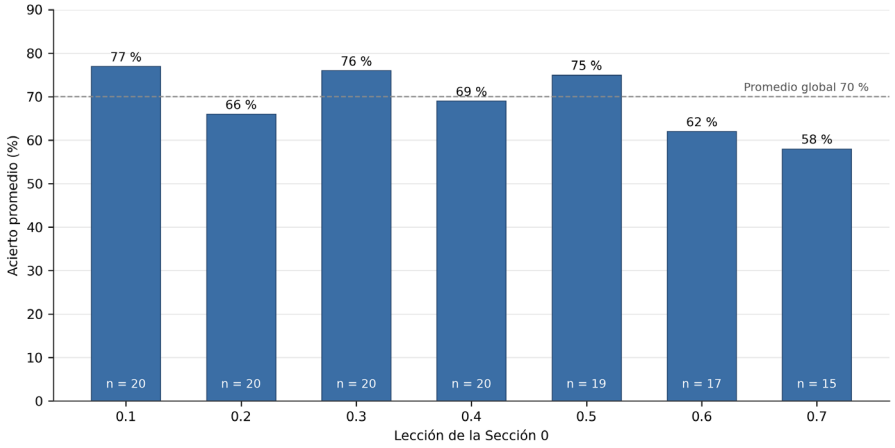
Lección	Estudiantes con intento	Interacciones por video	Acierto promedio	Conviene reforzar	En progreso	Sólido
0.1 Mezclar es decidir: el ciclo de trabajo	20	5	77 %	1	5	14
0.2 Tu oído miente: percepción y nivel de escucha	20	4	66 %	13	0	7
0.3 Monitores y auriculares: trabajar con lo que tienes	20	4	76 %	5	6	9
0.4 Anatomía universal del mixer: ruteo	20	4	69 %	8	7	5

Lección	Estudiantes con intento	Interacciones por video	Acierto promedio	Conviene reforzar	En progreso	Sólido
0.5 Gain staging: el cimiento de toda la cadena	19	4	75 %	7	4	8
0.6 Nativos frente a emulaciones analógicas	17	4	62 %	7	7	3
0.7 Checklist de sesión lista para mezclar	15	4	58 %	9	4	2

Nota. “Conviene reforzar” corresponde a menos de 60 % de acierto; “en progreso”, a valores entre 60 % y 79 %; y “desempeño esperado”, a 80 % o más. Las columnas muestran cuántos estudiantes quedaron en cada nivel. H5P registra aciertos y errores; no constituye una rúbrica analítica.

Figura 12

Acierto promedio en las evaluaciones interactivas por lección



Nota. Elaboración propia con los registros del piloto. Cada barra incluye el número de estudiantes que cursó la lección y representa el porcentaje de respuestas correctas.

Las actividades de las lecciones 0.6 y 0.7 requerían combinar varios criterios en una misma decisión. La primera requería distinguir entre plugins nativos y emulaciones, mientras que la segunda integraba lo estudiado en una lista de verificación. En la lección 0.2, trece estudiantes quedaron en el nivel de refuerzo y siete en desempeño esperado; no se registraron casos intermedios.

La Tabla 18 resume los niveles de logro de dos maneras: por cada intento de lección y por el promedio acumulado de cada estudiante.

Tabla 18

Distribución global de los niveles de logro

Nivel de logro	Intentos (n = 131)	% de intentos	Estudiantes (n = 20)	% de estudiantes
Conviene reforzar	50	38 %	7	35 %
En progreso	33	25 %	8	40 %
Sólido	48	37 %	5	25 %

Nota. La clasificación por estudiante se obtiene a partir del promedio de sus lecciones cursadas; la clasificación por intento, del puntaje de cada evaluación individual.

Los 50 intentos clasificados como «conviene reforzar» fueron los que pusieron en funcionamiento la orientación adaptativa. Para cada caso, el sistema localizó los conceptos con menor acierto y preparó una guía con el minuto, el recurso y la micropráctica asociados.

Conceptos con mayor necesidad de refuerzo

Cada interacción H5P quedó vinculada a un concepto, lo que permitió agrupar el desempeño por concepto. La Tabla 19 presenta los diez con mayor proporción de fallo entre los 26 evaluados en la Sección 0, y la Figura 13 los representa de forma gráfica.

Tabla 19

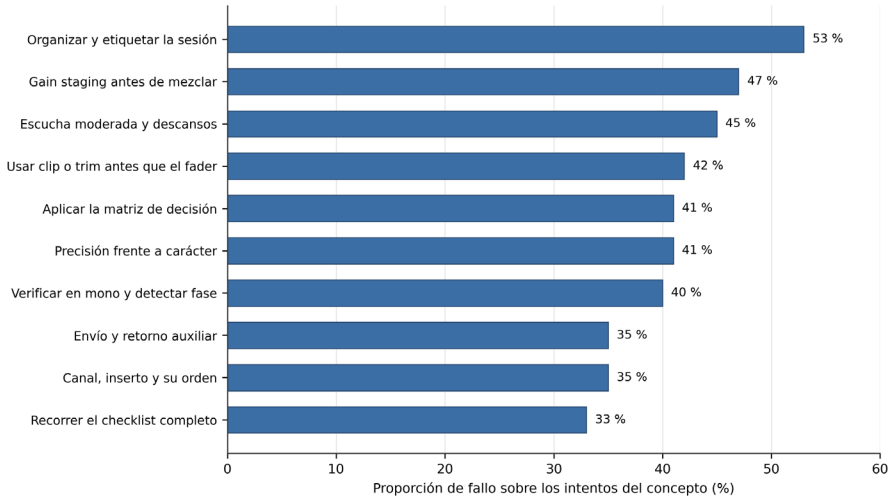
Conceptos con mayor proporción de fallo en la Sección 0

Concepto	Lección	Fallados / Respondidos	% de fallo
Organizar y etiquetar la sesión	0.7	8 / 15	53 %
Gain staging hecho antes de mezclar	0.7	7 / 15	47 %
Trabajar a nivel de escucha moderado y con descansos	0.2	9 / 20	45 %
Usar clip o trim de ganancia antes que el fader	0.5	8 / 19	42 %
Usar la matriz de decisión según el caso	0.6	14 / 34	41 %
Precisión (nativo) frente a carácter (emulación)	0.6	7 / 17	41 %
Verificar en mono y detectar problemas de fase	0.3	8 / 20	40 %
Envío y retorno auxiliar	0.4	7 / 20	35 %
Canal, inserto y su orden	0.4	7 / 20	35 %
Recorrer el checklist completo antes de empezar	0.7	10 / 30	33 %

Nota. El porcentaje se calcula con todos los intentos asociados a cada concepto. Como un concepto puede aparecer en más de una pregunta, su denominador puede superar el número de estudiantes.

Figura 13

Conceptos con mayor proporción de fallo en la Sección 0



Nota. Elaboración propia con registros del piloto. Se muestran los diez conceptos con más errores entre 26; el eje horizontal indica los fallos respecto de los intentos de cada concepto.

Los menores errores correspondieron a monitores frente a auriculares, con 10 %; 0 dBFS, clipping y headroom, con 11 %; y reconocimiento del sistema de escucha, con 15 %. El tutor prioriza hasta tres conceptos por estudiante.

Interacción con el tutor durante el piloto

Los veinte estudiantes dirigieron 99 preguntas al tutor de inteligencia artificial, repartidas en 81 conversaciones y con un promedio de cinco preguntas por estudiante. Las consultas se distribuyeron entre las siete lecciones con relativa uniformidad, desde once preguntas en la lección 0.6 hasta diecisiete en la 0.3. La Tabla 20 sintetiza los indicadores de esta interacción.

Tabla 20

Indicadores de la interacción con el tutor durante la prueba piloto

Indicador	Valor
Conversaciones registradas	81
Preguntas dirigidas al tutor	99
Mensajes totales del intercambio	198
Preguntas por estudiante (media)	5,0
Respuestas que incorporan las señales de aprendizaje del estudiante	99 de 99 (100 %)
Respuestas con anclaje clasificado como alto (clasificación automática)	99 de 99 (100 %)
Latencia media de respuesta	2,7 s
Latencia mediana de respuesta	2,5 s
Latencia máxima observada	7,5 s

Nota. El sistema calculó el anclaje frente al corpus. Entre las 99 respuestas con señales, 84 se atendieron a los conceptos, 7 a la optimización o la masterización, 4 a la orientación, 3 al diagnóstico y 1 a la decisión estética.

El tiempo medio fue de 2,7 segundos y el máximo de 7,5 segundos; estos valores contextualizan la valoración de la rapidez y las limitaciones del modelo local y de la memoria gráfica.

Calidad rubricada de las respuestas del tutor: matriz de evaluación docente

Diseño de la rúbrica y perfiles evaluadores

La batería reunió 28 preguntas, 4 por lección y distintas de las 99 consultas del piloto. En una corrida, cada criterio recibió de 0 a 2 puntos, hasta 12 por respuesta y 336 por evaluador. Audio revisó el contenido; docencia, su aplicación; e IA educativa, las fuentes. La Tabla 21 conserva el detalle. Los resultados no contienen información personal.

Resultados globales y por criterio

La Tabla 21 conserva los resultados por perfil. Los tres evaluadores superaron el umbral de 80 % y el tutor alcanzó 90,4 % del puntaje máximo.

Tabla 21

Puntaje global del tutor según cada perfil evaluador

Perfil evaluador	Criterio dominante	Puntaje	Máximo	% del máximo	Media por pregunta
Técnico (ingeniería de mezcla y masterización)	Precisión técnica	301	336	89,6 %	10,75
Pedagógico (docencia en producción musical)	Claridad didáctica	317	336	94,3 %	11,32
Evidencia (evaluación de IA educativa)	Anclaje y consistencia	293	336	87,2 %	10,46
Promedio de los tres evaluadores	—	303,7	336	90,4 %	10,85

Nota. Cada respuesta puede obtener hasta 12 puntos: seis criterios valorados de 0 a 2. Cada evaluador calificó 28 preguntas.

El evaluador pedagógico asignó el promedio más alto y el especialista en evidencia, el más bajo. La diferencia fue de siete puntos porcentuales y se concentró en la penalización aplicada cuando no se utilizó el recurso esperado, no en la corrección técnica. La Tabla 22 muestra los resultados de cada criterio y la Figura 14 los compara entre los tres perfiles evaluadores.

Tabla 22

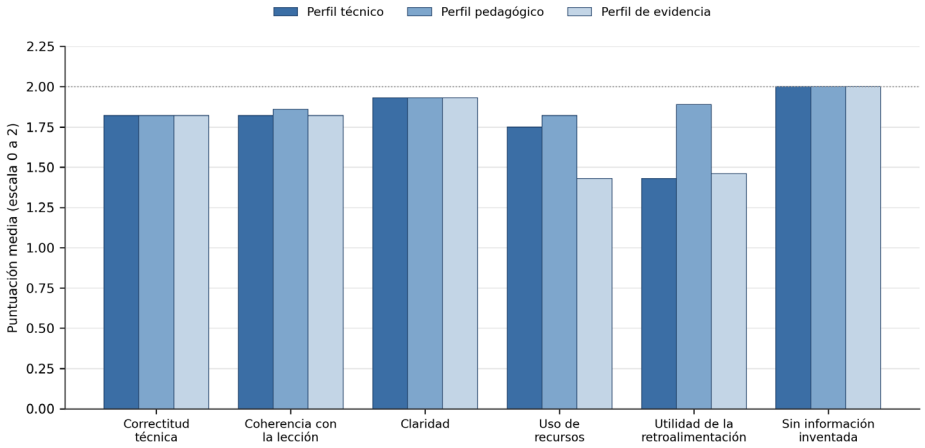
Puntuación media por criterio de la rúbrica según perfil evaluador

Criterio	Técnico	Pedagógico	Evidencia	Media
Correctitud técnica	1,82	1,82	1,82	1,82
Coherencia con la lección	1,82	1,86	1,82	1,83
Claridad	1,93	1,93	1,93	1,93
Uso de recursos	1,75	1,82	1,43	1,67
Utilidad de la retroalimentación	1,43	1,89	1,46	1,60
Sin información inventada	2,00	2,00	2,00	2,00

Nota. Cada criterio se valora de 0 a 2. Las medias se calcularon con las 28 preguntas de cada evaluador.

Figura 14

Puntuación media del tutor por criterio y perfil evaluador



Nota. Elaboración propia con la matriz de evaluación. Las barras se agrupan por perfil y la línea punteada marca el máximo de la escala. En correctitud, claridad y ausencia de información inventada, las tres barras coinciden.

El hallazgo más relevante es el acuerdo unánime en el criterio de ausencia de información inventada, que alcanzó la puntuación máxima en las 28 preguntas y en los tres perfiles: ninguna respuesta del tutor afirmó contenido fuera del corpus aprobado por el docente. Los 84 juicios sobre información inventada obtuvieron el máximo. En esta batería, las respuestas se mantuvieron vinculadas con el corpus. La claridad llegó a 1,93 y la corrección y coherencia a valores cercanos a 1,82. Utilidad obtuvo 1,60 y recursos 1,67. En utilidad, las puntuaciones por perfil fueron 1,89, 1,43 y 1,46. Algunas explicaciones aún necesitaban más detalle práctico. Las diferencias entre perfiles se observan en la Figura 14.

La diferencia entre perfiles no impidió un acuerdo global alto. El ICC individual de acuerdo absoluto fue 0,92 y su intervalo del 95 % quedó entre 0,77 y 0,97. La fiabilidad del promedio de los tres perfiles alcanzó 0,97. El intervalo es amplio porque la matriz contiene tres jueces y 28 preguntas; por tanto, el resultado debe interpretarse dentro de ese conjunto.

Puntaje por lección y casos observados

La Tabla 23 reúne los puntajes de los tres evaluadores por lección y permite identificar dónde se perdieron más puntos.

Tabla 23

Puntaje agregado de las respuestas del tutor por lección

Lección	Puntaje agregado	Máximo	% del máximo
0.1 Mezclar es decidir: el ciclo de trabajo	115	144	79,9 %
0.2 Tu oído miente: percepción y nivel de escucha	129	144	89,6 %
0.3 Monitores y auriculares: trabajar con lo que tienes	135	144	93,8 %
0.4 Anatomía universal del mixer: ruteo	144	144	100,0 %

Lección	Puntaje agregado	Máximo	% del máximo
0.5 Gain staging: el cimiento de toda la cadena	135	144	93,8 %
0.6 Nativos frente a emulaciones analógicas	112	144	77,8 %
0.7 Checklist de sesión lista para mezclar	141	144	97,9 %

Nota. Cada lección estuvo representada por cuatro preguntas. Los tres perfiles podían asignar hasta 12 puntos por respuesta, de manera que el máximo conjunto de una lección era de 144. Catorce de las 28 respuestas obtuvieron la calificación máxima de los tres evaluadores. En dos preguntas de las lecciones 0.1 y 0.6, el tutor pidió una aclaración que no hacía falta. Otras dos respuestas perdieron puntos: una dio prioridad a un fenómeno secundario y otra no recuperó todos los recursos pertinentes. La lección 0.4 obtuvo la puntuación máxima en los tres perfiles. Las puntuaciones más bajas se concentraron en una pregunta de la lección 0.1 y otra de la 0.6.

La principal incidencia fue una regla de clarificación demasiado restrictiva. Dos preguntas respondibles terminaron en una solicitud adicional porque no incluían un parámetro explícito, aunque su intención podía comprenderse. Cada caso obtuvo 10 de 36 puntos y redujo los resultados de las lecciones 0.1 y 0.6.

La segunda incidencia surgió en la lección 0.2. Ante una pregunta sobre el cambio perceptual al bajar el volumen de los monitores, el tutor resaltó la adaptación auditiva, cuando la explicación central correspondía a las curvas de igual sonoridad; la respuesta era parcialmente válida, pero priorizó un fenómeno secundario y obtuvo 26 de 36 puntos. En el tercer caso, el tutor explicó correctamente que el criterio técnico importa más que la marca del plugin, pero no recuperó el recurso con alternativas gratuitas. Esa omisión redujo el resultado a 30 de 36 puntos.

Las incidencias no presentaron información inventada ni contenido de otras lecciones. Dos se relacionaron con la regla de clarificación y la tercera con material incompleto en las lecciones 0.2 y 0.6. Los cambios pueden realizarse dentro de la arquitectura existente.

Resultados de la validación por percepción de usuarios

Una vez concluida la prueba piloto, los participantes respondieron el cuestionario correspondiente a su perfil. Los resultados se presentan por separado para cada grupo, en línea con la decisión metodológica de no agregar muestras que responden instrumentos distintos y cuyos tamaños no son comparables. La Tabla 24 describe la composición de los participantes.

Tabla 24

Participantes de la validación por percepción según perfil

Perfil	n	Caracterización declarada	Tratamiento analítico
Estudiantes	20	11 con nivel previo principiante, 8 intermedio y 1 avanzado	Estadística descriptiva por reactivo y dimensión; porcentaje de respuestas favorables
Docentes	3	Los 3 con experiencia avanzada en entornos virtuales de enseñanza	Análisis descriptivo por reactivo y dimensión, sin inferencia estadística
Administrador técnico	1	Administración de sistemas e infraestructura	Validación técnica exploratoria individual; sin expresión porcentual
Total	24	Muestreo no probabilístico por conveniencia	Análisis independiente por perfil

Nota. La caracterización corresponde a la autodeclaración de los participantes en el primer bloque de cada instrumento. No se recogieron datos identificativos.

Los tres instrumentos comparten la escala Likert de cinco puntos, pero organizan sus reactivos en dimensiones ajustadas al rol de cada perfil. La Tabla 25 muestra esa correspondencia, que permite leer los resultados de cada grupo dentro de su propio marco de referencia.

Tabla 25

Dimensiones evaluadas en cada instrumento de percepción

Instrumento	Dimensiones evaluadas	Reactivos cerrados	Preguntas abiertas
Estudiantes	Facilidad de uso; claridad del curso y recursos; videos interactivos; tutor de inteligencia artificial; retroalimentación y guía de refuerzo; satisfacción general	20	3
Docentes	Gestión del curso y lecciones; recursos y materiales docentes; videos interactivos y evaluación formativa; tutor de inteligencia artificial; retroalimentación al estudiante; viabilidad pedagógica	18	4
Administrador técnico	Administración de usuarios y roles; integración con Moodle; seguridad y privacidad; mantenimiento y operación; rendimiento y estabilidad; viabilidad institucional	18	4

Resultados del grupo de estudiantes

Los 20 estudiantes respondieron 400 reactivos cerrados. La media fue 4,33 sobre 5 (DE = 0,63) y el 91,5 % de las respuestas correspondió a “de acuerdo” o “totalmente de acuerdo”. No hubo puntuaciones de 1 o 2: se registraron 166 valores de 5, 200 de 4 y 34 de 3, de modo que las opiniones menos favorables fueron neutrales y no expresaron desacuerdo. La Tabla 26 presenta el detalle por dimensión y la Figura 15 muestra el perfil de las medias obtenidas.

Tabla 26

Resultados generales del grupo de estudiantes por dimensión (n = 20)

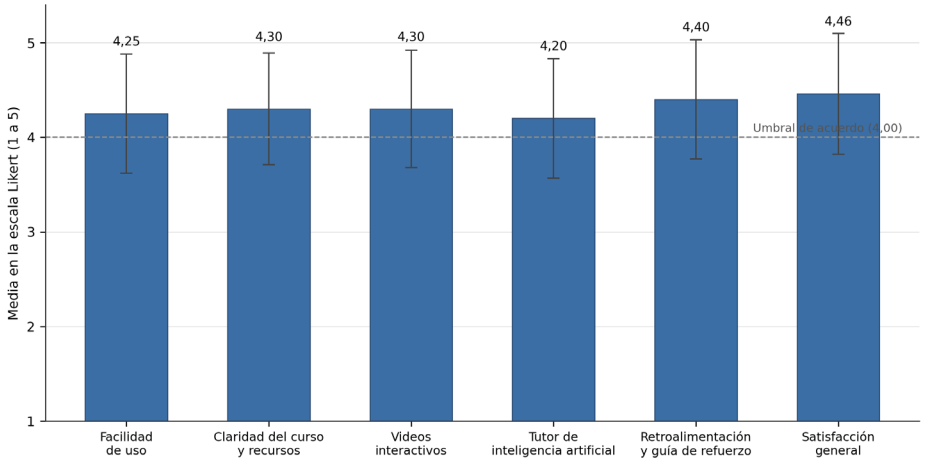
Dimensión	M	DE	Mdn	% favorable	Reactivo con menor media
Facilidad de uso	4,25	0,63	4,0	90,0 %	Funcionamiento fluido sin interrupciones (M = 4,15)

Dimensión	M	DE	Mdn	% favorable	Reactivo con menor media
Claridad del curso y recursos	4,30	0,59	4,0	93,3 %	Claridad del contenido y utilidad de los recursos (M = 4,25)
Videos interactivos	4,30	0,62	4,0	91,7 %	Mejora de la comprensión de los temas (M = 4,20)
Tutor de inteligencia artificial	4,20	0,63	4,0	88,3 %	Utilidad de las respuestas del tutor (M = 4,05)
Retroalimentación y guía de refuerzo	4,40	0,63	4,0	92,5 %	Utilidad de las recomendaciones de recursos (M = 4,30)
Satisfacción general	4,46	0,64	5,0	92,5 %	Sensación de acompañamiento (M = 4,25)
Global	4,33	0,63	4,0	91,5 %	—

Nota. M = media; DE = desviación estándar; Mdn = mediana. El porcentaje favorable reúne las respuestas 4 y 5. La escala va de 1, totalmente en desacuerdo, a 5, totalmente de acuerdo.

Figura 15

Percepción de los estudiantes por dimensión del cuestionario



Nota. Elaboración propia. El alfa de Cronbach de las respuestas estudiantiles fue 0,95. El valor indica una relación alta entre los 20 reactivos, aunque también puede estar influido por preguntas semejantes; por el tamaño de la muestra se considera exploratorio. La Tabla 26 y la

Figura 15 reúnen los resultados estudiantiles. La experiencia de aprendizaje obtuvo 4,65; la satisfacción, 4,46; y la guía, 4,40. La identificación del contenido y del minuto recibió 95 % de respuestas favorables. En el tutor, la claridad alcanzó 4,35; la facilidad, 4,25; el tiempo, 4,20; la fluidez, 4,15; y la utilidad, 4,05. La profundidad y la rapidez concentraron las observaciones menos favorables.

Resultados del grupo docente

Los tres docentes aportaron 54 respuestas: 30 tuvieron valor 5 y 24 tuvieron valor 4. La media fue de 4,56. Debido al tamaño del grupo, estos resultados describen únicamente a quienes participaron. La Tabla 27 recoge el detalle por dimensión.

Tabla 27

Resultados descriptivos del grupo docente por dimensión (n = 3)

Dimensión	M	Rango observado	Observación descriptiva
Gestión del curso y lecciones	4,50	4 – 5	La estructuración lógica del curso obtuvo la media más alta de la dimensión (M = 4,67).
Recursos y materiales docentes	4,44	4 – 5	La facilidad para proveer materiales complementarios destacó (M = 4,67).
Videos interactivos y evaluación formativa	4,56	4 – 5	El potencial didáctico de los videos y su adecuación como apoyo obtuvieron M = 4,67.
Tutor de inteligencia artificial	4,75	4 – 5	Dos reactivos (configurabilidad del tutor y refuerzo de contenidos) obtuvieron acuerdo unánime en el valor máximo (M = 5,00).
Retroalimentación al estudiante	4,44	4 – 5	La claridad y pertinencia de la retroalimentación fue el reactivo mejor valorado (M = 4,67).
Viabilidad pedagógica	4,56	4 – 5	La disposición a usar la plataforma en una asignatura real y el aporte metodológico obtuvieron M = 4,67.

Dimensión	M	Rango observado	Observación descriptiva
Global	4,56	4 – 5	Las 54 respuestas, todas en el rango de acuerdo de la escala.

Nota. M = media aritmética. Con tres informantes no se reportan desviaciones ni porcentajes, por carecer de sentido interpretativo. El rango observado indica los valores mínimo y máximo registrados en la dimensión.

La dimensión mejor valorada por los docentes fue la del tutor de inteligencia artificial (M = 4,75). La coherencia de la configuración y la función de refuerzo recibieron la máxima valoración docente. La confianza en el uso exclusivo de fuentes aprobadas alcanzó 4,33, lo que sugiere mostrar mejor la procedencia de las respuestas. Los recursos docentes y la retroalimentación obtuvieron medias de 4,44. Aunque fueron las dimensiones relativamente menores, todas las valoraciones quedaron dentro del acuerdo.

Resultado de la validación técnica del perfil administrativo

Un administrador de sistemas realizó la revisión técnica. Por tratarse de un solo participante, sus respuestas se presentan como un caso individual. De los 18 reactivos, cinco recibieron 5 y trece recibieron 4. El tercer reactivo de mantenimiento estaba formulado de manera inversa y se recodificó mediante 6 − x. Después de ese ajuste, la media fue 4,17 sobre 5. La Tabla 28 organiza la valoración por dimensión.

Tabla 28

Valoración técnica del administrador de sistemas (n = 1)

Dimensión	Puntuación por reactivo	Media	Apreciación registrada
Administración de usuarios y roles	4 · 4 · 4	4,00	Acuerdo con la claridad operativa, la gestión de roles y la delimitación de permisos entre perfiles.

Dimensión	Puntuación por reactivo	Media	Apreciación registrada
Integración con Moodle	5 · 4 · 4	4,33	Acuerdo pleno con la funcionalidad de la integración; acuerdo con la ausencia de conflictos y el control centralizado.
Seguridad y privacidad	5 · 4 · 4	4,33	Acuerdo pleno con el aporte de la ejecución local del tutor a la privacidad; acuerdo con los controles de acceso y la suficiencia de los registros.
Mantenimiento y operación	5 · 4 · 4	3,67	Acuerdo pleno con la mantenibilidad de la estructura; acuerdo con los mecanismos de respaldo. El tercer reactivo utiliza el sentido inverso: cuanto mayor es la respuesta, mayor es la necesidad de la documentación.
Rendimiento y estabilidad	4 · 4 · 4	4,00	Acuerdo con el rendimiento al procesar los videos y consultas al tutor, con la estabilidad del sistema y con la disponibilidad de los servicios.
Viabilidad institucional	4 · 5 · 5	4,67	Acuerdo pleno con la razonabilidad de los requerimientos de la infraestructura y con la conveniencia del despliegue institucional.

Nota. Escala de 1 (totalmente en desacuerdo) a 5 (totalmente de acuerdo). El reactivo inverso se transformó mediante 6 − x. Las respuestas 5 · 4 · 4 produjeron el 3,67 en la dimensión y un 4,17 en el resultado global.

La viabilidad institucional alcanzó un valor de 4,67. El administrador consideró que los recursos requeridos eran adecuados y que la implementación del sistema era favorable; sin embargo, destacó la importancia de contar con manuales de administración, respaldo, recuperación y monitoreo. Estos resultados respaldan la factibilidad del prototipo, pero también muestran las actividades operativas que todavía deben desarrollarse.

Análisis de las respuestas abiertas

Las preguntas abiertas de los tres instrumentos se analizaron mediante categorización temática. Los comentarios se analizaron por perfil y se agruparon en fortalezas, dificultades y propuestas. La Tabla 29 resume las categorías.

Tabla 29

Síntesis temática de las respuestas abiertas por perfil

Perfil	Categoría	Contenido de la categoría
Estudiantes	Fortaleza: orientación específica	Valoración de que el tutor indique el minuto exacto por revisar, el recurso concreto y la micropráctica sugerida, y de que explique qué reforzar en lugar de limitarse a calificar.
Estudiantes	Fortaleza: evaluación embebida	Las preguntas insertadas en el video se perciben como un apoyo para sostener la atención y comprobar la comprensión.
Estudiantes	Fortaleza: organización y recursos	Reconocimiento del orden lógico del curso y de la utilidad de los recursos complementarios de cada lección.
Estudiantes	Mejora: desempeño del tutor	Solicitud de reducir la latencia de la respuesta y de obtener las respuestas más breves y directas, con ejemplos mucho más concretos.
Estudiantes	Mejora: usabilidad y navegación	Propuestas de mayor visibilidad de botones y del acceso rápido a los recursos, incorporación de una barra de progreso y guardado automático de la guía del tutor.
Estudiantes	Mejora: ampliación de contenidos	Petición de más lecciones, más prácticas guiadas, ejercicios con audio real, ejemplos adicionales y material gráfico de apoyo.
Estudiantes	Dificultad: localización de elementos	Dificultades iniciales para ubicar los recursos de apoyo, la ventana del tutor y el resumen de respuestas; ausencia de dificultades relevantes en la mayoría de los casos.

Perfil	Categoría	Contenido de la categoría
Estudiantes	Dificultad: comprensión de la orientación	Confusión inicial ante el significado de la indicación de reforzar y ante la secuencia de acción esperada tras recibir la guía.
Docentes	Fortaleza: gobernanza y pertinencia	Valoración del control docente sobre el contenido que utiliza el tutor, de la organización de lecciones y recursos, y de la pertinencia de la retroalimentación por minuto y recurso.
Docentes	Fortaleza: apoyo a la evaluación formativa	Reconocimiento del aporte de los videos interactivos para evaluar la comprensión durante la clase y del rol del tutor como apoyo y no como reemplazo del profesor.
Docentes	Mejora: acompañamiento al profesorado	Necesidad de una guía rápida para los docentes nuevos y de una capacitación breve para configurar correctamente las actividades.
Docentes	Mejora: seguimiento agregado	Solicitud de un panel que permita observar el avance de varios estudiantes de una forma simultánea.
Docentes	Condición de adopción	La adopción se considera viable si se estandarizan la carga de los recursos y la revisión docente; se atribuye potencial a las asignaturas prácticas y a las virtuales.
Administrador	Fortaleza: viabilidad y escalabilidad	El sistema se considera viable y con capacidad de extensión hacia otros cursos, siempre que se estandaricen la gestión de recursos, usuarios y procesos de soporte.
Administrador	Mejora: documentación operativa	Solicitud de manuales para la administración, respaldo, recuperación y monitoreo del sistema.
Administrador	Mejora: operación y resiliencia	Recomendación de fortalecer el monitoreo, automatizar los procesos de respaldo y consolidar los registros de actividad junto con el manual operativo.

Nota. Las categorías se definieron mediante la agrupación temática de las respuestas. No se incluyen respuestas individuales ni información que permita identificar a los participantes.

Discusión de los resultados de percepción

Los participantes evaluaron de manera positiva la guía generada y las capacidades de control ofrecidas al docente. Desde la perspectiva estudiantil, se resaltó la importancia de mejorar la rapidez y la visibilidad de las orientaciones; los docentes valoraron el acompañamiento y las posibilidades de seguimiento; mientras que el administrador enfatizó la necesidad de fortalecer la documentación y la automatización. En la rúbrica de evaluación, el sistema obtuvo 1,93 puntos en claridad, 2,00 en ausencia de información inventada y 1,60 en utilidad. En general, las explicaciones fueron entendibles para los usuarios, aunque su aplicación práctica presentó algunas limitaciones.

Las percepciones recogidas deben interpretarse conjuntamente con los resultados de las pruebas aplicadas al sistema. Los tres perfiles evaluados manifestaron una valoración positiva de la guía generada mediante H5P, mientras que los docentes destacaron la capacidad de control sobre el corpus y el funcionamiento del mecanismo RAG. En conjunto, estos hallazgos fortalecen la evaluación inicial del prototipo; sin embargo, no permiten afirmar la existencia de mejoras en el aprendizaje ni generalizar los resultados obtenidos hacia otras poblaciones o contextos..

Los resultados coinciden con estudios que relacionan la utilidad del tutor con el contenido y la supervisión docente (Kasneci et al., 2023; Tobler & Köhler, 2026), y la retroalimentación con su oportunidad y capacidad de orientar acciones (Morris et al., 2021; Yin et al., 2025). En el piloto, claridad y anclaje superaron a utilidad práctica y rapidez. La percepción respaldó la arquitectura e identificó ajustes pendientes.

Limitaciones metodológicas de la validación con usuarios

Los resultados con usuarios deben leerse dentro de las restricciones de la Tabla 30. H5P, la rúbrica del tutor y los cuestionarios aportan evidencia distinta, pero todos provienen de

una muestra de conveniencia; describen, por tanto, a quienes participaron y no representan a toda la Carrera de Software ni a otras instituciones. Las encuestas se respondieron después del uso y recogen opiniones en un momento concreto, sin medir la adquisición de competencias en mezcla y masterización a lo largo del tiempo.

Tabla 30

Limitaciones metodológicas de la validación con usuarios

Limitación	**Implicación para la interpretación**
Muestreo no probabilístico por conveniencia	Los resultados no son generalizables a la población de estudiantes, docentes o personal técnico de la institución.
Tamaño reducido de los grupos docente (n = 3) y técnico (n = 1)	Únicamente permiten un análisis descriptivo; no se aplican estimaciones inferenciales ni se presentan porcentajes en el perfil técnico.
Medición transversal en un único momento	No representa una medición longitudinal del aprendizaje, ya que no se evalúan los cambios en el desempeño de los participantes a lo largo del tiempo.
Datos autorreportados	Posible presencia de sesgo por deseabilidad social y autoselección, debido a que los participantes conocían la naturaleza académica del proyecto.
Exposición limitada a la Sección 0 del curso piloto	Las percepciones se refieren a las 7 lecciones y a 1 sesión de uso, no al curso completo ni a un uso sostenido en el tiempo.
Ausencia de grupo de comparación	No es posible atribuir la valoración favorable al diseño del sistema frente a las otras modalidades de la enseñanza.

Limitación	Implicación para la interpretación
Instrumentos sin la validación psicométrica previa	Los cuestionarios fueron construidos para este estudio y no se evaluó su validez de constructo. El alfa de Cronbach del cuestionario estudiantil se reporta únicamente como la referencia exploratoria de la consistencia interna en los 20 participantes; no constituye en la validación psicométrica ni permite asumir la estabilidad en las otras muestras.
Registro de desempeño sin línea de base ni grupo de comparación	El desempeño observado en las evaluaciones interactivas refleja el nivel de logro alcanzado en las actividades formativas del curso; sin embargo, no permite determinar la ganancia de aprendizaje obtenida.
Descenso de la participación en las lecciones finales	Los promedios de las lecciones de 0.6 y de 0.7 se calculan sobre 17 y 15 estudiantes, lo que introduce un posible sesgo de supervivencia.
Matriz de precisión con los 3 evaluadores y las 28 preguntas	La concordancia interevaluador se cuantificó mediante un ICC de acuerdo absoluto de 0,92. Sin embargo, la estimación se apoya en solo los 3 jueces y las 28 preguntas, por lo que su intervalo de confianza es bastante amplio. La batería cubre las 7 lecciones, pero no agota el espacio de las posibles consultas.
Respuestas del tutor evaluadas en una única ejecución	El modelo local es parcialmente no determinista, por lo que la precisión podría variar entre las corridas del mismo conjunto de preguntas.

Nota. Los resultados constituyen la evidencia exploratoria sobre la experiencia, la utilidad y la viabilidad del prototipo.

En este piloto no se le midió la eficacia pedagógica. Lo que sí se comprobó fue el registro de las actividades y la generación de orientaciones. Los evaluadores valoraron bien las

respuestas revisadas y los participantes entendieron la propuesta. La revisión dejó, además, varios ajustes pendientes.

Áreas de mejora identificadas

Las mejoras se definieron al comparar los registros H5P, la rúbrica y los cuestionarios. La Tabla 31 relaciona cada problema con la evidencia encontrada y la acción correspondiente.

Tabla 31

Áreas de mejora priorizadas y evidencia que las sustenta

Área de mejora	Evidencia que la sustenta	Acción propuesta
Puerta de clarificación demasiado estricta	Dos de las 28 preguntas respondibles fueron derivadas a una solicitud de precisión innecesaria (matriz docente, lecciones 0.1 y 0.6)	Ajustar la condición que dispara el resguardo, con exigencia de ambigüedad efectiva del enunciado y no la mera ausencia de un parámetro explícito
Cobertura del corpus aprobado en las lecciones 0.2 y 0.6	Respuesta parcial sobre las curvas de igual sonoridad; recurso de alternativas gratuitas no recuperado	Ampliar y describir el material docente aprobado de esas lecciones y verificar la recuperación con preguntas de control específicas
Utilidad operativa de la retroalimentación	Media de 1,60 sobre 2 en la rúbrica; M = 4,05 en el reactivo de utilidad de las respuestas del tutor entre los estudiantes	Enriquecer las plantillas de respuesta con pasos concretos, valores de referencia y ejemplos verificables del propio curso
Latencia del tutor	Latencia máxima de 7,5 s en el piloto; M = 4,20 en el reactivo de tiempo de respuesta; menciones en las respuestas abiertas	Evaluar modelos locales de mayor eficiencia, respuesta en flujo y almacenamiento en caché del contexto de lección

Área de mejora	Evidencia que la sustenta	Acción propuesta
Retención en las lecciones finales	La participación desciende de 20 estudiantes en la lección 0.1 a 15 en la lección 0.7	Segmentar las lecciones de cierre, reforzar la micropráctica final y notificar el avance pendiente al estudiante
Dominio de los conceptos de integración	53 % de fallo en organizar y etiquetar la sesión, 47 % en gain staging previo y 41 % en la matriz de decisión de plugins	Reforzar la lección de cierre con práctica guiada y ejercicios de aplicación sobre sesiones reales
Documentación operativa y acompañamiento	Reactivo inverso de mantenimiento en el perfil técnico; respuestas abiertas de docentes y administrador	Elaborar manuales de administración, respaldo, recuperación y monitoreo, y una guía rápida para el profesorado
Usabilidad y navegación	Respuestas abiertas de los estudiantes; M = 4,15 en el reactivo de fluidez del funcionamiento	Aumentar la visibilidad de los controles y del acceso a recursos, incorporar indicador de progreso y clarificar la secuencia posterior a la guía

Nota. Las áreas se las ordenan según el efecto sobre el tutor y el ciclo de la orientación. Se trata de los ajustes sobre una arquitectura que ya funcionó durante el piloto.

La batería no mostró una información inventada, una mezcla de lecciones, una exposición de los recursos internos ni del contenido externo al corpus. Se observaron una aclaración innecesaria, una recuperación incompleta en 2 lecciones y un escaso detalle de lo práctico. La documentación y los ciertos ajustes de usabilidad continúan pendientes y no requieren la sustitución de la arquitectura.

Discusión de resultados

Lectura conjunta de la evidencia

Las evidencias procedieron de cuatro instrumentos. La suite de pruebas se utilizó para comprobar el software, H5P aportó los registros de las actividades, la rúbrica permitió revisar las respuestas del tutor y los cuestionarios recogieron la experiencia de los participantes. Los estudiantes asignaron 4,40 sobre 5 a las guías, mientras que los evaluadores otorgaron 1,60 sobre 2 a su utilidad. En las 99 consultas vinculadas con señales y en los 50 intentos de refuerzo, el sistema indicó un concepto, un minuto del video, un recurso y una práctica. Estos resultados coinciden con la importancia atribuida a una retroalimentación oportuna y aplicable a la tarea (Alfredo et al., 2025; Morris et al., 2021; Uzun et al., 2025; Yin et al., 2025). Sin una medición previa y un grupo de comparación, los datos no permiten calcular una ganancia de aprendizaje (Yermaganbetova et al., 2026).

Las lecciones 0.6 y 0.7 tuvieron 17 y 15 participantes. Los estudiantes que no continuaron presentaban resultados inferiores al promedio. En consecuencia, el 58 % de la última lección podría ser superior al que habría obtenido el grupo inicial completo.

Cumplimiento de requerimientos y de objetivos

La Sección 0 permitió administrar el curso, utilizar H5P, producir señales, configurar tutores por lección, controlar las fuentes, indexar recursos y generar guías. Las pruebas de estado, los contenedores, la suite automatizada y las trazas documentaron el funcionamiento. Los cuestionarios aportaron la experiencia de estudiantes y docentes.

La Tabla 32 conecta cada objetivo con su actividad, evidencia, resultado y limitación. Esa trazabilidad responde la pregunta respecto de la implementación y de la valoración inicial, no de una ganancia de aprendizaje.

Tabla 32

Trazabilidad de los objetivos específicos: actividad, evidencia, resultado y limitación

Objetivo	Actividad	Evidencia	Resultado	Limitación
1. Realizar una revisión sistemática de literatura aplicando la metodología PRISMA sobre enseñanza virtual de producción musical, IA aplicada a audio y plataformas LMS.	Búsqueda en Springer Link, IEEE Xplore y Scopus; cribado y elegibilidad por 6 razones	Matriz de revisión con 1865 registros y su razón de la exclusión; diagrama PRISMA	27 estudios incluidos; 6 ejes temáticos y 5 brechas identificadas	870 reportes no recuperados por depender de las vías de acceso abierto
2. Definir los requerimientos técnicos, pedagógicos y funcionales del curso en Moodle aplicando las mejores prácticas de la ingeniería de requisitos.	Derivación desde las brechas, el dominio y las restricciones institucionales	Matriz de requerimientos (Tablas 6, 7 y 8)	10 funcionales, 7 no funcionales y 6 pedagógicos, priorizados	Los requerimientos se validan sobre una sección del curso

Objetivo	Actividad	Evidencia	Resultado	Limitación
3. Implementar el curso virtual en Moodle y desarrollar los componentes interactivos basados en IA usando Ollama.	Arquitectura orientada a los servicios; el flujo teacher-driven RAG; el despliegue contenedorizado	263 pruebas de backend aprobadas; 3 contratos de interfaz; 21/21 en recuperación; 6/6 en recursos; la salud y el humo correctos	Sección 0 operativa de extremo a extremo; corpus canónico sustituido por el material docente	Solo la Sección 0 está poblada; la escala se sostiene en la arquitectura
4. Evaluar la efectividad del curso y la precisión del sistema de retroalimentación asistida por IA.	Prueba piloto con 20 estudiantes; rúbrica analítica aplicada por los 3 evaluadores a las 28 respuestas; cuestionarios de percepción	131 intentos y 544 interacciones H5P; matriz de precisión con ICC de acuerdo absoluto = 0,92, IC 95 % [0,77, 0,97]; encuestas de los estudiantes, de los docentes y del administrador	Efectividad exploratoria: 70 % de acierto global, 15 de 20 estudiantes completaron las 7 lecciones, percepción estudiantil de 4,33/5 y 91,5 % favorable. Puntaje del tutor: 90,4 % del puntaje máximo de la rúbrica y ausencia unánime de información inventada	Los indicadores describen desempeño formativo y percepción; no miden ganancia de aprendizaje ni admiten la inferencia causal

Objetivo	Actividad	Evidencia	Resultado	Limitación
5. Elaborar conclusiones, recomendaciones y trabajos futuros.	Análisis integral de los resultados técnicos, los formativos y los perceptuales obtenidos en los objetivos anteriores	Resultados consolidados del proyecto y del Capítulo V	Conclusiones alineadas con los objetivos, las recomendaciones priorizadas y las líneas de trabajo futuro	Muestreo por conveniencia; instrumentos sin la validación psicométrica previa

Nota. Elaboración propia. La formulación del objetivo 4 corresponde a la aprobada en el plan de titulación.

Para evaluar el objetivo 4 se utilizaron los registros H5P, la encuesta y la batería del tutor. Las actividades mostraron un 70 % de aciertos y el 37 % de los intentos quedó en el nivel de desempeño esperado. Las dificultades más frecuentes aparecieron en la organización de la sesión y el gain staging. La encuesta estudiantil alcanzó una media de 4,33 sobre 5 y un 91,5 % de respuestas favorables. En la batería de 28 preguntas, los tres evaluadores otorgaron al tutor el 90,4 % de los puntos disponibles. El ICC alcanzó 0,92 y no se identificó información inventada. La revisión también mostró que deben completarse ciertos contenidos, ajustarse la regla de clarificación y ampliarse el detalle de algunas respuestas. Estos datos corresponden al desempeño observado en el piloto y no a una ganancia frente a una línea base.

Limitaciones del sistema y deuda técnica

Actualmente, el recorrido completo solo está disponible en la Sección 0. Las demás secciones requieren incorporar sus materiales. El modelo local puede variar entre ejecuciones y tiene menor capacidad que ciertos servicios comerciales. El anclaje, las reglas y las verificaciones reducen la variación, pero no producen respuestas idénticas. La utilidad y la latencia fueron los aspectos relativamente menos valorados. También se observó una regla de

clarificación restrictiva y recuperación incompleta en las lecciones 0.2 y 0.6. Las dos incidencias pueden resolverse sin reemplazar la arquitectura y no incluyeron información inventada.

El sistema no asigna desempeño si no existe un intento H5P. La percepción se obtuvo de una muestra por conveniencia y no mide aprendizaje. Antes del uso institucional deben completarse los manuales, respaldos y monitoreo y revisarse los campos heredados. Son tareas operativas que no modifican los resultados ya registrados.

Capítulo V

Conclusiones

Este capítulo recoge las conclusiones asociadas a cada objetivo específico y a la pregunta de investigación, las recomendaciones para una aplicación posterior de la plataforma y las líneas de trabajo futuro.

El rastreo bibliográfico partió de 1865 registros procedentes de Springer Link, IEEE Xplore y Scopus. Después del cribado quedaron 27 estudios. Al organizarlos se obtuvieron seis ejes temáticos y cinco brechas, entre ellas la poca integración entre LMS, actividades interactivas, analítica y tutores que trabajen con fuentes controladas por el profesor. Este resultado cubrió el primer objetivo y sirvió para delimitar la propuesta.

Las brechas anteriores se tradujeron en 10 requerimientos funcionales, siete no funcionales y seis pedagógicos. A partir de ellos se preparó en Moodle la Sección 0, compuesta por siete lecciones de mezcla y masterización. Los videos H5P, MariaDB, ChromaDB y Ollama quedaron conectados dentro del mismo recorrido. El profesor puede elegir desde la interfaz los materiales que se indexan y que luego consulta el tutor.

El tercer objetivo se cumplió con la implementación del curso virtual en Moodle y con el desarrollo de los componentes interactivos apoyados en inteligencia artificial. La Sección 0 quedó desplegada en producción con siete lecciones, siete videos interactivos H5P, 29 interacciones evaluadas, un tutor por lección y un chat general; los modelos de lenguaje se ejecutaron con Ollama dentro del servidor del proyecto y el docente pudo sustituir el corpus inicial por sus propios recursos sin modificar archivos técnicos.

La evidencia técnica respaldó ese resultado: el backend aprobó 263 pruebas, los tres contratos de interfaz se verificaron sin errores, la batería de recuperación respondió 21 consultas de 21 y las preguntas sobre recursos alcanzaron 6 de 6, con la comprobación de salud y la prueba de humo correctas al cierre. El alcance completado de extremo a extremo se limitó a la Sección 0.

Durante el piloto participaron 20 estudiantes, de los cuales 15 lograron completar la última lección. En H5P se registraron 131 intentos y 544 respuestas, con un porcentaje de aciertos del 70 %. La percepción estudiantil alcanzó una media de 4,33 sobre 5, con un 91,5 % de valoraciones favorables. Asimismo, los tres especialistas evaluaron el desempeño del tutor con 911 puntos de un total de 1008 posibles, equivalente al 90,4 %. La evaluación de concordancia obtuvo un ICC de 0,92, con un intervalo de confianza del 95 % entre 0,77 y 0,97. Además, los evaluadores coincidieron en que las respuestas revisadas no incluyeron información inventada, lo que evidencia consistencia en la generación de orientaciones.

Con estos resultados se respondió la pregunta de investigación en cuanto al diseño, la integración y el funcionamiento inicial del curso. Moodle pudo recibir las señales de H5P, consultar material aprobado por el profesor y entregar una orientación relacionada con el desempeño registrado. Sin embargo, el 70 % de aciertos y las valoraciones de los participantes no deben interpretarse como una mejora causada por el tutor. La prueba se limitó a una sección, utilizó una muestra por disponibilidad y no contó con medición inicial ni grupo comparativo.

Recomendaciones

Una evaluación posterior necesitará una muestra más amplia. Deberá incluir un paralelo completo, docentes con experiencias diferentes y personal técnico. Antes de volver a aplicar los cuestionarios se requiere una revisión de expertos y un análisis de fiabilidad. La regla de clarificación deberá actuar solamente cuando la pregunta sea realmente ambigua. También falta completar las lecciones 0.2 y 0.6 con las curvas de igual sonoridad y el recurso sobre alternativas gratuitas. Una vez incorporados esos materiales, su recuperación deberá comprobarse mediante preguntas específicas.

Las demás secciones pueden incorporarse mediante la arquitectura ya implementada. Antes será necesario completar y reorganizar sus materiales. En la última lección faltan los ejercicios de organización, de etiquetado, de estructura de ganancia y de complementos,

además de una forma más clara de mostrar el progreso. La ampliación también requiere de manuales, de respaldos automáticos, de registros centralizados, de una guía docente y de una vista individual de cada uno de los estudiantes. Una vez resueltos esos puntos se podrá probar si un modelo local de una mayor capacidad o el uso de los ejemplos guiados mejora las respuestas sin sacar los datos del servidor.

Trabajos futuros

Como línea de trabajo futuro, se propone evaluar el sistema mediante mediciones realizadas antes y después de su implementación y, siempre que sea posible, considerar la inclusión de un grupo de comparación. Además, se plantea repetir la batería de pruebas del tutor en distintas ejecuciones, aumentar el número de preguntas y de evaluadores, extender el flujo de trabajo a otros cursos, perfeccionar el proceso de recuperación, incorporar un seguimiento longitudinal y analizar la capacidad del servidor cuando el tutor sea utilizado simultáneamente por varios cursos.

Referencias

Las referencias se presentan conforme a las normas APA, séptima edición, de acuerdo con las directrices institucionales aplicables.

ADL Initiative. (2025). *Experience API (xAPI) specification*. Advanced Distributed Learning Initiative. https://adlnet.gov/projects/xapi/

Ahangama, N. (2026). Designing assessments in the generative AI era: A tailored assessment framework for ICT tertiary education. *International Journal of Educational Technology in Higher Education, 23*, Article 20. https://doi.org/10.1186/s41239-026-00582-0

Akinpelu, G. A., Salman, M. F., Akinpelu, S. A., y Ameen, K. S. (2024). Effects of mastery learning strategy on senior school students' performance in mathematics in Osogbo, Nigeria. *Discover Education, 3(1)*, Article 197. https://doi.org/10.1007/s44217-024-00285-6

Alfredo, R., Milesi, M., Echeverria, V., Gašević, D., Buckingham Shum, S., Zhao, L., Yan, L., Jin, Y., Fan, J. X., Pammer-Schindler, V., Swiecki, Z., & Martinez-Maldonado, R. (2025). Co-designing AI-powered learning analytics: Bringing students and teachers together. *International Journal of Educational Technology in Higher Education, 22*, Article 20. https://doi.org/10.1186/s41239-025-00572-8

Braun, V., y Clarke, V. (2024). Supporting best practice in reflexive thematic analysis reporting in palliative medicine. *Palliative Medicine, 38*(6), 608–616. https://doi.org/10.1177/02692163241234800

Chang, C.-C., Shih, S.-C., & Kuo, B.-C. (2026). A dual-agent intelligent tutoring system for teaching two-step fraction word problems. *IEEE Transactions on Learning Technologies, 19*. https://doi.org/10.1109/TLT.2026.3708780

Chroma. (2025). *Chroma: The open-source embedding database* [Software]. https://www.trychroma.com/

Es, S., James, J., Espinosa-Anke, L., y Schockaert, S. (2024). RAGAS: Automated evaluation of retrieval augmented generation. En *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations* (pp. 150–158). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.eacl-demo.16

Ferrando, P. J., Morales-Vives, F., Casas, J. M., y Muñiz, J. (2025). Escalas Likert: una guía práctica para su diseño, construcción y uso. *Psicothema, 37*(4), 1–15. https://doi.org/10.70478/psicothema.2025.37.24

Fu, Z. (2025). Integrating reinforcement learning with dynamic knowledge tracing for personalized learning path optimization. *Scientific Reports, 15*, Article 23900. https://doi.org/10.1038/s41598-025-23900-4

Gaeta, A., Orciuoli, F., Pascuzzo, A., & Peduto, A. (2025). Enhancing traditional ITS architectures with large language models for generating motivational feedback. *Computers and Education: Artificial Intelligence, 9*, Article 100433. https://doi.org/10.1016/j.caeai.2025.100433

Gamage, S. H. P. W., Ayres, J. R., & Behrend, M. B. (2022). A systematic review on trends in using Moodle for teaching and learning. *International Journal of STEM Education, 9*, Article 9. https://doi.org/10.1186/s40594-021-00323-x

H5P Group. (2025). *H5P: Create, share and reuse interactive HTML5 content* [Software]. https://h5p.org/

Hilliger, I., Pérez-Sanagustín, M., González, C., Villalobos, E., & Celis, S. (2026). A GAI-based chatbot for scaffolding self-regulated learning: Insights from a design-based research approach. *Journal of Learning Analytics, 13*. https://doi.org/10.18608/jla.2026.9143

International Organization for Standardization. (2023). *ISO/IEC 25010:2023 Systems and software engineering—Systems and software Quality Requirements and Evaluation (SQuaRE)—Product quality model*. https://www.iso.org/standard/78176.html

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Koo, T. K., y Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Koszewski, D., Görne, T., Korvel, G., & Kostek, B. (2023). Automatic music signal mixing system based on one-dimensional Wave-U-Net autoencoders. *EURASIP Journal on Audio, Speech, and Music Processing, 2023*, Article 1. https://doi.org/10.1186/s13636-022-00266-3

Ladrón-de-Guevara Moreno, L., Mena-Bernal Rosales, I., & Nieto-Gómez, M. D. (2024). Sistema de gestión del aprendizaje en Educación Superior: Uso de herramientas de Moodle y experiencia docente. *Escuela Abierta, 27*, 109–121. https://revistascientificas.uspceu.com/EA/article/view/3442

Lee, D.-K., & Joe, I. (2025). A GPT-based code review system with accurate feedback for programming education. *IEEE Access, 13*. https://doi.org/10.1109/ACCESS.2025.3581139

Lee, Y. O., Bang, B., Lee, J., & Oh, S. (2026). Personalized auto-grading and feedback system for constructive geometry tasks using large language models on an online math platform. *IEEE Access, 14*. https://doi.org/10.1109/ACCESS.2026.3657726

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Advances in Neural Information Processing Systems*

(Vol. 33, pp. 9459–9474). Curran Associates. https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html

Ley Orgánica de Protección de Datos Personales. (2021, 26 de mayo). *Registro Oficial Suplemento No. 459 (Quinto Suplemento)*. Asamblea Nacional del Ecuador. Reglamento General expedido mediante Decreto Ejecutivo No. 904 (2023).

Liu, J., Keane, T., Sun, D., Yang, Y., & Yang, Y. (2025). Design and evaluation of ChatGPT-MWPS: An AI-enhanced learning system for improving primary students' mathematical word problem solving. *Smart Learning Environments, 12*, Article 19. https://doi.org/10.1186/s40561-025-00419-9

Liu, S., Guo, X., Hu, X., & Zhao, X. (2024). Advancing generative intelligent tutoring systems with GPT-4: Design, evaluation, and a modular framework for future learning platforms. *Electronics, 13*(24), Article 4876. https://doi.org/10.3390/electronics13244876

Marín, V. I., & Tur, G. (2023). La privacidad de los datos en Tecnología Educativa: Resultados de una revisión de alcance. *Edutec. Revista Electrónica de Tecnología Educativa, 83*, 7–23. https://doi.org/10.21556/edutec.2023.83.2701

Martínez-Araneda, C., Gutiérrez Valenzuela, M., Macías Aranda, J. L., Maldonado Montiel, D., Gómez Meneses, P., & Segura-Navarrete, A. (2025). Bridging code and timely feedback: Integrating generative AI into a programming platform. *PeerJ Computer Science, 11*, Article e3070. https://doi.org/10.7717/peerj-cs.3070

Martínez-Ramírez, M. A., Liao, W.-H., Fabbro, G., Uhlich, S., Nagashima, C., & Mitsufuji, Y. (2022). Automatic music mixing with deep learning and out-of-domain data. In *Proceedings of the 23rd International Society for Music Information Retrieval Conference* (pp. 411–418). ISMIR. https://archives.ismir.net/ismir2022/paper/000049.pdf

Maurya, K. K., Srivatsa, K. A., Petukhova, K., y Kochmar, E. (2025). Unifying AI tutor evaluation: An evaluation taxonomy for pedagogical ability assessment of LLM-powered AI tutors.

En *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.naacl-long.57

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods, 23*(3), 412–433. https://doi.org/10.1037/met0000144

Moodle. (2025). *Moodle documentation: Web services*. Moodle Pty Ltd. https://docs.moodle.org/en/Web_services

Morris, R., Perry, T., & Wardle, L. (2021). Formative assessment and feedback for learning in higher education: A systematic review. *Review of Education, 9*(3), Article e3292. https://doi.org/10.1002/rev3.3292

Nedrehagen, E., Vee, T. S., Stokstad, J. M., & Orm, S. (2025). A scoping review of dialogic formative feedback practices in higher education. *Frontiers in Education, 10*, Article 1696703. https://doi.org/10.3389/feduc.2025.1696703

Ollama. (2025). *Ollama: Get up and running with large language models locally* [Software]. https://ollama.com/

Organisation for Economic Co-operation and Development. (2015). *Frascati manual 2015: Guidelines for collecting and reporting data on research and experimental development*. OECD Publishing. https://doi.org/10.1787/9789264239012-en

Paavilainen, T., López-Pernas, S., Väisänen, S., Kontkanen, S., & Hirsto, L. (2026). Implementing learning design with learning analytics to scaffold the regulation of collaboration in primary education. *Smart Learning Environments, 13*, Article 3. https://doi.org/10.1186/s40561-026-00453-1

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E.,

McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, Article n71. https://doi.org/10.1136/bmj.n71

Pan, Z., Biegley, L., Taylor, A., y Zheng, H. (2024). A systematic review of learning analytics: Incorporated instructional interventions on learning management systems. *Journal of Learning Analytics, 11*(2), 52–72. https://doi.org/10.18608/jla.2023.8093

Papachristou, E., Troussas, C., Krouska, A., & Sgouropoulou, C. (2026). A fuzzy logic-driven system for interpretable and behavior-aware student assessment: E-Teacher Assistant case study. *Electronics, 15*(12), Article 2671. https://doi.org/10.3390/electronics15122671

Pressman, R. S., y Maxim, B. R. (2020). Software engineering: A practitioner's approach (9.ª ed.). McGraw-Hill Education.

Reddig, J. M., Arora, A., & MacLellan, C. J. (2025). Generating in-context, personalized feedback for intelligent tutors with large language models. *International Journal of Artificial Intelligence in Education, 35*, 3459–3500. https://doi.org/10.1007/s40593-025-00505-6

Saez, Y., Garcia, L. M., Mochon, A., & Isasi, P. (2026). Evaluating large language models for AI-assisted grading: A framework and case study in higher education. *Scientific Reports, 16*, Article 48656. https://doi.org/10.1038/s41598-026-48656-3

Sappa, A. (2025). Assessing the impact of large language models on the scalability and efficiency of automated feedback mechanisms in massive open online courses. *Indian Journal of Information Sources and Services, 15*(2), 35. https://doi.org/10.51983/ijiss-2025.IJISS.15.2.35

Scheu, S., Loeffler, S. N., & Maedche, A. (2026). Designing a mobile chatbot-based learning journaling system for intrinsic motivation and engagement. *International Journal of Educational Technology in Higher Education, 23*, Article 7. https://doi.org/10.1186/s41239-026-00589-7

Semerikov, S., Nechypurenko, P., Vakaliuk, T., Mintii, I., & Fadieieva, L. (2026). Development and validation of an adaptive learning readiness assessment framework for Moodle courses. *Journal of New Approaches in Educational Research, 15*, Article 69. https://doi.org/10.1007/s44322-026-00069-w

Sommerville, I. (2016). Software engineering (10.ª ed.). Pearson Education.

Steinert, S., Avila, K. E., Ruzika, S., Kuhn, J., & Küchemann, S. (2024). Harnessing large language models to develop research-based learning assistants for formative feedback. *Smart Learning Environments, 11*, Article 54. https://doi.org/10.1186/s40561-024-00354-1

Suen, H.-Y., & Hung, K.-E. (2026). An interpretable closed-loop intelligent tutoring system for multimodal affective feedback in asynchronous presentation training. *IEEE Transactions on Learning Technologies, 19*. https://doi.org/10.1109/TLT.2026.3693864

Tobler, S., & Köhler, K. (2026). A lecture-specific AI-based tutor for higher education: Pedagogical design and empirical evaluation. *Education Sciences, 16*(5), Article 812. https://doi.org/10.3390/educsci16050812

UNESCO. (2023). *Guidance for generative AI in education and research*. https://doi.org/10.54675/EWZM9535

Uzun, Y., Suraworachet, W., Zhou, Q., Gauthier, A., & Cukurova, M. (2025). Engagement with analytics feedback and its relationship to self-regulated learning competence and course performance. *International Journal of Educational Technology in Higher Education, 22*, Article 15. https://doi.org/10.1186/s41239-025-00515-3

Valente, A., Junior, W. M. V., & César Da Rocha Seruffo, M. (2026). An AI-driven conversational platform for the learning and assessment of programming logic. In *Proceedings of the ACM International Conference on Interactive Media Experiences (IMX 2026)*. Association for Computing Machinery. https://doi.org/10.1145/3788851.3815047

Villegas-Ch, W., Gutiérrez, R., García-Ortiz, J., & Guevara, V. (2025). Explainable educational assistant integrated in Moodle: Automated semantic assessment and adaptive tutoring based on NLP and XAI. *Discover Artificial Intelligence, 5*, Article 438. https://doi.org/10.1007/s44163-025-00438-y

Winget, M., y Persky, A. M. (2022). A practical review of mastery learning. *American Journal of Pharmaceutical Education, 86*(10), Article 8906. https://doi.org/10.5688/ajpe8906

Wu, F., Dang, Y., & Li, M. (2025). A systematic review of responses, attitudes, and utilization behaviors on generative AI for teaching and learning in higher education. *Behavioral Sciences, 15*(4), Article 467. https://doi.org/10.3390/bs15040467

Yermaganbetova, M., Ashimbekova, A., Kaibassova, D., Akhitova, R., & Dyussembina, E. (2026). Evaluating the impact of an AI-integrated learning platform on student performance: A quasi-experimental study. *Frontiers in Education, 11*, Article 1792353. https://doi.org/10.3389/feduc.2026.1792353

Yildirim-Erbasli, S. N., & Bulut, O. (2023). Conversation-based assessment: A novel approach to boosting test-taking effort in digital formative assessment. *Computers and Education: Artificial Intelligence, 4*, Article 100135. https://doi.org/10.1016/j.caeai.2023.100135

Yin, J., Xu, H., Pan, Y., & Hu, Y. (2025). Effects of different AI-driven chatbot feedback on learning outcomes and brain activity. *npj Science of Learning, 10*, Article 311. https://doi.org/10.1038/s41539-025-00311-8

Zhou, X., Zhang, Z., Xie, X., & Zhang, J. (2025). Deep learning based knowledge tracing in intelligent tutoring systems. *Scientific Reports, 15*, Article 7422. https://doi.org/10.1038/s41598-025-07422-7