

Daniel Esparza Barragán

Educación Impulsada por IA

2026



Universidad Autónoma de
Querétaro

Facultad de Ingeniería

**Educación Impulsada por IA:
Desarrollo de un Entorno Virtual de
Aprendizaje mediante un Modelo de
Lenguaje Ajustado**

Tesis

Que como parte de los requisitos para obtener el
Grado de

Maestro en Ciencias en Inteligencia Artificial

Presenta

Ing. Daniel Esparza Barragán

Dirigido por:
Dr. Sebastián Salazar Colores

Querétaro, Qro. a julio de 2026

La presente obra está bajo la licencia:
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>



CC BY-NC-ND 4.0 DEED

Atribución-NoComercial-SinDerivadas 4.0 Internacional

Usted es libre de:

Compartir — copiar y redistribuir el material en cualquier medio o formato

La licenciante no puede revocar estas libertades en tanto usted siga los términos de la licencia

Bajo los siguientes términos:



Atribución — Usted debe dar [crédito de manera adecuada](#), brindar un enlace a la licencia, e [indicar si se han realizado cambios](#). Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciante.



NoComercial — Usted no puede hacer uso del material con [propósitos comerciales](#).



SinDerivadas — Si [remezcla, transforma o crea a partir](#) del material, no podrá distribuir el material modificado.

No hay restricciones adicionales — No puede aplicar términos legales ni [medidas tecnológicas](#) que restrinjan legalmente a otras a hacer cualquier uso permitido por la licencia.

Avisos:

No tiene que cumplir con la licencia para elementos del material en el dominio público o cuando su uso esté permitido por una [excepción o limitación](#) aplicable.

No se dan garantías. La licencia podría no darle todos los permisos que necesita para el uso que tenga previsto. Por ejemplo, otros derechos como [publicidad, privacidad, o derechos morales](#) pueden limitar la forma en que utilice el material.



Universidad Autónoma de Querétaro
Facultad de Ingeniería
Maestría en Ciencias en Inteligencia Artificial

Educación Impulsada por IA: Desarrollo de un Entorno Virtual de Aprendizaje mediante un Modelo de Lenguaje Ajustado

Tesis

Que como parte de los requisitos para obtener el Grado de
Maestro en Ciencias en Inteligencia Artificial

Presenta

Ing. Daniel Esparza Barragán

Dirigido por

Dr. Sebastián Salazar Colores

Sínodo:

Dr. Sebastián Salazar Colores
Presidente

Dr. Andras Takacs
Suplente

Dr. Saúl Arriaga Tovar
Secretario

M.D.M. Ramón Torres Alonso
Suplente

M.C. Luis Roberto García Noguez
Vocal

UAQ Campus Aeropuerto, Querétaro, Qro. México
julio de 2026

Dedicatorias:

*A mis padres, por ser el cimiento de cada uno de mis pasos.
A mi novia, por caminar a mi lado con paciencia y amor incondicional.
A todos ellos, por creer en mí incluso cuando el camino parecía incierto.*

Agradecimientos

A la Universidad Autónoma de Querétaro y a la Facultad de Ingeniería, por ser mi casa de estudios y brindarme las herramientas necesarias para mi formación profesional en la Maestría en Ciencias en Inteligencia Artificial.

A la Secretaría de Ciencias, Humanidades, Tecnología e Innovación (SECIHTI), por el apoyo otorgado a través del programa de becas de posgrado, el cual fue fundamental para el desarrollo y la culminación de este trabajo de investigación.

Al Centro de Investigaciones en Óptica (CIO) en León, Guanajuato, por las facilidades brindadas durante mi estancia de investigación, así como por el acceso a la infraestructura tecnológica y los recursos de cómputo que permitieron el entrenamiento de los modelos presentados.

Al Dr. Sebastián Salazar Colores, por su invaluable dirección, pero sobre todo por su calidez humana y por facilitar todos los recursos necesarios para que este proyecto fuera posible. Gracias por la disponibilidad constante y por creer en mí desde el primer día.

Al Dr. Saúl Arriaga Tovar, por estar siempre a la distancia de un mensaje cuando más lo necesité; su disposición fue clave para resolver todas mis inquietudes durante este proceso.

A los integrantes de mi sínodo: Dr. Andras Takacs, M.C. Luis Roberto García Noguez y M.C. Ramón Torres Alonso, por sus valiosas observaciones que enriquecieron la calidad de este documento.

A Ángel Flores, por su excelente trabajo en el desarrollo del front-end de este sistema, tu apoyo técnico y personal fue invaluable.

A la directora de la Telesecundaria Jesús Urueta, la Mtra. Lorena Venegas Pacheco, por permitir que este proyecto cobrara un sentido social al ser aplicado con sus estudiantes.

A mis padres, por ser mi fuente de inspiración y el pilar fundamental de este éxito. Gracias por su amor incondicional, por el respeto a mis decisiones y por la confianza plena que siempre han depositado en mí. Gracias por el apoyo económico y moral que hizo posible este camino, pero sobre todo, por recordarme en cada momento cuánto me aman y que nunca estoy solo. Todo lo que soy se lo debo a ustedes.

A mi pareja, Daniela, por estos años de caminar juntos y por ser mi refugio incondicional durante cada etapa. Gracias por tu paciencia infinita, por creer en mí incluso en los momentos de mayor presión y por construir conmigo el hogar que compartimos. Gracias por amarme y por permitirme amarte.

A mi hermano Kevin, por ser una pieza clave en este camino, sin ti no hubiera concluido; gracias por creer siempre en mí y recordarme de lo que soy capaz.

A mi hermano Santiago, gracias por las risas, por tu compañía y por ser ese espacio de autenticidad donde simplemente pude ser yo mismo.

A mis profesores y compañeros, con quienes compartí retos y aprendizajes; gracias por el apoyo mutuo y por hacer de este proceso una experiencia compartida. A todos aquellos amigos que me brindaron su aliento y presencia a lo largo de estos años.

Abreviaturas y siglas

AI	Inteligencia Artificial (<i>Artificial Intelligence</i>)
API	Interfaz de Programación de Aplicaciones (<i>Application Programming Interface</i>)
CONALITEG	Comisión Nacional de Libros de Texto Gratuitos
DPO	Optimización Directa de Preferencias (<i>Direct Preference Optimization</i>)
EVA	Entorno Virtual de Aprendizaje
GenAI	Inteligencia Artificial Generativa
GPU	Unidad de Procesamiento Gráfico (<i>Graphics Processing Unit</i>)
HTTP	Protocolo de Transferencia de Hipertexto (<i>Hypertext Transfer Protocol</i>)
ISBN	Número Estándar Internacional de Libros (<i>International Standard Book Number</i>)
JSON	Notación de Objetos de JavaScript (<i>JavaScript Object Notation</i>)
JWT	Token de JSON para Web (<i>JSON Web Token</i>)
LLM	Modelo de Lenguaje Grande (<i>Large Language Model</i>)
LoRA	Adaptación de Bajo Rango (<i>Low-Rank Adaptation</i>)
MMLU	Comprensión de Lenguaje Masivo Multitarea (<i>Massive Multitask Language Understanding</i>)
NEM	Nueva Escuela Mexicana
NLP	Procesamiento de Lenguaje Natural (<i>Natural Language Processing</i>)
OCR	Reconocimiento Óptico de Caracteres
ORM	Mapeo Objeto-Relacional (<i>Object-Relational Mapping</i>)
PDF	Formato de Documento Portátil (<i>Portable Document Format</i>)
PEFT	Ajuste Fino Eficiente en Parámetros (<i>Parameter-Efficient Fine-Tuning</i>)
Q&A	Preguntas y Respuestas (<i>Questions and Answers</i>)
QLoRA	Adaptación de Bajo Rango Cuantizada (<i>Quantized Low-Rank Adaptation</i>)
RAG	Generación Aumentada por Recuperación (<i>Retrieval-Augmented Generation</i>)
RAM	Memoria de Acceso Aleatorio (<i>Random Access Memory</i>)
REST	Transferencia de Estado Representacional (<i>Representational State Transfer</i>)

RLHF	Aprendizaje por Refuerzo mediante Retroalimentación Humana (<i>Reinforcement Learning from Human Feedback</i>)
SFT	Ajuste Fino Supervisado (<i>Supervised Fine-Tuning</i>)
SQL	Lenguaje de Consulta Estructurada (<i>Structured Query Language</i>)
STT	Voz a Texto (<i>Speech-to-Text</i>)
TPU	Unidad de Procesamiento Tensorial (<i>Tensor Processing Unit</i>)
TTS	Texto a Voz (<i>Text-to-Speech</i>)
VARK	Modelo de estilos de aprendizaje (Visual, Auditivo, Lectura/Escritura, Kinestésico)
VRAM	Memoria de Acceso Aleatorio de Video (<i>Video Random Access Memory</i>)
XAI	Inteligencia Artificial Explicable (<i>Explainable Artificial Intelligence</i>)

Resumen

En las últimas dos décadas el desarrollo de modelos de inteligencia artificial ha experimentado un acelerado crecimiento, de entre éstos, los Modelos de Lenguaje Grande (LLM, por sus siglas en inglés) destacan por las habilidades emergentes que surgen de ellos y sus potenciales aplicaciones. El presente trabajo de investigación propone la implementación de un modelo de lenguaje grande ajustado específicamente en el dominio educativo, utilizando como base los contenidos de los libros de texto gratuitos del CONALITEG para secundaria y telesecundaria, con el fin de crear un prototipo de un Entorno de Aprendizaje Virtual; cuya principal característica sea el despliegue de un asistente educativo que ayude a los usuarios a desarrollar habilidades y adquirir conocimiento. El proceso para el desarrollo del proyecto consistió, en primer lugar, en la selección y preparación de un corpus educativo representativo del contexto mexicano, seguido del ajuste fino del modelo. Finalmente, se integró el modelo ajustado en una interfaz de usuario amigable que permite al asistente ofrecer orientación personalizada durante el proceso de aprendizaje. El objetivo del desarrollo de este prototipo fue proporcionar una herramienta educativa accesible y personalizada que busca mejorar la experiencia de aprendizaje, mitigando algunos de los problemas existentes que impiden alcanzar la equidad educativa en México.

Palabras clave: Modelos de Lenguaje Grande, Ajuste Fino, Entornos Virtuales de Aprendizaje, Inteligencia Artificial en Educación.

Abstract

In the last two decades the development of artificial intelligence models has experienced accelerated growth. Among these, Large Language Models (LLMs) stand out due to the emergent abilities that arise from them and their potential applications. The present research work proposes and implements a large language model specifically fine-tuned in the educational domain, using the contents of the free textbooks from CONALITEG for middle school and secondary distance education (Telesecundaria) as a base, with the purpose of creating a prototype for a Virtual Learning Environment; where its main characteristic is the deployment of an educational assistant that helps users develop skills and acquire knowledge. The process for the project development consisted, firstly, of the selection and preparation of an educational corpus representative of the Mexican context, followed by the fine-tuning of the model. Finally, the fine-tuned model was integrated into a user-friendly interface that allows the assistant to offer personalized guidance during the learning process. The objective of developing this prototype was to provide an accessible and personalized educational tool that seeks to improve the learning experience, mitigating some of the existing problems that impede achieving educational equity in Mexico.

Keywords: Large Language Models, Fine-Tuning, Virtual Learning Environments, Artificial Intelligence in Education.

Índice general

Índice de cuadros	III
Índice de figuras	1
1 Introducción	2
1.1 Contexto y Descripción del Problema	2
1.2 Propuesta de Solución y Alcance	3
1.3 Justificación	3
1.4 Hipótesis	4
1.5 Objetivos	4
1.6 Estructura de la Tesis	5
2 Marco Teórico y Antecedentes	7
2.1 El Paradigma Educativo Tradicional y sus Limitaciones	7
2.2 Fundamentación Teórica de los Grandes Modelos de Lenguaje	8
2.2.1 Arquitectura transformer	8
2.2.2 Modelos de Lenguaje Grandes (LLM)	8
2.2.3 Ajuste fino (Fine-Tuning) y Optimización	9
2.2.4 Técnicas de Optimización: PEFT y Cuantización	9
2.3 Estado del Arte de los LLM en la Educación	10
2.3.1 Chatbots y Asistentes Virtuales	10
2.3.2 Aprendizaje Personalizado y Tutores Virtuales	12
2.3.3 Diseño Curricular y Generación de Contenido	13
2.4 Análisis Crítico y Retos Estructurales	14
2.4.1 Ética, Sesgos y Equidad en la Personalización	14
2.4.2 Integridad Académica y Replanteamiento de la Evaluación	15
2.5 Pensamiento Crítico, Autonomía y Formación Docente	15
2.5.1 Riesgos Regulatorios y de Privacidad de Datos	16
3 Metodología de Desarrollo	17
3.1 Diseño y Enfoque de la investigación	17
3.2 Población objetivo y Muestra de Evaluación	17
3.3 Fases de la metodología experimental	17

4	Análisis de Contexto para Selección de LLM	19
5	Construcción del Corpus Educativo	21
5.1	Formato de Entrenamiento y División del Dataset	22
5.2	Consideraciones Éticas y de Propiedad Intelectual	23
6	Ajuste Fino Supervisado del LLM	24
6.1	Estrategias de Optimización y Eficiencia Computacional	24
6.2	Configuración del Experimento e Hiperparámetros	25
6.3	Preparación para el Despliegue	25
7	Arquitectura de Recuperación y Contextualización	26
7.1	El Paradigma RAG y análisis de viabilidad de implementación	26
7.2	Inyección de Contexto Dinámica	27
8	Desarrollo del Prototipo	29
8.1	Arquitectura y Flujo de Interacción	29
8.2	Capa de Servicios (Backend)	30
8.3	Capa de Datos	32
8.4	Capa de Presentación(Frontend)	33
9	Metodología de Evaluación y Diseño Experimental	37
9.1	Metodología de Evaluación Cuantitativa	37
9.1.1	Variantes del Modelo en Evaluación	37
9.1.2	Métricas de Procesamiento de Lenguaje Natural (NLP)	38
9.2	Metodología de Evaluación Cualitativa: Intervención en Campo	38
9.2.1	Sujetos de Estudio y Contexto de Implementación	39
9.2.2	Diseño de la Intervención Pedagógica (Proyectos Transversales) .	39
9.2.3	Protocolo de la Prueba Piloto	39
9.2.4	Instrumentos de Percepción del Usuario	40
9.3	Justificación del Marco de Evaluación	40
10	Resultados y Evaluación	41
10.1	Resultados Cuantitativos del Modelo Ajustado	41
10.1.1	Análisis de los resultados obtenidos	42
10.2	Resultados Cualitativos Pertinencia Pedagógica	44
10.3	Resultados de la Prueba Piloto en Campo	45
10.3.1	Análisis de la Intervención Pedagógica: Proyectos Transversales .	45
10.3.2	Análisis de resultados de la encuesta de satisfacción	46
10.3.3	Resultados de las preguntas abiertas	47
10.4	Conclusiones del capítulo de Resultados	47
11	Conclusiones y Trabajo Futuro	48
11.1	Conclusiones y Cumplimiento de Objetivos	48

11.2	Aportaciones y Trascendencia del Proyecto	48
11.3	Limitaciones y Desafíos	49
11.3.1	Reflexión sobre Mejora Continua y Perspectivas Futuras	49
	Referencias bibliográficas	51
A	Carta de Consentimiento Informado para Padres o Tutores	58
B	Carta de Asentimiento Informado para Estudiantes	60
C	Carta de Consentimiento Informado para el Docente	62
D	Instrumento de Evaluación Piloto 1: Cosecha Sostenible	64
E	Instrumento de Evaluación Piloto 2: Presupuesto Escolar Responsable	67
F	Cuestionario de Experiencia de Usuario y Usabilidad (Post-Prueba)	70

Índice de cuadros

4.1	Análisis comparativo de benchmarks para la selección del modelo base.	20
8.1	Endpoints Principales del Backend	30
8.2	Parámetros de Generación del LLM SFT	33
10.1	Comparativa extendida de métricas de desempeño entre las variantes del modelo.	41
10.2	Análisis cualitativo: Comparativa de pertinencia didáctica.	45
F.1	Escala de Usabilidad y Satisfacción del Entorno Virtual (Post-Prueba).	71

Índice de figuras

8.1	Arquitectura de Despliegue de Tres Capas del Prototipo TutorIA. . . .	30
8.2	Diagrama tablas relacionales (bases de datos)	32
8.3	Registro y Cuestionario VARK	34
8.4	Dashboard Principal con los distintos módulos de Aprendizaje.	35
8.5	Interfaz del Tutor Virtual de Aprendizaje con Temporizador Pomodoro Integrado.	35
8.6	Módulo de Gamificación	36
10.1	Gráfica comparativa de métricas de desempeño y longitud promedio de respuesta.	42

Capítulo 1

Introducción

La pandemia del COVID-19 evidenció algunas limitaciones inherentes de los paradigmas educativos tradicionales, resaltando la necesidad urgente de adaptar los enfoques educativos tradicionales a nuevas condiciones, en específico, la imposibilidad de continuar con los cursos presenciales. La implementación de herramientas digitales durante este periodo resultó ser efectiva para afrontar los retos que impuso la educación a distancia, demostrando así la capacidad de la tecnología para transformar y mejorar los métodos de enseñanza convencionales [1]. Este contexto plantea una oportunidad para aprovechar las nuevas tecnologías impulsadas por modelos de Inteligencia Artificial (AI, por sus siglas en inglés), con el objetivo de superar las limitaciones de los métodos tradicionales de enseñanza que a menudo representan obstáculos significativos en el proceso de enseñanza-aprendizaje [2].

1.1. Contexto y Descripción del Problema

Entrenar Modelos de Lenguaje Grandes (LLM, por sus siglas en inglés) desde cero es impracticable debido a los enormes recursos computacionales y la vasta cantidad de datos necesarios. En su lugar, se recurre al Ajuste Fino Supervisado (SFT, por sus siglas en inglés) de modelos preentrenados para adaptarlos a contextos específicos, como en este caso, el educativo. Sin embargo, este proceso enfrenta varios desafíos. La selección y preparación de datos específicos y de alta calidad es fundamental ya que datos incorrectos o sesgados pueden degradar el rendimiento del modelo. Además, el ajuste fino requiere un poder de cómputo considerable, utilizando hardware avanzado como Unidades de Procesamiento Gráfico (GPU, por sus siglas en inglés) o Unidades de Procesamiento Tensorial (TPU, por sus siglas en inglés), y un conocimiento técnico profundo en programación, *machine learning* y *deep learning*.

En el contexto educativo, es muy importante que los datos utilizados reflejen una amplia gama de contenidos y estilos pedagógicos para ofrecer una herramienta didáctica efectiva y personalizada. Problemas como el sobreajuste y el balance adecuado de los datos son críticos para evitar sesgos que perpetúen desigualdades. Asimismo, el

diseño del sistema debe facilitar una interacción intuitiva y útil para los estudiantes, adaptándose a diferentes ritmos y estilos de aprendizaje. Por último, la implementación responsable y ética de estos modelos es esencial para asegurar que sus beneficios sean equitativos y no perpetúen desinformación o injusticias, proporcionando así un entorno educativo inclusivo y efectivo.

1.2. Propuesta de Solución y Alcance

Para abordar estos desafíos se propone un prototipo inspirado en la técnica de Generación Aumentada por Recuperación (RAG, por sus siglas en inglés), utilizando un modelo de lenguaje ajustado con un corpus educativo curado y estructurado conforme a los planes de estudio de la Nueva Escuela Mexicana. La preparación del corpus incluirá limpieza, tokenización y segmentación temática de los recursos pedagógicos de la Comisión Nacional de Libros de Texto Gratuito (CONALITEG). El ajuste fino se realizará en entornos con GPU e implementando técnicas que reduzcan el tamaño en memoria del modelo y el sobreajuste. Finalmente, se desarrollará una interfaz interactiva que facilite consultas y fomente un aprendizaje personalizado.

1.3. Justificación

Los Modelos de Lenguaje Grande son una solución prometedora debido a sus capacidades de procesamiento del lenguaje natural y adaptación a contextos específicos, además de la accesibilidad y disponibilidad que ofrecen. Por otro lado, estos modelos mejoran continuamente mediante actualizaciones basadas en datos de interacción acumulados, lo que les permite ofrecer respuestas cada vez más pertinentes y ajustadas a las necesidades específicas del entorno educativo en el que se utilizan [3]. Estas características posicionan a los LLM como una opción viable para el desarrollo de asistentes educativos innovadores [4].

En esta investigación se desarrolla un prototipo con la funcionalidad de un asistente educativo con el fin de ofrecer una herramienta que se adapte a las necesidades de aprendizaje de estudiantes de nivel de educación básica en México. Para lograrlo, se utilizarán modelos de lenguaje ajustados con contenidos alineados al currículo nacional de la Nueva Escuela Mexicana y además empleando técnicas inspiradas en RAG.

A diferencia de otras propuestas internacionales, este proyecto se centra específicamente en el ajuste del LLM con un corpus educativo con contenidos oficiales del contexto educativo mexicano, mediante materiales desarrollados por la CONALITEG. Este enfoque poco explorado contribuye a garantizar la pertinencia curricular del prototipo y representa una aportación al área de tecnologías educativas en el país. Lo que se busca no es sólo generar respuestas automáticas, sino respuestas útiles y adecuadas al nivel educativo objetivo. Adicionalmente, se contempla implementar el prototipo en un entorno real con estudiantes y docentes con el fin de evaluar si efectivamente el prototipo cumple con los objetivos planteados en la presente investigación.

Por último, al tratarse de una herramienta con capacidad de funcionar de manera continua y remota, se espera que también tenga un impacto en la accesibilidad educativa al facilitar el acceso a materiales de apoyo en cualquier momento y lugar [5]. La combinación de estas características le confiere al prototipo la capacidad de ser una herramienta con potencial para mejorar el acceso a contenidos pedagógicos en contextos específicos. [6].

1.4. Hipótesis

El ajuste fino de Modelos de Lenguaje Grandes (LLM) con un corpus educativo mexicano de la CONALITEG mejorará la relevancia y personalización de un asistente educativo, incrementando la precisión y adecuación contextual de las respuestas, lo cual se reflejará en mejores puntuaciones en métricas de evaluación semántica y de alineación pedagógica así como en tasas de satisfacción de usuario en pruebas controladas, esperando que, en comparación con un modelo base, el modelo ajustado ofrezca una experiencia de aprendizaje más adaptativa y contribuya a la equidad educativa al ajustarse a diferentes ritmos y necesidades de aprendizaje.

1.5. Objetivos

a. Objetivo general:

Desarrollar un prototipo funcional de asistente educativo basado en un LLM ajustado mediante técnicas de *fine-tuning*, orientado a mejorar la personalización de los procesos de enseñanza-aprendizaje en el nivel de educación básica en México, promoviendo el acceso equitativo a recursos educativos de calidad y adaptándose a las necesidades individuales de los estudiantes.

b. Objetivos específicos:

- Seleccionar y preparar un corpus educativo representativo del contexto mexicano, utilizando como base los libros de texto de la CONALITEG, para garantizar la pertinencia cultural y curricular del modelo.
- Realizar el Ajuste Fino Supervisado de un Modelo de Lenguaje Grande con el corpus seleccionado, con el fin de especializarlo en contenidos educativos alineados al currículo oficial de la educación pública en México.
- Establecer una metodología de evaluación cuantitativa y cualitativa que permitan evaluar la exactitud contextual y relevancia temática del modelo ajustado, en comparación con modelos base no especializados.
- Diseñar e implementar una interfaz digital sencilla y funcional que permita a estudiantes y docentes interactuar con el asistente educativo, dándole priori-

dad a la usabilidad, accesibilidad y la adecuación al contexto de la educación pública en México.

- Analizar los resultados obtenidos en pruebas controladas con usuarios, enfocándose en la calidad percibida de las respuestas del modelo ajustado, así como en su utilidad y adecuación pedagógica, con el objetivo de valorar su impacto en la experiencia de aprendizaje.

1.6. Estructura de la Tesis

El presente trabajo de tesis está dividido en once capítulos. A continuación se muestra a modo de síntesis el contenido de cada uno de ellos.

- **Capítulo 1: Introducción.** Se establece el contexto de la investigación, cuyo eje principal gira en torno a las problemáticas educativas actuales y cómo los Modelos de Lenguaje Grande suponen una herramienta con el potencial de mitigarlas. Se exponen las dificultades técnicas que conlleva la implementación de LLM en el ámbito educativo y se presenta la hipótesis y objetivos del presente trabajo de tesis.
- **Capítulo 2: Marco teórico y antecedentes.** Se realiza una revisión sistemática de la literatura actual sobre los fundamentos teóricos de los LLM y se expone el estado del arte de la aplicación de estos en el ámbito educativo. Luego, se discuten las cuestiones éticas relativas a su uso y se expone la necesidad de establecer marcos regulatorios pertinentes.
- **Capítulo 3: Metodología.** Se define el enfoque de la investigación y el proceso ordenado y detallado para el desarrollo de la misma. Que va de la construcción de un corpus educativo contextualizado a la educación básica en México a partir de los recursos educativos de la CONALITEG, a la descripción del procedimiento de SFT de un LLM y el diseño de la evaluación experimental del prototipo producto de la investigación.
- **Capítulo 4: Análisis de Contexto para Selección de LLM.** Se realiza un análisis del contexto de implementación para justificar la elección del LLM que será el componente principal del prototipo de Entorno Virtual de Aprendizaje (EVA) a desarrollar.
- **Capítulo 5: Construcción del Corpus Educativo.** Se detalla el proceso de selección, preparación y curado de un corpus educativo contextualizado al Modelo de Educación pública actual en México.
- **Capítulo 6: Ajuste Fino Supervisado del LLM.** Se detalla el eje fundamental del proyecto de investigación; el proceso de entrenamiento (*Supervised Fine-Tuning*, SFT) del LLM seleccionado utilizando la técnica QLoRA, detallando cómo el modelo aprende el léxico y la pedagogía de la Nueva Escuela Mexicana.

- **Capítulo 7: Arquitectura de Recuperación y Contextualización.** Analiza el paradigma RAG (*Retrieval-Augmented Generation*) y justifica la transición hacia una arquitectura de **Inyección de Contexto Dinámica** con el fin de optimizar la latencia en entornos de baja conectividad.
- **Capítulo 8: Desarrollo del prototipo.** Se expone en detalle el proceso de desarrollo del EVA; se explica la arquitectura de software implementada, el desarrollo de la capa de servicios (backend) para la interacción con el LLM ajustado y el diseño y construcción de la Interfaz de usuario (frontend) haciendo énfasis en que esta sea amigable para el usuario.
- **Capítulo 9: Metodología de Evaluación y Diseño Experimental.** Se establece el protocolo de las pruebas piloto y se define el uso de una variante de control con inyección de referencia para medir el desempeño del modelo ajustado frente a un entorno con recuperación de contexto ideal.
- **Capítulo 10: Resultados y Evaluación.** Se presentan los resultados de la investigación; se muestran los resultados cuantitativos del desempeño del Modelo Ajustado, los resultados cualitativos de las pruebas piloto con usuarios y se discuten los hallazgos clave del trabajo en general.
- **Capítulo 11: Conclusiones y Trabajo Futuro.** Se resume el conocimiento inferido del trabajo de investigación, se muestra el cumplimiento de los objetivos y la validación de la hipótesis y destaca las aportaciones del proyecto. También se explican las limitaciones encontradas y la perspectiva de investigación futura que se plantea a partir del presente trabajo.

Capítulo 2

Marco Teórico y Antecedentes

2.1. El Paradigma Educativo Tradicional y sus Limitaciones

El paradigma tradicional de educación cuya principal característica es la modalidad presencial de un profesor frente a un grupo de estudiantes guiando el proceso de aprendizaje se ha mantenido prácticamente sin cambios a lo largo de los últimos siglos, principalmente porque funciona, sin embargo, este paradigma presenta algunas limitaciones que en ocasiones conducen a deficiencias en el aprendizaje del alumno [7][8].

Una de las principales limitantes es la inequidad de acceso a la educación debido a limitaciones geográficas y de recursos; el contexto social y económico en todo el mundo es el de la centralización de los recursos en las grandes ciudades, que aglomeran gran parte de los recursos económicos y humanos. Esto ha traído como consecuencia una serie de dificultades para integrar los sistemas de educación pública a toda la población, en particular las poblaciones rurales y comunidades marginadas, que en muchas ocasiones presentan limitaciones de infraestructura como aulas educativas, medios de transporte o insuficientes recursos pedagógicos y humanos son las principales afectadas, puesto que dichos factores influyen directamente dificultando el acceso a la educación de calidad [9]. Es por ello que en dichas comunidades los índices de abandono escolar en los niveles de educación básica y media superior suelen ser altos, y que muchas veces para poder acceder a educación de nivel superior la población se vea obligada a emigrar [10].

Otra dificultad reside en las deficiencias de integración y adaptabilidad de los métodos educativos; los paradigmas educativos tradicionales afrontan dificultades para implementar estrategias de enseñanza personalizada, es decir, estrategias que reconocen la diversidad de habilidades, estilos y ritmos de aprendizaje así como intereses individuales. En este contexto los alumnos tienen que adaptarse a los paradigmas educativos y en el proceso pueden presentar deficiencias en la adquisición de habilidades y conocimiento [2].

A esto se suman los problema relativos a la disponibilidad de horarios destinados para la enseñanza; los paradigmas educativos tradicionales comúnmente se desarrollan

bajo un esquema de horarios preestablecidos fijos. Sin embargo, el proceso de aprendizaje la mayoría de las veces requiere de mayor tiempo que aquel destinado en las sesiones de éstos, por lo que es habitual que se implementan actividades fuera de estos horarios para profundizar en el aprendizaje. En este contexto, en función de las habilidades y accesibilidad de recursos educativos, el desarrollo satisfactorio de las actividades mencionadas puede verse obstaculizado y en consecuencia no alcanzarse el refuerzo del aprendizaje deseado. Por otro lado, el esquema de horarios fijos puede ser problemático para aquellos individuos que tienen responsabilidades que comprometen su tiempo en determinados horarios del día [11].

Finalmente, existen dificultades relativas a la retroalimentación y evaluación continua de los estudiantes; la retroalimentación individual acerca del proceso de aprendizaje de los alumnos desempeña un papel significativo para alcanzar los objetivos educativos, así como la evaluación continua que refleja el progreso del estudiante. Sin embargo, los múltiples roles que desempeña el profesor en el proceso de enseñanza requieren de tiempo y dedicación, lo que pueden resultar agotador, principalmente cuando el grupo de estudiantes es numeroso. En este contexto la retroalimentación que reciben los alumnos en ocasiones es muy general o en ocasiones poco frecuente. Cuando esto ocurre el alumno puede enfrentar dificultades durante el proceso de aprendizaje [12].

En el contexto educativo, la implementación de herramientas computacionales y basadas en tecnologías de la información, ha demostrado ser efectiva al mitigar algunas de las dificultades que presentan los paradigmas tradicionales educativos, como los que se mencionaron anteriormente [13].

2.2. Fundamentación Teórica de los Grandes Modelos de Lenguaje

2.2.1. Arquitectura transformer

En 2017, en el contexto de la inteligencia artificial, el modelo de arquitectura transformer fue presentado en el artículo “Attention is All You Need” [14]. Esta arquitectura implementa un mecanismo de atención que permite que los modelos de lenguaje tomen en cuenta a los demás elementos de una secuencia de texto, facilitando que el modelo comprenda la relación entre palabras distantes, es decir, comprender el contexto de las palabras e inferir su significado a partir de dicho contexto. Esto condujo a la eliminación del uso de redes recurrentes y convolucionales en el Procesamiento del Lenguaje Natural (NLP, por sus siglas en inglés) lo que mejoró la eficiencia computacional [15].

2.2.2. Modelos de Lenguaje Grandes (LLM)

Los Modelos de Lenguaje Grande (*Large Language Model*) son sistemas de procesamiento de lenguaje natural basados en la arquitectura *transformer* que utilizan redes neuronales profundas para entender y generar texto [16].

El desarrollo de estos modelos implica dos etapas fundamentales; en la primera el modelo es entrenado con una gran cantidad de textos, lo que otorga al modelo una capacidad la comprensión de texto y la generación de texto de manera coherente. La segunda etapa es el ajuste del modelo previamente entrenado, esto se logra reentrenándolo con un conjunto de datos más pequeño, lo que le da al modelo la capacidad de especializarse en temas más específicos [16].

2.2.3. Ajuste fino (Fine-Tuning) y Optimización

El ajuste fino de grandes modelos de lenguaje refina las habilidades adquiridas por el modelo en la etapa de entrenamiento general, optimizándolo en tareas particulares, con lo que mejora el rendimiento en aplicaciones más concretas.

Para realizar el ajuste fino de un LLM se siguen los siguientes pasos [17][18]:

- Selección del conjunto de datos para el ajuste fino: Se selecciona y procesa el conjunto de datos relacionados con área de interés.
- Configuración del entrenamiento para el ajuste fino: Se establecen los hiperparámetros del modelo como el tamaño del batch y el número de épocas de entrenamiento.
- Entrenamiento supervisado: El modelo es entrenado con los datos etiquetados para minimizar el error en la tarea específica en la que se busca ajustar el modelo.
- Evaluación del modelo ajustado: Se utiliza un conjunto de datos que sirven para validar y monitorear el rendimiento del modelo ajustado.

2.2.4. Técnicas de Optimización: PEFT y Cuantización

Dada la alta demanda computacional asociada al entrenamiento y despliegue de los LLMs, han surgido técnicas de optimización para hacer el proceso más eficiente y accesible. El Ajuste Fino con Optimización de Parámetros (PEFT, por sus siglas en inglés), que incluye métodos como la Adaptación de bajo Rango (LoRA, por sus siglas en inglés), reduce drásticamente el número de parámetros que deben ser entrenados, manteniendo un rendimiento comparable al del ajuste fino completo.

Complementariamente, la cuantización es un proceso que reduce la precisión de los pesos del modelo (por ejemplo, de 16-bits a 4-bits) durante el entrenamiento o la inferencia. Esto disminuye significativamente los requisitos de Memoria de Acceso Aleatorio o Memoria de Acceso Aleatorio de Video (RAM o VRAM, por sus siglas en inglés) y la latencia, siendo crucial para el despliegue de modelos grandes en entornos con recursos computacionales limitados.

2.3. Estado del Arte de los LLM en la Educación

La integración de modelos de inteligencia artificial en el contexto educativo, en particular de modelos de lenguaje aplicados a sistemas de pregunta-respuesta (chatbots), ha demostrado ser una herramienta valiosa en el proceso de enseñanza-aprendizaje al facilitar interacciones naturales y personalizadas con los usuarios [19, 20, 21].

Diversos estudios recientes han analizado la implementación de la AI y LLM en la educación, destacando su potencial para mejorar la calidad del diálogo y de los contenidos educativos, así como los retos asociados a su uso en el aula [22, 23, 24]. Estas investigaciones coinciden en que la AI generativa puede favorecer nuevas estrategias pedagógicas centradas en el estudiante y en el desarrollo de habilidades de razonamiento, al tiempo que promueve una enseñanza más personalizada [25].

También se han explorado enfoques de ajuste fino y de instrucción curricular que emulan el aprendizaje humano y logran mejoras en el rendimiento de los modelos sin incrementar significativamente los recursos computacionales [26]. De igual modo, se han documentado avances en el uso de modelos de código abierto y técnicas como la Generación Aumentada por Recuperación (RAG), el Ajuste Fino Supervisado (SFT) y la Optimización Directa de Preferencias (DPO), que han mostrado incrementos en la precisión y coherencia de las respuestas de los asistentes educativos [27]. Algunos desarrollos han enfocado su aplicación en contextos científicos o de bajos recursos, como los prototipos de asistencia para la enseñanza de las ciencias en África o los sistemas de apoyo docente que integran varios agentes educativos virtuales para promover la participación estudiantil y mejorar la eficiencia del profesorado [28, 29]. Asimismo, se han implementado modelos como Galactica, entrenado con corpus científicos, que han superado a otros modelos en tareas especializadas y ofrecen oportunidades de libre acceso para la comunidad académica [30]. Finalmente, los prototipos de asistentes educativos virtuales desarrollados con herramientas como IBM Watson o Remy evidencian que la incorporación de LLM en entornos educativos fomenta el aprendizaje personalizado e incrementa la motivación demostrando su potencial en este contexto [31].

2.3.1. Chatbots y Asistentes Virtuales

Los *chatbots*, como asistentes virtuales en el proceso de enseñanza-aprendizaje, son la primera irrupción de los LLM en el contexto educativo. Estos representan una tecnología prometedora que ofrece posibilidades únicas como herramientas de comunicación e información para el aprendizaje digital, especialmente en términos de escalabilidad y accesibilidad [32]. Es por ello que los *chatbots* son una conexión clave entre la inteligencia artificial generativa y la pedagogía.

El estudio documentado en la literatura [33] expone como este tipo de sistemas inteligentes tienen la capacidad de asistir de manera interactiva a los alumnos en el proceso de aprendizaje, ofreciendo además evaluación continua y automatizada del proceso y brindar asistencia administrativa. Este es el primer punto diferenciador de este tipo de tecnologías en comparación con plataformas tradicionales. Un trabajo de revisión

sistemática [20], que analizó y sintetizó 53 publicaciones, destaca que las características más relevantes son la personalización de la enseñanza, la disponibilidad inmediata de asistencia y la reducción del trabajo administrativo por parte de los docentes. Por otro lado, también remarca algunas limitaciones relevantes inherentes a estos modelos, como la fiabilidad de la información desplegada y las cuestiones éticas relacionadas con la privacidad de los datos expuestos.

Algunos otros ejemplos de implementación que aportan a la discusión se sintetizan a continuación. Un artículo [31] evidencia la viabilidad de integración de sistemas configurables en entornos educativos de nivel universitario, donde son capaces de responder preguntas frecuentes y ofrecer recursos académicos pertinentes. De manera parecida, otra investigación [34], destaca como la integración de un *chatbot* en un portal institucional facilita información automática sobre horarios y servicios académicos, además de ofrecer orientación personalizada. Ambos ejemplos muestran como este tipo de chatbots no sólo benefician los procesos pedagógicos sino también los administrativos.

Analizando el tema desde el punto de vista de los profesores, un estudio [19] expone que los docentes afirman que estos *chatbots* pueden ser herramientas valiosas en las aulas, haciendo énfasis en la capacidad de responder preguntas rápidas o facilitar el proceso de elaboración de actividades. Los *chatbots* también pueden ayudar a los profesores con el proceso de evaluación y la generación de tareas, ahorrándoles tiempo y permitiéndoles concentrarse más en interactuar con los estudiantes. Por otro lado, la perspectiva de la comunidad estudiantil analizada en otro estudio concluye que los alumnos valoran la velocidad de respuesta y la disponibilidad temporal.

También se han documentado trabajos que destacan cómo los *chatbots* están transformando las dinámicas educativas de los paradigmas tradicionales. Un estudio [35] plantea que, como estos modelos tienen la capacidad para resolver tareas académicas, es necesario repensar los métodos de evaluación para garantizar el aprendizaje significativo. En este mismo tópico, un trabajo examina cómo estas tecnologías tienen la capacidad de mediar la interacción grupal, donde el marco de simulación SimClass demuestra que los agentes LLM pueden colaborar para crear interacciones dinámicas entre estudiantes y profesores. Otro eje relevante es la educación a distancia, donde diversos estudios [29] discuten cómo estos sistemas pueden promover la participación activa, la interacción y la comunidad. De la misma forma, un trabajo [36] explora cómo la integración de estas herramientas en plataformas educativas (como *Moodle*) permite acompañar a los alumnos de manera asincrónica, apoyando el aprendizaje autorregulado.

En resumen, los beneficios encontrados por los diversos estudios documentados coinciden en algunos puntos clave: 1) Accesibilidad y disponibilidad permanente [28] . 2) Reducción de la carga administrativa. 3) Capacidad de personalización de la experiencia de aprendizaje. 4) Mejora de la motivación de los estudiantes.

De manera análoga, las limitaciones también son evidentes: 1) Fiabilidad de las respuestas desplegadas. 2) Dependencia tecnológica. 3) Preocupaciones éticas concernientes a la integridad académica [37] . 4) Carencia de empatía real, no simulada. Es importante notar que, incluso con una alta capacidad de respuesta, la adopción por

parte de los estudiantes sigue una distribución no normal [38] . El uso de *AI Chatbot* también tiene potencial para desarrollar la capacidad de autoaprendizaje de los estudiantes.

2.3.2. Aprendizaje Personalizado y Tutores Virtuales

La idea de la educación personalizada como aquella capaz de adaptar contenidos, ritmos y metodologías de aprendizaje a las necesidades individuales de los estudiantes ha sido durante mucho tiempo un objetivo de gran relevancia para la investigación educativa. En este contexto, los LLM han abierto nuevas posibilidades para materializar este ideal dadas sus capacidades descritas anteriormente. Por ejemplo, un trabajo [39], muestra la capacidad de ChatGPT para proponer ejercicios que de manera dinámica estén ajustados específicamente al nivel educativo del alumno, así como ofrecer retroalimentación inmediata y explicaciones detalladas de los ejercicios propuestos. Este enfoque contrasta con los sistemas tradicionales que proponen ejercicios estáticos predefinidos sin ningún tipo de adaptabilidad contextual. El desarrollo de agentes multimodales es otra área clave. Un estudio [40] muestra cómo los sistemas multi-agente pueden automatizar la generación de contenido en formatos multimodales (como texto a imagen, texto a audio y texto a video) adaptados a diversos estilos de aprendizaje.

La personalización de la enseñanza de lenguas con LLM también ha sido documentada en trabajos como uno [41] donde se evaluó la implementación de un LLM como tutor de escritura en inglés, integrando principios pedagógicos para una evaluación efectiva y proporcionando retroalimentación en tiempo real. Asimismo, la investigación documentada [24] explora el papel de los chatbots en la generación de contenido educativo relevante y diverso, a través del uso de técnicas avanzadas de ingeniería de prompts.

Otra área emergente de la personalización de la educación impulsada por LLM es la creación de entornos de aulas virtuales como es descrito en un artículo [42] donde se recrean entornos educativos completos, simulando múltiples personajes estudiantiles para la resolución colaborativa de problemas en modelado matemático. También se han explorado ideas de implementación de elementos lúdicos ya que los sistemas de tutoría inteligente pueden transformar el aprendizaje en línea a través de una experiencia de aprendizaje activo y altamente personalizado que mejora significativamente los resultados de aprendizaje [43].

Por otro lado, también se han documentado trabajos que exploran los desafíos relevantes que existen entre la personalización y la autonomía de los estudiantes, como la investigación que analiza el impacto de la AI en la agencia del profesor, alertando que la Inteligencia Artificial Generativa (GenAI) puede afectar la capacidad del docente para tomar decisiones, lo cual aplica de manera análoga al dilema de la dependencia del estudiante[44] Para contrarrestar esta dependencia, se ha desarrollado un modelo [45], que utiliza LLM y la ingeniería de prompts para descubrir conceptos de conocimiento implícitos y mejorar la precisión del diagnóstico de la comprensión del estudiante. Finalmente, es importante tener en cuenta que la implementación de estos sistemas de tutoría virtual requiere de infraestructura tecnológica, acceso a internet y compe-

tencias digitales de parte de los alumnos como de los profesores. En ciertos contextos de desigualdad social y económica estas condiciones podrían aumentar la brecha en la equidad educativa en lugar de aminorarla. Por ello los proyectos de personalización de la educación impulsados por modelos de IA requieren de políticas públicas que garanticen el acceso equitativo y de alfabetización digital para los estudiantes, así como para los docentes. La necesidad de explorar cómo el uso no facilitado de LLM afecta los comportamientos de aprendizaje es una prioridad en este campo [46].

2.3.3. Diseño Curricular y Generación de Contenido

La digitalización de material didáctico y la adopción de plataformas digitales habían marcado una característica importante de la pedagogía en las últimas dos décadas. Sin embargo, la implementación de modelos de IA en la generación de contenido educativo representa una transformación significativa de este ámbito, dada la capacidad de los LLM para producir, organizar y adaptar contenidos en tiempo real. Con esta posibilidad los planes de estudio pueden ajustarse de manera dinámica a necesidades y contextos específicos con una velocidad no vista hasta ahora.

Un ejemplo de este potencial es el trabajo documentado en la literatura [47] donde se probó la capacidad de GPT-4 para generar objetivos de aprendizaje alineados con la taxonomía educativa de Bloom. Los resultados demostraron que el modelo pudo formular objetivos coherentes con los niveles cognitivos propuestos por Bloom y con el marco disciplinar de referencia. Con ello se demostró que los LLM más allá de redactar texto, pueden contribuir a la estructuración conceptual de la enseñanza, facilitando el diseño curricular. Sin embargo, el artículo también subraya la necesidad de una revisión de parte de expertos educativos, que aporten el juicio crítico y contextual que la experiencia humana es capaz de proporcionar. La idea de estructurar y generar contenido de manera sistemática, emulando el proceso secuencial de aprendizaje humano con preguntas que escalan en dificultad, ha demostrado mejoras sustanciales en el rendimiento de los modelos [26]. De manera complementaria, un estudio [48], muestra cómo los LLM tienen el potencial de agilizar el proceso de desarrollo curricular. No obstante, los autores del trabajo advierten que estos modelos suelen reproducir sesgos culturales presentes en los datos de entrenamiento, lo que conlleva al riesgo de alinear planes de estudio poco representativos. Por ello es que se recalca que la supervisión docente sobre los contenidos generados debe ser permanente para garantizar la pertinencia y adecuación a objetivos institucionales específicos. La creación de contenido a gran escala se facilita con sistemas como GLAN (*Generalized Instruction Tuning*), que utiliza una estructura jerárquica de conocimiento y programas de estudio generados por LLM para crear datos sintéticos de instrucción que cubren un amplio espectro de disciplinas [18]. Además, para la creación de programas de estudio, se ha propuesto el uso de herramientas de análisis de texto para reducir el tiempo y el esfuerzo requerido, sugiriendo temas y contenidos a los profesores [49].

Otros trabajos analizan qué papel pueden desempeñar los LLM en la reconfiguración de los planes de estudio universitarios. Un trabajo [50] concluye que, ante la

capacidad de ChatGPT de realizar tareas complejas, el método educativo y el plan de estudios deben ser reorganizados. De igual manera, un estudio [23] ofrece un conjunto de recomendaciones metodológicas para la integración de LLM en la práctica docente, haciendo hincapié en la necesidad de estrategias pedagógicas actualizadas que pongan especial atención en la descomposición estratégica de problemas y el desarrollo de nuevas habilidades. Por último, una revisión [51] reflexiona sobre el impacto de los LLM en la educación y cómo su introducción exige replantear el rol del profesorado, ya que la AI tiende a redistribuir la autoridad en el proceso de enseñanza.

Entre los beneficios que más destacan están: 1) Agilidad en la producción de materiales didácticos. 2) Alineación automática del diseño curricular. 3) La posibilidad de generar recursos adaptados a distintos perfiles de estudiantes. 4) La capacidad de estos sistemas para ir más allá de la generación de contenido y abordar tareas de apoyo específicas. Entre las limitaciones que más destacan se encuentran el riesgo de sesgos culturales y falta de contextualización y la dependencia excesiva de la generación automática. En conclusión, los LLM deben pensarse como asistentes creativos y no como sustitutos del trabajo docente. Su valor radica en la colaboración estratégica de docentes y modelos de AI generativos, donde la planificación educativa en un futuro dependerá de la capacidad de las instituciones académicas para integrar estas tecnologías en marcos metodológicos éticos y culturalmente pertinentes.

2.4. Análisis Crítico y Retos Estructurales

2.4.1. Ética, Sesgos y Equidad en la Personalización

El uso de LLM en entornos educativos ha despertado un creciente interés por sus posibilidades, pero también por los dilemas éticos que acompañan su integración. Estos modelos no solo median el acceso al conocimiento, sino que también lo producen de manera automática en forma de textos, explicaciones o recomendaciones personalizadas.

Diversos estudios advierten que los LLM pueden reproducir patrones discriminatorios presentes en los datos con los que fueron entrenados. En particular, un trabajo [52] señala que las respuestas pueden variar según el género, nivel socioeconómico o condiciones de discapacidad de los usuarios, lo que plantea interrogantes sobre la equidad en la personalización del aprendizaje. Frente a ello, distintas investigaciones [16] recomiendan la implementación de estrategias para mitigar sesgos y la adopción de prácticas más inclusivas. La implementación de estos modelos requiere de acceso a tecnologías avanzadas que no siempre son accesibles para todos los usuarios, lo que conlleva a la inequidad de acceso a este tipo de herramientas [32].

2.4.2. Integridad Académica y Replanteamiento de la Evaluación

Otro de los temas recurrentes en la literatura es la integridad académica. Como señalan los docentes entrevistados en [37], la generación automática de textos, códigos o argumentos complejos puede diluir la noción de autoría y fomentar prácticas de plagio, tanto consciente como inadvertido. En consecuencia, el uso de LLM obliga a repensar los sistemas de evaluación tradicionales. El estudio [53] expone cómo las pruebas escritas o los trabajos monográficos pierden validez como instrumentos de medición del aprendizaje, y plantea la necesidad de adoptar métodos alternativos. En esta misma línea, otro trabajo [54] subraya que la capacidad de ChatGPT para resolver tareas complejas de manera convincente demanda una reorganización profunda de los métodos de enseñanza y de los planes de estudio, de modo que el aprendizaje se enfoque en el razonamiento, la argumentación y la aplicación práctica del conocimiento. Las tendencias actuales apuntan hacia la diversificación de las estrategias de evaluación [35], incorporando actividades orales, proyectos prácticos y procesos de coevaluación entre pares, con el propósito de salvaguardar la autenticidad y promover la reflexión crítica.

2.5. Pensamiento Crítico, Autonomía y Formación Docente

Un tercer riesgo señalado por la literatura es la posible pérdida de pensamiento crítico y autonomía. El estudio [55] muestra que algunos estudiantes tienden a considerar a los LLM como una fuente de conocimiento autoritaria, en el sentido de que no se puede cuestionar el conocimiento que despliega, lo que puede generar dependencia y limitar la capacidad de análisis propio. La dependencia excesiva de herramientas de inteligencia artificial puede ocasionar una reducción de las habilidades de los usuarios lo que puede conducir a una reducción en el aprendizaje. Por ello es importante que el uso de estas herramientas sea complementario a los métodos tradicionales de enseñanza que fomentan activamente estas habilidades críticas para el aprendizaje [16]. Ante ello, se vuelve indispensable promover estrategias pedagógicas que fomenten la verificación de información, la argumentación y el contraste de ideas, integrando el uso de estas herramientas en un marco de aprendizaje reflexivo.

De este modo, el papel del docente adquiere una nueva relevancia. Ya no se trata solo de transmitir conocimiento, sino de guiar el proceso de interacción con la AI, ayudando a los estudiantes a discernir, cuestionar y comprender cómo se generan las respuestas. Este enfoque, vinculado con la alfabetización digital en IA [56], exige que los educadores se formen en competencias tecnológicas y éticas. Además, la transparencia en los procesos de toma de decisiones de la AI es fundamental para fortalecer la confianza y la apropiación crítica de la tecnología [57].

2.5.1. Riesgos Regulatorios y de Privacidad de Datos

Finalmente, la literatura revisada coincide en que los desafíos éticos y pedagógicos se entrelazan con problemas de gobernanza y regulación. La protección de los datos personales se ha convertido en una preocupación prioritaria. El estudio [50] plantea que el uso de modelos abiertos o locales puede representar una alternativa más segura, al permitir un mayor control sobre el tratamiento de los datos. Por otra parte, la estandarización de criterios de calidad y acreditación se ve comprometida por la incorporación de sistemas dinámicos y no siempre transparentes, donde la opacidad de los algoritmos dificulta que los evaluadores externos verifiquen los procesos pedagógicos [58]. Frente a ello, distintos autores [59] destacan la importancia de avanzar hacia marcos de Explicabilidad de la AI (XAI) que permitan entender cómo operan los modelos y cuáles son sus limitaciones. Solo a través de este tipo de esfuerzos será posible integrar los LLM en la educación de manera responsable, equitativa y alineada con los valores fundamentales del aprendizaje humano.

Capítulo 3

Metodología de Desarrollo

3.1. Diseño y Enfoque de la investigación

La investigación documentada en el presente proyecto de grado es de carácter científico aplicado; el enfoque combina una evaluación cuantitativa y una cualitativa. El diseño de la investigación es de tipo experimental, cuyo objetivo principal es la validación de la hipótesis mediante el desarrollo y prueba de un Modelo de Lenguaje Grande especializado en el currículo de la educación pública en México. Así, el Entorno Virtual de Aprendizaje se implementó como un ambiente controlado de prueba de dicho modelo, permitiendo la recolección de datos para la validación cualitativa.

3.2. Población objetivo y Muestra de Evaluación

- Población objetivo del Prototipo: El Asistente Educativo se diseñó para estudiantes de nivel secundaria en México, alineado con el currículo del modelo de la Nueva Escuela Mexicana.
- Muestra de Evaluación: La validación cualitativa se realizó a través de pruebas piloto con usuarios. Los resultados permitieron mostrar la utilidad percibida y adecuación pedagógica del modelo.

3.3. Fases de la metodología experimental

- Investigación bibliográfica y análisis del contexto de implementación:** A través de una revisión bibliográfica, se seleccionó el modelo de lenguaje más adecuado para las necesidades del proyecto, teniendo en cuenta el contexto educativo de México, las especificidades del currículo y la capacidad de cómputo disponible.
- Metodología de selección y preparación del corpus educativo:** Se definió la estrategia para la recopilación, organización y selección del corpus educativo, el

cual está basado en los materiales de CONALITEG, garantizando que el contenido esté alineado con el currículo de la educación básica en México y es culturalmente pertinente. Esta actividad aseguró que el asistente educativo se adapta a las necesidades del contexto escolar.

- c. **Procesamiento y selección de datos educativos:** Se extrajo el texto de los libros del CONALITEG y se limpió el texto para conseguir un corpus curado. Luego, se construyó un corpus de pares de preguntas y respuestas que sirvió como base de entrenamiento supervisado en el ajuste fino del modelo.
- d. **Ajuste fino del modelo de lenguaje:** Se llevó a cabo el *fine-tuning* del modelo seleccionado con el corpus, especializándolo en los contenidos educativos del currículo mexicano, lo que contribuyó a la personalización contextual de las respuestas generadas por el asistente educativo.
- e. **Establecimiento y aplicación de métricas de evaluación cuantitativas para el modelo ajustado:** Se utilizaron métricas como *BLEU*, *ROUGE* y otras medidas de coherencia semántica como *Meteor* y *BERTScore* para evaluar la calidad de las respuestas del modelo ajustado. Esto permitió comparar el rendimiento del modelo ajustado con el del modelo base y valorar la efectividad del ajuste fino.
- f. **Diseño e implementación de la interfaz gráfica:** Se desarrolló una interfaz accesible y fácil de usar, adecuada para el contexto de la educación pública mexicana, lo que facilitó la interacción de los estudiantes y docentes con el asistente educativo.
- g. **Pruebas piloto y análisis de resultados:** Se implementaron pruebas piloto con usuarios reales para evaluar la calidad pedagógica y el impacto del prototipo en la experiencia de aprendizaje. Los resultados obtenidos permitieron realizar ajustes en el prototipo y valorar su potencial en el contexto educativo.

Capítulo 4

Análisis de Contexto para Selección de LLM

Para el desarrollo del proyecto se contó con dos nodos de cómputo de alto rendimiento, donde las capacidades de esta infraestructura dictaron las restricciones para la selección del modelo de lenguaje grande que se usaría como base para asegurar la estabilidad del entrenamiento en el *hardware* disponible.

a. Entorno de Curación y Procesamiento del Corpus Educativo:

- **Arquitectura:** Servidor basado en **PPC64LE POWER 9**.
- **GPU:** 4 unidades **NVIDIA TESLA V100**.
- **Justificación:** Este entorno fue clave para tareas de procesamiento de datos en paralelo. Fue necesario utilizar un canal específico de la comunidad (*Open CE Channel*) para garantizar la compatibilidad de las librerías de *Transformers* con la arquitectura PPC64LE durante la fase de limpieza asistida del corpus.

b. Entorno de Ajuste Fino (SFT) y despliegue del Prototipo:

- **Arquitectura:** Servidor basado en **X86**.
- **GPU:** 4 unidades **NVIDIA RTX 3090** (24 GB VRAM c/u).
- **Justificación:** Se aprovechó la alta memoria de video de las RTX 3090, crucial para las operaciones de *fine-tuning* cuantizado (*QLoRA*), que demandaban hasta ≈ 22 GB de VRAM por GPU.

Mediante la revisión de la bibliografía disponible, se seleccionó el modelo de lenguaje que se consideró más adecuado para el proyecto. Como criterio principal, se buscó el equilibrio entre las capacidades generativas reportadas en las métricas de desempeño y la disponibilidad de recursos computacionales al momento del desarrollo del proyecto.

Con este análisis se llegó a la conclusión de implementar dos modelos de lenguaje: uno principal para impulsar el *chatbot* al cual se ajustó al currículo de la NEM y otro como herramienta de limpieza de textos para curar el corpus educativo.

- **Modelo para Ajuste Fino:** Se seleccionó **phi-4 14b** de Microsoft como modelo base por ser el modelo de código abierto más eficiente en número de parámetros y resultados de distintos *benchmarks* al momento de la comparación.

Cuadro 4.1: Análisis comparativo de benchmarks para la selección del modelo base.

Categoría	Benchmark	Phi-4 (14B)	Phi-3 (14B)	Qwen 2.5 (14B)	GPT-4o-mini	Llama-3.3 (70B)	Qwen 2.5 (72B)	GPT-4o
Popular Aggregated	MMLU	84.8	77.9	79.9	81.8	86.3	85.3	88.1
Science	GPQA	56.1	31.2	42.9	40.9	49.1	49.0	50.6
Math	MGSM / MATH	80.6 / 80.4	53.5 / 44.6	79.6 / 75.6	86.5 / 73.0	89.1 / 66.3	87.3 / 80.0	90.4 / 74.6
Code Generation	HumanEval	82.6	67.8	72.1	86.2	78.9	80.4	90.6
Factual Knowledge	SimpleQA	3.0	7.6	5.4	9.9	20.9	10.2	39.4
Reasoning	DROP	75.5	68.3	85.5	79.3	90.2	76.7	80.9

Fuente: Elaboración propia con datos de *Hugging Face – Microsoft/Phi-4*.

La justificación de esta elección se basa en tres puntos:

- Eficiencia y Recursos:** Los modelos de mayor tamaño (70B) superaban la capacidad de *fine-tuning* que podíamos alcanzar con el cómputo disponible ($4 \times$ RTX 3090).
 - Rendimiento en Tareas Clave:** En su categoría de tamaño, **phi-4** demostró ser el mejor en las métricas de interés: **MMLU (84.8)** y **Matemáticas ($\approx 80\%$)**.
 - Apertura y Flexibilidad:** Al ser de código abierto, permitió la aplicación de técnicas de optimización especializadas (*QLoRA*).
- **Modelo para Limpieza de Datos:** Se seleccionó **DeepSeek-R1-Distill-Qwen 7B** por el balance entre capacidades de razonamiento y tamaño reducido.

Capítulo 5

Construcción del Corpus Educativo

Con el objetivo de ajustar el modelo a contenidos pertinentes al contexto educativo mexicano, se procesó el texto plano de los libros de texto gratuitos de la **CONALITEG**. Este material se utilizó como insumo para la construcción de un corpus de preguntas y respuestas (Q&A) y una base de datos vectorial, los cuales constituyen los productos finales de esta fase. Específicamente, los contenidos integrados en el *pipeline* de procesamiento corresponden al nivel de educación **secundaria y telesecundaria**, lo que permitió delimitar el alcance del proyecto y garantizar la alineación con los objetivos de la NEM. El proceso de preparación y curación del corpus se llevó a cabo en cinco etapas fundamentales:

- a. **Extracción de texto:** Se empleó la librería PyMuPDF de Python para realizar la extracción inicial del texto plano a partir de los archivos en Formato de Documento Portátil (PDF, por sus siglas en inglés) originales. Esta herramienta permitió acceder a la capa de texto manteniendo la integridad de los caracteres especiales.
- b. **Limpieza Preliminar:** Se utilizaron expresiones regulares (librería **re**) para eliminar información irrelevante como números de página, encabezados, pies de página y errores comunes de extracción. Posteriormente, el texto fue segmentado utilizando el tokenizador de **Qwen-7B** con un tamaño de ventana máximo de 2,048 tokens para asegurar la consistencia en el procesamiento posterior.
- c. **Limpieza Asistida por LLM:** Se implementó el modelo **DeepSeek-R1-Distill-Qwen-7B** con un *prompt* especializado para la curación educativa. El modelo fue instruido para recuperar exclusivamente el contenido pedagógico, corregir errores de formato comunes en el Reconocimiento Óptico de Caracteres (OCR, por sus siglas en inglés) y reestructurar el texto en formato **Markdown**, incluyendo la representación de ecuaciones en **L^AT_EX**. El *prompt* utilizado se detalla a continuación:

El siguiente texto es parte de un libro educativo.
↪ Debes devolver el texto
completo, exactamente igual salvo que:

1. Corrija errores comunes de extracción:
 - Uniones incorrectas ('his - toria' -> 'historia').
 - Fragmentaciones de párrafos.
 - Errores OCR ('ci6n' -> 'ción', 'a±o' -> 'año').
2. Elimine exclusivamente:
 - Nombres de autores y créditos editoriales.
 - Notas legales e ISBN.
 - Referencias a imágenes (ej: 'Ver Figura 3').
 - Números de página y pies de página.
3. Modificaciones obligatorias:
 - Conserva títulos y subtítulos originales.
 - Normaliza nombres: 'Malinche' -> 'Malintzin (La ↪ Malinche)'.
 - Formatea citas con "> Texto" (Fuente, año).
 - Destaca datos clave en ****negritas****.

d. **Validación Manual:** Se realizó una revisión humana meticulosa sobre una muestra representativa para identificar y corregir posibles alucinaciones del modelo o secciones de contenido pedagógico omitidas erróneamente durante la limpieza automática.

e. **Estructuración del Dataset para SFT (Q&A):** Se generó un conjunto de entrenamiento de pares pregunta-respuesta. Se instruyó al modelo para actuar como un generador de datos para SFT, asegurando que cada Objeto en Notación JavaScript (JSON, por sus siglas en inglés) correspondiera a una sola idea o hecho del texto, incluyendo metadatos de capítulo y tema.

El resultado final fue un dataset de **11,850 pares Q&A**, equivalente a un volumen total del corpus de aproximadamente **1.185 millones de tokens**.

5.1. Formato de Entrenamiento y División del Dataset

Cada par de pregunta y respuesta se integró en un *template* de conversación diseñado para el modelo Phi-4, estructurado de la siguiente manera:

```
<|user|>
Pregunta: {pregunta}
Contexto:
- Grado: {grado}
- Asignatura: {asignatura}
- Tema: {tema}
- Capítulo: {capitulo}
```

```
- Fuente: {fuente}

<|assistant|>
Respuesta: {respuesta}
```

Para garantizar una evaluación robusta que cubra todas las áreas del conocimiento, el *dataset* se dividió por asignatura siguiendo una proporción del 95 % para entrenamiento y 5 % para validación:

- **Dataset de Entrenamiento (D_{train}):** 10,665 pares.
- **Dataset de Prueba/Validación (D_{test}):** 593 pares.
- **Criterio de consistencia:** Las asignaturas con una representación crítica (menos de 20 ejemplos) fueron asignadas íntegramente al conjunto de entrenamiento para no degradar la capacidad de aprendizaje en temas poco frecuentes.

5.2. Consideraciones Éticas y de Propiedad Intelectual

Con el objetivo de garantizar el uso responsable de la información y el cumplimiento de las normativas referentes a Derechos de Autor de las fuentes utilizadas para la construcción de los corpus, se declaran los siguientes lineamientos:

- **Limitaciones por Propiedad Intelectual:** El contenido de texto plano extraído de los materiales de los libros de la CONALITEG mencionados previamente, se utilizó exclusivamente como base para la creación del corpus final de instrucción (Q&A) y la base de datos vectorial; es decir, estos contenidos fungieron como un insumo en el *pipeline* de desarrollo y nunca como un componente destinado a la difusión pública, asegurando así la protección de la propiedad intelectual de la fuente original.
- **Análisis de Disponibilidad de Datos:** Con el fin de fomentar la transparencia en la reproductibilidad científica del proyecto, se identifica que el corpus de instrucción creado posee una naturaleza transformativa al ser un material derivado. Por ello, es posible compartir dicho recurso bajo licencias académicas restrictivas. Es pertinente recalcar que cada par de pregunta-respuesta de este corpus preserva la integridad de la fuente original mediante metadatos de proveniencia. Asimismo, se establece que el proyecto tiene una finalidad social y no comercial, por lo que la posible publicación de este corpus es viable siempre que se validen los lineamientos dados por los marcos de datos abiertos de la administración pública.

Capítulo 6

Ajuste Fino Supervisado del LLM

A continuación se describe el proceso de entrenamiento del modelo Phi-4 detallando las estrategias de optimización y la configuración de los hiperparámetros implementada para ello.

6.1. Estrategias de Optimización y Eficiencia Computacional

El modelo Phi-4 cuenta con 14B de parámetros, lo que en precisión total (*Full Fine-Tuning*) requeriría ≈ 56 GB de VRAM solo para cargar los pesos. Aunque la infraestructura disponible para el proceso de ajuste fino suma un total de **96 GB de VRAM** (4×24 GB), se implementaron técnicas de optimización para asegurar un entrenamiento ágil que redujera el consumo de recursos computacionales y energéticos, garantizando un manejo eficiente de los estados del optimizador y las activaciones durante las fases de mayor carga de memoria:

- a. **Cuantización (QLoRA):** Se aplicó una cuantización de **4 bits** (NF4), lo que redujo el peso del modelo en la memoria a ≈ 7 GB por dispositivo. Esta estrategia permitió liberar espacio en la VRAM para gestionar procesos de acumulación de gradiente sin riesgo de desbordamiento de memoria.
- b. **Ajuste Fino de Eficiencia de Parámetros (PEFT - LoRA):** Se utilizó la técnica *Low-Rank Adaptation* (LoRA) para actualizar únicamente matrices de bajo rango y así no reentrenar todo el modelo. Esto acelera el entrenamiento y al mismo tiempo evita la degradación de los pesos pre-entrenados del modelo base. La configuración de LoRA fue:
 - **LoRA Rank (r):** 8.
 - **LoRA Alpha (α):** 32.
 - **Módulos de Destino (*Target Modules*):** Matrices de atención (q_proj, k_proj, v_proj, y o_proj).

6.2. Configuración del Experimento e Hiperparámetros

El entrenamiento se realizó siguiendo un esquema de aprendizaje supervisado utilizando la función de pérdida de entropía cruzada (*Cross-Entropy Loss*). Los parámetros seleccionados fueron:

- **Optimizador:** AdamW, por su robustez en la convergencia de modelos de lenguaje.
- **Tasa de Aprendizaje (*Learning Rate*):** 2×10^{-4} .
- **Tamaño de Lote Global:** 16. Se configuró un *batch* por dispositivo de 1 con 16 pasos de acumulación de gradiente para estabilizar la actualización de los pesos.
- **Longitud Máxima de Secuencia:** 512 *tokens*.
- **Ciclos de Entrenamiento:** 1 Época.

Justificación del entrenamiento a una sola época: Dado el tamaño del corpus (11,850 pares) y su naturaleza específica se decidió realizar el entrenamiento por una única época para mitigar el riesgo de **sobreajuste**. Esto para garantizar que el modelo aprenda la información de la CONALITEG sin perder su capacidad de generalización.

6.3. Preparación para el Despliegue

Tras finalizar el proceso de *SFT*, los pesos adaptativos se mantuvieron en la configuración de inferencia cuantizada. El modelo resultante (LLM_{SFT}) con cuantización a 4 bits permitió que el prototipo final funcione con una latencia baja en el servidor de despliegue.

Capítulo 7

Arquitectura de Recuperación y Contextualización

7.1. El Paradigma RAG y análisis de viabilidad de implementación

El paradigma de Generación Aumentada por Recuperación surgió como una estrategia para reducir las alucinaciones de los LLM. La idea central de dicho paradigma consiste en la consulta y recuperación de fuentes de datos externos al modelo de lenguaje preentrenado; con ello se le confiere al modelo la capacidad de generar respuestas que contengan datos que no necesariamente estaban incluidos en la etapa de entrenamiento del modelo. Con la implementación de esta estrategia se consigue la trazabilidad y verificación de las fuentes que sustentan las respuestas inferidas por el modelo.

Inicialmente se planteó implementar este paradigma para establecer la relación entre las consultas de los usuarios del Entorno Virtual de Aprendizaje con los contenidos oficiales de la CONALITEG; se creó una base de datos vectorial de los contenidos educativos (*embeddings*) y se propuso la implementación de una Búsqueda de similitudes de Inteligencia Artificial de Facebook (FAISS, por sus siglas en inglés) para recuperar los contenidos educativos. Sin embargo, al hacer un análisis de viabilidad y restricciones técnicas del entorno real de despliegue del prototipo se identificaron limitaciones de infraestructura que dificultaban dicha implementación de recuperación de información. Es por ello que para la arquitectura de información del prototipo TutorIA se transicionó hacia una estrategia de inyección de contexto dinámico, inspirada en el paradigma RAG, con el fin de garantizar los requerimientos de eficiencia y estabilidad en contextos educativos rurales en México. A continuación se describen las principales limitaciones identificadas en el entorno real de implementación del prototipo:

- **Infraestructura de red limitada:** La implementación de la técnica RAG sigue una secuencia de pasos que van de la codificación de la consulta del usuario (*encoding*) a la búsqueda de similitud semántica en tiempo real entre la consulta

codificada y la base de datos vectorial de los contenidos educativos. Este proceso introduce un retardo en la generación de la respuesta a la consulta del usuario, lo que aunado al hecho de que el ancho de banda de la red en el contexto educativo rural en México suele ser limitado, podría comprometer la experiencia de usuario y por ende la capacidad pedagógica de un EVA.

- **Carga Computacional:** Adicionalmente, el proceso de implementación del paradigma RAG incrementa los requerimientos de hardware del servidor, dificultando la escalabilidad del sistema.

7.2. Inyección de Contexto Dinámica

Dadas las limitaciones mencionadas en la sección anterior, se decidió transicionar hacia un enfoque inspirado en la inyección de contexto dinámico de datos externos al modelo que se hace en paradigma RAG, pero prescindiendo de procesos de transformación y búsqueda vectoriales, con el fin de optimizar la ejecución de la recuperación dinámica de contexto. Se apostó por asumir que gracias al ajuste fino supervisado había conseguido que el modelo base interiorizara los contenidos educativos de la CONALITEG, por lo que no sería necesaria una búsqueda vectorial exhaustiva en tiempo real para cada consulta. En su lugar se implementó una técnica de inyección de contexto dinámico que funciona de la siguiente manera:

- Recuperación de Metadatos:** El sistema identifica metadatos desde la capa de datos (backend) del prototipo, que como se mencionó anteriormente, almacena datos del alumno en Lenguaje de Consulta Estructurado (SQL, por sus siglas en inglés).
- Construcción del Prompt Enriquecido:** Estos metadatos se inyectan automáticamente para construir un *prompt* enriquecido antes de la inferencia, dicho *prompt* utiliza los metadatos como restricciones de contexto para la inferencia
- Focalización Curricular:** Esta inyección actúa como una guía para el modelo ajustado con el fin de asegurar que las respuestas se mantengan dentro de los límites del currículo nacional sin la necesidad de consultar una base de datos externa en cada iteración.

Con esta estrategia de inyección dinámica de contexto se consigue lo siguiente:

- **Contextualización Curricular:** Se inyectan automáticamente el **grado académico** para que el modelo conozca el nivel de complejidad y los conceptos adecuados para el estudiante.
- **Adaptación Pedagógica:** Se integran las **preferencias de estilo de aprendizaje** del usuario, instruyendo al modelo para que adapte sus explicaciones, analogías y ejemplos.

- **Determinismo Curricular:** Se consigue que el modelo utilice información del nivel educativo correspondiente eliminando el riesgo de recuperar fragmentos de otros grados por similitud semántica errónea.
- **Análisis de Sentimiento:** Se incorpora el resultado del análisis de sentimiento de **DistilBERT**. Si se detecta frustración o confusión en el mensaje del alumno el prompt se ajusta dinámicamente para que el tono del tutor sea más paciente y alentador.

La transición de un sistema de recuperación vectorial hacia una arquitectura de inyección fue así una optimización de ingeniería basada en el contexto real de implementación. Este enfoque garantiza que el sistema sea lo suficientemente ligero para operar en condiciones de baja conectividad manteniendo la pertinencia pedagógica necesaria para la educación básica en México.

Capítulo 8

Desarrollo del Prototipo

A continuación se documenta a detalle la implementación de **TutorIA**, el prototipo de EVA desarrollado en tres componentes principales: la capa de servicios (*backend*), la interfaz de usuario (*frontend*) y la base de datos, de tal manera que se integra el modelo ajustado previamente en una plataforma educativa de tutoría personalizada con elementos de *gamificación*.

8.1. Arquitectura y Flujo de Interacción

La arquitectura de la plataforma está desarrollada siguiendo un esquema de microservicios, donde la inteligencia de tutoría (impulsada por el LLM ajustado) está contenida en el *backend* (FastAPI).

Diagrama de la Arquitectura

La Figura 8.1 ilustra la interconexión de las tres capas lógicas.

La arquitectura se divide en:

- **Capa de Presentación:** Aplicación web modular con React/TypeScript.
- **Capa de Servicios:** Servidor REST FastAPI con *endpoints* asincrónicos.
- **Capa de Datos:** Base relacional con SQLAlchemy.

Diagrama de Secuencia de Interacción

A continuación se muestra el flujo de una sesión de aprendizaje personalizada con el prototipo TutorIA.

El flujo de interacción es este:

Autenticación → Diagnóstico (VARK/Conocimiento) → Consulta al LLM
→ Respuesta y Gamificación → Registro de Progreso.

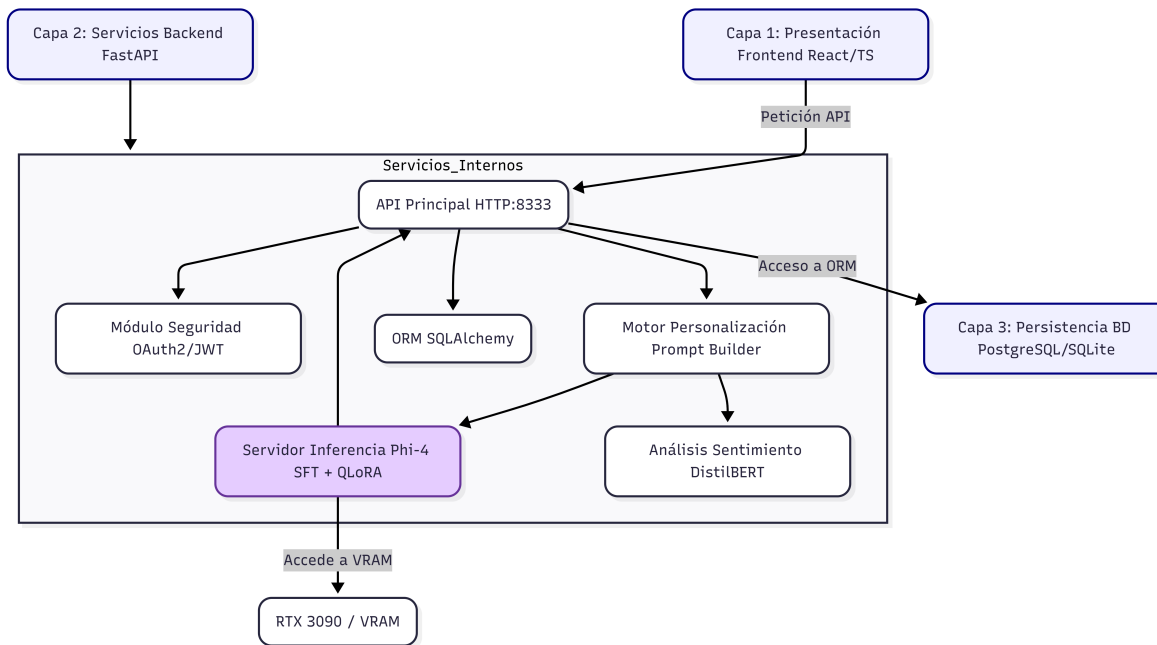


Figura 8.1: Arquitectura de Despliegue de Tres Capas del Prototipo TutorIA.

8.2. Capa de Servicios (Backend)

Endpoints y Arquitectura REST

La Capa de Servicios expone un conjunto de *endpoints* en el puerto **8333** para el manejo del flujo de datos y la inferencia del LLM.

Cuadro 8.1: Endpoints Principales del Backend

Endpoint	Método	Descripción	Input (schemas)	Seguridad
/token	POST	Autenticación y obtención de JWT.	OAuth2PasswordRequestForm	Pública
/users/me/	GET	Recuperación de perfil y estilos.	N/A	JWT
/chat	POST	Tutoría socrática y empática con LLM SFT.	TextInput	JWT
/analyze-sentiment	POST	Clasificación de la emoción del input.	TextInput	Pública
/generate-personalized-content	POST	Contenido adaptado al estilo VARK.	TextInput	JWT
/users/{id}/knowledge-assessment/	POST	Registro de evaluación inicial.	KnowledgeAssessmentCreate	JWT

Lógica de Personalización y Prompting

La personalización se realiza en el *endpoint* /chat a través de un *Prompt*, que considera los datos del usuario, estilo de aprendizaje y análisis de sentimiento.

```
@router.post("/chat")
async def chat(entrada: TextInput, usuario_actual: User =
    ↳ Depends(get_current_user)):
    # Perfil del usuario
```

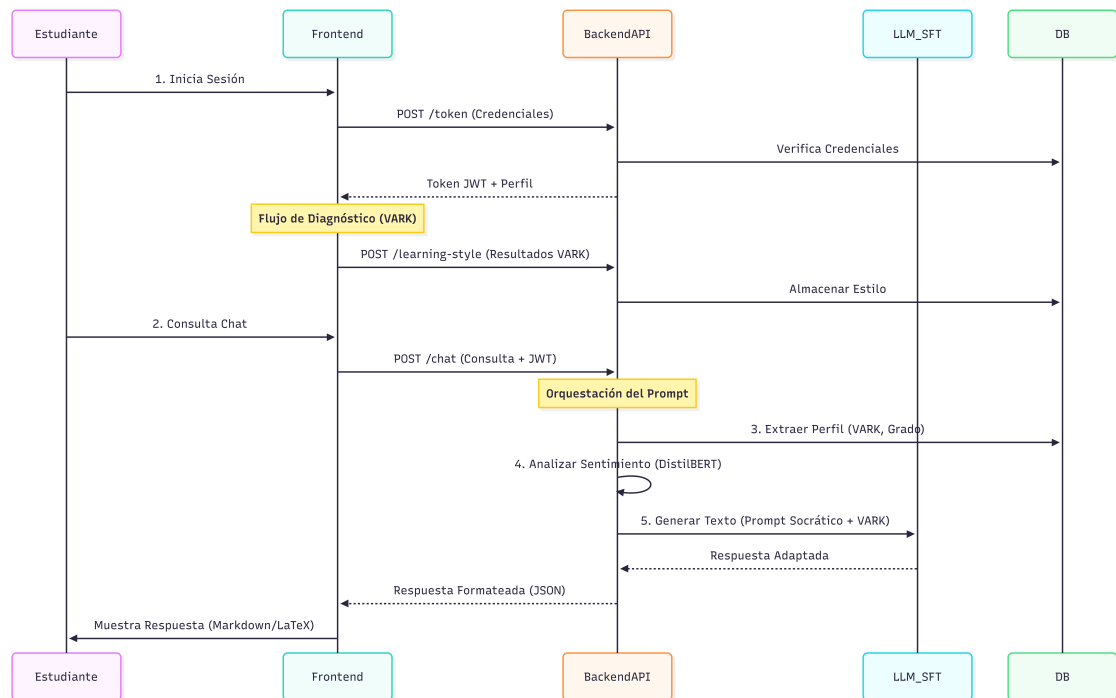


Diagrama de Secuencia: flujo estudiante → LLM → gamificación → progreso.

```

perfil = await crud.get_user_profile(usuario_actual.id)
estilo_aprendizaje = perfil.learning_style

# Estado de ánimo de la entrada del alumno
estado_animo = sentiment_pipeline(entrada.text)[0]['label']

# Prompt del tutor personalizado
prompt_sistema = f"""
Eres TutorIA, un profesor socrático, empático y experto en el plan
↳ de estudios de {user_escolaridad} de México (12-16 años).
Tu alumno se llama {user_name}, está en {user_grade}° grado y su
↳ estilo de aprendizaje dominante es {learning_style}.

REGLAS IMPORTANTES:
1. Seguridad: Nunca proporciones información inapropiada, violenta,
↳ sexual o sobre drogas. Si el alumno pregunta algo así,
↳ responde:
'Lo siento, como tutor académico, solo puedo ayudarte con temas
↳ educativos.'
2. Estilo Socrático: Guía al alumno mediante preguntas para que él
↳ deduzca la solución.

```

```

3. Personalización: Adapta la complejidad y los ejemplos al nivel
  ↳ de {user_grade}° grado y utiliza ejemplos que refuercen el
  ↳ estilo {learning_style}.
4. Explicación Detallada: Antes de plantear la pregunta socrática,
  ↳ proporciona una explicación clara y detallada (al menos 2
  ↳ párrafos de 3-4 líneas cada uno).
"""

# LLM
respuesta = llm.generate_text(prompt_sistema, temperature=0.01,
  ↳ do_sample=False)

# Retornar la respuesta al frontend
return {"respuesta": respuesta}

```

Manejo de Errores y Seguridad

- **Seguridad de Acceso:** Protección de *endpoints* críticos con OAuth2/JWT.
- **Manejo de Fallas:** Bloques `try-except` para dar *fallback* si el LLM no está funcionando.
- **Reglas de Seguridad:** Instrucciones de seguridad explícitas en el *Prompt* para prevenir contenido inapropiado.

8.3. Capa de Datos

Esquema de Tablas y Relaciones

A continuación se muestra el esquema de las bases de datos.

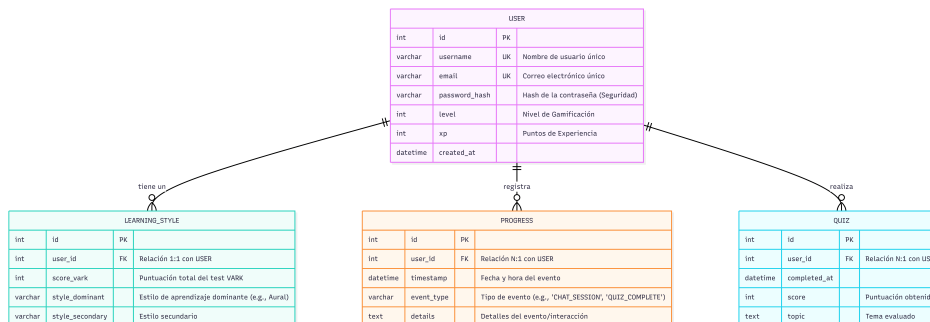


Figura 8.2: Diagrama tablas relacionales (bases de datos)

Parámetros de Inferencia

Los parámetros de generación del LLM SFT se ajustaron para garantizar **estabilidad, precisión y control de longitud de las respuestas**.

Cuadro 8.2: Parámetros de Generación del LLM SFT

Parámetro	Valor	Justificación
temperature	0.01	Minimiza la aleatoriedad (determinista).
do_sample	False	Desactiva el muestreo estocástico.
top_p	0.5	Limita la selección de tokens al conjunto altamente probable, manteniendo el contenido en el dominio curricular.
max_new_tokens	256	Controla la longitud de la respuesta.

8.4. Capa de Presentación(Frontend)

La capa de presentación es una **Aplicación Cliente Web** modular e interactiva para soportar el flujo de personalización y *gamificación*.

Flujo de Personalización y Captura de Datos

El Frontend gestiona la interacción del usuario a través de los módulos clave:

- **Módulos de Diagnóstico:** un módulo de conocimientos y un cuestionario VARK recopilan los datos para la personalización.

Módulos de Interacción

- **AITutor** (`AITutor.tsx`): Interfaz de conversación que muestra las respuestas socráticas y empáticas del LLM SFT.
- **Gamificación:** Módulos como `DailyChallenge`, `QuickQuiz` y `EscapeRoom` integran el sistema de puntos (`totalPoints`) para incentivar la práctica.

Gestión de Comunicación y Presentación de Contenido

- **Comunicación Asíncrona:** Uso de `fetch/axios` para manejar las llamadas al puerto **8333**.
- **Renderizado Técnico:** Se utilizan librerías de **React** para interpretar y presentar fielmente el formato **Markdown** y las fórmulas $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ generadas por el LLM_{SFT} .



Cuestionario de Estilo de Aprendizaje VARK

Hola , completa este cuestionario para que nuestra IA personalice tu experiencia de aprendizaje según tu estilo preferido.

Progreso

Pregunta 1 de 16

¿Cómo prefieres aprender sobre un tema nuevo?

☐  Viendo diagramas, gráficos o videos

☐  Escuchando explicaciones o podcasts

☐  Leyendo artículos o libros

☐  Practicando o experimentando directamente

Siguiente Pregunta →

 Visual

 Auditivo

 Lectura

 Kinestésico

Figura 8.3: Registro y Cuestionario VARK

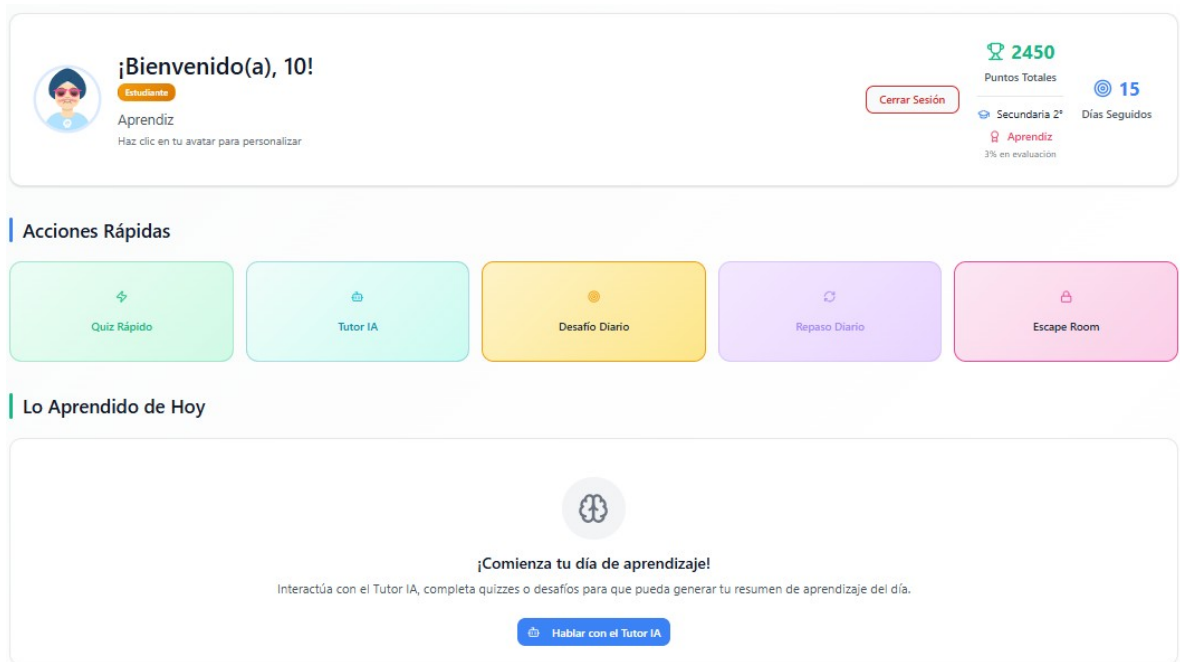


Figura 8.4: Dashboard Principal con los distintos módulos de Aprendizaje.



Figura 8.5: Interfaz del Tutor Virtual de Aprendizaje con Temporizador Pomodoro Integrado.

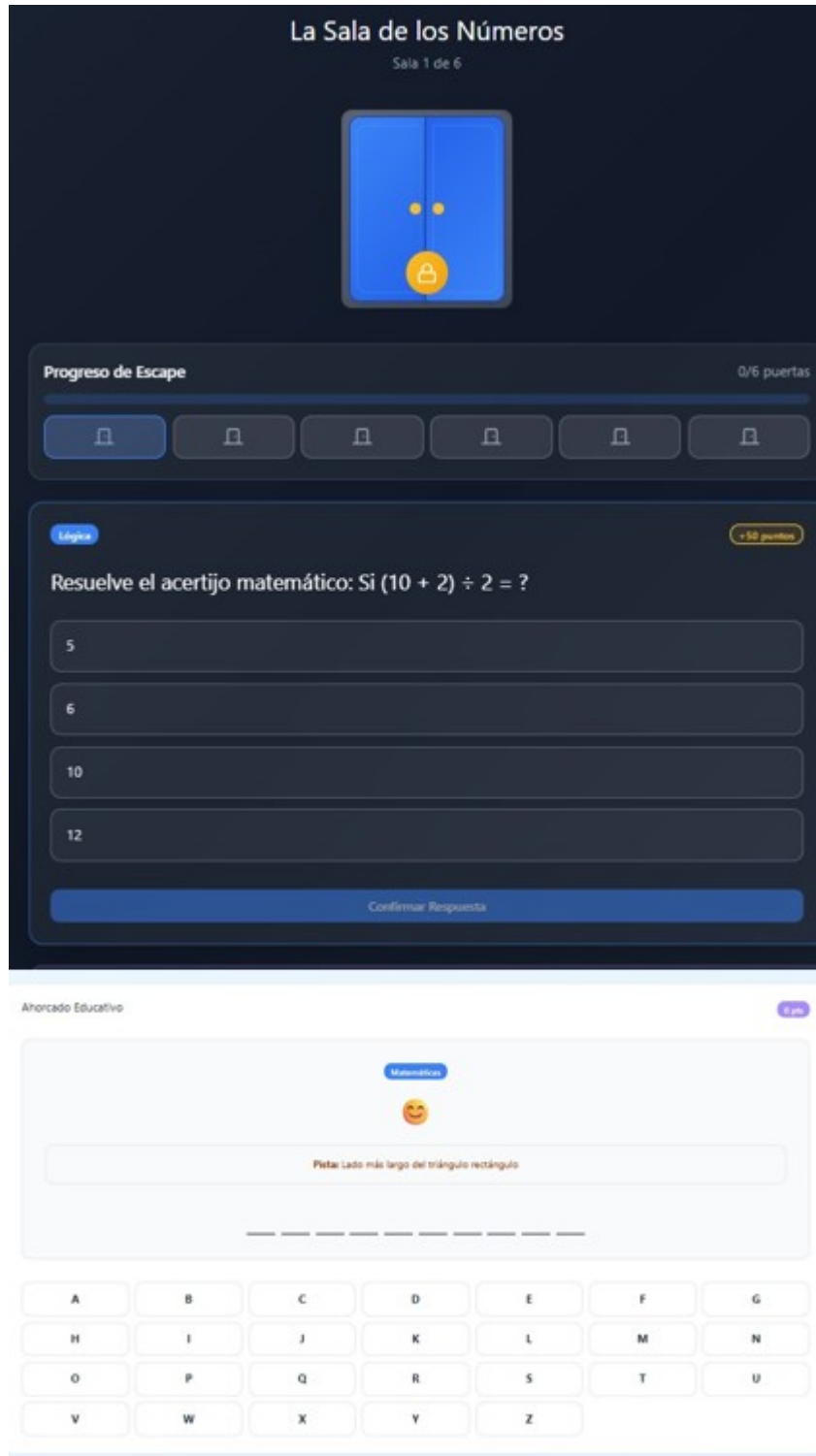


Figura 8.6: Módulo de Gamificación

Capítulo 9

Metodología de Evaluación y Diseño Experimental

Para validar la eficacia del prototipo desarrollado, se diseñó un marco metodológico estructurado en dos fases: la primera de carácter cuantitativo, mediante pruebas de *benchmarks* clásicos de Procesamiento de Lenguaje Natural (NLP) con el objetivo de medir el impacto que tuvo el Ajuste Fino Supervisado (SFT) en el modelo base. La segunda fase fue de carácter cualitativo y consistió en el desarrollo de pruebas piloto del prototipo en un entorno real de implementación; en dichas pruebas se diseñó una etapa de intervención pedagógica con el fin de recopilar información sobre la pertinencia del sistema en el contexto educativo y una etapa de uso libre del prototipo con el fin de recopilar información sobre la experiencia de usuario.

9.1. Metodología de Evaluación Cuantitativa

El objetivo de esta fase fue medir la precisión léxica y la calidad semántica del modelo después del proceso de Ajuste Fino Supervisado. Para ello se diseñó un experimento controlado para comparar el desempeño de cuatro variantes del modelo sobre un conjunto de validación de 593 pares de instrucción-respuesta reservados del corpus educativo original.

9.1.1. Variantes del Modelo en Evaluación

Con el objetivo de aislar el efecto del entrenamiento y la inyección de información se definieron las siguientes configuraciones de modelo a evaluar:

- **Phi-4 Base:** El modelo original sin modificaciones que sirve como control para establecer la línea base de conocimiento de los modelos.
- **Phi-4 Ajustado (SFT):** El modelo resultante tras el proceso de ajuste fino con el corpus de la CONALITEG para medir el conocimiento internalizado en los pesos del modelo especializado.

- **Phi-4 Ajustado (SFT) + Contexto:** Una variante también de control que simula un **entorno RAG ideal**. En esta configuración, se inyecta directamente la referencia bibliográfica exacta (*Ground Truth*) en el *prompt* para evaluar el techo de capacidad del modelo ajustado cuando dispone de información externa.
- **Llama-3 70B:** Un modelo de la empresa Meta de gran escala que permitió evaluar si un modelo especializado de tamaño medio (14B) puede competir en pertinencia pedagógica con uno mucho más grande pero generalista.

9.1.2. Métricas de Procesamiento de Lenguaje Natural (NLP)

Métricas de Evaluación cuantitativa seleccionadas

Se emplearon cuatro de las métricas más utilizadas en el área de Procesamiento de Lenguaje Natural (NLP):

- **BLEU:** Utilizada para medir la precisión léxica y observar la adopción del vocabulario específico de los recursos educativos del CONALITEG.
- **BERTScore:** Métrica que considera *embeddings* contextuales para medir la similitud, implementada para no penalizar las respuestas pedagógicamente correctas que utilizan sinónimos.
- **METEOR:** Utilizada para analizar la estructura secuencial.
- **ROUGE-L:** Utilizada para analizar la flexibilidad lingüística.
- **Eficiencia por Longitud:** Conteo de palabras promedio para analizar la capacidad de síntesis y el sesgo de verbosidad.

9.2. Metodología de Evaluación Cualitativa: Intervención en Campo

La evaluación cualitativa se diseñó como un estudio de caso en una escuela Telesecundaria rural para observar la interacción entre los alumnos y el sistema en condiciones reales de infraestructura. Con este fin, se definió una prueba experimental que permitiera evaluar la plataforma **TutorIA** como una herramienta de apoyo pedagógico.

Para dar cumplimiento a los criterios éticos de la investigación con menores de edad, previo al inicio de las actividades en el aula se gestionó la autorización de los padres de familia mediante una Carta de Consentimiento Informado (ver Apéndice A). De igual forma, se obtuvo el asentimiento voluntario de los alumnos participantes (ver Apéndice B) y la validación institucional del profesor del grupo a cargo de las sesiones (ver Apéndice C).

9.2.1. Sujetos de Estudio y Contexto de Implementación

Las pruebas piloto contaron con la participación de 46 alumnos de dos niveles educativos, distribuidos de la siguiente manera:

- N=46
- 21 Estudiantes de Primer grado de Secundaria.
- 25 Estudiantes de Tercer grado de Secundaria.

Dadas las condiciones del entorno, las pruebas se realizaron en lotes pequeños para mitigar las limitaciones de infraestructura de red; sin embargo, durante el primer día, una concurrencia de 11 alumnos simultáneos evidenció los límites críticos de la conectividad escolar.

Consideraciones Éticas y Privacidad de la Información

Para garantizar la integridad y privacidad de los menores participantes se implementó un protocolo de anonimización de datos asegurando que las interacciones almacenadas en la base de datos no pudieran ser vinculadas con alumnos específicos, cumpliendo con los estándares de protección de datos personales en entornos de investigación educativa.

9.2.2. Diseño de la Intervención Pedagógica (Proyectos Transversales)

Para evaluar la pertinencia didáctica del modelo de lenguaje se definieron dos secuencias de aprendizaje basado en proyectos transversales: *Cosecha Sostenible* y *Presupuesto Escolar Responsable*, que se pueden consultar en detalle en los Apéndices D y E, respectivamente. Es importante precisar que la actividad pedagógica estuvo centrada exclusivamente en la interacción con el **chatbot** para resolver dudas técnicas que surgieran en el proceso de desarrollo de los proyectos transversales.

9.2.3. Protocolo de la Prueba Piloto

Las pruebas piloto siguieron la siguiente secuencia de fases para su desarrollo:

- Asentimiento e Inducción:** Fase de explicación de los objetivos de la investigación.
- Evaluación Inicial (*Pre-test*):** Aplicación de un cuestionario diagnóstico (respondido a lápiz) para identificar conocimientos previos sobre los objetivos de aprendizaje de los proyectos propuestos, es decir, sobre el proyecto del huerto o la cooperativa.

- c. **Interacción Mediada por el Chatbot:** Fase de resolución de dudas específicas utilizando el chat de **TutorIA** para encontrar las explicaciones que facilitaran el desarrollo del proyecto transversal en cuestión.
- d. **Evaluación Final (*Post-test*):** Aplicación del mismo cuestionario (respondido a tinta) al finalizar el desarrollo del proyecto transversal para registrar cambios en la comprensión de los temas.
- e. **Exploración Libre del EVA:** Tras concluir la actividad dirigida, los alumnos dispusieron de una sesión libre de 60 minutos para interactuar de forma autónoma con todos los módulos de la plataforma (incluyendo herramientas de gamificación y navegación general) para probar la experiencia de usuario completa.

Es importante mencionar que hubo un receso entre la fase de los proyectos transversales y la actividad de exploración libre del prototipo.

9.2.4. Instrumentos de Percepción del Usuario

Posterior a la sesión de exploración libre se aplicaron encuestas basadas en la **Escala de Likert de 5 puntos** que puede consultarse en detalle en el Apéndice F. Este instrumento tuvo el objetivo de medir la usabilidad del sistema completo y la satisfacción de los usuarios acerca de la interfaz. Se agruparon los reactivos en tres dimensiones:

- **Diseño e Interfaz:** Navegación, estética y facilidad de uso del Prototipo. Esta sección estuvo comprendida por las preguntas 1, 2, 3 y 5 de la encuesta.
- **Eficacia Pedagógica y calidad del contenido:** Percepción de claridad y utilidad del contenido. Esta sección estuvo comprendida por las preguntas 6, 7, 8 y 9 de la encuesta.
- **Satisfacción Global y Usabilidad:** Intención de uso futuro, control de frustración y experiencia general con la plataforma. Esta sección estuvo comprendida por las preguntas 4, 10, 11 y 12 de la encuesta (tomando en consideración que el reactivo 4 correspondiente a la frustración opera como un ítem inverso).

9.3. Justificación del Marco de Evaluación

Este diseño experimental busca confrontar la solidez técnica del modelo en un entorno de laboratorio con su utilidad práctica en un entorno real de implementación ante las variables externas de un aula real. Al permitirle a los alumnos una fase de exploración libre del Prototipo de EVA previa a la encuesta de satisfacción se buscó que la percepción del usuario recolectada en la escala de Likert considere la plataforma en su totalidad para validar no solo la inteligencia del modelo sino la viabilidad del entorno virtual como una herramienta educativa para el contexto rural mexicano.

Capítulo 10

Resultados y Evaluación

A continuación se presentan los resultados obtenidos en la investigación tras el proceso de ajuste fino del modelo y tras implementar una arquitectura híbrida que combina el modelo ajustado con la técnica de inyección de contexto inspirada en el paradigma de recuperación aumentada (RAG). El objetivo es evaluar si el proceso de ajuste del modelo Phi-4 con el corpus educativo construido mejora la pertinencia educativa de las respuestas inferidas, así como analizar si la inyección de metadatos bibliográficos contribuye a ello.

10.1. Resultados Cuantitativos del Modelo Ajustado

A continuación se muestra la tabla comparativa de los resultados obtenidos para las variaciones evaluadas:

Cuadro 10.1: Comparativa extendida de métricas de desempeño entre las variantes del modelo.

Métrica	Phi-4 Base	Llama-3 70B	SFT	SFT + Contexto
BLEU (x10)	0.1700	0.1921	0.3182	0.1403
BERTScore	0.6752	0.6810	0.6778	0.6423
ROUGE-L	0.1273	0.1345	0.1169	0.1049
METEOR	0.3002	0.3115	0.1877	0.1949
Eficiencia				
Longitud (Palabras)	90.1	112.4	25.7	60.4

Análisis de los resultados obtenidos

Los resultados más destacados se describen a continuación:

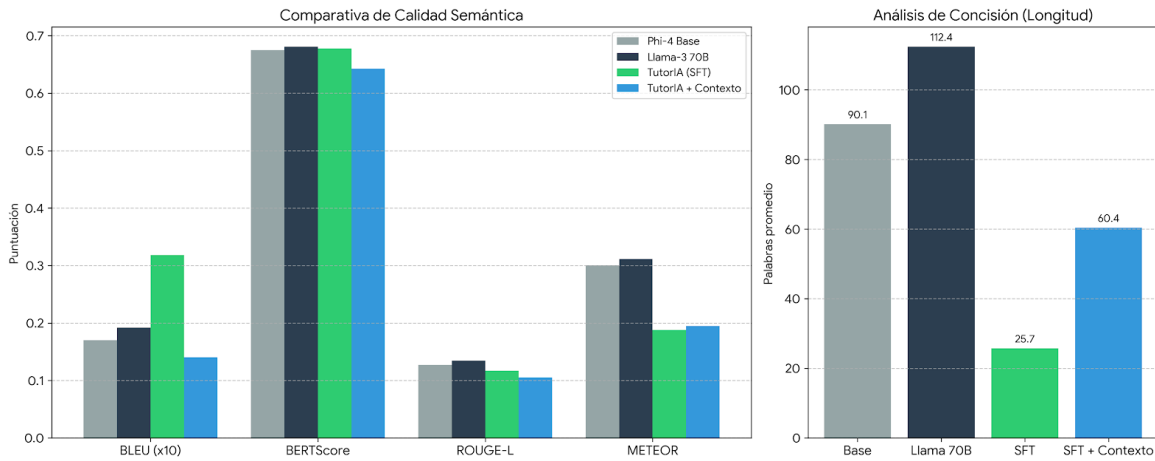


Figura 10.1: Gráfica comparativa de métricas de desempeño y longitud promedio de respuesta.

- Para la métrica BLEU, el modelo Ajustado (SFT) presenta el valor más alto (0.3182), superando incluso al modelo de gran escala Llama-3 70B (0.1921) y casi duplicando el puntaje del modelo base. Esto indica un dominio superior del léxico en el contexto de la Nueva Escuela Mexicana, demostrando que la especialización del dominio es más efectiva que el tamaño del modelo para la precisión terminológica.
- Para la métrica BERTScore, se mantiene un rango estrecho entre todas las variantes. Es notable que el modelo ajustado (SFT) iguala el desempeño semántico de Llama-3 70B (0.6778 vs 0.6810 respectivamente), lo que sugiere que la especialización compensa la diferencia en la escala de parámetros.
- Para las métricas ROUGE-L y METEOR, los modelos base y Llama-3 70B obtienen las puntuaciones más altas. Sin embargo, existe una correlación directa entre este resultado y la extensión de las respuestas, mientras que el modelo SFT + Contexto se posiciona como el límite superior de referencia para una estructura pedagógica controlada.
- En cuanto a la longitud promedio, Llama-3 70B y el modelo base generan las respuestas más extensas (superando las 90 palabras), mientras que el modelo ajustado logra una síntesis de aproximadamente 25 palabras, optimizando la entrega de información.

10.1.1. Análisis de los resultados obtenidos

El incremento del 87% en la métrica BLEU del modelo Ajustado con respecto al modelo base (y su superioridad ante un modelo de 70B parámetros) es un indicador

sólido de la efectividad del ajuste. Mientras que los modelos generalistas utilizan un vocabulario global, el modelo ajustado demuestra una internalización profunda de la terminología empleada en los recursos educativos del CONALITEG.

Por otro lado, la inclusión de metadatos bibliográficos exactos en el prompt de la variante SFT + Contexto provocó una alteración en la estrategia de decodificación del modelo. Mientras que el modelo SFT opera bajo un esquema de recuperación de los n-gramas guardados en sus pesos durante el entrenamiento, la inclusión del contexto explícito redirige la atención del modelo hacia la generación sintética. El modelo utiliza la información del contexto para dar una respuesta fluida, es decir, parafrasea. Es por esto que se observa una reducción en la métrica BLEU con respecto al modelo base.

Un punto crítico a discutir es la aparente superioridad de los modelos de gran escala en ROUGE-L y METEOR, lo cual se explica mediante dos factores:

- **Sesgo por verbosidad:** Los modelos base y Llama-3 70B generan respuestas considerablemente más extensas. Estadísticamente, esto aumenta la probabilidad de coincidencia de n-gramas con la referencia, inflando los resultados sin necesariamente mejorar la precisión pedagógica.
- **Eficiencia del modelo ajustado:** El modelo SFT logra una calidad semántica competitiva con el estado del arte (BERTScore) utilizando solo una fracción de las palabras y de los recursos computacionales. Esto permite ofrecer respuestas concisas que eliminan la carga cognitiva innecesaria sin perder información relevante.

Impacto de la inyección de contexto (Límite Superior)

La configuración SFT + Contexto se establece en esta investigación como el límite superior de capacidad. Aunque presenta un puntaje BLEU menor debido a que parafrasea la información para explicarla en lugar de repetirla textualmente, representa el equilibrio ideal para un tutor inteligente por las siguientes razones:

- **Calidad Semántica Estable:** Su BERTScore se mantiene en niveles óptimos, demostrando que la inyección de metadatos bibliográficos fundamenta la respuesta sin degradar la capacidad de razonamiento del modelo ajustado.
- **Función Didáctica Recuperada:** Con una longitud promedio de 60 palabras, esta variante deja de ser un motor de respuestas cortas para convertirse en una guía educativa. La inyección de contexto actúa como un ancla que permite estructurar explicaciones coherentes, contextualizadas y, sobre todo, libres de alucinaciones al restringir la generación al currículo oficial.

10.2. Resultados Cualitativos Pertinencia Pedagógica

A continuación se expone un ejemplo de inferencia comparativo entre el modelo base y la variante de control que se propuso como **estándar de referencia** para medir la eficiencia del ajuste fino y validar la robustez del sistema ante la inyección de metadatos.

Consulta (*prompt*): *Explica, usando ejemplos de México, cómo el movimiento de las placas tectónicas influyó en la formación de la Sierra Madre Oriental y el Golfo de California.*

- **Phi-4 Base:** *...Su formación está estrechamente relacionada con la subducción de la placa de **Farallon** bajo la placa Norteamericana... proceso que comenzó hace aproximadamente 200 millones de años durante el período Jurásico.*
- **Variante de Referencia (SFT + Contexto):** *La formación de ambas se debe a la interacción entre las siguientes placas:*
 - **-Placa Norteamericana:** *Donde se asienta la mayor parte del territorio mexicano.*
 - **-Placa del Pacífico:** *Su roce con la Norteamericana creó la Falla de San Andrés, formando el Golfo de California.*
 - **-Placa de Cocos y Placa Rivera:** *Se deslizan debajo de la Placa Norteamericana (subducción).*

Formación de la Sierra Madre Oriental: *Creada por el plegamiento de rocas debido a fuerzas de compresión, un proceso que dio forma al relieve actual del este de México.*

De este ejemplo es posible evidenciar los siguientes hallazgos:

- **Relevancia curricular:** El modelo base cita la placa **Farallon**; aunque es un dato correcto geológicamente, dicha información no se encuentra en los contenidos de geografía de los planes de estudio de la Nueva Escuela Mexicana. Por otro lado, la variante con contexto cita las placas **Cocos, Rivera y Pacífico**, las cuales son el eje central de los contenidos curriculares propuestos por la Secretaría de Educación Pública.
- **Formato pedagógico:** Se presenta la información mediante estructuras dinámicas (listas y negritas), lo que facilita la lectura rápida y reduce la carga cognitiva, cumpliendo con los objetivos de usabilidad del prototipo.

Esta comparación entre el enfoque de inyección de información y el modelo base se resume en la Tabla 10.2:

Cuadro 10.2: Análisis cualitativo: Comparativa de pertinencia didáctica.

Dimensión	Phi-4 Base	SFT + Contexto
Relevancia	Datos globales/históricos	Datos nacionales/curriculares
Estructura	Párrafo monolítico	Respuestas dinámicas contextuales
Estilo	Enciclopédico denso	Dialógico y pedagógico
Carga Cognitiva	Alta (Exceso de texto)	Optimizada (Conciso)

Conclusión del Análisis de Inferencia

El contraste entre las respuestas permite concluir que el modelo SFT logra capturar la esencia pedagógica del currículo mientras que la variante con contexto es una validación de que la arquitectura híbrida diseñada es capaz de escalar según la disponibilidad tecnológica de cada centro educativo. Es por eso que la variante de referencia (SFT + Contexto) no debe verse como un sistema ajeno sino como la prueba de que el proceso de ajuste fino le confirió al modelo de la **capacidad de razonar y estructurar información externa** bajo los principios de la NEM. Es decir, el ajuste actuó como un mecanismo de discriminación de relevancia: mientras el modelo base opera bajo un esquema enciclopédico global, el modelo ajustado demuestra una alineación estricta con los contenidos educativos públicos en México. Al omitir datos irrelevantes se confirma que el modelo adquirió la capacidad de **jerarquizar contenidos**, la cual es la característica esencial del tutor inteligente desarrollado: un sistema que no solo recupera información sino que posee el criterio pedagógico para guiar al estudiante dentro del contexto nacional.

10.3. Resultados de la Prueba Piloto en Campo

Cómo resultados de la etapa de intervención pedagógica en la prueba piloto del prototipo se encontraron hallazgos derivados de la interacción de los alumnos con el EVA. Como se mencionó en la metodología de evaluación cualitativa, los resultados se dividen en dos partes: las observaciones producto de la intervención pedagógica, y el análisis cuantitativo de la percepción de los usuarios reflejado en las encuestas de satisfacción.

10.3.1. Análisis de la Intervención Pedagógica: Proyectos Transversales

Durante el desarrollo de las actividades del **Huerto Escolar** y la **Gestión de la Cooperativa** se realizó una observación participativa que permitió identificar hallazgos críticos sobre la viabilidad del sistema en un entorno real:

- **Brecha de Alfabetización Digital y Hardware:** Se identificó que al menos la

mitad de la población estudiantil, con la que se trabajó, presentaba dificultades en el manejo de periféricos computacionales básicos. Este hallazgo muestra la importancia de que el prototipo genere respuestas directas y concisas para disminuir la carga cognitiva del alumno al interactuar con el prototipo.

- **Función de Regularización Académica:** Gracias a la evaluación diagnóstica realizada a los alumnos se observó un rezago en el conocimiento de conceptos básicos que se asumía que dominaban. Los alumnos deberían tener en el nivel educativo correspondiente. Por otro lado gracias a la interacción de los alumnos con el *ChatBot* del prototipo se demostró su capacidad como herramienta pedagógica de regularización, al proveer explicaciones de conceptos desde cero con un lenguaje accesible, lo cual se evidenció en la evaluación post proyecto transversal de aquellos alumnos que pudieron concluir la actividad pese a las limitaciones descritas anteriormente.
- **Impacto de la Infraestructura de Red:** La concurrencia de alumnos en el primer día de pruebas piloto evidenció las limitaciones de conectividad de la escuela telesecundaria. Sin embargo, este hallazgo permitió validar que la decisión del enfoque de inyección de contexto dinámico fue acertada, ya que al reducir la concurrencia en las siguientes pruebas la latencia disminuía, permitiendo a los alumnos retomar las consultas al *ChatBot* del prototipo.

10.3.2. Análisis de resultados de la encuesta de satisfacción

Con el objetivo de obtener una visión integral de la experiencia de los usuarios, se analizaron las respuestas a la encuesta de satisfacción agrupadas en las tres categorías definidas en la metodología:

Diseño de la interfaz, estética y navegación

- **Resultados:** Se registró un 86 % de valoración positiva (totalmente de acuerdo o de acuerdo), con un 12 % de respuestas neutrales y un 2 % de valoraciones negativas.
- **Análisis:** El diseño de la interfaz fue bien percibido. Sin embargo, la percepción neutral y negativa se asocia directamente a la baja alfabetización digital observada; las dificultades para interactuar con los periféricos del ordenador influyeron en la percepción de facilidad de navegación del entorno virtual.

Eficacia pedagógica y calidad del contenido

- **Resultados:** Esta categoría obtuvo la valoración más alta, con un 91 % de opiniones favorables.

- **Análisis:** Los resultados validan que el Ajuste Fino Supervisado del LLM fue exitoso. Los alumnos percibieron a TutorIA como una herramienta que facilita el aprendizaje. El punto mejor valorado fue la claridad del contenido, confirmando que el modelo logra “hablar el mismo idioma” que los alumnos de secundaria.

Satisfacción global

- **Resultados:** Esta sección obtuvo un 84 % de aceptación, un 6 % de valoraciones neutras y un 10 % de valoraciones negativas.
- **Análisis:** El factor crítico fue la pregunta referente a la frustración experimentada. La retroalimentación cualitativa confirmó que dicha frustración no derivaba del desempeño de la IA, sino de la latencia del internet en el aula. A pesar de las limitaciones de infraestructura, la intención de uso futuro y recomendación se mantuvieron altas.

10.3.3. Resultados de las preguntas abiertas

Al final de la encuesta, se recopilaron percepciones directas de los alumnos sobre su experiencia:

- **Lo que les gustó:** Las herramientas de gamificación. Casi todas las respuestas abiertas mencionaban “**los juegos**” como la mejor parte de utilizar TutorIA.
- **Lo que no les gustó:** La velocidad del sistema; los alumnos reportaron que la aplicación “**se trababa**”, haciendo referencia a las deficiencias de infraestructura de red de la telesecundaria.

10.4. Conclusiones del capítulo de Resultados

Los resultados expuestos en el presente capítulo confirman la viabilidad técnica y pedagógica del Entorno Virtual de Aprendizaje desarrollado. El análisis cuantitativo mostró que el Ajuste Fino Supervisado es capaz de “**nacionalizar**” un modelo de lenguaje global, logrando que este se adapte al léxico de la Nueva Escuela Mexicana y ofreciendo contenidos adaptados a este contexto.

Por otro lado, el estudio de caso evidenció que la integración de un sistema híbrido con inyección de contexto es clave para la pertinencia didáctica. Mientras que los modelos base tienden a la verbosidad e inclusión de datos poco relevantes para el currículo, este enfoque logra equilibrar la concisión y profundidad educativa en las respuestas generadas.

Finalmente, la prueba piloto permitió validar la herramienta en condiciones reales de implementación. Se demostró que, a pesar de los retos de infraestructura y alfabetización digital, los alumnos perciben el valor de una inteligencia artificial especializada como herramienta educativa y que el prototipo tiene potencial para mitigar el rezago educativo en entornos vulnerables.

Capítulo 11

Conclusiones y Trabajo Futuro

11.1. Conclusiones y Cumplimiento de Objetivos

La investigación desarrollada y documentada en la presente tesis demostró que es posible reducir la brecha educativa en México mediante el uso de avances de vanguardia en Inteligencia Artificial Generativa, siempre que estos sean adecuados a las necesidades específicas del sistema educativo público. A través del desarrollo de **TutorIA**, se validó la hipótesis central: el Ajuste Fino Supervisado (SFT), enriquecido mediante una arquitectura de inyección dinámica de contexto, logra transformar un modelo de lenguaje genérico en un tutor especializado con una precisión léxica y pertinencia curricular que supera a los modelos de propósito general.

Se cumplieron satisfactoriamente los objetivos propuestos en el protocolo de investigación, obteniendo como resultados principales:

- Un modelo de lenguaje especializado en el contexto de la educación básica en México y alineado a la Nueva Escuela Mexicana (NEM).
- Una interfaz de usuario intuitiva que facilita la interacción pedagógica con el modelo.
- Un prototipo de Entorno Virtual de Aprendizaje (EVA) funcional en condiciones de implementación exigentes, como lo es una Telesecundaria con limitaciones tecnológicas.

El éxito del proyecto radica en haber demostrado que la calidad educativa mediada por IA no depende estrictamente del tamaño del modelo (parámetros), sino de la calidad, el filtrado y el contexto de los datos utilizados para su especialización.

11.2. Aportaciones y Trascendencia del Proyecto

El impacto de esta investigación trasciende los resultados cuantitativos, ofreciendo aportaciones en tres niveles fundamentales:

- **Trascendencia Tecnológica y Científica:** Se propuso un flujo de trabajo replicable para la especialización de Modelos de Lenguaje Grandes (LLM) en dominios específicos. Esta aproximación es valiosa para países en desarrollo, ya que explora la viabilidad de ejecutar sistemas inteligentes de alta precisión con recursos computacionales limitados.
- **Trascendencia Educativa:** TutorIA demostró la capacidad de jerarquizar información con base en el currículo nacional. Al equilibrar la concisión de las respuestas y la profundidad pedagógica, se logra reducir la carga cognitiva del estudiante y garantizar que el aprendizaje sea guiado por fuentes oficiales.
- **Trascendencia Económica y Social:** La optimización del modelo permite una gestión eficiente de los recursos tecnológicos demostrando que es posible ofrecer tutoría inteligente sin incurrir en los costos prohibitivos de los modelos comerciales de gran escala. Esto sienta un precedente para el desarrollo de soluciones de soberanía tecnológica educativa en México.

11.3. Limitaciones y Desafíos

La fase de implementación en campo permitió identificar retos estructurales que definen el alcance actual del prototipo:

- **Brecha Digital:** La infraestructura de conectividad y la falta de alfabetización digital básica (manejo de periféricos) se evidenciaron como los principales obstáculos. Esto resalta que la innovación en software debe ir acompañada de un fortalecimiento del hardware y de habilidades operativas básicas en el aula.
- **Alcance del Corpus:** El dominio del modelo en esta fase inicial se limita al nivel secundaria, dado que el corpus educativo procesado representa solo una fracción del sistema educativo nacional.

11.3.1. Reflexión sobre Mejora Continua y Perspectivas Futuras

Durante el desarrollo del proyecto de investigación se abrieron diversas líneas de desarrollo que permitirían elevar el impacto de lo realizado hasta el momento:

- **Multimodalidad e Inclusión:** La literatura sugiere que el futuro del aprendizaje estará guiado por modelos que interactúan en distintos formatos como imágenes, audio y video. La creación de sistemas multimodales permitiría que TutorIA asistiera de manera más eficiente a estudiantes con distintos estilos de aprendizaje o discapacidades. Específicamente, la integración de interfaces de voz mitigaría la brecha de alfabetización digital detectada en el uso del teclado, aminorando los problemas de inclusión.

- **Aprendizaje continuo y local (Edge Computing):** Para resolver de manera definitiva las limitaciones de conectividad observadas, se propone la migración hacia arquitecturas locales con servidores físicos en el aula. Mediante técnicas de cuantización avanzada, el modelo podría ejecutarse localmente, eliminando la dependencia de una conexión remota, asegurando la continuidad del servicio y permitiendo un esquema de aprendizaje adaptativo en tiempo real.
- **Refinamiento Pedagógico y Capacitación:** El trabajo futuro debe contemplar la integración de docentes en activo para el refinamiento del modelo mediante retroalimentación humana (RLHF), alineando las respuestas con la didáctica específica del aula. Asimismo, se requiere de un marco ético y capacitación docente para que la inteligencia artificial generativa sea una herramienta complementaria estratégica que potencie la autonomía de los estudiantes.

En conclusión, este proyecto de investigación sienta las bases y evidencia que la soberanía tecnológica educativa en México es viable. El prototipo desarrollado demostró que es posible implementar herramientas de IA en la educación que no sólo sean técnicamente mejores, sino también culturalmente pertinentes, lo que representa una base sólida para una nueva era de aprendizaje personalizado y accesible en México.

Referencias bibliográficas

- [1] P. M. L. Domínguez and A. M. López, “Educación en línea: una revisión de las limitaciones en México ante la crisis del COVID-19,” *Tlatemoani: Revista Académica de Investigación*, vol. 12, no. 36, pp. 58–72, 2021. [Online]. Available: <https://dialnet.unirioja.es/servlet/articulo?codigo=7958805>
- [2] A. N. C. Torres, “La educación a distancia como respuesta a las necesidades educativas del siglo XXI,” *Academia y Virtualidad*, vol. 10, no. 1, pp. 23–41, 2017. [Online]. Available: <https://revistas.unimilitar.edu.co/index.php/ravi/article/view/2241>
- [3] J. Wei, Y. Tay, R. Bommasani, C. Raffel, B. Zoph, S. Borgeaud, D. Yogatama, M. Bosma, D. Zhou, D. Metzler, E. H. Chi, T. Hashimoto, O. Vinyals, P. Liang, J. Dean, and W. Fedus, “Emergent Abilities of Large Language Models,” 2022. [Online]. Available: <https://arxiv.org/abs/2206.07682>
- [4] H. Zhao, H. Chen, F. Yang, N. Liu, H. Deng, H. Cai, S. Wang, D. Yin, and M. Du, “Explainability for Large Language Models: A Survey,” *ACM Transactions on Intelligent Systems and Technology*, vol. 15, no. 2, pp. 1–38, 2024.
- [5] L. Labadze, M. Grigolia, and L. Machaidze, “Role of AI chatbots in education: systematic literature review,” *International Journal of Educational Technology in Higher Education*, vol. 20, no. 1, pp. 1–17, 2023. [Online]. Available: <https://doi.org/10.1186/s41239-023-00426-1>
- [6] W. D. Heaven, “ChatGPT is going to change education, not destroy it,” 2023. [Online]. Available: <https://www.technologyreview.com/2023/04/06/1071059/chatgpt-change-not-destroy-education-openai/>
- [7] S. Llinares, “Educación Matemática y COVID-19 en las Américas: limitaciones, adaptaciones, y lecciones aprendidas,” *Cuadernos de Investigación y Formación en Educación Matemática*, no. 20, pp. 12–28, 2021. [Online]. Available: <https://revistas.ucr.ac.cr/index.php/cifem/article/view/48472>
- [8] M. D. A. Trapero, “Importancia de las TIC para la educación,” *Innovación y Experiencias Educativas*, vol. 15, febrero 2009. [Online].

Available: https://archivos.csif.es/archivos/andalucia/ensenanza/revistas/csicsif/revista/pdf/Numero_15/MARIA_DOLORES_ALCANTARA_1.pdf

- [9] J. A. Fernández, “Desigualdad e inequidad en la educación rural mexicana: la experiencia del CONAFE en el estado de Chihuahua,” *Revista Iberoamericana de Educación*, vol. 91, no. 1, pp. 115–133, 2023. [Online]. Available: <https://rieoei.org/RIE/article/view/5568>
- [10] D. Trucco, *Educación y desigualdad en América Latina*. Comisión Económica para América Latina y el Caribe (CEPAL), 2014. [Online]. Available: <https://repositorio.cepal.org/handle/11362/36835>
- [11] J. G. Sacristán, *El valor del tiempo en educación*. Madrid, España: Ediciones Morata, 2008. [Online]. Available: <https://edmorata.es/producto/el-valor-del-tiempo-en-educacion/>
- [12] M. J. R. Villach, R. M. M. Saíz, M. N. Llanos, and C. C. Salvador, “Evaluación continua y ayuda al aprendizaje. Análisis de una experiencia de innovación en educación superior con apoyo de las TIC,” *Electronic Journal of Research in Educational Psychology*, vol. 5, pp. 783–804, 2007. [Online]. Available: <http://www.redalyc.org/articulo.oa?id=293121946014>
- [13] P. R. M. Graells, “Impacto de las TIC en la educación: Funciones y limitaciones,” *3C TIC: Cuadernos de Desarrollo Aplicados a las TIC*, vol. 2, no. 1, pp. 1–15, 2013. [Online]. Available: <https://dialnet.unirioja.es/servlet/articulo?codigo=4817326>
- [14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention Is All You Need,” *CoRR*, vol. abs/1706.03762, 2017. [Online]. Available: <https://arxiv.org/abs/1706.03762>
- [15] S. Bhattamishra, A. Patel, P. Blunsom, and V. Kanade, “Understanding In-Context Learning in Transformers and LLMs by Learning to Learn Discrete Functions,” 2023. [Online]. Available: <https://arxiv.org/abs/2310.03016>
- [16] M. U. Hadi, Q. Al-Tashi, R. Qureshi, A. Shah, A. Muneer, M. Irfan, A. Zafar, M. B. Shaikh, N. Akhtar, J. Wu, S. Mirjalili, and M. A. Al-Garadi, “Large Language Models: A Comprehensive Survey of its Applications, Challenges, Limitations, and Future Prospects,” *TechRxiv preprint*, 2023. [Online]. Available: <https://doi.org/10.36227/techrxiv.23589741.v4>
- [17] N. Ding, Y. Qin, G. Yang, F. Wei, Z. Yang, Y. Su, S. Hu, Y. Chen, C. M. Chan, W. Chen, J. Yi, W. Zhao, X. Wang, Z. Liu, H. T. Zheng, J. Chen, Y. Liu, J. Tang, J. Li, and M. Sun, “Parameter-efficient fine-tuning of large-scale pre-trained language models,” *Nature Machine Intelligence*, vol. 5, no. 3, pp. 220–235, 2023.

- [18] H. Li, Q. Dong, Z. Tang, C. Wang, X. Zhang, H. Huang, S. Huang, X. Huang, Z. Huang, D. Zhang, Y. Gu, X. Cheng, X. Wang, S.-Q. Chen, L. Dong, W. Lu, Z. Sui, B. Wang, W. Lam, and F. Wei, “Synthetic Data (Almost) from Scratch: Generalized Instruction Tuning for Language Models,” 2024. [Online]. Available: <https://aka.ms/GeneralAI>
- [19] N. Straková and J. Válek, “Chatbots as a Learning Tool: Artificial Intelligence in Education,” *R&E-SOURCE Open Online Journal for Research and Education*, vol. 2024, no. 1, pp. 245–265, 2024. [Online]. Available: <https://doi.org/10.53349/resource.2024.is1.a1259>
- [20] C. W. Okonkwo and A. Ade-Ibijola, “Chatbots applications in education: A systematic review,” *Computers and Education: Artificial Intelligence*, vol. 2, p. 100033, 2021. [Online]. Available: <https://doi.org/10.1016/j.caeai.2021.100033>
- [21] M. L. B. M. Khidir and S. N. bin Sa’ari, “Chatbot as an Educational Support System,” *EPRA International Journal of Multidisciplinary Research (IJMR)*, vol. 8, no. 5, pp. 182–185, 2022. [Online]. Available: https://www.academia.edu/80637784/CHATBOT_AS_AN_EDUCATIONAL_SUPPORT_SYSTEM
- [22] K. Tan, T. Pang, C. Fan, and S. Yu, “Towards Applying Powerful Large AI Models in Classroom Teaching: Opportunities, Challenges and Prospects,” *arXiv preprint arXiv:2305.03433*, 2023. [Online]. Available: <https://arxiv.org/abs/2305.03433>
- [23] P. Denny, J. Prather, B. A. Becker, J. Finnie-Ansley, A. Hellas, J. Leinonen, A. Luxton-Reilly, B. N. Reeves, E. A. Santos, and S. Sarsa, “Computing Education in the Era of Generative AI,” *Communications of the ACM*, vol. 67, no. 2, pp. 56–67, 2024.
- [24] D. Zarris and S. Sozos, “Educational Artificial Intelligent Chatbot: Teacher Assistant & Study Buddy,” Master’s thesis, Luleå University of Technology, Luleå, Sweden, julio 2023. [Online]. Available: <https://www.diva-portal.org/smash/get/diva2:1802805/FULLTEXT01.pdf>
- [25] F. St-Hilaire, D. D. Vu, A. Frau, N. Burns, F. Faraji, J. Potochny, S. Robert, A. Roussel, S. Zheng, T. Glazier, J. V. Romano, R. Belfer, M. Shayan, A. Smofsky, T. Delarosbil, S. Ahn, S. Eden-Walker, K. Sony, A. O. Ching, S. Elkins, A. Stepanyan, A. Matajova, V. Chen, H. Sahraei, R. Larson, N. Markova, A. Barkett, L. Charlin, Y. Bengio, I. V. Serban, and E. Kochmar, “A New Era: Intelligent Tutoring Systems Will Transform Online Learning for Millions,” *arXiv preprint arXiv:2203.03724*, 2022. [Online]. Available: <https://arxiv.org/abs/2203.03724>
- [26] B. W. Lee, H. Cho, and K. M. Yoo, “Instruction Tuning with Human Curriculum,” 2024. [Online]. Available: <https://arxiv.org/abs/2310.09518>

- [27] Y. Hicke, A. Agarwal, Q. Ma, and P. Denny, “AI-TA: Towards an Intelligent Question-Answer Teaching Assistant using Open-Source LLMs,” *arXiv preprint arXiv:2311.02775*, 2023. [Online]. Available: <https://arxiv.org/abs/2311.02775>
- [28] G. Boateng, S. John, A. Glago, S. Boateng, and V. Kumbol, “Kwame for Science: An AI Teaching Assistant Based on Sentence-BERT for Science Education in West Africa,” *arXiv preprint arXiv:2206.13703*, 2022. [Online]. Available: <https://arxiv.org/abs/2206.13703>
- [29] A. Goel, “AI-Powered Learning: Making Education Accessible, Affordable, and Achievable,” *arXiv preprint arXiv:2006.01908*, 2020. [Online]. Available: <https://arxiv.org/abs/2006.01908>
- [30] R. Taylor, M. Kardas, G. Cucurull, T. Scialom, A. Hartshorn, E. Saravia, A. Poulton, V. Kerkez, and R. Stojnic, “Galactica: A Large Language Model for Science,” *arXiv preprint arXiv:2211.09085*, 2022. [Online]. Available: <https://arxiv.org/abs/2211.09085>
- [31] L. Benedetto, P. Cremonesi, and M. Parenti, “A Virtual Teaching Assistant for Personalized Learning,” *CEUR Workshop Proceedings*, vol. 2482, pp. 233–241, 2019. [Online]. Available: <https://arxiv.org/abs/1902.09289>
- [32] S. Wollny, J. Schneider, D. D. Mitri, J. Weidlich, M. Rittberger, and H. Drachsler, “Are We There Yet? - A Systematic Literature Review on Chatbots in Education,” *Frontiers in Artificial Intelligence*, vol. 4, p. 654924, 2021.
- [33] M. L. B. M. Khidir and S. N. bin Sa’ari, “Chatbot as an Educational Support System,” *EPRA International Journal of Multidisciplinary Research (IJMR)*, vol. 8, no. 5, 2022.
- [34] P. F. Oliveira and P. Matos, “Introducing a Chatbot to the Web Portal of a Higher Education Institution to Enhance Student Interaction,” *Engineering Proceedings*, vol. 56, p. 128, 2023.
- [35] M. A. Alafnan, S. Dishari, M. Jovic, and K. Lomidze, “ChatGPT as an Educational Tool: Opportunities, Challenges, and Recommendations for Communication, Business Writing, and Composition Courses,” *Journal of Artificial Intelligence and Technology*, vol. 3, no. 2, pp. 60–68, 2023.
- [36] A. T. Neumann, Y. Yin, S. Sowe, S. Decker, and M. Jarke, “An LLM-Driven Chatbot in Higher Education for Databases and Information Systems,” *IEEE Transactions on Education*, vol. 68, no. 1, pp. 103–116, 2025.
- [37] M. M. V. Wyk, “Is ChatGPT an Opportunity or a Threat? Preventive Strategies Employed by Academics Related to a GenAI-Based LLM at a Faculty of Education,” *Journal of Applied Learning and Teaching*, vol. 7, no. 1, 2024.

- [38] J. Salminen, S. G. Jung, J. Medina, K. Aldous, J. Azem, W. Akhtar, and B. J. Jansen, “Using Cipherbot: An Exploratory Analysis of Student Interaction with an LLM-Based Educational Chatbot,” in *Proceedings of the 11th ACM Conference on Learning @ Scale (L@S 2024)*, Atlanta, GA, USA, 2024, pp. 279–283.
- [39] M. R. Kabir and F. Lin, “An LLM-Powered Adaptive Practicing System,” in *CEUR Workshop Proceedings*, 2023. [Online]. Available: <http://ceur-ws.org>
- [40] J. Ochoa-Luna, R. Lazo, and M. Meyer, “LLM Agents: Advancing Truly Personalized Online Education,” in *Education and New Developments 2025*, agosto 2025.
- [41] J. Han, H. Yoo, J. Myung, M. Kim, H. Lim, Y. Kim, T. Y. Lee, H. Hong, J. Kim, S.-Y. Ahn, and A. Oh, “LLM-as-a-Tutor in EFL Writing Education: Evaluation of Student-LLM Interaction,” 2024, arXiv preprint. [Online]. Available: <https://arxiv.org/abs/2310.05191>
- [42] M. Yue, W. Lyu, J. Suh, Y. Zhang, and Z. Yao, “MathVC: An LLM-Simulated Multi-Character Virtual Classroom for Mathematics Education,” 2025, arXiv preprint. [Online]. Available: <https://arxiv.org/abs/2404.06711>
- [43] F. St-Hilaire, D. D. Vu, A. Frau, N. Burns, F. Faraji, J. Potochny, S. Robert, A. Roussel, S. Zheng, T. Glazier, J. V. Romano, R. Belfer, M. Shayan, A. Smofsky, T. Delarosbil, S. Ahn, S. Eden-Walker, K. Sony, A. O. Ching *et al.*, “A New Era: Intelligent Tutoring Systems Will Transform Online Learning for Millions,” 2022, arXiv preprint arXiv:2203.03724. [Online]. Available: <https://arxiv.org/abs/2203.03724>
- [44] T. B. Frøsig and M. Romero, “Teacher Agency in the Age of Generative AI: Towards a Framework of Hybrid Intelligence for Learning Design,” 2024, arXiv preprint. [Online]. Available: <https://doi.org/10.48550/arXiv.2407.06655>
- [45] X. Chen, J. Zhang, T. Zhou, and F. Zhang, “LLM-CDM: A Large Language Model-Enhanced Cognitive Diagnosis Intelligent Education,” *IEEE Access*, vol. 13, pp. 47 165–47 180, 2025.
- [46] J. L. Weber, “Understanding and Developing Educational Tools in the LLM Era,” in *ACM Conference on International Computing Education Research V.2 (ICER 2025 Vol. 2)*. ACM, 2025, pp. 1–2.
- [47] P. Sridhar, A. Doyle, A. Agarwal, C. Bogart, J. Savelka, and M. Sakr, “Harnessing LLMs in Curricular Design: Using GPT-4 to Support Authoring of Learning Objectives,” 2023, arXiv preprint arXiv:2306.17459. [Online]. Available: <https://arxiv.org/abs/2306.17459>
- [48] L. Huovinen, “Assessing Usability of Large Language Models in Education,” Bachelor’s thesis, Metropolia University of Applied Sciences, 2024. [Online]. Available: <https://www.theseus.fi/handle/10024/850498>

- [49] R. Sodhi and J. Choudhary, “Text Analyzing Tool for Simplifying the Syllabus Creation Process,” *International Journal of Information and Education Technology*, vol. 13, no. 2, pp. 409–416, 2023.
- [50] E. Dimitriadou and A. Lanitis, “A Critical Evaluation, Challenges, and Future Perspectives of Using Artificial Intelligence and Emerging Technologies in Smart Classrooms,” *Smart Learning Environments*, vol. 10, no. 1, 2023.
- [51] M. Bearman, J. Ryan, and R. Ajjawi, “Discourses of Artificial Intelligence in Higher Education: A Critical Literature Review,” *Higher Education*, vol. 86, no. 2, pp. 369–385, 2023.
- [52] I. Weissburg, S. Anand, S. Levy, and H. Jeong, “LLMs Are Biased Teachers: Evaluating LLM Bias in Personalized Education,” 2025, arXiv preprint arXiv:2410.14012. [Online]. Available: <https://arxiv.org/abs/2410.14012>
- [53] M. Bernabei, S. Colabianchi, A. Falegnami, and F. Costantino, “Students’ Use of Large Language Models in Engineering Education: A Case Study on Technology Acceptance, Perceptions, Efficacy, and Detection Chances,” *Computers & Education: Artificial Intelligence*, vol. 5, p. 100172, 2023.
- [54] M. R. King, A. M. Abdulrahman, M. I. Petrovic, P. L. Poley, S. P. Hall, S. Kulapantana, and Z. E. Lamantia, “Incorporation of ChatGPT and Other Large Language Models into a Graduate-Level Computational Bioengineering Course,” *Cellular and Molecular Bioengineering*, vol. 17, no. 1, pp. 1–6, 2024.
- [55] V. Grande, N. Kiesler, and M. A. R. Francisco, “Student Perspectives on Using a Large Language Model (LLM) for an Assignment on Professional Ethics,” in *Proceedings of the 2024 Innovation and Technology in Computer Science Education (ITiCSE 2024)*. ACM, 2024, pp. 478–484.
- [56] I. Hagström, “Learning by Prompting: Investigating the Potential of Large Language Models as a Learning Tool in Education,” 2023. [Online]. Available: <https://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-199951>
- [57] A. Bilal, D. Ebert, and B. Lin, “LLMs for Explainable AI: A Comprehensive Survey,” 2025, arXiv preprint. [Online]. Available: <https://arxiv.org/abs/2504.00125>
- [58] L. Mew and W. H. Money, “Leveraging ChatGPT for Higher Education Course Offerings,” *Information Systems Education Journal*, vol. 22, no. 5, pp. 72–80, 2024.
- [59] T. Nazaretsky, M. Ariely, M. Cukurova, and G. Alexandron, “Teachers’ Trust in AI-Powered Educational Technology and a Professional Development Program to Improve It,” *British Journal of Educational Technology*, vol. 53, no. 4, pp. 914–931, 2025.

- [60] D. Baidoo-Anu and L. O. Ansah, “Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning,” *Journal of AI*, vol. 7, pp. 52–62, 2023.
- [61] L. Benedetto and P. Cremonesi, “Rexy, a configurable application for building virtual teaching assistants,” *Lecture Notes in Computer Science*, vol. 11747, pp. 233–241, 2019.
- [62] S. Sok and K. Heng, “ChatGPT for education and research: A review of benefits and risks,” *Cambodian Journal of Educational Research*, vol. 3, no. 1, pp. 110–122, 2023. [Online]. Available: <https://www.cambodianjr.org/>

Apéndice A

Carta de Consentimiento Informado para Padres o Tutores

Universidad Autónoma de Querétaro

Facultad de Ingeniería

Proyecto de investigación: Educación Impulsada por IA: Desarrollo de un Entorno Virtual de Aprendizaje mediante un Modelo de Lenguaje Ajustado.

Institución colaboradora: Centro de Investigaciones en Óptica (CIO), sede León, Guanajuato.

Estimado(a) padre, madre o tutor(a):

Su hijo(a) ha sido invitado(a) a participar voluntariamente en una prueba piloto de un prototipo de asistente educativo tipo chatbot desarrollado con fines académicos y de investigación. El propósito de este estudio es evaluar la utilidad de un modelo de lenguaje ajustado como apoyo en el aprendizaje en educación secundaria.

- **Actividades:** Los estudiantes interactuarán con el prototipo de asistente educativo en computadora durante sesiones breves, donde resolverán preguntas o actividades escolares, y después responderán una breve entrevista sobre su experiencia.
- **Lugar y duración:** Las sesiones se llevarán a cabo en el centro de cómputo y/o aulas de la escuela secundaria participante, con una duración aproximada de 30 a 45 minutos, bajo la supervisión de personal docente y del investigador responsable.
- **Riesgos y beneficios:** No se prevén riesgos físicos ni psicológicos. Las actividades se limitan a la interacción educativa y se realizarán en un entorno controlado. Los participantes no recibirán compensación económica ni material. El beneficio esperado es contribuir al desarrollo de herramientas tecnológicas que mejoren los procesos de enseñanza-aprendizaje.

- **Confidencialidad:** Toda la información recabada será anónima y tratada de forma confidencial. No se solicitarán datos personales identificables. Los registros se conservarán de forma segura durante 12 meses, y posteriormente serán eliminados. Los resultados se utilizarán únicamente con fines académicos.
- **Voluntariedad:** La participación es completamente voluntaria. Usted o su hijo(a) pueden retirarse en cualquier momento sin consecuencias.

Si desea más información, puede comunicarse al correo: desparza24@alumnos.uaq.mx.

Consentimiento

Por favor, marque la opción que corresponda y firme:

- ☐ **SÍ, autorizo** la participación de mi hijo/hija en el proyecto y acepto las condiciones de confidencialidad de la información.
- ☐ **NO, no autorizo** la participación.

Nombre y firma del padre/madre/tutor: _____

Nombre del menor: _____

Fecha: _____

Atentamente,

Daniel Esparza Barragán, investigador responsable del proyecto.

Estudiante de Maestría en Ciencias en Inteligencia Artificial, UAQ.

Director de tesis: Dr. Sebastián Salazar Colores, Investigador Titular, CIO-León.

Apéndice B

Carta de Asentimiento Informado para Estudiantes

Universidad Autónoma de Querétaro

Facultad de Ingeniería

Institución colaboradora: Centro de Investigaciones en Óptica (CIO), sede León, Guanajuato.

Proyecto: Educación Impulsada por IA: Desarrollo de un Entorno Virtual de Aprendizaje mediante un Modelo de Lenguaje Ajustado.

Investigador responsable: Daniel Esparza Barragán.

Correo electrónico: desparza24@alumnos.uaq.mx.

Hola, mi nombre es Daniel Esparza Barragán y estoy realizando una investigación sobre cómo un asistente educativo tipo chatbot que desarrollé utilizando un modelo de inteligencia artificial puede ayudarte a aprender mejor en la escuela.

- **Qué harás:** Usarás el chatbot en una computadora para responder algunas preguntas o repasar temas escolares. Luego te haré unas breves preguntas sobre lo que te pareció la herramienta.
- **Dónde y cuánto tiempo:** Esto se hará en el centro de cómputo y/o aula de clase de tu escuela secundaria, y tendrá una duración de alrededor de 30 a 45 minutos.
- **¿Hay riesgos?** No hay ningún riesgo. Si en algún momento no quieres participar o alguna pregunta no te gusta, puedes decidir no contestar o retirarte sin problema.

Tu información será confidencial, no se registrará tu nombre ni ningún dato que permita identificarte de manera personal. La información obtenida será utilizada exclusivamente con fines educativos y científicos, y se tratará bajo estrictos criterios de anonimato y confidencialidad. Todos los registros serán resguardados de forma segura durante un periodo máximo de doce meses, tras el cual serán eliminados de manera definitiva.

Si estás de acuerdo en participar, por favor firma abajo.

Firma del estudiante: _____

Nombre: _____ Fecha: _____

Apéndice C

Carta de Consentimiento Informado y Confidencialidad para el Docente a Cargo

Universidad Autónoma de Querétaro

Facultad de Ingeniería

Proyecto: Educación Impulsada por IA: Desarrollo de un Entorno Virtual de Aprendizaje mediante un Modelo de Lenguaje Ajustado.

Institución colaboradora: Centro de Investigaciones en Óptica (CIO), sede León, Guanajuato.

Por medio de la presente, manifiesto haber sido informado sobre el estudio titulado “Educación Impulsada por IA: Desarrollo de un Entorno Virtual de Aprendizaje mediante un Modelo de Lenguaje Ajustado”, cuyo propósito es evaluar el impacto educativo de un asistente virtual tipo chatbot en estudiantes de secundaria.

He sido informado de que:

- La participación de los estudiantes es voluntaria y se realiza con el consentimiento de padres o tutores.
- Las sesiones se llevarán a cabo en el centro de cómputo y/o aulas de la escuela, bajo mi supervisión y la del investigador.
- No existen riesgos físicos ni psicológicos previstos.
- No se otorgará compensación económica ni material.
- Los datos serán tratados de forma anónima y confidencial, con un periodo de resguardo de 12 meses.

Autorizo la realización de las pruebas piloto con los estudiantes de mi grupo, una vez que la escuela otorgue su autorización formal.

Firma del docente: _____

Nombre completo: _____

Escuela: _____ Fecha: _____

Investigador responsable: Daniel Esparza Barragán.

Estudiante de Maestría en Ciencias en Inteligencia Artificial, UAQ.

Director de tesis: Dr. Sebastián Salazar Colores, CIO-León.

Apéndice D

Instrumento de Evaluación Piloto 1: Cosecha Sostenible

Sección I: Evaluación de Conocimientos

Instrucción: Responde las siguientes preguntas seleccionando la opción correcta. Al lado de cada pregunta dispones de dos casillas de control para registrar tus respuestas en la evaluación inicial (T0) y en la evaluación final (T1).

a.	Un agricultor nota que sus plantas de maíz crecen muy poco. La causa más probable, relacionada con la fotosíntesis, es que les falta:	T0	T1
-----------	--	-----------	-----------

- a) Nitrógeno en el suelo.
- b) Dióxido de carbono en el aire.
- c) Glucosa artificial.
- d) Agua salada.

b.	Para lograr un crecimiento óptimo, una planta necesita que el suelo le proporcione nutrientes (factor abiótico). ¿Cuál es la forma más rápida en que un árbol devuelve nutrientes al suelo?	T0	T1
-----------	--	-----------	-----------

- a) Por la evaporación del agua.
- b) Por la sombra que da.
- c) Cuando sus hojas secas caen y se descomponen.
- d) Al absorber el agua.

c.	Si un campesino siembra en una zona de México donde llueve muy poco (clima seco), ¿qué tipo de suelo le ayudará a retener la poca agua disponible para las raíces?	T0	T1
----	---	-----------	-----------

- a) Suelo arenoso.
- b) Suelo arcilloso.
- c) Suelo rocoso.
- d) Suelo con mucha sal.

d.	La erosión del suelo por el viento hace que el suelo sea menos fértil. Una estrategia geográfica simple para proteger el suelo en una zona abierta es:	T0	T1
----	---	-----------	-----------

- a) Regar el suelo una vez al mes.
- b) Usar fertilizantes químicos.
- c) Sembrar árboles o barreras vivas alrededor del campo.
- d) Dejar el suelo completamente seco.

e.	En un “Plan de Cultivo Sostenible”, la Conclusión debe servir principalmente para:	T0	T1
----	---	-----------	-----------

- a) Contar cómo te sentiste al hacer el plan.
- b) Poner la fecha y tu nombre.
- c) Resumir el porqué tu plan funcionará y la recomendación final.
- d) Describir el problema inicial.

f.	Un argumento de autoridad es aquel que usa la opinión de un experto o un dato científico. ¿Cuál de estas frases es un argumento de autoridad en tu informe?	T0	T1
----	--	-----------	-----------

- a) “Creo que el suelo arcilloso es más bonito.”
- b) “Según estudios, la materia orgánica reduce la erosión.”
- c) “Me gusta sembrar maíz.”
- d) “La solución es muy fácil.”

Sección II: Actividad de Producción con la App

- **RETO:** El Plan de Cultivo Sostenible. Tu escuela quiere iniciar una pequeña cosecha de hortalizas en el patio trasero, pero el suelo no es muy bueno y el clima es seco.
- **TAREA:** Diseña un “Plan de Cultivo Sostenible” de 4 pasos. El plan debe detallar: el tipo de suelo ideal, la protección contra el clima, el abono necesario y la estructura de tu informe.
- **INSTRUCCIONES CLAVE PARA USAR LA APP:** Usa la app para preguntar sobre los conceptos de Biología, Geografía y Español que necesitas para resolver el reto. NO le pidas a la aplicación que te escriba el plan completo.

Escribe tu plan a continuación:

Apéndice E

Instrumento de Evaluación Piloto 2: Presupuesto Escolar Responsable

Sección I: Evaluación de Conocimientos

Instrucción: Responde las siguientes preguntas seleccionando la opción correcta. Al lado de cada pregunta dispones de dos casillas de control para registrar tus respuestas en la evaluación inicial (T0) y en la evaluación final (T1).

a.

La cooperativa tiene \$5,000. ¿Cuál es el cálculo correcto para saber cuánto dinero corresponde al 30 % de ese total?

T0	T1
----	----

- a) $\$5,000 + 30$
- b) $\$5,000 / 30$
- c) $\$5,000 \times 0.30$
- d) $\$5,000 \times 30$

b.

Si la categoría de “Materiales” es del 15 % y la de “Alimentos” es del 40 %, ¿cuál es el porcentaje restante para las otras categorías?

T0	T1
----	----

- a) 65 %
- b) 45 %
- c) 55 %
- d) 50 %

c. **El principio de Bien Común en la cooperativa significa que al tomar decisiones de gasto, el principal objetivo debe ser:**

T0	T1
----	----

- a) Beneficiar a los 3 alumnos más pobres.
- b) Asegurar que todos los estudiantes se beneficien del gasto.
- c) Guardar todo el dinero para el futuro.
- d) Solo comprar cosas que le gusten a la mayoría.

d. **¿Cuál es la razón más importante para que la cooperativa destine parte del dinero a un Fondo de Ahorro/Emergencia?**

T0	T1
----	----

- a) Para presumir que tienen dinero.
- b) Para estar preparados ante un gasto inesperado o una reparación urgente.
- c) Para no gastar el dinero.
- d) Para usarlo en dulces.

e. **Para que tu propuesta de presupuesto sea persuasiva (convencer), es vital que uses:**

T0	T1
----	----

- a) Adjetivos muy bonitos.
- b) Solo tu opinión personal.
- c) Los montos exactos y porcentajes como datos.
- d) Muchas mayúsculas.

f. **Si usas subtítulos como “Cálculo Total”, “Justificación” y “Conclusión”, ¿qué función cumplen principalmente?**

T0	T1
----	----

- a) Hacen el texto más corto.
- b) Decoran la hoja.
- c) Ayudan al lector a encontrar la información clave fácilmente.
- d) No tienen función importante.

Sección II: Actividad de Producción con la App

- **RETO:** Propuesta de Presupuesto Escolar responsable. La directora les ha asignado \$4,000 MXN a la cooperativa de tu grado.
- **TAREA:** Diseña una Propuesta de Presupuesto que incluya 4 categorías de gasto. Debes asignar un porcentaje a cada una, calcular el monto exacto y escribir una justificación de por qué esa distribución es la más responsable para toda la escuela.
- **INSTRUCCIONES CLAVE PARA USAR LA APP:** Usa la app para preguntar sobre los conceptos de Matemáticas, Formación Cívica y Español que necesitas para resolver el reto. NO le pidas a la aplicación que te escriba el plan completo.

Escribe tu plan a continuación:

Apéndice F

Cuestionario de Experiencia de Usuario y Usabilidad (Post-Prueba)

Instrucción: Evalúa los siguientes aspectos de tu experiencia interactuando con la aplicación usando la escala del 1 al 5.

Preguntas Abiertas (Retroalimentación de Mejora Cualitativa)

13. ¿Qué fue lo MÁS ÚTIL o lo que MÁS te gustó de la aplicación? (Sé específico sobre la ayuda que recibiste).

14. ¿Qué FUE LO PEOR o lo que necesita ser CAMBIADO?

#	Dimensión / Reactivo	1	2	3	4	5
1	Entendí rápidamente cómo funcionan los botones.	☐	☐	☐	☐	☐
2	La app era clara e intuitiva de usar.	☐	☐	☐	☐	☐
3	Los colores, dibujos y diseño de la app me gustaron.	☐	☐	☐	☐	☐
4	Frustración: Cuando usaba la app, me enojé o me sentí confundido(a).	☐	☐	☐	☐	☐
5	Me fue fácil encontrar la información que necesitaba.	☐	☐	☐	☐	☐
6	Claridad: Las explicaciones de la app sobre el tema fueron muy claras.	☐	☐	☐	☐	☐
7	Adaptación: La app me ayudó a aprender de la forma en que yo entiendo mejor.	☐	☐	☐	☐	☐
8	Siento que la aplicación me ayudó a aplicar los conceptos a la actividad.	☐	☐	☐	☐	☐
9	Las respuestas de la app me ayudaron a corregir mis dudas.	☐	☐	☐	☐	☐
10	Me gustó la experiencia de usar esta aplicación para resolver problemas.	☐	☐	☐	☐	☐
11	Si fuera posible, me gustaría volver a usar esta aplicación en la escuela o en casa.	☐	☐	☐	☐	☐
12	¿Qué tan probable es que le recomiendes esta app a un amigo para aprender?	☐	☐	☐	☐	☐

Cuadro F.1: Escala de Usabilidad y Satisfacción del Entorno Virtual. Escala: 1 = Nada de Acuerdo; 2 = Poco de Acuerdo; 3 = Neutro; 4 = De Acuerdo; 5 = Totalmente de Acuerdo.