



EDUCACIÓN

SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO
NACIONAL DE MÉXICO

Tecnológico Nacional de México

**Centro Nacional de Investigación
y Desarrollo Tecnológico**

Tesis de Maestría

**Desarrollo y evaluación de un Asistente Educativo
Holográfico basado en un
Modelo de Lenguaje Grande**

presentada por

Ing. Orvil Alexis Villarreal Molina

como requisito para la obtención del grado de
Maestro en Ciencias de la Computación

Director de tesis

Dr. Nimrod González Franco

Codirector de tesis

Dr. Noé Alejandro Castro Sánchez

Cuernavaca, Morelos, México. Mayo de 2025.



Cuernavaca, Mor., 28/Mayo/2025

OFICIO No. DCC/147/2025

Asunto: Aceptación de documento de tesis

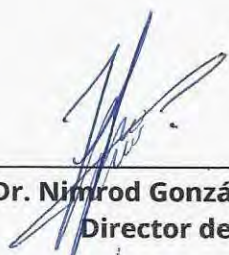
CENIDET-AC-004-M14-OFICIO

CARLOS MANUEL ASTORGA ZARAGOZA
SUBDIRECTOR ACADÉMICO
PRESENTE

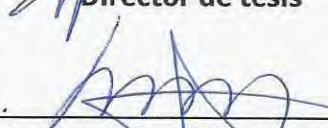
Por este conducto, los integrantes de Comité Tutorial de **ORVIL ALEXIS VILLARREAL MOLINA** con número de control **M23CE090**, de la Maestría en Ciencias de la Computación, le informamos que hemos revisado el trabajo de tesis de grado titulado **"Desarrollo y evaluación de un Asistente Educativo Holográfico basado en un Modelo de Lenguaje Grande"** y hemos encontrado que se han atendido todas las observaciones que se le indicaron, por lo que hemos acordado aceptar el documento de tesis y le solicitamos la autorización de impresión definitiva.

ATENTAMENTE

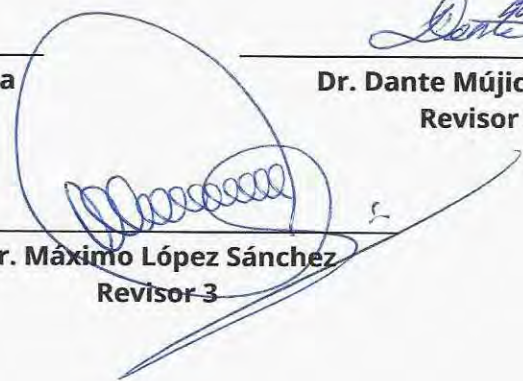
Excelencia en Educación Tecnológica®
"Conocimiento y Tecnología al Servicio de México"


Dr. Nimrod González Franco
Director de tesis


Dr. Noé Alejandro Castro Sánchez
Codirector de tesis


Dr. Juan Gabriel González Serna
Revisor 1


Dr. Dante Mújica Vargas
Revisor 2


Dr. Máximo López Sánchez
Revisor 3

C.c.p. Depto. Servicios Escolares.
Expediente / Estudiante



2025
Año de
La Mujer
Indígena

Interior Internado Palmira S/N, Col. Palmira,
C. P. 62490, Cuernavaca, Morelos Tel. 01 (777) 3627770, ext. 3201,
e-mail: dcc_cenidet@tecnm.mx tecnm.mx | cenidet.tecnm.mx





Centro Nacional de Investigación y Desarrollo tecnológico
Subdirección Académica

Cuernavaca Mor, 30/mayo/2025

Oficio No. SAC/117/2025

Asunto: Autorización de impresión de tesis

**ORVIL ALEXIS VILLARREAL MOLINA
CANDIDATO AL GRADO DE MAESTRO
EN CIENCIAS DE LA COMPUTACIÓN
P R E S E N T E**

Por este conducto, tengo el agrado de comunicarle que el Comité Tutorial asignado a su trabajo de tesis titulado **"Desarrollo y evaluación de un Asistente Educativo Holográfico basado en un Modelo de Lenguaje Grande"**, ha informado a esta Subdirección Académica, que están de acuerdo con el trabajo presentado. Por lo anterior, se le autoriza a que proceda con la impresión definitiva de su trabajo de tesis.

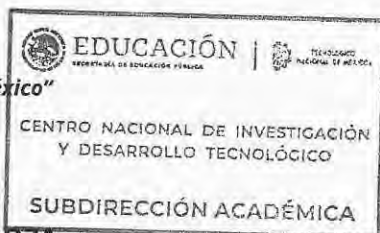
Esperando que el logro del mismo sea acorde con sus aspiraciones profesionales, reciba un cordial saludo.

ATENTAMENTE

Excelencia en Educación Tecnológica®

"Conocimiento y Tecnología al Servicio de México"

**CARLOS MANUEL ASTORGA ZARAGOZA
SUBDIRECTOR ACADÉMICO**



c.c.p. Departamento de Ciencias Computacionales
Departamento de Servicios Escolares

CMAZ/lmz



2025
Año de
La Mujer
Indígena

Interior Internado Palmira S/N, Col. Palmira,
C. P. 62490, Cuernavaca, Morelos Tel. 01 (777) 3627770, ext. 4104,
e-mail: acad_cenidet@tecnm.mx tecnm.mx | cenidet.tecnm.mx

cenidet
Centro Nacional de Investigación
y Desarrollo Tecnológico



*A mis padres, por ser mi guía constante, por su
amor incondicional, por darme la fuerza
necesaria para no rendirme y enseñarme con
el ejemplo el valor del esfuerzo y perseverancia.*

*A mis hermanos, por su calidez y por estar
siempre conmigo, compartiendo desafíos,
aprendizajes y logros personales.*

*A mi pareja, por su compañía, pero sobre
todo por su amor y paciencia incondicional
a lo largo de este trayecto.*

*A mis amigos, por brindarme su cercanía,
ánimo y comprensión, haciendo más ameno
y llevadero este proceso.*

*Y en especial, a mis abuelos, que aunque ya no
se encuentran entre nosotros, su memoria
vive en mí y me inspira en seguir adelante.
Gracias por su amor.*

Esta tesis es para ustedes.

Agradecimientos

Agradezco a la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI), por el respaldo económico brindado a través del programa de becas, el cual fue fundamental para la realización de este trabajo de investigación y para la continuidad de mi formación académica.

Al Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET) que me formo a lo largo de esta investigación, por brindarme las herramientas, el espacio y el respaldo necesario para crecer tanto profesional como personalmente.

A mis directores y mi comité tutorial por su guía, compromiso y paciencia durante todo el proceso de investigación. Sus conocimientos, sugerencias y apoyo fueron fundamentales para el desarrollo y culminación de esta investigación.

Y, por supuesto, a mi familia y amigos, quienes con su apoyo, comprensión y cariño me motivaron a seguir adelante, sobre todo en los momentos más desafiantes.

A cada uno de ustedes, mi más sincero agradecimiento.

Resumen

El presente trabajo tiene como objetivo evaluar cuantitativamente la calidad de un Asistente Educativo Holográfico (AEH) basado en un Modelo de Lenguaje Grande (MLG). Para ello, se propone una metodología que integra métricas de Procesamiento del Lenguaje Natural (PLN) y percepciones de Usabilidad obtenidas mediante cuestionarios aplicados a usuarios finales. La evaluación se realiza mediante un Modelo de Ecuaciones Estructurales (MEE), complementado con una regresión simbólica que permite derivar un indicador matemático de calidad.

Como parte de esta investigación, se desarrolló un prototipo funcional del AEH denominado NadIA 2.0, el cual permite la interacción mediante comandos de voz, procesamiento de lenguaje natural y respuestas audiovisuales. El sistema fue implementado utilizando una arquitectura cliente-servidor, integrando tecnologías derivadas de *OpenAI*, *Python*, *Java* y motores de síntesis de voz. Además, incorpora un módulo visual interactivo que representa distintas expresiones del asistente mediante secuencias animadas, mejorando la experiencia del usuario.

Los resultados muestran que el modelo alcanza un ajuste excelente ($CFI = 1.002$, $RMSEA = 0.000$), confirmando que las métricas explican adecuadamente la variabilidad en la percepción de los usuarios. La regresión simbólica permite encontrar una ecuación precisa y parsimoniosa que representa la calidad del AEH a partir de variables clave como la efectividad del asistente, la satisfacción del usuario, la calidad percibida de las respuestas y el uso posterior del sistema.

Este enfoque demuestra ser una alternativa viable para evaluar tecnologías educativas basadas en Inteligencia Artificial (IA), ofreciendo un marco metodológico cuantitativo, robusto y replicable. No obstante, se identifican limitaciones como el tamaño reducido de la muestra y la falta de inclusión de otras dimensiones de calidad del software.

Se sugiere como trabajo futuro ampliar la muestra a diferentes contextos educativos, incorporar subcaracterísticas adicionales definidas en la norma ISO/IEC 9126 y explorar nuevas técnicas de aprendizaje automático para implementar en el desarrollo de los AEH's. Se sugiere como trabajo futuro ampliar la muestra a diferentes contextos educativos, incorporar subcaracterísticas adicionales definidas en la norma ISO/IEC 9126 y explorar nuevas técnicas de aprendizaje automático para implementar en el desarrollo de los AEH's.

Abstract

This research aims to quantitatively evaluate the quality of a Holographic Educational Assistant (HEA) based on a Large Language Model (LLM). For this purpose, a methodology is proposed that integrates natural language processing (NLP) metrics and usability perceptions obtained through questionnaires applied to end users. The evaluation is performed by means of a Structural Equation Model (SEM), complemented with a symbolic regression that allows deriving a mathematical indicator of quality.

As part of this research, a functional prototype of the HEA, named NadIA 2.0, was developed. This system enables interaction through voice commands, natural language processing. The system was implemented using a client-server architecture, integrating technologies based on *OpenAI*, *Python*, *Java*, and speech synthesis engines. In addition, it incorporates an interactive visual module that displays different assistant expressions through animated sequences, enhancing the user experience. As part of this research, a functional prototype of the HEA, named NadIA 2.0, was developed. This system enables interaction through voice commands, natural language processing. The system was implemented using a client-server architecture, integrating technologies based on *OpenAI*, *Python*, *Java*, and speech synthesis engines. In addition, it incorporates an interactive visual module that displays different assistant expressions through animated sequences, enhancing the user experience.

The results show that the model achieves an excellent fit ($CFI = 1.002$, $RMSEA = 0.000$), confirming that the selected metrics adequately explain the variability in user perception. The symbolic regression generates a precise and parsimonious equation that models the quality of the HEA based on key variables such as assistant effectiveness, user satisfaction, perceived quality of responses and subsequent use of the system.

This approach proves to be a viable alternative to evaluate educational technologies based on Artificial Intelligence (AI), offering a quantitative, robust and replicable methodological framework. However, limitations such as the small sample size and the lack of inclusion of other software quality dimensions are identified.

It is suggested as future work to extend the sample to different educational contexts, incorporate additional sub-features defined in ISO/IEC 9126 standard and explore new machine learning techniques to implement in the development of HEA's.

Índice

Índice	VII
Índice de figuras	IX
Índice de tablas	X
Glosario	XI
1 Introducción	1
1.1 Planteamiento del problema	2
1.2 Hipótesis	2
1.2.1 Delimitación del problema	2
1.2.2 Complejidad del problema	2
1.3 Objetivos	3
1.3.1 Objetivo general	3
1.3.2 Objetivos específicos	3
1.4 Alcances y limitaciones	3
1.4.1 Alcances	3
1.4.2 Limitaciones	3
1.5 Justificación	4
1.6 Estructura de la tesis	5
2 Marco teórico	6
2.1 Asistente Educativo Holográfico	6
2.2 Holografía	7
2.3 Métricas para evaluar la calidad de texto	8
2.3.1 Coherencia	8
2.3.2 Cohesión	8
2.3.3 Gramaticalidad	8
2.3.4 Complejidad del léxico	9
2.3.5 Lecturabilidad de Fernández-Huerta	9
2.3.6 Legibilidad Inflesz	9
3 Estado del arte	10
3.1 Discusión del Estado del Arte	16
4 Metodología	19
4.1 Descripción de la propuesta de solución	19
4.1.1 Comprensión de la problemática	19
4.1.2 Exploración y evaluación de Modelos de Lenguaje Grandes	20
4.1.3 Formulación de heurísticas de calidad para AEH's	21
4.1.4 Diseño de un AEH	22
4.1.5 Desarrollo de un AEH	23
4.1.6 Pruebas funcionales de un AEH	26
4.1.7 Evaluación de aspectos asociados a la funcionalidad del AEH	26
4.1.8 Evaluación de aspectos de usabilidad percibidos por los usuarios de un AEH	27
4.1.9 Modelado de la calidad de un AEH	28
4.1.10 Formulación de un indicador de calidad para un AEH	29

5	Experimentación y resultados	31
5.1	Experimentación Corpus vs MLG	31
5.2	Diseño experimental	36
5.2.1	Metas, hipótesis y variables	36
5.2.2	Diseño	36
5.2.3	Sujetos	37
5.2.4	Objetos	37
5.2.5	Instrumentos	37
5.2.6	Proceso de adquisición de datos	37
5.2.7	Proceso de análisis de datos	38
5.2.8	Proceso de validación	38
5.3	Ejecución del experimento	38
5.3.1	Muestreo	38
5.3.2	Validación de datos	40
5.4	Análisis factorial confirmatorio	40
5.4.1	Construcción del primer modelo de ecuaciones estructurales (MEE)	41
5.4.2	Segundo modelo de ecuaciones estructurales	43
5.4.3	Tercer modelo de ecuaciones estructurales	43
5.4.4	Modelo final de ecuaciones estructurales	44
5.5	Obtención del indicador de calidad del asistente educativo holográfico .	46
5.5.1	Fundamentación y aplicación del modelo de ecuaciones estruc- turales	47
5.5.2	Uso de computación genética	47
5.5.3	Proceso de obtención de la ecuación	47
5.6	Validación de la hipótesis	52
6	Conclusiones	53
	Referencias	54
7	Anexos	56
7.1	Informe de pruebas de software	56
7.2	Implementación de la tesis por computadora	59

Índice de figuras

1	Prototipo de HIVA (Isaev et al., 2023).	10
2	Prototipo de HolograFX (Cerezo et al., 2019).	13
3	Proyector piramidal usado para mostrar la animación 3D (Hoon y Shaharuddin, 2019).	14
4	Modelo de Calidad basado en la ISO/IEC 9126 (Jaramillo y Pérez, 2016).	20
5	Arquitectura del sistema.	23
6	Vista frontal del avatar.	25
7	Vista lateral izquierda del avatar.	25
8	NadIA 2.0 en funcionamiento.	25
9	Portada del cuestionario de evaluación.	29
10	Fragmento del conjunto de datos de las métricas de calidad de texto.	39
11	Cantidad de variables del conjunto de datos de las métricas de calidad de texto.	39
12	Fragmento del conjunto de datos del cuestionario.	39
13	Cantidad de variables del conjunto de datos del cuestionario.	39
14	Fragmento del nuevo conjunto de datos del cuestionario.	39
15	Fragmento de conjunto de datos final.	40
16	Cantidad de datos por cada variable del conjunto de datos final.	40
17	Cantidad de variables del conjunto de datos final.	40
18	Matriz de correlación mediante el método de Spearman.	41
19	Representación gráfica del MEE final.	46
20	Resultados obtenidos de la primera regresión simbólica.	47
21	Resultados obtenidos de la segunda regresión simbólica.	49
22	Resultados obtenidos de la regresión simbólica final.	51

Índice de tablas

1	Tabla comparativa de Asistentes Educativos Holográficos.	17
2	Tabla comparativa de Asistentes Educativos Holográficos (continuación).	18
3	Tabla comparativa de modelos GPT.	20
4	Fragmento del conjunto de datos creado (Parte 1).	32
5	Fragmento del conjunto de datos creado (Parte 2).	32
6	Resultados de las pruebas de normalidad.	33
7	Objetos de la investigación	37
8	Adquisición de los datos	37
9	Índices de bondad de ajuste del modelo propuesto.	42
10	Índices de bondad de ajuste del segundo modelo.	43
11	Índices de bondad de ajuste del tercer modelo.	44
12	Índices de bondad de ajuste del modelo final.	45
13	Resultados del cálculo del indicador de calidad del AEH.	52
14	Resultados del cálculo del indicador de calidad del AEH (continuación).	52
15	Descripción del caso de prueba.	58
16	Resultados del caso de prueba.	59
17	Anomalías en el caso de prueba.	59

Glosario

Asistente educativo holográfico: herramienta tecnológica que utiliza hologramas para proporcionar instrucción y apoyo a los estudiantes. Estos hologramas pueden ser representaciones tridimensionales de personajes animados o avatares virtuales. Los asistentes educativos holográficos se pueden utilizar en una variedad de entornos educativos, incluidos en centros de tutoría o en los salones de clases presenciales y virtuales (Habboosh y Muhammad, 2022).

Modelos de lenguaje de grandes: también llamados MLG's, es un tipo de inteligencia artificial (IA) que puede procesar y generar texto. Los MLG's se entrenan en grandes cantidades de datos de texto, lo que les permite aprender las complejidades del lenguaje natural humano y realizar una variedad de tareas. Se basan típicamente en una arquitectura de red neuronal llamada transformador (*transformers*). Los transformadores son una poderosa herramienta para procesar lenguaje natural porque pueden aprender relaciones de largo alcance entre las palabras en una oración. Esto les permite comprender el contexto de una oración y generar texto que sea relevante y coherente (Makridakis et al., 2023).

Calidad de software: conjunto de características intrínsecas que hacen posible que el software satisfaga las necesidades de los usuarios y partes interesadas. Se identifican seis características principales de la calidad del software, cada una de las cuales se descompone en subcaracterísticas más específicas (Funcionalidad, Confiabilidad, Usabilidad, Eficiencia, Mantenibilidad y Portabilidad) (ISO/IEC 9126).

Usabilidad: grado en que el software puede ser utilizado de manera efectiva, eficiente y satisfactoria. Esto implica que el software sea fácil de aprender y usar, comprensible, intuitivo, con menús sencillos y una lectura de textos ágil. Además, las funciones deben ser claras y puntuales, facilitando así una experiencia positiva y sin complicaciones para el usuario, (ISO/IEC 9126).

Funcionalidad: evaluación la adecuación, el cumplimiento funcional, idoneidad, corrección, interoperabilidad, conformidad y seguridad de acceso, por lo que es posible afirmar que la funcionalidad determina la capacidad del software de funcionar en términos de lo que el usuario necesita, de interactuar con otros sistemas y que permita el acceso de diferentes personas pero que cumpla con las regulaciones de las leyes de protección de datos (ISO/IEC 9126).

Coherencia: es un proceso de la estructura semántica de un discurso, en la que una serie de conceptos primarios se relacionan entre ellos, así como un conjunto de conceptos secundarios que complementan a los primeros (Huerta, 2010).

Cohesión: es una propiedad del texto, inherente a la coherencia y reflejada en las pistas para la comprensión del texto (Huerta, 2010).

Gramaticalidad: representa la propiedad de pertenecer al sistema lingüístico de un determinado idioma, siendo una propiedad constitutiva e interna relativa a su naturaleza formal del texto (Muñoz et al., 2009).

Complejidad del léxico: se mide la complejidad léxica de un texto, determinada por la frecuencia de uso y la densidad léxica. Se considera que a mayor densidad léxica (mayor número de palabras diferentes por textos) mayor dificultad para la comprensión (López et al., 2018).

Lectorabilidad de Fernández-Huerta: siendo una adaptación al español de la prueba de legibilidad de Flesh y teniendo en cuenta que en español las palabras en promedio tienen más sílabas y las oraciones también son más largas. Esta métrica mide el promedio de las sílabas por palabra y el promedio de palabras por oración que hay en el texto (López et al., 2018).

Legibilidad Inflesz: es una escala para evaluar la legibilidad de textos en español, donde un valor más alto del índice indica mayor facilidad de lectura (Barrio Cantalejo et al., 2008).

Capítulo 1

1. Introducción

En el contexto educativo contemporáneo, los Asistentes Educativos (AE's) desempeñan un papel esencial como apoyo para educadores y alumnos. Se han convertido en una tendencia clave en la transformación digital de las aulas. Un AE es un profesional especializado que colabora con educadores para proporcionar apoyo y asistencia en el aula. Este rol ha evolucionado significativamente con los avances tecnológicos y la diversidad de las necesidades educativas. Los Asistentes Educativos no solo ayudan con las tareas administrativas, sino que también ofrecen apoyo individualizado a los estudiantes, facilitan el aprendizaje inclusivo y contribuyen al ambiente de aprendizaje colaborativo.

Si se implementara un asistente tutor usando IA, no tendría las deficiencias de un tutor de verdad (Kaya y Bulut, 2022). No tendría sentimientos ni pensamientos como el enojo, la impaciencia o el prejuicio, por lo tanto, no podrá reflejar estos sentimientos hacia los estudiantes. De esta manera, responderá pacientemente cada pregunta de los estudiantes, así, los estudiantes que tienen personalidades introvertidas y tímidas podrán escuchar al profesor sin sentirse preocupados y hacer preguntas fácilmente sobre los temas que no comprenden.

Por otra parte, el papel de la Inteligencia Artificial en el sector educativo se encuentra rezagado debido a que muchos educadores necesitan ser convencidos sobre su uso en la educación superior, además de que requiere un gran financiamiento para implementarse en todo el sector educativo superior (Suvrat y Roshita, 2019). También por medio de la revisión de la literatura, encontró un estudio realizado en Vermont que aproximadamente tres cuartas partes de las instrucciones de educación especial fueron impartidas por un Asistente Educativo y menos de una cuarta parte por profesores de educación especial u otros profesionales (Orchard, 2023). Los hallazgos señalaron que los estudiantes con discapacidad y perfiles de aprendizaje complejos no reciben tanta responsabilidad de maestros o profesionales, lo que pone en duda la equidad para los estudiantes con habilidades y discapacidades.

Por lo que una manera de fomentar y aumentar el interés para el uso de un AE, sería por medio de un Asistente Educativo Holográfico (AEH) el cual es una innovación que utiliza tecnología holográfica para crear una representación tridimensional de un asistente educativo en el aula. Esta representación holográfica permite una interacción inmersiva entre el Asistente Educativo y los estudiantes, creando un ambiente de aprendizaje único. Los Asistentes Educativos Holográficos están equipados con capacidades de Inteligencia Artificial, lo que les permite entender y responder a las consultas de los estudiantes de manera dinámica y envolvente. Como resultado, en el mundo tecnológico actual, la integración de la IA en el campo de la educación es una situación inevitable, por lo que, es crucial que los sistemas educativos de los países del mundo estén informados sobre las aportaciones de la IA en la educación.

Sin embargo, existe una necesidad de brindar nuevas herramientas para el de aprendizaje de los estudiantes, ya que la educación, por más avances tecnológicos que hay en la actualidad, su aplicación integrada en el aula de clase es limitada, por lo que sigue siendo en gran medida convencional.

1.1. Planteamiento del problema

En el contexto educativo actual, persiste la necesidad de implementar nuevas herramientas que potencien el aprendizaje de los estudiantes. Aunque se han logrado avances tecnológicos significativos, la interacción en entornos educativos continúa siendo, en gran medida, tradicional y limitada (Hoon y Shaharuddin, 2019). A pesar del desarrollo de tecnologías como la holografía y el Procesamiento de Lenguaje Natural (PLN), su integración efectiva en el aula es aún escasa.

En respuesta a esta necesidad, se han comenzado a incorporar **Asistentes Educativos Holográficos (AEH)** como una innovación en entornos académicos. Sin embargo, **el problema central radica en que dichos asistentes, en la mayoría de los casos, no ofrecen respuestas satisfactorias ni pertinentes para los usuarios.** Frecuentemente, sus respuestas no cumplen con lo solicitado, contienen información irrelevante o carecen de coherencia con la pregunta formulada.

Por tanto, **el problema que se busca atender en esta investigación es la falta de mecanismos que permitan evaluar objetivamente la calidad de un AEH, específicamente en términos de su funcionalidad y usabilidad.**

1.2. Hipótesis

Es posible evaluar cuantitativamente la Calidad de un AEH basado en un MLG comercial con base en métricas propias del PLN y aspectos de Usabilidad.

1.2.1. Delimitación del problema

En este trabajo de investigación se enfoca en el desarrollo de un proceso de evaluación de la calidad de un AEH, el cual está basado en un MLG (*gpt-3.5 Turbo*) para la generación de respuestas. Aunque la norma ISO/IEC 9126 contempla múltiples dimensiones para la evaluación de la calidad del software, esta investigación se delimita exclusivamente a dos de ellas: Usabilidad y Funcionalidad. Además de solo contemplar las métricas de coherencia, cohesión, gramaticalidad y complejidad del léxico, lecturabilidad de Fernández-Huerta y legibilidad Inflesz para la evaluación de la calidad del texto.

1.2.2. Complejidad del problema

- Evaluar la calidad de un AEH implica afrontar desafíos tecnológicos, educativos y metodológicos de forma simultánea.
- La dimensión de la calidad del software es amplia y compleja, ya que funcionalidad y usabilidad tienen subdimensiones propias.
- Evaluar la calidad de las salidas generadas por el AEH implica el desarrollo de ecuaciones matemáticas y computacionales específicas para cada métrica.
- Desarrollar una metodología adecuada para esta evaluación no es sencillo debido a la complejidad del enfoque.
- Desarrollar un AEH funcional con bajo costo computacional.

1.3. Objetivos

1.3.1. Objetivo general

Evaluar la Calidad de un Asistente Educativo Holográfico basado en un MLG en términos de Usabilidad y Funcionalidad.

1.3.2. Objetivos específicos

- Comprender conceptos de Asistente Educativo, Técnicas de holografía, Modelos de Lenguaje, principios de Semiótica y métricas de Calidad de texto.
- Desarrollar un corpus de contenidos educativos a partir de recursos existentes.
- Integrar principios de Semiótica Visual para proyectar en el AEH.
- Crear una evaluación del AEH considerando métricas de Calidad de Texto y evaluación de Modelos de Lenguaje.
- Explorar la evaluación propuesta para incrementar la Calidad de un AEH basado en un MLG.
- Realizar pruebas piloto con estudiantes de un nivel educativo determinado para evaluar la Calidad del AEH.

1.4. Alcances y limitaciones

1.4.1. Alcances

- Se optará por el MLG *gpt-3.5 Turbo*, habiéndolo seleccionado después de comparar varios modelos diferentes, destacando su superioridad con respecto a otros modelos de lenguaje, también soporta el idioma español para la generación de texto.
- Se usará un corpus etiquetado.
- Se utilizarán métricas basadas en coherencia, cohesión, gramaticalidad y complejidad del léxico, lecturabilidad de Fernández-Huerta y legibilidad Inflesz para la evaluación de la calidad de texto.
- Se reutilizarán herramientas de software existentes para el desarrollo de los módulos del AEH.
- Se implementará y construirá un prototipo de software de un AEH en lenguaje Android.

1.4.2. Limitaciones

- Solo se considerará el idioma español para el desarrollo del AEH.
- El corpus no será de contenido universal.
- El AEH solo podrá contestar en ambientes semi-controlados.
- Se limitará el uso del AEH a la interacción uno a uno.
- El AEH no estará capacitado para contestar preguntas ambiguas.

- No se considerarán otras métricas, exceptuando las que se mencionaron anteriormente.
- No se desarrollarán módulos o algoritmos propios de voz a texto, texto a voz, MLG, creación de avatar, prototipo de hardware y contenidos educativos.
- No se contempla tener el prototipo de software terminado para un uso comercial.

1.5. Justificación

En México, la población joven enfrenta desafíos significativos en el ámbito educativo. Según datos explicados por la Dra. Martínez Bordón y el Dr. Manuel López Pereyra, Académicos de la Universidad Iberoamericana (IBERO), en el Foro “La educación media superior en México siglo XXI” organizado por la Subsecretaría de Educación Media Superior de la Secretaría de Educación Pública (Universidad Iberoamericana, 2023), para el ciclo escolar 2023-2024, la tasa de abandono en la educación media superior alcanzó el 8.1 %, lo que representa aproximadamente 430 mil estudiantes que dejaron sus estudios en ese periodo . Esta situación es particularmente preocupante para el grupo de edad de 15 a 25 años, ya que la deserción escolar en esta etapa limita sus oportunidades de desarrollo personal y profesional.

Además, el rendimiento académico de los jóvenes ha mostrado signos de deterioro. Un informe de UNICEF destaca que, en países de ingresos altos, incluido México, se ha observado una disminución en la comprensión lectora y un aumento en la insatisfacción vital entre niños y adolescentes desde la pandemia. Estos factores afectan directamente la calidad del aprendizaje y la motivación de los estudiantes (Cadena SER, 2024).

En este contexto, la integración de tecnologías emergentes en el ámbito educativo se presenta como una estrategia para fortalecer la calidad de la educación. Los Asistentes Educativos Holográficos (AEH), que combinan tecnologías de holografía y Procesamiento de Lenguaje Natural (PLN), ofrecen una oportunidad para enriquecer la experiencia de aprendizaje.

Por lo tanto, esta investigación se propone desarrollar un modelo de evaluación que permita medir cuantitativamente la calidad de los AEH, enfocándose en su funcionalidad y usabilidad. Al establecer métricas claras y objetivas, se busca proporcionar una herramienta que facilite la mejora continua de estas tecnologías.

1.6. Estructura de la tesis

Este documento está compuesto de la siguiente manera:

En el Capítulo 2 se describen conceptos teóricos necesarios para facilitar la comprensión de este proyecto de investigación: cómo son las métricas de evaluación de las salidas de un AEH, cómo funcionan y componentes relacionados a un AEH.

En el Capítulo 3 se mencionan trabajos relacionados y de gran relevancia para esta investigación.

En el Capítulo 4 se describe la metodología que se desarrolló en esta investigación.

En el Capítulo 5 se explican la experimentación realizada y los resultados obtenidos.

Finalmente el Capítulo 6 se comentan las conclusiones obtenidas de esta investigación, los objetivos alcanzados, el producto desarrollado y propuestas para trabajos futuros.

Capítulo 2

2. Marco teórico

En este capítulo se abordan los conceptos fundamentales y especializados que resultan esenciales para el entendimiento, fundamentación y desarrollo de la presente investigación. Estos conceptos permiten contextualizar el estudio, clarificar los términos clave y establecer las bases teóricas sobre las que se sustenta el análisis del AEH.

2.1. Asistente Educativo Holográfico

Un Asistente Educativo Holográfico es una forma avanzada de tecnología educativa que utiliza proyecciones holográficas. Proporcionan asistencia y apoyo en el proceso de aprendizaje. Este tipo de asistente utiliza la realidad aumentada para interactuar con los estudiantes de manera más inmersiva y realista. La tecnología de hologramas mostró características que la habilitaron como una opción adecuada para ser utilizada en educación para mostrar contenido educativo en 3D (Ahmad et al., 2021).

Un asistente educativo holográfico puede estar compuesto por varias partes que trabajan en conjunto para ofrecer una experiencia educativa inmersiva y efectiva.

- **Proyector holográfico:** se encarga de proyectar imágenes holográficas en el espacio tridimensional. Las proyecciones son en forma 3D, utilizando técnicas holográficas para dar la impresión de objetos o personajes flotando en el aire.
- **Hardware de procesamiento:** son los componentes como procesadores, tarjetas gráficas, dispositivos móviles y memoria RAM que son necesarios para ejecutar la o las aplicaciones para los gráficos holográficos.
- **Software de inteligencia artificial:** Los asistentes educativos holográficos suelen estar equipados con software de inteligencia artificial que les permite comprender y procesar el lenguaje natural para adaptarse a las necesidades del usuario.
- **Base de datos de contenido educativo:** contiene una amplia variedad de contenido educativo, que puede incluir lecciones, información, y material visual de referencia. El asistente educativo holográfico accede a esta base de datos para proporcionar información relevante.
- **Interfaz de usuario:** la interfaz de usuario permite interactuar con el asistente educativo holográfico y el usuario de manera intuitiva. Esto puede incluir gestos de mano, comandos de voz, pantallas táctiles u otros métodos de entrada.

2.2. Holografía

La holografía fue inventada en 1948 por Dennis Gabor en un esfuerzo por mejorar la resolución del microscopio electrónico. La holografía es un método de recuperación integral de la información relacionada con el campo de irradiación generalizada de un objeto real, es decir, es un registro de los patrones de interferencia de la luz, que permite obtener imágenes ópticas tridimensionales de distintos objetos (Ribeiro et al., 2018). Esto permite crear imágenes que tienen la apariencia de estar tridimensionalmente presentes en el espacio, con altura, anchura y profundidad. Algunos métodos o técnicas usuales para proyectar un holograma son:

- **Efecto *Pepper's Ghost*:** tiene como propósito generar la presencia de un espectro o la ilusión óptica, que nos hace creer que es real, eso que solamente es una imagen. Esta técnica es una de las precursoras en la utilización de la imagen virtual en la escena y crea una relación entre la presencia virtual y la presencia “real”.
- **Ventilador holográfico:** a través de la rotación de imágenes LED permite visualizar efectos visuales 3D debido a la alta densidad del dispositivo. Mediante el ventilador se visualiza de manera clara, ya que la pantalla admite varias conexiones. Este dispositivo utiliza una combinación de tecnologías holográficas y ventiladores para crear imágenes tridimensionales en el aire.
- **Cortina de humo holográfica:** la cortina de humo holográfica, holograma aéreo, pantalla de vapor, o pantalla de humo, es capaz de crear una pantalla virtual y transparente de partículas de vapor de agua en la cual somos capaces de proyectar imágenes en 3D e incluso cualquier tipo de vídeo. Se puede colocar en cualquier lugar con ambientes de luz controlable. Utilizando proyectores láser de alta potencia y técnicas de difracción de luz, se forman imágenes tridimensionales en el espacio del humo, creando una experiencia visual inmersiva para los espectadores.
- **Pantallas OLED semitransparentes:** es una pantalla transparente que consta de dos áreas, una es un área emisora que emite luz debido a la presencia de un pixel y la otra es un área emisora que emite luz debido a la presencia de un pixel y la otra es un área transparente, que aumenta la transmitancia total debido a la ausencia de píxeles.
- **Redes holográficas:** las redes holográficas se reconstruyen con un espectrofotómetro de doble haz, obteniéndose la transmitancia de la red en función de la longitud de onda. Se puede ver como una cortina casi transparente de luces LED que muestran imágenes en una variedad de colores. Estos LED's están programados para crear imágenes que presentan un efecto holográfico 3D.

2.3. Métricas para evaluar la calidad de texto

Para evaluar la calidad del texto de una oración generada automáticamente, existen varias métricas que se pueden utilizar. Las métricas utilizadas en esta investigación son:

2.3.1. Coherencia

Es un proceso de la estructura semántica de un discurso, en la que una serie de conceptos primarios se relacionan entre ellos, así como un conjunto de conceptos secundarios que complementan a los primeros (Huerta, 2010). Por lo que podemos decir que la coherencia es el resultado y el proceso en sí mismo de la cohesión dentro de un texto. Para medir la coherencia de un texto, se puede hacer de la siguiente manera:

$$\text{Coherencia} = PCL + PCG + PEP \quad (1)$$

Donde; *PCL* es el Puntaje de Cohesión Léxica que evalúa la variedad de vocabulario y el uso adecuado de sinónimos y términos relacionados. Más variedad y uso adecuado de vocabulario podrían significar un puntaje más alto. *PCG* es el Puntaje Cohesión Gramatical que valúa el uso adecuado de conectores, pronombres y la estructura gramatical correcta. Un uso correcto y variado de conectores y pronombres podría significar un puntaje más alto y *PEP* es Puntaje Estructura de Párrafo que evalúa la organización lógica de las ideas en el texto. Una estructura de párrafo clara y lógica podría significar un puntaje más alto.

2.3.2. Cohesión

La cohesión es una propiedad del texto, inherente a la coherencia y reflejada en las pistas que el autor da para comprender el mensaje (Huerta, 2010). La cohesión nos propone el uso de indicadores textuales que sirven como nexo entre los términos que componen un texto. Si queremos medir la cohesión básica de un texto, se haría de la siguiente manera:

$$\text{Cohesión} = \frac{NPPC}{NTPP} \quad (2)$$

Donde; *NPPC* es el Número de Pares de Palabras Coherentes que se refiere al número de veces que palabras específicas o tipos de palabras (como pronombres) ocurren en una relación específica en el texto y *NTPP* es el Número Total de Pares de Palabras que podrían formar una relación coherente en el texto.

2.3.3. Gramaticalidad

La gramaticalidad de las expresiones representa una propiedad constitutiva e interna relativa a su naturaleza formal, la gramaticalidad de las consecuencias constituye una parte fundamental de la ciencia del lenguaje (Muñoz et al., 2009). Representa la propiedad de pertenecer al sistema lingüístico de un determinado idioma. Una ecuación simplificada para medir gramaticalidad sería:

$$\text{Gramaticalidad} = 1 - \frac{NEG}{NTEGT} \quad (3)$$

Donde; *NEG* es el Número de Errores Gramaticales que se refiere al número de errores gramaticales encontrados en el texto (esto podría incluir errores como discordancia de tiempo, errores de concordancia, uso incorrecto de preposiciones,

etc.) y *NTEGT* es el Número Total de Elementos Gramaticales en el Texto, como palabras y estructuras gramaticales.

2.3.4. Complejidad del léxico

Esta métrica se utiliza para medir la complejidad léxica de un texto, determinada por la frecuencia de uso y la densidad léxica. Se considera que a mayor densidad léxica (mayor número de palabras diferentes por textos) mayor dificultad para la comprensión (López et al., 2018). Para medir la complejidad de un texto sería con la siguiente ecuación:

$$\text{Complejidad}(T) = \frac{NPU}{NP} \quad (4)$$

Donde; *NPU* es el número de palabras únicas y *NP* es el número total de palabras.

2.3.5. Lecturabilidad de Fernández-Huerta

Siendo una adaptación al español de la prueba de legibilidad de Flesh y teniendo en cuenta que en español las palabras en promedio tienen más sílabas y las oraciones también son más largas. Esta métrica mide el promedio de las sílabas por palabra y el promedio de palabras por oración que hay en el texto (López et al., 2018). Para calcular este índice, lo podemos hacer con la siguiente ecuación:

$$\text{Fernández-Huerta} = 206.84 - (0.60 \cdot P) - (1.02 \cdot F) \quad (5)$$

Donde; *P* es el promedio de sílabas por 100 palabras y *F* es el promedio de palabras por frase.

2.3.6. Legibilidad Inflesz

Es una escala para evaluar la legibilidad de textos en español, donde un valor más alto del índice indica mayor facilidad de lectura (Barrio Cantalejo et al., 2008). Esto se debe a que toma en cuenta la cantidad de sílabas por palabra como la cantidad de palabras por frase para determinar la complejidad de un texto. Donde se representa con la siguiente ecuación:

$$\text{Inflesz} = 206.835 - (62.3 \cdot P) - F \quad (6)$$

Donde; *P* es el promedio de sílabas por palabra y *F* es el promedio de palabras por frase.

Capítulo 3

3. Estado del arte

En este capítulo se presentan trabajos de investigación relevantes y directamente relacionados con el tema de esta investigación. Se abordan conceptos claves, algoritmos y métodos que sirven como base para el desarrollo de un AEH, proporcionando el sustento teórico y técnico necesario para comprender los enfoques existentes y sustentar la propuesta planteada.

HIVA: Holographic Intellectual Voice Assistant (Isaev et al., 2023)

Este trabajo presenta a HIVA (Asistente Holográfico de Voz Intellectual), la cual es una tecnología innovadora que combina un avatar en 3D con un asistente de voz para proporcionar una experiencia audiovisual única para acceder a información de la Universidad Internacional de Ala-Too. HIVA proporciona información sobre admisión, problemas académicos, tarifas, departamentos, estructura e historia de la universidad, cafetería, recursos humanos, biblioteca, vida estudiantil y eventos, información sobre el país y la ciudad, etc. En el proyecto HIVA, se enfocaron en construir un efecto holográfico conocido como “*Pepper’s ghost*”. Los resultados obtenidos del PLN de HIVA fueron de un rango del 74 % al 97 % de precisión en diversas pruebas. Por lo tanto, la eficiencia del modelo aún puede mejorarse. En la Figura 1 se puede observar el prototipo de HIVA de manera funcional.



Figura 1: Prototipo de HIVA (Isaev et al., 2023).

Holographic Artificial Intelligence Assistance (Girish et al., 2019)

Este trabajo menciona que los sistemas de asistentes de Inteligencia Artificial actuales se utilizan para tomar el habla del usuario como entrada y procesarla para proporcionar la salida deseada. Sin embargo, los sistemas disponibles actualmente son Asistentes Virtuales Privados (del inglés *Virtual Private Assistant* o VPA). Lo que significa que se puede comunicar con el asistente, pero no es visible. Para desarrollar un asistente visible, los autores usaron una caja transparente con un monitor en la parte superior y un prisma de vidrio colocado dentro de la caja en un ángulo de 45 grados. Se usan Redes Neuronales Convolucionales (del inglés *Convolutional*

Neural Network o CNN) para usar filtros que se utilizan para detectar características específicas en una imagen, en este caso el holograma que será proyectado. Se determinó que la eficiencia y precisión eran medidas importantes, además de que, con el aumento de las funcionalidades, era más complejo implementarlas.

Cloud for Holography and Augmented Reality (Makris et al., 2021)

En este trabajo se presenta a CHARITY, una tecnología que aprovecha redes inteligentes para mejorar las experiencias holográficas, permitiendo aplicaciones de comunicación inmersiva como Realidad Virtual (del inglés *Virtual Reality* o VR), Realidad Aumentada (del inglés *Augment Reality* o AR) u holografía. CHARITY se compone de cuatro capas principales que son respaldadas por un marco de gestión de aplicaciones (del inglés *Application Management Framework* o AMF) y un Canal de Integración y Entrega Continua (del inglés **Continuous Integration/Continuous Delivery** o CI/CD). Como conclusión se da que es importante que las aplicaciones de Realidad Extendida (RE), en este CHARITY, en sus diferentes caracterizaciones de VR, AR e interfaces holográficas no deben de ser exageradas, referentes a la complejidad que supone implementar todas estas tecnologías en un solo dispositivo y aunque CHARITY se encuentra en una etapa inicial, se trata desarrollar nuevas tecnologías, algoritmos y arquitecturas de software para respaldar el diseño inicial de esta arquitectura.

Dispositivo Holográfico Interactivo para Mejorar el Aprendizaje en Estudiantes de Primer Grado de Primaria de la I.E.P Jan Komensky (Bartra y Tirado Miranda, 2018)

En esta investigación el objetivo principal fue mejorar el aprendizaje con el desarrollo e implementación de un dispositivo holográfico interactivo llamado Ana a través de imágenes tridimensionales proyectadas en un prisma generando un holograma visible a 360. Donde se implementó una arquitectura multicapa basada principalmente en 3 capas, donde cada capa a su vez puede contener más capas. Se hicieron pruebas con 26 estudiantes de primer grado de primaria en dos salones, uno donde se probaba el dispositivo holográfico y otro donde no. Los resultados antes de la implementación del dispositivo holográfico interactivo eran sólo un 53 % de aprobación de las respuestas correctas con un promedio de 2.12 de las respuestas. Mientras que con la implementación de dicho dispositivo se logró obtener un 99 % de las respuestas correctas correspondiente a un promedio de 3.96 de las respuestas, lo que incrementó el promedio de participaciones correctas en 46 %. Se concluyó que al implementar el dispositivo holográfico interactivo, se logró mejorar el aprendizaje en los estudiantes de primer grado de primaria de la I.E.P Jan Komensky.

Holographic display in future automotive smart cockpit: application scenarios, interaction modals, and VACP analysis (Wei et al., 2023)

En este trabajo habla acerca de una pantalla holográfica tridimensional que puede ofrecer más posibilidades para la interacción humano-máquina en la futura cabina inteligente, principalmente en cabinas de automóviles. Se utiliza el modelo VACP (del inglés Visual, Auditory, Cognitive, and Psychomotor) para validar los principios de interacción, basándose en el análisis y la evaluación de las tecnologías de pantalla holográfica existentes y potenciales de vanguardia. Durante la realización del artículo, se notó que los parámetros clave de las máquinas ópticas no pueden cumplir con los requisitos para su aplicación en la producción masiva. Por otro lado,

la luminosidad de la proyección es insuficiente, lo que conduce a una pobre calidad de visualización durante el día, especialmente en áreas con luz solar directa. Como resultado de lo anterior, se obtuvo que, en términos de tecnología, una pantalla holográfica no está lo suficientemente madura como para aparecer, incluyendo en vehículos de producción en masa. Por un lado, el obstáculo para el desarrollo proviene de las limitaciones del equipo óptico en sí mismo, y, por otro lado, de su alta demanda de potencia informática.

An Overview of 3D Holographic Visualization Technologies and Their Applications in Education (Keselj et al., 2023)

Este documento menciona que, la visualización tridimensional (3D) es el proceso de crear contenido gráfico utilizando tecnologías de software, donde el objeto se muestra utilizando una computadora o dispositivo móvil con el software adecuado, y el objeto 3D visualizado se puede ver desde diferentes ángulos y lados. Los objetos 3D se pueden visualizar utilizando una pantalla de computadora con gráficos 3D o mediante tecnologías como la realidad aumentada o los hologramas. Se desarrolló *HoloGeometry*, es una aplicación diseñada para ayudar a los estudiantes a aprender objetos geométricos, ya que proporciona una visualización de sólidos geométricos en el espacio con la posibilidad de interacción. Satisface las necesidades de los estudiantes que procesan la información principalmente de manera visual y mejora el aprendizaje para todos los estudiantes. Se utilizó el entorno de desarrollo *Unity* para crear la aplicación, usando los recursos de "Pirámide holográfica", que se utilizó para implementar el holograma en la aplicación. La aplicación incluye tres niveles: (aprender formas geométricas, reconocer formas geométricas y prueba tus conocimientos). Además de usar la técnica de *Pepper's Ghost* para la creación del holograma.

Evaluating a holographic mobile-based application for teaching basic oral English vocabulary to Spanish speaking children (Cerezo et al., 2019)

Este artículo describe el desarrollo, uso y evaluación de una aplicación móvil holográfica para enseñar inglés básico como segundo idioma, en particular, vocabulario oral a niños de habla española. Esta investigación aprovecha el poder de los estímulos multicanal (sonido, imagen e interacción) de una aplicación holográfica basada en móviles para motivar a los estudiantes y mejorar su experiencia de aprendizaje. Se adaptó un juego HolograFX y se desarrolló una nueva aplicación móvil para ayudar a practicar la pronunciación en inglés, además de un robot holográfico en 3D que actúa como profesor virtual se encarga de enseñar vocabulario oral en inglés. Como se aprecia en la Figura 2, HolograFX utiliza la técnica de proyección *Pepper's Ghost*.

3D Visualization through the Hologram for the Learning of Area and Volume Concepts (Orcos et al., 2019)

Este estudio tiene como objetivo implementar y evaluar una propuesta metodológica utilizando el holograma como medio de enseñanza para el aprendizaje de conceptos relacionados con áreas y volúmenes de cuerpos geométricos. El estudio se llevó a cabo con una muestra de 78 estudiantes del tercer año de educación secundaria de una escuela financiada por el estado pero de propiedad privada en Madrid. Para evaluar su eficacia, se utilizó la prueba *t de Student* para evaluar la existencia de diferencias significativas entre ambos grupos independientes bajo las suposiciones paramétricas, empleando el programa estadístico SPSS.

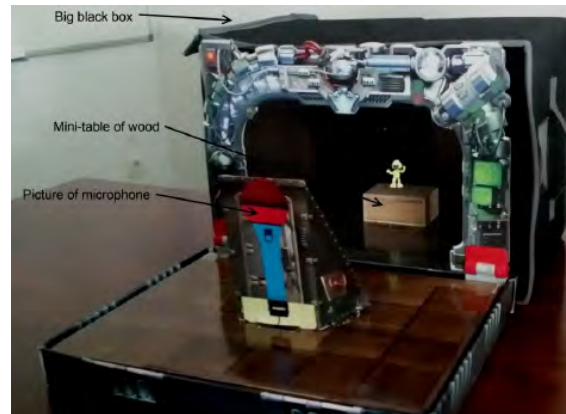


Figura 2: Prototipo de HolograFX (Cerezo et al., 2019).

The hologram as a teaching medium for the acquisition of STEM contents (Orcos y Magreñán, 2018)

El objetivo del artículo presentado es evaluar si el uso de un holograma mejora el aprendizaje de los contenidos de la división celular. Se llevó a cabo una fase piloto con una muestra de estudiantes de cuarto curso de educación secundaria en la ciudad de Logroño, España, trabajando con un grupo de control de manera tradicional y un grupo experimental con el uso del holograma. El pre-test consistió en diez preguntas sobre conceptos muy simples relacionados con la división celular. Dado que los estudiantes no habían profundizado en los procesos de mitosis y meiosis anteriormente, la calificación obtenida en el pre-test fue muy baja. Los resultados mostraron que el 47.6 % de los estudiantes sabían que una célula humana tiene 46 cromosomas y el 57.1 % sabía que cuando la célula se divide, el material genético se convierte en cromosomas. La hipótesis nula, H_0 , establecía que no había diferencias significativas entre los grupos y la hipótesis alternativa, H_1 , que las había. En caso de que el valor obtenido fuera menor que el valor p , para un nivel de significado de $\alpha = 0.05$.

Learning Effectiveness of 3D Hologram Animation on Primary School Learners (Hoon y Shaharuddin, 2019)

Este trabajo habla acerca de la aplicación de la tecnología de holograma tridimensional (3DH) para estudiantes de primero, segundo y tercer grado para determinar la efectividad del holograma 3D en el aula. Se utilizaron diseños de pre-test y post-test de acuerdo con el plan de estudios de la escuela primaria para recopilar datos, que fueron analizados estadísticamente utilizando el software SPSS. Un total de 50 estudiantes de SJK Chung Hua Kranji Bau, Sarawak, participaron en la evaluación. Los participantes se agruparon en 3 categorías de edad: estudiantes de primer grado (7 años), segundo grado (8 años) y tercer grado (9 años). En la Figura 3 se puede observar el prototipo diseñado.

Interactive Holograms for Construction Coordination and Quantification (Tayeh y Issa, 2020)

Este artículo presenta una nueva técnica de visualización para modelos de información de construcción y hologramas interactivos y valida su efectividad a través de un estudio experimental. Se desarrolló un complemento para automatizar la transferencia de modelos a motores de juegos, un software que permite una interacción



Figura 3: Proyector piramidal usado para mostrar la animación 3D (Hoon y Shaharuddin, 2019).

intuitiva con el holograma mediante gestos y comandos de voz, y una plataforma de visualización del holograma. El propósito principal de esta investigación es evaluar cómo la holografía puede mejorar la comprensión espacial y las habilidades para resolver problemas de los usuarios. Se planteó la hipótesis de que el holograma ayuda a los usuarios a tomar mejores decisiones gerenciales debido a que la visualización interactiva está disponible para múltiples usuarios al mismo tiempo. El estudio se llevó a cabo en la Escuela Rinker de Gestión de la Construcción de la Universidad de Florida. Un total de 55 estudiantes participaron y completaron el estudio, de los cuales el 64 % eran estudiantes de pregrado, desde primer año hasta último año, y el 36 % restante eran estudiantes de posgrado. La evaluación de los resultados se basó en los colores y texturas del edificio, así como en la información incorporada en sus elementos. De los 55 participantes, el 67 % percibió que usar el complemento era más fácil que 3DS MAX (el software que utilizan), y ninguno de los participantes sintió que el complemento era difícil o muy difícil de usar.

Corporate social responsibility (CSR) towards education: the application and possibility of 3D hologram to enhance cognitive skills of primary school learners (Loh y Shukhaila, 2019)

El trabajo presenta un análisis exhaustivo del estudio de caso del holograma tridimensional (3DH) como herramientas de aprendizaje digital y proporciona una visión general de la teoría de posibilidades según los usos de 3DH, para mejorar las habilidades cognitivas de aprendizaje para los niños de primaria. La publicación se dividió en dos subsecciones: desarrollo de habilidades cognitivas en el proceso de aprendizaje de los niños y ventajas de la aplicación de 3DH. Basándose en la observación, la animación de hologramas en 3D es capaz de captar el interés de los estudiantes y llamar su atención. Pueden entender fácilmente los conceptos de los temas a través de las ilustraciones de la animación de hologramas en 3D cuando están concentrados, se divierten y comprenden. Al final de la animación, los estudiantes solicitan ver más. Los resultados del metaanálisis y la observación muestran que el uso de la tecnología de hologramas en 3D puede captar la atención de los niños y mejorar su proceso de aprendizaje, la visualización de la información es el siguiente paso para el desarrollo de las habilidades de atención en los niños para comprender las lecciones.

A Virtual Assistant for Natural Interactions in Museums (Duguleană et al., 2020)

El artículo presenta el desarrollo de un asistente virtual inteligente para interacciones naturales en museos, con el objetivo de mejorar la accesibilidad a la información dentro de un museo histórico en Brasov, Rumania. El asistente permite la interacción en lenguaje hablado, utilizando un avatar en 3D que responde a las preguntas de los visitantes en rumano. El sistema se basa en inteligencia artificial para procesar el lenguaje natural y reconocer comandos de voz, empleando plataformas como Google Cloud Speech to Text y RASA NLU para entender y responder en tiempo real. Los desarrolladores se centraron en crear una experiencia interactiva, integrando el reconocimiento de voz y un avatar visual que puede sincronizar los movimientos de labios y gestos con las respuestas generadas. El estudio de usuarios incluyó la interacción de 250 estudiantes con el asistente en el Museo Casa Mureșenilor. Los resultados muestran una buena aceptación por parte de los visitantes, destacando la innovación del uso de avatares 3D y la utilidad de la IA para transferir conocimientos en contextos culturales.

Design of Seamless Multi-modal Interaction Framework for Intelligent Virtual Agents in Wearable Mixed Reality Environment (Ali et al., 2019)

Este trabajo presenta el diseño de un marco de interacción multimodal para agentes virtuales inteligentes en entornos de realidad mixta (MR) portátil, aplicable a lugares interactivos como museos y jardines botánicos. La tecnología combina el mapeo espacial, el reconocimiento de objetos, el reconocimiento de voz y las animaciones de personajes virtuales para maximizar la participación del usuario. La arquitectura del sistema permite a los agentes virtuales interactuar de manera fluida con los usuarios, reconociendo gestos y objetos, proporcionando información específica del dominio y mostrando gestos corporales y emociones acordes a la conversación. La modularidad del sistema permite su implementación en dispositivos con recursos limitados, delegando tareas intensivas a la nube, lo que resulta en tiempos de respuesta entre 2 y 4 segundos. Además, el sistema es adaptable a cualquier dispositivo de realidad mixta que soporte mapeo espacial. Durante las pruebas, los usuarios interactuaron sin interrupciones con los agentes virtuales, lo que subraya la efectividad del marco para mejorar las experiencias inmersivas en entornos de realidad aumentada.

Modelling the Interaction Levels in HCI Using an Intelligent Hybrid System with Interactive Agents: A Case Study of an Interactive Museum Exhibition Module in Mexico (Rosales et al., 2018)

En este documento se presenta un sistema para modelar los niveles de interacción en la Interacción Humano-Computadora (HCI) utilizando un sistema híbrido inteligente con agentes interactivos. La investigación se centra en evaluar y mejorar las experiencias de los usuarios midiendo los niveles de interacción en un entorno de exposición de museos en México. El sistema utiliza un sistema de inferencia difusa (FIS) para clasificar los niveles de interacción basándose en factores como presencia, interactividad, control, retroalimentación y creatividad. Se recopilieron y analizaron datos de 500 visitantes del museo utilizando una combinación de FIS empírico, árboles de decisión y sistemas neuro-difusos. Los resultados muestran que el sistema neuro-difuso proporcionó la mayor precisión en la predicción de los niveles de interacción. Esta investigación destaca el potencial para mejorar las interacciones

entre usuarios y agentes en entornos complejos como las exhibiciones interactivas de museos, con aplicaciones más amplias en el diseño de sistemas inteligentes.

Multimedia Learning Module (MLM) with Hologram Simulation to Improve Students' Mathematical Representation Ability (Mawardani et al., 2021)

En este documento se presenta un sistema para modelar los niveles de interacción en la Interacción Humano-Computadora (HCI) utilizando un sistema híbrido inteligente con agentes interactivos. La investigación se centra en evaluar y mejorar las experiencias de los usuarios, midiendo los niveles de interacción en un entorno de exposición de museos en México. El sistema utiliza un sistema de inferencia difusa (FIS) para clasificar los niveles de interacción basándose en factores como presencia, interactividad, control, retroalimentación y creatividad. Se recopilaron y analizaron datos de 500 visitantes del museo utilizando una combinación de FIS empírico, árboles de decisión y sistemas neuro-difusos. Los resultados muestran que el sistema neuro-difuso proporcionó la mayor precisión en la predicción de los niveles de interacción. Esta investigación destaca el potencial para mejorar las interacciones entre usuarios y agentes en entornos complejos, como las exhibiciones interactivas de museos, con aplicaciones más amplias en el diseño de sistemas inteligentes.

3.1. Discusión del Estado del Arte

El análisis de los trabajos evidencia avances en el desarrollo de AEH, sin embargo, se identifican deficiencias relevantes. La mayoría de estudios priorizan aspectos visuales, como la técnica de *Pepper's Ghost*, o el diseño de prototipos, pero pocos abordan de manera rigurosa la calidad de las respuestas generadas mediante modelos de lenguaje grandes. Asimismo, persiste una ausencia de métricas específicas de evaluación funcional y de usabilidad, lo que limita la posibilidad de medir objetivamente el desempeño de estos sistemas.

Por otro lado, varios prototipos presentan requerimientos técnicos elevados o dependen de infraestructura poco accesible para su implementación. Ante estas limitaciones, esta investigación propone un enfoque que integra el uso de modelos de lenguaje de grandes con técnicas de procesamiento de lenguaje natural, evaluando cuantitativamente la calidad del AEH a través de un modelo de ecuaciones estructurales. Esta propuesta permite atender vacíos metodológicos existentes.

Las Tablas 1 y 2 resumen los elementos clave de cada artículo, incluyendo cita, descripción, algoritmo, técnica o método utilizado, dominio, idioma, resultados y conclusiones.

Tabla 1: Tabla comparativa de Asistentes Educativos Holográficos.

Cita	Descripción	Algoritmo, técnica o método usado	Dominio e idioma	Resultados	Conclusión
Isaev et al., 2023.	Tecnología innovadora que combina un avatar en 3D con un asistente de voz para proporcionar una experiencia audiovisual.	Algoritmo <i>Naive Bayes</i> (MNB).	Información institucional e inglés.	Se obtuvieron resultados alrededor del 74 % y el 97 % de precisión.	Se descubrió que los asistentes virtuales servían como objetos de prueba para llegar a los mismos resultados como si se tratara de personas reales.
Girish et al., 2019.	Asistente de Inteligencia Artificial Holográfica, que permite a los usuarios interactuar con un asistente holográfico en 3D utilizando voz.	Redes Neuronales Convolucionales (CNN).	Asistencia educativa y médica e inglés.	La capa convolucional utiliza filtros que se utilizan para detectar estas características específicas en una imagen.	Se hizo una mejora en la precisión de su funcionamiento como el ser menos propenso al ruido y los acentos.
Makris et al., 2021.	Tecnología que aprovecha redes inteligentes para mejorar las experiencias holográficas, <i>CHARITY</i> .	Redes Neuronales Gráficas (GNN's).	Multidominio e inglés.	<i>CHARITY</i> hace una combinación de técnicas de redes, almacenamiento, comprensión y gestión dinámica.	Debido a la fase temprana, se espera que en el futuro se hallan nuevas tecnologías para respaldar el diseño inicial.
Bartra y Tirado Miranda, 2018.	Implementación de un dispositivo holográfico interactivo para mejorar el aprendizaje llamado Ana.	Arquitectura multicapas (<i>Main activities, Analyse problem domain, Desing</i> , etc).	Razonamiento matemático y español.	Se obtuvo que, la comprensión y atención de los niños hacia las vocales en un 27.6 % y números en un 22.6 % con el uso de Ana.	Se mejoró el aprendizaje, antes el porcentaje de aprobación era de 53 % y con la implementación del asistente se llegó a un 99 %.
Wei et al., 2023.	Exploración del potencial de las pantallas holográficas tridimensionales en el futuro de las cabinas de automóviles.	Modelo de datos basado en las respuestas visuales, auditivas (VACP)	Automovilismo e inglés.	Después de una serie de investigaciones con pruebas de concepto, se propusieron y analizaron las tendencias futuras de desarrollo.	La visualización holográfica actual no está lo suficientemente madura como para aparecer en vehículos de producción en masa a corto plazo.
Keselj et al., 2023.	Desarrollo de <i>HoloGeometry</i> para ayudar a los estudiantes a aprender objetos geométricos, ya que proporciona una visualización de sólidos geométricos.	No especificado.	Conceptos de geometría e inglés.	Debido a que se tiene un prototipo, no se tienen datos relevantes.	Proporciona una visión general de las tecnologías de visualización holográfica, sus capacidades de interacción e integración móvil.
Cerezo et al., 2019	Desarrollo de una aplicación móvil holográfica para enseñar inglés básico como segundo idioma, a niños de habla española.	Tecnología de proyección en 3D a cámara completa (<i>HolograFX</i>).	Práctica de la pronunciación y vocabulario oral en inglés.	Con una muestra de 70 niños, se dividieron en 3 grupos, uno sin la aplicación, otro con la aplicación sin holografía y el último con la aplicación holográfica, donde este último destacó con un 65 % de aprobación.	Los resultados del experimento de <i>pre-test</i> y <i>post-test</i> confirman que los niños tuvieron un mejor rendimiento al aprender y practicar utilizando el asistente holográfico.
Orcos et al., 2019.	Implementación de una metodológica utilizando el holograma como medio de enseñanza para el aprendizaje.	No especificado.	Áreas y volúmenes geométricos en inglés.	Una muestra de 78 alumnos, 38 para el proceso experimental, donde resultó con 3.9 % de aprobación con respecto al 0.05 % antes de la prueba.	Se obtuvieron porcentajes de 97.4 % que los estudiantes recomendarían al holograma como medio de enseñanza.

Tabla 2: Tabla comparativa de Asistentes Educativos Holográficos (continuación).

Cita	Descripción	Algoritmo, técnica o método usado	Dominio e idioma	Resultados	Conclusión
Orcos y Magreñán, 2018.	Implementación de hologramas para mejorar el aprendizaje de los contenidos de la división celular.	<i>Socratic®</i> y <i>Tea-cher®</i> .	Conceptos de división celular e inglés.	Los estudiantes se dividieron en dos fases, con y sin el uso del holograma. En el <i>pre-test</i> se obtenía un promedio de 0.05 de aprobación, mientras que el <i>post-test</i> un promedio de 2.55 de aprobación de respuestas.	Los resultados permiten considerar el uso de la holografía como un medio de enseñanza potencial, así como una herramienta altamente motivadora.
Hoon y Shaharudin, 2019.	Aplicación de tecnología holográfica tridimensional (3DH) para estudiantes de primero, segundo y tercer grado.	Holograma piramidal aplicando efecto “ <i>Pepper’s Ghost</i> ”.	Desarrollo vegetal e inglés.	Una muestra de 50 alumnos donde aumentó un 71.55 % en respuestas correctas en el <i>post-test</i> , en comparación con el 13.45 % del <i>pre-test</i> .	Los puntajes demuestran que la tecnología 3DH puede influir positivamente y mejorar el conocimiento adquirido durante el proceso de aprendizaje de los estudiantes.
Tayeh y Issa, 2020.	Técnica de visualización para modelos de construcción y hologramas interactivos y válida su efectividad.	<i>BIM-data</i> y complemento para <i>Autodesk Revit</i> .	Arquitectura, construcción e inglés.	Una muestra de 55 estudiantes, donde el 67 % se les hizo más fácil identificar las estructuras usando la aplicación holográfica.	La holografía en entornos educativos tiene el potencial de permitir que los profesores y alumnos tengan mayor y mejores oportunidades de interacción.
Loh y Shukhaila, 2019.	Análisis exhaustivo del estudio de caso del holograma tridimensional como herramientas de aprendizaje digital.	Holograma piramidal aplicando efecto “ <i>Pepper’s Ghost</i> ”.	Conceptos de holografía e inglés.	El uso de la tecnología de hologramas en 3D puede captar la atención de los niños y mejorar su proceso de aprendizaje.	Basándose en la observación, la animación de hologramas en 3D logra captar el interés de los estudiantes y atraer su atención.
Duguleană et al., 2020.	Desarrollo de un agente conversacional inteligente que se construyó para mejorar la accesibilidad a la información dentro de un museo de historia.	<i>RASA NLU</i> .	Asistente informativo e inglés.	Se hicieron 5 preguntas a 250 participantes, de edad entre los 13 y los 18 años, donde un promedio de 81 % dijeron que el asistente les parecía atractivo.	Se muestra buenos niveles de aceptación tanto de los visitantes como del personal del museo, dada la flexibilidad de la implementación del asistente holográfico.
Ali et al., 2019.	Marco multimodal para agentes virtuales inteligentes en entornos de realidad mixta portátiles, especialmente para aplicaciones interactivas en museos, jardines botánicos y lugares similares.	Interacción multimodal basada en la mirada y el habla.	Botánica e inglés.	Se hicieron 3 consultas donde la consulta A fue la carga del mapa de anclaje y requirió reconocimiento de objetos. La consulta B fue una conversación general. La consulta C fue sobre flores y requirió reconocimiento de objetos.	Los agentes virtuales respondieron dentro de los 2-4 segundos posteriores a la consulta del usuario.
Rosales et al., 2018.	Sistema híbrido inteligente con agentes evaluando los niveles de interacción de la Interacción Humano-Computadora (HCI).	Sistema de Inferencia Difuso (FIS), árboles de decisión y sistemas neurodifusos.	Asistente informativo e inglés.	Muestra de 500 usuarios que visitaron el museo interactivo “El Trompo”, donde el sistema neuro difuso tuvo una exactitud de 91.3 %.	Se explora la evaluación de la interacción en la HCI utilizando la Teoría de Niveles de Interacción de Gayer y Williams para mejorar la experiencia del usuario.
Mawardani et al., 2021.	Mejora de habilidades de representación matemática de los estudiantes después de usar el Módulo de Aprendizaje Multimedia (MLM) con simulación de hologramas.	Módulo de aprendizaje multimedia (MLM).	Análisis matemático e inglés.	Muestra de 35 estudiantes donde los estudiantes trabajan en preguntas sobre matemáticas antes y después de usar el MLM, el cual aumentó el promedio de aciertos en un 76 %, comparado al 36 % antes de usarlo.	El MLM holográfico facilita la comprensión de las habilidades de representación matemática de los estudiantes, al mismo tiempo que fomenta el aprendizaje independiente.

Capítulo 4

4. Metodología

En este capítulo se describe la metodología empleada para el desarrollo del presente trabajo de investigación. Se detallan las fases que conforman el proceso, desde la comprensión del problema, el diseño y desarrollo del AEH, hasta su evaluación funcional y de usabilidad. Además, se explican los métodos de recolección y análisis de datos, las herramientas tecnológicas utilizadas (como el modelo de lenguaje) y los criterios seguidos para construir un indicador cuantitativo de calidad, fundamentado en métricas de calidad de texto y percepción de los usuarios. A continuación se presenta la propuesta de solución empleada.



Propuesta de solución.

4.1. Descripción de la propuesta de solución

4.1.1. Comprensión de la problemática

Se realiza un análisis exhaustivo de la definición del problema, estableciendo con claridad sus alcances y limitaciones. Asimismo, se lleva a cabo un estudio de las características de los Asistentes Educativos Holográficos (AEH), de las métricas de calidad de texto y las dimensiones de Calidad de software.

Dimensiones de calidad de software. Debido a que el concepto de Calidad es muy abstracto, la norma ISO/IEC 9126 establece un marco para evaluar la calidad del software a través de diversas dimensiones, como se observa en la Figura 4. Entre ellas, la funcionalidad y la usabilidad, dado su impacto directo en el desempeño y la experiencia del usuario.



Figura 4: Modelo de Calidad basado en la ISO/IEC 9126 (Jaramillo y Pérez, 2016).

4.1.2. Exploración y evaluación de Modelos de Lenguaje Grandes

Se llevó a cabo un análisis comparativo de distintos modelos de lenguaje grandes, con el propósito de seleccionar el modelo más adecuado para el desarrollo del AEH. La evaluación se centró en tres modelos: *gpt-3.5 Turbo*, *gpt-3.5* y *gpt-2.5 Turbo*, donde se escogería el MLG con mejores características.

Tabla 3: Tabla comparativa de modelos GPT.

Modelo	Parámetros (miles de millones)	Numero de capas	Tamaño de los embeddings	Días de entre- namiento	Fecha de lan- zamiento
<i>gpt-3.5 Turbo</i>	250	64	4096	30	15/07/2023
<i>gpt-3.5</i>	175	48	3072	25	20/05/2023
<i>gpt-2.5 Turbo</i>	100	24	2048	20	10/03/2023

La Tabla 3 presenta las principales características de los modelos evaluados. Se observa que *gpt-3.5 Turbo* es el modelo con el mayor número de parámetros, la mayor cantidad de capas y el mayor tamaño de embeddings. Además, su tiempo de entrenamiento fue de 30 días, lo que indica un mayor nivel de refinamiento en comparación con los demás modelos. Por otro lado, *gpt-3.5* y *gpt-2.5 Turbo* muestran una configuración más limitada. Ambos modelos fueron entrenados en un menor período de tiempo (25 y 20 días, respectivamente), lo que puede repercutir en su desempeño en tareas complejas.

Tras la comparación de las especificaciones técnicas y considerando el impacto que estos factores tienen en la generación y comprensión del lenguaje, se determinó que *gpt-3.5 Turbo* representa la mejor opción, donde la mejor configuración de sus hiperparámetros fue con base en su temperatura de 0.8, el máximo de tokens de 1000, su rol asignado es de sistema y la instrucción para construir su prompt fue: “Dame información breve de 2 líneas”. Su mayor número de parámetros y capas, junto con un tamaño de embeddings superior, le otorgan una mayor capacidad de representación y generalización en tareas lingüísticas avanzadas. Además, al momento de iniciar la investigación, es el modelo más reciente con respecto a los otros modelos analizados, lo que lo convierte en la alternativa más robusta para el desarrollo del AEH.

4.1.3. Formulación de heurísticas de calidad para AEH's

La heurística se puede definir como una disciplina científica de amplio alcance, aplicable a diversas áreas del conocimiento para el desarrollo de métodos, principios y estrategias que permitan encontrar soluciones efectivas a problemas complejos. Los procedimientos heurísticos pueden clasificarse en distintas metodologías, cada una adaptada a contextos y aplicaciones específicas.

- Reglas heurísticas: establecen los medios para resolver un problema, proporcionando directrices específicas para la toma de decisiones.
- Estrategias heurísticas: organizan los materiales y recursos disponibles, facilitando la búsqueda de soluciones mediante enfoques estructurados.
- Principios heurísticos: ofrecen recomendaciones o sugerencias generales que orientan la identificación de la solución más adecuada al problema planteado.

En este contexto, para la formulación de heurísticas para AEH toma como referencia las propuestas de Nielsen y Molich (1990), así como las recomendaciones de Lolambean (2023) para el diseño de interfaces holográficas. El objetivo es establecer principios heurísticos para garantizar altos niveles de funcionalidad y usabilidad en los AEH, optimizando la experiencia de interacción con los usuarios.

Los principios heurísticos definidos a partir de este análisis buscan proporcionar un marco estructurado que facilite la evaluación y el desarrollo de interfaces educativas holográficas, asegurando su eficacia y accesibilidad. Los 9 principios heurísticos definidos en esta investigación son:

1. Visibilidad del estado del sistema. Asegurar que el asistente proporcione de manera clara señales visuales, auditivas que indiquen la interacción del usuario, reflejando estados como, por ejemplo, estar procesando una solicitud o estar en pausa. Esto garantiza que los usuarios sean conscientes del estado actual del sistema, promoviendo la confianza y reduciendo la incertidumbre en la interacción.

2. Relación con el mundo real. Utilizar íconos, ilustraciones y animaciones que sean reconocibles y que correspondan con objetos y conceptos del mundo real, para hacer que las explicaciones sean más intuitivas y fáciles de recordar.

3. Control y libertad del usuario. Ofrecer al usuario la capacidad de gestionar la interacción con el asistente holográfico de forma flexible, permitiendo, por ejemplo, pausar, retroceder, avanzar o repetir contenido en cualquier momento. Los controles deben ser fácilmente accesibles y visibles, proporcionando una experiencia interactiva en la que el usuario tenga el control total sobre el ritmo y la dirección de la sesión.

4. Consistencia y estándares. Asegurar que los términos utilizados en las respuestas y la forma de presentación sean consistentes a lo largo de la interacción, utilizando siempre un lenguaje y estilo visual coherentes.

5. Prevención de fallas. Implementar mecanismos que minimicen el riesgo de fallas en el sistema del asistente holográfico, asegurando que las interacciones sean claras y comprensibles para los usuarios. Así como proporcionar advertencias o mensajes de aclaración en caso de detectar potenciales flujos de riesgo en la interacción.

6. Diseño estético y minimalista. Utilizar un diseño visual minimalista en el holograma, con colores suaves y transiciones.

7. Identificación de las necesidades del usuario. Comprender el perfil, las necesidades y las expectativas de los usuarios para adaptar las funcionalidades del asistente holográfico de manera flexible y escalable.

8. Navegación intuitiva. Crear una interfaz holográfica que sea fácil de navegar, reduciendo al máximo la carga cognitiva del usuario y minimizando la posibilidad de frustración.

9. Capacidad conversacional. Incorporar técnicas avanzadas de procesamiento de lenguaje natural que permitan al asistente holográfico gestionar interacciones fluidas y naturales, adaptándose a diferentes niveles de complejidad en las consultas.

Se recomienda que el AEH a desarrollar cumpla con al menos 6 de las 9 heurísticas mencionadas, con el fin de garantizar altos niveles de intención de uso futuro. Esto permitirá mejorar la usabilidad y satisfacción del usuario, buscando que el AEH sea intuitivo y eficiente en sus funciones.

4.1.4. Diseño de un AEH

Para el desarrollo de un AEH se requiere la integración de diversas tecnologías que permitan ofrecer una experiencia interactiva, intuitiva y eficiente para los usuarios. En este contexto, se presenta NadIA 2.0, un AEH que incorpora múltiples módulos y elementos con el propósito de garantizar una interacción fluida entre el usuario y el sistema. Su diseño se basa en un enfoque cliente-servidor, integrando tecnologías avanzadas de procesamiento de lenguaje natural y soporte de servidores externos para optimizar el desempeño del sistema.

La arquitectura de NadIA 2.0 sigue un modelo distribuido compuesto por tres componentes principales: cliente, servidor y servicios externos. En la Figura 5 se ilustra la arquitectura y los flujos de comunicación.

- Cliente: gestiona la interacción directa con el usuario mediante módulos de entrada y salida, además de una interfaz holográfica que facilita la visualización de respuestas.
- Servidor: procesa las consultas enviadas por el cliente, evalúa la calidad del contenido generado y administra los datos de uso y rendimiento.
- Servicios externos: incluyen el servidor de Google, que convierte preguntas auditivas en texto y respuestas en audio, y el servidor de *OpenAI*, que procesa el texto con un MLG (*gpt-3.5 Turbo*).

El flujo de procesamiento de una consulta en NAdIA 2.0 se desarrolla en los siguientes pasos:

1. El usuario formula una petición en formato auditivo a través del cliente.
2. El cliente envía el audio al servidor de Google, que lo convierte en texto.
3. El texto se transmite al servidor principal, quien lo reenvía al servidor de *OpenAI* en forma de prompt para su procesamiento.
4. El Servidor de *OpenAI* selecciona al MLG (*gpt-3.5 Turbo*) y genera una respuesta basada en el prompt recibido.
5. La respuesta es enviada al servidor principal, donde se evalúa mediante métricas de calidad de texto antes de enviarse al cliente.

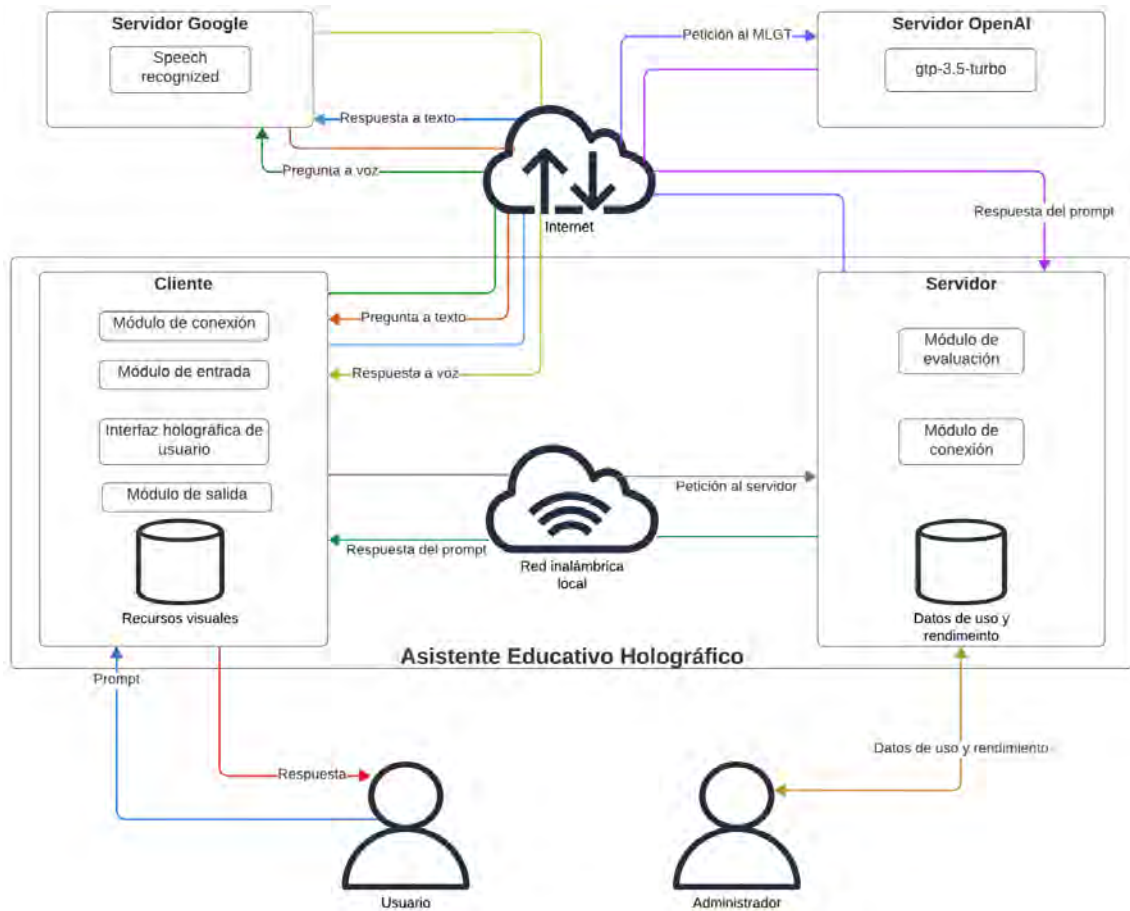


Figura 5: Arquitectura del sistema.

6. El cliente reenvía la respuesta al servidor de Google, quien transforma en audio y lo manda al cliente.
7. Finalmente, el cliente presenta la respuesta en formato auditivo y visual a través de la interfaz holográfica.

4.1.5. Desarrollo de un AEH

Para el desarrollo del asistente educativo holográfico NadIA 2.0, se emplean tecnologías de software y hardware que garantizan un funcionamiento eficiente y una interacción fluida con los usuarios.

Entorno de desarrollo y lenguajes de programación. El sistema se realizó utilizando *Python 3.12.3* como lenguaje principal para la implementación del servidor, dentro del entorno de desarrollo *Visual Studio* y bajo el sistema operativo Windows 11 Pro. Para el desarrollo de la interfaz del cliente, se empleó el IDE *Android Studio 2024.1.1*, utilizando Java como lenguaje de programación para la integración de los módulos de interacción holográfica.

Configuración de hardware. El sistema se implementa sobre una infraestructura de hardware que asegura la estabilidad y el rendimiento óptimo del AEH. Los componentes incluyen:

- **Servidor:** una laptop GIGABYTE CPU i5-12500H x64 con 16 GB de RAM, utilizada para el procesamiento de consultas y la gestión de los modelos de

lenguaje.

- Cliente: una tablet CUBOT con Android 13 x64 y 4 GB de RAM, utilizada para la interacción con el usuario mediante la interfaz holográfica.
- Cristal de proyección: elemento de proyección caracterizado por proporcionar una reproducción de color precisa y un alto nivel de brillo para una visibilidad óptima en diversas condiciones de iluminación con base en un ángulo de 45 grados con una resolución (720 x 1280 píxeles).
- Base montable de proyección: Base en forma de prisma rectangular para permitir la proyección de dispositivos móviles de 9 cm de ancho x 21 cm de largo x 15 cm de altura.
- Un mouse para facilitar la interacción manual y un micrófono incorporado para la recepción de comandos de voz.

Creación del avatar. Para la creación del avatar NadIA, se empleó la herramienta *MetaHuman Creator*, desarrollada por *Epic Games*. Esta tecnología permite la generación de personajes humanos digitales con un alto nivel de realismo, ampliamente utilizada en la industria de los videojuegos, el cine y otras aplicaciones digitales interactivas.

El uso de *MetaHuman Creator* facilita un diseño detallado y altamente personalizable, permitiendo la configuración precisa de rasgos faciales, peinados y vestimenta, garantizando una apariencia natural y expresiva en el avatar. Como se observa en las Figuras 6 y 7, esta herramienta se complementa con la avanzada tecnología de gráficos en tiempo real de *Unreal Engine 5.4*, lo que permite la generación de personajes con una fidelidad visual excepcional de manera rápida y eficiente. Además, se presenta en la Figura 8 el prototipo funcional de NadIA 2.0.



Figura 6: Vista frontal del avatar.



Figura 7: Vista lateral izquierda del avatar.



Figura 8: NadIA 2.0 en funcionamiento.

4.1.6. Pruebas funcionales de un AEH

Se aplicaron pruebas funcionales a NadIA 2.0, con el objetivo de verificar su correcto desempeño y capacidad de respuesta en un entorno real de uso. Para ello, se implementó un enfoque basado en pruebas de caja negra, permitiendo evaluar el sistema sin necesidad de analizar su código fuente, centrándose en su comportamiento desde la perspectiva del usuario.

El proceso de pruebas incluyó las siguientes etapas:

1. Diseño de casos de prueba: se definen escenarios específicos de interacción con el AEH para evaluar sus respuestas con base en las preguntas y dudas que le realice el usuario.
2. Reclutamiento de sujetos de prueba: se seleccionan participantes de maestría para interactuar con el sistema en condiciones controladas.
3. Ejecución de pruebas: los sujetos de prueba utilizan el AEH, ejecutando los casos diseñados y registrando su experiencia de uso.
4. Recopilación de datos: se registran métricas relacionadas con la funcionalidad del sistema.

Para información más detallada, se presenta en el **Informe de pruebas de software**.

4.1.7. Evaluación de aspectos asociados a la funcionalidad del AEH

Para evaluar la funcionalidad del Asistente Educativo Holográfico (AEH), se empleó un enfoque basado en el análisis de métricas de calidad de texto. Estas métricas permiten medir el desempeño del sistema en términos de coherencia, cohesión, gramaticalidad, complejidad léxica y legibilidad.

El proceso de evaluación consideró las siguientes etapas:

1. Generación de respuestas: el AEH procesa entradas de usuario y genera respuestas mediante el MLG.
2. Cálculo de métricas: tanto la entrada (pregunta) como la salida (respuesta) del sistema se someten a un análisis basado en métricas de calidad de texto, incluyendo:
 - Cohesión: mide la relación semántica entre palabras dentro del texto.
 - Gramaticalidad: evalúa la corrección lingüística mediante la detección de errores.
 - Complejidad léxica: cuantifica la proporción de palabras únicas con respecto al total de palabras del texto.
 - Índice Fernández-Huerta: estima la legibilidad del texto basándose en la cantidad de sílabas y la longitud promedio de las oraciones.
 - Índice Inflesz: evalúa la facilidad de lectura considerando la estructura del texto.
 - Coherencia: integra la cohesión, la gramaticalidad y la riqueza léxica para medir la fluidez del contenido generado.

3. Registro de resultados: los datos obtenidos se almacenan y analizan para optimizar la funcionalidad del AEH, asegurando que las respuestas generadas sean comprensibles, gramaticalmente correctas y adecuadas para el contexto educativo.

Este proceso permite evaluar de manera objetiva la capacidad del AEH para generar respuestas, garantizando su funcionalidad del sistema.

4.1.8. Evaluación de aspectos de usabilidad percibidos por los usuarios de un AEH

Para analizar la usabilidad del AEH desde la perspectiva del usuario, se diseñó un cuestionario específico con el fin de recopilar percepciones y experiencias de los participantes. Este instrumento permitió obtener información cuantitativa sobre distintos aspectos del sistema, identificando tanto fortalezas como áreas de mejora.

En el contexto de esta investigación se optó por diseñar un cuestionario propio en lugar de utilizar instrumentos estandarizados como el *Post-Study System Usability Questionnaire* (PSSUQ) o el *Computer System Usability Questionnaire* (CSUQ), dado que diversos estudios han demostrado que en contextos similares, como en el trabajo experimental de Bartra y Tirado Miranda (2018), se privilegia el uso de encuestas simplificadas con escalas Likert reducidas, para facilitar la comprensión y obtener respuestas más representativas por parte de los usuarios. Asimismo, se tomaron en cuenta referentes metodológicos orientados al uso educativo de tecnologías holográficas, como el estudio de Hoon y Shaharuddin (2019), quienes estructuran instrumentos post-test para evaluar la experiencia del usuario de forma simplificada y contextualizada.

Además de tomar como fundamento el modelo de evaluación propuesto en la tesis doctoral de Alejandres (2017), se adaptó y modificó dicho modelo para esta investigación. El instrumento resultante se organizó en bloques estructurados que evalúan las dimensiones de Sistema, Percepción, Experiencia e Interacción. Con base en lo anterior, resulta pertinente detallar los siguientes aspectos del cuestionario:

Objetivo del cuestionario. El propósito del cuestionario es evaluar la experiencia de los usuarios con el AEH, considerando los siguientes aspectos clave:

- Familiaridad con sistemas de inteligencia artificial: mide el nivel de conocimiento y experiencia previa de los participantes en el uso de tecnologías basadas en IA.
- Satisfacción: evalúa el grado de satisfacción de los usuarios tras su interacción con el AEH.
- Efectividad percibida: analiza la percepción de los usuarios sobre la precisión y utilidad de las respuestas proporcionadas por el asistente.
- Calidad percibida: valora la relevancia del contenido recibido a través del AEH.

Diseño del cuestionario. El cuestionario se diseña bajo principios metodológicos que garantizan la recolección de datos confiables y representativos. Para ello, se incorporan preguntas cerradas y se emplea una escala de Likert para responderlas.

Procedimiento de aplicación. Para garantizar la homogeneidad en la recolección de datos, el cuestionario se aplica siguiendo un protocolo estructurado compuesto por las siguientes etapas:

1. Los participantes interactúan con el AEH en un entorno supervisado, formulando preguntas académicas. Esta sesión permite evaluar el funcionamiento del sistema en condiciones reales de uso.
2. Aplicación inmediata del cuestionario. Tras la interacción, los participantes responden el cuestionario en formato digital, garantizando que sus respuestas reflejen de manera precisa su experiencia con el AEH.
3. Duración. El tiempo promedio para completar el cuestionario es de aproximadamente 6 minutos, asegurando un equilibrio entre la profundidad de la información recolectada y la comodidad de los participantes.

Análisis de datos. Para interpretar los resultados del cuestionario, se implementaron métodos de análisis cuantitativo:

- Análisis de contenido: las respuestas se clasifican y codifican en categorías temáticas para identificar patrones recurrentes en la percepción de los usuarios.

Limitaciones del instrumento. Al analizar los resultados del cuestionario, es fundamental considerar posibles limitaciones en su aplicación, tales como:

- Muestra reducida: un número limitado de participantes puede afectar la generalización de los resultados.
- Sesgo de respuesta: algunos usuarios podrían haber respondido de manera sesgada.
- Entorno controlado: las condiciones supervisadas de la prueba pueden no reflejar completamente el uso cotidiano del AEH.

En la Figura 9 se muestra la portada del cuestionario de evaluación de NadIA.

4.1.9. Modelado de la calidad de un AEH

Para evaluar la calidad del AEH, se adoptó un enfoque analítico basado en métricas cuantificables. Se implementó un análisis factorial confirmatorio con el propósito de examinar la relación entre variables latentes y observables, permitiendo la construcción de un modelo de ecuaciones estructurales (MEE). Este modelo representa de manera matemática la interacción entre funcionalidad y usabilidad dentro del AEH.

Análisis de datos. Para garantizar la validez del modelo, se aplican los siguientes procedimientos estadísticos:

- Pruebas de normalidad: se utiliza el test de Shapiro-Wilk para determinar si las variables analizadas siguen una distribución normal.
- Cálculo de correlaciones: se emplea el coeficiente de Spearman para identificar asociaciones entre las métricas de calidad de texto y las percepciones de los usuarios sobre el sistema.

Construcción del modelo de ecuaciones estructurales. El desarrollo del modelo se lleva a cabo en distintas fases:

1. Definición del modelo inicial: se establece una primera estructura basada en las correlaciones observadas.

Cuestionario de Evaluación de NadIA

Este cuestionario tiene como objetivo evaluar la experiencia de los usuarios con el **Asistente Educativo Holográfico NadIA**. A través de una serie de preguntas, buscamos conocer la percepción de los estudiantes en cuanto a la familiaridad con el uso de sistemas basados en inteligencia artificial, la satisfacción general con el uso de NadIA, la efectividad de las respuestas proporcionadas por el asistente, y la calidad percibida de la información recibida.

Instrucciones para responder:
Cada pregunta debe ser respondida seleccionando la opción que mejor describa tu experiencia personal con NadIA.

m23ce090@cenidet.tecnm.mx [Cambiar de cuenta](#)

No compartido

* Indica que la pregunta es obligatoria

Familiaridad con el uso de sistemas de Inteligencia Artificial

Estoy familiarizado/a con los sistemas de Inteligencia Artificial. *

☐ Totalmente de acuerdo

☐ De acuerdo

Figura 9: Portada del cuestionario de evaluación.

2. Evaluación de la bondad de ajuste: se analizan métricas como el Chi-cuadrado X^2 , el Índice de Ajuste Comparativo (CFI) y el Error Cuadrático Medio de Aproximación (RMSEA).
3. Ajustes progresivos: se refinan las relaciones entre las variables, mejorando la estructura del modelo y optimizando los índices de ajuste.
4. Modelo final optimizado: se obtiene un modelo robusto de la relación entre funcionalidad y usabilidad en el AEH.

4.1.10. Formulación de un indicador de calidad para un AEH

Una vez validado el MEE, se procede a la construcción de un indicador cuantitativo que permita evaluar objetivamente la calidad del Asistente Educativo Holográfico (AEH). Para ello, se emplea un enfoque basado en regresión simbólica, una técnica de computación genética que permite obtener una ecuación óptima sin imponer una forma funcional predeterminada.

Definición de variables predictoras. Para la formulación del indicador de calidad, se establecen dos variables latentes:

- Establecer dos variables latentes:
 - Funcionalidad: definida a partir de las métricas de calidad de texto

- Usabilidad: determinada mediante la efectividad del asistente, la calidad de la respuesta, el uso posterior del sistema y la satisfacción de los usuarios con NadIA 2.0.
- Adicionalmente, se incorporan variables de interacción con el objetivo de capturar relaciones más complejas identificadas en el MEE.

Generación de la ecuación. El proceso de construcción del indicador sigue los siguientes pasos:

1. División del conjunto de datos: el total de observaciones se separa en entrenamiento (80 %) y prueba (20 %).
2. Configuración del modelo: se implementa una regresión simbólica con penalización por complejidad para evitar sobreajustes en la ecuación resultante.
3. Optimización iterativa: la ecuación se ajusta a lo largo de múltiples generaciones e iteraciones, reduciendo progresivamente el error de ajuste.
4. Obtención de la ecuación final: se genera una representación matemática optimizada que representa de manera cuantitativa la relación entre funcionalidad y usabilidad en el AEH.

La ecuación obtenida proporciona una medida cuantitativa de calidad, combinando métricas de PLN y percepciones de los usuarios. Este indicador permite evaluar el desempeño del AEH de manera objetiva.

Capítulo 5

5. Experimentación y resultados

En este capítulo se presentan los experimentos realizados y los resultados obtenidos a partir de la implementación del AEH. Se expone la comparativa de la generación de respuestas entre un corpus generado a partir de una Red Neuronal Recurrente (CNN) y el Modelo de Lenguaje Grande (MLG), el proceso de validación de los métodos propuestos, la interpretación de los indicadores obtenidos y la discusión sobre el desempeño del AEH en términos de funcionalidad y usabilidad. Los resultados permiten sustentar la efectividad del enfoque metodológico y detectar oportunidades de mejora para desarrollos futuros.

5.1. Experimentación Corpus vs MLG

A partir del *corpus* obtenido de la tesis doctoral de Iturbe (2023), se llevó a cabo una comparación basada en las respuestas generadas por cada modelo. Para ello, se seleccionaron 100 preguntas al azar sobre diversos temas educativos. Estas preguntas se analizaron con el objetivo de calcular, a partir de sus respuestas, con base en las métricas de cohesión, gramaticalidad, complejidad léxica, lecturabilidad de Fernández-Huerta, legibilidad Inflesz y coherencia. Posteriormente, las mismas preguntas del corpus fueron planteadas al MLG, y se calcularon las mismas métricas para ambos casos, con el fin de establecer una comparación sistemática.

En las Tablas 4 y 5 presentan las métricas extraídas para cada uno de los casos.

- Cohesión representa el nivel de relación semántica entre palabras y oraciones dentro de la respuesta.
- Gramaticalidad mide el cumplimiento de las reglas sintácticas y gramaticales del idioma.
- Complejidad se refiere a la riqueza léxica y estructura morfosintáctica empleada en la redacción.
- Fernández-Huerta mide el promedio de las sílabas por palabra y el promedio de palabras por oración.
- Inflesz evalúa la legibilidad del texto tomando en cuenta la cantidad de sílabas por palabra y frases.
- Coherencia representa la estructura semántica de un texto.

Cada una de estas métricas se calculó tanto para el corpus original (sufijo “_C”) como para las respuestas generadas por el modelo de lenguaje (sufijo “_N”).

Tabla 4: Fragmento del conjunto de datos creado (Parte 1).

Cohesión_C	Cohesión_N	Gramaticalidad_C	Gramaticalidad_N	Complejidad_C	Complejidad_N
0.118	0.119	1	0.968	1	0.839
0.134	0.102	1	0.97	1	0.848
0.112	0.095	1	0.867	1	0.846
0.03	0.175	0	0.871	1	0.871
0.076	0.222	0.957	1	0.87	0.84
0	0.126	1	0.84	1	0.84
0.026	0.018	1	1	1	0.793
0.102	0.065	1	0.8	1	0.8
0.023	0.069	1	0.862	0.8	0.862
0.174	0.096	0.905	0.87	1	0.81
0.125	0.147	0.988	0.716	0.905	0.716

Tabla 5: Fragmento del conjunto de datos creado (Parte 2).

FernandezHuerta_C	FernandezHuerta_N	Inflesz_C	Inflesz_N	Coherencia_C	Coherencia_N
0.348	0.691	0.802	0.647	1	1
0.96	0.179	0.927	1	1	1
0.859	0.779	0.825	1	0.03	1
0.561	0.812	0.522	0.03	1	1
0.805	0.004	0.772	0.921	0.921	1
0.805	0.302	1	0.802	1	1
0.973	0.302	0.984	1	1	1
0.759	0.726	0.726	1	1	1
0.94	0.94	0.909	1	1	1
0.743	0.911	0.704	0.879	1	1
0.978	0.673	0.938	0.63	0.872	1

El análisis comparativo entre las respuestas del corpus y las generadas por el modelo de lenguaje grande permitió observar diferencias consistentes en distintas métricas de calidad textual. El análisis comparativo entre las respuestas del corpus y las generadas por el modelo de lenguaje grande permitió observar diferencias consistentes en distintas métricas de calidad textual. Los resultados muestran que el modelo de lenguaje grande presenta mejor desempeño en gramaticalidad y legibilidad (Fernández-Huerta e Inflesz), mientras que el corpus tiende a ser superior en coherencia. Las métricas de cohesión y complejidad son similares en ambos casos. En general, el modelo genera textos correctos y legibles.

Para verificar si los datos asociados a cada métrica seguían una distribución normal, se aplicaron las pruebas de normalidad de *Kolmogorov-Smirnov* y *Shapiro-Wilk* utilizando el software SPSS. Como se muestra en la Tabla 6, los valores de significancia obtenidos para todas las métricas fueron menores al nivel de significancia $\alpha = 0.05$, Esto indica que los datos no presentan una distribución normal. En consecuencia, se optó por utilizar pruebas estadísticas no paramétricas para el análisis de las métricas evaluadas.

Tabla 6: Resultados de las pruebas de normalidad.

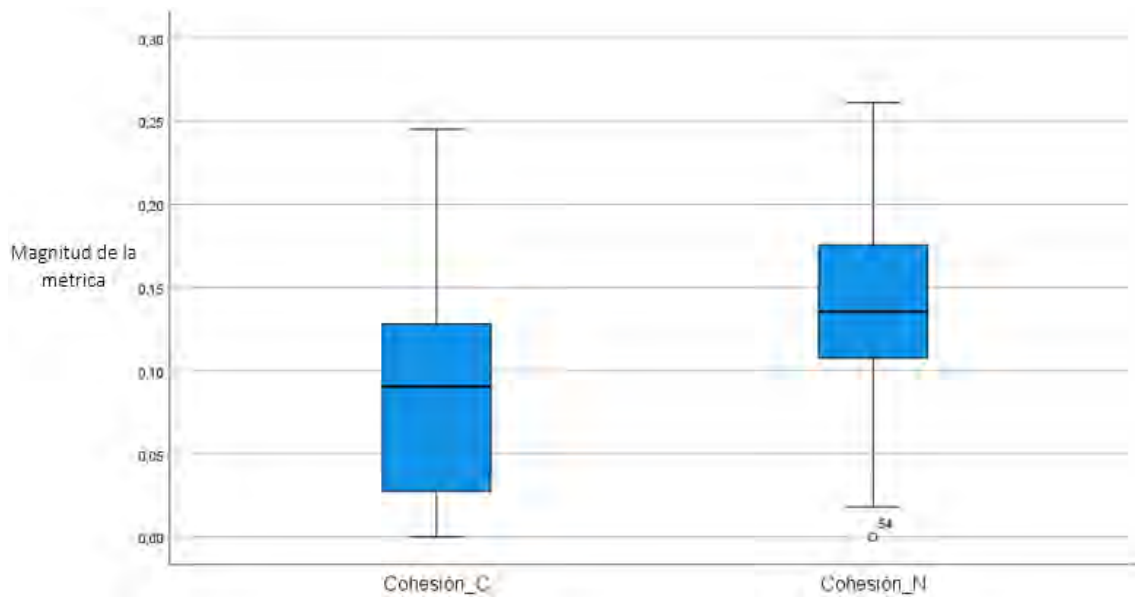
Variable	Kolmogorov-Smirnov			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
Cohesión	.094	100	.029	.943	100	.000
Gramaticalidad	.296	100	.000	.654	100	.000
Complejidad	.463	100	.000	.544	100	.000
fernandez_huerta	.223	100	.000	.806	100	.000
Inflesz	.200	100	.000	.830	100	.000
Coherencia	.317	100	.000	.579	100	.000

Los resultados de las pruebas de normalidad de Kolmogorov-Smirnov y Shapiro-Wilk indican que ninguna de las métricas evaluadas sigue una distribución normal, ya que en todos los casos los valores de significancia (Sig.) son menores al nivel de $\alpha = 0.05$.

Esto se observa tanto en las métricas de cohesión, gramaticalidad y coherencia como en las de legibilidad (Fernández-Huerta e Inflesz) y complejidad léxica. En consecuencia, se justifica el uso de pruebas estadísticas no paramétricas.

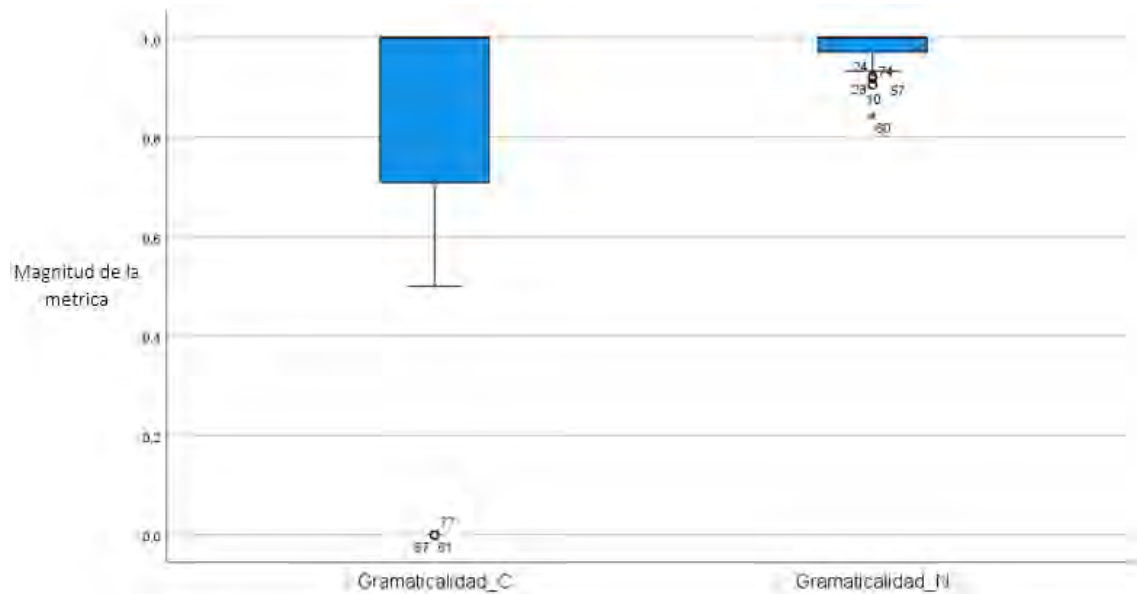
Se emplea la prueba de rangos con signo de Wilcoxon para para comparar los resultados entre el corpus y el modelo de lenguaje, analizar si existen diferencias significativas entre las métricas obtenidas por ambos modelos en función de las mismas preguntas. A continuación, se presenta la interpretación de los resultados:

Cohesión. Se muestra que el corpus presenta mayor dispersión, lo que indica una mayor variabilidad en sus respuestas. El MLG muestra una mayor distribución concentrada y valores medianos ligeramente superiores, lo que sugiere una mayor consistencia en esta métrica.



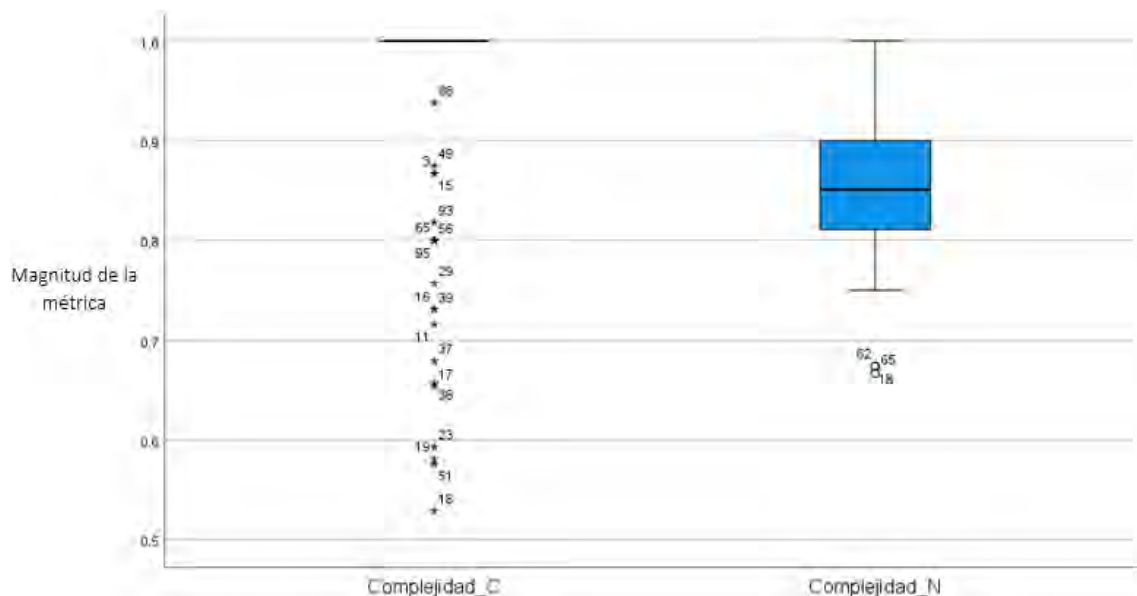
Boxplot de cohesión.

Gramaticalidad. Se evidencia muestra que las respuestas generadas por el MLG presentan una distribución altamente concentrada y cercana al valor máximo de la métrica (1). En contraste, el corpus presenta mayor variabilidad y varios valores atípicos cercanos al valor mínimo.



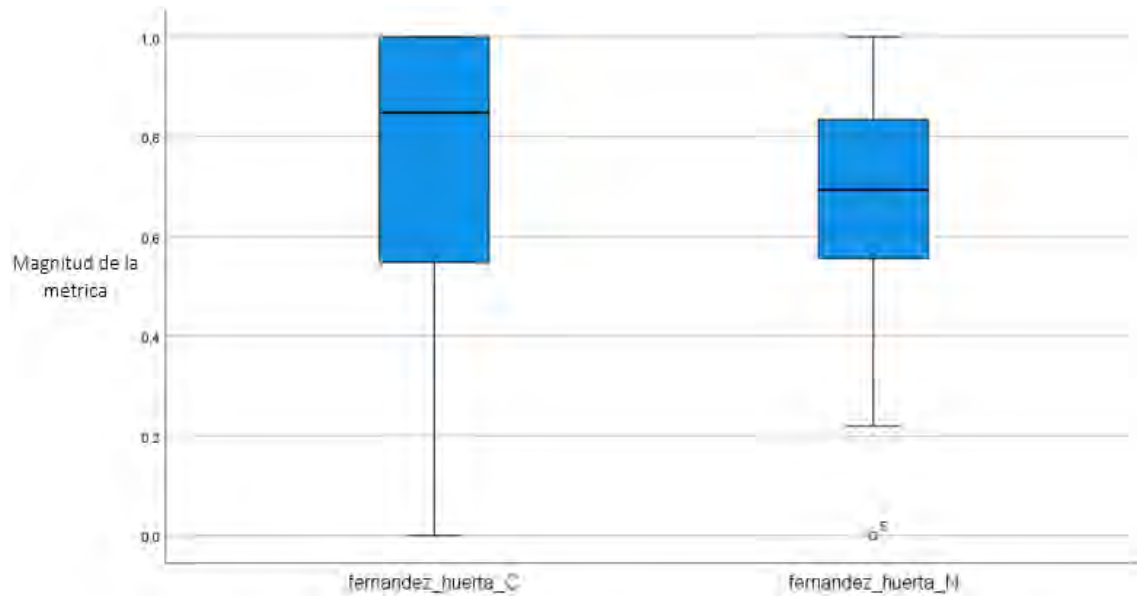
Boxplot de gramaticalidad.

Complejidad léxica. Se observa una mayor dispersión en los valores del corpus, con numerosos valores atípicos y una mediana elevada, lo que indica una complejidad más variable. Mientras que el MLG presenta una distribución más homogénea, con valores centrados y sin presencia notable de atípicos extremos.



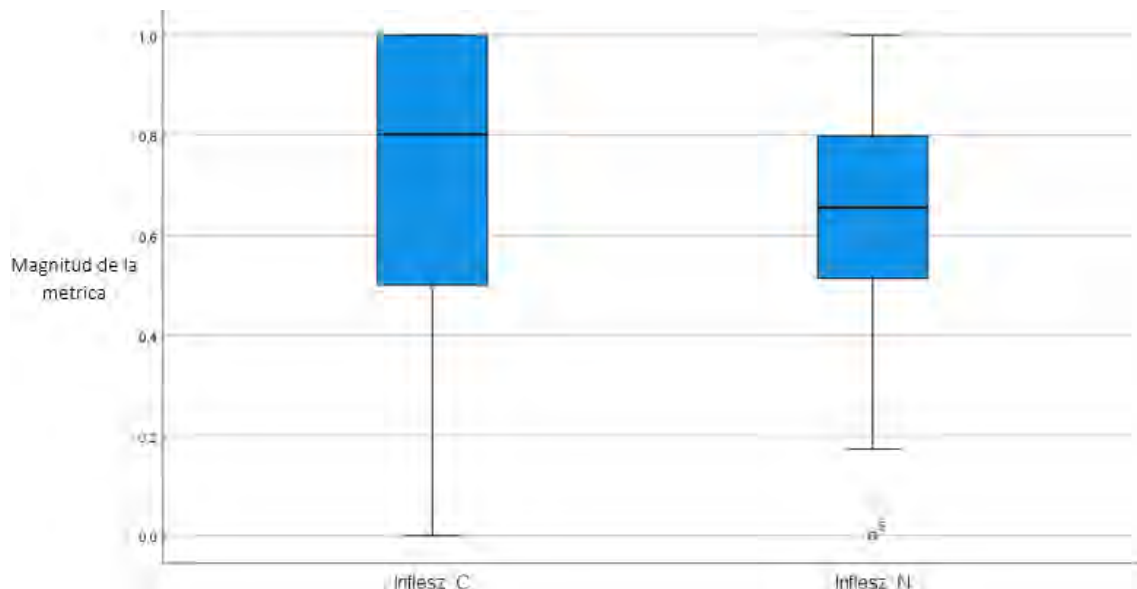
Boxplot de complejidad léxica.

Lectorabilidad de Fernández-Huerta. El corpus exhibe una mayor variabilidad en los puntajes, abarcando desde valores mínimos hasta máximos. En contraste, el MLG muestra una distribución más centrada y balanceada, lo que indica mayor uniformidad en la legibilidad de sus respuestas.



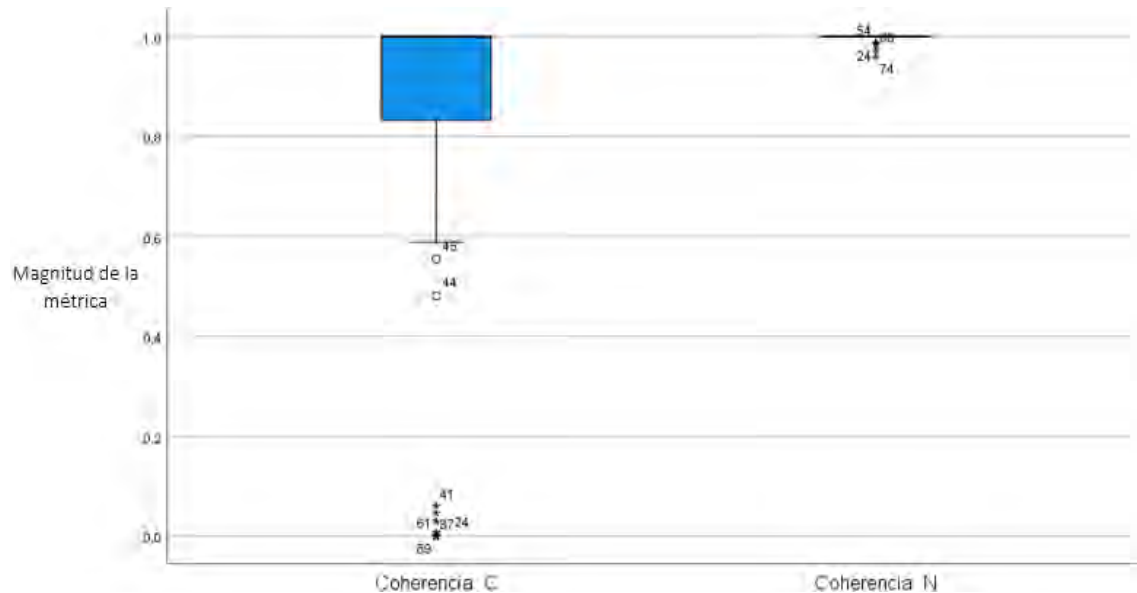
Boxplot de lectorabilidad de Fernández-Huerta.

Legibilidad Inflesz. Al igual que en Fernández-Huerta, el corpus muestra una mayor dispersión en los valores de legibilidad. El MLG presenta una distribución más concentrada, aunque con una mediana ligeramente inferior, sugiriendo una redacción más constante pero con menor legibilidad promedio.



Boxplot de legibilidad Inflesz.

Coherencia. Las respuestas del corpus presentan mayor dispersión, con presencia de valores bajos y múltiples atípicos. Mientras que el MLG muestra una distribución más uniforme y cercana al valor máximo, lo que indica una mayor coherencia en sus respuestas.



Boxplot de coherencia.

Los resultados obtenidos evidencian diferencias significativas entre los modelos evaluados, destacando las métricas de cohesión, gramaticalidad, complejidad y coherencia. El MLG demostró un desempeño más consistente, con distribuciones más homogéneas y valores centrales más elevados. Por lo que se respalda la superioridad del MLG en términos de calidad textual, al generar respuestas más claras, estables y entendibles. En consecuencia, se perfila como la opción más adecuada y prometedora para el desarrollo del AEH.

5.2. Diseño experimental

5.2.1. Metas, hipótesis y variables

Metas

- Identificar que métricas de usabilidad y funcionalidad representan de mejor manera el conjunto de datos.
- Obtener una ecuación matemática explícita que calcula la Calidad de un AEH.

Variables

- Seis métricas para medir calidad de texto.
- Seis campos que contiene preguntas relacionadas a la percepción de los usuarios del uso posterior del AEH.

5.2.2. Diseño

El diseño del experimento es de tipo cuasi-experimental, ya que no se realizó una asignación completamente aleatoria de los usuarios, sino que se seleccionaron con base en las aptitudes necesarias para la evaluación. Además, se considera un diseño

transversal, dado que la recolección de datos se llevará a cabo en un único momento, sin realizar mediciones repetidas a lo largo del tiempo.

5.2.3. Sujetos

Los sujetos de estudio seleccionados para este experimento son estudiantes de nivel maestría con conocimientos en áreas relacionadas con la computación, como ingeniería en computación, ciencias de la computación e inteligencia artificial. Sin embargo, no tenían conocimiento previo sobre el desarrollo, funcionamiento o intención del AEH. Su participación se limitó a responder las preguntas formuladas durante la evaluación.

5.2.4. Objetos

Los objetos utilizados en esta investigación se pueden observar en la Tabla 7.

Tabla 7: Objetos de la investigación

Objeto	Propósito	Enfoque	Perspectiva	Contexto
Estudiantes de nivel maestría	Evaluación	Uso del AEH	Investigador	Estudiantes de maestría con formación en áreas de computación, inteligencia artificial y disciplinas afines

5.2.5. Instrumentos

Se consideraron los datos recopilados a través de la API del servidor del AEH, los cuales proporcionan información sobre las métricas de calidad de texto. Además, estos datos fueron complementados con los valores obtenidos a partir del cuestionario aplicado a los usuarios después de su experiencia con el AEH. Esta combinación de fuentes permitió una evaluación integral, considerando tanto métricas objetivas extraídas del sistema como percepciones subjetivas de los participantes, presentada en la Tabla 8.

Tabla 8: Adquisición de los datos

Nombre de la tabla	Descripción de los datos	No. Atributos
Métricas	Métricas de calidad de texto	20 atributos
Percepciones de usabilidad	Preguntas relacionadas a la percepción de los usuarios de la calidad del AEH	20 atributos

5.2.6. Proceso de adquisición de datos

Los datos que conforman el conjunto de datos fueron obtenidos mediante pruebas realizadas con los sujetos de estudio. La dinámica consistió en que cada participante formulara 8 preguntas o dudas relacionadas con el contenido educativo de su elección. Para orientar el proceso, se les proporcionó un único ejemplo sobre los temas que podían y no podían abordar, sin restricciones adicionales en la formulación de sus preguntas.

Posteriormente, se aplicó un cuestionario diseñado para evaluar la experiencia de los usuarios con NadIA 2.0 (AEH). Este cuestionario constaba de 25 preguntas enfocadas en distintos aspectos clave, incluyendo su familiaridad con el uso de sistemas de Inteligencia Artificial, el nivel de dominio sobre el tema de la pregunta o duda planteada, su grado de satisfacción con el uso de NadIA 2.0, la efectividad percibida del asistente, la calidad de las respuestas proporcionadas y su intención de uso posterior. Esta metodología permitió recopilar tanto datos objetivos del desempeño del sistema como la percepción de los usuarios sobre su funcionalidad y usabilidad.

5.2.7. Proceso de análisis de datos

Para el procesamiento de los datos relacionados con las métricas de calidad de texto, se llevó a cabo una fase de limpieza que incluyó la eliminación de respuestas incompletas o erróneas generadas por los usuarios. Asimismo, se descartaron las columnas que contenían las preguntas formuladas por los participantes y las respuestas proporcionadas por NadIA 2.0, asegurando que únicamente se conservaran los datos relevantes para el análisis.

En cuanto a la información obtenida a partir del cuestionario, se eliminó la columna de hora y fecha, se registraron las preguntas de cada usuario junto con sus respectivas calificaciones, estableciendo una relación con el cálculo de las métricas correspondientes. Finalmente, los datos procesados fueron almacenados en hojas de cálculo (Excel), permitiendo su posterior tratamiento en las siguientes etapas de la investigación.

5.2.8. Proceso de validación

Para garantizar la fiabilidad de los datos recopilados, se llevó a cabo un proceso de validación el cual es el siguiente:

- Se eliminaron respuestas incompletas o erróneas registradas en el sistema, asegurando que solo se conservarán datos válidos para el análisis.
- Se descartaron columnas irrelevantes, como las preguntas formuladas por los usuarios y las respuestas proporcionadas por NadIA 2.0, manteniendo únicamente las métricas de calidad de texto y las calificaciones obtenidas del cuestionario.
- Se optó por calcular un promedio global de las métricas de calidad de texto de las respuestas, permitiendo una representación más homogénea de los datos.
- Se verificó la ausencia de valores nulos en el conjunto de datos.

5.3. Ejecución del experimento

5.3.1. Muestreo

Inicialmente, se obtuvieron dos conjuntos de datos. El primero corresponde a las métricas de calidad de texto generadas por la API del AEH. En la Figura 10 se presenta un fragmento de este conjunto de datos, mientras que en la Figura 11 se describen las variables incluidas en el análisis.

El segundo conjunto de datos se obtuvo a partir del cuestionario aplicado a los usuarios tras su interacción con NadIA 2.0. Este conjunto incluye todas las preguntas formuladas a los participantes, permitiendo evaluar distintos aspectos de su

	cohesion	gramaticalidad	complejidad	huerta	inflesz	coherencia
0	0.167	0.96	0.960	0.660	0.622	1.0
1	0.140	1.00	0.900	0.522	0.481	1.0
2	0.182	1.00	0.886	0.587	0.541	1.0
3	0.292	1.00	0.911	0.269	0.213	1.0
4	0.215	1.00	0.837	0.582	0.539	1.0

Figura 10: Fragmento del conjunto de datos de las métricas de calidad de texto.

	cohesion	gramaticalidad	complejidad	huerta	inflesz	coherencia
count	150.000000	150.000000	150.000000	150.000000	150.000000	150.000000

Figura 11: Cantidad de variables del conjunto de datos de las métricas de calidad de texto.

experiencia con el sistema. En la Figura 12 se presenta un fragmento del conjunto de datos, mientras que en la Figura 23 se detallan las variables consideradas en el análisis.

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16	R17	R18	R19	R20	R21	R22	R23	R24	R25
0	5	5	5	5	5	5	5	1	5	5	1	5	1	1	1	5	5	5	5	5	1	1	5	1	5
1	4	3	2	4	2	4	5	2	4	5	1	4	1	1	2	4	4	4	5	5	2	1	4	2	4
2	4	2	4	4	4	3	3	3	4	4	2	5	2	1	1	5	5	5	5	5	1	1	4	2	4
3	4	3	3	3	2	4	4	2	4	3	2	4	3	2	2	4	4	4	4	3	3	3	4	2	4
4	5	2	5	3	5	5	5	1	5	5	1	5	1	1	1	5	5	5	5	5	1	1	5	1	5

Figura 12: Fragmento del conjunto de datos del cuestionario.

	R1	R2	R3	R4	R5	...	R21	R22	R23	R24	R25
count	20.000000	20.000000	20.000000	20.000000	20.000000	...	20.000000	20.000000	20.000000	20.000000	20.000000

Figura 13: Cantidad de variables del conjunto de datos del cuestionario.

Posteriormente, se optó por promediar las variables dentro de cada categoría de evaluación (familiaridad con el uso de sistemas de Inteligencia Artificial, nivel de dominio sobre el tema de la pregunta o duda planteada, entre otras). Esta decisión se tomó con el propósito de obtener un valor representativo y homogéneo para cada usuario, reduciendo la variabilidad individual en las respuestas y permitiendo un análisis más estructurado. Al consolidar la información en un único valor por usuario, se facilita la comparación y el procesamiento estadístico de los datos. En la Figura 14 se presenta el nuevo conjunto de datos generado tras este procedimiento.

	familiaridad_ia	dominio_campo	satisfaccion_nadia	efectividad_asistente	calidad_respuesta	uso_posterior
0	5.0	5.0	5.0	5.0	5.0	5.0
1	3.2	4.0	3.4	4.0	3.4	2.7
2	4.6	4.4	3.0	3.7	3.4	4.0
3	4.0	5.0	4.0	4.2	4.2	4.0
4	3.0	4.0	3.6	4.0	3.4	4.0

Figura 14: Fragmento del nuevo conjunto de datos del cuestionario.

Una vez obtenidos ambos conjuntos de datos, correspondientes a las métricas de calidad de texto y a los resultados del cuestionario, se procedió a fusionarlos en un único conjunto. Sin embargo, se identificó una diferencia en la dimensionalidad de los datos, particularmente en el conjunto de métricas de calidad de texto, lo que dificultaba su integración directa.

Para solucionar esta discrepancia, se decidió promediar las variables de las métricas de calidad de texto, permitiendo obtener un valor representativo por usuario en dimensiones como cohesión, gramaticalidad, complejidad, entre otras. Este enfoque facilitó la unificación de los datos y optimizó su interpretación en el análisis posterior. En la Figura 15, 16 y 17 se presenta el conjunto de datos resultante tras este procedimiento y sus características.

	cohesion	gramaticalidad	complejidad	huerta	...	satisfaccion_nadia	efectividad_asistente	calidad_respuesta	uso_posterior
0	0.210429	0.989571	0.911286	0.537571	...	5.0	5.0	5.0	5.0
1	0.194375	1.000000	0.875750	0.638250	...	3.4	4.0	3.4	2.7
2	0.215333	0.987000	0.851444	0.502778	...	3.0	3.7	3.4	4.0
3	0.198000	0.882778	0.752556	0.470889	...	4.0	4.2	4.2	4.0
4	0.200250	0.991375	0.879750	0.534750	...	3.6	4.0	3.4	4.0

Figura 15: Fragmento de conjunto de datos final.

	cohesion	gramaticalidad	complejidad	...	efectividad_asistente	calidad_respuesta	uso_posterior
count	20.000000	20.000000	20.000000	...	20.000000	20.000000	20.000000

Figura 16: Cantidad de datos por cada variable del conjunto de datos final.

```
Index(['cohesion', 'gramaticalidad', 'complejidad', 'huerta', 'inflesz',
      'coherencia', 'familiaridad_ia', 'dominio_campo', 'satisfaccion_nadia',
      'efectividad_asistente', 'calidad_respuesta', 'uso_posterior'],
      dtype='object')
```

Figura 17: Cantidad de variables del conjunto de datos final.

5.3.2. Validación de datos

Se verificó que el conjunto de datos fuera almacenado correctamente y que mantuviera una dimensionalidad adecuada para el análisis. Además, se comprobó la ausencia de datos faltantes y se procedió a la identificación y eliminación de valores atípicos generados por errores, asegurando la robustez de los datos procesados.

5.4. Análisis factorial confirmatorio

Para avanzar en la investigación, se llevaron a cabo pruebas de normalidad de Shapiro-Wilk con el objetivo de evaluar si el conjunto de datos sigue una distribución normal. Los resultados revelaron que cuatro variables no cumplen con este criterio: gramaticalidad, coherencia, dominio campo y calidad respuesta, mientras que las demás siguen una distribución normal.

Para establecer las relaciones entre las variables del estudio, se calcularon las correlaciones utilizando el coeficiente de correlación de Spearman. Este método es particularmente adecuado para evaluar la relación entre variables en presencia de posibles no linealidades y valores atípicos, proporcionando una medida robusta de la asociación entre ellas. La matriz de correlaciones obtenida se muestra en la Figura 18, donde se observa la intensidad y dirección de las asociaciones entre los distintos constructos evaluados.

Los resultados indican diversas correlaciones significativas entre las variables, destacando asociaciones positivas y negativas de diferente magnitud. El análisis de correlaciones permitió justificar la inclusión de estas variables en el modelo de ecuaciones estructurales, asegurando que las relaciones consideradas fueran respaldadas

	cohesion	gramaticalidad	complejidad	huerta	inflesz	coherencia	familiaridad_ia
cohesion	1.00000000	0.256120981	0.03157895	-0.70375940	-0.73233083	0.09946434	0.220801586
gramaticalidad	0.25612098	1.000000000	0.34350343	0.07156322	0.04745771	0.29895548	-0.221217801
complejidad	0.03157895	0.343503434	1.00000000	0.38496241	0.34736842	0.05967860	-0.114207717
huerta	-0.70375940	0.071563215	0.38496241	1.00000000	0.99248120	-0.29839301	-0.347191460
inflesz	-0.73233083	0.047457711	0.34736842	0.99248120	1.00000000	-0.25860727	-0.343384536
coherencia	0.09946434	0.298955481	0.05967860	-0.29839301	-0.25860727	1.00000000	0.120866087
familiaridad_ia	0.22080159	-0.221217801	-0.11420772	-0.34719146	-0.34338454	0.12086609	1.000000000
dominio_campo	-0.01383351	0.023484304	0.08914931	0.22979002	0.25591999	0.16266682	-0.059146548
satisfaccion_nadia	0.23343413	-0.023310986	0.10451136	0.06407996	0.05034854	-0.14128302	-0.012746258
efectividad_asistente	-0.19858145	-0.002304507	0.04063636	0.45313373	0.47230182	-0.06085681	-0.310954796
calidad_respuesta	-0.33956243	-0.292434351	-0.26048625	0.35351705	0.37910052	-0.36920455	-0.069477571
uso_posterior	0.19259830	-0.287135717	0.14962816	-0.18799436	-0.19336563	-0.10150741	0.002331073
	dominio_campo	satisfaccion_nadia	efectividad_asistente	calidad_respuesta	uso_posterior		
cohesion	-0.01383351	0.23343413	-0.198581447	-0.33956243	0.192598301		
gramaticalidad	0.02348430	-0.02331099	-0.002304507	-0.29243435	-0.287135717		
complejidad	0.08914931	0.10451136	0.040636358	-0.26048625	0.149628162		
huerta	0.22979002	0.06407996	0.453133727	0.35351705	-0.187994358		
inflesz	0.25591999	0.05034854	0.472301820	0.37910052	-0.193365625		
coherencia	0.16266682	-0.14128302	-0.060856810	-0.36920455	-0.101507412		
familiaridad_ia	-0.05914655	-0.01274626	-0.310954796	-0.06947757	0.002331073		
dominio_campo	1.00000000	0.19532700	0.607760300	0.10341122	0.151372735		
satisfaccion_nadia	0.19532700	1.00000000	0.631279949	0.64184513	0.558981891		
efectividad_asistente	0.60776030	0.63127995	1.000000000	0.64865633	0.289514956		
calidad_respuesta	0.10341122	0.64184513	0.648656327	1.00000000	0.455720306		
uso_posterior	0.15137274	0.55898189	0.289514956	0.45572031	1.000000000		

Figura 18: Matriz de correlación mediante el método de Spearman.

empíricamente. Este proceso de selección garantiza que los factores latentes definidos cuenten con fundamentos estadísticos adecuados, lo que a su vez respalda la validez del modelo propuesto.

5.4.1. Construcción del primer modelo de ecuaciones estructurales (MEE)

Se observa el modelo de ecuaciones estructurales formado a partir de las correlaciones de las variables observadas con anterioridad.

funcionalidad $\sim$ cohesion + gramaticalidad + complejidad +
 huerta + inflesz + coherencia
 usabilidad $\sim$ familiaridad_ia + dominio_campo + satisfaccion_nadia +
 efectividad_asistente + calidad_respuesta + uso_posterior

funcionalidad $\sim$ usabilidad

cohesion $\sim$ familiaridad_ia
 cohesion $\sim$ dominio_campo
 gramaticalidad $\sim$ uso_posterior
 complejidad $\sim$ dominio_campo
 complejidad $\sim$ calidad_respuesta
 inflesz $\sim$ efectividad_asistente
 inflesz $\sim$ calidad_respuesta
 coherencia $\sim$ familiaridad_ia

Para evaluar la adecuación del MEE propuesto, se calcularon diversos índices, los cuales permiten determinar en qué medida el modelo representa adecuadamente los datos observados. Particularmente se presentan los valores de Chi-cuadrado X^2 , grados de libertad, Índice de ajuste comparativo (CFI), Error cuadrático medio de aproximación (RMSEA), Criterio de información de Akaike (AIC) y Criterio de información Bayesiano (BIC).

Los resultados de los índices de bondad de ajuste se muestran en la Tabla 9.

Tabla 9: Índices de bondad de ajuste del modelo propuesto.

Índice	Valor
Chi ²	315.331
Grados de libertad (DF)	45
CFI	0.575
RMSEA	0.562
AIC	34.47
BIC	67.33

El valor de X^2 obtenido es de 315.331 con 45 grados de libertad. Esta prueba evalúa si la matriz de covarianza estimada por el modelo difiere significativamente de la matriz observada. Valores altos de X^2 sugieren una falta de ajuste, mientras que valores bajos indican una mejor representación de los datos.

El CFI, mide la mejora del modelo en comparación con un modelo base de independencia, tiene un valor de 0.575. Para que un modelo se considere con un buen ajuste, el CFI debe ser superior a 0.90, con valores cercanos o superiores a 0.95, indicando un ajuste óptimo. En este caso, el valor obtenido sugiere un ajuste deficiente del modelo, lo que indica que este podría no estar representando adecuadamente los datos.

El RMSEA es un índice basado en la parsimonia del modelo, penalizando la complejidad excesiva. Un RMSEA inferior a 0.05 indica un buen ajuste, mientras que valores entre 0.05 y 0.08 reflejan un ajuste razonable. En este estudio, el RMSEA obtenido fue de 0.562, lo que está muy por encima del umbral aceptable, lo que sugiere un ajuste inadecuado del modelo.

El AIC y BIC se utilizan para comparar modelos alternativos, penalizando la complejidad del modelo para evitar sobreajuste. En este caso, los valores obtenidos son $AIC = 34.47$ y $BIC = 67.33$. Aunque estos valores no tienen umbrales fijos de calidad, se pueden utilizar para comparar con otros modelos candidatos, seleccionando aquel con valores más bajos.

Estos resultados sugieren que el primer modelo propuesto no alcanza un buen ajuste adecuado a los datos, dado un bajo CFI y un elevado RMSEA. Lo que indica que se necesita realizar ajustes tanto de las relaciones entre las variables latentes y observadas.

5.4.2. Segundo modelo de ecuaciones estructurales

Tras la evaluación del modelo inicial y considerando la necesidad de mejorar los índices de ajuste, se propuso una nueva especificación del MEE presentado en la Ecuación 7.

$$\begin{aligned}
 \text{funcionalidad} &\sim \text{cohesion} + \text{coherencia} + \text{complejidad} + \text{inflesz} \\
 \text{usabilidad} &\sim \text{efectividad_asistente} + \text{calidad_respuesta} + \text{uso_posterior} \\
 \text{complejidad} &\sim \text{calidad_respuesta} \\
 \text{inflesz} &\sim \text{efectividad_asistente} \\
 \text{inflesz} &\sim \text{calidad_respuesta}
 \end{aligned}
 \tag{7}$$

Se realizaron ajustes en las relaciones entre variables observadas y latentes. En esta versión, la variable latente **funcionalidad** se modeló en función de **cohesión**, **coherencia**, **complejidad** e **inflesz**, mientras que **usabilidad** se explicó mediante **efectividad del asistente**, **calidad de respuesta** y **uso posterior**. Adicionalmente, se incorporaron correlaciones entre las variables **complejidad** y **calidad de respuesta**, **inflesz** y **efectividad del asistente**, así como entre **inflesz** y **calidad de respuesta**, con el objetivo de mejorar la parsimonia del modelo y capturar mejor las relaciones observadas en los datos. Para determinar la validez del modelo ajustado, se calcularon nuevamente los índices de ajuste. Los resultados se muestran en la Tabla 10, donde se muestra una notable mejora en comparación con la versión anterior:

Tabla 10: Índices de bondad de ajuste del segundo modelo.

Índice	Valor
Chi ²	59.993
Grados de libertad (DF)	10
CFI	0.843
RMSEA	0.513
AIC	30.00
BIC	47.92

Los resultados indican que el segundo modelo representa una mejora con respecto al primero; sin embargo, aún existen indicadores que sugieren la necesidad de ajustar nuevamente, particularmente en la reducción del RMSEA y el incremento del CFI.

5.4.3. Tercer modelo de ecuaciones estructurales

En este modelo se propuso ajustar las variables latentes y las relaciones entre los constructos, la variable latente **funcionalidad** se definió a partir de **cohesión** y **huerta**, mientras que **usabilidad** se explicó mediante **efectividad del asistente**, **calidad de respuesta**, **uso posterior** y **satisfacción con Nadia**. Adicionalmente, se estableció una relación estructural entre las variables latentes de **funcionalidad** y **usabilidad**, así como correlaciones entre **cohesión** y **calidad de respuesta**, **huerta** y **efectividad del asistente**, y **huerta** y **calidad de respuesta**, representado en la Ecuación 8.

$$\begin{aligned}
&\text{funcionalidad} \sim\sim \text{cohesion} + \text{huerta} \\
&\text{usabilidad} \sim\sim \text{efectividad_asistente} + \text{calidad_respuesta} + \text{uso_posterior} + \text{satisfaccion_nadia} \\
\\
&\text{funcionalidad} \sim \text{usabilidad} \\
&\text{cohesion} \sim\sim \text{calidad_respuesta} \\
&\text{huerta} \sim\sim \text{efectividad_asistente} \\
&\text{huerta} \sim\sim \text{calidad_respuesta}
\end{aligned}$$

(8)

Los resultados obtenidos se presentan en la Tabla 11.

Tabla 11: Índices de bondad de ajuste del tercer modelo.

Índice	Valor
Chi ²	5.538
Grados de libertad (DF)	5
CFI	0.996
RMSEA	0.075
AIC	31.45
BIC	47.38

Los resultados obtenidos demuestran que el tercer modelo representa un ajuste significativamente mejor con respecto a sus predecesores, evidenciado por un CFI cercano a 1 y un RMSEA dentro del rango aceptable. La reducción del X^2 y la mejora en los criterios de información sugieren que la especificación actual representa adecuadamente las relaciones entre las variables observadas y latentes.

5.4.4. Modelo final de ecuaciones estructurales

Tras varias iteraciones en la construcción y refinamiento del MEE, se desarrolló una versión final que optimiza significativamente los índices de ajustes. En esta especificación, la variable latente **funcionalidad** se definió a partir de **cohesión** y **huerta**, mientras que **usabilidad** se explicó mediante **efectividad del asistente**, **calidad de respuesta**, **uso posterior** y **satisfacción con Nadia**. Además, se establecieron correlaciones adicionales entre **cohesión** y **calidad de respuesta**, **cohesión** y **efectividad del asistente**, **huerta** y **efectividad del asistente**, así como **huerta** y **calidad de respuesta**, lo que permitió capturar con mayor precisión las relaciones entre los constructos observados, representado en la Ecuación 9.

$$\begin{aligned}
&\text{funcionalidad} \sim\sim \text{cohesion} + \text{huerta} \\
&\text{usabilidad} \sim\sim \text{efectividad_asistente} + \text{calidad_respuesta} + \text{uso_posterior} + \text{satisfaccion_nadia} \\
\\
&\text{funcionalidad} \sim \text{usabilidad} \\
&\text{cohesion} \sim\sim \text{calidad_respuesta} \\
&\text{cohesion} \sim\sim \text{efectividad_asistente} \\
&\text{huerta} \sim\sim \text{efectividad_asistente} \\
&\text{huerta} \sim\sim \text{calidad_respuesta}
\end{aligned}$$

(9)

En la Tabla 12 se muestran los resultados de los índices de ajuste de bondad.

Tabla 12: Índices de bondad de ajuste del modelo final.

Índice	Valor
Chi ²	3.748
Grados de libertad (DF)	4
CFI	1.002
RMSEA	0.000
AIC	33.63
BIC	50.55

El X^2 , con un valor de 3.748 y 4 grados de libertad, es extremadamente bajo, indicando que la discrepancia entre la matriz de covarianza estimada y la observada es mínima. Un valor de X^2 cercano a los grados de libertad es un fuerte indicador de un modelo bien ajustado.

El CFI, que evalúa la mejora del modelo en relación con un modelo base de independencia, presenta un valor superior a 1 (1.002). Este resultado, aunque inusual, ocurre debido a la relación entre los grados de libertad y la varianza explicada, y confirma que el modelo ajusta de manera excelente a los datos, superando ampliamente el umbral recomendado de 0.95.

El RMSEA, que penaliza modelos con complejidad innecesaria, obtuvo un valor de 0.000, lo que indica una ausencia total de error de aproximación. Este resultado representa el mejor ajuste posible, ya que un RMSEA inferior a 0.05 se considera excelente y valores iguales a 0.00 indican una correspondencia perfecta entre la matriz de covarianza del modelo y la matriz observada.

Los criterios de información AIC (33.63) y BIC (50.55) son competitivos en comparación con los modelos previos. Aunque en términos absolutos estos valores no determinan la calidad del ajuste, permiten comparar modelos alternativos, y en este caso evidencian que el modelo final logra un equilibrio óptimo entre parsimonia y ajuste.

También el análisis realizado permitió la representación gráfica del MEE final, la cual refleja de manera clara las relaciones entre las variables latentes y observadas. El diagrama de la Figura 19 obtenido ilustra la estructura definitiva del modelo.

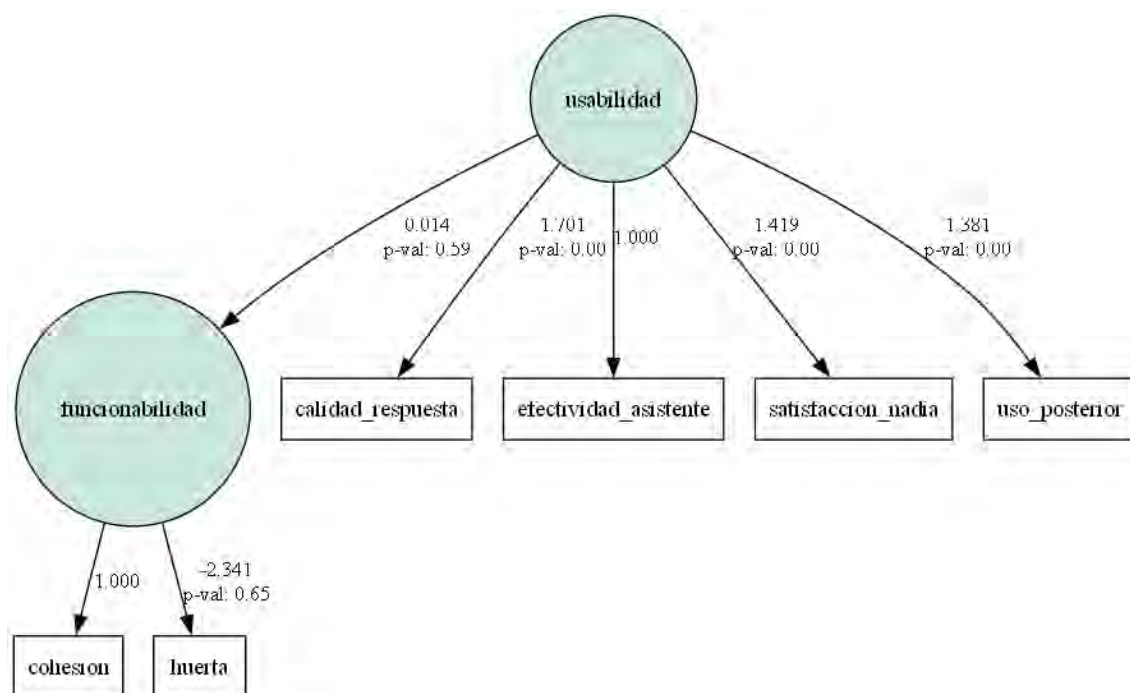


Figura 19: Representación gráfica del MEE final.

Los coeficientes de regresión proporcionados en el diagrama indican la fuerza de las relaciones entre las variables. En particular, se observa que **efectividad del asistente**, **satisfacción con Nadia** y **uso posterior** presentan efectos significativos sobre **usabilidad** ($p - val < 0.001$), lo que sugiere que estos factores desempeñan un papel clave en la percepción de la utilidad del sistema evaluado. Además, la relación entre **funcionalidad** y **calidad de respuesta** no resultó significativa ($p - val = 0.59$), lo que sugiere que esta asociación podría no ser determinante.

Como conclusión, el modelo final de ecuaciones estructurales representa la mejor versión entre todas las propuestas previas. Este ajuste final demuestra una mejora estadísticamente significativamente, respaldada por una sólida fundamentación teórica. Los resultados obtenidos confirman la validez del modelo y su aplicabilidad en contextos similares, estableciéndolo como una herramienta confiable para la evaluación de la funcionalidad y usabilidad de sistemas basados en modelos de lenguaje grandes.

5.5. Obtención del indicador de calidad del asistente educativo holográfico

Para la evaluación de la calidad del asistente educativo holográfico se requiere un enfoque riguroso que permita cuantificar su funcionalidad y usabilidad de manera objetiva. Para ello se desarrolló un indicador de calidad basado en el modelo de ecuaciones estructurales final, el cual permitió identificar las variables más relevantes. Sin embargo, con el propósito de obtener una expresión matemática explícita se aplicaron técnicas de computación genética, especialmente regresión simbólica, con el fin de derivar una ecuación que relacione las variables de entrada con una medida cuantitativa de calidad.

5.5.1. Fundamentación y aplicación del modelo de ecuaciones estructurales

El MEE final permitió estructurar las relaciones entre las métricas de calidad de texto y la percepción de los usuarios sobre su usabilidad del AEH. A través del análisis factorial confirmatorio, se identificaron las variables con mayor impacto, asegurando una alta significancia estadística. Lo que se representa como la base conceptual y teórica sobre la cual se definiría el indicador de calidad.

5.5.2. Uso de computación genética

Para obtener una ecuación matemática representativa del indicador de calidad del AEH, se empleó regresión simbólica, una técnica dentro de la computación genética basada en la programación evolutiva. Este enfoque permite encontrar estructuras matemáticas óptimas sin imponer una forma funcional predeterminada. La regresión simbólica se aplicó utilizando el conjunto de entrenamiento derivado de los datos procesados en el MEE, donde las variables predictoras correspondieron a las métricas de calidad de texto y las calificaciones obtenidas en el cuestionario aplicado a los usuarios.

5.5.3. Proceso de obtención de la ecuación

El proceso de generación del indicador de calidad fue a través de regresión simbólica.

Los resultados obtenidos se presentan en la Figura 20, donde el proceso de optimización se llevó a cabo a partir de 15 generaciones, observándose una reducción progresiva del error de ajuste (*Fitness*) en cada iteración. En la generación 0, la longitud promedio de las ecuaciones generadas fue de 9.39, con un error inicial de 16.0571 en la población. Conforme avanzó la evolución, la longitud de las ecuaciones disminuyó, logrando simplificaciones más eficientes sin sacrificar precisión.

Population Average			Best Individual			
Gen	Length	Fitness	Length	Fitness	OOB Fitness	Time Left
0	9.39	16.0571	15	0.200205	N/A	4.35s
1	6.15	3.18747	10	0.117003	N/A	4.29s
2	5.15	2.62389	8	0.106503	N/A	3.66s
3	3.91	2.56194	8	0.0976023	N/A	3.51s
4	4.35	2.62889	8	0.0976023	N/A	3.24s
5	5.22	3.06226	4	0.105453	N/A	2.84s
6	5.71	2.52131	4	0.105453	N/A	2.86s
7	5.38	3.42087	4	0.105453	N/A	2.22s
8	4.47	3.09792	4	0.105453	N/A	2.11s
9	3.99	2.62368	4	0.105453	N/A	1.79s
10	4.01	2.4425	4	0.105453	N/A	1.24s
11	3.99	2.56124	4	0.105453	N/A	1.02s
12	4.04	2.3935	4	0.105453	N/A	0.72s
13	3.96	2.54464	4	0.105453	N/A	0.44s
14	4.02	2.48173	4	0.105453	N/A	0.00s

Ecuación simbólica encontrada para el indicador de calidad del AEH:
`sqr(mul(X0, X1))`

Figura 20: Resultados obtenidos de la primera regresión simbólica.

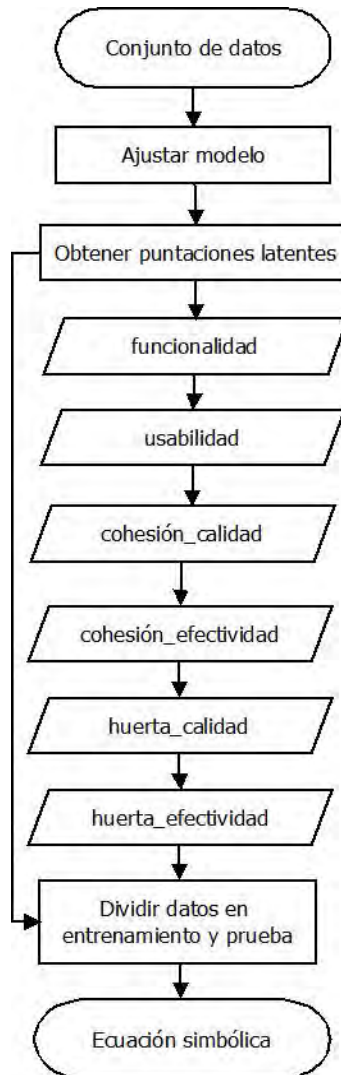
Desde la generación 6, se alcanzó una convergencia en la solución, donde el mejor individuo presentó un fitness de 0.105453 con una longitud reducida a 4 términos, lo que indica que el modelo identificó una estructura matemática estable y óptima. Esta convergencia demuestra que el algoritmo encontró una ecuación que balancea precisión y simplicidad, evitando sobreajustes o modelos excesivamente complejos. La ecuación resultante se representa de la siguiente manera:

$$Indicador_calidad_AEH = \sqrt{X_0 \times X_1} \quad (10)$$

Donde si hacemos la sustitución con base en al orden de las puntuaciones de variables predictoras, la ecuación quedaría de la siguiente manera:

$$Indicador_calidad_AEH = \sqrt{funcionalidad \times usabilidad} \quad (11)$$

Esta ecuación representa el indicador de calidad como la raíz cuadrada del producto entre las dos variables latentes. La selección de esta ecuación final sugiere que existe una relación multiplicativa entre estas variables y que su impacto en la calidad percibida del AEH es mejor representado a través de una transformación no lineal, como la raíz cuadrada.



Obtención del indicador de calidad.

En la Figura 21 se muestra el proceso de regresión simbólica que avanzó a lo largo de 14 generaciones, mostrando una disminución progresiva en el error de ajuste (*Fitness*). Inicialmente, en la generación 0, el error promedio en la población fue de 3.75218, con una longitud media de 9.15 términos. A medida que el modelo evolucionó, se logró una reducción significativa del error, alcanzando 0.0684648 en la mejor ecuación a partir de la generación 6, lo que indica que el algoritmo convergió rápidamente a una solución óptima.

El tamaño de la ecuación también se redujo de 9 términos en la generación 0 a solo 4 términos en la generación 6 en adelante, lo que demuestra que el modelo encontró una representación matemática más simple y eficiente para el indicador de calidad del AEH sin comprometer su precisión.

Population Average			Best Individual			
Gen	Length	Fitness	Length	Fitness	OOB Fitness	Time Left
0	9.15	3.75218	9	0.0616657	N/A	21.82s
1	6.75	1.6525	10	0.0586709	N/A	20.85s
2	5.78	1.28911	9	0.0365018	N/A	18.40s
3	4.90	1.1717	9	0.0365018	N/A	18.04s
4	4.40	0.863149	9	0.0365018	N/A	15.83s
5	4.00	0.50925	4	0.0684648	N/A	13.91s
6	4.02	0.49322	4	0.0684648	N/A	12.85s
7	4.01	0.465904	4	0.0684648	N/A	11.24s
8	4.05	0.487698	4	0.0684648	N/A	9.11s
9	4.00	0.479341	4	0.0684648	N/A	7.66s
10	4.04	0.478192	4	0.0684648	N/A	6.33s
11	3.97	0.49711	4	0.0684648	N/A	4.68s
12	4.02	0.490289	4	0.0684648	N/A	3.15s
13	4.00	0.479196	4	0.0684648	N/A	1.66s
14	4.02	0.483115	4	0.0684648	N/A	0.00s

Ecuación simbólica encontrada para el indicador de calidad del AEH:
`add(sqrt(X5), X2)`

Figura 21: Resultados obtenidos de la segunda regresión simbólica.

El algoritmo de regresión simbólica encontró la siguiente ecuación como la mejor representación para el indicador de Calidad del AEH:

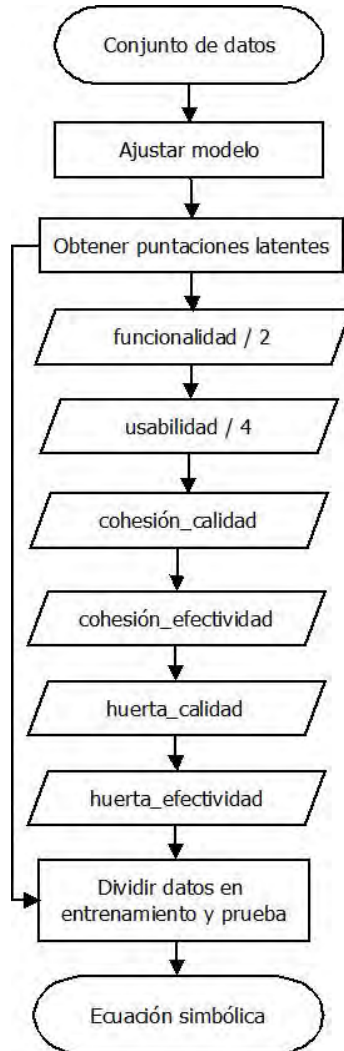
$$Indicador_calidad_AEH = \sqrt{X_5} + X_2 \quad (12)$$

Donde si hacemos la sustitución en base al orden de las puntuaciones de variables predictoras, la ecuación quedaría de la siguiente manera:

$$Indicador_calidad_AEH = \sqrt{huerta_calidad} + cohesion_calidad \quad (13)$$

La inclusión de una raíz cuadrada en *huerta_calidad* sugiere que esta variable tiene una relación decreciente con el indicador. Es decir, a medida que *huerta_calidad* aumenta, su impacto en el indicador crece, pero a un ritmo menor debido a la transformación de raíz cuadrada.

La presencia de *cohesion_calidad* sin transformaciones indica que esta variable tiene un impacto directo y proporcional en el indicador. La ecuación obtenida es mucho más simple que modelos más complejos de generaciones anteriores. Al reducir la longitud de la expresión matemática sin afectar significativamente el *fitness*, se garantiza un modelo más interpretable y eficiente para la evaluación del AEH.



Obtención del indicador de calidad.

Dentro de esta iteración del modelo de regresión simbólica, los resultados de la Figura 22 muestran una mejora significativa con respecto a los experimentos anteriores. La optimización alcanzada permite obtener una ecuación más representativa y con mejor ajuste estadístico, lo que asegura un indicador más robusto para la evaluación de la calidad del AEH.

El modelo evolucionó a lo largo de las generaciones, reduciendo progresivamente el error de ajuste y optimizando la estructura matemática de la ecuación generada. Inicialmente, en la generación 0, la población presentó un error promedio de 0.77922 con ecuaciones de longitud media de 12.51 términos. A medida que avanzó el entrenamiento, la ecuación se fue optimizando, alcanzando su mejor versión en la generación 13, con un error de 0.038462 y una longitud reducida a solo 8 términos, lo que indica una mejora significativa tanto en precisión como en simplicidad estructural.

warnings.warn(Population Average Best Individual						
Gen	Length	Fitness	Length	Fitness	OOB Fitness	Time Left
0	12.51	9.77922	9	0.0792276	N/A	23.86s
1	8.99	1.60337	26	0.038336	N/A	24.80s
2	9.72	1.2416	26	0.038336	N/A	46.26s
3	11.26	1.16872	27	0.0348331	N/A	40.95s
4	11.03	1.1694	27	0.0367735	N/A	39.34s
5	8.73	1.20954	24	0.033101	N/A	32.04s
6	8.05	1.0713	27	0.033101	N/A	27.97s
7	6.75	0.936681	25	0.033101	N/A	16.90s
8	6.17	0.981412	23	0.0315327	N/A	8.49s
9	6.92	1.11105	25	0.0288218	N/A	7.14s
10	7.94	1.03761	29	0.0288218	N/A	5.93s
11	7.92	0.924983	26	0.0324579	N/A	4.27s
12	8.01	0.882967	8	0.0384632	N/A	2.84s
13	7.95	0.872745	8	0.0384632	N/A	1.50s
14	7.97	0.89088	8	0.0384632	N/A	0.00s

Ecuación simbólica encontrada para el indicador de calidad del AEH:
 $\log(\text{mul}(\text{add}(\text{sqrt}(X_5), \text{sqrt}(X_2)), X_1))$

Figura 22: Resultados obtenidos de la regresión simbólica final.

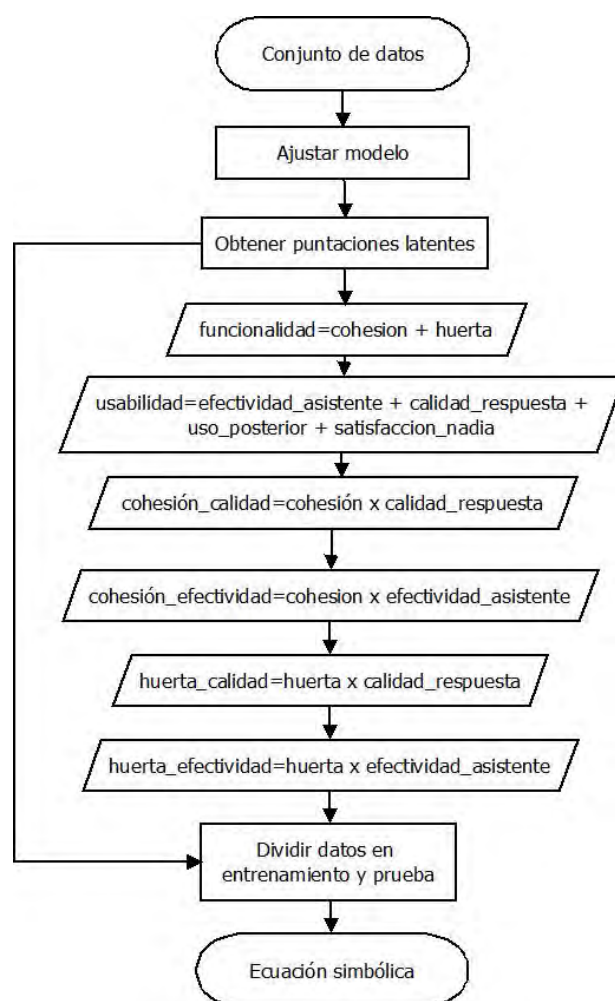
La ecuación final obtenida para el indicador de calidad del AEH es la siguiente:

$$\text{Indicador_calidad_AEH} = \log((\sqrt{X_5} + \sqrt{X_2}) \times X_1) \quad (14)$$

Donde si hacemos la sustitución en base al orden de las puntuaciones de variables predictoras, la ecuación quedaría de la siguiente manera:

$$\text{Indicador_calidad_AEH} = \log((\sqrt{\text{huerta_calidad}} + \sqrt{\text{cohesion_calidad}}) \times \text{usabilidad}) \quad (15)$$

El error de ajuste es inferior a los valores obtenidos en las anteriores iteraciones, lo que indica que la ecuación predice con mayor exactitud la calidad del AEH. La inclusión de operaciones logarítmicas y raíces cuadradas permite capturar relaciones más complejas entre las variables, a diferencia de ecuaciones previas que solo utilizaban sumas y multiplicaciones, además de que la ecuación final se optimizó a solo 8 términos, manteniendo un equilibrio entre precisión y parsimonia y se estabilizó con valores similares de *fitness*, lo que indica que la solución encontrada es robusta y no está sobreajustada a los datos de entrenamiento.



Obtención del indicador de calidad final.

5.6. Validación de la hipótesis

Tras aplicar la ecuación obtenida para el indicador de calidad del AEH (Ecuación 15) al conjunto de datos final, se obtuvieron los resultados que se presentan en la Tabla 13 y 14.

Tabla 13: Resultados del cálculo del indicador de calidad del AEH.

ID del usuario	1	2	3	4	5	6	7	8	9	10
Resultado del indicador	1.9028	1.6750	1.6986	1.7699	1.6795	1.8115	1.8033	1.7132	1.6932	1.7379

Tabla 14: Resultados del cálculo del indicador de calidad del AEH (continuación).

ID del usuario	11	12	13	14	15	16	17	18	19	20
Resultado del indicador	1.7815	1.7355	1.7251	1.74020	1.7425	1.7694	1.8185	1.6584	1.6141	1.8253

Capítulo 6

6. Conclusiones

Al calcular el promedio de todos los valores obtenidos en las tablas anteriores, se obtiene un resultado: **1.7448**. Aunque este valor proporciona una medición cuantitativa, su interpretación directa no resulta intuitiva. Para facilitar su comprensión, se normaliza el valor considerando los máximos y mínimos observados, permitiendo su expresión en una escala porcentual. Tras este ajuste, el resultado obtenido es **75.29 %**.

Los resultados obtenidos en esta investigación aportan una contribución significativa en la evaluación de la calidad de los Asistentes Educativos Holográficos (AEH's) basados en Modelos de Lenguaje Grandes (MLG). A través de una metodología estructurada, se logró desarrollar y validar un Modelo de Ecuaciones Estructurarles (MEE) que permite medir cuantitativamente la calidad del AEH, integrando métricas del Procesamiento del Lenguaje Natural (PLN) y aspectos de usabilidad.

Uno de los principales logros fue la formulación del indicador de calidad mediante la regresión simbólica, el cual permitió generar una expresión matemática que sintetiza de manera efectiva las dimensiones de usabilidad y funcionalidad del sistema. Este indicador proporcionó un valor cuantitativo que representa el desempeño general del AEH, facilitando su análisis y comparación.

El modelo validado alcanzó niveles óptimos de ajuste, evidenciando que las métricas empleadas explican adecuadamente la variabilidad en las métricas de calidad de texto y de la percepción de los usuarios. Además, se identificó que la efectividad del asistente, la calidad percibida de las respuestas, el uso posterior del sistema y la satisfacción con NadIA 2.0 son los principales predictores de la usabilidad.

Estos hallazgos respaldan el uso de enfoques cuantitativos para la evaluación de tecnologías educativas basadas en inteligencia artificial, aportando un marco metodológico robusto y replicable. Asimismo, se evidenció que modelos matemáticos simples pueden capturar de manera efectiva las relaciones clave entre variables sin comprometer su capacidad predictiva.

No obstante, el estudio presenta algunas limitaciones, como la cantidad de datos utilizados pueden no ser suficientes para generalizar los resultados en otros contextos educativos, además no se consideran otras variables que puedan influir en la percepción del sistema y la recolección de datos pudo haber estado sujeta a sesgos externos. Sin embargo, el enfoque propuesto resulta válido dentro del contexto de la investigación.

Como línea futura, se recomienda ampliar el estudio mediante la recolección de los datos en entornos educativos más diversos y con una muestra más amplia de sujetos, así como incorporar nuevas características y subcaracterísticas de calidad definidas en la norma ISO/IEC 9126. También se sugiere explorar nuevas técnicas avanzadas de aprendizaje automático que refuercen el indicador de calidad y realizar estudios comparativos con otros métodos de evaluación para validar la robustez del modelo propuesto. Adicionalmente, se plantea el uso de instrumentos estandarizados, como el PSSUQ (Post-Study System Usability Questionnaire) y el CSUQ (Computer System Usability Questionnaire), con el fin de complementar la evaluación subjetiva de la calidad percibida por los usuarios, ampliando el alcance de esta metodología para la evaluación de sistemas de inteligencia artificial aplicados a la educación.

Referencias

- Ahmad, A. S., Alomaier, A. T., Elmahal, D. M., Abdelfatah, R. F., & Ibrahim, D. M. (2021). EduGram: Education Development Based on Hologram Technology. *International Journal of Online & Biomedical Engineering*, 17(14).
- Alejandres, H. (2017). Evaluación centrada en el usuario de sistemas de recomendación sensibles al contexto: efecto de interfaces multimodales interactivas y esquemas de explicación en la experiencia del usuario. *Tesis de doctorado, Centro Nacional de Investigación y Desarrollo Tecnológico*. Repositorio Institucional del Tecnológico Nacional de México (RI - TecNM).
- Ali, G., Le, H.-Q., Kim, J., Hwang, S.-W., & Hwang, J.-I. (2019). Design of seamless multi-modal interaction framework for intelligent virtual agents in wearable mixed reality environment. *Proceedings of the 32nd International Conference on Computer Animation and Social Agents*, 47-52.
- Barrio Cantalejo, I., Simón-Lorda, P., Melguizo, M., Escalona, I., Marijuán, M., & Hernando, P. (2008). Validación de la Escala INFLESZ para evaluar la legibilidad de los textos dirigidos a pacientes Validation of the Escala INFLESZ to evaluate the readability of the texts directed to patient.
- Bartra, R., & Tirado Miranda, A. E. (2018). Dispositivo Holográfico Interactivo para Mejorar el Aprendizaje en Estudiantes de Primer Grado de Primaria de la IEP Jan Komensky.
- Cadena SER. (2024). *Los niños y niñas de los países más ricos tienen más sobrepeso, peores notas y más insatisfacción vital desde la pandemia* [Consultado en mayo de 2025]. <https://cadenaser.com/nacional/2025/05/14/los-ninos-y-ninas-de-los-paises-mas-ricos-tienen-mas-sobrepeso-peores-notas-y-mas-insatisfaccion-vital-desde-la-pandemia-cadena-ser/>
- Cerezo, R., Calderón, V., & Romero, C. (2019). A holographic mobile-based application for practicing pronunciation of basic English vocabulary for Spanish speaking children. *International Journal of Human-Computer Studies*, 124, 13-25. <https://doi.org/https://doi.org/10.1016/j.ijhcs.2018.11.009>
- Duguleană, M., Briciu, V.-A., Duduman, I.-A., & Machidon, O. M. (2020). A virtual assistant for natural interactions in museums. *Sustainability*, 12(17), 6958.
- Girish, P., Shivaji, S. C., Omkar, P., & Shweta, D. (2019). Holographic Artificial Intelligence Assistance. *International Journal of Computer Applications*, 975, 8887.
- Habboosh & Muhammad. (2022, enero). Education System in the Future: The Use of Hologram Technology. <https://www.skylineuniversity.ac.ae/knowledge-update/information-systems/education-system-in-the-future-the-use-of-hologram-technology>
- Hoon, L. N., & Shaharuddin, S. (2019). Learning effectiveness of 3D hologram animation on primary school learners. *Journal of visual art and design*, 11(2), 93-104.
- Huerta, S. (2010). Coherencia y cohesión. *Herencia: Estudios literarios, lingüísticos y creaciones artísticas*, 2(2), 76-80.
- Isaev, R., Gumerov, R., Esenalieva, G., Mekuria, R. R., & Doszhanov, E. (2023). HIVA: Holographic Intellectual Voice Assistant. *2023 17th International Conference on Electronics Computer and Computation (ICECCO)*, 1-6.
- Iturbe, A. (2023). Redes Neuronales Convolucionales y Recurrentes para la Extracción de Respuestas en Español. *Tesis de doctorado, Centro Nacional de Investigación y Desarrollo Tecnológico*. Repositorio Institucional del Tecnológico Nacional de México (RI - TecNM).
- Jaramillo, C. M. Z., & Pérez, Y. M. (2016). On the relationship of ISO/IEC 9126 metrics and the alpha states of the SEMAT kernel. *2016 4th International Conference in Software Engineering Research and Innovation (CONISOFT)*, 59-64.
- Kaya, N., & Bulut, S. (2022). Artificial intelligence and education: An overview. *Psychology & Psychological Research International Journal (PPRIJ)*.
- Keselj, A., Zubrinic, K., Milicevic, M., & Kuzman, M. (2023). An Overview of 3D Holographic Visualization Technologies and Their Applications in Education. *2023 46th MIPRO ICT and Electronics Convention (MIPRO)*, 454-459.
- Loh, N.-H., & Shukhaila, S. S. B. (2019). Corporate social responsibility (Csr) towards education: The application and possibility of 3d hologram to enhance cognitive skills of primary school learners. *International Journal of Business and Society*, 20(3), 1036-1047.
- Lolambean. (2023, julio). Diseño de contenido para la pantalla holográfica - Mixed Reality. <https://learn.microsoft.com/es-es/windows/mixed-reality/design/designing-content-for-holographic-display>

- López, R., Montejó-Ráez, A., Martínez-Santiago, F. J., & Díaz-Galiano, M. C. (2018). Legibilidad del texto, métricas de complejidad y la importancia de las palabras. *Procesamiento del lenguaje natural*, 61, 101-108.
- Makridakis, S., Petropoulos, F., & Kang, Y. (2023). Large Language Models: Their Success and Impact. *Forecasting*, 5 (3), 536-549.
- Makris, A., Boudi, A., Coppola, M., Cordeiro, L., Corsini, M., Dazzi, P., Andilla, F. D., Rozas, Y. G., Kamarianakis, M., Pateraki, M., et al. (2021). Cloud for holography and augmented reality. *2021 IEEE 10th international conference on cloud networking (CloudNet)*, 118-126.
- Mawardani, A., Dwandaru, W. S. B., Nugroho, M. P., et al. (2021). Multimedia Learning Module (MLM) with Hologram Simulation to Improve Students' Mathematical Representation Ability. *6th International Seminar on Science Education (ISSE 2020)*, 522-527.
- Muñoz, I. B., Bosque, I., & Gutiérrez-Rexach, J. (2009). *Fundamentos de sintaxis formal*. Ediciones Akal.
- Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 249-256.
- Orchard, J. (2023). Clarity of the Educational Assistant Role: A Look toward Policy and Practices. *Educ Res Appl*, 8, 220.
- Orcos, L., Jordán, C., & Magreñán, A. (2019). 3D visualization through the hologram for the learning of area and volume concepts. *Mathematics*, 7(3), 247.
- Orcos, L., & Magreñán, Á. A. (2018). The hologram as a teaching medium for the acquisition of STEM contents. *International Journal of Learning Technology*, 13(2), 163-177.
- Ribeiro, A. V., Godoy, G. C., Neto, L. B., & de Souza-Filho, M. P. (2018). Holografía y realidad virtual en la enseñanza de nanotecnología: Nuevos horizontes dirigido a educación secundaria. *MOMENTO*, (56E), 34-45.
- Rosales, R., Castañón-Puga, M., Lara-Rosano, F., Flores-Parra, J. M., Evans, R., Osuna-Millan, N., & Gaxiola-Pacheco, C. (2018). Modelling the interaction levels in HCI using an intelligent hybrid system with interactive agents: A case study of an interactive museum exhibition module in Mexico. *Applied Sciences*, 8(3), 446.
- Suvrat, J., & Roshita, J. (2019). Role of artificial intelligence in higher education—An empirical investigation. *IJRAR-International Journal of Research and Analytical Reviews*, 6(2), 144z-150z.
- Tayeh, R., & Issa, R. R. (2020). Interactive holograms for construction coordination and quantification. *Journal of Management in Engineering*, 36(6), 04020079.
- Universidad Iberoamericana. (2023). *Desertan de educación media superior 430 mil jóvenes en ciclo escolar 2023-2024* [Consultado en mayo de 2025]. <https://ibero.mx/prensa/desertan-de-educacion-media-superior-430-mil-jovenes-en-ciclo-escolar-2023-2024>
- Wei, H., Xu, J., Jiang, J., Liang, B., & Jia, Q. (2023). Holographic display in future automotive smart cockpit: application scenarios, interaction modals, and VACP analysis. *Advanced Fiber Laser Conference (AFL2022)*, 12595, 200-209.

7. Anexos

7.1. Informe de pruebas de software

El informe de pruebas tiene como objetivo evaluar el rendimiento del AEH en un entorno controlado, abarcando aspectos lógicos, físicos y humanos. Los módulos del sistema se componen de los siguientes:

Módulo de reconocimiento de voz

- Verificación de la precisión en el reconocimiento de preguntas formuladas por los usuarios.
- Pruebas con diferentes tonos y volúmenes de voz para evaluar la sensibilidad.
- Identificación de posibles problemas con palabras homónimas o ambiguas.

Módulo de conexión con el MLG

- Validación de la conexión con el modelo *gpt-3.5 Turbo*, para la generación de respuestas adecuadas.
- Evaluación de métricas de texto de las respuestas generadas.

Módulo de reproducción del holograma

- Ejecución del holograma al iniciar la aplicación.
- Verificación de la pausa automática del holograma al finalizar la interacción.
- Validación del correcto cambio al estado de espera tras una interacción completa.

Las pruebas del sistema se llevaron a cabo en un ambiente controlado, considerando los aspectos lógicos, físicos y humanos para garantizar la validez de los resultados.

Aspectos lógicos

El entorno lógico incluyó:

- Conexión a internet: todas las pruebas se realizaron en un entorno con conexión a internet, dado que el sistema depende de una conexión estable para interactuar con el modelo de lenguaje *gpt-3.5 Turbo* en la nube.
- Configuración del sistema: las pruebas se ejecutaron en una computadora con sistema operativo Windows 11 Pro, asegurando que los componentes necesarios, como el reconocimiento de voz y el modelo de lenguaje, estuvieran correctamente configurados. La aplicación se realizó en el IDE Android Studio 2024.1.1. Y la API para la conexión del MLG se utilizó Python 3.12.3.

Aspectos físicos

El ambiente físico fue controlado de la siguiente manera:

- Ubicación: las pruebas se llevaron a cabo en Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET).
- Dispositivos: se utilizaron una computadora GYGABYTE CPU i5-12500H x64 con 16 GB de RAM, una tablet CUBOT versión de Android 13 x64 con 4 GB de RAM y un mouse para la interacción manual con el micrófono. También se utilizó un cristal reflector para la representación holográfica de NadIA 2.0.

- Condiciones ambientales: la iluminación y el ruido se mantuvieron constantes y controlados para evitar que factores externos afectaran la calidad del holograma y la comprensión de las respuestas.

Aspectos humanos

El componente humano fue considerado en las pruebas de la siguiente manera:

- **Participantes:** las pruebas fueron realizadas por un grupo de 20 estudiantes de maestría, 13 hombres y 7 mujeres. Los usuarios interactuaron individualmente con el sistema realizando 8 preguntas o dudas relacionadas a sus conocimientos académicos.
- **Instrucciones:** se brindaron instrucciones claras a los usuarios antes de cada prueba, especificando cómo activar el micrófono, formular preguntas y utilizar la interfaz de la aplicación. Además de hablar en voz alta y clara para optimizar el reconocimiento de voz.
- **Evaluación:** se recopila tanto retroalimentación verbal como datos a través de un cuestionario posterior. También se observó la interacción de los usuarios con la interfaz gráfica y el nivel de satisfacción con las respuestas generadas.

Se diseñó y aplico un caso de prueba donde abarcan los principales módulos del sistema y están orientados a garantizar que todas las funcionalidades clave del AEH operen según lo esperado. En la Tabla 11 se dan los detalles completos de los casos de prueba proporcionando una descripción, precondiciones, pasos de la prueba, resultado esperado y estado.

Tabla 15: Descripción del caso de prueba.

ID del caso de prueba	Descripción	Precondiciones	Pasos de la prueba	Resultado esperado	Estado
CP 01	Prueba funcional de caja negra diseñada para comprobar que el sistema genere respuestas audibles y visuales a partir de preguntas realizadas por usuarios en un dominio específico.	1. Uso de internet. 2. Uso de un ambiente semi-controlado. 3. Selección de un usuario de prueba. 4. Configuración de una máquina utilizada como servidor. 5. Implementación de un dispositivo móvil utilizado como cliente. 6. Montaje del prototipo de hardware. 7. Instalación de un espejo reflejante. 8. Instrucción a un usuario con previo conocimiento de la pregunta o duda a formular. 9. Utilización de un mouse para interactuar con el micrófono. 10. Uso de conexión eléctrica para alimentar al cliente y al servidor.	1. El tester despliega el AEH en el cliente. 2. El usuario escucha el estado de saludo del AEH. 3. El usuario manipula el mouse para activar el micrófono del AEH. 4. El usuario dice su pregunta o duda académica. 5. El AEH responde la pregunta al usuario. 6. El AEH hace conexión al servidor del MLG. 7. El AEH queda en estado de espera.	Respuesta audible y visible del AEH.	Exitoso

El tiempo de ejecución promedio del caso de prueba fue de 6 minutos y fueron ejecutados con 20 usuarios, teniendo un total de 8 ejecuciones para cada caso de prueba. Los resultados obtenidos se muestran en la Tabla 12:

Tabla 16: Resultados del caso de prueba.

ID del caso de prueba	Casos de pruebas ejecutados	Casos de prueba exitosos	Casos de prueba fallidos	Casos de prueba bloqueados	Casos de prueba diferidos
01	161	150	11	0	0

Al ejecutar los casos de prueba se detectó 1 anomalía cuyo estado final se muestra en la Tabla 13:

Tabla 17: Anomalías en el caso de prueba.

Anomalías re-portadas	Anomalías en análisis	Anomalías des-cartadas	Anomalías en proceso	Anomalías corregidas
1	0	1	0	0

Dentro de los hallazgos derivados de la realización de las pruebas durante la evaluación del sistema, destacan diversas capacidades de NadIA 2.0 en la interacción con los usuarios. En primer lugar, se observó que el asistente es capaz de comprender preguntas y dudas formuladas tanto en primera como en tercera persona, lo que amplía su versatilidad en la interpretación del lenguaje. Asimismo, demostró un desempeño sólido en entornos con distintos niveles de ruido, logrando mantener una adecuada comprensión tanto en espacios silenciosos como en aquellos con ruido moderado.

Otro aspecto relevante es su capacidad para procesar el idioma inglés, destacándose especialmente su habilidad para entender y traducir palabras en inglés sin modificar ni mezclar los idiomas en la respuesta. Además, se evidenció un comportamiento cordial y proactivo al interactuar con los usuarios; cuando estos no formulaban correctamente una pregunta o la dejaban incompleta, ofreciendo la oportunidad de reformularla, mejorando así la experiencia conversacional.

A pesar de que NadIA 2.0 no cuenta con un sistema de memoria para recordar conversaciones previas, preguntas o dudas anteriores, algunos usuarios optaron por interactuar con ella en un formato conversacional más fluido. En estos casos, si bien su capacidad de respuesta fue limitada debido a la ausencia de contexto previo, logró responder, lo que resalta su potencial para futuras mejoras en términos de interactividad y continuidad en el diálogo.

7.2. Implementación de la tesis por computadora

A continuación, se proporciona el enlace al repositorio que contiene todos los códigos fuente desarrollados para la presente tesis. Este material incluye tanto los componentes del servidor como del cliente, así como las funciones encargadas del cálculo de métricas, conexión con la API y reproducción audiovisual del asistente.

En el lado del servidor, se implementó una API en *Python* para gestionar la conexión con el MLG de *OpenAI*, calcular las métricas de calidad de texto, manipular los *prompts* y controlar la persistencia de datos. Esta API se ejecuta en el puerto 8000 y se desarrolló en el entorno *Visual Studio*, utilizando diversas librerías para el manejo de tiempos, solicitudes HTTP, estructuras *JSON* y creación de *endpoints*. Las métricas fueron encapsuladas en funciones específicas, con funciones complementarias para el cálculo de sílabas e índices. Además, se creó una función especializada para

medir los tiempos de petición, manejar errores y ajustar hiperparámetros de la API de *OpenAI*.

En el lado del cliente, se utilizó *Android Studio* con el lenguaje *Java*. La lógica principal se organizó en dos clases: *SpeechRecognizer*, encargada del reconocimiento de voz en español mexicano, y *MainActivity*, que gestiona la interfaz, la conexión con la API y la reproducción de videos del asistente. El flujo contempla la conversión de la voz en texto, el envío del *prompt* a la API, la evaluación de la respuesta y la sincronización del video de NadIA con la salida generada. Al finalizar, se reproduce la respuesta mediante síntesis de voz y se activa el video de reposo, manteniéndose en espera de una nueva interacción.

Liga del documento: https://drive.google.com/file/d/1GUTLxm51WEax2mQPzhySbNs_BsSsFn5n/view?usp=sharing