



**UNIVERSIDAD ANDINA**

**NÉSTOR CÁCERES VELÁSQUEZ**

**FACULTAD DE INGENIERÍA DE SISTEMAS**

**ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**



**EMPLEO DEL LENGUAJE NATURAL PARA OPTIMIZAR  
LAS CONSULTAS SQL CON LANGCHAIN EN LA  
EMPRESA CONSDITEC JULIACA 2024**

**TESIS PRESENTADA POR:**

**Bach. CESAR AUGUSTO APAZA CACERES**

**PARA OPTAR EL TÍTULO PROFESIONAL DE:  
INGENIERO DE SISTEMAS**

**JULIACA – PERÚ**

**2025**



**UNIVERSIDAD ANDINA**

**NÉSTOR CÁCERES VELÁZQUEZ**

**FACULTAD DE INGENIERÍA DE SISTEMAS**

**ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**

**EMPLEO DEL LENGUAJE NATURAL PARA OPTIMIZAR  
LAS CONSULTAS SQL CON LANGCHAIN EN LA  
EMPRESA CONSDITEC JULIACA 2024**

TESIS PRESENTADA POR:

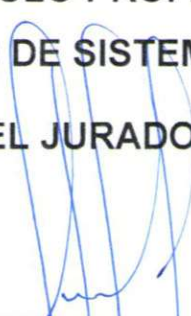
**Bach. CESAR AUGUSTO APAZA CACERES**

PARA OPTAR EL TÍTULO PROFESIONAL DE:

**INGENIERO DE SISTEMAS**

APROBADA POR EL JURADO REVISOR:

PRESIDENTE

:   
Dr. JUAN CARLOS HERRERA MIRANDA

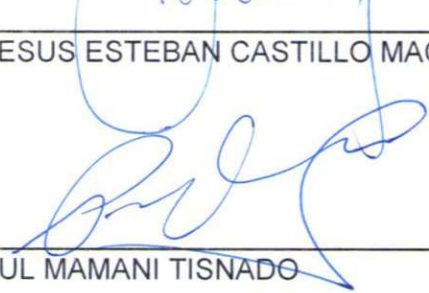
PRIMER MIEMBRO

:   
Mgtr. SALVADOR TEODORO VALDIVIA CARDENAS

SEGUNDO MIEMBRO

:   
Msc. JESUS ESTEBAN CASTILLO MACHACA

ASESOR DE TESIS

:   
Dr. PAUL MAMANI TISNADO

LÍNEA DE LA INVESTIGACIÓN : CIENCIA DE LOS ORDENADORES - P24



#### RESOLUCIÓN DECANAL N° 112-2025-D-FIS-UANCV

Juliaca, 29 de diciembre de 2025

**VISTOS;** el Expediente N° 2025-CU-13319, presentado por el señor (a) **CESAR AUGUSTO APAZA CACERES**, egresado de la Escuela Profesional de Ingeniería de Sistemas, de la Facultad de Ingeniería de Sistemas, quien solicita **REPROGRAMACIÓN DE FECHA Y HORA DE SUSTENTACIÓN**.

**CONSIDERANDO:**

Que, el (la) Bachiller **CESAR AUGUSTO APAZA CACERES**, quien solicita **REPROGRAMACIÓN DE FECHA Y HORA DE SUSTENTACIÓN** de la tesis titulada: **EMPLEO DEL LENGUAJE NATURAL PARA OPTIMIZAR LAS CONSULTAS SQL CON LANGCHAIN EN LA EMPRESA CONSDITEC JULIACA 2024**, la misma que pertenece a la línea de investigación, CIENCIA DE LOS ORDENADORES, para optar el título profesional de INGENIERO DE SISTEMAS.

Que, al haberse cumplido con los requisitos exigidos por el Reglamento Interno de Trabajo de Investigación conducente a Grados y Títulos mediante resolución 0294-2023-UANCV-CU-R y en concordancia con el dictamen de similitud.

De conformidad al Reglamento Interno de Trabajos de Investigación conducente a grafos y títulos, aprobado con resolución N° 0294-2023-UANCV-CU-R y en mérito al art. 24, Art. 28 del reglamento, con fines de obtención de obtención de Grados Académicos y Títulos Profesionales y en uso a las atribuciones que le concede la ley Universitaria N° 30220, Ley de creación de la UANCV N° 23738 y modificatoria N° 24661, y el estatuto de la UANCV, el Decano y Director de la Unidad de Investigación de la Facultad de Ingeniería de Sistemas.

**RESUELVE:**

**ARTÍCULO PRIMERO. - APROBAR** la **NOMINACION DE JURADOS**, integrado por los siguientes docentes:

- **Presidente : Dr. JUAN CARLOS HERRERA MIRANDA**
- **1er miembro : Mgtr. SALVADOR TEODORO VALDIVIA CARDENAS**
- **2do miembro : MSc. JESUS ESTEBAN CASTILLO MACHACA**

**ARTÍCULO SEGUNDO.- RECONOCER** como asesor de la investigación (tesis) de la Facultad de Ingeniería de Sistemas, al (a la) docente, **Dr. PAUL MAMANI TISNADO**

**ARTÍCULO TERCERO.- APROBAR** la **REPROGRAMACIÓN DE FECHA Y HORA DE SUSTENTACIÓN DE LA TESIS** del (la) Bachiller: **CESAR AUGUSTO APAZA CACERES**, del informe final de la investigación (tesis) titulado: **EMPLEO DEL LENGUAJE NATURAL PARA OPTIMIZAR LAS CONSULTAS SQL CON LANGCHAIN EN LA EMPRESA CONSDITEC JULIACA 2024**, para optar el título profesional de INGENIERO DE SISTEMAS, de acuerdo al siguiente detalle:

- **FECHA : martes, 30 de diciembre de 2025**
- **HORA : 15:45 PM.**
- **LUGAR : Sala de Docentes-FIS**

**ARTÍCULO CUARTO. - DISPONER** que, la Unidad de Investigación, Responsables del Comité de Investigación de la Facultad de Ingeniería de Sistemas y el Director de la Escuela profesional, Ingeniería de Sistemas, quedan encargados del cumplimiento de la presente Resolución.

Regístrese, Comuníquese y Archívese.



UNIVERSIDAD ANDINA  
"NÉSTOR CÁCERES VELÁSQUEZ"

Dr. Juan Carlos Herrera Miranda  
DECANO

**RESOLUCIÓN DECANAL Nº 023-2025-D-FIS-UANCV**

Juliaca, 16 de agosto de 2025

**VISTOS;** el Expediente Nº 2025-CU-6007, presentado por el (a) señor (a) **CESAR AUGUSTO APAZA CACERES**, quien solicita REVISION DEL INFORME FINAL DE LA INVESTIGACIÓN (borrador de tesis), el proveído Nº 00 -2025-UI-FIC-UANCV/J, del Director de la Unidad de Investigación de la Facultad de Ingeniería de Sistemas.

**CONSIDERANDO:**

Que, el (la) señor (a) **CESAR AUGUSTO APAZA CACERES**, ha presentado su informe final de la investigación (borrador de tesis), **titulado : EMPLEO DEL LENGUAJE NATURAL PARA OPTIMIZAR LAS CONSULTAS SQL CON LANGCHAIN EN LA EMPRESA CONSDITEC JULIACA 2024**, para optar el título profesional de **INGENIERO DE SISTEMAS**.

Que, al haberse cumplido con los requisitos exigidos por el Reglamento Interno de Trabajo de Investigación conducente a Grados y Títulos con fines de obtención de Grados Académicos y Títulos, el Director de la Unidad de Investigación de la Facultad de Ingeniería de Sistemas emitió opinión favorable al Expediente 2025-CU-6007, del informe final (borrador de tesis);, **titulado: EMPLEO DEL LENGUAJE NATURAL PARA OPTIMIZAR LAS CONSULTAS SQL CON LANGCHAIN EN LA EMPRESA CONSDITEC JULIACA 2024**, correspondiente a la línea de investigación, **CIENCIA DE LOS ORDENADORES**.

Que al haberse cumplido con los requisitos exigidos por el reglamento interno de trabajos de investigación conducentes a grados y títulos mediante resolución Nº 0294-2023-UANCV-CU-R, y estando a la opinión favorable de la Unidad de Investigación respecto al informe final de la investigación (borrador de tesis).

Estando, con la opinión favorable de la propuesta de investigación del Comité de investigación de la Facultad de Ingeniería de Sistemas y en concordancia con el reglamento interno de trabajos de investigación conducente a grafos y títulos, aprobado con resolución Nº 0294-2023-UANCV-CU-R y en mérito al art. 25 del reglamento, con fines de obtención de obtención de Grados Académicos y Títulos Profesionales y en uso a las atribuciones que le concede la ley Universitaria Nº 30220, Ley de creación de la UANCV Nº 23738 y modificatoria Nº 24661, y el estatuto de la UANCV, el Decano y Director de la Facultad de Ingeniería de Sistemas.

**RESUELVE:**

**ARTÍCULO PRIMERO.- APROBAR**, el INFORME FINAL DE LA INVESTIGACION (BORRADOR DE TESIS), para la REVISION DE SIMILITUD TURNITIN, presentado por el (la) señor (a): **CESAR AUGUSTO APAZA CACERES**, para optar el Título Profesional de **INGENIERO DE SISTEMAS** con el tema titulado: **EMPLEO DEL LENGUAJE NATURAL PARA OPTIMIZAR LAS CONSULTAS SQL CON LANGCHAIN EN LA EMPRESA CONSDITEC JULIACA 2024**, correspondiente a la línea de investigación, **CIENCIA DE LOS ORDENADORES**, en virtud a los considerandos expuestos.

**ARTÍCULO SEGUNDO. - RATIFICAR** como **ASESOR DE INVESTIGACION** al (a la) **Dr. PAUL MAMANI TISNADO**.

**ARTÍCULO TERCERO. - DISPONER** que, la Unidad de Investigación, Responsables del Comité de Investigación de la Facultad de Ingeniería de Sistemas y el Director de la Escuela profesional, **Ingeniería de Sistemas**, quedan encargados del cumplimiento de la presente Resolución.

Regístrese, Comuníquese y Archívese.



UNIVERSIDAD ANDINA  
"NÉSTOR CÁCERES VELÁSQUEZ"  
  
Dr. Rodolfo Fray Arpasi Chura  
DECANO (e)



## RESOLUCIÓN N° 291-2024-UI.P-D-FIS-UANCV-J

Juliaca, 19 de septiembre de 2024

### **VISTOS:**

El Expediente: 2024-CU-13376 de fecha 19 de septiembre de 2024, del (la) Bach. **CESAR AUGUSTO APAZA CACERES**; con el cual solicita Revisión de la Propuesta de Investigación y el Anexo (02 o 03) "Ficha de Opinión de la Propuesta de Investigación" que fue revisada por el Comité de Investigación de la Facultad de Ingeniería de Sistemas, Escuela Profesional de INGENIERÍA DE SISTEMAS.

### **CONSIDERANDO:**

**Que**, las Unidades de Investigación son unidades académicas que agrupan a docentes y estudiantes de diversas disciplinas, en razón del desarrollo de investigación científica, tecnológica y humanista de acuerdo al Estatuto Universitario Modificado 2020 de nuestra primera Casa Superior de Estudios.

**Que**, el (la) Bach. CESAR AUGUSTO APAZA CACERES, solicito la revisión y aprobación de la Propuesta de Investigación de la tesis titulada: EMPLEO DEL LENGUAJE NATURAL PARA OPTIMIZAR LAS CONSULTAS SQL CON LANGCHAIN EN LA EMPRESA CONSDITEC JULIACA 2024; conducente para optar el Título Profesional de INGENIERO DE SISTEMAS.

**Que**, al haberse cumplido con los requisitos exigidos por el Reglamento Interno de Trabajo de Investigación Conducente a Grados y Títulos plasmado en la Resolución N° 0294-2023-UANCV-CU-R.

**Que**, el Comité de Investigación ha emitido opinión favorable a la propuesta de investigación.

**Que**, el Director de la Unidad de Investigación de la Facultad de Ingeniería de Sistemas, Escuela Profesional de INGENIERÍA DE SISTEMAS, ratifico la propuesta del Asesor Dr. PAUL MAMANI TISNADO, quien debe estar acreditado y facultado para orientar y ayudar al asesorado en el proceso de elaboración del trabajo de investigación (Tesis).

**Estando**, la opinión favorable del comité de Investigación, en concordancia con el Reglamento Interno de Trabajo de Investigación Conducente a Grados y Títulos, Resolución N° 0294-2023-UANCV-CU-R, de conformidad a lo que establece la Ley Universitaria N° 30220, Ley de Creación de la UANCV N° 23738 y Modificatoria N° 24661 y el Estatuto de la UANCV, que confiere facultades al Decano de la Facultad de Ingeniería de Sistemas.

### **SE RESUELVE:**

**ARTÍCULO PRIMERO. - APROBAR Y AUTORIZAR LA EJECUCIÓN DE LA PROPUESTA DE INVESTIGACIÓN**, titulada: EMPLEO DEL LENGUAJE NATURAL PARA OPTIMIZAR LAS CONSULTAS SQL CON LANGCHAIN EN LA EMPRESA CONSDITEC JULIACA 2024, presentado por el (la) Bach. **CESAR AUGUSTO APAZA CACERES**, para optar el Título Profesional de INGENIERO DE SISTEMAS, en virtud de los considerandos expuestos.

**ARTÍCULO SEGUNDO. - RECONOCER**, como ASESOR al Dr. **PAUL MAMANI TISNADO**.

**ARTÍCULO TERCERO. - DISPONER** que la facultad, secretarías académicas y administrativas, quedan encargados del cumplimiento de la presente resolución.

Regístrese, comuníquese y archívese.



UNIVERSIDAD ANDINA  
"NÉSTOR CÁCERES VELÁSQUEZ"

M.Sc. Juan Carlos Herrera Miranda  
DECANO

C.c  
Arch 2024  
JCHM/ v1.1  
Distribución: Asesor de Tesis, Interesado

Ciudad Universitaria Urbanización Taparachi Km 4.5 Salida Puno - Juliaca



# CESAR AUGUSTO APAZA CACERES

## EMPLEO DEL LENGUAJE NATURAL PARA OPTIMIZAR LAS CONSULTAS SQL CON LANGCHAIN EN LA EMPRESA CONSDI...

 Trabajos de tesis para ingreso a repositorio

### Detalles del documento

Identificador de la entrega

trn:oid:::13016:596804192

Fecha de entrega

2 jun 2026, 9:31 GMT-5

Fecha de descarga

2 jun 2026, 9:57 GMT-5

Nombre del archivo

T036\_43190790\_T.docx

Tamaño del archivo

9.1 MB

94 páginas

14.483 palabras

82.820 caracteres



# TESIS UANCV

## 8% Similitud general

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

### Filtrado desde el informe

- Bibliografía
- Coincidencias menores (menos de 10 palabras)

### Fuentes principales

- 6%  Fuentes de Internet
- 2%  Publicaciones
- 6%  Trabajos entregados (trabajos del estudiante)

### Marcas de integridad

N.º de alertas de integridad para revisión

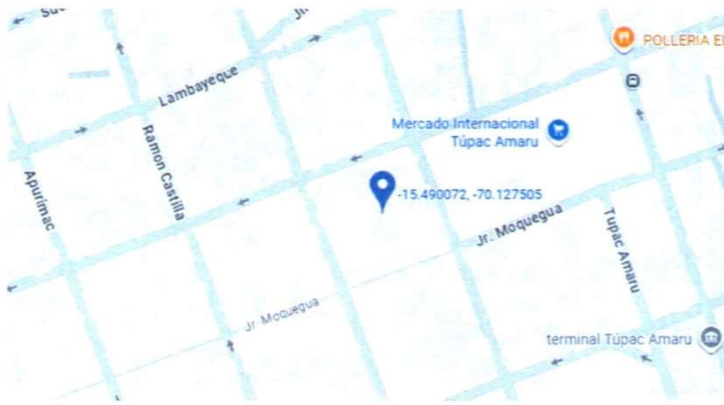
Los algoritmos de nuestro sistema analizan un documento en profundidad para buscar inconsistencias que permitirían distinguirlo de una entrega normal. Si advertimos algo extraño, lo marcamos como una alerta para que pueda revisarlo.

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



|                                                                                                                 |                                                                                           |
|-----------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------|
| EMPLEO DEL LENGUAJE NATURAL PARA OPTIMIZAR LAS CONSULTAS SQL CON LANGCHAIN EN LA EMPRESA CONSDITEC JULIACA 2024 |                                                                                           |
| <b>Datos de autor</b>                                                                                           |                                                                                           |
| Nombres y apellidos                                                                                             | Cesar Augusto Apaza Caceres                                                               |
| Tipo de documento de identidad                                                                                  | DNI                                                                                       |
| Número de documento de identidad                                                                                | 43190790                                                                                  |
| URL de ORCID                                                                                                    | <a href="https://orcid.org/0009-0002-4423-5760">https://orcid.org/0009-0002-4423-5760</a> |
| <b>Datos de asesor</b>                                                                                          |                                                                                           |
| Nombres y apellidos                                                                                             | Paul Mamani Tisnado                                                                       |
| Tipo de documento de identidad                                                                                  | DNI                                                                                       |
| Número de documento de identidad                                                                                | 01314987                                                                                  |
| URL de ORCID                                                                                                    | <a href="https://orcid.org/0000-0002-0287-7143">https://orcid.org/0000-0002-0287-7143</a> |
| <b>Datos del jurado</b>                                                                                         |                                                                                           |
| <b>Presidente del jurado</b>                                                                                    |                                                                                           |
| Nombres y apellidos                                                                                             | Juan Carlos Herrera Miranda                                                               |
| Tipo de documento                                                                                               | DNI                                                                                       |
| Número de documento de identidad                                                                                | 29606930                                                                                  |
| URL de ORCID                                                                                                    | <a href="https://orcid.org/0000-0002-5640-400X">https://orcid.org/0000-0002-5640-400X</a> |
| <b>Miembro del jurado 1</b>                                                                                     |                                                                                           |
| Nombres y apellidos                                                                                             | Salvador Teodoro Valdivia Cardenas                                                        |
| Tipo de documento                                                                                               | DNI                                                                                       |
| Número de documento de identidad                                                                                | 02383061                                                                                  |
| URL de ORCID                                                                                                    | <a href="https://orcid.org/0009-0008-8660-8733">https://orcid.org/0009-0008-8660-8733</a> |
| <b>Miembro del jurado 2</b>                                                                                     |                                                                                           |
| Nombres y apellidos                                                                                             | Jesus Esteban Castillo Machaca                                                            |
| Tipo de documento                                                                                               | DNI                                                                                       |
| Número de documento de identidad                                                                                | 01323821                                                                                  |
| URL de ORCID                                                                                                    | <a href="https://orcid.org/0000-0003-4595-7589">https://orcid.org/0000-0003-4595-7589</a> |



| Datos de investigación                                                                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                                   |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Línea de investigación                                                                                                                                                                         | Ciencia de los Ordenadores – P24                                                                                                                                                                                                                                                                                                                                  |
| Grupo de investigación                                                                                                                                                                         | No aplica                                                                                                                                                                                                                                                                                                                                                         |
| Agencia de financiamiento                                                                                                                                                                      | Sin financiamiento                                                                                                                                                                                                                                                                                                                                                |
| Ubicación geográfica de la investigación                                                                                                                                                       | <p>País: Perú<br/>Departamento: Puno<br/>Provincia: San Román<br/>Distrito: Juliaca<br/>Coordenadas:<br/>Longitud: -15.490185<br/>Latitud: -70.127516<br/>URL Maps:<br/><a href="https://maps.app.goo.gl/H8mdPvKLXwxaoKKh6">https://maps.app.goo.gl/H8mdPvKLXwxaoKKh6</a></p>  |
| Año o rango de años en que se realizó la investigación                                                                                                                                         | septiembre 2024 – junio 2025                                                                                                                                                                                                                                                                                                                                      |
| URL de disciplinas OCDE<br><a href="https://concytec-pe.github.io/Peru-CRIS/vocabularios/ocde_ford.html">https://concytec-pe.github.io/Peru-CRIS/vocabularios/ocde_ford.html</a><br>- Librería | <p>Ingeniería de Sistemas y Comunicaciones<br/><a href="https://purl.org/pe-repo/ocde/ford#2.02.04">https://purl.org/pe-repo/ocde/ford#2.02.04</a></p> <p>Ingeniería de Procesos<br/><a href="https://purl.org/pe-repo/ocde/ford#2.04.02">https://purl.org/pe-repo/ocde/ford#2.04.02</a></p>                                                                      |



UNIVERSIDAD ANDINA  
"NESTOR CACERES VELASQUEZ"

Dr. Juan Carlos Herrera Miranda  
DIRECTOR (e)  
Unidad de Investigación FIS



## DECLARACIÓN DE AUTENTICIDAD Y RESPONSABILIDAD

Yo, CESAR AUGUSTO APAZA CACERES identificado con DNI Nro. 43190790 en mi condición de egresado de:

- ☒ **Escuela Profesional**  
☐ **Programa de Segunda Especialidad,**  
☐ **Programa de Maestría o Doctorado**

INGENIERÍA DE SISTEMAS,

informo que he elaborado el/la ☒ **Tesis** o ☐ **Trabajo de Investigación,** ☐ **Trabajo Académico** denominada:

EMPLEO DEL LENGUAJE NATURAL PARA OPTIMIZAR LAS CONSULTAS SQL CON LANGCHAIN EN LA  
EMPRESA CONSDITEC JULIACA 2024

Asesorado por: Dr. PAUL MAMANI TISNADO

Es un tema original.

Declaro que el presente trabajo de tesis es elaborado por mi persona y **no existe plagio/copia** de ninguna naturaleza, en especial de otro documento de investigación (tesis, revista, texto, congreso, o similar) presentado por persona natural o jurídica alguna ante instituciones académicas, profesionales, de investigación o similares, en el país o en el extranjero.

Dejo constancia que las citas de otros autores han sido debidamente identificadas en el trabajo de investigación, por lo que no asumiré como tuyas las opiniones vertidas por terceros, ya sea de fuentes encontradas en medios escritos, digitales o Internet.

Asimismo, ratifico que soy plenamente consciente de todo el contenido de la tesis y asumo la responsabilidad de cualquier error u omisión en el documento, así como de las connotaciones éticas y legales involucradas.

El incumplimiento de lo declarado da lugar a responsabilidad del declarante, en consecuencia; a través del presente documento asumo frente a terceros, la Universidad Andina Néstor Cáceres Velásquez y/o la Administración Pública toda responsabilidad que pueda derivarse por el trabajo final presentado. Lo señalado incluye responsabilidad pecuniaria incluido el pago de multas u otros por los daños y perjuicios que se ocasionen.

Juliaca, 9 de ENERO del 20 26

  
FIRMA ASESOR

  
FIRMA TESISTA



Huella digital



## DEDICATORIA

To my mom, Ana Caceres Santa Cruz,  
who programmed the fundamental binaries of  
my firmware, shaping who I am.



### AGRADECIMIENTO

A los docentes de la Escuela Profesional de Ingeniería de Sistemas, de manera especial al Magíster Antonio Arroyo Paz, quien con sus terminales Linux y sus librerías "Panda" nos mostró nuevas fronteras.



## ÍNDICE GENERAL

|                      |      |
|----------------------|------|
| DEDICATORIA.....     | i    |
| AGRADECIMIENTO.....  | ii   |
| ÍNDICE GENERAL.....  | iii  |
| ÍNDICE TABLAS .....  | vi   |
| ÍNDICE FIGURAS ..... | viii |
| RESUMEN .....        | ix   |
| ABSTRACT .....       | x    |
| INTRODUCCIÓN .....   | xi   |

## CAPÍTULO I

### ASPECTOS GENERALES

|                                        |   |
|----------------------------------------|---|
| 1.1 Planteamiento del problema.....    | 1 |
| 1.2 Formulación del problema.....      | 3 |
| 1.2.1 Problema general .....           | 3 |
| 1.2.2 Problemas específicos.....       | 3 |
| 1.3 Justificación.....                 | 3 |
| 1.3.1 Teórica.....                     | 3 |
| 1.3.2 Metodológica .....               | 4 |
| 1.3.3 Teórica practica .....           | 5 |
| 1.4 Objetivos de la investigación..... | 5 |
| 1.4.1 Objetivo general.....            | 5 |
| 1.4.2 Objetivos específicos.....       | 6 |



|       |                                       |   |
|-------|---------------------------------------|---|
| 1.5   | Hipótesis .....                       | 6 |
| 1.5.1 | Hipótesis general .....               | 6 |
| 1.5.2 | Hipótesis específicas .....           | 6 |
| 1.6   | Variables .....                       | 6 |
| 1.6.1 | Variable independiente .....          | 6 |
| 1.6.2 | Variable dependiente .....            | 6 |
| 1.7   | Operacionalización de variables ..... | 7 |

## CAPÍTULO II

### MARCO TEÓRICO

|       |                                        |    |
|-------|----------------------------------------|----|
| 2.1   | Antecedentes de la investigación ..... | 8  |
| 2.1.1 | Antecedentes Internacionales .....     | 8  |
| 2.1.2 | Antecedentes Nacionales .....          | 9  |
| 2.2   | Marco teórico .....                    | 11 |
| 2.3   | Marco conceptual .....                 | 19 |

## CAPÍTULO III

### METODOLOGÍA

|       |                                |    |
|-------|--------------------------------|----|
| 3.1   | Métodos de investigación ..... | 27 |
| 3.1.1 | Tipo de investigación .....    | 27 |
| 3.1.2 | Nivel .....                    | 27 |
| 3.1.3 | Método .....                   | 28 |
| 3.2   | Población y muestra .....      | 28 |
| 3.2.1 | Población .....                | 28 |



|                                                                     |                                                            |    |
|---------------------------------------------------------------------|------------------------------------------------------------|----|
| 3.2.2                                                               | Muestra.....                                               | 28 |
| 3.3                                                                 | Técnicas e instrumentos de recogida de la información..... | 29 |
| 3.4                                                                 | Validación de la contrastación de hipótesis.....           | 29 |
| <br><b>CAPÍTULO IV</b><br><br><b>RESULTADOS Y DISCUSIÓN</b><br><br> |                                                            |    |
| 4.1                                                                 | Resultados .....                                           | 30 |
| 4.2                                                                 | Discusión de resultados .....                              | 58 |
| CONCLUSIONES.....                                                   |                                                            | 61 |
| RECOMENDACIONES .....                                               |                                                            | 62 |
| REFERENCIAS BIBLIOGRÁFICAS .....                                    |                                                            | 63 |
| ANEXOS .....                                                        |                                                            | 66 |



## ÍNDICE TABLAS

|                                                                                                                                                                   |    |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Tabla 1</b> Operacionalización de variables .....                                                                                                              | 7  |
| <b>Tabla 2</b> Datos estimados de modelos LLM para su entrenamiento .....                                                                                         | 13 |
| <b>Tabla 3</b> Resultados obtenidos ante la consulta: ¿Cuál fue el monto total de ventas del mes de mayo de 2025?.....                                            | 37 |
| <b>Tabla 4</b> Resultados obtenidos ante la consulta: ¿Cuál fue el monto total de las compras realizadas durante el mes de mayo de 2025? .....                    | 38 |
| <b>Tabla 5</b> Resultados obtenidos ante la consulta: ¿Cuál fue el producto más vendido durante el mes de febrero de 2025? .....                                  | 39 |
| <b>Tabla 6</b> Resultados ante la consulta ¿Cuál es el producto que menos se vendió en el mes de junio del 2025?.....                                             | 40 |
| <b>Tabla 7</b> Resultados obtenidos ante la consulta: ¿Cuál fue el total de ventas realizadas mediante pagos en efectivo durante el mes de junio de 2025? .....   | 41 |
| <b>Tabla 8</b> Resultados obtenidos ante la consulta: ¿Cuál fue el monto total de las ventas correspondientes a la última semana del mes de mayo? .....           | 42 |
| <b>Tabla 9</b> Resultados obtenidos ante la consulta: ¿Qué productos ingresados durante el mes de junio de 2025 presentaron un stock igual a cero? .....          | 43 |
| <b>Tabla 10</b> Resultados obtenidos ante la consulta: ¿Cuál fue el monto total de las compras realizadas durante el mes de junio de 2025? .....                  | 44 |
| <b>Tabla 11</b> Resultados obtenidos ante la consulta: ¿Cuál fue el monto total de las ventas realizadas mediante pagos Yape durante el mes de junio de 2025?.... | 45 |
| <b>Tabla 12</b> Resultados obtenidos ante la consulta: ¿Cuántos discos duros externos se vendieron durante el mes de junio de 2025? .....                         | 46 |
| <b>Tabla 13</b> ¿Con qué frecuencia necesitas obtener información específica desde el sistema de datos para realizar tu trabajo?.....                             | 50 |



|                                                                                                                                                               |    |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Tabla 14</b> ¿Qué tan cómodo/a te sientes utilizando herramientas digitales que te permiten consultar información (como filtros, menús o buscadores)?..... | 51 |
| <b>Tabla 15</b> ¿Qué tan fácil te resultó expresar lo que necesitabas mediante lenguaje natural? (escribiendo como lo dirías a otra persona).....             | 52 |
| <b>Tabla 16</b> ¿Qué tan útil fue poder escribir libremente lo que buscabas, en lugar de usar filtros o códigos como antes? .....                             | 53 |
| <b>Tabla 17</b> ¿En qué medida el nuevo asistente te ayudó a obtener los datos correctos sin necesidad de saber el código o ubicación exacta? .....           | 54 |
| <b>Tabla 18</b> ¿Qué impacto tuvo el uso del asistente en el tiempo que dedicaste a obtener la información que necesitabas? .....                             | 55 |



## ÍNDICE FIGURAS

|                  |                                                                                         |    |
|------------------|-----------------------------------------------------------------------------------------|----|
| <b>Figura 1</b>  | Diagrama de bloques de funcionamiento. ....                                             | 31 |
| <b>Figura 2</b>  | Diagrama de arquitectura de conexión implementada. ....                                 | 33 |
| <b>Figura 3</b>  | Frontend, realizando consulta de prueba. ....                                           | 34 |
| <b>Figura 4</b>  | Frontend, mostrando resultado de la consulta. ....                                      | 34 |
| <b>Figura 5</b>  | Gráfico de barras de los resultados de validez sintáctica y corrección por modelo ..... | 47 |
| <b>Figura 6</b>  | Gráfico de barras del tiempo medio de respuesta según el modelo LLM .....               | 48 |
| <b>Figura 7</b>  | Grafico de distribución de respuestas, frecuencia acceso.....                           | 50 |
| <b>Figura 8</b>  | Gráfico de distribución de respuestas; comodidad herramientas .....                     | 51 |
| <b>Figura 9</b>  | Gráfico de distribución de respuestas, facilidad expresión.....                         | 52 |
| <b>Figura 10</b> | Gráfico de distribución de respuestas, utilidad de respuestas.....                      | 53 |
| <b>Figura 11</b> | Gráfico de distribución de respuestas, grado facilitación.....                          | 54 |
| <b>Figura 12</b> | Gráfico de distribución de respuestas, ahorro tiempo. ....                              | 55 |
| <b>Figura 13</b> | Relación entre perfil de usuario y percepción del uso .....                             | 57 |
| <b>Figura 14</b> | Mapa de calor: correlaciones de Spearman. ....                                          | 58 |



## RESUMEN

Esta investigación se enfocó en el desarrollo de un asistente basado en un Modelo de Lenguaje a Gran Escala (LLM) desplegado de manera local en Ollama, orientado a la generación de consultas SQL a partir de lenguaje natural, para el personal no familiarizado con el manejo de base de datos y consultas SQL de la empresa Consditec. Se emplearon los frameworks LangChain y Streamlit, lo que permitió la integración local con la base de datos SQLite existente. Este enfoque desecha el uso de un API comercial, lo que preserva la privacidad de datos sensibles de la empresa, y como resultado permite optimizar el acceso a la información.

La evaluación técnica comparó cuatro modelos, LLaMA3, DeepSeek, DuckDB, Mistral y DuckDB, mediante 10 preguntas de uso recurrente contrastadas con la dataset Spyder, los resultados probaron que LLaMA3 alcanzó el mayor nivel de precisión con 7 respuestas correctas y un tiempo promedio de 12,687 ms. Seguido de Mistral; con 5 réplicas correctas y un lapso promedio de 59,833 ms, DeepSeek obtuvo un tiempo promedio significativamente mayor de 197,403 ms y 4 respuestas correctas, finalmente DuckDB registró 3 respuestas correctas y un tiempo promedio de 37,383 ms.

Para complementar el análisis se aplicó una encuesta de tipo Likert al personal de la empresa con el propósito de evaluar la percepción del asistente. Los resultados obtuvieron una correlación alta de Spearman ( $\rho = 0.942$ ,  $p < 0.001$ ) entre la necesidad de uso de datos por parte del usuario y su percepción respecto al empleo del lenguaje natural, lo que demuestra una aceptación favorable del uso de consultas en lenguaje natural.

**Palabras clave:** Langchain, Large Language Models, Text-to-SQL.



## ABSTRACT

This research focused on the development of a Large Language Model (LLM) based assistant deployed locally via Ollama, designed to generate SQL queries from natural language for Consditec staff who are unfamiliar with database management and SQL syntax. By utilizing the LangChain and Streamlit frameworks, local integration with an existing SQLite database was achieved. This approach eschews the use of commercial API, thereby preserving the privacy of sensitive corporate data and ultimately optimizing information accessibility.

The technical evaluation compared four models, LLaMA3, DeepSeek, Mistral, and DuckDB-NSQL using ten recurring queries benchmarked against the Spider dataset. The results demonstrated that LLaMA3 achieved the highest precision, with 7 correct responses and an average execution time of 12,687 ms. It was followed by Mistral, with 5 correct responses and an average time of 59,833 ms. DeepSeek recorded a significantly higher average latency of 197,403 ms with 4 correct responses, while DuckDB yielded 3 correct responses with an average time of 37,383 ms.

To complement the analysis, a Likert scale survey was administered to company personnel to evaluate user perception of the assistant. The results showed a high Spearman correlation ( $\rho=0.942$ ,  $p<0.001$ ) between the users' need for data and their perception of natural language utility, demonstrating favorable acceptance of natural language querying.

**Keywords:** LangChain, Large Language Models, Text-to-SQL, SQL queries.



## INTRODUCCIÓN

En el contexto actual, el acceso comprensible y eficiente a la información alojada en bases de datos se presenta como un factor determinante para procesos tal como, toma decisiones en operaciones cotidianas o procesos de gestión que deciden el rumbo del desarrollo de las empresas. Sin embargo, muchas pequeñas y medianas empresas presentan limitaciones estructurales o desconocimiento para convertir la información que disponen en conocimiento útil. Estas limitaciones no son solo restricciones técnicas, la escasez de personal con competencias especializadas, además el uso y la elección de herramientas poco intuitivas limitan el acceso a dicho conocimiento a usuarios no especializados.

Frente a esta problemática, la investigación tuvo como propósito evaluar la viabilidad de desarrollar un asistente con el que los usuarios no especializados puedan acceder a la información contenida en una repositorio de datos (BBDD), de forma que se permita el ingreso de consultas en lenguaje natural, mediante el framework LangChain como intermediario entre el modelo de lenguaje (LLM) y la base de datos, con el fin de realizar consultas y retornar datos de forma automática, sin requerir intervención directa del usuario.

Este enfoque tiene como propósito aplicar la disciplina Text-to-SQL, que busca la interpretación de consultas formuladas en lenguaje natural para su traducción a sentencias SQL válidas (PLN aplicado a lenguaje estructurado SQL). A diferencia de los primeros intentos que no lograban niveles de precisión aceptables por estar establecidas en técnicas basadas en tecnologías como los n-gramas, enfoques basados en reglas y algoritmos clásicos de aprendizaje automático, los nuevos enfoques se vieron fortalecidos con el desarrollo de los



modelos de lenguaje a gran escala actuales basados en arquitectura transformer y self-attention, lo que permitió una comprensión más fluida del lenguaje humano y posibilitó interpretar preguntas formuladas libremente por usuarios no técnicos, transformándolas en consultas estructuradas.

De este modo la propuesta planteada se integró en la empresa Consditec, la cual dispone de un sistema de gestión básicamente orientado solo al registro de ventas, compras e inventarios, como herramienta para la parte contable. Este sistema actual de la empresa, representa una barrera significativa para el personal que efectúa consultas principalmente mediante el uso de filtros predefinidos o la búsqueda a través de códigos de producto.

La investigación presenta en su primer capítulo el planteamiento de la problemática actual, centrada en las limitaciones de usabilidad asociadas a las consultas SQL.

En el segundo capítulo se desarrollaron los fundamentos teóricos que sustentaron el uso del lenguaje natural a través de modelos de lenguaje y su integración con la base de datos con LangChain, así también las investigaciones previas que sirvieron como respaldo para el planteamiento de la hipótesis de la investigación.

En el tercer capítulo se describió el diseño metodológico del estudio, así como las técnicas e instrumentos utilizados y la operacionalización de las variables.

Finalmente, en el cuarto capítulo se presentaron los resultados obtenidos y su respectiva discusión.

## CAPÍTULO I

### ASPECTOS GENERALES

#### 1.1 Planteamiento del problema

La tendencia digital actual, donde las organizaciones se ven obligadas a depender cada vez más de su capacidad para acceder y analizar de manera eficiente los datos recopilados. Según Olszak (2020), el 70% de las empresas a nivel mundial emplearon alguna plataformas de inteligencia de negocios para respaldar la toma de decisiones estratégicas, operando sobre bases de datos en crudo, como los *data lakes*, los cuales requieren habilidades especializadas para su adecuada limpieza y manipulación.

Según datos del INEI (2022), el predominio del 98,2 % de empresas del país corresponden al rubro de pequeñas y medianas empresas (PYME). Sin embargo, en conjunto comparten deficiencias en características clave, como el limitado desarrollo de marketing y competencias digitales. Además, se enfoca que solo el 25 % de las PYME utiliza algún tipo de herramientas de gestión de datos y apenas el 7 % cuenta con personal especializado en tecnologías de la información. Estas limitaciones muestran deficiencias en la toma de decisiones, acceso tardío a la información o pérdida de oportunidades e información, reduciendo su competitividad comparativa.

A nivel local la muestra se mantiene; ya que algunas empresas; si bien cuentan con herramientas básicas para el registro de compras y ventas, no



disponen de sistemas de consulta adecuados, y los proveedores de servicios para administrar negocios, no brindan facilidades de acceso a la información estructurada o no cuentan con herramientas capaces de procesar eficientemente esa información.

De esta manera, dentro de la empresa Consditec, surgió la necesidad de acceder a la información en su base de datos de manera directa y simple. Se planteó la adopción del enfoque correspondiente al campo de estudio Text-to-SQL, el cual busca interpretar lenguaje humano y convertirlo en instrucciones SQL válidas.

Con la aparición de modelos de lenguaje a gran escala más avanzados, como GPT, Gemini, LLaMA, DeepSeek, entre otros, estos propósitos son tareas técnicamente viables. Pero, para salvaguardar los datos de la empresa, el estudio consideró solo los modelos de lenguaje que se desplieguen de manera local, así el análisis del rendimiento de los distintos modelos evaluados pasó a ser un objetivo principal del estudio.

En este sentido, el objetivo consistió en la generación automática de consultas SQL a partir del lenguaje natural, evaluando la consulta generada, su validez semántica para la ejecución en la base de datos y la correcta traducción de información no estructurada hacia una representación final estructurada. Así, para validar esta propuesta; se buscó sustituir el uso de los mecanismos tradicionales basados en filtros o búsquedas por códigos de producto, los cuales generan una carga operativa adicional, retrasos en la obtención de información y restricciones en el acceso por parte de personal no capacitado.

Posteriormente, se midió el impacto del asistente en el personal; recopilando opiniones y percepciones sobre los aspectos abordados en la

investigación, las cuales puedan servir de referencia para futuros estudios relacionados con la temática.

## 1.2 Formulación del problema

### 1.2.1 Problema general

¿Cómo el empleo procesamiento de lenguaje natural y LangChain puede optimizar y simplificar la generación de consultas SQL en la empresa Consditec Juliaca 2024?

### 1.2.2 Problemas específicos

- ¿Es posible conectar la base de datos de la empresa Consditec mediante LangChain para realizar consultas SQL en lenguaje natural?
- ¿Cuál modelo de lenguaje natural tiene mejor desempeño para realizar consultas SQL en la empresa Consditec?
- ¿Cómo el uso del lenguaje natural influye en la simplificación de las consultas SQL por parte del personal de la empresa Consditec?

## 1.3 Justificación

### 1.3.1 Teórica

La investigación buscar revisar bases teóricas a partir del estudio del estado del arte en la rama de Text-to-SQL y del análisis de modelos de lenguaje a gran escala (LLM), que se vieron fortalecidos mediante innovaciones tecnológicas actuales como la arquitectura transformer (Leon et al., 2024). Además, se revisaron múltiples conceptos en el funcionamiento de los LLM, arquitectura y modelos, así como sus ámbitos de estudio globales que pertenecen a la rama de la inteligencia artificial, también se revisó la documentación de frameworks y librerías asociadas como complementos adicionales para la integración de las LLM en entornos locales.



En base a todos estos puntos, se planteó el desarrollo de un asistente basado en el uso de lenguaje natural mediante LLM integrados a través de frameworks, los cuales actuaron como un orquestador modular que facilitó la interacción entre el usuario y el modelo LLM. Para ello, se aplicó prompt engineering para adecuar la sintaxis de entrada para mejorar la precisión de las consultas sobre la base de datos SQLite. También se emuló un sistema de tipo Structured RAG con la base de datos; al extraer la estructura de tablas y pasarla directamente a la LLM pero no de manera vectorial como lo haría un RAG, en este esquema el modelo de lenguaje interpretó las preguntas y recuperó directamente los datos desde la base de datos, utilizando Streamlit como interfaz de salida web, la cual se integró con el backend de LangChain y el modelo de lenguaje.

### 1.3.2 Metodológica

La investigación se justificó en tres niveles desarrollando el enfoque técnico de los objetivos, primero se buscó demostrar la viabilidad y estabilidad de la construcción de un asistente local y funcional (on-premise).

Esta decisión responde a que la conexión de la base de datos con servicios externos como GPT, Gemini u otros modelos comerciales de mayor potencia implica riesgos a la privacidad y confidencialidad de los datos al exponer la información a servidores externos. También se elimina la dependencia de su disponibilidad del servicio y el costo mensual por derecho de uso de API. Por ello se optó por modelos open-source ejecutados de manera local mediante Ollama, herramienta que permite realizar pruebas en un entorno local y controlado.

Desde el punto de vista de la perspectiva operativa, se planteando una solución a las limitaciones a la extracción de datos y reduciendo la dependencia



del personal especializado. Con este fin, se analizaron opciones viables de modelos de lenguaje a gran escala compatibles con la capacidad computacional técnica real de los equipos informáticos de la empresa, evaluando y comparando de manera objetiva métricas de desempeño con el fin de identificar los modelos que ofrecieron mayor eficiencia y fluidez para el asistente.

Para el nivel organizacional, se promovió la democratización del acceso a los datos, contribuyendo a disminuir el desconocimiento de información clave para la toma de decisiones o desarrollo del trabajo habitual del personal.

### **1.3.3 Teórica practica**

Después de la identificación de la problemática en la empresa Consditec, caracterizada por la ausencia de una forma simplificada para el acceso a su información, la presente investigación no solo buscó evaluar técnicamente las condiciones para la implementación de un asistente de consultas en lenguaje natural, sino que planteó una solución a una problemática empresarial real, orientada a medir el impacto en la productividad, eficiencia, autonomía y comodidad del personal que tiene interacción regular con la base de datos de la empresa Consditec, al eliminar las barreras técnicas de acceso actuales preexistentes.

## **1.4 Objetivos de la investigación**

### **1.4.1 Objetivo general**

Aplicar el procesamiento de lenguaje natural y LangChain para optimizar y simplificar la generación de consultas SQL en la empresa Considetc Juliaca 2024



## 1.4.2 Objetivos específicos

- a. Conectar la base de datos de la empresa Consditec mediante LangChain para realizar consultas SQL en lenguaje natural.
- b. Evaluar los modelos de lenguaje natural óptimos para realizar consultas SQL en la empresa Consditec
- c. Describir como el uso del lenguaje natural influye en la simplificación de las consultas SQL por parte del personal de la empresa Consditec.

## 1.5 Hipótesis

### 1.5.1 Hipótesis general

El procesamiento de lenguaje natural y LangChain optimiza y simplifica la generación de consultas SQL en la empresa Consditec Juliaca 2024.

### 1.5.2 Hipótesis específicas

- a. Es factible La conexión de la base de datos de la empresa Consditec mediante LangChain para realizar consultas SQL en lenguaje natural.
- b. La evaluación de los modelos de lenguaje natural determina modelos óptimos para realizar consultas SQL en la empresa Consditec.
- c. El uso del lenguaje natural influye significativamente en la simplificación de las consultas SQL por parte del personal de la empresa Consditec.

## 1.6 Variables

### 1.6.1 Variable independiente

Uso del lenguaje natural

### 1.6.2 Variable dependiente

Optimización de consultas SQL

## 1.7 Operacionalización de variables

**Tabla 1**

*Operacionalización de variables*

| Variable                                                         | Dimensión                    | Indicador                                                      | Escala de medición    | Técnica e instrumento                                  |
|------------------------------------------------------------------|------------------------------|----------------------------------------------------------------|-----------------------|--------------------------------------------------------|
| Uso del lenguaje natural<br>( <i>Variable Independiente</i> )    | Interfaz de lenguaje natural | Nivel de facilidad para escribir consultas en lenguaje natural | Ordinal (Likert)      | Encuesta estructurada                                  |
|                                                                  | Experiencia de usuario       | Nivel de utilidad de respuestas generadas                      | Ordinal (Likert)      | Encuesta estructurada                                  |
|                                                                  | Utilidad percibida           | Grado de precisión del asistente sin usar SQL directamente     | Ordinal (Likert)      | Encuesta estructurada                                  |
|                                                                  | Ahorro de tiempo             | Percepción del tiempo ahorrado en consultas                    | Ordinal (Likert)      | Encuesta estructurada                                  |
| Optimización de consultas SQL<br>( <i>Variable Dependiente</i> ) | Tiempo de respuesta          | Milisegundos que tarda el modelo en generar la consulta        | Cuantitativa continua | Registro con time.perf_counter en Python               |
|                                                                  | Precisión sintáctica         | Nº de consultas con sintaxis SQL válidas                       | Cuantitativa discreta | Evaluación técnica: ejecución y comparación con Spider |
|                                                                  | Exactitud del resultado      | Nº de respuestas correctas frente a respuestas esperadas       | Cuantitativa discreta | Validación cruzada con respuestas esperadas            |

*Nota.* Descripción detallada de variables por dimensión, indicador, escala e instrumento. Fuente: Elaboración propia.



## CAPÍTULO II

### MARCO TEÓRICO

#### 2.1 Antecedentes de la investigación

##### 2.1.1 Antecedentes Internacionales

De acuerdo a Leon et al. (2024), en su trabajo titulado "Un estudio de Text-to-SQL con grandes modelos de lenguaje", el objetivo fue desarrollar y analizar un sistema capaz de convertir texto en consultas SQL empleando distintos LLM, el trabajo se orientó a mejorar la precisión de las consultas mediante estrategias como la generación aumentada por recuperación (RAG) y el aprendizaje de pocos ejemplos (few-shot learning). Los resultados obtenidos, se reportó una precisión máxima del 55%, concluyéndose que el uso de mayores recursos computacionales o modelos de mayor capacidad podría incrementar dicho desempeño, los métodos empleados incluyeron la evaluación de modelos sobre bases de datos etiquetadas y técnicas de prefiltrado mediante embeddings y bases de datos vectoriales.

Según Zhou et al. (2024), en su trabajo titulado "DB-GPT: Large Language Model Meets Database", se explora el uso de modelos de lenguaje a gran escala para la ejecución eficiente de tareas relacionadas con bases de datos, abordando diversos desafíos inherentes a los LLM, entre los que se destacaron la generación adecuada de prompts (prompt engineering). Adicionalmente, el estudio resalta aspectos relevantes como la obtención de datos, las



características estructurales de las bases de datos y la protección de la privacidad de la información. El estudio también sugiere introducción a conceptos orientados a la generación automática de prompts, así como al ajuste fino de modelos LLM especializados según la base de datos, su diseño y su proceso de preentrenamiento. En cuanto al rendimiento, los modelos DB-GPT mostraron mejoras en la reescritura de consultas y en la optimización de índices, superando los indicadores obtenidos por enfoques tradicionales. Adicionalmente, se plantean estrategias como la selección automática de ejemplos y la combinación de características textuales y no textuales para fortalecer la comprensión del modelo. Como cierre, los autores destacan la necesidad de continuar investigando aspectos relacionados con la validación de las salidas generadas y la gestión eficiente de los recursos del modelo.

## 2.1.2 Antecedente Nacionales

Camayteri et al. (2024), en su trabajo titulado "Métodos de Inteligencia Artificial (AI) en SQL Server para la eficiencia del sistema de auto calificación: Una Revisión sistemática literaria", busco la medición y análisis de un sistema de autoevaluación para entornos de educación. Los hallazgos de esta revisión muestran que la aplicación de inteligencia artificial mejora procesos en aspectos como exactitud y rapidez, superando los métodos tradicionales realizados por humanos y reduciendo tanto errores como tiempos de ejecución. Como conclusión, se señaló la necesidad de poner atención en mecanismos de ajuste en los procesos de evaluación para alcanzar los resultados esperados con la utilización de la IA.

Salazar (2020), en su tesis titulada "Chatbot para el aprendizaje de SQL básico", se planteó como objetivo el desarrollo de un chatbot capaz de generar



consultas SQL básicas, con el propósito de determinar el tiempo de interpretación de estas consultas y evaluar su eficiencia. El estudio adoptó un diseño no experimental con un enfoque cuantitativo y de carácter descriptivo-aplicado. Para su implementación, se empleó Telegram como canal de comunicación y Visual Studio Code para la programación, utilizando herramientas como Telegraf y Node.js. Las pruebas de campo arrojaron un tiempo promedio de interpretación de 0.85 segundos. Los resultados permitieron concluir que, a pesar de la funcionalidad del chatbot, su velocidad de respuesta es menor en comparación con sistemas más avanzados, se recomienda la incorporación de algoritmos más sofisticados y tecnologías actuales para optimizar su rapidez y efectividad.

Según Abanto (2021), desarrolló "Modelo de computación cognitiva basado en TOGAF 9.2 y Big Data Agile para el análisis de indicadores de gestión aplicando IBM Watson en la empresa Atento-Trujillo", el autor propone un modelo de computación cognitiva que se centra en conceptos de Big Data Agile y TOGAF 9.2. El estudio empleó IBM Watson con el objetivo de fortalecer el análisis de indicadores de gestión y mejorar la toma de decisiones, estructurando el proceso en seis fases validadas mediante casos prácticos de análisis de datos. Entre los principales hallazgos, se evidenció una mejora en la rapidez para la generación de indicadores y reportes, junto con una mayor adaptabilidad y estandarización de los procesos. El estudio concluye que la computación cognitiva potencia las capacidades de los usuarios al gestionar grandes volúmenes de datos, proporcionando una visión detallada que optimiza la eficiencia en la gestión de información gerencial.



## 2.2 Marco teórico

### 2.2.1 Procesamiento del lenguaje natural (PLN)

Es una rama de la IA que se orienta hacia la interpretación y generación de lenguaje humano en sistemas computacionales de tal forma que sea comprensible para personas en lenguaje nativo, para este propósito emplea unión de técnicas entre las cuales se encuentran la estadística, aprendizaje automático y lingüística computacional, mediante el cual permite realizar tareas como la comprensión y generación de respuestas contextualmente correctas y acertadas (Negueruela, 2024).

Entre los procedimientos que realiza un PLN consiste en desglosar el texto en unidades pequeñas o tokens, además identificar las partes relevantes y analizar las estructuras gramaticales que conecta el texto para la generación de significados más complejos. Para realizar estos propósitos se emplean técnicas que parten en la tokenización, procesos de análisis sintáctico o semántico con aprendizaje automático. La integración da como resultado la capacidad de realizar tareas como generación de texto estructurado, tareas de traducción o tareas multipropósito, los cuales son capaces de producir texto coherente, con contexto y con valor relevante.

Entre las aplicaciones actuales de los PLN se incluyen servicios de traducción automática, como Google Translate; asistentes virtuales, como modelos de HuggingFace o frameworks; así como la extracción de información personalizada mediante la identificación de entidades y relaciones relevantes dentro de textos. (Cargua & Francisco, 2025).

Los puntos aun limitantes de los sistemas de PLN radican en la complejidad y diversidad del lenguaje humano que incluye expresiones

coloquiales (slang), con sintaxis incorrecta o difíciles de interpretar, la dependencia del contexto, el humor y factores críticos cuando se busca traducir lenguaje natural a esquemas estructurados de SQL. Así como es importante considerar factores que limitan como el mismo sesgo que presenta cada modelo que se origina en la calidad de información utilizada durante el proceso de entrenamiento

### **2.2.2 Modelos de lenguaje (LM)**

Los modelos de lenguaje (LM) son la ejecución de los PLN, se definen como arquitecturas de red neuronal diseñadas para el reconocimiento y procesamiento de patrones de volúmenes de datos textuales mediante técnicas de entrenamiento que les permite predecir texto de manera sucesiva en base al contexto en el que se encuentra. La evolución de los LM tiene en sus comienzos enfoques basados en N-gramas, que se asocian a principios estadísticos que se orientaban a la frecuencia en la aparición de palabras. Estas técnicas tienen limitantes en cuanto a la compresión, al basarse en palabras y combinaciones frecuentes de aparición, lo que limita el procesamiento de dependencias a largo plazo o estructuras sintácticas ambiguas o de compresión de texto extensos. Para solventar el problema de los N-gramas, surgieron las RNN, cuya arquitectura permitió, obtener una retención secuencial de la información, lo que otorga la capacidad de retener información de manera mas extensa, lo que mejora la comprensión contextual (Bhushan et al., 2025). Entre los componentes clave para la comprensión de los LM se encuentran:

#### **a. Arquitectura**

La arquitectura interna determina su estructura funcional y la forma en la que procesa e interpreta secuencias de texto. Entre las principales

arquitecturas utilizadas se encuentran las redes neuronales recurrentes (RNN), las LSTM (Long Short-Term Memory) y los transformers, cada una con capacidades específicas para manejar diferentes niveles de complejidad y contexto en el lenguaje influyendo directamente con la carga computacional del LM.

## b. Entrenamiento

La fase de entrenamiento es el proceso durante el cual se ajustan los parámetros internos del modelo con el objetivo de que pueda ejecutar tareas específicas a partir de vastos corpus de texto. En base a estos volúmenes de texto permiten que el modelo aprenda patrones y relaciones lingüísticas que se almacenan en pesos en el modelo. A continuación, se presenta una tabla que detalla la cantidad de datos utilizados para el entrenamiento de los modelos LLM empleados en esta investigación y otras referencias, hasta el año 2025.

**Tabla 2**

*Datos estimados de modelos LLM para su entrenamiento*

| Modelo                 | Parámetros estimados               | Tokens pre entrenamiento o estimado | Hardware estimado                    | Costo estimado                                   |
|------------------------|------------------------------------|-------------------------------------|--------------------------------------|--------------------------------------------------|
| Llama 3.2 3B-Instruct  | ~3 000 millones (~3 B)             | ~9 trillones de tokens ("9 T")      | 4 000 – 8 000 GPUs tipo H100 / H800  | US\$1.5 a 3 millones                             |
| DuckDB-NS QL 7B-Q4_K_M | ~7 000 millones (~7 B)             | ~300 mil millones de tokens         | ~5 000 GPUs de consumo (tipo A100)   | US\$0.5 a 1 millones                             |
| OpenOrca-7 B_Q4_K_M    | ~7 000 millones (~7 B)             | ~200 mil millones de tokens         | ~2 000 GPUs tipo A6000/A100          | US\$0.3 a 0.6 millones                           |
| DeepSeek-R1            | ~671 B (MoE) o más (según versión) | ~6 trillones de tokens              | reportado: 2 048 H800 GPUs (o mayor) | reportado: US\$5.6 millones (solo entrenamiento) |

| Modelo                           | Parámetros estimados                           | Tokens pre entrenamiento o estimado | Hardware estimado              | Costo estimado      |
|----------------------------------|------------------------------------------------|-------------------------------------|--------------------------------|---------------------|
| GPT-4 (OpenAI)                   | ~1.8 trillones (1.8 T)                         | ~13 trillones (13 T)                | ~25,000 A100 GPUs              | US\$90-100 millones |
| PaLM 2 (Google)                  | ~340 miles de millones (340 B)                 | ~3-5 trillones                      | Decenas de miles GPUs          | US\$30-50 millones  |
| Llama 3 (Meta)                   | versiones hasta ~405 miles de millones (405 B) | hasta ~15 trillones                 | Hardware de gran escala        | US\$150 millones    |
| Claude 3 opus (Anthropic)        | ~200-300 miles de millones (200-300 B)         | ~3-6 trillones                      | Hardware estimado similar      | US\$60-80 millones  |
| Gemini 1.5 pro (Google DeepMind) | ~600 miles de millones (600 B)                 | ~8-10 trillones                     | Hardware grande (TPU v5/v6)    | US\$90-100 millones |
| DeepSeek R3                      | ~685 B (MoE)                                   | ~14.8 Trillones                     | varios miles de GPUs H800/H100 | US\$70-90 millones  |

*Nota.* Consolidado de datos estimados, obtenidos de diversas fuentes en el entrenamiento de modelos LLM. Fuente: Elaboración propia.

### c. Vocabulario

El vocabulario está conformado por el conjunto de palabras y subpalabras con las que fue entrenado (corpus), en base a este corpus el modelo es capaz de identificar y utilizar el léxico disponible, así uno a uno de los elementos del vocabulario recibe un identificador numérico único (token ID) que es transformado en un vector de alta dimensión (embedding), lo que permite su procesamiento computacional, es así como el modelo puede construir la comprensión del lenguaje permitiendo diferenciar matices de lenguaje y entablar relaciones semánticas precisas entre las palabras que procesa (Artasanchez, 2020), al acotar el espacio de búsqueda se reducen los tiempos de respuesta; sin

embargo. Un léxico demasiado amplio puede complicar el proceso de entrenamiento, en contrapuesta uno demasiado reducido puede limitar la capacidad del modelo para generalizar ante nuevos datos. Se debe tener en cuenta que, el tamaño, la calidad y la selección del vocabulario influyen directamente en el desempeño y la eficacia del modelo.

#### d. Embeddings

Los *embeddings* vienen a ser representaciones vectoriales de tokens que en su modo vectorial permiten comparar su magnitud mediante análisis estadístico, de esta forma tokens con significados similares se ubican cercanas en el espacio vectorial, de otra manera tokens con significados distintos se ubican a una mayor distancia, denotando una menor afinidad semántica. Todo esto se organiza en un espacio multidimensional que captura tanto las asociaciones semánticas directas como las relaciones sintácticas entre los vectores (Auffarth, 2023). Esta estructura permite realizar asociaciones entre tokens, por ejemplo, “princesa” se relaciona con “príncipe” de la misma manera que “mujer” se relaciona con “varón” manteniendo una distancia vectorial que, generalmente vectores con distancias vectoriales iguales pueden tener significancia parecida al embedding original; como se aprecia en el ejemplo anterior.

#### 2.2.3 Langchain

LangChain es un framework que se integra de manera natural con el ecosistema de Python, este framework funciona como un puente entre un LLM y datos externos que no están contenidos en el entrenamiento del modelo, un ejemplo de esto puede ser: documentos, bases de datos o vectores, este

proceso de vinculación se conoce como indexing (indexación). De esta manera, LangChain permite trabajar con la incorporación de prompts personalizados para interactuar con una LLM además que es posible proporcionar capacidades de memoria de interacción así como permitir conservar información histórica y contextual del usuario (Auffarth, 2023). Entre otras características de LangChain es que ofrece módulos para encadenar procesos (chains) cuya función permite generar texto, consulta de datos externos y la recuperación de información, optimizando flujos de trabajo complejos en el procesamiento con PLN puesto que es capaz de integrar una API de LLM. Entre los módulos más representativos se encuentran:

**a. LLM**

Si se realizara una analogía con el ser humano, el modelo de lenguaje podría considerarse como el cerebro, mientras que LangChain actuaría como el cuerpo, que estaría encargado de gestionar y procesar la interacción externa con la LLM.

**b. Prompts**

Son las instrucciones o directrices que se proporcionan a un LLM para que pueda generar una respuesta adecuada.

**c. Memorias**

Se refiere a su capacidad para retener información de interacciones pasadas, permite gestionar el estado de la conversación, evitando que el sistema pierda el hilo conductor en diálogos extensos.

**d. Herramientas**

Al integrarse a una LLM, puede ejecutar operaciones dinámicas, recuperar información actualizada y realizar cálculos sobre datos reales,

fortaleciendo la precisión, utilidad y aplicabilidad de sus respuestas en contextos específicos.

#### **2.2.4 Bases de datos.**

En el ámbito empresarial, las BBDD operan como el consolidado del almacenamiento y la gestión estructurada de la información, facilitando el registro, la organización, y la recuperación de los datos. Este enfoque se aplica a bases de datos pequeñas tanto como para sistemas distribuidos como aplicaciones en la nube. Las ventajas suman la gestión estable de grandes volúmenes de datos, reduciendo la redundancia y asegurando la integridad de los datos (Aguirre et al., 2022). El sistema de gestión de bases de datos (SGBD) actúa como interfaz de software encargado de administrar y manipular una base de datos mediante un conjunto de herramientas, lenguajes y protocolos que soportan las operaciones fundamentales de creación, lectura, actualización y eliminación de datos (CRUD). También integra mecanismos de seguridad, control de accesos, respaldo y recuperación de información, de este nodo se garantiza proteger los datos frente a accesos no autorizados, fallos del sistema y posibles pérdidas de información, garantizando así su disponibilidad e integridad.

##### **a. Base de datos relacionales (RDB)**

Propuestas por Edgar F. Codd en 1970, las bases de datos relacionales (RDB) se consolidaron como un estándar de gestión gracias a su eficacia para organizar información de forma estructurada y bajo principios de consistencia. Su arquitectura se fundamenta en la disposición de los datos en tablas bidimensionales; en este esquema, las columnas definen atributos específicos, mientras que las filas representan registros

individuales. Estas tablas se relacionan entre sí mediante claves, siendo las claves primarias las que permiten identificar de forma única cada registro, mientras que las claves foráneas establecen vínculos lógicos entre diferentes tablas asegurando la integridad referencial de los datos. No obstante, este modelo puede presentar problemas de rendimiento en aplicaciones que demandan baja latencia sobre macrodatos (big data), o en entornos que exigen una escalabilidad horizontal intensiva. Aun así, las RDB prevalecen en ecosistemas corporativos como los sistemas CRM, ERP y plataformas de transacciones financieras, donde la robustez y la integridad de los datos son requisitos críticos.

#### **2.2.5 Lenguaje SQL.**

Structured Query Language (SQL) es un lenguaje especializado en la gestión y manipulación de sistemas relacionales, tiene orígenes que se remontan a las investigaciones de IBM a principios de los años 1970. Con su estandarización se convirtió en el lenguaje más utilizado en la administración de datos, basado en álgebra relacional, su principal atributo se centra en la facilidad de realizar consultas flexibles denotando su eficiencia sobre indistinto volumen de datos y del tamaño de la base de datos.

El modo declarativo de SQL permite al usuario direccionar hacia los resultados deseados sin inferir en los mecanismos internos de recuperación, tarea que deja al motor de la base de datos, debido a la omisión de procesos mas complejos de los procesos subyacentes, de esta manera la estructuración de consultas se estandariza entre usuarios de formación técnica diversa. Otra característica ventajosa es la portabilidad entre distintos sistemas de gestores puesto que las sentencias generadas para un motor específico pueden utilizarse

en otro con ajustes mínimos, esta versatilidad ha consolidado su adopción tanto en el sector académico como en el entorno empresarial (Pérez, 2024).

### 2.2.6 Text-to-SQL

Es un enfoque de PLN que tiene como propósito convertir lenguaje humano en consultas SQL de forma automática permitiendo interactuar a personas sin conocimientos técnicos en SQL. Los sistemas Text-to-SQL se basan en modelos de machine learning, especialmente en arquitecturas basadas en transformers, tales como GPT, BERT entre otros. Estos modelos son entrenados utilizando corpus en base a pares de consultas en lenguaje natural y sus respectivas sentencias SQL correctas (gold standard) pero también en variantes y confusiones comunes en interpretación, este proceso les permite aprender patrones semánticos y estructurales entre la intención del usuario y la representación formal de la consulta, como resultado se tiene modelos que pueden inferir relaciones entre tablas, seleccionar atributos pertinentes y generar consultas coherentes (Leon et al., 2024).

Adicionalmente el modelo se requiere comprensión completa del esquema de la base de datos, así como las relaciones entre tablas y los tipos de datos. Los modelos actuales integran técnicas de grounding y contextualización para generar sentencias SQL más optimas y precisas, teniendo en cuenta restricciones como roles, permisos y límites de acceso propios de cada base de datos.

## 2.3 Marco conceptual

### 2.3.1 Fine-tuning

Es una técnica de ajuste supervisado, consiste en el reentrenamiento parcial del modelo de lenguaje que ya se encuentra preentrenado, utilizando

datos propios o específicos que son de interés, así se consigue especializar el comportamiento para mejorar el desempeño en tareas específicas. Este proceso permite ajustar los pesos del modelo a los patrones y estructuras propios del contexto de aplicación, lo que ayuda a reducir errores de generalización y procesamiento al momento de ejecutar una consulta con el modelo (Zhou et al., 2024).

### **2.3.2 Streamlit**

Es un framework desarrollado en Python cuya característica es la facilitación en la creación de interfaces web para aplicaciones de interacción con modelos de lenguaje; así es ampliamente utilizado en proyectos de ciencia de datos y aprendizaje automático, debido a su arquitectura simplificada y su alta capacidad de integración que permite desplegar interfaces gráficas sin requerir conocimientos avanzados en desarrollo web. Entre sus funcionalidades destaca la implementación de interfaces de tipo chat para el envío de prompts, donde el usuario interactúa con sistemas basados en modelos de lenguaje, lo que facilita su experimentación y desarrollo.

### **2.3.3 Retrieval-Augmented Generation (RAG)**

Es un enfoque híbrido entre el PLN que combina modelos generativos con mecanismos de recuperación de información. Inicialmente fue utilizada por Facebook en 2020 con el propósito de brindar mejor calidad y precisión en la generación de texto pero utilizando información externa al modelo que no se encuentra almacenada dentro de los parámetros del modelo. Este enfoque permite que los modelos de lenguaje accedan a fuentes de información actualizadas, y reduciendo la dependencia del conocimiento aprendido al que se encuentra parametrizado un modelo, mitigando problemas como la

desactualización o las alucinaciones del modelo, que son frecuentes ante la falta de datos (Negueruela, 2024).

Así mismo, RAG incorpora la capacidad de realizar búsquedas dinámicas sobre documentos almacenados en una base de datos o corpus, produciendo respuestas más precisas, requiere de dos componentes:

**a. Módulo de recuperación (retriever)**

Se encarga de buscar y recuperar fragmentos relevantes de información desde un corpus externo como documentos o bases de datos del usuario que requiere su consulta. Para ello genera índices, donde las consultas y los documentos se representan mediante embeddings. Estos vectores permiten medir la similitud semántica entre la consulta y los textos almacenados posibilitando la recuperación de información relevante aun cuando no exista coincidencia léxica exacta.

**b. Módulo generativo (generator)**

Este módulo se encarga de generar la respuesta final combinando la entrada del usuario con la información recuperada desde fuentes externas. El modelo de lenguaje integra ambos insumos dentro de su proceso de generación, lo que mejora la precisión y coherencia de la respuesta final.

#### **2.3.4 Transformers**

Esta arquitectura representa un cambio de paradigma en el procesamiento de lenguaje natural. Propuesto por Vaswani et al. (2017) en el artículo Attention is All You Need, los Transformers finalmente superaron las limitaciones de modelos previos, como las redes neuronales recurrentes (RNN), puesto que en su arquitectura implementaron mecanismos de atención en

paralelo, lo que les permite al modelo ponderar la relevancia de diferentes partes de una entrada de manera simultánea, este proceso es conocido como autoatención (self-attention). De esta forma, el modelo puede medir la importancia relativa de todas las palabras con respecto a las demás dentro de una secuencia revisando el contexto (Negueruela, 2024), Este modelo está conformado por dos componentes principales, los cuales se organizan en capas apiladas:

#### **a. Encoder**

El encoder analiza la secuencia de entrada para generar una representación vectorial basada en el contexto, esta secuencia analiza y captura la sintaxis y la semántica del texto en el espacio vectorial de alta dimensión donde cada dimensión representa distintos atributos del elemento analizado; por ejemplo una dimensión puede reflejar la frecuencia con la que una palabra o frase aparece junto a otras, mientras que otra puede representar su contexto o significado general dentro de una situación determinada entre otros.

#### **b. Decoder**

Usa representaciones contextuales para producir la secuencia de salida en cada capa del transformer, incorporando mecanismos de atención que asignan distintos niveles de relevancia para centrarse en las partes más significativas del contexto. Su relevancia reside en su capacidad para modelar la estructura del lenguaje humano de manera más eficiente que las arquitecturas anteriores.

### 2.3.5 Tokenización

Es un procedimiento fundamental en el procesamiento de lenguaje natural que consiste en dividir un corpus de texto en unidades más pequeñas denominadas tokens. Estos tokens pueden corresponder a palabras, caracteres, según el nivel de detalle o granularidad requerido por la acción solicitada. Este proceso constituye una etapa previa a la representación contextual del texto (Min et al., 2024). El procedimiento puede variar según influido por la tarea a realizar o por el idioma que se este utilizando, puesto que el procedimiento para tokenizar el español es diferente al japonés dada su diferencia en la estructura gramatical que maneja cada idioma. Entre los tipos de tokenización destacan:

#### a. Por palabra

Fragmenta el segmento de texto por palabras individuales, diferenciándolas mediante la presencia de espacios o signos de puntuación. Este enfoque puede generar un diccionario de gran tamaño.

#### b. Por subpalabra (Stemming)

Descompone los segmentos en unidades elementales como prefijos, sufijos y raíces léxicas. Este procedimiento ayuda a reducir el tamaño del vocabulario y a gestionar palabras compuestas o poco comunes.

#### c. Por caracteres

Este procedimiento segmenta el texto en caracteres individuales. Se utiliza en el reconocimiento óptico de caracteres (OCR) y donde es carente el espacio de separación que tienen algunos idiomas.

### 2.3.6 Lemmatization

Es una técnica que a diferencia del stemming, considera el contexto gramatical y semántico para obtener la raíz o la parte base de cada término a la

cual denominada lema. Este procedimiento simplifica las tareas de procesamiento al agrupar variaciones verbales bajo un mismo lema; por ejemplo, formas verbales como “estudiando”, “estudiado” o “estudiará” se reducen al lema “estudio”. Empleado para el ordenamiento de texto, análisis de sentimiento y procesos de traducción (Artasanchez, 2020).

### **2.3.7 Small language model (SLM)**

Los modelos de lenguaje de pequeña escala (SLM) son versiones optimizadas que a diferencia de los modelos a gran escala, están pensados para tareas específicas en entornos con recursos de hardware limitados, con una arquitectura orientada a la rapidez y la eficiencia operativa (Brito, 2024), por ejemplo Phi-2 cuenta con aproximadamente 2.7 millones de parámetros

### **2.3.8 Corpus**

Se define como la base fundamental para el aprendizaje de patrones sintácticos y semánticos, el Corpus es un conjunto seleccionado, estructurado y representativo de datos lingüísticos, que son recopilados de forma sistemática para entrenar y evaluar modelos de procesamiento de lenguaje natural. Entre estos conjuntos de datos se puede incluir textos, consultas, diálogos o documentos que son organizados bajo técnicas y criterios lingüísticos específicos (Uma et al., 2019).

### **2.3.9 Ollama**

Es una plataforma multiplataforma diseñada para el despliegue local de modelos de lenguaje. Uno de sus puntos mas resaltantes es que su arquitectura permite la ejecución sin conexión a internet y creado en un enfoque de código abierto (Alvear, 2025). Permite la descarga y ejecución de diversos modelos optimizados como: LLaMA, Mistral, Qwen, etc, mediante una interfaz sencilla

basada en terminal o línea de comandos. Ollama se caracteriza para poner atención a la privacidad, la portabilidad y el control total de la ejecución, lo que la hace adecuada para la utilización para desarrolladores e investigadores.

### 2.3.10 Dataset

Un dataset es una colección de datos organizada y sistematizada que sirve como insumo para el entrenamiento o la validación de un modelo. En el desarrollo de sistemas Text-to-SQL, estos datos enlazan preguntas en lenguaje natural y sus respectivas sentencias lógicas. La importancia de un dataset se encuentra en la calidad y variedad de las muestras, conteniendo una muestra representativa de datos y casos, de esto depende la capacidad del modelo para generalizar respuestas ante consultas (Mollocondo, 2025).

### 2.3.11 Prompt

Se define como la entrada de texto que establece el punto de partida para la generación de contenido por parte de un LLM. Consiste en una instrucción o pregunta que activa una respuesta del sistema; que puede variar desde una forma simple, como una sola letra hasta entradas complejas con múltiples instrucciones, restricciones o parámetros, definiendo el alcance de la salida o acción que se espera del modelo (Castro & Muñoz, 2023).

Tipos:

#### a. Prompt abierto

Es un tipo de instrucción que permite formular peticiones con libertad creativa, sin imponer restricciones ni parámetros específicos.

#### b. Prompt cerrado

Es un prompt más específico y delimitado, parametriza la respuesta esperada.

**c. Prompt con restricciones**

Impone restricciones o parámetros sobre la salida generada; ejemplo,

Escribe un cuento acerca de un perro sin utilizar la palabra “tristeza”.

**2.3.12 Query**

Una query es una sentencia formal que en el contexto de un modelo de base de datos permite obtener una respuesta específica a partir de una petición estructurada (Jacinto, 2022). En la estructura del query puede incorporar operadores lógicos, comparadores o funciones agregadas, con el propósito de manipular datos o aplicar filtros sobre los registros.



## **CAPÍTULO III**

### **METODOLOGÍA**

#### **2.4 Métodos de investigación**

Esta investigación adoptó un diseño no experimental de tipo transeccional y un enfoque cuantitativo. El estudio se centró en la medición objetiva de las variables sin realizar una manipulación directa de las mismas, adicionando la medición de la percepción de los usuarios que fue evaluada en un momento único mediante escalas de tipo Likert.

##### **2.4.1 Tipo de investigación**

La investigación se clasifica como aplicada, dado que estuvo orientada a la solución de un problema específico en un contexto empresarial real y local, mediante el empleo de herramientas tecnológicas evaluando resultados prácticos (Caballero, 2010). El enfoque es cuantitativo puesto que se basa en la recopilación y el análisis de datos numéricos sobre el desempeño de los modelos de lenguaje y las valoraciones del personal.

##### **2.4.2 Nivel**

La investigación es de nivel descriptivo-explicativo, es descriptiva al caracterizar el proceso de integración del sistema con la infraestructura de la empresa y la mecánica de generación de consultas, explicativo al analizar cómo influye el uso del lenguaje natural en la eficiencia de las sentencias SQL y en la experiencia de los usuarios (Dieterich, 2010).

### **2.4.3 Método**

Para los dos primeros objetivos específicos se empleó el método hipotético-deductivo, puesto que partió de la revisión teórica, al problema particular planteado y su comprobación de su operatividad, y para el tercer objetivo se utilizó el método inductivo, el cual infiere conclusiones generales a partir de la observación de datos y patrones (Dieterich, 2010). Al no realizarse la manipulación de variables, la investigación correspondió a un diseño no experimental y transeccional, dado que la recolección de datos se efectuó en un único momento temporal. Bajo este método se observaron y analizaron tanto el rendimiento de los modelos como las percepciones de los usuarios bajo condiciones reales de uso (Caballero, 2010).

## **2.5 Población y muestra**

### **2.5.1 Población**

La población de estudio estuvo constituida por la totalidad de 11 trabajadores de la empresa Consditec que requieren acceder a la base de datos para el desarrollo de sus actividades habituales, tales como la consulta de ventas, la revisión de stock de productos, etc.

### **2.5.2 Muestra**

Dado el tamaño reducido del grupo y el acceso total de la población, se optó por un muestreo de tipo censal. Al considerar a todos los integrantes sin un proceso de selección aleatoria, se asegura una representatividad completa de los datos. El muestreo censal se define como la recolección de la información de la totalidad de los elementos que integran la población de interés y sin excluir a ningún individuo, con el propósito de obtener datos completos, evitando la pérdida de información. (Hernandez, 2018).

## 2.6 Técnicas e instrumentos de recogida de la información

En el desarrollo de la investigación se utilizaron dos técnicas principales para la recolección de datos: observación estructurada-técnica, y encuesta.

### 2.6.1 Observación estructurada y técnica

Para el objetivo específico 1, la observación técnica utilizó como instrumento una lista de cotejo para supervisar los requerimientos y la ejecución de la implementación. Para el objetivo específico 2 Esta técnica se aplicó mediante el registro sistemático del desempeño de distintos modelos de lenguaje natural (LLM) en la generación de consultas SQL. La observación estructurada implicó el registro planificado y controlado de situaciones específicas, estableciendo criterios previamente definidos con el fin de garantizar la uniformidad y la imparcialidad en la recolección de los datos (Caballero, 2010).

### 2.6.2 Encuesta

Para el tercer objetivo específico se empleó la técnica de la encuesta mediante un cuestionario estructurado de seis preguntas cerradas de tipo Likert. Esta técnica de recolección de información consiste en la aplicación de un conjunto de preguntas previamente diseñadas a la población de estudio para obtener datos relevantes sobre las percepciones de los participantes, realizando así una evaluación perceptual (Hernandez, 2018).

## 2.7 Validación de la contrastación de hipótesis

Las respuestas obtenidas fueron consolidadas en una base de datos en formato CSV para su posterior análisis estadístico, aplicándose medidas de tendencia central, correlación de Spearman y gráficos de distribución. El análisis se llevó a cabo mediante el uso de las librerías especializadas Pandas y NumPy.

## CAPÍTULO IV

### RESULTADOS Y DISCUSIÓN

#### 3.1 Resultados

##### 3.1.1 Conexión de la LLM y la base de datos con LangChain para emplear lenguaje natural en consultas SQL

En primera instancia se elaboró un diagrama de bloques del funcionamiento general del asistente en base a la revisión bibliográfica Figura 1, en el cual se muestra las acciones generales que debe de ejecutar el asistente, desde el ingreso de la consulta del usuario hasta el resultado estructurado, incluyendo los bloques en los cuales funcional un modelo de lenguaje.

Para la implementación del asistente, se estableció que la interacción usuario-asistente se realizará en una interfaz con el framework Streamlit, se conectó la base de datos de la empresa Consditec en SQLite con un modelo LLM mediante el framework LangChain. Para esto, se utilizaron herramientas de conexión del módulo SQLiteDatabase, que permite operar bases de datos relacionales tipo SQLite y un modelo de lenguaje (LLM), desplegado de manera local a través de la plataforma Ollama en el servidor local de la empresa.

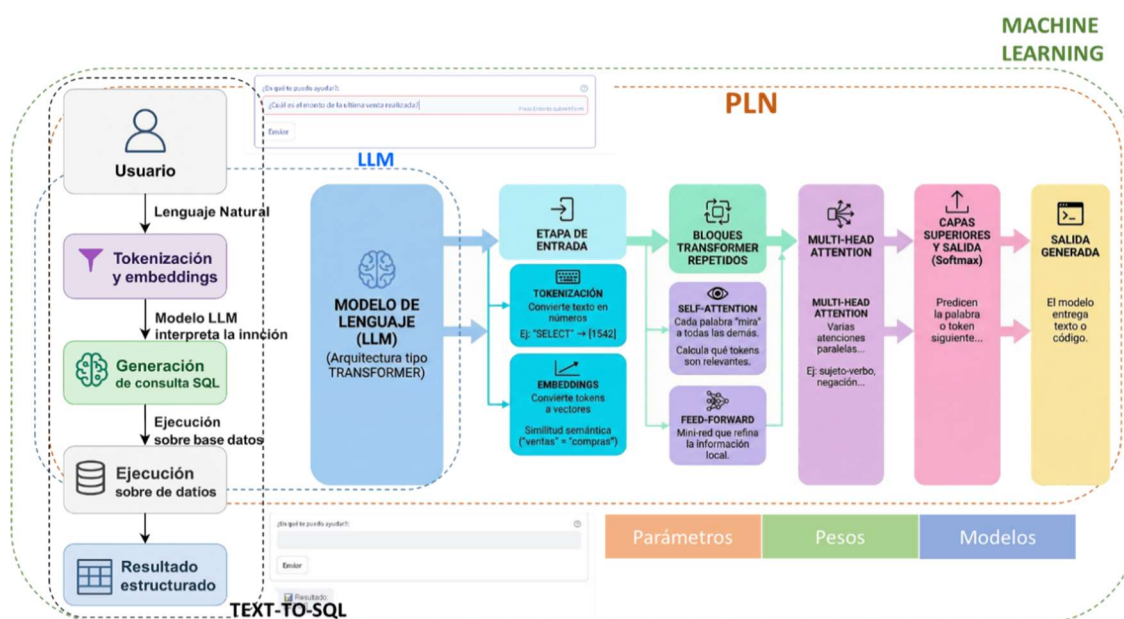
Para la conexión directa con la base de datos, se utilizó la clase SQLiteDatabase del módulo langchain\_community.utilities. Además, esta clase permite extraer la estructura de la base de datos, donde se adiciona al prompt

de entrada y permite parametrar la LLM, para luego enviarlo al modelo de lenguaje para la generación de la salida.

```
schema = db.get_table_info()
```

**Figura 1**

*Diagrama de bloques de funcionamiento.*



*Nota.* Diagrama entre la interacción usuario SQLite, mediante Langchain y LLM. Fuente: Elaboración propia.

En base a esto, se construyó la función PromptTemplate, en la cual se le parametriza al modelo explícitamente; que genere una sentencia SQL válida en función de la estructura del esquema; ante la petición del usuario ingresada en Streamlit, sin embargo, se obtuvieron resultados óptimos ingresando directamente la estructura de la base de datos implícita a través de una cadena de texto mediante la variable ddl. Así mismo, LangChain permite ajustar la temperatura de los modelos, lo que es relevante puesto que para el uso de la LLM se necesita que el modelo responda de la manera más puntual y específica; que la modalidad creativa que implicaría posibles alucinaciones con la generación de sentencias SQL.

```
llm = ChatOllama(model="llama3.2:3b-instruct-q4_K_M", temperature=0)
```

Para el procesamiento del lenguaje natural, se realizó una preselección de modelos que se amolden a la capacidad computacional del servidor, descartando moldeos que funcionen con más de 16GB de memoria RAM y también modelos que mostraban pobres desempeños en tiempos de respuesta y sintaxis, se evaluaron modelos como gemma3:4b-it-q4\_K\_M, phi3 o sqlcoder:7b-q4\_K\_M siendo este último entrenado específicamente para Text-to-SQL, no obstante, no se apreciaron rendimientos óptimos, mostrando salidas con errores o contenido textual no SQL incluso después de la parametrización del modelo mediante LangChain, los modelos seleccionados fueron: llama3.2:3b-instruct-q4\_K\_M, mistral-openorca:7b-q4\_K\_M, deepseek-r1:7b-qwen-distill-q4\_K\_M, y duckdb-nsq:7b-q4\_K\_M.

Adicionalmente se creó el LLMChain, esta clase orquesta el engineering prompt, la instancia de ChatOllama y la memoria opcional.

```
cadena_llm = LLMChain(llm=llm, prompt=prompt, memory=memoria)
```

El modelo de lenguaje fue encapsulado mediante la clase de LangChain, ChatOllama. La comunicación entre el modelo, la cadena de entrada del usuario y el esquema de la base de datos se gestionó a través de una LLMChain, la cual permitió generar la consulta SQL como salida, sin requerir intervención humana. LangChain fue seleccionado por tener una arquitectura modular, la cual permitió integrar varios modelos de lenguaje de una forma flexible con fuentes de datos estructurados. ChatOllama fue elegido por su compatibilidad con varios modelos LLM desplegados de forma local y por brindar un control directo sobre los modelos utilizados en el ámbito de desarrollo de software o de investigación. La base de datos SQLite corresponde a la base de datos utilizada actualmente por

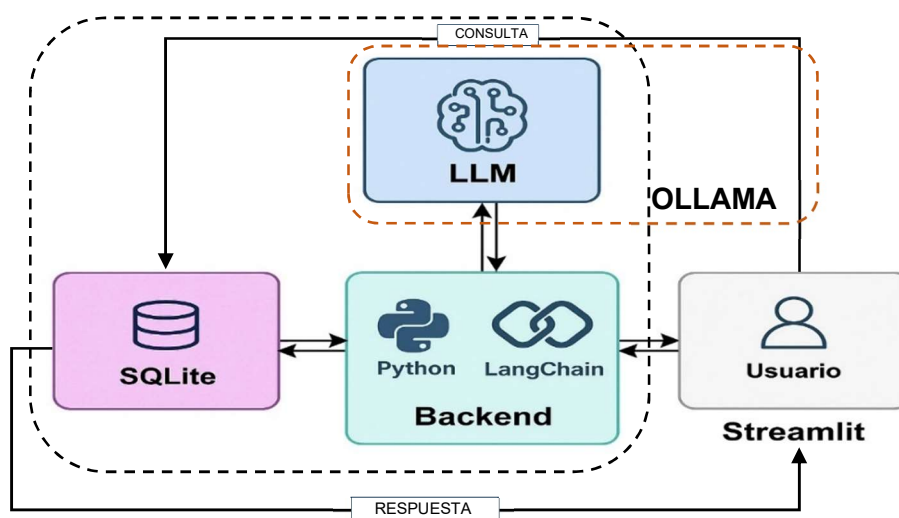
la empresa Consditec, lo que facilitó una conexión directa sin realizar modificaciones en el entorno de datos existente o alterar el funcionamiento normal del sistema de registro de inventario y stock con él se trabaja en la empresa.

Posteriormente, se integró con interfaz web; desarrollada con Streamlit, la cual actuó como frontend para la interacción con los demás componentes del asistente. Los usuarios ingresan consultas a través de un campo de texto, la cual es enviada al backend, donde se genera la sentencia SQL, este se ejecuta sobre la base de datos y se retorna automáticamente el resultado al frontend, donde se visualiza.

El siguiente diagrama muestra la arquitectura de conexión implementada:

**Figura 2**

*Diagrama de arquitectura de conexión implementada.*



*Nota.* Diagrama entre la interacción usuario SQLite, mediante Langchain y LLM. Fuente: Elaboración propia.

Se evidenció que la conexión se mantuvo estable durante el proceso de pruebas, observándose el uso de recursos adicionales por parte del modelo LLM en el servidor, lo que se tradujo en una carga adicional sobre la memoria RAM y

el CPU. No se presentaron errores durante la comunicación entre los distintos módulos del asistente.

### Figura 3

*Frontend, realizando consulta de prueba.*

*Nota.* Pagina de consulta principal, donde el usuario ingresa el prompt.  
*Fuente:* Elaboración propia.

### Figura 4

*Frontend, mostrando resultado de la consulta.*

*Nota.* Pagina de consulta principal, la LLM responde despues de haber realizado la consulta a SQLite. *Fuente:* Elaboración propia

En cuanto a las limitaciones, se identificó una relación directa entre el uso del modelo LLM y la capacidad de procesamiento de la máquina host. De igual manera, la escalabilidad del asistente dependerá tanto del tamaño de la base de datos como de la potencia del modelo LLM utilizada para procesar dicha información.

El servidor local tiene las siguientes características: Procesador Intel® Core™ i5-10400 caché de 12 M hasta 4,30 GHz, memoria RAM DDR3 16GB y tarjeta grafica de NVIDIA GeForce GTX 1650 4GB.

La implementación evidenció la viabilidad del uso de LangChain como enlace entre el lenguaje natural y la base de datos de la empresa, cumpliéndose de este modo el primer objetivo planteado en la investigación.

### **3.1.2 Evaluación, desempeño de los modelos LLM seleccionados para realizar consultas SQL**

Se evaluó el desempeño de los cuatro modelos de lenguaje seleccionados para la generación de consultas SQL. Para ello, se diseñó un procedimiento comparativo con base en un conjunto representativo de requerimientos que el personal ejecuta con mayor frecuencia sobre el motor SQLite. A partir de entrevistas con los usuarios, se consolidó un listado de diez peticiones frecuentes. Estas, expresadas en lenguaje natural, sirvieron como entradas estandarizadas para las pruebas controladas con los distintos modelos LLM. El entorno de evaluación se configuró sobre la arquitectura de LangChain, integrando la base de datos mediante el módulo SQLDatabase. Los modelos fueron seleccionados en función de su compatibilidad con la capacidad de procesamiento del equipo informático de la empresa.

Los modelos que cumplieron con dichos requisitos fueron: duckdb-  
mysql:7b-q4\_K\_M, mistral-openorca:7b-q4\_K\_M, deepseek-r1:7b-qwen-distill-  
q4\_K\_M y llama3.2:3b-instruct-q4\_K\_M. Posteriormente, se evaluaron dos  
dimensiones clave del método de observación estructurada aplicado:

- a. Eficiencia:** La primera dimensión midió el tiempo de respuesta del  
modelo, calculado automáticamente en milisegundos desde el momento  
en que se envió la consulta hasta la generación de la sentencia SQL.  
Para este propósito, se utilizó la librería `time` de Python.

```
inicio = time.perf_counter()  
print(f"Tiempo de respuesta del modelo: {tiempo_ms:.2f} ms")
```

- b. Calidad de la respuesta:** estimó dos niveles de evaluación para las  
sentencias SQL generadas:

- **Sintaxis válida:** Se comprueba si la consulta SQL generada fue  
sintácticamente correcta y pudo ejecutarse sin errores en el motor de  
SQLite, independientemente de que la respuesta obtenida no fuera  
necesariamente correcta.
- **Exactitud del resultado:** Se verifica si la ejecución de la sentencia SQL  
generada entrega el mismo resultado que el valor de referencia (Gold  
Standard), mediante su comparación con el dataset Spider.

Los resultados obtenidos a las 10 preguntas con más frecuencia fueron:

**Pregunta 1** ¿Cuál fue el monto total de las ventas registradas durante el mes de  
mayo de 2025?

El mes y el año indicados en la pregunta se consideraron referenciales, la  
consulta SQL Gold Standard del dataset Spider es:

```
SELECT SUM(total_venta) FROM ventas WHERE strftime('%Y-%m',  
fecha_venta) = '2025-05';
```

Resultado de la consulta a la base de datos: 71107.30

**Tabla 3**

*Resultados obtenidos ante la consulta: ¿Cuál fue el monto total de ventas del mes de mayo de 2025?*

| LLM | SQL generado                                                                                                                                      | Sintaxis<br>valida | Resultado<br>correcto | Tiempo<br>(ms) |
|-----|---------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|-----------------------|----------------|
| 1   | SELECT SUM(total_venta)<br>FROM ventas WHERE<br>strftime('%Y-%m', fecha_venta) =<br>'2025-05';                                                    | Si                 | Si                    | 8846.01        |
| 2   | SELECT SUM(total_venta) as<br>monto_total_ventas<br>FROM ventas<br>WHERE strftime('%Y-%m',<br>fecha_venta) = '2025-05';                           | Si                 | Si                    | 39954.61       |
| 3   | SELECT SUM(total_venta)<br>FROM ventas WHERE<br>fecha_venta LIKE '2025-05%';                                                                      | Si                 | Si                    | 26798.30       |
| 4   | SELECT SUM(total_venta) AS<br>monto_total<br>FROM ventas<br>WHERE strftime('%Y-%m-%d',<br>fecha_venta) BETWEEN '2025-<br>05-01' AND '2025-05-31'; | Si                 | Si                    | 36798.30       |

*Nota.* Donde: 1 llama3.2:3b, 2 mistral-openorca:7b, 3 duckdb-nsql:7b, 4 deepseek-r1:7b. Fuente: Elaboración propia.

Los resultados encontrados en la tabla 3 muestran una efectividad del 100% en validez sintáctica y precisión lógica. De todos los modelos, sin embargo, el modelo 1 resolvió la consulta en 8.8 segundos, superando ampliamente a los demás modelos.

**Pregunta 2** ¿Cuál fue el monto total de las compras realizadas durante el mes de mayo de 2025?

El mes y el año indicados en la pregunta se consideraron referenciales, la consulta SQL Gold Standard del dataset Spider es:

```
SELECT SUM(cantidad_comprada * precio_compra) AS  
monto_total_compras FROM compras WHERE strftime('%Y-%m',  
fecha_compra) = '2025-05';
```

Resultado de la consulta a la base de datos: 118232.88

**Tabla 4**

*Resultados obtenidos ante la consulta: ¿Cuál fue el monto total de las compras realizadas durante el mes de mayo de 2025?*

| LLM | SQL generado                                                                                                                                                                                                 | Sintaxis<br>valida | Resultado<br>correcto | Tiempo<br>(ms) |
|-----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|-----------------------|----------------|
| 1   | SELECT SUM(precio_compra)<br>FROM compras WHERE<br>fecha_compra LIKE '%2025-05%'                                                                                                                             | Si                 | No                    | 10302.29       |
| 2   | SELECT SUM(precio_compra) as<br>monto_compras_totales<br>FROM compras<br>WHERE strftime('%Y-%m',<br>fecha_compra) = '2025-05';                                                                               | Si                 | No                    | 11576.57       |
| 3   | SELECT<br>SUM(cantidad_comprada) FROM<br>compras WHERE fecha_compra<br>LIKE '2025-05%';                                                                                                                      | Si                 | No                    | 30767.71       |
| 4   | SELECT SUM(precio_compra) as<br>monto_total<br>FROM compras<br>WHERE strftime('%Y-%m-%d',<br>substr(fecha_compra, 0, 7),<br>substr(fecha_compra, 8, 3),<br>substr(fecha_compra, 12, 2))<br>LIKE '2025-05-%'; | No                 | No                    | 144737.6<br>6  |

*Nota.* Donde: 1 llama3.2:3b, 2 mistral-openorca:7b, 3 duckdb-nsql:7b, 4 deepseek-r1:7b. Fuente: Elaboración propia.

La tabla 4 evidencia un desempeño crítico, todos los modelos generaron una sintaxis válida, pero ninguno logró obtener el resultado correcto. Destaca que los modelos 1, 2 y 3 fallaron en la lógica de cálculo en cantidad y precio.

**Pregunta 3:** ¿Cuál fue el producto más vendido durante el mes de febrero de 2025?

El mes y el año indicados en la pregunta se consideraron referenciales, la consulta SQL Gold Standard del dataset Spider es:

```
SELECT producto, SUM(cantidad_vendida) FROM ventas WHERE  
strftime('%Y-%m', fecha_venta) = '2025-02' GROUP BY producto  
ORDER BY total_vendida DESC;
```

Resultado de la consulta a la base de datos: 'USB', 360

**Tabla 5**

*Resultados obtenidos ante la consulta: ¿Cuál fue el producto más vendido durante el mes de febrero de 2025?*

| LLM | SQL generado                                                                                                                                                                                                                   | Sintaxis<br>valida | Resultado<br>correcto | Tiempo<br>(ms) |
|-----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|-----------------------|----------------|
| 1   | SELECT producto FROM ventas<br>WHERE fecha_venta =<br>date('now', 'first day of 2025')<br>AND cantidad_vendida = (<br>SELECT MAX(cantidad_vendida)<br>FROM ventas WHERE<br>fecha_venta = date('now', 'first<br>day of 2025') ) | No                 | No                    | 15634.62       |
| 2   | SELECT producto,<br>SUM(cantidad_vendida) AS<br>cantidad_total<br>FROM ventas<br>WHERE strftime('%Y-%m',<br>fecha_venta) = '2025-02'<br>GROUP BY producto<br>ORDER BY cantidad_total DESC<br>LIMIT 1;                          | Si                 | Si                    | 46351.92       |
| 3   | SELECT producto FROM ventas<br>WHERE fecha_venta LIKE '2025-<br>02%' GROUP BY producto<br>ORDER BY COUNT(*) DESC<br>LIMIT 1;                                                                                                   | Si                 | No                    | 32244.25       |
| 4   | SELECT producto, total_venta<br>FROM ventas<br>WHERE strftime('%Y-%m-%d',<br>fecha_venta) BETWEEN '2025-<br>02-01' AND '2025-02-28'<br>GROUP BY producto<br>ORDER BY total_venta DESC<br>LIMIT 1;                              | Si                 | No                    | 157461.9<br>1  |

*Nota.* Donde: 1 llama3.2:3b, 2 mistral-openorca:7b, 3 duckdb-nsql:7b, 4 deepseek-r1:7b. Fuente: Elaboración propia.

La tabla 5 muestra que solo el modelo 2 logró un resultado correcto y sintaxis válida. En contraste, los demás modelos fallaron en la precisión o la sintaxis, con modelo 4 registrando el tiempo de respuesta más alto.

**Pregunta 4** ¿Cuál es el producto que menos se vendió en el mes de junio del 2025?

El mes y el año indicados en la pregunta se consideraron referenciales, la consulta SQL Gold Standard del dataset Spider es:

```
SELECT producto, SUM(cantidad_vendida) FROM ventas WHERE  
strftime('%Y-%m', fecha_venta) = '2025-06' GROUP BY producto  
ORDER BY total_vendida ASC LIMIT 1;
```

Resultado de la consulta: 'ADAPTADOR DISPLAY PORT A HDMI', 1

### Tabla 6

Resultados ante la consulta ¿Cuál es el producto que menos se vendió en el mes de junio del 2025?

| LLM | SQL generado                                                                                                                                                                                                    | Sintaxis<br>valida | Resultado<br>correcto | Tiempo<br>(ms) |
|-----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|-----------------------|----------------|
| 1   | SELECT producto FROM ventas<br>WHERE strftime('%Y-%m',<br>fecha_venta) = '2025-06' GROUP<br>BY producto ORDER BY<br>SUM(cantidad_vendida) LIMIT 1                                                               | Si                 | Si                    | 10490.27       |
| 2   | SELECT producto,<br>SUM(cantidad_vendida) AS<br>cantidad_total_ventas FROM<br>ventas WHERE strftime('%Y-%m',<br>fecha_venta) = '2025-06' GROUP<br>BY producto ORDER BY<br>cantidad_total_ventas ASC LIMIT<br>1; | Si                 | Si                    | 47563.51       |
| 3   | SELECT producto FROM ventas<br>WHERE id_venta = 1005;                                                                                                                                                           | Si                 | No                    | 23780.39       |
| 4   | SELECT producto,<br>MIN(total_venta) AS<br>ventas_menores FROM ventas<br>WHERE fecha_venta LIKE '2025-<br>06-%' GROUP BY producto<br>ORDER BY ventas_menores<br>ASC;                                            | Si                 | No                    | 199254.4<br>8  |

*Nota.* Donde: 1 llama3.2:3b, 2 mistral-openorca:7b, 3 duckdb-nsql:7b, 4 deepseek-r1:7b. Fuente: Elaboración propia.

**Pregunta 5** ¿Cuál fue el total de ventas realizadas mediante pagos en efectivo durante el mes de junio de 2025?

El mes y el año indicados en la pregunta se consideraron referenciales, la consulta SQL Gold Standard del dataset Spider es:

```
SELECT SUM(total_venta) FROM ventas WHERE strftime('%Y-%m',  
fecha_venta) = '2025-06' AND modo_pago = 'efectivo';
```

Resultado de la consulta a la base de datos: 54867.4

### Tabla 7

*Resultados obtenidos ante la consulta: ¿Cuál fue el total de ventas realizadas mediante pagos en efectivo durante el mes de junio de 2025?*

| LLM | SQL generado                                                                                                                                                                                 | Sintaxis<br>valida | Resultado<br>correcto | Tiempo<br>(ms) |
|-----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|-----------------------|----------------|
| 1   | SELECT SUM(total_venta)<br>FROM ventas WHERE<br>strftime('%Y-%m', fecha_venta) =<br>'2025-06' AND modo_pago =<br>'efectivo';                                                                 | Si                 | Si                    | 13576.39       |
| 2   | SELECT SUM(total_venta) as<br>total_ventas_efectivo FROM<br>ventas WHERE modo_pago =<br>'efectivo' AND strftime('%m',<br>fecha_venta) = '06' AND<br>strftime('%Y', fecha_venta) =<br>'2025'; | Si                 | Si                    | 55039.12       |
| 3   | SELECT SUM(cantidad_vendida)<br>FROM ventas WHERE<br>modo_pago = 'efectivo' AND<br>strftime('%m', fecha_venta) = '06'<br>AND strftime('%Y', fecha_venta)<br>= '2025';                        | Si                 | Si                    | 43779.87       |
| 4   | SELECT SUM(total_venta)<br>FROM ventas WHERE<br>strftime('%Y', fecha_venta) = '2025'<br>AND strftime('%m', fecha_venta) =<br>'06' AND modo_pago = 'efectivo';                                | Si                 | Si                    | 206459.1<br>0  |

*Nota.* Donde: 1 llama3.2:3b, 2 mistral-openorca:7b, 3 duckdb-nsql:7b, 4 deepseek-r1:7b. Fuente: Elaboración propia.

**Pregunta 6** ¿Cuál fue el monto total de las ventas correspondientes a la última semana del mes de mayo?

El mes y el año indicados en la pregunta se consideraron referenciales, la consulta SQL Gold Standard del dataset Spider es:

```
SELECT SUM(total_venta)  
FROM ventas WHERE fecha_venta BETWEEN '2025-05-25' AND '2025-  
05-31';
```

Resultado de la consulta a la base de datos: 13700.10

**Tabla 8**

*Resultados obtenidos ante la consulta: ¿Cuál fue el monto total de las ventas correspondientes a la última semana del mes de mayo?*

| LLM | SQL generado                                                                                                                                                                            | Sintaxis valida | Resultado correcto | Tiempo (ms)   |
|-----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------|--------------------|---------------|
| 1   | SELECT SUM(total_venta)<br>FROM ventas WHERE<br>strftime('%Y-%m', fecha_venta) =<br>'2023-05' AND strftime('%W',<br>fecha_venta) = 22                                                   | No              | No                 | 14689.39      |
| 2   | SELECT SUM(total_venta) as<br>monto_total_ventas<br>FROM ventas<br>WHERE fecha_venta >=<br>date('now', 'start of week', '-1')<br>AND fecha_venta < date('now',<br>'end of week', '-1'); | No              | No                 | 44074.41      |
| 3   | SELECT SUM(total_venta)<br>FROM ventas WHERE<br>strftime(fecha_venta, '%Y-%m') =<br>'2022-05' AND cantidad_vendida<br>> 0;                                                              | No              | No                 | 37551.98      |
| 4   | SELECT SUM(total_venta) as<br>monto_total<br>FROM ventas<br>WHERE fecha_venta BETWEEN<br>'2025-05-25' AND '2025-05-31';                                                                 | Si              | Si                 | 112925.0<br>0 |

*Nota.* Donde: 1 llama3.2:3b, 2 mistral-openorca:7b, 3 duckdb-nsql:7b, 4 deepseek-r1:7b. Fuente: Elaboración propia.

**Pregunta 7** ¿Qué productos ingresados durante el mes de junio de 2025 presentaron un stock igual a cero?

El mes y el año indicados en la pregunta se consideraron referenciales, la consulta SQL Gold Standard del dataset Spider es:

```
SELECT producto, marca, modelo, capacidad, descripcion_producto  
FROM compras WHERE strftime('%Y-%m', fecha_compra) = '2025-06'  
AND stock_actual = 0;
```

Resultado de la consulta a la base de datos: ('WEB CAM',), ('HOLDER LAPTOP',), ('SMARTBAND',), ('MOUSE GAMER',), ('AUDIFONO PC',),

('MOUSE PAD'), ('SMARTWATCH'), ('TECLADO MECANICO'), ('DISCO DURO INTERNO'),)

**Tabla 9**

*Resultados obtenidos ante la consulta: ¿Qué productos ingresados durante el mes de junio de 2025 presentaron un stock igual a cero?*

| LLM | SQL generado                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | Sintaxis valida | Resultado correcto | Tiempo (ms)   |
|-----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------|--------------------|---------------|
| 1   | SELECT DISTINCT T1.producto<br>FROM compras AS T1 INNER<br>JOIN ventas AS T2 ON<br>T1.id_codigo = T2.id_codigo<br>WHERE strftime('%Y-%m',<br>T1.fecha_compra) = '2025-06'<br>AND T1.stock_actual = 0                                                                                                                                                                                                                                                                           | Si              | Si                 | 18711.46      |
| 2   | SELECT p.producto, p.marca,<br>p.modelo, p.capacidad,<br>p.descripcion_producto,<br>pv.stock_actual FROM compras<br>AS c JOIN productos AS p ON<br>c.id_codigo = p.id_codigo<br>LEFT JOIN ( SELECT id_venta,<br>SUM(cantidad_vendida) AS<br>cantidad_total_vendida FROM<br>ventas GROUP BY id_venta<br>) AS pv ON c.id_codigo =<br>pv.id_codigo WHERE<br>strftime('%Y-%m',<br>c.fecha_compra) = '2025-06' AND<br>pv.cantidad_total_vendida IS<br>NULL AND pv.stock_actual = 0; | No              | No                 | 100850.3<br>5 |
| 3   | SELECT * FROM ventas WHERE<br>cantidad_vendida = 0 AND<br>strftime(fecha_venta, '%Y-%m') =<br>'2025-06';                                                                                                                                                                                                                                                                                                                                                                       | No              | No                 | 35830.76      |
| 4   | SELECT producto FROM<br>compras WHERE stock_actual =<br>0 AND fecha_compra >= '2025-<br>06-01' AND fecha_compra <=<br>'2025-06-30' iflag=1                                                                                                                                                                                                                                                                                                                                     | No              | No                 | 265835.1<br>5 |

*Nota.* Donde: 1 llama3.2:3b, 2 mistral-openorca:7b, 3 duckdb-nsql:7b, 4 deepseek-r1:7b. Fuente: Elaboración propia.

**Pregunta 8** ¿Cuál fue el monto total de las compras realizadas durante el mes de junio de 2025?



El mes y el año indicados en la pregunta se consideraron referenciales, la consulta SQL Gold Standard del dataset Spider es:

```
SELECT SUM(cantidad_comprada * precio_compra) AS  
monto_total_compras_junio FROM compras WHERE strftime('%Y-%m',  
fecha_compra) = '2025-06';
```

Resultado de la consulta a la base de datos: 107543.44

**Tabla 10**

*Resultados obtenidos ante la consulta: ¿Cuál fue el monto total de las compras realizadas durante el mes de junio de 2025?*

| LLM | SQL generado                                                                                                                                          | Sintaxis<br>valida | Resultado<br>correcto | Tiempo<br>(ms) |
|-----|-------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|-----------------------|----------------|
| 1   | SELECT SUM(c.precio_compra *<br>c.cantidad_comprada) FROM<br>compras c WHERE strftime('%Y-<br>%m', c.fecha_compra) = '2025-06'                        | Si                 | Si                    | 14785.46       |
| 2   | SELECT SUM(precio_compra)<br>AS monto_total<br>FROM compras<br>WHERE strftime('%Y-%m',<br>fecha_compra) = '2025-06';                                  | Si                 | No                    | 42403.34       |
| 3   | SELECT<br>SUM(cantidad_comprada) FROM<br>compras WHERE fecha_compra<br>LIKE '2025-06%';                                                               | Si                 | No                    | 30779.19       |
| 4   | SELECT SUM(precio_compra *<br>cantidad_comprada) AS<br>monto_total<br>FROM compras<br>WHERE fecha_compra<br>BETWEEN '2025-06-01' AND<br>'2025-06-30'; | Si                 | Si                    | 162586.7<br>1  |

*Nota.* Donde: 1 llama3.2:3b, 2 mistral-openorca:7b, 3 duckdb-nsql:7b, 4 deepseek-r1:7b. Fuente: Elaboración propia.

Los resultados de la Tabla 10 destacan que a diferencia del primer intento, esta vez 1 y 4 si lograron el resultado correcto al integrar la operación aritmética de precio por cantidad. No obstante, el modelo 1 sobresale nuevamente por su eficiencia, resolviendo la consulta en apenas 14.7 segundos frente a los 162.5 segundos que requirió el modelo 4.

**Pregunta 9** ¿Cuál fue el monto total de las ventas realizadas mediante pagos con Yape durante el mes de junio de 2025?

El mes y el año indicados en la pregunta se consideraron referenciales, la consulta SQL Gold Standard del dataset Spider es:

```
SELECT SUM(total_venta) FROM ventas
WHERE strftime('%Y-%m', fecha_venta) = '2025-06' AND
LOWER(modo_pago) = 'yape';
```

Resultado de la consulta a la base de datos: 16564.0

**Tabla 11**

*Resultados obtenidos ante la consulta: ¿Cuál fue el monto total de las ventas realizadas mediante pagos con Yape durante el mes de junio de 2025?*

| LLM | SQL generado                                                                                                                                                                          | Sintaxis valida | Resultado correcto | Tiempo (ms)   |
|-----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------|--------------------|---------------|
| 1   | SELECT SUM(total_venta)<br>FROM ventas WHERE<br>strftime('%Y-%m', fecha_venta) =<br>'2025-06' AND modo_pago =<br>'yape'                                                               | Si              | Si                 | 12382.44      |
| 2   | SELECT SUM(total_venta) AS<br>monto_total_ventas<br>FROM ventas<br>WHERE proveedor = 'YAPE' AND<br>strftime('%m', fecha_venta) = '06'<br>AND strftime('%Y', fecha_venta)<br>= '2025'; | No              | No                 | 50590.92      |
| 3   | SELECT SUM(total_venta)<br>FROM ventas WHERE<br>fecha_venta LIKE '2025-06%'<br>AND modo_pago = 'yape';                                                                                | Si              | Si                 | 33000.38      |
| 4   | SELECT total_venta FROM<br>ventas WHERE mode_pago =<br>'YAPE' AND<br>substr(fecha_venta,7,2) = '06'<br>AND substr(fecha_venta,4,3) =<br>'2025';                                       | No              | No                 | 560578.5<br>2 |

*Nota.* Donde: 1 llama3.2:3b, 2 mistral-openorca:7b, 3 duckdb-nsql:7b, 4 deepseek-r1:7b. Fuente: Elaboración propia.

**Pregunta 10** ¿Cuántos discos duros externos se vendieron durante el mes de junio de 2025?

El mes y el año indicados en la pregunta se consideraron referenciales, la consulta SQL Gold Standard del dataset Spider es:

```
SELECT SUM(cantidad_vendida) FROM ventas WHERE strftime('%Y-%m', fecha_venta) = '2025-06' AND LOWER(producto) = 'disco duro externo';
```

Resultado de la consulta a la base de datos: 75

**Tabla 12**

*Resultados obtenidos ante la consulta: ¿Cuántos discos duros externos se vendieron durante el mes de junio de 2025?*

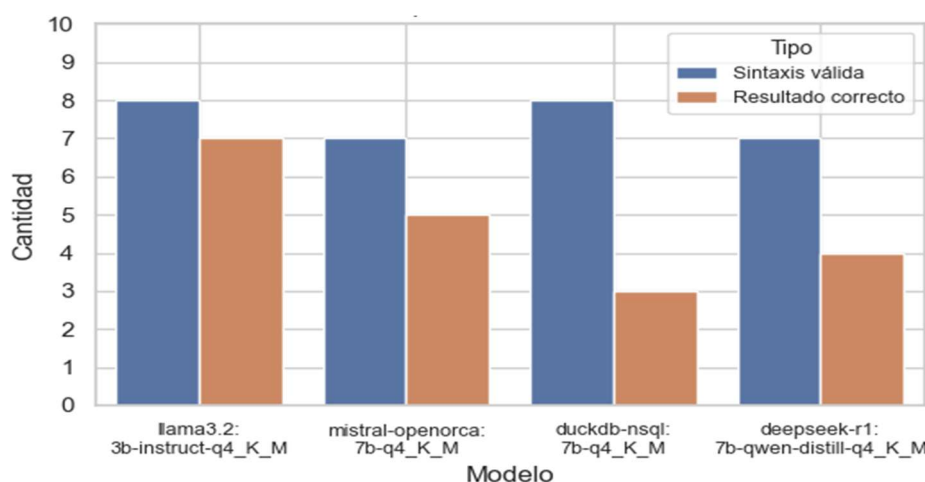
| LLM | SQL generado                                                                                                                                                                                     | Sintaxis valida | Resultado correcto | Tiempo (ms) |
|-----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------|--------------------|-------------|
| 1   | SELECT COUNT(id_codigo) FROM ventas WHERE strftime('%Y-%m', fecha_venta) = '2025-06' AND producto = 'disco duro externo';                                                                        | Si              | Si                 | 16544.85    |
| 2   | SELECT COUNT(*) FROM ventas WHERE producto = 'disco duro externo' AND strftime('%m', fecha_venta) = '06' AND strftime('%Y', fecha_venta) = '2025';                                               | Si              | Si                 | 50880.21    |
| 3   | SELECT COUNT(*) FROM ventas WHERE producto = 'disco duro externo' AND strftime(fecha_venta, '%Y-%m') = '2025-06';                                                                                | Si              | No                 | 40299.96    |
| 4   | SELECT COUNT(id_venta) AS cnt_disco_duro_exterior FROM ventas WHERE strftime('%Y', fecha_venta) = 2025 AND strftime('%m', fecha_venta) = 6 AND descripcion_producto LIKE '%disco duro externo%'; | Si              | No                 | 171442.38   |

*Nota.* Donde: 1 llama3.2:3b, 2 mistral-openorca:7b, 3 duckdb-nsql:7b, 4 deepseek-r1:7b. Fuente: Elaboración propia.

En base a los datos obtenidos, se realizó una tabla de resultados comparativos.

**Figura 5**

*Gráfico de barras de los resultados de validez sintáctica y corrección por modelo*



*Nota.* Resultados correctos y validos de los 4 modelos LLM evaluados en la escala de 0 a 10. Fuente: Elaboración propia.

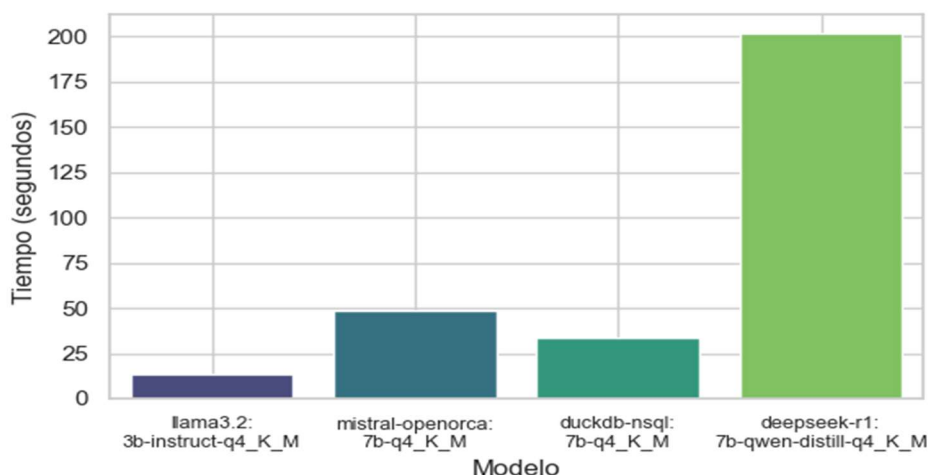
De acuerdo con los resultados ilustrados en la Figura 5, las arquitecturas llama3.2:3b-instruct-q4\_K\_M y duckdb-nsq:7b-q4\_K\_M mostraron la mayor eficacia en la construcción de sentencias con sintaxis SQL válida, logrando una tasa de éxito del 80% (8 de 10 ejecuciones), con la diferencia de que llama3.2:3b-instruct-q4\_K\_M denota una superioridad cualitativa al presentar una combinación de resultados correctos y validez sintáctica. Por otro lado, los modelos mistral-openorca:7b-q4\_K\_M y deepseek-r1:7b-qwen-distill-q4\_K\_M mostraron una precisión más reducida en la entrega de resultados correctos; si bien muestran un rendimiento estable de validez sintáctica con 7 de 10 intentos, su capacidad para interpretar la intención de la consulta fue menor en resultados correctos.

En conjunto, los datos muestran que el modelo llama3.2:3b-instruct-q4\_K\_M proporcionó el desempeño más consistente al traducir lenguaje natural a sentencias SQL bajo el entorno de prueba. No obstante, cabe señalar que estos resultados no alcanzaron niveles óptimos puesto que el modelo más

eficiente solo logro una tasa de éxito de un 70 % de precisión semántica, esto indica la necesidad de priorizar la fidelidad semántica como un indicador crítico de desempeño, más allá del simple cumplimiento de las reglas gramaticales del lenguaje estructurado.

**Figura 6**

*Gráfico de barras del tiempo medio de respuesta según el modelo LLM*



*Nota.* Tiempo promedio en milisegundos de respuesta brindada por cada modelo. Fuente: Elaboración propia

La figura 6 presenta la tabla de comparación de desempeño de los tiempos de respuesta promedio por modelo, donde el modelo llama3.2:3b-instruct-q4\_K\_M se posicionó como la opción más ágil con un promedio de 13.20s, superando ampliamente a duckdb-nsql:7b-q4\_K\_M, que registró 37.03s. En el extremo opuesto, mistral-openorca:7b-q4\_K\_M y deepseek-r1:7b-qwen-distill-q4\_K\_M demandaron tiempos significativamente mayores; este último sobrepasó los tres minutos de promedio por respuesta con una marcada incertidumbre en su rendimiento. Estas medidas indican que la elección del modelo influye directamente en la viabilidad operativa, puesto que una configuración incorrecta resulta en latencias prolongadas e inasumibles para entornos de producción en tiempo real.

Asimismo se observa que modelos como `mistral-openorca:7b-q4_K_M`, `deepseek-r1:7b-qwen-distill-q4_K_M` tienden a generar funciones incompatibles con el motor SQLite, tales como `EXTRACT` o `INTERVAL`, aun cuando se establecieron restricciones de parámetros con LangChain para la tarea, estos modelos produjeron sentencias con una estructura lógica coherente pero incompatible con SQLite. Esto destaca que la compatibilidad con el motor de la base de datos es un factor tan relevante para tomar en cuenta en el diseño como la correcta interpretación de la intención del usuario.

De esta manera, en relación con el objetivo de optimización del asistente de consultas para la empresa Consditec, la latencia de respuesta se consolidó como un factor determinante para la satisfacción del usuario, considerado en el tercer objetivo. Tras contrastar los niveles de precisión frente a los tiempos de ejecución, se concluyó que el modelo `llama3.2:3b-instruct-q4_K_M` proporciona el equilibrio más eficaz entre fidelidad semántica y eficiencia. Estos hallazgos confirman la viabilidad técnica de implementar una solución basada en dicha arquitectura y el marco de trabajo LangChain, para la mejora del acceso a los datos de la empresa Consditec dando cumplimiento al segundo objetivo específico.

### **3.1.3 Descripción del uso del lenguaje natural y su influencia en la simplificación de consultas SQL**

En relación con el objetivo específico 3, se realizó el análisis de la información obtenida a partir del instrumento de recolección de datos, en base a la implementación del asistente para realizar consultas SQL en lenguaje natural, a partir del cual se encontraron los siguientes resultados:

**Tabla 13**

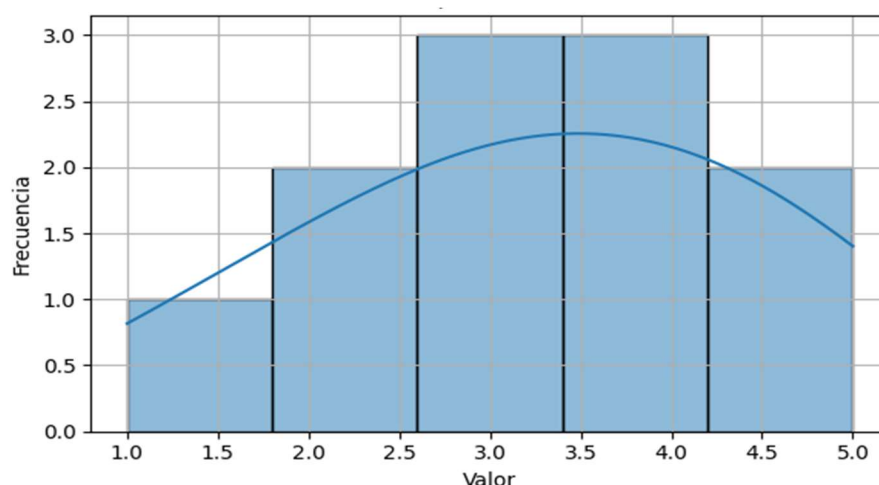
*¿Con qué frecuencia necesitas obtener información específica desde el sistema de datos para realizar tu trabajo?*

| Frecuencia Acceso  | Frecuencia | Porcentaje | Porcentaje válido | Porcentaje acumulado |
|--------------------|------------|------------|-------------------|----------------------|
| Nunca              | 1          | 9.09       | 9.09              | 9.09                 |
| Raramente          | 2          | 18.18      | 18.18             | 27.27                |
| Ocasionalmente     | 3          | 27.27      | 27.27             | 54.55                |
| Frecuentemente     | 3          | 27.27      | 27.27             | 81.82                |
| Muy frecuentemente | 2          | 18.18      | 18.18             | 100                  |

*Nota.* Porcentaje valido frecuencia de acceso. Fuente: Elaboración propia.

**Figura 7**

*Grafico de distribución de respuestas, frecuencia acceso.*



*Nota.* Distribución de las respuestas, frecuencia de acceso. Fuente: Elaboración propia.

Los datos recolectados en la Tabla 13 revelan que más de la mitad de los colaboradores (54.6%) integran las consultas al sistema como una práctica recurrente en sus funciones, frente a un reducido 9.1% que reporta un uso marginal. En sintonía con esto, la distribución de frecuencias de la Figura 6 exhibe una concentración en los niveles superiores de la escala (valores 3 y 4) y un sesgo positivo, factores que validan una disposición favorable y una adopción progresiva de estas herramientas para la gestión de información interna.

**Tabla 14**

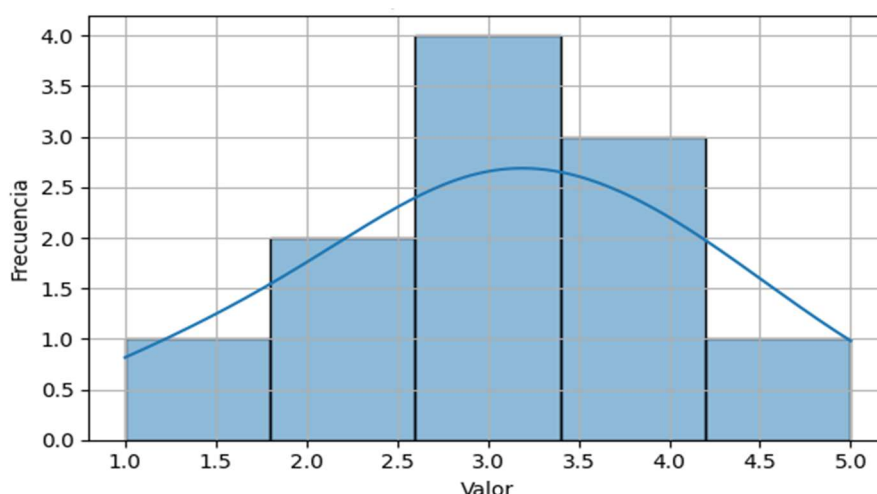
*¿Qué tan cómodo/a te sientes utilizando herramientas digitales que te permiten consultar información (como filtros, menús o buscadores)?*

| Comodidad Herramientas | Frecuencia | Porcentaje | Porcentaje válido | Porcentaje acumulado |
|------------------------|------------|------------|-------------------|----------------------|
| Muy incómodo/a         | 1          | 9.09       | 9.09              | 9.09                 |
| Incómodo/a             | 2          | 18.18      | 18.18             | 27.27                |
| Neutral                | 4          | 36.36      | 36.36             | 63.64                |
| Cómodo/a               | 3          | 27.27      | 27.27             | 90.91                |
| Muy cómodo/a           | 1          | 9.09       | 9.09              | 100                  |

*Nota.* Porcentaje válido del nivel de comodidad en el uso de herramientas digitales. Fuente: Elaboración propia.

**Figura 8**

*Gráfico de distribución de respuestas; comodidad herramientas digitales*



*Nota.* Gráfico de distribución del nivel de comodidad en el uso de herramientas digitales. Fuente: Elaboración propia.

La Tabla 14 indica que una mayoría significativa de los trabajadores (63.87%) percibe un nivel de comodidad y neutralidad frente al manejo de los recursos digitales actuales, como el filtrado de datos o la búsqueda por códigos y solo el 9.1% expresó una disconformidad total con el asistente. Por su parte, la Figura 8 revela una densidad de respuestas centrada en el valor intermedio de

la escala, con un ligero desplazamiento hacia los rangos superiores, lo que ratifica una predisposición aceptable y estable hacia el entorno tecnológico.

**Tabla 15**

*¿Qué tan fácil te resultó expresar lo que necesitabas mediante lenguaje natural? (escribiendo como lo dirías a otra persona)*

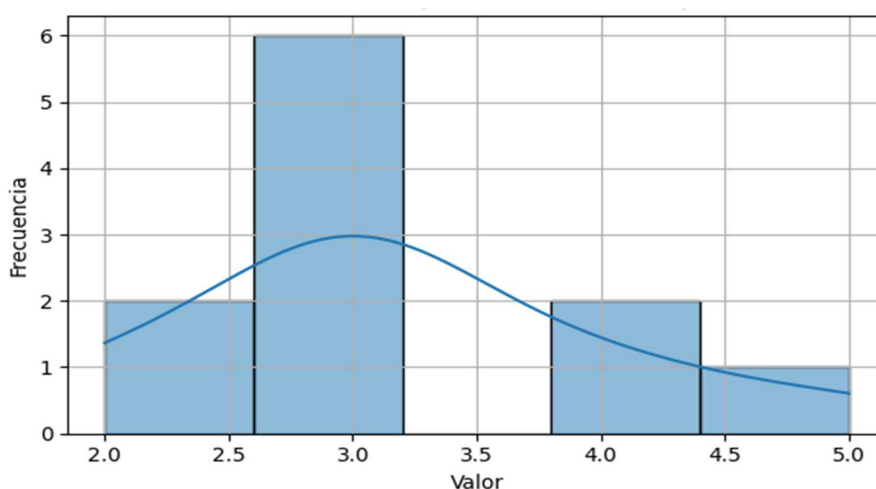
| Facilidad Expresión | Frecuencia | Porcentaje | Porcentaje válido | Porcentaje acumulado |
|---------------------|------------|------------|-------------------|----------------------|
| Difícil             | 2          | 18.18      | 18.18             | 18.18                |
| Neutral             | 6          | 54.55      | 54.55             | 72.73                |
| Fácil               | 2          | 18.18      | 18.18             | 90.91                |
| Muy fácil           | 1          | 9.09       | 9.09              | 100                  |

*Nota.* Porcentaje válido, facilidad de expresión de uso, lenguaje natural.

Fuente: Elaboración propia.

**Figura 9**

*Gráfico de distribución de respuestas, facilidad expresión de uso*



*Nota.* Gráfico de distribución, facilidad de expresión de uso, lenguaje natural. Fuente: Elaboración propia.

Los datos de la Tabla 15 revelan que el 54.55% de los colaboradores mantiene una postura neutral respecto a la interacción mediante lenguaje natural. Esta percepción se atribuye a que, si bien la herramienta simplifica la redacción de peticiones, las pruebas del segundo objetivo específico confirmaron que la precisión semántica no supera el umbral del 70%. Por otro lado, un 27%

de los usuarios calificó la experiencia de forma positiva. Finalmente, la distribución de frecuencias en la Figura 9 confirma esta tendencia al concentrar la mayoría de las valoraciones en el valor central de la escala, reflejando una aceptación moderada condicionada por el margen de error del sistema.

**Tabla 16**

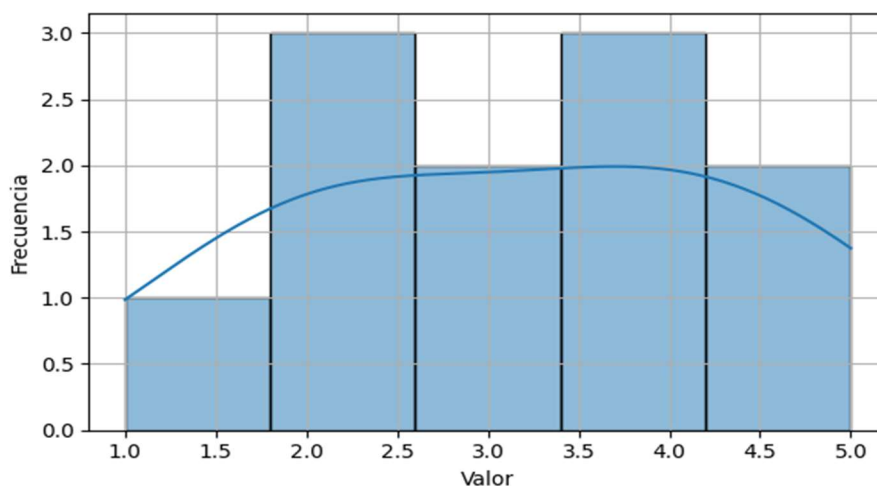
*¿Qué tan útil fue poder escribir libremente lo que buscabas, en lugar de usar filtros o códigos como antes?*

| Utilidad<br>Respuestas | Frecuencia | Porcentaje | Porcentaje<br>válido | Porcentaje<br>acumulado |
|------------------------|------------|------------|----------------------|-------------------------|
| Nada útiles            | 1          | 9.09       | 9.09                 | 9.09                    |
| Poco útiles            | 3          | 27.27      | 27.27                | 36.36                   |
| Medianamente<br>útiles | 2          | 18.18      | 18.18                | 54.55                   |
| Útiles                 | 3          | 27.27      | 27.27                | 81.82                   |
| Muy útiles             | 2          | 18.18      | 18.18                | 100                     |

*Nota.* Porcentaje válido de la utilidad de respuestas proporcionadas por el asistente. Fuente: Elaboración propia.

**Figura 10**

*Gráfico de distribución de respuestas, utilidad de respuestas*



*Nota.* Gráfico de distribución del nivel de utilidad de respuestas. Fuente: Elaboración propia.

De acuerdo con la Tabla 16, la percepción de utilidad se dividió equitativamente: un 45.45% de los usuarios calificó las respuestas como

altamente beneficiosas, mientras que un porcentaje idéntico las consideró de utilidad moderada. Esta dualidad se refleja en la Figura 10, donde la distribución de frecuencias muestra un comportamiento bimodal con concentraciones en los rangos intermedio y superior. En conjunto, este patrón evidencia una valoración favorable hacia el desempeño de la herramienta, validando la relevancia práctica de los resultados generados mediante lenguaje natural.

**Tabla 17**

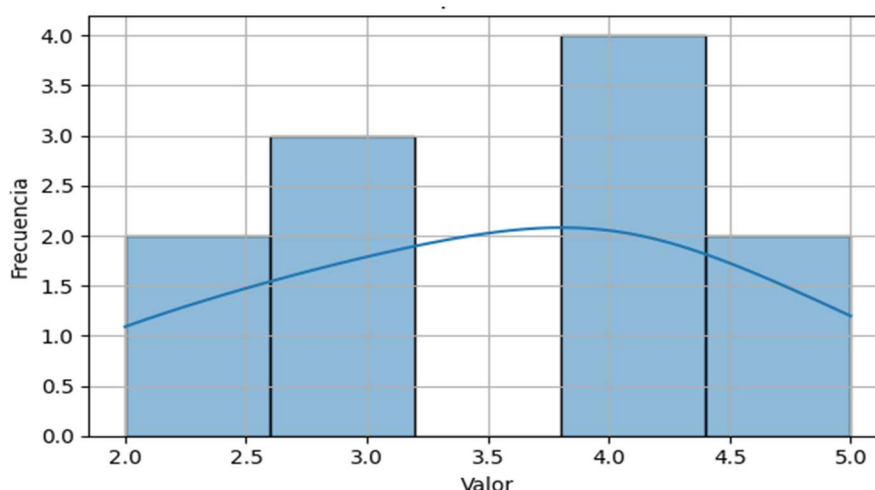
*¿En qué medida el nuevo asistente te ayudó a obtener los datos correctos sin necesidad de saber el código o ubicación exacta?*

| Grado Facilitación | Frecuencia | Porcentaje | Porcentaje válido | Porcentaje acumulado |
|--------------------|------------|------------|-------------------|----------------------|
| Muy poco           | 2          | 18.18      | 18.18             | 18.18                |
| Algo               | 3          | 27.27      | 27.27             | 45.45                |
| Bastante           | 4          | 36.36      | 36.36             | 81.82                |
| Mucho              | 2          | 18.18      | 18.18             | 100                  |

*Nota.* Porcentaje valido del grado de utilidad. Fuente: Elaboración propia.

**Figura 11**

*Gráfico de distribución de respuestas, grado de utilidad.*



*Nota.* Distribución del nivel de utilidad de las respuestas. Fuente: Elaboración propia.

La Tabla 17 muestra que el 54.54% de los trabajadores percibió una mejora significativa en la accesibilidad a la información frente a los métodos

convencionales de consulta en la base de datos. La Figura 11 exhibe una distribución con asimetría positiva y un valor modal de 5; este comportamiento ratifica una valoración favorable, confirmando que la implementación del asistente simplifica la interacción con la BBDD de la empresa.

**Tabla 18**

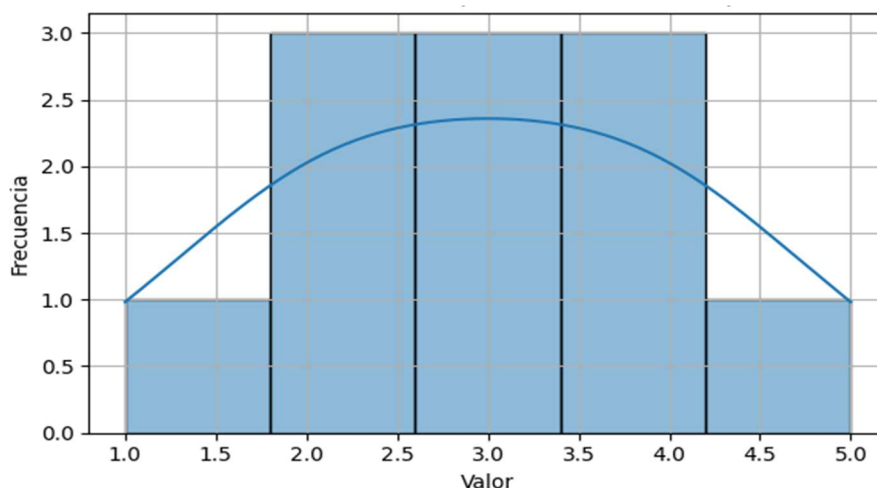
*¿Qué impacto tuvo el uso del asistente en el tiempo que dedicaste a obtener la información que necesitabas?*

| Ahorro Tiempo | Frecuencia | Porcentaje | Porcentaje válido | Porcentaje acumulado |
|---------------|------------|------------|-------------------|----------------------|
| Nada          | 1          | 9.09       | 9.09              | 9.09                 |
| Poco          | 3          | 27.27      | 27.27             | 36.36                |
| Moderado      | 3          | 27.27      | 27.27             | 63.64                |
| Bastante      | 3          | 27.27      | 27.27             | 90.91                |
| Mucho         | 1          | 9.09       | 9.09              | 100                  |

*Nota.* Porcentaje válido del ahorro de tiempo en la realización de consultas. Fuente: Elaboración propia.

**Figura 12**

*Gráfico de distribución de respuestas, ahorro tiempo.*



*Nota.* Gráfico de distribución del ahorro de tiempo en la realización de consultas. Fuente: Elaboración propia.

Los hallazgos detallados en la Tabla 18 indican que el 81.81% de los colaboradores mantiene una postura neutral frente al ahorro de tiempo percibido con el nuevo sistema de consulta. Esta valoración sugiere que, si bien el método

es funcional, la eficiencia temporal no fue percibida como un cambio disruptivo por la gran mayoría. En concordancia, la Figura 12 presenta una distribución de frecuencias con una clara concentración en los rangos intermedios (valores 2, 3 y 4), lo que ratifica estadísticamente la tendencia de estabilidad y neutralidad observada en la tabla.

### 3.1.4 Prueba contrastación de hipótesis específica 3.

El uso del lenguaje natural influye significativamente en la simplificación de las consultas SQL por parte del personal de la empresa.

**a. Hipótesis Nula (H0):** El uso del lenguaje natural NO influye significativamente en la simplificación de las consultas SQL por parte del personal de la empresa.

**b. Hipótesis Alterna (H1):** El uso del lenguaje natural SI influye significativamente en la simplificación de las consultas SQL por parte del personal de la empresa.

Para determinar la relación entre el perfil del usuario, considerando la frecuencia de acceso y el nivel de comodidad en el uso de herramientas digitales, y la percepción del uso del lenguaje natural, evaluada a través de la facilidad, utilidad, grado de facilitación y ahorro de tiempo, se utilizó el coeficiente de correlación de Spearman ( $\rho$ ). La elección de esta prueba no paramétrica responde a la naturaleza ordinal de las escalas de medición empleadas y a la ausencia de una distribución normal en los datos recolectados. El resultado obtenido fue:

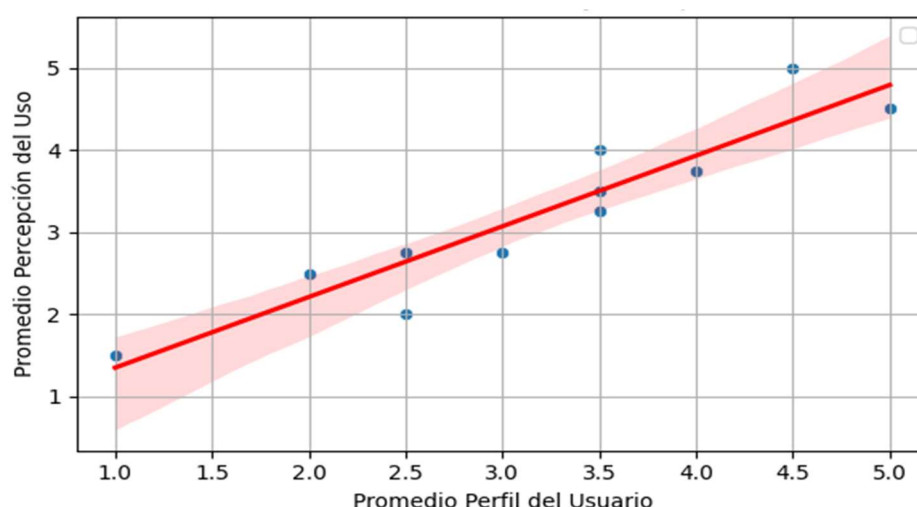
```
rho, pval = spearmanr(df['PerfilUsuario'], df['PercepcionUso'])  
Correlación  $\rho = 0.942$ , Significancia  $p = 0.0000$ 
```

Los resultados obtenidos evidenciaron la existencia de una correlación positiva muy fuerte, con un coeficiente de Spearman de  $\rho = 0.942$  y un nivel de

significancia de  $p = 0.0000 < 0.05$ , lo cual indica que a mayor necesidad y uso de los datos, más favorable resulta la percepción del método de obtención de información basado en lenguaje natural para la generación de consultas SQL. En consecuencia, se rechazó la hipótesis nula y se aceptó la hipótesis alterna, confirmando que el uso del lenguaje natural influye de manera significativa en la simplificación de las consultas SQL realizadas por el personal de la empresa.

**Figura 13**

*Relación entre perfil de usuario y percepción del uso*



*Nota.* Distribución de las respuestas correspondientes a la variable frecuencia de acceso. Fuente: Elaboración propia.

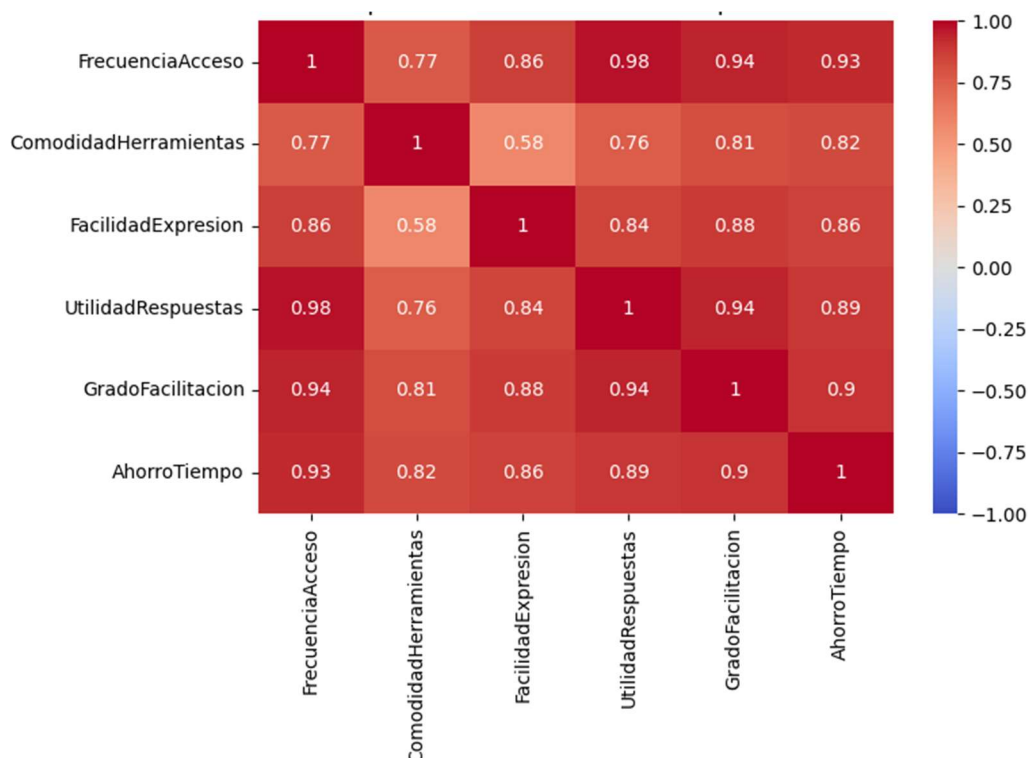
Asimismo, la Figura 13 correspondiente al gráfico de dispersión con línea de regresión evidenció una tendencia ascendente, lo que indica que, conforme aumenta la variable perfil del usuario, también se incrementa la percepción del uso del lenguaje natural, reflejando una relación directa y muy fuerte entre ambas variables.

El mapa de calor de correlaciones expuesto en la Figura 14, revela asociaciones positivas y consistentes entre cada componente del perfil del usuario y la percepción del uso del asistente. Estos hallazgos corroboraron que la formulación del instrumento estuvo adecuadamente orientada, tal como se

demuestra en el Anexo 2 con la evaluación de la confiabilidad a través de la prueba alfa de Cronbach aplicada al instrumento.

**Figura 14**

*Mapa de calor: correlaciones de Spearman.*



*Nota.* Distribución de las respuestas de la variable frecuencia de acceso.  
Fuente: Elaboración propia.

La evidencia recolectada valida la hipótesis de que la integración del lenguaje natural impacta positivamente en la simplificación del acceso a bases de datos para el personal de la organización. De igual manera, los hallazgos confirman que la experiencia de usuario se ve optimizada al alinearse con el grado de competencia digital y los hábitos tecnológicos de cada colaborador.

### 3.2 Discusión de resultados

Al contrastar los resultados alcanzados con los del estudio desarrollado por Leon et al. (2024), Se observan convergencias y contrastes significativos en la arquitectura y el rendimiento de las soluciones Text-to-SQL fundamentadas en modelos de lenguaje. Mientras que las metodologías de implementación



presentan similitudes en el uso de marcos de trabajo y conectores de datos, la capacidad operativa varía drásticamente en función de la optimización de cada modelo.

Leon et al. (2024) En su estudio referido planteó un enfoque basado en un mecanismo de prefiltrado de tablas mediante el uso de *embeddings* vectoriales, lo cual permitió evidenciar una mejora en la precisión de los modelos SQLCoder2 y SQLCoder-7b-2 de hasta un 7.3%. De este modo los autores buscaron reducir el contexto irrelevante y orientar a los modelos hacia las tablas más pertinentes para cada consulta. En el caso de la empresa Consditec, no se implementó un prefiltrado vectorial; en su lugar, se optó por un ajuste fino del *prompt* con restricciones explícitas, considerando que la base de datos no presenta una estructura compleja y que se proporcionó el [DDL] de la tabla. Esta estrategia permitió que el modelo de lenguaje dispusiera permanentemente de la estructura correcta, favoreciendo la generación de respuestas y produciendo mejoras significativas en la precisión obtenida.

Estos resultados sugieren que, si bien el prefiltrado constituye una estrategia útil, el diseño del *prompt* y su adecuación al motor SQL específico; en este caso SQLite, también cumplen un rol determinante en el desempeño del asistente. En ambos estudios se evidencian comportamientos similares respecto al rendimiento de modelos de lenguaje abiertos, de los cuales aunque resultan prometedores, aún no pueden alcanzar los niveles de precisión requeridos para su adopción en entornos productivos estables sin la aplicación de ajustes adicionales o sin la intervención humana, como sería la elaboración de *prompts* con conocimiento previo de la estructura de la base de datos. En esta línea, Leon et al. (2024), señala que modelos optimizados como SQLCoder-7b-2 alcanzan



tasas de precisión de apenas el 55 % al 60 % en entornos de bases de datos con alta complejidad estructural. Estas métricas contrastan significativamente con el desempeño del 90 % reportado originalmente por los desarrolladores, evidenciando una brecha considerable entre los resultados controlados y los escenarios de aplicación real. De manera similar las pruebas realizadas en la empresa Consditec mostraron que el modelo llama3.2:3b-instruct-q4\_K\_M alcanzó únicamente un 70% de precisión incluso en bases de datos de estructura sencilla, lo que evidencia que aún existe un amplio margen de desarrollo para lograr niveles de precisión superiores al 90%, sin embargo las limitantes de hardware de la maquina host no permitió evaluar modelos más potentes como si se hicieron en otros estudios.

## CONCLUSIONES

**Primero.** Se evidencia la viabilidad técnica, de la implementación de un asistente de consultas empleando lenguaje natural y modelos de lenguaje natural mediante LangChain para la generación de consultas SQL a partir de BBDD existentes, La investigación confirmó que el uso técnicas de prompt engineering y filtros previos, permite generar soluciones prácticas y aplicables a problemas empresariales sin depender de una API externa ni requerir el uso de grandes recursos computacionales ejecutados de manera local.

**Segundo.** La evaluación comparativa de los modelos de LLM disponibles resulta un factor relevante al igual que su integración al proyecto. En este sentido, el modelo llama3.2:3b-instruct-q4\_K\_M entrenado con fines multipropósito y del tipo "instruct" evidencio un mejor desempeño que duckdb-nsql:7b-q4\_K\_M o sqlcoder:7b-q4\_K\_M que son entrenados con el propósito de realizar tareas con BBDD, demostrando que la selección del modelo correcto ofrece un balance superior antes que el uso de un modelo específico.

**Tercero.** El análisis estadístico para el desempeño del asistente de los usuarios, confirmó una relación positiva muy fuerte, obteniéndose un coeficiente de correlación de Spearman de 0.942 ( $p = 0.000$ ) entre la su necesidad de uso de datos y el perfil del usuario, así como su percepción del asistente. Este resultado muestra una relación positiva fuerte y cuya interpretación es que, a mayor necesidad de acceder y utilizar datos de manera simple, mayor es la valoración del uso del lenguaje natural para la generación de consultas.

## RECOMENDACIONES

**Primero.** Se recomienda realizar el proceso de *fine-tuning* al modelo de lenguaje con mayor desempeño, empleando un corpus personalizado de consultas en lenguaje natural vinculadas a la base de datos de la empresa, junto con sus respectivas sentencias SQL de referencia (*gold standard*), además también se recomienda incluir la estructura de la base de datos, puesto que su estructura no cambia con respecto al tiempo. Esta estrategia podría fortalecer la capacidad del modelo para generar consultas SQL más precisas y coherentes ajustadas a las estructuras y patrones propios del caso de estudio, favoreciendo así un desempeño contextualizado.

**Segundo.** En este trabajo se priorizo la salvaguarda de la privacidad de los datos de la empresa, sin embargo, se recomienda evaluar el uso de una API comercial, pero sin proporcionar acceso a los datos, mediante el proceso de aislamiento al enviar solo estructura de las tablas en el prompt para la conexión con un servicio externo sin vulnerar privacidad. Puesto que durante las pruebas se observó que muchos LLM solo necesitan la estructura general de las tablas para generar una consulta SQL y no implica el acceso a los datos, de los resultados dependería evaluar el costo de tokens por consulta.



## REFERENCIAS BIBLIOGRÁFICAS

- Abanto, H. G. (2021). Modelo de computación cognitiva basado en TOGAF 9.2 y Big Data Agile para el análisis de indicadores de gestión aplicando IBM Watson en la empresa Atento-Trujillo [Universidad Privada Antenor Orrego]. oai:repositorio.upao.edu.pe:20.500.12759/8169
- Aguirre, I., Argomedeo, G. Y., Monzon, J. A., & Tuesta, C. A. (2022). Impacto de la adopción de inteligencia artificial como estrategia de negocio en las empresas del sector servicios durante la época de pandemia en el Perú. <https://tesis.pucp.edu.pe/repositorio//handle/20.500.12404/21241>
- Alvear, V. (2025). Aplicación impulsada por modelos de lenguaje de gran tamaño para el diseño de consultas SQL mediante ingeniería de prompt. <https://repositorio.usm.cl/handle/123456789/76882>
- Artasanchez, A. (with Joshi, P.). (2020). Artificial intelligence with Python: Your complete guide to building intelligent apps using Python 3.x and TensorFlow 2 (Second edition.). Packt Publishing.
- Auffarth, B. (2023). Generative AI with LangChain: Build large language model (LLM) apps with Python, ChatGPT, and other LLMs. Packt Publishing Ltd.
- Bhushan, S., Thomas, D., Borchers, C., & Raghuvanshi, I. (2025, junio 25). Detecting LLM-Generated Short Answers and Effects on Learner Performance | Request PDF. ResearchGate. <https://doi.org/10.48550/arXiv.2506.17196>
- Brito, B. (2024). Sistema de Generación de Consultas SQL basado en Peticiones de Lenguaje Natural a partir de un Chat-Bot. <https://titula.universidadeuropea.com/handle/20.500.12880/8877>
- Caballero, A. (2010). Metodología de la investigación científica: Diseños con hipótesis explicativas. Udegraf.
- Camayteri, A., Condori, O., Duran, A., Esparza, F., Farias, J., Torres, R., & Vasquez, M. (2024). Métodos de inteligencia artificial (AI) en SQL Server



para la eficiencia del sistema de auto calificación: Una revisión sistemática literaria. Universidad César Vallejo.

<https://repositorioacademico.ucv.edu.pe/handle/10787/684791>

Cargua, R., & Francisco, J. (2025). Uso los modelos de lenguaje grandes (LLM) como asistentes virtuales de salud en el Ecuador. ResearchGate.  
[https://www.researchgate.net/publication/392455172\\_Uso\\_los\\_modelos\\_de\\_lenguaje\\_grandes\\_LLM\\_como\\_asistentes\\_virtuales\\_de\\_salud\\_en\\_el\\_Ecuador](https://www.researchgate.net/publication/392455172_Uso_los_modelos_de_lenguaje_grandes_LLM_como_asistentes_virtuales_de_salud_en_el_Ecuador)

Castro, M., & Muñoz, F. (2023). Optimizing Prompt Engineering for Improved Generative AI Content [Universidad Pontificia Comillas].  
<https://repositorio.comillas.edu/xmlui/handle/11531/80629>

Dieterich, H. (2010). Nueva guía para la investigación científica. Editorial Planeta Mexicana.

Hernandez, R. (2018). Metodología de la Investigación: Las rutas Cuantitativa, Cualitativa y Mixta. McGraw-Hill Education.

INEI. (2022). Compendio Estadístico. Perú 2022.  
[https://www.inei.gob.pe/media/MenuRecursivo/publicaciones\\_digitales/Est/Lib1872/COMPENDIO2022.html](https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1872/COMPENDIO2022.html)

Jacinto, E. (2022). Análisis de los sistemas de gestión de base de datos relacionales con marcos de trabajo para procesamiento de datos masivos. Repositorio Institucional - USS.  
<https://repositorio.uss.edu.pe/handle/20.500.12802/10234>

Leon, G., Ljubicic, M., & Oldán, J. (2024a). Un Estudio de Text-to-SQL con Grandes [Universidad ORT Uruguay].  
<https://hdl.handle.net/20.500.11968/7029>

Leon, G., Ljubicic, M., & Oldán, J. (2024b). Un Estudio de Text-to-SQL con Grandes [Universidad ORT Uruguay].  
<https://hdl.handle.net/20.500.11968/7029>



- Min, B., Ross, H., Sulem, E., Veyseh, A. P. B., Nguyen, T. H., Sainz, O., Agirre, E., Heintz, I., & Roth, D. (2024). Recent Advances in Natural Language Processing via Large Pre-trained Language Models: A Survey. *ACM Computing Surveys*, 56(2), 1-40. <https://doi.org/10.1145/3605943>
- Mollocondo, A. (2025). Implementación del Modelo de Lenguaje Largo (LLM) en el Módulo de Ventas del ERP Odoo en Medianas Empresas del Perú. Repositorio Institucional - UTP. <http://repositorio.utp.edu.pe/handle/20.500.12867/13015>
- Negueruela, I. (2024). Análisis y evaluación de LLMs con técnicas RAG para el desarrollo y despliegue de un asistente virtual. <https://repositorio.unican.es/xmlui/handle/10902/33837>
- Olszak, C. M. (2020). *Business Intelligence and Big Data: Drivers of Organizational Success*. CRC Press.
- Pérez, D. (2024). Modelos de lenguaje grande (LLM) y contratación pública: Una propuesta emergente para la elaboración de condiciones técnicas. <https://hdl.handle.net/20.500.14625/36991>
- Salazar, O. F. (2020). Chatbot para el aprendizaje de SQL básico [Universidad César Vallejo]. <https://hdl.handle.net/20.500.12692/66340>
- Uma, M., Sneha, V., Sneha, G., Bhuvana, J., & Bharathi, B. (2019). Formation of SQL from Natural Language Query using NLP. 1-5. <https://doi.org/10.1109/ICCIDS.2019.8862080>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. ukasz, & Polosukhin, I. (2017). Attention is All you Need. *Advances in Neural Information Processing Systems*, 30. [https://papers.nips.cc/paper\\_files/paper/2017/hash/3f5ee243547dee91fb-d053c1c4a845aa-Abstract.html](https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fb-d053c1c4a845aa-Abstract.html)
- Zhou, X., Sun, Z., & Li, G. (2024). DB-GPT: Large Language Model Meets Database. *Data Science and Engineering*, 9(1), 102-111. <https://doi.org/10.1007/s41019-023-00235-6>



## ANEXOS



Anexo 1: Matriz de consistencia

EMPLEO DEL LENGUAJE NATURAL PARA OPTIMIZAR LAS CONSULTAS SQL CON LANGCHAIN PARA LA EMPRESA CONSDITEC JULIACA 2024

| PROBLEMA                                                                                                                                                                            | OBJETIVOS                                                                                                                                                               | HIPOTESIS                                                                                                                                                                   | VARIABLES                                            | DIMENSIONES                   | INDICADORES                                                                            | METODOLOGÍA                                                                                                                                                                                             |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------|-------------------------------|----------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Problema general</b><br>¿Cómo el empleo de procesamiento de lenguaje natural y LangChain puede optimizar y simplificar la generación de consultas SQL para la empresa Consditec? | <b>Objetivo general</b><br>Aplicar el procesamiento de lenguaje natural y LangChain para optimizar y simplificar la generación de consultas SQL en la empresa Considetc | <b>Hipótesis general</b><br>El procesamiento de lenguaje natural y LangChain optimiza y simplifica la generación de consultas SQL en la empresa Consditec.2                 | Uso del lenguaje natural (Variable Independiente)    | *Interfaz de lenguaje natural | *Nivel de facilidad para escribir consultas en lenguaje natural                        | Método:<br>Cuantitativo.<br>Tipo de investigación:<br>Aplicada.<br>Diseño de investigación:<br>Transeccional, descriptivo-explicativo.<br>Técnicas:<br>Observación Estructurada, Encuesta, Cuestionario |
| <b>Problemas específicos:</b><br>a. ¿Es posible conectar la base de datos de la empresa Consditec mediante LangChain para realizar consultas SQL en lenguaje natural?               | <b>Objetivos específicos:</b><br>a. Conectar la base de datos de la empresa Consditec mediante LangChain para realizar consultas SQL en lenguaje natural.               | <b>Hipótesis específicas:</b><br>a. Es factible La conexión de la base de datos de la empresa Consditec mediante LangChain para realizar consultas SQL en lenguaje natural. |                                                      | *Experiencia de usuario       | *Nivel de utilidad de respuestas generadas                                             |                                                                                                                                                                                                         |
|                                                                                                                                                                                     |                                                                                                                                                                         |                                                                                                                                                                             |                                                      | *Utilidad percibida           | *Grado de precisión del sistema sin usar SQL directamente                              |                                                                                                                                                                                                         |
| b. ¿Cuál modelo de lenguaje natural tiene mejor desempeño para realizar consultas SQL para la empresa Consditec?                                                                    | b. Evaluar los modelos de lenguaje natural óptimos para realizar consultas SQL para la empresa Consditec                                                                | b. La evaluación de los modelos de lenguaje natural determina modelos óptimos para realizar consultas SQL para la empresa Consditec.                                        | Optimización de consultas SQL (Variable Dependiente) | *Ahorro de tiempo             | *Percepción del tiempo ahorrado en consultas                                           |                                                                                                                                                                                                         |
| c. ¿Cómo el uso del lenguaje natural influye en la simplificación de las consultas SQL por parte del personal de la empresa?                                                        | c. Describir como el uso del lenguaje natural influye en la simplificación de las consultas SQL por parte del personal de la empresa.                                   | c. El uso del lenguaje natural influye significativamente en la simplificación de las consultas SQL por parte del personal de la empresa.                                   |                                                      | *Tiempo de respuesta          | *Milisegundos que tarda el modelo en generar la consulta                               |                                                                                                                                                                                                         |
|                                                                                                                                                                                     |                                                                                                                                                                         |                                                                                                                                                                             |                                                      | *Precisión sintáctica         | *Nº de consultas con sintaxis SQL válidas                                              |                                                                                                                                                                                                         |
|                                                                                                                                                                                     |                                                                                                                                                                         |                                                                                                                                                                             |                                                      | *Exactitud del resultado      | *Nº de respuestas correctas frente a respuestas esperadas cotejadas con dataset Spider |                                                                                                                                                                                                         |



### Anexo 2: Instrumentos

La presente encuesta tiene el objetivo de evaluar la influencia del uso del lenguaje natural y la simplificación en consultas SQL por parte del personal de la empresa Consditec. la encuesta es anónima, se agradece su colaboración y sinceridad al momento de responder las preguntas.

Responda con un aspa (x) o cruz (+) la alternativa que manifiesta mejor su opinión.

#### Bloque A: Perfil del usuario (necesidad y uso de datos)

1. ¿Con qué frecuencia necesitas obtener información específica desde el sistema de datos para realizar tu trabajo?

- a) Nunca
- b) Raramente
- c) Ocasionalmente
- d) Frecuentemente
- e) Muy frecuentemente

2. ¿Qué tan cómodo/a te sientes utilizando herramientas digitales que te permiten consultar información (como filtros, menús o buscadores)?

- a) Muy incómodo/a
- b) Incómodo/a
- c) Neutral
- d) Cómodo/a
- e) Muy cómodo/a

#### Bloque B: Percepción del uso del lenguaje natural para consultas

3. ¿Qué tan fácil te resultó expresar lo que necesitabas mediante lenguaje natural (escribiendo como lo dirías a otra persona)?

- a) Muy difícil
- b) Difícil
- c) Neutral
- d) Fácil
- e) Muy fácil

4. ¿Qué tan útil fue poder escribir libremente lo que buscabas, en lugar de usar filtros o códigos como antes?

- a) Nada útiles
- b) Poco útiles
- c) Medianamente útiles
- d) Útiles
- e) Muy útiles

5. ¿En qué medida el nuevo asistente te ayudó a obtener los datos correctos sin necesidad de saber el código o ubicación exacta?

- a) No ayudó nada
- b) Ayudó muy poco
- c) Ayudó algo
- d) Ayudó bastante
- e) Ayudó mucho

6. ¿Qué impacto tuvo el uso del asistente en el tiempo que dedicaste a obtener la información que necesitabas?

- a) Aumento mucho el tiempo
- b) Aumento el tiempo
- c) El tiempo fue el mismo
- d) Redujo el tiempo
- e) Redujo mucho el tiempo



### Anexo 3: Validación de instrumentos

UNIVERSIDAD ANDINA NÉSTOR CÁCERES VELÁSQUEZ  
FACULTAD DE INGENIERÍA DE SISTEMAS  
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS



#### FICHA DE VALIDACIÓN DE INSTRUMENTO DE INVESTIGACIÓN

##### JUICIO DE EXPERTOS

#### VII. REFERENCIAS

- f. Experto/Nombres : Yban Vilchez Navarro  
g. Especialidad : Desarrollo de Software  
h. Cargo Actual : Docente Post Grado  
i. Grado Académico : Doctor  
j. Fecha : 17 mayo 2025

#### VIII. TEST DE LIKERT DE: EMPLEO DEL LENGUAJE NATURAL PARA OPTIMIZAR LAS CONSULTAS SQL CON LANGCHAIN EN LA EMPRESA CONSDITEC JULIACA 2024

#### IX. AUTOR DEL INSTRUMENTO DE INVESTIGACIÓN:

Bach. CESAR AUGUSTO APAZA CACERES

#### X. ASPECTOS DE VALIDACIÓN

(1 = Deficiente, 2 = Regular, 3 = Buena, 4 = Muy buena, 5 = Excelente)

| INDICADORES        | CRITERIOS                                                    | 1 | 2 | 3 | 4 | 5 |
|--------------------|--------------------------------------------------------------|---|---|---|---|---|
| a. Claridad        | Está redactado con lenguaje apropiado                        |   |   | X |   |   |
| b. Objetividad     | Está expresado en capacidades observables                    |   |   | X |   |   |
| c. Actualidad      | Está adecuado al avance de la ciencia                        |   |   |   | X |   |
| d. Organización    | Existe una organización lógica de los ítems y variables      |   |   |   |   | X |
| e. Suficiencia     | Valora las dimensiones en cantidad y calidad suficientes     |   |   |   | X |   |
| f. Intencionalidad | Está adecuada para cumplir los objetivos de la investigación |   |   |   |   | X |
| g. Consistencia    | Está basada en aspectos teóricos científicos                 |   |   |   | X |   |
| h. Coherencia      | Muestra coherencia entre dimensiones, indicadores e ítems    |   |   |   |   | X |
| i. Metodología     | Responde al propósito de la investigación                    |   |   | X |   |   |
| j. Pertinencia     | Es útil y adecuado para la investigación                     |   |   |   | X |   |

Coefficiente de valoración porcentual.  $C = \text{Total}/50$

#### XI. OBSERVACIONES Y RECOMENDACIONES

#### XII. RESOLUCION DEL EXPERTO

- Aprobado ( $C > 75\% = 0.75$ ) ☒  
Desaprobado ( $C < 75\% = 0.75$ ) ☐

Apellidos y Nombres YBAN VILCHEZ NAVARRO



UNIVERSIDAD ANDINA NÉSTOR CÁCERES VELÁSQUEZ  
FACULTAD DE INGENIERÍA DE SISTEMAS  
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS



### FICHA DE VALIDACIÓN DE INSTRUMENTO DE INVESTIGACIÓN

#### JUICIO DE EXPERTOS

##### I. REFERENCIAS

- a. Experto/Nombres : Alfredo Carlos Castro Quispe  
b. Especialidad : Investigación Científica  
c. Cargo Actual : Docente Post Grado  
d. Grado Académico : Magister  
e. Fecha : 23 mayo 2025

##### II. TEST DE LIKERT DE: EMPLEO DEL LENGUAJE NATURAL PARA OPTIMIZAR LAS CONSULTAS SQL CON LANGCHAIN EN LA EMPRESA CONSDITEC JULIACA 2024

##### III. AUTOR DEL INSTRUMENTO DE INVESTIGACIÓN:

Bach. CESAR AUGUSTO APAZA CACERES

##### IV. ASPECTOS DE VALIDACIÓN

(1 = Deficiente, 2 = Regular, 3 = Buena, 4 = Muy buena, 5 = Excelente)

| INDICADORES        | CRITERIOS                                                    | 1 | 2 | 3 | 4 | 5 |
|--------------------|--------------------------------------------------------------|---|---|---|---|---|
| a. Claridad        | Está redactado con lenguaje apropiado                        |   |   |   | X |   |
| b. Objetividad     | Está expresado en capacidades observables                    |   |   | X |   |   |
| c. Actualidad      | Está adecuado al avance de la ciencia                        |   |   |   | X |   |
| d. Organización    | Existe una organización lógica de los ítems y variables      |   |   |   | X |   |
| e. Suficiencia     | Valora las dimensiones en cantidad y calidad suficientes     |   |   |   | X |   |
| f. Intencionalidad | Está adecuada para cumplir los objetivos de la investigación |   |   |   |   | X |
| g. Consistencia    | Está basada en aspectos teóricos científicos                 |   |   | X |   |   |
| h. Coherencia      | Muestra coherencia entre dimensiones, indicadores e ítems    |   |   | X |   |   |
| i. Metodología     | Responde al propósito de la investigación                    |   |   |   | X |   |
| j. Pertinencia     | Es útil y adecuado para la investigación                     |   |   |   | X |   |

Coefficiente de valoración porcentual.  $C = \text{Total}/50$

##### V. OBSERVACIONES Y RECOMENDACIONES

##### VI. RESOLUCION DEL EXPERTO

- Aprobado ( $C > 75\% = 0.75$ ) ☒  
Desaprobado ( $C < 75\% = 0.75$ ) ☐

  
Dr. Alfredo Carlos Castro Quispe  
M.Sc. EN GESTIÓN EMPRESARIAL

#### Anexo 4: Confiabilidad del instrumento

El alfa de Cronbach ( $\alpha$ ) es una medida estadística que evalúa la confiabilidad interna de un instrumento, el nivel de coherencia de las preguntas entre sí cuando miden una misma dimensión (Hernandez, 2018). Mediante la siguiente escala

|                         |                                                                |
|-------------------------|----------------------------------------------------------------|
| $\alpha \geq 0.9$       | Excelente (instrumento debería revisarse, posible redundancia) |
| $0.8 \leq \alpha < 0.9$ | Buena                                                          |
| $0.7 \leq \alpha < 0.8$ | Aceptable                                                      |
| $0.6 \leq \alpha < 0.7$ | Cuestionable                                                   |
| $\alpha < 0.6$          | Débil (el instrumento debería reformularse) (Hernandez, 2018)  |

Para cada bloque, se puede calcular el alfa de Cronbach para verificar si esas preguntas están midiendo de forma consistente un mismo concepto.

1. Perfil del usuario:
  - Frecuencia de acceso
  - Comodidad con herramientas
2. Percepción del uso del lenguaje natural:
  - Facilidad de expresión
  - Utilidad de respuestas
  - Grado de facilitación
  - Ahorro de tiempo

Los resultados obtenidos, procesados con *numpy*, por dimensión, fueron:

Alfa de Cronbach - PerfilUsuario: 0.893

Alfa de Cronbach - PercepcionUso: 0.861

De acuerdo a los resultados obtenidos y comparados con la tabla, se considera que el instrumento es confiable para fines de investigación cuantitativa.



ANEXO 1  
FORMULARIO DE AUTORIZACIÓN

AUTORIZACIÓN PARA LA INCORPORACIÓN DE LOS  
TRABAJOS DE INVESTIGACIÓN  
EN EL REPOSITORIO INSTITUCIONAL UANCV

Formato digital ☒

Fecha de entrega: 29-04-2026

1. Datos del autor (es):

Nombres y Apellidos: CESAR AUGUSTO APAZA CACERES

Dirección: Pje. Machallata 110 - Puno

DNI/Carné de Extranjería/Pasaporte N°: 43190790

Teléfono: 900986943 email: saicesaraugusto@gmail.com

Nombres y Apellidos:

Dirección:

DNI/Carné de Extranjería/Pasaporte N°:

Teléfono: email:

Facultad y/o Escuela de Posgrado: INGENIERÍA DE SISTEMAS

Escuela Profesional o Mención: INGENIERÍA DE SISTEMAS

Título o Grado Académico a optar: INGENIERO DE SISTEMAS

Asesor: Dr. PAUL MAMANI TISNADO

Esta obra se encuentra dentro de las siguientes denominaciones:

Trabajo de Investigación ☐ Tesis ☒ Trabajo de Suficiencia Profesional ☐ Trabajo Académico ☐

Título: EMPLEO DEL LENGUAJE NATURAL PARA OPTIMIZAR LAS CONSULTAS SQL

CON LANGCHAIN EN LA EMPRESA CONSDITEC JULIACA 2024

Palabras claves, (3 a 5 términos): Langchain, Large Language Models, Text-to-SQL

¿Esta obra se desarrolló en la UANCV <sup>1, 2</sup>?

2

<sup>1</sup> Indicar si su producción intelectual ha empleado recursos tales como, instalaciones, laboratorios, insumos, equipos, bases de datos, asesoría técnica por parte del personal de la UANCV, financiamiento, entre otros relacionados.

<sup>2</sup> Si su producción intelectual se desarrolló en la UANCV totalmente o parcialmente, deberá autorizar el depósito en el Repositorio de manera obligatoria.



2. Referencia de tesis:

☐ Bachiller ☒ Título ☐ 2da Especialidad ☐ Maestría ☐ Doctorado

3. Licencias:

a) Licencia estándar:

**Bajo los siguientes términos, autorizo el depósito de mi tesis en el Repositorio Digital de la UANCV.**

Con la autorización de depósito de mi producción intelectual, otorgo a la Universidad Andina "Néstor Cáceres Velásquez" una licencia no exclusiva para reproducir, distribuir, comunicar al público, transformar (únicamente mediante su traducción a otros idiomas) y poner a disposición del público mi producción intelectual (incluido el resumen), en formato físico o digital, en cualquier medio, conocido o por conocerse, a través de los diversos servicios por la Universidad, creados o por crearse, tales como el Repositorio Digital de tesis UANCV, colección de producción intelectual, entre otros, en el Perú y en el extranjero por el tiempo y veces que considere necesarias, y libres de remuneraciones.

En virtud de dicha licencia, la Universidad Andina "Néstor Cáceres Velásquez" podrá reproducir mi producción intelectual en cualquier tipo de soporte y en más de un ejemplar, sin modificar su contenido, solo con propósitos de seguridad, respaldo y preservación.

Declaro que la producción intelectual es una creación de mi autoría y exclusiva titularidad, coautoría con titularidad compartida, y me encuentro facultado a conceder la presente licencia y, asimismo, garantizo que dicha producción intelectual no infringe derechos de autor de terceras personas.

La Universidad Andina "Néstor Cáceres Velásquez" consignará el nombre del y/o los autor(es) de la producción intelectual, y no le hará ninguna modificación más que la permitida en la licencia.

**Autorizo su publicación (marque con una X)**

- ☒ Sí, autorizo que se deposite inmediatamente.
- ☐ Sí, autorizo que se deposite a partir de la fecha (d/m/a): \_\_\_\_\_
- ☐ No autorizo.

b) Licencia CREATIVE COMMONS 4.0 INTERNACIONAL:

Si usted concede una licencia CREATIVE COMMONS sobre su producción intelectual, mantiene la titularidad de los derechos de autor de esta y, a la vez, permite que otras personas puedan reproducirla, comunicarla al público y distribuir ejemplares de esta, bajo las condiciones siguientes:

**¿Quiere permitir usos comerciales de su producción intelectual?**

**Sí:** significa que usted permite la reproducción, distribución y comunicación pública de la producción intelectual incluso con fines comerciales.

**No:** significa que usted permite la reproducción, y comunicación pública de la producción intelectual, pero sin fines comerciales.

- ☐ Sí autorizo
- ☒ No autorizo



### Jurisdicción de su Licencia

Todas las licencias CREATIVE COMMONS son de ámbito mundial, sin embargo, usted puede elegir entre la opción “internacional” o una adaptada a su jurisdicción, como para el caso peruano.

La opción “internacional” emplea el lenguaje y la terminología de los tratados internacionales; en cambio, la adaptada a su jurisdicción, recoge las particularidades de la legislación peruana.

En consecuencia, **la opción “internacional” goza de una mayor eficacia a nivel mundial, gracias a que tiene jurisdicción neutral.** Mientras que la opción adaptada a la jurisdicción del Perú goza de una mayor eficacia ante los tribunales peruanos.

- ☐ Internacional  
☒ Nacional

Línea de investigación: CIENCIA DE LOS ORDENADORES - P24

Firma de Autor

huella digital

29 - ABRIL - 2026

Fecha