

**UNIVERSIDAD NACIONAL DE SAN CRISTÓBAL
DE HUAMANGA**

FACULTAD DE INGENIERÍA DE MINAS, GEOLOGÍA Y CIVIL

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS



TESIS:

“Implementación de un chatbot para la optimización de la atención de consultas académicas y administrativas en la Escuela Profesional de Ingeniería de Sistemas de la UNSCH, 2025”

Para optar el título profesional de:
INGENIERA DE SISTEMAS

PRESENTADO POR:

Bach. Melanie Stephany HERRERA SANTA CRUZ

ASESORA:

Ing. Celia Edith MARTÍNEZ CÓRDOVA

AYACUCHO - PERÚ

2026

DEDICATORIA

A Dios, por brindarme la fortaleza y la fe necesarias para seguir adelante en cada momento de dificultad.

A mi padre, por su amor incondicional, su sacrificio y su acompañamiento constante a lo largo de toda mi formación profesional.

A mi novio Jaime, por su comprensión, motivación y valiosa colaboración durante todo el proceso de investigación.

AGRADECIMIENTO

A mi asesora, la Ing. Celia Edith Martínez Córdova, por su orientación en las diferentes etapas de la investigación.

Al personal administrativo de la EPIS y la FIMGC de la UNSCH, por su disposición y colaboración durante la fase exploratoria de esta investigación, al facilitar información clave sobre las consultas más frecuentes de los estudiantes.

A mi hermana, quien con su compañía y palabras de aliento me motivó a seguir adelante y no rendirme.

ÍNDICE DE CONTENIDOS

DEDICATORIA.....	2
AGRADECIMIENTO.....	3
ÍNDICE DE TABLAS.....	9
ÍNDICE DE FIGURAS.....	10
RESUMEN.....	11
ABSTRACT.....	12
INTRODUCCIÓN	13
CAPÍTULO I PLANTEAMIENTO DEL PROBLEMA	16
1.1 Descripción de la realidad problemática.....	16
1.2 Formulación del problema	17
1.2.1 Problema general	17
1.2.2 Problemas específicos	18
1.3 Objetivos de la investigación	18
1.3.1 Objetivo general.....	18
1.3.2 Objetivos específicos	18
1.4 Hipótesis de la investigación	19
1.4.1 Hipótesis general.....	19
1.4.2 Hipótesis específicas.....	19
1.5 Justificación de la investigación	19
1.5.1 Justificación teórica	19
1.5.2 Justificación práctica.....	20
1.5.3 Justificación tecnológica.....	20
1.5.4 Justificación social	21

1.6	Delimitación de la investigación.....	21
1.6.1	Delimitación espacial.....	21
1.6.2	Delimitación temporal	21
1.7	Importancia de la investigación	22
1.8	Limitaciones de la investigación.....	23
CAPÍTULO II MARCO TEÓRICO		24
2.1	Antecedentes de la investigación	24
2.1.1	Antecedentes internacionales.....	24
2.1.2	Antecedentes nacionales	25
2.2	Bases teóricas.....	27
2.2.1	Chatbot.....	27
2.2.2	Funcionalidad.....	28
2.2.3	Fiabilidad	29
2.2.4	Usabilidad	30
2.2.5	Optimización de la atención de consultas académicas y administrativas.	30
2.2.6	Eficacia	31
2.2.7	Eficiencia	32
2.2.8	Satisfacción del usuario	33
2.3	Definiciones básicas.....	33
2.3.1	Procesamiento de lenguaje natural (NLP)	33
2.3.2	Inteligencia artificial generativa	34
2.3.3	<i>Fuzzy matching</i>	35
2.3.4	Modelos de lenguaje grande (LLM)	35
2.3.5	FastAPI	36
2.3.6	Vue.js	37

2.3.7	Supabase	38
2.3.8	Tavily API	38
2.3.9	Norma ISO/IEC 25010	39
CAPÍTULO III MATERIAL Y MÉTODOS		41
3.1	Tipo de investigación	41
3.2	Nivel de investigación.....	41
3.3	Diseño de investigación	42
3.4	Población y muestra.....	43
3.4.1	Población.....	43
3.4.2	Muestra	44
3.5	Variables y dimensiones.....	45
3.5.1	Variable independiente.....	45
3.5.2	Dimensiones de la variable independiente.....	45
3.5.3	Variable dependiente	46
3.5.4	Dimensiones de la variable dependiente.....	46
3.6	Técnicas e instrumentos de recolección de datos	46
3.6.1	Técnicas para recolectar datos	46
3.6.2	Instrumentos para recolectar datos.....	47
3.7	Técnicas de procesamiento y análisis de datos	47
3.7.1	Técnicas de procesamiento de datos	47
3.7.2	Herramientas de procesamiento de datos.....	50
3.8	Requisitos de infraestructura.....	53
3.9	Arquitectura del chatbot.....	54
3.9.1	Interacción multimodal del sistema	61
3.10	Componentes del sistema.....	61

3.10.1	<i>Frontend</i> : interfaz de usuario	61
3.10.2	<i>Backend</i> : lógica de procesamiento	62
3.10.3	Base de datos: persistencia y analítica	63
CAPÍTULO IV RESULTADOS Y DISCUSIÓN		64
4.1	Caracterización de la muestra	64
4.2	Análisis de fiabilidad del instrumento	64
4.3	Prueba de normalidad	66
4.4	Estadísticos descriptivos	68
4.4.1	Análisis por ítem	68
4.4.2	Análisis por dimensión	71
4.5	Distribución de respuestas por categoría	74
4.6	Análisis inferencial: prueba de Wilcoxon	78
4.7	Discusión.....	82
4.8	Análisis de las pruebas del chatbot	86
4.8.1	Resultados generales	87
4.8.2	Precisión por nivel de la arquitectura.....	88
4.8.3	Análisis del tiempo de respuesta.....	90
4.8.4	Score de similitud semántica	92
4.8.5	Análisis de fallos.....	94
4.8.6	Precisión por categoría temática	96
4.8.7	Síntesis del análisis de pruebas	98
CAPÍTULO V CONCLUSIONES Y RECOMENDACIONES		100
5.1	Conclusiones	100
5.2	Recomendaciones	101
BIBLIOGRAFÍA		103

ANEXOS	108
MATRIZ DE CONSISTENCIA.....	108
MATRIZ DE OPERACIONALIZACIÓN DE VARIABLES.....	110
IMÁGENES DE LA INTERFAZ DEL CHATBOT Y DEL EVALUADOR	111
CUESTIONARIO BASADO EN LA ISO/IEC 25010 (PRETEST).....	114
CUESTIONARIO BASADO EN LA ISO/IEC 25010 (POSTEST).....	116
GUÍA DE ENTREVISTA.....	118

ÍNDICE DE TABLAS

Tabla 1	Etapas del proceso de recolección y procesamiento de datos	47
Tabla 2	Fase de preprocesamiento de datos	48
Tabla 3	Fase de preparación de la base de conocimiento.....	49
Tabla 4	Fase de evaluación e interpretación	49
Tabla 5	Herramientas tecnológicas para el procesamiento de datos	50
Tabla 6	Requisitos de infraestructura para la construcción del chatbot	53
Tabla 7	Coefficiente alfa de Cronbach por dimensión e instrumento completo	65
Tabla 8	Prueba de normalidad Shapiro-Wilk por dimensión y medición (N = 220)	67
Tabla 9	Estadísticos descriptivos por ítem: pretest y postest (N = 220)	68
Tabla 10	Estadísticos descriptivos dimensionales: pretest y postest (N = 220).....	72
Tabla 11	Porcentajes de acuerdo y desacuerdo por dimensión: pretest y postest	76
Tabla 12	Resultados de la prueba de Wilcoxon para muestras relacionadas (N = 220).....	79
Tabla 13	Resultados de evaluación por nivel del chatbot (N = 600)	88
Tabla 14	Causas de fallo en las consultas no resueltas satisfactoriamente (N = 94)	94
Tabla 15	Precisión del chatbot por categoría temática de consulta.....	96

ÍNDICE DE FIGURAS

Figura 1 Estructura del proyecto.....	56
Figura 2 Arquitectura del Chatbot EPIS-UNSCH.....	57
Figura 3 Caracterización de la muestra por género y semestre académico (N = 220)	64
Figura 4 Coeficiente alfa de Cronbach por dimensión en pretest y posttest.....	66
Figura 5 Medias por ítem en el pretest (sistema tradicional) y posttest (chatbot)	70
Figura 6 Puntajes dimensionales (Variable Dependiente)	73
Figura 7 Puntajes dimensionales (Variable Independiente).....	74
Figura 8 Distribución porcentual de respuestas por ítem (%)	75
Figura 9 Porcentaje de acuerdo por dimensión.....	78
Figura 10 Distribución de puntajes por dimensión (Variable Dependiente).....	82
Figura 11 Precisión global del chatbot (N=600).....	88
Figura 12 Precisión por nivel del chatbot	90
Figura 13 Distribución de tiempos de respuesta del chatbot (N = 600)	91
Figura 14 Score promedio de similitud semántica por nivel	93
Figura 15 Causas de fallo en las consultas (N = 94 preguntas fallidas)	95
Figura 16 Precisión por categoría temática.....	97
Figura 17 Vista inicial del chatbot (PC).....	111
Figura 18 Vista de la interacción con el chatbot (PC)	111
Figura 19 Vista inicial del chatbot (Móviles)	112
Figura 20 Vista del Reporte del Evaluador de Consultas del Chatbot.....	113

RESUMEN

La Escuela Profesional de Ingeniería de Sistemas (EPIS) de la Universidad Nacional de San Cristóbal de Huamanga (UNSCH) presenta limitaciones en la atención de consultas académicas y administrativas, debido a la dispersión de canales de comunicación y la dependencia de la disponibilidad del personal, lo que genera demoras y restringe el acceso rápido a la información. De esta forma, esta investigación tuvo la finalidad de evaluar la implementación de un chatbot para mejorar la atención de consultas.

El estudio aplica un diseño de investigación preexperimental, de nivel explicativo, prospectivo y longitudinal. Además, se utiliza un único grupo de medición. Tanto para el pretest, como para el postest ($O_1 \rightarrow X \rightarrow O_2$). La muestra estuvo conformada por 220 estudiantes matriculados en el semestre 2025-II. En el tema de recolección de datos, se realizó un cuestionario basado en la norma ISO/IEC 25010; que engloba las dimensiones de ambas variables (VI y VD), que son: funcionalidad, fiabilidad, usabilidad, eficacia, eficiencia y satisfacción.

La arquitectura del chatbot es original de la autora. Ésta incluye 4 niveles e incorpora búsqueda exacta, coincidencia aproximada al texto, consulta web en tiempo real y generación de nuevas respuestas con IA. Las tecnologías utilizadas fueron FastAPI, Supabase, Railway y Vue.js.

Los resultados más significativos mostraron diferencias agrupadas estadísticamente significativas ($p < .001$) en todas las dimensiones, además de tamaños del efecto grandes ($r \geq 0.68$), lo que puede ser interpretado como un efecto concreto. Además, la evaluación automatizada por medio de consultas de estudiantes mostró ser capaz de realizar un 84,3 % de las decisiones correctas en el total de las consultas, con un tiempo medio de respuesta de 7.50 segundos, mostrando un rendimiento aceptable a nivel técnico.

Palabras clave: chatbot, arquitectura multinivel, PLN, ISO/IEC 25010 y UNSCH.

ABSTRACT

The Professional School of Systems Engineering (EPIS) of the National University of San Cristóbal de Huamanga (UNSCH) presents limitations in the attention of academic and administrative queries, due to the dispersion of communication channels and the dependence on the availability of personnel, which generates delays and restricts quick access to information. In this way, this research had the purpose of evaluating the implementation of a chatbot to improve the attention of queries.

The study applies a pre-experimental, explanatory, prospective and longitudinal research design. Furthermore, a single measurement group is used. For both the pretest and the posttest ($O_1 \rightarrow X \rightarrow O_2$). The sample was made up of 220 students enrolled in the 2025-II semester. Regarding data collection, a questionnaire was carried out based on the ISO/IEC 25010 standard; which encompasses the dimensions of both variables (VI and VD), which are: functionality, reliability, usability, effectiveness, efficiency and satisfaction.

The chatbot architecture is original to the author. This includes 4 levels and incorporates exact search, fuzzy text matching, real-time web query and generation of new answers with AI. The technologies used were FastAPI, Supabase, Railway, and Vue.js.

The most significant results showed statistically significant pooled differences ($p < .001$) across all dimensions, as well as large effect sizes ($r \geq 0.68$), which can be interpreted as a concrete effect. Furthermore, the automated evaluation using student queries demonstrated an 84.3% accuracy rate in making correct decisions across all queries, with an average response time of 7.50 seconds, indicating acceptable technical performance.

Keywords: chatbot, multi-level architecture, NLP, ISO/IEC 25010, and UNSCH.

INTRODUCCIÓN

La transformación digital ha modificado de manera significativa la forma en que las instituciones educativas gestionan la información y se relacionan con su comunidad académica.

En este panorama, los chatbots han emergido como una herramienta tecnológica de alto potencial, cuyo objetivo principal es el automatizar la atención de consultas frecuentes, disminuir tiempos de respuesta y dar una mejor experiencia de usuario. Muchos estudios han mostrado que los sistemas conversacionales implementados con inteligencia artificial, posibilitan una asistencia más rápida, continua y adaptada a las necesidades del usuario, lo que facilita el acceso a la información y mejora los procesos de soporte en entornos educativos (Winkler & Söllner, 2018; Shawar & Atwell, 2007). Sin embargo, en gran parte de las universidades públicas peruanas, generalmente en aquellas que están ubicadas en regiones descentralizadas como Ayacucho, la incorporación de este tipo de soluciones sigue siendo escasa, por lo que pone en manifiesto una brecha pendiente de modernización de servicios académicos.

La Escuela Profesional de Ingeniería de Sistemas (EPIS) de la Universidad Nacional de San Cristóbal de Huamanga (UNSCH) también comparte esta realidad. En la actualidad, los estudiantes dependen principalmente de la atención presencial para atender consultas vinculadas a trámites administrativos, requisitos académicos, horarios e información normativa. Además, suelen recurrir a canales no oficiales como búsquedas en internet, redes sociales como Facebook, mensajes de grupos de WhatsApp o llamadas telefónicas a conocidos. Esto evidencia de manera clara, la ausencia de un sistema centralizado, confiable y permanentemente disponible.

Por tanto, esta situación dificulta el acceso oportuno a la información, sobre todo cuando el personal administrativo no se encuentra en las oficinas, ya sea porque se encuentra fuera de horario o están en diligencias propias de los procesos administrativos. Además de ello, la

problemática se agrava más cuando los estudiantes viven fuera de la ciudad, por lo que es necesario que tengan que desplazarse para obtener información que perfectamente podría proporcionarse de forma automatizada.

Ante esta realidad, la presente investigación plantea el desarrollo e implementación de un chatbot inteligente destinado a optimizar la atención de consultas académicas y administrativas en la EPIS-UNSCH. Este sistema se apoya en una arquitectura multinivel que combina búsqueda exacta, coincidencia aproximada de texto, consultas web en tiempo real y generación de respuestas mediante inteligencia artificial, tomando como base de conocimiento los documentos institucionales oficiales. El sistema tiene un número de formas de interactuar (texto, voz o imagen), lo que hace más fácil llegar a la información.

El sistema tiene una estructura que responde a preguntas difíciles, propiciando una mayor precisión y calidad. El funcionamiento y diseño del sistema se reflejan en el Capítulo III.

Para conocer el funcionamiento de la solución, se realizó un estudio en el que participaron 220 estudiantes del semestre 2025-II. A los estudiantes que participaron se les pidió que respondieran encuestas antes y después de usar la solución. Las encuestas fueron orientadas por la norma ISO/IEC 25010, que evalúa si el sistema es fácil de usar, funciona como se espera, es rápido y si los usuarios son felices con la experiencia.

También se tomaron medidas del desempeño del sistema con datos objetivos, como las métricas de la precisión de las respuestas o el tiempo de respuesta, las cuales estaban basadas en las conversaciones de los estudiantes con el chatbot.

Este documento tiene un total de cinco capítulos. En el Capítulo I se explica la problemática que se analiza, los objetivos, las hipótesis e importancias del estudio. El Capítulo II desarrolla el marco teórico que sustenta el estudio. El Capítulo III describe la metodología empleada y el diseño

del sistema propuesto. El Capítulo IV presenta los resultados obtenidos y su discusión. Finalmente, el Capítulo V expone las conclusiones y recomendaciones derivadas de la investigación.

CAPÍTULO I

PLANTEAMIENTO DEL PROBLEMA

1.1 Descripción de la realidad problemática

En el marco de la transformación digital que atraviesan las instituciones de educación superior a nivel mundial, las universidades públicas peruanas enfrentan el reto de modernizar sus servicios académicos y administrativos con recursos limitados. La utilización de tecnologías relacionadas con inteligencia artificial, como las aplicaciones de atención automatizada mediante chatbots, son una opción viable para poder brindar un buen servicio y sin gestionar la carga laboral del personal.

En la Escuela Profesional de Ingeniería de Sistemas (EPIS) de la Universidad Nacional de San Cristóbal de Huamanga (UNSCH), la atención de las consultas académicas y administrativas destaca por no disponer de un canal de atención que favorezca centralizar la información institucional de manera accesible y actualizada. Cuando los estudiantes necesitan resolver dudas sobre trámites, normativas, requisitos, horarios o procedimientos, habitualmente acuden a la atención presencial o utilizan otros medios como búsquedas en internet, redes sociales o llamadas telefónicas a contactos de confianza. Esta dispersión de fuentes produce inconsistencias, lo que complica la verificación de la información y restringe el acceso oportuno a respuestas confiables.

Esta forma de gestión de información arrastra consigo diversas limitaciones que repercuten negativamente en la experiencia de los usuarios. Primero porque la información institucional está repartida entre distintos canales, por lo que la tarea de identificar una fuente única y confiable se vuelve difícil. Esta división de información se puede traducir en inconsistencias o desactualización en las respuestas que reciben los estudiantes. Segundo porque la disponibilidad del personal administrativo está anclada a un horario laboral, lo que limita a realizar las consultas a sólo dicho

horario. Y tercero es que, un número importante de estudiantes de la EPIS proviene de distritos o provincias distantes a Ayacucho, por lo que trasladarse físicamente para obtener información básica supone un costo adicional de tiempo y recursos que podría evitarse con soluciones digitales de atención automatizada.

Frente a esta situación, resulta necesario contar con un mecanismo alternativo de atención que centralice la información institucional y ofrezca respuestas accesibles, confiables y disponibles en cualquier momento. Este sistema debería funcionar desde cualquier dispositivo con acceso a internet y resolver consultas frecuentes de forma automática, disminuyendo la dependencia de la atención presencial y facilitando el acceso oportuno a la información académica y administrativa.

Como respuesta a esta necesidad, la investigación propone diseñar, desarrollar e implementar un chatbot inteligente para la EPIS-UNSCH. El sistema se sustenta en una arquitectura multinivel que integra recuperación de información estructurada, coincidencia aproximada de texto, búsqueda web en tiempo real e inteligencia artificial generativa, con el propósito de ofrecer respuestas contextualizadas a partir de los documentos institucionales oficiales.

Mediante un diseño preexperimental con medición pretest y posttest, se busca demostrar que la incorporación de esta herramienta tecnológica contribuye a optimizar la atención de consultas académicas y administrativas en la EPIS-UNSCH, mejorando los niveles de eficacia, eficiencia y satisfacción de los usuarios.

1.2 Formulación del problema

1.2.1 Problema general

¿En qué medida la implementación de un chatbot optimiza la atención de consultas académicas y administrativas en la Escuela Profesional de Ingeniería de Sistemas de la UNSCH, 2025?

1.2.2 Problemas específicos

1. ¿En qué medida la funcionalidad del chatbot optimiza la atención de consultas académicas y administrativas?
2. ¿En qué medida la fiabilidad del chatbot optimiza la atención de consultas académicas y administrativas?
3. ¿En qué medida la usabilidad del chatbot optimiza la atención de consultas académicas y administrativas?

1.3 Objetivos de la investigación

1.3.1 Objetivo general

Evaluar el efecto de implementar un chatbot que optimice la atención de consultas académicas y administrativas en la Escuela Profesional de Ingeniería de Sistemas de la UNSCH, 2025.

1.3.2 Objetivos específicos

1. Evaluar si la funcionalidad del chatbot optimiza la atención de consultas académicas y administrativas.
2. Evaluar si la fiabilidad del chatbot optimiza la atención de consultas académicas y administrativas.
3. Evaluar si la usabilidad del chatbot optimiza la atención de consultas académicas y administrativas.

1.4 Hipótesis de la investigación

1.4.1 Hipótesis general

La implementación de un chatbot optimiza la atención de consultas académicas y administrativas en la Escuela Profesional de Ingeniería de Sistemas de la UNSCH, 2025.

1.4.2 Hipótesis específicas

1. La funcionalidad del chatbot optimiza la atención de consultas académicas y administrativas.
2. La fiabilidad del chatbot optimiza la atención de consultas académicas y administrativas.
3. La usabilidad del chatbot optimiza la atención de consultas académicas y administrativas.

1.5 Justificación de la investigación

La atención a las consultas académicas y administrativas es un elemento clave de la vida universitaria, ya que incide de forma directa en la capacidad del alumno/a de acceder a información en el tiempo adecuado y en el de tomar decisiones bien fundamentadas acerca de su trayectoria académica. En la Escuela Profesional de Ingeniería de Sistemas (EPIS) de la UNSCH, la atención a los alumnos(as) se apoya casi por completo en la presencia física del personal administrativo, lo que limita mucho el acceso a la misma, y establece un desajuste entre las necesidades informativas de los alumnos(as) y la capacidad de respuesta de la misma.

1.5.1 Justificación teórica

La investigación aporta al conocimiento acerca de la implementación de la inteligencia artificial en la mejora de los servicios de atención al alumnado en instituciones de educación superior, muy particularmente en las universidades públicas ubicadas en lugares descentralizados

de Perú. La investigación conduce en su esencia a producir evidencia empírica del potencial que presentan los sistemas conversacionales en la mejora del acceso a la información académica y administrativa, favoreciendo de esta forma la disponibilidad y la oportunidad de la atención a las consultas de los/las alumnos(as). La evaluación del sistema se hace de acuerdo con los criterios de la norma ISO/IEC 25010, lo cual permite analizar la calidad de la solución desde dimensiones reconocidas a nivel internacional, como funcionalidad, fiabilidad, usabilidad, eficacia, eficiencia y satisfacción.

1.5.2 Justificación práctica

La necesidad de implementar un chatbot para la Escuela Profesional de Ingeniería de Sistemas de UNSCH es concreta, porque permite mediante la automatización, ofrecer respuestas a las consultas frecuentes, así como amortiguar y aliviar la carga de trabajo y operacional del personal administrativo así como ampliar la disponibilidad del servicio, permitiendo al estudiante poseer un acceso a la información académica y administrativa en cualquier momento y en todos los lugares, al mismo tiempo que posibilita el acceso a información institucional de manera oportuna sin que el estudiante tenga que trasladarse de forma física a la oficina de atención, situación que llega a ser más apreciada por estudiantes provenientes de distritos y provincias tan alejadas de Ayacucho, que muchas veces no poseen el tiempo ni los recursos necesarios.

1.5.3 Justificación tecnológica

Esta investigación propone una solución tecnológica que fusiona las herramientas modernas para el procesamiento del lenguaje natural junto a generación de respuestas contextualizadas, como son FastAPI, Google Gemini, Tavily API, Supabase, Vue.js en un entorno institucional real, lo que demuestra que es posible la producción de sistemas de atención automatizada intrínsecamente ligados a la inteligencia artificial utilizando tecnologías asequibles

de bajo coste que hacen posible el poder implementarlos y utilizarlos en universidades públicas que tienen recursos limitados, de forma que la propuesta sienta un precedente tecnológico replicable en otras escuelas profesionales de la UNSCH y en otras Instituciones de educación superior que presentan características similares.

1.5.4 Justificación social

La investigación contribuye a reducir las inequidades en el acceso a la información institucional dentro de la comunidad universitaria. Al ofrecer un sistema disponible las veinticuatro horas, accesible desde dispositivos móviles e interactivo en lenguaje natural, se facilita la consulta de información académica y administrativa de manera oportuna y sin barreras de lugar. Esta disponibilidad beneficia especialmente a quienes enfrentan mayores obstáculos geográficos, económicos o de disponibilidad horaria, promoviendo una mayor equidad en el acceso a los servicios informativos de la institución.

1.6 Delimitación de la investigación

1.6.1 Delimitación espacial

La presente investigación se desarrolla en la Escuela Profesional de Ingeniería de Sistemas (EPIS) de la Universidad Nacional de San Cristóbal de Huamanga (UNSCH), ubicada en el distrito de Ayacucho, provincia de Huamanga, región Ayacucho, Perú. El estudio se enfoca en los estudiantes matriculados en dicha Escuela, quienes constituyen el ámbito institucional en el que se implementa y evalúa el sistema propuesto.

1.6.2 Delimitación temporal

Esta investigación se llevó a cabo durante el semestre académico 2025-II, período en el que se completó el diseño e implementación del chatbot. La recogida de datos por medio de los instrumentos pretest y posttest fue ejecutada con estudiantes que se matricularon en la EPIS-

UNSCH durante ese semestre, mediante la realización de una encuesta entre finales de diciembre de 2025 y los primeros meses de 2026.

1.7 Importancia de la investigación

La importancia de esta investigación se evidenció en diversos planos. En el plano institucional, contribuyó a mejorar los mecanismos de atención a las consultas académicas y administrativas en la Escuela Profesional de Ingeniería de Sistemas de la UNSCH, impulsando el uso de herramientas digitales que permitan ofrecer información de forma más accesible, rápida y organizada para los estudiantes.

En lo académico, la investigación genera evidencia empírica sobre la efectividad de una arquitectura multinivel para la atención automatizada de consultas académicas y administrativas. En contextos de universidades públicas, supone también un referente metodológico que puede orientar muchas investigaciones futuras en el ámbito de la inteligencia artificial aplicada a la educación superior en el Perú.

En una dimensión social, la solución brindada favorece la disminución de las barreras que se interponen en el acceso a la información institucional, con beneficios directos para los 513 estudiantes que se matricularon durante el periodo 2025-II, en particular para aquellos que tienen problemas geográficos o temporales para recibir atención en la modalidad presencial.

En lo tecnológico, la investigación pone de manifiesto que es posible diseñar e implementar soluciones de inteligencia artificial funcionales y de bajo costo en universidades públicas de regiones descentralizadas, y constituir un precedente replicable en otras escuelas o facultades de la UNSCH o en instituciones semejantes.

1.8 Limitaciones de la investigación

Cuando se desarrolló el chatbot para brindar atención a las consultas académicas y administrativas en la EPIS-UNSCH se reconocieron las siguientes limitaciones:

- **Limitación de alcance y generalización.** La muestra estuvo conformada por estudiantes matriculados en la Escuela Profesional de Ingeniería de Sistemas de la UNSCH durante el semestre académico 2025-II. Por ello, los resultados no son directamente extrapolables a otras escuelas profesionales ni a períodos académicos distintos sin previamente replicar estudios similares en esos contextos.
- **Limitación por dispersión y falta de digitalización de la información institucional.** En la universidad, distintas áreas manejan información relevante para los estudiantes, tales como cronogramas de trámites, requisitos administrativos o normativas internas. Sin embargo, gran parte de esta información aún no se encuentra sistematizada ni integrada en un repositorio institucional centralizado. Esta situación limita el alcance del chatbot, ya que el sistema no puede responder con precisión a consultas que dependen de información que no está digitalizada o disponible de manera estructurada.
- **Limitación por dependencia de servicios externos.** El funcionamiento del sistema en los niveles de búsqueda web y generación de respuestas con inteligencia artificial depende de plataformas externas, como Tavily API y Google Gemini. Por ello, posibles interrupciones del servicio, cambios en sus condiciones de uso o problemas de disponibilidad podrían afectar temporalmente la capacidad de respuesta del chatbot.
- **Limitación del instrumento de evaluación.** La evaluación del sistema se realizó mediante encuestas de percepción basadas en los criterios de la norma ISO/IEC 25010, las cuales recogen la valoración subjetiva de los usuarios. Aunque este método es adecuado para medir

aspectos como usabilidad y satisfacción, los resultados pueden verse influenciados por factores como la novedad de la herramienta o la actitud de los usuarios frente al uso de nuevas tecnologías.

CAPÍTULO II

MARCO TEÓRICO

2.1 Antecedentes de la investigación

2.1.1 Antecedentes internacionales

Serrano Franco y Zamudio García (2025) desarrollaron un chatbot con el objetivo de complementar los procesos de tutoría y atender consultas frecuentes de estudiantes de los programas de Ingeniería en Tecnologías de la Información e Ingeniería en Animación y Efectos Visuales de la Universidad Politécnica Metropolitana de Hidalgo, México. La investigación adoptó un enfoque cuantitativo con diseño no experimental y transversal, aplicado a una muestra de 213 estudiantes seleccionados mediante muestreo probabilístico. Los resultados permitieron mejorar el funcionamiento del chatbot a partir de la retroalimentación de los usuarios, con ganancias en la eficiencia para resolver dudas académicas y en la calidad del servicio personalizado. Esta investigación pone de manifiesto el potencial de los chatbots para optimizar la atención de consultas en programas universitarios de ingeniería.

Múnera, Salazar y Osorio (2022) llevaron a cabo un estudio con una finalidad: evaluar el uso de un chatbot como forma de realizar la atención a las consultas de este tipo de estudiantes en la modalidad virtual de la Escuela Interamericana de Bibliotecología con sede en Colombia. Este trabajo, que fue publicado en la revista Investigación Bibliotecológica, se ocupó de evaluar la percepción de los usuarios sobre la utilidad y la facilidad de uso del sistema, a partir de la implementación de instrumentos para evaluar la experiencia de usuario, y fue aceptado por las

críticas. Como resultado de la investigación, mostramos como los estudiantes valoran positivamente el uso del chatbot debido a la posibilidad de tener una atención permanente y a la rapidez en la respuesta. Esta investigación da cuenta de la aceptación de los chatbots como herramienta de soporte frente a la atención de consultas en la educación superior.

La obra colectiva coordinada por Fernández-Ferrer (2023) titulada Chatbots en educación: tendencias actuales y retos en el futuro, publicada por el grupo Learning, Media & Social Interactions de la Universitat de Barcelona, estudia las aplicaciones de los chatbots en el escenario universitario; evaluación que tiene en cuenta sus bases tecnológicas, fundamentadas en el procesamiento de lenguaje natural y la deep learning, así como los aspectos pedagógicos para su utilización. Este trabajo emergente presenta un estado de la cuestión para la comprensión del papel de los chatbots en contextos educativos.

Mayon y Ruiz (2024) construyen un chatbot de orientación vocacional para bachilleres con inteligencia artificial, a partir de su implementación en la Universidad Técnica de Machala, en Ecuador, que se basa en una escala tipo Likert orientada a una experiencia personalizada del usuario en la construcción de respuestas adaptadas a su perfil. Este trabajo ofrece altos índices de aceptación y satisfacción entre los estudiantes en la que los chatbots constituyen una práctica más personalizada y eficaz a la atención tradicional. Su metodología de evaluación basada en la percepción del usuario guarda una relación directa con el enfoque adoptado en esta investigación para medir la usabilidad y satisfacción del sistema implementado.

2.1.2 Antecedentes nacionales

Espinosa-Luna et al. (2023) implementaron un chatbot basado en el modelo de lenguaje GPT-3.5-Turbo con el objetivo de mejorar la atención de consultas de estudiantes de primer ciclo de Ingeniería de Sistemas de la Universidad Nacional de Trujillo. La investigación adoptó un

diseño preexperimental y utilizó el cuestionario Chatbot Usability Questionnaire (CUQ) para evaluar la usabilidad del sistema. Los resultados evidenciaron un puntaje promedio de 73.4 en el CUQ y que el 93 % de los estudiantes consideró útiles las respuestas proporcionadas por el chatbot. Este estudio demuestra la aplicabilidad de los chatbots en escuelas de ingeniería dentro del contexto universitario peruano.

Veliz Yamunaque y Velásquez Granda (2024) implementaron un asistente virtual basado en chatbot para la atención de consultas relacionadas con la gestión académica en la Universidad Nacional Pedro Ruiz Gallo. El desarrollo del sistema se realizó utilizando Python como lenguaje de programación y el marco de trabajo Scrum. Los resultados evidenciaron un incremento significativo en la satisfacción de los usuarios respecto al acceso a la información, pasando del 48.7 % antes de la implementación al 89.45 % después de su uso. Este estudio demuestra el impacto positivo de los chatbots en la mejora de la atención de consultas en universidades públicas peruanas.

López Aguirre y Segura Carmona (2023) desarrollaron un chatbot con inteligencia artificial para mejorar la atención de consultas de estudiantes y egresados en el área de Grados y Títulos de la Universidad de San Martín de Porres. El sistema fue diseñado para automatizar la respuesta a consultas frecuentes, reduciendo los tiempos de espera y la carga operativa del personal administrativo. Los resultados mostraron mejoras en los indicadores de calidad del servicio después de la implementación del sistema. Este antecedente evidencia la utilidad de los chatbots para optimizar procesos administrativos universitarios.

Quispe Cruz (2024) implementó un chatbot orientado a mejorar la calidad del servicio de atención al egresado en la Universidad Peruana de Ciencias e Informática. La evaluación del sistema se realizó mediante un cuestionario de satisfacción con escala Likert, cuyos resultados

fueron analizados utilizando IBM SPSS y la prueba de Wilcoxon para muestras relacionadas. Los resultados apuntan a una buena tendencia hacia una mejora de los scores en la media de la calidad del servicio, de fiabilidad, de la capacidad de respuesta y de la seguridad tras la implementación. El trabajo aporta evidencia cuantitativa en lo que respecta a la utilización de chatbots para la mejora del servicio universitario.

En su conjunto, los antecedentes revisados reflejan el creciente protagonismo de los chatbots basados en inteligencia artificial como herramientas para mejorar la atención de consultas y el acceso a la información en entornos universitarios. Tanto en el ámbito internacional como en el nacional, los estudios reportan resultados favorables en usabilidad, satisfacción de los usuarios y eficiencia en la gestión de consultas académicas y administrativas. No obstante, se advierte que la mayoría de estas implementaciones se concentran en universidades concretas o bajo enfoques tecnológicos particulares, con escaso desarrollo de soluciones adaptadas a las necesidades de universidades públicas en regiones descentralizadas del Perú. Ante ello, esta investigación busca aportar mediante el desarrollo e implementación de un chatbot inteligente para optimizar la atención de consultas en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional de San Cristóbal de Huamanga.

2.2 Bases teóricas

2.2.1 Chatbot

Un chatbot, agente conversacional o asistente virtual es un programa computacional diseñado para simular una conversación con personas utilizando lenguaje natural, que puede ser escrito u oral. Su funcionamiento radica en la interpretación de los mensajes del usuario, el procesamiento de la información existente y la producción de mensajes de respuesta de forma coherente y relevante en tiempo real. De acuerdo a Adamopoulou & Moussiades (2020), los

chatbots son una de las aplicaciones más evidentes de la inteligencia artificial en el ámbito de la interacción humano-máquina, y los mismos se utilizan en el comercio electrónico, la salud, la educación o los servicios públicos con el fin de automatizar la atención sobre las consultas más frecuentes y la dependencia de personal humano en tareas repetitivas.

Desde una perspectiva arquitectónica, los chatbots se clasifican en dos bloques de chatbots: aquellos basados en reglas, pues trabajan con los árboles de decisión y únicamente responden a patrones específicos, y los chatbots basados en inteligencia artificial, que se basan en modelos de machine learning y de procesamiento de lenguaje natural para entender el contexto de la consulta y producir una respuesta adaptativa. En la actualidad los sistemas tienden hacia este segundo tipo de chatbots, aprovechando modelos de lenguaje de mayor escala que incrementan enormemente esa capacidad de comprensión y de respuesta.

Introducir un chatbot en un entorno institucional implica llevar a cabo el cuerpo de actividades técnicas y metodológicas necesarias para su diseño, desarrollo, despliegue y puesta en operación, de un sistema de diálogo automatizado que satisfaga las necesidades particulares de la organización. Este cuerpo de actividades incluye la definición de la arquitectura del sistema, la construcción y estructuración de la base de conocimiento institucional, la integración de los diferentes componentes tecnológicos (*backend*, *frontend* y base de datos), la configuración de los mecanismos de respuesta y la validación mediante pruebas con usuarios reales. Implementar el chatbot en la presente investigación para la EPIS-UNSCH implicó desarrollar una arquitectura multinivel propia, desplegarlo en infraestructura en la nube y evaluar su efectividad a través de un diseño preexperimental con medición pretest-postest.

2.2.2 Funcionalidad

En el ámbito de la calidad del software, la funcionalidad hace referencia al grado en que un determinado sistema proporciona las funciones que cubren las necesidades del usuario, tanto explícitas como implícitas, en unas determinadas condiciones de uso. La funcionalidad es una de las características principales del modelo de calidad del producto del software, tal y como expresa la norma ISO/IEC 25010 (2011), e incluye subcaracterísticas como la completitud funcional, la corrección funcional y la pertinencia funcional. El ámbito de los sistemas conversacionales estudia la funcionalidad en función de la capacidad de un chatbot para dar respuesta correcta y completa a las preguntas de los usuarios, es decir, a la gama de temas.

En el presente estudio, la funcionalidad es operacionalizada como variable independiente X1 y se obtiene a partir del porcentaje de respuestas correctas que recibe el chatbot al contestar las consultas planteadas por los estudiantes. El porcentaje de respuestas correctas se halla mediante un sistema automático que establece la coincidencia entre las respuestas generadas por la máquina y los criterios de validación. Con ello, se persigue garantizar la evaluación desde una perspectiva de consistencia y de reproducibilidad de la misma.

2.2.3 Fiabilidad

La fiabilidad es la característica de calidad del software que indica el grado en que un sistema en particular mantiene su nivel de rendimiento bajo condiciones específicas y en un período de tiempo determinado. De acuerdo con la ISO/IEC 25010 (2011) la fiabilidad engloba subcaracterísticas como la madurez del sistema, la disponibilidad, la tolerancia a fallos y la capacidad de recuperación. En cuanto a los sistemas de atención automatizada la fiabilidad llega a ser comparable a la probabilidad de que un sistema de atención automatizada no falle en su tarea, y se manifiesta en la consistencia de las respuestas generadas, la estabilidad del servicio ante grandes volúmenes de consultas simultáneas y la ausencia de interrupciones no programadas.

En esta investigación la fiabilidad es operacionalizada como variable independiente X2, medida a partir de la tasa de acierto obtenida en las respuestas (entendida como el cociente entre las respuestas correctas y el total de las consultas respondidas), y el tiempo medio entre fallos del sistema. Esta variable permite evaluar la consistencia y la estabilidad del chatbot implementado bajo situaciones reales de uso.

2.2.4 Usabilidad

La expresividad de la usabilidad recoge el grado en que un sistema puede ser utilizado por un grupo de usuarios para llevar a cabo unos objetivos determinados con eficacia, eficiencia y satisfacción en un contexto de uso dado. Según la norma ISO/IEC 25010 (2011), la usabilidad como característica de calidad del software abarca subcaracterísticas como la reconocibilidad de la idoneidad, la facilidad de aprendizaje, la operabilidad, la protección frente a errores del usuario y la accesibilidad. En el contexto de los chatbots educativos, la usabilidad cobra especial importancia porque los usuarios (estudiantes universitarios con distintos niveles de familiaridad tecnológica) deben poder interactuar con el sistema de forma intuitiva y sin necesidad de capacitación previa.

En el proyecto, la usabilidad es operacionalizada como la variable independiente X3, a partir de la percepción del usuario sobre la facilidad de uso del chatbot; su medición se realiza mediante un cuestionario tipo Likert de 5 puntos (1 = totalmente en desacuerdo, 5 = totalmente de acuerdo), considerando dimensiones como facilidad de uso, claridad de la información, intuitividad en la interacción y facilidad para completar tareas.

2.2.5 Optimización de la atención de consultas académicas y administrativas.

La optimización de la atención de consultas académicas y administrativas implica la mejora sistemática del proceso mediante el cual una institución educativa organiza, gestiona y responde a

las solicitudes de información de sus usuarios, con el objetivo de elevar la calidad del servicio en términos de velocidad de respuesta, precisión de la información y satisfacción del usuario.

Dentro del contexto de las universidades, el proceso mencionado puede englobar la atención de consultas vinculadas a trámites administrativos, normativas institucionales, horarios y procedimientos académicos. Históricamente, la operativa de dicho proceso ha dependido de la presencia física del personal, lo que conlleva limitaciones de accesibilidad, disponibilidad condicionada a los horarios laborales, tiempos de respuesta prolongados, etc.

La presencia de un chatbot permite transformar este proceso mediante una automatización de las respuestas y un acceso inmediato a la información institucional. En este trabajo, esta variable es la variable dependiente (Y) operativizada a través de las dimensiones de la eficacia, de la eficiencia y de la satisfacción del usuario.

2.2.6 Eficacia

La eficacia se refiere al grado en el que un sistema logra los resultados para los cuales fue diseñado independientemente de los medios que emplea. En el caso de la calidad del software, la norma ISO/IEC 25010 (2011) define la eficacia como la capacidad del sistema de proporcionar al usuario una información correcta, completa y relevante en respuesta a su consulta. En el caso de los chatbots universitarios, la eficacia se evalúa evaluando la pertinencia y la exactitud de las respuestas que el sistema genera en función de las necesidades informativas del usuario.

En este caso, la eficacia es la dimensión Y1 de la variable dependiente y se mide a partir de un cuestionario, con las características de un cuestionario tipo Likert de cinco puntos, es decir, gradúa las alternativas de respuesta (1 = totalmente en desacuerdo, 5 = totalmente de acuerdo) con relación a la precisión de la respuesta, a la utilidad de la respuesta, a la completitud de la respuesta.

Cabe decir que esta dimensión también comparte indicadores con la dimensión de funcionalidad. Sin embargo, se tratan de dimensiones analíticamente distintas: la funcionalidad tiene que ver con una medida de la calidad técnica del sistema (corrección, completitud y finura de las respuestas: se refiere a las preguntas, las cualidades e ingredientes técnicos que se deben cumplir para obtener una respuesta de calidad); la eficacia tiene que ver con la calidad desde la perspectiva del usuario, en función de la medida que la respuesta satisface a la necesidad informativa y que se ajusta al objetivo de la consulta.

2.2.7 Eficiencia

La eficiencia hace referencia a la relación entre los resultados obtenidos y los recursos invertidos para llegar a esos resultados. Para el caso del servicio prestado, constituye una medida reservada al tiempo y esfuerzo para satisfacer una necesidad del usuario. La norma ISO/IEC 25010 (2011) considera que la eficiencia, como atributo del software, hace referencia, tanto al comportamiento temporal del sistema (tiempos de respuesta y tiempos de procesamiento) como al uso de recursos computacionales. En los sistemas automatizados de atención, esta característica se traduce para la persona usuaria en la rapidez de la producción de las respuestas con respecto a los tiempos de espera por los medios de atención convencionales, que son considerablemente más altos.

En esta investigación, la eficiencia constituye la dimensión Y2 de la variable dependiente y es operacionalizada mediante un enfoque mixto: por un lado, a través del tiempo promedio de atención por consulta registrado automáticamente por el sistema; y por otro, mediante la percepción subjetiva del usuario evaluada con una escala Likert de cinco puntos (1 = totalmente en desacuerdo, 5 = totalmente de acuerdo), considerando aspectos como rapidez en la atención y esfuerzo requerido para obtener la información buscada.

2.2.8 Satisfacción del usuario

La satisfacción del usuario mide el grado en que un sistema o servicio logra cumplir las expectativas y necesidades de quienes lo utilizan. Aunque es una medida de naturaleza subjetiva, constituye un indicador esencial de la calidad percibida del servicio. En el campo de los sistemas de información, ha sido ampliamente estudiada como uno de los principales determinantes de la aceptación tecnológica y del uso continuado de herramientas digitales (DeLone & McLean, 2003). En el contexto de los chatbots universitarios, la satisfacción se relaciona con la naturalidad de la interacción, la relevancia de las respuestas recibidas, la facilidad de uso del sistema y la percepción de mejora respecto al servicio de atención previo.

En el presente trabajo, la satisfacción del usuario constituye la dimensión Y3 de la variable dependiente. Es evaluada mediante un cuestionario tipo Likert de cinco puntos (1 = totalmente en desacuerdo, 5 = totalmente de acuerdo), considerando aspectos como la satisfacción general con el servicio, la calidad de la experiencia de uso y la disposición del usuario para recomendar el sistema a otros estudiantes.

2.3 Definiciones básicas

2.3.1 Procesamiento de lenguaje natural (NLP)

El procesamiento de lenguaje natural (PLN) o NLP (Natural Language Processing) en inglés, es una subdisciplina de la inteligencia artificial dedicada al estudio de la interacción de las computadoras con el lenguaje humano de forma que las computadoras puedan llevar a cabo una representación del lenguaje humano capaz de interpretar, comprender y generar texto o voz de manera significativa. De acuerdo con Manning et al. (2009), el NLP integra técnicas provenientes de la lingüística computacional, el aprendizaje automático y la estadística para abordar tareas como

el reconocimiento de entidades, el análisis de sentimientos, la traducción automática, la clasificación de textos y la generación de respuestas en lenguaje natural.

En el contexto de los sistemas de chatbot, el NLP cumple una función central: permite al sistema transformar la consulta del usuario (expresada en lenguaje coloquial, con variaciones ortográficas o sintácticas), en una representación estructurada que pueda ser procesada computacionalmente. Técnicas como la tokenización, la lematización, la eliminación de palabras vacías y la vectorización de texto son componentes habituales de los pipelines de NLP aplicados en agentes conversacionales. En el sistema desarrollado para la presente investigación, el NLP se aplica en la fase de normalización de texto, que incluye la conversión a minúsculas, la supresión de caracteres especiales y la eliminación de tildes, como paso previo al procesamiento por los cuatro niveles del sistema.

2.3.2 Inteligencia artificial generativa

La inteligencia artificial generativa (IAG) es una rama de la inteligencia artificial dedicada al desarrollo de modelos capaces de producir contenido nuevo (texto, imágenes, audio o código) a partir del aprendizaje de patrones presentes en grandes volúmenes de datos de entrenamiento. Los modelos generativos, a diferencia de los sistemas de IA discriminativos, que sólo son sistemas de clasificación/ predicción basados en datos previos, proporcionan salidas originales no como un output ya almacenado en su base de conocimiento, sino como resultado del proceso de inferencia estadística sobre patrones aprendidos (Ray, 2023).

La IAG en el ámbito del procesamiento de lenguaje natural adoptaría el carácter de los modelos de lenguaje de gran escala como la familia GPT de OpenAI y Gemini de Google que han mostrado capacidades sobresalientes para generar texto que sea correcto, contextualizado y coherente al contestar instrucciones en lenguaje natural, su uso en sistemas de chatbot presenta

ventajas frente a los métodos que se basan en reglas o en la recuperación de información estática, ya que se tenderán a generar respuestas apropiadas de manera específica por cada consulta. En este trabajo, se introduce la inteligencia artificial generativa en el nivel 4 del sistema con el modelo Google Gemini 2.5 Flash como un mecanismo de respaldo para las respuestas que no pueden ser procesadas por los niveles anteriores.

2.3.3 *Fuzzy matching*

La coincidencia aproximada o *fuzzy matching* es una técnica computacional capaz de detectar similitudes entre pares de cadenas de texto que no son iguales, aunque sí tienen un interés común que les da un carácter lexical o semántico muy similar. A diferencia de la coincidencia exacta que exige la correspondencia literal de los términos que se están consultando y los que están almacenados, el *fuzzy matching* toma en cuenta los errores ortográficos, la conjugación de las palabras, los sinónimos, las reformulaciones de una misma pregunta, etc., para determinar el grado de coincidencia a partir de la distancia de Levenshtein, el coeficiente de Jaro-Winkler o la ratio de coincidencia de los caracteres (Naumann & Herschel, 2010).

Esta técnica es particularmente adecuada para la atención a consultas institucionales, puesto que los usuarios consultan la misma pregunta de muchas maneras y con distintas dosis de formalidad en el lenguaje. En el sistema desarrollado para la EPIS-UNSCH, el *fuzzy matching* es el mecanismo central del Nivel 2, que compara la consulta normalizada del usuario con más de 235 pares pregunta-respuesta de la base de conocimiento. Cuando el puntaje de similitud supera el umbral del 70 %, el sistema activa la respuesta correspondiente, lo que le permite atender eficazmente consultas con variaciones lingüísticas respecto a las formulaciones originales almacenadas.

2.3.4 Modelos de lenguaje grande (LLM)

Los modelos de lenguaje grande, conocidos en inglés como LLM (Large Language Models), son sistemas de inteligencia artificial entrenados sobre corpus textuales de gran escala mediante arquitecturas de redes neuronales de tipo Transformer. Su capacidad para comprender y generar texto en lenguaje natural con alto grado de coherencia, fluidez y pertinencia contextual los ha convertido en la base tecnológica de las aplicaciones más avanzadas de procesamiento de lenguaje natural en la actualidad. Modelos como el de GPT-4 de OpenAI, Gemini de Google y LLaMA de Meta muestran un adecuado rendimiento en tareas de respuesta a preguntas, de resumen, de traducción, de razonamiento lógico o de generación de código (Eloundou et al., 2023)

Desde la perspectiva arquitectónica, los LLM utilizan el mecanismo de atención (attention mechanism) introducido por Vaswani et al. (2017), que facilita que cada token pueda dar un peso relativamente mayor a otros tokens de la secuencia, abarcando dependencias de largo alcance en el texto. Su modo de entrenamiento en dos fases (preentrenamiento no supervisado sobre grandes cantidades de texto y fine-tuning con retroalimentación humana - RLHF) les dota de la capacidad de responder adecuadamente de modo contextual incluso ante preguntas no estrictamente contempladas durante la concepción del sistema. En el contexto de este estudio, el LLM utilizado es Google Gemini 2.5 Flash, el cual se inserta en la capa 4 del sistema como el generador en caso de falla del mecanismo.

2.3.5 FastAPI

FastAPI es un moderno y alto rendimiento marco de trabajo para la creación de interfaces de programación de aplicaciones (API) con Python, con una especial atención a la rapidez de ejecución, la facilidad de uso y la creación automática de documentación. Desarrollado por Sebastián Ramírez y publicado en 2018, FastAPI se basa en los estándares abiertos OpenAPI y JSON Schema, y aprovecha las capacidades asíncronas del lenguaje Python a través de la librería

ASGI (*Asynchronous Server Gateway Interface*) para gestionar múltiples solicitudes concurrentes con baja latencia (Ramírez, 2018). Su adopción en entornos de producción ha crecido significativamente en los últimos años, dada su velocidad equiparable a la de frameworks desarrollados en Node.js o Go, combinada con la productividad y legibilidad propias del ecosistema Python.

En este análisis, FastAPI constituye el corazón del *backend* del sistema, que recopila las siguientes peticiones HTTP vía la interfaz de usuario, ejecuta la lógica de procesamiento via del orquestador multinivel y devuelve la respuesta expuesta al cliente en formato JSON. Su despliegue y la cultura en la plataforma Railway le otorgan disponibilidad y escalado automáticamente ante variaciones en la demanda de consultas.

2.3.6 Vue.js

Vue.js es un framework progresivo de JavaScript que se utiliza para desarrollar interfaces de usuario interactivas y aplicaciones web de tipo página única (Single Page Applications, SPA). Creado por Evan You y publicado en 2014, Vue.js se caracteriza por su arquitectura basada en componentes reutilizables, su sistema de enlace de datos reactivo y su curva de aprendizaje más accesible en comparación con otros frameworks de *frontend* como Angular o React (You, 2022). Su condición de progresivo le permite ser adoptado de manera gradual y formar parte de proyectos pequeños o aplicaciones de grandes dimensiones sin necesidad de reformular el código "heredado".

En este trabajo, se utilizó Vue.js 3 para implementar la interfaz conversacional del chatbot, a través de la que el usuario puede ingresar preguntas en lenguaje natural, observar las respuestas del sistema y, a la vez, poder ofrecer *feedback* relativo a la calidad del servicio prestado. La

aplicación se implementó en la plataforma Vercel, garantizando la distribución a nivel mundial a través de CDN, así como tiempos de carga para los usuarios de EPIS-UNSCH reducidos.

2.3.7 Supabase

Supabase es una plataforma de *backend* como servicio (*Backend as a Service*, BaaS) de código abierto que proporciona una base de datos relacional PostgreSQL gestionada en la plataforma de la nube, combinando servicios integrados de autenticación, almacenamiento de archivos, funciones serverless y suscripciones en tiempo real. Tomado como alternativa abierta a Firebase de Google, Supabase permite a los equipos de desarrollo configurar *backends* completos sin tener que gestionar infraestructura de servidores, disminuyendo los tiempos de desarrollo y los gastos operacionales (Supabase, 2023). Su capacidad de adaptación al estándar SQL y la autoexposición de una API RESTful sobre las tablas de la base de datos hacen que la conexión de este sistema con cualquier framework de *frontend* o *backend* sea trivial.

En nuestra investigación, Supabase también gestiona la persistencia de datos del sistema, que consiste en almacenar las conversaciones, sesiones de usuario, valoraciones de *feedback* y métrica de uso en un día. Estos datos sirven de input para un panel de analítica del sistema, que proporciona información sobre los indicadores de eficacia, eficiencia y satisfacción del chatbot implementado en tiempo real.

2.3.8 Tavily API

Tavily es una API específica para búsqueda web hecha para sistemas de IA. A diferencia de los motores de búsqueda convencionales, Tavily está ideada para devolver resultados relevantes, organizados y directamente procesables por los modelos de lenguaje, dejando de lado el ruido informativo típico de los resultados de búsqueda convencionales y facilitando trabajar con información actualizada en el flujo de generación de respuestas de los agentes de IA (Tavily, 2024).

Su arquitectura permite realizar búsquedas temáticas con gran precisión al filtrar contenido irrelevante y dar prioridad a fuentes confiables.

El modelo desarrollado para la EPIS-UNSCH incluye la Tavily API en el Nivel 3 del flujo del procesamiento lo que permite que el chatbot tenga acceso a información institucional actualizada disponible en la página web (avisos, noticias y comunicados oficiales de la UNSCH que no forman parte de la base de conocimiento estática del sistema), dándole así capacidad de respuesta ante preguntas sobre información dinámica y actual.

2.3.9 Norma ISO/IEC 25010

La norma ISO/IEC 25010, publicada en 2011 dentro de la familia de normas SQuaRE (Systems and Software Quality Requirements and Evaluation) define el modelo de calidad del producto de software más ampliamente aceptado a nivel internacional. Define dos modelos complementarios: el de calidad del producto, que especifica ocho características (adecuación funcional, eficiencia de desempeño, compatibilidad, usabilidad, fiabilidad, seguridad, mantenibilidad y portabilidad), y el de calidad en uso, que evalúa la calidad desde la perspectiva del usuario en términos de eficacia, eficiencia, satisfacción, libertad de riesgo y cobertura del contexto (ISO/IEC, 2011).

En la presente investigación, la norma ISO/IEC 25010 constituye el marco normativo de referencia para la evaluación del chatbot implementado. Las dimensiones seleccionadas para la variable independiente (funcionalidad, fiabilidad y usabilidad), corresponden a características del modelo de calidad del producto, mientras que las dimensiones de la variable dependiente (eficacia, eficiencia y satisfacción) se alinean con el modelo de calidad en uso. La adopción de este estándar internacional dota a la investigación de rigor metodológico y permite comparar los resultados con los de otras investigaciones que emplean el mismo marco de evaluación.

También se precisa que en el año 2023 se publicó la versión actualizada ISO/IEC 25010:2023. Sin embargo, la presente investigación adopta la versión 2011 dado que las dimensiones seleccionadas (adecuación funcional, fiabilidad, usabilidad, eficiencia de desempeño y satisfacción) se mantienen conceptualmente equivalentes en ambas versiones, y la mayoría de los antecedentes revisados emplean igualmente la versión 2011 como marco de referencia, lo que facilita la comparabilidad de resultados.

CAPÍTULO III

MATERIAL Y MÉTODOS

3.1 Tipo de investigación

El tipo de investigación es aplicada, también denominada tecnológica, dado que su propósito es generar conocimiento con aplicación directa a la resolución de un problema concreto: la deficiente atención de consultas académicas y administrativas en la EPIS-UNSCH. Esta investigación no busca generar teoría de manera aislada, sino producir y evaluar una solución tecnológica (un chatbot con arquitectura multinivel) en un contexto real de uso.

Asimismo, la investigación es de carácter prospectivo, dado que los datos fueron recolectados a partir de la implementación del chatbot; es decir, los resultados se generaron hacia adelante en el tiempo y no a partir de datos históricos. También es longitudinal, dado que el efecto del chatbot fue evaluado en dos momentos distintos: antes y después de la intervención.

3.2 Nivel de investigación

El nivel de investigación es **explicativo**, ya que se busca determinar el efecto de la implementación de un chatbot en la optimización de la atención de consultas académicas y administrativas en los estudiantes de la EPIS-UNSCH. Para ello, se comparan los resultados obtenidos antes y después de la implementación de la herramienta tecnológica mediante un esquema **pretest–postest aplicado al mismo grupo de participantes**, lo que permite analizar en qué medida el uso del chatbot influye en la mejora de la eficacia, eficiencia y satisfacción de los usuarios.

Cabe precisar que, al tratarse de un **diseño preexperimental sin asignación aleatoria ni grupo de control**, los resultados no permiten establecer relaciones de causalidad con plena certeza.

La mejora observada se interpreta, por tanto, como evidencia de asociación entre la intervención y los cambios en la percepción de los usuarios, más que como demostración de causalidad estricta.

3.3 Diseño de investigación

Esta investigación adopta un diseño preexperimental de tipo pretest–posttest con un solo grupo, sin asignación aleatoria de participantes ni grupo de control. Este diseño consiste en evaluar a un mismo grupo antes y después de aplicar una intervención, con el fin de analizar los cambios producidos por esta.

La intervención consiste en la implementación de un chatbot como herramienta tecnológica para la atención de consultas académicas y administrativas en la Escuela Profesional de Ingeniería de Sistemas (EPIS-UNSCH), desplegado en un entorno real de uso.

La medición se efectúa en dos momentos: un pretest, en el que se registra el estado inicial del proceso de atención mediante los canales tradicionales, y un posttest, aplicado tras la implementación del chatbot, en el que se evalúa el desempeño del nuevo sistema.

Los efectos de la intervención se analizan a través de la variable dependiente "optimización de la atención de consultas académicas y administrativas", considerando las dimensiones de eficacia, eficiencia y satisfacción del usuario. Para ello se combinan indicadores objetivos del sistema (como el tiempo de respuesta y las métricas de desempeño) con indicadores subjetivos basados en la percepción del usuario a través de encuestas tipo Likert.

Este tipo de diseño permite identificar diferencias que pueden explicarse por la introducción del chatbot pero, a su vez, adolece de una fuerte limitación en el control de las variables externas del tipo de estudios sin grupo de control, como es el caso del nivel de familiaridad tecnológica del alumnado, la variabilidad en la complejidad de las consultas formuladas, la calidad y la actualización de información que puede encontrarse en la base de

conocimiento, y cuestiones contextuales como el acceso a internet con la velocidad que se podría desear o la carga académica, factores todos ellos que pueden incidir en la percepción y el funcionamiento del sistema. La ausencia de grupo de control es justificada por cuestiones éticas y por otras cuestiones logísticas propias del contexto universitario público: como el chatbot estaba implementado en una institución dirigida a la totalidad del alumnado matriculado, el acceso restringido a un subgrupo de alumnado habría supuesto limitar artificialmente el acceso a un recurso que ya es disponible por parte de ese colectivo. Este tipo de diseño de investigación es el habitual en las investigaciones aplicadas a la tecnología educativa en el contexto de universidades reales en las que no habría posibilidad de implementar la aleatoriedad (Mayon & Ruiz, 2024; Espinosa-Luna et al., 2023). La uniformidad de los resultados obtenidos en cada una de las dimensiones que fueron evaluadas, así como por los grandes tamaños del efecto que se registraron, refuerzan la contundencia de las conclusiones a la que se llega a partir de las limitaciones del diseño por el que se ha optado.

En esta misma línea, cuando en la definición de los objetivos e hipótesis de la investigación se recurre al término "optimizar", este se debe considerar con el sentido evaluativo-comparativo que le concede la propia evaluación: se está evaluando si la incorporación del chatbot produce algún tipo de mejora de este sistema medible en las dimensiones evaluadas en comparación al sistema tradicional, y no se establece el establecimiento de una causalidad como tal. Esta precisión que se hace corresponde a diseños preexperimentales, donde los resultados nos permiten afirmar que los cambios que se producen están asociados a la intervención, aportando evidencia empírica de mejora en las condiciones del estudio.

3.4 Población y muestra

3.4.1 Población

La población es el conjunto de todos los casos que concuerdan con determinadas especificaciones (Hernández-Sampieri et al., 2014).

En el presente estudio, la población está constituida por los 513 estudiantes matriculados en la Escuela Profesional de Ingeniería de Sistemas durante el semestre académico 2025-II. Si bien el chatbot ha sido diseñado para atender a diversos actores institucionales, como docentes y personal administrativo, la investigación se delimita exclusivamente a los estudiantes, por constituir el principal grupo de usuarios y el foco de la intervención implementada.

3.4.2 Muestra

La muestra es un subgrupo de la población del cual se recolectan los datos y debe ser representativo de esta (Hernández-Sampieri et al., 2014).

Para obtener el tamaño muestral en el estudio, se aplicó la fórmula para poblaciones finitas, considerando un nivel de confianza del 95%, una proporción esperada de 0.5 y un margen de error del 5%.

El cálculo anterior se explica de la siguiente manera:

Fórmula para población finita:

$$n = \frac{N \cdot Z^2 \cdot p \cdot (1 - p)}{(e^2 \cdot (N - 1) + Z^2 \cdot p \cdot (1 - p))}$$

Donde:

N = 513 estudiantes

Z = 1.96 (confianza 95%)

p = 0.5 (proporción esperada)

e = 0.05 (margen de error)

Entonces:

$$n = \frac{513 \cdot (1.96)^2 \cdot (0.5) \cdot (1 - 0.5)}{((0.05)^2 \cdot (513 - 1) + (1.96)^2 \cdot (0.5) \cdot (1 - 0.5))} = \frac{492}{2.2404} = 220$$

En consecuencia, el tamaño de la muestra quedó establecido en 220 estudiantes.

La fórmula para poblaciones finitas se empleó únicamente como referencia para establecer el tamaño mínimo requerido de muestra ($n = 220$), y no como mecanismo de justificación de representatividad probabilística. La selección de los participantes se realizó de manera no probabilística por conveniencia, dado que el acceso a los estudiantes dependió de su disponibilidad y disposición a participar durante el periodo de recolección de datos. En consecuencia, los resultados deben interpretarse con cautela respecto a su generalización, dado que el muestreo no garantiza representatividad probabilística estricta.

Dado que el estudio sigue un diseño preexperimental con un único grupo, la muestra de 220 estudiantes participó en su totalidad en la experiencia de uso del chatbot. No se estableció una división entre grupo experimental y grupo de control; en su lugar, todos los participantes fueron sometidos a una medición pretest (antes del uso del chatbot) y a una medición posttest (después de interactuar con el sistema), siguiendo el esquema $O_1 \rightarrow X \rightarrow O_2$. Este diseño permitió evaluar el impacto directo de la implementación del chatbot sobre la atención de consultas académicas y administrativas, mediante la comparación de los resultados obtenidos antes y después de la intervención.

3.5 Variables y dimensiones

3.5.1 Variable independiente

X: Implementación de un chatbot

3.5.2 Dimensiones de la variable independiente

X1: Funcionalidad

X2: Fiabilidad

X3: Usabilidad

3.5.3 Variable dependiente

Y: Optimización de la atención de consultas académicas y administrativas

3.5.4 Dimensiones de la variable dependiente

Y1: Eficacia

Y2: Eficiencia

Y3: Satisfacción

Nota: Si bien el instrumento de evaluación recoge información sobre dimensiones tanto de la variable independiente como de la variable dependiente mediante los mismos ítems en diferentes momentos, estas dimensiones se diferencian en su rol analítico. Las dimensiones de la variable independiente (funcionalidad, fiabilidad y usabilidad) reflejan atributos de calidad del sistema implementado, mientras que las dimensiones de la variable dependiente (eficacia, eficiencia y satisfacción) expresan los resultados percibidos por el usuario en la atención de consultas. La comparación pretest–posttest permite distinguir empíricamente el estado anterior a la intervención del estado posterior, otorgando sentido diferenciado a ambas variables dentro del mismo instrumento.

3.6 Técnicas e instrumentos de recolección de datos

3.6.1 Técnicas para recolectar datos

Las técnicas utilizadas en la presente investigación son las siguientes:

- Análisis documental: para la revisión de documentos institucionales y fuentes de información relevantes.
- Entrevista: aplicada al personal administrativo con el fin de identificar las consultas más frecuentes de los estudiantes.

- Encuesta: dirigida a los estudiantes para evaluar la percepción del servicio de atención antes y después de la implementación del chatbot.
- Evaluación automatizada: utilizada para medir el desempeño del chatbot mediante el procesamiento de consultas reales.

3.6.2 Instrumentos para recolectar datos

Los instrumentos utilizados en la presente investigación son los siguientes:

- Ficha de análisis documental: empleada para sistematizar la revisión de documentos institucionales y contenido web.
- Guía de entrevista: aplicada al personal administrativo para identificar las consultas más frecuentes.
- Cuestionario: basado en el modelo de calidad ISO/IEC 25010, con escala tipo Likert de 5 puntos (1 = totalmente en desacuerdo, 5 = totalmente de acuerdo), aplicado en modalidad pretest y posttest a un mismo grupo de estudiantes.
- Script de evaluación automatizada: desarrollado como instrumento para medir el desempeño del chatbot, permitiendo calcular indicadores como el porcentaje de respuestas correctas, la tasa de acierto y otros parámetros de funcionamiento a partir de las consultas.

3.7 Técnicas de procesamiento y análisis de datos

3.7.1 Técnicas de procesamiento de datos

Tabla 1

Etapas del proceso de recolección y procesamiento de datos

ETAPA	TÉCNICA	PROCESAMIENTO
Etapa exploratoria	Entrevista semiestructurada a	Análisis de contenido y categorización de
	personal administrativo	las consultas más frecuentes

	(secretario de la EPIS y secretario de la FIMGC)	
Etapa de recolección pre-intervención	Encuesta (pretest) enviada por correo masivo a estudiantes matriculados	Estadística descriptiva (media, desviación estándar, frecuencias)
Etapa de intervención (chatbot)	Uso del chatbot por parte de los estudiantes	Registro de interacciones en Supabase
Etapa de recolección post-intervención	Encuesta (postest) enviada por correo masivo a estudiantes matriculados	Estadística descriptiva e inferencial (pruebas de Wilcoxon – no paramétrica)

Nota. Elaboración propia a partir del diseño metodológico de la investigación.

Tabla 2

Fase de preprocesamiento de datos

FASE	TAREA	DESCRIPCIÓN	HERRAMIENTA
Preprocesamiento	Limpiar el texto	Eliminación de caracteres especiales, espacios innecesarios y errores de formato en los documentos de la base de conocimiento.	Python (re, pandas)
		División del contenido en archivos Markdown separados por tema (FAQ, malla curricular, documentos oficiales, información crítica).	Edición manual + Python scripts

Preparar datos estructurados	Formatear datos críticos en JSON	Información de alta frecuencia (fechas, costos, contactos) almacenada en formato JSON para recuperación directa.	Python (json)
------------------------------	----------------------------------	--	---------------

Nota. Elaboración propia a partir del diseño metodológico de la investigación.

Tabla 3

Fase de preparación de la base de conocimiento

FASE	TAREA	DESCRIPCIÓN	HERRAMIENTA
Estructuración y carga	Organizar la base de conocimiento	Estructuración de los archivos Markdown y JSON en una jerarquía de carpetas por nivel de procesamiento (datos/nivel1/, datos/nivel2/)	Sistema de archivos + Python
	Cargar documentos en memoria	Lectura de los archivos al iniciar el servidor para permitir búsqueda inmediata sin consultas a disco en cada petición.	FastAPI

Nota. Elaboración propia a partir del diseño metodológico de la investigación.

Tabla 4

Fase de evaluación e interpretación

FASE	TAREA	DESCRIPCIÓN	MÉTRICAS CLAVE
------	-------	-------------	----------------

Evaluación de precisión	Pruebas de precisión	Evaluación de la exactitud de las respuestas generadas en comparación con respuestas esperadas.	Exactitud, Precisión, Recall, F1
Evaluación de velocidad	Pruebas de velocidad	Medición del tiempo de respuesta promedio del chatbot ante consultas.	Tiempo medio de respuesta (ms)
Evaluación de facilidad de uso	Pruebas de facilidad de uso	Aplicación de cuestionario a usuarios para evaluar la experiencia de interacción con el sistema.	Satisfacción del usuario (escala Likert)
Validación funcional	Ajustes post-pruebas	Ajustes al chatbot según <i>feedback</i> de usuarios y resultados de pruebas.	Registro de cambios
Interpretación de datos	Análisis e informe	Análisis de los resultados obtenidos durante las pruebas.	Informe validado por la asesora

Nota. Elaboración propia a partir del diseño metodológico de la investigación.

3.7.2 Herramientas de procesamiento de datos

Tabla 5

Herramientas tecnológicas para el procesamiento de datos

HERRAMIENTA	FUNCIÓN	JUSTIFICACIÓN
	Lenguaje	
Python	principal para scripts, preprocesamiento y desarrollo del	Versatilidad, comunidad activa, soporte para bibliotecas IA (google-generativeai, fuzzywuzzy, FastAPI)

	<i>backend</i> del chatbot	
FastAPI	Framework para el desarrollo de la API REST del <i>backend</i>	Permite construir endpoints de alto rendimiento con tipado estático, documentación automática y arranque asíncrono
FuzzyWuzzy	Librería de coincidencia aproximada de cadenas de texto para el Nivel 2 del sistema	Permite recuperar respuestas relevantes ante variaciones ortográficas o formulaciones distintas de una misma consulta, sin necesidad de embeddings ni modelos vectoriales
Tavily API	Motor de búsqueda web para el Nivel 3 del sistema	Proporciona resultados de búsqueda en tiempo real orientados a casos de uso de IA, con respuestas estructuradas y filtrado por relevancia
Supabase	Base de datos PostgreSQL en la nube para almacenamiento de conversaciones,	Ofrece acceso mediante API REST y cliente Python, con gestión de roles y autenticación integrada

	<i>feedback</i> y métricas	
Railway	Plataforma de despliegue en la nube para el <i>backend</i> FastAPI	Permite despliegue continuo desde GitHub con servidor activo permanente, sin suspensión automática por inactividad
Vue.js 3	Framework <i>frontend</i> para interfaz dinámica e interactiva	Permite crear una experiencia de usuario responsiva y modular
Bootstrap	Librería de diseño responsivo y componentes visuales	Facilita la construcción de interfaces limpias y adaptativas
Vercel	Plataforma de despliegue para el <i>frontend</i> Vue 3	Integración directa con repositorios GitHub y despliegue automático ante cada actualización
Google Gemini API	Generación de respuestas mediante modelo LLM	Proporciona respuestas contextualizadas de alta calidad, soporte multilingüe y personalización

Nota. Elaboración propia a partir del diseño metodológico de la investigación.

3.8 Requisitos de infraestructura

Tabla 6

Requisitos de infraestructura para la construcción del chatbot

CATEGORÍA	COMPONENTE	REQ. MÍNIMO	JUSTIFICACIÓN
Hardware	Procesador	Intel Core i7 de 8 núcleos o equivalente	Necesario para ejecutar simultáneamente procesos de recuperación, interfaz y preprocesamiento
	Memoria RAM	32 GB	Permite gestionar múltiples procesos en paralelo durante el desarrollo y las pruebas del sistema
	Almacenamiento	SSD 2 TB	Permite acceso rápido a documentos indexados y almacenamiento de vectores y registros temporales
Despliegue	<i>Backend</i>	Railway (cloud)	Servidor FastAPI activo de forma permanente, con despliegue continuo desde GitHub
	<i>Frontend</i>	Vercel (cloud)	Despliegue automático de la aplicación Vue 3 ante cada actualización del repositorio

			PostgreSQL administrado en la
	Base de datos	Supabase (cloud)	nube con acceso mediante API
			REST y cliente Python
			El sistema requiere conexión a
			internet activa para consultar
Conectividad	APIs externas	Google Gemini	los modelos generativos y el
		API, Tavily API	motor de búsqueda web en los
			niveles 3 y 4

Nota. Elaboración propia a partir del diseño metodológico de la investigación.

3.9 Arquitectura del chatbot

El chatbot implementado sigue una arquitectura multinivel de cuatro capas de procesamiento secuencial y progresivo, diseñada para proporcionar respuestas precisas y oportunas a consultas de naturaleza académica y administrativa. La propuesta arquitectónica combina técnicas de recuperación de información estructurada, coincidencia aproximada de texto, búsqueda web en tiempo real e inteligencia artificial generativa, conformando un sistema robusto capaz de adaptarse a distintos niveles de complejidad en las consultas recibidas.

Cada nivel actúa como un filtro de confianza: **si el nivel actual no supera el umbral mínimo requerido, la consulta es escalada al siguiente nivel de mayor capacidad generativa.**

Este tipo de diseño permite que las respuestas más frecuentes y críticas sean respondidas en la máxima precisión y en un latencia mínima, al tiempo que las acciones de entrada más complejas o ambiguas son resueltas con fuentes de información externas o con la capacidad generativa de modelos de lenguaje de gran escala (LLM).

La siguiente figura explica la estructura de directorios del sistema que se ha implementado. El repositorio se encuentra estructurado en dos componentes de despliegue; el directorio *backend/*, con la configuración adecuada que establece la raíz del servicio en Railway albergando toda la lógica del servidor, así como también la implementación de los módulos Python que cubren cada nivel de procesamiento de consultas, el orquestador principal del chatbot, el cliente de conexión a Supabase y la base de conocimiento completa instalada en datos; esta última incluye los ficheros JSON de la información crítica, los doce ficheros Markdown del FAQ de temática, las mallas curriculares formalizadas en repositorios estructurados y los documentos oficiales convertidos. El directorio *frontend/*, desplegado de forma independiente en Vercel, alberga la aplicación Vue 3 con sus componentes de interfaz de chat y de panel de métricas. En la raíz del repositorio se encuentran solamente los ficheros transversales: el esquema SQL de la base de datos en Supabase, el *.gitignore* y la documentación general del proyecto. Esta estructura del repositorio responde al principio de la cohesión por despliegue, esto es, cada capa puede actualizarse de manera aislada sin que las demás capas se vean afectadas.

La base de datos del chatbot se encuentra estructurada de la siguiente manera: el total del contenido del mismo se encuentra dentro de la carpeta *datos/* de la zona *backend* del repositorio y se organiza en torno a tres subfolders, que contienen la información por su naturaleza y por su nivel de criticidad.

El primero de estos subfolders es el *nivel1/*, donde se encuentra el archivo *info_critica.json*, también estructurado en formato JSON, que recoge la información institucional de mayor volatilidad y la exactitud exigida: fechas de los procesos académicos, costos de trámites vigentes y links con fuentes oficiales de la EPIS-UNSCH. La forma en que está diseñado responde a la exigencia de mantener datos concretos que tienen que ser exactos y actualizados.

El segundo subfolder es el nivel2/ que constituye la base del conocimiento y se halla dividido a su vez en tres componentes. El primero de estos componentes es la subcarpeta `faq/`, en la que se encuentran doce archivos en formato Markdown que cubren las categorías temáticas de mayor interés entre los estudiantes de la escuela: información institucional general, docentes y profesores, planes de estudio y mallas curriculares, costos y trámites del TUPA, admisión e inscripción, matrícula y reserva de matrícula, grados y títulos, ciclo vacacional, calendario académico y preguntas frecuentes de tipo general. La subcarpeta `malla_curricular/` constituye el segundo componente, y en ella se encuentran los documentos `malla_2005.json` y `malla_2018.json` que, en formato JSON, representan la estructura de los dos planes de estudio actuales que rigen en la EPIS-UNSCH. El tercer componente `documentos_oficiales/`, que contiene en formato Markdown los documentos de orden normativo de la escuela: el reglamento de grados y títulos, el TUPA en su totalidad, la directiva de nivelación y las normas de convalidación.

Figura 1

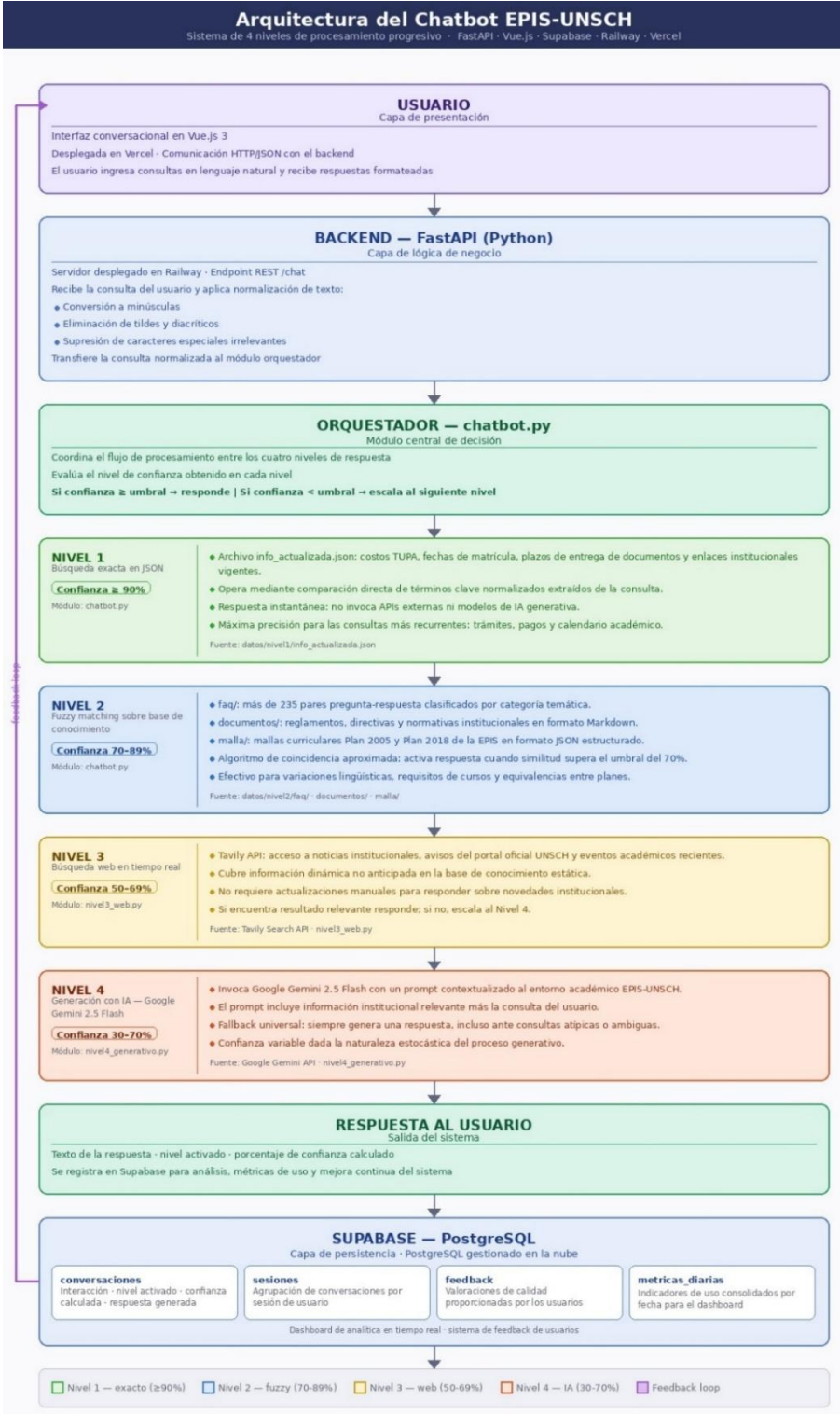
Estructura del proyecto



El núcleo del sistema reside en el módulo orquestador denominado `chatbot.py`, responsable de coordinar el flujo de procesamiento entre los cuatro niveles de respuesta. Cuando el usuario envía una consulta desde la interfaz web, el texto es recibido por la API REST del *backend* y sometido a un proceso de normalización que incluye conversión a minúsculas, eliminación de signos diacríticos y supresión de caracteres especiales irrelevantes. Esta fase de preprocesamiento proporciona una homogeneidad de las comparaciones a realizar en cada uno de los niveles. En el gráfico que sigue a continuación se puede observar el flujo de procesamiento por niveles:

Figura 2

Arquitectura del Chatbot EPIS-UNSCH



- Nivel 1: Búsqueda exacta en fuentes estructuradas

El primer nivel corresponde a la búsqueda de coincidencia exacta sobre un fichero de datos estructurados (en JSON) denominado `info_actualizada.json`, en el que centralizan información de consulta frecuente y de alta sensibilidad en el tiempo regular (costos de los tramites fijados en el TUPA UNSCH, fechas de matrícula, periodos de entrega de documentos, enlaces a los portales institucionales vigentes, etc.). La búsqueda se produce mediante la comparación directa de los términos claves normalizados obtenidos de la consulta efectuada por el usuario contra los índices contenidos en el fichero. Si el sistema halla una coincidencia con un nivel de confianza igual o superior a 90%, la respuesta se produce totalmente a partir de los datos estructurados sin invocar en ningún caso componentes de mayor coste computacional. Este nivel supone unos tiempos de respuesta mínimos y la certeza de una respuesta absolutamente precisa para las consultas más habituales y recurrentes.

- Nivel 2: Coincidencia aproximada sobre base de conocimiento

Cuando el Nivel 1 no consigue validar la condición del umbral de confianza estipulado, la consulta es reconducida al nivel 2, que representará una coincidencia aproximada o fuzzy matching sobre una base de conocimiento estática. Esta base de datos se encuentra organizada en tres carpetas dentro de la carpeta `datos/nivel2/`: la carpeta `faq/`, que almacena más de 235 pares de pregunta-respuesta que son ordenados por la categoría de temática; la carpeta `documentos/`, que conserva reglamentos, directivas y normativas institucionales en la sintaxis de Markdown; y la carpeta `mallla/`, que guarda las mallas curriculares de los estudios de la EPIS de 2005 y 2018 en la sintaxis JSON estructurado.

El algoritmo de coincidencia aproximada obtiene la similaridad entre la consulta normalizada y los términos indexados de la base de conocimiento, y activa la respuesta sólo en el momento en el que el puntaje de similaridad sea mayor al 70 %. El rango de confianza resultante

para este nivel oscila entre el 70 % y el 89 %. Este mecanismo resulta especialmente efectivo para atender variaciones lingüísticas y reformulaciones de preguntas recurrentes, así como consultas sobre estructura curricular, requisitos de cursos, equivalencias entre planes de estudio y procedimientos administrativos contemplados en los documentos normativos.

- Nivel 3: Búsqueda web en tiempo real

El conjunto de consultas que no alcanza el curso del Nivel 2 es derivado a la tercera operación, el cual suma a la búsqueda de respuestas de la API de Tavily un componente que hace que el sistema busque información actualizada en la web en dónde se puede encontrar las siguientes respuestas: noticias institucionales, avisos publicados en los portales de la UNSCH, o información relacionada con las actividades curriculares más recientes. El nivel de confianza de este mecanismo se sitúa entre un 50% y un 69%, dependiendo del tipo de relevancia y calidad de la información de los resultados recuperados. El agregado de este nivel satisface la máquina de responder consultas que tengan información sobre el estado dinámico que no está contemplada en la base de conocimiento estática con la que fue alimentada de forma previa, así que se amplía el rango temático del chatbot sin necesidad de actualizaciones continuas.

- Nivel 4: Generación mediante inteligencia artificial

El cuarto y último nivel funciona como mecanismo de respaldo generativo del sistema. Ante consultas que los niveles anteriores no han podido resolver de manera satisfactoria, el módulo nivel4_generativo.py invoca el modelo de lenguaje Google Gemini 2.5 Flash mediante la API correspondiente. El modelo acepta un prompt contextualizado que integra información institucional pertinente sobre el modelo EPIS-UNSCH y la consulta que realiza el usuario, y devuelve una respuesta que resulta coherente y adecuada en el marco académico. El rango de confianza asignado al nivel que aquí se presenta es de 30 % a 70 %, a causa de la naturaleza

estocástica de este proceso generativo. Este nivel actúa como fallback universal, garantizando que el usuario reciba siempre una respuesta, incluso ante consultas atípicas o formulaciones muy ambiguas.

3.9.1 Interacción multimodal del sistema

El sistema desarrollado va más allá de la interacción textual, ya que incorpora capacidades multimodales que posibilitan a los usuarios realizar consultas mediante voz e imágenes.

Para la entrada por voz, se emplea un mecanismo de reconocimiento automático de voz que transcribe el audio a texto, el cual es procesado a continuación por el chatbot a través del modelo de lenguaje. Esto completa una interacción más fluida y natural.

Ahora bien, a pesar de que el sistema tiene impresas estas capacidades multimodales, la evaluación empírica presentada en la presente investigación se ocupó exclusivamente de consultas textuales, de modo tal que se garantizara la estandarización y la comparabilidad del análisis. La evaluación de las alternativas de voz e imagen debe considerarse una línea futura de investigación, como se indica en las recomendaciones del Capítulo V.

Estas capacidades ofrecen una importante ampliación de la accesibilidad y de la flexibilidad del sistema, de acuerdo con los enfoques existentes de interacción humano-máquina codificados en inteligencia artificial multimodal.

3.10 Componentes del sistema

El sistema está conformado por tres módulos funcionales diferenciados que operan de forma integrada: la capa de presentación (*frontend*), la capa de lógica de negocio (*backend*) y la capa de persistencia de datos (base de datos). A continuación, se describe cada uno de ellos.

3.10.1 *Frontend*: interfaz de usuario

La interfaz de usuario del chatbot fue construida con Vue.js 3, un framework progresivo de JavaScript orientado al desarrollo de interfaces de usuario reactivas y componibles. La aplicación se despliega en la plataforma Vercel, que ofrece alojamiento estático con distribución global mediante red de entrega de contenidos (CDN), asegurando bajas latencias de carga para los usuarios ubicados en la región de Ayacucho y en el resto del país.

La interfaz adopta un diseño conversacional que permite al usuario ingresar consultas en lenguaje natural y recibir respuestas formateadas. Incorpora además un sistema de retroalimentación mediante el cual el usuario puede valorar la calidad de cada respuesta recibida, dato que queda registrado en la base de datos para su posterior análisis. La comunicación con el *backend* se realiza a través de llamadas a la API REST mediante peticiones HTTP POST al endpoint /chat, con envío y recepción de datos en formato JSON.

3.10.2 *Backend*: lógica de procesamiento

El *backend* del sistema fue implementado desde el framework FastAPI, este es un framework de alto rendimiento para APIs en Python que incorpora capacidades asíncronas del lenguaje para facilitar múltiples solicitudes en paralelo. El servidor está desplegado en la plataforma Railway, que proporciona un entorno de ejecución en la nube con soporte para contenedores Docker y escalado automático.

El módulo principal main.py gestiona las rutas de la API y delega el procesamiento de las consultas al orquestador chatbot.py. La estructura del proyecto *backend* se organiza de la siguiente manera: config.py centraliza las variables de entorno y los parámetros de configuración; nivel3_web.py y nivel4_generativo.py encapsulan la lógica específica de los niveles 3 y 4 respectivamente; y el directorio datos/ almacena los archivos de la base de conocimiento organizados por nivel. Las

credenciales de acceso a las APIs externas (Tavily y Google Gemini) se gestionan como variables de entorno, siguiendo buenas prácticas en cuanto al desarrollo del software.

3.10.3 Base de datos: persistencia y analítica

La capa de persistencia de los datos está construida sobre Supabase, un backend como servicio (BaaS) que presenta una base de datos relacional PostgreSQL administrada en la nube. El esquema está construido por sólo cuatro tablas: conversaciones, que almacena cada interacción entre el usuario y el sistema, agregando el nivel activado, confianza calculada y la respuesta devuelta; sesiones, que son las que agrupan las conversaciones por las sesiones del usuario; feedback, que almacena las valoraciones de los usuarios respecto de la calidad de las respuestas; y metricas_diarias, que recoge los indicadores de uso agregados por fecha para la visualización en el panel de analítica.

A partir del panel de analítica instaurado, el sistema, a partir del nivel de administrador, permite contabilizar el volumen de consultas que han sido atendidas, el nivel de distribuciones por canales de activación, los índices de satisfacción de los usuarios, las temáticas que se presentan con mayor frecuencia, etc. Esta información es la que permite la mejora del sistema a partir de un conocimiento empírico que orienta las actualizaciones en la base de conocimiento y los ajustes en los umbrales de confianza por nivel.

CAPÍTULO IV

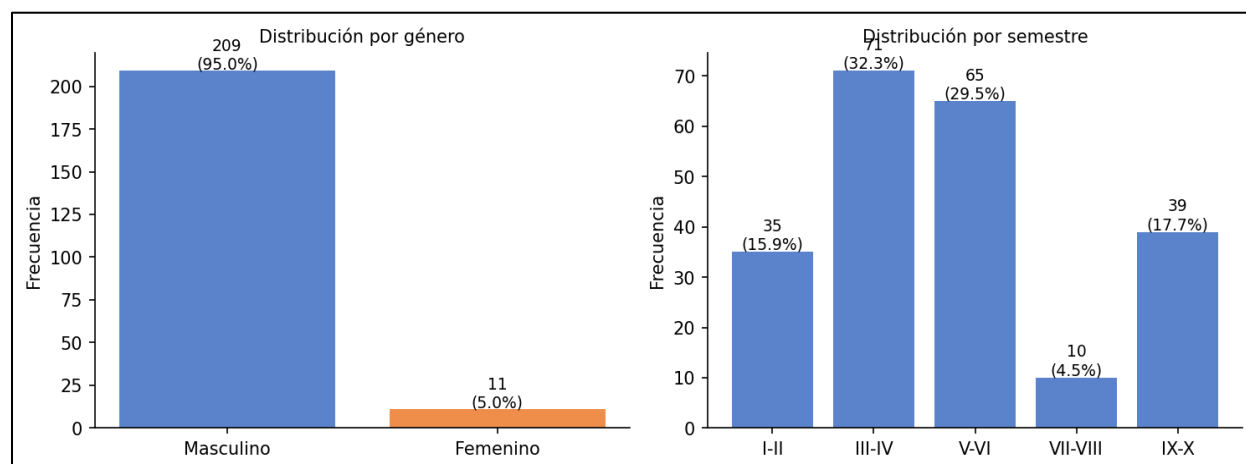
RESULTADOS Y DISCUSIÓN

4.1 Caracterización de la muestra

La muestra del estudio estuvo conformada por 220 estudiantes de la Escuela Profesional de Ingeniería de Sistemas (EPIS) de la UNSCH, quienes participaron en ambas mediciones del diseño pretest-postest ($O_1 \rightarrow X \rightarrow O_2$). Respecto a la distribución por género, 209 estudiantes fueron de género masculino (95.0 %) y 11 de género femenino (5.0 %), distribución que refleja la composición histórica de la carrera. En cuanto al semestre académico, la muestra incluyó estudiantes de todos los ciclos: 35 de primer semestre (I–II, 15.9 %), 71 de tercero y cuarto (III–IV, 32.3 %), 65 de quinto y sexto (V–VI, 29.5 %), 10 de séptimo y octavo (VII–VIII, 4.5 %) y 39 de noveno y décimo (IX–X, 17.7 %), garantizando la representatividad transversal de toda la carrera.

Figura 3

Caracterización de la muestra por género y semestre académico (N = 220)



Nota. Elaboración propia.

4.2 Análisis de fiabilidad del instrumento

El instrumento utilizado en ambas encuestas constó de 16 ítems con escala Likert de cinco puntos (1 = Totalmente en desacuerdo; 5 = Totalmente de acuerdo), organizados en seis dimensiones: funcionalidad (P1–P4), fiabilidad (P5–P8), usabilidad (P13–P16), eficacia (P1–P4), eficiencia (P5–P7) y satisfacción (P9–P12), alineadas con el modelo de calidad ISO/IEC 25010. La consistencia interna del instrumento fue evaluada mediante el coeficiente alfa de Cronbach (α), indicador ampliamente utilizado en investigación cuantitativa para medir la fiabilidad de escalas psicométricas. Según George y Mallery (2003), valores de α superiores a 0.90 se consideran excelentes, entre 0.80 y 0.90 buenos, y entre 0.70 y 0.80 aceptables.

Tabla 7

Coeficiente alfa de Cronbach por dimensión e instrumento completo

Dimensión	Ítems	Nº ítems	α Pretest	α Posttest
Funcionalidad (X1)	P1, P2, P3, P4	4	0.708	0.708
Fiabilidad (X2)	P5, P6, P7, P8	4	0.707	0.706
Usabilidad (X3)	P13, P14, P15, P16	4	0.705	0.707
Eficacia (Y1)	P1, P2, P3, P4	4	0.708	0.708
Eficiencia (Y2)	P5, P6, P7	3	0.719	0.710
Satisfacción (Y3)	P9, P10, P11, P12	4	0.707	0.709
Instrumento completo	P1–P16	16	0.799	0.791

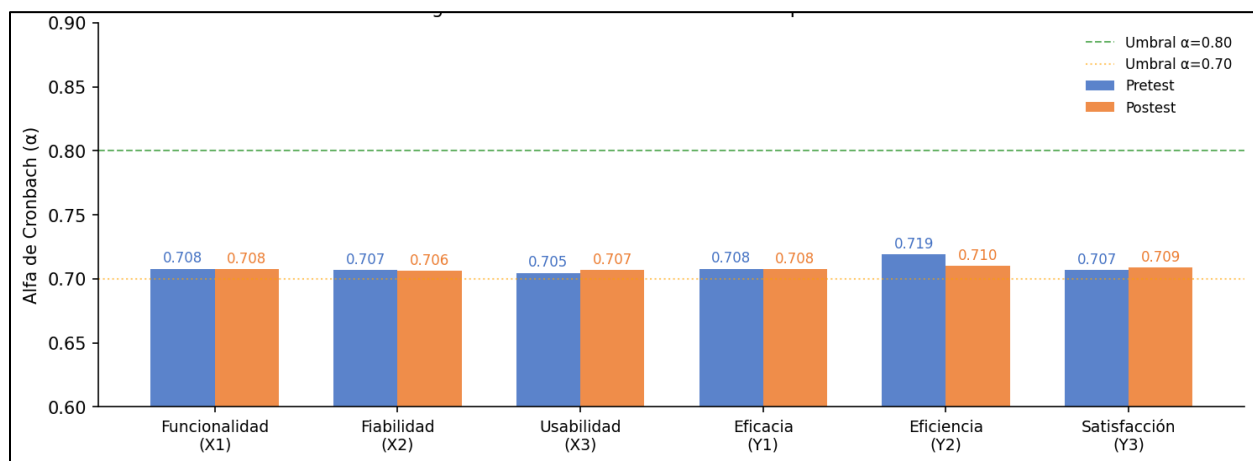
Nota. α = coeficiente alfa de Cronbach. Clasificación: 0.70–0.79 = aceptable; 0.80–0.89 = bueno (George & Mallery, 2003).

Como se observa en la Tabla 7, el instrumento completo obtuvo valores de $\alpha = 0.799$ en el pretest y $\alpha = 0.791$ en el posttest, ambos clasificados como aceptables/buenos. A nivel dimensional, los coeficientes oscilaron entre $\alpha = 0.705$ y $\alpha = 0.719$ en el pretest, y entre $\alpha = 0.706$ y $\alpha = 0.710$

en el postest, todos dentro del rango aceptable (0.70–0.79). Estos resultados confirman que el instrumento presenta una consistencia interna adecuada para medir los constructos de interés, siendo los valores obtenidos consistentes con estudios similares que emplean escalas multidimensionales de calidad del servicio. La Figura 4 muestra gráficamente los valores por dimensión.

Figura 4

Coefficiente alfa de Cronbach por dimensión en pretest y postest



Nota. Las líneas discontinuas indican los umbrales de excelente ($\alpha = 0.90$) y aceptable ($\alpha = 0.70$).

4.3 Prueba de normalidad

Previo a la selección de las pruebas inferenciales, se evaluó la distribución de los puntajes dimensionales mediante la prueba de Shapiro-Wilk. Si bien esta prueba tiende a ser más conservadora con muestras medianas y grandes (lo que aumenta la probabilidad de rechazar la normalidad), su aplicación constituye una práctica habitual en estudios de evaluación de sistemas de información y calidad de software con muestras de tamaño similar (Razali & Wah, 2011). Los resultados se presentan en la Tabla 8. Adicionalmente, la inspección visual de los histogramas y diagramas de caja confirmó la ausencia de distribución normal en todas las dimensiones, lo que valida de manera convergente la selección de pruebas no paramétricas.

Tabla 8

Prueba de normalidad Shapiro-Wilk por dimensión y medición (N = 220)

Dimensión	Medición	W	p-valor	Conclusión
Funcionalidad (X1)	Pretest	0.9075	< 0.001	No normal
	Posttest	0.8938	< 0.001	No normal
Fiabilidad (X2)	Pretest	0.9265	< 0.001	No normal
	Posttest	0.8704	< 0.001	No normal
Usabilidad (X3)	Pretest	0.9256	< 0.001	No normal
	Posttest	0.8913	< 0.001	No normal
Eficacia (Y1)	Pretest	0.9075	< 0.001	No normal
	Posttest	0.8938	< 0.001	No normal
Eficiencia (Y2)	Pretest	0.9263	< 0.001	No normal
	Posttest	0.8686	< 0.001	No normal
Satisfacción (Y3)	Pretest	0.9160	< 0.001	No normal
	Posttest	0.8968	< 0.001	No normal

Nota. W = estadístico de Shapiro-Wilk. Hipótesis nula: los datos siguen una distribución normal.

Se rechaza H_0 cuando $p < .05$.

Los resultados de la Tabla 8 indican que, en todas las dimensiones y en ambas mediciones, el p-valor asociado al estadístico W fue inferior a 0.001, por lo que se rechaza la hipótesis de distribución normal en todos los casos. Dado que los datos no cumplen el supuesto de normalidad, se descartó el uso de la prueba t de Student para muestras relacionadas y se optó por la prueba no paramétrica de Wilcoxon para muestras relacionadas, la cual no requiere el supuesto de normalidad y es adecuada para datos ordinales o continuos no normales con medidas pareadas.

4.4 Estadísticos descriptivos

4.4.1 Análisis por ítem

La Tabla 9 recoge los estadísticos descriptivos de los 16 ítems evaluados en el pretest y postest. En el pretest, las medias de los ítems se situaron entre 2,355 (P4) y 2,664 (P16), es decir, por debajo del punto neutro de la escala (valor 3), lo que indica una percepción generalmente desfavorable del sistema de atención tradicional. Las desviaciones estándar fueron desde 1,21 a 1,33, lo que nos conduce a una variabilidad moderada en las respuestas.

En el postest, las medias se incrementaron de forma relevante, tomando valores que oscilan entre 3,89 (P2) y 4,03 (P6 y P15). De nuevo, todos los valores estaban por encima del punto neutro, lo que evidencia una mejora en la percepción del servicio una vez que se implementó el chatbot. Las desviaciones estándar que fueron desde 1,04 hasta 1,13 indican a su vez una mayor homogeneidad en las respuestas, lo que haría pensar que hay una mayor coincidencia de opiniones entre los participantes.

Y por último, las diferencias de medias (ΔM) fueron positivas en todos los ítems y corrieron entre +1,43 (P7) hasta +1,65 (P15), confirmando así la mejora y del mismo tamaño en todos los ítems que constatamos, constituyendo el resultado un efecto del chatbot sobre el funcionamiento de la calidad del servicio de atención de peticiones.

Tabla 9

Estadísticos descriptivos por ítem: pretest y postest (N = 220)

Ítem	M pre	DE pre	M post	DE post	ΔM	Contenido resumido
P1	2.609	1.304	3.850	1.235	+1.241	Información correcta y precisa
P2	2.573	1.292	3.664	1.207	+1.091	Respuesta útil y relevante
P3	2.495	1.298	3.750	1.259	+1.255	Información completa

P4	2.355	1.279	3.709	1.288	+1.354	Veracidad de la información
P5	2.527	1.360	3.905	1.168	+1.378	Resolución rápida
P6	2.495	1.360	3.823	1.220	+1.328	Sin mucho esfuerzo
P7	2.632	1.312	3.755	1.120	+1.123	Tiempo razonable
P8	2.655	1.388	3.814	1.234	+1.159	Pasos sencillos
P9	2.536	1.280	3.814	1.200	+1.278	Satisfacción general
P10	2.491	1.315	3.736	1.164	+1.245	Experiencia satisfactoria
P11	2.441	1.293	3.714	1.211	+1.273	Recomendaría el servicio
P12	2.518	1.277	3.845	1.180	+1.327	Cumple expectativas
P13	2.573	1.289	3.864	1.201	+1.291	Fácil de usar
P14	2.741	1.355	3.777	1.235	+1.036	Respuestas claras
P15	2.577	1.371	3.777	1.155	+1.200	Inicio intuitivo
P16	2.664	1.347	3.973	1.077	+1.309	Fácil corregir rumbo

Nota. M = media aritmética; DE = desviación estándar; ΔM = diferencia posttest – pretest. Escala Likert 1–5. El ítem P8 ("Los pasos para realizar una consulta son sencillos y directos") fue incluido en el bloque de eficiencia del cuestionario y mantenido en esa dimensión durante todo el análisis estadístico. Aunque se puede decir que guarda relación desde el punto de vista conceptual con la usabilidad, queda justificado su encuadre dentro de la eficiencia porque la facilidad de cada uno de los pasos que tiene el usuario para poder realizar la consulta (la adecuación al usuario) también tiene que ver en cuanto al esfuerzo que tiene que realizar el usuario para obtener una respuesta, cosa que la norma ISO/IEC 25010 (2011) asocia al comportamiento temporal y a los recursos que se utilizan en el sistema. De esta forma, la doble pertenencia conceptual es suficiente para ubicarlo en eficiencia, sin poner en riesgo la validez del instrumento. También se necesita aclarar que las

dimensiones funcionalidad y eficacia, pese a que comparten los ítems P1-P4, difieren en interpretación: la funcionalidad evalúa la calidad técnica del sistema, mientras que la eficacia evalúa el cumplimiento del objetivo de la consulta desde la perspectiva del usuario.

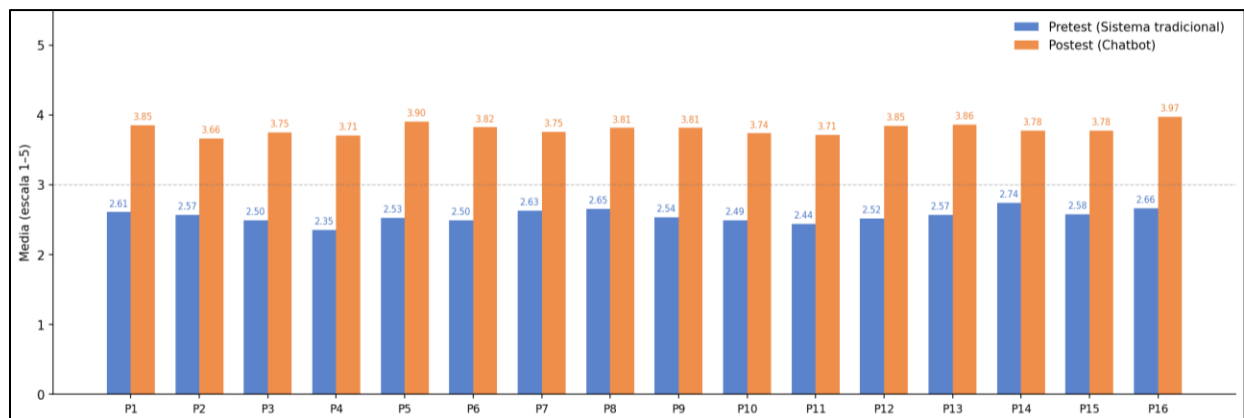
La Figura 5 muestra la comparación de las medias para cada uno de los 16 ítems del pretest (sistema tradicional) y del posttest (sistema chatbot). Se puede observar que en todos los casos las medias del posttest son superiores a los resultados del pretest, lo que demuestra una mejora general en la percepción del servicio tras la implementación del chatbot.

En el pretest, todas las medias se encuentran por debajo del punto neutro de la escala (valor = 3), mientras que en el posttest superan dicho valor en todos los ítems, lo que indica un cambio de una percepción desfavorable a una favorable. Asimismo, la diferencia entre ambas mediciones es relativamente uniforme, lo que sugiere un impacto consistente del chatbot en los distintos aspectos evaluados.

Cabe señalar que las dimensiones de funcionalidad y eficacia comparten indicadores, pero difieren en su interpretación: la primera se centra en la calidad técnica del sistema, mientras que la segunda evalúa el cumplimiento del objetivo de la consulta desde la percepción del usuario.

Figura 5

Medias por ítem en el pretest (sistema tradicional) y posttest (chatbot)



Nota. La línea discontinua indica el punto neutro de la escala (valor = 3).

4.4.2 Análisis por dimensión

La Tabla 10 ofrece los puntajes medios (totales por dimensión), de tal modo que se pueden comparar directamente el sistema tradicional (pretest) y el chatbot (postest) en las seis dimensiones.

Los puntajes medios en el pretest se encuentran entre el 47.8 % (satisfacción) y el 50.0 % (eficiencia), encontrándose en un rango del valor del puntaje próximo al desacuerdo o la neutralidad en la escala Likert. Los resultados evidencian la percepción negativa del sistema tradicional.

En el postest, los puntajes medios alcanzan valores entre el 74,9 % (funcionalidad y eficacia) y el 77,0 % (usabilidad), situándose muy cercano al acuerdo en todas las dimensiones evaluadas.

Las medianas del postest, que oscilan entre 16 y 17 puntos sobre un máximo de 20 en las dimensiones de cuatro ítems, muestran una concentración de respuestas en los niveles superiores de la escala. Este comportamiento confirma una mejora consistente en la percepción del servicio.

Por último, las diferencias de medias (ΔM) fueron positivas en todas las dimensiones, con incrementos que van desde +3,83 (eficiencia) hasta +5,12 (satisfacción), lo que reafirma el impacto favorable de la implementación del chatbot en cada uno de los aspectos evaluados. En conjunto, estos resultados dan cuenta de una mejora significativa y generalizada en la optimización de la atención de consultas tras la implementación del sistema.

Tabla 10

Estadísticos descriptivos dimensionales: pretest y postest (N = 220)

Dimensión	Ítems	Máx	M pre	DE pre	Md pre	M post	DE post	Md post	%máx pre	%máx post	ΔM
Funcionalidad (X1)	P1–P4	20	10.03	3.78	10	14.97	3.64	16	50.2%	74.9%	+4.94
Fiabilidad (X2)	P5–P8	20	10.31	3.95	10	15.30	3.46	16	51.6%	76.5%	+4.99
Usabilidad (X3)	P13–P16	20	10.55	3.91	10	15.39	3.41	16	52.8%	77.0%	+4.84
Eficacia (Y1)	P1–P4	20	10.03	3.78	10	14.97	3.64	16	50.2%	74.9%	+4.94
Eficiencia (Y2)	P5–P7	15	7.65	3.23	7	11.48	2.79	12	51%	76.5%	+3.83
Satisfacción (Y3)	P9–P12	20	9.99	3.77	9	15.11	3.47	16	50%	75.5%	+5.12

Nota. M = media del puntaje total dimensional; DE = desviación estándar; Md = mediana; %máx pre = porcentaje del puntaje máximo posible alcanzado en el pretest; %máx post = porcentaje del puntaje máximo posible alcanzado en el postest; ΔM = diferencia de medias.

La Figura 6 permite comparar las medias obtenidas en el pretest y el postest para las dimensiones de la variable dependiente: eficacia, eficiencia y satisfacción.

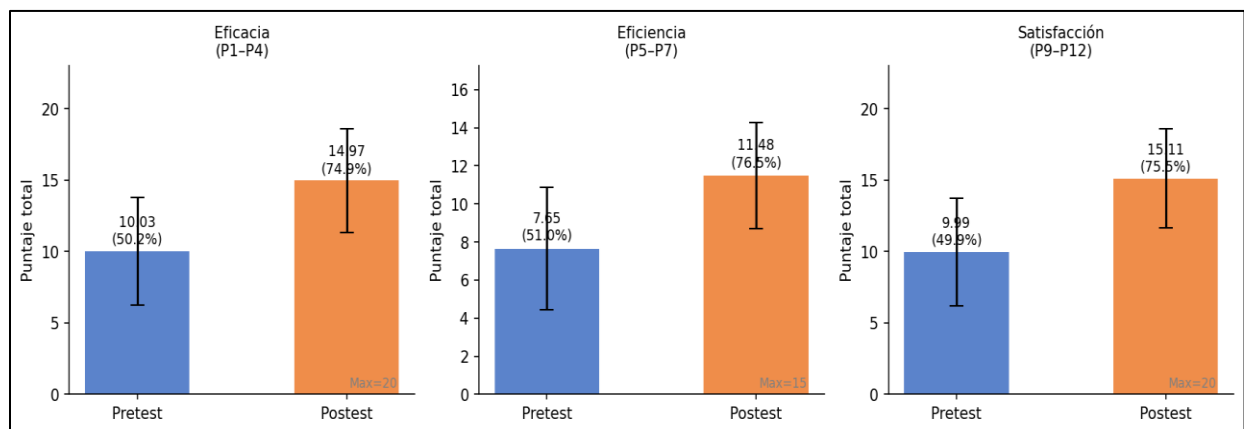
Aparece un incremento claro en los puntajes en las tres dimensiones posteriores en los puntajes después de la implementación del chatbot. En el pretest, los valores se sitúan alrededor de la mitad del puntaje máximo posible, reflejando una percepción moderada o desfavorable del

sistema tradicional. Con el posttest se observó que los puntajes se acercan a puntuaciones de 15 sobre 20 (y de 11 sobre 15 en eficiencia), lo que es también un dato que indica una mejora relevante a niveles de acuerdo.

Las barras de error hacen ver que también hay una leve reducción de la dispersión de los datos en el posttest, lo que hace suponer una mayor homogeneidad de las repuestas del conjunto de los participantes. Estos resultados en conjunto hacen constar que, tras la implementación del chatbot, se llevaron a cabo mejoras claras y homogéneas en la eficacia, la eficiencia y la satisfacción.

Figura 6

Puntajes dimensionales (Variable Dependiente)



Nota. Puntajes medios de la variable dependiente por dimensión antes (pretest) y después (posttest) de la implementación del chatbot. Las barras de error representan ± 1 DE.

La Figura 7 muestra la comparación de los puntajes medios del pretest y posttest para las dimensiones de la variable independiente: funcionalidad, fiabilidad y usabilidad del chatbot.

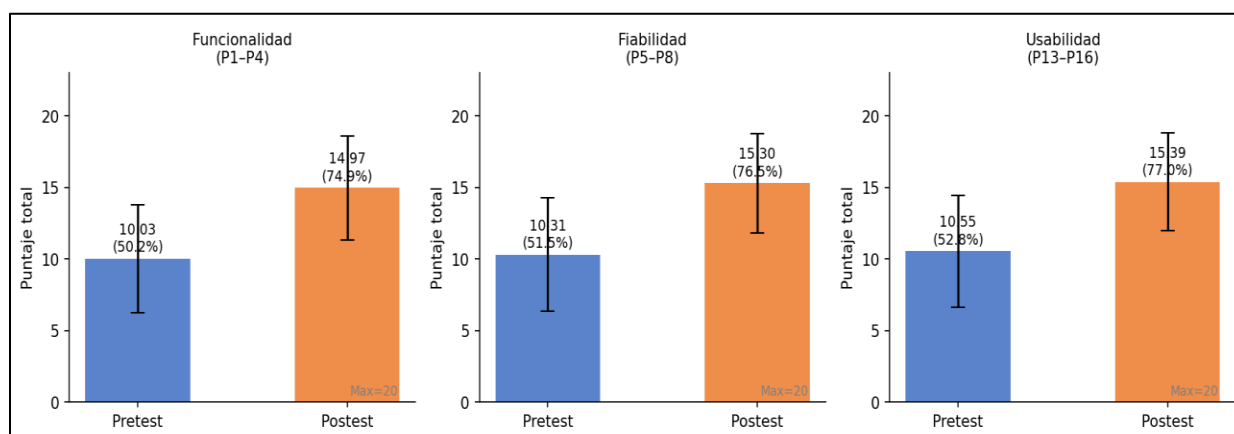
Se observa que, en todas las dimensiones, los puntajes del posttest superan ampliamente a los del pretest. Mientras que en el pretest los valores se ubican alrededor de la mitad del puntaje

máximo, en el posttest se incrementan hasta niveles cercanos a 15 sobre 20, lo que refleja una valoración positiva del sistema implementado.

Las barras de error indican una reducción en la variabilidad de las respuestas en el posttest, lo que sugiere una mayor homogeneidad en la percepción de los usuarios respecto al funcionamiento del chatbot. En general, los resultados evidencian una mejora consistente en la funcionalidad, fiabilidad y usabilidad del sistema.

Figura 7

Puntajes dimensionales (Variable Independiente)



Nota. Puntajes medios de la variable independiente por dimensión antes (pretest) y después (posttest) de la implementación del chatbot. Las barras de error representan ± 1 DE.

4.5 Distribución de respuestas por categoría

El análisis de la distribución porcentual de respuestas por ítem (Figura 8) muestra un cambio claro en el patrón de respuestas entre el pretest y el posttest.

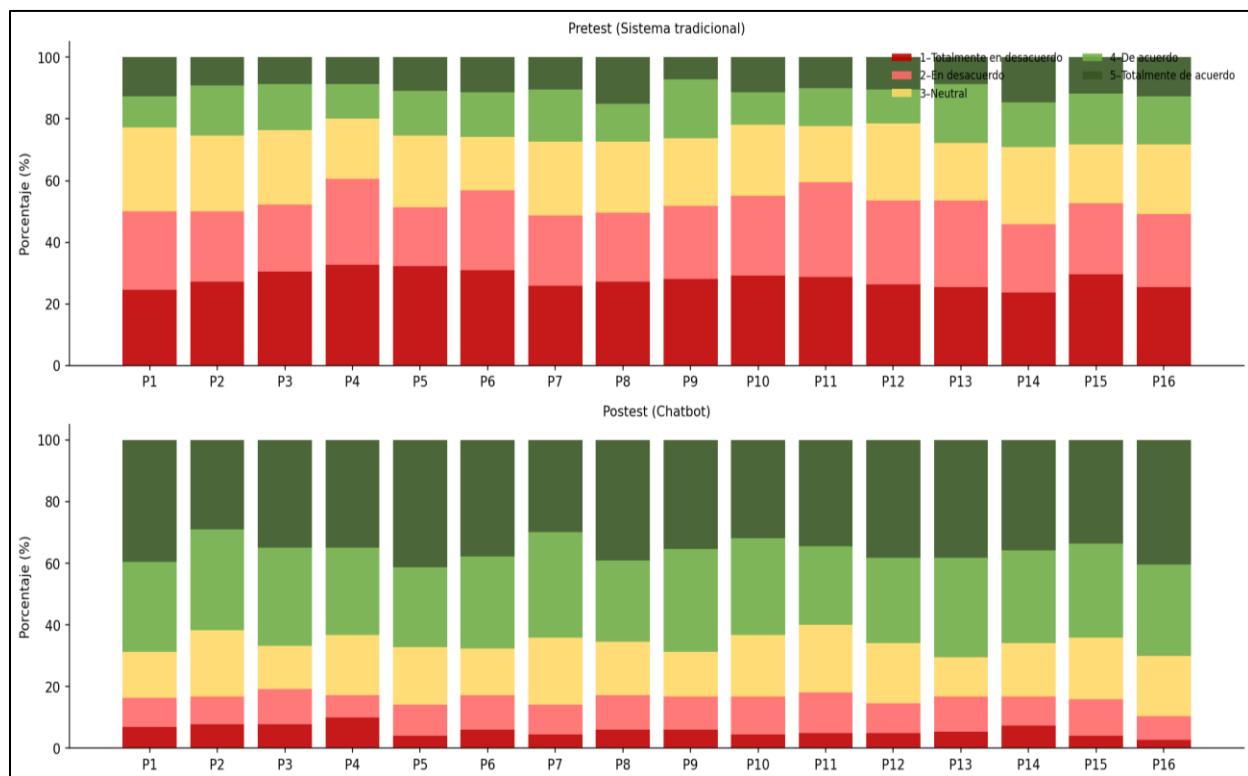
En el pretest, predomina una percepción desfavorable del sistema tradicional, evidenciada por la alta concentración de respuestas en las categorías de desacuerdo (opciones 1 y 2), las cuales representan entre el 52.3 % y el 60.0 % según el ítem. En contraste, las categorías de acuerdo (opciones 4 y 5) presentan una baja participación, sin superar el 25.5 % en ningún caso.

En el postest, se observa una inversión significativa de este patrón. Las respuestas se concentran mayoritariamente en las categorías de acuerdo (opciones 4 y 5), alcanzando valores entre el 67.7 % y el 74.5 %, mientras que las categorías de desacuerdo se reducen considerablemente, situándose entre el 8.6 % y el 12.7 %.

Este cambio evidencia una mejora sustancial en la percepción de los estudiantes tras la implementación del chatbot, pasando de una valoración predominantemente negativa a una claramente positiva en todos los ítems evaluados. La Tabla 11 complementa este análisis al presentar los porcentajes agregados por dimensión.

Figura 8

Distribución porcentual de respuestas por ítem (%)



Nota. Distribución porcentual de respuestas por ítem según escala Likert en el pretest y postest. Los colores representan las categorías de respuesta: rojo (1 = totalmente en desacuerdo), rojo claro

(2 = en desacuerdo), amarillo (3 = neutral), verde claro (4 = de acuerdo) y verde oscuro (5 = totalmente de acuerdo).

La Tabla 11 presenta los porcentajes agrupados de respuestas en las categorías de desacuerdo (opciones 1 y 2) y acuerdo (opciones 4 y 5) para cada dimensión, tanto en el pretest como en el postest.

En el pretest, se observa un predominio de respuestas en desacuerdo en todas las dimensiones, con valores que oscilan entre 55,1 % (eficiencia) y 58,0 % (funcionalidad y eficacia). En contraste, las respuestas de acuerdo presentan porcentajes considerablemente menores, situándose entre 21,2 % (satisfacción) y 23,6 % (eficiencia), lo que evidencia una percepción mayoritariamente negativa del sistema tradicional. Cabe precisar que estos porcentajes se calculan sobre el total de respuestas por dimensión ($N \times$ número de ítems), por lo que difieren de los valores mostrados en la Figura 9, que refleja el porcentaje de participantes con respuesta de acuerdo en el ítem representativo de cada dimensión.

En el postest esa tendencia se invierte consistentemente en las dimensiones. Los porcentajes de desacuerdo descienden considerablemente entre 10.1 % y 11.8 %, mientras que los porcentajes de acuerdo aumentan considerablemente entre el 69.1 % y el 73.8 %.

Este cambio representa una mejora general en la percepción que han mostrado los estudiantes tras la implementación del chatbot, dado que se produce un cambio de una puntuación muy negativa a puntuaciones claramente favorables.

Tabla 11

Porcentajes de acuerdo y desacuerdo por dimensión: pretest y postest

Dimensión	% Desacuerdo	% Acuerdo	% Desacuerdo	% Acuerdo
	pretest (1-2)	pretest (4-5)	postest (1-2)	postest (4-5)

Funcionalidad (X1)	58.0%	21.4%	11.8%	69.1%
Fiabilidad (X2)	55.6%	23.1%	10.1%	73.4%
Usabilidad (X3)	55.2%	22.0%	10.1%	70.9%
Eficacia (Y1)	58.0%	21.4%	11.8%	69.1%
Eficiencia (Y2)	55.1%	23.6%	10.2%	73.8%
Satisfacción (Y3)	57.7%	21.2%	10.5%	69.3%

Nota. Porcentajes calculados sobre el total de respuestas por dimensión ($N \times$ número de ítems por dimensión), agrupando las categorías 1 y 2 como desacuerdo, y 4 y 5 como acuerdo.

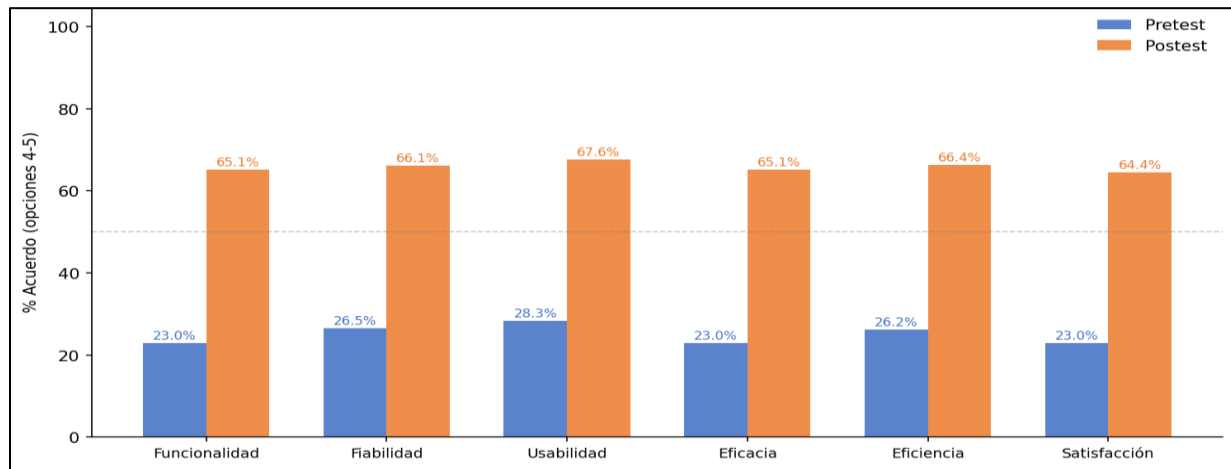
La figura número 9 se refiere al porcentaje de respuestas de acuerdo (opciones 4 y 5) por dimensión del pretest y del postest. En ella, se puede ver que, en el pretest, los niveles de acuerdo son bajos para todas las dimensiones, donde van de un 23.0 % (funcionalidad, eficacia y satisfacción) a un 28.3 % (usabilidad) por debajo del umbral de acuerdo del 50 %, lo que reflejaría, tal como se indicó, una percepción de sistema tradicional negativamente mayoritaria.

En el postest, se puede ver un aumento significativo de los niveles de acuerdo para todas las dimensiones, que varían entre un 64.4 % (satisfacción) y un 67.6 % (usabilidad) que superan con creces ese umbral del 50 % lo que muestra, por tanto, una percepción mayoritaria favorable a la intervención del chatbot.

La variación entre ambas mediciones demuestra que existe una mejora en las dimensiones evaluadas: está confirmada la relación positiva del chatbot en la mejora de la atención a las consultas académicas y administrativas para el sistema tradicional.

Figura 9

Porcentaje de acuerdo por dimensión



Nota. Porcentaje de acuerdo (opciones 4 y 5) por dimensión en el pretest y posttest. La línea discontinua indica el umbral del 50 %, utilizado como referencia para distinguir entre predominio de desacuerdo (< 50 %) y acuerdo (> 50 %).

4.6 Análisis inferencial: prueba de Wilcoxon

Dado que los puntajes dimensionales no se ajustan al supuesto de normalidad, se realizó la aplicación de la prueba no paramétrica de Wilcoxon para muestras relacionadas, que es la prueba con la que se busca contrastar la hipótesis general de la investigación.

La hipótesis nula (H_0) establece que no existen diferencias estadísticamente significativas entre los puntajes del pretest y del posttest para las dimensiones evaluadas. La hipótesis alternativa (H_1) establece que al comparar los puntajes del posttest con los del pretest, se prevé que los puntajes del primer posttest sean significativamente superiores a los del primer pretest. El nivel de significancia adoptado fue $\alpha = 0.05$.

Adicionalmente, se calculó el tamaño del efecto (r) con el propósito de estimar la magnitud práctica de las diferencias observadas entre el pretest y el posttest. Este indicador permite

complementar el análisis de significancia estadística, mostrando no solo si existen diferencias, sino también qué tan grandes son dichas diferencias en términos reales.

El tamaño del efecto se obtuvo mediante la expresión $r = |Z| / \sqrt{N}$, donde Z corresponde al valor estandarizado de la prueba de Wilcoxon y N al tamaño de la muestra. De acuerdo con los criterios propuestos por Cohen (1988), valores de r iguales o superiores a 0,50 se interpretan como un efecto grande, lo que indica un impacto considerable de la intervención. Los resultados se presentan en la Tabla 12.

Tabla 12

Resultados de la prueba de Wilcoxon para muestras relacionadas (N = 220)

Dimensión	W	Valor p	Z aprox.	r efecto	Magnitud	Decisión
Funcionalidad (X1)	1484.0	1.06×10^{-25}	-10.491	0.707	Grande	$p < .001$
Fiabilidad (X2)	1762.0	3.87×10^{-25}	-10.322	0.696	Grande	$p < .001$
Usabilidad (X3)	1734.5	4.65×10^{-25}	-10.302	0.695	Grande	$p < .001$
Eficacia (Y1)	1484.0	1.06×10^{-25}	-10.491	0.707	Grande	$p < .001$
Eficiencia (Y2)	1645.0	4.33×10^{-24}	-10.083	0,680	Grande	$p < .001$
Satisfacción (Y3)	1410.5	6.89×10^{-27}	-10,682	0.720	Grande	$p < .001$

Nota. W = estadístico de Wilcoxon; r = tamaño del efecto ($r \geq 0,50$ = efecto grande, Cohen, 1988).

Nivel de significación $\alpha = .05$.

Los resultados presentados en la Tabla 12 evidencian que, en todas las dimensiones evaluadas, los valores p son considerablemente inferiores al nivel de significancia establecido ($\alpha = 0.05$), situándose en rangos del orden de 10^{-24} a 10^{-27} . Esto implica que la probabilidad de que las diferencias observadas entre el pretest y el posttest se deban al azar es prácticamente nula.

En consecuencia, se rechaza la hipótesis nula en todas las dimensiones analizadas y se acepta la hipótesis alternativa. Esto permite afirmar que existen diferencias estadísticamente significativas entre los puntajes del pretest y el posttest.

Específicamente, se evidencia una mejora significativa tanto en la variable independiente (representada por las dimensiones funcionalidad, fiabilidad y usabilidad) como en la variable dependiente (compuesta por eficacia, eficiencia y satisfacción), lo que indica que la implementación del chatbot produjo cambios positivos en el desempeño del sistema y en la percepción de los usuarios.

El tamaño del efecto r se movió entre 0,680 (eficiencia) y 0,720 (satisfacción), lo que se traduce en un efecto grande según los umbrales de Cohen (1988); lo que significa que las diferencias no sólo fueron significativas en términos estadísticos, sino que además mostraron una gran magnitud a un nivel práctico. En otras palabras, en términos prácticos, el cambio que se derivó de la implementación del chatbot no fue menor ni marginal, sino importante, y a su vez, relevante en el contexto de aplicación.

Desde una perspectiva práctica, los tamaños del efecto obtenidos indican que el impacto de la implementación del chatbot no fue marginal, sino sustancial. Un valor de $r \geq 0,68$ en todas las dimensiones, clasificado como efecto grande según Cohen (1988), significa que la intervención produjo cambios perceptibles y consistentes en la experiencia de los usuarios, más allá de la significancia estadística.

Los anteriores resultados en conjunto, confirman que la implementación del chatbot ha estado relacionada con el presente resultado, de gran magnitud en la atención de consultas académicas y administrativas de satisfacción de los participantes respecto de la solución, tanto a nivel estadístico como práctico.

Podemos ver que en la Figura 10 se encuentran distribuidos los puntajes de cada dimensión de la variable dependiente (eficacia, eficiencia y satisfacción) en forma de diagramas de caja, donde se las comparan las respuestas obtenidas en la evaluación inicial o pretest y la evaluación final o postest.

En las tres dimensiones que estamos analizando en este caso se aprecia un claro desplazamiento de las distribuciones hacia valores superiores del postest con respecto al pretest, lo cual se nota sobre todo por el incremento de las medianas que se aproximan del pretest a valores cercanos a 7-10 en el pretest al postest.

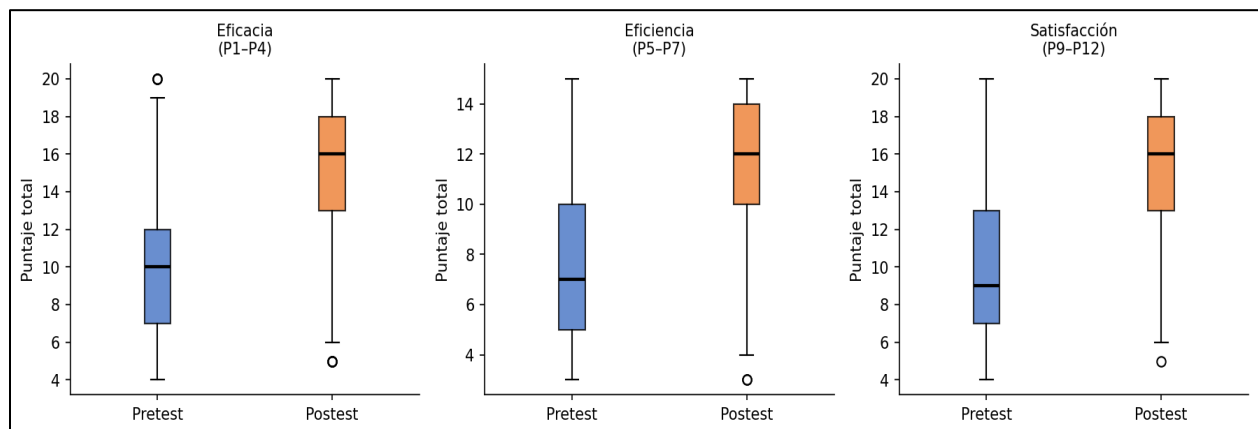
A la vez, las cajas del postest que desglosan el rango intercuartílico se sitúan totalmente por encima de aquellas que representan el pretest, lo cual indica que la mayoría de los valores del postest suponen un incremento del pretest del mismo tipo. Esta separación de las distribuciones dan una idea de un cambio global por niveles de respuesta, no si se quiere referir a un caso particular.

En cuanto a la variabilidad, podemos decir que el aumento de la variabilidad en el postest es un tanto inferior, con lo que se podría justificar una mayor homogeneidad en las respuestas por parte de los estudiantes tras las intervenciones. A la par, también podemos observar que los resultados obtenidos en el pretest y el postest se ven acompañados por algunos valores atípicos que, a pesar de existir, no alteran la tendencia de mejora que se persigue.

El conjunto de los resultados muestra que la implementación del chatbot produce un incremento sostenido de los puntajes de eficacia, eficiencia y satisfacción, lo que pone de relieve a la vez que da cuenta del hecho de que existe una mejora sostenida en la optimización de la atención de consultas académicas y administrativas.

Figura 10

Distribución de puntajes por dimensión (Variable Dependiente)



Nota. Diagramas de caja de los puntajes dimensionales de la variable dependiente: pretest vs. posttest. La línea central de cada caja representa la mediana; los bordes de la caja corresponden a los cuartiles Q1 y Q3; los “bigotes” indican el rango de valores sin considerar valores atípicos, y los puntos representan observaciones atípicas. La separación de las cajas del pretest y del posttest indica claramente que, tras la implementación del chatbot, las distribuciones se situaron en valores claramente superiores a los iniciales.

4.7 Discusión

Los análisis de los resultados demostraron que la utilización del chatbot logró mejoras estadísticamente significativas ($p < .001$) y de gran tamaño ($r \geq 0,68$) en todas las dimensiones puestas a prueba, lo que permite validar empíricamente la hipótesis general de nuestra investigación. Lógicamente, estos resultados no sólo indican que el desempeño técnico del sistema ha mejorado, sino que también hay un cambio importante en las percepciones del servicio de atención habitual por parte de los usuarios. A continuación, se presentan los resultados por dimensión en contraste con la evidencia de estudios previos:

- **Funcionalidad y eficacia:** La funcionalidad registró un incremento de +4,94 puntos (de 10,03 a 14,97, equivalente al 74,9 % del puntaje máximo), lo que indica una mejora en la percepción de precisión, utilidad y completitud de las respuestas. Es decir, desde un punto de vista de mejora de la efectividad en la tutoría online, estas cifras sugieren que el chatbot atendió de forma más significativa las necesidades informativas de los estudiantes que el sistema tradicional.

Este resultado concuerda con lo encontrado por Espinosa-Luna et al. (2023), quienes hallaron niveles muy altos de satisfacción en la percepción de la utilidad en los mismos sistemas conversacionales basados en inteligencia artificial. En este trabajo, la arquitectura multinivel (en especial el componente de generación del Nivel 4) garantizó la generación de respuesta en casos no explícitamente contemplados por la base de conocimiento, además de la percibida como una mejora en la completitud funcional.

Como limitación del instrumento, resulta importante especificar que las dimensiones Funcionalidad (X1) y Eficacia (Y1) se encuentran agrupadas en los mismos ítems (P1–P4) y que obtienen el mismo valor alfa de Cronbach, lo que significa una superposición tal que la distinción entre dimensiones es analítica o interpretativa, pero no empírica porque se miden en un mismo momento y con los mismos reactivos. En futuras investigaciones se debería intentar implementar ítems para cada dimensión bien diferenciados.

- **Fiabilidad:** La dimensión puede apreciarse con un incremento de +4,99 puntos (de 10,31 a 15,30, lo que representa un 76,5 % de la puntuación máxima), lo que indica una mejora en la consistencia y disponibilidad del servicio. El incremento antes descrito tiene como principal explicación la eliminación de las restricciones que ofrece el sistema tradicional:

dependencia de horarios de atención y variabilidad en la capacidad de respuesta del personal administrativo.

Los resultados se alinean con lo señalado por Veliz Yamunaque y Velásquez Granda (2024), quienes mostraron sustanciales mejoras de acceso a la información a través del uso de chatbots en la universidad. La disponibilidad continua del sistema es el principal determinante de mejora del constructo fiabilidad.

- **Usabilidad:** La dimensión tiene un incremento de +4,84 puntos (de 10,55 a 15,39, lo que representa el 77,0 % de la puntuación máxima), lo que indica que la interfaz fue considerada como intuitiva y fácil. Este hallazgo adquiere gran relevancia, ya que la usabilidad es uno de los elementos de gran peso en el uso de las tecnologías en el ámbito escolar.

Un estudio posterior a la presente investigación, por Serrano Franco y Zamudio García (2025), determinaría que las barreras de usabilidad son uno de los obstáculos más enlazados en el uso de sistemas digitales en las universidades. En la investigación presente y en este ámbito, el diseño conversacional elegido permitió que los estudiantes interactuasen con el sistema sin formación previa, favoreciendo el uso natural del mismo.

- **Eficiencia:** La dimensión tuvo una mejora de +3,83 puntos (de 7,65 a 11,48, representando un 76,5 % de la máxima puntuación). Este hallazgo tiene que ser interpretado, puesto que, además, la dimensión tiene un rango de puntuaciones menor que la de las demás (máxima puntuación de 15 puntos).

Desde la práctica, la mejora se debe a la capacidad del chatbot en dar respuestas inmediatas, eliminando el tiempo de espera y reduciendo el esfuerzo del usuario respecto del sistema tradicional.

- **Satisfacción:** La dimensión tuvo una mejora de +5,12 puntos (de 9,99 a 15,11, representando un 75,5 % de la máxima puntuación). Tal resultado se traduce por un juicio positivo del servicio de la que hacen los alumnos e incluye de la misma forma la percepción de mejora en el resto de dimensiones, traduciéndose en una experiencia global más positiva.

El resultado mencionado es coincidente con otros como los de Quispe Cruz (2024) y López Aguirre y Segura Carmona (2023) en los que se experimentaban mejoras significativas en la satisfacción del usuario a partir de la implementación de chatbots en contextos educativos. La mejora percibida sugiere que el mismo cumple no solo con la función de informar, sino que a su vez genera una experiencia buena para el usuario.

Sobre la fiabilidad del instrumento obtenido, los coeficientes alfa de Cronbach (α entre 0.705 y 0.719), mediante las dos mediciones, nos otorgan una consistencia interna aceptable, la cual otorga mucha confianza en la validez de los datos obtenidos. Los valores obtenidos son muy superiores a los de otros estudios similares, realizados con instrumentos de menor número de ítems, y en este sentido agudizan la calidad metodológica del presente trabajo. Por otro lado, la inclusión de capacidades multimodales (voz e imagen) podría ser una posible ventaja del mismo que podría incrementar aún más su accesibilidad y uso en un entorno natural.

En otras palabras, los resultados obtenidos revelan, tal y como es el caso en otras investigaciones anteriores existentes, que tienden a incrementar notablemente la atención prestada a las consultas administrativas y académicas en la EPIS-UNSCH respecto a todos los parámetros analizados, alcanzándose diferencias que son estadísticamente significativas ($p < .001$) y tamaños de efecto grandes (r entre 0,680 y 0,720). De igual modo, la disponibilidad ininterrumpida del

sistema, su arquitectura multinivel y la estructuración de la base de conocimiento institucional explican de forma más cerrada los cambios experimentados.

No obstante, los resultados deben ser considerados con cuidado, si se tiene en cuenta que se han obtenido el marco de un diseño exclusivamente preexperimental y que no se cuenta con un grupo de control en la investigación, ya que no se puede llegar a aislar correctamente las variables personales externas a la ejecución de la tarea. Aún así, el hallazgo de resultados consistentes en las dimensiones analizadas ofrece evidencias concordantes del efecto positivo que el uso de esta intervención instrumental tiene.

Por último, cabe señalar que el evaluador semántico utilizado para la evaluación automatizada de las 600 consultas forma parte de la misma familia que los modelos generativos del componente del Nivel 4 del sistema (Google Gemini). Aunque el evaluador opera de forma independiente (recibiendo únicamente la respuesta del chatbot y una respuesta de referencia predefinida, sin acceso al proceso interno de generación), esta coincidencia constituye una limitación metodológica que podría introducir cierto sesgo favorable en los puntajes de similitud semántica reportados.

4.8 Análisis de las pruebas del chatbot

Para evaluar el funcionamiento técnico del sistema, se evaluó de forma automatizada un total de 600 consultas realizadas por parte del alumnado de la EPIS-UNSCH durante la aplicación del uso del chatbot. Las consultas que se han considerado son las preguntas académicas y administrativas que el sistema había podido resolver en condiciones del funcionamiento real del sistema. Es importante aclarar que el chatbot que se ha evaluado añade capacidades multimodales dentro de las que se incluye la consulta por texto, voz e imágenes; sin embargo, esta evaluación se entenderá en el marco de las consultas por texto, para garantizar la estandarización del análisis.

La evaluación se realizó a través de un evaluador semántico del tipo Google Gemini, que contrasta cada respuesta generada por el chatbot contra una respuesta de referencia, la cual ha sido predefinida, computando un valor de similitud semántica expresado en porcentajes (0 %, 33 %, 67 % o 100 %). Aunque el evaluador y el chatbot, Nivel 4, pertenecen a la misma familia de modelos (Google Gemini), este hecho no introduce un sesgo entre evaluador y sistema como punto determinante, dado que el evaluador trabaja de forma independiente: recibe como entrada la respuesta del chatbot y una respuesta de referencia definida por la investigadora, y emite un juicio de similitud semántica sin acceso al proceso interno de generación. Este mecanismo es acorde con la práctica establecida en la evaluación automatizada de sistemas conversacionales (Chang et al., 2024), en la que los LLM sirven como jueces semánticos dado que poseen el manejo de matices lingüísticos que las métricas léxicas tradicionales no describen.

4.8.1 Resultados generales

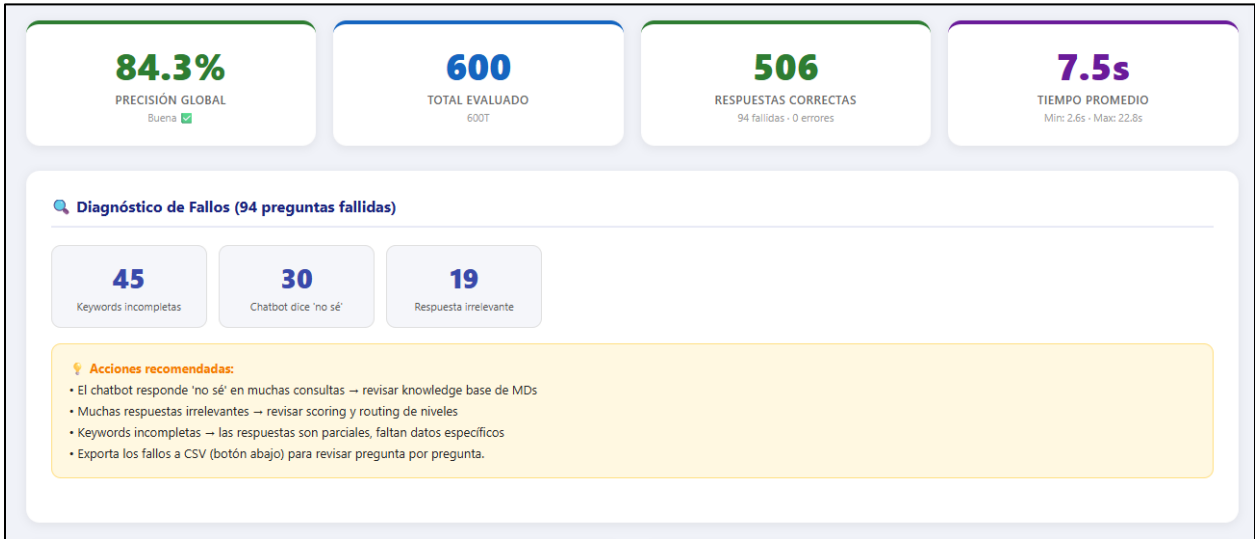
El sistema consigue procesar 506 consultas de una muestra de 600 consultas evaluadas, lo cual se traduce en una precisión global del 84,3 %. Igualmente, tampoco se producen errores técnicos de ejecución (0 errores), lo cual pone en evidencia una fijación adecuada de lo que el sistema hace en condiciones de estabilidad en los procesos relacionados con la evaluación. En términos de rendimiento, el tiempo de respuesta promedio fue de 7,500 segundos ($DE = 3.13$ s), con valores mínimos de 2.6 s y máximos de 22.8 s, lo que indica una respuesta relativamente rápida y consistente para los usuarios.

En relación con los errores identificados, se registraron 94 consultas fallidas, las cuales se distribuyen principalmente en tres categorías: consultas con palabras clave incompletas, respuestas de tipo “no sé” y respuestas irrelevantes. Este diagnóstico permite identificar oportunidades de

mejora en la ampliación de la base de conocimiento y en los mecanismos de recuperación de información.

Figura 11

Precisión global del chatbot (N=600)



Nota. Precisión global del chatbot (N = 600 consultas evaluadas). La figura incluye el número total de respuestas correctas, el tiempo promedio de respuesta y la distribución de fallos clasificados en tres categorías: palabras clave incompletas, respuestas “no sé” y respuestas irrelevantes.

4.8.2 Precisión por nivel de la arquitectura

La Tabla 13 y la Figura 12 presentan los resultados de precisión desagregados por nivel de la arquitectura multinivel del sistema.

Tabla 13

Resultados de evaluación por nivel del chatbot (N = 600)

Nivel	N	Correctas	Precisión	M tiempo (s)	DE tiempo (s)	Score M (%)
-------	---	-----------	-----------	-----------------	------------------	----------------

N1 - Búsqueda exacta	49	41	83.7%	7.08	3.48	68.0%
N2 - <i>Fuzzy matching</i>	423	349	82.5%	7.40	2.92	65.7%
N3 - Web (Tavily)	4	2	50.0%	7.47	4.36	43.8%
N4 - IA generativa (Gemini)	124	114	91.9%	8.01	3.57	73.8%
TOTAL	600	506	84.3%	7,500	3.13	67.4%

Nota. Precisión = porcentaje de consultas correctas por nivel; Score M = media del score de similitud semántica; DE = desviación estándar del tiempo.

El Nivel 4 (IA generativa) muestra un máximo valor de precisión (91.9 %), lo que pone de manifiesto una notable capacidad del modelo para responder consultas complejas o bien no recogidas en la base de conocimiento estructurada. Este resultado pone de manifiesto una relevante actuación del componente generativo como mecanismo de backup para garantizar la cobertura total de respuestas. El Nivel 2 (fuzzy matching) ha procesado el mayor número de consultas (n =423; 70.5 % del total), con un valor de precisión del 82.5 %, lo que evidencia un buen funcionamiento en términos de recuperación de información a partir de la similitud textual contenida en la base de conocimiento.

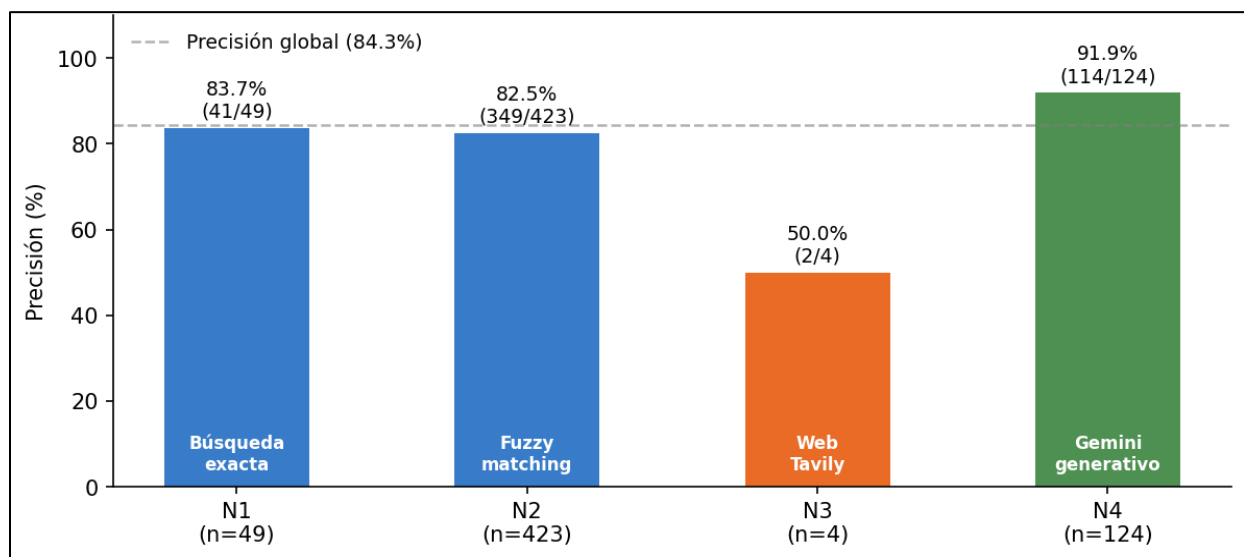
El Nivel 1 (búsqueda exacta) ha alcanzado un porcentaje de precisión del 83.7 %, dando muestras de un buen comportamiento en las consultas directas y bien formuladas por los usuarios.

Finalmente, el Nivel 3 (búsqueda web) ha alcanzado el porcentaje más bajo (50.0 %), si bien hay que tener en cuenta que este nivel ha procesado un total de sólo 4 consultas, por lo que su resultado no es estadísticamente representativo y puede estar condicionado también por la disponibilidad de la información externa en tiempo real.

En definitiva, los resultados muestran que la arquitectura multinivel permite distribuir de forma eficiente las consultas en función de su complejidad, combinando de este modo mecanismos de recuperación estructurada y generación de lenguaje, de forma que se obtenga la máxima precisión del sistema.

Figura 12

Precisión por nivel del chatbot



Nota. Precisión por nivel del chatbot. La línea discontinua representa la precisión global del sistema (84.3 %), permitiendo comparar el rendimiento relativo de cada nivel respecto al desempeño general.

4.8.3 Análisis del tiempo de respuesta

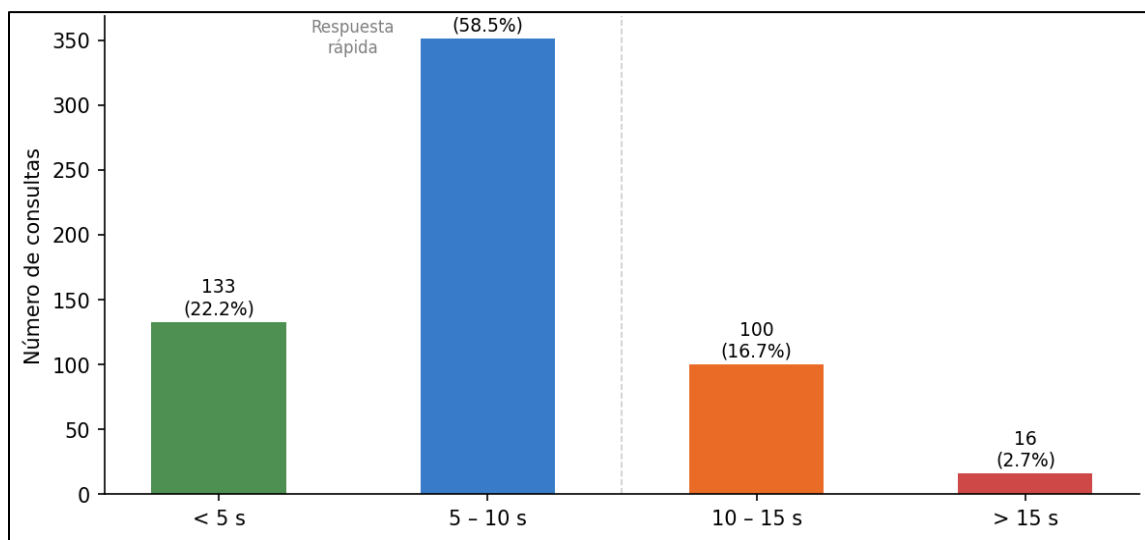
El tiempo medio de respuesta del sistema se estableció en 7.500 segundos, un valor que es perfectamente acorde con la arquitectura multinivel que se implementa (una arquitectura en la que la respuesta se va obteniendo a través de una secuencia de niveles), de lo que se deduce que es un tiempo aceptable.

Este resultado está dentro de los límites de lo considerado aceptable para los sistemas de asistencia conversacional, ya que, de acuerdo con Nielsen (1993), los tiempos de respuesta de menos de 10 segundos mantienen la atención de un usuario en una tarea; en este sentido, el hecho que el 80,7 % de las consultas se respondan en menos de 10 segundos nos puede resultar adecuado. El tiempo medio registrado también es coherente con lo reportado en implementaciones similares de chatbots universitarios que incorporan inteligencia artificial (Espinosa-Luna et al., 2023), que indican que el sistema se encuentra dentro de parámetros de rendimiento comparables al estado del arte en el contexto educativo peruano.

La Figura 13 muestra la distribución de los tiempos de respuesta para las 600 consultas que fueron evaluadas.

Figura 13

Distribución de tiempos de respuesta del chatbot (N = 600)



Nota. Distribución de tiempos de respuesta del chatbot (N = 600 consultas). Se presentan cuatro intervalos: menor a 5 s, entre 5–10 s, 10–15 s y mayor a 15 s, junto con sus respectivos porcentajes sobre el total de consultas.

Como se observa, el 80.7 % de las consultas fue resuelto en menos de 10 segundos, lo que evidencia un desempeño eficiente en términos de rapidez. En particular, 133 consultas (22,2 %) se respondieron en menos de 5 segundos y 351 (58,5 %), entre 5 y 10 segundos, concentrándose la mayoría de las consultas en intervalos de tiempos que son de respuesta rápida.

Por otro lado, los tiempos de respuesta por encima de los 15 segundos fueron más raros (2,7 %; 16 consultas) y siempre estaban asociados con el Generativo de nivel 4, donde la invocación a servicios externos supone una latencia que hace que el tiempo de respuesta sea más elevado.

En el ámbito de la arquitectura, se observó, por el registro del tiempo de respuesta, que el Nivel 4 ($M = 8.01$ s) tenía el mayor tiempo medio, seguido del Nivel 3 ($M = 7.47$ s), Nivel 2 ($M = 7.40$ s) y finalmente Nivel 1 ($M = 7.08$ s). Esta tendencia está alineada con la complejidad del procesamiento de cada uno de los niveles y pone de manifiesto que hay un equilibrio eficaz entre el valor de la precisión obtenido y el tiempo transcurrido para contestar la consulta.

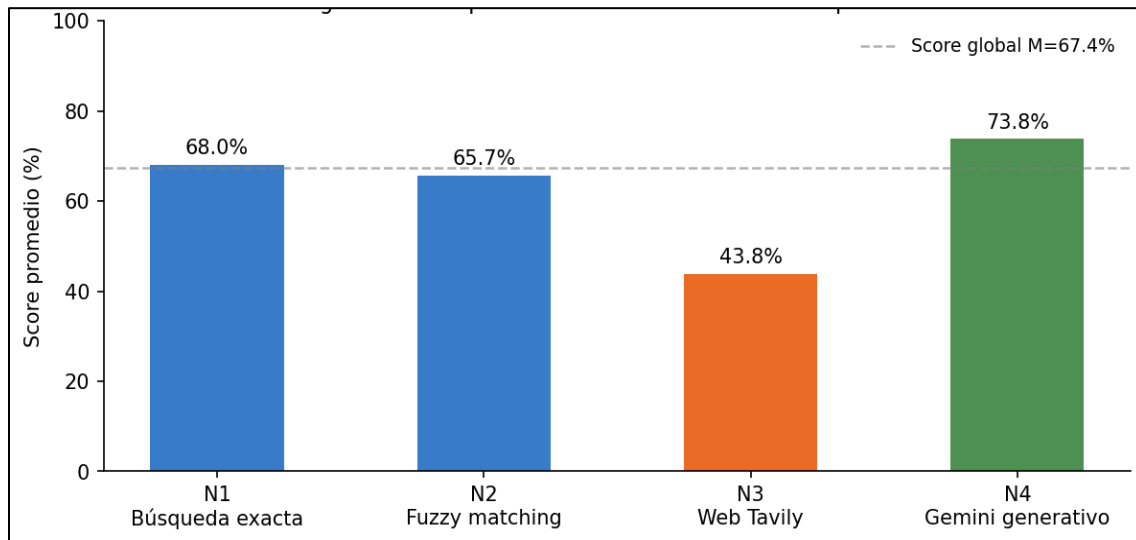
4.8.4 Score de similitud semántica

Además de la clasificación binaria de las respuestas (correcto/incorrecto), se analizó el score de similitud semántica, que permite evaluar el grado de correspondencia entre la respuesta generada por el chatbot y la respuesta de referencia. A nivel global, el sistema obtuvo un score promedio de 67.4 % ($Md = 75$ %, $DE = 27.9$ %), lo que indica un nivel intermedio-alto de coincidencia semántica en las respuestas generadas.

La Figura 14 presenta los scores promedio desagregados por nivel del sistema.

Figura 14

Score promedio de similitud semántica por nivel



Nota. Score promedio de similitud semántica por nivel del chatbot. La línea discontinua representa el promedio global (67.4 %), permitiendo comparar el desempeño relativo de cada nivel.

El Nivel 4 generativo obtuvo el mayor score promedio (73.8 %), evidenciando una mayor capacidad para generar respuestas semánticamente completas en consultas complejas. Le siguen el Nivel 1 (68.0 %) y el Nivel 2 (65.7 %), con desempeños similares, mientras que el Nivel 3 (43.8 %) presentó el menor valor, debido a su dependencia de fuentes externas. En cuanto a la distribución de scores, se observa que:

- El 25.8 % de las consultas alcanzó el score máximo (100 %), indicando respuestas completamente correctas.
- El 15.5 % obtuvo 67 %, reflejando respuestas mayormente correctas.
- El 4.0 % obtuvo 33 %, correspondiente a respuestas parcialmente correctas.
- El 7.2 % obtuvo 0 %, representando respuestas incorrectas.

Estos resultados muestran que, además de un buen nivel de acierto, el sistema logra en una proporción importante respuestas completas y semánticamente alineadas con lo esperado.

4.8.5 Análisis de fallos

De las 600 consultas evaluadas, 94 (15.7 %) no fueron resueltas satisfactoriamente. Estas fueron clasificadas según su causa principal, como se muestra en la Tabla 14 y la Figura 15.

Tabla 14

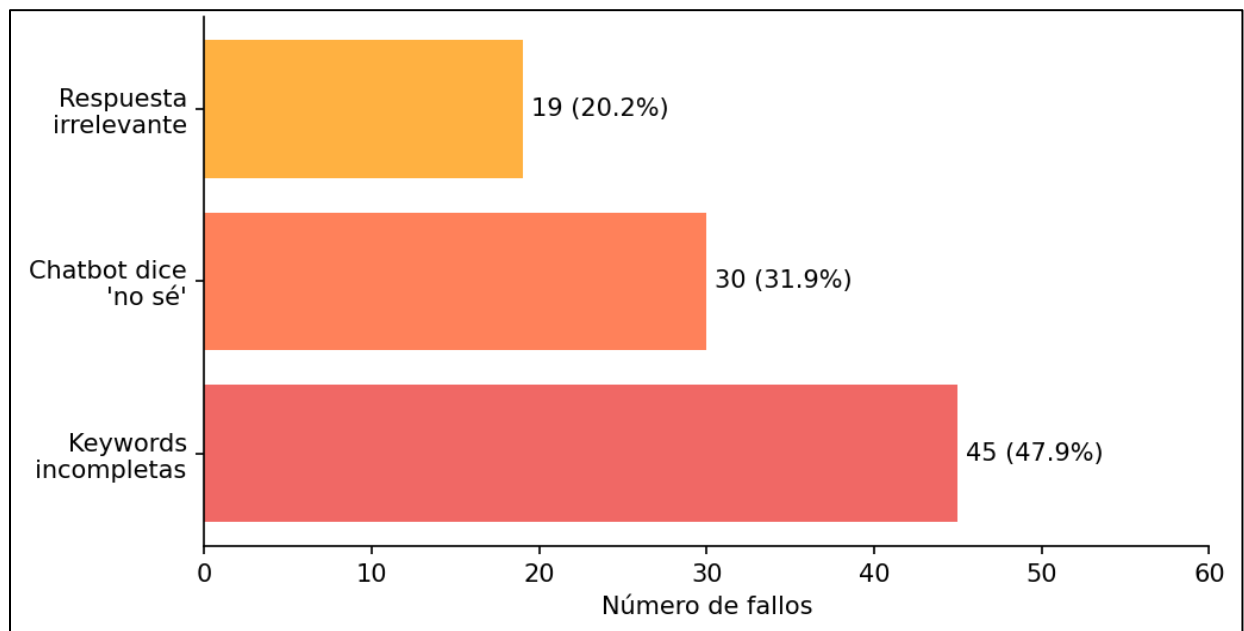
Causas de fallo en las consultas no resueltas satisfactoriamente (N = 94)

Causa	N	%	Descripción
Keywords incompletas	45	47.9%	La respuesta es parcialmente correcta, pero omite datos específicos clave (montos, códigos de cursos, fechas exactas)
Chatbot dice 'no sé'	30	31.9%	El sistema no encontró información en ningún nivel y respondió con mensaje de desconocimiento
Respuesta irrelevante	19	20.2%	La respuesta no guarda relación con la consulta formulada
TOTAL	94	100%	

Nota. Porcentajes calculados sobre el total de consultas fallidas (N = 94). Cada categoría representa la causa principal del fallo identificada en la respuesta del chatbot.

Figura 15

Causas de fallo en las consultas (N = 94 preguntas fallidas)



Nota. Distribución de causas de fallo en las consultas no resueltas (N = 94). Se incluyen tres categorías: respuestas con información incompleta (keywords faltantes), respuestas de desconocimiento (“no sé”) y respuestas irrelevantes.

El análisis muestra que la causa más frecuente corresponde a keywords incompletas (45 casos; 47.9 %), lo que indica que el sistema proporciona respuestas parcialmente correctas, pero omite información específica clave, como montos, códigos de cursos o fechas exactas. La segunda causa corresponde a respuestas de tipo “no sé” (30 casos; 31.9 %), lo que ocurre cuando el sistema no encuentra información suficiente en ninguno de sus niveles de procesamiento. Este resultado sugiere la necesidad de ampliar la base de conocimiento institucional.

Finalmente, 19 consultas (20.2 %) fueron clasificadas como respuestas irrelevantes, generalmente asociadas a consultas ambiguas o mal formuladas, lo que evidencia limitaciones en

la interpretación semántica en ciertos casos. Un análisis más detallado revela que las principales carencias de información están relacionadas con:

- Montos específicos de trámites (S/190, S/15, S/50)
- Sistemas institucionales (SIIGE)
- Códigos de asignaturas (IS181, IS282, IS483)
- Planes de estudio (500 par, 400 impar, etc.)

Estos hallazgos permiten identificar líneas claras de mejora, especialmente en la actualización y enriquecimiento de la base de conocimiento, lo que podría incrementar de manera significativa la precisión del sistema en futuras iteraciones.

4.8.6 Precisión por categoría temática

El análisis de la precisión por categoría temática identifica las áreas específicas en las que el chatbot muestra mayor y menor desempeño, aportando información clave para la mejora del sistema. La Tabla 15 y la Figura 16 presentan los resultados obtenidos.

Tabla 15

Precisión del chatbot por categoría temática de consulta

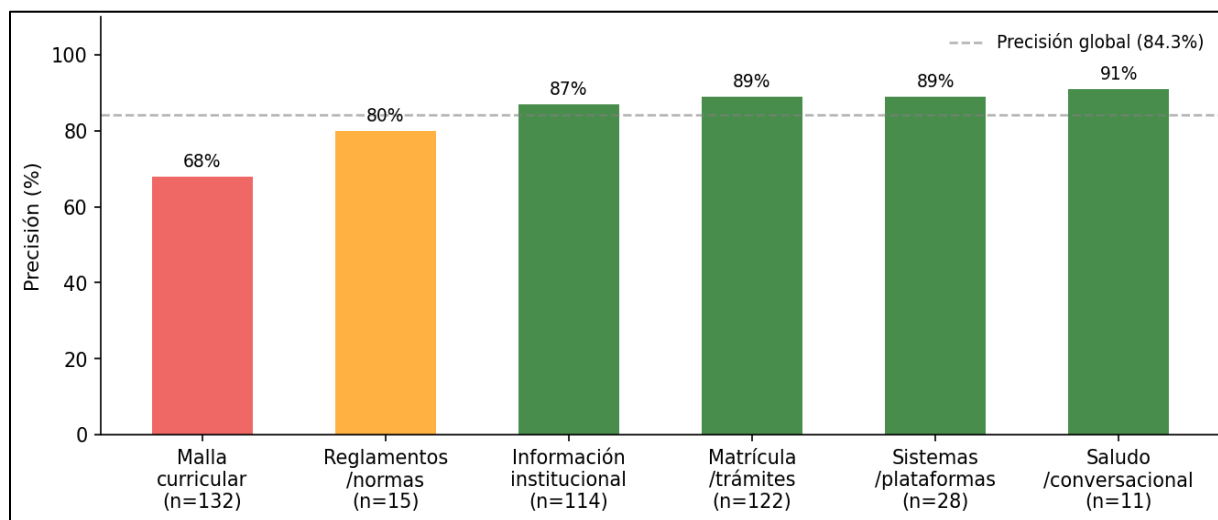
Categoría	N	%	Correctas	Precisión	Descripción
Malla curricular	132	22.0%	90	68%	Cursos, créditos, planes de estudio, semestres, ciclo vacacional
Matrícula y trámites	122	20.3%	109	89%	Costos TUPA, plazos, requisitos de matrícula, pagos
Información institucional	114	19.0%	99	87%	EPIS, UNSCH, ubicación, docentes, carrera

Reglamentos y normas	15	2.5%	12	80%	Separación, condicionada, reingreso, reglamento
Sistemas y plataformas	28	4.7%	25	89%	SIIGE, OTI, portales virtuales, acceso digital
Saludo / conversacional	11	1.8%	10	91%	Saludos, despedidas, preguntas sobre el chatbot
Otras consultas	178	29.7%	151	85%	Consultas mixtas y otras temáticas
TOTAL	600	100%	506	84.3%	

Nota. N = número de consultas por categoría; % = proporción respecto al total; precisión = porcentaje de respuestas correctas.

Figura 16

Precisión por categoría temática



Nota. Precisión por categoría temática. La línea discontinua representa la precisión global del sistema (84.3 %), lo que permite comparar el rendimiento relativo de cada categoría.

Los resultados evidencian que las categorías con mejor desempeño fueron saludo/conversacional (91 %), sistemas y plataformas (89 %), matrícula y trámites (89 %) e información institucional (87 %), todas por encima de la precisión global. Esto indica que el sistema responde de manera más efectiva en consultas frecuentes, estructuradas y con información bien definida dentro de la base de conocimiento. En contraste, la categoría malla curricular registró la menor precisión (68 %), considerablemente por debajo del promedio general. Este resultado puede explicarse por la complejidad de este tipo de consultas, ya que suelen requerir información muy específica, normalmente códigos de asignaturas, la equivalencia entre planes de estudio y características de la currícula que la base de conocimiento no siempre satisface o no se encuentra en una versión totalmente estructurada. Asimismo, destaca que la categoría “otras consultas” tuvo una precisión del 85 %, valor que señala un rendimiento aceptable en vistas a su tipología heterogénea y a la pluralidad de temas que permiten agrupar.

Estos hallazgos se alinean con las causas de fallos, que se mostraban mediante el análisis previamente presentado, en el que se puede observar que una buena parte de los errores observados están relacionados con la inexistencia de información específica (keywords) especialmente en aquellas temáticas referidas a la malla curricular. Así, en este sentido, la mejora del sistema debe ser orientado claramente al enriquecimiento de la base de conocimiento para estas temáticas.

4.8.7 Síntesis del análisis de pruebas

En conjunto, la evaluación automatizada generada en base a las 600 consultas de alumnos desvela que el chatbot que se ha puesto en marcha tiene una precisión global del 84,3 %, nivel que puede corresponder a alto dentro del universo de los sistemas conversacionales aplicados a sistemas educativos, este último aspecto dado que, además, el sistema presenta una elevada estabilidad técnica al no detectar errores de ejecución, así como permitir tiempos de respuesta

adecuados ($M = 7,500$ s), garantizando así una experiencia de uso fluida. El proceso de análisis riguroso hace presente que el chatbot muestra un buen rendimiento en la mayor parte de las categorías temáticas, especialmente en las que se pueden incluir las relativas a trámites, información institucional y uso de los sistemas, en las que la información es, por su propia naturaleza, más estructurada. Asimismo, la arquitectura multinivel contribuye a mantener un equilibrio entre cobertura y precisión, destacando el aporte del componente generativo para resolver consultas complejas. No obstante, se identifican oportunidades de mejora importantes, principalmente en:

- La ampliación y actualización de la base de conocimiento, en especial en los temas de malla curricular y planes de estudio.
- La incorporación de información específica y detallada (montos, códigos, fechas).
- El refinamiento de los mecanismos de enrutamiento, particularmente en el Nivel 3, para mejorar su utilización y precisión.

En resumen, los resultados alcanzados demuestran que el sistema es funcional y es estable, y que puede constituir una buena solución, como un sistema alternativo de optimización de la atención de consultas académico-administrativas y con posibilidad de mejora incremental como consecuencia de los resultados encontrados.

CAPÍTULO V

CONCLUSIONES Y RECOMENDACIONES

5.1 Conclusiones

Se evidencia una mejora significativa en la atención de consultas académicas y administrativas mediante el chatbot implementado en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional de San Cristóbal de Huamanga (UNSCH) durante el semestre 2025-II. Los resultados de la prueba de Wilcoxon muestran diferencias estadísticamente significativas ($p < .001$) en todas las dimensiones evaluadas (funcionalidad, fiabilidad, usabilidad, eficacia, eficiencia y satisfacción), con tamaños del efecto grandes (r entre 0,680 y 0,720). Se trata de un resultado que, además de corroborar que el chatbot se produce una superación frente al canal de atención presencial tradicional desde sus medidas estadísticas, también evidencia una incidencia práctica importante en la experiencia de los usuarios (los estudiantes). La evaluación técnica automatizada desvela, además, una tasa de aciertos del 84,3 % y un tiempo medio de respuesta de 7,50 segundos, lo que acaba por poner en valor el rendimiento técnico de todo el sistema.

La funcionalidad del chatbot experimentó un aumento notable de +4,94 puntos (de 10,03 a 14,97, es decir, un avance del 74,9 % del puntaje máximo; $\Delta M = +4,94$; $p < .001$; $r = 0,707$), por lo que se pudo afirmar que aquel ofrece respuestas correctas, útiles, completas y confiables en lugar de las prestaciones del sistema tradicional; dicho cambio valida la primera hipótesis específica. Este hecho se encuentra en la misma línea que el tipo de evaluación que se realizó de manera técnica y automatizada (84,3 % de aciertos; puntaje medio de similitud semántica de 67,4 %).

Respecto a la fiabilidad, la dimensión incrementó significativamente un +4,99 puntos (de 10,31 a 15,30, es decir, del 76,5 % del puntaje máximo; $\Delta M = +4,99$; $p < .001$; $r = 0,696$), lo que refleja el incremento de la consistencia y la disponibilidad percibida del servicio. La ruptura de la prohibición de una restricción horaria y que el servicio esté disponible para todos en todo momento explican en gran parte que la fiabilidad haya incrementado y esto a su vez valida la segunda hipótesis específica.

La dimensión de usabilidad alcanzó el 77,0 % del puntaje máximo en el posttest, con un incremento significativo de +4,84 puntos ($\Delta M = +4,84$; $p < .001$; $r = 0,695$). La mayoría del alumnado consideró el sistema como intuitivo, claro y fácil de utilizar sin necesidad de capacitarse previamente, lo que contribuye a la adopción del sistema validando así la tercera hipótesis específica.

La dimensión de la eficiencia incrementa con +3,83 puntos (de 7,65 a 11,48, es decir, el 76,5 % del puntaje máximo; $\Delta M = +3,83$; $p < .001$; $r = 0,680$), lo que significa una mejora relativa a la rapidez de acceso a la información y a la facilidad de aclarar dudas en comparación con la atención tradicional, es decir, valida la cuarta hipótesis específica.

Finalmente, la satisfacción del usuario fue la que presentó la mayor variación entre todas las dimensiones de la variable dependiente: +5,12 puntos desde 9,99 a 15,11 alcanzando el 75,5 % de la puntuación máxima; $\Delta M = +5,12$; $p < .001$; $r = 0,720$ -por lo que refleja de una forma favorable una experiencia global con el sistema que está implantado, por lo que se valida la quinta hipótesis específica.

5.2 Recomendaciones

Se sugiere un uso paulatino del chatbot como otra forma de atención al estudiante; en primer lugar integrarlo en el portal web institucional y hacer que los usuarios puedan utilizarlo a

través de los canales de comunicación de la escuela. Esta metodología de uso permitirá hacer pruebas del chatbot en condiciones de funcionamiento real y facilitará que la base de conocimiento crezca poco a poco y a la vez se expandan los temas hacia otras áreas de la institución.

De forma análoga, se considera también que sea necesario priorizar la actualización de la base de conocimiento sobre el área de la malla curricular, ya que fue la que obtuvo el menor índice de precisión en el estudio (68 %). En particular, debe ser actualizada con el detalle de los códigos de las asignaturas, equivalencias de los planes de estudios y la información del ciclo vacacional, los cuales explican parte de los errores de respuesta que presenta el sistema. También es aconsejable revisar el mecanismo de enrutamiento del sistema, en especial lo relacionado a los temas de Nivel 3, y ajustar los umbrales de confianza en los temas de Nivel 2 con el objetivo de reducir la cantidad de consultas que no recibieron respuesta satisfactoria.

Del mismo modo, se propone iniciar estudios concretos que analicen empíricamente las capacidades multimodales del sistema (entrada por voz e imagen) y que evalúen su impacto en la accesibilidad, la usabilidad y la satisfacción del usuario, sobre todo en contextos donde el alumnado presenta problemas de interacción a partir de texto escrito en dispositivos móviles.

Por último, se sugiere la posibilidad de extender el estudio hacia las diferentes escuelas profesionales de la UNSCH, haciendo uso de la contextualización que admite la arquitectura multinivel. Como futuro objeto de investigación, podrían investigarse nuevas modalidades de interacción o la aplicación de técnicas de aprendizaje automático que permitan garantizar la mejora del mismo y, en el mismo sentido, la incorporación de estudios longitudinales que analicen la satisfacción del usuario en el transcurso del tiempo.

BIBLIOGRAFÍA

Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. *Machine Learning with Applications*, 2, 100006.

<https://doi.org/10.1016/j.mlwa.2020.100006>

Arana, J., Collantes, R., & Mamani, R. (2021). *Modelo de chatbot basado en inteligencia artificial para incrementar la satisfacción del cliente en empresas de venta de alimentos, Callao 2021* [Tesis de pregrado, Universidad Nacional del Callao]. Repositorio Institucional UNAC.

<https://repositorio.unac.edu.pe/handle/20,500.12952/6087>

Bommasani, R., Hudson, D. A., Adeli, E., et al. (2021). On the opportunities and risks of foundation models. *Center for Research on Foundation Models*. <https://arxiv.org/pdf/2108.07258>

Casazola, O., Alfaro, G., Burgos, J., & Ramos, O. (2021). La usabilidad percibida de los chatbots sobre la atención al cliente en las organizaciones: una revisión de la literatura. *Interfases*, 14, 184–204. <https://doi.org/10.26439/INTERFASES2021.N014.5401>

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., ... & Xie, X. (2024). A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology*, 15(3), 1–45. <https://doi.org/10.1145/3641289>.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2.^a ed.). Lawrence Erlbaum Associates.

DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: A ten-year update. *Journal of Management Information Systems*, 19(4), 9–30.

<https://doi.org/10.1080/07421222.2003.11045748>

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). GPTs are GPTs: An early look at the labor market impact potential of large language models. *arXiv*. <https://arxiv.org/abs/2303.10130>

Espinosa Peña, R. A., & Castillo Diaz, M. A. (2024). Implementación un chatbot basado en analítica de datos para la disminución de tiempos de respuesta y mejorar la calidad del servicio de la Corporación Unificada Nacional de Educación Superior [Tesis de pregrado, Universidad Cooperativa de Colombia]. Repositorio Institucional UCC. <https://hdl.handle.net/20,500.12494/55165>.

Espinosa-Luna, B. H., Mendoza-De-los-Santos, A., et al. (2023). Implementación de un chatbot basado en modelo de lenguaje de inteligencia artificial. *Revista Científica de Sistemas e Informática*, 3(2), e570.

George, D., & Mallery, P. (2003). *SPSS for Windows step by step* (4.^a ed.). Allyn & Bacon.

Hernández-Sampieri, R., Fernández-Collado, C., & Baptista-Lucio, P. (2014). Metodología de la investigación (6.^a ed.). McGraw-Hill.

International Organization for Standardization. (2011). *ISO/IEC 25010: Systems and software quality models*. Geneva: ISO.

International Organization for Standardization. (2023). ISO/IEC 25010:2023 — Systems and software quality requirements and evaluation (SQuaRE) — Product quality model. ISO.

Johnson, J., Douze, M., & Jégou, H. (2021). Billion-scale similarity search with GPUs. *IEEE Transactions on Big Data*, 7(3), 535–547.

Lewis, P., Perez, E., Piktus, A., et al. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *NeurIPS*, 33, 9459–9474.

López Aguirre, J. A., & Segura Carmona, F. A. (2023). Chatbot para mejorar la calidad de atención de consultas de grados y títulos de la USMP-ICED [Tesis de pregrado, Universidad de San Martín de Porres]. Repositorio Institucional USMP. <https://repositorio.usmp.edu.pe/handle/20,500.12727/12388>.

Luger, E., & Sellen, A. (2016). Like having a really bad PA. *CHI Proceedings*, 5286–5297.

Manning, C. D., Raghavan, P., & Schütze, H. (2009). *Introduction to information retrieval*. Cambridge University Press.

Mayon, J., & Ruiz, M. (2024). Desarrollo de una aplicación de orientación vocacional para bachilleres mediante test utilizando inteligencia artificial para la toma de decisiones [Trabajo de titulación, Universidad Técnica de Machala]. Repositorio Institucional UTMACH. https://repositorio.utmachala.edu.ec/bitstream/48000/23575/1/Trabajo_Titulacion_3439.pdf

Múnera, M., Salazar, L., & Osorio, A. (2022). Estudio inicial de un chatbot. *Investigación Bibliotecológica*, 36(90).

Naumann, F., & Herschel, M. (2010). An introduction to duplicate detection. Morgan & Claypool Publishers. <https://doi.org/10.2200/S00262ED1V01Y201003DTM003>.

Nielsen, J. (1993). Usability engineering. Academic Press.

Quispe Cruz, G. (2024). Chatbot y calidad de servicio de atención al egresado de la Universidad Peruana de Ciencias e Informática, Lima-2024 [Tesis de pregrado, Universidad Peruana de Ciencias e Informática]. Repositorio Institucional UPCI. <https://repositorio.upci.edu.pe/handle/upci/1163>.

Ramírez, S. (2018). FastAPI [Framework de software]. <https://fastapi.tiangolo.com>.

Ray, P. P. (2023). ChatGPT: A comprehensive review. *Internet of Things and Cyber-Physical Systems*, 3, 121–154.

Razali, N. M., & Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, 2(1), 21–33.

Saranya, A., & Subhashini, R. (2023). A systematic review of Explainable Artificial Intelligence models and applications: Recent developments and future trends. *Decision Analytics Journal*, 7, 100230. <https://doi.org/10.1016/j.dajour.2023.100230>.

Serrano Franco, G. (2025). Desarrollo de un Chatbot para Mejorar la Tutoría en Carreras de Ingeniería: Un enfoque en la Universidad Politécnica Metropolitana de Hidalgo. *RIDE Revista Iberoamericana Para La Investigación Y El Desarrollo Educativo*, 16(31). <https://doi.org/10.23913/ride.v16i31.2496>.

Shawar, B. A., & Atwell, E. (2007). Chatbots: Are they really useful? *LDV Forum*, 22(1), 29–49. <https://jclcl.org/article/view/88/86>

Supabase. (2023). Supabase documentation. <https://supabase.com/docs>

Tavily. (2024). Tavily AI search API documentation. <https://docs.tavily.com>

Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. *NeurIPS*, 30.

Veliz Yamunaque, F. D., & Velásquez Granda, J. J. (2024). Implementación de un asistente virtual mediante chatbots para la atención de consultas de la gestión académica en la UNPRG - 2023 [Tesis de pregrado, Universidad Nacional Pedro Ruiz Gallo]. Repositorio Institucional UNPRG. <https://repositorio.unprg.edu.pe/handle/20,500.12893/13307>.

Villegas Espinoza, M. G., & Pitot Galarreta, A. S. (2023). Uso de chatbots para la digitalización del canal de atención a clientes B2B [Trabajo de suficiencia profesional, Universidad Peruana de Ciencias Aplicadas]. Repositorio Académico UPC. <http://hdl.handle.net/10757/671772>.

Winkler, M., & Söllner, M. (2018). Unleashing the potential of chatbots in education: A state-of-the-art analysis. *Academy of Management Annual Meeting Proceedings*, 1–6.

<https://journals.aom.org/doi/10.5465/AMBPP.2018.15903abstract>

You, E. (2022). Vue.js documentation. <https://vuejs.org>

ANEXOS

MATRIZ DE CONSISTENCIA

“Implementación de un chatbot para la optimización de la atención de consultas académicas y administrativas en la Escuela Profesional de Ingeniería de Sistemas de la UNSCH, 2025”				
PROBLEMAS	OBJETIVOS	HIPÓTESIS	VARIABLES	METODOLOGÍA
PROBLEMA GENERAL ¿En qué medida la implementación de un chatbot optimiza la atención de consultas académicas y administrativas en la Escuela Profesional de Ingeniería de Sistemas de la UNSCH, 2025?	OBJETIVO GENERAL Evaluar el efecto de implementar un chatbot que optimice la atención de consultas académicas y administrativas en la Escuela Profesional de Ingeniería de Sistemas de la UNSCH, 2025.	HIPÓTESIS GENERAL La implementación de un chatbot optimiza la atención de consultas académicas y administrativas en la Escuela Profesional de Ingeniería de Sistemas de la UNSCH, 2025.	VARIABLE INDEPENDIENTE X: Implementación de un chatbot DIMENSIONES X1: Funcionalidad X2: Fiabilidad X3: Usabilidad VARIABLE DEPENDIENTE Y: Optimización de la atención de consultas académicas y administrativas DIMENSIONES Y1: Eficacia Y2: Eficiencia Y3: Satisfacción	TIPO DE INVESTIGACIÓN Aplicada (tecnológica) NIVEL DE INVESTIGACIÓN Explicativo DISEÑO Pre-experimental, con un único grupo y medición pretest-posttest POBLACIÓN Está compuesta por 513 estudiantes matriculados (2025-II) MUESTRA La muestra total es de 220 estudiantes. TÉCNICA - Análisis Documental - Entrevista
PROBLEMAS ESPECÍFICOS 1. ¿En qué medida la funcionalidad del chatbot optimiza la atención de consultas académicas y administrativas?	OBJETIVOS ESPECÍFICOS 1. Evaluar si la funcionalidad del chatbot optimiza la atención de consultas académicas y administrativas. 2. Evaluar si la fiabilidad del chatbot optimiza la	HIPÓTESIS ESPECÍFICAS 1. La funcionalidad del chatbot optimiza la atención de consultas académicas y administrativas.		

2. ¿En qué medida la fiabilidad del chatbot optimiza la atención de consultas académicas y administrativas?	atención de consultas académicas y administrativas.	2. La fiabilidad del chatbot optimiza la atención de consultas académicas y administrativas.	- Encuesta -
3. ¿En qué medida la usabilidad del chatbot optimiza la atención de consultas académicas y administrativas?	3. Evaluar si la usabilidad del chatbot optimiza la atención de consultas académicas y administrativas.	3. La usabilidad del chatbot optimiza la atención de consultas académicas y administrativas.	<i>INSTRUMENTO</i> - Ficha de análisis documental - Guía de entrevista - Cuestionario

MATRIZ DE OPERACIONALIZACIÓN DE VARIABLES

“Implementación de un chatbot para la optimización de la atención de consultas académicas y administrativas en la Escuela Profesional de Ingeniería de Sistemas de la UNSCH, 2025”			
VARIABLE DE ESTUDIO	DEFINICIÓN OPERACIONAL	DIMENSIONES	INDICADORES
X: Implementación de un chatbot	Grado en que el chatbot cumple con criterios de calidad del software, evaluado mediante métricas de precisión de respuestas, estabilidad del sistema y facilidad de uso percibida por los usuarios.	Funcionalidad (X1)	0–100%. Porcentaje de respuestas correctas frente a consultas reales
		Fiabilidad (X2)	0–100%. Tasa de acierto en las respuestas (respuestas correctas vs. total de consultas)
			Tiempo medio entre fallos (medido en segundos o minutos)
Y: Optimización de la atención de consultas académicas y administrativas	Nivel de mejora en el proceso de atención de consultas, medido a través del cumplimiento del objetivo de la consulta, el tiempo de respuesta y la satisfacción del usuario.	Usabilidad (X3)	Nivel de percepción de usabilidad (facilidad, claridad, intuitividad). 1–5 (Totalmente en desacuerdo – Totalmente de acuerdo)
		Eficacia (Y1)	Nivel de cumplimiento de la consulta (precisión, utilidad, completitud). 1–5 (Totalmente en desacuerdo – Totalmente de acuerdo)
		Eficiencia (Y2)	Tiempo promedio de atención por consulta (medido en segundos o minutos)
			Percepción de rapidez y esfuerzo en la atención. 1–5 (Totalmente en desacuerdo – Totalmente de acuerdo)
		Satisfacción (Y3)	Nivel de satisfacción del usuario (experiencia, recomendación, cumplimiento de expectativas) 1–5 (Totalmente en desacuerdo – Totalmente de acuerdo)

IMÁGENES DE LA INTERFAZ DEL CHATBOT Y DEL EVALUADOR

Figura 17

Vista inicial del chatbot (PC)

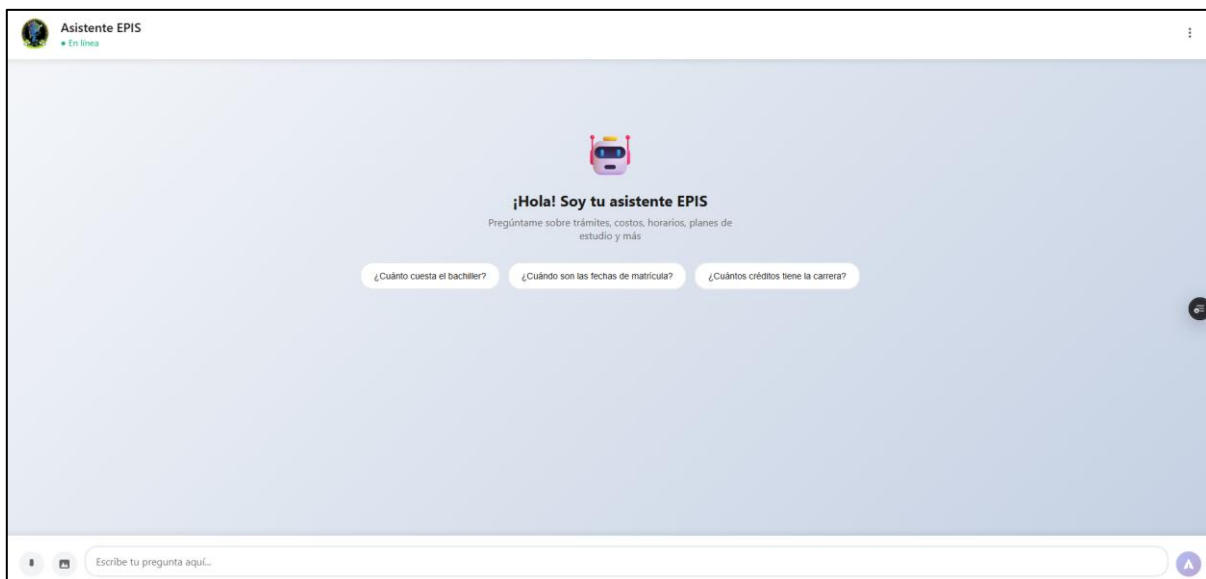


Figura 18

Vista de la interacción con el chatbot (PC)



Figura 19

Vista inicial del chatbot (Móviles)

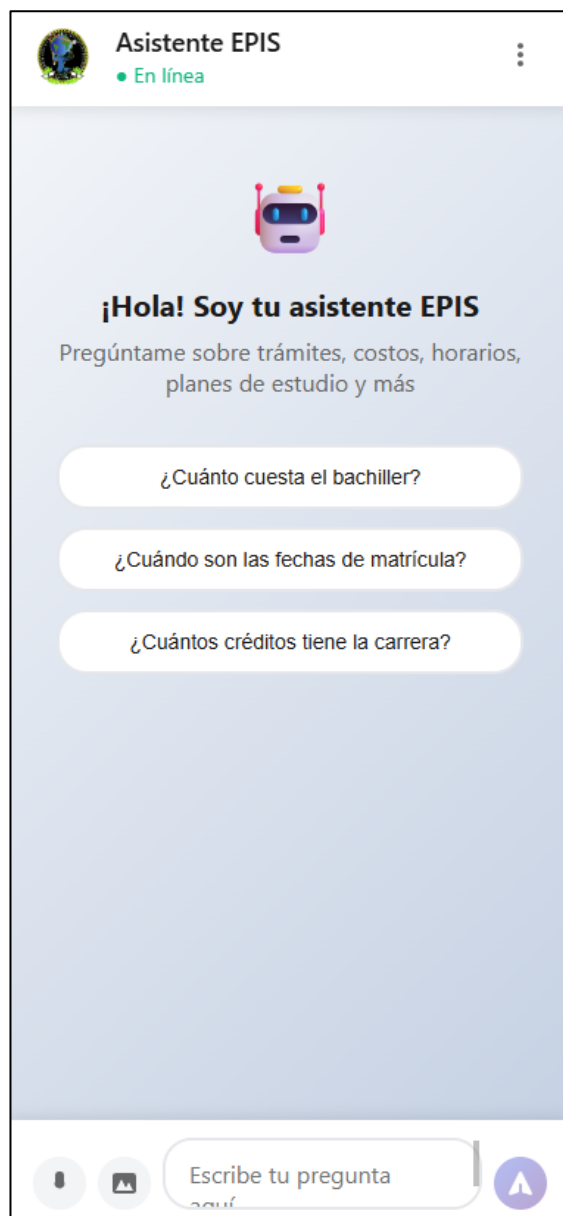
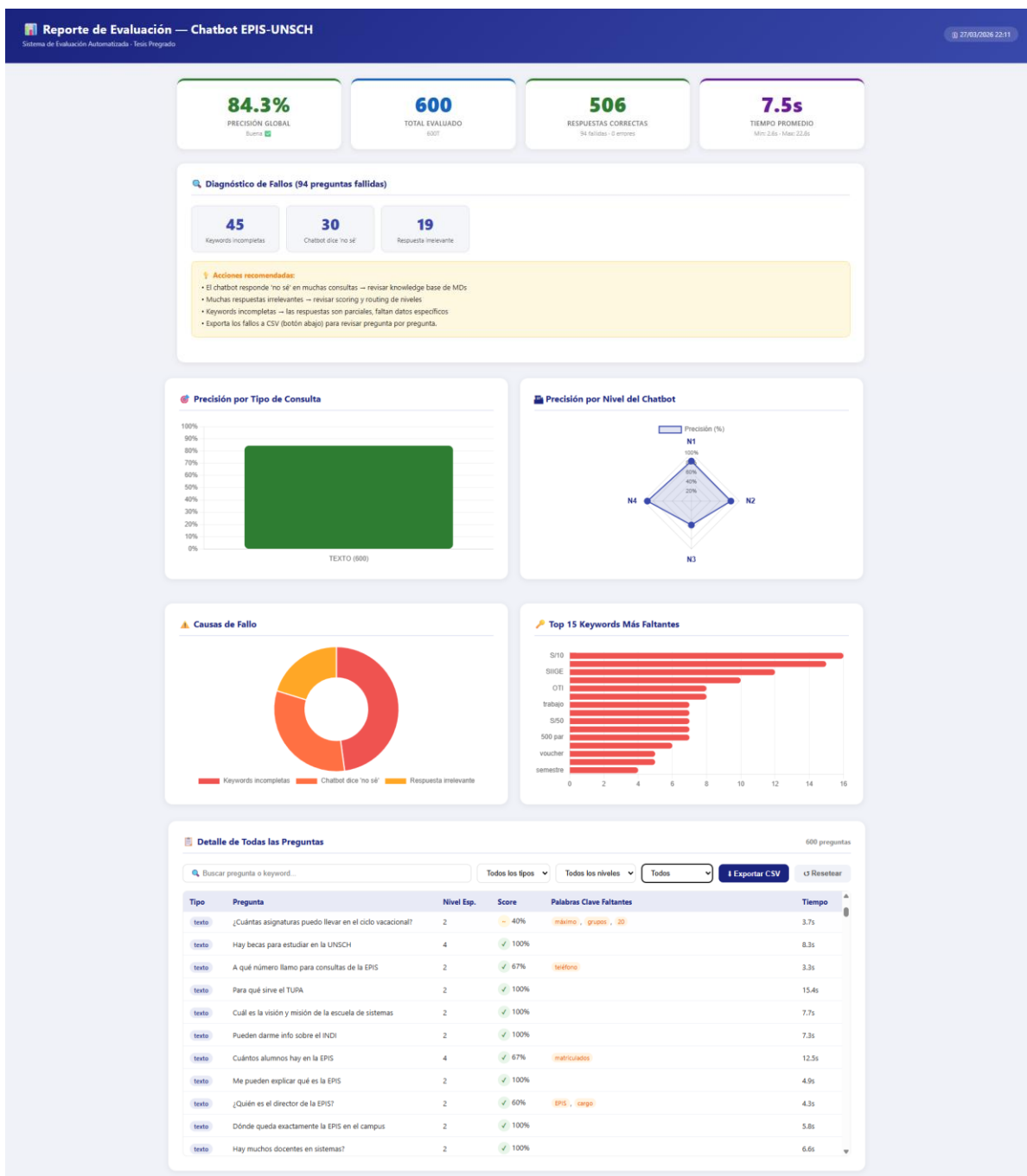


Figura 20

Vista del Reporte del Evaluador de Consultas del Chatbot



CUESTIONARIO BASADO EN LA ISO/IEC 25010 (PRETEST)

Evaluación de Canales de Atención Tradicionales - EPIS

IMPORTANTE: Esta encuesta evalúa tu experiencia ACTUAL al buscar información o resolver consultas académicas y administrativas en la EPIS, considerando los canales de atención que habitualmente utilizas.

Las mismas preguntas se aplicarán nuevamente en la segunda encuesta, por lo que es fundamental que respondas basándote únicamente en tu experiencia con los servicios que utilizas actualmente.

CORREO (Obligatorio):

SECCIÓN 1: DATOS GENERALES

GÉNERO (Marca solo uno):

☐	Masculino
☐	Femenino

SEMESTRE ACADÉMICO QUE CURSA ACTUALMENTE (Marca solo uno):

☐	100
☐	200
☐	300
☐	400
☐	500

SECCIÓN 2: PERCEPCIÓN SOBRE LOS CANALES ACTUALES DE ATENCIÓN

Por favor, responde las siguientes afirmaciones según tu grado de acuerdo.

Escala:

1 = Totalmente en desacuerdo, 5 = Totalmente de acuerdo.

Nota:

- "Académicas" se refiere a temas como horarios, cursos, docentes, prácticas, etc.
- "Administrativas" incluye trámites como matrículas, constancias, certificados, pagos, etc.
- "Institucionales" comprende toda la información oficial de la EPIS disponible en páginas web.

Dimensión	Nº	Ítem / Afirmación	1	2	3	4	5
Eficacia	1	La información que recibo al hacer una consulta es correcta y precisa.					
	2	La respuesta que obtengo es útil y relevante para mi pregunta.					
	3	Consigo la información completa que necesito para resolver mi duda.					
	4	Confío en la veracidad de la información que se me proporciona.					
Eficiencia	5	Resolver mis dudas es rápido.					
	6	Encuentro la información que busco sin invertir mucho tiempo ni esfuerzo.					
	7	El tiempo que invierto para obtener una respuesta es razonable.					
	8	Los pasos para realizar una consulta son sencillos y directos.					
Satisfacción	9	En general, estoy satisfecho(a) con el servicio de atención a consultas.					
	10	La experiencia de buscar ayuda o información es satisfactoria.					
	11	Recomendaría este servicio de consultas a otros estudiantes.					
	12	Este servicio cumple con mis expectativas para resolver consultas académicas y administrativas.					
Usabilidad	13	El servicio que utilizo para hacer consultas es fácil de usar.					
	14	Las instrucciones o respuestas que recibo son claras y fáciles de entender.					
	15	Iniciar una consulta es un proceso claro e intuitivo.					
	16	Cuando la consulta no avanza como esperaba, es fácil corregir el rumbo o intentar de nuevo.					

CUESTIONARIO BASADO EN LA ISO/IEC 25010 (POSTEST)

Evaluación del Servicio de Atención con Chatbot - EPIS

IMPORTANTE: Esta encuesta evalúa tu experiencia usando el chatbot de la EPIS para consultas académicas y administrativas.

Por favor, responde basándote únicamente en tu experiencia con el chatbot, considerando las consultas que hayas realizado a través de esta nueva herramienta.

CORREO (Obligatorio):

SECCIÓN 1: DATOS GENERALES

GÉNERO (Marca solo uno):

☐	Masculino
☐	Femenino

SEMESTRE ACADÉMICO QUE CURSA ACTUALMENTE (Marca solo uno):

☐	100
☐	200
☐	300
☐	400
☐	500

SECCIÓN 2: PERCEPCIÓN SOBRE LA ATENCIÓN CON EL CHATBOT

Por favor, responde las siguientes afirmaciones según tu grado de acuerdo.

Escala:

1 = Totalmente en desacuerdo, 5 = Totalmente de acuerdo.

Nota:

- "Académicas" se refiere a temas como horarios, cursos, docentes, prácticas, etc.
- "Administrativas" incluye trámites como matrículas, constancias, certificados, pagos, etc.
- "Institucionales" comprende toda la información oficial de la EPIS disponible en páginas web.

Dimensión	Nº	Ítem / Afirmación	1	2	3	4	5
Eficacia	1	La información que recibo al hacer una consulta es correcta y precisa.					
	2	La respuesta que obtengo es útil y relevante para mi pregunta.					
	3	Consigo la información completa que necesito para resolver mi duda.					
	4	Confío en la veracidad de la información que se me proporciona.					
Eficiencia	5	Resolver mis dudas es rápido.					
	6	Encuentro la información que busco sin invertir mucho tiempo ni esfuerzo.					
	7	El tiempo que invierto para obtener una respuesta es razonable.					
	8	Los pasos para realizar una consulta son sencillos y directos.					
Satisfacción	9	En general, estoy satisfecho(a) con el servicio de atención a consultas.					
	10	La experiencia de buscar ayuda o información es satisfactoria.					
	11	Recomendaría este servicio de consultas a otros estudiantes.					
	12	Este servicio cumple con mis expectativas para resolver consultas académicas y administrativas.					
Usabilidad	13	El servicio que utilizo para hacer consultas es fácil de usar.					
	14	Las instrucciones o respuestas que recibo son claras y fáciles de entender.					
	15	Iniciar una consulta es un proceso claro e intuitivo.					
	16	Cuando la consulta no avanza como esperaba, es fácil corregir el rumbo o intentar de nuevo.					

GUÍA DE ENTREVISTA

Título: “Implementación de un chatbot para la optimización de la atención de consultas académicas y administrativas en la Escuela Profesional de Ingeniería de Sistemas de la UNSCH, 2025”

Indicaciones: El presente instrumento busca recopilar su opinión sobre los procesos de atención académica y administrativa de la Escuela de Ingeniería de Sistemas, con el fin de identificar oportunidades de mejora mediante la implementación de un chatbot.

Entrevistado/a: _____

Cargo: _____

Objetivo General:

Evaluar el efecto de implementar un chatbot que optimice la atención de consultas académicas y administrativas en la Escuela Profesional de Ingeniería de Sistemas de la UNSCH, 2025.

Objetivos Específicos:

1. Evaluar si la precisión de las respuestas del chatbot optimiza la atención de consultas académicas y administrativas.
2. Evaluar si el tiempo de respuesta del chatbot optimiza la atención de consultas académicas y administrativas.
3. Evaluar si la facilidad de uso del chatbot optimiza la atención de consultas académicas y administrativas.

Bloques de preguntas:

Objetivo Específico 1: Precisión de las respuestas

1. ¿Qué tipo de información solicitan con mayor frecuencia los estudiantes y docentes en secretaría?

2. ¿Qué dudas suelen repetirse y podrían responderse de forma automatizada?
3. ¿Qué tan importante considera que sea la exactitud de la información brindada por un chatbot en este contexto?

Objetivo Específico 2: Tiempo de respuesta

4. ¿Cuánto tiempo en promedio le toma responder las consultas más comunes de los estudiantes?
5. ¿Cuáles son los trámites o consultas que generan más demoras?
6. ¿Cree que un sistema automatizado reduciría los tiempos de espera de los estudiantes?

Objetivo Específico 3: Facilidad de uso

7. ¿Qué herramientas o medios de contacto emplean los estudiantes actualmente para solicitar información (por ejemplo, correo, ventanilla, redes sociales, otros)?

8. ¿Qué dificultades tienen los estudiantes para acceder a información administrativa?
9. ¿Qué características debería tener un chatbot para que sea fácil de usar para los estudiantes y docentes?

Expectativas y riesgos

10. ¿Qué beneficios espera que el chatbot traiga a la secretaría y a los estudiantes?
11. ¿Qué riesgos o limitaciones cree que podrían presentarse?
12. Si pudiera elegir tres funciones prioritarias que el chatbot debería ofrecer, ¿cuáles serían?

ENTREVISTADOR

DNI:

ENTREVISTADO

DNI:



ACTA DE SUSTENTACIÓN DE TESIS N° 43-2026-FIMGC

PARA OPTAR EL TÍTULO PROFESIONAL DE INGENIERA DE SISTEMAS

En la Universidad Nacional de San Cristóbal de Huamanga, en la ciudad de Ayacucho, en cumplimiento a la **Resolución Decanal No 225-2026-FIMGC-D**, a los ocho días del mes de julio de **2026**, siendo las **10:33 a.m.**, reunidos en el **Auditorio de la Escuela Profesional de Ingeniería de Minas**, bajo la presidencia del **MSc. José Ernesto ESTRADA CARDENAS**, y los miembros: **Mg. Ing^a. Celia Edith MARTINEZ CORDOVA** y **Dr. Manuel Avelino LAGOS BARZOLA**, actuando como secretario docente el **Ing. Saul Walter RETAMOZO FERNANDEZ**, para proceder a la sustentación de tesis para optar el **Título Profesional de Ingeniera de Sistemas**, de la Bachiller en Ingeniería de Sistemas:

MELANIE STEPHANY HERRERA SANTA CRUZ

Quien presentó la tesis denominada:

"Implementación de un chatbot para la optimización de la atención de consultas académicas y administrativas en la Escuela Profesional de Ingeniería de Sistemas de la UNSCH, 2025"

Los señores miembros del jurado luego de expuesta la tesis y absueltas las preguntas, deliberaron y declararon:

Aprobado con 17 (diecisiete)

Siendo las **11:46 a.m.** del día **8 de julio del 2026**, culmina el acto de sustentación de tesis, y en conformidad de lo actuado los miembros del jurado firmamos al pie del presente.

MSc. José Ernesto ESTRADA CARDENAS
Presidente

Dr. Manuel Avelino LAGOS BARZOLA
Miembro

Mg. Ing^a. Celia Edith MARTINEZ CORDOVA
Miembro - Asesor

Ing. Saul Walter RETAMOZO FERNANDEZ
Secretario docente de la FIMGC



UNSCH

FACULTAD DE
INGENIERÍA
DE MINAS, GEOLOGÍA Y CIVIL



CONSTANCIA DE ORIGINALIDAD DE TRABAJO DE INVESTIGACIÓN

CONSTANCIA N° 039-2026-KPS-FIMGC/UNSCH

El que suscribe; responsable verificador de originalidad de trabajos de tesis de pregrado con el software Turnitin, en segunda instancia para las **Escuelas Profesionales** de la **Facultad de Ingeniería de Minas, Geología y Civil**; en cumplimiento a la **Resolución de Consejo Universitario N° 039-2021-UNSCH-CU**, Reglamento de Originalidad de Trabajos de Investigación de la Universidad Nacional San Cristóbal de Huamanga y **Resolución Decanal N° 697-2024-FIMGC-D**, deja constancia de originalidad de trabajo de investigación, que el/la Sr./Srta.

Nombres y Apellidos : Melanie Stephany Herrera Santa Cruz

Escuela Profesional : INGENIERÍA DE SISTEMAS

Título de la Tesis : "Implementación de un chatbot para la optimización de la atención de consultas académicas y administrativas en la Escuela Profesional de Ingeniería de Sistemas de la UNSCH, 2025"

Evaluación de la Originalidad : 5% Índice de Similitud

Identificador de la entrega : 2999311000

Por tanto, según los Artículos 12, 13 y 17 del Reglamento de Originalidad de Trabajos de Investigación, es **PROCEDENTE** otorgar la **Constancia de Originalidad** para los fines que crea conveniente.

En señal de conformidad y verificación se firma la presente constancia

Ayacucho, 23 de julio de 2026



Firmado digitalmente por:
PERALTA SOTOMAYOR Karel
FAU 20143680754 soft
Motivo: Soy el autor del documento
Fecha: 23/07/2026 08:23:20-0500

“Implementación de un chatbot para la optimización de la atención de consultas académicas y administrativas en la Escuela Profesional de Ingeniería de Sistemas de la UNSCH, 2025”

por Melanie Stephany Herrera Santa Cruz

"Implementación de un chatbot para la optimización de la atención de consultas académicas y administrativas en la Escuela Profesional de Ingeniería de Sistemas de la UNSCH, 2025"

INFORME DE ORIGINALIDAD

5%	4%	4%	4%
INDICE DE SIMILITUD	FUENTES DE INTERNET	PUBLICACIONES	TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1	Submitted to Universidad Nacional de San Cristóbal de Huamanga	1%
	Trabajo del estudiante	
2	repositorio.uns.edu.pe	<1%
	Fuente de Internet	
3	Submitted to Universidad Privada del Norte	<1%
	Trabajo del estudiante	
4	sedici.unlp.edu.ar	<1%
	Fuente de Internet	
5	repositorio.uap.edu.pe	<1%
	Fuente de Internet	
6	repositorio.upci.edu.pe	<1%
	Fuente de Internet	
7	repositorio.eesppjsco.edu.pe	<1%
	Fuente de Internet	
8	Ernesto Díaz-Roncero, Carlos Joel C. Inocente-Accilio. "Power Apps para gestión de reportes y mantenimiento en el sector agrícola de Perú: eficiencia y transformación digital", AiBi Revista de Investigación, Administración e Ingeniería, 2026	<1%
	Publicación	
9	repositorio.upt.edu.pe	<1%
	Fuente de Internet	
10	Submitted to Universidad Peruana Los Andes	<1%
	Trabajo del estudiante	
11	www.ride.org.mx	<1%
	Fuente de Internet	
12	repositorio.unach.edu.pe	<1%
	Fuente de Internet	
13	Melgarejo Bolivar, Romel Percy. "Modelo especializado de soporte a repositorios institucionales del Perú, basado en la experiencia de sus administradores para la	<1%

indexación nacional e internacional", Universidad Nacional del Altiplano de Puno (Peru)

Publicación

14	Submitted to Pontificia Universidad Catolica del Peru	<1 %
	Trabajo del estudiante	
15	Submitted to Universidad Tecnologica del Peru	<1 %
	Trabajo del estudiante	
16	rad.ort.edu.uy	<1 %
	Fuente de Internet	
17	Submitted to Universidad TecMilenio	<1 %
	Trabajo del estudiante	
18	riunet.upv.es	<1 %
	Fuente de Internet	

Excluir citas

Activo

Excluir coincidencias

< 30 words

Excluir bibliografía

Activo