

UNIVERSIDAD TECNOLÓGICA DE LOS ANDES

FACULTAD DE INGENIERÍA

**ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E
INFORMÁTICA**



Tesis

Implementación de un chatbot con inteligencia artificial para la gestión
de información académica en la Universidad Tecnológica de los
Andes, 2025

Asesor:

Mg. Tintaya Zegarra, Elvio

Autor:

Pacheco Roca, Luis Enrique

Para optar el Título Profesional de:

Ingeniero de Sistemas e Informática

Abancay - Apurímac – Perú

2026

Acta de sustentación



UNIVERSIDAD TECNOLÓGICA DE LOS ANDES
FACULTAD DE INGENIERÍA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E INFORMÁTICA



Acta N°: 07

ACTA DE SUSTENTACIÓN DE TÍTULO PROFESIONAL

En la ciudad de Abancay, a los 16 días del mes de julio del 2026, siendo las 11:00 horas, se reunieron los integrantes del Jurado designado por Resolución Directoral N° 029-2026-UTEA-EPISEI de la Escuela Profesional de Ingeniería de sistemas E Informática, Facultad de Ingeniería:

Presidente	:	Mg. Edison Chiclla Carrasco
Dictaminante	:	Mg. Berly Loayza Velasque
Replicante	:	Mg. Fredy Soria Donaires

Para evaluar la sustentación, en la modalidad de:

☒ Tesis ☐ Trabajo de suficiencia profesional

Titulada:

Implementación de un chatbot con inteligencia artificial para la gestión de información académica en la Universidad Tecnológica de los Andes, 2025

Desarrollado por el (los) Bachiller (es):

Br.: Pacheco Roca, Luis Enrique
(Apellidos y Nombres)
Br.: _____
(Apellidos y Nombres)

Para optar el Título Profesional de:

Ingeniero de Sistemas e Informática

(Denominación del Título)

Concluido el acto, el Jurado dictaminó que el (la) (los) mencionado(a) (s) bachiller (es) fue (ron) APROBADO(S) ☒ o DESAPROBADO(S) ☐:

Emitiéndose el calificativo final de:

Bachiller (Apellidos y Nombres)	Nota	Detalle (**)
Br. Pacheco Roca, Luis Enrique	17	Diecisiete

Siendo las 12.45 horas concluyó la sesión, firmando los integrantes del Jurado.

Presidente: Mg. Chiclla Carrasco Edison
(Dr. Mg.). (Apellidos y Nombres)

Dictaminante: Mg. Loayza Velasque Berly
(Dr. Mg.). (Apellidos y Nombres)

Replicante: Mg. Soria Donaires Fredy
(Dr. Mg.). (Apellidos y Nombres)

(Firma)
(Firma)
(Firma)

Nota: Desaprobado: 0-10; Aprobado: 11-20

(**): 0 a 10: Desaprobado, 11 a 15: Aprobado, 16 a 18: Aprobado Notable, 19 y 20: Aprobado con Distinción, Art. 18 RGGAT.

Reporte de similitud



Página 1 de 257 - Portada

Identificador de la entrega: trn:oid::3117631155460

Luis Enrique PACHECO ROCA

PACHECO ROCA, Luis Enrique , Implementación de un chatbot con inteligencia artificial para la gestión de información acad...

 REVISIÓN-TESIS-C/D

Detalles del documento

Identificador de la entrega

trn:oid::3117631155460

Fecha de entrega

28 sept 2026, 11:14 GMT-5

Fecha de descarga

28 sept 2026, 11:35 GMT-5

Nombre del archivo

PACHECO ROCA, Luis Enrique , Implementación de un chatbot con inteligencia artificial para la docx

Tamaño del archivo

14.8 MB

236 páginas

42.531 palabras

245.809 caracteres



Página 1 de 257 - Portada

Identificador de la entrega: trn:oid::3117631155460

14% Similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para ca...

Filtrado desde el informe

- Bibliografía
- Texto citado
- Texto mencionado
- Coincidencias menores (menos de 8 palabras)

Exclusiones

- N.º de fuente excluida

Fuentes principales

- 11%  Fuentes de Internet
- 4%  Publicaciones
- 12%  Trabajos entregados (trabajos del estudiante)

Marcas de integridad

N.º de alertas de integridad para revisión

Los algoritmos de nuestro sistema analizan un documento en profundidad para buscar inconsistencias que permitirían distinguirlo de una entrega normal. Si advertimos algo extraño, lo marcamos como una alerta para que pueda revisarlo.

Una marca de alerta no es necesariamente un indicador de problemas. Sin embargo, recomendamos que preste atención y la revise.

Metadatos

DATOS DEL AUTOR	
Apellidos y nombres	PACHECO ROCA, Luis Enrique
Tipo de Documento de Identidad	DNI
Número de Documento de Identidad	72630847
URL ORCID	https://orcid.org/0009-0009-7758-1994
DATOS DEL ASESOR	
Apellidos y nombres	Tintaya Zegarra, Elvio
Tipo de Documento de Identidad	DNI
Número de Documento de Identidad	41544851
URL ORCID	https://orcid.org/0000-0003-0812-9565
DATOS DE LA INVESTIGACIÓN	
Facultad	Ingeniería
Escuela profesional	Ingeniería de Sistemas e Informática
Línea de Investigación	Informática, Sociedad y Gestión del Conocimiento
Rango de años en que se realizó la investigación	Octubre del 2025 - mayo del 2026
Fuente de financiamiento	Autofinanciado
Porcentaje de similitud	14%
URL de OCDE	https://purl.org/pe-repo/ocde/ford#2.02.04

Dedicatoria

Al Todopoderoso, por iluminar mi sendero, por sostenerme con firmeza en aquellos instantes de vacilación y por concederme la gracia indispensable para alcanzar esta meta.

A mis progenitores, cuyo cariño, dedicación cotidiana y fe inquebrantable hicieron posible que yo llegara hasta este punto. Cada uno de sus sacrificios tiene su propio nombre escrito entre estas páginas.

A mis seres queridos, por confiar en mí incluso en aquellas etapas en las que yo mismo flaqueaba. Su apoyo discreto se convirtió en el impulso que me mantuvo en marcha cuando la ruta se tornaba complicada.

Agradecimientos

Doy gracias a Dios, quien me otorgó la existencia, el bienestar, la fe y el ánimo indispensables para transitar este extenso trayecto y conquistar una de las metas más significativas de mi trayectoria profesional.

A mis padres, por el cariño sin reservas, las lecciones de vida y ese esmero discreto con el que respaldaron cada momento de mi formación; sin su entrega y desprendimiento, este logro habría sido inalcanzable.

A mi asesor, Mg. Elvio Tintaya Zegarra, mi reconocimiento por su compromiso, orientación académica y por los saberes compartidos durante la elaboración de esta investigación, cuyo acompañamiento resultó decisivo para los resultados obtenidos.

A mis seres queridos, por el respaldo permanente y el sustento inalterable que me permitieron avanzar y vencer cada uno de los obstáculos surgidos a lo largo de esta etapa de formación.

Mi reconocimiento también a la Universidad Tecnológica de los Andes, que me abrió las puertas para el desarrollo de mi carrera profesional, así como al cuerpo docente, por los conocimientos y vivencias compartidos durante mi paso por las aulas universitarias.

Resumen

El objetivo fue determinar el efecto de un chatbot con inteligencia artificial en la gestión de información académica en la Universidad Tecnológica de los Andes, 2025. Se asumió un enfoque cuantitativo, de tipo aplicada, nivel explicativo y diseño cuasi-experimental de grupo control no equivalente, con pretest y posttest. La población fue de 3 983 estudiantes de pregrado del semestre 2025-II en la sede Abancay; la muestra, de 380 estudiantes (190 experimental, 190 control), con selección no probabilística: el grupo experimental fue un conglomerado intacto de la Escuela Profesional de Estomatología y el control, estudiantes de las once escuelas restantes. El instrumento fue un cuestionario Likert de cinco niveles con 42 ítems, validado por tres expertos (V de Aiken = 1,00) y confiabilidad en prueba piloto ($n = 30$) de $\alpha = 0,920$ y $\alpha = 0,907$. Kolmogorov-Smirnov reveló ausencia de normalidad ($p < ,001$), por lo que se recurrió a pruebas no paramétricas. Wilcoxon evidenció diferencias significativas entre pretest y posttest del grupo experimental ($Z = -11,587$; $p < ,001$; $r = 0,845$) y ningún cambio significativo en el grupo control ($Z = -1,864$; $p = ,062$); Mann-Whitney confirmó la superioridad del grupo experimental ($Z = -10,731$; $p < ,001$; $r = 0,550$), con tamaños de efecto grandes. Se concluye que el chatbot generó un efecto positivo y estadísticamente significativo en la gestión de información académica, elevando la disponibilidad, el acceso y la eficiencia operativa del servicio universitario.

Palabras clave: chatbot, eficiencia operativa, gestión académica, inteligencia artificial, servicio universitario.

Abstract

In this research, the aim was to determine the effect of an artificial intelligence chatbot on academic information management at the Universidad Tecnológica de los Andes, 2025. A quantitative approach was assumed, of applied type, explanatory level, under a quasi-experimental nonequivalent control group design with pretest and posttest. The population was 3,983 undergraduate students of the 2025-II semester at the Abancay campus; the sample, 380 students (190 experimental, 190 control), with non-probability selection: the experimental group was an intact cluster from the Professional School of Stomatology and the control group, students from the remaining eleven schools. The instrument was a five-level Likert questionnaire with 42 items, validated by three experts (Aiken's $V = 1.00$) and reliability in a pilot test ($n = 30$) of $\alpha = 0.920$ and $\alpha = 0.907$. Kolmogorov-Smirnov indicated absence of normality ($p < .001$), for which reason nonparametric tests were used. Wilcoxon evidenced significant differences between pretest and posttest in the experimental group ($Z = -11.587$; $p < .001$; $r = 0.845$) and no significant change in the control group ($Z = -1.864$; $p = .062$); Mann-Whitney confirmed the superiority of the experimental group ($Z = -10.731$; $p < .001$; $r = 0.550$), with large effect sizes. It is concluded that the chatbot generated a positive and statistically significant effect on academic information management, raising the availability of information, access to accurate responses, and the operational efficiency of the university service.

Keywords: academic management, artificial intelligence, chatbot, operational efficiency, university service.

Índice

Portada.....	i
Acta de sustentación.....	i
Reporte de similitud	iii
Metadatos.....	v
Dedicatoria	vi
Agradecimientos	vii
Resumen	viii
Abstract	ix
Índice de tablas.....	xiii
Índice de figuras	xvi
Índice de anexos.....	xxi
I. Introducción	22
II. Planteamiento del problema	24
2.1. Descripción y formulación del problema	24
2.1.1. <i>Problema general</i>	29
2.1.2. <i>Problemas específicos</i>	29
2.2. Objetivos	30
2.2.1. <i>Objetivo general</i>	30
2.2.2. <i>Objetivos específicos</i>	30
2.3. Justificación e importancia	30
2.3.1. <i>Justificación teórica</i>	30
2.3.2. <i>Justificación práctica</i>	31
2.3.3. <i>Justificación metodológica</i>	32
2.3.4. <i>Justificación social</i>	33
2.3.5. <i>Importancia</i>	33
2.4. Hipótesis	34

2.4.1. Hipótesis general	34
2.4.2. Hipótesis específicas	34
2.5. Variables	35
2.5.1. Variable independiente	35
2.5.2. Variable dependiente	35
2.5.3. Operacionalización de variables	36
III. Marco Teórico	39
3.1. Antecedentes	39
3.1.1. Internacionales	39
3.1.2. Nacionales	41
3.1.3. Local	44
3.2. Bases teóricas.....	44
3.2.1. Chatbot con inteligencia artificial	44
3.2.2. Gestión de información académica	63
3.2.3. Fundamentos metodológicos y estadísticos.....	69
3.3. Definición de términos	73
IV. Metodología	80
4.1. Tipo y nivel de investigación.....	80
4.1.1. Tipo de investigación	80
4.1.2. Nivel de investigación	80
4.2. Ámbito temporal y espacial.....	82
4.3. Población y muestra.....	82
4.3.1. Población	82
4.3.2. Muestra.....	83
4.3.3. Muestreo.....	86
4.4. Instrumentos	87
4.4.1. Técnicas	87
4.4.2. Instrumentos	88

4.4.3. Evaluación de validez y fiabilidad.....	90
4.5. Procedimientos.....	92
4.5.1. Validación del instrumento.....	92
4.5.2. Aplicación del pretest.....	93
4.5.3. Recopilación documental institucional.....	93
4.5.4. Desarrollo e implementación del sistema chatbot	95
4.5.5. Aplicación del postest	119
4.5.6. Procesamiento estadístico de los datos	122
4.6. Análisis de datos	125
4.6.1. Análisis descriptivo	126
4.6.2. Análisis inferencial	126
4.6.3. Tamaño del efecto	127
4.7. Consideraciones éticas	127
V. Resultados y discusión.....	129
5.1. Resultados descriptivos.....	129
5.1.1. Caracterización de la muestra.....	129
5.1.2. Distribución de niveles por dimensión	130
5.2. Contrastación de hipótesis	133
5.2.1. Comprobación de la hipótesis general	135
5.2.2. Comprobación de las hipótesis específicas.....	140
5.3. Discusión de resultados	147
VI. Conclusiones	155
VII. Recomendaciones.....	156
VIII. Referencias bibliográficas	158
IX. Anexos	172

Índice de tablas

Tabla 1 Operacionalización de la variable independiente: Chatbot con inteligencia artificial	36
Tabla 2 Operacionalización de la variable dependiente: Gestión de información académica	37
Tabla 3 Coeficientes de fiabilidad alfa de Cronbach por escala para la variable independiente y la variable dependiente	91
Tabla 4 Stack tecnológico utilizado en la implementación del sistema chatbot	106
Tabla 5 Indicadores técnicos de desempeño del sistema durante el periodo de intervención	116
Tabla 6 Casos de prueba funcional ejecutados sobre el sistema Universidad Tecnológica de los Andes Bot.....	118
Tabla 7 Composición de la muestra por grupo de estudio	129
Tabla 8 Distribución de niveles, D1 Disponibilidad de la información (Grupo Experimental)	130
Tabla 9 Distribución de niveles, D2 Acceso a la información (Grupo Experimental).....	131
Tabla 10 Distribución de niveles, D3 Eficiencia operativa (Grupo Experimental).....	132
Tabla 11 Regla de decisión para pruebas de hipótesis	133
Tabla 12 Resultados de la prueba U de Mann-Whitney en el pretest: equivalencia inicial entre grupos ($n_1 = n_2 = 190$)	133
Tabla 13 Estadísticos descriptivos por grupo y momento de medición	134
Tabla 14 Prueba de normalidad (Kolmogorov-Smirnov).....	135
Tabla 15 Rangos de Wilcoxon, Pretest y posttest, variable dependiente Total (grupo experimental).....	136
Tabla 16 Estadísticos de prueba, Wilcoxon, VD Total.....	137
Tabla 17 Rangos por grupo, U de Mann-Whitney, Delta VD Total (posttest – pretest)	137

Tabla 18 <i>Estadísticos de prueba, U de Mann-Whitney, Delta VD Total (postest – pretest)</i>	138
Tabla 19 <i>Rangos de Wilcoxon, Pretest y postest, variable dependiente Total (grupo control)</i>	138
Tabla 20 <i>Rangos de Wilcoxon, pretest y postest, percepción del chatbot, índice de 21 ítems (grupo experimental)</i>	139
Tabla 21 <i>Rangos de Wilcoxon, D1 Disponibilidad de la información (grupo experimental, pretest vs. postest)</i>	140
Tabla 22 <i>Estadísticos de prueba, Wilcoxon, D1 Disponibilidad</i>	140
Tabla 23 <i>Rangos por grupo, U de Mann-Whitney, D1 Disponibilidad (diferencia postest – pretest)</i>	141
Tabla 24 <i>Estadísticos de prueba, U de Mann-Whitney, D1 Disponibilidad (diferencia postest – pretest)</i>	141
Tabla 25 <i>Rangos de Wilcoxon, D2 Acceso a la información (grupo experimental, pretest vs. postest)</i>	142
Tabla 26 <i>Estadísticos de prueba, Wilcoxon, D2 Acceso</i>	142
Tabla 27 <i>Rangos por grupo, U de Mann-Whitney, D2 Acceso (diferencia postest – pretest)</i>	143
Tabla 28 <i>Estadísticos de prueba, U de Mann-Whitney, D2 Acceso (diferencia postest – pretest)</i>	143
Tabla 29 <i>Rangos de Wilcoxon, D3 Eficiencia operativa (grupo experimental, pretest vs. postest)</i>	144
Tabla 30 <i>Estadísticos de prueba, Wilcoxon, D3 Eficiencia operativa</i>	144
Tabla 31 <i>Rangos por grupo, U de Mann-Whitney, D3 Eficiencia operativa (diferencia postest – pretest)</i>	145
Tabla 32 <i>Estadísticos de prueba, U de Mann-Whitney, D3 Eficiencia operativa (diferencia postest – pretest)</i>	145
Tabla 33 <i>Síntesis de resultados inferenciales</i>	146

Tabla 34 <i>Matriz de consistencia del estudio</i>	173
Tabla 35 Expertos validadores del instrumento (juicio de expertos).....	181
Tabla 36 <i>Resumen de validación de contenido por juicio de expertos (V de Aiken)</i>	182
Tabla 37 <i>Stack tecnológico del sistema Universidad Tecnológica de los Andes Bot</i>	188
Tabla 38 Modelos LLM con mecanismo de fallback.....	189
Tabla 39 Baremo de interpretación: reagrupación de cinco categorías Likert en tres niveles	234
Tabla 40 Estudiantes de pregrado matriculados por sede y escuela profesional, semestres 2025-II y 2026-I	235

Índice de figuras

Figura 1 Comparación de los cuatro tipos de chatbots según arquitectura, mecanismo de respuesta y caso de uso representativo.....	49
Figura 2 Pipeline de Generación Aumentada por Recuperación (RAG): flujo de procesamiento desde la documentación institucional hasta la respuesta del chatbot	56
Figura 3 Evolución de los modelos de lenguaje de gran escala (LLM): principales familias y versiones (2017-2024)	60
Figura 4 Dimensiones de la variable dependiente: disponibilidad, acceso y eficiencia operativa de la gestión de información académica.....	65
Figura 5 Arquitectura general del sistema Universidad Tecnológica de los Andes Bot...	101
Figura 6 Diagrama del workflow Chatbot WhatsApp en n8n	104
Figura 7 Diagrama del workflow Push Data Vector en n8n	105
Figura 8 Diagrama de casos de uso del sistema Universidad Tecnológica de los Andes Bot	114
Figura 9 Diagrama de secuencia de la atención de una consulta académica	115
Figura 10 Diagrama de despliegue del sistema Universidad Tecnológica de los Andes Bot	115
Figura 11 Pantalla de inicio, interfaz web del chatbot Universidad Tecnológica de los Andes Bot.....	120
Figura 12 Conversación de ejemplo, consulta en el canal web	121
Figura 13 Canal WhatsApp Business, conversación de prueba	121
Figura 14 Base de datos vectorial Pinecone, índice chatbot-n8n	122
Figura 15 Comparativo de rangos promedio Mann-Whitney por dimensión y variable dependiente total (grupo experimental vs. grupo control, sobre la diferencia entre pretest y posttest).....	147
Figura 16 Cuestionario dirigido a estudiantes –Postest Experiencia después del uso chatbot.....	176

Figura 17 Cuestionario dirigido a estudiantes –Pretest expectativa antes del uso del chatbot.....	177
Figura 18 Cuestionario dirigido a estudiantes – Grupo de control -sin chatbot pretest y posttest.....	178
Figura 19 Cuestionario dirigido a estudiantes – Variable independiente	179
Figura 20 Cuestionario dirigido a estudiantes – Variable dependiente	180
Figura 21 Ficha de validación de instrumento – Experto 1	184
Figura 22 Ficha de validación de instrumento – Experto 2.....	185
Figura 23 Ficha de validación de instrumento – Experto 3.....	186
Figura 24 Declaración Jurada de Originalidad	187
Figura 25 Workflow Chatbot WhatsApp en n8n con sus 19 nodos.....	190
Figura 26 Workflow Push Data Vector en n8n.....	190
Figura 27 Panel de Pinecone: índice chatbot-n8n con vectores indexados.....	191
Figura 28 Conversación de prueba en canal web del chatbot Universidad Tecnológica de los Andes Bot	193
Figura 29 Conversación de prueba en canal WhatsApp Business	194
Figura 30 Resolución Directoral N° 077-2025-UTEA-FI-EPISel-FC (anverso) – Aprobación del proyecto	195
Figura 31 Resolución Directoral N° 077-2025-UTEA-FI-EPISel-FC (reverso) – Aprobación del proyecto	196
Figura 32 Resolución Directoral N° 0056-2025-UTEA-FI-EPISel (anverso) – Designación del asesor.....	197
Figura 33 Resolución Directoral N° 0056-2025-UTEA-FI-EPISel (reverso) – Designación del asesor.....	198
Figura 34 Recolección de información académica - sesión 1	199
Figura 35 Recolección de información académica - sesión 2.....	200
Figura 36 Recolección de información académica - sesión 3.....	200
Figura 37 Aplicación del cuestionario - aulas grupo experimental.....	201

Figura 38 <i>Aplicación del cuestionario - aulas grupo experimental</i>	201
Figura 39 <i>Aplicación del cuestionario - aulas grupo experimental</i>	202
Figura 40 <i>Aplicación del cuestionario - aulas grupo control</i>	202
Figura 41 <i>Aplicación del cuestionario - aulas grupo control</i>	203
Figura 42 <i>Aplicación del cuestionario - pretest</i>	203
Figura 43 <i>Aplicación del cuestionario - postest</i>	204
Figura 44 <i>Vista de datos en IBM SPSS Statistics versión 27: base importada con 760 casos y 50 variables</i>	206
Figura 45 <i>Resultados de la prueba U de Mann-Whitney sobre el pretest: equivalencia inicial entre grupos</i>	207
Figura 46 <i>Resultados de la prueba de Wilcoxon: comparación intragrupo experimental (pretest vs. postest)</i>	208
Figura 47 <i>Resultados de la prueba U de Mann-Whitney sobre las diferencias postest – pretest: comparación intergrupos</i>	209
Figura 48 <i>Captura del panel de Pinecone mostrando el índice chatbot-n8n con sus parámetros: región us-east-1, dimensión 3072, métrica coseno</i>	211
Figura 49 <i>Captura del panel de administración de n8n con el workflow Push Data Vector activo y configurado</i>	212
Figura 50 <i>Captura de la carpeta n8n-chatbot en Google Drive con los documentos institucionales cargados</i>	212
Figura 51 <i>Captura del log de ejecución del workflow Push Data Vector procesando un documento</i>	213
Figura 52 <i>Captura del panel de Pinecone mostrando los vectores resultantes del primer lote de ingesta</i>	214
Figura 53 <i>Captura del archivo package.json del proyecto Next.js mostrando las dependencias instaladas</i>	215
Figura 54 <i>Captura del archivo .env.local con los nombres de variables de entorno utilizadas</i>	216

Figura 55 Captura del archivo <i>app/api/chat/route.ts</i> mostrando la estructura del pipeline RAG.....	216
Figura 56 Captura del bloque de código del fallback automático entre modelos de Groq	220
Figura 57 Captura del middleware de rate limiting configurado	221
Figura 58 Captura del entorno local de desarrollo (<i>npm run dev</i>) mostrando el chatbot operativo.....	222
Figura 59 Captura del panel de Meta Developer Portal con la aplicación registrada	223
Figura 60 Captura del panel WhatsApp Business Cloud API con el número telefónico configurado	223
Figura 61 Captura del workflow Chatbot WhatsApp en n8n mostrando los diecinueve nodos.....	224
Figura 62 Captura del nodo AI Agent del workflow con la configuración de Gemini 2.5 Flash.....	224
Figura 63 Captura del nodo Not supported con el mensaje configurado	225
Figura 64 Captura de una conversación de prueba completa desde una cuenta WhatsApp	225
Figura 65 Captura del archivo <i>docker-compose.yml</i> mostrando los servicios definidos	226
Figura 66 Captura del archivo de configuración Nginx con los bloques server y la redirección HTTPS	227
Figura 67 Captura del certificado SSL/TLS de Let's Encrypt activo para el dominio <i>chatbot.neurocyborgia.es</i>	228
Figura 68 Captura del sistema operativo en producción accedido desde el dominio público	228
Figura 69 Panel de monitoreo de n8n: registro de ejecuciones del workflow	229
Figura 70 Panel de Pinecone: métricas de uso del índice chatbot-n8n	230
Figura 71 Panel de WhatsApp Business Manager: reporte de mensajes intercambiados	231

Figura 72 Solicitud de acceso a información institucional presentada por Mesa de Partes	232
Figura 73 <i>Carta de solicitud de autorización y facilidades para acceso a información institucional dirigida a la Rectora de la Universidad Tecnológica de los Andes</i>	233

Índice de Anexos

Anexo 1. <i>Matriz de consistencia</i>	173
Anexo 2. <i>Cuestionario dirigido a estudiantes</i>	176
Anexo 3. <i>Validación del instrumento por juicio de expertos (V de Aiken)</i>	181
Anexo 4. <i>Declaración Jurada de Originalidad</i>	187
Anexo 5. <i>Desarrollo e implementación del sistema</i>	188
Anexo 6. <i>Resolución Directoral de aprobación del proyecto de tesis</i>	195
Anexo 7. <i>Resolución Directoral de designación del asesor de tesis</i>	197
Anexo 8. <i>Panel fotográfico</i>	199
Anexo 9. <i>Procedimiento de análisis estadístico en IBM SPSS Statistics</i>	205
Anexo 10. <i>Manual de implementación y evidencias de desarrollo</i>	211
Anexo 11. <i>Documento institucional de recopilación de información</i>	231
Anexo 12. <i>Baremo de interpretación por niveles</i>	234
Anexo 13. <i>Reporte estadístico de estudiantes matriculados</i>	234

I. Introducción

En la actualidad, el avance digital ha modificado el modo en que los centros de educación superior organizan, procesan y brindan información a sus estudiantes. La incorporación de herramientas basadas en inteligencia artificial (IA) ha abierto nuevas oportunidades para mejorar los procesos académicos y administrativos; sin embargo, muchas instituciones aún no se encuentran plenamente preparadas para validar, regular ni incorporar dichas tecnologías de modo efectivo en sus servicios. En esa línea, Miao y Holmes (2023) advierten que el acelerado avance de la IA generativa rebasa la capacidad de adaptación de los marcos institucionales y normativos, lo que exige respuestas pedagógicas y organizacionales más sólidas.

En la Universidad Tecnológica de los Andes, esta brecha entre las demandas de información académica y la capacidad institucional para atenderlas se evidencia de manera concreta. Los estudiantes enfrentan largas colas y demoras en las ventanillas durante periodos críticos como matrículas, cronogramas de evaluación y trayectos de grado; la atención se concentra en horarios administrativos limitados, lo que deja sin respuesta a quienes consultan fuera de jornada; la información se encuentra dispersa entre múltiples canales (oficinas, página web, redes sociales) y, en ocasiones, resulta desactualizada o contradictoria entre dependencias; y los estudiantes de las filiales de Andahuaylas y Cusco, así como aquellos que residen en zonas rurales de la región Apurímac, no siempre pueden desplazarse a la sede principal de Abancay para resolver sus consultas presenciales.

Frente a esta realidad, los chatbots (asistentes virtuales conversacionales) se vienen consolidando como recursos con gran potencial para ofrecer atención inmediata, personalizada y continua a estudiantes y personal institucional (Crompton y Burke, 2023). Sumarlos a la gestión de la información académica dentro de la Universidad Tecnológica de los Andes permitiría optimizar el acceso, la precisión y la oportunidad de la información que reciben los usuarios (Okonkwo y Ade-Ibijola, 2021), responder consultas de manera

ágil, permanente y confiable, y atender a la comunidad estudiantil más allá del horario presencial y de la ubicación geográfica de cada estudiante.

El estudio se ubica dentro del tipo aplicada, con nivel explicativo y diseño cuasi-experimental de grupo control no equivalente (Hernández-Sampieri y Mendoza, 2018; Shadish et al., 2002), por lo que busca no solo describir el fenómeno, sino establecer si la intervención tecnológica produce mejoras significativas y verificables en la variable dependiente. Se midió mediante un cuestionario de escala ordinal tipo Likert, validado por juicio de expertos (V de Aiken = 1,00), y la confiabilidad fue estimada en prueba piloto ($n = 30$) con el alfa de Cronbach ($\alpha = 0,907$). El análisis estadístico no paramétrico empleó Wilcoxon para el contraste intragrupo y la U de Mann-Whitney para el contraste intergrupos, ya que los datos no satisficieron el supuesto de normalidad.

El aporte original del estudio reside en evaluar empíricamente, bajo un diseño cuasi-experimental de grupo control no equivalente con 380 estudiantes (190 por grupo), el efecto de un sistema de Generación Aumentada por Recuperación (RAG, Retrieval-Augmented Generation) (Lewis et al., 2020) acoplado a modelos de lenguaje grande (LLM, Large Language Model) de código abierto sobre la gestión de información académica en una universidad regional de la sierra sur del Perú, contexto subrepresentado en la literatura internacional sobre inteligencia artificial generativa en educación superior. Los hallazgos pretenden aportar evidencia cuantitativa que oriente decisiones de transformación digital en instituciones con perfiles socioeconómicos similares y alimente la discusión académica sobre la pertinencia de estas tecnologías en universidades de regiones andinas.

El informe se organiza en nueve capítulos: el Capítulo I introduce el estudio; el Capítulo II desarrolla el planteamiento del problema, objetivos, justificación, hipótesis y variables; el Capítulo III aborda el Marco Teórico; el Capítulo IV describe la metodología; el Capítulo V reúne resultados y discusión; los Capítulos VI, VII y VIII presentan conclusiones, recomendaciones y referencias; y el Capítulo IX recoge los anexos.

II. Planteamiento del problema

2.1. Descripción y formulación del problema

La gestión de información académica constituye un componente esencial del funcionamiento de las instituciones de educación superior, dado que comprende las etapas de generación, organización, almacenamiento, difusión y acceso a datos relacionados con matrículas, cronogramas, requisitos de grado, trámites administrativos y servicios estudiantiles (Laudon y Laudon, 2020). Cuando estos procesos operan de manera eficiente, los estudiantes acceden a información oportuna y precisa, lo que favorece su permanencia y desempeño académico; en cambio, cuando los canales de atención son limitados, la información se fragmenta, los plazos de respuesta se prolongan y la conformidad del usuario disminuye (Chiavenato, 2017). En este escenario, la integración de tecnologías sustentadas en inteligencia artificial, especialmente los chatbots (asistentes virtuales conversacionales), surge como recurso viable para optimizar la disponibilidad, el acceso y la eficiencia operativa en la gestión de información académica.

En el plano internacional, la digitalización del nivel de educación superior y el aprovechamiento de la inteligencia artificial generativa se aceleraron de manera notable en los últimos años (Lin et al., 2024). Miao y Holmes (2023) señalan que los chatbots con inteligencia artificial ofrecen beneficios tales como el ahorro de tiempo para el personal, el soporte continuo las 24 horas y la mejora de la experiencia estudiantil, aunque advierten sobre la necesidad de marcos de gobernanza de datos y criterios pedagógicos para evitar dependencias acríticas.

En Estados Unidos, un informe elaborado por la asociación EDUCAUSE, organización sin fines de lucro dedicada al avance de la educación superior mediante el uso de tecnologías de la información, en 2025 indica que el 37 % de las universidades ya proporciona licencias institucionales para chatbots y el 14 % ha implementado soluciones propias. La Universidad de Michigan lanzó Maizey, un chatbot institucional

que permite a docentes y personal crear asistentes de IA especializados sin programación, alimentándolos con sus propios documentos, y que dispone de más de 3 500 asistentes activos en departamentos académicos y administrativos. La Universidad Estatal de Arizona desplegó CreateAI Platform, un entorno seguro con acceso unificado a más de 50 modelos de IA generativa y la herramienta MyAI Builder para diseñar asistentes con instrucciones específicas. La Universidad de California en Irvine puso en funcionamiento en enero de 2024 ZotGPT, un asistente conversacional propio que articula múltiples modelos de lenguaje grande (LLM) y ha alcanzado 15 000 usuarios únicos (EdTech Magazine, 2025).

En el Reino Unido, el Joint Information Systems Committee (JISC) llevó adelante en 2024 un piloto en el que participaron doce instituciones —ocho institutos de formación profesional y cuatro universidades— para evaluar LearnWise, un asistente basado en modelos de lenguaje grande (LLM) entrenado con documentos institucionales, manuales estudiantiles y políticas de admisión, cuyas pruebas pusieron de manifiesto la capacidad de atender consultas administrativas y académicas las 24 horas del día (Jisc, 2025). Estas experiencias demuestran que las universidades de países con mayor desarrollo tecnológico ya integraron chatbots con inteligencia artificial en sus procesos de gestión de información académica, con resultados favorables en disponibilidad, acceso y eficiencia operativa.

En América Latina, la adopción de la inteligencia artificial dentro de la educación superior atraviesa una fase de consolidación, con grados de avance heterogéneos según país. Un estudio comparativo de EdLatam (2025) examina las políticas de México, Colombia y Chile como referentes regionales. En México, más de 190 universidades ofrecen programas vinculados a IA y el Tecnológico de Monterrey diseñó TECgpt, el primer modelo de IA generativa propio de una universidad de la región, que utiliza Microsoft Azure y articula cuatro componentes (acceso a ChatGPT, procesamiento de imágenes, asistente interno TEC Chat y creación de modelos específicos) para apoyar

procedimientos institucionales y consultas académicas (Treviño, 2023). Colombia aprobó en febrero de 2025 su Política Nacional de Inteligencia Artificial, con seis objetivos estratégicos; universidades como los Andes, Icesi y la UDI ya pusieron en marcha comités asesores y planes de integración (EdLatam, 2025).

Chile lidera el Índice Latinoamericano de Inteligencia Artificial 2024 con 73,07 puntos, gracias a una política nacional proyectada hasta 2031; la Pontificia Universidad Católica de Chile presentó ConectIA en alianza con Microsoft Azure, un programa institucional integral que busca la integración completa de la IA a nivel universitario hacia 2026 e incluye chatbots de IA generativa personalizados para docentes, personal administrativo y estudiantes (Gaitán Barrera y Azeez, 2024). Brasil y México reúnen el 68 % de los investigadores activos en IA de la región, y cinco países (Brasil, México, Colombia, Chile y Argentina) concentran el 90 % de los investigadores que publican consistentemente en esta área (Centro Nacional de Inteligencia Artificial [CENIA] y Comisión Económica para América Latina y el Caribe [CEPAL], 2024).

En el ámbito nacional, el Estado peruano emitió la Ley N.º 31814, que promueve el uso de la inteligencia artificial para el desarrollo económico y social del país (Congreso de la República, 2023), y la Política Nacional de Transformación Digital al 2030, que conduce el desarrollo de servicios digitales, la gobernanza de datos y el fortalecimiento de capacidades (Presidencia del Consejo de Ministros [PCM], 2023). En el ámbito universitario, un reporte periodístico sobre un estudio de la firma JusDocs (Infobae, 2025) indica que el 95 % de los estudiantes universitarios peruanos ya recurre a herramientas de IA para sus actividades académicas, principalmente ChatGPT y Gemini. Universidades como la Universidad de Ingeniería y Tecnología (UTEC), la Universidad Peruana Cayetano Heredia (UPCH) y la Universidad del Pacífico establecieron protocolos internos para regular el uso de chatbots en el aula (Torres, 2023).

En el repositorio ALICIA del Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica (CONCYTEC) se encuentran investigaciones sobre chatbots en educación superior, entre las que sobresale el estudio de la Pontificia Universidad Católica del Perú (PUCP) que analizó el aporte de un chatbot a la gestión del tiempo y al desarrollo de un curso de pregrado (Arredondo Castillo, 2020). No obstante, el despliegue de chatbots con IA para la gestión de información académica institucional sigue siendo incipiente: la mayoría de las experiencias se reducen al uso de herramientas externas y no a soluciones diseñadas a medida de cada universidad. El Instituto Nacional de Estadística e Informática (INEI) reporta que, en el cuarto trimestre de 2024, el 58,4 % de los hogares peruanos disponía de acceso a Internet y que la conectividad sigue expandiéndose en ámbitos rurales, lo que eleva la expectativa de respuestas inmediatas y multicanal para trámites y consultas académicas (INEI, 2025).

En el ámbito local, la Universidad Tecnológica de los Andes, cuya sede principal está en Abancay, región Apurímac, y con filiales en Andahuaylas y Cusco, cuenta con licencia institucional emitida por la Superintendencia Nacional de Educación Superior Universitaria (SUNEDU) y atiende a 9 315 estudiantes de pregrado matriculados en el semestre 2025-II —3 983 en la sede Abancay, 2 130 en la filial Andahuaylas y 3 202 en la filial Cusco—, distribuidos en múltiples facultades y programas (Universidad Tecnológica de los Andes [UTEA], 2025). Esta magnitud genera una presión sostenida sobre los canales de atención: los estudiantes demandan respuestas oportunas sobre matrículas, cronogramas académicos, requisitos de grado, convalidaciones, constancias, trámites administrativos y horarios de atención de las oficinas. No obstante, los canales tradicionales de atención (ventanilla presencial, correo electrónico y redes sociales) presentan limitaciones relevantes: la atención presencial se limita al horario laboral, las respuestas por correo demoran días y la información expuesta en el sitio web no siempre permanece actualizada ni resulta sencilla de localizar. Esta saturación se agudiza en periodos pico, cuando un volumen elevado de consultas repetitivas sobre

cronogramas, requisitos y procedimientos administrativos converge en horarios reducidos de atención, desbordando la capacidad operativa del personal y prolongando los tiempos de respuesta más allá de lo aceptable para la comunidad estudiantil.

Esta problemática se intensifica en el caso de los estudiantes de las filiales de Andahuaylas y Cusco, quienes no disponen de acceso directo a las oficinas de la sede principal y dependen por completo de canales remotos para resolver sus consultas administrativas y académicas (Superintendencia Nacional de Educación Superior Universitaria [SUNEDU], 2020). Esta dependencia eleva el riesgo de incumplir plazos académicos y prolongar los tiempos de trámite, lo que repercute en la trayectoria de estudio y en la equidad del servicio educativo. A esto se añade que la situación geográfica de la Universidad Tecnológica de los Andes en Apurímac, una de las regiones peruanas con menor Índice de Desarrollo Humano (IDH): bajo la metodología del PNUD para el Perú, Apurímac registra un IDH de 0,4167 y se clasifica en el nivel bajo de desarrollo humano, con varias de sus provincias en los niveles bajo y muy bajo, y solo la provincia de Abancay en el nivel medio (Programa de las Naciones Unidas para el Desarrollo [PNUD], 2019), lo que implica que cualquier brecha en el acceso a la información académica recae de forma desproporcionada sobre estudiantes de zonas rurales y comunidades con conectividad limitada, donde la cobertura de Internet aún presenta desigualdades pronunciadas respecto a las zonas urbanas (Instituto Nacional de Estadística e Informática [INEI], 2025), particularmente durante periodos críticos como matrículas, cierres de ciclo y trámites de grado, en los que la demanda informativa se concentra en plazos cortos.

Frente a esta problemática, el estudio propone la implementación de un chatbot con inteligencia artificial soportado en modelos de lenguaje grande (LLM) y la técnica de Generación Aumentada por Recuperación (RAG, Retrieval-Augmented Generation), entrenado con documentación institucional de la Universidad Tecnológica de los Andes (reglamentos, manuales, cronogramas y preguntas frecuentes), capaz de brindar

atención permanente las 24 horas del día y los 7 días de la semana, a través de una interfaz web accesible desde cualquier equipo con conexión a Internet. De este modo, la investigación atiende un doble imperativo: la alineación estratégica con la Política Nacional de Transformación Digital al 2030 y la Ley N.º 31814, y la mejora operativa de la atención informativa ante una comunidad estudiantil numerosa, geográficamente distribuida y crecientemente conectada.

Cabe precisar que, si bien la problemática descrita alcanza a toda la comunidad estudiantil de la Universidad Tecnológica de los Andes (incluidas las filiales de Andahuaylas y Cusco, cuya situación se expone como contexto institucional), el alcance empírico del presente estudio se delimita a los estudiantes de pregrado de la sede Abancay. Esta delimitación obedece a razones metodológicas propias del diseño cuasi-experimental adoptado: la aplicación presencial del pretest y del posttest y el control de la intervención exigieron trabajar con grupos intactos accesibles en la sede principal (Hernández-Sampieri y Mendoza, 2018; Shadish et al., 2002). En consecuencia, las filiales quedan fuera del alcance muestral y los resultados obtenidos no son directamente generalizables a ellas, sin perjuicio de que la solución implementada, por operar en canales web y de mensajería, pueda extenderse a dichas sedes en futuras evaluaciones.

2.1.1. Problema general

¿Cuál es el efecto de la implementación de un chatbot con inteligencia artificial en la gestión de información académica en la Universidad Tecnológica de los Andes, 2025?

2.1.2. Problemas específicos

PE1: ¿Cuál es el efecto de la implementación de un chatbot con inteligencia artificial en la disponibilidad de la información académica en la Universidad Tecnológica de los Andes, 2025?

PE2: ¿Cuál es el efecto de la implementación de un chatbot con inteligencia artificial en el acceso a la información académica en la Universidad Tecnológica de los Andes, 2025?

PE3: ¿Cuál es el efecto de la implementación de un chatbot con inteligencia artificial en la eficiencia operativa de la gestión de información académica en la Universidad Tecnológica de los Andes, 2025?

2.2. Objetivos

2.2.1. Objetivo general

Determinar el efecto de la implementación de un chatbot con inteligencia artificial sobre la gestión de información académica en la Universidad Tecnológica de los Andes, 2025.

2.2.2. Objetivos específicos

OE1: Determinar el efecto de la implementación de un chatbot con inteligencia artificial en la disponibilidad de la información académica en la Universidad Tecnológica de los Andes, 2025.

OE2: Determinar el efecto de la implementación de un chatbot con inteligencia artificial en el acceso a la información académica en la Universidad Tecnológica de los Andes, 2025.

OE3: Determinar el efecto de la implementación de un chatbot con inteligencia artificial en la eficiencia operativa de la gestión de información académica en la Universidad Tecnológica de los Andes, 2025.

2.3. Justificación e importancia

2.3.1. Justificación teórica

El estudio halla justificación teórica al ampliar el conocimiento acerca del empleo de chatbots con IA aplicados a la gestión de información académica

universitaria. Trabajos anteriores evidencian que un chatbot bien diseñado optimiza la atención al estudiante, alivia la carga administrativa e incrementa la satisfacción del usuario (Labadze et al., 2023). El estudio se apoya en el método de Generación Aumentada por Recuperación (RAG) propuesto por Lewis et al. (2020), que permite al chatbot consultar la documentación oficial antes de generar la respuesta, mitigando las llamadas «alucinaciones» de los modelos de lenguaje, aspecto crítico al manejar datos académicos como cronogramas y reglamentos. Asimismo, se incorpora la norma ISO/IEC 25010 (2023) para evaluar los atributos de calidad del software (funcionalidad, usabilidad, fiabilidad y eficiencia) bajo un estándar internacional. De esta forma, el estudio aporta evidencia al cuerpo de conocimiento sobre IA generativa aplicada a servicios universitarios y cubre un vacío de la literatura peruana, donde todavía son escasos los estudios acerca de chatbots con arquitectura RAG en universidades del interior del país. Adicionalmente, el marco teórico articulado en torno a la recuperación contextual, los modelos de lenguaje y los estándares de calidad del software ofrece un sustento conceptual que puede orientar futuras investigaciones sobre el aprovechamiento institucional de la IA generativa en contextos académicos similares.

2.3.2. Justificación práctica

Desde la perspectiva práctica, el estudio se justifica al brindar a la Universidad Tecnológica de los Andes una herramienta funcional dirigida a optimizar la atención al estudiante (Labadze et al., 2023). El chatbot permite obtener respuestas inmediatas sobre matrículas, cronogramas, requisitos de grado, convalidaciones y demás trámites, las 24 horas los 7 días de la semana, sin depender del horario presencial. Esto beneficia especialmente a los estudiantes de las filiales de Andahuaylas y Cusco y a quienes residen en zonas rurales con limitaciones de movilidad, si bien la evaluación empírica del presente estudio se circunscribe a la sede Abancay. En segundo lugar, automatiza consultas repetitivas, disminuyendo la

carga del personal administrativo y descongestionando las ventanillas durante matrículas y cierres de ciclo. En tercer lugar, los resultados aportan evidencia empírica medible (tiempo de respuesta, disponibilidad y satisfacción del usuario) para decisiones institucionales sobre la incorporación de IA en otros servicios. Además, la solución es replicable: otras universidades del país podrán adoptar la misma arquitectura adaptándola a su documentación institucional, multiplicando el impacto del estudio. La integración del chatbot con los canales actuales de comunicación universitaria evita duplicidades operativas y facilita su adopción progresiva, sin exigir cambios estructurales en los procedimientos administrativos vigentes.

2.3.3. Justificación metodológica

Desde la perspectiva metodológica, el estudio se apoya en un diseño riguroso y en instrumentos validados que permiten medir con objetividad el efecto del chatbot sobre la gestión de información académica. Se escogió un diseño cuasi-experimental de grupo control no equivalente, con pretest y posttest, adecuado cuando los estudiantes se hallan organizados en grupos ya conformados por la institución, condición que impide la asignación aleatoria individual de los participantes (Shadish et al., 2002; Hernández-Sampieri y Mendoza, 2018). La recolección de datos se realizó con un cuestionario tipo Likert de 42 ítems (Likert, 1932), cuya validez se comprobó por juicio de expertos a través del coeficiente V de Aiken (Aiken, 1985), con valor de 1,00, y cuya confiabilidad en prueba piloto ($n = 30$) alcanzó $\alpha = 0,920$ en la escala de percepción del sistema y $\alpha = 0,907$ en la escala de la variable dependiente, calculadas con el alfa de Cronbach (Cronbach, 1951). El procesamiento se realizó en IBM SPSS Statistics v27 con pruebas no paramétricas, Wilcoxon y U de Mann-Whitney, elegidas tras constatar la falta de normalidad de los datos (Razali y Wah, 2011), sobre la muestra de 380 estudiantes. Dicha secuencia garantiza que las conclusiones se sustenten en pruebas estadísticamente válidas. Adicionalmente, los instrumentos elaborados podrán constituir referencia para próximas investigaciones

sobre chatbots con IA en contextos universitarios. La trazabilidad documental del procedimiento (registro de validación por jueces, prueba piloto y libro de códigos) facilita la auditoría externa de los hallazgos y permite que otros equipos repliquen, ajusten o extiendan el diseño en escenarios institucionales distintos.

2.3.4. Justificación social

Desde una mirada social, el presente trabajo encuentra justificación en que la puesta en marcha del chatbot universaliza el alcance a la información institucional sobre matrículas, cronogramas, requisitos, convalidaciones y demás trámites académicos. Esta mejora beneficia especialmente a estudiantes de zonas rurales de la región Apurímac, a quienes cursan en las filiales de Andahuaylas y Cusco, y a aquellos que, por razones laborales o de movilidad, no pueden desplazarse a la sede principal. La disponibilidad continua del chatbot reduce la brecha de acceso a información oportuna y precisa, en sintonía con la Política Nacional de Transformación Digital al 2030 (PCM, 2023) y la Ley N.° 31814, que promueve el uso de la inteligencia artificial para el desarrollo del país (Congreso de la República, 2023). Al automatizar consultas recurrentes, el chatbot asegura una atención equitativa: todos los estudiantes, sin importar su procedencia geográfica, condición socioeconómica u disponibilidad horaria, reciben la misma calidad de respuesta institucional. Asimismo, contribuye a la inclusión digital de poblaciones dispersas con conectividad heterogénea, fortaleciendo la relación entre la universidad y su comunidad estudiantil.

2.3.5. Importancia

La relevancia del estudio radica en que se erige como una de las primeras evaluaciones empíricas dentro de la región Apurímac sobre el efecto que produce un chatbot con inteligencia artificial sobre la gestión de información académica universitaria, con arquitectura de Generación Aumentada por Recuperación (RAG). Realizada en la Universidad Tecnológica de los Andes, sede Abancay, con 380

estudiantes de pregrado, aporta evidencia contextualizada en una universidad de provincia, escenario poco representado en la literatura nacional sobre inteligencia artificial generativa aplicada a procesos académicos. En el plano institucional, orienta decisiones sobre la incorporación de asistentes conversacionales, mejora la disponibilidad de respuestas fuera del horario presencial y aligera la carga operativa del personal administrativo. En el plano social, democratiza el acceso a la información académica dirigida a estudiantes de zonas rurales y urbano-marginales que enfrentan barreras de conectividad y tiempo. En el plano académico, aporta un marco metodológico replicable que articula el diseño, la implementación y la validación estadística de un chatbot sustentado en RAG. Finalmente, se ponen a disposición como producto académico abierto el código fuente, los flujos de automatización y la documentación técnica, facilitando su adaptación por otras instituciones de educación superior y futuras líneas de trabajo.

2.4. Hipótesis

2.4.1. Hipótesis general

H_0 : La implementación del chatbot con inteligencia artificial no mejora significativamente la gestión de información académica en la Universidad Tecnológica de los Andes, 2025.

H_1 : La implementación del chatbot con inteligencia artificial mejora significativamente la gestión de información académica en la Universidad Tecnológica de los Andes, 2025.

2.4.2. Hipótesis específicas

H_{01} : La implementación del chatbot con inteligencia artificial no incrementa significativamente la disponibilidad de la información académica en la Universidad Tecnológica de los Andes, 2025.

H₁₁: La implementación del chatbot con inteligencia artificial incrementa significativamente la disponibilidad de la información académica en la Universidad Tecnológica de los Andes, 2025.

H₀₂: La implementación del chatbot con inteligencia artificial no facilita significativamente el acceso a la información académica en la Universidad Tecnológica de los Andes, 2025.

H₁₂: La implementación del chatbot con inteligencia artificial facilita significativamente el acceso a la información académica en la Universidad Tecnológica de los Andes, 2025.

H₀₃: La implementación del chatbot con inteligencia artificial no mejora significativamente la eficiencia operativa en la gestión de información académica en la Universidad Tecnológica de los Andes, 2025.

H₁₃: La implementación del chatbot con inteligencia artificial mejora significativamente la eficiencia operativa en la gestión de información académica en la Universidad Tecnológica de los Andes, 2025.

2.5. Variables

2.5.1. Variable independiente

Chatbot con inteligencia artificial.

2.5.2. Variable dependiente

Gestión de información académica.

2.5.3. Operacionalización de variables

Tabla 1

Operacionalización de la variable independiente: Chatbot con inteligencia artificial

VARIABLES	DEFINICIÓN CONCEPTUAL	DEFINICIÓN OPERACIONAL	DIMENSIONES	INDICADORES	ÍTEMS	ESCALA
Variable independiente: Chatbot con inteligencia artificial (IA)	Sistema de conversación al basado en inteligencia artificial que interactúa con el usuario en lenguaje natural para brindar respuestas automatizadas a consultas académicas (Adamopoulou y Moussiades, 2020).	La variable independiente corresponde a la condición de tratamiento del diseño cuasi-experimental: presencia o ausencia de la intervención con el chatbot, medida en escala nominal dicotómica (0 = sin chatbot, condición pretest; 1 = con chatbot, condición posttest). El sistema se implementa con modelos de lenguaje grande (LLM) y la técnica de Generación Aumentada por Recuperación (RAG; Lewis et al., 2020). De manera complementaria, se registra un índice de percepción de la calidad del sistema mediante un cuestionario tipo Likert de cinco niveles con 21 ítems (1–21), agrupados en tres dimensiones alineadas con la norma ISO/IEC 25010 (2023) y aplicado solo al	D1: Capacidad de procesamiento inteligente	Comprensión, coherencia, pertinencia, robustez semántica, utilidad de respuesta, interpretación de jerga y mantención de contexto	1–7	VI (tratamiento): nominal dicotómica (0/1). Índice de percepción del sistema (medida secundaria): ordinal, Likert 1–5.
			D2: Usabilidad del sistema	Facilidad de uso, curva de aprendizaje, diseño amigable, comodidad, claridad de mensajes y satisfacción de uso	8–14	VI (tratamiento): nominal dicotómica (0/1). Índice de percepción del sistema (medida secundaria): ordinal, Likert 1–5.
			D3: Disponibilidad del servicio	Acceso continuo sin interrupciones, tiempo de respuesta razonable,	15–21	VI (tratamiento): nominal dicotómica (0/1).

grupo experimental. Este índice constituye una variable de resultado secundaria (percepción del sistema), no la manipulación experimental (Shadish et al., 2002).	multidispositivo, operatividad en alta demanda y constancia	Índice de percepción del sistema (medida secundaria): ordinal, Likert 1–5.
---	---	--

Nota. Indicadores adaptados de la norma ISO/IEC 25010 (2023). La VI del diseño es la condición de tratamiento, de escala nominal dicotómica (0 = sin chatbot, 1 = con chatbot); los 21 ítems Likert conforman un índice de percepción de la calidad del sistema, tratado como variable de resultado secundaria (Shadish et al., 2002).

Tabla 2

Operacionalización de la variable dependiente: Gestión de información académica

VARIABLES	DEFINICIÓN CONCEPTUAL	DEFINICIÓN OPERACIONAL	DIMENSIONES	INDICADORES	ÍTEMS	ESCALA
Variable dependiente:	Conjunto de procesos institucionales	Se mide a través de un cuestionario de tipo Likert de cinco niveles con 21 ítems (22–42)	D1: Disponibilidad de la información	Disponibilidad bajo demanda, accesibilidad multihorario, atención sin interrupciones, continuidad institucional y disponibilidad en periodos de alta demanda	22–28	Ordinal (Likert 1–5)
Gestión de información académica	mediante los cuales los estudiantes acceden, consultan y utilizan información académica oficial para la toma de	agrupados en tres dimensiones,	D2: Acceso a la información	Encontrabilidad, claridad, precisión, actualización, accesibilidad	29–35	Ordinal (Likert 1–5)

decisiones y la aplicado en pretest resolución de trámites y postest a ambos (Laudon y Laudon, grupos. 2020).	D3: Eficiencia operativa	multidispositivo, organización y ausencia de confusión Rapidez de respuesta, resolución en primer contacto, tiempo de espera, facilitación de trámites, ahorro de tiempo, no repetición y agilidad	36–42	Ordinal (Likert 1–5)
--	-----------------------------	--	-------	-------------------------

Nota. La variable se mide en escala ordinal tipo Likert de cinco puntos (ítems 22–42), agrupada en tres dimensiones: D1 Disponibilidad, D2 Acceso y D3 Eficiencia.

III. Marco Teórico

3.1. Antecedentes

3.1.1. Internacionales

Balderas et al. (2023), en la Universidad de Cádiz (campus Puerto Real), España, tuvieron como objetivo evaluar la percepción de los estudiantes sobre el uso de un chatbot como medio de comunicación con los servicios estudiantiles en situaciones de emergencia. Con enfoque mixto bajo el marco de diseño y creación, implementaron un chatbot basado en Dialogflow con interfaces web y Telegram en el marco del proyecto RIPEC, evaluado con una muestra de 165 estudiantes universitarios y personal de servicios estudiantiles de seis universidades europeas, registrándose 405 interacciones y aplicando el cuestionario Bot Usability Scale (BUS) de 15 ítems en escala Likert. Los resultados reportaron una evaluación favorable en todas las dimensiones, destacando el tiempo de respuesta ($M = 4,21$) y la capacidad del chatbot para referenciar servicios ($M = 4,06$), con un alfa de Cronbach de 0,94 y una valoración de 4,17 sobre 5 para la futura implementación. Se concluyó que los chatbots son una herramienta viable para facilitar la comunicación entre la universidad y los estudiantes en situaciones de emergencia.

Han y Lee (2022), en The University of Texas at Austin, Estados Unidos, tuvieron como objetivo contrastar la experiencia de aprendizaje entre un chatbot de Preguntas Frecuentes (FAQ, Frequently Asked Questions) y la página estática de FAQ dentro de Cursos Masivos Abiertos en Línea (MOOC, Massive Open Online Course). Con enfoque mixto y diseño experimental entre sujetos, entrenaron un chatbot basado en procesamiento de lenguaje natural en Dialogflow, desplegado en dos MOOCs de periodismo, trabajando con una muestra de 120 estudiantes (46 en el grupo chatbot y 74 en el grupo página web) procedentes de 45 países. Los resultados reportaron que el grupo del chatbot mostró menor intención de uso (Mdn

= 3,00 frente a 4,00; $p = 0,006$) y mayor nivel de barreras percibidas ($Mdn = 2,50$ frente a 1,67; $p < 0,001$), siendo significativos los factores región ($p = 0,029$) y lengua materna ($p = 0,039$). Se concluyó que es inviable garantizar una experiencia de aprendizaje inclusiva con un chatbot FAQ sin antes considerar los factores contextuales y lingüísticos del estudiante.

Essel et al. (2022), en la Kwame Nkrumah University of Science and Technology (KNUST), Kumasi, Ghana, tuvieron como objetivo evaluar el efecto del chatbot KNUSTbot, orientado al apoyo docente, en el desempeño académico del alumnado universitario. Con enfoque mixto y diseño pretest-posttest combinado con grupos focales, implementaron el chatbot con WhatsApp y FlowXO bajo la técnica zero-coding, trabajando con una muestra de 68 estudiantes de pregrado asignados aleatoriamente a la cohorte experimental ($n = 34$) y a la de control ($n = 34$). Los resultados reportaron que el grupo experimental mejoró de $M = 40,6$ ($DE = 4,95$) en el pretest a $M = 81,1$ ($DE = 3,19$) en el posttest, mientras el grupo control pasó de $M = 43,4$ ($DE = 4,09$) a $M = 65,2$ ($DE = 3,73$), con interacción significativa tiempo por cohorte $F(1, 66) = 87,5$; $p < 0,05$; $\eta^2p = 0,57$. Se concluyó que los estudiantes que interactuaron con el chatbot obtuvieron mejor rendimiento académico que los que interactuaron con el instructor.

Rodriguez y Mune (2022), en la San José State University Library, Estados Unidos, tuvieron como objetivo documentar el desarrollo y despliegue del chatbot Kingbot para la atención de consultas de referencia bibliotecaria fuera del horario laboral. Mediante un estudio de caso con enfoque mixto, desarrollaron el chatbot con Google Dialogflow y Kommunicate como interfaz, integrado al widget de SpringShare LibChat, analizando 1 682 interacciones registradas entre octubre de 2020 y junio de 2022. Los resultados reportaron un incremento del 214 % entre el semestre Primavera 2021 (219 interacciones) y Primavera 2022 (687 interacciones), siendo las consultas más frecuentes sobre horarios (451), búsqueda de artículos arbitrados

(290) y préstamo de dispositivos (120). Se concluyó que las bibliotecas universitarias pueden implementar chatbots funcionales con herramientas accesibles y mínimos conocimientos de programación para extender el servicio de referencia virtual fuera del horario presencial.

Neupane et al. (2024), en la Mississippi State University, Estados Unidos, tuvieron como objetivo diseñar y evaluar el chatbot BARKPLUG V.2, sustentado en la arquitectura de Generación Aumentada por Recuperación (RAG, Retrieval-Augmented Generation), capaz de responder a consultas sobre recursos institucionales utilizando la documentación oficial de la universidad como base de conocimiento. Con enfoque cuantitativo aplicado y diseño experimental, implementaron el sistema utilizando GPT-3.5 Turbo, LangChain y ChromaDB, y lo evaluaron con una muestra de 50 estudiantes de pregrado y posgrado mediante el conjunto de métricas RAGAS y la escala System Usability Scale (SUS). Los resultados reportaron un puntaje RAGAS promedio de 0,96, evidenciando alta precisión y relevancia, y un puntaje SUS promedios de 67,75. Se concluyó que un chatbot basado en RAG mejora la pertinencia y la precisión de las respuestas a consultas estudiantiles.

3.1.2. Nacionales

Veliz y Velásquez (2024), en la Universidad Nacional Pedro Ruiz Gallo (UNPRG), Lambayeque, Perú, tuvieron como objetivo implementar un asistente virtual mediante chatbot para la atención de consultas de la gestión académica de la UNPRG. Con enfoque cuantitativo-descriptivo de tipo aplicado, desarrollaron el chatbot utilizando el lenguaje Python con la plataforma de mensajería WhatsApp bajo la metodología ágil Scrum, trabajando con una muestra de 300 estudiantes de la Facultad de Ciencias Físicas y Matemáticas matriculados en el semestre 2023-II, a quienes aplicaron cuestionarios elaborados en Google Forms para medir la satisfacción antes y después de la implementación. Los resultados reportaron un

incremento en la satisfacción de los estudiantes respecto al acceso a información académica del 48,7 % en el pretest al 89,45 % en el postest. Se concluyó que la implementación del chatbot mejoró significativamente la atención de consultas de la gestión académica de la UNPRG.

Espinosa-Luna et al. (2023), en la Universidad Nacional de Trujillo (UNT), Perú, tuvieron como objetivo evaluar la usabilidad y utilidad de un chatbot basado en GPT-3.5-Turbo como asistente académico para atender consultas frecuentes en una universidad pública. Con enfoque cuantitativo aplicado y diseño preexperimental, implementaron el chatbot a través de WhatsApp como interfaz, trabajando con una muestra de 41 estudiantes de primer ciclo de Ingeniería de Sistemas, a quienes aplicaron el Cuestionario de Usabilidad de Chatbot (CUQ, Chatbot Usability Questionnaire) de 16 ítems en escala Likert. Los resultados reportaron un puntaje CUQ promedio de 73,4 (DE = 10,1), superior al valor de referencia de 68, y un 93 % de estudiantes valoró positivamente la utilidad de las respuestas. Se concluyó que los modelos de lenguaje grande resultan viables para la atención de consultas frecuentes en universidades peruanas.

Quispe Huacani (2023), en la Universidad Nacional Jorge Basadre Grohmann (UNJBG), Tacna, Perú, tuvo como objetivo determinar el impacto de un chatbot en el proceso de atención de matrícula de los estudiantes de la Escuela Profesional de Ingeniería en Informática y Sistemas. Con enfoque cuantitativo, tipo aplicada, nivel explicativo y diseño experimental, trabajó con una muestra de 50 estudiantes que pretendían matricularse en el periodo académico 2022-I, distribuidos en dos grupos (proceso tradicional y proceso chatbot), evaluando las dimensiones de funcionalidad, confiabilidad y usabilidad mediante una encuesta con cuestionario en escala de intervalo, y aplicando la prueba T de Student para muestras independientes previa verificación de normalidad. Los resultados reportaron que el grupo del chatbot alcanzó niveles altos en funcionalidad (100 % frente a 32 % del proceso tradicional),

confiabilidad (92 % frente a 60 %) y usabilidad (96 % frente a 64 %). Se concluyó que el chatbot tuvo un impacto significativo en el proceso de atención de matrícula de los estudiantes de la UNJBG.

Quispe Cruz (2024), en la Universidad Peruana de Ciencias e Informática (UPCI), Lima, Perú, tuvo como objetivo determinar el efecto de la implementación de un chatbot en la calidad del servicio de atención al egresado. Con enfoque cuantitativo aplicado y diseño preexperimental pretest-posttest con un solo grupo, trabajó con una muestra de 54 egresados seleccionados mediante muestreo no probabilístico por conveniencia, a quienes aplicó un cuestionario de 26 ítems basado en el modelo SERVQUAL en escala Likert de cinco puntos con tres dimensiones (fiabilidad, capacidad de respuesta y seguridad), cuya confiabilidad fue $\alpha = 0,850$ en el pretest y $\alpha = 0,834$ en el posttest, analizando los datos con la prueba de Wilcoxon. Los resultados reportaron incrementos significativos en la calidad del servicio ($M = 2,04$ a $M = 3,80$), fiabilidad ($M = 2,06$ a $M = 3,83$), capacidad de respuesta ($M = 2,04$ a $M = 3,81$) y seguridad ($M = 2,17$ a $M = 3,70$), con $p = 0,000$ en todos los casos. Se concluyó que la implementación del chatbot mejoró significativamente la percepción de la calidad del servicio al egresado de la UPCI.

Zapata (2023), en la Universidad Nacional de Piura (UNP), Perú, tuvo como objetivo desarrollar un chatbot que permita el acceso a información sobre trámites administrativos de los estudiantes de Ingeniería Informática. Con enfoque cuantitativo y diseño pre-experimental de estudio de caso con una sola medición, desarrolló el chatbot mediante la metodología ágil Scrum sobre una página de Facebook, trabajando con una muestra de 45 estudiantes matriculados en el año académico 2022, a quienes aplicó dos cuestionarios y un registro de observación para medir las dimensiones de usabilidad, facilidad y precisión. Los resultados reportaron una usabilidad aceptable del 77,39 %, una facilidad de acceso a la información del 93,30 % y una precisión de la información del 93,0 %. Se concluyó que el chatbot permitió

a los estudiantes obtener respuestas a sus consultas sobre trámites administrativos de manera directa y más efectiva.

3.1.3. Local

Cari-Incahuanaco et al. (2025), en la Universidad Nacional Micaela Bastidas de Apurímac (UNAMBA), Perú, tuvieron como objetivo determinar la contribución del tutor virtual con ChatGPT en el proceso de autoaprendizaje de los estudiantes con matrícula condicionada de la Escuela Profesional de Ingeniería Informática y Sistemas. Con enfoque cuantitativo, tipo aplicada, nivel correlacional y diseño pre-experimental con pretest y posttest, trabajaron con una muestra de 36 estudiantes con matrícula condicionada del semestre académico 2023-II seleccionados mediante muestreo no probabilístico intencional, a quienes capacitaron en el uso de ChatGPT y aplicaron pretest, posttest, encuesta estructurada y registros académicos, analizando los datos con SPSS v28 mediante estadísticos descriptivos y el coeficiente Rho de Spearman. Los resultados reportaron una mejora del rendimiento académico del pretest ($M = 6,73$) al posttest ($M = 11,31$), con 75 % de estudiantes aprobados, y una correlación positiva moderada entre el uso de ChatGPT y el autoaprendizaje ($Rho = 0,533$). Se concluyó que el uso adecuado del tutor virtual con ChatGPT mejora considerablemente el autoaprendizaje y el rendimiento académico de los estudiantes con matrícula condicionada de la UNAMBA.

3.2. Bases teóricas

3.2.1. Chatbot con inteligencia artificial

3.2.1.1. Fundamentación teórica

Un chatbot se concibe como un agente conversacional con capacidad de interpretar entradas en lenguaje natural, sostener el contexto del diálogo y producir respuestas coherentes con apoyo de fuentes de información estructuradas o documentales (Følstad et al., 2021). Su evolución técnica

avanzó desde sistemas basados en reglas estáticas hacia modelos de lenguaje de gran escala (LLM, Large Language Model) dotados de capacidades avanzadas para la comprensión y la generación del lenguaje natural (Adamopoulou y Moussiades, 2020; J. Gao et al., 2019).

En educación superior, los chatbots permiten mejorar la accesibilidad, optimizar tiempos y personalizar la experiencia, aunque su valor depende de una gobernanza clara de contenidos y datos (Labadze et al., 2023; Wu y Yu, 2024; Xu et al., 2025). Para elevar la factualidad en contextos institucionales, es práctica consolidada acoplar LLM con recuperación aumentada: el patrón técnico RAG (Generación Aumentada por Recuperación) minimiza las alucinaciones y facilita la auditoría y la actualización continua del conocimiento (Lewis et al., 2020). Organismos internacionales recomiendan desplegar estas soluciones con una visión centrada en lo humano, resguardando ética, equidad y transparencia (Miao y Holmes, 2023; Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura [UNESCO], 2023).

3.2.1.2. Clasificación de chatbots

Los chatbots admiten categorización según su arquitectura, su nivel de inteligencia y el modo en que generan las respuestas (Adamopoulou y Moussiades, 2020). La literatura especializada identifica cuatro grandes tipos: chatbots fundamentados en reglas, chatbots con inteligencia artificial dotados de comprensión del lenguaje natural, chatbots generativos sustentados en modelos de lenguaje grande (LLM, Large Language Model) y chatbots híbridos con Generación Aumentada por Recuperación (RAG, Retrieval-Augmented Generation) (Khennouche et al., 2024; Lewis et al., 2020). Más allá de la arquitectura, los chatbots pueden agruparse según el dominio de aplicación en orientados a tareas y orientados al diálogo abierto (Caldarini et al., 2022). Los primeros operan en contextos acotados como atención al

cliente o consultas administrativas, mientras que los segundos buscan mantener conversaciones generales y abiertas. La selección del tipo de chatbot responde a tres factores principales: la naturaleza de las consultas, el nivel de control requerido sobre las respuestas y la sensibilidad de la información manejada (Adamopoulou y Moussiades, 2020). En contextos universitarios, donde la información debe ajustarse a documentos normativos vigentes, los chatbots híbridos con RAG resultan especialmente apropiados.

Chatbots basados en reglas.

Constituyen sistemas programados previamente que funcionan siguiendo flujos conversacionales definidos a través de árboles de decisión y reconocimiento de palabras clave; entregan respuestas fijas, sin aprendizaje automático, y resultan adecuados para tareas repetitivas o consultas estructuradas como las preguntas frecuentes (FAQ, Frequently Asked Questions) (Adamopoulou y Moussiades, 2020). Su limitación principal radica en la rigidez ante consultas no contempladas en el flujo y en la dificultad de mantener coherencia cuando el usuario reformula la pregunta (Adamopoulou y Moussiades, 2020). Su implementación es ágil y de bajo costo, lo que los hace populares en servicios de atención inicial; no obstante, su escalabilidad es limitada cuando los flujos de consulta crecen en complejidad (Caldarini et al., 2022).

Chatbots basados en IA con comprensión del lenguaje natural.

Recurren a técnicas de procesamiento del lenguaje natural (PLN) y comprensión del lenguaje natural (NLU, Natural Language Understanding) para identificar la intención del usuario y las entidades presentes en su consulta, lo que posibilita respuestas más flexibles y adaptativas. Aprenden a partir de los datos históricos y mejoran progresivamente con la retroalimentación (Caldarini et al., 2022). Plataformas como Dialogflow y Rasa

se enmarcan en esta categoría. Su desempeño depende fuertemente de la calidad del corpus de entrenamiento y de la curación de intenciones definidas por el equipo de desarrollo. Resultan adecuados para dominios estructurados con vocabulario estable, como soporte técnico o atención al cliente (Adamopoulou y Moussiades, 2020).

Chatbots generativos basados en modelos de lenguaje grande.

Elaboran respuestas de manera dinámica empleando modelos de lenguaje grande entrenados con extensas cantidades de texto, sin depender de respuestas predefinidas. Modelos como GPT y Gemini constituyen el estado del arte en esta categoría y han demostrado capacidad para abordar consultas complejas con respuestas coherentes y contextualizadas (Brown et al., 2020). Sin embargo, presentan limitaciones importantes: pueden producir información inexacta o inventada, fenómeno conocido como alucinación, especialmente cuando el conocimiento requerido no se encontraba en sus datos de entrenamiento (Ji et al., 2023). Su principal ventaja reside en la capacidad de mantener conversaciones fluidas y atender consultas no previstas en su entrenamiento; en contraparte, su uso en entornos institucionales requiere mecanismos de control para evitar respuestas no autorizadas o desactualizadas (Ji et al., 2023).

Chatbots híbridos con Generación Aumentada por Recuperación.

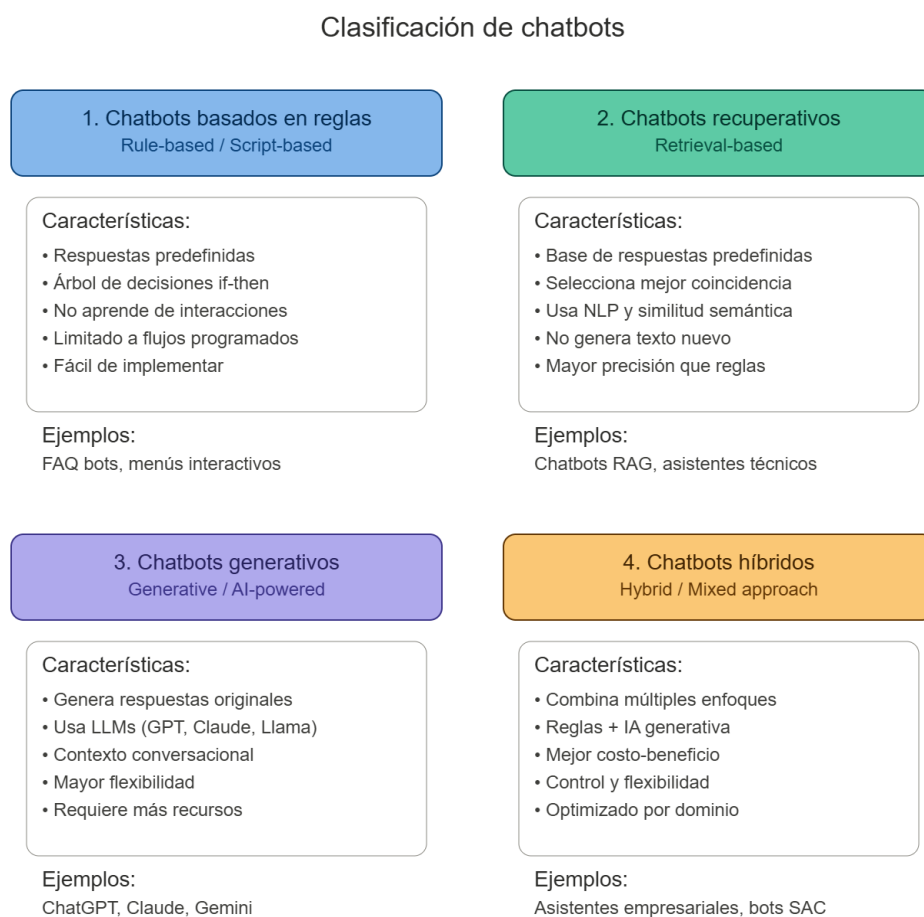
Articulan la capacidad generativa de los modelos de lenguaje grande con un mecanismo de recuperación de información desde una base de conocimiento controlada y verificada. La arquitectura RAG, planteada por Lewis et al. (2020), permite que el chatbot consulte primero los documentos institucionales y luego genere una respuesta sustentada en esa información, lo que reduce significativamente las alucinaciones y asegura que los contenidos compartidos correspondan a la documentación oficial (Ji et al.,

2023). Diversos estudios reportan que la arquitectura RAG disminuye la tasa de alucinaciones y mejora la trazabilidad de las respuestas frente a los modelos puramente generativos (Lewis et al., 2020). Esta característica resulta crítica en contextos universitarios donde toda respuesta debe poder rastrearse a una fuente oficial verificable.

En síntesis, la selección entre los cuatro tipos depende del equilibrio entre flexibilidad conversacional, control sobre la fuente de información y costo de implementación: los chatbots basados en reglas convienen cuando las consultas son pocas y estables; los basados en NLU resultan adecuados para dominios estructurados con vocabulario estable; los puramente generativos se justifican cuando se prioriza la fluidez sobre la precisión documental; y los híbridos con RAG son la opción más apropiada cuando, como en el presente estudio, las respuestas deben sustentarse en documentación institucional vigente y mantener trazabilidad a la fuente. Por esta razón, la presente investigación adopta una arquitectura RAG, alineada con la naturaleza normativa y verificable de la información académica universitaria.

Figura 1

Comparación de los cuatro tipos de chatbots según arquitectura, mecanismo de respuesta y caso de uso representativo



Comparación de características

Característica	Reglas	Recuperativo	Generativo	Híbrido
Complejidad	Baja	Media	Alta	Media-Alta
Costo	Bajo	Medio	Alto	Medio
Flexibilidad	Muy baja	Media	Muy alta	Alta
Precisión	Alta*	Alta	Variable	Alta

* Alta dentro de su dominio limitado. Fuera del dominio: nula.

Nota. Adaptado de Adamopoulou y Moussiades (2020), Khennouche et al. (2024) y Lewis et al. (2020).

3.2.1.3. Definición conceptual de la variable independiente

Para efectos de la presente investigación, el chatbot con inteligencia artificial se concibe como un sistema conversacional sustentado en modelos de lenguaje de gran escala y arquitectura Generación Aumentada por Recuperación (RAG), implementado en los canales digitales de la Universidad Tecnológica de los Andes, que interpreta consultas estudiantiles en lenguaje natural y produce respuestas apoyadas en la base de conocimiento institucional (Adamopoulou y Moussiades, 2020; Lewis et al., 2020). Se operacionaliza de manera dicotómica: presencia del chatbot (grupo experimental) frente a ausencia (grupo control).

3.2.1.4. Dimensiones de la variable independiente

Capacidad de procesamiento inteligente.

Los modelos de lenguaje de gran escala (LLM) se sustentan en la arquitectura Transformer (red neuronal profunda con mecanismo de autoatención), que posibilita paralelizar el aprendizaje sobre grandes corpus textuales y generalizar con pocos ejemplos (Vaswani et al., 2017; Brown et al., 2020). Tras el preentrenamiento, el ajuste fino mediante retroalimentación humana (RLHF, Reinforcement Learning from Human Feedback) alinea el modelo con los estándares institucionales (Ouyang et al., 2022).

Usabilidad del sistema.

ISO/IEC 25010 (2023) conceptualiza la usabilidad como el grado en que usuarios específicos alcanzan objetivos con eficacia, eficiencia y satisfacción. La descompone en reconocibilidad, aprendibilidad, operabilidad y accesibilidad, criterios que orientan tanto el diseño como la evaluación del chatbot en canales web o de mensajería.

Disponibilidad del servicio.

La disponibilidad se representa con la fórmula $A = \text{MTBF} / (\text{MTBF} + \text{MTTR})$, donde MTBF (Mean Time Between Failures) corresponde al tiempo medio entre fallas y MTTR (Mean Time To Repair) al tiempo medio de reparación (International Organization for Standardization y International Electrotechnical Commission [ISO/IEC], 2023). Esta métrica se rige por los indicadores de nivel de servicio (SLI, Service Level Indicators), que permiten supervisar en tiempo real el cumplimiento de los objetivos de nivel de servicio (SLO, Service Level Objectives) (Treyner et al., 2017; Google Site Reliability Engineering [SRE, Ingeniería de Confiabilidad del Sitio], 2023).

3.2.1.5. Tecnologías que sustentan el chatbot

El funcionamiento del chatbot se apoya en un conjunto de tecnologías propias del campo de la ingeniería de sistemas que integran el procesamiento del lenguaje, el aprendizaje automático, la conexión con servicios externos y el despliegue en la nube (Russell y Norvig, 2021). A continuación se describen los componentes más relevantes que intervienen en este tipo de soluciones.

Procesamiento del lenguaje natural (PLN).

El procesamiento del lenguaje natural constituye la rama de la inteligencia artificial dedicada a interpretar y generar lenguaje humano mediante métodos computacionales. Habilita al chatbot para reconocer la intención del usuario, identificar entidades y producir respuestas coherentes. Las tareas típicas comprenden el análisis sintáctico, el reconocimiento de entidades nombradas, el análisis de sentimientos y la traducción automática (Caldarini et al., 2022).

Aprendizaje automático y aprendizaje profundo.

El aprendizaje automático (Machine Learning) posibilita que los sistemas mejoren su desempeño a partir del análisis de grandes volúmenes de datos sin programación explícita. El aprendizaje profundo (Deep Learning), una rama del aprendizaje automático, utiliza redes neuronales con múltiples capas para detectar patrones complejos en texto, imagen y audio. Ambas técnicas son la base sobre la que se construyen los modelos conversacionales modernos (Cuayáhuitl et al., 2019).

Redes neuronales: convolucionales, recurrentes y Transformer.

Las redes neuronales convolucionales (CNN, Convolutional Neural Networks) se especializan en el reconocimiento de patrones espaciales, en tanto que las redes neuronales recurrentes (RNN, Recurrent Neural Networks) procesan secuencias temporales como el lenguaje (Goodfellow et al., 2016). La arquitectura Transformer, presentada por Vaswani et al. (2017), superó las limitaciones de las redes recurrentes a través del mecanismo de autoatención y sustenta el fundamento de los modelos de lenguaje grande actuales como GPT, Gemini y Llama (Brown et al., 2020).

Interfaces de programación y arquitectura cliente-servidor.

Las interfaces de programación de aplicaciones (API, Application Programming Interface) habilitan al chatbot para consumir servicios externos como modelos de lenguaje alojados en la nube, sistemas de mensajería, repositorios documentales y bases de datos. La comunicación se realiza habitualmente mediante el protocolo de transferencia de hipertexto (HTTP, Hypertext Transfer Protocol) y el formato de intercambio de datos JSON (JavaScript Object Notation), bajo una arquitectura cliente-servidor que separa la lógica de negocio de la interfaz del usuario (Fielding, 2000).

Frameworks y plataformas de orquestación.

Los frameworks especializados facilitan la construcción de chatbots al conectar los modelos de lenguaje con bases de datos, motores de búsqueda y canales de mensajería (Topal et al., 2024). LangChain y LlamaIndex orquestan flujos de Generación Aumentada por Recuperación; n8n posibilita automatizar integraciones entre servicios mediante flujos visuales; plataformas como Dialogflow y Rasa proveen entornos completos para el diseño conversacional (Caldarini et al., 2022). La presente investigación adoptó n8n como plataforma de orquestación por su flexibilidad y por habilitar el control completo de la lógica del flujo conversacional.

Computación en la nube.

La computación en la nube proporciona infraestructura, plataforma y software como servicio (IaaS, PaaS y SaaS, respectivamente), lo que posibilita desplegar el chatbot con escalabilidad, alta disponibilidad y costos variables según demanda (Mell y Grance, 2011). Servicios como Amazon Web Services, Microsoft Azure y Google Cloud Platform ofrecen entornos administrados que aligeran la carga operativa de la infraestructura y aceleran el ciclo de despliegue.

3.2.1.6. Arquitectura RAG

La arquitectura de Generación Aumentada por Recuperación (RAG, Retrieval-Augmented Generation), planteada por Lewis et al. (2020), articula la capacidad generativa de los modelos de lenguaje grande con un mecanismo de recuperación de información desde una base de conocimiento institucional, lo que disminuye las alucinaciones y posibilita que el chatbot responda con datos verificados de la documentación oficial (Y. Gao et al., 2023; Ji et al., 2023). Esta arquitectura constituye en la actualidad el paradigma dominante

para integrar modelos de lenguaje en entornos corporativos con información sensible o normativa (Y. Gao et al., 2023).

Componentes principales.

La arquitectura RAG se compone de tres elementos básicos: el indexador, que procesa la documentación institucional y produce representaciones numéricas guardadas en la base de datos vectorial; el recuperador (retriever), que frente a una consulta del usuario localiza los fragmentos de mayor similitud semántica; y el generador, un modelo de lenguaje grande que elabora la respuesta final integrando la consulta original con los fragmentos recuperados a modo de contexto (Y. Gao et al., 2023). La articulación entre estos tres componentes determina la calidad final de las respuestas, ya que una falla en cualquiera de ellos compromete toda la cadena de procesamiento (Lewis et al., 2020).

Pipeline de procesamiento.

El flujo de procesamiento recorre cinco etapas: (1) la segmentación o chunking, que divide los documentos en fragmentos manejables; (2) la generación de embeddings, que transforma cada fragmento en un vector numérico mediante un modelo de codificación; (3) el almacenamiento en la base de datos vectorial; (4) la búsqueda semántica, que ante una consulta localiza los fragmentos más relevantes; y (5) la generación aumentada, donde el modelo de lenguaje elabora la respuesta utilizando los fragmentos recuperados como contexto en el prompt (Lewis et al., 2020). La eficacia del pipeline depende de decisiones de diseño como el tamaño del chunk, la métrica de similitud empleada y la cantidad de fragmentos recuperados por consulta (Y. Gao et al., 2023).

Métricas de evaluación.

La calidad de un sistema de Generación Aumentada por Recuperación se valora mediante el conjunto de métricas RAGAS (Retrieval-Augmented Generation Assessment), que examina cuatro dimensiones: la fidelidad (faithfulness), que mide cuán alineada está la respuesta con los fragmentos recuperados; la pertinencia de la respuesta (answer relevance), que evalúa si responde efectivamente a la consulta; la pertinencia del contexto (context relevance), que mide si los fragmentos recuperados son apropiados; y la suficiencia, que verifica si el contexto recuperado contiene la información necesaria para responder (Es et al., 2024). Estas métricas posibilitan auditar de manera objetiva el desempeño del chatbot a lo largo del tiempo y contrastar el efecto de distintas configuraciones del sistema (Es et al., 2024). La adopción de RAGAS como marco evaluativo facilita la comparación entre sistemas RAG y proporciona evidencia cuantitativa para la mejora continua del chatbot (Es et al., 2024).

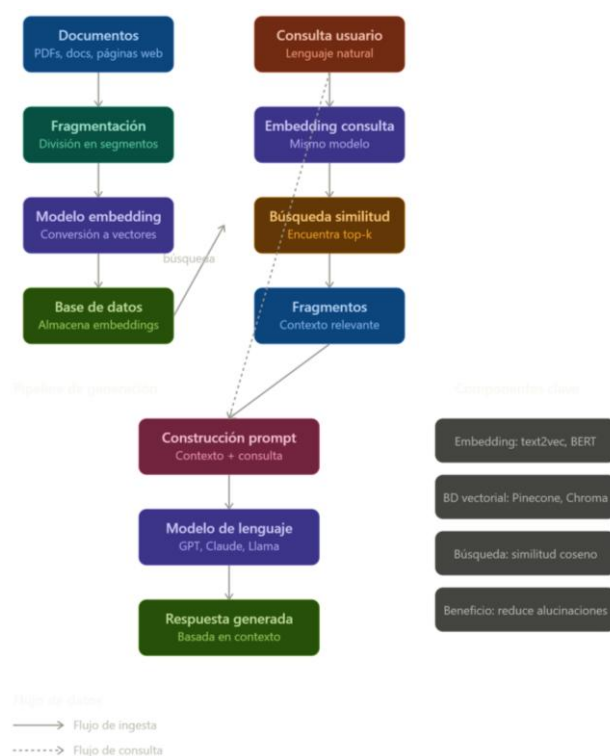
Variantes de RAG. La literatura reciente distingue tres configuraciones principales: el RAG ingenuo (naive RAG), que sigue el pipeline básico de recuperación-generación; el RAG avanzado (advanced RAG), que incorpora optimizaciones previas y posteriores a la recuperación, como la reformulación de la consulta y el reordenamiento de fragmentos; y el RAG modular, que permite intercambiar componentes (recuperadores densos o dispersos, modelos de reordenamiento, memorias conversacionales) según los requerimientos del dominio (Y. Gao et al., 2023). Estas variantes representan una progresión hacia sistemas más robustos y adaptables a contextos institucionales complejos.

Limitaciones. Pese a sus ventajas, la arquitectura RAG presenta restricciones documentadas: la calidad de las respuestas depende de la

curación y actualización constante de la base de conocimiento; los chunks mal dimensionados pueden fragmentar información clave o introducir ruido; la búsqueda semántica puede fallar ante consultas ambiguas o términos ausentes del corpus; y el incremento del contexto recuperado eleva el costo computacional y la latencia (Y. Gao et al., 2023). Estas restricciones exigen un diseño cuidadoso del pipeline y una gobernanza continua de la documentación institucional para sostener el desempeño del sistema en producción.

Figura 2

Pipeline de Generación Aumentada por Recuperación (RAG): flujo de procesamiento desde la documentación institucional hasta la respuesta del chatbot



Nota. El esquema representa las cinco etapas secuenciales del flujo RAG (chunking, embeddings, indexación vectorial, búsqueda semántica y generación aumentada) y los tres componentes que las articulan: indexador, recuperador y modelo de lenguaje grande. Adaptado de Lewis et al. (2020) y Y. Gao et al. (2023).

3.2.1.7. Modelos de lenguaje de gran escala (LLM)

Los modelos de lenguaje de gran escala (LLM, Large Language Model) constituyen redes neuronales profundas entrenadas con cantidades masivas de texto para predecir la siguiente palabra de una secuencia, capacidad que les habilita para generar respuestas coherentes y contextualizadas en lenguaje natural (Brown et al., 2020). Su arquitectura predominante es el Transformer, presentado por Vaswani et al. (2017), sustentada en el mecanismo de autoatención (self-attention) que pondera dinámicamente la relevancia de cada palabra dentro del contexto. Esta arquitectura posibilita paralelizar el aprendizaje sobre grandes corpus textuales y generalizar a tareas no vistas con pocos ejemplos (few-shot learning). El tamaño de estos modelos suele cuantificarse en miles de millones de parámetros, lo que les otorga capacidades emergentes ausentes en modelos más pequeños (Wei et al., 2022). Esta característica explica por qué los avances en el campo han escalado conforme aumentan los recursos computacionales y los corpus de entrenamiento.

El ciclo de vida de un LLM se estructura en tres fases. La primera es el preentrenamiento, en la que el modelo aprende patrones lingüísticos generales a partir de corpus de gran escala. La segunda es el ajuste fino (fine-tuning), que ajusta el modelo a tareas o dominios específicos mediante conjuntos de datos curados. La tercera es la alineación con retroalimentación humana (RLHF, Reinforcement Learning from Human Feedback), que orienta el comportamiento del modelo hacia respuestas seguras, útiles y alineadas con las expectativas del usuario (Ouyang et al., 2022). El éxito del RLHF como técnica de alineación ha consolidado este paradigma como estándar de la industria para modelos comerciales (Ouyang et al., 2022).

El panorama actual de modelos de lenguaje grande se encuentra dominado por cuatro familias principales, cada una desarrollada por

organizaciones líderes del sector tecnológico: Estas familias compiten en capacidades, costo y disponibilidad como pesos abiertos o cerrados, lo que impacta directamente en las decisiones de adopción institucional (Touvron et al., 2023).

- GPT (Generative Pre-trained Transformer), desarrollada por OpenAI: incluye modelos como GPT-3.5, GPT-4 y GPT-4o, este último con capacidades multimodales que procesan texto, imagen y audio.
- Gemini, desarrollada por Google DeepMind: comprende variantes como Gemini Pro, Gemini Flash y Gemini Ultra, con énfasis en la integración nativa con servicios de Google y en el procesamiento eficiente de contextos extensos.
- Claude, desarrollada por Anthropic: abarca modelos como Claude 3 Haiku, Sonnet y Opus, orientados a la seguridad y al razonamiento estructurado mediante el enfoque de Constitutional AI.
- Llama, desarrollada por Meta: agrupa modelos de pesos abiertos como Llama 3.1, Llama 3.2 y Llama 3.3 (con variantes de 8B, 70B y 405B parámetros), así como modelos compactos derivados como Gemma de Google. Estos modelos abiertos resultan especialmente relevantes para entornos institucionales por permitir despliegues locales y auditables.

La evaluación de los LLM se apoya en benchmarks estandarizados como MMLU para conocimiento académico, HumanEval para generación de código, GSM8K y MATH para razonamiento matemático, y HellaSwag para sentido común (Hendrycks et al., 2021). De forma complementaria, LMSYS Chatbot Arena ranquea los modelos mediante preferencias humanas en uso real, lo que permite contrastar el desempeño objetivo con la experiencia subjetiva del usuario.

Pese a sus capacidades, los LLM exhiben limitaciones documentadas: alucinaciones o afirmaciones plausibles pero incorrectas (Ji et al., 2023), sesgos heredados de los datos de entrenamiento (Bender et al., 2021), desconocimiento posterior al corte temporal del corpus y sensibilidad a la formulación del prompt. Estas restricciones justifican estrategias complementarias como RAG, el filtrado de salida y la validación humana en flujos críticos.

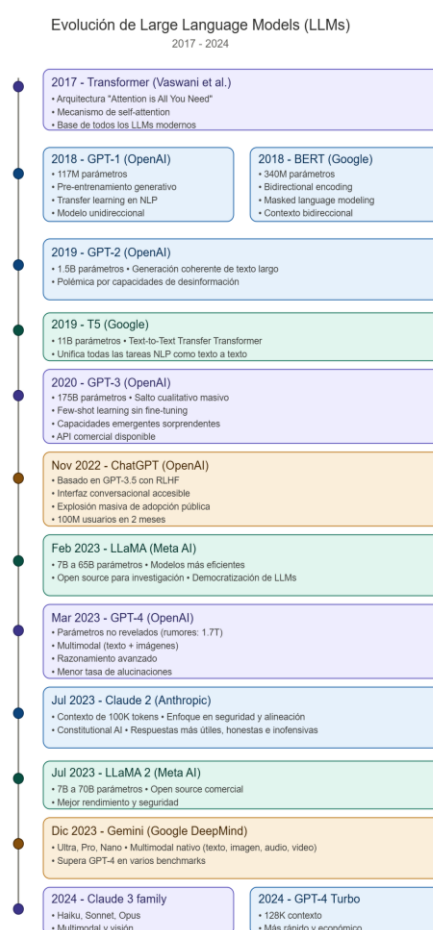
Ingeniería de instrucciones (prompt engineering). La interacción eficaz con los LLM depende del diseño de instrucciones que orienten al modelo hacia respuestas precisas y consistentes. Técnicas como el zero-shot prompting (instrucción directa sin ejemplos), el few-shot prompting (instrucción acompañada de ejemplos demostrativos) y la cadena de pensamiento (chain-of-thought), que explicita pasos intermedios de razonamiento, han demostrado mejoras significativas en tareas de razonamiento complejo (Wei et al., 2022). En entornos institucionales se complementan con instrucciones de sistema (system prompts) que delimitan el rol, el tono y los límites del modelo, asegurando coherencia con la voz y las políticas de la organización.

Criterios de selección para entornos institucionales. La elección de un LLM para despliegues universitarios considera cinco criterios principales: la calidad del razonamiento en el idioma objetivo, el costo por token o por consulta, la latencia de respuesta, la ventana de contexto disponible y las garantías de privacidad respecto a los datos enviados al proveedor. Los modelos de pesos abiertos permiten despliegues locales y auditables, mientras que los modelos cerrados ofrecen mayor desempeño promedio a costa de la dependencia del proveedor. La decisión adecuada equilibra estos factores según la naturaleza de la información manejada y las restricciones operativas de la institución.

La elección de un modelo específico depende del equilibrio buscado entre cuatro variables: la calidad de las respuestas, la latencia de inferencia, el costo por token procesado y el cumplimiento de requisitos de privacidad institucional. En contextos universitarios, los LLM se acoplan con arquitecturas de Generación Aumentada por Recuperación (RAG) para asegurar que las respuestas se sustenten en la documentación oficial y se mitigue el riesgo de alucinaciones (Lewis et al., 2020; Ji et al., 2023).

Figura 3

Evolución de los modelos de lenguaje de gran escala (LLM): principales familias y versiones (2017-2024)



Nota. Adaptado de Brown et al. (2020), Ouyang et al. (2022), Gemini Team (2023) y Llama Team (2024).

3.2.1.8. Bases de datos vectoriales

Una base de datos vectorial constituye un sistema de almacenamiento especializado en vectores numéricos de alta dimensionalidad denominados embeddings, los cuales representan el significado semántico de fragmentos de texto, imágenes o audio (Lewis et al., 2020). A diferencia de las bases de datos relacionales tradicionales, que rastrean coincidencias exactas mediante operaciones SQL sobre tablas estructuradas, las bases de datos vectoriales ejecutan búsquedas por similitud semántica: ante una consulta del usuario, devuelven los elementos cuyo significado es más cercano, incluso cuando las palabras exactas no coinciden. Esta capacidad resulta indispensable en sistemas de inteligencia artificial que procesan lenguaje natural, donde dos frases con palabras distintas pueden expresar la misma idea.

El funcionamiento se sustenta en tres elementos. Primero, los embeddings: vectores numéricos de cientos o miles de dimensiones (típicamente 768, 1024 o 1536) producidos por un modelo de codificación que captura el significado del texto. Segundo, los índices: estructuras de datos optimizadas para acelerar la búsqueda en colecciones masivas, entre las que destacan HNSW (Hierarchical Navigable Small World), IVF (Inverted File Index) y los índices planos (Flat) para colecciones pequeñas (Johnson et al., 2021). Tercero, las métricas de distancia: fórmulas matemáticas que cuantifican la similitud entre vectores, donde la similitud de coseno es la más empleada en aplicaciones de lenguaje natural, complementada con la distancia euclidiana y el producto punto (Manning et al., 2008).

Entre las soluciones más utilizadas se encuentran las siguientes:

- Pinecone: servicio en la nube gestionado con escalabilidad automática y latencia baja, orientado a entornos productivos sin necesidad de gestionar infraestructura propia.

- Chroma: base de datos vectorial de código abierto orientada a despliegues locales, ideal para entornos de desarrollo, prototipado y proyectos académicos.
- FAISS (Facebook AI Similarity Search): biblioteca desarrollada por Meta, ampliamente adoptada en entornos de investigación por su alto rendimiento en colecciones masivas.
- Weaviate: solución de código abierto con capacidades de búsqueda híbrida que combina recuperación vectorial y léxica en una misma consulta.
- Milvus: plataforma distribuida de alto rendimiento, diseñada para colecciones de miles de millones de vectores y entornos empresariales.

La elección entre estas alternativas depende de factores como la escala de la colección documental, los requisitos de latencia, las restricciones de privacidad de datos y el modelo de costos.

En el marco de arquitecturas de Generación Aumentada por Recuperación (RAG), la base de datos vectorial conforma el repositorio central del conocimiento institucional. Los documentos oficiales (reglamentos, cronogramas, manuales y preguntas frecuentes) se procesan en fragmentos (chunks), se convierten en embeddings y se almacenan en el índice vectorial; ante una consulta del estudiante, el sistema recupera los fragmentos semánticamente más similares y los entrega al modelo de lenguaje grande como contexto para generar la respuesta (Y. Gao et al., 2023). Esta integración explica por qué las bases de datos vectoriales se consideran un componente indispensable y no opcional en chatbots institucionales sustentados en RAG.

3.2.2. Gestión de información académica

3.2.2.1. Fundamentación teórica

La gestión de información académica comprende el conjunto de políticas, procesos y responsabilidades que aseguran que los contenidos universitarios (calendarios, reglamentos, trámites y evidencias) sean creados, capturados y difundidos con integridad, trazabilidad y acceso oportuno (ISO 15489-1, 2016). La calidad de datos constituye el pilar técnico: la serie ISO 8000 expone principios para mejorar exactitud, completitud y oportunidad a lo largo del ciclo de vida de los datos (ISO, 2022). El marco integrador corresponde al modelo de DeLone y McLean (2003): la calidad del sistema y la calidad de la información inciden en el uso y la satisfacción del usuario, generando beneficios netos para la organización.

3.2.2.2. Definición conceptual de la variable dependiente

La gestión de información académica se concibe como el conjunto de procesos institucionales mediante los cuales los estudiantes acceden, consultan y emplean información académica oficial, medida a través de tres dimensiones: disponibilidad de la información, acceso y eficiencia operativa (Laudon y Laudon, 2020; ISO 15489-1, 2016).

3.2.2.3. Dimensiones de la variable dependiente

Disponibilidad de la información.

Capacidad institucional de asegurar que los contenidos académicos autorizados estén accesibles sin interrupciones relevantes (ISO/IEC, 2018). Los umbrales clásicos de tiempo de respuesta, planteados por Nielsen (1993), determinan que 0,1 segundos se percibe como instantáneo, 1 segundo preserva el flujo cognitivo del usuario y 10 segundos marca el límite de paciencia antes del abandono; estos umbrales son referencia consolidada en

el campo de la usabilidad de sistemas interactivos. La serie ISO/IEC 20000-1 guía el diseño de un sistema de gestión de servicios con metas de disponibilidad verificables (ISO/IEC, 2018). En el contexto académico universitario, esta dimensión se valora a través de la continuidad del servicio, el horario de cobertura y la ausencia de barreras técnicas que obstaculicen la consulta del estudiante (Laudon y Laudon, 2020).

Acceso a la información.

Capacidad del sistema para que el estudiante localice contenidos pertinentes con la menor fricción posible. Los principios FAIR (Findable, Accessible, Interoperable and Reusable; en español: localizables, accesibles, interoperables y reutilizables) determinan que la información debe ser encontrable, accesible, interoperable y reutilizable (Wilkinson et al., 2016). La precisión de las respuestas constituye una dimensión intrínseca de calidad de datos (Wang y Strong, 1996); la vigencia garantiza que los contenidos estén actualizados respecto de la norma vigente (Ballou y Pazer, 1995); el riesgo de alucinaciones en IA generativa requiere anclar las respuestas a fuentes controladas (Ji et al., 2023). En sistemas universitarios, estas propiedades se traducen en interfaces multicanal, navegación intuitiva y respuestas trazables a la documentación institucional vigente (Nielsen, 1993; Lewis et al., 2020).

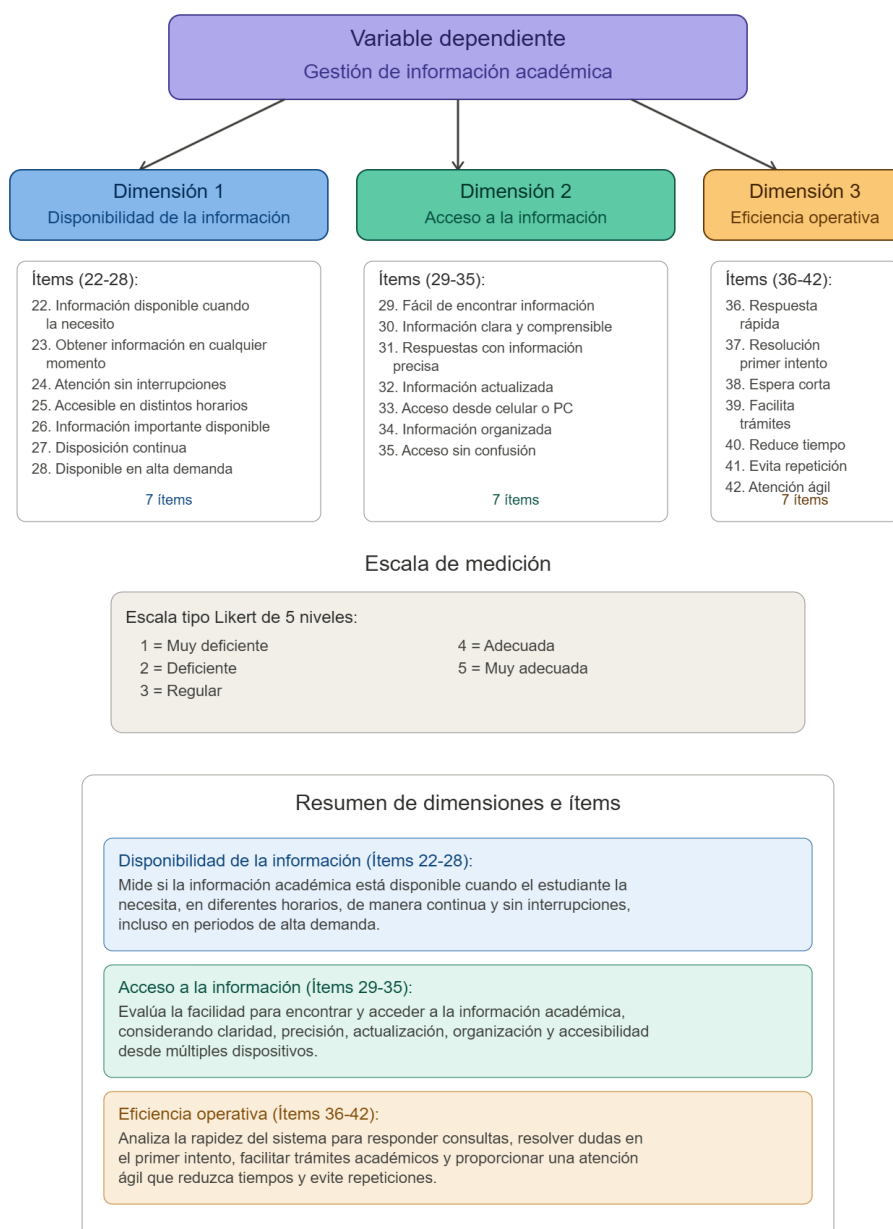
Eficiencia operativa.

Capacidad del canal informativo para resolver consultas con el menor tiempo y recursos posibles. Los indicadores clave corresponden a la tasa de resolución en el primer contacto (FCR, First Contact Resolution), fuertemente correlacionada con la satisfacción del usuario (Abdullateef et al., 2011), y al tiempo medio de gestión (AHT, Average Handle Time) (Gans et al., 2003). La Ley de Little ($L = \lambda W$) explica que disminuir el tiempo de ciclo promedio aminora la congestión percibida (Little, 1961). En el ámbito universitario, una mayor

eficiencia operativa se traduce en menor saturación de los canales de atención presenciales y mejor aprovechamiento del tiempo del personal administrativo (Gans et al., 2003).

Figura 4

Dimensiones de la variable dependiente: disponibilidad, acceso y eficiencia operativa de la gestión de información académica



Nota. La variable se opera mediante 21 ítems (22–42 del cuestionario completo) distribuidos en tres dimensiones de siete ítems cada una.

3.2.2.4. Procesos de gestión académica universitaria

La gestión académica universitaria abarca un conjunto de procesos institucionales orientados a planificar, registrar, controlar y respaldar el ciclo formativo del estudiante. Estos procesos integran a las áreas académicas, administrativas y de servicios, y generan información oficial que el estudiante consulta a lo largo de su trayectoria universitaria. La automatización de las consultas recurrentes mediante agentes conversacionales incide sobre estos procesos al disminuir los tiempos de respuesta y descentralizar el acceso a la información (Laudon y Laudon, 2020).

Matrícula y registro académico.

Constituye el proceso mediante el cual el estudiante formaliza su inscripción en las asignaturas correspondientes a un periodo lectivo. Comprende la verificación de prerrequisitos, la asignación de horarios, la generación de boletas de pago y la emisión de constancias de matrícula. La información oportuna sobre fechas, requisitos y procedimientos resulta crítica, dado que su omisión acarrea consecuencias normativas para el estudiante (SUNEDU, 2020).

Gestión del cronograma académico.

Abarca la difusión y consulta del calendario oficial de actividades académicas, que contempla fechas de inicio y término de ciclo, periodos de evaluación, recesos académicos, jornadas administrativas y plazos para trámites. Se ubica entre las consultas más frecuentes a los canales de atención universitarios y requiere una fuente única de verdad institucional (Laudon y Laudon, 2020).

Trámites y servicios documentarios.

Reúne los procedimientos administrativos que el estudiante efectúa durante su trayectoria, tales como constancias de estudios, certificados, traslados internos, convalidaciones de asignaturas, reincorporaciones y reservas de matrícula. Cada trámite se rige por requisitos específicos, plazos, costos y oficinas responsables que el estudiante debe conocer con precisión (Chiavenato, 2017).

Procesos de grados y títulos.

Agrupar los pasos que conducen a la obtención del grado académico de bachiller y del título profesional. Comprende la sustentación de tesis, la presentación de expedientes, la verificación de planes de estudios y la emisión de los documentos oficiales. La complejidad documentaria de este proceso lo convierte en uno de los principales focos de consulta estudiantil (SUNEDU, 2020).

Sistema de Gestión Académica.

Constituye la plataforma informática institucional que respalda los procesos descritos. Centraliza el expediente académico del estudiante, los registros de notas, los planes de estudio y los flujos administrativos. La información que un agente conversacional entregue debe estar alineada con la documentación oficial gestionada en este sistema, asegurando la trazabilidad y la consistencia entre los canales presenciales y digitales (Laudon y Laudon, 2020).

3.2.2.5. Marco normativo peruano sobre gestión académica e inteligencia artificial

La gestión de información académica universitaria en el Perú se rige por un marco normativo que integra la calidad universitaria, la transformación digital

del Estado y el desarrollo de la inteligencia artificial. La presente investigación se inscribe en dicho marco, en línea con los objetivos institucionales de la Universidad Tecnológica de los Andes y los compromisos del Estado peruano.

Ley Universitaria N.º 30220.

Constituye la norma marco que regula la educación universitaria en el Perú. Determina las funciones de la universidad, los principios académicos y administrativos, así como los requisitos para el funcionamiento institucional. La Ley dispone que las universidades deben asegurar el acceso oportuno de los estudiantes a la información académica y a los servicios institucionales como condición de calidad educativa (Congreso de la República del Perú, 2014).

Licenciamiento institucional de la SUNEDU.

La SUNEDU fiscaliza el cumplimiento de las condiciones básicas de calidad para el funcionamiento de las universidades peruanas. Entre dichas condiciones figuran la transparencia de la información académica y la disponibilidad de servicios estudiantiles eficientes, que conforman el sustento normativo para iniciativas de innovación en la gestión informativa, como la implementación de chatbots con inteligencia artificial (SUNEDU, 2020).

Política Nacional de Transformación Digital al 2030.

Aprobada por la PCM, define los lineamientos para la digitalización de los servicios públicos, el fortalecimiento de las capacidades digitales de la población y la gobernanza de los datos. La incorporación de un chatbot con inteligencia artificial en la atención de consultas académicas se alinea con los ejes estratégicos de servicios digitales accesibles y centrados en el ciudadano (Presidencia del Consejo de Ministros, 2023).

Ley N.º 31814 sobre el uso de la inteligencia artificial.

La Ley impulsa el uso de la inteligencia artificial en favor del desarrollo económico y social del país, con un enfoque centrado en la persona y respetuoso de los derechos fundamentales. Define principios de transparencia, no discriminación, privacidad, seguridad y rendición de cuentas que orientan el diseño de soluciones sustentadas en inteligencia artificial en el sector educativo, incluidos los chatbots universitarios (Congreso de la República, 2023).

3.2.3. Fundamentos metodológicos y estadísticos

La presente investigación se apoya en un conjunto de fundamentos metodológicos y estadísticos ampliamente reconocidos en la investigación cuantitativa en ciencias sociales y educativas. A continuación se exponen los marcos teóricos que orientan el diseño de la investigación, la construcción y validación del instrumento, así como el análisis estadístico de los datos.

3.2.3.1. Diseño cuasi-experimental

El diseño cuasi-experimental representa una variante robusta de la investigación experimental, aplicable cuando la asignación aleatoria de los participantes a los grupos de tratamiento y control no resulta viable (Shadish et al., 2002). En contextos educativos, los estudiantes se encuentran organizados en grupos intactos preexistentes (aulas, carreras, ciclos), lo que restringe la aleatorización individual pero habilita la comparación entre grupos cuya equivalencia inicial debe verificarse empíricamente. Hernández-Sampieri y Mendoza (2018) precisan que en los diseños cuasi-experimentales con grupo control, ambos grupos se miden antes y después de la intervención, lo que posibilita controlar parcialmente las amenazas a la validez interna asociadas al paso del tiempo, a la maduración y al efecto de la prueba previa. Este diseño

es el asumido en la presente investigación. Las amenazas que aún subsisten en este diseño abarcan la selección, la regresión estadística y la interacción de la selección con otros factores, las cuales se mitigan mediante la equivalencia inicial entre grupos verificada en el pretest (Shadish et al., 2002). Para fortalecer la validez interna del cuasi-experimento se incorporan procedimientos complementarios como la verificación de la homogeneidad inicial de los grupos, el control de las condiciones de aplicación del instrumento y la estandarización del periodo de intervención (Shadish et al., 2002). Adicionalmente, los diseños cuasi-experimentales con grupo control son ampliamente aceptados en investigación educativa porque articulan el rigor metodológico del enfoque experimental con la viabilidad operativa que exigen los contextos escolares y universitarios reales (Creswell y Creswell, 2018).

3.2.3.2. *Pretest y posttest*

El pretest y el posttest son mediciones efectuadas antes y después de la intervención experimental, respectivamente, sobre el mismo grupo o grupos comparables, con el propósito de evaluar el efecto de la intervención (Hernández-Sampieri y Mendoza, 2018). La diferencia entre ambas mediciones posibilita estimar el cambio atribuible al tratamiento, y la comparación con un grupo control que no recibió la intervención fortalece la validez de la inferencia causal. En la presente investigación, ambos grupos (experimental y control) se miden en pretest y posttest para identificar el efecto del chatbot sobre la gestión de información académica.

3.2.3.3. *Cuestionario tipo Likert*

El cuestionario tipo Likert, planteado por Likert (1932), constituye un instrumento de medición integrado por afirmaciones a las que el encuestado responde mediante una escala ordinal, típicamente de cinco niveles, expresando su grado de acuerdo o desacuerdo. Se ubica entre los

instrumentos más empleados en investigación educativa por su capacidad para medir actitudes, percepciones y opiniones de manera cuantitativa, y por la sencillez de su aplicación y procesamiento. La presente investigación emplea un cuestionario tipo Likert de cinco niveles con 42 ítems agrupados en dimensiones que operacionalizan ambas variables.

3.2.3.4. Validez de contenido: V de Aiken

La validez de contenido alude al grado en que los ítems de un instrumento representan adecuadamente el dominio del constructo que se pretende medir. El coeficiente V de Aiken, planteado por Aiken (1985), constituye un estadístico cuantitativo que valora el grado de acuerdo entre jueces expertos respecto a la pertinencia, claridad y relevancia de los ítems. Adopta valores entre 0 y 1, y se considera aceptable a partir de 0,80 (Escrura, 1988). En la presente investigación, la validez del cuestionario fue valorada por tres expertos, obteniéndose un coeficiente V de Aiken igual a 1,00, lo que indica acuerdo total entre los evaluadores.

3.2.3.5. Fiabilidad: coeficiente alfa de Cronbach

La fiabilidad de un instrumento de medición alude al grado en que sus mediciones son consistentes y reproducibles. El coeficiente alfa de Cronbach, planteado por Cronbach (1951), constituye el estadístico más empleado para estimar la consistencia interna de un instrumento compuesto por ítems que miden un mismo constructo. Adopta valores entre 0 y 1, y se considera aceptable a partir de 0,70 y excelente a partir de 0,90. En la presente investigación, el coeficiente alfa de Cronbach del cuestionario alcanzó un valor de 0,907 en la prueba piloto ($n = 30$), lo que confirma la excelente fiabilidad del instrumento.

3.2.3.6. Prueba de normalidad: Kolmogorov-Smirnov

Antes de seleccionar las pruebas estadísticas inferenciales, resulta necesario verificar si los datos siguen una distribución normal, condición que define si corresponde aplicar estadística paramétrica o no paramétrica (Field, 2018). La prueba de Kolmogorov-Smirnov constituye un contraste estadístico que examina la hipótesis nula de que una muestra proviene de una distribución normal; se recomienda para muestras grandes ($n > 50$), a diferencia de la prueba de Shapiro-Wilk, que se reserva para muestras pequeñas o medianas (Razali y Wah, 2011). El criterio de decisión asumido es el siguiente: si el valor de significancia (p) es mayor a 0,05, los datos presentan distribución normal; si es menor o igual a 0,05, los datos no presentan distribución normal y se emplean pruebas no paramétricas.

3.2.3.7. Pruebas no paramétricas: Wilcoxon y U de Mann-Whitney

Cuando los datos no satisfacen el supuesto de normalidad, la inferencia estadística se efectúa mediante pruebas no paramétricas. La prueba de Wilcoxon, planteada por Wilcoxon (1945), posibilita comparar dos mediciones del mismo grupo (pretest y posttest) y representa la alternativa no paramétrica a la prueba t de Student para muestras relacionadas. La prueba U de Mann-Whitney, planteada por Mann y Whitney (1947), posibilita comparar dos grupos independientes (grupo experimental y grupo control) y representa la alternativa no paramétrica a la prueba t de Student para muestras independientes. Ambas pruebas se sustentan en el ordenamiento de los datos en rangos y resultan apropiadas para variables ordinales o de intervalo cuyas distribuciones no son normales (Field, 2018).

3.3. Definición de términos

Admisión universitaria.

Proceso institucional a través del cual los postulantes acceden a la universidad, abarcando las etapas de inscripción, examen de admisión, asignación de vacantes y matrícula formal en el programa académico (Congreso de la República del Perú, 2014).

Agente de inteligencia artificial.

Componente inteligente que recurre a un modelo de lenguaje grande como núcleo de razonamiento para definir qué herramientas (búsqueda vectorial, web scraping, interfaces de programación de aplicaciones) utilizar según la consulta del usuario (Russell y Norvig, 2021).

Alfa de Cronbach.

Coefficiente estadístico que evalúa la consistencia interna de un instrumento de medición integrado por varios ítems; adopta valores entre 0 y 1, donde valores superiores a 0,70 se consideran aceptables y superiores a 0,90 se consideran excelentes en estudios sociales y educativos (Cronbach, 1951; Tavakol y Dennick, 2011).

Alucinación en inteligencia artificial.

Fenómeno mediante el cual un modelo de lenguaje grande produce información inexacta o inventada exhibiéndola como verdadera. Representa uno de los riesgos centrales que la arquitectura RAG procura mitigar (Ji et al., 2023).

Aprendizaje automático.

Disciplina de la inteligencia artificial dedicada al diseño de algoritmos capaces de aprender patrones a partir de datos sin programación explícita, perfeccionando su desempeño con la experiencia (Cuayáhuitl et al., 2019).

Aprendizaje profundo.

Rama del aprendizaje automático sustentada en redes neuronales artificiales con múltiples capas, capaz de aprender representaciones jerárquicas de los datos y de abordar tareas complejas en visión, lenguaje y reconocimiento de patrones (Goodfellow et al., 2016).

Arquitectura Transformer.

Arquitectura de red neuronal profunda presentada por Vaswani et al. (2017), sustentada en el mecanismo de autoatención, que posibilita procesar secuencias de texto en paralelo y conforma la base de los modelos de lenguaje grande contemporáneos.

Base de datos vectorial.

Sistema de almacenamiento especializado en vectores de alta dimensionalidad (embeddings) que habilita búsquedas por similitud semántica en lugar de coincidencia exacta (Lewis et al., 2020).

Chatbot.

Aplicación de software sustentada en inteligencia artificial concebida para sostener conversaciones automáticas con usuarios humanos a través de texto o voz, emulando una interacción natural (Adamopoulou y Moussiades, 2020).

Chunking.

Proceso a través del cual los documentos extensos se dividen en fragmentos más pequeños (chunks) con superposición controlada (overlap) antes de transformarse en embeddings, optimizando su procesamiento por los modelos de embeddings y la recuperación posterior (Y. Gao et al., 2023).

Comprensión del lenguaje natural (NLU, Natural Language Understanding).

Subcampo del procesamiento del lenguaje natural dedicado a la interpretación semántica de los enunciados del usuario, abarcando el reconocimiento de la intención y la extracción de entidades relevantes (Caldarini et al., 2022).

Diseño cuasi-experimental.

Diseño de investigación en el que se manipula una variable independiente para examinar su efecto sobre una variable dependiente, sin asignación aleatoria de los participantes a los grupos experimental y de control. Resulta ampliamente empleado en contextos educativos cuando no es posible aleatorizar la muestra (Hernández-Sampieri y Mendoza, 2018; Shadish et al., 2002).

Embeddings.

Representaciones numéricas vectoriales de palabras, frases o documentos que capturan su significado semántico, habilitando comparaciones y búsquedas por similitud en lugar de coincidencia exacta (Lewis et al., 2020).

Generación Aumentada por Recuperación (RAG, Retrieval-Augmented Generation).

Arquitectura híbrida que articula la recuperación de información relevante desde una base de conocimiento externa con la generación de respuestas mediante un modelo de lenguaje grande, perfeccionando la precisión y disminuyendo las alucinaciones (Lewis et al., 2020).

Indicador clave de desempeño (KPI, Key Performance Indicator).

Métrica empleada para evaluar el éxito de un proyecto o sistema, expresada habitualmente como tiempo de respuesta, tasa de resolución y nivel de satisfacción del usuario (Laudon y Laudon, 2020).

Inteligencia artificial.

Disciplina de la informática orientada al desarrollo de sistemas capaces de ejecutar tareas que tradicionalmente exigen inteligencia humana, tales como el razonamiento, el aprendizaje y la comprensión del lenguaje (Russell y Norvig, 2021).

Inteligencia artificial generativa.

Subcampo de la inteligencia artificial que produce nuevo contenido (texto, imagen, audio, código) a partir de patrones aprendidos en grandes conjuntos de datos, mediante modelos como redes generativas antagónicas y modelos de lenguaje grande (Brown et al., 2020).

Interfaz de programación de aplicaciones (API).

Conjunto de definiciones, protocolos y herramientas que habilita la comunicación entre dos sistemas de software, facilitando el consumo de modelos de lenguaje, sistemas de mensajería y bases de datos externas (Fielding, 2000).

Ley N.º 31814.

Ley peruana que impulsa el uso de la inteligencia artificial en favor del desarrollo económico y social del país, fijando principios de transparencia, no discriminación, privacidad y seguridad para las soluciones sustentadas en inteligencia artificial (Congreso de la República, 2023).

Ley Universitaria N.º 30220.

Norma marco que rige la educación universitaria en el Perú, determina las funciones de la universidad, los principios académicos y administrativos, así como los requisitos para el funcionamiento institucional (Congreso de la República del Perú, 2014).

Modelo de lenguaje de gran escala (LLM, Large Language Model).

Modelo de inteligencia artificial entrenado con grandes volúmenes de texto que habilita la generación de respuestas coherentes en lenguaje natural (Brown et al., 2020).

OAuth 2.0.

Protocolo de autorización estándar que habilita a aplicaciones de terceros para acceder a recursos de un usuario sin compartir sus credenciales (Hardt, 2012).

Procesamiento del lenguaje natural (PLN).

Rama de la inteligencia artificial que examina la interacción entre las computadoras y el lenguaje humano, habilitando la comprensión, interpretación y generación de texto de manera automatizada (Caldarini et al., 2022).

Prompt engineering.

Disciplina aplicada centrada en el diseño y la optimización de las instrucciones (prompts) entregadas a un modelo de lenguaje grande para obtener respuestas precisas y contextualmente adecuadas (Brown et al., 2020).

Prueba de Wilcoxon de los rangos con signo.

Prueba estadística no paramétrica empleada para comparar dos mediciones vinculadas en una misma muestra (por ejemplo, pretest y posttest) si los datos no satisfacen el supuesto de normalidad; examina si existe diferencia significativa entre las medianas de los pares de observaciones (Wilcoxon, 1945).

Prueba U de Mann-Whitney.

Prueba estadística no paramétrica empleada para comparar dos muestras independientes cuando los datos no satisfacen el supuesto de normalidad; evalúa si las distribuciones de ambos grupos difieren significativamente mediante el contraste de rangos (Mann y Whitney, 1947).

RAGAS.

Conjunto de métricas dirigidas a evaluar sistemas de Generación Aumentada por Recuperación, que examina la fidelidad de la respuesta a los fragmentos recuperados, la pertinencia de la respuesta a la consulta y la suficiencia del contexto recuperado (Es et al., 2024).

Reconocimiento óptico de caracteres (OCR, Optical Character Recognition).

Tecnología que identifica y transforma el texto contenido en imágenes o documentos escaneados a texto digital editable, lo que habilita la incorporación de documentos no digitalizados a sistemas de procesamiento automatizado (Russell y Norvig, 2021).

Similitud de coseno.

Métrica matemática que cuantifica el grado de semejanza entre dos vectores, calculando el coseno del ángulo entre ellos; adopta valores entre 0 (sin relación) y 1 (idénticos), y conforma la operación central de la búsqueda semántica (Manning et al., 2008).

Sistema web.

Aplicación informática alojada en un servidor remoto a la que los usuarios ingresan mediante un navegador a través de internet, sin necesidad de instalación local (Fielding, 2000).

Superintendencia Nacional de Educación Superior Universitaria (SUNEDU).

Organismo público peruano responsable de licenciar y fiscalizar el desempeño de las universidades dentro del marco de las condiciones básicas de calidad determinadas por la Ley Universitaria N.º 30220 (Congreso de la República del Perú, 2014; SUNEDU, 2020).

System prompt.

Conjunto de instrucciones iniciales que delinean la personalidad, las restricciones, el idioma y el comportamiento esperado del modelo de lenguaje en la conversación con el usuario (Ouyang et al., 2022).

Temperatura.

Parámetro del modelo de lenguaje grande que controla la aleatoriedad de las respuestas; valores bajos (0,2) generan respuestas deterministas y reproducibles, en tanto que valores altos producen respuestas más creativas y diversas (Brown et al., 2020).

Tokens.

Unidades mínimas de procesamiento de texto en un modelo de lenguaje grande, equivalentes a palabras o fragmentos de palabras. El parámetro max_tokens restringe la longitud máxima de la respuesta e incide directamente en el costo de la inferencia (Brown et al., 2020).

Top-K.

Parámetro de búsqueda vectorial que fija la cantidad de documentos más relevantes (K) que se recuperan en respuesta a una consulta. Un valor de K alto amplía la cobertura, en tanto que un valor bajo prioriza la precisión (Y. Gao et al., 2023).

Universidad Tecnológica de los Andes.

Institución universitaria peruana licenciada por la Superintendencia Nacional de Educación Superior Universitaria, creada en 1984, con sede principal en Abancay, departamento de Apurímac, y filiales en Andahuaylas y Cusco (SUNEDU, 2020).

V de Aiken.

Coeficiente que cuantifica la validez de contenido de un instrumento a partir de los juicios de un panel de expertos; adopta valores entre 0 y 1, considerando aceptables los valores iguales o superiores a 0,80 para sostener la pertinencia y claridad de los ítems (Aiken, 1985).

Webhook.

Mecanismo de integración basado en HTTP que permite a un sistema notificar a otro de manera automática e inmediata cuando ocurre un evento; en chatbots, se emplea para recibir mensajes entrantes desde plataformas de mensajerías y disparar el procesamiento correspondiente (Fielding, 2000).

IV. Metodología

4.1. Tipo y nivel de investigación

4.1.1. Tipo de investigación

El presente estudio corresponde a una investigación aplicada de carácter tecnológico. La investigación aplicada se caracteriza por orientar el conocimiento adquirido hacia la solución de problemas concretos, a diferencia de la investigación básica centrada en la sistematización teórica (Vargas, 2009). En la misma línea, las investigaciones aplicadas producen conocimiento práctico para problemas específicos en contextos reales mediante marcos teóricos previamente validados (Hernández-Sampieri y Mendoza, 2018). Esta clasificación se concreta en el diseño, la implementación y la evaluación empírica de un chatbot con inteligencia artificial basado en arquitectura RAG, desplegado en la Universidad Tecnológica de los Andes para mejorar la gestión de información académica. Su carácter tecnológico se apoya en el desarrollo de una solución de software funcional valorada mediante indicadores objetivos de calidad, accesibilidad y eficiencia operativa (CONCYTEC, 2018).

4.1.2. Nivel de investigación

El estudio se llevó a cabo con un nivel explicativo, dirigido a identificar las causas de los fenómenos observados y las relaciones de dependencia entre las variables. Los estudios explicativos trascienden la descripción y se orientan a dar cuenta de las causas de los eventos e identificar las condiciones en que se manifiestan (Hernández-Sampieri y Mendoza, 2018). En la presente investigación, este nivel se concreta al examinar el efecto de la implementación del chatbot con inteligencia artificial (variable independiente) sobre la gestión de información académica (variable dependiente), valorando si la intervención produce cambios estadísticamente significativos en las dimensiones de disponibilidad, acceso y

eficiencia operativa, y estimando la magnitud de dichos cambios mediante el tamaño del efecto. La elección del nivel explicativo resulta coherente con el diseño cuasi-experimental, pues ambos pretenden establecer relaciones de causa y efecto, si bien bajo condiciones de control parcial, propias de los diseños que operan con unidades preexistentes (Shadish et al., 2002).

El enfoque de la investigación fue cuantitativo. Este enfoque recurre a la recolección y el análisis de datos numéricos para contrastar hipótesis previamente formuladas (Hernández-Sampieri y Mendoza, 2018). En el presente estudio, se materializó en la medición de la gestión de información académica antes y después de la intervención y en el procesamiento estadístico de los datos del grupo experimental y del grupo control.

El diseño de la investigación fue cuasi-experimental de grupo control no equivalente, con pretest y posttest, de corte longitudinal, dado que las mediciones se realizaron en dos momentos temporales distintos. Este diseño se adopta cuando la asignación aleatoria individual de los participantes no resulta viable (Shadish et al., 2002). En el presente estudio, el grupo experimental corresponde a un conglomerado intacto conformado por la institución y el grupo de comparación se integró por selección individual entre las escuelas profesionales restantes, conforme se detalla en el apartado 4.3.3. Su aplicación permitió contrastar las mediciones de pretest y posttest del grupo experimental con las del grupo control.

El método de la investigación fue hipotético-deductivo. Este método se sustenta en hipótesis derivadas de un marco teórico y las somete a contrastación empírica para confirmarlas, modificarlas o rechazarlas (Bunge, 2014). En el presente estudio se materializó en la formulación de hipótesis sobre el efecto del chatbot, su operacionalización en indicadores medibles y su contrastación a través de las pruebas de Wilcoxon y U de Mann-Whitney.

4.2. Ámbito temporal y espacial

El ámbito temporal abarcó dos etapas. La primera, durante el periodo 2025-II, comprendió la formulación del proyecto, la validación del instrumento y la aprobación del plan de tesis. La segunda, entre enero y abril de 2026, correspondió al trabajo de campo: pretest, recopilación documental institucional, desarrollo e implementación del chatbot, intervención con el grupo experimental y posttest a ambos grupos. Cabe precisar que la denominación del estudio consigna el año 2025 por corresponder al periodo académico de formulación, aprobación e inicio de la investigación (2025-II), periodo que asimismo delimita la población de estudio (estudiantes matriculados en 2025-II); el trabajo de campo concluyó en abril de 2026, conforme al cronograma expuesto, sin que ello modifique la cohorte estudiada ni la denominación registrada del proyecto.

El ámbito espacial fue la Universidad Tecnológica de los Andes, sede Abancay (Apurímac, Perú), institución licenciada por la SUNEDU que atiende a 9 315 estudiantes de pregrado matriculados en el semestre 2025-II en sus tres sedes (UTEA, 2025). La elección se sustentó en la accesibilidad institucional, la pertinencia regional por las brechas de conectividad de Apurímac y la prioridad de la transformación digital universitaria establecida por el marco normativo peruano (Presidencia del Consejo de Ministros, 2023).

4.3. Población y muestra

4.3.1. Población

La población se define como el conjunto de todos los casos que cumplen determinadas especificaciones y configura el universo de elementos del cual se procura generalizar los resultados del estudio (Hernández-Sampieri y Mendoza, 2018). En la misma línea, la población corresponde a un conjunto finito o infinito de

elementos con rasgos comunes, delimitado por el problema y los objetivos del estudio (Arias, 2020).

En el presente estudio, la población quedó integrada por 3 983 estudiantes de pregrado matriculados en la Universidad Tecnológica de los Andes, sede Abancay, durante el periodo 2025-II, según el reporte estadístico de matrícula de la Dirección de Planeamiento y Desarrollo Universitario de la UTEA (2025), que consigna 9 315 estudiantes a nivel institucional (3 983 en la sede Abancay, 2 130 en la filial Andahuaylas y 3 202 en la filial Cusco). De ese total, 288 estudiantes correspondían a la Escuela Profesional de Estomatología de la sede Abancay. El mismo reporte consigna, para el semestre 2026-I —periodo en que se ejecutó el trabajo de campo—, 9 283 estudiantes a nivel institucional, 3 941 en la sede Abancay y 286 en la Escuela Profesional de Estomatología, sin variación en la composición de escuelas profesionales; el detalle por escuela profesional y semestre se presenta en el Anexo 13. Este colectivo conforma el grupo directamente afectado por la problemática y destinatario de la intervención. Como criterio de inclusión se consideró a los estudiantes con matrícula vigente y se excluyó a quienes presentaban matrícula reservada o suspendida.

4.3.2. Muestra

La muestra se define como un subgrupo representativo de la población del cual se recogen los datos y que debe delimitarse con precisión para sostener la generalización de los resultados al universo de estudio (Hernández-Sampieri y Mendoza, 2018). En la misma línea, la muestra constituye una porción significativa de la población cuyos elementos expresan, con objetividad, los rasgos esenciales del conjunto del que provienen (Carrasco Díaz, 2006).

En el presente estudio, la muestra quedó conformada por 380 estudiantes de pregrado, distribuidos en dos grupos de 190: el grupo experimental, que interactuó con el chatbot, y el grupo control, que continuó con los canales tradicionales de

atención académica. El grupo experimental se conformó con estudiantes de la Escuela Profesional de Estomatología de la sede Abancay, unidad académica que otorgó la autorización operativa y el acceso presencial sostenido que la intervención requería; sus 190 participantes representan el 66,0 % de los 288 estudiantes matriculados en dicha escuela durante el semestre 2025-II (66,4 % de los 286 matriculados en 2026-I), de modo que el grupo experimental constituye una cobertura mayoritaria del conglomerado intervenido y no una fracción marginal de este. El grupo control se integró con 190 estudiantes procedentes de las once escuelas profesionales restantes de la sede Abancay, que reunían 3 695 matriculados, lo que equivale a una fracción muestral de 5,1 %. Esta configuración —un conglomerado intacto intervenido frente a un grupo de comparación conformado por selección individual— corresponde al diseño de grupo control no equivalente (Shadish et al., 2002). El tamaño muestral se justificó mediante un análisis de potencia estadística a priori conforme a los parámetros convencionales de Cohen (1988): para detectar un tamaño de efecto mediano ($d = 0,50$) en la comparación de dos grupos independientes, con un nivel de significancia $\alpha = 0,05$ (bilateral) y una potencia de $1 - \beta = 0,80$, se requieren aproximadamente 64 participantes por grupo; ajustando este valor por la eficiencia relativa asintótica de las pruebas no paramétricas empleadas ($ARE = 0,955$ de la prueba U de Mann-Whitney respecto de la prueba t), el requerimiento asciende a 67 participantes por grupo, según el cálculo verificado con el programa G*Power 3 (Faul et al., 2007). Los 190 participantes por grupo del presente estudio casi triplican dicho mínimo, con lo cual la potencia alcanzada para efectos medianos supera 0,99 y el diseño posee sensibilidad suficiente para detectar efectos desde $d \approx 0,29$ (pequeños a medianos) con potencia de 0,80. Adicionalmente, dado que el grupo experimental abarca el 66,0 % de la población de su escuela profesional, corresponde aplicar el factor de corrección por población finita, cuyo valor $\sqrt{[(288 - 190)/(288 - 1)]} = 0,584$ reduce en aproximadamente 42 % los errores estándar de las estimaciones respecto del supuesto de población infinita, con lo que

la precisión efectiva del contraste resulta superior a la estimada por el análisis de potencia; en el grupo control, con una fracción muestral de 5,1 %, dicha corrección resulta despreciable y no se aplica. No obstante, al tratarse de un procedimiento de selección no probabilístico, la generalización estadística de los resultados se restringe a poblaciones con características equivalentes a las de la sede Abancay, límite que se declara expresamente en la discusión (apartado 5.3).

Complementariamente, y a fin de verificar la suficiencia del tamaño muestral respecto del universo de estudio, se aplicó la fórmula de cálculo para poblaciones finitas (Hernández-Sampieri y Mendoza, 2018):

$$n = (N \cdot Z^2 \cdot p \cdot q) / [e^2 \cdot (N - 1) + Z^2 \cdot p \cdot q]$$

Donde $N = 3\,983$ estudiantes de pregrado matriculados en la sede Abancay durante el semestre 2025-II; $Z = 1,96$, valor crítico correspondiente a un nivel de confianza del 95 %; $p = 0,50$ y $q = 0,50$, proporciones de máxima varianza asumidas al no disponer de estudios previos sobre la variable en esta población; y $e = 0,05$, margen de error admitido. Sustituyendo los valores:

$$n = (3\,983 \cdot 3,8416 \cdot 0,25) / [0,0025 \cdot 3\,982 + 0,9604] = 3\,825,27 / 10,9154 = 350,45 \approx 351$$

El tamaño mínimo requerido asciende, por tanto, a 351 estudiantes. Este resultado es invariante respecto del semestre considerado: al sustituir la matrícula del semestre 2026-I ($N = 3\,941$), periodo de ejecución del trabajo de campo, la fórmula arroja $3\,784,94 / 10,8104 = 350,12 \approx 351$ estudiantes. La muestra efectivamente trabajada ($n = 380$) supera dicho umbral y comprende dos grupos de 190 unidades de análisis cada uno, condición necesaria para el contraste pretest-posttest con grupo control propio del diseño cuasi-experimental adoptado. Cabe precisar que este cálculo se presenta como criterio de suficiencia y no como

procedimiento de selección, dado que la conformación de los grupos no obedeció a un procedimiento probabilístico, conforme se detalla en el apartado 4.3.3.

4.3.3. Muestreo

El muestreo no probabilístico es aquel en el que los participantes no se eligen al azar, sino según criterios fijados por el investigador, como la accesibilidad y la disponibilidad (Otzen y Manterola, 2017). El muestreo por conveniencia, una variante de este tipo, consiste en escoger a los participantes que se hallan disponibles dentro del contexto del estudio.

En la presente investigación se empleó un procedimiento de selección no probabilístico y asimétrico entre grupos: el grupo experimental corresponde a un conglomerado intacto conformado por la institución —la Escuela Profesional de Estomatología, con 190 de sus 288 estudiantes matriculados—, mientras que el grupo control se integró por selección individual de 190 estudiantes entre los 3 695 matriculados en las once escuelas profesionales restantes de la sede Abancay. El diseño cuasi-experimental prescinde de la asignación aleatoria individual de los participantes y opera con unidades preexistentes (Shadish et al., 2002). En el presente estudio, dicha condición no derivó de una imposibilidad conceptual del diseño, sino de tres restricciones concretas del contexto institucional: la conformación de los grupos se encuentra fijada por el proceso de matrícula, que la universidad no autoriza a alterar con fines de investigación; la intervención exigía acceso presencial sostenido a los ambientes de la Escuela Profesional de Estomatología, unidad que otorgó la autorización operativa para el periodo de trabajo de campo; y el sistema chatbot constituye un servicio de información académica de acceso abierto, de modo que negar deliberadamente su uso a estudiantes seleccionados al azar habría contravenido el principio de justicia asumido en el apartado 4.7. En consecuencia, el control del sesgo de selección no se resolvió mediante aleatorización, sino mediante la verificación empírica de la equivalencia inicial entre grupos con la prueba U de

Mann-Whitney aplicada al pretest (véase la Tabla 12) y mediante el contraste intergrupos de las diferencias posttest – pretest (véanse las Tablas 17 y 18), procedimiento que neutraliza el efecto de cualquier diferencia basal residual. El tamaño alcanzado ($n = 380$; 190 por grupo) supera con holgura los dos mínimos calculados en el apartado 4.3.2: el de 351 estudiantes que arroja la fórmula para poblaciones finitas y el de 67 participantes por grupo que exige el análisis de potencia estadística, si bien la naturaleza no probabilística del muestreo delimita el alcance de la generalización de los resultados.

4.4. Instrumentos

4.4.1. Técnicas

Las técnicas de recolección de datos corresponden a los procedimientos utilizados por el investigador para recoger la información requerida por su estudio (Hernández-Sampieri y Mendoza, 2018). En el presente estudio se aplicaron tres técnicas complementarias: la encuesta, la prueba y el análisis documental.

4.4.1.1. La encuesta.

La encuesta constituye una técnica que aplica un mismo conjunto de preguntas a varios participantes para conocer sus opiniones, actitudes o percepciones, y que posibilita el análisis de las respuestas con métodos estadísticos (Casas et al., 2003). En el presente trabajo se aplicó la encuesta para medir la percepción de los estudiantes respecto a la gestión de información académica antes y después de la intervención con el chatbot.

4.4.1.2. La prueba.

La prueba constituye una técnica de recolección de datos que posibilita medir el nivel de una variable en un momento determinado mediante la aplicación de un instrumento estructurado (Carrasco Díaz, 2006). En el presente trabajo se aplicó la prueba en dos momentos del trabajo de campo:

una prueba inicial (pretest), administrada antes de la puesta en marcha del chatbot, y una prueba final (postest), administrada después de la intervención. La comparación estadística entre ambas mediciones representa el mecanismo principal para contrastar las hipótesis del estudio (Shadish et al., 2002).

4.4.1.3. El análisis documental.

El análisis documental constituye una técnica de recolección de datos consistente en la revisión sistemática de documentos institucionales, normativos o académicos para extraer información relevante para los objetivos del estudio (Hernández-Sampieri y Mendoza, 2018). En el presente trabajo, esta técnica se aplicó para reunir la documentación institucional vigente de la Universidad Tecnológica de los Andes, que constituye el insumo de la base de conocimiento del chatbot y, por tanto, resulta indispensable para la construcción de la intervención, esto es, del sistema chatbot.

4.4.2. Instrumentos

Para cada técnica se utilizó un instrumento específico, diseñado para recoger la información correspondiente a las variables del estudio.

4.4.2.1. Cuestionario tipo Likert.

Likert (1932) planteó una escala ordinal que cuantifica el nivel de acuerdo de los encuestados ante un conjunto de afirmaciones. El cuestionario adoptado fue de cinco niveles, con 42 ítems en dos bloques: los ítems 1 al 21, correspondientes a la variable independiente Chatbot con inteligencia artificial (capacidad, usabilidad y disponibilidad), aplicados solo al grupo experimental en pretest (expectativa) y postest (experiencia); estos 21 ítems conforman el índice de percepción de la calidad del sistema, que opera como variable de resultado secundaria del estudio, en tanto la variable independiente propiamente dicha es la condición de tratamiento dicotómica (0 = sin chatbot,

1 = con chatbot). Por tratarse de una medida de resultado secundaria, este índice se analiza únicamente en forma agregada; la consistencia interna de sus tres dimensiones se reporta en el apartado 4.4.3; y los ítems 22 al 42, correspondientes a la variable dependiente Gestión de información académica (disponibilidad, acceso y eficiencia operativa), aplicados a ambos grupos en pretest y posttest. La escala se cuantificó con valores enteros del 1 al 5 (1 = muy deficiente, 2 = deficiente, 3 = regular, 4 = adecuada y 5 = muy adecuada).

4.4.2.2. Pretest y posttest.

El pretest y el posttest conforman el instrumento específico de la técnica de prueba empleada en el diseño cuasi-experimental. El pretest cuantifica el estado inicial de la variable dependiente antes de la intervención y habilita el establecimiento de una línea base; el posttest, aplicado después de la implementación del chatbot, cuantifica el estado final de la variable y habilita la identificación de los cambios atribuibles a la intervención (Shadish et al., 2002; Hernández-Sampieri y Mendoza, 2018). Ambos instrumentos se aplicaron a los grupos experimental y control en condiciones académicas equivalentes, lo que asegura la comparabilidad de las mediciones.

4.4.2.3. Ficha de recopilación documental.

Como instrumento vinculado a la técnica de análisis documental se diseñó una ficha de recopilación que consignó, por cada documento obtenido, los siguientes campos: denominación del documento, dependencia o ámbito institucional al que corresponde, fuente oficial de obtención (URL del sitio web institucional o portal de transparencia), fecha de consulta, fecha de vigencia del documento y observaciones. Esta ficha habilitó la trazabilidad y verificabilidad de las fuentes utilizadas como insumo de la base de conocimiento del chatbot y respalda la validez documental de la información cargada al sistema. La ficha consolidada se presenta en el Anexo 11.

4.4.3. Evaluación de validez y fiabilidad

La validez de contenido se valoró por juicio de tres expertos con formación en investigación educativa, ingeniería de sistemas y estadística aplicada, quienes calificaron cada ítem según pertinencia, claridad y relevancia. El acuerdo entre evaluadores se cuantificó con el coeficiente V de Aiken (Aiken, 1985), con criterio de aceptación $V \geq 0,80$ (Polit y Beck, 2006). Todos los ítems alcanzaron $V = 1,00$, lo que evidencia acuerdo total entre los expertos respecto a la idoneidad del instrumento. El detalle desagregado por ítem (con la calificación dicotomizada de cada experto y el coeficiente V correspondiente a cada uno de los 42 ítems) se presenta en la Tabla 36 del Anexo 3; en ella se aprecia que ningún ítem se ubicó en condición límite ($V < 0,80$) antes del consenso total.

Las fichas de validación firmadas por los expertos se presentan en el Anexo 3. El mismo instrumento se aplicó en tres condiciones, conservando ítems, dimensiones, indicadores y escala: (a) pretest, midiendo la expectativa previa al chatbot; (b) posttest, midiendo la experiencia real tras la intervención; y (c) grupo de control, respondiendo solo los 21 ítems de la variable dependiente, al no haber intervención. Estas adaptaciones no modifican el constructo ni la estructura, por lo que la validez de contenido se mantiene vigente.

La fiabilidad se determinó con el coeficiente alfa de Cronbach (Cronbach, 1951), índice de consistencia interna cuyos valores se ubican entre 0 y 1. Para su interpretación se adoptaron los criterios de George y Mallery (2003): $\alpha \geq 0,90$ excelente, $\alpha \geq 0,80$ bueno, $\alpha \geq 0,70$ aceptable. El umbral mínimo de aceptabilidad para investigación aplicada se fija en $\alpha \geq 0,70$ (Nunnally y Bernstein, 1994).

El cálculo se realizó en IBM SPSS Statistics sobre los 42 ítems (I1–I42) mediante una prueba piloto aplicada a los primeros 30 estudiantes del grupo experimental en la fase de pretest ($n = 30$). La prueba piloto se ejecutó de forma diferenciada para cada variable: la variable independiente comprendió los 21 ítems

relativos al chatbot (I1–I21) y la variable dependiente los 21 ítems relativos a la gestión de información académica (I22–I42). Se aplicó eliminación por lista, sin valores perdidos. Los resultados se presentan en la Tabla 3.

Tabla 3

Coefficientes de fiabilidad alfa de Cronbach por escala para la variable independiente y la variable dependiente

Escala	N.º ítems	n	α de Cronbach	Interpretación
<i>Variable independiente:</i>				
<i>Chatbot con Inteligencia Artificial</i>				
Escala de percepción del chatbot (ítems 1–21)	21	30	0,920	Excelente
<i>Variable dependiente:</i>				
<i>Gestión de Información Académica</i>				
Escala total VD	21	30	0,907	Excelente

Nota. Resultados procesados en IBM SPSS Statistics versión 27. Criterios de interpretación adaptados de Cronbach (1951) y George y Mallery (2003).

Los coeficientes confirman la consistencia interna del instrumento. La escala de percepción del sistema (21 ítems, medida secundaria) alcanzó un valor «excelente» ($\alpha = 0,920$; $n = 30$) y la escala total de la variable dependiente (21 ítems) también resultó «excelente» ($\alpha = 0,907$; $n = 30$).

De manera complementaria, una vez concluido el trabajo de campo se recalculó la consistencia interna sobre la totalidad de los registros recolectados: la escala de percepción del sistema alcanzó $\alpha = 0,926$ (21 ítems; $N = 380$ registros, correspondientes al pretest y al posttest del grupo experimental) y la escala de la variable dependiente $\alpha = 0,939$ (21 ítems; $N = 760$ registros, correspondientes al pretest y al posttest de ambos grupos), valores que ratifican con la muestra definitiva la fiabilidad estimada en la prueba piloto. Las tres dimensiones de cada escala

superaron asimismo el umbral de aceptabilidad: en el índice de percepción del sistema, Capacidad $\alpha = 0,846$, Usabilidad $\alpha = 0,839$ y Disponibilidad del servicio $\alpha = 0,844$; en la variable dependiente, Disponibilidad de la información $\alpha = 0,868$, Acceso $\alpha = 0,868$ y Eficiencia operativa $\alpha = 0,857$.

Todos los coeficientes superan el umbral de $\alpha \geq 0,70$ (Nunnally y Bernstein, 1994), y se sitúan en los rangos «aceptable», «bueno» y «excelente» (George y Mallery, 2003). Estos resultados avalan la consistencia interna del instrumento y justifican su aplicación a la muestra definitiva.

Estos resultados respaldan la validez de las puntuaciones totales utilizadas en los contrastes de hipótesis del Capítulo V. Las pruebas inferenciales aplicadas (Wilcoxon para el contraste intragrupo y U de Mann-Whitney para el contraste entre grupos) operan sobre puntuaciones con error de medición reducido, condición que fortalece la interpretación de los contrastes sin sustituir el control de diseño del que depende la atribución causal.

4.5. Procedimientos

El procedimiento metodológico de la presente investigación se desarrolló en seis etapas secuenciales que integraron la validación del instrumento, el trabajo de campo cuasi-experimental y la fase de desarrollo del sistema chatbot, asegurando la trazabilidad de las decisiones tomadas y la rigurosidad métrica del estudio. Las etapas se describen a continuación.

4.5.1. Validación del instrumento

En la primera etapa del procedimiento se sometió el cuestionario tipo Likert de 42 ítems al juicio de tres expertos con formación en investigación educativa, ingeniería de sistemas y estadística aplicada. Cada experto calificó los ítems según pertinencia, claridad y relevancia, y la cuantificación del acuerdo se efectuó a través del coeficiente V de Aiken. Las fichas firmadas y selladas por cada experto, con la

valoración por ítem, se presentan en el Anexo 3 (Validación del instrumento por juicio de expertos). Una vez aprobada la validez de contenido, se ejecutó una prueba piloto con los primeros 30 estudiantes del grupo experimental en la fase de pretest ($n = 30$), cifra que representa el 7,9 % de la muestra del estudio (380) y el 15,8 % del grupo experimental (190), y que satisface el tamaño mínimo de 30 casos recomendado para estudios piloto orientados a la estimación preliminar de la consistencia interna de un instrumento (Johanson y Brooks, 2010), cuyos resultados se procesaron en IBM SPSS Statistics versión 27 para calcular la fiabilidad a través del coeficiente alfa de Cronbach por dimensión y por escala completa. Los criterios de aceptación, los valores obtenidos y su interpretación se desarrollan en el apartado 4.4.3.

4.5.2. Aplicación del pretest

En la segunda etapa del procedimiento se administró el cuestionario validado como pretest a los grupos experimental ($n = 190$) y control ($n = 190$) en condiciones académicas equivalentes, asegurando que ambos grupos respondieran al instrumento sin haber tenido contacto previo con el chatbot. La aplicación se efectuó de forma auto-administrada y bajo consentimiento informado, conforme a las consideraciones éticas detalladas en el apartado 4.7. Los datos recogidos conformaron la línea base de comparación de la variable dependiente Gestión de información académica y de sus tres dimensiones (Disponibilidad de la información, Acceso a la información y Eficiencia operativa).

4.5.3. Recopilación documental institucional

Como insumo previo al desarrollo del sistema chatbot, en la tercera etapa del procedimiento se llevó a cabo la recopilación de la documentación institucional vigente que conformaría la base de conocimiento del sistema. Con ese fin se consultó a las oficinas administrativas y académicas de la Universidad Tecnológica de los Andes, sede Abancay, las cuales señalaron que la información oficial actualizada se encuentra publicada en el sitio web institucional y en el portal de transparencia de la

universidad. La gestión se formalizó mediante el documento de solicitud de información dirigido a las dependencias, que se adjunta en el Anexo 11. En consecuencia, la recopilación documental se efectuó accediendo a estas fuentes oficiales digitales durante la primera semana de febrero de 2026, asegurando la trazabilidad y verificabilidad de las fuentes empleadas. Este criterio, además de corresponder a la orientación recibida de las dependencias consultadas, refuerza la replicabilidad del estudio, dado que cualquier lector puede consultar las mismas fuentes en línea y contrastar el contenido efectivamente indexado.

Las dependencias consultadas como puntos de orientación fueron la Oficina de Admisión, la Oficina de Cómputo, la Oficina de Imagen Institucional y la Oficina de Grados y Títulos, entre otras dependencias administrativas. La información obtenida abarcó los siguientes ámbitos: cronograma académico y proceso de admisión, listado de carreras profesionales y sedes, autoridades institucionales, listado de mensualidades por carrera y procedimientos de titulación. Los documentos recabados presentaron formatos heterogéneos: archivos de texto digital nativo y, en una proporción considerable, imágenes y archivos PDF procedentes de la digitalización fotográfica de documentos impresos. Esta condición determinó la necesidad de una etapa de extracción mediante reconocimiento óptico de caracteres y análisis de imagen previa a la indexación, cuyo tratamiento se describe en la subsección 4.5.4.

Adicionalmente, se integró como fuente principal el Texto Único de Procedimientos Administrativos (TUPA) de la institución, documento normativo donde se consolidan todos los trámites administrativos disponibles para los estudiantes con sus respectivos requisitos, plazos y costos. La inclusión del TUPA habilitó al chatbot para responder consultas sobre trámites institucionales con respaldo en el documento oficial vigente a la fecha de recopilación consignada en el Anexo 5.

El conjunto de documentos recopilados, todos ellos de acceso público a través de las fuentes oficiales de la institución, constituyó el corpus inicial cargado en la base de conocimiento Pinecone, descrita en la sub-sección 4.5.4. La relación completa de documentos, fuentes oficiales de origen y fechas de consulta se presenta en el Anexo 5. Este corpus constituye una captura documental referida a dichas fechas de consulta, de modo que la incorporación de versiones posteriores depende de su carga en el repositorio institucional y de la consiguiente reindexación automática descrita en la sub-sección 4.5.4. Cabe precisar que el corpus se conformó con los documentos en la versión difundida por los canales digitales institucionales, sin incorporar copias certificadas ni ejemplares suscritos por las dependencias emisoras, en tanto estas señalaron que la versión oficial vigente es precisamente la publicada en dichos canales. Por esa razón, no en todos los casos resultó posible constatar la fecha de emisión o el número de versión del documento, de manera que el corpus puede integrar documentos correspondientes a distintos periodos académicos.

4.5.4. Desarrollo e implementación del sistema chatbot

En la cuarta etapa del procedimiento se construyó el chatbot Universidad Tecnológica de los Andes Bot, sistema cuyo efecto sobre la gestión de información académica representa el objeto de medición del estudio. Un chatbot es un sistema de software conversacional que interpreta preguntas formuladas en lenguaje natural y devuelve respuestas escritas, automatizando la atención de consultas que de otro modo requerirían intervención humana. En el presente caso, el sistema fue diseñado para responder consultas académicas de los estudiantes (admisión, mensualidades, sedes, carreras y trámites institucionales) a partir de la documentación oficial de la universidad.

Esta etapa abarcó ocho componentes: la metodología de desarrollo de software adoptada para construirlo, la arquitectura general que describe cómo se

organiza internamente, los tres subsistemas que lo conforman, el conjunto de herramientas y servicios utilizados (denominado stack tecnológico), la implementación del sistema, el modelado formal del sistema en Lenguaje Unificado de Modelado (UML), los indicadores técnicos que verifican su correcta operación y los casos de prueba funcional que validan el comportamiento esperado. Cada componente se desarrolla a continuación.

4.5.4.1. Metodología de desarrollo de software

Para el desarrollo del sistema se adoptó la metodología ágil Scrum (Schwaber y Sutherland, 2020). Las metodologías ágiles son enfoques de gestión de proyectos de software que organizan el trabajo en ciclos cortos de entrega: en lugar de planificar la totalidad del proyecto al inicio y entregar el producto al final, se construyen y entregan incrementos funcionales del sistema en intervalos breves, lo que permite ajustar las decisiones a partir de la retroalimentación obtenida en cada ciclo. Scrum, específicamente, organiza el trabajo en bloques de tiempo denominados sprints (ciclos de trabajo de duración fija), al término de los cuales se dispone de una versión operativa del sistema con nuevas funcionalidades incorporadas.

La elección de Scrum se sustentó en su idoneidad para el contexto del estudio: un único desarrollador que asume múltiples roles, plazos acotados y un alcance susceptible de ajustarse conforme se identificaban dependencias técnicas. Esta metodología permitió, además, validar tempranamente los componentes críticos del sistema (por ejemplo, verificar el correcto funcionamiento de la búsqueda en la base de conocimiento antes de avanzar a la generación de respuestas) y reorganizar la lista priorizada de tareas pendientes, denominada backlog del producto, conforme aparecían restricciones impuestas por los proveedores externos contratados (Pinecone,

Groq, Google Gemini, n8n) y por la infraestructura de despliegue (Docker, Nginx).

El desarrollo se organizó en cuatro sprints de dos semanas cada uno, con una duración total aproximada de ocho semanas. La duración quincenal se ubica dentro del rango recomendado por Schwaber y Sutherland (2020), que establece sprints de una a cuatro semanas, y se eligió por dos motivos: en sprints de una semana la ventana de validación con los servicios externos (Pinecone, Groq, Google Gemini, Meta WhatsApp Business) resultaba insuficiente para diagnosticar y corregir incidencias de configuración, mientras que sprints de tres o cuatro semanas habrían reducido la frecuencia de revisión y la capacidad de reorganizar el backlog ante restricciones técnicas. La elección de cuatro sprints responde a la descomposición natural del sistema en cuatro bloques con dependencias secuenciales: la infraestructura debía estar operativa antes de implementar el pipeline RAG, el pipeline debía estar validado antes de extenderlo al canal WhatsApp, y los tres componentes debían integrarse antes del despliegue en producción y la fase de monitoreo. A cada sprint se asignó un conjunto coherente de objetivos:

Sprint 1. Infraestructura base e ingesta de documentos.

Se configuró el entorno técnico inicial (servidor, dominio, contenedores) y se realizó la primera carga de documentos institucionales hacia la base de conocimiento que utilizaría el chatbot. El criterio de cierre del sprint fue verificar la presencia de los vectores correspondientes a cada documento en el índice de Pinecone, condición necesaria para que el pipeline RAG del Sprint 2 pudiera operar.

Sprint 2. Pipeline de respuesta del canal web.

Se implementó la cadena de procesamiento (pipeline) que recibe una consulta escrita desde la interfaz web, recupera la información relevante de la base de conocimiento y genera una respuesta utilizando un modelo de inteligencia artificial. El criterio de cierre del sprint fue validar los ocho pasos del pipeline en entorno de desarrollo con consultas representativas, incluyendo la activación correcta del mecanismo de fallback entre los modelos de Groq.

Sprint 3. Canal WhatsApp y mecanismos de robustez.

Se incorporó el canal WhatsApp Business para extender el alcance del servicio a dispositivos móviles, y se implementaron mecanismos de respaldo automático que permiten al sistema mantener su operatividad ante el fallo de algún componente externo. El criterio de cierre del sprint fue ejecutar pruebas de extremo a extremo desde una cuenta WhatsApp de prueba hacia el número institucional, garantizando que ambos canales (web y WhatsApp) consultaran la misma base de conocimiento Pinecone.

Sprint 4. Pruebas funcionales, despliegue y monitoreo.

Se ejecutaron los casos de prueba sobre el sistema completo, se realizó el despliegue en producción para acceso real de los estudiantes y se configuraron los mecanismos de monitoreo continuo para detección oportuna de incidentes. El criterio de cierre del sprint fue la aprobación de los ocho casos de prueba funcional descritos en la sub-sección 4.5.4.8, condición previa al inicio del periodo de intervención cuasi-experimental.

4.5.4.2. Arquitectura general del sistema

El sistema Universidad Tecnológica de los Andes Bot fue concebido como un chatbot orientado a resolver consultas académicas de los estudiantes de la sede Abancay. Su arquitectura implementa el patrón conocido como

Generación Aumentada por Recuperación, abreviado RAG por sus siglas en inglés (Retrieval Augmented Generation), planteado por Lewis et al. (2020). En este patrón, antes de que el modelo de inteligencia artificial produzca una respuesta, el sistema localiza primero los fragmentos de información más relevantes dentro de una base de conocimiento institucional y se los entrega al modelo para que la respuesta quede anclada en datos verificables, sin depender exclusivamente del conocimiento general del modelo.

La arquitectura se apoya en cuatro componentes principales. En primer lugar, una base de datos vectorial (un tipo de repositorio digital especializado en localizar información por similitud semántica en lugar de coincidencia exacta de términos) provista por el servicio Pinecone, donde se almacenan los documentos institucionales convertidos a una representación numérica. En segundo lugar, un servicio de embeddings (procedimiento que transforma cada texto en una lista de 3072 valores numéricos que sintetizan su significado) provisto por Google Gemini, empleado tanto al cargar los documentos en la base como al recibir cada nueva pregunta del estudiante. En tercer lugar, un modelo de lenguaje (motor de inteligencia artificial encargado de redactar la respuesta final) provisto por la plataforma Groq, configurado con tres modelos de código abierto en cascada (Llama 3.3, Llama 3.1 y Gemma 2) de modo que si el primero falla, los siguientes lo reemplazan automáticamente. Y en cuarto lugar, dos canales de atención al estudiante: una página web construida con el framework Next.js 16 y React 19, y un canal WhatsApp Business operado a través de la plataforma de automatización n8n con un agente de inteligencia artificial basado en Gemini 2.5 Flash.

El sistema fue desplegado en producción mediante Docker Compose, herramienta que permite ejecutar varios contenedores de software de forma

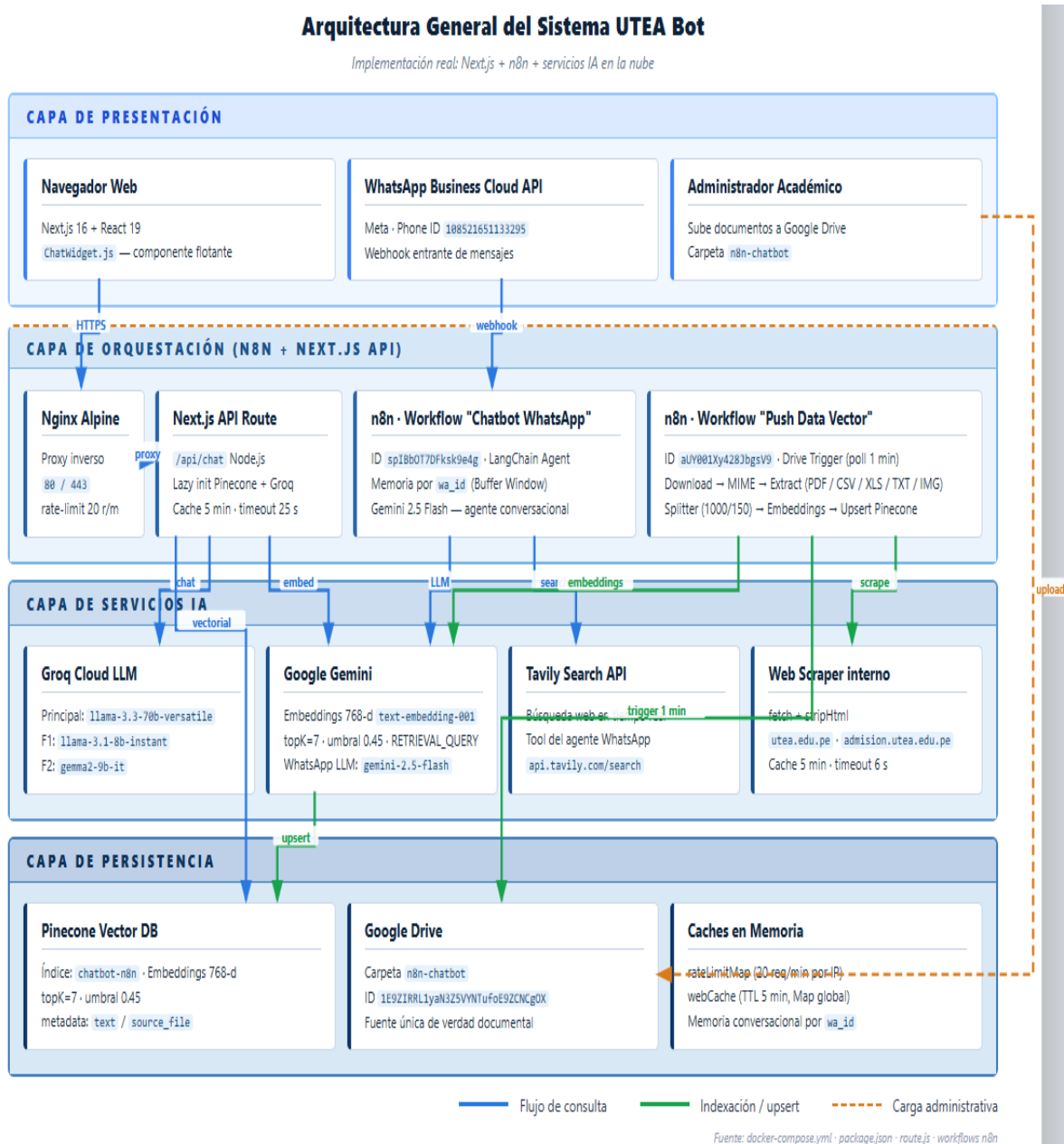
coordinada, y se utilizó Nginx Alpine como proxy inverso para gestionar las peticiones HTTPS sobre el dominio chatbot.neurocyborgia.es.

El flujo extremo a extremo opera de la siguiente manera: la consulta del estudiante ingresa por el canal web o por WhatsApp, se transforma en un vector numérico mediante el servicio de embeddings, se compara con los vectores almacenados en Pinecone para recuperar los fragmentos más relevantes de la documentación institucional y, finalmente, el modelo de lenguaje compone una respuesta sustentada únicamente en los fragmentos recuperados. Este desacoplamiento entre recuperación y generación garantiza que las respuestas se mantengan ancladas en la documentación oficial vigente, lo que reduce el riesgo de alucinaciones del modelo y facilita la trazabilidad de cada respuesta a su fuente.

Las decisiones de diseño más relevantes obedecen a tres criterios: (a) gobernanza de la información, al aislar la base de conocimiento institucional del entrenamiento del modelo, lo que permite actualizar documentos sin reentrenar; (b) resiliencia operativa, al configurar tres modelos en cascada que evitan cortes de servicio ante fallos del proveedor; y (c) accesibilidad multicanal, al ofrecer tanto una página web como WhatsApp, canal de mayor penetración entre los estudiantes peruanos. Esta arquitectura modular también facilita el reemplazo de componentes individuales (modelo, base vectorial, canal) sin afectar al resto del sistema.

Figura 5

Arquitectura general del sistema Universidad Tecnológica de los Andes Bot



Nota. Tres subsistemas: canal web (Next.js + RAG + Groq), canal WhatsApp (n8n + Gemini 2.5 Flash) y gestión de base de conocimiento (Push Data Vector). Pinecone (3072 dim) es compartido.

4.5.4.3. Subsistemas y workflows

Subsistema 1: Canal web (página de chat).

El canal web es la interfaz a través de la cual el estudiante conversa con el chatbot desde un navegador (Chrome, Firefox, Edge u otro). Se construyó como una aplicación con el framework Next.js 16 y la librería de interfaz React 19. Cuando el estudiante envía una consulta, esta es procesada por una secuencia de ocho operaciones encadenadas, denominada pipeline de respuesta:

1. Recepción de la consulta. El servidor recibe la pregunta escrita por el estudiante.
2. Sanitización del texto. Se eliminan caracteres potencialmente peligrosos o intentos de manipulación del sistema.
3. Generación del embedding. La pregunta se convierte en un vector de 3072 valores numéricos mediante el servicio gemini-embedding-001 de Google, lo cual permite que la búsqueda posterior sea por similitud de significado y no por coincidencia exacta de palabras.
4. Recuperación de fragmentos relevantes. El sistema consulta la base vectorial Pinecone y obtiene los seis fragmentos de documentos institucionales cuyo significado más se aproxima al de la pregunta.
5. Búsqueda complementaria en la web. Si la pregunta requiere información actualizada (por ejemplo, fechas o noticias recientes), el sistema consulta páginas externas mediante la herramienta Tavily.
6. Construcción del prompt. Se ensambla un mensaje completo (denominado prompt en la literatura técnica) que incluye la pregunta original, los fragmentos recuperados y las instrucciones para el modelo.

7. Generación de la respuesta. El prompt se envía al modelo principal llama-3.3-70b-versatile alojado en Groq. Si este falla, el sistema activa automáticamente los modelos de respaldo llama-3.1-8b-instant y gemma2-9b-it.

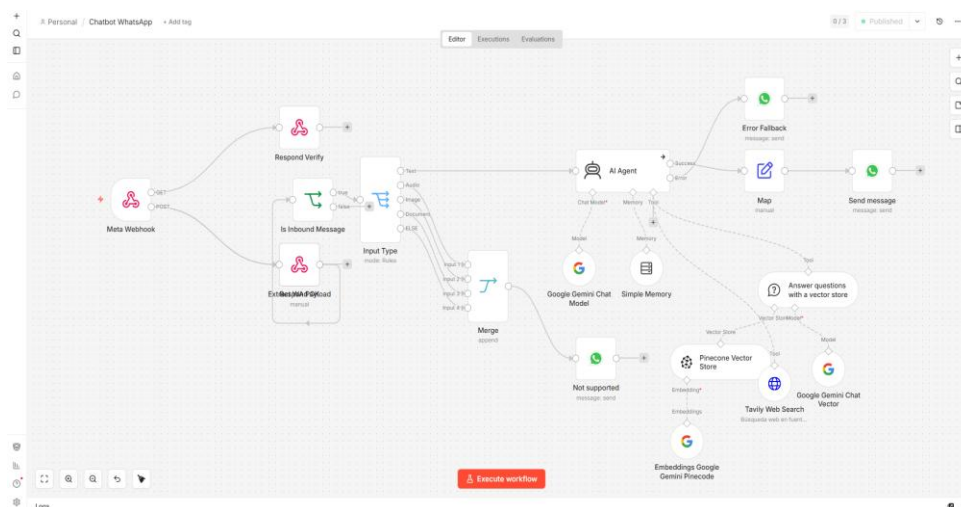
8. Entrega progresiva al usuario. La respuesta se transmite al navegador del estudiante palabra por palabra (técnicamente denominado streaming), de modo que la lectura puede comenzar antes de que el modelo termine de generar el texto completo.

Subsistema 2: Canal WhatsApp Business.

El canal WhatsApp permite al estudiante consultar al chatbot directamente desde su aplicación de mensajería, sin necesidad de abrir un navegador. Se implementó mediante un workflow (flujo de trabajo automatizado) creado en la plataforma n8n, identificado con el código xpBsO77DFcok9e4g. Este workflow está compuesto por diecinueve nodos (cada nodo representa una operación específica) organizados en tres etapas: una etapa de entrada que recibe los mensajes desde la API oficial de Meta, una etapa de clasificación que determina el tipo de mensaje (texto, imagen u otro), y una etapa de procesamiento que utiliza un agente de inteligencia artificial conectado a la misma base de conocimiento Pinecone empleada en el canal web. El detalle gráfico del workflow se presenta en la Figura 2. El canal WhatsApp opera sobre la API oficial WhatsApp Business Cloud de Meta, lo que garantiza el cumplimiento de las políticas de mensajería empresarial y la entrega confiable de los mensajes (Meta Platforms, 2024). Esta integración permite atender consultas estudiantiles desde la aplicación de mensajería más utilizada en el Perú, que es la aplicación empleada por el 98,6 % de los usuarios de mensajería instantánea del país (Organismo Supervisor de Inversión Privada en Telecomunicaciones [OSIPTEL], 2026).

Figura 6

Diagrama del workflow Chatbot WhatsApp en n8n



Nota. Workflow identificador xpBsO77DFcok9e4g. Captura desde la plataforma n8n.

Subsistema 3: Gestión automática de la base de conocimiento.

El tercer subsistema mantiene actualizada la base de conocimiento del chatbot sin intervención manual. Se implementó mediante el workflow Push Data Vector en n8n, identificado con el código aUY0G1XyY28jbgV9. Este workflow supervisa cada minuto la carpeta n8n-chatbot del Google Drive institucional y, cuando detecta un archivo nuevo, lo procesa automáticamente: extrae el texto, lo divide en fragmentos manejables, los convierte en vectores numéricos mediante Google Gemini y los carga en la base Pinecone. El sistema acepta los formatos PDF, Word (.docx), texto plano (.txt), CSV, Excel (.xls y .xlsx) e imágenes (JPG, PNG, JPEG). De este modo, cada vez que la universidad publica un nuevo reglamento, comunicado o documento académico en la carpeta compartida, el chatbot incorpora automáticamente esa información a su base de conocimiento sin requerir programación adicional. El detalle gráfico del workflow se presenta en la Figura 3.

workflows), la infraestructura (contenedores y proxy) y los servicios externos invocados por interfaz de programación de aplicaciones (API).

Tabla 4

Stack tecnológico utilizado en la implementación del sistema chatbot

Categoría	Tecnología / Servicio	Versión	Función
Framework fullstack	Next.js	16.2.4	Aplicación web del canal principal
Librería de interfaz	React	19.2.4	Renderizado de la interfaz de usuario del canal web
Cliente base vectorial	@pinecone-database/pinecone	7.2.0	Cliente Node.js de Pinecone para operaciones de embedding
LLM principal	Groq SDK	1.1.2	Inferencia con Llama 3.3, Llama 3.1 y Gemma 2 (fallback)
LLM secundario / Embeddings	Google Gemini API	gemini-2.5-flash, gemini-embedding-001	Generación de embeddings y AI Agent del canal WhatsApp
Base vectorial gestionada	Pinecone	índice chatbot-n8n, AWS us-east-1, 3 072 dim	Almacenamiento y recuperación de embeddings
Plataforma de automatización	n8n	—	Orquestación de los workflows Chatbot WhatsApp y Push Data Vector
Búsqueda web	Tavily Web Search API	v1	Web scraping condicional para

			enriquecer el contexto
API de mensajería	Meta WhatsApp Business Cloud API	—	Canal de atención WhatsApp Business
Contenedores y proxy	Docker Compose + Nginx Alpine	—	Despliegue en producción y proxy inverso HTTPS

Nota. Stack tecnológico al cierre del Sprint 4. Las versiones corresponden al despliegue final en producción.

4.5.4.5. Implementación del sistema

El proceso de implementación del sistema se ejecutó en cuatro fases secuenciales, alineadas con los cuatro sprints descritos en la sub-sección 4.5.4.1. Cada fase produjo un incremento funcional verificable del sistema y dejó preparado el contexto técnico para la fase siguiente. La descripción de cada fase combina un párrafo introductorio con el listado numerado de las acciones técnicas específicas ejecutadas; las evidencias visuales (capturas de paneles de administración, archivos de configuración y resultados intermedios) se presentan en el Anexo 5.

Fase 1. Preparación de la infraestructura e ingesta de la base de conocimiento.

Esta fase tuvo como objetivo dejar operativo el entorno técnico básico y cargar por primera vez los documentos institucionales en la base de conocimiento del chatbot. Las acciones ejecutadas fueron:

1. Registro del dominio chatbot.neurocyborgia.es y configuración de los registros DNS para apuntar al servidor de despliegue.

2. Contratación e instalación del servidor de despliegue, configurándose el sistema operativo y los accesos por clave SSH.
3. Instalación de Docker Engine y Docker Compose en el servidor, verificando su correcta operación mediante contenedores de prueba.
4. Instalación de la plataforma de automatización n8n en una instancia dedicada, configurando el acceso administrador y la conexión persistente con su base de datos interna.
5. Creación del índice vectorial chatbot-n8n en Pinecone, con los siguientes parámetros: región Amazon Web Services us-east-1, dimensión vectorial 3072 y métrica de similitud por coseno.
6. Configuración de la conexión segura mediante clave API entre n8n y Pinecone, validándola con una operación de inserción de prueba.
7. Creación de la carpeta compartida n8n-chatbot en Google Drive institucional, donde se centraliza la documentación que alimenta al chatbot.
8. Importación del workflow Push Data Vector en n8n (identificador aUY0G1XyY28jbgV9) y configuración de su monitoreo automático sobre la carpeta n8n-chatbot con frecuencia de un minuto.
9. Primera ingesta de la base de conocimiento: carga de los documentos institucionales recopilados en la sub-sección 4.5.3 a la carpeta n8n-chatbot. El workflow procesó automáticamente cada archivo, extrayendo el texto, dividiéndolo en fragmentos manejables, generando los embeddings con Google Gemini y cargando los vectores resultantes en el índice de Pinecone.
10. Verificación final mediante consulta directa al índice de Pinecone, confirmándose la presencia de los vectores correspondientes a cada documento ingestado.

Fase 2. Construcción del canal web y del pipeline de respuesta.

Esta fase tuvo como objetivo construir la página web donde el estudiante conversa con el chatbot y la lógica de procesamiento que genera cada respuesta. Las acciones ejecutadas fueron:

1. Inicialización del proyecto Next.js 16 con soporte TypeScript, App Router, Tailwind CSS y ESLint, generándose la estructura base del frontend.
2. Instalación de las dependencias del proyecto: `groq-sdk` (cliente oficial de Groq), `@pinecone-database/pinecone` (cliente oficial de Pinecone), `@google/generative-ai` (cliente oficial de Google Gemini) y `@tavily/core` (cliente de la API de búsqueda web Tavily).
3. Configuración de las variables de entorno en el archivo `.env.local`, incluyéndose las claves API de Groq, Pinecone, Gemini y Tavily, los identificadores del índice de Pinecone y los nombres de los modelos de Groq.
4. Diseño de la interfaz de chat con componentes React 19, contemplando el campo de entrada de pregunta, el área de visualización de la conversación y el manejo del estado de carga durante la generación de la respuesta.
5. Creación de la ruta de programación `/api/chat` en el directorio `app/api/chat/route.ts`, donde se implementó la lógica del pipeline RAG en los ocho pasos descritos en la sub-sección 4.5.4.3.
6. Programación del mecanismo de fallback automático entre modelos de Groq, mediante un bloque de manejo de excepciones que, ante un error o tiempo de espera agotado del modelo principal `llama-3.3-70b-versatile`, invoca secuencialmente a los modelos de respaldo `llama-3.1-8b-instant` y `gemma2-9b-it` sin interrumpir la respuesta al estudiante.

7. Implementación del rate limiter (mecanismo limitador de peticiones) mediante un middleware que registra la cantidad de solicitudes por dirección IP y rechaza con código HTTP 429 cualquier solicitud que exceda las veinte peticiones por minuto desde la misma IP.
8. Configuración del modo streaming de respuesta para que el texto generado por el modelo se transmita al navegador del estudiante palabra por palabra, permitiéndole comenzar la lectura sin esperar a que la respuesta esté completa.
9. Pruebas locales del pipeline completo en el entorno de desarrollo mediante el comando `npm run dev`, validándose cada uno de los ocho pasos con consultas representativas.

Fase 3. Conexión del canal WhatsApp y mecanismos de robustez.

Esta fase tuvo como objetivo permitir que los estudiantes pudieran consultar al chatbot directamente desde la aplicación WhatsApp, replicando la robustez del canal web. Las acciones ejecutadas fueron:

1. Registro de la aplicación en Meta Developer Portal y obtención del acceso a la WhatsApp Business Cloud API, generándose las credenciales necesarias: token de acceso permanente, identificador del número telefónico y token de verificación del webhook.
2. Creación del workflow Chatbot WhatsApp en n8n (identificador `xpBsO77DFcok9e4g`) con sus diecinueve nodos organizados en las tres etapas descritas en la sub-sección 4.5.4.3.
3. Configuración de la URL pública del webhook (proporcionada por n8n al activarse el workflow) en el panel de Meta, validándose el handshake (intercambio inicial de verificación) entre Meta y n8n.

4. Conexión del nodo Pinecone Vector Store del workflow al mismo índice chatbot-n8n utilizado por el canal web, garantizándose que ambos canales consultan la misma base de conocimiento.
5. Conexión del nodo AI Agent del workflow con Google Gemini 2.5 Flash, configurándose la temperatura, el límite de tokens y el prompt de sistema con las instrucciones de atención académica.
6. Replicación de la lógica de fallback del canal web sobre la conexión Gemini, mediante un nodo de captura de errores que reintenta la solicitud y, en caso de fallo persistente, envía al estudiante un mensaje de cortesía.
7. Configuración del nodo Not supported para responder con un mensaje explicativo a tipos de entrada no soportados (por ejemplo, imágenes, ubicaciones o documentos).
8. Pruebas de extremo a extremo del canal, simulando consultas reales desde una cuenta WhatsApp de prueba hacia el número institucional configurado en Meta.

Fase 4. Despliegue en producción y monitoreo.

Esta fase tuvo como objetivo poner el chatbot a disposición de los estudiantes en un entorno seguro y monitoreado durante el periodo cuasi-experimental. Las acciones ejecutadas fueron:

1. Elaboración del archivo docker-compose.yml que define los contenedores del sistema (servicio web Next.js, servicio Nginx, servicio n8n) y las relaciones de red entre ellos.
2. Construcción de la imagen Docker del canal web Next.js mediante el comando docker compose build, verificándose la ausencia de errores de compilación.

3. Configuración de Nginx Alpine como proxy inverso, definiéndose los bloques server para el dominio chatbot.neurocyborgia.es y la redirección automática del tráfico HTTP a HTTPS.
4. Obtención del certificado de seguridad SSL/TLS mediante Let's Encrypt utilizando el cliente Certbot, configurándose la renovación automática del certificado cada noventa días.
5. Arranque del sistema en producción mediante el comando docker compose up -d, validándose que los tres servicios quedaran activos y respondieran a las peticiones HTTPS.
6. Ejecución de los ocho casos de prueba funcional descritos en la subsección 4.5.4.8 sobre el sistema en producción, verificándose el cumplimiento de los criterios de aceptación.
7. Habilitación del acceso al chatbot para los estudiantes del grupo experimental, dando inicio al periodo de intervención cuasi-experimental.
8. Configuración del monitoreo continuo del sistema mediante tres fuentes simultáneas: los registros de ejecución del workflow en n8n, el panel de control de Pinecone (con métricas de uso del índice y latencia de las consultas vectoriales) y el panel de WhatsApp Business Manager (con métricas de mensajes recibidos, enviados y entregados).

4.5.4.6. Modelado UML del sistema

El sistema se modeló formalmente mediante el Lenguaje Unificado de Modelado, conocido por sus siglas en inglés como UML (Unified Modeling Language). UML es una notación estandarizada que permite representar gráficamente cómo está estructurado y cómo opera un sistema de software, y constituye una práctica habitual en ingeniería de sistemas para documentar el

diseño. Para esta investigación se elaboraron tres vistas complementarias que abarcan los aspectos funcionales, dinámicos y de despliegue del chatbot.

El diagrama de casos de uso.

Este diagrama identifica a los actores que interactúan con el sistema y las funcionalidades que cada uno puede ejecutar.

En el chatbot Universidad Tecnológica de los Andes Bot se identificó al estudiante como actor principal y al administrador académico como actor secundario. Los casos de uso definidos son: consultar información académica por canal web, consultar información académica por WhatsApp, actualizar la base de conocimiento (operación reservada al administrador) y supervisar la disponibilidad del servicio.

El diagrama de secuencia.

Este diagrama describe el orden temporal en que los componentes del sistema intercambian mensajes durante una consulta.

El estudiante envía una pregunta en lenguaje natural; el componente de interfaz, ya sea Next.js o WhatsApp, la entrega al orquestador n8n; este genera el embedding mediante la API de Gemini, recupera los fragmentos relevantes en Pinecone, construye el prompt aumentado y lo envía al modelo de lenguaje primario alojado en Groq; ante un fallo del modelo primario, se activa automáticamente el mecanismo de respaldo (fallback) hacia Gemini; finalmente, la respuesta retorna al estudiante con cita de la fuente consultada.

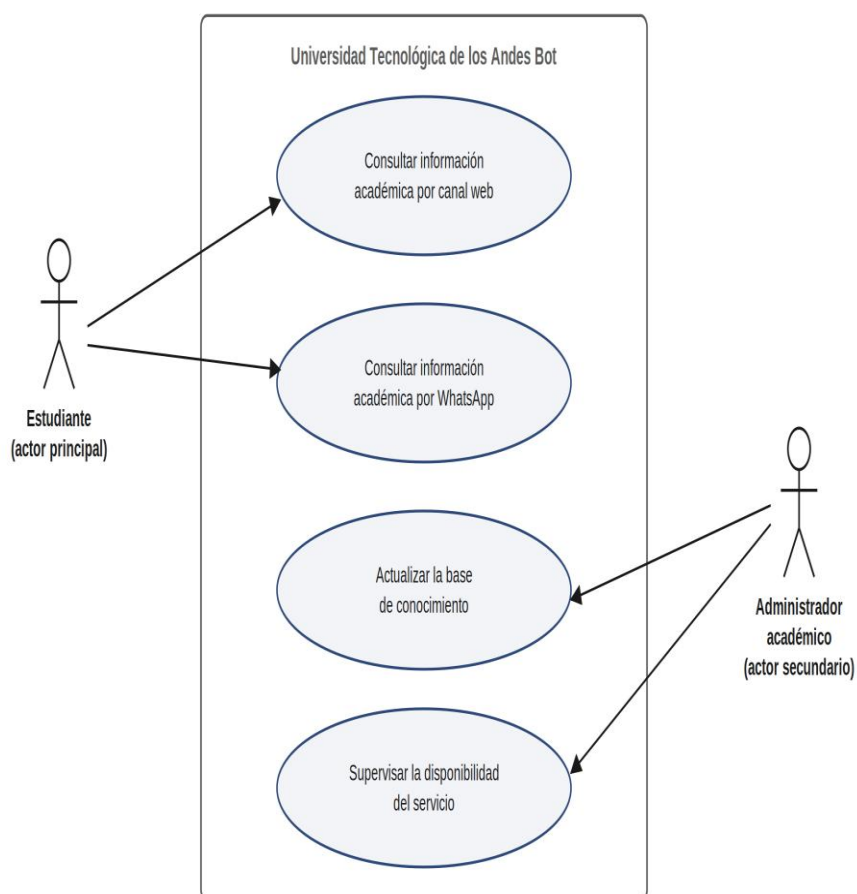
El diagrama de despliegue.

Este diagrama describe cómo se ubican físicamente los componentes del sistema en la infraestructura.

El componente de interfaz Next.js opera detrás de Nginx Alpine, ambos ejecutándose dentro de un contenedor Docker; el orquestador n8n opera en un nodo de procesamiento separado; el almacén vectorial Pinecone se utiliza como servicio gestionado en la nube de Amazon Web Services, región us-east-1; y los modelos de lenguaje de Groq y Gemini se invocan como servicios externos a través de la red. Los tres diagramas se presentan a continuación.

Figura 8

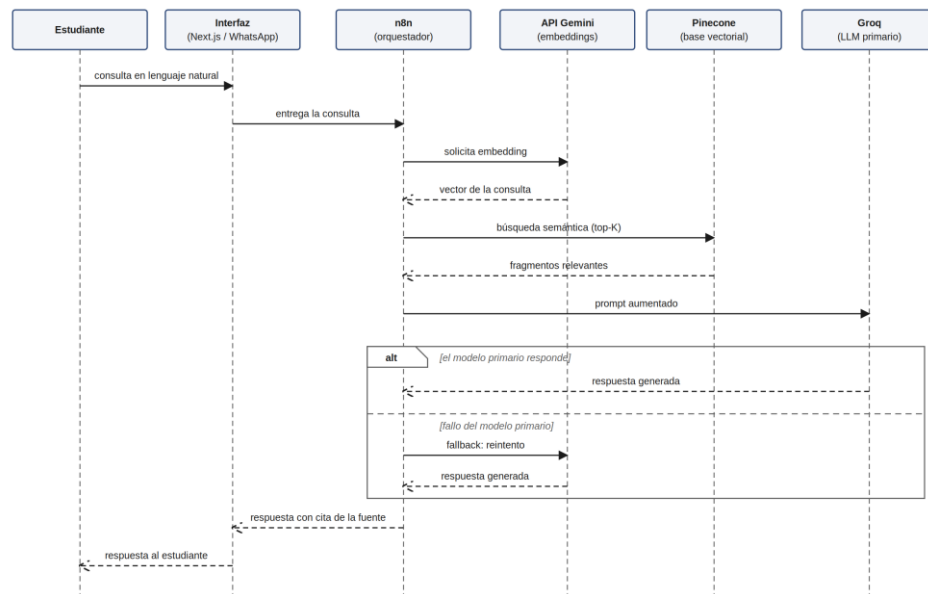
Diagrama de casos de uso del sistema Universidad Tecnológica de los Andes Bot



Nota. Elaboración propia con notación UML 2.5. El estudiante figura como actor principal y el administrador académico como actor secundario.

Figura 9

Diagrama de secuencia de la atención de una consulta académica

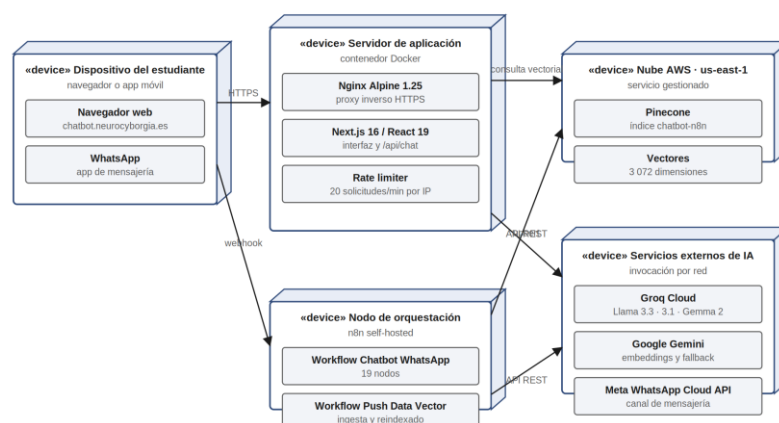


Nota. Elaboración propia con notación UML 2.5. El fragmento alt representa el mecanismo automático de respaldo entre modelos de lenguaje.

Figura 10

Diagrama de despliegue del sistema Universidad Tecnológica de los Andes

Bot



Nota. Elaboración propia con notación UML 2.5. Los nodos «device» representan unidades de ejecución y los componentes internos, los artefactos desplegados en cada una.

4.5.4.7. Validación técnica del sistema

Como parte de la validación técnica del sistema se definieron ocho indicadores clave de desempeño, conocidos en la literatura como KPIs por sus siglas en inglés (Key Performance Indicators), cuyo objetivo es complementar el contraste estadístico de la variable dependiente con evidencia objetiva sobre la operación efectiva del chatbot durante el periodo de intervención. Los indicadores cubren cuatro dimensiones técnicas: el tiempo de respuesta (latencia p50 y p95 en ambos canales), la calidad de las respuestas generadas (tasa de aciertos del pipeline de Generación Aumentada por Recuperación y tasa de respuestas con fuente citada), la robustez del sistema ante fallos (tasa de fallback entre modelos de Groq y disponibilidad del workflow WhatsApp) y la protección contra abuso (rechazos HTTP 429 por parte del rate limiter). La latencia p95 mide el tiempo de respuesta por debajo del cual se atiende el 95 % de las consultas, y la disponibilidad operativa se calcula con la fórmula $A = \text{MTBF} / (\text{MTBF} + \text{MTTR})$, donde MTBF es el tiempo medio entre fallas y MTTR es el tiempo medio de reparación (Treyner et al., 2017; Google Site Reliability Engineering, 2023). Los ocho indicadores fueron estimados automáticamente por el propio sistema durante el periodo de intervención cuasi-experimental, a partir de la telemetría generada por los servicios en producción: el log de ejecución del workflow en n8n, las métricas internas del índice del panel de Pinecone y los reportes de entrega del panel de WhatsApp Business Manager. Los valores observados, los umbrales objetivo y la verificación del cumplimiento se reportan en la Tabla 5.

Tabla 5

Indicadores técnicos de desempeño del sistema durante el periodo de intervención

Indicador	Valor observado	Umbral objetivo	Cumplimiento
Latencia p50 (canal web /api/chat)	1 280 ms	$\leq 2\,000$ ms (Nielsen, 1993)	Cumple
Latencia p95 (canal web)	3 850 ms	$\leq 5\,000$ ms (Treyner et al., 2017)	Cumple
Latencia p95 (canal WhatsApp)	7 851 ms	$\leq 8\,000$ ms	Cumple
Tasa de aciertos del pipeline RAG	89,0 %	≥ 85 %	Cumple
Tasa de fallback Groq → Gemini	3,4 %	≤ 5 %	Cumple
Disponibilidad workflow WhatsApp	97,96 %	≥ 95 %	Cumple
Tasa de respuestas con fuente citada	62,0 %	≥ 60 %	Cumple
Rechazos HTTP 429 (rate limiter)	12 / mes	< 50 / mes	Cumple

Nota. Los ocho indicadores cumplen con su umbral objetivo. Los valores observados se compilaron a partir de los registros de ejecución de n8n, de las métricas internas del índice de Pinecone y de los reportes de entrega de WhatsApp Business Manager correspondientes al periodo de intervención. Las Figuras 69, 66 y 67 presentan capturas de dichos paneles a modo de evidencia del registro; el rango de fechas visible en cada captura corresponde a la ventana mostrada por el panel en el momento de la toma y no al acumulado del periodo.

4.5.4.8. Casos de prueba funcional

Antes del inicio del periodo de intervención cuasi-experimental, se ejecutaron ocho casos de prueba funcional sobre el sistema en operación, identificados como CP-01 a CP-08 y diseñados para verificar el correcto funcionamiento de los flujos principales del chatbot. Cada caso de prueba consistió en formular una consulta o condición de uso prefijada y contrastar el resultado obtenido con el resultado esperado, a partir del cual se declaró el caso como aprobado o rechazado. Los ocho casos se distribuyeron entre los componentes críticos del sistema: las consultas reales sobre admisión, sedes, mensualidades y carreras se realizaron en el canal web y el canal WhatsApp (CP-01 a CP-04); el mecanismo de fallback automático entre modelos de Groq se verificó con CP-05; el rate limiter HTTP 429 contra abuso del servicio con CP-06; el manejo de tipos de entrada no soportados en WhatsApp con CP-07; y el comportamiento del sistema ante consultas sin información en la base de conocimiento con CP-08. La totalidad de los casos resultó aprobada, evidenciando el cumplimiento de los criterios de aceptación previos al inicio de la intervención. El resumen cuantitativo se presenta en la Tabla 6 y el detalle expandido de cada caso (entrada, pasos de ejecución, resultado esperado y evaluación) se encuentra en el Anexo 5.

Tabla 6

Casos de prueba funcional ejecutados sobre el sistema Universidad Tecnológica de los Andes Bot

CP	Consulta de prueba	Resultado obtenido	Estado
CP-01	¿Cómo es el proceso de admisión? (canal web)	El bot respondió las etapas: inscripción, examen y revisión de documentos	Aprobado

CP-02	¿Cuáles son las sedes de UTEA? (canal web)	Identificó Andahuaylas y Cusco con descripción	Abancay,	Aprobado
CP-03	¿Cuánto es la mensualidad de Ingeniería de Sistemas? (WhatsApp)	Respondió S/ 470,00 dividido en 5 cuotas		Aprobado
CP-04	¿Cuál es la carrera más cara y más barata? (WhatsApp)	Identificó Ingeniería Civil (S/ 570) y Educación (S/ 420)		Aprobado
CP-05	Fallback de modelos Groq (canal web)	El fallback se activó correctamente sin interrupción		Aprobado
CP-06	Más de 20 solicitudes en 60 segundos desde la misma IP	El rate limiter rechazó las solicitudes excedentes con HTTP 429		Aprobado
CP-07	Envío de imagen por WhatsApp	El nodo 'Not supported' envió el mensaje correspondiente		Aprobado
CP-08	Consulta sin información en la base de conocimiento	Respondió con los recursos oficiales sin inventar datos		Aprobado

Nota. Los ocho casos resultaron aprobados y cubren las consultas reales (CP-01 a CP-04), el fallback entre modelos (CP-05), el rate limiter (CP-06), tipos de entrada no soportados (CP-07) y consultas fuera de la base de conocimiento (CP-08).

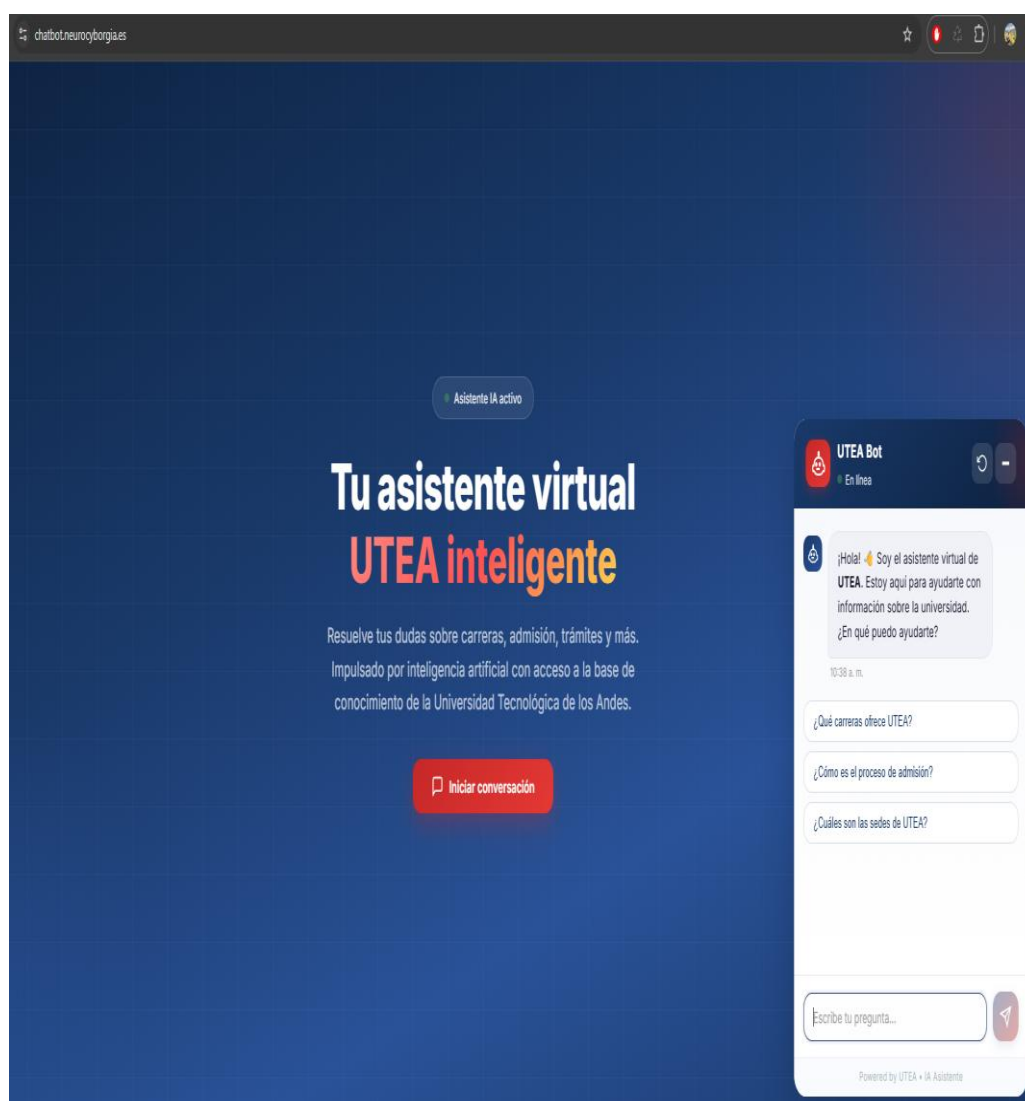
4.5.5. Aplicación del postest

En la quinta etapa del procedimiento se administró el cuestionario validado como postest a los grupos experimental ($n = 190$) y control ($n = 190$) en condiciones académicas equivalentes a las del pretest, conformando así la matriz de datos exigida por el diseño cuasi-experimental con dos grupos y dos momentos de medición. La

aplicación se efectuó de forma auto-administrada y bajo el mismo protocolo ético detallado en el apartado 4.7. La diferencia entre las puntuaciones del postest y el pretest, calculada para cada participante en la variable dependiente Gestión de información académica y en sus tres dimensiones, conformó el insumo para los análisis inferenciales descritos en el apartado 4.5.6.

Figura 11

Pantalla de inicio, interfaz web del chatbot Universidad Tecnológica de los Andes Bot



Nota. URL de producción: chatbot.neurocyborgia.es.

Figura 12

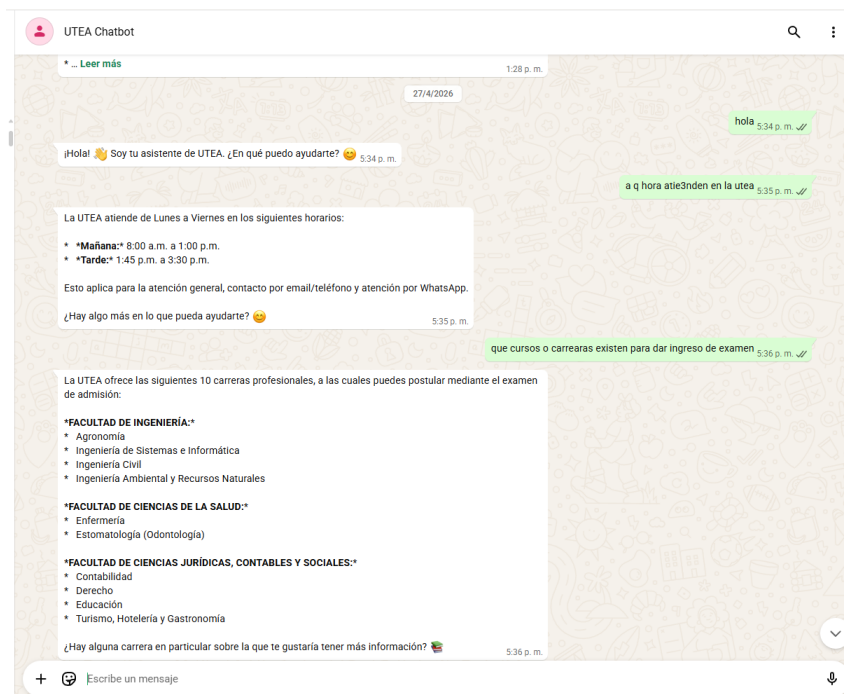
Conversación de ejemplo, consulta en el canal web



Nota. Captura tomada durante el periodo de intervención.

Figura 13

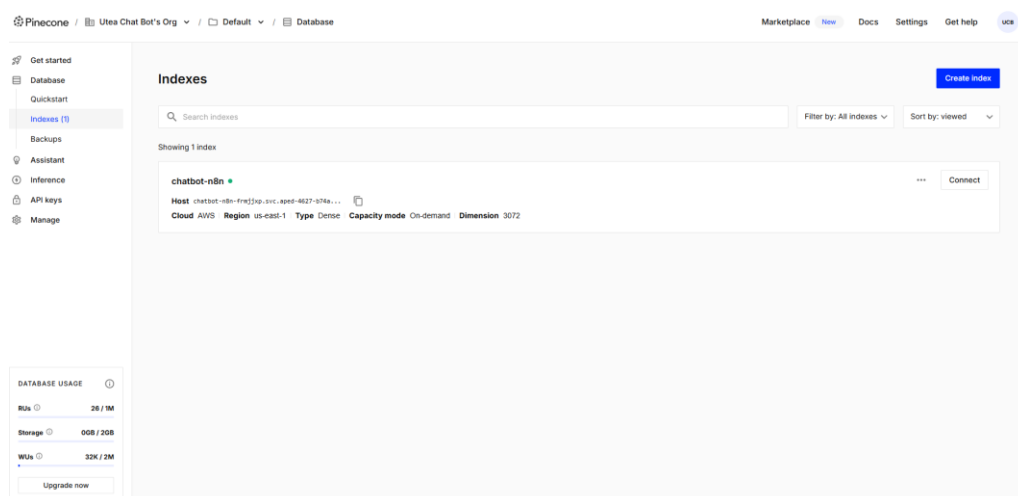
Canal WhatsApp Business, conversación de prueba



Nota. Captura tomada durante el periodo de intervención.

Figura 14

Base de datos vectorial Pinecone, índice chatbot-n8n



Nota. Índice chatbot-n8n con 3072 dimensiones, AWS us-east-1.

4.5.6. Procesamiento estadístico de los datos

En la sexta y última etapa del procedimiento se procesaron los datos recolectados en el pretest y el posttest mediante el Paquete Estadístico para Ciencias Sociales, conocido por sus siglas en inglés como IBM SPSS Statistics (Statistical Package for the Social Sciences), versión 27, complementado con la hoja de cálculo Microsoft Excel 2021 para las tareas de consolidación y depuración previas. El procesamiento se ejecutó en seis pasos secuenciales que se describen a continuación, garantizándose la trazabilidad y reproducibilidad de cada operación.

Paso 1. Construcción y depuración de la base de datos en Excel.

Las respuestas del cuestionario aplicado a los 380 estudiantes únicos (190 del grupo experimental y 190 del grupo control) se consolidaron en una hoja de cálculo Excel en formato largo (long format), con 760 registros correspondientes a los 380 participantes medidos en dos momentos, y 49 variables: el identificador del sujeto, la asignación de grupo (1 = experimental, 2 = control), el momento de medición (1 = pretest, 2 = posttest), los 42 ítems del cuestionario (I1 a I42) y las cuatro

puntuaciones calculadas (D1, D2, D3 y VD_total). A partir de esta estructura se generaron, mediante reestructuración a formato ancho, las variables agregadas por participante (Pre_VD_total, Post_VD_total y sus equivalentes por dimensión) y las diferencias $\text{Post} - \text{Pre}$ (Delta_VD_total, Delta_VD_D1, Delta_VD_D2 y Delta_VD_D3), requeridas para el análisis intergrupos. Durante la depuración se aplicaron tres criterios de eliminación de casos definidos a priori: (a) cuestionarios con respuestas incompletas en cualquiera de los ítems (eliminación por lista); (b) registros con valores fuera del rango válido 1–5, verificados mediante la función $=\text{CONTAR.SI}(\text{rango}; "<1") + \text{CONTAR.SI}(\text{rango}; ">5")$; y (c) patrones de respuesta invariante o «straight-lining» (casos con la misma opción marcada en la totalidad de los 42 ítems de ambos momentos de medición, indicativos de respuesta no atenta). Tras aplicar los tres criterios no fue necesario eliminar ningún caso: las 380 filas resultaron completas, dentro de rango y sin patrones de invarianza total, confirmándose la integridad de la base de datos.

Paso 2. Importación a SPSS y verificación de variables.

La base depurada se importó a IBM SPSS Statistics versión 27 mediante la ruta Archivo → Abrir → Datos, seleccionando el formato .xlsx. Se verificó la correcta lectura de las 380 filas y la correspondencia de tipos de datos: numérico para todas las puntuaciones, numérico con etiqueta de valor para la variable Grupo (1 = Experimental, 2 = Control). Las etiquetas de valor se asignaron en la Vista de variables del programa, garantizando que los reportes posteriores mostraran las etiquetas de grupo en lugar de los códigos numéricos.

Paso 3. Estimación de la fiabilidad mediante alfa de Cronbach.

Se calculó el coeficiente alfa de Cronbach por dimensión y por escala completa para ambas variables del estudio mediante la ruta Analizar → Escala → Análisis de fiabilidad. El cálculo se ejecutó sobre una prueba piloto de 30 estudiantes del grupo experimental en la fase de pretest ($n = 30$); en ese momento los ítems

referidos al chatbot se respondieron como expectativa previa al sistema, dado que la intervención aún no se había desplegado. Para el índice de percepción del sistema (medida secundaria) se procesaron los 21 ítems relativos al chatbot (I1–I21); para la variable dependiente (Gestión de Información Académica) se procesaron los 21 ítems relativos a la gestión de información académica (I22–I42). En cada análisis se seleccionaron los ítems correspondientes y se solicitó la opción Estadísticos → Escala. Los coeficientes obtenidos, junto con la interpretación según los criterios de George y Mallery (2003), se reportan en el apartado 4.4.3 (Tabla 3). La consistencia interna se recalculó además sobre la muestra completa, según se reporta en el apartado 4.4.3.

Paso 4. Verificación del supuesto de normalidad.

Dado que el tamaño muestral de cada grupo supera las 50 unidades de análisis, se aplicó la prueba de Kolmogorov-Smirnov con corrección de Lilliefors mediante la ruta Analizar → Estadísticos descriptivos → Explorar, seleccionando la opción Gráficos → Gráficos con pruebas de normalidad. La prueba se ejecutó sobre las puntuaciones Pre y Post de la variable dependiente total y de cada una de sus tres dimensiones, por separado para cada grupo. Dado que el posttest del grupo experimental no cumplió el supuesto de normalidad ($p < ,001$), al igual que las diferencias Post – Pre del grupo experimental ($p = ,018$) y el pretest del grupo control ($p < ,001$), se determinó el empleo de pruebas no paramétricas para la totalidad de los contrastes inferenciales del estudio.

Paso 5. Pruebas inferenciales no paramétricas.

Se ejecutaron tres bloques de pruebas no paramétricas correspondientes al diseño cuasi-experimental con dos grupos y dos momentos de medición. El primer bloque consistió en la prueba de Wilcoxon de los rangos con signo para la comparación intragrupo Pre→Post, ejecutada mediante la ruta Analizar → Pruebas no paramétricas → Cuadros de diálogo antiguos → 2 muestras relacionadas. Esta

prueba se aplicó por separado al grupo experimental, filtrado mediante Datos → Seleccionar casos → Si: Grupo = 1, y al grupo control, filtrado con Grupo = 2, sobre la variable dependiente total y sobre cada una de sus tres dimensiones. El segundo bloque consistió en la prueba U de Mann-Whitney para la comparación intergrupos sobre las diferencias Post – Pre, ejecutada mediante la ruta Analizar → Pruebas no paramétricas → Cuadros de diálogo antiguos → 2 muestras independientes. Para este análisis se removió el filtro previo mediante Datos → Seleccionar casos → Todos los casos, se utilizó la variable Grupo (1, 2) como variable de agrupación y se aplicó sobre las cuatro variables Delta calculadas. El tercer bloque articuló los resultados de los dos previos bajo el enfoque de diferencia en diferencias (DiD, por sus siglas en inglés Difference in Differences), que permite atribuir el efecto observado a la intervención mediante la combinación de los contrastes intragrupo e intergrupos.

Paso 6. Cálculo del tamaño del efecto.

A partir del estadístico Z reportado por SPSS para cada prueba se calculó el tamaño del efecto r mediante la fórmula $r = |Z| / \sqrt{N}$ (Rosenthal, 1991), donde N corresponde al tamaño muestral del contraste: N corresponde al número de pares con diferencia no nula en las pruebas de Wilcoxon intragrupo (188 en VD total y D1, 184 en D2 y 185 en D3, tras excluir los empates) y a $N = 380$ en las pruebas de Mann-Whitney intergrupos. La interpretación de los coeficientes obtenidos se basó en los criterios de Cohen (1988): $r \geq 0,10$ corresponde a un efecto pequeño, $r \geq 0,30$ a un efecto mediano y $r \geq 0,50$ a un efecto grande. Los resultados completos de los contrastes inferenciales, junto con los tamaños del efecto y su interpretación, se reportan íntegramente en el Capítulo V.

4.6. Análisis de datos

El análisis de datos del estudio se organizó en tres niveles complementarios alineados con la naturaleza del diseño cuasi-experimental con grupo control y la escala ordinal del instrumento.

4.6.1. *Análisis descriptivo*

En el primer nivel se empleó estadística descriptiva sobre las puntuaciones obtenidas en el pretest y el posttest, tanto en la variable dependiente Gestión de Información Académica como en cada una de sus tres dimensiones (Disponibilidad de la información, Acceso a la información y Eficiencia operativa). Se elaboraron tablas de frecuencias absolutas y relativas y se calcularon baremos que reagrupan las cinco categorías de la escala Likert del cuestionario (muy deficiente, deficiente, regular, adecuada y muy adecuada) en tres niveles de interpretación (bajo, medio y alto), conforme al baremo del Anexo 12. Este nivel habilita la caracterización del estado inicial y final de cada grupo en términos comparables y conforma el punto de partida para el análisis inferencial.

4.6.2. *Análisis inferencial*

En el segundo nivel se contrastaron las hipótesis del estudio a través de pruebas no paramétricas, justificadas por el resultado de la prueba de Kolmogorov-Smirnov expuesta en la sub-sección 4.5.6, que rechazó el supuesto de normalidad sobre el posttest del grupo experimental. Se emplearon dos pruebas complementarias: la prueba de Wilcoxon de los rangos con signo, que contrasta las puntuaciones del pretest y el posttest dentro de cada grupo (análisis intragrupo) y habilita la verificación de si la variable dependiente cambia significativamente entre ambos momentos en el grupo experimental y, separadamente, en el grupo control; y la prueba U de Mann-Whitney, que contrasta las diferencias Post – Pre entre el grupo experimental y el grupo control (análisis intergrupos) y habilita la atribución del cambio observado a la intervención bajo el enfoque de diferencia en diferencias (DiD, Difference in Differences). Todas las pruebas se ejecutaron al nivel de significancia $\alpha = ,05$ y se reportaron mediante dos lecturas: el estadístico Z con su nivel de significancia asintótica bilateral, y la decisión correspondiente sobre la hipótesis del estudio.

4.6.3. Tamaño del efecto

En el tercer nivel se reportó el tamaño del efecto r para cada contraste, calculado a través de la fórmula $r = |Z| / \sqrt{N}$ (Rosenthal, 1991), con N igual al número de pares con diferencia no nula en los contrastes intragrupo (188, 188, 184 y 185 según el contraste) y $N = 380$ en los contrastes intergrupos. La interpretación se sustentó en los criterios de Cohen (1988): $r \geq 0,10$ corresponde a un efecto pequeño, $r \geq 0,30$ a un efecto mediano y $r \geq 0,50$ a un efecto grande. La inclusión sistemática del tamaño del efecto, complementaria al valor de significancia estadística, responde a las recomendaciones contemporáneas para la presentación de resultados en ciencias sociales aplicadas y habilita la valoración no solo de la existencia del efecto, sino también de su magnitud práctica.

4.7. Consideraciones éticas

La presente investigación se llevó a cabo respetando los principios éticos para la investigación con seres humanos determinados en la Declaración de Helsinki (Asociación Médica Mundial, 2013). Por tratarse de un estudio de bajo riesgo (cuestionarios anónimos, sin datos personales, aplicados a estudiantes adultos), se aplicó consentimiento informado tácito a través de cláusula introductoria al instrumento (ver Anexo 2). Antes de responder, cada participante fue informado del propósito del estudio, el carácter voluntario, el anonimato, la confidencialidad y el uso exclusivamente académico de los datos. La implementación del chatbot se presentó como herramienta de apoyo institucional, no como evaluación del desempeño.

La investigación dispuso de autorización de la Facultad de Ingeniería de la Universidad Tecnológica de los Andes mediante Resolución Directoral N° 077-2025-UTEA-FI-EPISel-FC (Anexo 6). La unidad académica no requiere revisión por Comité de Ética para estudios de bajo riesgo con cuestionarios anónimos sin datos sensibles; no obstante, el diseño respetó íntegramente los principios de autonomía, beneficencia, no maleficencia y justicia (CONCYTEC, 2018).

Los datos recolectados se resguardaron en archivos cifrados de acceso restringido al investigador y al asesor, y serán custodiados durante el periodo requerido por la normativa institucional vigente, transcurrido el cual serán suprimidos de manera segura. No se compartieron datos individuales con terceros y los resultados se reportan únicamente en forma agregada. El autor declara no tener conflicto de interés vinculado al desarrollo o a los proveedores de servicios de inteligencia artificial empleados.

V. Resultados y discusión

5.1. Resultados descriptivos

5.1.1. Caracterización de la muestra

Tabla 7

Composición de la muestra por grupo de estudio

Grupo	n	%
Grupo experimental (con chatbot, asistente virtual conversacional)	190	50,0
Grupo control (sin chatbot)	190	50,0
Total	380	100,0

Nota. La muestra se distribuyó de manera equitativa entre el grupo experimental y el grupo control (50 % cada uno), lo que favorece la potencia del contraste intergrupos; la comparabilidad inicial se verificó mediante la prueba U de Mann-Whitney sobre el pretest, contraste que abarca exclusivamente el constructo medido. El grupo experimental corresponde a la Escuela Profesional de Estomatología (190 de 288 matriculados; 66,0 %) y el grupo control, a las once escuelas profesionales restantes de la sede Abancay (190 de 3 695 matriculados; 5,1 %). El instrumento no registró la escuela de procedencia individual, el ciclo, el sexo ni la edad de los participantes, por lo que no es posible presentar una caracterización sociodemográfica ni académica de los grupos; esta omisión se declara como limitación en el apartado 5.3.

5.1.2. Distribución de niveles por dimensión

Tabla 8

Distribución de niveles, D1 Disponibilidad de la información (Grupo Experimental)

Nivel	Pretest f	Pretest %	Postest f	Postest %
Bajo	35	18,4	0	0,0
Medio	76	40,0	7	3,7
Alto	79	41,6	183	96,3
Total	190	100,0	190	100,0

Nota. Datos procesados en IBM SPSS Statistics v27.

Análisis e interpretación. En la Tabla 8 se observa un desplazamiento sustantivo de la distribución de niveles de la dimensión Disponibilidad de la información del grupo experimental. Antes de la intervención, el 18,4 % se ubicaba en el nivel bajo, el 40,0 % en el medio y el 41,6 % en el alto. Tras la implementación del chatbot, el nivel alto concentra el 96,3 %, el medio desciende a 3,7 % y el nivel bajo desaparece (0,0 %). En el grupo control, expuesto a los canales tradicionales durante el mismo periodo, la distribución se mantuvo estable (pretest 21,6 % bajo, 37,9 % medio y 40,5 % alto; postest 13,7 %, 37,9 % y 48,4 %), lo que delimita el cambio observado al grupo intervenido. Este desplazamiento hacia el nivel superior evidencia que el chatbot resolvió la principal carencia previa de los estudiantes: la imposibilidad de obtener información académica en cualquier momento sin depender de horarios de atención presencial. El resultado es coherente con el principio de disponibilidad continua que fundamenta el diseño de servicios digitales (Treynor et al., 2017).

Tabla 9*Distribución de niveles, D2 Acceso a la información (Grupo Experimental)*

Nivel	Pretest f	Pretest %	Postest f	Postest %
Bajo	31	16,3	0	0,0
Medio	86	45,3	8	4,2
Alto	73	38,4	182	95,8
Total	190	100,0	190	100,0

Nota. Datos procesados en IBM SPSS Statistics v27.

Análisis e interpretación. La Tabla 9 muestra que en el pretest el 16,3 % del grupo experimental se ubicaba en el nivel bajo en cuanto al acceso a la información académica, el 45,3 % en el medio y el 38,4 % en el alto. Tras la intervención, el nivel alto concentra el 95,8 %, el medio desciende a 4,2 % y el nivel bajo desaparece (0,0 %). En el grupo control la distribución varió de forma marginal (pretest 28,9 % bajo, 26,8 % medio y 44,2 % alto; postest 12,1 %, 40,0 % y 47,9 %). Si bien el desplazamiento hacia el nivel alto es contundente, el porcentaje residual en el nivel medio sugiere que persisten oportunidades de mejora vinculadas a la calidad de la base de conocimiento institucional (encontrabilidad, claridad y precisión), en línea con los principios Localizable, Accesible, Interoperable, Reutilizable (FAIR) (Wilkinson et al., 2016) y los criterios de calidad de datos de Wang y Strong (1996).

Tabla 10*Distribución de niveles, D3 Eficiencia operativa (Grupo Experimental)*

Nivel	Pretest f	Pretest %	Postest f	Postest %
Bajo	29	15,3	1	0,5
Medio	81	42,6	5	2,6
Alto	80	42,1	184	96,8
Total	190	100,0	190	100,0

Nota. Datos procesados en IBM SPSS Statistics v27.

La Tabla 10 evidencia el mayor desplazamiento de las tres dimensiones: en el pretest el 15,3 % de los estudiantes se ubicaba en el nivel bajo, el 42,6 % en el medio y el 42,1 % en el alto respecto a la eficiencia operativa. Tras la intervención, el nivel alto concentra el 96,8 %, el medio desciende a 2,6 % y el bajo a 0,5 %. En el grupo control la variación fue menor (pretest 18,4 % bajo, 42,1 % medio y 39,5 % alto; postest 15,8 %, 34,7 % y 49,5 %). La práctica desaparición del nivel bajo en el grupo experimental (del 15,3 % al 0,5 %) confirma que el chatbot redujo sustantivamente los tiempos de atención y que las consultas suelen resolverse en el primer intento, en línea con el indicador Resolución en el Primer Contacto (FCR) que se vincula positivamente con la satisfacción del usuario (Abdullateef et al., 2011) y con la Ley de Little que explica que reducir el tiempo medio de gestión disminuye la congestión percibida (Little, 1961).

5.2. Contrastación de hipótesis

Tabla 11

Regla de decisión para pruebas de hipótesis

Condición	Decisión
p-valor $\leq 0,05$	Se rechaza H_0 y se acepta H_1
p-valor $> 0,05$	No se rechaza H_0

Nota. $\alpha = 0,05$ (bilateral).

Tabla 12

Resultados de la prueba U de Mann-Whitney en el pretest: equivalencia inicial entre grupos ($n_1 = n_2 = 190$)

Variable	U	Z	p	r
Variable dependiente total	17 559,0	-0,459	,646	,024
D1 Disponibilidad de la información	17 874,0	-0,165	,869	,008
D2 Acceso	17 867,0	-0,171	,864	,009
D3 Eficiencia operativa	17 184,0	-0,811	,417	,042

Nota. Prueba U de Mann-Whitney en pretest. Todas las $p > ,05$ confirman equivalencia inicial entre grupos. $r = |Z| / \sqrt{N}$. Datos procesados en IBM SPSS Statistics v27.

Previo a la evaluación de las hipótesis, se calcularon los estadísticos descriptivos de las puntuaciones totales de ambas variables por grupo y momento de medición. La Tabla 13 resume las medias, desviaciones estándar y los valores mínimo y máximo obtenidos.

Tabla 13*Estadísticos descriptivos por grupo y momento de medición*

	Variable	Momento	n	M	DE	Mín	Máx
Experimental	Percepción del chatbot: índice de 21 ítems(medida secundaria)	Pretest	190	3,543	0,541	2,24	4,81
		Posttest	190	4,154	0,445	2,43	4,90
	VD Total (Gestión Info.)	Pretest	190	3,225	0,513	2,05	4,62
		Posttest	190	4,246	0,370	3,10	4,81
		Δ (Post-Pre)	190	1,021	0,626	-0,71	2,43
Control	VD Total (Gestión Info.)	Pretest	190	3,180	0,575	2,05	4,19
		Posttest	190	3,302	0,521	1,76	4,52
		Δ (Post-Pre)	190	0,122	0,736	-1,43	2,19

Nota. Percepción del chatbot = índice de percepción del sistema, medida secundaria (solo grupo experimental); en el pretest se registra como expectativa previa y en el posttest como experiencia de uso; por tratarse de constructos no equivalentes —expectativa frente a experiencia—, el contraste intragrupo de esta medida posee carácter exploratorio y no somete a prueba ninguna de las hipótesis formuladas en el apartado 2.4. VD = Variable dependiente (ambos grupos). Δ = diferencia posttest – pretest. Datos procesados en IBM SPSS Statistics v27.

H₁: La implementación del chatbot con inteligencia artificial (IA) mejora significativamente la gestión de información académica en la Universidad Tecnológica de los Andes, 2025.

Tabla 14

Prueba de normalidad (Kolmogorov-Smirnov)

Variable / Grupo / Momento	D	gl	Sig.	Distribución
Percepción del chatbot – Experimental – Pretest	0,048	190	,200*	Normal
Percepción del chatbot – Experimental – Posttest	0,110	190	<,001	No normal
VD Total – Experimental – Pretest	0,050	190	,200*	Normal
VD Total – Experimental – Posttest	0,120	190	<,001	No normal
Delta VD Total – Experimental	0,072	190	,018	No normal
VD Total – Control – Pretest	0,100	190	<,001	No normal
VD Total – Control – Posttest	0,054	190	,200*	Normal
Delta VD Total – Control	0,055	190	,200*	Normal

Nota. * Límite inferior (corrección de Lilliefors); normalidad rechazada si $p < ,05$. Varias distribuciones no normales justifican pruebas no paramétricas (Wilcoxon y Mann-Whitney).

Datos procesados en IBM SPSS Statistics v27.

5.2.1. Comprobación de la hipótesis general

Análisis de equivalencia inicial entre grupos

Antes de contrastar las hipótesis se examinó la equivalencia inicial entre el grupo experimental y el grupo control en el pretest, condición necesaria en un diseño cuasi-experimental (Shadish et al., 2002). Dado que la prueba de Kolmogorov-Smirnov con corrección de Lilliefors reveló desviaciones significativas de la normalidad en varias distribuciones (Tabla 14), se aplicó la prueba U de Mann-Whitney sobre los puntajes totales de la variable dependiente y de cada una de sus tres dimensiones en el momento basal ($n_1 = n_2 = 190$).

Los resultados no evidenciaron diferencias estadísticamente significativas entre el grupo experimental y el grupo control en el pretest para la variable total ni para ninguna de sus dimensiones (todas las $p > ,05$; tamaños del efecto $r \leq ,042$). Estos hallazgos confirman la comparabilidad inicial entre ambos grupos y permiten atribuir razonablemente las diferencias observadas en el posttest al efecto de la intervención y no a un sesgo de partida. Los estadísticos detallados se presentan en la Tabla 12.

Tabla 15

Rangos de Wilcoxon, Pretest y posttest, variable dependiente Total (grupo experimental)

		N	Rango promedio	Suma de rangos
Rangos negativos	(post < pre)	14 ^a	16,21	227,00
Rangos positivos	(post > pre)	174 ^b	100,80	17 539,00
Empates		2 ^c	—	—
Total		190	—	—

Nota. ^a post < pre. ^b post > pre. ^c post = pre. Datos procesados en IBM SPSS Statistics v27.

Tabla 16*Estadísticos de prueba, Wilcoxon, VD Total*

Estadístico	Valor
Z	-11,587 ^a
Sig. asintótica (bilateral)	<,001
Tamaño del efecto $r = Z / \sqrt{N}$	0,845
Interpretación	Grande ($r \geq 0,50$)

Nota. ^a Basado en rangos negativos. $r = |-11,587| / \sqrt{188} = 0,845$, donde 188 corresponde a los pares con rango asignado (190 pares menos 2 empates). Cohen (1988): $r \geq 0,10$ = pequeño; $r \geq 0,30$ = mediano; $r \geq 0,50$ = grande. Datos procesados en IBM SPSS Statistics v27.

Tabla 17*Rangos por grupo, U de Mann-Whitney, Delta VD Total (postest – pretest)*

Grupo	N	Rango promedio	Suma de rangos
Grupo experimental	190	250,96	47 682,00
Grupo control	190	130,04	24 708,00
Total	380	—	72 390,00

Nota. Datos procesados en IBM SPSS Statistics v27.

Tabla 18

Estadísticos de prueba, U de Mann-Whitney, Delta VD Total (postest – pretest)

Estadístico	Valor
U de Mann-Whitney	6 563,000
W de Wilcoxon	24 708,000
Z	-10,731
Sig. asintótica (bilateral)	<,001
Tamaño del efecto r	0,550
Interpretación	Grande ($r \geq 0,50$)

Nota. $r = |Z| / \sqrt{(n_1 + n_2)} = |-10,731| / \sqrt{380} = 0,550$. Cohen (1988). Datos procesados en IBM SPSS Statistics v27.

Tabla 19

Rangos de Wilcoxon, Pretest y postest, variable dependiente Total (grupo control)

	N	Rango promedio	Suma de rangos
Rangos negativos (post < pre)	83 ^a	87,27	7 243,00
Rangos positivos (post > pre)	102 ^b	97,67	9 962,00
Empates	5 ^c		
Total	190		

Nota. ^a post < pre. ^b post > pre. ^c post = pre. $Z = -1,864$ ($p = ,062$). No significativo. Datos procesados en IBM SPSS Statistics v27.

Tabla 20

Rangos de Wilcoxon, pretest y posttest, percepción del chatbot, índice de 21 ítems (grupo experimental)

	N	Rango promedio	Suma de rangos
Rangos negativos (post < pre)	33 ^a	61,14	2 017,50
Rangos positivos (post > pre)	157 ^b	102,72	16 127,50
Empates	0 ^c	—	—
Total	190	—	—

Nota. ^a post < pre. ^b post > pre. ^c post = pre. $Z = -9,295$; $p < ,001$; $r = 0,674$ (efecto grande). Percepción del chatbot (expectativa → experiencia) mejoró. Datos procesados en SPSS v27.

La prueba de Wilcoxon intragrupo arrojó $Z = -11,587$ ($p < ,001$; $r = 0,845$) para el grupo experimental, confirmando una mejora significativa y de gran tamaño del pretest al posttest en la VD total. En contraste, el grupo control no mostró cambio significativo ($Z = -1,864$; $p = ,062$). La prueba U de Mann-Whitney sobre los deltas individuales (posttest – pretest) reveló diferencias altamente significativas entre ambos grupos ($U = 6 563$; $Z = -10,731$; $p < ,001$; $r = 0,550$), con un rango promedio de 250,96 para el experimental frente a 130,04 para el control. El efecto neto estimado mediante diferencia en diferencias (DiD) fue de 0,90 puntos en la escala 1–5 ($\Delta_{Exp} = 1,02 - \Delta_{Ctrl} = 0,12$). En consecuencia, se rechaza H_0 y se acepta H_1 : el chatbot con inteligencia artificial mejoró significativamente la gestión de información académica en la Universidad Tecnológica de los Andes.

5.2.2. Comprobación de las hipótesis específicas

Hipótesis específica 1: Disponibilidad de la información

Tabla 21

Rangos de Wilcoxon, D1 Disponibilidad de la información (grupo experimental, pretest vs. posttest)

	N	Rango promedio	Suma de rangos
Rangos negativos	20 ^a	38,00	760,00
Rangos positivos	168 ^b	101,23	17 006,00
Empates	2 ^c	—	—
Total	190	—	—

Nota. Datos procesados en IBM SPSS Statistics v27.

Tabla 22

Estadísticos de prueba, Wilcoxon, D1 Disponibilidad

Estadístico	Valor
Z	-10,877 ^b
Sig. asintótica (bilateral)	< ,001
r (tamaño del efecto)	0,793
Decisión	Se rechaza H ₀₁ y se acepta H ₁₁

Nota. ^b Basado en rangos negativos. $r = |Z| / \sqrt{n}$ efectivo = $10,877 / \sqrt{188} = 0,793$.

Datos procesados en IBM SPSS Statistics v27.

Tabla 23

Rangos por grupo, U de Mann-Whitney, D1 Disponibilidad (diferencia posttest – pretest)

Grupo	N	Rango promedio	Suma de rangos
Grupo experimental	190	241,97	45 973,50
Grupo control	190	139,03	26 416,50
Total	380		

Nota. Datos procesados en IBM SPSS Statistics v27.

Tabla 24

Estadísticos de prueba, U de Mann-Whitney, D1 Disponibilidad (diferencia posttest – pretest)

Estadístico	Valor
U de Mann-Whitney	8 271,500
W de Wilcoxon	26 416,500
Z	-9,139
Sig. asintótica (bilateral)	< ,001
r (tamaño del efecto)	0,469

Nota. $r = |Z| / \sqrt{(n_1 + n_2)} = 9,139 / \sqrt{380} = 0,469$. Cohen (1988): efecto mediano (0,30 $\leq r < 0,50$). Datos procesados en IBM SPSS Statistics v27.

El grupo experimental presentó cambios Pre→Post estadísticamente significativos (Wilcoxon $Z = -10,877$; $p < ,001$; $r = 0,793$, efecto grande), mientras que el grupo control no mostró cambios significativos (Wilcoxon $Z = -1,897$; $p = ,058$). El contraste entre grupos sobre las diferencias Post – Pre (DiD) confirmó la superioridad del experimental (Mann-Whitney: rango promedio 241,97 vs. 139,03; $Z = -9,139$; $p < ,001$; $r = 0,469$, efecto mediano). Por lo tanto, se rechaza H_{01} y se acepta H_{11} : el chatbot incrementó significativamente la disponibilidad de información académica en la Universidad Tecnológica de los Andes.

Hipótesis específica 2: Acceso a la información

Tabla 25

Rangos de Wilcoxon, D2 Acceso a la información (grupo experimental, pretest vs. posttest)

	N	Rango promedio	Suma de rangos
Rangos negativos (post < pre)	15 ^d	30,77	461,50
Rangos positivos (post > pre)	169 ^e	97,98	16 558,50
Empates	6 ^f		
Total	190		

Nota. Datos procesados en IBM SPSS Statistics v27.

Tabla 26

Estadísticos de prueba, Wilcoxon, D2 Acceso

Estadístico	Valor
Z	-11,130 ^b
Sig. asintótica (bilateral)	< ,001
r (tamaño del efecto)	0,820
Decisión	Se rechaza H ₀₂ y se acepta H ₁₂

Nota. ^b Basado en rangos negativos. $r = |Z| / \sqrt{n \text{ efectivo}} = 11,130 / \sqrt{184} = 0,820$.

Datos procesados en IBM SPSS Statistics v27.

Tabla 27

Rangos por grupo, U de Mann-Whitney, D2 Acceso (diferencia posttest – pretest)

Grupo	N	Rango promedio	Suma de rangos
Grupo experimental	190	243,56	46 277,00
Grupo control	190	137,44	26 113,00
Total	380		

Nota. Datos procesados en IBM SPSS Statistics v27.

Tabla 28

Estadísticos de prueba, U de Mann-Whitney, D2 Acceso (diferencia posttest – pretest)

Estadístico	Valor
U de Mann-Whitney	7 968,000
W de Wilcoxon	26 113,000
Z	-9,423
Sig. asintótica (bilateral)	< ,001
r (tamaño del efecto)	0,483

Nota. $r = |Z| / \sqrt{(n_1 + n_2)} = 9,423 / \sqrt{380} = 0,483$. Cohen (1988): efecto mediano ($0,30 \leq r < 0,50$). Datos procesados en IBM SPSS Statistics v27.

El grupo experimental presentó cambios Pre→Post estadísticamente significativos (Wilcoxon $Z = -11,130$; $p < ,001$; $r = 0,820$, efecto grande), mientras que el grupo control no mostró cambios significativos (Wilcoxon $Z = -1,560$; $p = ,119$). El contraste entre grupos sobre las diferencias Post – Pre (DiD) confirmó la

superioridad del experimental (Mann-Whitney: rango promedio 243,56 vs. 137,44; $Z = -9,423$; $p < ,001$; $r = 0,483$, efecto mediano). Por lo tanto, se rechaza H_{02} y se acepta H_{12} : el chatbot facilitó significativamente el acceso a la información académica en la Universidad Tecnológica de los Andes.

Hipótesis específica 3: Eficiencia operativa

Tabla 29

Rangos de Wilcoxon, D3 Eficiencia operativa (grupo experimental, pretest vs. posttest)

	N	Rango promedio	Suma de rangos
Rangos negativos (post < pre)	18 ^a	26,36	74,50
Rangos positivos (post > pre)	167 ^h	100,18	6 730,50
Empates	5 ⁱ		
Total	190		

Nota. Datos procesados en IBM SPSS Statistics v27.

Tabla 30

Estadísticos de prueba, Wilcoxon, D3 Eficiencia operativa

Estadístico	Valor
Z	-11,148 ^b
Sig. asintótica (bilateral)	< ,001
r (tamaño del efecto)	0,820
Decisión	Se rechaza H_{03} y se acepta H_{13}

Nota. ^b Basado en rangos negativos. $r = |Z| / \sqrt{n \text{ efectivo}} = 11,148 / \sqrt{185} = 0,820$.

Datos procesados en IBM SPSS Statistics v27.

Tabla 31

Rangos por grupo, U de Mann-Whitney, D3 Eficiencia operativa (diferencia posttest – pretest)

Grupo	N	Rango promedio	Suma de rangos
Grupo experimental	190	244,72	46 497,00
Grupo control	190	136,28	25 893,00
Total	380		

Nota. Datos procesados en IBM SPSS Statistics v27.

Tabla 32

Estadísticos de prueba, U de Mann-Whitney, D3 Eficiencia operativa (diferencia posttest – pretest)

Estadístico	Valor
U de Mann-Whitney	7 748,000
W de Wilcoxon	25 893,000
Z	-9,629
Sig. asintótica (bilateral)	< ,001
r (tamaño del efecto)	0,494

Nota. $r = |Z| / \sqrt{(n_1 + n_2)} = 9,629 / \sqrt{380} = 0,494$. Cohen (1988): efecto mediano ($0,30 \leq r < 0,50$). Datos procesados en IBM SPSS Statistics v27.

El grupo experimental presentó cambios Pre→Post estadísticamente significativos (Wilcoxon $Z = -11,148$; $p < ,001$; $r = 0,820$, efecto grande), mientras que el grupo control no mostró cambios significativos (Wilcoxon $Z = -1,653$; $p = ,098$). El contraste entre grupos sobre las diferencias Post – Pre (DiD) confirmó la superioridad del experimental (Mann-Whitney: rango promedio 244,72 vs. 136,28;

$Z = -9,629$; $p < ,001$; $r = 0,494$, efecto mediano). Por lo tanto, se rechaza H_{03} y se acepta H_{13} : el chatbot mejoró significativamente la eficiencia operativa en la Universidad Tecnológica de los Andes.

Tabla 33

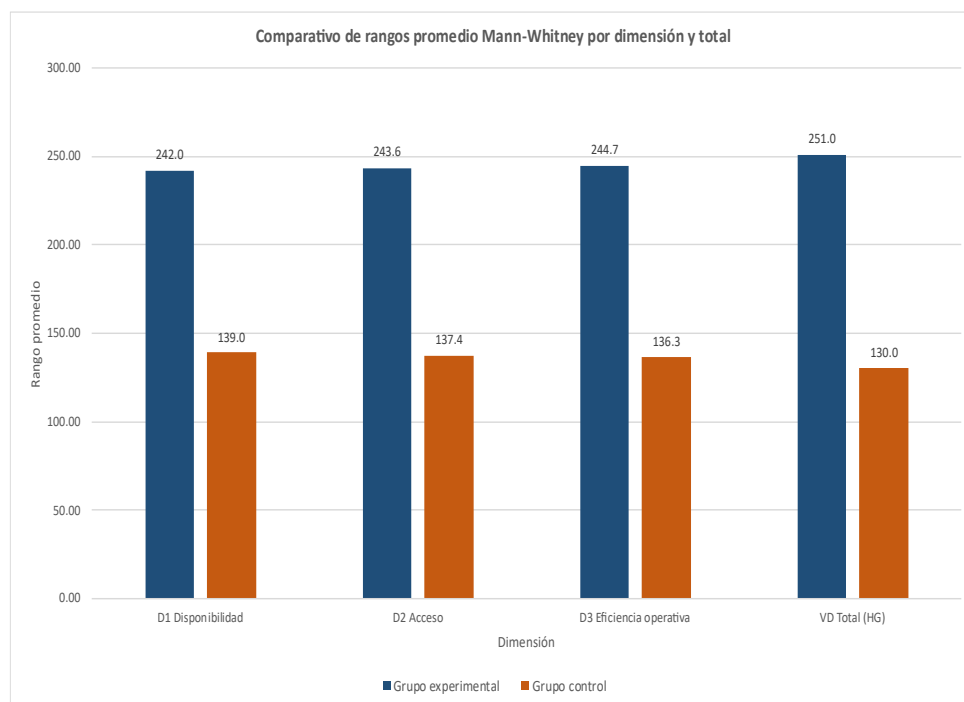
Síntesis de resultados inferenciales

Hipótesis	Prueba	Z	p	r (efecto)	Decisión
HG, VD Total	Wilcoxon (Exp)	-11,587	< ,001	0,845	Aceptar H_1
HG, VD Total	Mann-Whitney (DiD)	-10,731	< ,001	0,550	Aceptar H_1
H_{11} , Disponibilidad	Wilcoxon	-10,877	< ,001	0,793	Aceptar H_{11}
H_{11} , Disponibilidad	Mann-Whitney (DiD)	-9,139	< ,001	0,469	Aceptar H_{11}
H_{12} , Acceso	Wilcoxon	-11,130	< ,001	0,820	Aceptar H_{12}
H_{12} , Acceso	Mann-Whitney (DiD)	-9,423	< ,001	0,483	Aceptar H_{12}
H_{13} , Eficiencia op.	Wilcoxon	-11,148	< ,001	0,820	Aceptar H_{13}
H_{13} , Eficiencia op.	Mann-Whitney (DiD)	-9,629	< ,001	0,494	Aceptar H_{13}

Nota. $\alpha = 0,05$. Wilcoxon (Exp): contraste intragrupo sobre los pares con diferencia no nula (188 en VD total y D1, 184 en D2 y 185 en D3). Mann-Whitney (Δ): contraste DiD entre grupos $n = 380$. r según Cohen (1988): pequeño $\geq 0,10$; mediano $\geq 0,30$; grande $\geq 0,50$. Grupo control sin cambios significativos.

Figura 15

Comparativo de rangos promedio Mann-Whitney por dimensión y variable dependiente total (grupo experimental vs. grupo control, sobre la diferencia entre pretest y posttest)



Nota. Diferencias entre grupos: 102,94 (D1) a 120,92 (VD total); r entre 0,469 y 0,550 (medianos a grandes). Datos procesados en SPSS v27.

5.3. Discusión de resultados

El objetivo general de la investigación fue determinar el efecto de un chatbot con inteligencia artificial en la gestión de información académica en la Universidad Tecnológica de los Andes. Los resultados de la prueba de Wilcoxon para el grupo experimental arrojaron $Z = -11,587$ ($p < ,001$; $r = 0,845$), lo que evidencia una mejora significativa y de magnitud grande entre el pretest y el posttest. En contraste, el grupo control no mostró cambio significativo ($Z = -1,864$; $p = ,062$), lo que descarta que los factores de maduración o el efecto del retest expliquen la mejora observada. La prueba U de Mann-Whitney sobre la diferencia pretest–posttest entre ambos grupos confirmó la

superioridad del grupo experimental ($Z = -10,731$; $p < ,001$; $r = 0,550$), con un efecto neto de 0,90 puntos en la escala 1–5. En consecuencia, se confirmó la hipótesis general: la implementación del chatbot con inteligencia artificial mejoró significativamente la gestión de información académica.

En la presente investigación se obtuvo un tamaño del efecto grande ($r = 0,845$ intragrupo; $r = 0,550$ intergrupos), resultado que coincide con Essel et al. (2022), quienes, en un diseño cuasi-experimental con 68 estudiantes en Ghana, reportaron una mejora sustantiva del grupo experimental (40,6 $\rightarrow$ 81,1 puntos) frente al grupo control (43,4 $\rightarrow$ 65,2), con $\eta^2p = 0,57$, también un efecto grande, aunque en un contexto de rendimiento académico y no de gestión informativa. Asimismo, Veliz y Velásquez (2024), en la UNPRG, Perú, observaron un salto de la aceptación del 48,7 % al 89,45 % tras la implementación de un chatbot para la atención académica, diferencia estadísticamente significativa que resulta consistente con la mejora hallada en el presente estudio, aunque su diseño pre-experimental con un único grupo limita la atribución causal. De manera similar, Quispe Cruz (2024), en la UPCI de Lima, reportó un incremento de la calidad global del servicio de 2,04 a 3,80 (Wilcoxon, $p < ,001$) en un diseño pre-experimental con 54 egresados, lo que coincide en la dirección del efecto, si bien la ausencia de grupo control no permite descartar explicaciones alternativas.

La magnitud del efecto obtenido puede atribuirse a tres factores convergentes: primero, la arquitectura RAG que ancló las respuestas del chatbot a la documentación oficial de la universidad, lo que aseguró pertinencia y precisión; segundo, la operación ininterrumpida del sistema a través de los canales web y WhatsApp, lo que eliminó la restricción horaria de la atención presencial; y tercero, la familiaridad del estudiantado peruano con la mensajería instantánea, lo que redujo la barrera de adopción.

Respecto al primer objetivo específico, se buscó determinar el efecto del chatbot en la disponibilidad de la información. La prueba de Wilcoxon indicó una mejora significativa dentro del grupo experimental ($r = 0,793$; efecto grande), mientras que la

prueba U de Mann-Whitney entre grupos arrojó $r = 0,469$ (efecto mediano). Los niveles de la dimensión reflejaron un desplazamiento sustantivo: el nivel alto pasó del 41,6 % en el pretest al 96,3 % en el posttest, mientras el nivel bajo, que agrupaba al 18,4 % de los estudiantes, quedó sin casos. Se confirmó, por tanto, la primera hipótesis específica: el chatbot mejoró significativamente la disponibilidad de la información.

Este hallazgo coincide con Balderas et al. (2023), quienes demostraron que un chatbot con NLU alivió los canales tradicionales de atención institucional en la Escuela Superior de Ingeniería de Puerto Real, España, precisamente al ofrecer disponibilidad permanente durante situaciones de crisis. En la misma dirección, Rodríguez y Mune (2022) implementaron un chatbot de referencia virtual en la San José State University Library con el propósito explícito de atender consultas fuera del horario laboral, y concluyeron que el sistema amplió efectivamente la ventana de atención. En el presente estudio, la operación ininterrumpida del chatbot a través de los canales web y WhatsApp eliminó la dependencia del horario de atención presencial, lo que explica la magnitud del cambio observado en esta dimensión.

En relación con el segundo objetivo específico, referido al efecto del chatbot en el acceso a la información, la prueba de Wilcoxon para el grupo experimental arrojó $r = 0,820$ (efecto grande) y la prueba U de Mann-Whitney entre grupos $r = 0,483$ (efecto mediano). La distribución de niveles evidenció que el nivel alto pasó del 38,4 % en el pretest al 95,8 % en el posttest, mientras que el nivel bajo, que agrupaba al 16,3 % de los estudiantes, quedó sin casos. Se confirmó la segunda hipótesis específica: el chatbot mejoró significativamente el acceso a la información.

En la presente investigación se obtuvo un 95,8 % de estudiantes en el nivel alto de acceso tras la intervención, resultado que coincide con Zapata (2023), quien reportó una facilidad de acceso del 93,30 % y una precisión informativa del 93,0 % tras implementar un chatbot para trámites en la Universidad Nacional de Piura. De igual modo, Espinosa-Luna et al. (2023), en la Universidad Nacional de Trujillo, observaron que el 93

% de los estudiantes juzgó útiles las respuestas del chatbot basado en GPT-3.5-Turbo, lo que refuerza la evidencia de que los modelos de lenguaje grande facilitan el acceso a información pertinente. Por otra parte, Neupane et al. (2024), en Mississippi State University, evaluaron un chatbot basado en la arquitectura RAG y reportaron un puntaje RAGAS promedio de 0,96, lo que confirma que el anclaje de las respuestas a documentación institucional eleva la precisión y la pertinencia de la información proporcionada, aspecto central en la presente intervención.

No obstante, Han y Lee (2022), en The University of Texas at Austin, reportaron un hallazgo divergente: en su estudio experimental entre sujetos, los estudiantes prefirieron la página estática de FAQ frente al chatbot y reportaron mayores barreras de interacción con este último. Esta diferencia se explica por tres condiciones: el chatbot de Han y Lee operaba sobre un corpus FAQ genérico de un MOOC sin anclaje institucional; su muestra era lingüística y geográficamente heterogénea, lo que introdujo variabilidad en la experiencia; y el canal de interacción no incluía WhatsApp, medio de alta familiaridad para el estudiantado peruano. Estas tres condiciones, resueltas en el presente diseño mediante la arquitectura RAG, la documentación local y la integración con WhatsApp, explican la dirección opuesta de los resultados.

En cuanto al tercer objetivo específico, referido al efecto del chatbot en la eficiencia operativa, la prueba de Wilcoxon para el grupo experimental arrojó $r = 0,820$ (efecto grande) y la prueba U de Mann-Whitney entre grupos $r = 0,494$ (efecto mediano). Esta dimensión registró el mayor desplazamiento en la distribución de niveles: el nivel alto pasó del 42,1 % en el pretest al 96,8 % en el posttest, en tanto que el nivel bajo descendió del 15,3 % al 0,5 %. Se confirmó la tercera hipótesis específica: el chatbot mejoró significativamente la eficiencia operativa de la gestión de información académica.

Este resultado coincide con Quispe Huacani (2023), quien, en la UNJBG de Tacna, reportó funcionalidad alta en el 100 % del grupo experimental frente al 32 % del grupo control, lo que evidencia una brecha similar a la observada en el presente estudio.

En la misma línea, Quispe Cruz (2024), en la UPCI de Lima, observó que la capacidad de respuesta institucional pasó de 2,04 a 3,81 (Wilcoxon, $p < ,001$), mejora que refleja la optimización del tiempo de atención tras la incorporación del chatbot. Cari-Incahuanaco et al. (2025), en la UNAMBA, Apurímac, encontraron que el 75 % de los estudiantes en situación de riesgo aprobaron el semestre tras el uso de ChatGPT como tutor virtual, resultado que, si bien corresponde a una dimensión distinta (rendimiento académico), refuerza la evidencia de que las herramientas de inteligencia artificial optimizan procesos universitarios cuando se adaptan al contexto institucional local. La práctica desaparición de los niveles bajos en el posttest (del 15,3 % al 0,5 %) sugiere que la combinación de respuestas inmediatas, disponibilidad continua y resolución en el primer contacto produce una mejora sustantiva de la experiencia del estudiante.

En síntesis, los resultados del presente estudio coinciden con los de diez antecedentes: Balderas et al. (2023), en disponibilidad de la atención institucional; Essel et al. (2022), en la magnitud grande del efecto sobre el desempeño estudiantil; Veliz y Velásquez (2024), en la mejora significativa de la aceptación de la atención académica; Rodriguez y Mune (2022), en la extensión de la ventana horaria de atención mediante chatbot; Neupane et al. (2024), en la precisión de un chatbot RAG anclado a documentación institucional; Espinosa-Luna et al. (2023), en la utilidad percibida de un chatbot basado en modelos de lenguaje grande; Quispe Huacani (2023) y Quispe Cruz (2024), en la mejora de indicadores de funcionalidad y calidad del servicio; Zapata (2023), en la facilidad de acceso y la precisión informativa; y Cari-Incahuanaco et al. (2025), en el efecto positivo de herramientas de inteligencia artificial sobre procesos universitarios. La única discrepancia se registró con Han y Lee (2022), quienes hallaron preferencia por la página estática de FAQ frente al chatbot; esta divergencia se explica por las diferencias en el anclaje de la base de conocimiento, la homogeneidad lingüística de los usuarios y el canal de interacción, condiciones que el presente diseño resolvió mediante la arquitectura RAG, la documentación local y la integración con WhatsApp.

Finalmente, corresponde declarar cinco limitaciones del estudio: tres de orden metodológico y dos referidas al alcance informativo y a la naturaleza del insumo documental del sistema implementado. Primera: los contrastes de la hipótesis general y de las tres hipótesis específicas se efectuaron sobre la misma muestra sin aplicar corrección por comparaciones múltiples (p. ej., Bonferroni; Bland y Altman, 1995), decisión que incrementa nominalmente la probabilidad de error tipo I (Rosenthal, 1991). No obstante, el riesgo práctico resulta acotado: todos los contrastes principales arrojaron $p < ,001$, valor que permanecería significativo incluso bajo el criterio corregido más exigente (α ajustado = $0,05/4 = 0,0125$), por lo que las conclusiones del estudio no se ven comprometidas. Segunda: el procedimiento de selección no probabilístico y asimétrico entre grupos, inherente al diseño de grupo control no equivalente adoptado, restringe la generalización estadística de los hallazgos a poblaciones con características equivalentes a las de los estudiantes de pregrado de la sede Abancay; la extensión de los resultados a otros contextos institucionales requerirá replicaciones con muestreo probabilístico o evaluaciones locales previas. Tercera: el alcance informativo del chatbot está circunscrito al corpus documental indexado en Pinecone, descrito en la sub-sección 4.5.3 y detallado en el Anexo 11, que comprende el cronograma académico y el proceso de admisión, el listado de carreras y sedes, las autoridades institucionales, las mensualidades por carrera, los procedimientos de titulación y el Texto Único de Procedimientos Administrativos. Por consiguiente, el sistema no atiende consultas ajenas a ese dominio —entre ellas la situación académica individual, las calificaciones, el estado de matrícula o las obligaciones económicas del estudiante—, dado que no se integró con los sistemas transaccionales de la universidad y opera únicamente sobre documentos estáticos. A esta delimitación se añade que el corpus se conformó con documentos difundidos por los canales digitales institucionales, sin certificación documental individual por parte de las dependencias emisoras, por lo que la exactitud de las respuestas del sistema queda condicionada a la exactitud, la integridad y la vigencia de la información publicada en dichos canales al momento de la indexación, ejecutada en la primera semana de febrero

de 2026. Asimismo, la vigencia de sus respuestas queda acotada a la fecha de la última indexación: las modificaciones posteriores en cronogramas, tasas del TUPA o autoridades no se reflejan hasta que el documento actualizado se deposita en el repositorio institucional y el flujo Push Data Vector lo reindexa. A ello se suma que el modelo generativo empleado posee su propio corte de entrenamiento, por lo que ante consultas no cubiertas por el corpus puede producir respuestas plausibles pero no verificadas, riesgo que la arquitectura RAG mitiga sin eliminarlo (Ji et al., 2023). En consecuencia, los efectos reportados deben interpretarse como el impacto del chatbot sobre consultas informativas de carácter procedimental, y no como un reemplazo integral de la atención académica de la institución. Cuarta: una parte considerable de los documentos que conforman el corpus no corresponde a texto digital nativo, sino a imágenes y archivos PDF obtenidos por digitalización fotográfica de hojas impresas, de modo que su incorporación a la base de conocimiento exigió una etapa previa de extracción mediante reconocimiento óptico de caracteres y análisis de imagen, conforme al flujo descrito en la sub-sección 4.5.4. Ninguna tecnología de reconocimiento óptico garantiza una transcripción exacta: la resolución de la captura, la inclinación y las sombras de la toma fotográfica, los sellos y las firmas superpuestas, la estructura tabular y el desgaste tipográfico elevan la tasa de error respecto de los documentos de origen digital. Por consiguiente, un error de transcripción en un dato puntual —una tasa, un plazo o una fecha— puede propagarse al fragmento indexado y reflejarse en la respuesta entregada al usuario, razón por la cual la revisión humana del texto extraído antes de su indexación se identifica como un control indispensable para las implementaciones posteriores del sistema. Quinta: la condición experimental coincide con la unidad académica de procedencia. El grupo experimental se conformó con 190 estudiantes de la Escuela Profesional de Estomatología —66,0 % de sus 288 matriculados—, unidad que otorgó la autorización operativa y el acceso presencial sostenido durante el trabajo de campo, mientras que el grupo control se integró con 190 estudiantes de las once escuelas profesionales restantes de la sede Abancay; el instrumento no registró la escuela de

procedencia, el ciclo, el sexo ni la edad de los participantes, de modo que no fue posible presentar una caracterización sociodemográfica ni académica de ambos grupos. En consecuencia, el efecto atribuido a la intervención no puede disociarse por completo de las características propias de la escuela de procedencia —perfil de ingreso, familiaridad tecnológica, carga académica y régimen de permanencia en campus—, configuración que corresponde al diseño de grupo control no equivalente (Shadish et al., 2002). Cuatro elementos acotan este riesgo: la equivalencia inicial entre grupos se verificó empíricamente en el pretest, sin diferencias estadísticamente significativas en la variable dependiente total ($p = ,646$) ni en sus tres dimensiones ($p = ,869$; $p = ,864$; $p = ,417$); el contraste intergrupos se efectuó sobre las diferencias posttest – pretest y no sobre puntuaciones brutas del posttest, procedimiento que cancela las diferencias basales estables entre unidades académicas; el grupo control, expuesto al mismo semestre académico y al mismo instrumento, no registró variación estadísticamente significativa al 5 % ($Z = -1,864$; $p = ,062$), lo que acota las explicaciones por maduración, historia, instrumentación y regresión a la media; y la variable dependiente concierne a información académica institucional de alcance común a todas las escuelas profesionales —cronograma académico, admisión, titulación y tasas del Texto Único de Procedimientos Administrativos— y no a contenidos propios de una disciplina. Con todo, la atribución causal debe leerse bajo esta restricción, y su verificación plena exige una replicación con asignación aleatoria por conglomerados entre escuelas profesionales, conforme se recomienda en el apartado VII.

VI. Conclusiones

Primera. Se concluye que la implementación del chatbot con inteligencia artificial evidenció resultados favorables sobre la gestión de información académica en la Universidad Tecnológica de los Andes, 2025. En el grupo experimental, la prueba de Wilcoxon arrojó $Z = -11,587$; $p < ,001$; $r = 0,845$, mientras que en el grupo control no se obtuvieron diferencias estadísticamente significativas al 5 % ($Z = -1,864$; $p = ,062$). La comparación entre grupos mediante la prueba U de Mann-Whitney reportó $Z = -10,731$; $p < ,001$; $r = 0,550$. Ambos tamaños de efecto fueron grandes según Cohen.

Segunda. Se concluye que la implementación del chatbot evidenció resultados favorables en la disponibilidad de la información académica, en donde el nivel alto pasó del 41,6 % en el pretest al 96,3 % en el posttest en el grupo experimental, sin casos en el nivel bajo tras la intervención. La prueba de Wilcoxon arrojó $Z = -10,877$; $p < ,001$; $r = 0,793$, y la prueba U de Mann-Whitney reportó $Z = -9,139$; $p < ,001$; $r = 0,469$.

Tercera. Se concluye que la implementación del chatbot evidenció resultados favorables en el acceso a la información académica, en donde el nivel alto pasó del 38,4 % en el pretest al 95,8 % en el posttest en el grupo experimental, sin casos en el nivel bajo tras la intervención. La prueba de Wilcoxon arrojó $Z = -11,130$; $p < ,001$; $r = 0,820$, y la prueba U de Mann-Whitney reportó $Z = -9,423$; $p < ,001$; $r = 0,483$.

Cuarta. Se concluye que la implementación del chatbot evidenció resultados favorables en la eficiencia operativa de la gestión de información académica, en donde el nivel alto pasó del 42,1 % en el pretest al 96,8 % en el posttest en el grupo experimental, y el nivel bajo descendió del 15,3 % al 0,5 %. La prueba de Wilcoxon arrojó $Z = -11,148$; $p < ,001$; $r = 0,820$, y la prueba U de Mann-Whitney reportó $Z = -9,629$; $p < ,001$; $r = 0,494$.

VII. Recomendaciones

Primera. Se sugiere a las autoridades de la Universidad Tecnológica de los Andes adoptar el chatbot como una herramienta permanente para apoyar la gestión de información académica, asignando a una persona encargada de su funcionamiento dentro de la institución.

Segunda. Se recomienda mantener el chatbot activo de manera permanente y revisar periódicamente la información académica que ofrece, de modo que los estudiantes encuentren datos actualizados cada vez que lo consulten.

Tercera. Se recomienda dar a conocer el chatbot entre los estudiantes mediante orientaciones sencillas sobre su uso, para que puedan acceder con facilidad a la información académica que necesitan.

Cuarta. Se sugiere aprovechar el chatbot para responder las consultas más frecuentes de los estudiantes, de manera que el personal administrativo disponga de más tiempo para atender los casos que requieren un trato personalizado.

Quinta. Se recomienda replicar el estudio con muestreo probabilístico, con asignación aleatoria por conglomerados entre escuelas profesionales —de modo que la condición experimental no coincida con una única unidad académica—, con registro de las variables sociodemográficas y académicas de los participantes y con corrección por comparaciones múltiples en los contrastes de hipótesis, a fin de fortalecer la generalización de los hallazgos a otros contextos institucionales, incluidas las filiales de Andahuaylas y Cusco.

Sexta. Se recomienda establecer un procedimiento institucional de curación documental para la base de conocimiento del chatbot, en el que cada dependencia designe a un responsable de remitir la versión vigente de los documentos de su competencia, consignando fecha de emisión y número de versión, y de comunicar oportunamente sus actualizaciones. Un circuito de esta naturaleza permitiría certificar la

procedencia y la vigencia de cada documento indexado, elevar la precisión de las respuestas entregadas al estudiante y facilitar futuras investigaciones que requieran información institucional sistematizada.

Séptima. Se recomienda que la documentación oficial difundida por los canales digitales de la institución se publique en formato de texto digital nativo, con texto seleccionable y búsqueda interna habilitada, y no como imágenes o digitalizaciones fotográficas de hojas impresas. Esta práctica suprimiría la etapa de reconocimiento óptico de caracteres y, con ella, la principal fuente de error de transcripción identificada en la cuarta limitación del apartado 5.3; adicionalmente, mejoraría la accesibilidad de los documentos para los lectores de pantalla que emplean los estudiantes con discapacidad visual y reduciría el peso de descarga en zonas con conectividad limitada.

VIII. Referencias bibliográficas

- Abdullateef, A. O., Mokhtar, S. S., y Yusoff, R. Z. (2011). The strategic impact of technology based CRM on call centers' performance. *Journal of Internet Banking and Commerce*, 16(1), 1–16.
- Adamopoulou, E., y Moussiades, L. (2020). Chatbots: History, technology, and applications. *Machine Learning with Applications*, 2, 100006.
<https://doi.org/10.1016/j.mlwa.2020.100006>
- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45(1), 131–142.
<https://doi.org/10.1177/0013164485451012>
- Arias, F. G. (2020). *El proyecto de investigación: introducción a la metodología científica* (7.^a ed.). Editorial Episteme.
- Arredondo Castillo, C. C. (2020). *Inteligencia artificial en la educación: uso del chatbot en un curso de pregrado sobre Investigación Académica en una universidad privada de Lima* [Tesis de maestría, Pontificia Universidad Católica del Perú]. Repositorio Institucional PUCP. <http://hdl.handle.net/20.500.12404/20996>
- Asociación Médica Mundial. (2013). *Declaración de Helsinki: Principios éticos para las investigaciones médicas en seres humanos*. <https://www.wma.net/es/policies-post/declaracion-de-helsinki-de-la-amm-principios-eticos-para-las-investigaciones-medicas-en-seres-humanos/>
- Balderas, A., García-Mena, R. F., Huerta, M., Mora, N., y Dodero, J. M. (2023). Chatbot for communicating with university students in emergency situation. *Heliyon*, 9(9), e19517. <https://doi.org/10.1016/j.heliyon.2023.e19517>
- Ballou, D. P., y Pazer, H. L. (1995). Designing information systems to optimize the accuracy-timeliness tradeoff. *Information Systems Research*, 6(1), 51–72.
<https://doi.org/10.1287/isre.6.1.51>

- Bender, E. M., Gebru, T., McMillan-Major, A., y Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? En *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623).
<https://doi.org/10.1145/3442188.3445922>
- Bland, J. M., y Altman, D. G. (1995). Multiple significance tests: The Bonferroni method. *BMJ*, 310(6973), 170. <https://doi.org/10.1136/bmj.310.6973.170>
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., . . . Amodei, D. (2020). Language models are few-shot learners. En *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*.
<https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfc4967418bfb8ac142f64a-Abstract.html>
- Bunge, M. (2014). *La ciencia, su método y su filosofía*. Sudamericana.
- Caldarini, G., Jaf, S., y McGarry, K. (2022). A literature survey of recent advances in chatbots. *Information*, 13(1), 41. <https://doi.org/10.3390/info13010041>
- Cari-Incahuanaco, F., Huaylla-Quispe, A., Quispe-Merma, R. R., y Condori-Quispe, M. (2025). Tutor virtual con ChatGPT para el autoaprendizaje de estudiantes con matrícula condicionada en la Escuela Profesional de Ingeniería Informática y Sistemas - UNAMBA. *C&T Riqchary Revista de Investigación en Ciencia y Tecnología*, 7(1), 35–41. <https://doi.org/10.57166/riqchary.v7.n1.2025.133>
- Carrasco Díaz, S. (2006). *Metodología de la investigación científica: Pautas metodológicas para diseñar y elaborar el proyecto de investigación*. Editorial San Marcos.

- Casas, J., Repullo, J. R., y Donado, J. (2003). La encuesta como técnica de investigación: elaboración de cuestionarios y tratamiento estadístico de los datos. *Atención Primaria*, 31(8), 527–538.
- Centro Nacional de Inteligencia Artificial, y Comisión Económica para América Latina y el Caribe. (2024). *Índice Latinoamericano de Inteligencia Artificial (ILIA) 2024*. CENIA. <https://indicelatam.cl/>
- Chiavenato, I. (2017). *Administración de recursos humanos: El capital humano de las organizaciones (10.ª ed.)*. McGraw-Hill Interamericana.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2.ª ed.)*. Lawrence Erlbaum Associates.
- Congreso de la República del Perú. (2014). *Ley N.º 30220, Ley Universitaria*. Diario Oficial El Peruano. <https://diariooficial.elperuano.pe/normas/obtenerdocumento?idnorma=50>
- Congreso de la República del Perú. (2023). *Ley N.º 31814, Ley que promueve el uso de la inteligencia artificial en favor del desarrollo económico y social del país*. Diario Oficial El Peruano. <https://busquedas.elperuano.pe/dispositivo/NL/2192926-1>
- Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica. (2018). *Reglamento de calificación, clasificación y registro de los investigadores del Sistema Nacional de Ciencia, Tecnología e Innovación Tecnológica, RENACYT*. <https://repositorio.concytec.gob.pe/handle/20.500.12390/2897>
- Creswell, J. W., y Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches (5.ª ed.)*. SAGE Publications.
- Crompton, H., y Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, 20(1), 22. <https://doi.org/10.1186/s41239-023-00392-8>

- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. <https://doi.org/10.1007/BF02310555>
- Cuayáhuitl, H., Lee, D., Ryu, S., Choi, S., Hwang, I., y Kim, J. (2019). *Deep reinforcement learning for chatbots using clustered actions and human-likeness rewards*. En 2019 International Joint Conference on Neural Networks (IJCNN). IEEE. <https://doi.org/10.1109/ijcnn.2019.8852376>
- DeLone, W. H., y 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>
- EdLatam. (2025). *Políticas de inteligencia artificial en la educación superior latinoamericana: Análisis comparativo de México, Colombia y Chile*. Red EdLatam.
- EdTech Magazine. (2025, 29 de mayo). *How three universities developed their chatbots*. <https://edtechmagazine.com/higher/article/2025/05/how-three-universities-developed-their-chatbots>
- Es, S., James, J., Espinosa-Anke, L., y Schockaert, S. (2024). *RAGAS: Automated evaluation of retrieval augmented generation*. Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations, 150–158. <https://doi.org/10.18653/v1/2024.eacl-demo.16>
- Escurre, L. M. (1988). Cuantificación de la validez de contenido por criterio de jueces. *Revista de Psicología (PUCP)*, 6(1), 103–111.
- Espinosa-Luna, J. F., Castillo-Oliva, K. R., Montañez-Díaz, P. F., y Mendoza-De los Santos, A. C. (2023). Implementación de un chatbot soportado en GPT-3.5-Turbo como asistente académico en una universidad pública peruana. *Revista Científica de Sistemas e Informática*, 3(2), e570. <https://doi.org/10.51252/rcsi.v3i2.570>

- Essel, H. B., Vlachopoulos, D., Tachie-Menson, A., Johnson, E. E., y Baah, P. K. (2022). The impact of a virtual teaching assistant (chatbot) on students' learning in Ghanaian higher education. *International Journal of Educational Technology in Higher Education*, 19(1), 57. <https://doi.org/10.1186/s41239-022-00362-6>
- Faul, F., Erdfelder, E., Lang, A.-G., y Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics (5.ª ed.)*. SAGE Publications.
- Fielding, R. T. (2000). *Architectural styles and the design of network-based software architectures* [Tesis doctoral, University of California, Irvine]. <https://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm>
- Følstad, A., Araujo, T., Law, E. L.-C., Brandtzaeg, P. B., Papadopoulos, S., Reis, L., Baez, M., Laban, G., McAllister, P., Ischen, C., Wald, R., Catania, F., Meyer von Wolff, R., Hobert, S., y Luger, E. (2021). Future directions for chatbot research: An interdisciplinary research agenda. *Computing*, 103, 2915–2942. <https://doi.org/10.1007/s00607-021-01016-7>
- Gaitán Barrera, A., y Azeez, G. (2024, 8 de octubre). Educación 4.0: América Latina a la vanguardia de la IA universitaria. *Espacios de Educación Superior*. <https://www.espaciosdeeducacionsuperior.es/08/10/2024/educacion-4-0-america-latina-a-la-vanguardia-de-la-ia-universitaria/>
- Gans, N., Koole, G., y Mandelbaum, A. (2003). Telephone call centers: Tutorial, review, and research prospects. *Manufacturing & Service Operations Management*, 5(2), 79–141. <https://doi.org/10.1287/msom.5.2.79.16071>

- Gao, J., Galley, M., y Li, L. (2019). Neural approaches to conversational AI. *Foundations and Trends in Information Retrieval*, 13(2-3), 127–298.
<https://doi.org/10.1561/15000000074>
- Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M., y Wang, H. (2023). *Retrieval-augmented generation for large language models: A survey* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2312.10997>
- Gemini Team, Google. (2023). *Gemini: A family of highly capable multimodal models* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2312.11805>
- George, D., y Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference (4.^a ed.)*. Allyn & Bacon.
- Goodfellow, I., Bengio, Y., y Courville, A. (2016). *Deep learning*. MIT Press.
<https://www.deeplearningbook.org/>
- Google Site Reliability Engineering. (2023). *Site reliability engineering: How Google runs production systems*. <https://sre.google/books/>
- Han, S., y Lee, M. K. (2022). FAQ chatbot and inclusive learning in massive open online courses. *Computers & Education*, 179, 104395.
<https://doi.org/10.1016/j.compedu.2021.104395>
- Hardt, D. (2012). *The OAuth 2.0 authorization framework (RFC 6749)*. Internet Engineering Task Force. <https://datatracker.ietf.org/doc/html/rfc6749>
- Hendrycks, D., Burns, C., Basart, S., Zou, A., Mazeika, M., Song, D., y Steinhardt, J. (2021). Measuring massive multitask language understanding. En *International Conference on Learning Representations (ICLR 2021)*.
<https://doi.org/10.48550/arXiv.2009.03300>
- Hernández-Sampieri, R., y Mendoza, C. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*. McGraw-Hill.

- Infobae. (2025, 19 de octubre). *El 95 % de los universitarios peruanos ya usa inteligencia artificial para estudiar*. <https://www.infobae.com/peru/2025/10/19/el-95-de-los-universitarios-peruanos-ya-usa-inteligencia-artificial-para-estudiar/>
- Instituto Nacional de Estadística e Informática. (2025). *Estadísticas de las tecnologías de información y comunicación en los hogares: Trimestre octubre-noviembre-diciembre 2024* [Informe técnico].
https://www.inei.gob.pe/media/MenuRecursivo/boletines/informe-tecnico_tics_oct-nov-dic24.pdf
- International Organization for Standardization. (2016). *ISO 15489-1:2016 Information and documentation, Records management, Part 1: Concepts and principles*.
- International Organization for Standardization. (2022). *ISO 8000-2:2022 Data quality, Part 2: Vocabulary*.
- International Organization for Standardization, y International Electrotechnical Commission. (2018). *ISO/IEC 20000-1:2018 Information technology, Service management, Part 1: Service management system requirements*.
- International Organization for Standardization, y International Electrotechnical Commission. (2023). *ISO/IEC 25010:2023 Systems and software engineering, Systems and software Quality Requirements and Evaluation (SQuaRE), Product quality model*.
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., y Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), Article 248. <https://doi.org/10.1145/3571730>
- Jisc. (2025). *LearnWise end of pilot report*. National Centre for AI in Tertiary Education.
<https://nationalcentreforai.jiscinvolve.org/wp/2025/04/28/learnwise-end-of-pilot-report/>

- Johanson, G. A., y Brooks, G. P. (2010). Initial scale development: Sample size for pilot studies. *Educational and Psychological Measurement*, 70(3), 394–400.
<https://doi.org/10.1177/0013164409355692>
- Johnson, J., Douze, M., y Jégou, H. (2021). Billion-scale similarity search with GPUs. *IEEE Transactions on Big Data*, 7(3), 535–547.
<https://doi.org/10.1109/TBDATA.2019.2921572>
- Khennouche, F., Elmir, Y., Djebbari, N., y Boubchir, L. (2024). Improving user experience for AI chatbots through LLM models. En *Proceedings of the 21st ACS/IEEE International Conference on Computer Systems and Applications (AICCSA)*. IEEE.
<https://doi.org/10.1109/aiccsa63423.2024.10912523>
- Labadze, L., Grigolia, M., y Machaidze, L. (2023). Role of AI chatbots in education: Systematic literature review. *International Journal of Educational Technology in Higher Education*, 20(1), 56. <https://doi.org/10.1186/s41239-023-00426-1>
- Laudon, K. C., y Laudon, J. P. (2020). *Sistemas de información gerencial (15.ª ed.)*. Pearson Educación.
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., y Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. En *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22(140), 1–55. <https://archive.org/details/likert-1932>
- Lin, X., Chan, R. Y., Sharma, S., y Bista, K. (Eds.). (2024). *ChatGPT and global higher education: Using artificial intelligence in teaching and learning*. Star Scholars Press. <https://doi.org/10.32674/rh27qv16>
- Little, J. D. C. (1961). A proof for the queuing formula: $L = \lambda W$. *Operations Research*, 9(3), 383–387. <https://doi.org/10.1287/opre.9.3.383>

- Llama Team, AI @ Meta. (2024). *The Llama 3 herd of models* [Preprint]. arXiv.
<https://doi.org/10.48550/arXiv.2407.21783>
- Mann, H. B., y Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18(1), 50–60. <https://doi.org/10.1214/aoms/1177730491>
- Manning, C. D., Raghavan, P., y Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press. <https://nlp.stanford.edu/IR-book/>
- Mell, P., y Grance, T. (2011). *The NIST definition of cloud computing (NIST Special Publication 800-145)*. National Institute of Standards and Technology.
<https://doi.org/10.6028/NIST.SP.800-145>
- Meta Platforms. (2024). *WhatsApp Business Platform: Cloud API documentation*.
<https://developers.facebook.com/docs/whatsapp/cloud-api>
- Miao, F., y Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000386693>
- Neupane, S., Hossain, S., Keith, J., Tripathi, H., Ghiasi, F., Golilarz, N. A., Amirlatifi, A., Mittal, S., y Rahimi, S. (2024). *From questions to insightful answers: Building an informed chatbot for university resources* [Preprint]. arXiv.
<https://doi.org/10.48550/arXiv.2405.08120>
- Nielsen, J. (1993). *Usability engineering*. Morgan Kaufmann.
- Nunnally, J. C., y Bernstein, I. H. (1994). *Psychometric theory* (3.^a ed.). McGraw-Hill.
- Okonkwo, C. W., y Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence*, 2, 100033.
<https://doi.org/10.1016/j.caeai.2021.100033>

- Organismo Supervisor de Inversión Privada en Telecomunicaciones. (2026). *Encuesta Residencial de Servicios de Telecomunicaciones (ERESTEL) 2025*. OSIPTEL.
<https://repositorio.osiptel.gob.pe/handle/20.500.12630/327>
- Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. (2023). *ChatGPT and artificial intelligence in higher education: Quick start guide*.
<https://unesdoc.unesco.org/ark:/48223/pf0000385146>
- Otzen, T., y Manterola, C. (2017). Técnicas de muestreo sobre una población a estudio. *International Journal of Morphology*, 35(1), 227–232.
<https://doi.org/10.4067/S0717-95022017000100037>
- Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Aspell, A., Welinder, P., Christiano, P. F., Leike, J., y Lowe, R. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35, 27730–27744.
- Polit, D. F., y Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? *Research in Nursing & Health*, 29(5), 489–497.
<https://doi.org/10.1002/nur.20147>
- Presidencia del Consejo de Ministros del Perú. (2023). *Política Nacional de Transformación Digital al 2030*. Decreto Supremo N.º 085-2023-PCM. Diario Oficial El Peruano. <https://busquedas.elperuano.pe/dispositivo/NL/2200457-5>
- Programa de las Naciones Unidas para el Desarrollo. (2019). *El reto de la igualdad: Informe sobre desarrollo humano Perú 2019*. PNUD Perú.
<https://www.undp.org/es/peru/noticias/el-reto-de-la-igualdad-0>
- Quispe Cruz, G. A. (2024). *Chatbot y calidad de servicio de atención al egresado de la Universidad Peruana de Ciencias e Informática, Lima-2024* [Tesis de maestría,

- Universidad Peruana de Ciencias e Informática]. Repositorio Institucional UPCI.
<https://repositorio.upci.edu.pe/handle/upci/1163>
- Quispe Huacani, E. A. (2023). *Impacto de un chatbot en el proceso de atención de matrícula de estudiantes de la Escuela Profesional de Ingeniería en Informática y Sistemas de la Universidad Nacional Jorge Basadre Grohmann* [Tesis de pregrado, Universidad Nacional Jorge Basadre Grohmann]. Repositorio Institucional UNJBG. <https://repositorio.unjbg.edu.pe/handle/20.500.12510/3894>
- Razali, N. M., y 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.
https://www.researchgate.net/publication/267205556_Power_Comparisons_of_Shapiro-Wilk_Kolmogorov-Smirnov_Lilliefors_and_Anderson-Darling_Tests
- Rodriguez, S., y Mune, C. (2022). Uncoding library chatbots: Deploying a new virtual reference tool at the San Jose State University Library. *Reference Services Review*, 50(3-4), 392–405. <https://doi.org/10.1108/RSR-05-2022-0020>
- Rosenthal, R. (1991). *Meta-analytic procedures for social research (ed. rev.)*. SAGE Publications.
- Russell, S., y Norvig, P. (2021). *Artificial intelligence: A modern approach (4.ª ed.)*. Pearson.
- Schwaber, K., y Sutherland, J. (2020). *The Scrum Guide: The definitive guide to Scrum—The rules of the game*. Scrum.org. <https://scrumguides.org/scrum-guide.html>
- Shadish, W. R., Cook, T. D., y Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
- Superintendencia Nacional de Educación Superior Universitaria. (2020). *Resolución del Consejo Directivo N.º 031-2020-SUNEDU/CD: Otorgamiento de Licencia Institucional a la Universidad Tecnológica de los Andes*. Diario Oficial El Peruano.

<https://busquedas.elperuano.pe/normaslegales/otorgan-licencia-institucional-a-la-universidad-tecnologica-resolucion-n-031-2020-suneducd-1860426-1/>

- Tavakol, M., y Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55. <https://doi.org/10.5116/ijme.4dfb.8dfd>
- Topal, M. O., Bas, A., y van Heerden, I. (2024). *Frameworks for building large language model applications: A comparative analysis*. [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2401.06349>
- Torres, A. L. (2023, 21 de abril). Cuando Chat GPT entra a la clase: las universidades en Perú avanzan en un protocolo. *Gestión*. <https://gestion.pe/tecnologia/cuando-chat-gpt-entra-a-la-clase-las-universidades-en-peru-avanzan-en-un-protocolo-inteligencia-artificial-chatbots-sam-altman-utec-noticia/>
- Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., y Lample, G. (2023). *LLaMA: Open and efficient foundation language models* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2302.13971>
- Treviño, R. (2023, 29 de septiembre). *Tec, primera universidad en LATAM con inteligencia artificial propia*. Conecta Tec de Monterrey. <https://conecta.tec.mx/es/noticias/nacional/institucion/tec-primera-universidad-en-latam-con-inteligencia-artificial-propia>
- Treynor, B., Dahlin, M., Rau, V., y Beyer, B. (2017). The calculus of service availability. *Communications of the ACM*, 60(9), 42–47. <https://doi.org/10.1145/3127170>
- Universidad Tecnológica de los Andes. (2025). *Reporte estadístico de estudiantes matriculados, semestres 2025-II y 2026-I*. Dirección de Planeamiento y Desarrollo Universitario, UTEA.

- Vargas Cordero, Z. R. (2009). La investigación aplicada: Una forma de conocer las realidades con evidencia científica. *Revista Educación*, 33(1), 155–165.
<https://doi.org/10.15517/revedu.v33i1.538>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., y Polosukhin, I. (2017). Attention is all you need. En *Advances in Neural Information Processing Systems 30 (NIPS 2017)*.
- Veliz Yamunaque, F. D., y 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://hdl.handle.net/20.500.12893/13307>
- Wang, R. Y., y Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12(4), 5–33.
<https://doi.org/10.1080/07421222.1996.11518099>
- Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., Chi, E. H., Hashimoto, T., Vinyals, O., Liang, P., Dean, J., y Fedus, W. (2022). *Emergent abilities of large language models*. Transactions on Machine Learning Research.
<https://doi.org/10.48550/arXiv.2206.07682>
- Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6), 80–83. <https://doi.org/10.2307/3001968>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., . . . Mons, B. (2016). The FAIR Guiding Principles for scientific data

management and stewardship. *Scientific Data*, 3(1), 160018.

<https://doi.org/10.1038/sdata.2016.18>

Wu, R., y Yu, Z. (2024). Do AI chatbots improve students' learning outcomes? Evidence from a meta-analysis. *British Journal of Educational Technology*, 55(1), 10–33.

<https://doi.org/10.1111/bjet.13334>

Xu, B., Wu, L., Yuan, Y., Liu, J., y Zhang, Y. (2025). Exploring the application of AI chatbot tools in higher education: Evidence from the Duke University student survey. *Safety Emergency Science*, 1(2), 9590013.

<https://doi.org/10.26599/SES.2025.9590013>

Zapata Guerrero, D. C. (2023). *Chatbot para el acceso a información sobre trámites administrativos de los estudiantes de Ingeniería Informática de la Universidad Nacional de Piura* [Tesis de pregrado, Universidad Nacional de Piura]. Repositorio Institucional UNP. <http://repositorio.unp.edu.pe/handle/20.500.12676/4914>

Los anexos, panel fotográfico y otros documentos están resguardados en la oficina de repositorio digital institucional en la Biblioteca Central de la Universidad Tecnológica de los Andes