

Implementación de QueryBot UNT: un chatbot RAG híbrido para consultas académicas y administrativas universitarias

Implementation of QueryBot UNT: a hybrid RAG chatbot for university academic and administrative queries



Universidad Nacional Micaela Bastidas de Apurímac – Perú

Riqchary, revista de investigación en ciencia y tecnología

ISSN: 2810-8124 (en línea) / ISSN: 2706-543x

Vol. 8 Núm. 1 (2026) - Publicado: 30/07/2026

doi.org/10.57166/riqchary/v8.n1.2026

Páginas: 71- 79

Recibido 19/06/2026 ; Aceptado 24/07/2026

<https://doi.org/10.57166/riqchary/v8.n1.2026.9>

Dejerbakers Luis Josmell Cruz-Esquivel

Universidad Nacional de Trujillo

<https://orcid.org/0009-0001-6273-6844>

dlcruzes@unitru.edu.pe

Germaín Alexander Cruz-Vargas

Universidad Nacional de Trujillo

<https://orcid.org/0009-0009-9601-5234>

gacruzva@unitru.edu.pe

Stiven Adrian López-Malca

Universidad Nacional de Trujillo

<https://orcid.org/0009-0002-2483-5646>

salopezma@unitru.edu.pe

Alberto Carlos Mendoza-de-los-Santos

Universidad Nacional de Trujillo

<https://orcid.org/0000-0002-0469-915X>

amendezad@unitru.edu.pe

Resumen— En las universidades mucha información académica y administrativa está dispersa en reglamentos, directivas y comunicados institucionales, lo que dificulta su acceso oportuno por parte de los estudiantes. Esta investigación tuvo como fin implementar y evaluar QueryBot UNT, chatbot basado en arquitectura híbrida de generación aumentada por recuperación (RAG) para responder consultas universitarias mediante documentos oficiales como fuente de conocimiento. El sistema incluye mecanismos de recuperación semántica y generación de respuestas sustentadas documentalmente, disminuyendo la necesidad de realizar búsquedas manuales en múltiples archivos institucionales. Se utilizó el instrumento Chatbot Usability Questionnaire (CUQ) para evaluar la solución mediante un estudio con 30 estudiantes del séptimo ciclo de Ingeniería de Sistemas de la Universidad Nacional de Trujillo. Los resultados mostraron una percepción favorable de usabilidad, con un puntaje promedio superior al umbral de aceptabilidad establecido para este tipo de sistemas. Los resultados mostraron que QueryBot UNT fue bien aceptada y confirmaron que sirve como una herramienta de apoyo eficaz para facilitar el acceso a la información académica y administrativa de la universidad.

Palabras clave: chatbot, documentos oficiales, rag, usabilidad, educación superior, procesamiento de lenguaje natural

Abstract— In universities, much academic and administrative information is scattered across regulations, directives, and institutional communications, making it difficult for students to access it in a timely manner. This research aimed to implement and evaluate QueryBot UNT, a chatbot based on a hybrid architecture of augmented generation by retrieval (RAG) to answer university queries using official documents as a source of knowledge. The system includes semantic retrieval mechanisms and generates document-supported responses, reducing the need for manual searches across multiple institutional archives. The Chatbot Usability Questionnaire (CUQ) was used to evaluate the solution through a study with 30 seventh-semester Systems Engineering students at the National University of Trujillo. The results showed a favorable perception of usability, with an average score exceeding the established acceptability threshold for this type of system. The results demonstrated that QueryBot UNT was well-received and confirmed its effectiveness as a support tool for facilitating access to the university's academic and administrative information.

Keywords: chatbot, official documents, rag, usability, higher education, natural language processing

1 INTRODUCCIÓN

Los chatbots son sistemas automatizados que se comunican con los usuarios en lenguaje natural, brindando respuestas de manera eficaz y rápida [1]. Su uso ha aumentado en diferentes áreas y, durante la pandemia de COVID-19, ayudaron a facilitar el acceso a servicios de salud mental y a llegar a un mayor número de personas que necesitaban atención [2].

Con el paso de los años, los chatbots han dejado de depender únicamente de reglas predefinidas para responder preguntas sencillas [3]. Actualmente, incorporan tecnologías como RAG, que les permite obtener información de fuentes externas y generar respuestas más completas y confiables [4]. Gracias a estos avances, los usuarios pueden acceder a la información de una manera más sencilla y tener una mejor experiencia durante su interacción con estos asistentes [5].

Uno de los principales retos que enfrentan las universidades es brindar a los estudiantes un acceso rápido y confiable a la información institucional [6]. En la Universidad Nacional de Trujillo, la información académica y administrativa se encuentra distribuida en distintos documentos y canales, lo que dificulta que los estudiantes encuentren reglamentos, normas o procedimientos necesarios. Esta dispersión ya ha sido identificada como un problema en otras dependencias de la institución, como el caso de las bibliotecas de la UNT, donde los estudiantes debían consultar cada una por separado para ubicar la información requerida, con el riesgo de invertir tiempo sin obtener resultados [7]. Una situación similar se presenta con la información académica y administrativa, lo que aumenta la dependencia de la atención presencial o de fuentes no oficiales.

Para los estudiantes universitarios, el acceso rápido y sencillo a la información institucional resulta fundamental para el adecuado desarrollo de sus actividades académicas. Por ello, el presente proyecto propone el desarrollo de un chatbot (asistente virtual) que realice búsquedas en los documentos oficiales de la institución, con el fin de proporcionar respuestas claras y precisas ante las consultas de los estudiantes relacionadas con los servicios universitarios [8].

La presente investigación tuvo como objetivo implementar y evaluar QueryBot UNT, un chatbot basado en una arquitectura híbrida RAG, con la finalidad de facilitar el acceso a la información institucional de la Universidad Nacional de Trujillo. En ese sentido, se formuló la siguiente pregunta de investigación: ¿en qué

medida un chatbot basado en arquitectura híbrida RAG, aplicado sobre los documentos oficiales de la UNT, ofrece a los estudiantes una experiencia de uso usable y respuestas útiles, como alternativa a la búsqueda dispersa en documentos y canales no centralizados? Para responder a esta pregunta, se planteó como hipótesis nula (H_0) que la media del puntaje CUQ de QueryBot UNT es menor o igual a 68 puntos ($\mu \leq 68$), valor considerado como umbral de usabilidad aceptable, frente a la hipótesis alterna (H_1), que sostiene que dicha media es mayor a 68 puntos ($\mu > 68$).

2 METODOLOGÍA

2.1 Arquitectura y componentes del sistema

QueryBot UNT es un chatbot que trabaja con arquitectura RAG (Retrieval-Augmented Generation), y que combina la recuperación de información de documentos oficiales y la generación de respuestas en lenguaje natural. Con este enfoque mejoramos la precisión de las respuestas y reducimos el riesgo de información no sustentada [9].

La interfaz de usuario está desarrollada en Next.js 14 con Tailwind CSS para ofrecer una experiencia tipo chat, mantener un historial persistente y mostrar de forma explícita las fuentes. En los sistemas de atención digital, una interfaz conversacional facilita la interacción, ya que organiza la consulta en un flujo más natural para el usuario y mejora la percepción de accesibilidad y utilidad del servicio [10].

El backend está desarrollado en Python 3.11 con FastAPI, con una arquitectura liviana y escalable para procesar solicitudes web de manera eficiente. Se agregó autenticación con JWT (JSON Web Token) para resguardar el acceso al sistema, una práctica habitual para verificar identidad y manejar permisos en apps web actuales [11].

Se utilizó PostgreSQL 16 con las extensiones pgvector y pg_trgm para búsquedas semánticas y léxicas respectivamente. La combinación de las dos técnicas mejora la precisión de la recuperación de información documental [12], [13].

El pipeline de ingesta documental consistió en la extracción de texto desde archivos PDF mediante pypdf y pdfplumber, seguida de limpieza y normalización del contenido para reducir ruido OCR y caracteres residuales. Este paso es importante porque la calidad de los fragmentos indexados, y por tanto la pertinencia de las respuestas generadas, dependen de la calidad del texto recuperado [9].

Posteriormente se procedió a realizar un fragmentado

coherente del texto o chunking, teniendo en cuenta los títulos, las viñetas y las secciones. Esta decisión metodológica permite conservar el contexto semántico de cada bloque, facilitando así una recuperación más precisa cuando el estudiante plantea una pregunta específica sobre trámites, reglamentos o comunicados institucionales [10].

Cada fragmento se convirtió en embeddings, representaciones numéricas del significado del texto, y se indexó junto con metadatos como documento, página y categoría temática. En sistemas de recuperación semántica, el uso de embeddings es una estrategia clave, ya que permite que el motor de búsqueda encuentre información relacionada incluso cuando la consulta utiliza palabras distintas a las del documento original [9], [12].

Se introdujo un mecanismo anti-alucinación que limita las respuestas del modelo al contexto recuperado, devolviendo un mensaje de desconocimiento cuando no hay suficiente información. Esta estrategia tiene como prioridad la fiabilidad y la trazabilidad de las respuestas en los entornos educativos [10], [14].

La arquitectura híbrida puesta en práctica incorpora mecanismos de recuperación léxica y semántica con el fin de maximizar la relevancia de los documentos recuperados con anterioridad a la generación de la respuesta. Este enfoque se muestra especialmente adecuado en los entornos universitarios, donde la información se dispersa a través de reglamentos, directivas y comunicados que utilizan terminología especializada y estructuras documentales heterogéneas. De esta forma, el sistema combina las ventajas de la búsqueda basada en similitud semántica con la recuperación textual tradicional, disminuyendo la probabilidad de que se pierda información relevante durante el proceso de consulta.

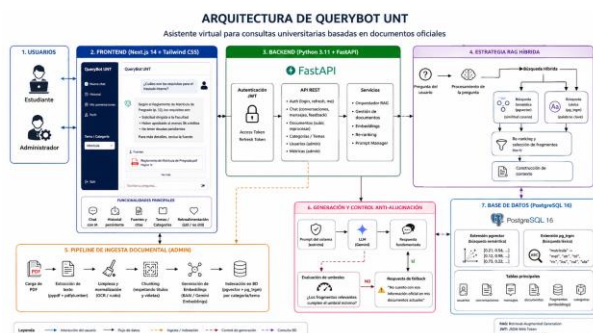


Fig. 1. Arquitectura del chatbot "QueryBot"

2.2 Sujetos de estudio y muestra

La población de estudio estuvo formada por los estudiantes de Ingeniería de Sistemas de la Universidad Nacional de Trujillo. Se escogió una muestra homogénea

de 30 alumnos del séptimo ciclo de la escuela profesional de ingeniería de sistemas de la universidad nacional de Trujillo mediante un muestreo no probabilístico por conveniencia por tratarse de una investigación de carácter aplicado y exploratorio, cuyo fin principal fue evaluar, de forma preliminar, la percepción de usabilidad de un prototipo funcional, antes de una implementación a mayor escala. El acceso directo a estudiantes con disponibilidad para participar también permitió realizar las pruebas de interacción en condiciones controladas y obtener retroalimentación temprana sobre el desempeño del sistema.

Se eligieron estudiantes de ciclos superiores por sus competencias técnicas para que evaluaran de manera crítica la arquitectura RAG, la interfaz y la calidad de las respuestas del chatbot. De igual forma, la literatura destaca que los usuarios con alta competencia digital contribuyen a identificar limitaciones y oportunidades de mejora en sistemas conversacionales [15], [10]. Con el propósito de mitigar sesgos se aplicaron estrictamente los siguientes criterios de elegibilidad:

Criterios de inclusión:

- Manifestación explícita de conformidad y participación voluntaria a través de la suscripción de un consentimiento informado digital previo al inicio de la prueba.
- Disponibilidad de un dispositivo electrónico (computadora portátil o móvil) con acceso estable a internet para interactuar con la interfaz del chatbot de manera independiente.
- Conocimientos mínimos sobre el uso de plataformas web y herramientas digitales de consulta, utilizadas en el entorno universitario.

Criterios de exclusión:

- Estudiantes de otras escuelas profesionales o ciclos académicos de la universidad que no sea el del grupo en estudio.
- Alumnos que presenten alguna limitación o discapacidad visual no asistida que les impida la lectura e interacción fluida con la interfaz del agente conversacional.

Los datos se recogieron durante el mes de mayo de 2026 en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional de Trujillo, ubicada en Trujillo, Perú.

2.3 Instrumento de recolección de datos

Para evaluar la percepción de los usuarios sobre su experiencia con QueryBot UNT, se utilizó el Chatbot Usability Questionnaire (CUQ) [16], un instrumento derivado de la System Usability Scale (SUS) y adaptado para evaluar agentes conversacionales. Se seleccionó porque permite medir aspectos propios de los chatbots, como la claridad de las respuestas, la personalidad del asistente y el manejo de errores.

Debido a que el instrumento original fue desarrollado en inglés, se utilizó una versión en español previamente adaptada en la literatura científica. A partir de esta, se realizaron ajustes en la redacción de algunos enunciados para adecuarlos al contexto de la Universidad Nacional de Trujillo, incorporando términos familiares para los estudiantes, como "trámites", "ciclos académicos" y "reglamentos". Estos cambios fueron únicamente de carácter semántico y no modificaron el significado, la estructura ni las dimensiones originales del instrumento [16].

El cuestionario está conformado por 16 enunciados, distribuidos entre afirmaciones positivas y negativas, las cuales ayudan a reducir el sesgo en las respuestas. Cada ítem se evaluó mediante una escala Likert de cinco puntos, desde "Totalmente en desacuerdo" (1) hasta "Totalmente de acuerdo" (5).

Las puntuaciones obtenidas se transformaron a una escala de 0 a 100, siguiendo el procedimiento establecido por los autores del instrumento [16], lo que facilitó la interpretación de los resultados y su comparación con investigaciones previas.

Finalmente, la confiabilidad del CUQ se evaluó mediante el coeficiente alfa de Cronbach, utilizando las respuestas de los 30 estudiantes participantes y considerando la inversión de los ítems negativos. Se obtuvo un valor de $\alpha = 0,81$, lo que evidencia una alta confiabilidad para este estudio. Asimismo, no fue necesario realizar una nueva validación de contenido, ya que el instrumento cuenta con una validación previa para su aplicación en entornos conversacionales [16].

2.4 Procedimiento estadístico de datos

La presente investigación tiene un enfoque cuantitativo, es de tipo aplicada, de nivel descriptivo y presenta un diseño no experimental de corte transversal. Se desarrolló sin manipular las variables de estudio y los datos se recopilaron en un único momento, con el propósito de describir la percepción de usabilidad de QueryBot UNT.

Una vez recopiladas las respuestas de los 30 participantes, los datos se registraron en Microsoft Excel y se procesaron con la herramienta oficial del CUQ para obtener la puntuación individual de cada estudiante. A partir de estos resultados, se calcularon estadísticos descriptivos como la media, mediana, desviación estándar, así como los valores mínimo y máximo, con el fin de describir el comportamiento de las puntuaciones.

El análisis inferencial se realizó en Python utilizando las bibliotecas SciPy y NumPy. Primero, se evaluó la normalidad de los puntajes mediante la prueba de Shapiro-Wilk con un nivel de significancia de 0,05. Al comprobarse este supuesto, se aplicó la prueba t de Student para una muestra, comparando la media obtenida con el valor de referencia de 68 puntos, considerado el umbral de aceptabilidad para las escalas SUS y CUQ [17].

Con base en este análisis, se plantearon las siguientes hipótesis estadísticas para determinar si el puntaje promedio de QueryBot UNT supera el nivel de aceptabilidad establecido:

- **Hipótesis nula (H_0):** La media del puntaje CUQ de QueryBot UNT es menor o igual a 68 puntos ($\mu \leq 68$).
- **Hipótesis alterna (H_1):** La media del puntaje CUQ de QueryBot UNT es mayor a 68 puntos ($\mu > 68$).

2.5 Consideraciones éticas

La investigación se desarrolló bajo los principios éticos de voluntariedad, anonimato y confidencialidad, obteniendo el consentimiento informado digital de cada participante previo a las pruebas.

3 RESULTADOS

3.1 Implementación funcional y flujo conversacional

Tras la implementación de QueryBot UNT sobre una arquitectura híbrida de generación aumentada por recuperación (RAG), se llevaron a cabo sesiones de prueba con los estudiantes participantes, quienes interactuaron con el asistente virtual en un entorno controlado. Durante estas sesiones, se monitoreó el comportamiento del sistema para verificar la pertinencia de las respuestas generadas y su desempeño ante consultas que no se encontraban respaldadas por la base documental.

La Figura 2 presenta un ejemplo representativo de interacción satisfactoria, en el que un estudiante realiza una consulta relacionada con un trámite académico y el chatbot proporciona una respuesta estructurada basada en la información contenida en los documentos oficiales indexados. Este resultado evidencia la capacidad del sistema para recuperar información relevante y

presentarla de manera coherente dentro del contexto de la consulta realizada.

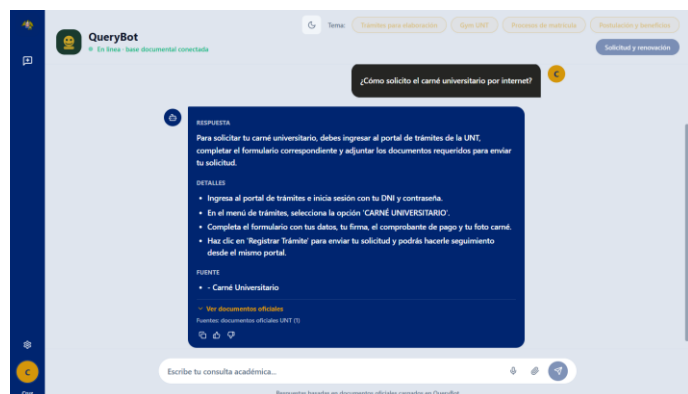


Fig. 2. Interacción exitosa entre el estudiante y el chatbot ante una consulta de trámite académico.

Asimismo, se evaluó el comportamiento del sistema ante consultas que no se encontraban cubiertas por la base documental. Como se muestra en la Figura 3, en estos casos el chatbot no generó respuestas sin respaldo documental y, en su lugar, informó al usuario sobre las limitaciones de la información disponible. Este comportamiento permitió mantener la consistencia de las respuestas proporcionadas y evitar la entrega de información no sustentada por las fuentes incorporadas al sistema.

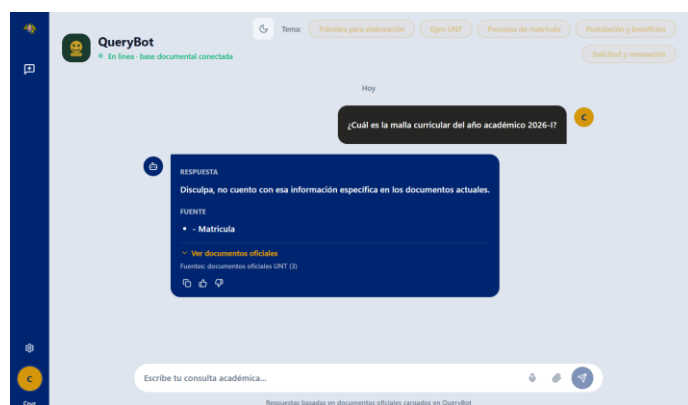


Fig. 3. Respuesta del chatbot ante una consulta fuera del alcance de su base de conocimientos.

3.2 Análisis descriptivo de usabilidad

El análisis de las treinta respuestas válidas registró una puntuación media de usabilidad de 73,6 puntos y una desviación estándar de 12,0. La mediana alcanzó un valor de 75,0, mientras que las puntuaciones mínima y máxima fueron de 42,2 y 98,4, respectivamente. La Tabla 1 presenta el resumen de los principales estadísticos descriptivos obtenidos:

TABLA 1
Estadísticos descriptivos de la puntuación CUQ

Estadístico	Valor
Media	73,6
Desviación estándar	12,0
Mediana	75,0
Mínimo	42,2
Máximo	98,4

Al comparar la media obtenida con el valor de referencia de 68 puntos propuesto [17], se observa que QueryBot UNT alcanzó una puntuación superior al umbral histórico de aceptabilidad establecido para este tipo de instrumentos.

La Figura 4 muestra la distribución de los treinta puntajes del CUQ mediante un diagrama de cajas. En ella se observa que la mediana y la mayoría de los resultados superan el umbral de aceptabilidad de 68 puntos.

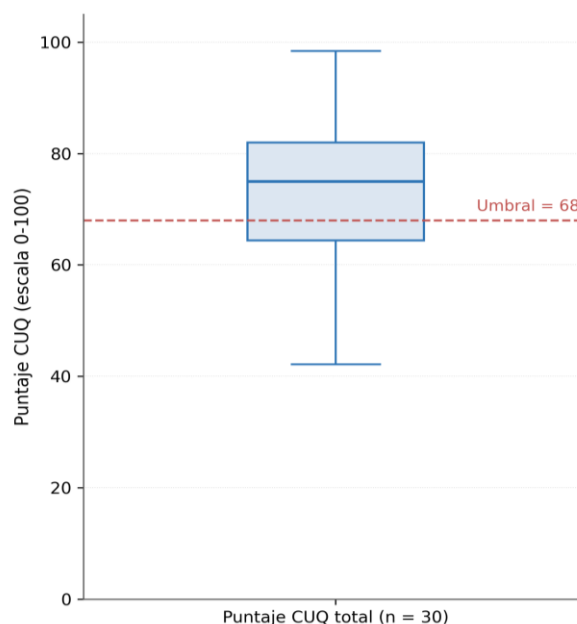


Fig. 4. Distribución de los puntajes CUQ obtenidos por los participantes, en contraste con el umbral de aceptabilidad.

La Figura 5 muestra la distribución de las 30 puntuaciones del CUQ mediante un histograma con su curva de densidad (KDE). La mayoría de los puntajes se concentra entre 70 y 80 puntos, mientras que solo dos participantes obtuvieron valores inferiores al umbral de aceptabilidad de 68 puntos. Asimismo, la media muestral (73,6) se ubicó por encima de dicho umbral, lo que refleja una percepción favorable de la usabilidad de QueryBot UNT y una variabilidad moderada ($DE = 12,0$).

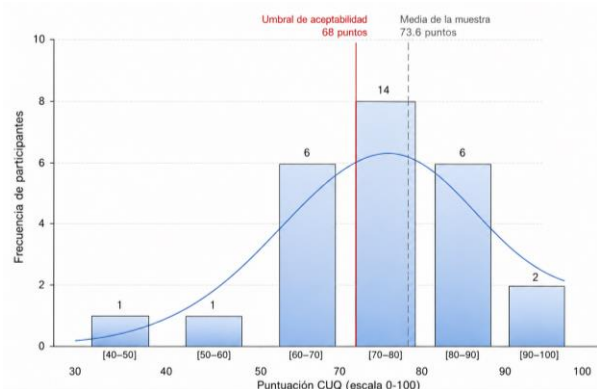


Fig. 5. Histograma de frecuencias de las puntuaciones CUQ de los 30 participantes.

El análisis de las respuestas por ítem permitió identificar diferencias en la valoración de los distintos aspectos evaluados. Entre los enunciados de contenido positivo, las puntuaciones más elevadas correspondieron a la facilidad de navegación (Q7: $4,5 \pm 0,9$) y a la facilidad general de uso (Q15: $4,4 \pm 0,8$). Asimismo, se registraron valoraciones favorables en la experiencia de configuración inicial (Q3: $4,1 \pm 0,9$), la personalidad del asistente (Q1: $4,0 \pm 0,9$) y la claridad con la que explicó su propósito (Q5: $4,0 \pm 1,1$).

Por su parte, los enunciados de contenido negativo presentaron puntuaciones bajas. Los valores más reducidos se observaron en la percepción de complejidad del sistema (Q16: $1,4 \pm 0,6$), la falta de claridad respecto a su propósito (Q6: $1,7 \pm 0,9$) y la posibilidad de experimentar confusión durante su uso (Q8: $2,0 \pm 1,1$). En contraste, los ítems que registraron las puntuaciones relativamente más altas dentro de este grupo negativo fueron la apariencia robótica (Q2: $3,2 \pm 1,0$) y el reconocimiento de entradas del usuario (Q10: $2,4 \pm 1,0$) junto con la incapacidad percibida para manejar errores (Q14: $2,4 \pm 0,9$). La Tabla 2 presenta el detalle de las medias y desviaciones estándar obtenidas para cada ítem, agrupados por su tipo de enunciado.

TABLA 2

Medias y desviaciones estándar de los ítems del CUQ agrupados por tipo de enunciado

Ítems positivos		
Ítem	Enunciado	Media $\pm$ DE
Q1	Personalidad realista y atractiva	$4,0 \pm 0,9$
Q3	Acogedor durante la configuración inicial	$4,1 \pm 0,9$
Q5	Explicó bien su alcance y propósito	$4,0 \pm 1,1$
Q7	Fácil de navegar	$4,5 \pm 0,9$
Q9	Entendió bien al usuario	$3,8 \pm 0,9$
Q11	Respuestas útiles, apropiadas e informativas	$3,9 \pm 0,9$
Q13	Manejó bien los errores	$3,6 \pm 1,1$

Q15 Muy fácil de usar $4,4 \pm 0,8$

Ítems negativos (puntuaciones invertidas)

Ítem	Enunciado	Media $\pm$ DE
Q2	Apariencia demasiado robótica	$3,2 \pm 1,0$
Q4	Aspecto poco amigable	$2,1 \pm 1,0$
Q6	No dio indicaciones sobre su propósito	$1,7 \pm 0,9$
Q8	Fácil de confundirse al usarlo	$2,0 \pm 1,1$
Q10	No reconoció muchas entradas	$2,4 \pm 1,0$
Q12	Respuestas irrelevantes	$2,0 \pm 0,9$
Q14	Incapaz de manejar errores	$2,4 \pm 0,9$
Q16	Muy complejo	$1,4 \pm 0,6$

La Figura 6 muestra las medias de los 16 ítems del CUQ con sus respectivas desviaciones estándar. Los enunciados se agrupan en ítems positivos y negativos para facilitar su interpretación. Los aspectos mejor valorados corresponden a la facilidad de navegación y al uso general del sistema, mientras que las principales oportunidades de mejora se relacionan con la gestión de errores y el reconocimiento de entradas complejas.

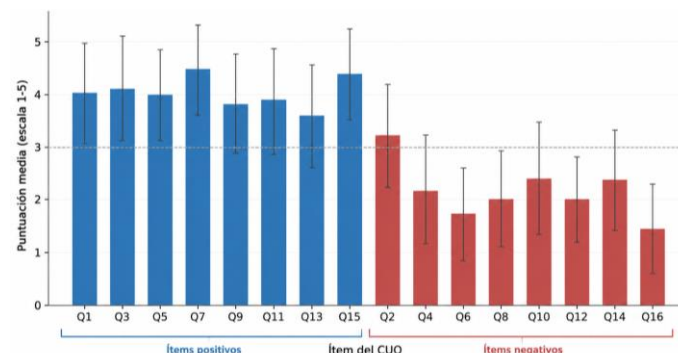


Fig. 6. Medias y desviaciones estándar de los ítems del Chatbot Usability Questionnaire (CUQ), agrupados en enunciados positivos y negativos.

Respecto a la utilidad percibida de las respuestas, el análisis del ítem Q11 mostró que el 73,3 % de los participantes indicó estar "De acuerdo" o "Totalmente de acuerdo" con la afirmación de que el chatbot proporcionó respuestas útiles, apropiadas e informativas. La Tabla 3 presenta la distribución de frecuencias correspondiente a este ítem.

TABLA 3

Aspecto "Utilidad de las respuestas" del chatbot

Escala	Frecuencia	Porcentaje
Totalmente en desacuerdo	1	3,3 %
En desacuerdo	0	0,0 %
Neutral	7	23,3 %
De acuerdo	14	46,7 %
Totalmente de acuerdo	8	26,7 %
Total	30	100 %

La Figura 7 muestra la distribución porcentual de las respuestas del ítem Q11, lo que permite complementar la información presentada en la Tabla 3. Como se aprecia, el 73,3 % de los participantes evaluó las respuestas del chatbot como útiles, apropiadas e informativas.

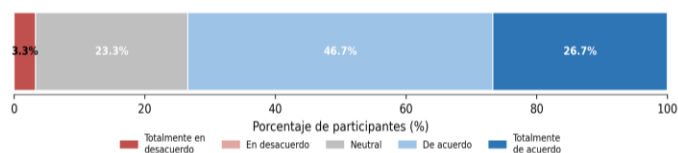


Fig. 7. Distribución porcentual de las respuestas al ítem Q11 del CUQ.

3.3 Prueba de hipótesis

Tal como se planteó en la sección 2.4, el propósito de esta prueba fue determinar si la usabilidad percibida de QueryBot UNT supera el umbral histórico de aceptabilidad (68 puntos).

En primer lugar, se verificó el supuesto de normalidad mediante la prueba de Shapiro-Wilk, la cual arrojó un estadístico $W = 0,9769$ y un p-valor de $0,7372$. Dado que este último es superior al nivel de significancia establecido ($\alpha = 0,05$), no se rechazó la hipótesis de normalidad, asumiendo que los datos son compatibles con una distribución normal.

En consecuencia, se aplicó la prueba t de Student para una muestra. Los resultados, resumidos en la Tabla 4, indicaron un estadístico $t = 2,547$ (con 29 grados de libertad) y un p-valor bilateral de $0,016$. Al ser el p-valor inferior a $0,05$ y la media muestral ($73,6$) superior al valor de referencia (68), se rechazó la hipótesis nula (H_0) y se aceptó la hipótesis alterna (H_1) con un nivel de confianza del 95 %.

Estos resultados evidencian que la puntuación media de usabilidad obtenida por QueryBot UNT es estadísticamente superior al umbral de aceptabilidad considerado para este tipo de sistemas.

TABLA 4

Resultados de las pruebas de normalidad e hipótesis			
Prueba	Estadístico	p-valor	Decisión
Shapiro-Wilk	0,9769	0,7372	Distribución normal ($\alpha = 0,05$)
T de Student (una muestra)	2,547	0,016	Se rechaza H_0 y se acepta H_1

4 DISCUSIÓN

4.1 Usabilidad y experiencia del usuario

Los resultados obtenidos evidencian que QueryBot UNT alcanzó un nivel de usabilidad aceptable, con una

media de $73,6$ sobre la escala CUQ, superando el umbral de referencia de 68 puntos. Esto es coherente con lo reportado por Holmes et al. (2019) y Espinosa-Luna et al. (2023), quienes también observaron puntuaciones favorables en chatbots universitarios con arquitecturas basadas en recuperación de información. Los aspectos mejor valorados, facilidad de navegación y facilidad general de uso, sugieren que una interfaz conversacional sencilla contribuye a la experiencia positiva del usuario, aunque parte de esta valoración podría deberse también al efecto de novedad propio de una herramienta recién introducida, y no solo a su calidad intrínseca.

El 73,3 % de los estudiantes consideró que las respuestas fueron útiles, adecuadas e informativas, lo que respalda la capacidad de la arquitectura RAG para recuperar información pertinente a partir de documentos institucionales. Esto refuerza la idea de que anclar las respuestas en fuentes verificables reduce la percepción de imprecisión, en comparación con modelos generativos sin mecanismos de recuperación. Entre las limitaciones, además del tamaño reducido de la muestra, existe un posible sesgo de selección, pues participaron estudiantes con disposición voluntaria a probar la herramienta. Tampoco puede descartarse deseabilidad social en las respuestas, ya que el cuestionario fue aplicado por los mismos investigadores del proyecto.

4.2 Limitaciones del sistema y trabajos futuros

Si bien los resultados obtenidos resultan alentadores, el sistema presenta aspectos susceptibles de mejora, tales como una mayor capacidad de comprensión ante consultas complejas y el manejo adecuado de situaciones de error o respuestas inesperadas. Asimismo, es importante señalar limitaciones metodológicas del estudio: la muestra de estudiantes empleada fue reducida y estuvo conformada por participantes voluntarios, lo que introduce un posible sesgo de selección, ya que es probable que estudiantes con mayor familiaridad o interés en herramientas tecnológicas hayan estado más dispuestos a participar. Adicionalmente, al haber sido la evaluación administrada por el propio equipo de investigación, no puede descartarse un efecto de deseabilidad social en las respuestas obtenidas. En consecuencia, como trabajo futuro se recomienda ampliar la evaluación a una muestra más numerosa y diversa, seleccionada de forma aleatoria y evaluada de manera anónima, así como fortalecer la arquitectura del sistema con el fin de garantizar su escalabilidad y su capacidad para atender un mayor volumen de solicitudes de manera eficiente.

5 CONCLUSIONES

El chatbot QueryBot UNT, basado en una arquitectura híbrida de generación aumentada por recuperación (RAG) para la atención de estudiantes universitarios, obtuvo un puntaje promedio de 73,6 en la escala CUQ, valor que se ubica por encima del umbral de referencia (68) y que permite calificar su usabilidad como aceptable dentro del contexto evaluado, aunque cercano al límite inferior de un desempeño considerado bueno. Asimismo, el 73,3 % de los participantes consideró que las respuestas del sistema fueron útiles, apropiadas e informativas. Entre los aspectos mejor valorados se encuentran la facilidad de navegación y la facilidad general de uso de la interfaz conversacional.

Pese a estos resultados favorables, se identificaron dificultades del chatbot para interpretar consultas ambiguas y para comunicar de forma clara cuando ocurre un error. Para atender este problema, se recomienda incorporar un módulo de reformulación de preguntas que solicite al usuario aclarar su consulta cuando el sistema detecte baja confianza en la respuesta recuperada, así como estandarizar mensajes de error que indiquen al estudiante qué ocurrió y qué acción tomar a continuación.

En el plano técnico, se recomienda implementar una gestión de hilos independiente por usuario, lo que permitiría procesar solicitudes concurrentes sin generar caídas o errores de API en momentos de alta demanda, como periodos de matrícula. A nivel institucional, estos hallazgos sugieren que la Universidad Nacional de Trujillo podría integrar este tipo de sistema como un canal complementario de atención en oficinas de trámites académicos y administrativos, reduciendo la carga de consultas repetitivas sobre el personal y permitiendo una atención más rápida en horarios fuera de oficina.

REFERENCIAS

- [1] E. M. Molina Castellano y J. M. Ocaña, «Eficacia de los Chatbots para la depresión en adultos. Revisión sistemática», *rrrc*, vol. 10, n.º 21, pp. 17-29, dic. 2025. <https://doi.org/10.53877/rc10.21-606>
- [2] C. S. Orduz-Cortes, P. Y. Niño-Pérez, E. Pérez-Sandoval, y A. N. Padilla-Jaimes, «Impacto de la inteligencia artificial en la práctica clínica en salud mental: una revisión sistemática: Impact of Artificial Intelligence on Clinical Practice in Mental Health: A Systematic Review», *Health care & global health*, vol. 10, n.º 2, jun. 2026. <https://doi.org/10.22258/hgh.v10i2.396>
- [3] M. Garzón-Quiroz, G. del Campo-Saltos, y B. Loo-Ávila, "Análisis sistemático sobre la eficiencia comunicativa entre chatbots basados en reglas y modelos de lenguaje natural," 2025. <https://doi.org/10.17163/uni.n42.2025.07>
- [4] I. Hernández Bas, Desarrollo de un chatbot especializado en enfermedades raras basado en técnicas RAG (Retrieval-Augmented Generation), Trabajo Fin de Máster, Universidad Pontificia Comillas, Madrid, España, 2026. Disponible en: <http://hdl.handle.net/11531/98413>
- [5] J. L. León Carrión y H. M. Quinde Pasato, Desarrollo de un sistema integrado de captura, procesamiento y consulta de información nutricional de productos procesados y ultra procesados: una aplicación móvil con asistente virtual basado en RAG, Trabajo de Titulación para obtener el título de Ingeniero de Sistemas, Universidad de Cuenca, Cuenca, Ecuador, ene. 2025. Disponible en: <https://dspace.ucuenca.edu.ec/handle/123456789/46228>
- [6] K. B. A. Ábalos Ramírez, "Diseño de un sistema inteligente de agente conversacional para consultas académicas en entornos universitarios," Tesis, 2025. [En línea]. Disponible: <https://hdl.handle.net/20.500.12880/14092>
- [7] G. M. Chaffo Aguilar y J. C. Mariños Urquiaga, "Modelado e implementación de un sistema distribuido para la gestión de consultas en las bibliotecas de la Universidad Nacional de Trujillo," Tesis de pregrado, Escuela Académico Profesional de Informática, Facultad de Ciencias Físicas y Matemáticas, Universidad Nacional de Trujillo, Trujillo, Perú, 2013.
- [8] C. A. Ortiz Valentin, H. M. Padilla Osuna, H. J. Escobar Cuevas, and M. Camacho Martinez, "Asistente inteligente para consultas académicas y administrativas de la Facultad de Informática de la UAS," *International Journal of Information Science and Technological Applications-UAS IJISTA*, vol. 2, no. 1, pp. 36-46, May 2026. Disponible en: <https://revistas.uas.edu.mx/index.php/IJISTA/article/view/1825>
- [9] R. V. K. Bevara, B. D. Lund, N. R. Mannuru, S. P. Karedla, Y. Mohammed, S. T. Kolapudi y A. Mannuru, "Prospects of Retrieval Augmented Generation (RAG) for Academic Library Search and Retrieval," *Information Technology and Libraries*, vol. 44, no. 2, 2025. <https://doi.org/10.5860/ital.v44i2.17361>
- [10] B. H. Espinosa-Luna, J. Castillo-Oliva, B. A. Montañez-Díaz y A. Mendoza-De-los-Santos, "Implementación de un chatbot basado en modelo de lenguaje de inteligencia artificial para responder preguntas frecuentes de estudiantes universitarios," *Revista Científica de Sistemas e Informática*, vol. 3, no. 2, p. e570, 2023. <https://doi.org/10.51252/rcsi.v3i2.570>
- [11] S. Sotnik, T. Shakurova y V. Lyashenko, "Development features web-applications," *International Journal of Academic and Applied Research*, vol. 7, no. 1, pp. 79-85, 2023. Disponible en:

<https://openarchive.nure.ua/handle/document/21600>

- [12] Google Cloud, "What is pgvector?," 2026. Disponible en: <https://cloud.google.com/discover/what-is-pgvector?hl=es-419>
- [13] ServBay, "Guía para utilizar la extensión pgvector de PostgreSQL en ServBay," s. f. Disponible en: <https://support.servbay.com/es/database-management/postgresql-extensions/pgvector>
- [14] N. E. Sánchez, G. C. Michay y J. V. Calderón, "Inteligencia artificial generativa en educación superior: una revisión sistemática de literatura hispanohablante," *Espacios*, vol. 46, no. 6, pp. 14-25, 2025. <https://doi.org/10.48082/espacios-a25v46n06p02>
- [15] V. C. Contreras Cortez, Usabilidad de un chatbot basado en inteligencia artificial y satisfacción del cliente en una entidad financiera, Lima - 2025, tesis de licenciatura, Universidad Autónoma del Perú, Lima, Perú, 2025. Disponible en: <https://hdl.handle.net/20.500.13067/3923>
- [16] S. Holmes, A. Moorhead, R. Bond, H. Zheng, V. Coates, and M. McTear, "Usability testing of a healthcare chatbot: Can we use conventional methods to assess conversational user interfaces?" in *Proc. 31st Eur. Conf. Cogn. Ergonomics*, 2019, pp. 207-214. <https://doi.org/10.1145/3335082.3335094>
- [17] J. R. Lewis and J. Sauro, "Item benchmarks for the System Usability Scale," *Journal of Usability Studies*, vol. 13, no. 3, pp. 158-167, May 2018.