

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INFORMÁTICA
SANTIAGO - CHILE



**“DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA
DE IA GENERATIVA PARA LA CREACIÓN DE
EVALUACIONES ACADÉMICAS EN EDUCACIÓN
SUPERIOR: ENFOQUE EN INGENIERÍA CIVIL
INFORMÁTICA UTFSM”**

CLAUDIO IGNACIO AROS AGUIRRE

**MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL EN INFORMÁTICA**

Profesora Guía: Dayana Carrillo
Profesora Correferente: Liubov Dombrovskaja

Julio - 2025



CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): ☒ Memoria o trabajo de título; ☐ Tesis de Postgrado;

Título del trabajo: DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS EN EDUCACIÓN SUPERIOR: ENFOQUE EN INGENIERÍA CIVIL INFORMÁTICA UTFSM

Nombre del candidato(a): Claudio Ignacio Aros Aguirre

Carrera / Grado: Ingeniería Civil en informática

Campus: Santiago San Joaquín ; **Departamento:** Informática

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Dayana Carillo Daboin, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL

☒ El trabajo **NO contiene información que amerite confidencialidad** y puede ser publicado de inmediato en repositorio con acceso abierto.

☐ El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (embargo) por:

☐ 6 meses; ☐ 12 meses; ☐ 2 años; ☐ 3 años; ☐ 5 años; ☐ 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha: 01/08/2025

; Firma:

Dayana Carillo

Estudiante o Candidato(a):

Fecha: 01/08/2025

; Firma:

Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

DEDICATORIA

A todos mis compañeros que siguen persiguiendo el sueño de su título, sepan que sus esfuerzos siempre tienen frutos. ¡Si yo pude ustedes pueden!

A mi Padre, que me enseñó a no rendirme y con su esfuerzo hizo posible mi camino.

A mi madre, quien ha sido mi pilar cada día, en esta etapa y en toda mi vida.

A mi Hermano, que ha sido mi compañero de vida y espero lo siga siendo.

A mi familia, que siempre me apoya, me recuerda lo orgullosos que están de mí y me impulsa a esforzarme cada día por ser un buen ejemplo para mis primos.

Y a mí, que caí mil veces, pero me levanté mil y una.

AGRADECIMIENTOS

A lo largo de este largo proceso, lo que más he aprendido es que ningún proyecto se realiza en soledad. Desde las actividades grupales que me tocó hacer con personas que ni conocía, hasta este proyecto de Feria de Software.

Gracias a todos quienes me ayudaron a lo largo de mi carrera; al excepcional grupo de Feria del que fui parte —Giuliana, Simón, Orlando, Natalia y Esteban—; a mi profesora guía, Dayana Carrillo, quien fue un pilar en el desarrollo de esta tesis; a los profesores que se dieron el tiempo de participar y aportar en la validación de la plataforma; y a mi familia, que me apoyó a seguir adelante. En especial a mi madre, que siempre estuvo cuando la necesité, ya fuera con un desayuno, una acercada al metro o simplemente un abrazo.

Gracias especiales a mi pilar social, mi compañero de universidad Cristóbal Martínez. A pesar de haber hecho muchos amigos en este camino —algunos que se fueron, otros que iban y volvían—, tú siempre estuviste ahí: para hacer una tarea juntos, estudiar o simplemente pasar el rato. Gracias por estar siempre y por ser mi pana.

Y por último, pero no menos importante, gracias a mi novia Ari, quien me impulsó a seguir adelante cuando no me sentía capaz de lograr todo lo que he conseguido. Gracias por sacarme una sonrisa cuando más lo necesitaba y por hacer que mi día a día sea más feliz.

RESUMEN

Este trabajo presenta el diseño, desarrollo y validación de una plataforma basada en inteligencia artificial generativa (IAg) para automatizar la creación de evaluaciones académicas. La solución, orientada a docentes de Ingeniería Civil Informática de la UTFSM, utiliza la API de GPT-4 para generar preguntas alineadas con contenidos y resultados de aprendizaje definidos por el usuario. El desarrollo se realizó mediante Investigación Basada en Diseño (DBR) y metodología ágil Scrum.

La validación incluyó encuestas, entrevistas, el cuestionario SUS y una revisión técnico-pedagógica de los ítems generados. Los resultados evidencian una alta aceptación por parte de los usuarios (4,44/5) y una usabilidad destacada (88/100). Se concluye que la plataforma es efectiva para apoyar la evaluación formativa, especialmente en niveles fáciles e intermedios, aunque se identifican oportunidades de mejora en particular en ítems para el nivel avanzado. La plataforma se proyecta como una herramienta eficaz para la creación de bancos de preguntas y la automatización de evaluaciones formativas, siempre que se complemente con procesos de curaduría pedagógica, alineación curricular y mejora continua.

Palabras clave: Inteligencia artificial generativa, GPT, usabilidad, evaluación automatizada, alineación pedagógica, educación en ingeniería.

ABSTRACT

This study presents the design, development, and validation of a platform based on generative artificial intelligence (GAI) aimed at automating the creation of academic assessments. The solution, designed for faculty members in the Computer Engineering program at UTFSM, leverages the GPT-4 API to generate questions aligned with user-defined content and learning outcomes. The development followed a Design-Based Research (DBR) approach combined with the agile Scrum methodology.

Validation included surveys, interviews, the SUS usability questionnaire, and a technical-pedagogical review of the generated items. The results indicate high user acceptance (4.44/5) and an outstanding usability score (88/100). The platform proved effective in supporting formative assessment, particularly at basic and intermediate levels, though improvements are needed for advanced-level items. It shows strong potential as a tool for building question banks and automating formative assessments, provided it is complemented by pedagogical curation, curricular alignment, and continuous improvement processes.

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

Keywords: Generative Artificial Intelligence, GPT, usability, automated assessment, pedagogical alignment, engineering education.

GLOSARIO

API (Application Programming Interface): conjunto de funciones y protocolos que permiten la comunicación entre diferentes sistemas. En este trabajo, se utiliza una API para conectar la plataforma con los modelos de IA generativa de OpenAI.

Backend: parte del desarrollo de software que se encarga de la lógica interna de la aplicación, el acceso a la base de datos y la comunicación con servicios externos. En este proyecto, el backend está implementado con el framework FastAPI.

Corpus (plural: corpora): es una colección estructurada y organizada de textos escritos o transcripciones de habla que se utiliza para el análisis lingüístico y el procesamiento del lenguaje natural (NLP). Estos textos pueden incluir libros, artículos, conversaciones, publicaciones en redes sociales, y cualquier otro tipo de contenido textual.

Frontend: interfaz gráfica de usuario (GUI) con la que interactúan directamente los usuarios del sistema. Permite visualizar la información procesada por el backend y ejecutar acciones como cargar material o generar evaluaciones.

GPT (Generative Pre-trained Transformer): modelo de lenguaje desarrollado por OpenAI, basado en la arquitectura de transformers, que permite generar texto de forma autónoma a partir de una entrada previa. GPT-4 es la versión utilizada en este proyecto.

Inteligencia Artificial (IA): campo de la informática que se enfoca en desarrollar sistemas capaces de realizar tareas que requieren inteligencia humana, tales como el razonamiento, el aprendizaje, la toma de decisiones o el procesamiento del lenguaje natural.

Investigación Basada en Diseño (DBR): enfoque metodológico que combina la intervención práctica con la generación de conocimiento teórico. Se utiliza en contextos educativos reales para diseñar, implementar y evaluar soluciones tecnológicas.

Procesamiento de Lenguaje Natural (PLN): subcampo de la inteligencia artificial que permite a las computadoras comprender, interpretar y generar lenguaje humano de manera útil y significativa. Incluye tareas como análisis sintáctico, traducción automática y generación de texto.

Prompt / Prompting: término utilizado en el campo del procesamiento de lenguaje natural (PLN) e inteligencia artificial generativa para referirse a la instrucción, pregunta o conjunto de datos de entrada que se entrega a un modelo de lenguaje (como GPT-4) con el fin de generar una respuesta coherente, contextual o específica. El prompting es la técnica mediante la cual se diseña estratégicamente esta entrada para obtener resultados precisos, relevantes o creativos. En el contexto de esta plataforma, los prompts estructurados son fundamentales para controlar el tipo, formato y dificultad de las preguntas generadas.

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

Scrum: metodología ágil de gestión de proyectos que promueve el desarrollo iterativo e incremental de productos. Se basa en roles definidos, reuniones periódicas y entregas constantes de valor funcional.

Usabilidad: grado en que un sistema puede ser utilizado por usuarios específicos para lograr objetivos específicos con efectividad, eficiencia y satisfacción. Se evalúa mediante pruebas prácticas con usuarios reales, como el cuestionario SUS.

ÍNDICE DE CONTENIDOS

RESUMEN.....	4
ABSTRACT	4
ÍNDICE DE CONTENIDOS	8
ÍNDICE DE FIGURAS.....	11
ÍNDICE DE TABLAS.....	12
INTRODUCCIÓN	13
CAPÍTULO 1: DEFINICIÓN DEL PROBLEMA	13
1.1 Contexto y línea de investigación.....	13
1.2 Situación Actual.....	14
1.3 Objetivos	17
1.3.1 Objetivo general.....	17
1.3.2 Objetivos específicos	17
1.4 Justificación	17
1.5 Limitación y alcance	18
1.5.1 Limitaciones	18
1.5.2 Alcance.....	19
CAPÍTULO 2: MARCO TEORICO	19
2.1 Inteligencia Artificial Generativa (IAG) en Educación	19
2.2 GPT y su aplicación educativa	20
2.3 Plataformas tecnológicas basadas en IA aplicadas a la educación	21
2.4 Evaluación educativa automatizada con IA	22
2.5. Consideraciones éticas y pedagógicas de la IA en educación.....	24
2.6 Usabilidad y Experiencia del Usuario (UX)	25

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

2.6.1 System Usability Scale (SUS) e interpretación de resultados	25
2.7 Relación entre el marco conceptual y el diseño de la plataforma automatizada de evaluaciones.....	26
CAPÍTULO 3: METODOLOGÍA.....	26
3.1 Investigación Basada en Diseño (DBR).....	27
3.1.1 Fases de la metodología DBR.....	27
3.2 Principios de Scrum aplicados al proyecto.....	29
3.3 Roles de Scrum	29
3.4 Integración de DBR con la metodología Scrum.....	30
CAPÍTULO 4: DISEÑO E IMPLEMENTACIÓN DE LA PLATAFORMA	32
4.1 Proceso de Desarrollo	32
4.2 Arquitectura de la plataforma.....	33
4.2.1 Backend	34
4.2.2 Middleware	35
4.2.3 Autenticación y Seguridad	35
4.2.4 Integración con Servicios Externos	35
4.2.5 Funcionalidades Adicionales.....	35
4.2.6 Frontend	35
4.3 Flujo general de la plataforma	36
CAPÍTULO 5: VALIDACION DE LA PLATAFORMA.....	38
5.1 Pruebas de validación técnica	38
5.1.1 Prueba de integración.....	38
5.2 Pruebas de validación con usuarios	38
5.2.1 Prueba de funcionalidad, calidad técnica y alineación pedagógica	38
5.2.2 Prueba de usabilidad	40

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

5.2.3 Prueba de eficiencia.....	40
5.2.4 Validación en la Feria de Software 2024	41
CAPÍTULO 6: RESULTADOS	43
6.1 Pruebas de integración	43
6.2 Prueba de funcionalidad, calidad técnica y alineación pedagógica de las preguntas.	46
6.3 Validación con usuarios de la Feria de Software de la UTFSM	50
6.4 Mejoras implementadas tras la Feria del Software 2024	55
6.4.1 Impacto de las mejoras.....	56
6.5 Prueba de Usabilidad (SUS)	56
6.6 Prueba de eficiencia	58
6.7 Producto obtenido	59
CAPÍTULO 7: CONCLUSIONES Y TRABAJO FUTURO	60
7.1 CONCLUSIONES	60
7.2 RECOMENDACIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN	62
7.3 TRABAJOS FUTUROS.....	63
ANEXO 1: CÓDIGO FUENTE	65
ANEXO 2: PRODUCTO OBTENIDO CAPTURAS DE LA PLATAFORMA.....	67
REFERENCIAS	70

ÍNDICE DE FIGURAS

Figura 1: Fases del proceso de investigación basada en diseño.	28
Figura 2: Proceso de planificación empleada en el proyecto.....	30
Figura 3: Esquema de las fases del desarrollo de la plataforma.	32
Figura 4: Arquitectura de la plataforma.	34
Figura 5: flujo del aplicativo.	37
Figura 6: Área de especialización de los encuestados.	50
Figura 7 Familiaridad del uso de la IA en el ámbito educativo.	51
Figura 8: percepción acerca de los beneficios de la IA.....	52
Figura 9: Percepción de usabilidad de la plataforma.	52
Figura 10: Capacidad de personalización.	53
Figura 11: Velocidad de generación de evaluaciones.	53
Figura 12: intención de uso de la plataforma (mejora de evaluaciones).....	54
Figura 13: intención de uso de la plataforma (uso en instituciones).	55
Figura 14: Puntaje de usabilidad (SUS) obtenido por los usuarios.	57
Figura 16: Pantalla de inicio.....	67
Figura 17: Dashboard.	67
Figura 18: Módulo de cursos.	68
Figura 19: Modulo para generar evaluaciones.....	68
Figura 20: Evaluación generada.....	69

ÍNDICE DE TABLAS

Tabla 1: Niveles de usabilidad percibida según el rango de puntaje del cuestionario SUS. Fuente: Elaboración propia.	26
Tabla 2: Integración entre las fases del enfoque de investigación basado en diseño (DBR y los ciclos iterativos del marco ágil Scrum. Fuente: Elaboración Propia.	31
Tabla 3: Rúbrica de valoración de preguntas. Fuente: Elaboración Propia.	39
Tabla 4: Rúbrica de evaluación de la calidad pedagógica de las preguntas generadas. Fuente: Elaboración propia.	40
Tabla 5: Resultado de las pruebas de integración. Fuente: Elaboración Propia.	45
Tabla 6: Resultado de las pruebas de alineación pedagógica. Nivel de dificultad Fácil. Fuente: Elaboración Propia.	46
Tabla 7: Resultado de las pruebas de alineación pedagógica. Nivel de dificultad intermedio. Fuente: Elaboración propia.	47
Tabla 8: Resultado de las pruebas de alineación pedagógica. Nivel de dificultad avanzado. Fuente: Elaboración propia.	48
Tabla 9: Síntesis de resultados. Fuente: Elaboración propia.	49
Tabla 10: Resultados en tiempos de prueba de eficiencia. Fuente: Elaboración Propia.	59

INTRODUCCIÓN

En la Universidad Técnica Federico Santa María (UTFSM), los docentes del departamento de Ingeniería Civil Informática enfrentan una alta carga laboral asociada a la creación de evaluaciones. Este proceso, aunque crucial para medir y garantizar el aprendizaje de los estudiantes, resulta tedioso y demandante, consumiendo una cantidad significativa de tiempo. Dicho tiempo podría ser redirigido hacia otras actividades igualmente relevantes, como la implementación de estrategias pedagógicas innovadoras, la orientación a los estudiantes o la investigación académica y científica.

La tecnología de inteligencia artificial (IA) generativa ofrece una solución innovadora para automatizar la creación de evaluaciones académicas, favoreciendo la personalización y eficiencia del proceso evaluativo en educación superior (Popenici y Kerr, 2017; Luckin et al., 2016). El desarrollo de una plataforma basada en IA Generativa puede ayudar a los docentes a reducir significativamente el tiempo dedicado a esta tarea, mientras se garantiza la calidad y diversidad de las evaluaciones generadas.

Esta memoria propone el diseño e implementación de una plataforma que, mediante el uso de Inteligencia Artificial Generativa —específicamente a través de la API de GPT—, permite generar preguntas y evaluaciones a partir del contenido académico entregado por los docentes. El sistema está orientado a optimizar la gestión del proceso evaluativo, liberando a los profesores de tareas repetitivas y permitiéndoles enfocarse en aspectos estratégicos de la enseñanza, como la planificación pedagógica, la retroalimentación personalizada y la innovación metodológica.

CAPÍTULO 1: DEFINICIÓN DEL PROBLEMA

1.1 Contexto y línea de investigación

El desarrollo de herramientas tecnológicas en el ámbito de la educación ha sido una prioridad para mejorar la eficacia de los procesos pedagógicos y optimizar los recursos disponibles. En particular, la informática educativa ha permitido la creación de plataformas que apoyan la labor docente, facilitando tareas fundamentales como la evaluación del aprendizaje de los estudiantes.

Uno de los mayores desafíos a los que se enfrentan los docentes en la actualidad es la elaboración de pruebas, un proceso que exige no solo la selección adecuada de contenidos y la redacción de preguntas, sino también la construcción de alternativas de respuesta y la posterior corrección de las evaluaciones. Este conjunto de actividades resulta fundamental para medir el progreso académico de los estudiantes, pero a su vez puede ser una tarea repetitiva, demandante y altamente dependiente del tiempo disponible de los profesores. Según datos recogidos en estudios previos, como el realizado por la Universidad de Santiago de Chile (Villarroel Encina et al., 2018), la falta de tiempo es una de las principales causas

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

de estrés entre los docentes, afectando tanto su calidad de vida personal como la calidad de la enseñanza impartida. Este fenómeno está estrechamente relacionado con la carga laboral, la cual incluye no solo las responsabilidades académicas, sino también tareas administrativas y la preparación curricular. Según Castilla-Gutiérrez et al. (2021), los docentes enfrentan altos niveles de carga laboral asociados a múltiples responsabilidades, lo que genera una falta de tiempo para actividades personales, riesgos psicosociales y problemas de salud física. Además, el estrés asociado a estas condiciones laborales repercute directamente en la calidad de vida de los docentes, con impactos negativos tanto a nivel físico como emocional.

De manera similar, Retamozo Ávalos et al. (2024) destacan que la presión derivada de los plazos ajustados, las demandas institucionales y los cambios en las políticas educativas contribuyen significativamente al estrés laboral docente. Estas tensiones afectan no solo su bienestar emocional, sino también su capacidad para mantener un entorno educativo de alta calidad, lo que refuerza la necesidad de implementar estrategias para abordar este problema en el contexto educativo.

La necesidad de reducir los tiempos destinados a la creación de evaluaciones sin comprometer la calidad de estas plantea un reto significativo en el sector educativo. En este sentido, la inteligencia artificial (IA) emerge como una tecnología con el potencial de transformar este proceso, automatizando la elaboración de pruebas y personalizándolas según el material entregado por el docente. Este enfoque permite no solo reducir la carga de trabajo, sino también adaptar las evaluaciones a las necesidades específicas de cada grupo de estudiantes, lo que contribuye a una mayor efectividad pedagógica.

Este trabajo se enmarca en esta línea de investigación, proponiendo el desarrollo de una plataforma web para evaluaciones automatizada basada en la tecnología de IA generativa GPT-4. Esta plataforma permitirá a los docentes crear evaluaciones personalizadas de forma automática, a partir de un corpus de material previamente proporcionado. De esta manera, se busca optimizar el tiempo invertido en la elaboración de pruebas y mejorar la calidad del proceso evaluativo, facilitando a los docentes una herramienta que se alinee con sus metodologías y los contenidos impartidos en sus asignaturas.

1.2 Situación Actual

En el escenario educativo actual, la evaluación del aprendizaje continúa siendo una de las tareas más demandantes y fundamentales del quehacer docente. En el contexto de la educación superior, esta labor se realiza mayoritariamente a través de evaluaciones escritas diseñadas manualmente por los propios profesores. Si bien estos instrumentos son esenciales para medir competencias, conocimientos y niveles de logro, su elaboración implica una secuencia de tareas complejas y exigentes: selección de contenidos, redacción de ítems, creación de alternativas de respuesta y, finalmente, corrección y retroalimentación.

El procedimiento tradicional que los docentes aplican para elaborar pruebas implica varias etapas: la selección de los contenidos a evaluar, la redacción de preguntas adecuadas, la

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

creación de alternativas de respuesta en el caso de preguntas de selección múltiple, y finalmente, la corrección de las pruebas. Cada una de estas fases demanda una considerable inversión de tiempo y esfuerzo por parte del docente, lo que resulta especialmente problemático en instituciones educativas con una alta carga de trabajo o cuando los recursos disponibles son limitados.

En particular, se identifican tres factores críticos que obstaculizan la generación de evaluaciones pertinentes y bien construidas:

- **Falta de tiempo:** Los docentes deben equilibrar diversas tareas, como la planificación de clases, la atención a los estudiantes, la corrección de evaluaciones y la gestión administrativa, lo que les deja poco margen para la creación de pruebas de calidad. La sobrecarga laboral genera un impacto negativo tanto en su bienestar personal como en la calidad del material evaluativo que producen.
- **Falta de recursos:** Muchas instituciones educativas carecen de las herramientas tecnológicas o el apoyo administrativo necesario para asistir a los docentes en la elaboración de evaluaciones. Esta carencia obliga a los profesores a realizar todo el proceso manualmente, limitando la capacidad de implementar evaluaciones que verdaderamente reflejen el aprendizaje y las necesidades de los estudiantes.
- **Ausencia de herramientas adecuadas:** Actualmente, las herramientas tecnológicas accesibles para la creación de evaluaciones no cumplen con las expectativas de los docentes en cuanto a efectividad y facilidad de uso. Las plataformas existentes suelen ofrecer funcionalidades genéricas y poco adaptables, sin responder con precisión a las necesidades pedagógicas específicas de cada asignatura o estilo de enseñanza.

Diversos estudios evidencian las consecuencias de este panorama. Estos muestran como los docentes universitarios enfrentan una carga laboral compleja y creciente, caracterizada por la exigencia de múltiples funciones que trascienden la docencia directa: preparación de clases, creación de evaluaciones, tareas administrativas, investigación, extensión y vinculación con el medio. De acuerdo con Castilla-Gutiérrez, Colihuil-Catrileo, Bruneau-Chávez, y Lagos-Hernández (2021), existen varios factores que generan efectos adversos sobre la calidad de vida del profesorado, al limitar el tiempo para la vida personal y el descanso adecuado. Entre las principales consecuencias identificadas se encuentra el incremento del estrés laboral, manifestado en síntomas como agotamiento emocional, alteraciones del sueño y presencia del síndrome de burnout, el cual afecta tanto la salud física como psicológica de los docentes. Estudios como el de Parra y Felli (2017) en la educación superior chilena, ya venían advirtiendo que docentes universitarios presentan sobrecarga de trabajo y bajos ingresos o que les obliga a dedicar tiempo de su vida personal al trabajo, aunque el ambiente entre colegas es favorable.

Asimismo, otras investigaciones recientes han demostrado la magnitud de este fenómeno. Zúñiga-Jara y Pizarro-León (2021) identificaron que un 28% de los docentes universitarios en ciencias de la salud presentan síntomas del síndrome de burnout, una condición que

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

afecta el desempeño, la motivación y el bienestar emocional del profesorado. Por su parte, Hernández y Lagos (2021) señalaron que la alta carga laboral está directamente asociada a una disminución en la calidad de vida de los docentes universitarios, producto de la presión institucional, la falta de tiempo para la vida personal y el aumento de tareas no pedagógicas.

Además, el estrés docente no sólo afecta a los individuos, sino que también puede comprometer la calidad del proceso educativo. Pérez y Martínez (2017) encontraron que casi un 20% de los académicos universitarios presenta alta tensión laboral, lo que puede incidir negativamente en la preparación de materiales de enseñanza, la innovación en la docencia y, particularmente, en la generación de instrumentos evaluativos de calidad. Este último aspecto —la elaboración de evaluaciones— se ha transformado en una tarea crítica, tanto por su complejidad como por su impacto directo en el proceso de aprendizaje.

Estas dificultades afectan a una amplia proporción del profesorado, generando altos niveles de estrés y repercutiendo negativamente en la calidad de la enseñanza. De acuerdo con Frías Ortega et al. (2018), la reducción y mejor gestión de los tiempos dedicados a tareas administrativas y de planificación se reconoce como una estrategia para mejorar tanto la calidad de vida de los docentes como la calidad de la enseñanza impartida.

Ante esta situación, se hace evidente la necesidad de una solución tecnológica que no solo automatice la elaboración de evaluaciones, sino que también permita adaptarlas a los contenidos y metodologías específicas de cada docente. En este contexto, el uso de Inteligencia Artificial Generativa (IAG), especialmente mediante modelos como GPT-4, representa un camino prometedor. La implementación de un sistema basado en esta tecnología, alimentado por un corpus de material personalizado, puede tener el potencial de reducir significativamente el tiempo invertido por los docentes en la creación de evaluaciones, al tiempo que mejora su precisión, diversidad, pertinencia y personalización. Esto permitiría a los docentes y académicos de la institución universitaria redirigir esfuerzos hacia actividades de mayor valor pedagógico, como la retroalimentación formativa, la innovación metodológica y el acompañamiento estudiantil.

El presente trabajo propone una intervención tecnológica educativa, basada en IAG como herramienta de apoyo docente. No obstante, se reconoce que la automatización no sustituye el juicio pedagógico, sino que lo complementa. Por ello, se considera fundamental que futuras investigaciones profundicen en el impacto longitudinal de estas herramientas, incluyendo variables como el estrés percibido, la carga administrativa, la satisfacción profesional y, eventualmente, el efecto en el aprendizaje de los estudiantes evaluados mediante instrumentos generados por IA.

En este contexto, y considerando la necesidad de reducir la carga docente mediante herramientas tecnológicas innovadoras, esta investigación se orienta a responder la siguiente pregunta:

¿Es posible diseñar e implementar una plataforma basada en inteligencia artificial generativa que permita a los docentes de Ingeniería Civil Informática crear evaluaciones académicas personalizadas de manera eficiente, efectiva y con altos niveles de usabilidad?

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

1.3 Objetivos

1.3.1 Objetivo general

Diseñar y desarrollar una plataforma de Inteligencia Artificial Generativa (IAG) para la creación automatizada de evaluaciones académicas en el campo de la Ingeniería Civil Informática, dirigida a docentes de la Universidad Federico Santa María de Chile.

1.3.2 Objetivos específicos

1. Analizar las necesidades de los docentes en cuanto a la creación de evaluaciones académicas.
2. Diseñar una arquitectura técnica de la plataforma de IA Generativa, considerando los requisitos de usabilidad, seguridad y escalabilidad.
3. Implementar prototipos iterativos de la plataforma y realizar pruebas de usabilidad sistemáticas, recopilando retroalimentación de usuarios clave para optimizar su funcionalidad y experiencia de uso.
4. Evaluar la adecuación pedagógica de las evaluaciones generadas por la plataforma, con énfasis en su coherencia con los contenidos académicos, su funcionalidad para los docentes y su valoración como recurso de apoyo al aprendizaje.

1.4 Justificación

La educación es un pilar fundamental para el desarrollo social y económico de cualquier país, y la calidad de la enseñanza impartida es un factor determinante en la formación de futuros profesionales. En este contexto, el proceso evaluativo desempeña un papel central, pues permite medir el aprendizaje y las competencias adquiridas por los estudiantes. Sin embargo, la elaboración manual de pruebas representa una carga significativa para los docentes, quienes deben destinar una considerable cantidad de tiempo y esfuerzo a esta tarea, en detrimento de otras actividades pedagógicas esenciales.

El presente trabajo se justifica en la necesidad de optimizar el tiempo y los recursos destinados a la creación de evaluaciones. Los estudios realizados en el ámbito educativo, como el llevado a cabo Frías Ortega et al. (2018), han demostrado que una de las principales causas de estrés en los docentes es la falta de tiempo para cumplir con sus diversas responsabilidades, incluyendo la elaboración de pruebas. Este problema no solo afecta la calidad de vida de los profesores, sino también la calidad de la educación impartida, ya que las evaluaciones, al ser realizadas de manera apresurada o bajo presión, pueden no reflejar con precisión el nivel de aprendizaje de los estudiantes.

En respuesta a esta problemática, la inteligencia artificial (IA) se presenta como una herramienta innovadora con el potencial de transformar el proceso evaluativo. Específicamente, la IA generativa, mediante el uso de modelos como GPT-4, ofrece la posibilidad de automatizar la creación de pruebas, adaptándolas de manera precisa y personalizada a los contenidos y metodologías de cada docente. Esta capacidad de personalización es particularmente relevante, ya que permite que las evaluaciones se

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

ajusten a las necesidades específicas de cada asignatura y grupo de estudiantes, mejorando así la calidad del proceso pedagógico.

El desarrollo de la plataforma de IA_g para crear evaluaciones automatizada responde a esta necesidad al proporcionar a los docentes una herramienta accesible, eficiente y personalizable que no solo reduce los tiempos destinados a la creación de evaluaciones, sino que también permite generar pruebas alineadas con los objetivos de aprendizaje y los contenidos seleccionados por los profesores. Además, esta tecnología puede contribuir a reducir el estrés laboral de los docentes, permitiéndoles dedicar más tiempo a la planificación de actividades pedagógicas, la atención a los estudiantes y el desarrollo profesional. Según Seldon y Abidoye (2018), la IA puede asumir tareas repetitivas, liberando espacio para que los docentes se enfoquen en aspectos humanos e innovadores de la enseñanza.

1.5 Limitación y alcance

1.5.1 Limitaciones

Este estudio está limitado al desarrollo de una plataforma web de inteligencia artificial generativa para la creación de evaluaciones personalizadas, dirigida a docentes de nivel medio y superior. La plataforma, “De IA_g para crear evaluaciones automatizada”, se centrará en la integración de la API GPT-4, permitiendo generar pruebas a partir de un corpus de material proporcionado por los docentes. El sistema funcionará únicamente en un entorno web accesible desde dispositivos con conexión a internet.

Las limitaciones del proyecto incluyen:

- El proyecto se restringe al diseño e implementación de un sistema basado en la API GPT-4, sin desarrollar o modificar el modelo de inteligencia artificial subyacente. Se empleará una IA preexistente, limitándose a su integración en la plataforma.
- El corpus de material que utilizará la IA para generar las evaluaciones será aportado por los docentes. No se contempla que la plataforma cree automáticamente dicho material ni que valide su calidad o adecuación.
- Las pruebas del sistema se llevarán a cabo con docentes de la Universidad Técnica Federico Santa María (UTFSM) y se evaluará con diferentes usuarios, entre ellos, docentes y asistentes de la feria de software de 2024. No se considerará una implementación fuera de este contexto en el marco de este estudio.
- La evaluación se centrará en la precisión y relevancia de las pruebas generadas, y en la percepción de los docentes usuarios. No se analizará el impacto en el rendimiento académico de los estudiantes, ni respecto a la calidad ni eficiencia. Para ello, en trabajos futuros se hará un seguimiento a largo plazo de los resultados.

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

1.5.2 Alcance

En cuanto al alcance, el proyecto tiene como objetivo diseñar e implementar una plataforma que permita a los docentes generar evaluaciones personalizadas utilizando inteligencia artificial generativa. Su desarrollo se llevará a cabo conforme a los siguientes lineamientos:

- Se integrará la API GPT-4 para que el sistema pueda generar evaluaciones y/o preguntas a partir de un corpus de material académico proporcionado por los docentes.
- Se desarrollará un backend que gestione los datos de entrada y salida (material docente y evaluaciones generadas) y un frontend que proporcione una interfaz intuitiva para que los docentes puedan interactuar fácilmente con la plataforma.
- La plataforma será evaluada en dos fases: la primera con docentes de la UTFSM y la segunda con asistentes a la Feria de Software 2024. Las percepciones recopiladas en torno a su efectividad, facilidad de uso y aplicabilidad constituirán la base de una evaluación preliminar, contribuyendo al cumplimiento de uno de los objetivos principales de esta memoria.

Esta memoria está enfocada en validar la funcionalidad técnica de la plataforma y su viabilidad como herramienta de apoyo docente, sin contemplar su escalabilidad, implementación en otros niveles educativos ni un análisis exhaustivo del impacto pedagógico.

CAPÍTULO 2: MARCO TEORICO

La incorporación de inteligencia artificial (IA) en los procesos educativos ha generado una transformación significativa en la forma en que se piensan y gestionan tanto la enseñanza como la evaluación. En particular, los avances en Inteligencia Artificial Generativa (IAG) y el desarrollo de modelos de lenguaje como GPT han abierto posibilidades inéditas para automatizar tareas docentes, entre ellas, la creación de evaluaciones académicas. Este marco teórico explora los conceptos clave que sustentan la presente memoria, profundizando en el uso de IAG, el modelo GPT, la evaluación automatizada, la usabilidad de sistemas educativos y las implicancias éticas y pedagógicas de esta tecnología.

2.1 Inteligencia Artificial Generativa (IAG) en Educación

La inteligencia artificial generativa (IAG) es una rama avanzada de la inteligencia artificial que utiliza modelos de redes neuronales profundas, como las redes transformer, para crear contenido original (como texto, imágenes, audio o código) a partir del análisis de grandes volúmenes de datos. A diferencia de la IA tradicional basada en reglas explícitas, la IAG es capaz de generar información contextual y coherente de forma autónoma, gracias a su

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

capacidad de aprendizaje no supervisado y a la inferencia probabilística sobre patrones lingüísticos y semánticos complejos (Brown et al., 2020).

En el ámbito educativo, la IAG ha abierto nuevas posibilidades para la creación automatizada de recursos didácticos personalizados, tales como explicaciones temáticas adaptadas, materiales de apoyo, preguntas evaluativas alineadas con los objetivos de aprendizaje y resúmenes de textos extensos. Estas funcionalidades no solo enriquecen los entornos virtuales de aprendizaje, sino que también promueven experiencias más inclusivas y centradas en el estudiante (García-Peñalvo, Llorens-Largo, y Vidal, 2024).

Diversos estudios coinciden en que la incorporación de herramientas basadas en IAG puede reducir significativamente la carga operativa del profesorado, permitiendo a los docentes enfocar sus esfuerzos en aspectos pedagógicos de mayor valor, como la retroalimentación cualitativa, la mentoría y el diseño instruccional (García-Peñalvo et al., 2024; Luckin et al., 2016). Por ejemplo, el uso de asistentes de planificación de clases, generadores de evaluaciones automáticas y sistemas de tutoría virtual ha demostrado facilitar la personalización del aprendizaje y mejorar la toma de decisiones educativas basadas en datos (Molina et al., 2024). Estos autores señalan que la IA debe entenderse como un medio para potenciar y ampliar el juicio profesional de los docentes, no como un sustituto de su labor.

No obstante, el uso de la IAG en educación requiere de una integración crítica, ética y pedagógicamente fundamentada. Su implementación efectiva depende de la participación del profesorado en el diseño y supervisión del uso de estas herramientas, asegurando que complementen, y no reemplacen, el juicio profesional docente. Como advierte Caballero Toro (2023), no basta con incorporar tecnología en el aula; es esencial repensar el rol docente en un contexto mediado por sistemas inteligentes, promoviendo una alfabetización digital crítica que garantice el uso ético, inclusivo y responsable de la IA.

2.2 GPT y su aplicación educativa

Los Generative Pre-trained Transformers (GPT), desarrollados por OpenAI, son modelos de lenguaje de gran escala entrenados con billones de parámetros y grandes corpus textuales. GPT-4, por ejemplo, puede generar textos coherentes, responder preguntas complejas, traducir, resumir y redactar contenido académico (OpenAI, 2023). Su arquitectura permite interpretar la semántica y el contexto de manera más precisa que modelos anteriores, lo que lo hace útil para contextos educativos.

En la educación superior, GPT ha sido explorado para aplicaciones como el soporte en tutorías virtuales, la generación de explicaciones personalizadas y, especialmente, la automatización de evaluaciones. Lundgren (2024) demostró que GPT-4 puede calificar ensayos universitarios con una concordancia comparable a evaluadores humanos, aunque señala que su uso debe complementarse con revisión docente para evitar errores interpretativos o sesgos.

Yan et al. (2023) advierten sobre los desafíos éticos y técnicos del uso de modelos como GPT en la educación, especialmente en relación con la transparencia, los sesgos

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

algorítmicos y la precisión contextual. Aun así, su capacidad para generar preguntas alineadas con objetivos de aprendizaje, niveles cognitivos y áreas temáticas lo convierte en una herramienta valiosa para el diseño de evaluaciones formativas y sumativas.

2.3 Plataformas tecnológicas basadas en IA aplicadas a la educación

En el ámbito educativo, existen diversas plataformas que incorporan inteligencia artificial para facilitar la creación de evaluaciones y apoyar el proceso de enseñanza. Por ejemplo, **Quizlet** (<https://quizlet.com/>) es una herramienta popular que permite a los usuarios crear tarjetas de estudio y cuestionarios de forma sencilla, pero su capacidad de personalización y generación automática avanzada de evaluaciones es limitada y no está orientada a contextos técnicos específicos ni a adaptaciones curriculares locales. Por otro lado, **Knewton** (<https://www.knewton.com/>) ofrece sistemas de aprendizaje adaptativo que ajustan el contenido según el desempeño del estudiante; sin embargo, su enfoque está dirigido a mercados amplios y globales, sin una integración concreta con los currículos nacionales ni particularidades pedagógicas propias de la educación chilena.

Asimismo, plataformas como **Edulastic** (<https://edulastic.com/>) facilitan la creación y administración de evaluaciones personalizadas, incluyendo algunas funciones automatizadas y reportes detallados. No obstante, su grado de adaptación a materias técnicas específicas, como las que se imparten en carreras de ingeniería, y su capacidad para manejar contenidos altamente especializados es aún limitada. Además, estas plataformas no contemplan una personalización profunda que permita al docente controlar detalladamente la dificultad y los tipos de preguntas según la progresión de su curso y las características de su alumnado.

Testualia (<https://testualia.com>) es otra plataforma que utiliza inteligencia artificial para la creación automática de evaluaciones. Aunque permite la personalización de las preguntas, su enfoque no está tan específicamente orientado a las necesidades de materias técnicas como Ingeniería Civil Informática. Su personalización, aunque avanzada, no cuenta con la opción de ajustar la dificultad de las preguntas de manera estructurada como lo permite la taxonomía de Bloom en el presente trabajo. Además, Testualia es una plataforma de pago, lo que limita su acceso a un público más amplio, en comparación con una plataforma gratuita como la que se propone en este estudio.

Por su parte, **Examica** (<https://examica.com>) ofrece una solución similar en términos de personalización de las preguntas, pero con la limitación de no contar con una integración directa con Moodle, la plataforma de gestión de aprendizaje más utilizada en la USM y en otras universidades. La posibilidad de exportar las evaluaciones generadas en formatos como Word, PDF o XML directamente a Moodle es una característica clave de la plataforma propuesta en el presente proyecto, lo que simplifica el proceso de gestión de evaluaciones y resultados, y evita la necesidad de realizar ajustes manuales o de cargar las evaluaciones a la plataforma por separado.

Por otro lado, mientras que plataformas como Examica solo permiten la distribución de las evaluaciones dentro de su propia plataforma, lo que limita la flexibilidad del docente al

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

momento de gestionar las pruebas, la solución propuesta en este estudio permite descargar las evaluaciones generadas en distintos formatos, permitiendo así una mayor libertad y facilidad de uso en el entorno educativo.

Considerando estas limitaciones, se evidencia la necesidad de desarrollar soluciones que integren inteligencia artificial generativa con un enfoque pedagógico y técnico adaptado a la realidad de la educación superior chilena, en particular para docentes de Ingeniería Civil Informática. Una de las ventajas destacables de este tipo de soluciones es su capacidad para generar evaluaciones con distintos niveles de dificultad, alineados con los niveles taxonómicos propuestos en la taxonomía de Bloom, lo que permite una mejor correspondencia con los resultados de aprendizaje esperados (Anderson y Krathwohl, 2001).

Por ello, esta tesis propone y valida una plataforma que permite a los profesores generar evaluaciones ajustadas a su material académico y a los objetivos formativos, con una personalización que supera la ofrecida por las herramientas generalistas existentes. Esta propuesta se estudia en detalle en los capítulos siguientes, donde se analizan sus características técnicas, metodológicas y los resultados de su implementación en contextos reales.

2.4 Evaluación educativa automatizada con IA

La evaluación constituye un pilar esencial del proceso educativo, ya que permite valorar el nivel de logro de los aprendizajes, orientar la retroalimentación oportuna y fundamentar decisiones pedagógicas que mejoren la enseñanza (Santos Guerra, 2003). En este contexto, la automatización de evaluaciones mediante inteligencia artificial (IA) ha emergido como una solución prometedora frente a la creciente carga laboral del profesorado y a la necesidad de ofrecer retroalimentación más oportuna y adaptada a las características de los estudiantes.

Según Luckin et al. (2016), la IA posee el potencial de transformar profundamente los sistemas educativos, no solo al automatizar tareas administrativas o evaluativas, sino también al apoyar la toma de decisiones pedagógicas mediante el análisis inteligente de datos y la adaptación del aprendizaje a necesidades individuales. Lejos de reemplazar al docente, la IA busca empoderarlo al reducir tareas repetitivas, aunque su adopción efectiva exige una participación del profesorado en su diseño e implementación.

La evaluación automatizada con IA puede abarcar diversas funcionalidades, entre ellas:

- Generación automática de ítems de evaluación (de selección múltiple, verdadero/falso, ensayo breve).
- Calificación automática, especialmente en preguntas de tipo objetivo.
- Entrega de retroalimentación inmediata y personalizada.

Estas capacidades permiten mejorar la eficiencia y escalabilidad del proceso evaluativo, especialmente en contextos con alta matrícula, sin necesariamente comprometer su calidad pedagógica (Díaz et al., 2021). Al integrar modelos generativos con corpus

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

personalizados, como contenidos del curso o materiales del docente, es posible generar preguntas contextualizadas, alineadas con los objetivos curriculares. Sin embargo, también se han señalado limitaciones, como la tendencia a producir ítems poco diversos, problemas de coherencia semántica y la necesidad de una supervisión humana constante para asegurar la calidad y validez del contenido generado (Lindner et al., 2023).

El estudio reciente de Tan, Armoush, Mazzullo, y Gierl (2024) ha abordado el uso de modelos de lenguaje de gran escala (LLMs), tales como GPT, BERT y T5, los cuales están revolucionando la generación automática de preguntas para evaluaciones educativas, permitiendo la creación masiva de ítems en distintos idiomas, dominios temáticos y formatos. Si bien se reconoce la efectividad y flexibilidad de estas tecnologías, los autores advierten que muchos estudios carecen de fundamentos educativos sólidos, ya que no consideran teorías de aprendizaje ni propiedades de medición como validez o confiabilidad. Por ello, proponen que la futura investigación integre enfoques interdisciplinarios, con participación de expertos en evaluación educativa, para asegurar que los ítems generados no solo sean gramaticalmente correctos, sino también pedagógicamente sólidos y útiles para la enseñanza y la evaluación real.

Desde una perspectiva institucional, el Departamento de Educación de los Estados Unidos ha subrayado que la IA puede automatizar procesos evaluativos y liberar tiempo para que los docentes se concentren en tareas humanas como el acompañamiento emocional o la enseñanza diferenciada (U.S. Department of Education, 2023). En la misma línea, Chan y Tsi (2023) afirman que la IA resulta más efectiva como herramienta de apoyo que como reemplazo del profesorado, dado que carece de habilidades esenciales como el juicio pedagógico o la empatía, por lo que debe ser incorporada bajo un modelo colaborativo y supervisado.

Adicionalmente, la usabilidad de las plataformas de evaluación es un aspecto clave para su adopción efectiva por parte del profesorado. De acuerdo con la norma ISO 9241-11 (2018), la usabilidad se define como el grado en que un sistema permite a los usuarios alcanzar sus objetivos con eficiencia, efectividad y satisfacción. En esta línea, Brooke (1996) desarrolló el System Usability Scale (SUS), una herramienta ampliamente validada para evaluar la percepción de usabilidad, la cual ha sido aplicada en entornos educativos para valorar prototipos de software, entornos virtuales de aprendizaje y aplicaciones móviles.

Hernández y Rodríguez (2022) concluyen que la usabilidad en plataformas docentes influye directamente en la frecuencia, profundidad y calidad del uso de sus funcionalidades. Una herramienta con alta potencia técnica puede resultar ineficaz si no es percibida como clara, accesible o adaptable a las dinámicas del aula. Por tanto, en la presente memoria se ha considerado la evaluación de usabilidad como un componente central del desarrollo iterativo de la plataforma, buscando asegurar que el diseño técnico responda auténticamente a las necesidades del profesorado en contexto.

2.5. Consideraciones éticas y pedagógicas de la IA en educación

La incorporación de inteligencia artificial (IA), y en particular de Inteligencia Artificial Generativa (IAGen), en contextos educativos plantea importantes desafíos éticos y pedagógicos que deben ser abordados con profundidad y responsabilidad. Como señala la UNESCO (2021), es fundamental establecer marcos normativos claros que garanticen la transparencia, la rendición de cuentas, la equidad y la protección de los derechos humanos en el uso de tecnologías de IA en el ámbito educativo. Estos marcos deben incluir principios orientadores que aseguren que el diseño y la implementación de herramientas inteligentes respondan a objetivos pedagógicos legítimos y estén alineados con valores de justicia social, inclusión y bienestar estudiantil.

Uno de los principales desafíos éticos identificados es la denominada opacidad algorítmica, es decir, la dificultad que tienen docentes y estudiantes para comprender cómo los sistemas de IA generan respuestas o recomendaciones. Esto puede reducir la confianza y el sentido de agencia de los usuarios, especialmente en procesos sensibles como la retroalimentación automatizada o la calificación (Porayska-Pomsta, Holmes y Nemorin, 2024). La falta de explicabilidad en los modelos generativos puede llevar a que los usuarios acepten respuestas sin un análisis crítico, debilitando así competencias fundamentales como la reflexión, el razonamiento lógico y la toma de decisiones informada.

Desde una perspectiva pedagógica, se plantea el riesgo de que el uso indiscriminado de tecnologías de IA conduzca a un aprendizaje superficial o pasivo si no se acompaña de estrategias didácticas reflexivas. En este sentido, Poyatos y Pujol (2025) advierten que, en contextos bien diseñados, la IAG puede ser una aliada para fomentar el pensamiento crítico y la autonomía del estudiante. Sin embargo, si se utiliza de manera mecánica o como sustituto del esfuerzo cognitivo, puede contribuir a una dependencia tecnológica que debilita los procesos metacognitivos.

El proyecto Artificial Intelligence and the Futures of Learning de la UNESCO (2024) refuerza esta visión al enfatizar que las tecnologías basadas en IA deben ser comprendidas como herramientas complementarias que fortalecen el juicio pedagógico del docente, en lugar de reemplazarlo. Bajo un enfoque centrado en el ser humano, se promueve el desarrollo de marcos de competencias digitales e inteligencia artificial tanto para estudiantes como para educadores, de modo que puedan utilizar estas herramientas de manera informada, crítica y ética (UNESCO, 2024).

Adicionalmente, estudios recientes coinciden en que el uso responsable de la IA en educación requiere no solo marcos éticos institucionales, sino también espacios formativos que promuevan la alfabetización crítica en IA. Por ejemplo, Holmes, Bialik y Fadel (2019) proponen integrar la ética de la tecnología y la pedagogía basada en datos en la formación inicial docente, mientras que Chan y Tsi (2023) subrayan la importancia de que el profesorado participe activamente en el diseño, implementación y evaluación de estas herramientas para garantizar su alineación con las necesidades reales del aula.

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

En síntesis, el uso de IA en educación debe estar mediado por una reflexión crítica constante que considere no solo lo que es técnicamente posible, sino también lo que es pedagógica y éticamente deseable. Este enfoque ético-pedagógico es fundamental para asegurar que herramientas como las desarrolladas en este proyecto contribuyan a una educación de calidad, centrada en el desarrollo integral de los estudiantes y el respeto a los derechos y valores humanos.

2.6 Usabilidad y Experiencia del Usuario (UX)

La usabilidad y la experiencia del usuario (UX) son aspectos fundamentales en el diseño y desarrollo de plataformas tecnológicas, especialmente cuando se dirigen a usuarios con diversos niveles de habilidad tecnológica, como es el caso de los docentes universitarios. Para asegurar que la plataforma sea efectiva y ampliamente adoptada, se implementaron principios de diseño centrado en el usuario (DCU), que priorizan la comprensión profunda de las necesidades, limitaciones y contextos de uso de los destinatarios finales.

De acuerdo con Nielsen (1994), la usabilidad comprende cinco componentes esenciales: la facilidad de aprendizaje, la eficiencia de uso, la memorabilidad, la reducción de errores y la satisfacción del usuario. Estos principios guían el desarrollo iterativo de la plataforma, orientando tanto el diseño de la interfaz como la implementación de funcionalidades que faciliten su uso y reduzcan la curva de aprendizaje.

En línea con estos principios, la norma internacional ISO 9241-210 (2019) establece que el diseño centrado en el usuario debe incluir la participación de los usuarios en todas las fases del desarrollo, la evaluación continua del producto mediante métodos empíricos y la adaptación del diseño según los resultados obtenidos. Este enfoque asegura que la solución tecnológica sea útil, usable y satisfactoria, incrementando su efectividad y aceptación.

Para medir objetivamente la usabilidad de la plataforma, se aplicó el cuestionario System Usability Scale (SUS), una herramienta validada que permite evaluar de forma rápida y confiable la percepción de facilidad de uso y satisfacción del usuario (Brooke, 1996). Los resultados obtenidos a través de este instrumento orientaron ajustes iterativos en el diseño de la plataforma, lo que permitió mejorar aspectos identificados como la organización visual, la claridad de los menús y la personalización de algunas funcionalidades, en línea con las buenas prácticas del diseño centrado en el usuario.

2.6.1 System Usability Scale (SUS) e interpretación de resultados

El System Usability Scale (SUS) es un instrumento ampliamente utilizado para evaluar la usabilidad de sistemas interactivos. Consta de 10 afirmaciones que los usuarios responden utilizando una escala de Likert de 5 puntos, que va desde **"totalmente en desacuerdo"** hasta **"totalmente de acuerdo"** (Brooke, 1996). La combinación de las respuestas permite obtener un puntaje global de usabilidad que varía entre **0 y 100**.

Una de las principales ventajas del SUS es su simplicidad y rapidez de aplicación, lo que lo convierte en una herramienta ideal para evaluaciones iterativas durante el desarrollo de sistemas. Cabe destacar que la escala no mide aspectos técnicos en profundidad, sino que

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

se enfoca en la **percepción subjetiva del usuario** respecto a la facilidad de uso, la complejidad, la consistencia y la confianza al interactuar con el sistema.

El puntaje total obtenido a partir del SUS no representa un porcentaje, pero puede interpretarse según rangos estandarizados que permiten categorizar la usabilidad percibida. Bangor, Kortum y Miller (2009), propusieron una escala adjetiva para interpretar los puntajes de la SUS, asignando descripciones cualitativas a diferentes rangos de puntajes. Por ejemplo, puntajes en el rango de 85 a 100 se consideran "mejor imaginable", de 70 a 84 como "bueno", de 50 a 69 como "aceptable/promedio", de 25 a 49 como "deficiente" y de 0 a 24 como "inaceptable".

Tabla 1: Niveles de usabilidad percibida según el rango de puntaje del cuestionario SUS.
Fuente: Elaboración propia.

Rango de puntaje SUS	Nivel de usabilidad percibida
85 – 100	Excelente
70 – 84	Bueno
50 – 69	Aceptable / Promedio
25 – 49	Deficiente
0 – 24	Inaceptable

2.7 Relación entre el marco conceptual y el diseño de la plataforma automatizada de evaluaciones

Los conceptos desarrollados en el marco teórico no solo orientan la comprensión del problema, sino que también sirvieron como base para las decisiones técnicas y funcionales tomadas durante el desarrollo de la plataforma. Por ejemplo, el enfoque de **Inteligencia Artificial Generativa**, fundamentado en el uso de modelos de lenguaje como GPT-4, motivó la inclusión de un módulo de generación automática de preguntas basado en corpus cargado por el docente. Asimismo, los principios de **Personalización del aprendizaje** se reflejan en la posibilidad de adaptar los ítems generados según el nivel cognitivo o el contenido específico de una asignatura. Por otra parte, la noción de **usabilidad**, entendida como la facilidad con que los usuarios pueden interactuar eficazmente con el sistema (Nielsen, 1994), fue central al momento de definir la interfaz. Las decisiones sobre navegación, colores y organización de menús se validaron mediante herramientas como el cuestionario SUS, asegurando que la experiencia docente fuese intuitiva y satisfactoria.

CAPÍTULO 3: METODOLOGÍA

El desarrollo de la plataforma para evaluaciones automatizadas se fundamenta en un enfoque metodológico mixto que combina la Investigación Basada en Diseño (DBR) con la metodología ágil Scrum. Esta integración permite abordar el proyecto desde dos perspectivas complementarias:

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

- **DBR** se centra en el diseño, análisis y evaluación de la plataforma como una intervención educativa, generando conocimiento teórico y práctico sobre su uso en contextos reales.
- **Scrum** organiza y gestiona el proceso técnico de desarrollo, garantizando la entrega incremental de funcionalidades y la adaptabilidad a los cambios.

Esta combinación metodológica permite el desarrollo de un producto funcional, en conjunto con una investigación rigurosa sobre su efectividad y relevancia en el contexto educativo.

3.1 Investigación Basada en Diseño (DBR)

La Design-Based Research (DBR) es una metodología de investigación aplicada que tiene como propósito diseñar, desarrollar, implementar y evaluar soluciones prácticas a problemas reales en contextos educativos específicos, al mismo tiempo que genera conocimientos teóricos que contribuyen al avance de la investigación en educación (Benito y Salinas, 2016). Se trata de un enfoque iterativo, participativo y flexible, que integra estrechamente el diseño y la indagación empírica, permitiendo comprender fenómenos complejos a través de la intervención sistemática en entornos auténticos.

A diferencia de otros enfoques más tradicionales, la DBR no se limita a observar o describir una realidad educativa, sino que busca transformarla mediante la implementación de innovaciones pedagógicas o tecnológicas. Estas intervenciones son refinadas en ciclos sucesivos de diseño, prueba, análisis y rediseño, lo que favorece la mejora continua tanto de la solución desarrollada como de la teoría subyacente.

Una característica distintiva de la DBR es su compromiso con la validez, ya que las soluciones se desarrollan y validan en escenarios reales, por ejemplo, aulas universitarias o plataformas educativas, con la participación de los actores implicados, como docentes, estudiantes y diseñadores instruccionales. Asimismo, esta metodología combina métodos cuantitativos y cualitativos para capturar tanto el impacto de la intervención como las dinámicas del proceso.

3.1.1 Fases de la metodología DBR

Según Benito y Salinas (2016), el proceso de investigación basado en DBR se estructura en varias fases interrelacionadas.

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS



Figura 1: Fases del proceso de investigación basada en diseño.

Análisis de la situación y detección del problema educativo: En esta primera fase se identifican las necesidades y desafíos reales del contexto en que se desea intervenir, recurriendo a diagnósticos preliminares, revisión de literatura y consulta con actores clave. Esta etapa permite definir los objetivos de la intervención y justificar su pertinencia.

Desarrollo de la solución: Basándose en principios teóricos y en el análisis contextual, se diseña una intervención pedagógica o tecnológica que responda al problema identificado. Este diseño inicial puede incluir prototipos, guías, plataformas o recursos instruccionales.

Implementación en contexto real: La solución es puesta en práctica en el entorno educativo real, permitiendo observar su funcionamiento, sus fortalezas y debilidades. La participación de los usuarios es fundamental para captar información rica y contextualizada.

Validación (Evaluación formativa e iteración): Se recogen datos cualitativos y cuantitativos mediante instrumentos como encuestas, entrevistas, observación directa y análisis de desempeño. Los resultados permiten refinar el diseño de manera iterativa, optimizando su efectividad y aceptabilidad.

Producción de documentación: A partir de los resultados obtenidos, se extraen principios de diseño, patrones de interacción o modelos conceptuales que trascienden el caso específico, contribuyendo así al conocimiento generalizable en el campo de la tecnología educativa o la innovación pedagógica.

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

En este proyecto, la DBR se aplica para:

1. **Identificar el problema y las necesidades:** A través de entrevistas y encuestas a docentes de la UTFSM, se identificaron las principales dificultades en la creación de evaluaciones académicas, como la falta de tiempo y herramientas adecuadas.
2. **Diseñar la intervención:** La plataforma de IAg para crear evaluaciones automatizada fue conceptualizada como una solución a estos problemas, basándose en principios pedagógicos y tecnológicos. La interacción constante con los usuarios permitió ajustar el diseño según sus necesidades y contextos.
3. **Implementar y evaluar en ciclos iterativos:** Esto mediante la implementación de pruebas de usabilidad con docentes, alumnos y otros expertos en el campo, recopilando retroalimentación en cada ciclo. Este proceso de mejora continua es un pilar de la DBR, permitiendo refinar la plataforma a lo largo del tiempo.
4. **Generar conocimiento teórico-práctico:** Los resultados de la implementación y evaluación contribuyen al conocimiento sobre el uso de IA generativa en la educación, ofreciendo indicadores sobre su impacto y aceptación.

3.2 Principios de Scrum aplicados al proyecto

1. **Entregas incrementales y adaptativas:** El desarrollo de la plataforma se realizará en iteraciones conocidas como *sprints*, en las cuales se abordarán pequeñas porciones de la funcionalidad total. Esto permitirá realizar ajustes y mejorar el producto con base en la retroalimentación obtenida en cada iteración.
2. **Colaboración continua:** Dado que el proyecto abarca aspectos técnicos como el desarrollo de la arquitectura tecnológica (backend y frontend), así como consideraciones de usabilidad, los integrantes del equipo de desarrollo de la plataforma se organizaron para colaborar en tareas multidisciplinarias, manteniendo reuniones frecuentes de planificación y revisión.
3. **Retroalimentación de usuarios:** Al finalizar cada sprint, los prototipos iterativos de la plataforma deben probarse con usuarios reales, incluyendo docentes y otros colaboradores de la Universidad Federico Santa María. La retroalimentación recibida permitirá ajustar el desarrollo de la plataforma a las necesidades de los usuarios finales.

3.3 Roles de Scrum

1. **Product Owner:** Este rol será responsable de definir la visión del producto y priorizar las funcionalidades en el *Product Backlog*. Se encargará de representar los intereses de los docentes de la UTFSM y garantizar que el desarrollo de la plataforma esté alineado con sus necesidades.
2. **Scrum Master:** Actuará como facilitador del equipo, asegurándose de que se sigan las prácticas de Scrum. Su función principal será eliminar obstáculos que puedan

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

afectar el progreso y promover la comunicación eficiente dentro del equipo de desarrollo.

3. **Equipo de Desarrollo:** El equipo estará compuesto por desarrolladores especializados en backend, frontend y en integración de APIs de IA. Estos integrantes serán responsables de llevar a cabo las tareas necesarias para completar cada sprint y alcanzar los objetivos del proyecto.

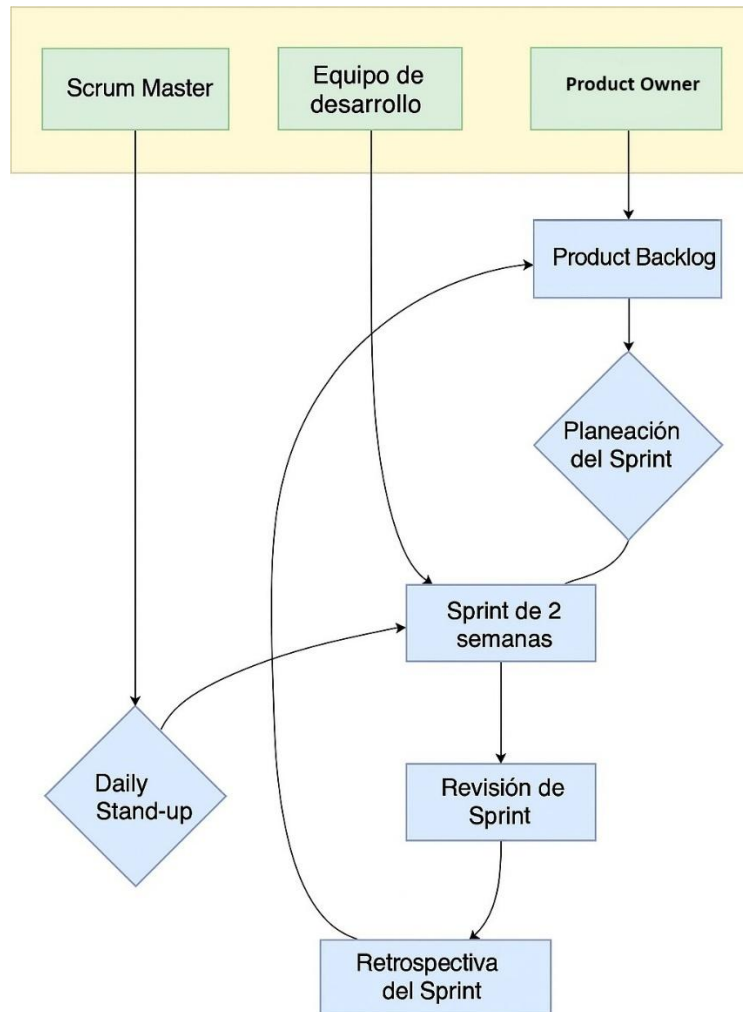


Figura 2: Proceso de planificación empleada en el proyecto.

3.4 Integración de DBR con la metodología Scrum

Para llevar a cabo este proyecto, se decidió combinar el enfoque de investigación basada en diseño (DBR) con la metodología ágil Scrum. Esta integración permitió estructurar el desarrollo de la plataforma en ciclos iterativos, con espacios definidos para el diseño,

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

prueba, ajuste y validación del sistema. A través de esta combinación, fue posible avanzar de manera ordenada pero flexible, respondiendo a las necesidades reales de los usuarios y ajustando continuamente el prototipo. La siguiente tabla muestra cómo se alinearon las fases de DBR con los distintos Sprint definidos dentro de la metodología Scrum.

Tabla 2: Integración entre las fases del enfoque de investigación basado en diseño (DBR y los ciclos iterativos del marco ágil Scrum.

Fuente: Elaboración Propia.

Fase DBR	Sprint Scrum	Actividad principal
Análisis del problema	Sprint 0 (exploración)	Levantamiento de necesidades con docentes
Diseño de la intervención	Sprint 1	Primer prototipo: arquitectura y funcionalidad base
Implementación iterativa	Sprint 2 y 3	Desarrollo, pruebas y ajustes
Evaluación y rediseño	Sprint 4	Validación, retroalimentación y mejoras finales

CAPÍTULO 4: DISEÑO E IMPLEMENTACIÓN DE LA PLATAFORMA

4.1 Proceso de Desarrollo

El proceso de desarrollo de la plataforma se estructuró en los siguientes pasos durante cada sprint:



Figura 3: Esquema de las fases del desarrollo de la plataforma.

Análisis y definición de Requerimientos (Product Backlog)

En una primera fase, el equipo realizó un análisis detallado de los requisitos y necesidades de los docentes de Ingeniería Civil Informática en la Universidad Federico Santa María. A partir de entrevistas y sesiones de recolección de datos, el **Product Owner** documentó las funcionalidades esenciales en el *Product Backlog*. Esta lista priorizada sirvió como base para planificar los sprints y comprender las expectativas de los usuarios finales. Los requisitos abarcaron aspectos como la personalización de las evaluaciones y la facilidad de uso, que guiaron la construcción de la plataforma para cumplir con los objetivos definidos.

Planificación del Sprint Inicial (Sprint Planning)

El equipo inició la metodología Scrum con una reunión de **Sprint Planning**, en la cual, junto al Product Owner, se seleccionaron tareas específicas para el primer sprint. Durante esta fase, se decidió abordar los aspectos de infraestructura técnica de la plataforma, así como la configuración preliminar de la conexión con la API de GPT-4. Las tareas fueron distribuidas en función de la complejidad y la especialización del equipo de desarrollo, con un enfoque en el avance incremental de funcionalidades. La duración de cada sprint se fijó en dos semanas, lo que permitió establecer metas claras y un ritmo de trabajo constante.

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

Ejecución del Sprint (Desarrollo)

Durante la ejecución de cada sprint, el **Equipo de Desarrollo** comenzó a trabajar en las tareas definidas. En la primera iteración, se enfocaron en la creación del entorno de desarrollo y la implementación del backend que gestionaría los datos entre la API de GPT-4 y la plataforma. Posteriormente, cada iteración fue abordando distintos componentes, como el diseño del frontend y las pruebas de conectividad con la base de datos.

Para asegurar la comunicación y el monitoreo constante, el equipo realizó **Daily Stand-ups**. Estas reuniones permitieron a los desarrolladores compartir el avance diario, identificar posibles obstáculos y coordinar los esfuerzos en tiempo real, manteniendo el proyecto alineado con el calendario acordado.

Revisión del Sprint y Validación de Resultados (Sprint Review)

Al concluir cada sprint, el equipo celebró una **reunión de revisión del sprint** para presentar el incremento de la plataforma a los stakeholders y obtener retroalimentación. En las primeras iteraciones, se desarrollaron versiones preliminares del backend y el frontend, las cuales fueron evaluadas para determinar su efectividad en la creación de pruebas y en la interfaz con el usuario. Esta retroalimentación inicial resultó fundamental para realizar ajustes en las funcionalidades básicas y mejorar la usabilidad de la plataforma desde sus primeras etapas.

Retroalimentación y Mejoras Continuas (Sprint Retrospective)

Finalizado cada sprint, el equipo llevó a cabo una **retrospectiva** para identificar las fortalezas y las áreas de mejora en el proceso de desarrollo. Esta sesión permitió un análisis crítico de las dificultades encontradas, como la optimización del tiempo en el Daily Stand-up y la eficacia en la integración de la API. La retroalimentación interna llevó a mejorar los tiempos de carga y a refinar la interfaz del usuario.

Iteraciones y Ajustes Basados en las Pruebas de Usabilidad

Conforme el desarrollo avanzó y el equipo presentó versiones cada vez más completas de la plataforma, se realizaron **pruebas de usabilidad** con docentes de la Universidad. La retroalimentación permitió implementar ajustes adicionales, asegurando que la plataforma cumpliera con las expectativas en términos de personalización y adaptabilidad a diversas asignaturas. Además, el equipo ajustó la arquitectura técnica para asegurar que el sistema fuera seguro y escalable.

4.2 Arquitectura de la plataforma

La arquitectura de la plataforma está organizada en varias capas que trabajan juntas para ofrecer una experiencia completa y eficiente al usuario. En primer lugar, el **frontend** es la parte visible de la plataforma, donde los usuarios interactúan directamente a través de una interfaz amigable y sencilla, desarrollada con tecnologías modernas como React o Vue.js. Esta capa se encarga de mostrar la información y recibir las acciones que el usuario realiza.

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

El **backend** es el motor que procesa todas las solicitudes hechas desde el frontend. Utiliza el framework FastAPI para gestionar y responder a las peticiones, administrar datos y garantizar la seguridad. En esta capa se conecta a la base de datos MySQL, donde se almacena toda la información de usuarios, cursos y evaluaciones.

Entre el frontend y backend está el **middleware**, que facilita la comunicación entre ambos y asegura que las peticiones se realicen de forma segura y eficiente, permitiendo que el sistema funcione correctamente incluso cuando se usan diferentes dominios o servidores.

Finalmente, la plataforma también se integra con servicios externos, como APIs de inteligencia artificial, que enriquecen sus funcionalidades para generar automáticamente evaluaciones personalizadas, apoyando así el trabajo de los docentes.

Esta estructura modular permite que cada parte del sistema se enfoque en tareas específicas, facilitando su desarrollo, mantenimiento y futuras mejoras, además de entregar una experiencia fluida y confiable al usuario final.

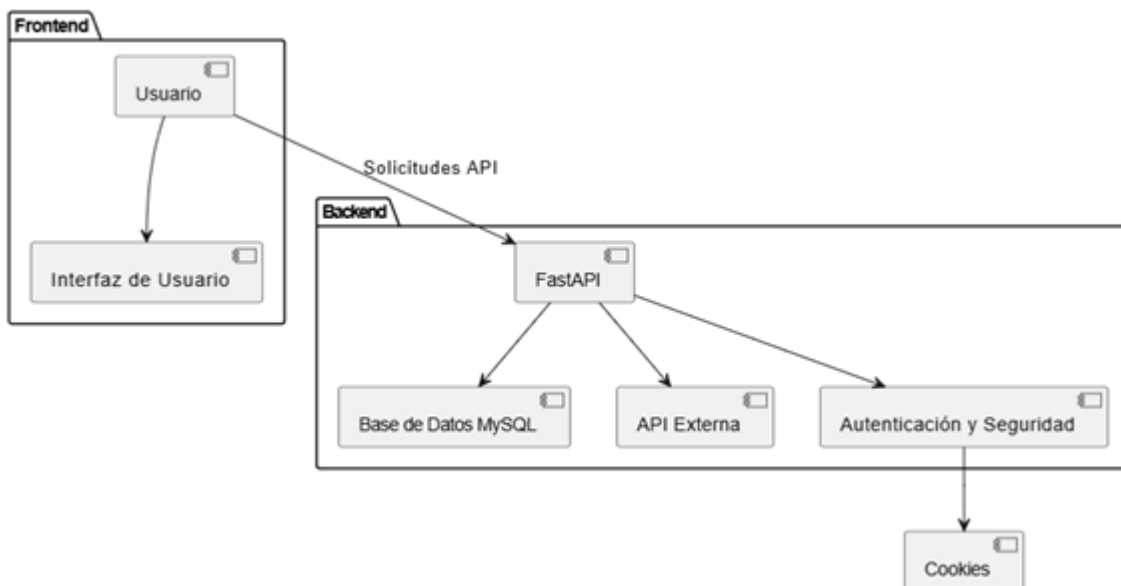


Figura 4: Arquitectura de la plataforma.

4.2.1 Backend

- **Framework:** FastAPI
 - Proporciona un rendimiento alto y es fácil de usar para crear APIs RESTful.
- **Manejo de Rutas:**
 - Rutas para la gestión de usuarios (registro, inicio de sesión, actualización de perfil).
 - Rutas para la gestión de cursos y unidades (crear, actualizar, eliminar).
 - Rutas para obtener información sobre cursos y actividades.

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

- **Base de Datos:**
 - **MySQL:** Utilizado para almacenar datos de usuarios, cursos y unidades.
 - Conexiones a la base de datos gestionadas mediante funciones dedicadas (connect_db).
- **Modelo de Datos:**
 - Uso de Pydantic para definir modelos como UsuarioCreate para validar los datos entrantes.

4.2.2 Middleware

- **CORS (Cross-Origin Resource Sharing):**
 - Configurado para permitir solicitudes desde diferentes orígenes, lo que es esencial para aplicaciones que tienen un frontend separado.

4.2.3 Autenticación y Seguridad

- **Sesiones y Cookies:**
 - Manejo de autenticación mediante cookies para mantener la sesión del usuario.
- **Manejo de Errores:**
 - Uso de excepciones HTTP para manejar errores en las rutas, proporcionando mensajes claros en caso de fallos.

4.2.4 Integración con Servicios Externos

- **API Externa:**
 - Integración con una API externa para crear asistentes y vectores relacionados con los cursos.

4.2.5 Funcionalidades Adicionales

- **Envío de Correos:**
 - Configuración para enviar correos electrónicos, lo que puede ser útil para notificaciones o recuperación de contraseñas.

4.2.6 Frontend

- **Frameworks:**
 - React o Vue.js, que permiten construir interfaces dinámicas y responsivas.
- **Funciones:**
 - Presentar la interfaz de usuario para interacción con la plataforma, gestionar formularios, mostrar resultados y manejar la navegación.

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

- **Comunicación:**
 - Se comunica con el backend a través de las rutas definidas en FastAPI, usando peticiones RESTful protegidas y autorizadas.
- **Diseño:**
 - Enfocado en la usabilidad y experiencia del usuario, asegurando accesibilidad y una curva de aprendizaje baja para los docentes.

4.3 Flujo general de la plataforma

El flujo general de la plataforma corresponde a la secuencia lógica de acciones e interacciones que se desarrollan dentro del sistema, desde el ingreso del usuario hasta la ejecución de las funcionalidades clave. Este flujo detalla cómo los usuarios navegan y utilizan las distintas herramientas disponibles, garantizando que cada etapa del proceso se lleve a cabo de forma ordenada, segura y eficiente, en concordancia con los objetivos de la plataforma. A continuación, se describe dicho flujo.

1. Registro e inicio de sesión:

- Los usuarios pueden registrarse proporcionando información personal.
- Al iniciar sesión, se valida la información y se establece una sesión.

2. Gestión de cursos:

- Los usuarios pueden crear nuevos cursos y unidades asociadas.
- Cada curso puede tener múltiples unidades y actividades.

3. Interacción con el Frontend:

- El frontend se comunica con el backend a través de las rutas definidas en FastAPI.
- Cuando el usuario realiza acciones que requieren generación automática de evaluaciones, el frontend envía solicitudes a una API interna que a su vez se comunica
- El backend recibe la respuesta generada por GPT, la procesa y la envía nuevamente al frontend para mostrarla al usuario.
- Este flujo permite manejar la lógica compleja y la interacción con servicios externos sin exponer detalles al usuario, garantizando una experiencia fluida y segura.
- El frontend está desarrollado con un framework moderno como React o Vue.js, aprovechando el middleware CORS para una comunicación segura y eficiente.

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

4. Manejo de Actividades:

- Los usuarios pueden ver y gestionar actividades dentro de las unidades.

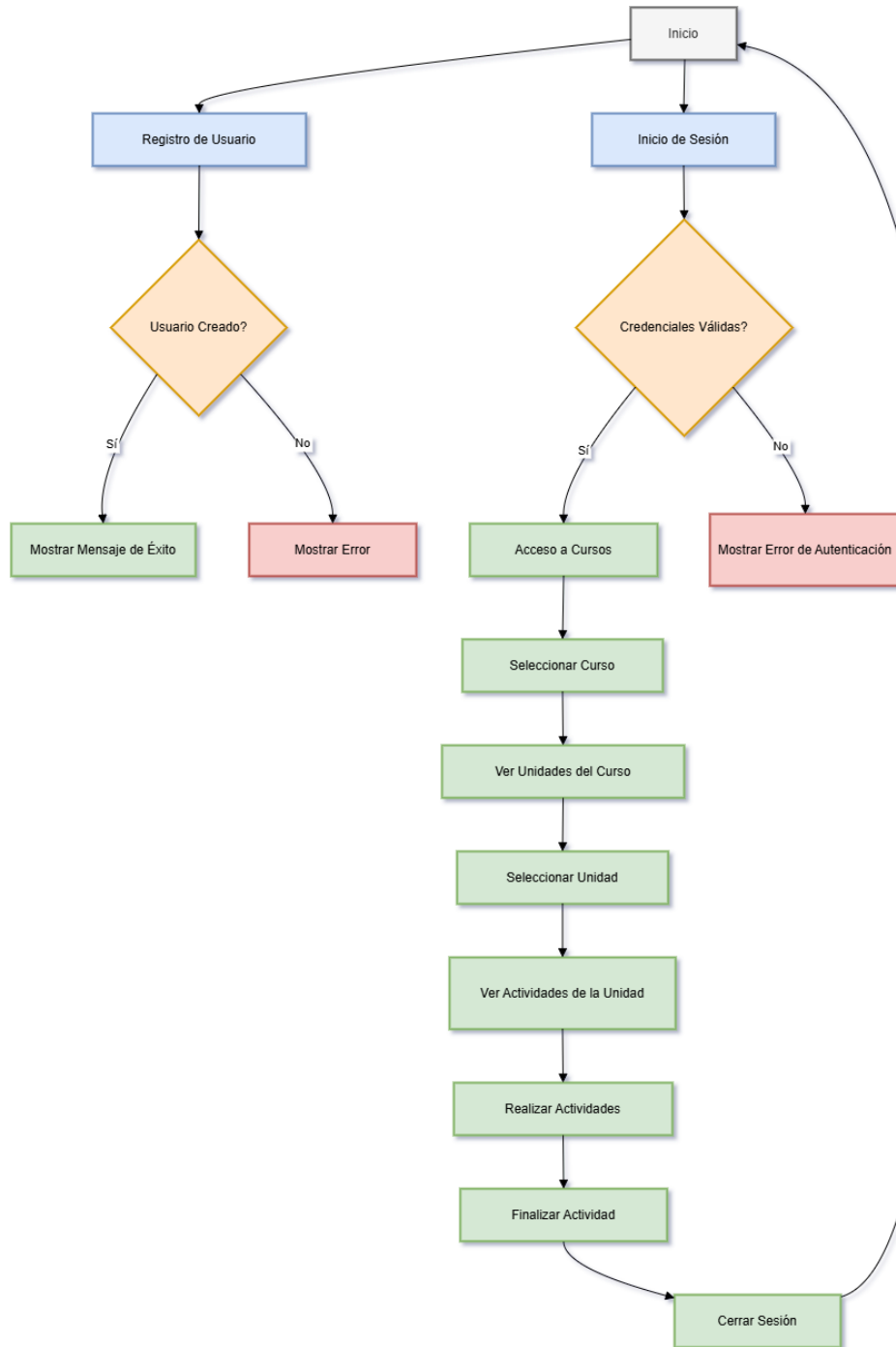


Figura 5: flujo del aplicativo.

CAPÍTULO 5: VALIDACION DE LA PLATAFORMA

Este capítulo presenta el proceso de validación de la plataforma desarrollada. Se diferencia entre pruebas de **integración técnica**, orientadas a asegurar el correcto funcionamiento de los componentes del sistema, y pruebas de **validación con usuarios**, destinadas a evaluar la calidad, utilidad, usabilidad y pertinencia pedagógica del producto final desde la perspectiva de sus usuarios reales.

5.1 Pruebas de validación técnica

5.1.1 Prueba de integración

La prueba de integración tiene como finalidad validar el funcionamiento conjunto de los principales componentes del sistema —frontend, backend y API externa de GPT-4— asegurando una interacción fluida, coherente y libre de errores. Los criterios de evaluación considerados incluyeron la correcta operación de las rutas RESTful, la comunicación sin fallos con servicios externos, un manejo robusto de errores y excepciones, y la persistencia adecuada de los datos.

Método. Se implementaron pruebas automatizadas utilizando las librerías *pytest* y *httpx*, las cuales simulan solicitudes reales desde la interfaz de usuario hacia el servidor. Estas pruebas abarcan funcionalidades críticas como la gestión de usuarios, cursos y evaluaciones, así como la generación automática de preguntas mediante la API de GPT-4. Con el fin de preservar la integridad del entorno de producción, se empleó una base de datos aislada durante las pruebas.

Criterios de evaluación. Los criterios de evaluación considerados incluyeron la correcta operación de las rutas RESTful, la comunicación sin fallos con servicios externos, un manejo robusto de errores y excepciones, y la persistencia adecuada de los datos.

5.2 Pruebas de validación con usuarios

5.2.1 Prueba de funcionalidad, calidad técnica y alineación pedagógica

Esta etapa tuvo como propósito verificar que la plataforma cumpliera con los requerimientos funcionales definidos durante el diseño, poniendo especial énfasis en la calidad de las evaluaciones y preguntas generadas. Una plataforma de generación de evaluaciones no puede ser considerada exitosa únicamente por su capacidad de operar sin errores, sino que debe demostrar que los ítems que genera, en este caso, las preguntas y pruebas son válidos desde el punto de vista pedagógico. Esto significa que deben estar alineados con los contenidos proporcionados por el docente, representar distintos niveles cognitivos según marcos como la taxonomía de Bloom (Anderson y Krathwohl, 2001), y responder a los criterios de relevancia, claridad, coherencia semántica y nivel de dificultad apropiado para el contexto educativo en el que se utilizarán.

En el caso particular de plataformas que generan evaluaciones automatizadas, como la desarrollada en este proyecto, resulta indispensable evaluar tanto la solidez técnica de la solución como la coherencia y pertinencia de los ítems generados con respecto a los contenidos curriculares, objetivos de aprendizaje y enfoques pedagógicos del contexto

**DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA
CREACIÓN DE EVALUACIONES ACADÉMICAS**

disciplinar específico. Si bien, los modelos de Lenguaje (LLMs), como GPT, entre otros, ofrecen ventajas técnicas importantes, la generación automática de ítems debe mantenerse como un enfoque que complementa, pero no reemplaza la experiencia humana. La calidad, validez y aplicabilidad de los ítems dependen de integrar expertos humanos y fundamentos educativos sólidos en todo el proceso. (Tan et al., 2024).

Por ello, esta etapa de validación incluyó no solo pruebas en el entorno técnico de desarrollo, sino también la revisión cualitativa por parte de un docente experto, que permitió evaluar la calidad pedagógica de las preguntas generadas por la plataforma.

Método. Se realizaron pruebas en el entorno de desarrollo para validar el correcto funcionamiento de las principales características de la plataforma. Además, un docente experto revisó las evaluaciones generadas para comprobar su calidad y alineación pedagógica con los contenidos académicos y resultados de aprendizaje. Se utilizó la siguiente rúbrica de evaluación para cada pregunta.

Tabla 3: Rúbrica de valoración de preguntas.
Fuente: Elaboración Propia.

Ítem	Criterio	Puntaje (1–4)	Observaciones / Justificación
1	Relevancia temática (Correspondencia con el contenido)		
2	Alineación con objetivos de aprendizaje (Relación con el nivel y objetivo del curso)		
3	Nivel cognitivo (Taxonomía Bloom)		
4	Claridad y redacción (Comprensible, sin ambigüedad)		
5	Pertinencia del tipo de ítem (Adecuado al nivel)		

- **Criterios de evaluación.** Los criterios de evaluación considerados incluyeron la precisión en la generación de evaluaciones según el material proporcionado y la opinión experta sobre la pertinencia y calidad pedagógica de las preguntas.

Para la rúbrica de evaluación que se muestra en la tabla 2, se utilizó la siguiente escala de valoración:

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

Tabla 4: Rúbrica de evaluación de la calidad pedagógica de las preguntas generadas.
Fuente: Elaboración propia.

Puntaje	Nivel de cumplimiento	Descripción
4	Excelente	Cumple completamente con el criterio de manera precisa, clara y coherente.
3	Bueno	Cumple en general con el criterio, aunque podría beneficiarse de pequeñas mejoras.
2	Insuficiente	Cumple parcialmente; requiere ajustes importantes para ser pedagógicamente válido.
1	Deficiente	No cumple con el criterio; la pregunta presenta errores graves o carece de alineación o claridad.

5.2.2 Prueba de usabilidad

Objetivo: Medir la facilidad de uso de la plataforma y la percepción de los usuarios respecto a su diseño y experiencia de usuario.

Método: Para las pruebas de usabilidad de la plataforma, se aplicó el cuestionario estándar de usabilidad del sistema (SUS) y se seleccionó una muestra compuesta por seis docentes de Ingeniería Civil Informática. Esta decisión responde a tres criterios principales. En primer lugar, desde un enfoque metodológico, se ha demostrado que una muestra pequeña, de entre 5 y 8 usuarios, es suficiente para identificar al menos el 80% de los problemas de usabilidad en un sistema (Nielsen, 1994). En segundo lugar, se trató de una muestra intencionada y pertinente, seleccionada entre docentes que efectivamente diseñan y aplican evaluaciones en sus cursos y en distintas áreas, lo cual aseguraba su experiencia directa con la problemática abordada. Finalmente, se optó por esta cantidad debido a razones prácticas de disponibilidad y recursos, propias de estudios exploratorios con enfoque iterativo, donde la retroalimentación obtenida se incorpora rápidamente al diseño y mejora del prototipo.

Criterios de evaluación:

- Puntuación obtenida en la escala SUS.
- Comentarios y observaciones de los participantes sobre la interfaz y experiencia de usuario.

5.2.3 Prueba de eficiencia

La eficiencia es un componente clave en la evaluación de sistemas interactivos, especialmente en contextos educativos donde el tiempo de los docentes es un recurso crítico y limitado. De acuerdo con la norma ISO 9241-11 (2018), la eficiencia se refiere al grado en que los usuarios pueden lograr sus objetivos con una inversión adecuada de

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

recursos (principalmente tiempo y esfuerzo). Según (Castilla-Gutiérrez et al., 2021; Frías Ortega et al., 2018) concluyen que la alta carga laboral genera falta de tiempo personal, aumenta riesgos psicosociales (como estrés) y afecta la salud física de los docentes, sin diferencias significativas entre países (Europa, Asia, América). Por tanto, esta prueba busca evidenciar el potencial de la plataforma para liberar tiempo valioso que los docentes podrían redirigir hacia actividades pedagógicas de mayor valor añadido.

Objetivo: Determinar el tiempo que los docentes tardan en realizar tareas clave dentro de la plataforma y su percepción sobre la eficiencia del sistema.

Método: Se seleccionó la misma muestra (seis profesores de la UTFSM) y se les asignaron dos tareas específicas dentro de la plataforma. Se midió el tiempo necesario para completar cada tarea y se aplicó un cuestionario breve sobre la percepción del tiempo empleado en comparación con sus expectativas.

Criterios de evaluación:

- Tiempo promedio en completar las tareas.
- Diferencia entre el tiempo esperado y el tiempo real.
- Nivel de satisfacción respecto a la rapidez del sistema.

5.2.4 Validación en la Feria de Software 2024

Como parte del enfoque iterativo de validación adoptado en este proyecto, se llevó a cabo una instancia de evaluación en la Feria de Software 2024 de la Universidad Técnica Federico Santa María (FESW). Esta actividad permitió exponer la plataforma a un grupo diverso de usuarios en un entorno real, generando así evidencia empírica sobre su desempeño, usabilidad y percepción general.

Durante la FESW, se aplicó un cuestionario especialmente diseñado para recoger impresiones sobre aspectos clave de la plataforma, tales como:

- Facilidad de uso e intuición de la interfaz.
- Claridad en la generación de evaluaciones.
- Utilidad percibida en contextos docentes reales.
- Velocidad y eficiencia del sistema.
- Nivel de personalización disponible.
- Potencial de adopción en instituciones educativas.

La muestra incluyó estudiantes, docentes y profesionales del área tecnológica, lo que permitió obtener retroalimentación desde múltiples perspectivas y contextos de uso.

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

Además de evaluar la plataforma, esta instancia sirvió como un espacio valioso para la recolección de sugerencias cualitativas, muchas de las cuales fueron **consideradas e implementadas** en una nueva iteración del sistema. Este ejercicio de validación en terreno aportó información clave para fortalecer tanto la experiencia de usuario como la pertinencia pedagógica de la herramienta.

Los resultados específicos de esta evaluación, así como las mejoras aplicadas a partir de ella, se presentan con mayor detalle en el Capítulo 6.

CAPÍTULO 6: RESULTADOS

El presente capítulo expone los resultados obtenidos a partir de la implementación y validación de la plataforma desarrollada. Se presentan los hallazgos obtenidos de las pruebas funcionales, de integración y de usabilidad, así como el análisis cualitativo de la revisión experta de las evaluaciones generadas. Estos resultados permiten valorar el grado de cumplimiento de los objetivos del proyecto, identificar fortalezas del sistema y proponer mejoras basadas en la retroalimentación recibida.

6.1 Pruebas de integración

Las pruebas de integración tuvieron como propósito verificar el funcionamiento coordinado entre los distintos componentes del sistema. Se busca asegurar que las operaciones de creación, actualización, eliminación y recuperación de datos se ejecuten sin errores, garantizando así la consistencia y estabilidad del sistema.

A continuación, se presentan los resultados correspondientes a cada una de las pruebas realizadas:

- Prueba 1: Conexión y Autenticación con la API GPT-4
- Prueba 2: Comunicación entre Backend y Frontend
- Prueba 3: Gestión de Usuarios y Seguridad
- Prueba 4: Usabilidad y Rendimiento

Configuración del entorno de pruebas

- **Framework de pruebas:** Se utilizó *pytest* como marco principal para la ejecución automatizada de pruebas.
- **Cliente HTTP:** Se empleó *httpx* para realizar solicitudes HTTP asíncronas al API, simulando interacciones reales del frontend con el backend.
- **Base de datos:** Se configuró una base de datos independiente exclusivamente para pruebas, con el fin de evitar cualquier interferencia con los datos del entorno de producción y garantizar un entorno controlado para la validación del sistema.

Instalación de dependencias

Como parte del proceso de preparación del entorno de desarrollo, fue necesario verificar que las siguientes dependencias estuvieran correctamente instaladas, asegurando así el funcionamiento adecuado de los componentes y herramientas requeridas para las etapas posteriores de implementación y prueba.

```
pip install fastapi pytest httpx
```

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

Estructura del código de pruebas

Para el detalle completo del código de pruebas automatizadas, se remite al lector al [Anexo 1](#).

Descripción de los casos de prueba

Como parte del proceso de validación técnica de la plataforma, se definieron e implementaron diversos casos de prueba orientados a verificar el correcto funcionamiento e integración de los módulos principales del sistema. Estos casos fueron diseñados para simular operaciones clave que los usuarios o el frontend podrían ejecutar durante el uso normal de la plataforma. De este modo, fue posible comprobar que las funcionalidades básicas y críticas respondieran conforme a lo esperado, contribuyendo así a garantizar la estabilidad operativa antes de avanzar hacia pruebas en entornos reales.

- **test_create_assistant:** Verifica que se pueda crear un asistente y que la respuesta contenga el ID del asistente creado.
- **test_update_assistant:** Comprueba que se pueda actualizar un asistente existente y que se devuelva el ID actualizado.
- **test_delete_assistant:** Asegura que se pueda eliminar un asistente y que se reciba un mensaje de confirmación.
- **test_create_vector:** Valida la creación de un vector asociado a un asistente y verifica que se devuelva el ID del vector creado.
- **test_upload_file:** Comprueba que se pueda subir un archivo a OpenAI y que se reciba el ID del archivo subido.
- **test_delete_file:** Verifica que se pueda eliminar un archivo y que se reciba un mensaje de confirmación.

Detalles de ejecución

Esta sección describe brevemente cada uno de los casos de prueba ejecutados, especificando las entradas proporcionadas, los resultados esperados y los criterios de verificación para asegurar que cada función opere correctamente dentro del sistema. Estas descripciones permiten documentar de forma clara y precisa cómo se evaluó la funcionalidad de los distintos componentes, garantizando la calidad y confiabilidad de la plataforma.

A. Crear asistente

- a. **Entrada:** Nombre del asistente.
- b. **Salida:** ID del asistente creado.
- c. **Verificación:** La respuesta contiene el ID y el código de estado es 200.

B. Actualizar asistente

- a. **Entrada:** ID del asistente y del vector.
- b. **Salida:** ID del asistente actualizado.

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

- c. **Verificación:** La respuesta contiene el ID actualizado y el código de estado es 200.
- C. **Eliminar asistente**
 - a. **Entrada:** ID del asistente a eliminar.
 - b. **Salida:** Mensaje de confirmación.
 - c. **Verificación:** Confirmación recibida con código de estado 200.
- D. **Crear vector**
 - a. **Entrada:** ID del asistente.
 - b. **Salida:** ID del vector creado.
 - c. **Verificación:** La respuesta contiene el ID y el código de estado es 200.
- E. **Subir archivo**
 - a. **Entrada:** Archivo a subir, junto con los IDs del asistente y vector.
 - b. **Salida:** ID del archivo subido.
 - c. **Verificación:** La respuesta contiene el ID del archivo y el código de estado es 200.
- F. **Eliminar archivo**
 - a. **Entrada:** ID del archivo a eliminar.
 - b. **Salida:** Mensaje de confirmación.
 - c. **Verificación:** Confirmación recibida con código de estado 200.

Resultados de las pruebas

Una vez definidos e implementados los casos de prueba, se procedió a su ejecución en un entorno controlado, utilizando una base de datos independiente para evitar interferencias con el entorno de producción. Los resultados obtenidos permitieron identificar el comportamiento del sistema frente a diversas operaciones críticas, tales como la gestión de usuarios, la generación automática de evaluaciones y la comunicación con la API de GPT-4. La siguiente tabla resume los principales resultados, indicando para cada caso de prueba si fue superado satisfactoriamente, si presentó incidencias, y los ajustes realizados en caso de ser necesario.

Tabla 5: Resultado de las pruebas de integración.
Fuente: Elaboración Propia.

Caso de Prueba	Resultado	Detalles
Crear Asistente	Éxito	Se creó un asistente correctamente. ID recibido.
Actualizar Asistente	Éxito	Se actualizó el asistente con el vector indicado. ID recibido.
Eliminar Asistente	Éxito	El asistente fue eliminado exitosamente.

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

Crear Vector	Éxito	Se creó un vector asociado al asistente. ID recibido.
Subir Archivo	Éxito	Archivo subido correctamente. ID del archivo recibido.
Eliminar Archivo	Éxito	El archivo fue eliminado exitosamente.

6.2 Prueba de funcionalidad, calidad técnica y alineación pedagógica de las preguntas.

Para evaluar la funcionalidad de la plataforma, así como la calidad técnica y la coherencia pedagógica de las preguntas generadas, se diseñaron pruebas aplicadas a un corpus de contenido perteneciente a una asignatura específica: Teoría de Sistemas. En esta etapa, se generaron dos versiones de preguntas para cada unidad temática y por cada nivel de dificultad declarado (Fácil, Intermedio y Avanzado).

La revisión de los ítems producidos por la plataforma se llevó a cabo mediante una rúbrica analítica compuesta por cinco criterios fundamentales: (1) relevancia temática, (2) alineación con los resultados de aprendizaje, (3) nivel cognitivo según la taxonomía de Bloom, (4) claridad y redacción, y (5) pertinencia del tipo de ítem. Cada criterio fue valorado en una escala de 1 (deficiente) a 4 (excelente). El análisis se centró inicialmente en dos preguntas representativas del nivel fácil correspondientes a la Unidad 1, ampliándose posteriormente a los niveles intermedio y avanzado.

Resultados de las pruebas

Unidad 1 – Teorías de Sistemas y Pensamiento Sistémico

Resultado de aprendizaje asociado:

- Define el concepto de sistema, reconociendo los atributos y componentes que los distinguen.
- Reconoce los fundamentos de la Tectología, Teoría General de Sistemas y Cibernética, distinguiendo las características de cada una de esas teorías del Pensamiento Sistémico.

Nivel de dificultad declarado en la plataforma: Fácil

Nivel cognitivo esperado: Conocimiento / Memoria (Taxonomía de Bloom – Nivel 1)

Tabla 6: Resultado de las pruebas de alineación pedagógica. Nivel de dificultad Fácil.
Fuente: Elaboración Propia.

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

Pregunta	Promedio Total	Evaluación General
1. ¿Qué término describe un proceso en el cual un sistema se autorregula...?	3.8	Muy buena alineación
2. En el análisis FODA, ¿qué significa la "F"?	2.8	Alineación parcial
Promedio general (fácil)	3.3	Aceptable / con mejoras

Análisis del experto (Nivel básico).

- Se observa que la primera pregunta obtuvo puntuaciones altas en todos los criterios, destacando su alineación con los resultados de aprendizaje y su claridad. Por el contrario, la segunda pregunta, si bien es técnicamente correcta, mostró una menor congruencia pedagógica con la unidad evaluada, lo que evidencia la necesidad de una mejor curaduría del contenido generado automáticamente.
- La pregunta 2 es técnicamente correcta y adecuada para un cuestionario básico o tipo trivia, pero **no está alineada directamente con los resultados de aprendizaje de la Unidad 1** sobre teorías de sistemas. Evalúa solo recuerdo factual (repetición de un dato aprendido, sin comprenderlo o aplicarlo); útil para diagnóstico, pero limitado en profundidad.

Nivel de dificultad declarado: Intermedio

Nivel cognitivo esperado: Comprensión / Aplicación conceptual (Bloom Nivel 2–3)

Tabla 7: Resultado de las pruebas de alineación pedagógica. Nivel de dificultad intermedio.
Fuente: Elaboración propia.

Pregunta	Promedio Total	Evaluación General
1. ¿Qué características tiene un sistema invariante en el tiempo?	3.6	Buena alineación técnica y pedagógica
2. ¿Qué componente es considerado la "exportación" del sistema al medio ambiente?	3.8	Excelente claridad y alineación pedagógica
Promedio general (intermedio)	3.7	Muy buen desempeño general

Análisis del experto (Nivel intermedio)

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

- Ambas preguntas muestran una alta coherencia con los resultados de aprendizaje definidos para la unidad, especialmente con el objetivo de reconocer componentes y propiedades de los sistemas.
- Los ítems evalúan correctamente niveles **intermedios de la taxonomía de Bloom**, como la **comprensión profunda** y el **reconocimiento aplicado** de conceptos técnicos, aunque se recomienda avanzar hacia niveles de *aplicación contextualizada* en versiones posteriores.
- La pregunta sobre sistemas invariantes en el tiempo destaca por su valor técnico, aunque su redacción podría enriquecerse con ejemplos para mejorar la interpretación del estudiante.
- Se recomienda reforzar el nivel de aplicación (incorporar contexto) incluyendo situaciones problemáticas reales o simuladas en la formulación de ítems.

Nivel de dificultad declarado: Avanzado

Nivel cognitivo esperado: Nivel Bloom 3–6 (aplicar-evaluación y creación).

Tabla 8: Resultado de las pruebas de alineación pedagógica. Nivel de dificultad avanzado.
Fuente: Elaboración propia.

Pregunta	Promedio Total	Evaluación General
1. ¿Cómo se puede optimizar un proceso industrial mediante automatización avanzada?	3.0	Alta complejidad, pero débil alineación pedagógica
2. ¿Cuál es el rol del sistema MES y su contribución a la eficiencia productiva?	3.4	Muy buena redacción, pero fuera del alcance de la unidad
Promedio general (Avanzado)	3.2	Preguntas con valor pedagógico, pero requieren ser reubicadas o complementadas con ajustes curriculares

Análisis del experto (Nivel avanzado)

- Se analizaron dos ítems clasificados como nivel avanzado, enfocados en la aplicación de conceptos sistémicos en contextos industriales reales, como la automatización de procesos y el uso de sistemas MES. Se evidencia que su inclusión en la Unidad 1 resulta inconsistente, ya que los objetivos de aprendizaje definidos no abarcan explícitamente automatización avanzada ni integración tecnológica.
- La pregunta 1 del nivel avanzado, tiene alto potencial formativo, ya que exige una integración de conocimientos técnicos y sistémicos para su resolución. Sin embargo,

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

su alineación directa con los resultados de la Unidad 1 es baja, por lo que sería más adecuada en una unidad de *aplicación de pensamiento sistémico a contextos industriales o cibernética aplicada*. Incluir un caso práctico o escenario industrial concreto, para evitar respuestas excesivamente genéricas.

Síntesis de resultados de la prueba de calidad técnica y alineación pedagógica de las preguntas.

Tabla 9: Síntesis de resultados.
Fuente: Elaboración propia.

Nivel de dificultad	Puntaje promedio (escala 1–4)	Evaluación general
Fácil	3.3	Aceptable/ con mejoras
Intermedio	3.7	Muy buen desempeño pedagógico
Avanzado	3.2	Buena calidad técnica, baja alineación curricular

Hallazgos principales de las pruebas:

- Se comprobó la correcta carga e integración de materiales en distintos formatos (PDF, PPTX, Word, TXT).
- No se detectaron errores críticos en el flujo de generación de evaluaciones ni en la descarga en formato PDF.
- En cuanto a la calidad técnica y pedagógica de los ítem o preguntas, se obtuvieron los siguientes resultados:
 - La plataforma muestra una alta capacidad para generar ítems técnicamente válidos en términos de estructura, redacción y distractores aceptables. Es decir, las alternativas incorrectas no son triviales ni evidentes, sino que están diseñadas para discriminar entre estudiantes que dominan el contenido y aquellos que no.
 - Los ítems de nivel fácil e intermedio se alinean bien con los resultados de aprendizaje específicos de la unidad, siendo aptos para evaluaciones diagnósticas y formativas.
 - En el nivel avanzado, varias preguntas reflejan conocimientos transversales no contemplados en los resultados de aprendizaje de la unidad analizada, lo que afecta la coherencia pedagógica.

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

- Algunas preguntas abiertas carecen de contextualización, lo que limita su capacidad para evaluar adecuadamente niveles superiores de la taxonomía de Bloom.
- Aún es indispensable la revisión docente para garantizar la pertinencia curricular y la profundidad conceptual de los ítems generados.

6.3 Validación con usuarios de la Feria de Software de la UTFSM

Con el objetivo de profundizar en la experiencia de uso de la plataforma, se aplicó una encuesta de percepción a 41 asistentes a la Feria de Software 2024 de la UTFSM. La muestra incluyó docentes, estudiantes y profesionales del área tecnológica, lo que permitió recoger retroalimentación desde diversas perspectivas. La evaluación se centró en aspectos clave como la facilidad de navegación, la claridad en la generación de evaluaciones y la utilidad general del prototipo. La encuesta fue diseñada específicamente para esta actividad, con el objetivo de captar impresiones tanto cuantitativas como cualitativas sobre la primera versión funcional del sistema, en un entorno auténtico de validación.

A continuación, se detallan los principales resultados obtenidos en esta instancia de evaluación.

• Perfil de los usuarios

La encuesta fue respondida por un total de 41 participantes, cuyo perfil resulta altamente representativo del público objetivo de la plataforma. La mayoría corresponde a estudiantes (80%), seguidos por profesionales del área tecnológica (10%) y docentes universitarios (7,5%). Este perfil mixto aporta una perspectiva diversa, pero alineada con los usuarios finales previstos, lo que refuerza la validez de los resultados obtenidos durante esta fase de evaluación.

Pregunta: ¿Cuál es su área de especialización o a qué se dedica actualmente?

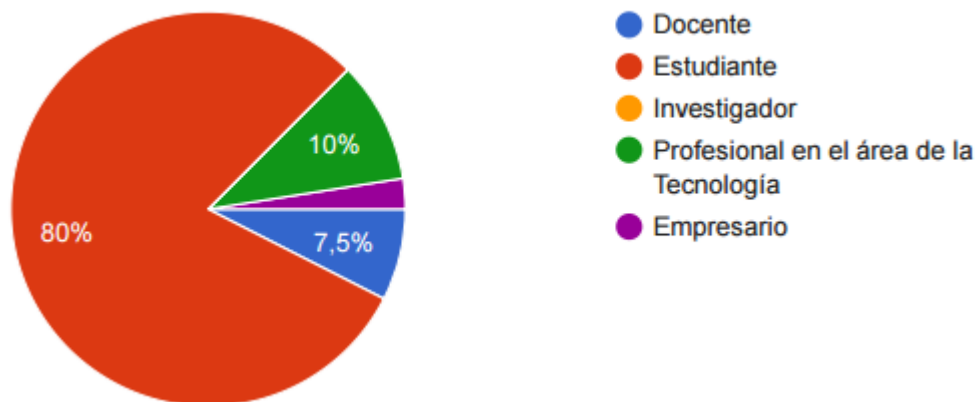


Figura 6: Área de especialización de los encuestados.

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

- **Familiaridad con la IA en el Ámbito Educativo**

Pregunta: ¿Está familiarizado/a con el uso de inteligencia artificial en el ámbito educativo?

Resultados:

El 48,8% de los participantes indicó poseer un “conocimiento básico” sobre el uso de inteligencia artificial en contextos educativos, mientras que un 24,4% afirmó estar “moderadamente familiarizado”. El resto manifestó tener poca o nula experiencia previa con esta tecnología.

Interpretación:

Estos resultados evidencian que, si bien una parte considerable de los usuarios ya posee cierta familiaridad con la inteligencia artificial, aún existe una proporción significativa que requiere mayor orientación. Esto refuerza la necesidad de incorporar recursos introductorios, como tutoriales, guías prácticas o materiales explicativos, que faciliten una apropiación progresiva y significativa de la plataforma por parte de usuarios con distintos niveles de experiencia.

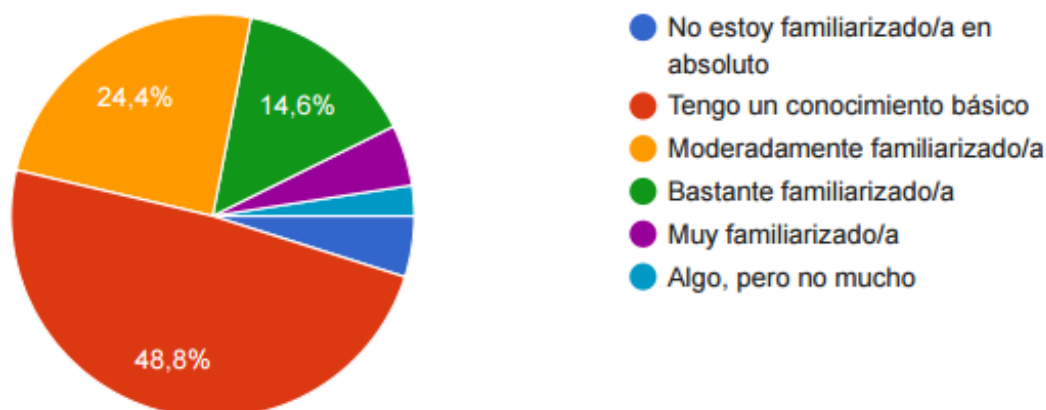


Figura 7 Familiaridad del uso de la IA en el ámbito educativo.

- **Percepción de beneficios del uso de IA para crear evaluaciones**

Pregunta: ¿Considera que el uso de IA para crear evaluaciones podría beneficiar a los docentes?

Resultados: el 78% consideró que la IA sí sería beneficiosa o muy beneficiosa; un 12,2% se mostró indeciso y solo un 7,3% cree que el beneficio sería limitado.

Interpretación: La percepción general es positiva, destacando el potencial de la IA para optimizar el proceso de creación de evaluaciones y aliviar la carga laboral de los docentes, lo que avala la pertinencia de su aplicación en este campo.

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

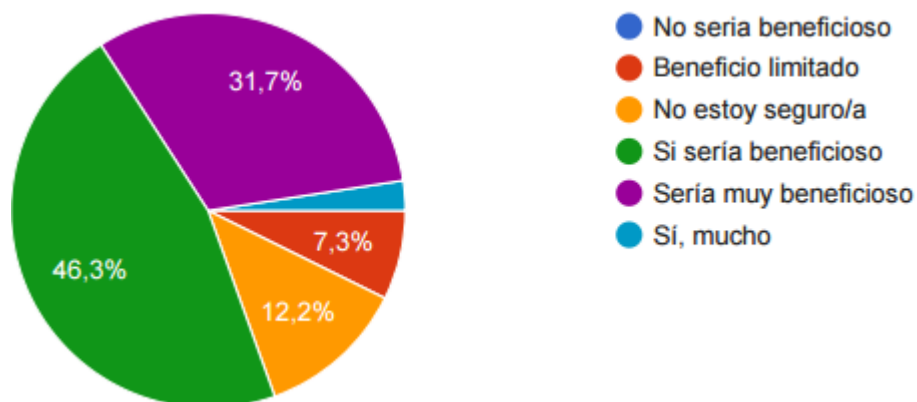


Figura 8: percepción acerca de los beneficios de la IA.

- **Facilidad de uso de la plataforma**

Pregunta: ¿Qué tan fácil le pareció entender la funcionalidad de la plataforma?

Resultados: El 68,3% calificó la plataforma como fácil de entender y un 24,4% como muy fácil de entender. Por otra parte, solo un 7,3% encontró dificultades.

Interpretación: Estos resultados indican que el diseño de la interfaz y el flujo de navegación de la plataforma resultaron mayoritariamente comprensibles para la mayoría de los usuarios.

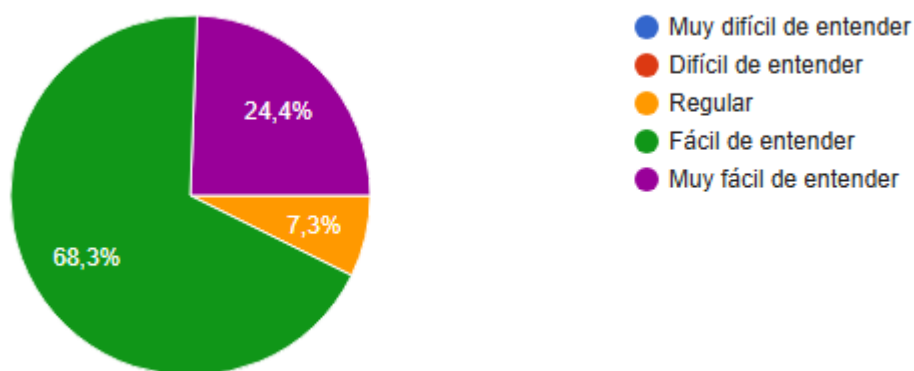


Figura 9: Percepción de usabilidad de la plataforma.

- **Capacidad de personalización de las evaluaciones**

Pregunta: ¿Cómo calificaría la capacidad de la plataforma para personalizar las evaluaciones según sus necesidades?

Resultados: El 39% calificó la personalización como adecuada, pero mejorable, mientras que un 31,7% la consideró alta. Solo un 26,8% sugirió que era muy alta.

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

Interpretación: La plataforma logra satisfacer las necesidades de personalización en gran medida, aunque existen áreas de mejora para adaptar mejor las evaluaciones a distintas asignaturas y metodologías.

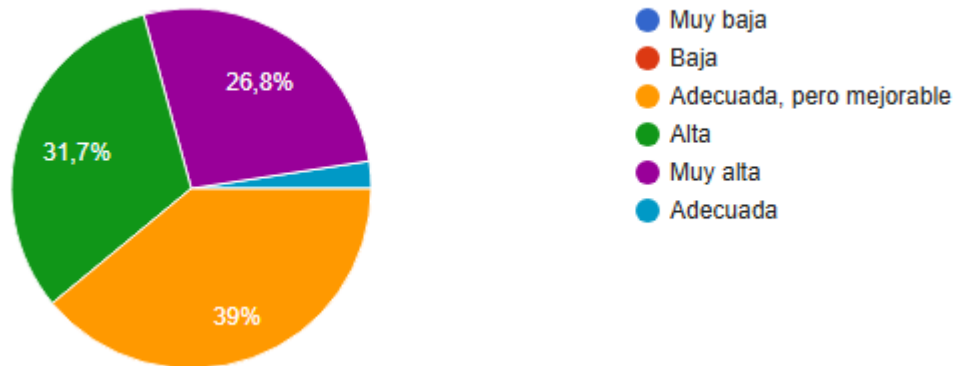


Figura 10: Capacidad de personalización.

- **Rapidez en la generación de evaluaciones**

Pregunta: ¿Que le pareció la rapidez de la plataforma para completar una tarea o evaluación?

Resultados: El 55% señaló que la plataforma era rápida; un 25% la calificó como muy rápida; un 17,5% la encontró moderada, y solo un 2.5% la considera lenta.

Interpretación: El rendimiento de la plataforma es considerado adecuado, lo que contribuye a una experiencia positiva. Sin embargo, la optimización continua es necesaria para mantener estos niveles.

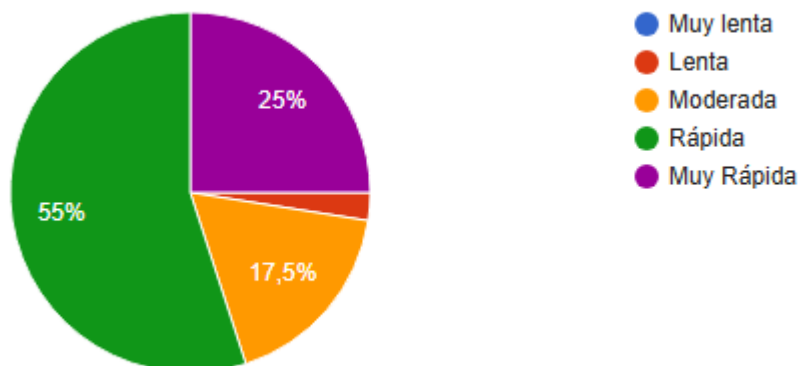


Figura 11: Velocidad de generación de evaluaciones.

- **Intención de uso y recomendaciones**

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

Pregunta: ¿Cree que esta plataforma podría mejorar el proceso de creación de evaluaciones?

Resultados: El 85,4% de los encuestados afirmó que la plataforma mejora el proceso de creación de evaluaciones y que usaría la plataforma en su institución si estuviera bien integrada. El impacto percibido es muy positivo, lo que refuerza el valor práctico de la solución propuesta.

Sugerencias destacadas: Algunos usuarios sugirieron mejoras en el diseño y la incorporación de funcionalidades adicionales, como feedback personalizado para ejercicios de desarrollo.

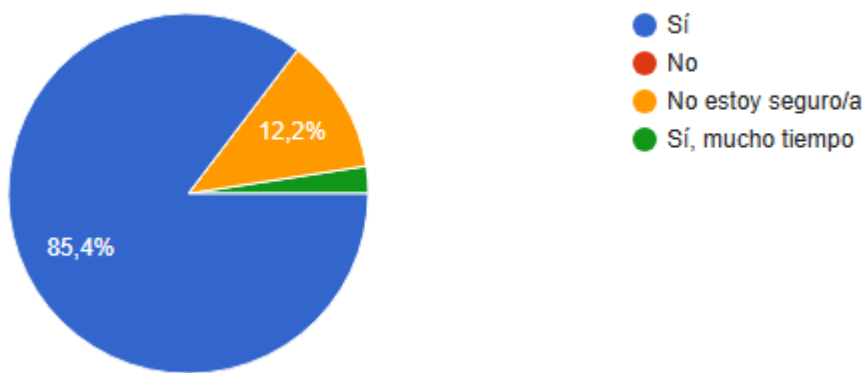


Figura 12: intención de uso de la plataforma (mejora de evaluaciones).

Pregunta: ¿Usaría esta plataforma en su institución si fuera gratuita y estuviera bien integrada?

Resultado: el 85,4% manifestó que **sí utilizaría la plataforma** si fuera gratuita y estuviera bien integrada y el 14,6% dijeron que **tal vez**, dependiendo de mejoras futuras.

Interpretación: Existe una alta predisposición a la adopción, siempre que se garantice una buena integración institucional y una experiencia de uso sólida que dependerá de futuras mejoras

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

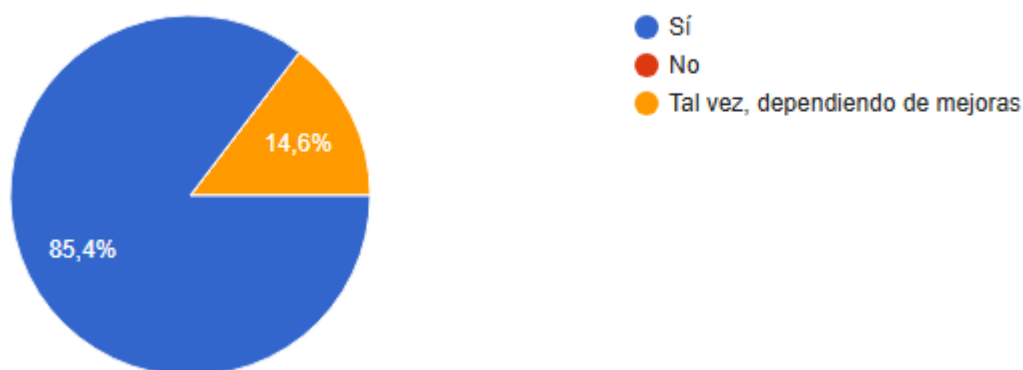


Figura 13: intención de uso de la plataforma (uso en instituciones).

Estos hallazgos no solo reflejan una alta aceptación inicial del sistema, sino que también sirvieron como insumo directo para la siguiente iteración de desarrollo, como se detalla en la sección 6.4.

6.4 Mejoras implementadas tras la Feria del Software 2024

Como parte del enfoque metodológico adoptado, específicamente, la Investigación Basada en Diseño (Design-Based Research, DBR), la plataforma fue sometida a un ciclo iterativo de validación y rediseño, en el que los datos obtenidos de la evaluación con usuarios reales fueron utilizados para refinar y fortalecer su funcionalidad (Benito y Salinas, 2016; Wang y Hannafin, 2005). En este sentido, los resultados recopilados durante la Feria del Software 2024 representaron una fase clave del proceso, al proporcionar evidencia empírica sobre la experiencia de uso, las percepciones de utilidad y las oportunidades de mejora identificadas por estudiantes, docentes y profesionales del área tecnológica.

Durante la Feria del Software 2024 se recogieron sugerencias clave de los asistentes, las cuales permitieron identificar áreas de mejora relevantes en la plataforma. Como respuesta a esta retroalimentación, se implementaron ajustes funcionales y visuales que dieron forma a una segunda iteración del sistema, más robusta y centrada en las necesidades reales del usuario docente. Esta práctica responde directamente al principio fundamental de la DBR: diseñar soluciones educativas efectivas a partir de la interacción con los propios actores del entorno en el que se implementarán. Entre las mejoras más relevantes implementadas se destacan:

Ajuste de dificultad en preguntas: uno de los aspectos señalados por los usuarios fue la necesidad de mayor control sobre el nivel de dificultad de las preguntas generadas. Como respuesta, se integró una opción para seleccionar entre tres niveles de complejidad: **fácil**, **intermedio** y **avanzado**, definidos en función de los niveles cognitivos de la taxonomía revisada de Bloom (Anderson y Krathwohl, 2001). Esta mejora permite adaptar la evaluación al momento del curso y al nivel de los estudiantes, otorgando mayor control sobre la complejidad del instrumento. Las dificultades se definen de la siguiente forma:

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

- *Fácil*: preguntas centradas en **recordar** o **comprender** información básica.
- *Intermedio*: preguntas orientadas al **análisis** o **aplicación** de conceptos.
- *Avanzado*: preguntas enfocadas en la **evaluación** crítica o la **creación** de soluciones.

Generador de preguntas independiente: Otra mejora significativa consistió en la incorporación de un generador de preguntas que opera de manera autónoma respecto al generador de evaluaciones. Esta nueva funcionalidad permite a los docentes explorar, filtrar, editar y combinar preguntas individuales según sus necesidades específicas (basadas en el material cargado, filtrando por tipo y cantidad), ampliando el rango de posibilidades de uso. Las preguntas generadas pueden ser copiadas, editadas o combinadas libremente por el docente, facilitando la construcción de evaluaciones a la medida, o incluso sirviendo como inspiración para su diseño manual. Esta mejora entrega mayor autonomía y control, equilibrando la generación automática con una experiencia flexible y creativa.

6.4.1 Impacto de las mejoras

Estas modificaciones fueron implementadas antes de la segunda fase de validación con docentes, quienes valoraron positivamente los cambios realizados. Se destacó la mayor claridad, flexibilidad y control sobre el proceso de generación de evaluaciones, lo que se tradujo en una mejora en la percepción de utilidad y pertinencia de la plataforma. Este proceso refuerza la importancia de iterar el diseño en contextos reales, combinando evidencia empírica con fundamentos teóricos para la mejora continua del sistema.

6.5 Prueba de Usabilidad (SUS)

Como parte del proceso de validación de la plataforma, se aplicó el cuestionario **System Usability Scale (SUS)** a un grupo focal compuesto por seis docentes del Departamento de Informática de la carrera de Ingeniería Civil Informática de la Universidad Técnica Federico Santa María (UTFSM). Los participantes tienen edades comprendidas principalmente entre los 30 y 60 años, incluyendo un docente en el rango de 20 a 30 años. Todos cuentan con amplia experiencia en docencia universitaria, lo que garantiza una evaluación experta y representativa desde distintas generaciones y niveles de manejo tecnológico. Esta prueba tuvo como objetivo medir la percepción general de usabilidad del sistema, abordando aspectos como la facilidad de uso, la claridad de la interfaz, la eficiencia en la ejecución de tareas y el grado de satisfacción experimentado por los usuarios.

Resultados generales:

- El puntaje promedio del cuestionario SUS fue de **88/100**, lo que, según los estándares del SUS, corresponde a una calificación sobresaliente en términos de usabilidad
- La mayoría de los docentes participantes consideró que la plataforma es **intuitiva y fácil de usar**.

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

- Se destacó la claridad del diseño, aunque algunos participantes sugirieron **mejorar la organización y color de los menús y etiquetas**.
- Se observó una curva de aprendizaje rápida, ya que los docentes lograron completar las tareas sin necesidad de instrucciones adicionales y poca ayuda externa.

De acuerdo con la escala System Usability Scale (SUS), un puntaje de 88 se considera sobresaliente y corresponde a una experiencia de uso altamente satisfactoria. Esto indica que los usuarios perciben la plataforma como **intuitiva, cómoda de usar, fácil de aprender, libre de frustraciones y que transmite seguridad durante su uso** (Brooke et al. 1996)

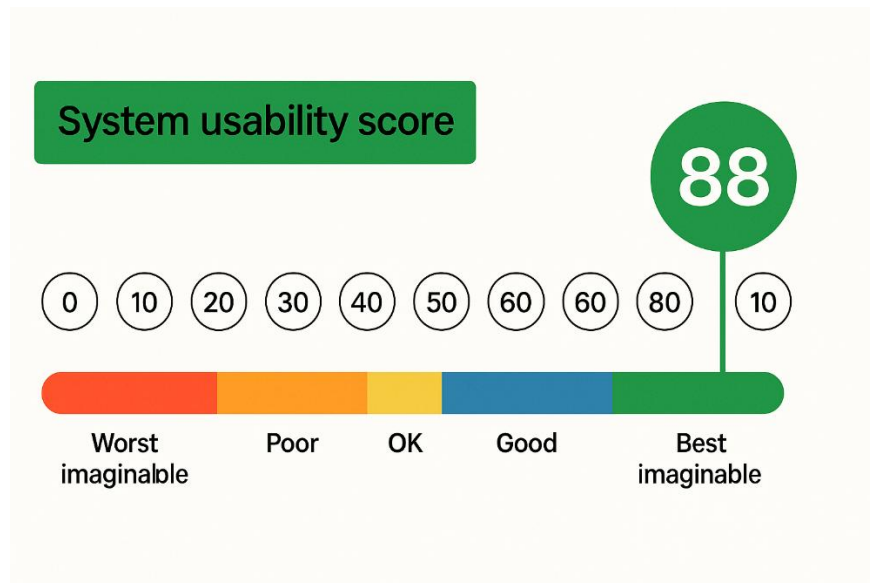


Figura 14: Puntaje de usabilidad (SUS) obtenido por los usuarios.

Comentarios destacados de los participantes:

- *"La plataforma facilita bastante la generación de evaluaciones y ahorra tiempo."*
- *"Me encantaría verlo aplicado en un LMS para usarlo en mi asignatura."*
- *"Sería útil agregar más opciones de personalización en las preguntas."*

Observaciones adicionales: Si bien la mayoría de los resultados y comentarios fueron positivos, se identificaron algunas observaciones relevantes por parte de los usuarios. Por ejemplo, se señaló que algunas preguntas generadas se enfocaban excesivamente en fragmentos específicos del material de entrada, basándose en ejemplos particulares del contenido más que en evaluar conocimientos generales de la asignatura. Esto limita la aplicabilidad de la evaluación a contextos más amplios y plantea un desafío en la generación de preguntas con un mayor nivel de abstracción.

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

También se observó que, en ciertos casos, se generaba un mayor número de preguntas desde un archivo específico, mientras que otros archivos cargados recibían menor representación. Esta distribución no equilibrada puede afectar la cobertura del contenido deseado, por lo que se propone, como mejora futura, implementar un selector que permita a los docentes decidir explícitamente desde qué archivos se desea generar preguntas. Este cambio requeriría una modificación en el flujo actual de la plataforma y un ajuste en su arquitectura.

Por último, algunos docentes notaron una tendencia en la ubicación de las respuestas correctas, concentrándose en ciertas letras como la B o la C en las preguntas de alternativas. Esto puede ser corregido mediante ajustes manuales posteriores a la generación o mediante la mejora del *prompting* utilizado en la construcción de las evaluaciones.

6.6 Prueba de eficiencia

Con el objetivo de evaluar el rendimiento de la plataforma en condiciones de uso real, se llevó a cabo una prueba de eficiencia centrada en la medición del tiempo requerido por los docentes para completar tareas específicas dentro del sistema (generar evaluaciones, preguntas). Esta prueba buscó comparar el tiempo real de ejecución con las expectativas iniciales de los profesores, así como con los tiempos estimados que implicaría realizar estas mismas acciones mediante métodos tradicionales o manuales.

Resultados generales:

- **Tarea 1:** Creación de una asignatura, subida de material y generación de una evaluación
 - **Tiempo promedio:** 4,28 minutos + 6 minutos de revisión (esperado: 20 minutos).
- **Tarea 2:** Uso del generador de preguntas
 - **Tiempo promedio:** 1,93 minutos (esperado: 10 minutos).
- **Satisfacción con la eficiencia:** El 100% de los docentes consideró que la plataforma fue rápida y eficiente en comparación con métodos manuales.

Los resultados sugieren que la plataforma reduce de forma significativa el tiempo requerido para la creación de evaluaciones y la administración del contenido académico. En entrevistas posteriores a la prueba, los docentes señalaron que, en promedio, la creación manual de una evaluación les toma al menos dos horas, considerando la lectura del material, redacción de preguntas y organización del formato. En cambio, utilizando la plataforma, lograron realizar tareas similares en menos de 6 minutos promedio, lo que representa una reducción del tiempo de trabajo de más del 95% en tareas específicas, incluso tratándose de su primera vez usando la herramienta.

Para respaldar estos hallazgos, se recopilaron los tiempos efectivos que cada docente tardó en realizar tareas reales dentro de la plataforma:

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

Tabla 10: Resultados en tiempos de prueba de eficiencia.
Fuente: Elaboración Propia.

Profesor/a	Tarea 1 (min)	Tarea 2 (min)
Profesor 1	4.10	2.05
Profesor 2	4.05	2.00
Profesor 3	3.90	1.85
Profesor 4	4.45	2.10
Profesor 5	5.00	2.00
Profesor 6	4.20	1.60
Promedio	4.28	1.93

Este análisis cuantitativo refuerza lo indicado por los profesores posterior a la prueba, al demostrar que las tareas que normalmente requerían una considerable inversión de tiempo ahora pueden realizarse en **menos de 5 minutos**, con una experiencia que fue valorada positivamente por todos los docentes participantes. Aunque algunos usuarios recomendaron mejoras menores, como la velocidad de carga de ciertos elementos, la eficiencia general fue percibida como muy alta, con un **potencial real de reducir significativamente la carga operativa del profesorado universitario**.

6.7 Producto obtenido

Para una visualización detallada de la plataforma, sus interfaces y funcionalidades, se remite al lector al [ANEXO 2](#), donde se presentan capturas de pantalla y figuras ilustrativas del producto final.

CAPÍTULO 7: CONCLUSIONES Y TRABAJO FUTURO

7.1 CONCLUSIONES

El desarrollo de esta memoria se llevó a cabo mediante una combinación de la Investigación Basada en Diseño (DBR) y la metodología ágil Scrum. Se logró diseñar e implementar una plataforma que automatiza la creación de evaluaciones académicas personalizadas. Utilizando la API de GPT-4 y adaptando los contenidos proporcionados por los docentes, se generaron evaluaciones alineadas con los objetivos de aprendizaje. La combinación de DBR y Scrum permitió no solo crear una herramienta funcional, sino también documentar y validar su efectividad en escenarios educativos reales.

La aplicación de DBR permitió validar empíricamente la funcionalidad y pertinencia de la herramienta en contextos reales, mientras que el enfoque ágil facilitó el desarrollo iterativo y centrado en el usuario. A través de entrevistas y encuestas aplicadas a docentes de Ingeniería Civil Informática de la Universidad Técnica Federico Santa María, se identificaron las principales dificultades en el diseño de evaluaciones, destacando la falta de tiempo y recursos como los principales obstáculos. Estos hallazgos fundamentaron la necesidad y relevancia de la solución propuesta.

La arquitectura de la plataforma fue diseñada con una lógica modular: un backend robusto que gestiona la interacción con la API de GPT-4, y un frontend intuitivo que permite a los docentes cargar contenidos, generar preguntas y administrar las evaluaciones. Se consideraron criterios de usabilidad, escalabilidad y seguridad, asegurando así la adaptabilidad del sistema a distintos contextos educativos.

Durante la Feria de Software, la plataforma fue evaluada por usuarios reales, obteniendo una calificación promedio de 4,44 sobre 5, con énfasis positivo en su facilidad de uso, rapidez en la generación de preguntas y utilidad para optimizar el tiempo docente. La mayoría expresó su disposición a utilizar la plataforma en sus instituciones si estuviera bien integrada y gratuita.

Adicionalmente, se aplicó el cuestionario de usabilidad SUS a seis docentes de la UTFSM, obteniendo un puntaje promedio de 88/100, lo que se considera un nivel de usabilidad sobresaliente. Este resultado cuantitativo respalda de forma objetiva que la plataforma es clara, intuitiva, libre de frustraciones y efectiva para usuarios sin necesidad de un mayor entrenamiento previo.

Como parte de la validación pedagógica, se realizó una revisión crítica de ítems generados por la plataforma en tres niveles de dificultad (fácil, intermedio y avanzado), utilizando criterios como relevancia temática, alineación con los resultados de aprendizaje, nivel cognitivo (según la Taxonomía de Bloom), claridad y tipo de ítem. La evaluación reveló que la plataforma tiene una alta capacidad para generar preguntas técnicamente válidas,

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

especialmente en los niveles fácil e intermedio, donde mostró una adecuada coherencia curricular. Sin embargo, en el nivel avanzado, algunas preguntas evidenciaron desalineación pedagógica, al integrar contenidos transversales no contemplados en los resultados de aprendizaje de la unidad evaluada.

Desde una perspectiva cualitativa, tanto los asistentes a la Feria del Software como los profesores participantes en las pruebas finales destacaron aspectos similares: el valor de la plataforma en términos de reducción de tiempo y su utilidad como apoyo en la labor docente. A la vez, se identificaron áreas de mejora comunes, como el equilibrio en el uso del material cargado, la necesidad de una mejor comprensión semántica al generar preguntas y la incorporación de tutoriales visuales para facilitar el uso inicial.

Lo anterior valida que el ciclo iterativo seguido, donde una *primera versión fue evaluada en la feria, luego mejorada y validada nuevamente con profesores en un entorno más controlado, fue efectivo para refinar la propuesta*. Las mejoras introducidas entre ambas fases (**generador de preguntas independiente, selección de dificultad, ajustes visuales**) fueron bien valoradas y respondieron directamente a las necesidades detectadas en el uso real de la herramienta.

El enfoque iterativo aplicado, donde una primera versión fue validada públicamente y luego refinada en función del feedback recibido, demostró ser efectivo para mejorar la plataforma. Las mejoras incorporadas entre ambas fases (como el generador de preguntas independiente, la selección de nivel de dificultad, y ajustes en la interfaz) fueron valoradas positivamente y respondieron directamente a las necesidades detectadas en su uso real.

En resumen, la plataforma de creación automatizada de evaluaciones mediante el uso de inteligencia artificial generativa (IAg) logró cumplir con los objetivos propuestos: entregar una solución funcional, adaptable y validada empíricamente tanto desde el punto de vista técnico como pedagógico. Los resultados obtenidos en las pruebas de validación demuestran que la herramienta es eficaz para apoyar la generación inicial de bancos de preguntas, especialmente en los niveles de dificultad fácil e intermedio, donde se evidenció una coherencia con los resultados de aprendizaje. Tiene el potencial de contribuir significativamente a la automatización de la evaluación formativa, siempre que se complemente con el proceso de curaduría pedagógica y alineación curricular. Su desempeño es satisfactorio y prometedor, pero requiere ajustes estratégicos para garantizar una coherencia plena con los objetivos de aprendizaje en todos los niveles.

Si bien aún se encuentra en una etapa inicial de implementación, constituye un avance significativo en la incorporación de IA en contextos educativos reales, particularmente en el ámbito de la Ingeniería Civil Informática. La plataforma se proyecta como una herramienta eficaz para la creación de bancos de preguntas y la automatización de evaluaciones formativas, siempre que se complemente con procesos de curaduría pedagógica, alineación curricular y mejora continua.

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

Su desempeño es satisfactorio y prometedor, y con ajustes estratégicos, tiene el potencial de consolidarse como una solución escalable para instituciones que busquen integrar tecnología avanzada en la planificación y gestión de sus procesos evaluativos.

7.2 RECOMENDACIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN

En base a los hallazgos obtenidos durante las distintas etapas del proceso de validación de la plataforma, se identificaron diversas oportunidades de mejora orientadas tanto a facilitar su implementación en contextos reales como a maximizar su utilidad pedagógica y técnica. A continuación, se presentan recomendaciones concretas que buscan guiar futuras implementaciones y despliegues de la plataforma en instituciones educativas.

1. Apoyo y formación inicial a usuarios docentes. Si bien la plataforma fue valorada como fácil de usar, se identificó la necesidad de brindar apoyo adicional a usuarios primerizos, especialmente considerando que muchos docentes no tienen experiencia previa con herramientas de inteligencia artificial. Se recomienda:

- Incorporar un sistema de ayuda guiada al primer uso, que acompañe al usuario en su recorrido inicial.
- Habilitar una sección de ayuda permanente, con explicaciones claras sobre el flujo de uso, funcionalidades clave y resolución de errores comunes.
- Desarrollar cápsulas audiovisuales, manuales y sesiones de capacitación que introduzcan al usuario en el uso pedagógico de IA generativa.
- Mostrar mensajes claros cuando se produzcan errores en la generación de preguntas, como respuestas vacías o fallos del modelo, para evitar confusión y mejorar la transparencia del sistema.

2. Incorporación de feedback automatizado en preguntas de desarrollo. Para ampliar el aporte pedagógico de la plataforma, se sugiere explorar mecanismos de retroalimentación automática en preguntas abiertas, lo que puede facilitar el aprendizaje autónomo de los estudiantes y apoyar la labor docente. Se recomienda:

- Utilizar análisis de palabras clave y respuestas modelo como referencia para identificar conceptos centrales.
- Generar comentarios automáticos que orienten al estudiante, refuercen lo aprendido o sugieran aspectos a mejorar.
- Mantener esta funcionalidad como un complemento, sin reemplazar la corrección manual del docente.

3. Buenas prácticas para la carga y segmentación del material. Durante la validación se observaron casos en los que las preguntas se concentraban desproporcionadamente en un solo archivo, afectando el equilibrio temático. Para mitigar este problema, se recomienda:

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

- Orientar a los usuarios sobre cómo organizar y etiquetar los archivos de forma clara y modular.
- Incluir advertencias automáticas cuando se detecte un desequilibrio en la cantidad de preguntas generadas por archivo.
- Presentar un resumen visual o gráfico que permita revisar la distribución temática de las preguntas generadas.

4. Garantizar la escalabilidad del sistema. Dado el potencial de crecimiento de usuarios, es fundamental asegurar el rendimiento de la plataforma en escenarios de alta concurrencia. Se recomienda:

- Optimizar el backend mediante almacenamiento en caché y balanceo de carga para mantener tiempos de respuesta adecuados.
- Preparar la infraestructura para soportar múltiples usuarios simultáneos sin afectar el desempeño.
- Implementar monitoreo de uso en tiempo real, que permita anticipar cuellos de botella y ajustar los recursos según demanda.

7.3 TRABAJOS FUTUROS

Si bien la plataforma ha demostrado ser efectiva y bien valorada por los usuarios en pruebas reales, aún existen líneas de desarrollo que podrían abordarse en futuras versiones para ampliar su impacto y consolidar su utilidad en el ámbito académico de la UTFSM.

- **Evaluación longitudinal en contexto real.** Una de las principales proyecciones es implementar la plataforma en un curso universitario durante un semestre completo, lo que permitiría evaluar su impacto sostenido en la planificación docente, la generación de evaluaciones, y la percepción de estudiantes y profesores. Esta instancia permitiría recopilar datos cualitativos y cuantitativos que respalden su adopción institucional.
- **Interoperabilidad con plataformas educativas (LMS).** Para facilitar la integración de la plataforma con los flujos académicos existentes, se propone explorar su conexión con sistemas de gestión del aprendizaje como Moodle. Esto permitiría la exportación directa de evaluaciones, sincronización con calendarios y gestión de notas, ampliando su funcionalidad como herramienta docente centralizada.
- **Validación semántica y control de calidad de las preguntas.** Se sugiere desarrollar mecanismos automáticos de revisión semántica del contenido generado por la IA, de manera que se detecten inconsistencias, ambigüedades o errores de contexto antes de ser presentados al docente. Esta validación serviría como filtro previo para garantizar la pertinencia y coherencia pedagógica de las preguntas con los Resultados de aprendizaje y contenidos.
- **Expansión a otras disciplinas y niveles educativos.** Actualmente, la plataforma ha sido validada en contextos relacionados con ingeniería e informática. Un trabajo futuro relevante sería adaptar sus funcionalidades y tipos de preguntas para otras

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

áreas del conocimiento, como ciencias sociales, salud o educación básica, lo que permitiría evaluar su versatilidad y escalabilidad.

Los hallazgos de esta investigación demuestran que el uso de inteligencia artificial generativa en la educación puede ser una herramienta valiosa para reducir la carga de trabajo docente y mejorar la calidad de las evaluaciones. Tal como advierten Luckin et al. (2016), la IA tiene el potencial de transformar los procesos educativos si se implementa con un enfoque pedagógico adecuado y centrado en el usuario.

ANEXO 1: CÓDIGO FUENTE

El siguiente código muestra la estructura y casos principales de las pruebas de integración automatizadas implementadas para la plataforma, utilizando pytest y httpx para validar la comunicación entre backend y API externa.

```
import pytest
from fastapi.testclient import TestClient
from main import app

client = TestClient(app)

@pytest.fixture(scope="module")
def test_app():
    yield client

def test_create_assistant(test_app):
    response = test_app.post("/create-assistant/", json={"name": "Asistente Prueba"})
    assert response.status_code == 200
    assert "id" in response.json()

def test_update_assistant(test_app):
    assistant_id = "assistant_test_id"
    vector_id = "vector_test_id"
    response = test_app.put(f"/update-assistant/{assistant_id}/{vector_id}")
    assert response.status_code == 200
    assert "id" in response.json()

def test_delete_assistant(test_app):
    assistant_id = "assistant_test_id"
    response = test_app.delete(f"/delete-assistant/{assistant_id}")
    assert response.status_code == 200
    assert "message" in response.json()

def test_create_vector(test_app):
    assistant_id = "assistant_test_id"
    response = test_app.post(f"/create-vector/{assistant_id}")
    assert response.status_code == 200
    assert "id" in response.json()

def test_upload_file(test_app):
    assistant_id = "assistant_test_id"
    vector_id = "vector_test_id"

    with open("test_file.txt", "rb") as file:
        response = test_app.post(f"/upload-file/{assistant_id}/{vector_id}", files={"archivo": file})
        assert response.status_code == 200
        assert "file_id" in response.json()
```

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

```
def test_delete_file(test_app):  
    file_id = "file_test_id"  
    response = test_app.delete(f"/delete-file/{file_id}")  
    assert response.status_code == 200  
    assert "message" in response.json()
```

ANEXO 2: PRODUCTO OBTENIDO CAPTURAS DE LA PLATAFORMA

En este anexo se presentan las capturas de pantalla y figuras que ilustran la interfaz y funcionalidades principales de la plataforma desarrollada, evidenciando el diseño, flujo y experiencia de usuario logrados tras la implementación y ajustes iterativos.

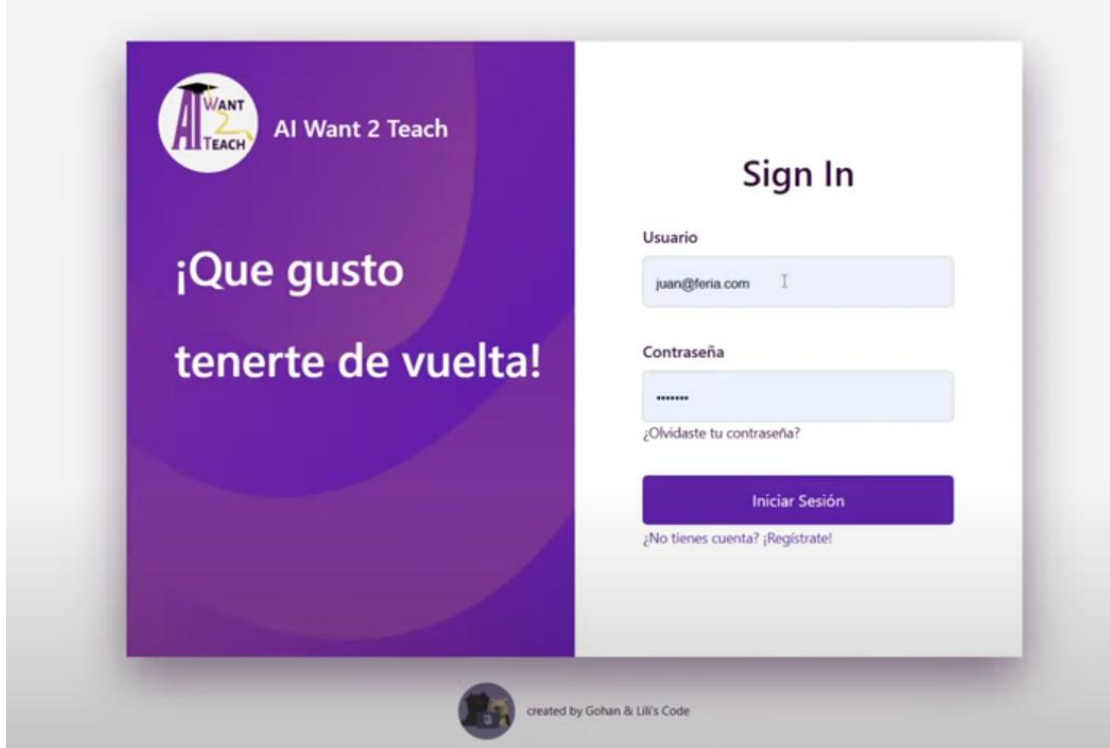


Figura 16: Pantalla de inicio.

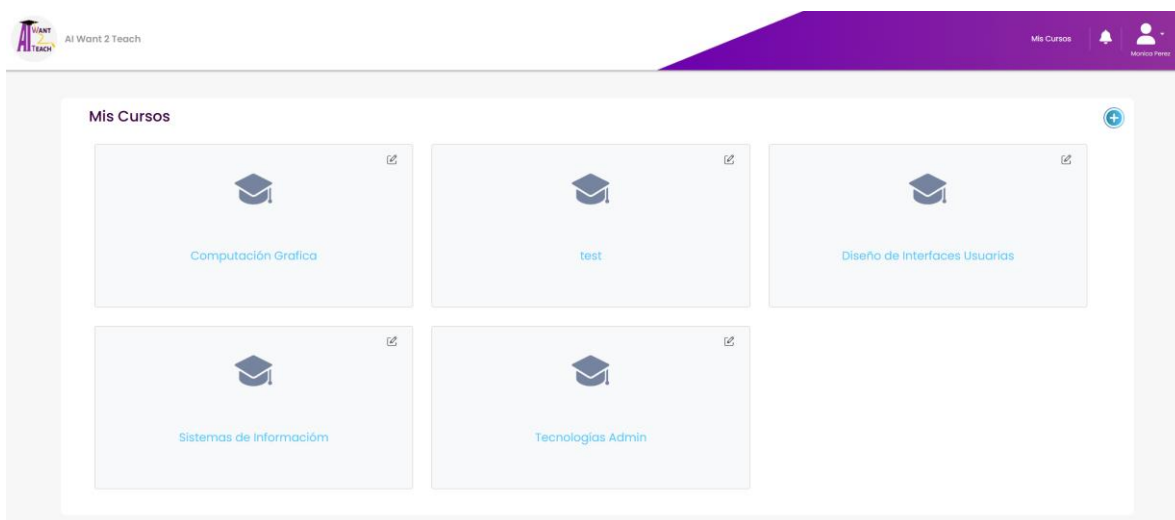


Figura 17: Dashboard.

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

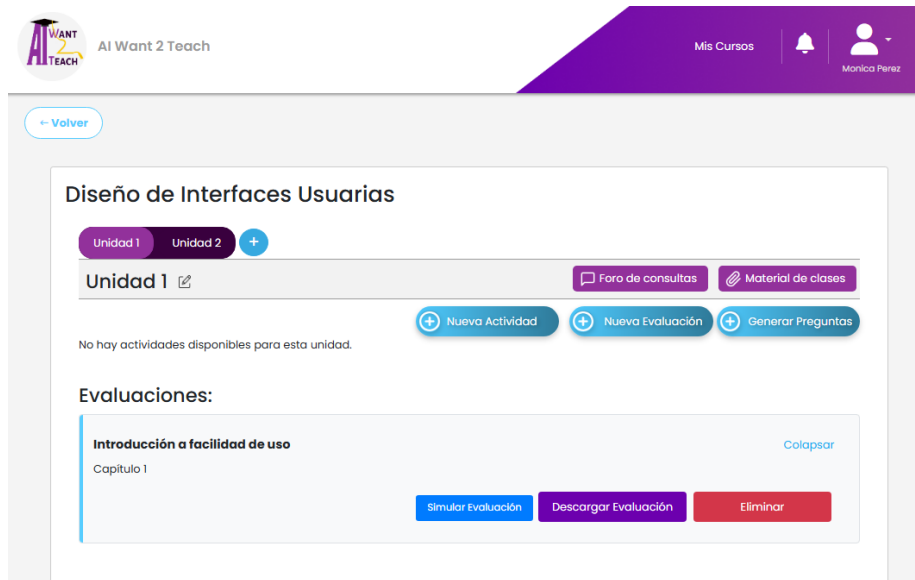


Figura 18: Módulo de cursos.

Nueva Evaluación

Título:
Título de la Evaluación

Descripción:
Descripción de la Evaluación

Cantidad de preguntas de Desarrollo:
Cantidad de preguntas de Desarrollo

Cantidad de preguntas de Alternativas:
Cantidad de preguntas de Alternativas

Cantidad de preguntas de Verdadero/Falso:
Cantidad de preguntas de Verdadero/Falso

Vista Previa Cancelar

Figura 19: Modulo para generar evaluaciones.

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

Evaluaciones:

Introducción a facilidad de uso

[Colapsar](#)

Capítulo 1

Simular Evaluación

Descargar Evaluación

Eliminar

Evaluación: Introducción a facilidad de uso

1. ¿Qué es un modelo mental y qué importancia tiene en el diseño de interfaces?

Puntaje: 0

2. ¿Cómo influyen las características del usuario en el diseño de interfaces de usuario?

Puntaje: 0

3. ¿Qué representa un escenario de tarea en el diseño de interfaces?

- A. Una tabla de cursos
- B. Un resumen académico
- C. Un proceso de inicio a fin visto desde los ojos del usuario
- D. Una interfaz gráfica con botones
- E. Una gráfica con datos de usuario

Puntaje: 0

4. ¿Cuáles son algunas de las necesidades de diseño cuando se considera a los usuarios de Internet en Chile según los materiales revisados?

- A. Consideración especial de las heurísticas de Nielsen
- B. Relación más directa con entornos empresariales
- C. Todas las anteriores
- D. Atención a diferencial de género en consumo online
- E. Mayor enfoque en diseño minimalista

Puntaje: 0

Figura 20: Evaluación generada.

REFERENCIAS

- Anderson, L. W., y Krathwohl, D. R. (Eds.). (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman.
- Banco Mundial. (2023). Revolución de la IA en la educación: Lo que necesita saber. <https://documents1.worldbank.org/curated/en/099734306182493324/pdf/IDU152823b13109c514ebd19c241a289470b6902.pdf>
- Benito, B., y Salinas, J. (2016). La metodología de investigación basada en diseño (Design-Based Research) en Tecnología Educativa. RED. Revista de Educación a Distancia, 50(4). <https://revistas.um.es/riite/article/view/260631/195691>
- Brooke, J. (1996). SUS: A quick and dirty usability scale. En P. W. Jordan, B. Thomas, B. A. Weerdmeester y A. L. McClelland (Eds.), Usability evaluation in industry (pp. 189–194). Taylor y Francis. https://www.researchgate.net/publication/228593520_SUS_A_quick_and_dirty_usability_scale
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... y Amodei, D. (2020). Language models are few-shot learners. arXiv. <https://doi.org/10.48550/arXiv.2005.14165>
- Caballero Toro, Á. (2023). Inteligencia artificial generativa en educación: Impulsando la creatividad y la interactividad en el aula. Universidad Miguel Hernández. <https://dspace.umh.es/handle/11000/32744>
- Caballero Toro, J. (2023). La inteligencia artificial en la educación: retos éticos y oportunidades pedagógicas. Revista Iberoamericana de Tecnología Educativa, 18(2), 45–61.
- Castilla-Gutiérrez, S., Colihuil-Catrileo, R., Bruneau-Chávez, J., y Lagos-Hernández, R. (2021). Carga laboral y efectos en la calidad de vida de docentes universitarios y de enseñanza media. Revista Chakiñan de Ciencias Sociales y Humanidades, 15, 166–179. <https://doi.org/10.37135/chk.002.15.11>
- Chan, C. H., y Tsi, K. Y. (2023). The AI revolution in education: Will AI replace or assist teachers in higher education? arXiv. <https://arxiv.org/abs/2305.01185>
- Díaz, J., Pérez, M., y Rodríguez, L. (2021). Aplicación de la inteligencia artificial en educación: una revisión sistemática. Revista Código Científico, 4(2), 45–60.
- Frías Ortega, C., Gutiérrez Monclus, P., Espinosa, A., Mella Díaz, S., Villarroel Encina, V., y Morrison Jara, R. (2018). Carga laboral académica y síndrome de burnout en docentes universitarios: Un análisis por género. [Tesis de licenciatura, Universidad de Chile]. <https://repositorio.uchile.cl/handle/2250/153568>

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

- García-Peñalvo, F. J., Llorens-Largo, F., y Vidal, G. (2024). Impacto de la inteligencia artificial generativa en los entornos de enseñanza-aprendizaje: Perspectivas docentes. *Education in the Knowledge Society*, 25 (1), e13903.
- García-Peñalvo, F. J., Llorens-Largo, F., y Vidal, J. (2024). La nueva realidad de la educación ante los avances de la inteligencia artificial generativa. *RIED: Revista Iberoamericana de Educación a Distancia*, 27 (1). <https://doi.org/10.5944/ried.27.1>
- Hernández, A., y Rodríguez, D. (2022). Evaluación de la usabilidad en entornos de aprendizaje basados en IA: un enfoque desde el usuario. *Revista Iberoamericana de Tecnología Educativa*, 18(3), 22–37.
- Holmes, W., Bialik, M., y Fadel, C. (2019). Artificial intelligence in education: Promises and implications for teaching and learning. Center for Curriculum Redesign.
- International Organization for Standardization. (2019). Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems (ISO 9241-210:2019). ISO.
- ISO. (2018). Ergonomics of human-system interaction. Part 11: Usability: Definitions and concepts (ISO 9241-11:2018). International Organization for Standardization.
- Lindner, M. A., et al. (2023). Challenges in automated item generation: Limitations of GPT-based approaches. *Educational Technology Research and Development*, 71, 1423–1440. <https://doi.org/10.1007/s11423-023-10257-6>
- Luckin, R., Holmes, W., Griffiths, M., y Forcier, L. B. (2016). Intelligence unleashed: An argument for AI in education. Pearson. <https://oro.open.ac.uk/50104/>
- Lundgren, M. (2024). Large language models in student assessment: Comparing ChatGPT and human graders. SSRN. <https://ssrn.com/abstract=4874359>
- Nielsen, J. (1994). Usability engineering. Academic Press.
- OpenAI. (2023). GPT-4 technical report. arXiv. <https://doi.org/10.48550/arXiv.2303.08774>
- Parra, D., y Felli, V. E. A. (2017). Work process of nursing professors. *Revista Latino-Americana de Enfermagem*, 25. <https://doi.org/10.1590/1518-8345.1941.2946>
- Pérez, C., y Martínez, L. (2017). Condiciones de trabajo y estrés en académicos universitarios. *Revista Chilena de Salud Pública*, 21(1), 49–58. https://www.scielo.cl/scielo.php?pid=S0718-24492017000100049&script=sci_arttext
- Pérez Morales, J. I. (2007). La evaluación como instrumento de mejora de la calidad del aprendizaje: Propuesta de intervención psicopedagógica para el aprendizaje del idioma inglés. [Tesis doctoral, Universidad Autónoma de Barcelona]. Currículum Nacional. https://www.curriculumnacional.cl/estudiante/621/articles-216096_recurso_pdf.pdf

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

- Popenici, S. A. D., y Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, 12*(1), 1–13. <https://doi.org/10.1186/s41039-017-0062-8>
- Porayska-Pomsta, K., Holmes, W., y Nemorin, S. (2024a). Ethical and practical challenges of AI in education. En B. Luckin (Ed.), **AI and education: The importance of teacher agency*. Routledge.
- Porayska-Pomsta, K., Holmes, W., y Nemorin, S. (2024b). The ethics of AI in education. arXiv. <https://arxiv.org/abs/2406.11842>
- Poyatos, R., y Pujol, D. (2025). Reflexión pedagógica y tecnologías emergentes: IA en el aula universitaria. *Revista de Innovación Educativa*, 29(1), 55–72.
- Retamozo Ávalos, A., Astocondor Gonzales, C., Juárez Tamayo, N., y Ramos Cevallos, M. (2024). Perspectivas actuales sobre el estrés y su relación con el desempeño docente: una revisión sistemática. *Mendive: Revista de Educación*, 22(2), e3635. <https://mendive.upr.edu.cu/index.php/MendiveUPR/article/view/3635>
- Rodríguez, M., y Soto, P. (2023). Relación del estrés laboral y habilidades emocionales en docentes universitarios. *Revista Ciencias de la Actividad Física UCM*, 25(1), 11–20. <https://www.scielo.cl/pdf/rcaf/v25n1/0719-4013-rcaf-25-01-11.pdf>
- Santos Guerra, M. Á. (2003). *La evaluación: Un proceso de diálogo, comprensión y mejora* Morata. https://www.cucs.udg.mx/avisos/Martha_Pacheco/Software%20e%20hipertexto/Antologia_Electronica_pa121/Santos%20G.Eval.PDF
- Seldon, A., y Abidoye, O. (2018). *The fourth education revolution: Will artificial intelligence liberate or infantilise humanity?* University of Buckingham Press. <https://www.buckingham.ac.uk/research/publications/the-fourth-education-revolution>
- Tan, B., Armoush, N., Mazzullo, E., y Gierl, M. J. (2024). A review of automatic item generation techniques leveraging large language models. ResearchGate. <https://www.researchgate.net/publication/379664254>
- UNESCO. (2021). *Recomendación sobre la ética de la inteligencia artificial*. <https://www.unesco.org/es/artificial-intelligence/recommendation-ethics>
- UNESCO. (2024a). *Artificial intelligence and the futures of learning*. <https://www.unesco.org/en/digital-education/ai-future-learning>
- UNESCO. (2024b). *Guía para el uso de IA generativa en educación e investigación*. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. <https://www.unesco.org/es/digital-education/ai-future-learning>

DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA DE IA GENERATIVA PARA LA CREACIÓN DE EVALUACIONES ACADÉMICAS

- U.S. Department of Education. (2023). Artificial intelligence and the future of teaching and learning: Insights and recommendations. Office of Educational Technology. <https://www.ed.gov/sites/ed/files/documents/ai-report/ai-report.pdf>
- Wang, F., y Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development*, 53(4), 5–23. <https://doi.org/10.1007/BF02504682>
- Yan, L., et al. (2023). Practical and ethical challenges of large language models in education: A systematic scoping review. *arXiv*. <https://arxiv.org/abs/2303.13379>
- Zúñiga-Jara, S., y Pizarro-León, M. (2021). Síndrome de burnout en docentes de ciencias de la salud de universidades en Chile durante la pandemia de COVID-19. *Revista de Salud Pública*, 23(4), 495–503. https://scielo.isciii.es/scielo.php?pid=S3020-11602021000400495&script=sci_arttext