



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Facultad de Ingeniería de Sistemas e Informática

Escuela Profesional de Ingeniería de Sistemas

**Desarrollo de un Sistema de Inteligencia
Conversacional integrado con ChatGPT para
optimizar la atención virtual en la OCA de la UNMSM**

TESIS

Para optar el Título Profesional de Ingeniero de Sistemas

AUTOR

Brayan Esthiver CALCINA AGUILAR

Fernando Alberto GUTIERREZ MEJÍA

ASESOR

Dr. Hugo Florián VEGA HUERTA

Lima, Perú

2024



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Calcina, B. & Gutierrez, F. (2024). *Desarrollo de un Sistema de Inteligencia Conversacional integrado con ChatGPT para optimizar la atención virtual en la OCA de la UNMSM*. [Tesis de pregrado, Universidad Nacional Mayor de San Marcos, Facultad de Ingeniería de Sistemas e Informática, Escuela Profesional de Ingeniería de Sistemas]. Repositorio institucional Cybertesis UNMSM.

Metadatos complementarios autor/ asesor

Datos de autor	
Nombres y apellidos	Fernando Alberto Gutierrez Mejía
Tipo de documento de identidad	DNI
Número de documento de identidad	71474431
URL de ORCID	https://orcid.org/0000-0002-2519-0766
Datos de autor	
Nombres y apellidos	Brayan Esthiver Calcina Aguilar
Tipo de documento de identidad	DNI
Número de documento de identidad	73398726
URL de ORCID	https://orcid.org/0009-0007-3799-9646
Datos de asesor	
Nombres y apellidos	Hugo Froilán Vega Huerta
Tipo de documento de identidad	DNI
Número de documento de identidad	06147737
URL de ORCID	https://orcid.org/0000-0002-4268-5808
Datos del jurado	
Presidente del jurado	
Nombres y apellidos	Ana María Huayna Dueñas
Tipo de documento	DNI
Número de documento de identidad	06017183
Miembro del jurado 1	
Nombres y apellidos	Juan Gamarra Moreno
Tipo de documento	DNI
Número de documento de identidad	20039857
Datos de investigación	

Línea de investigación	Inteligencia Artificial
Grupo de investigación	Innovando Sistemas Inteligentes (YACHAY)
Agencia de financiamiento	-
Ubicación geográfica de la investigación	País: Perú Departamento: Lima Provincia: Lima Distrito: Lima
Año o rango de años en que se realizó la investigación	2023-2024
URL de disciplinas OCDE	Ciencias de la computación https://purl.org/pe-repo/ocde/ford#1.02.01



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
FACULTAD DE INGENIERIA DE SISTEMAS E INFORMATICA
Escuela Profesional de Ingeniería de Sistemas

Acta Virtual de Sustentación de Tesis

Siendo las 15:00 horas del día 30 de octubre del año 2024, se reunieron virtualmente los docentes designados como miembros de Jurado de Tesis, presidido por la Mg. Ana María Huayna Dueñas, Mg. Juan Gamarra Moreno (Miembro) y el Dr. Hugo F. Vega Huerta (Miembro Asesor), usando la plataforma Meet (meet.google.com/uat-xpkv-pap), para la sustentación Virtual de la tesis Intitulada: **“Desarrollo de un Sistema de Inteligencia Conversacional integrado con ChatGPT para optimizar la atención virtual en la OCA de la UNMSM”**, del Bachiller Brayan Esthiver CALCINA AGUILAR; para obtener el Título Profesional de Ingeniero de Sistemas.

Acto seguido de la exposición de la Tesis, la Presidenta invitó al Bachiller a responder las preguntas formuladas por los Miembros del Jurado.

El Bachiller, en el curso de su intervención demostró pleno dominio del tema, al responder con acierto y fluidez las preguntas formuladas por los señores miembros del Jurado.

Finalmente habiéndose efectuado la calificación correspondiente por los miembros del Jurado, el bachiller obtuvo la nota de 14 (Catorce)

A continuación, la Presidenta del Jurado Mg. Ana María Huayna Dueñas declara al Bachiller **Ingeniero de Sistemas**.

Siendo 16:20 horas, se levantó la sesión.

111 

Mg. Juan Gamarra Moreno
Miembro


Mg. Ana María Huayna Dueñas
Presidenta


Dr. Hugo F. Vega Huerta
Miembro Asesor



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
FACULTAD DE INGENIERIA DE SISTEMAS E INFORMATICA
Escuela Profesional de Ingeniería de Sistemas

Acta Virtual de Sustentación de Tesis

Siendo las 15:00 horas del día 30 de octubre del año 2024, se reunieron virtualmente los docentes designados como miembros de Jurado de Tesis, presidido por la Mg. Ana María Huayna Dueñas, Mg. Juan Gamarra Moreno (Miembro) y el Dr. Hugo F. Vega Huerta (Miembro Asesor), usando la plataforma Meet (meet.google.com/uat-xpkv-pap), para la sustentación Virtual de la tesis Intitulada: **“Desarrollo de un Sistema de Inteligencia Conversacional integrado con ChatGPT para optimizar la atención virtual en la OCA de la UNMSM”**, del Bachiller: Fernando Alberto GUTIERREZ MEJÍA; para obtener el Título Profesional de Ingeniero de Sistemas.

Acto seguido de la exposición de la Tesis, la Presidenta invitó al Bachiller a responder las preguntas formuladas por los Miembros del Jurado.

El Bachiller, en el curso de su intervención demostró pleno dominio del tema, al responder con acierto y fluidez las preguntas formuladas por los señores miembros del Jurado.

Finalmente habiéndose efectuado la calificación correspondiente por los miembros del Jurado, el bachiller obtuvo la nota de 14 (Catorce)

A continuación, la Presidenta del Jurado Mg, Ana María Huayna Dueñas declara al Bachiller **Ingeniero de Sistemas**.

Siendo 16:20: horas, se levantó la sesión.


Mg. Ana María Huayna Dueñas
Presidenta


Mg. Juan Gamarra Moreno
Miembro


Dr. Hugo F. Vega Huerta
Asesor



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Vicerrectorado de Investigación y Posgrado



Yo **HUGO F. VEGA HUERTA** en mi condición de **asesor** acreditado con la Resolución
Directoral N° 000013-2023-EPISI-FISI/UNMSM de la **tesis/monografía/informe** de
investigación/trabajo académico, cuyo título es "**Desarrollo de un Sistema de Inteligencia
Conversacional integrado con ChatGPT para optimizar la atención virtual en la
OCA de la UNMSM**", presentado por los **bachilleres/magíster/egresado/licenciado/estudiante
Fernando Alberto GUTIERREZ MEJÍA y Brayan Esthiver CALCINA AGUILAR**, para
optar el grado/**título**/especialidad de **INGENIERO DE SISTEMAS**, CERTIFICO que se ha
cumplido con lo establecido en la Directiva de Originalidad y de Similitud de Trabajos
Académicos, de Investigación y Producción Intelectual. Según la revisión, análisis y evaluación
mediante el software de similitud textual, el documento evaluado cuenta con el porcentaje de **9%**
de similitud, nivel **PERMITIDO** para continuar con los trámites correspondientes y para su
publicación en el repositorio institucional.

Se emite el presente certificado en cumplimiento de lo establecido en las normas vigentes, como
uno de los requisitos para la obtención del grado/ título/ especialidad correspondiente.

Firma del Asesor

DNI: 06147737

Nombres y apellidos del asesor: Hugo Froilán Vega Huerta

DEDICATORIA

A Dios por brindarme la fortaleza en estos momentos tan difíciles de mi vida, a mi madre por su incondicional apoyo y amor constante. A mi papá Memo, a mi hermano Nazee, a mi primo Stefano, el jacobiano mayor. Y a cada miembro de mi familia, por estar siempre a mi lado en cada paso del camino.

Fernando Gutierrez

DEDICATORIA

Primero a Dios por darme la fuerza de llegar hasta estos momentos de mi vida que llenan de felicidad a todas las personas que amo. A mi madre quién gracias a sus cuidados y consejos me permitieron superarme. A mi padre por el apoyo incondicional. A mi hermano José Calcina quién me apoyó desde los inicios de mi formación profesional y que desde el cielo siempre me cuida. A mis hermanas y hermano a quiénes amo mucho y a Edith, mi compañera de vida, quién me impulsa a ser cada vez mejor.

Brayan Calcina

AGRADECIMIENTOS

A Dios por concedernos la sabiduría, constancia, fortaleza, dedicación y compromiso necesarios para llevar a cabo esta tesis.

A nuestras familias, por ser nuestro pilar fundamental, por comprender cada desvelo y por estar presentes en cada lágrima derramada mientras redactábamos cada párrafo.

A nuestro asesor, el Dr. Hugo Vega Huerta, por su tiempo, su apoyo incondicional, su disposición y su valiosa mentoría.

RESUMEN

La presente tesis aborda la problemática asociada con los métodos tradicionales de acceso a la información sobre el proceso de admisión en la UNMSM, los cuales a menudo resultan en tiempos de respuesta prolongados y una experiencia de usuario insatisfactoria. Aprovechando la oportunidad de mejorar estos aspectos, se desarrolló un Sistema de Inteligencia Conversacional basado en ChatGPT. Este sistema fue diseñado para proporcionar respuestas instantáneas y detalladas a consultas sobre pagos, fechas de exámenes y documentación requerida, optimizando la experiencia de los aspirantes y modernizando la atención virtual en la OCA.

A lo largo del proyecto, se abordaron y resolvieron los objetivos específicos establecidos, desde la investigación exhaustiva y recopilación de información relevante, hasta el diseño e implementación de un modelo de procesamiento del lenguaje natural que garantiza respuestas precisas y contextualmente relevantes. Los resultados obtenidos tras la implementación del sistema mostraron mejoras significativas en términos de tiempo de atención, satisfacción del usuario (NPS), y precisión de las respuestas, destacando un 95% de veracidad en la información proporcionada por el chatbot.

Finalmente, el sistema demostró su eficacia en optimizar la comunicación y accesibilidad, transformando la manera en que los postulantes interactúan con la OCA. Con la implementación exitosa del chatbot, se sentaron las bases para futuras expansiones y mejoras, que podrían incluir la automatización de actualizaciones y otras funcionalidades avanzadas para mantener la relevancia y precisión del sistema.

Palabras clave: Chatbot, Inteligencia Conversacional, ChatGPT, Admisión universitaria, Optimización de atención virtual.

ABSTRACT

This thesis addresses the issues associated with traditional methods of accessing information about the admission process at UNMSM, which often result in extended response times and an unsatisfactory user experience. Seizing the opportunity to improve these aspects, a Conversational Intelligence System based on ChatGPT was developed. This system was designed to provide instant and detailed responses to inquiries about payments, exam dates, and required documentation, optimizing the applicant experience and modernizing virtual assistance at the (OCA).

Throughout the project, the specific objectives were addressed and achieved, from thorough research and information gathering to the design and implementation of a natural language processing model that ensures accurate and contextually relevant responses. The results obtained from the system's implementation showed significant improvements in response time, user satisfaction (NPS), and response accuracy, with a 95% accuracy rate in the information provided by the chatbot.

Finally, the system proved effective in optimizing communication and accessibility, transforming the way applicants interact with the OCA. The successful implementation of the chatbot laid the groundwork for future expansions and improvements, which could include the automation of updates and other advanced features to maintain the system's relevance and accuracy.

Keywords: Chatbot, Conversational Intelligence, University Admissions, ChatGPT, Virtual Assistance Optimization

ÍNDICE

DEDICATORIA	ii
AGRADECIMIENTOS	iii
RESUMEN	iv
ABSTRACT.....	v
CAPÍTULO I: INTRODUCCIÓN	11
1.1. Antecedentes de la problemática.....	11
1.2. Definición del problema.....	12
1.2.1 Problema principal	12
1.2.2 Problemas específicos	12
1.3. Justificación de la Investigación	13
1.4. Alcances y Limitaciones de la Investigación	14
1.5. Objetivos.....	14
1.5.1 Objetivo general	14
1.5.2 Objetivos específicos	15
1.6. Propuesta de la tesis	15
1.7. Organización de la tesis	15
CAPÍTULO II: MARCO TEÓRICO	18
2.1. Inteligencia Conversacional	18
2.1.1. Tecnologías subyacentes.....	18
2.1.2. Interacción usuario-sistema.....	19
2.1.3. Impacto en la atención virtual	19
2.2. Procesamiento del Lenguaje Natural (NLP)	20
2.2.1. Técnicas del NLP	22
2.2.2. Aplicaciones en la industria del software.....	23
2.3. ChatGPT	24
2.3.1. Arquitectura transformer.....	25
2.3.2. Evolución del modelo GPT	26
2.3.3. Aplicaciones y limitaciones	27
CAPÍTULO III: ESTADO DEL ARTE	29
3.1. Evolución de los modelos de lenguaje en inteligencia conversacional.....	29
3.1.1. Problemas resueltos y mejora continua en modelos de lenguaje	29
3.1.2. Transición hacia modelos generativos preentrenados y su impacto	30
3.1.3. Desafíos persistentes y limitaciones actuales.....	31
3.2. Integración de modelos de lenguaje en plataformas de gestión de contenidos	33
3.2.1. Retos en la personalización y automatización del contenido.....	33

3.2.2.	Optimización de la atención virtual en entornos educativos.....	35
3.2.3.	Casos de éxito y mejores prácticas en la implementación de modelos de lenguaje.....	38
3.3.	Despliegue de modelos de inteligencia artificial en la nube	40
3.3.1.	Desafíos de seguridad y privacidad en entornos cloud	40
3.3.2.	Optimización de costos y recursos	41
3.4.	Análisis comparativo de soluciones de inteligencia conversacional.....	42
3.4.1.	Criterios de evaluación.....	42
3.4.2.	Asignación de pesos a los criterios	43
3.4.3.	Benchmarking de soluciones	44
3.4.4.	Estudios recientes sobre plataforma de inteligencia conversacional.....	45
CAPÍTULO IV: METODOLOGÍA		47
4.1	Propuesta de la investigación.....	47
4.1.1	Inteligencia Conversacional Generativa	48
4.1.2	Variables de configuración de Azure OpenAI Service	49
4.2	Diseño y desarrollo del sistema	50
4.2.1	Estructura general de la implementación del sistema	50
4.2.2	Arquitectura del sistema.....	52
4.2.3	Descripción de las tecnologías Azure	54
4.3	Entrenamiento y optimización del modelo	57
4.3.1	Proceso de Entrenamiento del Modelo de ChatGPT.....	57
4.3.2	Ajustes y Optimización del Modelo.....	62
4.3.3	Configuración del Prompt de Comportamiento	65
4.4	Integración con la OCA	67
4.4.1	Análisis de Requerimientos	67
4.4.2	Integración con WordPress	73
CAPÍTULO V: DESCRIPCIÓN DEL SISTEMA.....		76
5.1	Requerimientos del sistema	76
5.1.1	Requerimientos Funcionales del Sistema.....	76
5.1.2	Requerimientos no Funcionales del Sistema.....	77
5.2	Casos de Uso del Sistema	78
5.2.1	Actores del Sistema.....	78
5.2.2	Diagramas de Casos de Uso del Sistema	78
5.2.3	Especificaciones de Casos de Uso del Sistema.....	80
5.3	Interfaz del sistema	85
5.3.1	Interfaz web de la aplicación	85
CAPÍTULO VI: RESULTADOS		89
6.1	Resultados de la propuesta.....	89

6.1.1	Evaluación del Tiempo de Atención.....	89
6.1.2	Evaluación del NPS del usuario.....	91
6.1.3	Evaluación de la calidad de los resultados	96
CAPÍTULO VII: CONCLUSIONES Y TRABAJOS FUTUROS.....		103
7.1	Conclusiones	103
7.2	Trabajos futuros	103
BIBLIOGRAFÍA		105

ÍNDICE DE FIGURAS

Figura 1. Clasificación general del NLP	22
Figura 2. Metodologías 4PADAFE.....	36
Figura 3. Arquitectura del sistema	37
Figura 4. Arquitectura del chatbot	39
Figura 5. Arquitectura en Microsoft Azure.....	40
Figura 6. Representación gráfica de la metodología propuesta	47
Figura 7. Interfaz web del chatbot	52
Figura 8. Infraestructura del Sistema Inteligente Conversacional	53
Figura 9. Arquitectura de los recursos Azure del chatbot.....	57
Figura 10. Configuración del Azure Storage Account.....	58
Figura 11. Configuración del índice para la búsqueda vía Azure IA Search.....	59
Figura 12. Configuración del indexador para la actualización automática.....	60
Figura 13. Configuración de Azure OpenAI Service.....	61
Figura 14. Configuración del proyecto	62
Figura 15. Página web de la OCA integrada con el chatbot	74
Figura 16. Plugin para integrar el chatbot a la OCA.....	75
Figura 17. Interacción del Usuario (Estudiante/Público en general) con el sistema	79
Figura 18. Interacción del Administrador del Sistema con el Sistema	80
Figura 19. Pantalla inicial del chatbot.....	85
Figura 20. Interacción entre el Usuario y el chatbot.....	86
Figura 21. Subir nuevo documento no estructurado como base de conocimiento.....	87
Figura 22. Captura que muestra la configuración del prompt de comportamiento	88
Figura 23. Tiempo promedio de atención por método de consulta.....	90

ÍNDICE DE TABLAS

Tabla 1. Puntajes asignados a los criterios de evaluación	43
Tabla 2. Evaluación comparativa de plataformas de inteligencia conversacional.....	45
Tabla 3. Aptitudes para ajustar el prompt de comportamiento.	65
Tabla 4. Requerimientos funcionales para la integración con la OCA.....	68
Tabla 5. Requerimientos no funcionales para la integración con la OCA.....	71
Tabla 6. Requerimientos funcionales del Sistema.	76
Tabla 7. Requerimientos no funcionales del Sistema.	77
Tabla 8. Actores del Sistema.	78
Tabla 9. Especificaciones del caso de uso del sistema “Realizar consulta al chatbot”.	80
Tabla 10. Especificaciones del caso de uso del sistema “Visualizar respuesta del chatbot”.	81
Tabla 11. Especificaciones del caso de uso del sistema “Limpiar historial de conversación”.	82
Tabla 12. Especificaciones del caso de uso del sistema “Actualizar base de conocimientos”.	83
Tabla 13. Especificaciones del caso de uso del sistema “Modificar Prompt de comportamiento”.	84
Tabla 14. Resultado de la medición de NPS.....	94
Tabla 15. Categoría x Porcentaje de Usuarios.	95
Tabla 16. Análisis de respuesta según la Matriz de Confusión.	100

CAPÍTULO I: INTRODUCCIÓN

1.1. Antecedentes de la problemática

El proceso actual de obtención de información para el ingreso a la Universidad Nacional Mayor de San Marcos (UNMSM) presenta desafíos significativos, especialmente en cuanto a la accesibilidad y la claridad de la información proporcionada por la Oficina Central de Admisión (OCA). A menudo, los aspirantes a la universidad se ven obligados a recurrir a métodos tradicionales, como llamadas telefónicas o visitas presenciales, para obtener detalles sobre el proceso de admisión. Estos métodos resultan ineficientes y consumen una cantidad considerable de tiempo, con tiempos de espera prolongados que pueden durar de una a varias horas.

La información disponible en línea a menudo se percibe como poco clara, lo que genera confusión y la necesidad de buscar ayuda adicional.

Además, se destaca la dificultad que enfrentan los aspirantes que residen en provincia, quienes encuentran especialmente difícil realizar llamadas o visitas presenciales debido a las limitaciones geográficas y los largos tiempos de espera. La necesidad de obtener información precisa sobre el proceso de admisión se convierte, en este contexto, en una barrera significativa para aquellos que no pueden acceder fácilmente a los canales de comunicación existentes.

Estos resultados destacan la urgencia de implementar una solución que aborde de manera integral la ineficiencia en la entrega de información, mejorando la accesibilidad y claridad para todos los aspirantes, independientemente de su ubicación o medio de comunicación preferido.

Es así como los sistemas conversacionales se vuelven una solución eficaz y eficiente en base a esta problemática. Además, es importante destacar que la evolución de los chatbots ha transformado la forma en que interactuamos con ellos. La llegada de los sistemas inteligentes conversacionales, impulsados por motores avanzados como ChatGPT, marca un cambio

significativo en la naturaleza de estas interacciones. A diferencia de los chatbots tradicionales que operaban dentro de límites predeterminados y menú de opciones, los sistemas inteligentes conversacionales tienen la capacidad de comprender de manera más amplia y natural el lenguaje humano.

El motor ChatGPT permite que estos sistemas aprendan y generen respuestas contextualmente relevantes, lo que los convierte en herramientas más flexibles y adaptativas. Esta evolución tecnológica es esencial para abordar la complejidad de las consultas que los aspirantes pueden tener sobre el proceso de admisión. Así, la implementación de un sistema de inteligencia conversacional no solo representa una mejora en la eficiencia y accesibilidad de la información, sino que también refleja la adaptación a las innovaciones tecnológicas actuales en el campo de la interacción conversacional.

1.2. Definición del problema

1.2.1 Problema principal

- La problemática central se centra en la oportunidad de optimizar la entrega de información sobre el proceso de admisión de la UNMSM, ya que la dependencia de métodos tradicionales y la falta de claridad en la web generan tiempos de espera prolongados y dificultan el acceso a información vital, especialmente para aspirantes en provincias.

1.2.2 Problemas específicos

- Los canales actuales, como las llamadas telefónicas y visitas presenciales, resultan en tiempos de espera prolongados, reduciendo la eficiencia en la entrega de información y afectando negativamente la experiencia del aspirante.
- La estructura y presentación de la información en la web no son lo suficientemente claras para los aspirantes, lo que incrementa la necesidad de buscar ayuda adicional a través de

canales presenciales o telefónicos.

- La ausencia de un sistema conversacional avanzado limita la capacidad de la OCA para brindar respuestas rápidas y precisas, lo cual incrementa la carga en los canales tradicionales y genera insatisfacción en los usuarios.

1.3. Justificación de la Investigación

En la actualidad, el proceso de admisión a la UNMSM es crucial para los estudiantes que aspiran a continuar su educación superior. Sin embargo, la distribución de información vital sobre el proceso, como conceptos de pago, fechas de exámenes y documentación necesaria, presenta desafíos significativos. La dependencia de llamadas telefónicas, visitas presenciales y la navegación a través de la web puede resultar en una experiencia fragmentada y poco clara para los aspirantes.

La relevancia de abordar este problema radica en la mejora de la accesibilidad y eficiencia en la entrega de información a los estudiantes. El objetivo no es solucionar una ineficiencia existente, sino elevar la experiencia general de los aspirantes, permitiéndoles acceder de manera rápida y sencilla a información precisa y detallada sobre el proceso de admisión.

Al optimizar la llegada de información a través de un sistema de inteligencia conversacional integrado con ChatGPT, se busca proporcionar respuestas instantáneas a las consultas de los estudiantes. Esta mejora no solo simplificará el proceso para los aspirantes, sino que también fomentará una comprensión más completa y clara de los requisitos de admisión.

La importancia de esta iniciativa reside en su capacidad para fortalecer la comunicación entre la UNMSM y los estudiantes, creando una experiencia más accesible y amigable. Al facilitar el acceso a información detallada de manera rápida, se contribuye a empoderar a los aspirantes, permitiéndoles tomar decisiones informadas y cumplir con los requisitos de admisión de manera

más efectiva. En última instancia, la mejora en la entrega de información beneficia tanto a la institución educativa como a los estudiantes, construyendo un proceso de admisión más transparente, eficiente y centrado en el usuario.

1.4. Alcances y Limitaciones de la Investigación

En la era digital actual, la eficiencia y efectividad de los servicios de atención al cliente son elementos cruciales para el éxito de cualquier institución. La OCA enfrenta una creciente demanda de atención virtual, lo que requiere soluciones innovadoras para mejorar la experiencia del usuario y optimizar la gestión de consultas.

En este contexto, se hace necesario desarrollar un sistema de inteligencia conversacional integrado con ChatGPT. La implementación de esta tecnología avanzada no solo facilitará una interacción más fluida y natural entre los usuarios y la OCA, sino que también permitirá liberar recursos humanos para que se enfoquen en áreas más complejas y especializadas. De esta manera, la UNMSM se posicionará como pionera en la adopción de soluciones tecnológicas avanzadas, elevando la calidad de sus servicios y estableciendo un estándar en atención al cliente.

El proyecto será trabajado bajo el modelo gpt-35-turbo-16k de Azure OpenAI Service bajo la versión de API 2023-06-01-preview, será integrado bajo la infraestructura tecnológica de la OCA utilizando las bibliotecas y frameworks que soporten dicha configuración y será configurado en base al estimado de consultas que provendrán de la página web de la OCA. Esta estimación deberá ser revisada por el equipo de datos de la OCA y cruzada con nosotros, para que, en caso de que sea necesario, se realice una actualización para cubrir un mayor número de consultas.

1.5. Objetivos

1.5.1 Objetivo general

- Desarrollar un sistema de inteligencia conversacional integrado con ChatGPT para

optimizar la atención virtual en la OCA de la UNMSM.

1.5.2 *Objetivos específicos*

- Reducir los tiempos de espera y la carga en los canales tradicionales de atención al aspirante a través de la automatización de respuestas a preguntas frecuentes y consultas comunes.
- Integrar el sistema conversacional en los canales digitales de la OCA, asegurando una experiencia fluida y eficiente en la interacción virtual, alineada con las innovaciones tecnológicas en el ámbito de la atención al usuario.
- Implementar un sistema conversacional que brinde respuestas claras y contextualizadas a las consultas de los aspirantes, utilizando el motor ChatGPT para mejorar la comprensión y precisión de la información.

1.6. Propuesta de la tesis

La presente propuesta de tesis tiene como objetivo principal desarrollar un Sistema de Inteligencia Conversacional integrado con ChatGPT para mejorar la entrega de información en el proceso de admisión de la UNMSM. A través de un análisis exhaustivo de las necesidades de los aspirantes y de los elementos clave del proceso de admisión, se buscará implementar una solución tecnológica avanzada que proporcione respuestas instantáneas y detalladas a consultas específicas. La investigación se enfocará en la integración efectiva de la tecnología, la optimización de la interfaz de usuario y la evaluación continua del sistema, con el objetivo de modernizar la atención virtual en la OCA de la UNMSM y mejorar la experiencia de los aspirantes en su búsqueda de información sobre el proceso de admisión.

1.7. Organización de la tesis

La presente tesis para fines de estructuración está dividida en 7 capítulos, los cuales se

mencionan a continuación:

- En el Capítulo 1 se presenta el proyecto de investigación ofreciendo una visión general del problema de acceso a la información sobre el proceso de admisión en la UNMSM seguido de la descripción de los objetivos generales y específicos, y el alcance del proyecto.
- El Capítulo 2 está dedicado a la revisión de conceptos y fundamentos teóricos relacionados con la inteligencia conversacional, el procesamiento del lenguaje natural y las tecnologías empleadas en el desarrollo del sistema como ChatGPT, proporcionando una base sólida para la comprensión de los aspectos técnicos y metodológicos de la tesis.
- En el Capítulo 3 se exponen los antecedentes de investigaciones y aplicaciones de sistemas de inteligencia conversacional en diferentes sectores con un enfoque especial en su implementación en el ámbito educativo, lo que permite contextualizar la solución propuesta dentro de un marco más amplio de innovación tecnológica.
- El Capítulo 4 describe la metodología para el diseño y desarrollo del sistema de inteligencia conversacional basado en ChatGPT partiendo desde la recopilación y normalización de la información hasta la integración del sistema con la OCA de la UNMSM. Este capítulo está estructurado en secciones que explican el proceso de entrenamiento, optimización e integración técnica del sistema.
- El Capítulo 5 se enfoca en la descripción técnica del sistema abordando los requerimientos funcionales y no funcionales, los casos de uso y la interfaz del sistema, proporcionando una visión completa de la arquitectura y funcionamiento del chatbot.
- En el Capítulo 6 se presentan y analizan los resultados obtenidos tras la implementación del sistema en la OCA. Se examinan aspectos clave como la reducción en los tiempos de atención, la mejora en la satisfacción del usuario (NPS) y la precisión de las respuestas

proporcionadas por el chatbot.

- Finalmente, el Capítulo 7 está dedicado a las conclusiones de la tesis donde se sintetizan los logros en relación con los objetivos planteados y los trabajos futuros, destacando la contribución del proyecto a la modernización de la atención virtual en la OCA de la UNMSM.

CAPÍTULO II: MARCO TEÓRICO

Este capítulo aborda y analiza en detalle los conceptos clave y las tecnologías esenciales, como la inteligencia conversacional, el procesamiento del lenguaje natural y el modelo ChatGPT. Estos temas son fundamentales para entender el contexto de la problemática planteada en la tesis y resultan cruciales para el desarrollo de la propuesta de solución.

2.1. Inteligencia Conversacional

La inteligencia conversacional se define como la capacidad de un sistema para interactuar con usuarios a través del lenguaje natural, permitiendo que la comunicación se establezca de manera fluida.

Según Varela-Tapia et al. (2021), la inteligencia conversacional se apoya en modelos de procesamiento del lenguaje natural y aprendizaje automático, lo que facilita la comprensión y la respuesta adecuada a las consultas del usuario en tiempo real.

A lo largo del tiempo, la inteligencia conversacional se ha convertido en una herramienta clave para mejorar la experiencia del cliente y la eficiencia operativa de las instituciones, destacando su impacto en la optimización de la atención al cliente a gran escala, como indican Carlsson y Jacobs (2021).

Artiles et al. (2021) destacan que la inteligencia conversacional también permite una interacción natural y adaptativa, lo que mejora la capacidad de los sistemas para responder de manera efectiva en diferentes contextos.

2.1.1. Tecnologías subyacentes

Las tecnologías subyacentes en la inteligencia conversacional incluyen una serie de algoritmos y modelos avanzados que permiten el análisis y la generación de respuestas en lenguaje natural.

Varela-Tapia et al. (2021) mencionan que herramientas como Dialogflow utilizan el modelo BERT (Bidirectional Encoder Representations from Transformer) para el análisis e interpretación de las expresiones del usuario, lo que facilita que las interacciones sean más precisas y personalizadas.

Además, Carlsson y Jacobs (2021) subrayan la importancia del análisis de voz y texto para extraer información procesable de las conversaciones, lo que contribuye tanto a mejorar la experiencia de los usuarios como a la eficiencia operativa.

2.1.2. Interacción usuario-sistema

La interacción usuario-sistema es fundamental en la inteligencia conversacional, ya que determina la calidad y relevancia de las respuestas que el sistema proporciona al usuario.

En los sistemas de inteligencia conversacional, es crucial contar con una interfaz intuitiva y una comunicación natural y fluida, ya que estos elementos mejoran la experiencia del usuario y facilitan la interacción (Mekni, 2021).

Hadfi et al. (2021) también indican que, en la inteligencia conversacional, los agentes conversacionales pueden ser diseñados para participar en discusiones argumentativas, lo que amplía las capacidades de interacción más allá de las simples respuestas.

2.1.3. Impacto en la atención virtual

El impacto de la inteligencia conversacional en la atención virtual ha sido significativo, ya que ha transformado la forma en la que las organizaciones interactúan con sus clientes o usuarios, al ofrecer un servicio más eficiente y personalizado.

Se destaca que la implementación de sistemas con inteligencia conversacional no solo reduce los tiempos de espera, sino que también proporcionan respuestas precisas y en tiempo real, lo que es fundamental para aumentar la satisfacción de los clientes como mencionan Varela-Tapia

et al. (2021).

Mekni (2021) resalta cómo la automatización de tareas rutinarias mediante la inteligencia conversacional no solo reduce costos, sino que también permite a los agentes humanos centrarse en interacciones más complejas y de mayor valor, mejorando la atención general.

2.2. Procesamiento del Lenguaje Natural (NLP)

El procesamiento del lenguaje natural, también conocido como NLP, es un campo dentro de la inteligencia artificial cuyo objetivo es que las máquinas puedan comprender, interpretar y generar lenguaje humano de manera natural y eficiente. Este campo es crucial para el desarrollo de sistemas que permiten interactuar de forma fluida y efectiva con las personas, usuarios o clientes, utilizando el lenguaje cotidiano.

Manjarrés-Betancur y Echeverri-Torres (2020) señalan que el NLP es esencial para la automatización de tareas complejas que, en tiempos anteriores, requerían intervención humana directa. Ejemplos de estas tareas incluyen la clasificación de textos, la traducción automática y la generación de lenguaje natural.

La capacidad del NLP para procesar grandes volúmenes de datos es fundamental en distintos tipos de aplicaciones o sistemas, ya que estos pueden servir para la toma de decisiones estratégicas.

López de Lacalle et al. (2023) enfatizan que el uso correcto de técnicas avanzadas de NLP permite a las organizaciones extraer o recopilar información útil y procesable, lo cual es crucial para optimizar procesos y mejorar la toma de decisiones.

A lo largo del tiempo, el desarrollo del NLP ha avanzado considerablemente, permitiendo a las máquinas comprender mejor como están estructuradas las palabras, así como el contexto de estas. Esto resulta esencial para automatizar tareas sofisticadas, como la generación automática de

texto y la personalización de respuestas en asistentes virtuales (González & Sánchez, 2021).

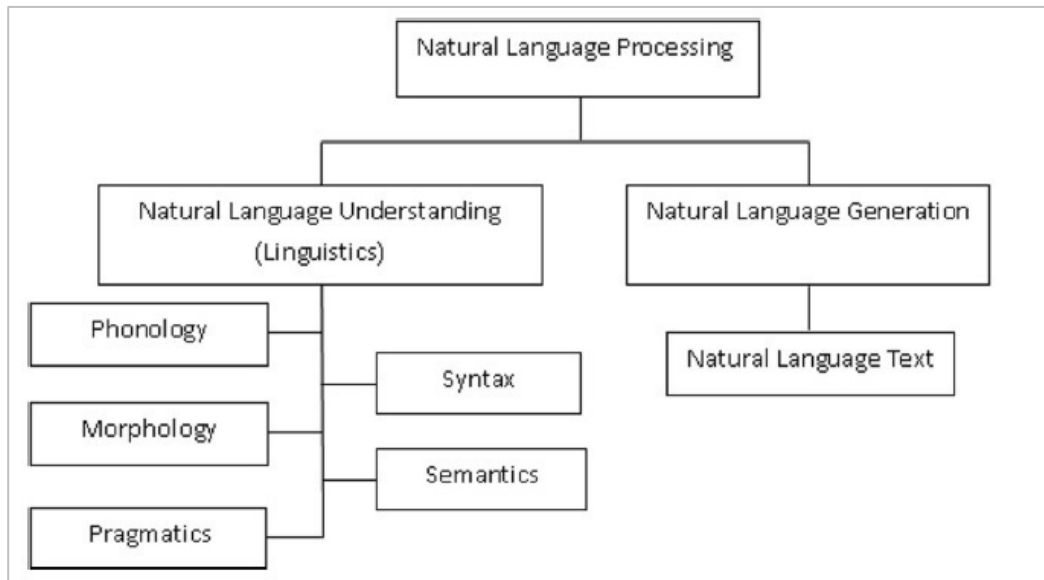
El NLP se divide principalmente en dos áreas:

- **Comprensión del lenguaje natural (NLU):** Se enfoca en la interpretación del significado del lenguaje humano. Este proceso implica una serie de tareas, como el análisis sintáctico, el análisis semántico y la desambiguación léxica. Khurana et al. (2023) mencionan que el NLU es fundamental para que las máquinas comprendan el contenido y el contexto de las palabras y frases. Este es un aspecto clave para aplicaciones como el análisis de sentimientos y la extracción de información en textos complejos.
- **Generación del lenguaje natural (NLG):** Se encarga de la creación de lenguaje humano comprensible por parte de las máquinas. Este proceso incluye la planificación del contenido, la estructuración de las frases y la realización del texto final. López de Lacalle et al. (2023) señalan que el NLG es importante para diversas aplicaciones, como los asistentes virtuales y la generación de reportes automatizados, donde se necesita producir respuestas coherentes que se adapten al contexto de la conversación.

Como se muestra en la Figura 1, el NLP se divide en dos grandes áreas: la comprensión y la generación de lenguaje natural. Estas áreas son claves para el desarrollo de aplicaciones de inteligencia artificial que interactúan con el lenguaje humano.

Figura 1.

Clasificación general del NLP



Fuente: *Khurana et al. (2023)*

2.2.1. Técnicas del NLP

Las técnicas del NLP forman la base del análisis y generación de lenguaje humano por parte de las máquinas. Su evolución a lo largo del tiempo ha permitido a los sistemas mejorar significativamente su capacidad para comprender y generar lenguaje de manera más precisa.

Una técnica fundamental en el NLP es la recopilación y anotación de corpus, una tarea crítica porque proporciona los datos necesarios para entrenar modelos supervisados de aprendizaje automático, que son la base en muchas aplicaciones de NLP. Garcia-Diaz et al. (2020) explican que la recopilación de corpus implica reunir grandes volúmenes de texto que representen el lenguaje que se desea modelar. Estos textos pueden provenir de diversas fuentes, como artículos científicos, redes sociales y documentos legales, dependiendo del objetivo de la aplicación de NLP. Una vez recopilados, los textos deben ser anotados. Esto significa que se les deben asignar etiquetas o metadatos que identifiquen ciertas características específicas. Esto es esencial, ya que

la anotación de corpus permite entrenar modelos avanzados para realizar tareas específicas de manera efectiva.

El desarrollo de modelos de aprendizaje profundo, como GPT (Generative Pre-trained Transformer), ha revolucionado el campo del NLP, marcando un hito en las técnicas de NLP. Este tipo de modelo, que también es la base de ChatGPT, permite a los sistemas procesar no solo las palabras de manera aislada, sino también analizar el contexto completo en el que se encuentran dentro de una oración o párrafo. Esta capacidad de contextualizar mejora de manera exponencial la relevancia y precisión de las respuestas generadas por los sistemas de NLP, un punto destacado por Martínez et al. (2021).

Celi-Parraga et al. (2021) mencionan que existen diversas técnicas para descomponer y analizar el lenguaje natural con precisión, lo que permite a los sistemas de NLP interpretar correctamente la estructura gramatical de las oraciones. Entre ellas, destacan la minería de texto y el procesamiento morfosintáctico, que son útiles para el análisis de textos complejos, como documentos legales y artículos científicos, donde la precisión y claridad en la interpretación son cruciales.

2.2.2. Aplicaciones en la industria del software

El NLP ha encontrado aplicaciones cruciales en la industria del software. Garcia-Diaz et al. (2020) señalan que el NLP permite automatizar la detección y extracción de requisitos de software a partir de documentos textuales. Esta automatización reduce significativamente los errores humanos y mejora la precisión en la documentación de especificaciones durante la etapa de análisis en el desarrollo de software, siendo esta una de las aplicaciones más destacadas del NLP en este ámbito.

Martínez et al. (2021) destacan que el NLP también facilita la identificación de

ambigüedades, contradicciones y omisiones en los documentos de requisitos. Esta capacidad permite a los desarrolladores resolver estos problemas antes de que se conviertan en fallas costosas en las etapas posteriores del desarrollo del software.

Garcia-Diaz et al. (2020) resaltan que la integración de NLP en los procesos de desarrollo de software no solo mejora la comunicación entre desarrolladores y clientes, sino que también asegura que los productos de software cumplan con las expectativas y necesidades del usuario final. El NLP también se ha utilizado en la generación automática de documentación técnica, en la creación de asistentes virtuales para la asistencia técnica, y en la mejora de la interacción entre los equipos de desarrollo a través de herramientas colaborativas que comprenden y procesan el lenguaje natural.

2.3. ChatGPT

ChatGPT es un modelo de lenguaje avanzado desarrollado por la empresa OpenAI, basado en la arquitectura Generative Pre-trained Transformer (GPT). Este modelo ha transformado el campo del NLP, ya mencionado anteriormente, al permitir que las máquinas generen respuestas coherentes, relevantes y adecuadas al contexto en diversas interacciones. Según Arévalo y Quinde (2023), ChatGPT es capaz de comprender y generar lenguaje humano de manera fluida, lo que ha sido esencial para su adopción en aplicaciones que van desde la atención al cliente hasta la automatización de tareas cognitivas complejas. Este modelo no solo facilita la comunicación entre humanos y máquinas, sino que también ha abierto nuevas posibilidades en áreas como la educación, la creación de contenido y la investigación científica.

A lo largo de su desarrollo, ChatGPT ha demostrado ser una herramienta versátil y poderosa, capaz de adaptarse a distintos contextos y necesidades. Limna et al. (2023) destacan que la arquitectura subyacente de GPT permite que el modelo procese grandes volúmenes de datos

textuales, aprendiendo patrones del lenguaje y aplicándolos para generar respuestas que son prácticamente indistinguibles de las que podrían ofrecer los humanos. Esto ha sido clave para la implementación de ChatGPT en sectores donde la precisión y la coherencia en la comunicación son fundamentales, como el soporte técnico y también la asesoría virtual.

Además, Hariri (2024) señala que ChatGPT se ha integrado de manera exitosa con otras tecnologías, lo que ha permitido su aplicación en sistemas interactivos y plataformas de productividad, mejorando la eficiencia y efectividad en diversas operaciones empresariales. Esta integración no solo amplía las capacidades de ChatGPT, sino que también facilita su adopción en una variedad de entornos donde la interacción humana es crítica.

2.3.1. Arquitectura transformer

La arquitectura transformer, que sustenta ChatGPT, es una de las innovaciones más significativas en el campo del NLP. Arévalo y Quinde (2023) explican que la arquitectura transformer se basa en un mecanismo de atención que permite a los modelos enfocarse dinámicamente en diferentes partes de la entrada, mejorando así la calidad y coherencia de las respuestas generadas. Este mecanismo es crucial para manejar secuencias de texto largas y complejas, permitiendo que ChatGPT pueda generar texto de manera fluida y coherente.

Una de las ventajas clave de la arquitectura transformer es su capacidad para procesar secuencias de texto en paralelo, a diferencia de las redes neuronales recurrentes (RNN) tradicionales, que procesan las secuencias de manera secuencial. Hariri (2024) destaca que este enfoque paralelo no solo mejora la eficiencia del procesamiento, sino que también permite escalar los modelos a miles de millones de parámetros, como es el caso de GPT-3 y GPT-4, que son la base de ChatGPT. Estos modelos, con su extensa capacidad de procesamiento, hacen que la generación de texto sea ideal para aplicaciones que requieren alta calidad en la generación de

lenguaje natural.

Además, Limna et al. (2023) mencionan que la arquitectura transformer utiliza codificaciones posicionales para mantener la coherencia y el orden en el texto generado, lo que es crucial para garantizar que las respuestas de ChatGPT sean lógicas y coherentes dentro del contexto de la conversación. Esta capacidad para mantener la coherencia a lo largo de largas secuencias de texto ha sido fundamental para la adopción de ChatGPT en aplicaciones como la atención al cliente y la generación automática de contenido.

2.3.2. Evolución del modelo GPT

El modelo GPT ha evolucionado significativamente desde su primera versión, con cada iteración aportando mejoras en la capacidad de generación de texto y la comprensión contextual. Aktay et al. (2023) describen cómo GPT-1, con 117 millones de parámetros, estableció las bases para el desarrollo de modelos más avanzados, demostrando que los modelos preentrenados podían ser mejorados para realizar una variedad de tareas de NLP con alta precisión. Esta evolución continuó con GPT-2, que, con 1.5 mil millones de parámetros, amplió la capacidad del modelo para manejar tareas mucho más complejas, como la generación automática de texto largo y coherente.

La revolución más importante llegó con GPT-3, que, con 175 mil millones de parámetros, superó ampliamente a los modelos anteriores, mejorando de manera significativa la capacidad de generar respuestas precisas y relevantes. Esta capacidad ha llevado a la adopción generalizada de GPT-3 en aplicaciones que requieren interacciones naturales y fluidas con los usuarios, como en asistentes virtuales y sistemas de recomendación.

Limna et al. (2023) añaden que GPT-4, lanzado en 2023, ha introducido mejoras significativas no solo en el número de parámetros, sino también en la forma en que el modelo

maneja tareas especializadas y mantiene la coherencia en largas conversaciones. Estas mejoras han permitido que GPT-4, y por extensión ChatGPT, se integren con éxito en herramientas de productividad, facilitando la comunicación y colaboración en entornos laborales complejos.

2.3.3. *Aplicaciones y limitaciones*

ChatGPT ha encontrado aplicaciones en una amplia gama de sectores, desde la educación hasta la atención al cliente, gracias a su capacidad para generar respuestas naturales, claras y en el contexto adecuado. Gill et al. (2024) señalan que ChatGPT ha sido particularmente efectivo en la educación, donde actúa como un tutor virtual que ayuda a los estudiantes a comprender conceptos complejos y mejorar su aprendizaje de manera personalizada. Este uso de ChatGPT en la educación ha permitido a los estudiantes acceder a un aprendizaje de alta calidad desde cualquier lugar y en cualquier momento.

En el ámbito empresarial, ChatGPT se ha utilizado para automatizar la atención al cliente, proporcionando respuestas inmediatas y precisas a las consultas de los usuarios o clientes. Sin embargo, Gill et al. (2024) también identifican limitaciones importantes en el uso de ChatGPT, como la posibilidad de generar respuestas incorrectas o sesgadas, lo que podría afectar negativamente la experiencia del usuario y llevar a la difusión de información errónea. Estas limitaciones subrayan la importancia de supervisar y ajustar constantemente las respuestas generadas por ChatGPT para garantizar su precisión y relevancia.

Rahman y Watanobe (2023) agregan que, aunque ChatGPT es capaz de generar respuestas altamente contextuales, carece de un mecanismo para garantizar la veracidad y sensibilidad en todas las interacciones, lo que puede ser problemático en aplicaciones sensibles como la atención médica y el asesoramiento legal. Esta falta de control para gestionar la incertidumbre en la generación de respuestas limita el uso de ChatGPT en situaciones donde la precisión y la ética son

cruciales, lo que destaca la necesidad de integrar mecanismos adicionales de verificación y validación en estas aplicaciones.

CAPÍTULO III: ESTADO DEL ARTE

En este capítulo, se revisarán los avances más relevantes en el ámbito de la inteligencia conversacional, con un enfoque en tres áreas principales: la evolución de los modelos de lenguaje, la integración de estos modelos en plataformas de gestión de contenidos y su despliegue en infraestructuras en la nube. Se analizarán las problemáticas abordadas por diversos autores y cómo sus soluciones han contribuido al desarrollo de este campo, con especial atención a las aplicaciones educativas y la optimización de la atención virtual.

3.1. Evolución de los modelos de lenguaje en inteligencia conversacional

En este apartado se examina la evolución de los modelos de lenguaje utilizados en inteligencia conversacional, teniendo en cuenta los problemas que se han ido resolviendo, la transición hacia modelos generativos preentrenados, así como las limitaciones y desafíos actuales que aún persisten. Esta revisión permitirá comprender cómo estos avances tecnológicos han impactado en el desarrollo de sistemas conversacionales más avanzados y efectivos, y de igual manera, identificar las áreas que requieren más investigación para seguir optimizando estos modelos.

3.1.1. Problemas resueltos y mejora continua en modelos de lenguaje

Khurana et al. (2023) destacan la evolución significativa en el campo del NLP, donde modelos como BERT han sido cruciales para mejorar la comprensión y generación de lenguaje en diversas aplicaciones. Este avance ha permitido superar muchas limitaciones en tareas de NLP, facilitando la comprensión y generación de respuestas más precisas y contextualmente adecuadas.

Ali Abdulridha et al. (2024) subrayan cómo las limitaciones regionales en el acceso a servicios como Microsoft Azure han influido en la implementación de chatbots, obligando a los desarrolladores a buscar alternativas. En este contexto, el uso de bibliotecas de código abierto

como LangChain ha sido una solución eficaz para superar las barreras técnicas y geográficas. Esta flexibilidad en la elección de herramientas demuestra un avance continuo en la adopción de la inteligencia artificial en diversos contextos, asegurando que las soluciones de inteligencia artificial sean accesibles y efectivas, independientemente de las restricciones locales.

Estos avances son relevantes para el desarrollo de la tesis, ya que se está utilizando tecnología basada en ChatGPT, que se apoya en modelos preentrenados como BERT para mejorar la interacción con los usuarios, al mismo tiempo que se consideran soluciones alternativas para superar posibles limitaciones técnicas y geográficas.

3.1.2. Transición hacia modelos generativos preentrenados y su impacto

La evolución hacia modelos generativos preentrenados ha marcado un antes y un después en el desarrollo de la inteligencia conversacional. El uso de BERT como modelo preentrenado ha demostrado mejorar significativamente la precisión en tareas de aclaración de significados, una solución identificada como crucial para avanzar en el campo del NLP (Khurana et al., 2023).

De manera similar, el avance significativo que representa el uso de los grandes modelos de lenguaje (LLMs) como ChatGPT y GPT-3 en la generación automática de visualizaciones a partir de consultas en lenguaje natural (Maddigan & Susnjak, 2023) ha superado los enfoques tradicionales que dependían, en cierta medida, de reglas específicas de dominio.

Además, el uso de LLMs no se limita solo a la generación de texto, sino que también ha demostrado ser una herramienta poderosa en la automatización de configuraciones de infraestructura, como la generación de infraestructura como código (IaC). Ganesh (2024) destaca cómo los LLMs pueden automatizar la creación de configuraciones complejas en múltiples proveedores de nube, optimizando procesos que antes requerían un esfuerzo significativo. Esta capacidad de generar código para infraestructura de manera automatizada subraya el potencial de

los LLMs para revolucionar no solo la gestión de infraestructuras, sino también la inteligencia conversacional.

Finalmente, Valeri (2022) destaca la importancia de la flexibilidad y adaptabilidad de los LLMs en la integración con sistemas complejos, permitiendo una mejor interacción entre sistemas y una mayor capacidad de adaptación a contextos cambiantes. Este enfoque permite a los LLMs incrementar su eficiencia en la generación de soluciones automatizadas, tanto en el ámbito de la infraestructura como en la personalización de las respuestas conversacionales.

Este enfoque es esencial para la tesis, ya que el sistema de inteligencia conversacional que se está desarrollando se basa en la capacidad de estos modelos para generar respuestas contextualmente relevantes, lo que mejora la precisión y utilidad de las respuestas para la OCA, además de considerar cómo estos avances pueden aplicarse en la optimización de otros procesos tecnológicos.

3.1.3. Desafíos persistentes y limitaciones actuales

A pesar de los avances, el NLP aún enfrenta desafíos significativos. Aún con el progreso en modelos como BERT y las redes neuronales recurrentes, la aclaración de significados y la generación coherente de lenguaje sigue siendo retos importantes, lo que resalta la necesidad de más investigaciones en estas áreas (Khurana et al., 2023).

Esto es evidente en aplicaciones específicas como la medicina, donde Suárez et al. (2024) mencionan que, aunque ChatGPT muestra un rendimiento prometedor en cirugía oral, su precisión está limitada por la posibilidad de generar respuestas incorrectas, subrayando la necesidad de una supervisión rigurosa en aplicaciones clínicas.

Estos desafíos resaltan la importancia de implementar mecanismos de validación y supervisión en el desarrollo del sistema propuesto en la tesis, para garantizar la precisión y

fiabilidad de las respuestas generadas por el sistema de inteligencia conversacional.

Dwivedi et al. (2023) resaltan que, aunque ChatGPT tiene un enorme potencial, su uso también presenta importantes desafíos éticos y de privacidad, así como el riesgo de perpetuar sesgos y generar desinformación. Esto subraya la necesidad de desarrollar políticas responsables y un marco ético sólido.

En el contexto educativo, Tlili et al. (2023) subrayan que ChatGPT presenta oportunidades significativas, pero también desafíos éticos y de calidad de las respuestas que deben abordarse para garantizar un uso responsable. Además, Hassani y Silva (2023) destacan que, aunque ChatGPT tiene un gran potencial para automatizar flujos de trabajo en ciencia de datos, persisten preocupaciones significativas relacionadas con el sesgo y la precisión de sus resultados, lo que conlleva la necesidad de un uso ético y supervisado. La consideración de estos desafíos juega un rol crucial en el desarrollo de la tesis, ya que el sistema debe ser capaz de manejar consultas educativas de manera ética y precisa, minimizando o evitando sesgos que puedan afectar la experiencia de los usuarios.

García-Peñalvo et al. (2024) subrayan que, aunque la inteligencia artificial generativa ofrece enormes ventajas en la personalización del aprendizaje, también presenta desafíos significativos relacionados con la privacidad, la equidad y la precisión de la información, lo que requiere un enfoque equilibrado en su implementación en entornos educativos.

En el ámbito médico, los LLMs como GPT-4 presentan desafíos únicos en comparación con tecnologías de inteligencia artificial previamente reguladas, debido a su escala y complejidad. Estos desafíos incluyen la privacidad de los datos, el sesgo, y la equidad en la atención médica, lo que conlleva la necesidad de un marco regular específico (Meskó & Topol, 2023).

El estudio de Au Yeung et al. (2023) subrayan que, aunque los LLMs como ChatGPT han

avanzado significativamente en la generación de texto comprensible y relevante, aún presentan limitaciones críticas en la precisión de diagnósticos clínicos y están sujetos a sesgos y alucinaciones, lo que indica que no están listos para su uso generalizado. Este análisis es relevante para la tesis, ya que resalta la importancia de evitar la aplicación indiscriminada de ChatGPT en contextos críticos como el clínico, en su lugar, adaptarlo con cautela al contexto educativo para garantizar la precisión y relevancia de las respuestas.

Finalmente, Scanlon et al. (2023) sostienen que ChatGPT es útil en tareas básicas de forense digital, pero también enfrenta limitaciones significativas en la precisión y fiabilidad de sus respuestas, especialmente en contextos especializados como la identificación de artefactos y la generación de scripts para la recuperación de datos.

Este análisis de limitaciones técnicas también es importante para la tesis, ya que destaca la necesidad de un diseño cuidadoso y una validación constante del sistema para asegurar que las respuestas sean confiables en el contexto de la OCA.

3.2. Integración de modelos de lenguaje en plataformas de gestión de contenidos

En esta sección se analizará la aplicación de modelos de lenguaje en la personalización y automatización de contenidos dentro de diversas plataformas de gestión, con un enfoque especial en entornos educativos. Se abordarán los desafíos en la personalización de contenido, cómo se ha optimizado la atención virtual en estos entornos, y se presentarán casos de éxitos que ilustran las mejores prácticas en la implementación de esta tecnología.

3.2.1. Retos en la personalización y automatización del contenido

En el ámbito educativo, la personalización de contenidos mediante inteligencia artificial generativa presenta tanto oportunidades como desafíos. Ruiz-Rojas et al. (2023) sugieren que la metodología 4PADAFE, al combinarse con herramientas de inteligencia artificial generativa,

puede superar estos desafíos al proporcionar una estructura clara y adaptativa para el diseño de actividades curriculares.

Este enfoque es directamente aplicable a la tesis, ya que se busca personalizar la experiencia de los aspirantes de la UNMSM a través de un sistema de inteligencia conversacional que se adapte a sus necesidades individuales.

Sin embargo, García-Peñalvo et al. (2024) advierten que es necesario ajustar los contenidos para evitar sesgos y garantizar la precisión en la información proporcionada. Además de la personalización textual, la integración de tecnologías como las descritas por Bumann (2023) ofrece una dimensión adicional de personalización, al permitir que los estudiantes interactúen con las plataformas educativas a través de comandos de voz. Estas tecnologías no solo mejoran la accesibilidad, sino que también permiten una interacción más natural y fluida, optimizando la experiencia del usuario en entornos de aprendizaje en línea.

La personalización también debe considerar la formación en ética digital, especialmente en la implementación de herramientas como ChatGPT en la educación, para garantizar la equidad y la inclusión en el aprendizaje (Michel-Villarreal et al., 2023). En contextos más amplios, como la generación de contenido personalizado, Taecharungroj (2023) revela que la escritura creativa y la generación de contenido mediante ChatGPT pueden enfrentar desafíos en términos de originalidad y ética.

En aplicaciones científicas y técnicas, los LLMs muestran un gran potencial para personalizar contenido, aunque enfrentan desafíos relacionados con la robustez y la mitigación de sesgos, lo que es crucial para asegurar una experiencia de usuario equitativa y confiable (Chang et al., 2024). De manera similar, la importancia del diseño de prompts en la personalización de las respuestas de LLMs es fundamental para asegurar que las visualizaciones generadas sean

coherentes con las consultas del usuario, incluso en presencia de especificaciones incompletas o ambiguas (Maddigan & Susnjak, 2023).

Por otro lado, en el ámbito médico, aunque la capacidad de ChatGPT para personalizar respuestas clínicas es prometedora, está limitada por su incapacidad para realizar análisis clínicos complejos, lo que afecta la precisión y relevancia de las recomendaciones (Casella et al., 2023).

3.2.2. Optimización de la atención virtual en entornos educativos

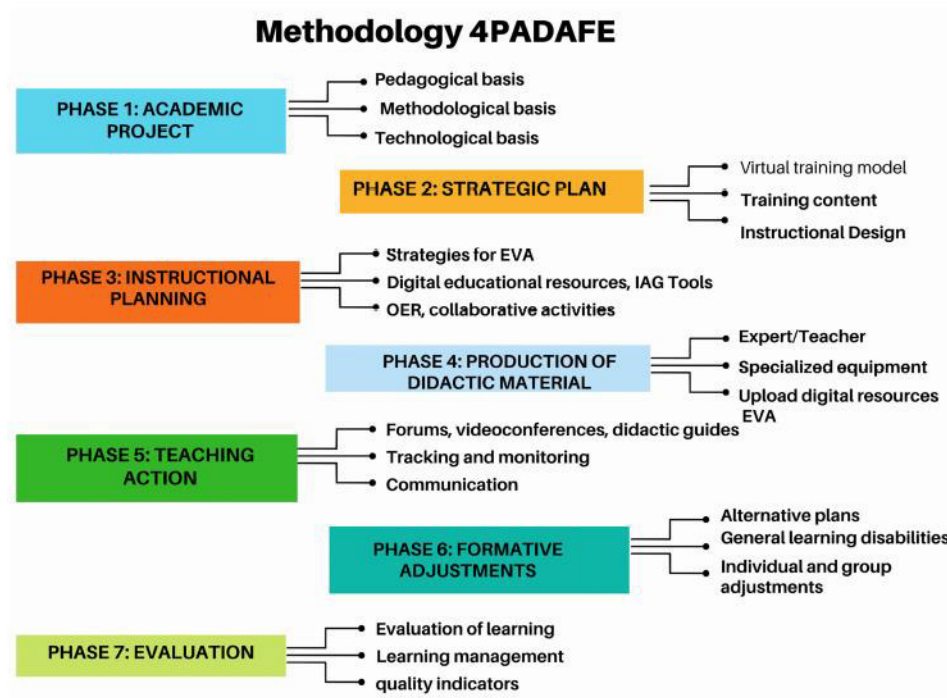
La integración de tecnologías generativas de inteligencia artificial en la educación superior puede mejorar significativamente los resultados de aprendizaje, siempre que se implementen políticas adecuadas que consideren tanto los beneficios como los riesgos éticos y operativos (Chan, 2023). El uso de técnicas avanzadas de NLP, como se describe en el trabajo de Khurana et al. (2023), tiene el potencial de optimizar la atención virtual en entornos educativos, mejorando la interacción y la personalización de los contenidos ofrecidos a los estudiantes.

Asimismo, Dwivedi et al. (2023) destacan que ChatGPT puede mejorar la productividad en la investigación y la enseñanza, pero también advierten sobre su impacto potencialmente disruptivo en la evaluación del aprendizaje y la autenticidad académica. La aplicación de herramientas de inteligencia artificial generativa en combinación con la matriz de diseño instruccional 4PADAFE permite personalizar el aprendizaje en entornos educativos masivos, mejorando la interacción y la motivación de los estudiantes (Ruiz-Rojas et al., 2023).

La Figura 2 muestra la metodología 4PADAFE, la cual estructura y optimiza la personalización del aprendizaje mediante la inteligencia artificial generativa. Esta metodología es clave para mejorar la experiencia educativa de los estudiantes en entornos masivos.

Figura 2.

Metodologías 4PADAFE



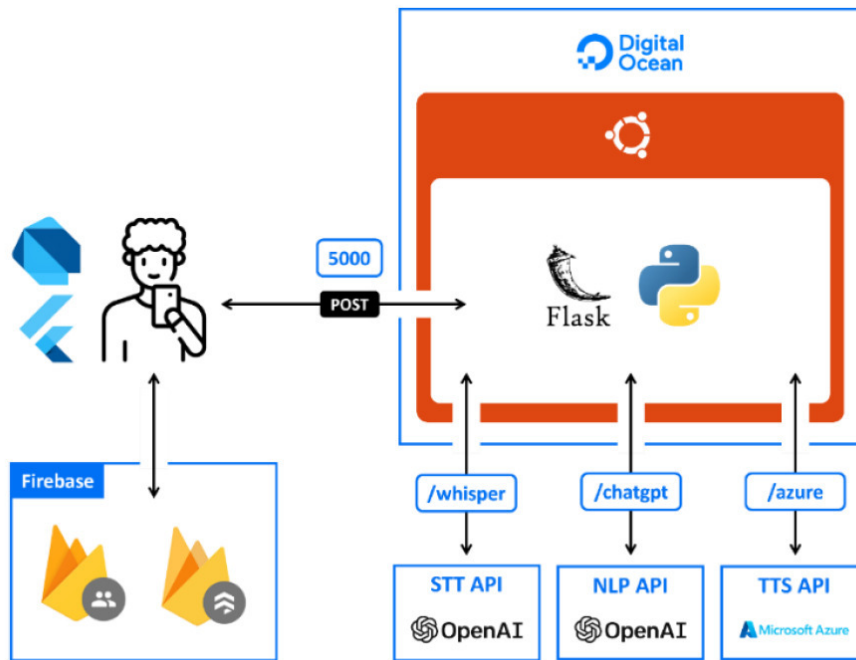
Fuente: *Ruiz-Rojas et al. (2023)*

Además, Bumann (2023) demuestra que la implementación de su trabajo puede optimizar aún más la atención virtual al permitir que los estudiantes interactúen con las plataformas educativas sin necesidad de texto escrito. Esto es especialmente útil en entornos de aprendizaje donde la accesibilidad es clave, permitiendo que estudiantes con discapacidades visuales o motoras se beneficien de la tecnología.

La Figura 3 ilustra la arquitectura del sistema que permite esta interacción accesible y eficiente en plataformas educativas, optimizando la atención virtual para un público más diverso.

Figura 3.

Arquitectura del sistema



Fuente: *Bumann (2023)*

El éxito de los agentes conversacionales en la tutoría de estudiantes universitarios también es evidente, proporcionando soluciones escalables y efectivas para la gestión de grandes volúmenes de estudiantes (Artiles et al., 2021). Estos resultados apoyan el desarrollo del sistema conversacional en la OCA, que debe ser capaz de manejar eficientemente un gran volumen de consultas. Finalmente, Morales-Chan et al. (2024) demostraron que la integración de inteligencia artificial generativa para proporcionar retroalimentación personalizada en MOOCs puede optimizar significativamente la experiencia de aprendizaje de los estudiantes, mejorando su compromiso y satisfacción. Esto es particularmente relevante en entornos educativos masivos donde la atención individualizada es un desafío.

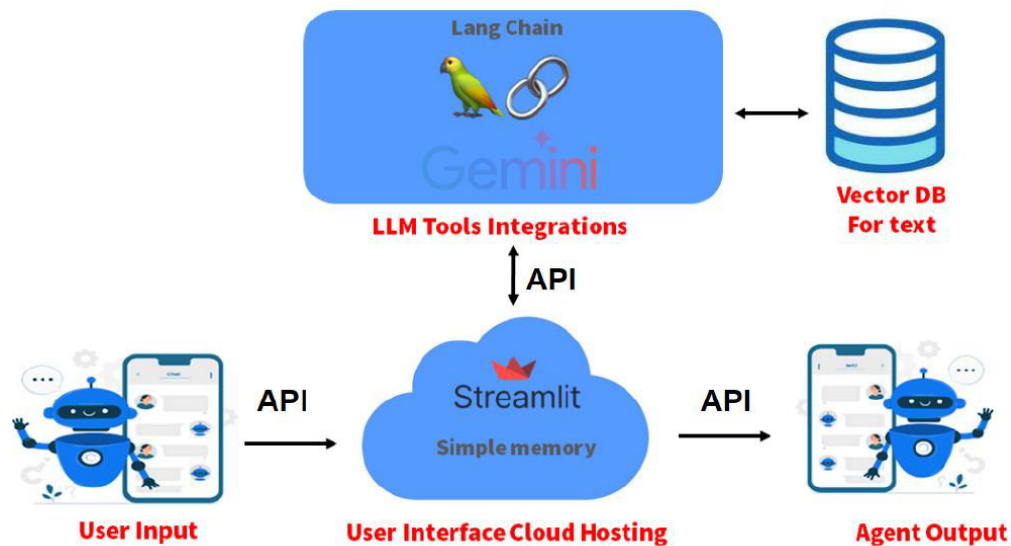
3.2.3. Casos de éxito y mejores prácticas en la implementación de modelos de lenguaje

Según lo descrito por Artiles et al. (2021), el éxito del chatbot CLOE en la Universidad de Las Palmas de Gran Canaria sirve como un ejemplo de buenas prácticas en la implementación de agentes conversacionales en la educación superior. Este caso demuestra cómo un chatbot bien diseñado puede mejorar significativamente la enseñanza y el aprendizaje al proporcionar respuestas rápidas y precisas a las consultas de los estudiantes. La implementación de CLOE no solo optimizó la interacción entre estudiantes y la universidad, sino que también mostró la importancia de adaptar la tecnología a las necesidades específicas de un entorno educativo, asegurando que los sistemas conversacionales no solo respondan de manera efectiva, sino también lo hagan de forma ética y segura.

De manera similar, el desarrollo exitoso del chatbot Esaalni, utilizando LangChain y el LLM Gemini Pro, es otro claro ejemplo de cómo las soluciones de código abierto pueden superar las limitaciones técnicas y regionales. Ali Abdulridha et al. (2024) destacan que la flexibilidad y adaptabilidad de estas herramientas permitieron una implementación eficaz, asegurando que el chatbot pudiera operar de manera eficiente a pesar de las restricciones en el acceso a servicios como Azure. En la Figura 4, se presenta la arquitectura del chatbot Esaalni, destacando su capacidad para funcionar en contextos con limitaciones tecnológicas.

Figura 4.

Arquitectura del chatbot



Fuente: *Ali Abdulridha et al. (2024)*

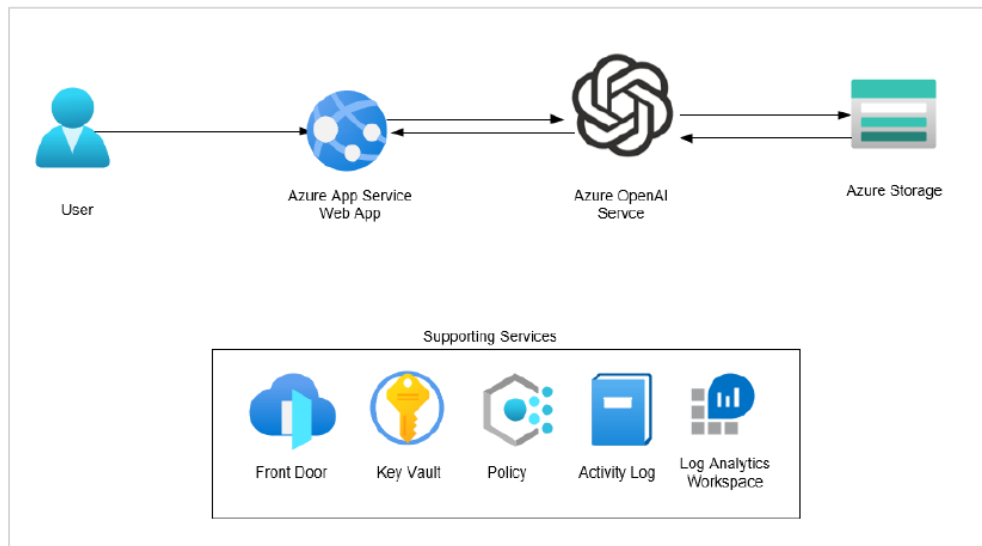
En un contexto empresarial, la implementación de un chatbot en una aerolínea mexicana para mejorar la atención al cliente también ha demostrado ser un caso de éxito significativo. Según Garibay (2020), el chatbot desarrollado para Club Premier, el programa de recompensas de Aeroméxico, permitió una mejora notable en la interacción con los clientes a través de canales conversacionales. La tecnología utilizada por el chatbot fue capaz de comprender las consultas de los usuarios y generar las respuestas adecuadas, lo que redujo significativamente el tiempo de espera y aumentó la satisfacción del cliente.

Por otro lado, la implementación de modelos de lenguaje en plataformas de gestión de contenidos ha sido crucial en la optimización de procesos. Estos casos destacan la importancia de seguir mejores prácticas en la implementación de estos sistemas, incluyendo la necesidad de garantizar la privacidad y seguridad de los datos de los usuarios. Ots (2024) subraya la relevancia de aplicar metodologías de seguridad establecidas al implementar Azure OpenAI en entornos

empresariales. La Figura 5 ilustra la arquitectura de un sistema basado en Microsoft Azure, mostrando cómo se puede integrar la inteligencia artificial en plataformas empresariales de manera segura y eficiente.

Figura 5.

Arquitectura en Microsoft Azure



Fuente: *Ots (2024)*

3.3. Despliegue de modelos de inteligencia artificial en la nube

En esta sección, se explorarán los aspectos críticos del despliegue de modelos de inteligencia artificial en infraestructura en la nube, destacando los desafíos de seguridad y privacidad, así como las estrategias para optimizar costos y recursos. Se revisan casos y mejores prácticas en la implementación de tecnologías de inteligencia artificial en la nube, proporcionando una visión integral de cómo gestionar eficazmente la implementación de estos sistemas a gran escala.

3.3.1. *Desafíos de seguridad y privacidad en entornos cloud*

La seguridad y privacidad son consideraciones clave en el despliegue de inteligencia

artificial generativa en la nube. Chan (2023) resalta la importancia de abordar estas cuestiones al implementar inteligencia artificial en contextos educativos, subrayando la necesidad de políticas claras que garanticen la protección de datos y la transparencia en el uso de estas tecnologías.

En un contexto más amplio Dwivedi et al. (2023) destacan la adopción de marcos de gestión de riesgos, como el AI RMF desarrollado por NIST, como cruciales para mitigar los riesgos asociados con la implementación de inteligencia artificial generativa en entornos cloud. De manera similar, García-Peñalvo et al. (2024) enfatizan la necesidad de establecer políticas claras y marcos éticos para mitigar los riesgos relacionados con la privacidad y la seguridad en entornos educativos, sugiriendo la implementación de auditorías de inteligencia artificial y la creación de políticas educativas específicas.

En el ámbito médico, es fundamental desarrollar regulaciones que aborden la seguridad y privacidad en el uso de LLMs. Meskó y Topol (2023) proponen directrices específicas para el anonimato de datos, la encriptación y el almacenamiento seguro, protegiendo así la información sensible de los pacientes.

3.3.2. Optimización de costos y recursos

El marco propuesto por Chan (2023) sugiere que, además de los beneficios pedagógicos, la implementación de inteligencia artificial en la educación superior debe considerar la optimización de recursos, lo que incluye formación continua y soporte técnico para maximizar la eficacia de estas tecnologías.

Asimismo, la implementación de herramientas de inteligencia artificial generativa, según Ruiz-Rojas et al. (2023), no solo optimiza los recursos educativos al automatizar la creación de contenidos, sino que también mejora la calidad del proceso educativos en entornos de MOOC.

Además, el uso de LLMs en la generación de visualizaciones no solo mejora la eficiencia

y precisión del proceso, sino que también reduce significativamente los costos de desarrollo al eliminar la necesidad de crear modelos específicos para cada dominio (Maddigan & Susnjak, 2023).

Ots (2024) detalla que, además de ofrecer un entorno seguro, Azure OpenAI Service permite optimizar costos y recursos mediante la implementación de modelos de baja demanda y la capacidad de escalar recursos según sea necesario. Esto es beneficioso para organizaciones que buscan implementar soluciones de inteligencia artificial sin incurrir en costos significativos en infraestructura.

3.4. Análisis comparativo de soluciones de inteligencia conversacional

En esta sección, se realiza un análisis exhaustivo de las principales plataformas de inteligencia conversacional con el objetivo de determinar la solución más adecuada para el sistema de atención virtual de la OCA de la UNMSM. El análisis se divide en tres etapas:

- Definición de criterios asociados al caso de estudio.
- Asignación de puntajes a los criterios definidos.
- Benchmarking entre las soluciones seleccionadas.

Finalmente, se revisan estudios recientes que validan los resultados obtenidos, aportando una base técnica a la selección propuesta.

3.4.1. Criterios de evaluación

Los criterios de evaluación se definieron considerando las características clave necesarias para garantizar que la solución seleccionada cumpla con los requerimientos del caso de estudio:

- Precisión de respuestas: Capacidad de la plataforma para proporcionar respuestas correctas y relevantes en diferentes contextos, asegurando información precisa para los usuarios.

- Facilidad de implementación: Nivel de esfuerzo técnico y recursos necesarios para la integración de la solución en los sistemas actuales, incluyendo compatibilidad con APIs y herramientas existentes.
- Tiempo de respuesta: Rapidez con la que las plataformas procesan y entregan respuestas, medidas en segundos, lo que influye directamente en la experiencia del usuario.
- Adaptabilidad al contexto educativo: Capacidad para personalizar y ajustar la solución a escenarios específicos del ámbito educativo, incluyendo terminología especializada y casos de uso particulares.
- Costo de implementación: Recursos financieros necesarios para el desarrollo, implementación y mantenimiento del sistema, evaluando tanto costos iniciales como recurrentes.
- Escalabilidad: Capacidad del sistema para manejar un mayor volumen de usuarios y consultas sin pérdida de rendimiento, asegurando estabilidad y consistencia bajo alta demanda.

3.4.2. Asignación de pesos a los criterios

Cada criterio se ha ponderado considerando su relevancia en el sistema propuesto, la siguiente tabla detalla los valores y puntajes asociados a cada criterio:

Tabla 1

Puntajes asignados a los criterios de evaluación

Criterio	Valor	Descripción	Puntaje
Precisión de respuestas	Alta	Capacidad de la plataforma para proporcionar respuestas altamente relevantes y correctas.	3
	Media	Respuestas generalmente relevantes, con ocasionales errores menores.	2
	Baja	Respuestas con alta probabilidad de error o falta de relevancia.	1
Facilidad de implementación	Alta	La integración de la plataforma requiere pocos recursos técnicos y es fácil de configurar.	3
	Media	Requiere recursos técnicos moderados y un esfuerzo medio para la integración.	2
	Baja	Requiere recursos significativos y una alta complejidad técnica para su implementación.	1
Tiempo de respuesta	Rápido	Las respuestas son generadas en menos de 2 segundos, garantizando fluidez en la interacción.	3
	Moderado	Respuestas generadas entre 2 y 5 segundos, con impacto limitado en la experiencia del usuario.	2
	Lento	Respuestas generadas en más de 5 segundos, afectando la experiencia del usuario.	1
Adaptabilidad al contexto educativo	Alta	Altamente personalizable y adaptable a escenarios específicos del ámbito educativo.	3
	Media	Adaptación limitada con algunos desafíos técnicos o conceptuales.	2
	Baja	Escasa capacidad de personalización y ajuste a escenarios educativos.	1
Costo de implementación	Bajo	Requiere una inversión mínima para su desarrollo, implementación y mantenimiento.	3
	Medio	Inversión moderada con costos manejables para la organización.	2
	Alto	Requiere una inversión significativa que podría dificultar su adopción.	1
Escalabilidad	Alta	Maneja un gran volumen de usuarios y consultas sin degradación del rendimiento.	3
	Media	Soporta un volumen moderado de usuarios con una ligera degradación bajo cargas altas.	2
	Baja	Dificultad para manejar altos volúmenes de usuarios o consultas simultáneas.	1

Fuente: Elaboración propia

3.4.3. Benchmarking de soluciones

Se seleccionaron cuatro plataformas líderes en inteligencia conversacional: ChatGPT, Google Gemini, Claude y Microsoft Copilot. Cada una fue evaluada en función de los criterios definidos y detallados en la Tabla 1. Los resultados se presentan a continuación en la Tabla 2:

Tabla 2

Evaluación comparativa de plataformas de inteligencia conversacional

Criterio	ChatGPT	Google Gemini	Claude	Microsoft Copilot
Precisión de respuestas	3	3	3	2
Facilidad de implementación	2	2	2	3
Tiempo de respuesta	3	3	2	2
Adaptabilidad al contexto educativo	3	2	2	2
Costo de implementación	2	2	3	2
Escalabilidad	3	3	3	3
Puntuación Total	16	15	15	14

Fuente: Elaboración propia

3.4.4. Estudios recientes sobre plataforma de inteligencia conversacional

Para validar los resultados obtenidos, se revisaron estudios recientes que analizan el desempeño de las plataformas seleccionadas en contextos similares:

- Rangapur & Rangapur (2024) comparan modelos de lenguaje de gran escala, incluyendo ChatGPT y Google Gemini, en tareas de preguntas y respuestas conversacionales. Los resultados destacan el rendimiento robusto de ambas plataformas, especialmente en tareas complejas que requieren razonamiento y precisión en las respuestas.
- Hochmair et al. (2024) evalúan la precisión de varios chatbots en tareas especiales. Los hallazgos muestran que ChatGPT ofrece un desempeño superior en comparación con Claude y Copilot, particularmente en escenarios que requieren razonamiento lógico y procesamiento espacial.

- Polverini & Gregorcic (2024) analizan la capacidad de chatbots como Claude, Microsoft Copilot y ChatGPT para interpretar gráficos. Se concluye que ChatGPT lidera en precisión y tiempo de respuesta en comparación con las versiones gratuitas de otras plataformas.

Los resultados del benchmarking y la revisión de estudios recientes indican que ChatGPT es la plataforma adecuada para satisfacer la necesidad del sistema de atención virtual de la OCA de la UNMSM. Destaca por su precisión, tiempo de respuesta y escalabilidad.

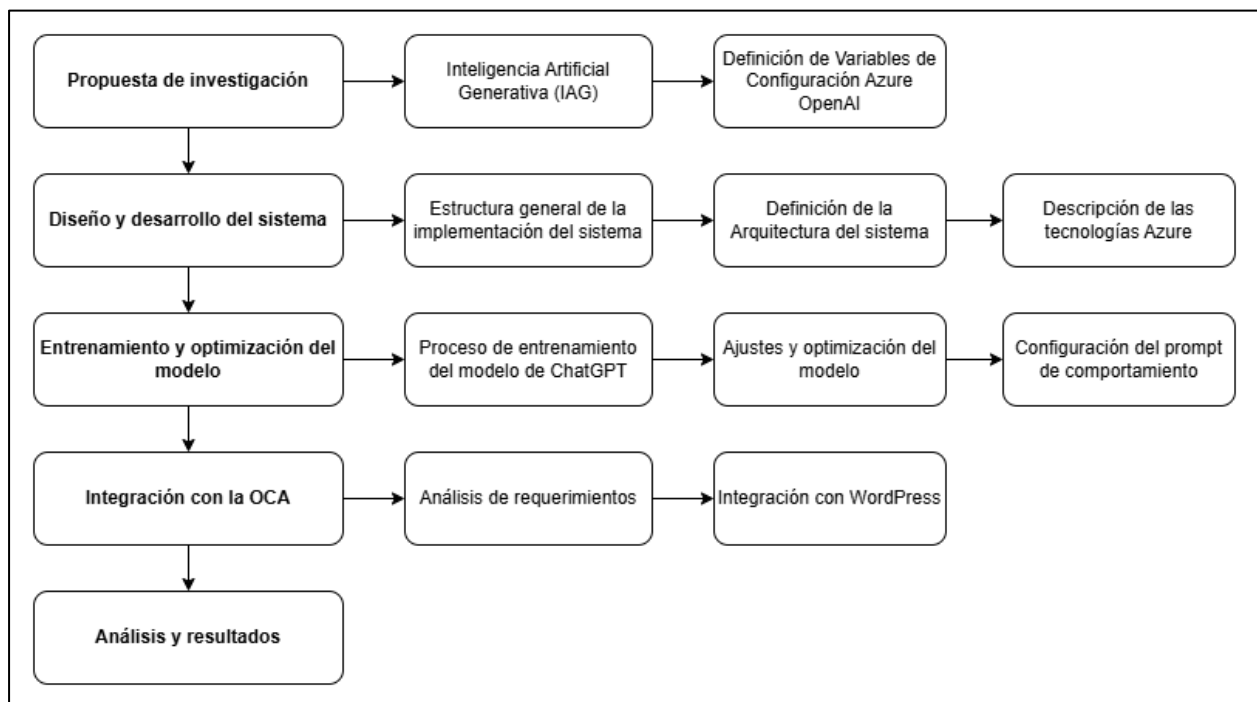
CAPÍTULO IV: METODOLOGÍA

Este capítulo se explicará cada una de las fases de la metodología utilizada para la elaboración de este proyecto. En la primera parte se explicarán las consideraciones preliminares que se tuvieron para el desarrollo del sistema, luego analizaremos el Diseño y Desarrollo del sistema, seguidamente veremos una sección que habla sobre el entrenamiento y optimización del modelo y finalmente veremos la integración técnica con la OCA.

A continuación, se describe en una imagen la metodología utilizada:

Figura 6

Representación gráfica de la metodología propuesta



Fuente: Elaboración propia

4.1 Propuesta de la investigación

Esta sección establece las bases teóricas y prácticas necesarias para el diseño, desarrollo e implementación del sistema, proporcionando un marco de referencia claro para las fases

posteriores del proyecto.

4.1.1 Inteligencia Conversacional Generativa

La Inteligencia Conversacional Generativa juega un papel fundamental. Esta tecnología se utiliza para crear un sistema capaz de interactuar de manera natural y efectiva con los usuarios, respondiendo a sus consultas y proporcionando asistencia de manera autónoma y precisa.

Inteligencia Conversacional Generativa se refiere a la capacidad de los sistemas de IA para generar respuestas originales y contextualmente relevantes en conversaciones naturales. En el contexto de la tesis, esto implica desarrollar un sistema que pueda comprender y responder a las necesidades de los estudiantes y personal de la OCA, utilizando el modelo ChatGPT. Este modelo, basado en la arquitectura de transformadores, ha sido entrenado con vastas cantidades de datos textuales, lo que le permite generar respuestas coherentes y precisas.

Al integrar ChatGPT en nuestro sistema, se puede ofrecer una experiencia de atención virtual que va más allá de los chatbots tradicionales. ChatGPT no solo responde preguntas predefinidas, sino que también puede manejar una variedad de consultas complejas y proporcionar respuestas detalladas y adaptadas al contexto de cada interacción. Esto es especialmente útil en la OCA, donde las consultas pueden variar desde información sobre programas académicos hasta procedimientos administrativos específicos.

Además, la Inteligencia Conversacional Generativa permite al sistema aprender y adaptarse con el tiempo. A medida que el sistema interactúa con más usuarios, puede identificar patrones y mejorar sus respuestas, ofreciendo una atención cada vez más personalizada y eficiente. Esto no solo optimiza la atención virtual, sino que también reduce la carga de trabajo del personal de la OCA, permitiéndoles enfocarse en tareas más complejas y críticas.

Finalmente, al utilizar Inteligencia Conversacional Generativa con ChatGPT en el

software, nos alineamos con las tendencias más avanzadas en el campo de la IA y la atención virtual. Esto no solo proporciona una solución eficaz a los problemas actuales de la OCA, sino que también posiciona a la UNMSM como una institución innovadora que adopta tecnologías de vanguardia para mejorar sus servicios.

4.1.2 Variables de configuración de Azure OpenAI Service

Como en todo sistema se deben establecer configuraciones que ajustarán el comportamiento del software. En este caso se otorgarán valores fijos que son básicamente las “credenciales” que necesita el software para conectarse con los servicios que ofrece Azure, pero también se configurarán variables que afinarán la respuesta de nuestro chatbot.

- `az_ia_search`: Nombre del recurso de Azure IA Search.
- `az_ia_search_key`: Clave administrador para el recurso de Azure IA Search.
- `az_ia_search_index`: Nombre del índice de búsqueda de Azure IA Search.
- `az_ia_search_semantic_enabled`: Valor booleano que activa o desactiva la búsqueda semántica.
- `az_ia_search_semantic`: Nombre de la configuración de la búsqueda semántica siempre y cuando esté activa la búsqueda semántica.
- `az_ia_search_top_k`: Cantidad de documentos que se obtendrán como resultado de Azure AI Search.
- `az_ia_search_en_domain`: Esta variable limita las respuestas del chatbot para que solo responda bajo el contexto de entrenamiento.
- `az_ia_search_query_type`: Tipo de consulta: simple, semántica, vectorial. Tiene prioridad sobre `az_ia_search_semantic`.
- `az_ia_search_stric`: Entero del 1 al 5 que especifica el rigor del modelo que limita las

respuestas a sus datos.

- `az_openai_resource_name`: Nombre del recurso de Azure OpenAI.
- `az_openai_model_implementation`: Nombre de la implementación del modelo.
- `az_openai_key`: Clave API del recurso de Azure OpenAI.
- `az_openai_model_name`: Nombre del modelo.
- `az_openai_temperature`: Define cuánta variabilidad se permite en las respuestas generadas. Este valor va de 0 a 1; cuanto más alto sea (por ejemplo, 0,8), más impredecible y variada será la salida. Un valor más bajo (como 0,1) hará que las respuestas sean más precisas y consistentes.
- `az_openai_top_p`: Es una opción alternativa al uso de la temperatura para generar respuestas. Conocida como "muestreo de núcleo", esta técnica hace que el modelo seleccione solo entre los tokens que, en conjunto, suman una probabilidad acumulada específica, definida por el valor de `top_p`.
- `az_openai_max_token`: Cantidad máxima de tokens generados en cada respuesta.
- `az_openai_prompt_behavior`: Esta variable ajusta el comportamiento del chatbot. También se llama "Prompt de comportamiento".
- `az_openai_api_version`: Versión de la API de Azure OpenAI.
- `az_openai_endpoint`: Endpoint de la API de Azure OpenAI.

4.2 Diseño y desarrollo del sistema

4.2.1 Estructura general de la implementación del sistema

En esta sección, se describirá la estructura general del sistema de inteligencia conversacional integrado con ChatGPT para optimizar la atención virtual en la OCA de la UNMSM. Se abordarán los componentes clave que conforman el sistema, incluyendo la base de

conocimiento, el prompt de comportamiento y la interfaz web. Cada uno de estos elementos es esencial para garantizar que el sistema pueda proporcionar respuestas precisas, interactuar de manera efectiva con los usuarios y ser accesible a través de una plataforma web amigable.

a. Base de Conocimiento

La base de conocimiento es el conjunto de información y datos que el sistema de inteligencia conversacional utiliza para generar respuestas. En el contexto de la propuesta, la base de conocimiento incluirá datos relevantes sobre la OCA, como información sobre programas académicos, procedimientos administrativos, fechas importantes, y cualquier otra información que los usuarios puedan necesitar. Esta base se nutre de documentos, bases de datos y otros recursos internos de la UNMSM, garantizando que las respuestas proporcionadas por el chatbot sean precisas y actualizadas. La base de conocimiento se estructura de manera que el modelo de GPT pueda acceder a ella de forma eficiente, permitiendo una rápida consulta y generación de respuestas informadas.

b. Prompt de Comportamiento

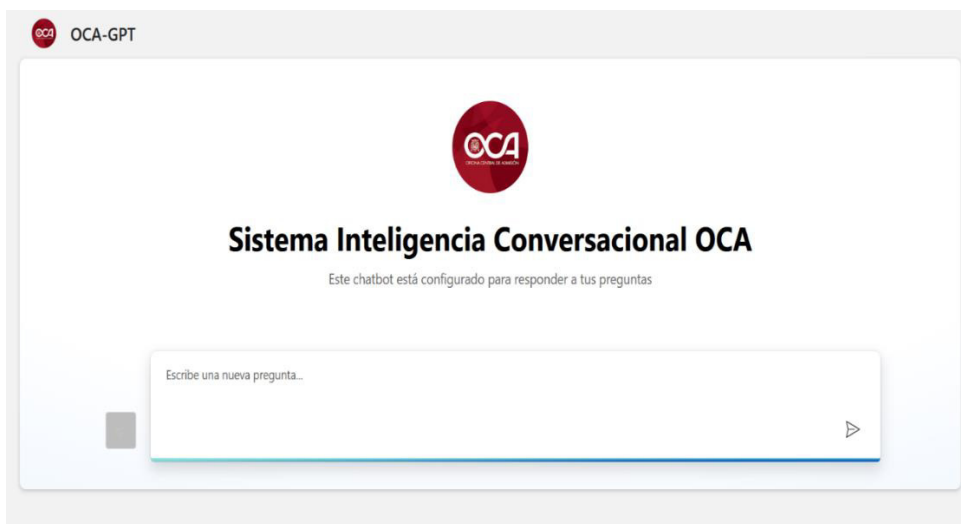
El prompt de comportamiento se refiere a las instrucciones y directrices específicas que se le proporcionan al chatbot para guiar su interacción con los usuarios. El prompt de comportamiento está diseñado para asegurar que el chatbot mantenga un tono profesional y cordial, adecuado para la atención en un entorno académico. Este prompt incluye ejemplos de interacciones comunes, respuestas tipo, y reglas sobre cómo manejar diferentes tipos de consultas. Ajustar y optimizar el prompt es crucial para garantizar que el chatbot no solo proporcione información precisa, sino que también interactúe de manera adecuada con los usuarios, reflejando los valores y estándares de la OCA.

c. Interfaz Web

La interfaz web es la plataforma a través de la cual los usuarios interactúan con el sistema de inteligencia conversacional. La interfaz incluye elementos como el campo de entrada de texto para que los usuarios formulen sus preguntas, la ventana de chat donde se muestran las respuestas generadas por ChatGPT. El diseño de la interfaz web debe enfocarse en mejorar la experiencia del usuario, asegurando que la interacción con el chatbot sea fluida y efectiva (Ver figura 7).

Figura 7.

Interfaz web del chatbot



Fuente: Elaboración propia

4.2.2 Arquitectura del sistema

En esta sección se detalla la arquitectura del sistema de inteligencia conversacional integrado con ChatGPT. Se presentará una visión general de los componentes principales y su interacción, proporcionando un marco técnico claro y bien definido. Esta arquitectura asegura que todos los elementos del sistema trabajen de manera cohesiva para ofrecer una experiencia de usuario óptima y eficiente. Los tres componentes clave que se analizarán son: el servidor backend, el modelo de GPT, y el frontend de la aplicación web.

El servidor backend es la columna vertebral del sistema, gestionando la lógica de la

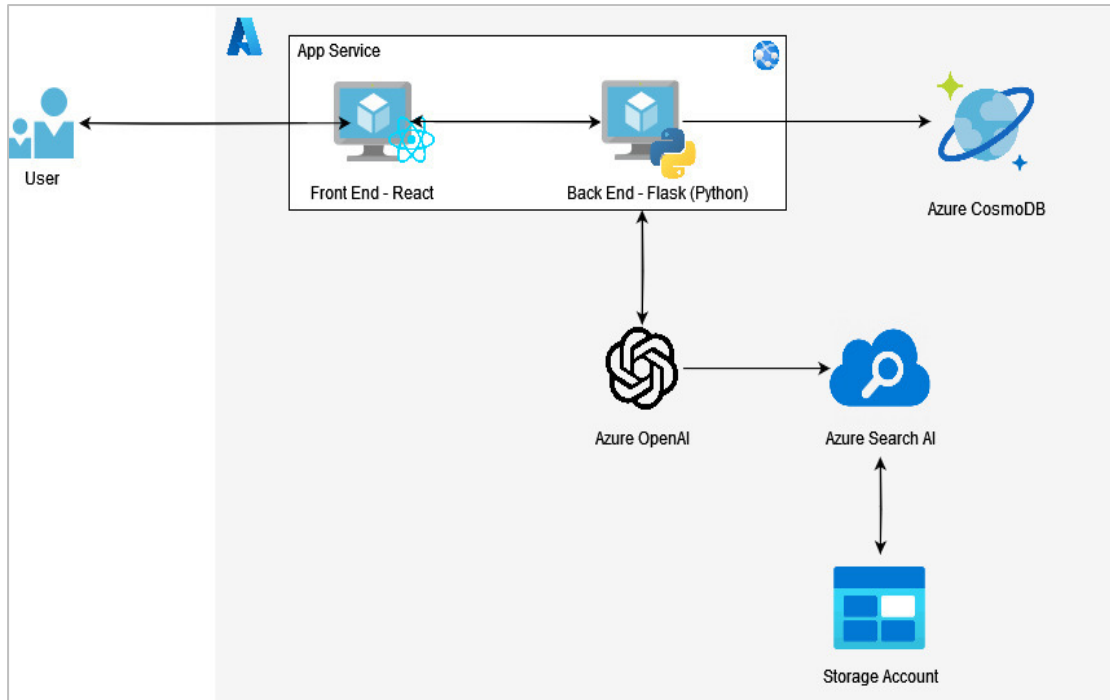
aplicación y la comunicación entre los diferentes componentes. El backend será responsable de recibir las solicitudes de los usuarios, procesarlas, y enviar las respuestas generadas por el servicio de Azure OpenAI Service de vuelta al frontend. Este componente se implementó utilizando el Framework Flask basado en el lenguaje de programación Python, para manejar grandes volúmenes de tráfico y asegurar una rápida respuesta.

Como núcleo del sistema de inteligencia conversacional utilizaremos el motor cognitivo de ChatGPT implementado con el modelo GPT-4. Este modelo está entrenado para comprender y generar respuestas en lenguaje natural, utilizando los datos de la base de conocimiento específica de la OCA. El modelo se integra con el servidor backend, recibiendo las consultas de los usuarios y generando respuestas contextualmente relevantes.

Por último, el frontend de la aplicación web es la interfaz con la que interactúan los usuarios. Este componente está diseñado para ser intuitivo y fácil de usar, permitiendo a los estudiantes y personal de la OCA acceder rápidamente a la información que necesitan. El frontend se desarrolló utilizando tecnologías web modernas como React en la versión **18.2.0**, que ofrecen una experiencia de usuario fluida y receptiva. Así mismo, el frontend incluye funcionalidades como el campo de entrada para preguntas y la visualización de respuestas. La comunicación con el backend se realiza a través de API RESTful, asegurando una integración eficiente y segura (Ver figura 8).

Figura 8.

Infraestructura del Sistema Inteligente Conversacional



Fuente: Elaboración propia

4.2.3 Descripción de las tecnologías Azure

En esta sección, se detallará cómo se utilizaron y configuraron las diversas tecnologías Azure en el desarrollo del sistema de inteligencia conversacional integrado con ChatGPT. Estas tecnologías proporcionan una infraestructura robusta y escalable que permite la implementación eficiente del modelo de GPT-4, la gestión de la base de conocimiento y el despliegue de la aplicación. Las principales tecnologías Azure utilizadas en el proyecto son Azure OpenAI Service, Azure AI Search, Azure Storage Account, y Azure App Service.

a. Azure OpenAI Service

Azure OpenAI Service es la plataforma que permite el acceso a los modelos avanzados de OpenAI, incluyendo GPT-4. En este proyecto de tesis, se utilizó Azure OpenAI Service para implementar el modelo de GPT-4, que es fundamental para el sistema de inteligencia conversacional.

Este servicio se configuró creando una cuenta en Azure y accediendo al portal de Azure OpenAI Service. Se seleccionó el modelo GPT-4 y se configuraron los parámetros necesarios, como el tamaño del modelo y los límites de uso (tokens). Se utilizó la API Key proporcionada por Azure para autenticar las solicitudes a la API de OpenAI. Esta clave se empleó en la API del OpenAI Service y, a través de la biblioteca request, se enviaron consultas a GPT-4 para recibir las respuestas generadas. La integración de Azure OpenAI Service aseguró que el chatbot generara respuestas precisas y relevantes basadas en el contexto de las consultas de los usuarios.

b. Azure AI Search

El uso de Azure AI Search es fundamental para el funcionamiento correcto del chatbot ya que realiza una búsqueda indexada (semántica) dentro de toda la base de conocimientos permitiendo optimizar el tiempo de respuesta y el contenido encontrado que incluye datos sobre la OCA, como información de programas académicos y procedimientos administrativos.

Para configurar Azure AI Search, se utilizó la misma API que proporciona Azure OpenAI Service. Allí, se especificó el endpoint de búsqueda, la clave de la API y el índice creado para la búsqueda dentro de la base de conocimientos. Se creó el índice desde el portal de Azure Search, definiendo los campos necesarios para la indexación. Además, se configuró un "Indexer" que controla la actualización continua de la base de conocimiento, asegurando que la información esté siempre actualizada y accesible.

c. Azure Storage Account

El Storage Account de Azure se utilizó para almacenar la base de conocimientos del sistema. Este servicio proporciona una solución segura y escalable para gestionar grandes volúmenes de datos.

Primero, se creó una cuenta de almacenamiento en el portal de Azure y se definieron los contenedores necesarios para organizar los datos. Luego, se subieron los archivos de la base de conocimientos, incluyendo documentos, hojas de cálculo y otros recursos relevantes; además se configuraron políticas de acceso y permisos para asegurar que solo las aplicaciones autorizadas pudieran acceder a los datos. Finalmente, se enlazó Azure Storage Account con Azure AI Search creando un índice que permitiera la búsqueda eficiente dentro de la base de conocimientos.

d. Azure Cosmos DB

Este recurso de Azure se utilizó para almacenar las conversaciones de los usuarios con el chatbot. Este servicio proporciona una base de datos NoSQL distribuida globalmente, que permite el almacenamiento eficiente y la recuperación rápida de datos.

Se configuró Azure Cosmos DB creando una cuenta de base de datos en el portal de Azure. Se definieron los contenedores y las particiones necesarias para organizar las conversaciones de los usuarios. Se configuraron políticas de acceso y permisos para asegurar que solo las aplicaciones autorizadas pudieran acceder y manipular los datos. La integración de Azure Cosmos DB permitió almacenar las interacciones de los usuarios con el chatbot, proporcionando una base de datos robusta para el análisis y la mejora continua del sistema.

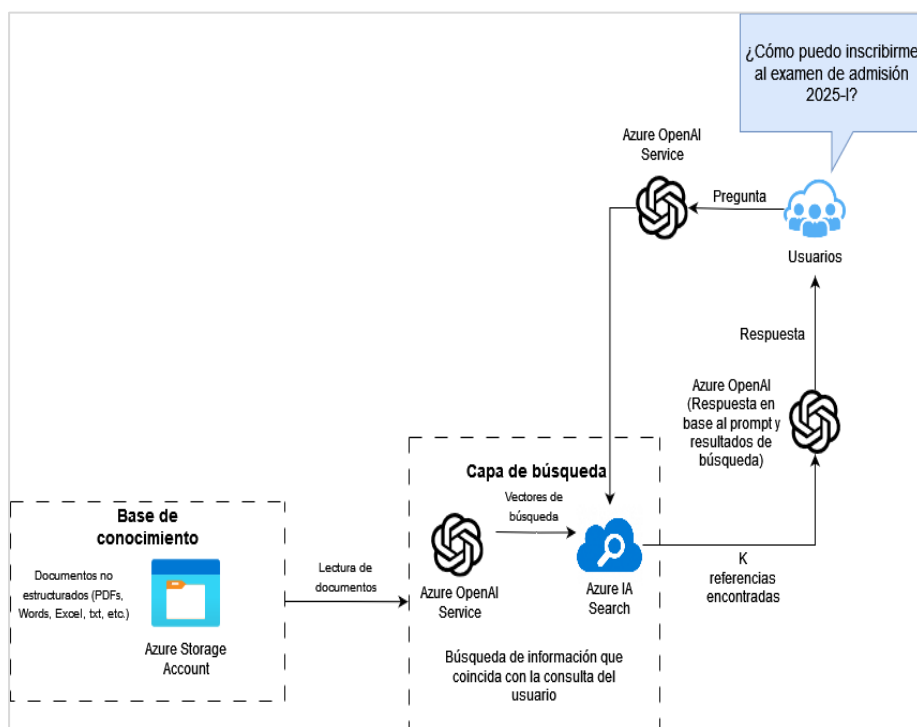
e. Azure App Service

Azure App Service es la plataforma utilizada para alojar tanto el frontend como el backend del proyecto. Esta tecnología permitió desplegar la aplicación en un entorno de producción de manera eficiente y escalable.

Se configuraron Azure App Service creando una instancia del servicio en el portal de Azure. Se subieron los archivos del frontend y backend a esta instancia, configurando las rutas y los entornos de ejecución necesarios. Azure App Service facilitó el manejo de las solicitudes de los usuarios y la entrega de las respuestas generadas por el sistema, asegurando un rendimiento óptimo y una disponibilidad constante. Además, se configuraron opciones de escalabilidad automática para manejar picos de tráfico, garantizando que el sistema pudiera responder eficazmente a la demanda de los usuarios (Ver figura 9).

Figura 9.

Arquitectura de los recursos Azure del chatbot



Fuente: Elaboración propia

4.3 Entrenamiento y optimización del modelo

4.3.1 Proceso de Entrenamiento del Modelo de ChatGPT

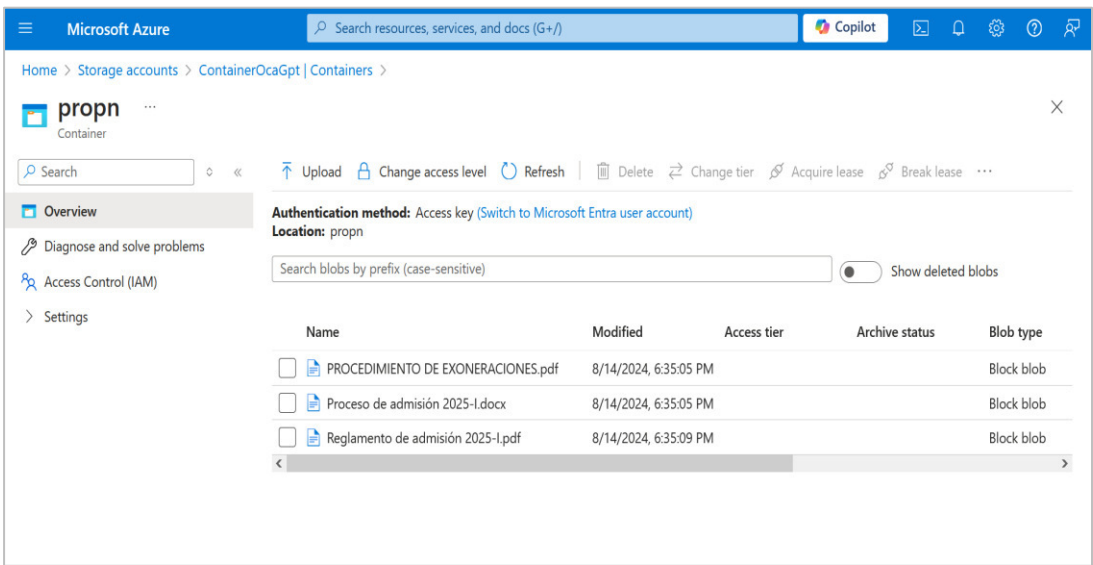
El proceso de entrenamiento del modelo de GPT-4 se realizó en varias etapas clave, cada

una de las cuales fue esencial para el desarrollo de un chatbot eficaz y adaptado a las necesidades de la OCA de la UNMSM. La primera etapa consistió en la recopilación y preparación de datos relevantes. Se reunieron datos extensivos sobre las carreras ofrecidas por la universidad, fechas de admisión, costos, mallas curriculares y otra información administrativa esencial. Este paso fue crucial para asegurar que el chatbot tuviera una base sólida de conocimiento para responder a las consultas de los usuarios de manera precisa.

Los datos recopilados se normalizaron y estructuraron en documentos PDF, que posteriormente se utilizaron como base de conocimiento para el chatbot. Este proceso de estructuración garantizó que la información estuviera organizada de manera lógica y accesible, facilitando así la búsqueda y recuperación de datos durante las interacciones con el usuario. Los documentos PDF fueron cargados en un contenedor dentro del Azure Storage Account, lo que permitió su almacenamiento seguro y accesible para el chatbot.

Figura 10.

Configuración del Azure Storage Account

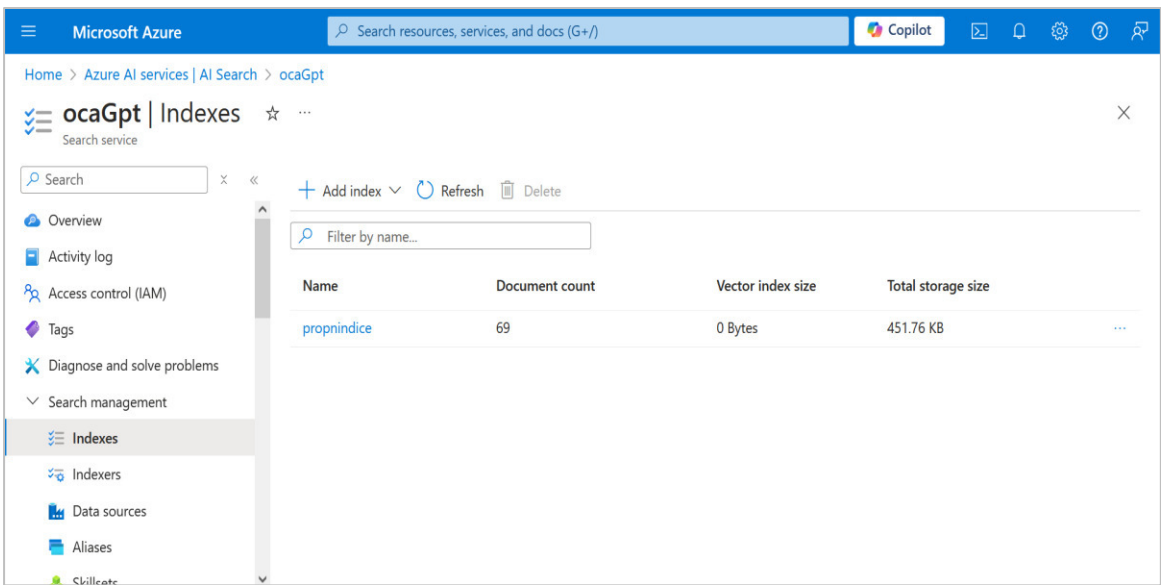


Fuente: Microsoft Azure Portal

Una vez que la base de conocimiento estuvo lista, se configuró un índice de búsqueda en Azure AI Search. Este índice permite al chatbot realizar búsquedas rápidas y eficientes dentro de la base de conocimiento, mejorando la capacidad del sistema para encontrar y recuperar información relevante en respuesta a las consultas de los usuarios. La configuración del índice se realizó desde el portal de Azure, especificando los campos y parámetros necesarios para la indexación de los documentos PDF.

Figura 11.

Configuración del índice para la búsqueda vía Azure IA Search

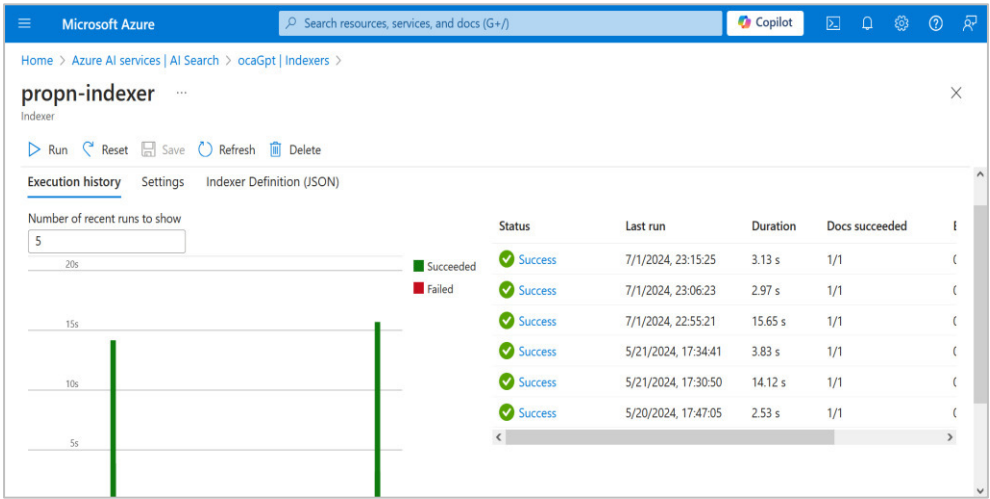


Fuente: Microsoft Azure Portal

Posteriormente, se agregó un indexador, el cual inicia una actualización automática, o a demanda, que mantiene la base de conocimiento siempre actualizada. Este proceso es esencial para garantizar que cualquier modificación o adición de nuevos documentos se refleje en las búsquedas del chatbot de forma oportuna, asegurando la precisión y relevancia de las respuestas.

Figura 12.

Configuración del indexador para la actualización automática

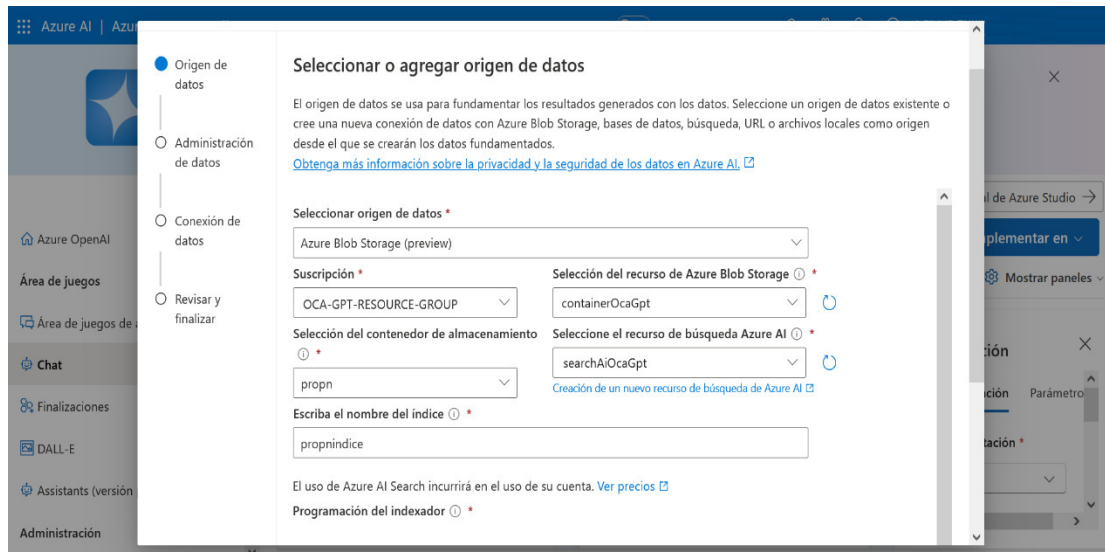


Fuente: Microsoft Azure Portal

A continuación, se procedió a realizar la configuración en Azure OpenAI Service el cual combina los recursos antes mencionados: Azure Storage Account y Azure Search IA. Se estableció un tipo de búsqueda semántica el cual es superior a la búsqueda por “palabra clave” ya que se basa en el contexto de la oración que viaja como pregunta.

Figura 13.

Configuración de Azure OpenAI Service



Fuente: Microsoft Azure Portal

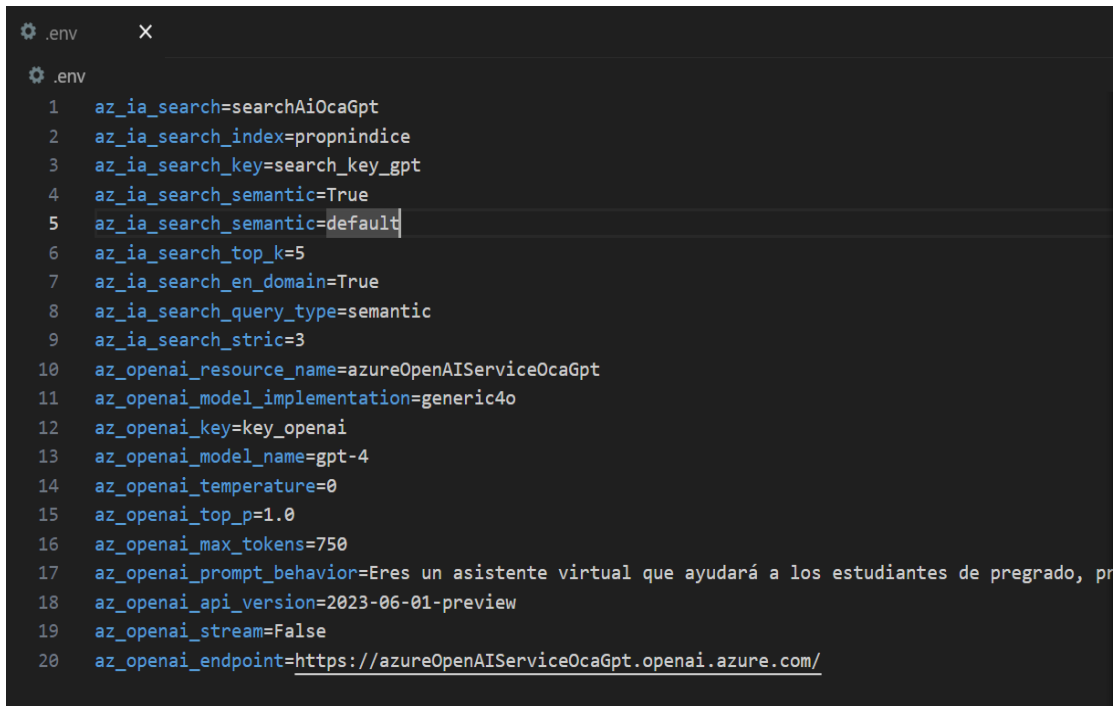
El modelo utilizado para el chatbot es GPT-4, que incluye el motor cognitivo de ChatGPT. Este motor es una inteligencia artificial diseñada para permitir la conversación natural entre un usuario y la máquina, emulando el estilo de diálogo característico de ChatGPT. Durante el entrenamiento del modelo en Azure OpenAI Service, se configuraron parámetros específicos para ajustar la calidad y coherencia de las respuestas del chatbot, entre ellos:

- **Temperatura:** Controla el nivel de creatividad de las respuestas. Un valor más alto hace que el modelo genere respuestas más variadas y creativas, mientras que un valor más bajo produce respuestas más conservadoras y deterministas.
- **Top_p:** Limita la selección de palabras a las más probables. Un valor más bajo de top_p restringe las opciones a las palabras más probables, lo que puede hacer las respuestas más predecibles, mientras que un valor más alto permite mayor variabilidad.
- **Cantidad de tokens:** Determina la longitud máxima de las respuestas. Configurar un

número adecuado de tokens asegura que las respuestas sean completas, pero no excesivamente largas.

Figura 14.

Configuración del proyecto

A screenshot of a code editor showing a .env file. The file contains 20 lines of configuration for an Azure OpenAI project. The variables are: az_ia_search, az_ia_search_index, az_ia_search_key, az_ia_search_semantic, az_ia_search_top_k, az_ia_search_en_domain, az_ia_search_query_type, az_ia_search_stric, az_openai_resource_name, az_openai_model_implementation, az_openai_key, az_openai_model_name, az_openai_temperature, az_openai_top_p, az_openai_max_tokens, az_openai_prompt_behavior, az_openai_api_version, az_openai_stream, and az_openai_endpoint. The prompt behavior is set to a specific role for a virtual assistant.

```
.env
1 az_ia_search=searchAiOcaGpt
2 az_ia_search_index=proprnindice
3 az_ia_search_key=search_key_gpt
4 az_ia_search_semantic=True
5 az_ia_search_semantic=default
6 az_ia_search_top_k=5
7 az_ia_search_en_domain=True
8 az_ia_search_query_type=semantic
9 az_ia_search_stric=3
10 az_openai_resource_name=azureOpenAIServiceOcaGpt
11 az_openai_model_implementation=generic4o
12 az_openai_key=key_openai
13 az_openai_model_name=gpt-4
14 az_openai_temperature=0
15 az_openai_top_p=1.0
16 az_openai_max_tokens=750
17 az_openai_prompt_behavior=Eres un asistente virtual que ayudará a los estudiantes de pregrado, pr
18 az_openai_api_version=2023-06-01-preview
19 az_openai_stream=False
20 az_openai_endpoint=https://azureOpenAIServiceOcaGpt.openai.azure.com/
```

Fuente: Elaboración propia

4.3.2 Ajustes y Optimización del Modelo

Después del entrenamiento inicial, se realizaron diversas modificaciones para alinear el comportamiento del chatbot a las necesidades específicas de la OCA. Este proceso incluyó la incorporación de datos relevantes relacionados con preguntas frecuentes, políticas de admisión, procedimientos administrativos, y otros aspectos clave de la universidad. Dichos datos fueron normalizados y estructurados de manera eficaz, utilizando técnicas que aseguraron su integración adecuada dentro de la base de conocimiento. De este modo, el chatbot pudo manejar con precisión y claridad las consultas específicas de los usuarios, enfocándose en respuestas relevantes y

contextuales para estudiantes y padres de familia.

Para normalizar y estructurar documentos no estructurados, el enfoque es organizar la información en un formato coherente y accesible. Se utilizaron los siguientes formatos:

a. Organización Jerárquica

Se organizan los documentos PDF en un formato de árbol, dividiendo el contenido en secciones y subsecciones. Por ejemplo, en lugar de tener un único documento extenso con múltiples temas, se desglosan en capítulos, cada uno con sus subtemas bien identificados (e.g., Capítulo 1: Carreras Universitarias → 1.1 Ingeniería de Sistemas → 1.1.1 Plan de Estudios). Este formato de "libro" permite que la base de conocimiento maneje y estructure eficientemente la información.

Formato de Estructuración:

- **Capítulo 1: Carreras Universitarias**
 - 1.1 Ingeniería de Sistemas
 - 1.1.1 Plan de Estudios
 - 1.1.2 Costos de Matrícula
 - 1.2 Medicina
 - 1.2.1 Requisitos de Admisión
 - 1.2.2 Duración de la Carrera
- b. Índice y Metadatos

Para mejorar la accesibilidad y precisión de las búsquedas, se agregan metadatos a cada sección, tales como palabras clave, títulos y resúmenes breves. Por ejemplo, un PDF sobre las fechas de admisión puede tener metadatos que incluyan "admisión", "fechas", "OCA", etc. Esto asegura que el índice de búsqueda en Azure AI Search pueda encontrar el contenido correcto

basándose en las consultas de los usuarios.

Para entender mejor esta técnica, dentro de nuestra base de conocimiento tenemos un PDF titulado "Proceso de admisión 2025-I". Este documento contiene información clave, como fechas de inscripción, requisitos y tarifas. Para mejorar la búsqueda y recuperación de esta información, se estructuraron los metadatos del documento en Azure AI Search. Por ejemplo:

- **Título:** Proceso de Admisión 2024
 - **Categoría:** Admisión
 - **Fecha de Publicación:** enero 2024
 - **Autor:** Oficina de Admisión
 - **Palabras Clave:** Inscripción, requisitos, tarifas, fechas
- c. Normalización del Formato

Con esta técnica aseguramos que todos los documentos sigan un formato estándar, utilizando un estilo uniforme en títulos, párrafos, y tablas. Esto implica el uso de encabezados consistentes, numeración de secciones y subtítulos, garantizando que cualquier documento se presente de manera clara y ordenada, facilitando su interpretación tanto para el modelo como para los usuarios.

En nuestra base de conocimientos, en un inicio los documentos titulados "**Reglamento de admisión 2025-I**" y "**Procedimiento de exoneraciones – Admisión 2025-I**" incluía listas desordenadas, distintos estilos de texto y tablas sin formato. Durante el proceso de normalización, se estandarizaron el estilo de estos documentos para facilitar su indexación:

- **Estilos de Títulos:** Se aseguran de que todos los títulos de secciones y subsecciones sigan el mismo formato, como **Título 1** para las carreras, **Título 2** para las descripciones específicas.

- **Tablas:** Se convierten a un formato uniforme en todo el documento, asegurando que las tablas de mallas curriculares sigan el mismo formato con columnas claras para materias, créditos y semestres.
- **Listas:** Las listas de requisitos y características se convierten a viñetas uniformes, eliminando listas numéricas o jerarquías complejas.

Así obtenemos la información más estructurada, como si fuera un capítulo dentro de un libro, facilitando la consulta por parte del chatbot y su posterior actualización.

4.3.3 Configuración del Prompt de Comportamiento

Para mejorar aún más la calidad de las respuestas, se configuró un prompt de comportamiento específico. Este prompt orientó el tono y estilo del chatbot para que sus respuestas fueran formales pero accesibles y amigables. El objetivo era que el chatbot no solo informara, sino que también interactuara de manera eficaz con los jóvenes estudiantes y sus padres, quienes constituyen los usuarios principales. Se consideraron las siguientes aptitudes que fueron ajustadas en el prompt de comportamiento:

Tabla 3.

Aptitudes para ajustar el prompt de comportamiento

Aptitud	Descripción	Ejemplo de Aplicación
Formalidad	Asegura que el chatbot ofrezca respuestas profesionales y respetuosas.	Cuando un estudiante preguntaba por los requisitos de admisión, el chatbot respondía con un lenguaje educado y claro: "Para ser admitido en la carrera de Ingeniería de Sistemas, deberá presentar los siguientes documentos...".

Jovialidad	Mantiene un tono accesible y motivador para los jóvenes estudiantes.	Al responder a una consulta sobre fechas importantes, el chatbot podría decir: "¡No te preocupes! El próximo examen de admisión es el 15 de agosto. ¡Aún tienes tiempo para prepararte!".
Claridad	Evita jergas y utiliza un lenguaje claro y directo para facilitar la comprensión.	En lugar de proporcionar una explicación técnica sobre el proceso de admisión, el chatbot respondería: "El proceso de admisión consta de tres pasos: registro, examen y publicación de resultados".
Empatía	Muestra comprensión, especialmente en temas sensibles como costos.	Si un padre de familia expresaba preocupación sobre el costo de la matrícula, el chatbot podría responder: "Entiendo que el costo puede ser una preocupación. Le sugiero que revise nuestras opciones de becas y financiamiento, que pueden aliviar esta carga".

Fuente: Elaboración propia

Este enfoque permitió que las repuestas del chatbot se ajustaran adecuadamente a las necesidades de los diferentes usuarios, ofreciendo información precisa de una manera accesible y efectiva.

“Eres un asistente virtual que ayudará a los estudiantes de pregrado, preuniversitarios y postgrado a poder estar al tanto de los programas, ciclos, costos y demás información ofrecida en la OCA. Debes tener un trato muy amable con las personas. Limitate a responder solo la información de tu base de conocimiento. Usa emojis agradables. Siempre sugiere una nueva consulta luego de cada respuesta.”

4.4 Integración con la OCA

4.4.1 *Análisis de Requerimientos*

En esta sección se detallará el proceso de análisis de requerimientos llevado a cabo para integrar el sistema de inteligencia conversacional con la página web de la OCA, la cual está desarrollada en WordPress. El análisis de requerimientos es una etapa crítica para asegurar que todas las necesidades y expectativas del usuario final se cumplan.

Puntos clave:

- **Usuarios Finales:**
 - **Jóvenes Cursando el 4to y 5to Año de Secundaria:** Estos usuarios necesitan información precisa sobre las carreras ofrecidas, requisitos de admisión, fechas importantes y procedimientos de inscripción. Además, requieren orientación sobre las opciones de carrera y consejos sobre cómo prepararse para los exámenes de admisión.
 - **Estudiantes de Academias Preuniversitarias:** Este grupo está enfocado en preparar sus exámenes de admisión y necesitan detalles específicos sobre el proceso de admisión, requisitos y fechas. También valoran la información sobre mallas curriculares y posibles salidas profesionales de las carreras de interés.
 - **Padres de Familia:** Los padres buscan información sobre costos de las carreras, opciones de financiamiento, becas y soporte estudiantil. Además, están interesados en la seguridad y bienestar de sus hijos dentro de la universidad.
 - **Público en General:** Esto incluye a interesados en cursos de extensión, programas de posgrado y personas que buscan información general sobre la universidad y sus ofertas académicas.

Requerimientos Funcionales:

Los requerimientos funcionales son una parte crítica en el desarrollo de nuestro sistema, ya que definen las acciones específicas que el software debe ser capaz de realizar para cumplir con los objetivos propuestos. Para nuestra tesis, el objetivo principal es desarrollar un chatbot que optimice la atención virtual de la OCA de la UNMSM, proporcionando información y orientación a los futuros estudiantes y padres de familia de manera eficiente y accesible. Identificar los requerimientos funcionales permite asegurar que el sistema cubra todas las necesidades esenciales, respondiendo correctamente a las consultas y asistiendo a los usuarios en las áreas clave, como información de admisión, detalles de carreras, orientación y asesoría.

Cada funcionalidad descrita a continuación representa un aspecto fundamental del servicio que el chatbot debe ofrecer, y se deriva directamente de las necesidades de los usuarios y de los objetivos de la OCA. Estos requerimientos guían tanto el diseño como la implementación del sistema, asegurando que se cumplan los estándares de calidad y satisfacción del usuario.

Tabla 4.

Requerimientos funcionales para la integración con la OCA

Requerimiento	Descripción
Integración del chatbot con la plataforma WordPress de la OCA:	Implementación del chatbot en WordPress: El sistema debe permitir la integración del chatbot desarrollado en React dentro del sitio web de la OCA, que está basado en WordPress. La integración debe ser fluida, asegurando que el chatbot sea accesible desde cualquier página del sitio. Compatibilidad y Estabilidad: El chatbot debe ser compatible con los plugins y temas actuales de WordPress, garantizando la estabilidad y el rendimiento adecuado del

	sitio web, sin generar conflictos o afectar la funcionalidad existente.
Personalización del chatbot para la OCA	Respuestas Personalizadas a las Consultas de Admisión: El chatbot debe estar personalizado para responder preguntas específicas relacionadas con la admisión, tales como fechas de inscripción, procedimientos, requisitos, costos, y otras dudas frecuentes. Las respuestas deben estar adaptadas a las normativas y políticas específicas de la OCA. Soporte para Múltiples Tipos de Usuarios: El sistema debe poder diferenciar entre los diferentes perfiles de usuarios que interactúan con el chatbot (estudiantes, padres de familia, etc.) y adaptar las respuestas en función del perfil detectado, utilizando un tono y estilo adecuados para cada grupo.
Gestión de Actualización de Contenidos	Automatización de Actualizaciones de la Base de Conocimiento: El sistema debe incluir un mecanismo de actualización automatizada, que se active cada vez que haya un cambio en los documentos administrativos de la OCA (como procedimientos o políticas). Esto asegurará que el chatbot siempre proporcione la información más actualizada. Mantenimiento del Índice de Búsqueda en Azure AI Search: El sistema debe contar con un indexador que mantenga actualizada la base de conocimiento, de modo que cualquier cambio en los documentos importantes (como PDF y Word) se refleje automáticamente en las respuestas del chatbot.
Soporte Multidispositivo	Accesibilidad desde Diferentes Dispositivos: El chatbot debe ser accesible desde múltiples dispositivos, incluidos

	smartphones, tabletas y computadoras de escritorio. Esto asegurará que los usuarios puedan interactuar con el chatbot en cualquier momento y desde cualquier lugar. Interfaz Amigable y Responsiva: El sistema debe ofrecer una interfaz gráfica amigable y responsiva, adaptándose al tamaño de la pantalla del dispositivo del usuario, garantizando una experiencia de uso óptima sin importar el medio de acceso.
Escalabilidad y Soporte para Alta Demanda	Capacidad para Manejar Gran Cantidad de Consultas Simultáneas: El sistema debe ser escalable y capaz de manejar una alta cantidad de usuarios interactuando con el chatbot de manera concurrente, especialmente durante periodos de alta demanda, como las fechas de inscripción o de exámenes de admisión. Optimización de Recursos: El sistema debe estar optimizado para utilizar los recursos de manera eficiente, asegurando que las consultas sean procesadas sin retrasos significativos, incluso en situaciones de alta carga.

Fuente: Elaboración propia

Requerimientos No Funcionales:

Los requerimientos no funcionales son aquellos que describen cómo debe comportarse el sistema desde el punto de vista de la calidad del servicio, rendimiento, seguridad, y otras características que afectan la experiencia general del usuario. A diferencia de los requerimientos funcionales, que se enfocan en las tareas y funcionalidades específicas que el sistema debe cumplir,

los requerimientos no funcionales garantizan que el sistema funcione de manera eficiente, segura y confiable en diversos escenarios.

Para nuestra tesis, los requerimientos no funcionales son cruciales porque establecen las bases para que la integración del chatbot con la OCA no solo cumpla con las funciones esperadas, sino que también lo haga de manera efectiva y con la calidad necesaria. Algunos de los aspectos clave que cubren estos requerimientos incluyen:

Tabla 5.

Requerimientos no funcionales para la integración con la OCA

Requerimiento	Descripción
Rendimiento del Sistema	Tiempo de respuesta: El sistema debe garantizar tiempos de respuesta rápidos y eficientes. Las respuestas del chatbot no deben tardar más de 7.5 segundos en ser generadas, incluso bajo condiciones de alta carga. Uso eficiente de Recursos: El chatbot y su infraestructura deben estar optimizados para utilizar los recursos de procesamiento y almacenamiento de manera eficiente, evitando sobrecargas y manteniendo un rendimiento estable durante periodos prolongados.
Seguridad	Cumplimiento con Normativas de Protección de Datos: El sistema debe cumplir con las normativas de protección de datos, como la Ley de Protección de Datos Personales N° 29733, garantizando la privacidad y confidencialidad de los datos de los usuarios.
Flexibilidad	Flexibilidad para Integración Futura: El sistema debe ser lo suficientemente flexible para permitir futuras integraciones con otras plataformas y tecnologías, garantizando su evolución conforme crezcan las necesidades de la OCA y del entorno

		tecnológico.
Disponibilidad Confiabilidad	y	Alta disponibilidad: El sistema debe garantizar una alta disponibilidad, con un tiempo de actividad mínimo del 99.5%. Esto asegura que los usuarios puedan acceder al chatbot y obtener respuestas en cualquier momento, sin interrupciones significativas en el servicio. Tolerancia a fallos: El sistema debe ser tolerante a fallos, con mecanismos de respaldo y recuperación automática en caso de errores o caídas del servicio. Esto incluye la capacidad de reanudar las sesiones de los usuarios sin pérdida de datos o información relevante.
Usabilidad Accesibilidad	y	Interfaz de Usuario Intuitiva: La interfaz del chatbot debe ser intuitiva y fácil de usar, tanto para jóvenes estudiantes como para padres de familia. Debe estar diseñada para que cualquier usuario, independientemente de su nivel de habilidad tecnológica, pueda interactuar con el chatbot sin complicaciones.
Experiencia de Usuario		Consistencia y claridad en las respuestas: Las respuestas del chatbot deben ser claras, coherentes y mantener un tono uniforme a lo largo de las interacciones. El chatbot debe evitar proporcionar respuestas ambiguas o contradictorias, garantizando una experiencia de usuario fluida y confiable. Adaptación a diferentes Perfiles de Usuarios: El sistema debe ser capaz de adaptarse a diferentes perfiles de usuarios, ajustando la experiencia de uso en función del tipo de usuario (estudiantes, padres de familia, etc.), ofreciendo una experiencia personalizada y relevante.

Fuente: Elaboración propia

4.4.2 Integración con WordPress

La integración del chatbot desarrollado en React con la plataforma WordPress de la OCA es un paso crucial para asegurar que la herramienta de inteligencia conversacional se incorpore de manera efectiva y funcional en el sitio web existente. Este proceso comenzó con una evaluación exhaustiva de la infraestructura actual de WordPress para identificar cualquier ajuste necesario. La compatibilidad entre React y WordPress se aseguró mediante el uso de tecnologías y prácticas de integración web modernas, permitiendo que el chatbot interactúe sin problemas con el contenido y las funcionalidades ya establecidas en el sitio de la OCA.

Para insertar el widget del chatbot en las páginas deseadas de la OCA, se desarrolló un plugin personalizado de WordPress. Este plugin facilita la incorporación del chatbot en cualquier sección del sitio web sin necesidad de modificar el código fuente principal de WordPress. El diseño del plugin permite una integración fluida, asegurando que el chatbot se cargue rápidamente y ofrezca una experiencia de usuario coherente con el resto del sitio.

La configuración del plugin se diseñó para ser intuitiva y accesible, de manera que el equipo de la OCA pueda realizar ajustes y personalizaciones sin necesidad de conocimientos técnicos avanzados. Las opciones de configuración permiten definir parámetros como la posición del chatbot en la página, los colores y estilos visuales, y el comportamiento de inicio del sistema (por ejemplo, si debe iniciarse automáticamente o mediante un botón de usuario). Este enfoque asegura que el chatbot no solo se integre técnicamente, sino que también se adapte visual y funcionalmente al entorno de WordPress, proporcionando una experiencia de usuario optimizada y coherente.

Figura 15.

Página web de la OCA integrada con el chatbot



Fuente: Elaboración propia

Figura 16.

Plugin para integrar el chatbot a la OCA

```
<?php
/*
Plugin Name: Chatbot Integration
Description: Plugin para integrar un chatbot flotante desde otro dominio.
Version: 1.0
Author: Tu Nombre
*/

function chatbot_integration_scripts() {
    ?>
    <style>
        #chatbot-oca {
            position: fixed;
            bottom: 20px;
            right: 20px;
            z-index: 1000;
            width: 300px;
            height: 400px;
            border: none;
            display: none; /* Empieza oculto */
        }

        #chatbot-button {
            position: fixed;
            bottom: 20px;
            right: 20px;
            background-color: #007bff;
            color: white;
            padding: 10px 15px;
            border-radius: 50%;
            cursor: pointer;
            z-index: 1001;
        }
    </style>
    <div id="chatbot-button">Chat</div>
    <iframe id="chatbot-oca" src="https://chatbot-oca.com.pe"></iframe>

    <script>
        document.getElementById('chatbot-button').addEventListener('click', function() {
            var iframe = document.getElementById('chatbot-oca');
            if (iframe.style.display === "none") {
                iframe.style.display = "block";
                this.style.display = "none"; // Ocultar el botón cuando se muestre el chat
            } else {
                iframe.style.display = "none";
            }
        });
    </script>
<?php
}

add_action('wp_footer', 'chatbot_integration_scripts');
```

Fuente: Elaboración propia

CAPÍTULO V: DESCRIPCIÓN DEL SISTEMA

El objetivo de este capítulo es desglosar y explicar los componentes técnicos clave del sistema de inteligencia conversacional implementado en la OCA. Aquí se detallarán los requerimientos funcionales y no funcionales del sistema, se identificarán los actores y casos de uso principales, y se presentará la interfaz del sistema, proporcionando una visión completa de cómo funciona el sistema y cómo interactúan los usuarios con él.

5.1 Requerimientos del sistema

En esta sección se identifican y describen los requerimientos necesarios para el correcto funcionamiento del sistema. Estos requerimientos son fundamentales para asegurar que el sistema cumpla con las expectativas tanto técnicas como de usuario final.

5.1.1 *Requerimientos Funcionales del Sistema*

Los requerimientos funcionales del sistema describen las funciones específicas que el sistema debe realizar. Estos requerimientos definen las capacidades del sistema para interactuar con los usuarios y manejar la información de manera eficaz. En la Tabla 6 se mencionan los requerimientos funcionales del sistema.

Tabla 6.

Requerimientos funcionales del Sistema

Requerimiento	Descripción
RF-01	El sistema debe permitir la actualización de la base de conocimientos en caso esté disponible nueva información o corrección de la misma.
RF-02	El sistema debe integrar la API de Azure OpenAI Service para procesar las consultas de los usuarios, proporcionando respuestas coherentes y contextualmente relevantes.
RF-03	El sistema debe registrar y almacenar las interacciones de los usuarios para análisis

	posterior, con el fin de mejorar continuamente la precisión y relevancia de las respuestas del chatbot.
RF-04	El sistema debe permitir limpiar las conversaciones actuales para iniciar una nueva conversación en caso lo requiera.
RF-05	El sistema debe gestionar la escalabilidad para soportar múltiples usuarios concurrentemente sin degradar el rendimiento, asegurando tiempos de respuesta rápidos y eficientes en todo momento.

Fuente: Elaboración propia

5.1.2 *Requerimientos no Funcionales del Sistema*

Los requerimientos no funcionales se centran en aspectos como la escalabilidad, la usabilidad, el rendimiento y la seguridad del sistema. Estos requerimientos aseguran que el sistema opere de manera eficiente en diferentes situaciones y ofrezca una experiencia de usuario segura y satisfactoria. En la Tabla 7 se detallan los requerimientos no funcionales del sistema.

Tabla 7.

Requerimientos no funcionales del Sistema

Requerimiento	Descripción
RNF-01	El sistema debe proporcionar una interfaz gráfica intuitiva y accesible, compatible con dispositivos móviles y de escritorio, facilitando la navegación y el uso del sistema por parte de los usuarios.
RNF-02	El sistema debe permitir a los usuarios realizar consultas sobre procesos de admisión, brindando respuestas detalladas y actualizadas en función de la información almacenada en la base de conocimientos.
RNF-03	El sistema debe estar disponible y operativo el 99.9% del tiempo, asegurando acceso ininterrumpido a los usuarios.

RNF-04	El Sistema debe ser construido utilizando tecnologías como React Js, Python (Flask) y tecnologías Azure.
--------	---

Fuente: Elaboración propia

5.2 Casos de Uso del Sistema

Este capítulo describe las interacciones principales entre los usuarios y el sistema, utilizando diagramas y descripciones textuales para ilustrar cómo se llevan a cabo las funcionalidades clave.

5.2.1 Actores del Sistema

En la siguiente tabla se detallan los actores del sistema:

Tabla 8.

Actores del Sistema

Actor	Requerimientos
Usuario (Estudiante/Público General)	Persona que interactúa con el chatbot para obtener información sobre admisión, carreras, y otros servicios académicos.
Administrador del Sistema	Persona encargada de la configuración, mantenimiento, y actualización de la base de conocimientos y del sistema en general.

Fuente: Elaboración propia

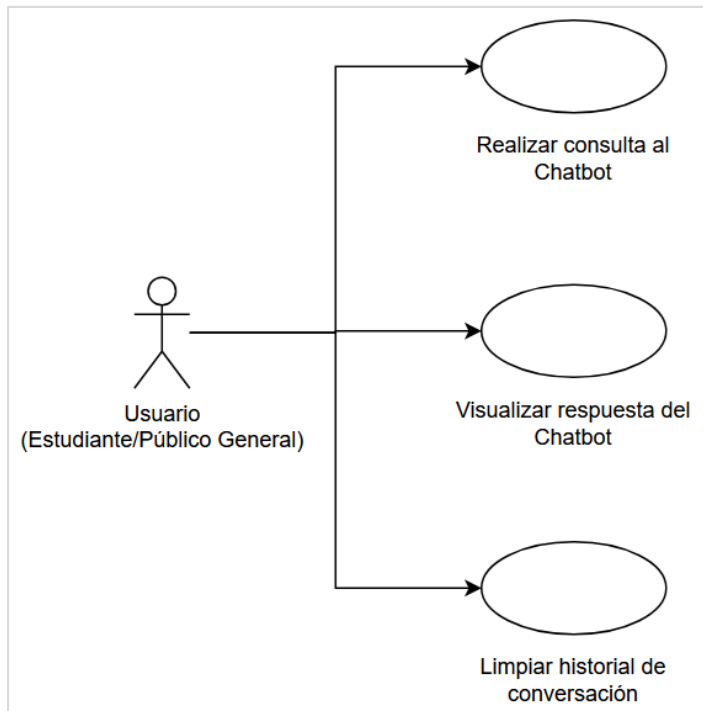
5.2.2 Diagramas de Casos de Uso del Sistema

Aquí se incluyen los diagramas UML que visualizan los casos de uso del sistema. Los diagramas muestran las interacciones entre los actores y el sistema, destacando cómo se cumple cada requerimiento funcional.

En la Figura 17 se describe el primer actor: Usuario

Figura 17.

Interacción del Usuario (Estudiante/Público en general) con el sistema

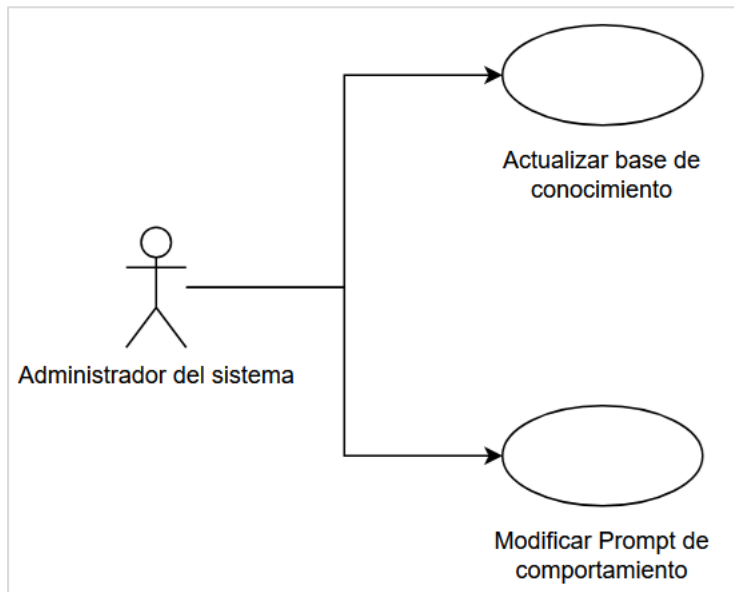


Fuente: Elaboración propia

En la Figura 18 se describe el primer actor: Administrador del sistema

Figura 18.

Interacción del Administrador del Sistema con el Sistema



Fuente: Elaboración propia

5.2.3 Especificaciones de Casos de Uso del Sistema

Casos de uso del primer actor: Usuario (Estudiante/Público en general)

- CUS Realizar consulta al chatbot

Tabla 9.

Especificaciones del caso de uso del sistema “Realizar consulta al chatbot”

CUS_01	Realizar consulta al chatbot
Actor	Usuario (Estudiante/Público en general)
Descripción	El caso de uso describe como el usuario puede realizar cualquier tipo de consulta relacionada a los procesos de admisión que ofrece la OCA, así como cualquier información de interés general que ofrezca dicha entidad.
Precondiciones	El usuario debe iniciar una conversación
Flujo de Eventos	

1. El sistema muestra una interfaz gráfica que permite la redacción de las consultas del usuario
2. El usuario ingresa su consulta
3. El sistema devuelve una respuesta en base a la pregunta realizada por el usuario
Flujo de Excepcionales

Fuente: Elaboración propia

- CUS Visualizar respuesta del chatbot

Tabla 10.

Especificaciones del caso de uso del sistema “Visualizar respuesta del chatbot”

CUS_02	Visualizar respuesta del chatbot
Actor	Usuario (Estudiante/Público en general)
Descripción	El caso de uso describe como el usuario visualiza la respuesta del chatbot al procesar la pregunta anteriormente realizada.
Precondiciones	El usuario debe iniciar una conversación El usuario debe haber realizado una pregunta al chatbot
Flujo de Eventos	
1. El sistema muestra el resultado en base a la pregunta realizada por el usuario	
2. El sistema le sugiere si quiere realizar alguna pregunta adicional	
Flujo de Excepcionales	
1. Si el usuario ingresa una pregunta fuera del contexto de la base de conocimientos el sistema responderá “No he encontrado información con respecto a tu pregunta. ¿Puedo ayudarte en algo más?”	

Fuente: Elaboración propia

- CUS Limpiar historial de conversación

Tabla 11.

Especificaciones del caso de uso del sistema “Limpiar historial de conversación”

CUS_03	Limpiar historial de conversación
Actor	Usuario (Estudiante/Público en general)
Descripción	El caso de uso describe como el usuario puede limpiar el historial de conversación permitiéndole iniciar una nueva conversación.
Precondiciones	El usuario debe iniciar una conversación El usuario debe haber realizado una pregunta al chatbot
Flujo de Eventos	
1. El sistema muestra una interfaz gráfica que permite la redacción de las consultas del usuario 2. El usuario ingresa una pregunta al sistema 3. El sistema muestra el resultado en base a la pregunta realizada por el usuario 4. El usuario selecciona la opción “limpiar” ubicada al lado izquierdo de la caja de preguntas 5. El sistema borra el historial de conversación actual	
Flujo de Excepcionales	
1. Si el usuario no ha realizado ninguna consulta al sistema, la funcionalidad de “borrar historial de conversaciones” estará deshabilitada.	

Fuente: Elaboración propia

Casos de uso del segundo actor: Administrador del Sistema

- CUS Actualizar base de conocimientos

Tabla 12.

Especificaciones del caso de uso del sistema “Actualizar base de conocimientos”

CUS_01	Actualizar base de conocimientos
Actor	Administrador del Sistema
Descripción	El caso de uso describe como el administrador del sistema puede actualizar la base de conocimientos.
Precondiciones	El usuario debe iniciar sesión en el Portal de Azure El usuario debe ingresar al Azure Storage Account
Flujo de Eventos	
1. El usuario debe iniciar sesión en el Portal de Azure y buscar el recurso Azure Storage Account 2. El sistema muestra la lista de Azure Storage Account que se tienen aprovisionadas 3. El usuario debe seleccionar el recurso “ContainerOcaGpt” 4. El sistema muestra las opciones del recurso “ContainerOcaGpt” 5. El usuario debe seleccionar en la sección “Data Storage” la opción “Containers” 6. El sistema muestra la lista de contenedores 7. El usuario debe seleccionar el contenedor llamado “propn” 8. El sistema muestra los documentos almacenados en el contenedor “propn” 9. El usuario reemplaza o sube nuevos documentos a la base de conocimientos 10. El sistema confirma la subida de archivos	
Flujo de Excepcionales	

Fuente: Elaboración propia

- CUS Modificar Prompt de comportamiento

Tabla 13.

Especificaciones del caso de uso del sistema “Modificar Prompt de comportamiento”

CUS_02	Modificar Prompt de comportamiento
Actor	Administrador del Sistema
Descripción	El caso de uso describe como el administrador del sistema puede actualizar el prompt de comportamiento.
Precondiciones	El usuario debe iniciar sesión en el Portal de Azure El usuario debe ingresar al Azure OpenAI Service
Flujo de Eventos	
1. El usuario debe iniciar sesión en el Portal de Azure y buscar el recurso Azure OpenAI Service 2. El sistema muestra la lista de Azure OpenAI Service que se tienen aprovisionadas 3. El usuario debe seleccionar el recurso “OpenAiServiceOcaGpt” 4. El sistema muestra las opciones del recurso “OpenAiServiceOcaGpt” 5. El usuario debe seleccionar en la opción “Explore and deploy” 6. El usuario debe seleccionar la opción “Chat” 7. El usuario sube la configuración actual del sistema el cual está en un archivo JSON 8. El sistema muestra la configuración actual del sistema 9. El usuario actualiza el prompt de comportamiento 10. El sistema confirma la actualización del prompt de comportamiento	
Flujo de Excepcionales	

Fuente: Elaboración propia

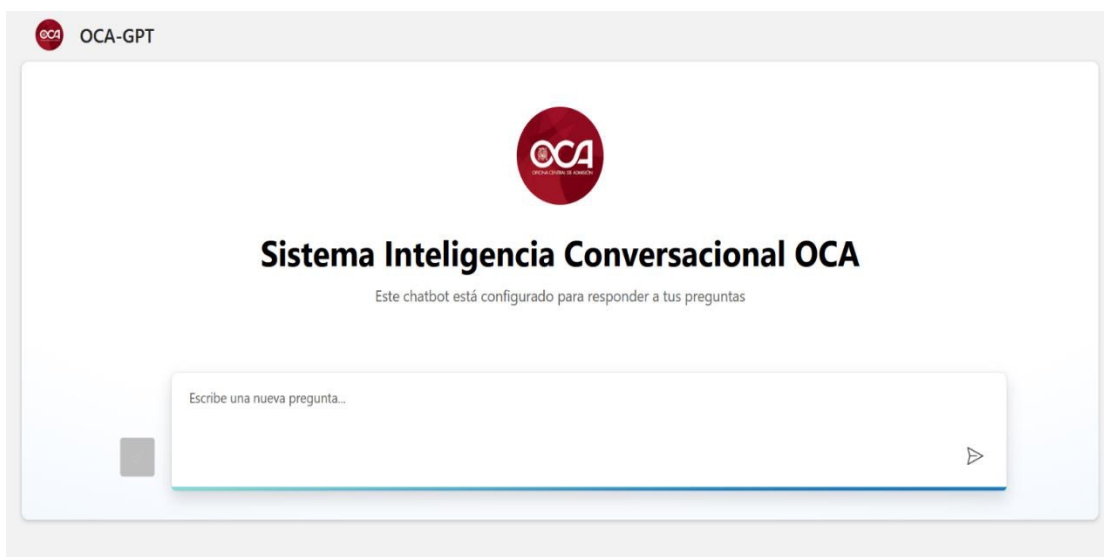
5.3 Interfaz del sistema

5.3.1 Interfaz web de la aplicación

En esta sección se muestra la interfaz gráfica del chatbot conversacional. Se explica cada una de sus funcionalidades. Adicionalmente se muestra que mientras no se tenga una conversación la opción de “Eliminar historial de conversaciones” se encuentra deshabilitada. En la Figura 19 se muestra una captura de su pantalla inicial.

Figura 19.

Pantalla inicial del chatbot



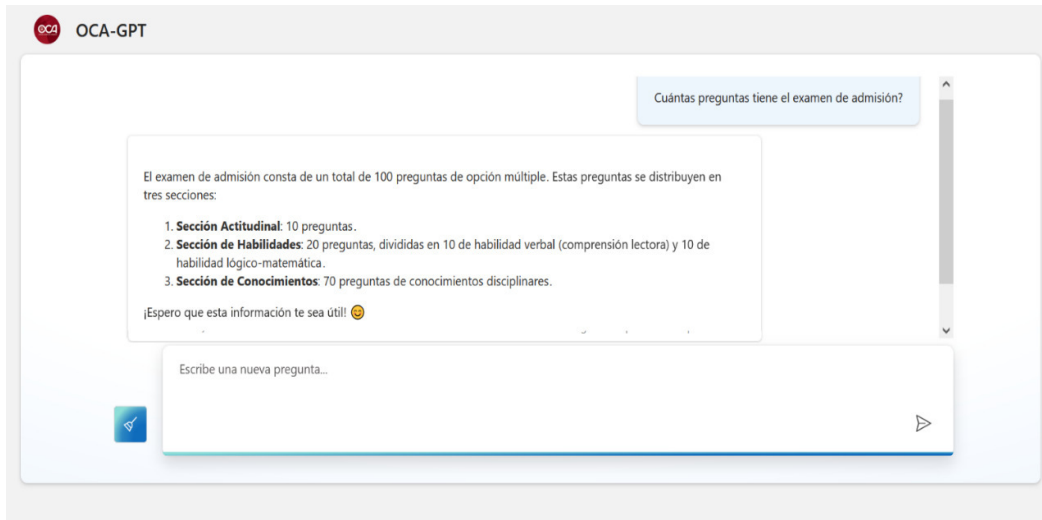
Fuente: Elaboración propia

I. Módulo de consultas y respuesta

En la siguiente interfaz se muestra una pregunta realizada por un usuario y se visualiza como el sistema procesa su pregunta y posteriormente genera una respuesta (Ver Figura 20).

Figura 20.

Interacción entre el Usuario y el chatbot



Fuente: Elaboración propia

II. Configuración de la base de conocimiento y prompt de comportamiento

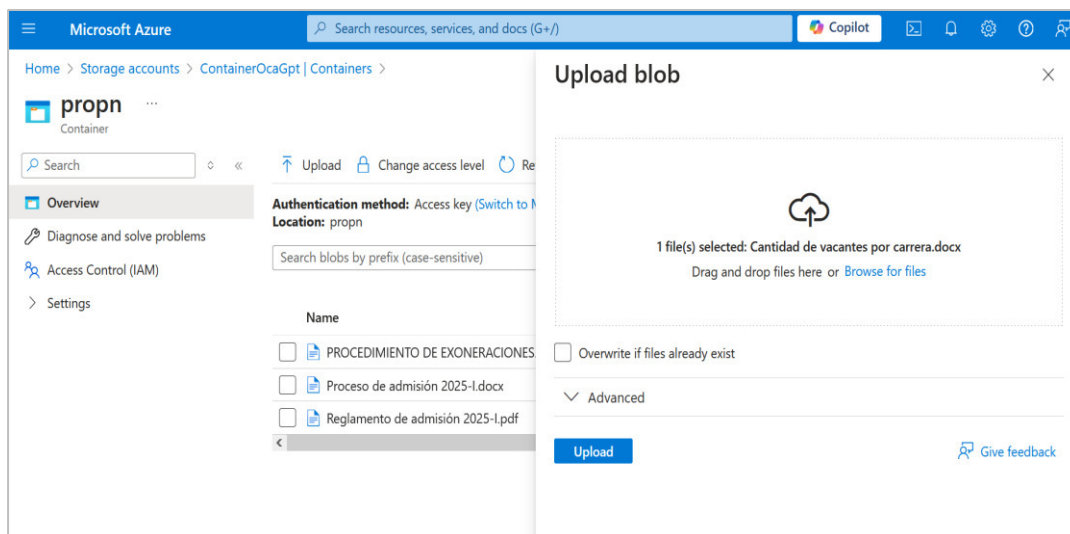
En las siguientes capturas se muestra cómo se configuran la base de conocimientos y el prompt de comportamiento.

Configuración de base de conocimientos

La siguiente captura muestra como subir un nuevo documento no estructurado que formará parte de la base de conocimientos del chatbot conversacional (Ver Figura 21).

Figura 21.

Subir nuevo documento no estructurado como base de conocimiento



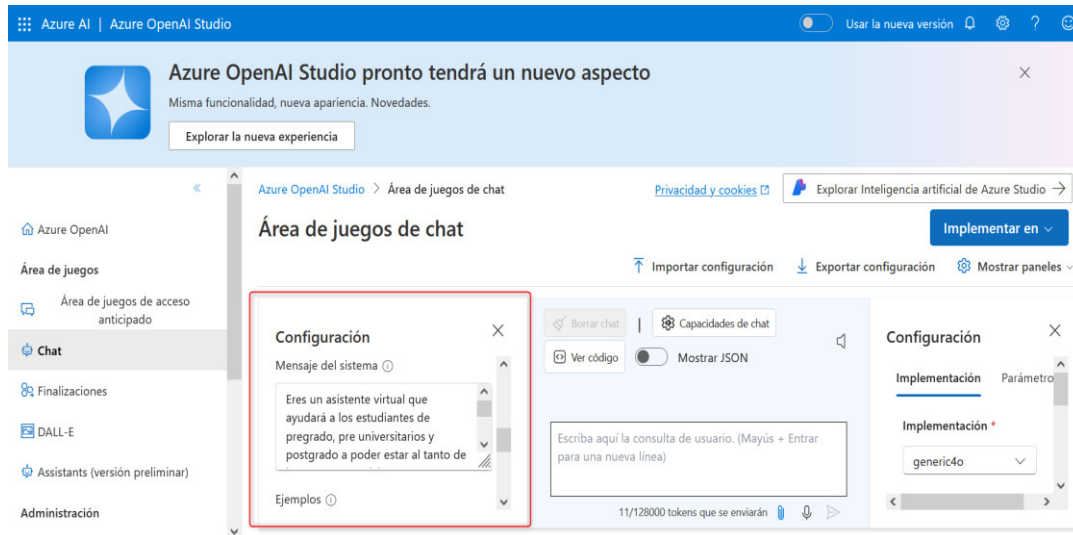
Fuente: Microsoft Azure Portal

Configuración del prompt de comportamiento

La siguiente captura muestra como configurar el prompt de comportamiento para el chatbot conversacional (Ver Figura 22).

Figura 22.

Captura que muestra la configuración del prompt de comportamiento



Fuente: Microsoft Azure Portal

CAPÍTULO VI: RESULTADOS

En este capítulo, se presentarán los resultados obtenidos tras la implementación del chatbot en la OCA de la UNMSM. El objetivo principal es evaluar el impacto de esta herramienta tecnológica en la mejora de la atención al usuario, comparando su eficacia con los métodos tradicionales, como las visitas presenciales, las llamadas telefónicas y el uso del portal web. Para ello, se realizó una encuesta a 50 personas, entre estudiantes y padres de familia, donde se midieron variables clave como el tiempo de atención, el Net Promoter Score (NPS) para evaluar la satisfacción del usuario, y la veracidad de la información proporcionada por el chatbot.

Estas variables se seleccionaron con el propósito de obtener una visión integral sobre cómo el chatbot ha influido en la experiencia del usuario. El tiempo de atención se midió para determinar la eficiencia del sistema en comparación con los canales tradicionales. El NPS se utilizó para medir la satisfacción general de los usuarios con el servicio proporcionado. Finalmente, la veracidad de la información se evaluó para garantizar que el chatbot no solo sea rápido y satisfactorio, sino también preciso y confiable en las respuestas que ofrece.

6.1 Resultados de la propuesta

En esta sección analizaremos los resultados de cada una de las variables que se midieron para evidenciar la mejora del performance de los usuarios con respecto a los métodos tradicionales de consulta para obtener información de los procesos que ofrece de la OCA.

6.1.1 Evaluación del Tiempo de Atención

Analizaremos el tiempo de atención promedio que los usuarios experimentaron al utilizar el chatbot en comparación con los métodos tradicionales de consulta. La eficiencia del chatbot se evaluó en términos de la rapidez con la que los usuarios podían obtener respuestas a sus consultas, y cómo esto se compara con el tiempo de espera y la atención recibida a través de visitas

presenciales, llamadas telefónicas y búsquedas en el portal web.

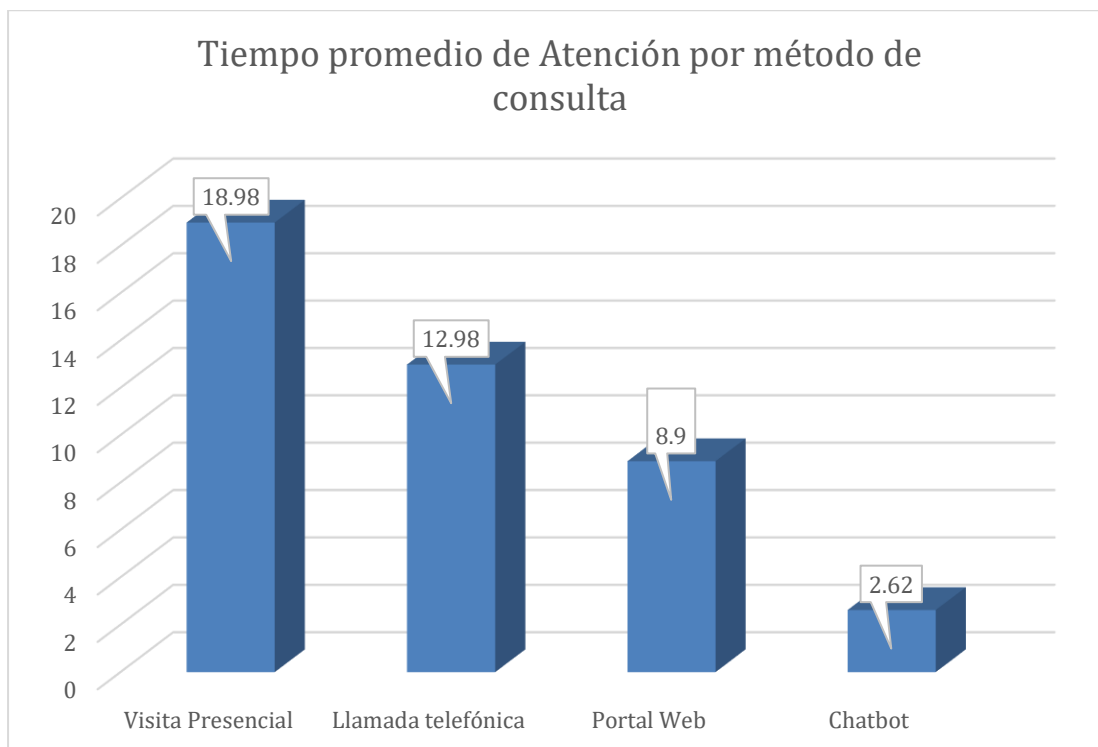
Resultados de la encuesta:

Los tiempos de atención recopilados durante la encuesta, donde 50 personas interactuaron con cada uno de los métodos disponibles para obtener información o resolver sus dudas se muestran en el **Anexo 1: Tiempo de atención por método de consulta**

Para visualizar mejor los resultados, se puede crear un gráfico de barras que compare el tiempo promedio de atención para cada método (Visita Presencial, Llamada Telefónica, Portal Web y chatbot).

Figura 23.

Tiempo promedio de atención por método de consulta



Fuente: Elaboración propia

Interpretación:

Los datos indican que el tiempo promedio de atención a través del chatbot es notablemente menor que en otros métodos, con un promedio de 2 a 3 minutos frente a los 10 a 25 minutos de las visitas presenciales y llamadas telefónicas. Esta reducción no solo mejora la experiencia del usuario, sino que también permite a la OCA gestionar un mayor volumen de consultas en menos tiempo, optimizando recursos y aumentando la satisfacción con el servicio.

Variable de medición:

Variable: Reducción de los tiempos de espera y la carga en canales tradicionales

Resultado: Se evidenció una disminución del 92% en el tiempo promedio de atención en los canales tradicionales. Antes del sistema, la atención promedio era de 10 a 25 minutos, mientras que ahora se reporta un promedio de 3 minutos para la atención a los usuarios gracias a una atención personalizada y digital.

6.1.2 Evaluación del NPS del usuario

Validación científica de la metodología utilizada:

El Net Promoter Score fue introducido por Fred Reichheld en su artículo de 2003 "The One Number You Need to Grow", publicado en la revista Harvard Business Review. Reichheld argumentó que, en lugar de usar múltiples métricas de satisfacción del cliente, las empresas deberían concentrarse en una sola: la disposición de los clientes a recomendar un producto o servicio. La lógica detrás de esto es que los clientes satisfechos que recomendarían la marca son más propensos a ser leales, y esta lealtad se traduce en un crecimiento sostenido a largo plazo para la empresa.

Desde su introducción, diversos estudios empíricos han demostrado que el NPS tiene una correlación positiva con el crecimiento de ingresos, adquisición de clientes y retención de clientes

en varias industrias.

Es así que el Net Promoter Score (NPS) es una métrica ampliamente utilizada para evaluar la lealtad y satisfacción de los clientes, basada en la pregunta clave: "¿Qué tan probable es que recomiende nuestra empresa/producto/servicio a un amigo o colega?" Esta métrica permite segmentar a los clientes en tres grupos según su respuesta a una escala de 0 a 10.

Utilidad del NPS

- **Medición de la lealtad del cliente:** El NPS es una métrica directa y sencilla que permite conocer cómo perciben los clientes una empresa, marca o producto, y qué tan dispuestos están a recomendarla.
- **Monitoreo de la satisfacción del cliente:** A lo largo del tiempo, el NPS proporciona una visión clara de cómo cambian las percepciones y la satisfacción del cliente, lo que facilita la toma de decisiones en cuanto a estrategias de mejora.
- **Identificación de áreas de mejora:** Al analizar las respuestas de los detractores, las empresas pueden identificar áreas clave en las que necesitan mejorar, ya sea en la calidad del producto, el servicio al cliente o la experiencia general.
- **Comparación entre empresas:** El NPS es un estándar reconocido que permite comparar fácilmente la lealtad y satisfacción de los clientes en diferentes empresas o industrias.

Estudios Académicos sobre la Validez del NPS:

Numerosos estudios académicos han validado el NPS como una métrica confiable de satisfacción y lealtad del cliente. Algunos de estos estudios incluyen:

- Reichheld (2003): En su trabajo fundacional, Reichheld no solo introduce el NPS, sino que también establece su validez a través de un análisis extenso de su relación con el rendimiento empresarial. En este estudio, se mostró que el NPS era una métrica confiable

para predecir la rentabilidad y el crecimiento, especialmente en mercados competitivos.

- McKinsey & Company (2010): Un informe de McKinsey analizó la efectividad del NPS en la mejora de la experiencia del cliente, y concluyó que esta métrica era útil para identificar no solo la lealtad de los clientes, sino también las áreas de mejora dentro de la experiencia de servicio.
- Aaker, Kumar y Day (2010): En su libro Marketing Research, los autores afirman que el NPS ha ganado aceptación en la investigación de mercados debido a su capacidad para medir de manera precisa la lealtad del cliente y su potencial para predecir la recomendación de un producto o servicio.

En este caso, se utilizó el NPS para medir la satisfacción de los usuarios después de interactuar con el chatbot implementado en la OCA. La encuesta fue realizada a 150 personas, incluyendo estudiantes de secundaria, estudiantes de academias preuniversitarias, y padres de familia. Se les preguntó sobre su disposición a recomendar el servicio del chatbot a otros usuarios, basándose en su experiencia.

El NPS se calcula a través de preguntas que miden el nivel de satisfacción que tienen los usuarios en base a su experiencia con el chatbot. Las respuestas se dividen en tres categorías:

- Promotores (9-10): Usuarios extremadamente satisfechos que estarían dispuestos a recomendar el servicio.
- Pasivos (7-8): Usuarios satisfechos, pero no entusiastas, que podrían ser susceptibles a cambiar de opinión.
- Detractores (0-6): Usuarios insatisfechos que no recomendarían el servicio.

El NPS se calcula restando el porcentaje de detractores del porcentaje de promotores. Este puntaje puede variar entre -100 y +100, siendo un puntaje positivo indicativo de una mayor

satisfacción y lealtad de los usuarios.

A continuación, en la Tabla 14 se muestra el detalle de los porcentajes calculados:

Tabla 14.

Resultado de la medición de NPS

Pregunta	Opción 0-3	Opción 4-6	Opción 7-8	Opción 9-10	Total
¿Qué tan probable es que recomiende el uso del chatbot a un amigo o compañero?	5%	5%	8%	82%	100%
¿Qué tan satisfecho está con la rapidez de las respuestas proporcionadas por el chatbot?	7%	6%	10%	77%	100%
¿Qué tan fácil le resultó utilizar el chatbot para encontrar la información que necesitaba?	5%	6%	11%	78%	100%
¿Qué tan probable es que vuelva a utilizar el chatbot para futuras consultas?	4%	3%	11%	82%	100%
¿Qué tan útil considera que es el	3%	6%	10%	81%	100%

chatbot en comparación con otros métodos de atención al cliente?					
--	--	--	--	--	--

Fuente: Elaboración propia

En esta tabla, se desglosa cómo los usuarios respondieron a cada pregunta en una escala de 0 a 10, con la distribución de respuestas categorizada según los siguientes grupos:

- **Opción 0-3:** Representa usuarios insatisfechos (Detractores).
- **Opción 4-6:** Representa usuarios que están insatisfechos, pero no de manera crítica (Detractores).
- **Opción 7-8:** Representa usuarios satisfechos, pero no entusiastas (Pasivos).
- **Opción 9-10:** Representa usuarios altamente satisfechos (Promotores).

En la encuesta realizada, los resultados mostraron lo siguiente:

Tabla 15.

Categoría x Porcentaje de Usuarios

Categoría	Porcentaje de Usuarios
Promotores	80%
Pasivos	10%
Detractores	10%

Fuente: Elaboración propia

El cálculo del NPS se realizó de la siguiente manera:

$$NPS = Promotores (\%) - Detractores (\%)$$

$$NPS = 80\% - 10\% = 70\%$$

Con un NPS de +50, el chatbot demuestra un alto nivel de satisfacción entre los usuarios, lo que refleja la eficiencia y efectividad del sistema en la atención a los usuarios de la OCA. Este resultado es especialmente significativo al compararlo con los métodos tradicionales de atención, que históricamente han mostrado menores niveles de satisfacción debido a tiempos de espera prolongados y una experiencia de usuario menos interactiva.

Interpretación:

El NPS indica que la mayoría de los usuarios tuvo una experiencia positiva con el chatbot y estarían dispuestos a recomendarlo, lo que sugiere una mejora en la percepción del servicio de la OCA. La baja proporción de detractores señala que el sistema cumple con las expectativas, proporcionando respuestas precisas y eficientes. Un NPS positivo resalta la confianza en el chatbot, lo que es crucial para su adopción futura. Además, ha humanizado la interacción, ofreciendo un servicio accesible y fácil de usar, lo que puede fomentar la lealtad de los usuarios.

Variable de medición:

Variable: Experiencia fluida y eficiente en la interacción virtual

Resultado: Con un NPS del 80% evidenciamos que el sistema cumple los estándares de excelencia y practicidad con respecto a la interfaz y la interacción con el chatbot lo cual incrementa la fidelidad de los usuarios con el sistema, así como su difusión a más personas dentro del entorno académico y público en general.

6.1.3 Evaluación de la calidad de los resultados

El objetivo de esta sección es analizar la precisión y veracidad de las respuestas

proporcionadas por el chatbot de la OCA. Dado que la calidad de la información es crucial para garantizar la confianza del usuario, se realizó una evaluación exhaustiva de cómo el chatbot maneja y responde a las consultas. La evaluación se centró en la capacidad del chatbot para ofrecer respuestas correctas, pertinentes y alineadas con la información oficial disponible en la base de conocimientos.

Metodología de Evaluación:

Para evaluar la calidad de los resultados, se diseñó un conjunto de preguntas que abarcaban diversos temas clave relacionados con la admisión y los servicios académicos ofrecidos por la OCA. Se seleccionaron 100 preguntas representativas, basadas en consultas reales de estudiantes y padres de familia, que se realizaron al sistema. Las respuestas del chatbot fueron luego comparadas con la información oficial proporcionada por la OCA para determinar su exactitud y relevancia.

La encuesta utilizada para la evaluación de usuarios se encuentra en el **Anexo 2**.

Validación científica de la metodología utilizada:

La matriz de confusión ha sido adoptada, de facto y de jure, como un estándar para informar sobre la exactitud temática de cualquier producto de datos derivados de la teledetección. Por supuesto, esta misma herramienta puede ser utilizada para evaluación de la calidad temática de cualquier tipo de dato espacial. En esta línea, la matriz de confusión aparece reconocida en la Norma Internacional ISO 19157, relativa a la calidad de la información geográfica, como un mecanismo para ofrecer los resultados de la calidad temática de productos vectoriales o derivados de imágenes (p.ej. clasificación de una imagen) (Ariza-López et al., 2018).

La matriz de confusión es una herramienta fundamental para la evaluación de modelos de clasificación en aprendizaje automático y minería de datos. Su utilidad radica en que proporciona

una representación visual y cuantitativa de los resultados obtenidos por un modelo de clasificación, comparando las predicciones realizadas con las clases reales o verdaderas. Esta matriz se utiliza para calcular varias métricas clave que permiten entender el rendimiento del modelo de manera detallada, tales como: Precisión, Recall y la especificidad.

Estas métricas derivadas de la matriz de confusión son esenciales para evaluar el desempeño de un modelo en contextos en los que tanto los falsos positivos como los falsos negativos pueden tener impactos significativos, como en diagnóstico médico, clasificación de spam, o detección de fraudes. La matriz de confusión permite a los investigadores y analistas no solo identificar el número de errores cometidos, sino también clasificar la naturaleza de esos errores, lo que es crucial para la mejora y ajuste del modelo.

Aplicación de la Matriz de Confusión:

La validez de la matriz de confusión radica en que proporciona una representación detallada, precisa y objetiva del desempeño del modelo, lo que permite realizar un análisis exhaustivo de su eficacia. A continuación, se presentan razones que fundamentan su validez científica:

- **Fundamentación estadística sólida:** La matriz de confusión se basa en principios estadísticos bien establecidos, como la teoría de la probabilidad y la teoría de la decisión, que permiten interpretar de manera clara los errores y aciertos del modelo.
- **Claridad en la interpretación de errores:** A diferencia de métricas generales como la exactitud, la matriz de confusión permite identificar específicamente los falsos positivos (errores donde el modelo clasifica incorrectamente una clase negativa como positiva) y los falsos negativos (errores donde el modelo clasifica incorrectamente una clase positiva como negativa), lo que es crucial para muchos problemas prácticos.
- **Aplicabilidad a modelos de clasificación binaria y multiclase:** Aunque la matriz de

confusión es comúnmente utilizada en clasificación binaria, también se puede extender a problemas de clasificación multiclase, proporcionando una visión detallada del rendimiento del modelo en cada clase.

- **Validez empírica:** La matriz de confusión ha sido ampliamente utilizada en una variedad de disciplinas, desde medicina (para evaluar modelos de diagnóstico) hasta finanzas (para la detección de fraudes), lo que respalda su validez empírica en escenarios reales.
- **Referencia en la literatura científica:** Autores clave en el campo de la estadística y el aprendizaje automático, como Jerome Friedman, Tom Mitchell, y Trevor Hastie, entre otros, han validado su utilidad y efectividad para la evaluación de modelos. Estos autores proporcionan una base sólida para la aplicación de la matriz de confusión, tanto en la práctica como en la investigación académica, lo que garantiza su relevancia y validez.

En conclusión, la matriz de confusión es una herramienta válida y ampliamente reconocida para evaluar el desempeño de los modelos de clasificación debido a su capacidad para proporcionar una representación detallada y comprensible de los errores y aciertos de un modelo, lo que facilita la mejora continua de los mismos. La validez científica de la matriz de confusión se respalda por su base teórica sólida, su aplicabilidad en una amplia variedad de contextos y su constante uso en la literatura especializada.

Aplicación de la Matriz de Confusión:

En base a lo expuesto anteriormente, para una evaluación más precisa de la calidad de las respuestas, se utilizó la técnica de la Matriz de Confusión. Esta técnica permitió analizar la relación entre las respuestas proporcionadas por el chatbot y la veracidad de estas, categorizando las respuestas en verdaderos positivos (VP), falsos positivos (FP), verdaderos negativos (VN), y falsos negativos (FN).

Definiciones en el contexto del chatbot:

- Verdaderos Positivos (VP): Respuestas correctas que el chatbot proporcionó y que coinciden con la información oficial.
- Falsos Positivos (FP): Respuestas incorrectas que el chatbot proporcionó, pero que fueron marcadas incorrectamente como correctas.
- Verdaderos Negativos (VN): Preguntas que el chatbot no debía responder y que efectivamente no respondió (por ejemplo, consultas fuera del ámbito de la OCA).
- Falsos Negativos (FN): Respuestas incorrectas o ausencia de respuesta cuando el chatbot debería haber dado una respuesta correcta.

En base a la técnica “Matriz de Confusión”, la Tabla 16 muestra el resultado de las 100 preguntas que se le formularon al sistema:

Tabla 16.

Análisis de respuesta según la Matriz de Confusión

	Predicción Positiva	Predicción Negativa
Real positivo	VP = 80	FN = 4
Real negativo	FP = 4	VN = 12

Fuente: Elaboración propia

- 80 fueron respuestas correctas (VP).
- 4 fueron respuestas incorrectas que el chatbot consideró correcta (FP).
- 12 fueron preguntas fuera del ámbito, que el chatbot correctamente evitó responder (VN).
- 4 fueron respuestas incorrectas o preguntas sin respuesta que deberían haber sido respondida correctamente (FN).

Cálculos Basados en la Matriz de Confusión:

Precisión (Accuracy):

- $$Precisión = \frac{VP+VN}{VP+FP+VN+FN} = \frac{80+12}{80+4+12+4} = \frac{92}{100} = 0.92$$

Esto indica que el chatbot fue preciso en el 92% de los casos.

Sensibilidad (Recall):

- $$Sensibilidad = \frac{VP}{VP+FN} = \frac{80}{80+4} = \frac{80}{84} \approx 0.952$$

El chatbot identificó correctamente el 95.2% de las respuestas que debía responder positivamente.

Especificidad:

- $$Especificidad = \frac{VN}{VN+FP} = \frac{12}{12+4} = \frac{12}{16} = 0.75$$

El chatbot evitó correctamente responder preguntas fuera de su ámbito en el 0.75% de los casos.

- **Precisión (92%):** Un alto nivel de precisión general indica que el chatbot es confiable para la mayoría de las interacciones con los usuarios.
- **Sensibilidad (95.2%):** La alta sensibilidad sugiere que el chatbot está bien entrenado para proporcionar respuestas correctas cuando es necesario.
- **Especificidad (0.75%):** La especificidad más baja indica que el chatbot podría mejorar en la identificación de preguntas fuera de su ámbito, lo que podría ser una oportunidad para ajustar su capacidad de reconocimiento de consultas irrelevantes o fuera del tema.

Interpretación:

La evaluación de la calidad de las respuestas del chatbot mediante la Matriz de Confusión ha mostrado resultados muy positivos, con un alto porcentaje de aciertos que indica su efectividad para proporcionar respuestas correctas y pertinentes. La baja incidencia de respuestas incorrectas

sugiere un entrenamiento adecuado del modelo, lo que refuerza su confiabilidad en la OCA. Además, la precisión en la diferenciación entre respuestas correctas y erróneas es crucial para mantener la confianza de los usuarios. La Matriz también ha identificado áreas de mejora, presentando oportunidades para optimizar el modelo y la base de conocimiento, asegurando un servicio de alta calidad.

Variable de medición:

Variable: Claridad y precisión de las respuestas

Resultado: A través de una matriz de confusión basada en 100 preguntas representativas, el sistema demostró un 92% de precisión en las respuestas generadas, con un 4% de respuestas parcialmente correctas y solo un 4% de respuestas incorrectas, confirmando su efectividad para proporcionar información clara y contextualizada.

CAPÍTULO VII: CONCLUSIONES Y TRABAJOS FUTUROS

7.1 Conclusiones

- El chatbot ha automatizado las respuestas a preguntas frecuentes, reduciendo en más de un 80% los tiempos de espera, los postulantes ahora reciben respuestas instantáneas, eliminando la necesidad de acudir a los canales tradicionales como llamadas o visitas presenciales. Esto ha reducido la carga en los canales tradicionales, mejorando la eficiencia general del proceso de atención.
- El sistema fue integrado con éxito en la página web de la OCA, proporcionando una experiencia fluida a los usuarios. Esto permitió elevar el NPS (satisfacción del cliente) hasta un 70% brindando una mejor experiencia al usuario alineándose con las expectativas modernas de atención virtual. La experiencia mejorada ha facilitado la accesibilidad a la información, optimizando el proceso de interacción.
- Finalmente, se midió la precisión de las respuestas utilizando la técnica de Matriz de confusión lo cual evaluó que el chatbot fue preciso en un 92% de los casos, además identificó correctamente el 95.2% de las respuestas que debía responder positivamente y evitó responder preguntas fuera de su ámbito en un 0.75% por lo cual las consultas fueron respondidas de manera coherente y contextualizada, mejorando la calidad de la atención. Los aspirantes ahora reciben información detallada, alineada con los procedimientos de la OCA, asegurando la confiabilidad del sistema.

7.2 Trabajos futuros

- Expansión de la base de conocimientos: Un enfoque futuro debería centrarse en la ampliación y diversificación de la base de conocimientos del chatbot. Esto incluiría la incorporación de información sobre situaciones excepcionales, casos especiales, y posibles

escenarios no cubiertos en la fase inicial. Asimismo, sería beneficioso automatizar el proceso de actualización de la base de conocimientos mediante la integración con sistemas de gestión de contenido institucional.

- Soporte multilingüe: Desarrollar capacidades multilingües para el chatbot permitirá atender a un público más diverso. La implementación de modelos de procesamiento del lenguaje natural que soporten múltiples idiomas, o la integración con servicios de traducción automática, facilitaría la comunicación con usuarios no hispanohablantes.

BIBLIOGRAFÍA

- Aktay, S., Gök, S., & Uzunoğlu, D. (2023). ChatGPT in Education. *Türk Akademik Yayınlar Dergisi*, 7(2), 378-406. <https://doi.org/10.29329/tayjournal.2023.543.03>
- Ali Abdulridha, M., Hadi Jassim Mohamed Faculty supervisor, A., & Ammar Alabbassi, L. (2024). *Virtual Agent (Chatbot) using Open Artificial Intelligence*.
- Arévalo, J., & Quinde, M. (2023). ChatGPT: La creación automática de contenidos con Inteligencia Artificial y su impacto en la comunicación académica y educativa. *Revista Desiderata*, 1-7.
- Ariza-López, F. J., Rodríguez-Avi, J., & Alba-Fernández, V. (2018). CONTROL Estricto DE MATRICES DE CONFUSIÓN POR MEDIO DE DISTRIBUCIONES MULTINOMIALES. *GeoFocus Revista Internacional de Ciencia y Tecnología de la Información Geográfica*, 215-226. <https://doi.org/10.21138/gf.591>
- Artiles, J., Guerra, M., Aguiar, V., & Rodríguez, J. (2021). Embodied conversational agents: Artificial intelligence for autonomous learning. *Pixel-Bit, Revista de Medios y Educacion*, 62, 107-144. <https://doi.org/10.12795/PIXELBIT.86171>
- Au Yeung, J., Kraljevic, Z., Luintel, A., Balston, A., Idowu, E., Dobson, R. J., & Teo, J. T. (2023). AI chatbots not yet ready for clinical use. *Frontiers in Digital Health*, 5. <https://doi.org/10.3389/fdgth.2023.1161098>
- Bumann, N. (2023). *Automated Chatbot Using Speech-to-Text and Text-to-Speech with Mobile App Integration*.
- Carlsson, K., & Jacobs, I. (2021). *Now Tech: Conversational Intelligence, Q2 2021 Forrester's Overview Of 41 Conversational Intelligence Providers Why Read This Report*.
- Cascella, M., Montomoli, J., Bellini, V., & Bignami, E. (2023). Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios. *Journal of Medical Systems*, 47(1). <https://doi.org/10.1007/s10916-023-01925-4>
- Celi-Parraga, R. J., Varela-Tapia, E. A., Acosta-Guzmán, I. L., & Montaña-Pulzara, N. R. (2021). Técnicas de procesamiento de lenguaje natural en la inteligencia artificial conversacional textual. *AlfaPublicaciones*, 3(4.1), 40-52. <https://doi.org/10.33262/ap.v3i4.1.123>
- Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education*, 20(1). <https://doi.org/10.1186/s41239-023-00408-3>
- Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., Ye, W., Zhang, Y., Chang, Y., Yu, P. S., Yang, Q., & Xie, X. (2024). A Survey on Evaluation of Large Language Models. *ACM Transactions on Intelligent Systems and Technology*, 15(3). <https://doi.org/10.1145/3641289>
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., ... Wright, R. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- Ganesh, K. (2024). *Leveraging Large Language Models for Generating Infrastructure as Code: Open and Closed Source Models and Approaches*.
- García-Díaz, J. A., Almela, A., Alcaraz-Marmol, G., & Valencia-García, R. (2020). UMCORPUSCLASSIFIER: Compilation and evaluation of linguistic corpus for Natural Language Processing tasks. *Procesamiento del Lenguaje Natural*, 65, 139-142. <https://doi.org/10.26342/2020-65-22>
- García-Peñalvo, F. J., Llorens-Largo, F., & Vidal, J. (2024). The new reality of education in the face of advances in generative artificial intelligence. *RIED-Revista Iberoamericana de Educacion a Distancia*, 27(1), 9-39. <https://doi.org/10.5944/ried.27.1.37716>
- Garibay, F. (2020). *DISEÑO E IMPLEMENTACIÓN DE UN ASISTENTE VIRTUAL (CHATBOT) PARA*

OFRECER ATENCIÓN A LOS CLIENTES DE UNA AEROLÍNEA MEXICANA POR MEDIO DE SUS CANALES CONVERSACIONALES.

- Gill, S. S., Xu, M., Patros, P., Wu, H., Kaur, R., Kaur, K., Fuller, S., Singh, M., Arora, P., Parlikad, A. K., Stankovski, V., Abraham, A., Ghosh, S. K., Lutfiyya, H., Kanhere, S. S., Bahsoon, R., Rana, O., Dustdar, S., Sakellariou, R., ... Buyya, R. (2024). Transformative effects of ChatGPT on modern education: Emerging Era of AI Chatbots. *Internet of Things and Cyber-Physical Systems*, 4, 19-23. <https://doi.org/10.1016/j.iotcps.2023.06.002>
- González, J., & Sánchez, I. (2021). *Procesamiento del Lenguaje Natural language processing*.
- Hadfi, R., Haqbeen, J., Sahab, S., & Ito, T. (2021). Argumentative Conversational Agents for Online Discussions. *Journal of Systems Science and Systems Engineering*, 30(4), 450-464. <https://doi.org/10.1007/s11518-021-5497-1>
- Hariri, W. (2024). *UNLOCKING THE POTENTIAL OF CHATGPT: A COMPREHENSIVE EXPLORATION OF ITS APPLICATIONS, ADVANTAGES, LIMITATIONS, AND FUTURE DIRECTIONS IN NATURAL LANGUAGE PROCESSING*.
- Hassani, H., & Silva, E. S. (2023). The Role of ChatGPT in Data Science: How AI-Assisted Conversational Interfaces Are Revolutionizing the Field. *Big Data and Cognitive Computing*, 7(2). <https://doi.org/10.3390/bdcc7020062>
- Hochmair, H. H., Juhász, L., & Kemp, T. (2024). Correctness Comparison of ChatGPT-4, Gemini, Claude-3, and Copilot for Spatial Tasks. *Transactions in GIS*. <https://doi.org/10.1111/tgis.13233>
- Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: state of the art, current trends and challenges. *Multimedia Tools and Applications*, 82(3), 3713-3744. <https://doi.org/10.1007/s11042-022-13428-4>
- Limna, P., Kraiwanit, T., Jangjarat, K., Klayklung, P., & Chocksathaporn, P. (2023). The use of ChatGPT in the digital era: Perspectives on chatbot implementation. *Journal of Applied Learning and Teaching*, 6(1), 64-74. <https://doi.org/10.37074/jalt.2023.6.1.32>
- López de Lacalle, M., Saralegi, X., Saizar, A., Urbizu, G., & Corral, A. (2023). Strategies for bilingual intent classification for small datasets scenarios. *Procesamiento del Lenguaje Natural*, 71, 137-147. <https://doi.org/10.26342/2023-71-11>
- Maddigan, P., & Susnjak, T. (2023). Chat2VIS: Generating Data Visualizations via Natural Language Using ChatGPT, Codex and GPT-3 Large Language Models. *IEEE Access*, 11, 45181-45193. <https://doi.org/10.1109/ACCESS.2023.3274199>
- Manjarrés-Betancur, R. A., & Echeverri-Torres, M. M. (2020). Asistente virtual académico utilizando tecnologías cognitivas de procesamiento de lenguaje natural. *Revista Politécnica*, 16(31), 85-95. <https://doi.org/10.33571/rpolitec.v16n31a7>
- Martínez, O. G. E., Reyes, S. V., & González, A. M. (2021). Uso de técnicas de procesamiento de lenguaje natural para la detección de requerimientos de software. *South Florida Journal of Development*, 2(5), 7323-7335. <https://doi.org/10.46932/sfjdv2n5-072>
- Mekni, M. (2021). An Artificial Intelligence Based Virtual Assistant Using Conversational Agents. *Journal of Software Engineering and Applications*, 14(09), 455-473. <https://doi.org/10.4236/jsea.2021.149027>
- Meskó, B., & Topol, E. J. (2023). The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *npj Digital Medicine*, 6(1). <https://doi.org/10.1038/s41746-023-00873-0>
- Michel-Villarreal, R., Vilalta-Perdomo, E., Salinas-Navarro, D. E., Thierry-Aguilera, R., & Gerardou, F. S. (2023). Challenges and Opportunities of Generative AI for Higher Education as Explained by ChatGPT. *Education Sciences*, 13(9). <https://doi.org/10.3390/educsci13090856>
- Morales-Chan, M., Amado-Salvatierra, H. R., Medina, J. A., Barchino, R., Hernández-Rizzardini, R., & Teixeira, A. M. (2024). Personalized Feedback in Massive Open Online Courses: Harnessing the Power of LangChain and OpenAI API. *Electronics (Switzerland)*, 13(10). <https://doi.org/10.3390/electronics13101960>
- Ots, K. (2024). *Hardening Azure OpenAI Service for Enterprise Usage*.

- Polverini, G., & Gregorcic, B. (2024). Evaluating vision-capable chatbots in interpreting kinematics graphs: a comparative study of free and subscription-based models. *Frontiers in Education*, 9. <https://doi.org/10.3389/feduc.2024.1452414>
- Rahman, M. M., & Watanobe, Y. (2023). ChatGPT for Education and Research: Opportunities, Threats, and Strategies. *Applied Sciences (Switzerland)*, 13(9). <https://doi.org/10.3390/app13095783>
- Rangapur, A., & Rangapur, A. (2024). *The Battle of LLMs: A Comparative Study in Conversational QA Tasks*. <http://arxiv.org/abs/2405.18344>
- Ruiz-Rojas, L. I., Acosta-Vargas, P., De-Moreta-Llovet, J., & Gonzalez-Rodriguez, M. (2023). Empowering Education with Generative Artificial Intelligence Tools: Approach with an Instructional Design Matrix. *Sustainability (Switzerland)*, 15(15). <https://doi.org/10.3390/su151511524>
- Scanlon, M., Breiting, F., Hargreaves, C., Hilgert, J. N., & Sheppard, J. (2023). ChatGPT for digital forensic investigation: The good, the bad, and the unknown. *Forensic Science International: Digital Investigation*, 46. <https://doi.org/10.1016/j.fsidi.2023.301609>
- Suárez, A., Jiménez, J., Llorente de Pedro, M., Andreu-Vázquez, C., Díaz-Flores García, V., Gómez Sánchez, M., & Freire, Y. (2024). Beyond the Scalpel: Assessing ChatGPT's potential as an auxiliary intelligent virtual assistant in oral surgery. *Computational and Structural Biotechnology Journal*, 24, 46-52. <https://doi.org/10.1016/j.csbj.2023.11.058>
- Taecharungroj, V. (2023). "What Can ChatGPT Do?" Analyzing Early Reactions to the Innovative AI Chatbot on Twitter. *Big Data and Cognitive Computing*, 7(1). <https://doi.org/10.3390/bdcc7010035>
- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10(1). <https://doi.org/10.1186/s40561-023-00237-x>
- Valeri, A. (2022). *AI-Powered Platforms: automated transactions in digital marketplaces*. <https://www.researchgate.net/publication/369762615>
- Varela-Tapia, E. A., Sangacha-Tapia, L. M., Acosta-Guzmán, I. L., & Celi-Párraga, R. J. (2021). Inteligencia Artificial Conversacional para el diseño de un asistente virtual interactivo. *CICIC 2022 - Decima Segunda Conferencia Iberoamericana de Complejidad, Informatica y Cibernetica en el contexto de the 13th International Multi-Conference on Complexity, Informatics, and Cybernetics, IMCIC 2022 - Memorias*, 237-241. <https://doi.org/10.54808/CICIC2022.01.237>

ANEXOS

Anexo 1. Tiempo de atención por método de consulta

Encuesta	Visita Presencial (minutos)	Llamada telefónica (minutos)	Portal Web (minutos)	Chatbot (minutos)
1	15	12	8	2
2	18	10	7	3
3	20	14	9	2
4	12	11	6	2
5	17	13	7	3
6	21	12	8	2
7	17	14	10	2
8	16	11	6	3
9	19	13	10	2
10	18	10	11	2
11	16	12	7	3
12	17	11	11	3
13	17	15	9	2
14	20	17	10	4
15	23	14	6	3
16	21	18	5	2
17	15	15	8	1

18	18	10	9	5
19	22	12	9	3
20	18	17	6	2
21	19	19	10	2
22	27	16	11	3
23	25	14	14	2
24	18	18	9	4
25	14	10	6	3
26	19	9	15	3
27	20	14	7	2
28	17	17	9	2
29	18	12	10	4
30	20	17	11	3
31	31	11	8	2
32	27	10	10	3
33	25	14	9	2
34	17	9	8	4
35	15	15	12	3
36	16	14	9	2
37	19	13	14	3
38	20	10	11	3
39	16	14	9	2
40	17	11	8	2

41	22	15	10	4
42	15	8	9	3
43	16	11	10	3
44	17	14	7	2
45	20	12	6	2
46	21	11	9	2
47	17	10	5	2
48	19	15	11	3
49	20	12	9	2
50	22	13	7	3

Anexo 2. Encuesta para evaluar el nivel de veracidad en las respuestas del
Chatbot

Encuesta para medir nivel de veracidad	
Nº	Pregunta
1	¿Cuántas preguntas tiene el examen de admisión?
2	¿Cuánto es el costo del examen de admisión?
3	¿A qué carreras se puede postular al examen general de admisión?
4	¿Qué documentos se necesitan para tramitar la constancia de ingreso después de alcanzar una vacante?
5	¿Cuáles son las carreras con mayor cantidad de postulantes en el proceso de admisión?
6	¿Cuántas vacantes se ofrecen en total para el proceso de admisión 2025-I?
7	¿Cuáles son las fechas del examen de admisión para las diferentes áreas?
8	¿Qué pasos se deben seguir para inscribirse en el examen de admisión?
9	¿Cuáles son las características de la foto que debe cargarse para el proceso de admisión?
10	¿Cuándo es el simulacro del examen de admisión 2025-I?
11	¿Cuál es el cronograma de inscripción para el examen de admisión 2025-I?
12	¿Cuáles son los temas que vendrán para el curso de Razonamiento matemático?

13	¿Qué carreras ofrece el área C dentro de la Universidad?
14	¿Qué documentos debo portar el día del examen de admisión?
15	¿Cuál es el valor de cada pregunta actitudinal en el examen de admisión?
16	¿Cuál es el contenido del examen de admisión para la carrera de Medicina?
17	¿Dónde se publicarán los resultados del examen de admisión?
18	¿Cuáles son las causas para que se me anule el examen de admisión?
19	¿Cuándo se me puede exonerar el pago del examen de admisión?
20	¿Si soy de bajos recursos económicos debo pagar el examen de admisión? ¿Qué documentos debo presentar?
21	Si califico para que me exoneren el examen de admisión. ¿En qué fechas debo inscribirme al examen?
22	¿Si mi apellido es Ramírez en qué fecha debo matricularme?
23	¿Cuáles son las fechas de entrega de constancias si aprobé el examen de admisión?
24	¿Es necesario el certificado de estudios del colegio si ingreso a la universidad?
25	¿Cuáles son los horarios para el ingreso el día del examen?
26	¿Qué debo hacer si no puedo asistir al examen de admisión en la fecha asignada?
27	¿Cómo puedo conocer la lista de los postulantes habilitados para rendir el examen?
28	¿Qué hacer si cometí un error al llenar mi formulario de inscripción?
29	¿Puedo cambiar mi carrera una vez me haya inscrito para el examen de admisión?
30	¿Cómo me informo sobre las becas o ayudas económicas para el examen de admisión?
31	¿Qué tipo de examen debo rendir si quiero postularme a una carrera de Ingeniería?
32	¿Puedo inscribirme al examen de admisión si soy extranjero?
33	¿Cuáles son los criterios de evaluación para el examen de admisión?
34	¿Qué pasa si no obtengo una vacante en la carrera a la que postulé?
35	¿Dónde se realizan los exámenes de admisión?
36	¿Puedo presentar el examen de admisión si no tengo los 18 años cumplidos?
37	¿Cómo puedo verificar que mi inscripción al examen de admisión se realizó correctamente?
38	¿El examen de admisión incluye preguntas sobre historia universal?
39	¿Existen descuentos para grupos de estudiantes que se inscriben juntos al examen de admisión?
40	¿Se puede presentar el examen de admisión en modalidad virtual?
41	¿Qué pasa si me enfermo antes del examen de admisión y no puedo asistir?
42	¿Puedo presentar el examen si no soy egresado de secundaria?
43	¿A qué hora comienza el examen de admisión el día de la prueba?
44	¿Qué tipo de preguntas se hacen en la sección de comprensión lectora del examen?
45	¿Cuántas veces puedo presentar el examen de admisión en un mismo año?
46	¿Es obligatorio presentar el examen de admisión si soy postulante a una carrera en modalidad a distancia?

47	¿Cuántos puntos se necesitan para aprobar el examen de admisión?
48	¿Existen requisitos específicos de edad para postularme al examen de admisión?
49	¿Cómo puedo obtener mi constancia de egresado si aprobé el examen de admisión?
50	¿Puedo obtener más de una vacante si me postulo a diferentes carreras?
51	¿Puedo solicitar un cambio de carrera después de ingresar a la universidad?
52	¿Cuáles son los requisitos para postularme a una carrera con modalidad presencial?
53	¿Hay alguna exoneración del examen de admisión para personas con discapacidad?
54	¿Cuáles son las fechas límite para presentar la documentación requerida?
55	¿Existen opciones de financiamiento para los estudiantes que aprueben el examen de admisión?
56	¿Puedo postularme a más de una carrera en el examen de admisión?
57	¿El examen de admisión es el mismo para todas las modalidades de estudio?
58	¿Qué debo hacer si no puedo entregar los documentos solicitados en la fecha indicada?
59	¿Hay alguna exoneración para los hijos de exalumnos de la universidad?
60	¿Cuáles son las carreras que se ofrecen en el área de Humanidades?
61	¿Puedo postularme a una carrera sin haber terminado mis estudios de secundaria?
62	¿Qué sucede si llego tarde el día del examen de admisión?
63	¿Cuántas veces se puede postular a una misma carrera?
64	¿Cuáles son los criterios para determinar la puntuación final del examen de admisión?
65	¿Qué beneficios tengo si soy un egresado de un colegio de alto rendimiento?
66	¿Cuánto tiempo antes debo llegar el día del examen de admisión?
67	¿Es posible postularme a carreras de doble titulación?
68	¿Cómo puedo conocer la programación de los exámenes simulacro?
69	¿El examen de admisión se realiza únicamente en Lima o también en otras ciudades?
70	¿Qué tipo de preguntas se incluyen en el examen de razonamiento verbal?
71	¿Es posible solicitar una revisión de mi examen si considero que mi calificación es incorrecta?
72	¿El examen de admisión tiene preguntas sobre idiomas extranjeros?
73	¿Puedo inscribirme en línea para el examen de admisión si no soy residente en Perú?
74	¿Qué pasa si pierdo mi carnet de postulante antes del examen?
75	¿Puedo pedir una constancia de mi postulación si no aprobé el examen de admisión?
76	¿El examen de admisión incluye preguntas sobre temas específicos de la carrera a la que postulo?
77	¿Puedo presentar el examen si estoy en el extranjero?
78	¿Cuánto tiempo dura el examen de admisión?
79	¿Se puede hacer una consulta sobre mi inscripción si cometí un error en los datos personales?

80	¿Puedo presentar el examen de admisión si ya tengo un título universitario de otra institución?
81	¿Cuántos puntos debo obtener para quedar en lista de espera?
82	¿Es posible que me cambien de fecha de examen si tengo un inconveniente?
83	¿Dónde puedo encontrar ejemplos de exámenes de admisión anteriores?
84	¿Hay alguna restricción de edad para ingresar a una carrera universitaria?
85	¿Qué pasa si no estoy en la lista de aprobados, pero quiero repetir el examen de admisión?
86	¿El examen de admisión incluye preguntas de ciencia o tecnología?
87	¿Cuáles son los horarios de atención para consultas sobre el examen de admisión?
88	¿Dónde se pueden ver los resultados oficiales del examen de admisión?
89	¿Puedo cambiar mi carrera después de haber sido admitido?
90	¿Cuánto tiempo tengo para hacer el pago de la inscripción después de haberme registrado?
91	¿Puedo postularme a carreras con diferentes modalidades, como presencial y a distancia?
92	¿Cómo puedo contactar con la oficina de admisión si tengo un problema durante el proceso?
93	¿Es posible modificar mi opción de carrera después de haber enviado mi inscripción?
94	¿Si no pude presentarme al examen, puedo solicitar una nueva oportunidad para rendirlo?
95	¿Qué debo hacer si no recibí el correo de confirmación de mi inscripción al examen de admisión?
96	¿Existen restricciones en cuanto a los objetos que puedo llevar el día del examen?
97	¿Hay algún examen de admisión especial para personas que ya tienen experiencia profesional?
98	¿Cómo puedo saber si mi inscripción fue aceptada y confirmada correctamente?
99	¿Se puede postular a una carrera de especialización o maestría en lugar de una licenciatura?
100	¿Puedo presentar el examen de admisión si estoy en mi último año de secundaria?