



UNIVERSIDAD METROPOLITANA DE CIENCIAS DE LA EDUCACIÓN  
FACULTAD DE CIENCIAS BÁSICAS  
DEPARTAMENTO DE FÍSICA

**CONSTRUCCIÓN DE CHATBOT BASADO EN UN GRAN MODELO DE  
LENGUAJE PARA PROMOVER EL APRENDIZAJE SIGNIFICATIVO DEL  
MODELO ONDULATORIO**

TESIS PARA OPTAR AL GRADO DE LICENCIADO EN EDUCACIÓN EN FÍSICA Y  
TÍTULO DE PROFESOR EN FÍSICA CON MENCIÓN EDUCACIÓN EN TECNOLOGÍA

AUTOR: JUAN PABLO PIZARRO DROLETT

PROFESOR GUÍA: NELSON ESTEBAN SEPÚLVEDA NAVARRO

SANTIAGO DE CHILE, DICIEMBRE 2025



UNIVERSIDAD METROPOLITANA DE CIENCIAS DE LA EDUCACIÓN  
FACULTAD DE CIENCIAS BÁSICAS  
DEPARTAMENTO DE FÍSICA

**CONSTRUCCIÓN DE CHATBOT BASADO EN UN GRAN MODELO DE  
LENGUAJE PARA PROMOVER EL APRENDIZAJE SIGNIFICATIVO DEL  
MODELO ONDULATORIO**

TESIS PARA OPTAR AL GRADO DE LICENCIADO EN EDUCACIÓN EN FÍSICA Y  
TÍTULO DE PROFESOR EN FÍSICA CON MENCIÓN EDUCACIÓN EN TECNOLOGÍA

AUTOR: JUAN PABLO PIZARRO DROLETT  
PROFESOR GUÍA: NELSON ESTEBAN SEPÚLVEDA NAVARRO

SANTIAGO DE CHILE, DICIEMBRE 2025

Autorizado para

**Sibumce Digital**

UNIVERSIDAD METROPOLITANA DE CIENCIAS DE LA EDUCACIÓN

FACULTAD DE CIENCIAS BÁSICAS

DEPARTAMENTO DE FÍSICA

Informe de Aprobación

TESINA Y EXAMEN DE TÍTULO

Se informa a la Dirección del Departamento de Física de la Facultad de Ciencias Básicas que la Tesina y Examen de Título presentados por el candidato

JUAN PABLO PIZARRO DROLETT

Ha sido aprobada por la comisión informante de Tesina y Examen de Título como requisito para optar al Grado de Licenciado en Educación en Física y Título de Profesor de Física, mención en Educación en Tecnología, en el Examen de Defensa de Tesina rendido el día ..... de ..... de 2025.

.....  
Profesora Guía Tesina

.....  
Profesor Informante Tesina

**Dedicatoria:**

A mi madre, quien dio su vida y esfuerzo constante por mi felicidad y futuro. Esto es por ti y  
ahora podré ser yo quien te sostenga y dé esperanza.

A mi padre, que espero estés orgulloso de mí en tu descanso eterno.

A Francisca, por tu fortaleza, amor y compañía.

A mi hija Emma, por quien vivo y amo la vida.

## **Agradecimientos:**

Agradezco profundamente a la profesora Isabel Quintrileo, por su empatía y escucha en uno de los momentos más críticos de mi formación docente. Gracias por creer en mí.

A la profesora Pamela Palomera, que a través de su trabajo se convirtió en un punto de inflexión en mi comprensión de la labor pedagógica. Es un ejemplo para mí como futuro profesor.

Al profesor Nelson Sepúlveda, por aceptar trabajar conmigo cuando nadie más lo hizo y por su constante amabilidad.

A los estudiantes del colegio Boston College Maipú que tuve el privilegio de conocer y que, con sus palabras y acciones, me devolvieron la esperanza y reafirmaron mi vocación.

## Tabla de contenido

|   |           |
|---|-----------|
| <b>Resumen</b>  | <b>1</b>  |
| <b>Capítulo I: Planteamiento del Problema</b>             | <b>2</b>  |
| <b>Capítulo II: Objetivos de la Investigación</b>         | <b>8</b>  |
| <b>Capítulo III: Marco Teórico</b>                        | <b>10</b> |
| <b>Capítulo IV: Marco Metodológico</b>                    | <b>15</b> |
| <b>Capítulo V: Presentación de Resultados y Discusión</b> | <b>26</b> |
| <b>Capítulo VI: Conclusiones y Proyecciones</b>           | <b>51</b> |
| <b>Referencias Bibliográficas.</b>                        | <b>61</b> |

## Resumen

El aprendizaje del modelo ondulatorio en educación formal presenta dificultades persistentes, asociadas a concepciones alternativas y a desafíos para sostener procesos de aprendizaje significativo con retroalimentación oportuna. En este contexto, esta tesis tuvo como propósito diseñar, desarrollar y validar, mediante juicio de expertos, un prototipo de chatbot educativo basado en modelos de lenguaje y un sistema de generación aumentada por recuperación (RAG), orientado a promover el aprendizaje significativo del modelo ondulatorio. El estudio se organizó bajo la metodología Design and Development Research (DDR), abarcando fases de análisis, diseño, desarrollo y evaluación del prototipo. El sistema fue desplegado como aplicación web e integró recuperación sobre documentos, trazabilidad de interacciones y soporte multimodal. La validación consideró la participación de tres expertos en física y educación, quienes interactuaron con el sistema en un escenario de estudio de caso y respondieron un instrumento con escala Likert (1–5) para usabilidad, utilidad y confianza, complementado con preguntas abiertas. Los resultados evidencian una evaluación global favorable, con fortalezas en la experiencia de uso, la claridad del acompañamiento y la percepción de confiabilidad. En paralelo, el análisis cualitativo identificó oportunidades de mejora vinculadas a consistencia conceptual puntual, continuidad conversacional y aspectos de presentación de respuestas. En conjunto, los hallazgos respaldan la viabilidad del prototipo y orientan mejoras focalizadas para iteraciones posteriores.

**Palabras claves:** Chatbot educativo; Modelos de lenguaje de gran tamaño; Recuperación aumentada por recuperación (RAG); Aprendizaje significativo; Modelo ondulatorio; Validación por juicio de expertos.

**Abstract:**

Learning about the wave model in formal education presents persistent difficulties, often associated with alternative conceptions and with challenges in sustaining meaningful learning processes supported by timely feedback. In this context, this thesis aimed to design, develop, and validate—through expert judgment—a prototype educational chatbot based on large language models with Retrieval-Augmented Generation (RAG), intended to promote meaningful learning of the wave model. The study was organized under the Design and Development Research (DDR) methodology, covering analysis, design, development, and a descriptive evaluation of the prototype. The system was deployed as a web application and incorporated document-based retrieval, interaction traceability, and multimodal support. Validation involved three experts in physics and education, who interacted with the system using a shared case-study scenario and then completed an instrument including a Likert-scale (1–5) assessment of usability, usefulness, and trust, complemented by open-ended questions. Results indicate an overall favorable evaluation, highlighting strengths in user experience, clarity of guidance, and perceived trustworthiness. In parallel, qualitative evidence identified opportunities for improvement related to occasional conceptual consistency, conversational continuity, and response presentation. Overall, the findings support the feasibility of the prototype and inform targeted refinements for subsequent iterations.

**Keywords:** Educational chatbot; Large language models; Retrieval-Augmented Generation (RAG); Meaningful learning; Wave model; Expert judgment validation.

## Capítulo I: Planteamiento del Problema

El aprendizaje del modelo ondulatorio y de fenómenos asociados a ondas en la educación formal continúa siendo un desafío persistente en diversos contextos nacionales y presentes en distintos niveles educativos (Rico et al., 2021; Planinić et al., 2024; Mešić et al., 2019). Estudios recientes en ondas mecánicas muestran que, incluso tras la enseñanza formal, el alumnado de secundaria alcanza niveles de comprensión limitados sobre la propagación, la superposición, la reflexión y la formación de ondas estacionarias, con porcentajes reducidos de estudiantes que logran explicaciones consistentes con el modelo científico y con una resistencia particular en la superposición y en la comprensión de cómo se propaga la perturbación en el medio (Kanyesigye et al., 2022; Kurniawan et al., 2023). En acústica, análisis de progresiones de aprendizaje sobre la propagación del sonido en el aire indican que los niveles explicativos más avanzados, como el caracterizar el sonido como una onda mecánica de presión que transporta energía sin transporte neto de materia, solo se alcanzan después de intervenciones específicas, sin embargo muchos estudiantes se mantienen en descripciones superficiales del proceso (Costa et al., 2024). La evidencia empírica reciente en el ámbito de la óptica ondulatoria muestra que el estudiantado presenta dificultades persistentes para interpretar fenómenos como la interferencia y la difracción de la luz, incluso después de haber recibido instrucción formal sobre estos contenidos. En particular, los resultados de evaluaciones conceptuales reportados por Balta et al., (2022) evidencian rendimientos bajos y razonamientos inconsistentes en tareas que requieren explicar la superposición de ondas luminosas, observándose el uso simultáneo de elementos propios de la óptica geométrica y de la óptica ondulatoria sin una integración conceptual coherente. De manera convergente, el estudio de Planinić et al., (2024) muestra que numerosos estudiantes mantienen modelos explicativos híbridos, en los que la propagación rectilínea de la luz coexiste con interpretaciones ondulatorias parciales, lo que limita la comprensión de los patrones de interferencia y difracción. En conjunto, estos trabajos indican que las dificultades observadas se asocian principalmente a una comprensión conceptual incompleta y a la persistencia de esquemas explicativos alternativos, más que a una falta de cobertura curricular de los contenidos de óptica ondulatoria.

Un componente crítico, como se mencionó anteriormente, es la persistencia de concepciones alternativas, documentada en distintos dominios de la física como en óptica, donde se han identificado concepciones estables e inestables sobre la propagación de la luz y la visibilidad de los objetos, como también en la comprensión y explicación de fenómenos sonoros (Chu & Treagust, 2014; Hrepic et al., 2010; Medina & Ramírez, 2016). En acústica, diversos estudios muestran que el alumnado tiende a imaginar el sonido como una entidad u “objeto” que viaja de forma independiente del medio, o bien a suponer que puede propagarse en el vacío o a través de huecos entre las partículas, en lugar de concebirlo como una perturbación del medio que se transmite mediante vibraciones y cambios de presión (Hrepic et al., 2010; Medina & Ramírez, 2016). Estas representaciones se traducen en explicaciones incompletas o erróneas sobre procesos como la reflexión y la refracción del sonido, así como sobre las magnitudes que cambian durante estos fenómenos, y dificultan la construcción de un modelo de onda mecánica capaz de dar cuenta de la transmisión de energía sin transporte neto de materia (Medina & Ramírez, 2016). En educación secundaria y en los primeros cursos universitarios, la investigación en didáctica de la física ha documentado de manera consistente la persistencia de concepciones alternativas en torno al modelo ondulatorio de la luz, particularmente en fenómenos de óptica ondulatoria como la difracción, la interferencia y la propagación luminosa. Estudios empíricos basados en instrumentos diagnósticos y análisis de respuestas estudiantiles muestran que una proporción significativa del alumnado aplica esquemas propios de la óptica geométrica en situaciones que requieren el uso del modelo ondulatorio, por ejemplo al explicar la difracción en rendijas estrechas o al predecir cualitativamente la forma de los patrones observados (Mešić et al., 2019; Matejak et al., 2022). Asimismo, se ha evidenciado que muchos estudiantes no conceptualizan la luz como una entidad física que existe y se propaga en el espacio independientemente de su fuente o de sus efectos visibles, manifestando ideas como que la luz sale hasta que golpea un objeto o que solo está presente en la fuente o en la pantalla de observación (Mešić et al., 2019; Chu & Treagust, 2014). En conjunto, estos resultados sugieren que las dificultades en óptica ondulatoria no se reducen a errores puntuales o a una cobertura insuficiente de contenidos, sino que responden a la coexistencia y persistencia de modelos explicativos alternativos, que interfieren con la apropiación del modelo ondulatorio y con la apropiación de representaciones clave para la comprensión profunda de los fenómenos luminosos.

La investigación en educación científica ha documentado de manera consistente que las metodologías de enseñanza predominantemente expositivas, centradas en la transmisión unidireccional de información, presentan limitaciones relevantes para promover aprendizajes profundos y significativos. Estudios empíricos de amplio alcance muestran que este tipo de enfoques suele asociarse a aprendizajes superficiales, con bajos niveles de comprensión conceptual integrada y limitada transferencia del conocimiento a nuevas situaciones, particularmente en contextos donde el estudiantado presenta diversidad de saberes previos y requiere apoyo para articularlos con nuevos contenidos (Wieman & Perkins, 2005; Freeman et al., 2014). En contraste, las metodologías activas, tales como la instrucción interactiva, el aprendizaje colaborativo estructurado y la indagación guiada, han demostrado generar mejoras sustantivas en el rendimiento académico y en la comprensión conceptual, al favorecer una participación cognitiva más intensa y una integración significativa entre conocimientos previos y nuevos aprendizajes (Freeman et al., 2014). Desde esta perspectiva, el valor pedagógico de las metodologías activas no radica únicamente en su carácter participativo, sino en su potencial para crear condiciones que favorezcan el aprendizaje significativo, entendido como un proceso de construcción de significados coherentes y funcionales, más allá de la memorización de información.

No obstante, la literatura también advierte que la implementación sostenida de metodologías activas orientadas al aprendizaje significativo enfrenta importantes obstáculos en los contextos educativos reales. Entre las dificultades más recurrentes se identifican la presión por la cobertura curricular, las restricciones de tiempo para el diseño e implementación de actividades complejas, la heterogeneidad de los saberes previos del estudiantado y la elevada demanda de retroalimentación formativa oportuna y de calidad que estas metodologías requieren para desplegar su potencial pedagógico (Wieman & Perkins, 2005). En el contexto chileno, investigaciones recientes han evidenciado que las condiciones institucionales, los marcos de desarrollo profesional docente y los sistemas de evaluación y rendición de cuentas pueden actuar tanto como facilitadores como barreras para la innovación pedagógica, afectando la posibilidad de sostener prácticas activas en el tiempo (Carrasco-Aguilar et al., 2023). Asimismo, estudios empíricos realizados en instituciones de educación superior chilenas han mostrado que, en ausencia de apoyos estructurales adecuados (organizacionales, tecnológicos o pedagógicos), se debilitan los procesos de acompañamiento y

retroalimentación, lo que limita el impacto efectivo de enfoques centrados en el estudiante y tensiona su contribución al logro de aprendizajes significativos (Villarroel et al., 2021).

La expansión reciente de los chatbots educativos basados en grandes modelos de lenguaje ha suscitado un interés creciente en el ámbito educativo debido a su capacidad para sostener interacciones conversacionales complejas y adaptativas. Estos sistemas, entrenados sobre grandes volúmenes de datos textuales, han comenzado a integrarse en contextos formativos como tutores conversacionales y asistentes pedagógicos con la promesa de apoyar el aprendizaje individualizado y de responder a limitaciones estructurales de la enseñanza tradicional, particularmente en escenarios caracterizados por alta carga docente y heterogeneidad de aprendizajes (Guizani et al., 2025; Li et al., 2024). No obstante, la literatura coincide en que este potencial no se traduce automáticamente en mejoras educativas, ya que el impacto de estas herramientas depende de manera crítica del diseño pedagógico que orienta su uso, del rol que se les asigna en la actividad de aprendizaje y de las condiciones institucionales, éticas y curriculares que regulan su integración en contextos escolares (Miao & Holmes, 2024).

La evidencia empírica disponible muestra resultados favorables del uso de chatbots basados en LLMs cuando estos se emplean con objetivos didácticos explícitos y bajo estrategias de acompañamiento estructurado. Estudios de revisión y meta-análisis reportan mejoras en el rendimiento académico y en la percepción de aprendizaje del estudiantado, así como efectos positivos moderados en el desarrollo de habilidades cognitivas de orden superior, especialmente cuando el chatbot actúa como un tutor dialógico que promueve la explicación, la justificación de razonamientos y la revisión progresiva de errores (Wang & Fan, 2025; Laun & Wolff, 2025). Sin embargo, estos resultados se concentran mayoritariamente en experiencias desarrolladas en educación superior, lo que plantea interrogantes respecto de su transferibilidad a la enseñanza media, donde las exigencias curriculares, las condiciones institucionales y las características del estudiantado difieren sustantivamente (Guizani et al., 2025).

De forma paralela, diversos estudios advierten que el uso no mediado de chatbots basados en LLMs puede generar efectos contraproducentes en los procesos de aprendizaje. La aceptación acrítica de respuestas generadas por estos sistemas, junto con la sobreconfianza en su

coherencia lingüística aparente, puede favorecer estrategias de estudio pasivas y disminuir la reflexión metacognitiva, incluso en estudiantes con formación disciplinar avanzada. Investigaciones en el ámbito de la educación en física muestran que este tipo de interacción puede conducir a prácticas de copia directa y a una reducción en la verificación de la información, reforzando errores conceptuales en lugar de promover una comprensión profunda (Krupp et al., 2023). Estos hallazgos ponen de relieve que los LLMs no son epistemológicamente neutrales y que su incorporación educativa requiere mediación pedagógica explícita y alfabetización crítica en inteligencia artificial (Porayska-Pomsta et al., 2024; Miao & Holmes, 2024).

Ante este escenario, la literatura reciente ha propuesto enfoques técnicos como la generación aumentada por recuperación como una vía para reducir algunos de los riesgos asociados al uso educativo de LLMs, al permitir que las respuestas del chatbot se basen en fuentes curriculares verificadas y alineadas con los objetivos de aprendizaje. Si bien este tipo de arquitectura contribuye a mejorar la trazabilidad y la coherencia curricular de las respuestas, no resuelve por sí sola las tensiones pedagógicas vinculadas a la evaluación de los aprendizajes, la mediación docente y la gobernanza institucional del uso de estas tecnologías (Li et al., 2024; Neumann et al., 2025). En el contexto chileno, estas problemáticas adquieren especial relevancia, dado que las orientaciones nacionales enfatizan que la inteligencia artificial generativa debe incorporarse como apoyo al aprendizaje activo y al pensamiento crítico, y no como sustituto del proceso educativo, lo que evidencia la necesidad de investigaciones situadas que examinen cómo diseñar e implementar chatbots educativos coherentes con el currículo escolar y con los marcos éticos y normativos vigentes definidos por el Ministerio de Educación de Chile y por el Ministerio de Ciencia, Tecnología, Conocimiento e Innovación (Centro de Innovación, Ministerio de Educación de Chile, 2023, 2025; Ministerio de Ciencia, Tecnología, Conocimiento e Innovación [MinCiencia], 2024).

### **Problema y necesidad de intervención**

En síntesis, confluyen tres hechos problemáticos: i) el aprendizaje del modelo ondulatorio presenta dificultades persistentes y concepciones resistentes aún después de la enseñanza formal; ii) las prácticas expositivas hegemónicas no ofrecen, por sí solas, oportunidades suficientes para modelizar, argumentar y transformar esas concepciones; iii) existe una

oportunidad de innovación con chatbots basados en LLM, pero su potencial depende de un diseño y una evaluación pedagógica rigurosos, alineados con el currículo y con lineamientos éticos.

### **Propósito de la investigación**

En este marco, se reconoce la posibilidad de desarrollar y validar una herramienta didáctica innovadora mediante la creación de un software basado en LLMs para abordar las dificultades de aprendizaje mencionadas. La presente investigación tiene como propósito diseñar un chatbot educativo basado en un modelo de lenguaje de gran tamaño, orientado a promover el aprendizaje significativo del modelo ondulatorio, conduciendo la interacción con un enfoque socrático de micro-diálogos que privilegia el desempeño cognitivo del usuario y su comportamiento activo (explicar, justificar, modelar) por sobre la entrega directa de respuestas. Su diseño integra metacognición, atención a concepciones alternativas y un tratamiento gradual de los problemas cuantitativos, articulando teoría y práctica, dentro de un marco contextual definido. La evaluación del prototipo se realizará a través de un juicio de expertos en física y educación, quienes interactuarán con el chatbot dentro de dicho marco y lo valorarán en términos de usabilidad, utilidad y confianza mediante una pauta breve en escala Likert de 1 a 5, complementada con dos preguntas abiertas para identificar fortalezas y sugerencias de mejora. Los resultados cuantitativos y cualitativos se integrarán con fines descriptivos para fundamentar orientaciones de diseño y condiciones de implementación futuras, considerando dimensiones didácticas, técnicas y éticas, manteniendo el software sin modificaciones durante el desarrollo de esta tesis. Con ello, se busca aportar evidencia situada en el contexto de validación experta del prototipo sobre criterios y resguardos necesarios para integrar críticamente chatbots basados en modelos de lenguaje en la enseñanza de la física.

## **Capítulo II: Objetivos de la Investigación**

### **2.1 Pregunta de Investigación**

En coherencia con el problema planteado y con el objetivo general de la investigación, se formula una pregunta de investigación general que orienta el desarrollo completo del estudio. Esta pregunta busca articular, en un mismo enunciado, el proceso de diseño del recurso y su evaluación mediante juicio de expertos. De este modo, la presente tesis se guía por la siguiente interrogante general:

¿Cuáles son las características basales que debe contener un chatbot basado en LLM creado para promover el aprendizaje significativo del modelo ondulatorio?

Dado que esta pregunta integra dimensiones de naturaleza distinta (diseño pedagógico y tecnológico por un lado y validación mediante juicio de expertos por otro) resulta necesario desglosarla en preguntas más acotadas que permitan organizar el estudio de manera coherente con los objetivos específicos y con el enfoque metodológico adoptado.

- a) ¿Qué criterios pedagógicos, disciplinares y tecnológicos deben considerarse para el diseño de un chatbot basado en modelos de lenguaje que favorezca el aprendizaje significativo del modelo ondulatorio en estudiantes?
- b) ¿Cómo valoran expertos en física y educación la usabilidad, la utilidad y la confianza del prototipo de chatbot al interactuar con él en el contexto del estudio de caso de la sala de ensayos?
- c) ¿Qué fortalezas, debilidades y sugerencias de mejora identifican los expertos respecto del chatbot, y qué orientaciones entregan para su futura implementación en contextos de educación formal?

### **2.2 Objetivo General**

Construir un chatbot educativo basado en un modelo de lenguaje de gran tamaño, orientado a promover el aprendizaje significativo del modelo ondulatorio en estudiantes de educación formal.

### **2.2.1 Objetivos específicos.**

1. Diseñar un prototipo de chatbot educativo, definiendo su fundamentación pedagógica, disciplinar y tecnológica a partir de la revisión de la literatura y del contexto del estudio de caso de la sala de ensayos.
2. Construir el chatbot educativo mediante la implementación de su arquitectura técnica, interfaz de interacción y comportamiento deseado, documentando su funcionamiento como recurso de apoyo al aprendizaje significativo del modelo ondulatorio.
3. Validar el prototipo de chatbot mediante juicio de tres expertos en física y educación, recogiendo sus valoraciones sobre usabilidad, utilidad y confianza, así como sus observaciones para la mejora y futura implementación del recurso en contextos de aula.

### **Capítulo III: Marco Teórico**

A fin de sustentar conceptualmente el problema de investigación y orientar el diseño del recurso propuesto, este marco teórico se organiza en subtemas que abordan, de manera articulada, tanto el fundamento tecnológico-pedagógico del uso de inteligencia artificial en educación como las particularidades didácticas del contenido disciplinar en estudio. En primer lugar, se examina la evolución conceptual y técnica de la inteligencia artificial y de los modelos de lenguaje, con énfasis en su incorporación reciente en contextos educativos mediante chatbots y sistemas conversacionales, atendiendo a la evidencia empírica, los aportes pedagógicos y las advertencias asociadas a su uso. En segundo término, se analizan las dificultades persistentes en el aprendizaje del modelo ondulatorio presentes en diferentes niveles de educación formal, considerando resultados provenientes de la investigación en didáctica de la física que documentan concepciones alternativas, modelos híbridos y limitaciones asociadas a prácticas de enseñanza tradicionales. En conjunto, estos subtemas permiten establecer un marco integrado que justifica la necesidad de diseñar y validar un chatbot educativo basado en modelos de lenguaje, orientado a apoyar el aprendizaje del modelo ondulatorio, explorando de forma práctica el potencial de los grandes modelos de lenguaje en educación.

#### **3.1 Inteligencia Artificial y Modelos de Lenguaje en Educación**

##### **Inteligencia Artificial**

La inteligencia artificial (IA) ha experimentado una evolución sustantiva tanto en su base tecnológica como en su definición conceptual desde mediados del siglo XX. Uno de los hitos fundacionales en la formalización del concepto se sitúa en la propuesta presentada en la Conferencia de Dartmouth, donde McCarthy et al. (1955) introducen de manera explícita el término *artificial intelligence* y delimitan el problema central del campo como el desarrollo de máquinas capaces de exhibir comportamientos que, en un contexto humano, serían calificados como inteligentes (McCarthy et al., 1955). Esta formulación se sustenta en la conjetura de que los procesos asociados al aprendizaje y a otras manifestaciones de la inteligencia pueden ser descritos con suficiente precisión como para permitir su simulación mediante sistemas computacionales. En conjunto, esta postura refleja una concepción racionalista dominante en la época, basada en la idea de que la inteligencia podía ser formalizada mediante reglas

explícitas y procedimientos computables, así como un marcado optimismo respecto de la posibilidad de reproducirla artificialmente.

De manera previa, Alan Turing había cuestionado la pertinencia de abordar directamente la pregunta sobre si las máquinas pueden pensar, señalando las dificultades conceptuales asociadas a definir términos como pensar o máquina a partir del uso cotidiano del lenguaje. En lugar de ello, propuso un criterio evaluativo alternativo, conocido como el juego de la imitación, en el cual la inteligencia de una máquina se evalúa a partir de su desempeño en una interacción comunicativa mediada, particularmente en el uso del lenguaje escrito (Turing, 1950). Este enfoque desplaza el análisis desde la naturaleza interna de los procesos mentales hacia la observación del comportamiento funcional del sistema en un contexto específico, sentando las bases para una comprensión operativa de la inteligencia artificial.

No obstante, este criterio ha sido objeto de revisiones posteriores, principalmente debido a su dependencia de la comparación directa con capacidades humanas y a sus limitaciones para abarcar formas de inteligencia que no se ajustan a un marco antropocéntrico. En este contexto, Pei Wang propone una definición alternativa de inteligencia, concebida como la capacidad de adaptación en condiciones caracterizadas por conocimiento y recursos insuficientes (Wang, 2019). Esta definición busca establecer un marco operativo general que no dependa de rasgos específicos de la cognición humana, sino de la capacidad de un sistema para desenvolverse eficazmente en entornos inciertos, dinámicos y con restricciones informacionales o computacionales. Desde esta perspectiva, la inteligencia se entiende como un fenómeno funcional y gradual, aplicable tanto a sistemas biológicos como artificiales.

Esta concepción resulta coherente con la definición contemporánea adoptada por la OCDE, la cual define los sistemas de IA como sistemas basados en máquinas que, orientados por objetivos explícitos o implícitos, infieren a partir de los datos que reciben cómo generar resultados tales como predicciones, contenidos, recomendaciones o decisiones capaces de influir en entornos físicos o virtuales (OECD, 2023). Asimismo, esta definición reconoce que los sistemas de IA presentan distintos niveles de autonomía y adaptabilidad tras su implementación, enfatizando el carácter inferencial, contextual y orientado a objetivos que caracteriza a las aproximaciones actuales en el campo de la inteligencia artificial.

## **Evolución de los Modelos de Lenguaje**

En la actualidad, existen diversas tecnologías basadas en IA diseñadas para abordar tareas específicas, entre las cuales destacan los Modelos de Lenguaje (ML). De acuerdo con la revisión amplia y actualizada realizada por Zhao et al. (2025), los ML se caracterizan por estimar la probabilidad generativa de secuencias de palabras, lo que les permite predecir la probabilidad de tokens futuros o faltantes dentro de un contexto lingüístico dado. Estos modelos constituyen un componente central para el desarrollo del Procesamiento del Lenguaje Natural (PLN), entendido como un subcampo de la informática y la inteligencia artificial que emplea técnicas de aprendizaje automático para posibilitar que los sistemas computacionales comprendan, interpreten y generen lenguaje humano (Stryker & Holdsworth, n.d.).

De acuerdo con la revisión de Zhao et al. (2025), la evolución de los ML puede organizarse en cuatro etapas principales, las cuales reflejan transformaciones progresivas en los mecanismos de representación del lenguaje, en la capacidad de modelar el contexto y en la escala del aprendizaje a partir de grandes volúmenes de datos textuales:

En una primera etapa se sitúan los Modelos de Lenguaje Estadísticos, los cuales estiman la probabilidad de secuencias de palabras a partir de su frecuencia de aparición en corpus extensos. Su implementación más representativa corresponde a los modelos n-grama, que operan bajo el supuesto de que la probabilidad de una palabra depende únicamente de un número fijo y reducido de palabras previas. Este enfoque permitió los primeros avances sistemáticos en tareas de procesamiento del lenguaje natural, al ofrecer un marco formal y computacionalmente tratable para la modelación del lenguaje. Sin embargo, presenta limitaciones estructurales relevantes, entre ellas el crecimiento exponencial del número de parámetros con el tamaño del vocabulario y la longitud del contexto, así como una capacidad limitada para generalizar y capturar dependencias de largo alcance. Estas restricciones evidenciaron la necesidad de representaciones más compactas y expresivas.

Una segunda etapa corresponde a los Modelos de Lenguaje Neuronales, los cuales introducen el uso de redes neuronales y representaciones vectoriales continuas del lenguaje (embeddings). Estas representaciones permiten codificar regularidades sintácticas y semánticas en espacios

de menor dimensión, reduciendo los problemas de escalabilidad asociados a los modelos puramente estadísticos. Para incorporar información contextual, estos modelos emplean arquitecturas capaces de procesar secuencias, lo que mejora la modelación de dependencias temporales entre palabras. No obstante, su desempeño se ve condicionado por dificultades relacionadas con la estabilidad del entrenamiento, el manejo eficiente de contextos extensos y una elevada demanda computacional, factores que limitan la profundidad de los modelos y su alcance práctico en escenarios de gran escala.

En una tercera etapa emergen los Modelos de Lenguaje Preentrenados, caracterizados por la separación entre el aprendizaje general del lenguaje y su adaptación a tareas específicas. Estos modelos se entrenan inicialmente mediante aprendizaje autosupervisado sobre grandes volúmenes de texto no etiquetado, adquiriendo representaciones lingüísticas generales que luego pueden ajustarse a distintos usos mediante procedimientos de fine-tuning o aprendizaje con pocos ejemplos. Un rasgo distintivo de esta etapa es la incorporación de representaciones contextualizadas profundas y el uso de arquitecturas basadas en mecanismos de atención, lo que permite capturar dependencias de largo alcance de manera más eficiente. Como resultado, estos modelos logran mejoras sustantivas en una amplia variedad de tareas, aunque aún requieren procesos de adaptación posteriores para un desempeño óptimo en contextos específicos.

La cuarta etapa corresponde a los Modelos de Lenguaje de Gran Escala (LLMs por sus siglas en inglés), los cuales amplían los enfoques anteriores mediante un incremento masivo en el número de parámetros, la cantidad de datos de entrenamiento y los recursos computacionales disponibles. Este escalamiento produce mejoras sistemáticas en el rendimiento y da lugar a la aparición de habilidades emergentes, tales como el aprendizaje en contexto y el seguimiento de instrucciones sin necesidad de reentrenamiento explícito. En este marco, la adaptación del modelo tiende a desplazarse desde el ajuste de parámetros hacia el diseño de instrucciones o prompts, lo que incrementa la flexibilidad de uso y reduce los costos asociados a la implementación y personalización.

En este contexto evolutivo, las características de los modelos de lenguaje de gran escala los posicionan como una base tecnológica especialmente adecuada para el desarrollo de sistemas conversacionales avanzados. La combinación de escala, capacidad de adaptación y generación de diálogo coherente permite superar las limitaciones de los chatbots basados en reglas o en modelos de menor capacidad. En particular, estos sistemas pueden reformular explicaciones, ofrecer ejemplos contextualizados y responder de manera flexible a las intervenciones del usuario, lo que los convierte en candidatos idóneos para aplicaciones de apoyo pedagógico orientadas a la mediación del aprendizaje.

### **Chatbots en educación**

La aparición y rápida difusión de los LLMs ha generado un punto de inflexión en la investigación y práctica educativa contemporánea. Como anteriormente se menciona, estos sistemas basados principalmente en arquitecturas transformer entrenadas sobre grandes volúmenes de datos textuales, poseen la capacidad de generar respuestas lingüísticamente coherentes, contextualizadas y adaptativas, lo que ha impulsado su integración en múltiples ámbitos educativos bajo la forma de chatbots, tutores inteligentes y asistentes pedagógicos conversacionales (Guizani et al., 2025; Li et al., 2024). A diferencia de tecnologías educativas previas, los LLMs permiten sostener interacciones dialógicas complejas, lo que los posiciona como herramientas con potencial para mediar procesos cognitivos de alto nivel, tales como la elaboración conceptual, la argumentación y la autorregulación del aprendizaje. No obstante, la literatura coincide en que su valor educativo no es intrínseco ni automático, sino que depende críticamente del diseño pedagógico, del rol asignado al sistema y de las condiciones institucionales, éticas y curriculares que regulan su uso (Miao & Holmes, 2024; OECD, 2025).

Uno de los aportes más robustamente documentados de los chatbots educativos basados en LLMs es su capacidad para escalar el acompañamiento individual en contextos caracterizados por alta carga docente y heterogeneidad de aprendizajes. La revisión sistemática de Guizani et al. (2025) muestra que los LLMs pueden asumir funciones de apoyo instruccional, resolución de dudas frecuentes, entrega de ejemplos, explicaciones iniciales y retroalimentación inmediata, permitiendo redistribuir el tiempo docente hacia tareas de mayor complejidad pedagógica. Desde una perspectiva cuantitativa, el meta-análisis de Wang y Fan (2025) reporta

un efecto positivo grande sobre el rendimiento académico del estudiantado, junto con efectos moderados en la percepción de aprendizaje y en el desarrollo de habilidades cognitivas de orden superior. Estos resultados, sin embargo, se observan principalmente en escenarios donde el uso del chatbot se encuentra guiado por objetivos didácticos explícitos y no como sustituto directo del trabajo cognitivo del estudiante.

El potencial distintivo de los LLMs en educación no reside en la generación automática de respuestas, sino en su capacidad para sostener andamiajes conversacionales que promuevan la reflexión, la justificación y la revisión de errores. La evidencia sintetizada por Guizani et al. (2025) y por el meta-análisis de Laun y Wolff (2025) converge en que los mayores efectos de los chatbots se observan cuando estos actúan como tutores dialógicos, orientando el razonamiento del estudiante mediante preguntas guiadas y retroalimentación formativa. Este tipo de interacción resulta particularmente coherente con enfoques de aprendizaje activo, como la indagación científica o el aprendizaje basado en proyectos, donde el diálogo, la metacognición y la construcción progresiva de explicaciones constituyen elementos centrales del aprendizaje profundo. En este marco, el chatbot no reemplaza la actividad del estudiante, sino que la apoya y estructura. Asimismo, la literatura señala que los chatbots educativos pueden favorecer el aprendizaje autorregulado al ofrecer un entorno de interacción constante, privada y no evaluativa, que reduce barreras para la búsqueda de ayuda académica y promueve la práctica autónoma (Neumann et al., 2025).

A pesar de la evidencia favorable, diversos estudios advierten que el uso no mediado de LLMs puede generar efectos contraproducentes. En el ámbito de la educación en física, Krupp et al. (2023) demostraron que estudiantes con formación avanzada tienden a sobreconfiar en respuestas incorrectas generadas por ChatGPT, adoptando estrategias de copia y pegado y disminuyendo los niveles de reflexión metacognitiva. Este fenómeno pone de relieve que la sofisticación lingüística de los LLMs puede ocultar errores conceptuales relevantes. Estos resultados refuerzan la idea de que los LLMs no son epistemológicamente neutrales y que su integración en educación debe contemplar explícitamente estrategias de alfabetización en IA, verificación de información y desarrollo del juicio crítico, tal como recomiendan organismos internacionales (Miao & Holmes, 2024).

Una de las estrategias más relevantes para abordar las limitaciones de los LLMs en contextos educativos es el enfoque de generación aumentada por recuperación (Retrieval-Augmented Generation, RAG). Este enfoque combina la capacidad generativa del modelo con mecanismos de recuperación de información desde fuentes externas verificadas, tales como documentos curriculares, materiales de clase o bases de conocimiento institucionales. La revisión sistemática de Li et al. (2024) muestra que los sistemas RAG reducen significativamente la incidencia de errores factuales, mejoran la trazabilidad de las respuestas y permiten una mayor alineación curricular. Un ejemplo concreto de esta arquitectura es MoodleBot, descrito por Neumann et al. (2025), el cual utiliza RAG para responder exclusivamente en función de los contenidos del curso, logrando altos niveles de precisión y aceptación estudiantil. Desde una perspectiva pedagógica, los sistemas RAG permiten diseñar chatbots educativos contextualizados, controlados y coherentes con los objetivos de aprendizaje, lo que resulta especialmente pertinente para su implementación en educación escolar y superior. Los documentos normativos recientes coinciden en que la integración de IA generativa en educación requiere marcos de gobernanza claros. Miao y Holmes (2024) enfatiza la necesidad de proteger la privacidad, garantizar la equidad, prevenir sesgos y promover usos pedagógicos responsables. En la misma línea, los resultados de TALIS 2024 muestran que, si bien una proporción creciente del profesorado utiliza herramientas de IA, persisten preocupaciones significativas en torno al plagio, la evaluación y la formación docente (OECD, 2025). En el contexto chileno, el Ministerio de Educación ha planteado que la IA generativa debe incorporarse como una herramienta de apoyo al aprendizaje activo y al pensamiento crítico, y no como un sustituto del proceso educativo humano, subrayando la importancia de orientaciones pedagógicas explícitas para su uso en aula (Ministerio de Educación de Chile, 2023; 2025).

En conjunto, la evidencia empírica, técnica y normativa permite sostener que los LLMs y chatbots educativos poseen un alto potencial pedagógico, particularmente para ampliar el acompañamiento individual, apoyar el aprendizaje autorregulado y sostener procesos de aprendizaje activo. Sin embargo, este potencial solo se materializa bajo condiciones de diseño pedagógico cuidadoso, integración curricular explícita y gobernanza institucional responsable. La discusión educativa contemporánea, por tanto, no se sitúa en la dicotomía de adoptar o

rechazar estas tecnologías, sino en cómo diseñarlas, regularlas y evaluarlas para que contribuyan efectivamente a la comprensión profunda, al desarrollo del pensamiento crítico y a la equidad educativa.

### **3.2 Dificultades en el aprendizaje del modelo ondulatorio en la educación media**

La investigación en didáctica de la física ha sostenido, con suficiente consistencia, que el aprendizaje del modelo ondulatorio no progresa de manera lineal ni automática con la instrucción formal. Más bien, tiende a configurarse como un proceso de reorganización conceptual que enfrenta obstáculos persistentes en distintos dominios y niveles educativos. En revisiones de literatura orientadas al sonido, se ha mostrado que las dificultades no se limitan a errores aislados, sino que se organizan alrededor de núcleos conceptuales que el estudiantado suele articular de manera parcial o con significados no científicos, particularmente al intentar relacionar propiedades del sonido con magnitudes ondulatorias y al interpretar la propagación como un proceso materializado o sustancializado. Esta regularidad se reporta como transversal a etapas educativas, lo que sugiere que la enseñanza frecuente no logra estabilizar un modelo explicativo que integre coherentemente la naturaleza del fenómeno, propagación y formalización ondulatoria (Rico et al., 2021).

En óptica ondulatoria, el patrón es convergente: aun cuando el currículo introduce interferencia y difracción, la apropiación del modelo ondulatorio suele competir con marcos alternativos, especialmente el razonamiento propio de la óptica geométrica, que permanece como recurso explicativo dominante cuando se requiere justificar cualitativamente patrones y condiciones de observación. La literatura empírica ha insistido en que este tipo de coexistencia no se explica solo por cobertura curricular o falta de ejercicios, sino por la estabilidad funcional de modelos previos en tareas escolares típicas y por la dificultad de construir criterios de aplicación del modelo ondulatorio en situaciones concretas (Mešić et al., 2019; Planinić et al., 2024).

En el ámbito de ondas mecánicas, varios estudios recientes describen desempeños limitados incluso después de experiencias de enseñanza, con especial resistencia en ideas estructurantes como la superposición y la propagación de la perturbación en el medio. En investigaciones centradas en ondas transversales y estacionarias, se ha utilizado análisis tipo Rasch para

estimar niveles de comprensión y localizar conceptos de mayor dificultad. En ese marco, se reporta que una fracción relevante del estudiantado presenta ausencia de comprensión o concepciones erróneas en tópicos clave, destacándose precisamente propagación y superposición como focos de dificultad y como barreras para interpretar formación de ondas estacionarias más allá de descripciones memorísticas (Kurniawan et al., 2023; Kanyesigye et al., 2022).

Este tipo de resultados es pedagógicamente relevante porque la superposición funciona como principio transversal del modelo ondulatorio: no solo permite explicar interferencia y batidos, sino también comprender que la perturbación no es un objeto que se desplaza, sino un patrón de cambio que se transmite a través de la dinámica del medio. Cuando la superposición se aprende como regla de cálculo descontextualizada o se confunde con colisión de entidades, se debilita la posibilidad de construir explicaciones sobre reflexión, transmisión y resonancia que conserven coherencia causal (Kanyesigye et al., 2022).

En la misma línea, investigaciones que reportan dificultades en ondas mecánicas suelen describir que el estudiantado separa la representación gráfica de la onda del comportamiento del medio. Es decir, interpreta la curva como trayectoria de una “cosa” que viaja, o bien atribuye movimiento neto de materia a lo largo de la dirección de propagación. Desde el punto de vista del modelo, esto es crítico porque impide distinguir con claridad entre oscilación local de las partículas del medio y propagación del estado de perturbación, distinción necesaria para explicar transferencia de energía sin transporte neto de materia (Rico et al., 2021).

Un aporte particularmente útil para el marco teórico proviene de trabajos que conceptualizan el aprendizaje como avance por niveles de sofisticación explicativa. En estudios sobre propagación del sonido en el aire, se ha validado una progresión de aprendizaje mediante pre y post test tras una intervención de aproximadamente un mes con estudiantes de secundaria. Los resultados describen que los niveles explicativos más avanzados, aquellos que caracterizan el sonido como onda mecánica de presión y enfatizan transporte de energía sin transporte neto de materia, no emergen como resultado espontáneo de la enseñanza habitual, sino que tienden a aparecer en mayor proporción tras intervenciones focalizadas, mientras que una parte del estudiantado permanece en explicaciones descriptivas o incompletas (Costa et al., 2024).

Desde una lectura pedagógica, este hallazgo permite sostener que la comprensión del sonido exige algo más que reconocer definiciones. Requiere coordinar, al menos, cuatro elementos: (1) un mecanismo de interacción local entre partículas del medio, (2) una descripción de la perturbación como patrón que se propaga, (3) la naturaleza longitudinal de la onda en aire y su relación con compresiones y rarefacciones, y (4) un criterio explícito para diferenciar energía transferida de materia transportada. Cuando cualquiera de estos elementos queda implícito, el estudiantado suele recurrir a explicaciones intuitivas de tipo sustancialista, lo que explica por qué el avance hacia niveles altos de la progresión no es mayoritario sin mediaciones didácticas específicas (Costa et al., 2024).

La evidencia empírica en óptica ondulatoria ofrece un respaldo especialmente robusto porque se apoya en instrumentos diagnósticos validados y en tareas de explicación asociadas a experimentos estándar. En primer lugar, se han desarrollado pruebas específicas para nivel secundario que cubren aspectos conceptuales de interferencia, difracción y polarización, permitiendo estimar con mayor precisión qué dimensiones resultan más frágiles. Por ejemplo, la construcción y validación de instrumentos tipo test en educación media reporta rendimientos globales limitados y patrones de respuesta que revelan razonamientos inconsistentes en ítems que exigen interpretar superposición de ondas luminosas y condiciones para formar patrones (Balta et al., 2022).

En segundo lugar, estudios en el contexto croata han producido un instrumento diagnóstico ampliamente citado para óptica ondulatoria, con un proceso de construcción que parte de bancos de ítems y culmina en una versión final que cubre interferencia, difracción y polarización. Este tipo de instrumento es relevante para el marco teórico porque muestra que las dificultades no se restringen a un concepto particular, sino que atraviesan el uso del modelo en diferentes fenómenos y representaciones. En términos educativos, el aporte central es que permite sostener con evidencia que el problema no es solo ejecutar una receta, sino aplicar un marco explicativo ondulatorio para justificar cualitativamente lo observado (Mešić et al., 2019). Complementariamente, investigaciones basadas en entrevistas y en secuencias de predicción, observación y explicación con montajes típicos han documentado cómo el estudiantado interpreta patrones desde recursos híbridos. Se observa que, frente a interferencia y difracción, muchos estudiantes alternan entre la idea de propagación rectilínea propia de la

óptica geométrica y fragmentos de razonamiento ondulatorio, sin establecer reglas claras sobre cuándo corresponde cada modelo. Este resultado es crucial porque explica por qué, aun recordando definiciones de difracción o interferencia, la explicación del patrón no se estabiliza: la base conceptual utilizada para decidir qué es relevante del experimento sigue siendo mixta y dependiente del contexto (Mešić et al., 2019; Planinić et al., 2024).

Desde un punto de vista didáctico, también es significativo que una propuesta de enseñanza por indagación guiada, diseñada para varias sesiones, reporte mejoras en comprensión cuando el diseño instruccional estructura explícitamente la exploración, la toma de decisiones y la justificación basada en observación. El valor del hallazgo no es únicamente que el rendimiento suba, sino que sugiere que el modelo ondulatorio se fortalece cuando la actividad obliga a coordinar evidencias experimentales con explicaciones, en lugar de limitarse a exposición y resolución rutinaria de problemas (Planinić et al., 2024).

Una explicación de fondo para la persistencia de dificultades es la estabilidad funcional de concepciones alternativas. En sonido, se ha descrito de manera consistente que el estudiantado utiliza modelos mentales donde el sonido funciona como entidad que viaja separada del medio, en contraste con el modelo de onda. Estudios con metodología fenomenográfica han identificado no solo un modelo alternativo dominante de tipo entidad, sino también modelos híbridos que combinan ingredientes del modelo ondulatorio con suposiciones sustancialistas, lo que ayuda a explicar por qué algunas respuestas parecen parcialmente correctas pero mantienen errores estructurales (Hrepic et al., 2010; Medina & Ramírez, 2016).

En óptica, se han investigado concepciones estables e inestables sobre propagación de la luz y visibilidad de objetos usando preguntas contextualizadas a lo largo de varios años escolares. Esta línea aporta un matiz clave para el marco teórico: no todas las concepciones alternativas tienen el mismo grado de estabilidad, y su activación depende del contexto del problema. En consecuencia, el progreso conceptual puede ser aparente si se evalúa solo con tareas familiares, pero volverse frágil cuando cambia la situación, el tipo de representación o el foco de la pregunta. En términos pedagógicos, esto respalda la necesidad de diseños didácticos que trabajen con variación sistemática de contextos y en los que se mencione explícitamente los criterios de aplicación del modelo (Chu & Treagust, 2014).

En óptica ondulatoria, la persistencia de modelos híbridos se expresa, por ejemplo, cuando el estudiantado intenta explicar difracción como si fuese solo desvío de un rayo, o interpreta interferencia como suma de intensidades sin justificar superposición de amplitudes y relación de fase. La literatura empírica basada en instrumentos diagnósticos y en análisis de explicaciones estudiantiles coincide en que estos híbridos no son simples confusiones terminológicas, sino estructuras explicativas que compiten con el modelo científico en tareas escolares (Mešić et al., 2019; Planinić et al., 2024).

En coherencia con estos hallazgos, la literatura sobre enseñanza de la física ha argumentado que enfoques predominantemente expositivos tienden a promover aprendizajes frágiles, especialmente cuando la meta es la reorganización conceptual y no solo la retención de definiciones. En contraste, la evidencia sobre metodologías activas en educación STEM reporta mejoras en rendimiento y comprensión conceptual, en parte porque fuerza al estudiantado a hacer explícitos sus razonamientos, contrastarlos con evidencia y recibir retroalimentación durante el proceso (Wieman & Perkins, 2005; Freeman et al., 2014).

Sin embargo, el marco teórico también requiere reconocer que el potencial de metodologías activas depende de condiciones de implementación. En Chile, se ha documentado cómo marcos de rendición de cuentas con altas consecuencias y configuraciones institucionales asociadas pueden tensionar trayectorias docentes y decisiones pedagógicas, afectando la sostenibilidad de innovaciones en el tiempo. Esto es relevante para interpretar por qué, aun con evidencia a favor de prácticas activas, la enseñanza cotidiana puede permanecer centrada en cobertura y preparación para evaluación externa, con menor espacio para secuencias de indagación o trabajo conceptual prolongado (Carrasco-Aguilar et al., 2023).

Asimismo, en contextos de educación superior se ha mostrado que programas de formación docente y cambios pedagógicos hacia evaluación auténtica y prácticas retroalimentadas requieren apoyos y condiciones sistemáticas para sostenerse, lo que refuerza la idea de que el enfoque centrado en el estudiante no se reduce a una técnica, sino que implica capacidades docentes, tiempo, coordinación curricular y criterios claros de evaluación formativa (Villarroel et al., 2021).

En síntesis, la evidencia revisada permite fundamentar que las dificultades de aprendizaje del modelo ondulatorio se explican por la convergencia de tres factores: la complejidad conceptual transversal del modelo, la persistencia de concepciones alternativas y modelos híbridos sensibles al contexto, y las restricciones pedagógicas e institucionales que limitan la implementación sostenida de experiencias que favorezcan reestructuración conceptual basada en evidencia (Rico et al., 2021; Mešić et al., 2019; Costa et al., 2024; Planinić et al., 2024; Carrasco-Aguilar et al., 2023).

## **Capítulo IV: Marco Metodológico**

### **4.1 Metodología**

Este estudio se orienta al diseño, construcción y validación de un prototipo de chatbot educativo basado en LLM, con soporte de recuperación aumentada (RAG), destinado a promover el aprendizaje significativo del modelo ondulatorio en contextos de educación formal. En coherencia con la pregunta de investigación, el trabajo se organiza bajo la metodología Design and Development Research (DDR) (Richey & Klein, 2007), estructurándose en cuatro fases articuladas: análisis, diseño, desarrollo y evaluación del prototipo.

En la fase de análisis, se fundamenta empírica y conceptualmente la necesidad del recurso y se acotan los requerimientos pedagógicos, disciplinares y tecnológicos del chatbot. En la fase de diseño, se define la fundamentación pedagógica y la arquitectura conceptual del recurso, estableciendo criterios para su comportamiento deseado y su rol como apoyo al aprendizaje activo del modelo ondulatorio. En la fase de desarrollo, se construye el prototipo mediante la implementación de su arquitectura técnica, su interfaz de interacción y los componentes que operacionalizan los criterios definidos, documentando su funcionamiento.

Finalmente, la fase de evaluación se centra en la validación mediante juicio de expertos, tal como lo establece el tercer objetivo específico. Esta validación se realizará con tres expertos en física y educación, quienes interactuarán con el prototipo y emitirán valoraciones sobre usabilidad, utilidad y confianza, junto con observaciones cualitativas orientadas a identificar fortalezas, debilidades y oportunidades de mejora del recurso.

## **4.2 Enfoque de la investigación**

La investigación adopta un enfoque mixto orientado a integrar evidencia de distinta naturaleza para responder las preguntas específicas del estudio. En particular, la validación por expertos contempla: (a) evidencia cuantitativa derivada de valoraciones estructuradas sobre usabilidad, utilidad y confianza, y (b) evidencia cualitativa proveniente de observaciones, comentarios fundamentados y/o una matriz de análisis (por ejemplo, FODA), centradas en el desempeño del prototipo y en orientaciones para su mejora e implementación futura.

La integración de ambas fuentes se realizará a nivel de análisis e interpretación, buscando producir conclusiones convergentes que permitan caracterizar el prototipo respecto de las dimensiones evaluadas y, a la vez, levantar recomendaciones justificadas para su perfeccionamiento en ciclos posteriores. En coherencia con el objetivo del estudio, la evaluación se implementa como un juicio experto en un único momento de aplicación (corte transversal), en el marco del estudio de caso de la sala de ensayos, sin intervención pedagógica en aula ni manipulación de variables.

## **4.3 Nivel o alcance de la investigación**

El alcance de este trabajo es descriptivo y analítico, centrado en la validación del prototipo mediante juicio de expertos, con el propósito de medir y caracterizar el desempeño del chatbot en las dimensiones de usabilidad, utilidad y confianza, y de identificar fortalezas, debilidades y sugerencias de mejora relevantes para su implementación futura.

El diseño es no experimental, de corte transversal, con muestreo intencional basado en criterios de expertos. El nivel de inferencia es analítico: no se busca generalización estadística, sino producir conclusiones justificadas y transferibles como orientaciones para el diseño y evaluación de recursos similares. Para ello, se desarrollará una descripción densa del contexto del estudio, del perfil de los expertos y del protocolo de aplicación, fortaleciendo la trazabilidad metodológica del proceso.

#### 4.4 Diseño de la investigación: Design and Development Research (DDR)

Una parte significativa del esfuerzo realizado tiene que ver con el diseño y desarrollo mismo del chatbot, como solución a un panorama previamente investigado y latente de discusión en el contexto educativo actual, considerando la perfectibilidad de cualquier herramienta digital, material o sistema que se pueda elaborar, implicando procesos de mejora continua.

Richey y Klein (2007) describen, específicamente para investigaciones centradas en el diseño y desarrollo de productos y soluciones necesarias para contextos reales, la estrategia metodológica DDR, útil para enmarcar el trabajo y las pretensiones del proyecto, desde la etapa de análisis previo, hasta la evaluación del producto o prototipo. Este estudio es proyectado hacia una implementación final en aula como una herramienta útil para el contexto escolar, para lo cual debe ser evaluada y desarrollada continuamente, con tal de alcanzar la mayor calidad ante la tremenda relevancia de los procesos formativos en la escuela. Si bien este diseño de investigación contempla una última etapa de implementación, ésta será relegada posterior a una segunda instancia de prototipado a evaluar, para aspirar finalmente a la implementación eficaz en aula. En cuanto a las acciones pertinentes a esta tesis de pregrado, se adaptarán las etapas propuestas por Richey y Klein para investigaciones de desarrollo de productos, en 4 fases:

- a) **Fase de análisis:** Se investiga y delimita la problemática a abordar, indagando las distintas necesidades y posibles soluciones, mediante una intervención diagnóstica o una revisión de la literatura actualizada. Basado en lo anterior, se establecen los objetivos priorizados. Se puede acompañar de un listado de principios, criterios u orientaciones para el diseño e implementación del producto o solución a crear.
- b) **Fase de diseño:** En base al análisis anterior se diseña la estructura de la solución para responder a los objetivos propuestos. Se especifican las características del producto y cómo es que aportan a la solución de los problemas, describiendo componentes, arquitecturas, flujos de acciones y criterios a verificar.
- c) **Fase de desarrollo:** Corresponde al proceso de construcción de la solución o producto. Usando el registro de la evolución y versiones de la propuesta, se describe posteriormente cómo fluyó el proceso de creación hasta la versión o prototipo a evaluar en la siguiente fase. También se desarrollan los instrumentos de evaluación o plan de pruebas para validar la solución propuesta.

- d) Fase de evaluación:** Se somete la solución o producto a la aplicación de instrumentos diseñados para su verificación o validación, con el fin de recopilar evidencia empírica sobre su desempeño y pertinencia. La evaluación se realiza mediante interacción con la solución en condiciones definidas para su uso, obteniendo datos cuantitativos y cualitativos que se analizan para sustentar la discusión de resultados e identificar acciones de ajuste o recomendaciones orientadas a la mejora continua del producto.

## **4.5 Procedimiento**

A continuación se describe el detalle del trabajo realizado en cada fase para cumplir con los objetivos propuestos

### **4.5.1 Análisis**

La fase de análisis tuvo por propósito fundamentar empírica y conceptualmente la necesidad del prototipo de chatbot educativo y delimitar con precisión el problema educativo, el contexto disciplinar y curricular, así como los condicionantes pedagógicos, técnicos y éticos que orientan las fases posteriores de diseño, desarrollo y evaluación. Esta fase se concibió como un proceso de carácter diagnóstico y de sistematización, orientado a construir insumos analíticos que permitieran justificar la intervención propuesta y acotar su alcance metodológico, en coherencia con la metodología Design and Development Research adoptada en este estudio.

En primer lugar, se abordó la identificación del problema educativo asociado al aprendizaje del modelo ondulatorio en la educación formal. A partir de la experiencia docente y del análisis de prácticas habituales de enseñanza, se reconoció que el estudiantado presenta dificultades persistentes para construir explicaciones coherentes sobre fenómenos ondulatorios, particularmente en relación con la propagación, la superposición y la interpretación de magnitudes como velocidad, frecuencia, período y amplitud. Estas dificultades no se manifiestan como errores aislados, sino como patrones explicativos relativamente estables que obstaculizan la apropiación del modelo ondulatorio como marco conceptual integrador.

En este contexto, se procedió a caracterizar concepciones alternativas relevantes documentadas en la literatura especializada. Entre ellas, se identificó como núcleo problemático la tendencia a concebir la velocidad de propagación de una onda como dependiente de parámetros propios de la onda, como la frecuencia, el período o la amplitud, en lugar de comprenderla como una propiedad del medio en condiciones dadas. Asimismo, se reconocieron modelos híbridos en los que se confunden propiedades de la perturbación con características del movimiento de las partículas del medio, lo que conduce a explicaciones inconsistentes sobre la transmisión de energía. Estas concepciones fueron entendidas como obstáculos estructurales para la reorganización conceptual y no como simples errores procedimentales.

Con el fin de contrastar y afinar esta delimitación del problema, se realizó una revisión de investigaciones en didáctica de la física que abordan el aprendizaje del modelo ondulatorio en distintos dominios, incluyendo ondas mecánicas, acústica y óptica ondulatoria. Este análisis permitió constatar que las dificultades descritas se reportan de manera consistente en diversos contextos educativos y niveles de formación, incluso después de procesos de enseñanza formal. Se revisaron los métodos diagnósticos utilizados por los estudios, así como los resultados empíricos asociados a la persistencia de concepciones alternativas y a la coexistencia de modelos explicativos, lo que reforzó la necesidad de intervenciones pedagógicas orientadas explícitamente a la mediación conceptual.

De forma articulada con lo anterior, se analizó la literatura que problematiza las limitaciones de enfoques de enseñanza predominantemente expositivos para promover aprendizajes profundos y significativos en ciencias. La evidencia revisada muestra que este tipo de prácticas tiende a favorecer aprendizajes superficiales y una integración débil entre conocimientos previos y nuevos contenidos, especialmente cuando se requiere reorganización conceptual. En contraste, se identificaron enfoques de aprendizaje activo, como el diálogo socrático, el andamiaje progresivo, la retroalimentación formativa y la metacognición, que presentan potencial para apoyar procesos de construcción de explicaciones y revisión de errores cuando se implementan bajo condiciones pedagógicas adecuadas.

En este marco, se desarrolló una revisión de estudios empíricos y revisiones recientes sobre el uso de chatbots y grandes modelos de lenguaje en educación STEM. El análisis permitió identificar oportunidades pedagógicas asociadas a estos sistemas, particularmente su capacidad para sostener interacciones conversacionales y ofrecer apoyo individualizado en contextos de alta heterogeneidad. No obstante, la literatura también advierte riesgos relevantes, tales como la aceptación acrítica de respuestas generadas, la reducción de la reflexión metacognitiva y la reproducción de errores conceptuales, lo que evidencia que el valor educativo de estas tecnologías depende críticamente de su diseño pedagógico y de las condiciones de uso.

Como parte de este análisis diagnóstico, se revisaron propuestas técnicas reportadas en la literatura, como la generación aumentada por recuperación, entendida como una estrategia para mejorar la coherencia curricular y la trazabilidad de las respuestas generadas por sistemas basados en modelos de lenguaje. Estas aproximaciones fueron consideradas como antecedentes relevantes para comprender el estado del arte y las posibilidades técnicas existentes, sin constituir decisiones de implementación en esta fase del estudio.

Paralelamente, se analizó el marco curricular chileno asociado al estudio del modelo ondulatorio en educación media con el objetivo de asegurar la coherencia curricular del prototipo propuesto. Este análisis se centró en identificar los objetivos de aprendizaje y las habilidades científicas vinculadas a la comprensión de fenómenos ondulatorios, entendiendo el currículo como un marco de referencia que orienta el diseño pedagógico del recurso sin circunscribir el análisis a un curso o nivel específico.

De manera complementaria, se revisaron documentos de políticas educativas y marcos normativos nacionales e internacionales relativos al uso de tecnologías digitales e inteligencia artificial en educación. Este análisis permitió identificar principios de uso responsable, resguardos éticos y consideraciones de gobernanza que condicionan la integración de herramientas basadas en inteligencia artificial en contextos educativos formales. A partir de estas orientaciones, se delinearon criterios éticos preliminares que acotan el alcance del

prototipo y que serán desarrollados con mayor detalle en la sección correspondiente a consideraciones éticas.

Con base en la síntesis del problema educativo, la evidencia empírica revisada y los marcos curriculares y normativos analizados, se procedió a identificar las necesidades pedagógicas que el sistema debería ser capaz de abordar. Estas se formularon en términos generales como requerimientos de mediación y apoyo al aprendizaje, tales como el fomento de la reflexión, la argumentación y la revisión progresiva de explicaciones, sin definir aún estrategias concretas de implementación.

Asimismo, se identificaron requerimientos funcionales generales que condicionan la viabilidad del prototipo como objeto de estudio, tales como la necesidad de interacción conversacional coherente con el currículo, la trazabilidad de las respuestas y la accesibilidad en contextos educativos diversos. Estos requerimientos se establecieron como criterios orientadores para las fases posteriores, sin describir arquitecturas técnicas ni decisiones de desarrollo.

En esta fase también se reconocieron explícitamente las principales limitaciones y condicionantes del estudio. Entre ellas se consideraron restricciones metodológicas, técnicas y éticas, así como los riesgos asociados al uso no mediado de sistemas basados en modelos de lenguaje y las limitaciones institucionales para una implementación directa en aula. Estas condiciones fundamentan la decisión de acotar el alcance de la investigación a una validación experta del prototipo y no a una intervención con estudiantes.

Finalmente, toda la información recabada durante la fase de análisis fue organizada y sistematizada mediante esquemas y matrices de síntesis que articulan el problema educativo, la evidencia empírica, el marco curricular, los criterios éticos y los requerimientos generales del sistema. Esta sistematización permitió verificar la coherencia interna entre la pregunta de investigación, los objetivos del estudio y el enfoque metodológico adoptado, estableciendo una base sólida para el desarrollo de las fases posteriores de diseño, desarrollo y evaluación del prototipo.

#### 4.5.2 Diseño

La fase de diseño tuvo por propósito transformar los requerimientos y condicionantes identificados en el análisis en una estructura operativa del prototipo, especificando con claridad qué características debía presentar el chatbot y por qué dichas características resultan pertinentes para promover aprendizaje activo del modelo ondulatorio en educación formal. En coherencia con la DDR, esta fase se centró en definir componentes, arquitecturas, flujos de acciones y criterios a verificar, sin confundirlos todavía con decisiones de implementación o detalles de programación.

En primer lugar, el diseño estableció el rol pedagógico del sistema como un apoyo conversacional que no sustituye el trabajo intelectual del usuario, sino que lo media mediante interacción guiada. Esto implicó traducir los requerimientos generales de mediación identificados en la fase de análisis (fomento de reflexión, argumentación y revisión progresiva de explicaciones) en una lógica conversacional concreta. En este marco, el comportamiento deseado se formalizó como una dinámica de micro-diálogos, donde cada consulta inicia con reformulación, diagnóstico breve de conocimientos previos y una tarea acotada para que el usuario intente un primer paso antes de recibir explicaciones extensas.

Un énfasis específico del diseño fue asegurar coherencia con el propósito de identificación conceptual de fenómenos cuando el usuario se enfrenta a tareas de clasificación o reconocimiento (por ejemplo, identificar el fenómeno responsable de máximos y mínimos de intensidad sonora). Para estos casos, se definió como criterio de comportamiento que el chatbot describa primero el comportamiento ondulatorio esperado y promueva hipótesis del usuario antes de entregar etiquetas como interferencia, resonancia u ondas estacionarias, salvo que el usuario ya haya propuesto una alternativa o declare bloqueo. Esto se complementó con el uso sistemático de preguntas metacognitivas distribuidas en el diálogo, con el fin de favorecer autorregulación, monitoreo de coherencia y transferencia del aprendizaje a contextos cercanos.

En segundo lugar, el diseño definió el marco de contexto de interacción para la validación, articulando el prototipo con un estudio de caso principal: la sala de ensayos donde se observan picos y valles de intensidad sonora, diferencias marcadas de volumen por ubicación y la

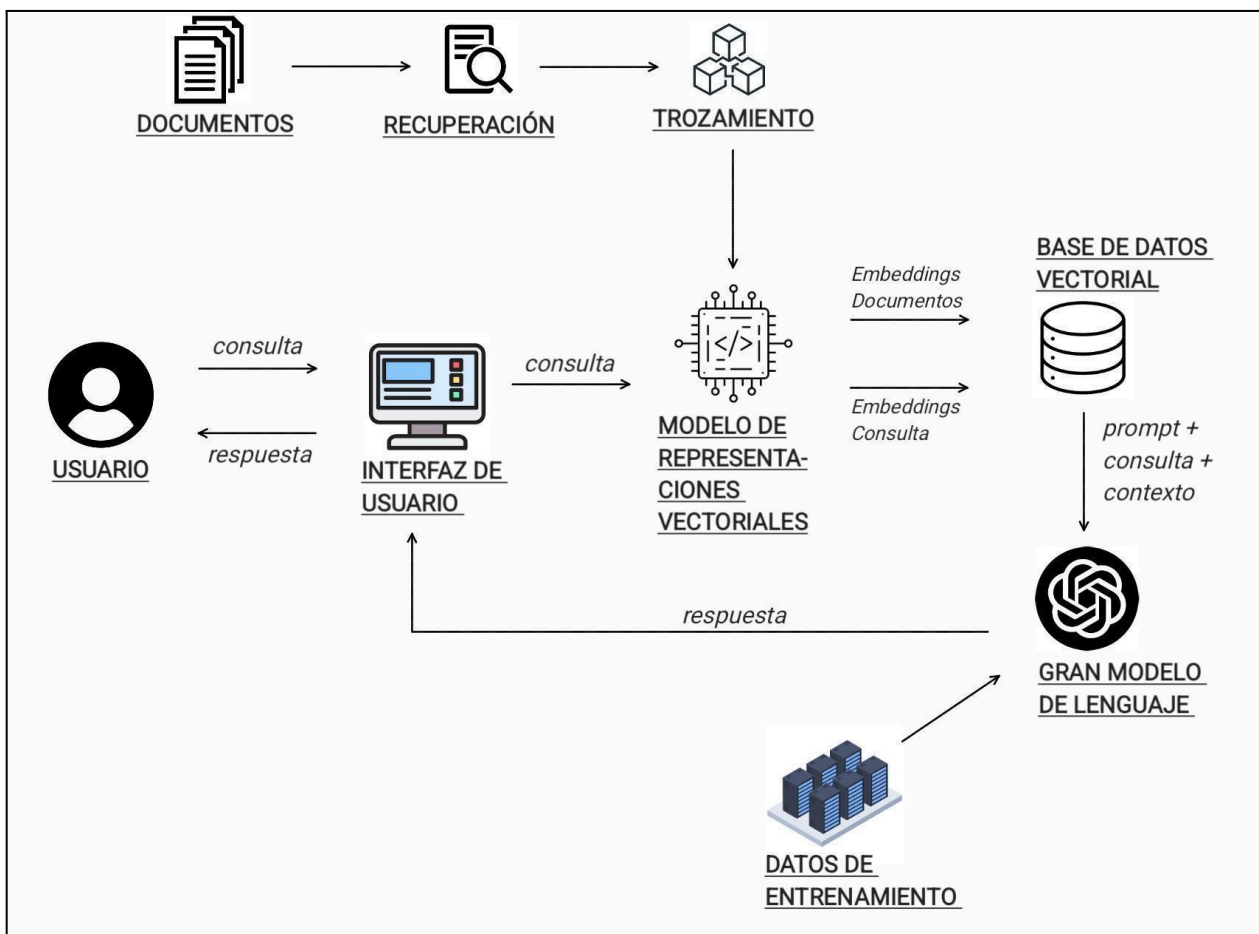
necesidad de proponer mejoras acústicas basadas en principios físicos. Este estudio de caso se adoptó como escenario central de uso durante la evaluación experta, de modo que la interacción con el chatbot se realizara bajo un contexto suficientemente denso y estandarizable, sin restringir el dominio conceptual del sistema únicamente a ese problema. En términos de diseño, esto exigió definir criterios para que el chatbot pueda, por una parte, anclar sus orientaciones en los datos del caso cuando la consulta sea vaga o incompleta y, por otra, aceptar otros contextos cuando el usuario explicita que está trabajando una situación distinta dentro del modelo ondulatorio.

En tercer lugar, se diseñó la arquitectura conceptual del recurso considerando que el chatbot debía mantener trazabilidad y coherencia disciplinar, y que su desempeño debía ser validable bajo dimensiones como usabilidad, utilidad y confianza mediante juicio experto. Esto llevó a definir una arquitectura que integra un modelo conversacional con acceso controlado a un conjunto de documentos de apoyo, privilegiando la consistencia con material disciplinar y orientador del proyecto. Adicionalmente, se incorporó como criterio explícito que el sistema no entregue productos listos para evaluación académica ni fomente plagio, resguardando integridad académica como condición de confianza.

Finalmente, y como condición de diseño asociada a gobernanza y uso responsable, se delimitó lo que el sistema debe comunicar durante la interacción: el chatbot puede declarar límites de dominio y sugerir contraste con docente o material oficial, pero debe evitar revelar detalles internos de su implementación. Este criterio se entiende como parte del diseño del comportamiento y no como un aspecto meramente técnico, pues condiciona la experiencia de uso y la forma en que se construye confianza en el recurso.

### **Arquitectura conceptual y técnica del sistema**

La Figura 1 representa la arquitectura general del sistema diseñada para este prototipo, mostrando el flujo desde los documentos de apoyo hasta la generación de respuestas del chatbot, mediado por un componente de recuperación de información y un modelo de lenguaje:



**Figura 1.** Arquitectura general que grafica el flujo interno, desde la consulta del usuario hasta la respuesta obtenida.

Desde lo conceptual, la arquitectura se organiza como un sistema de tutoría conversacional que combina dos capas:

**Capa pedagógica de mediación conversacional:** Aquí se ubican las reglas de interacción ( micro-diálogos, diagnóstico breve, exigencia de participación del usuario, metacognición y tratamiento de concepciones alternativas). Esta capa es la que define cómo apoyar al usuario (el chatbot guía, solicita justificaciones, detecta errores típicos y promueve reformulaciones, priorizando comprensión significativa por sobre respuestas inmediatas).

**Capa de soporte disciplinar con recuperación:** Esta capa busca sostener la pertinencia conceptual y reducir respuestas inconsistentes al recuperar fragmentos relevantes desde un corpus de documentos del proyecto (por ejemplo, estudio de caso y materiales

disciplinarios). La recuperación permite que el chatbot fundamente sus explicaciones y ejemplos en insumos previamente seleccionados, manteniendo coherencia con el marco curricular y con el escenario de validación.

Desde lo técnico, la arquitectura se describe como un pipeline con los siguientes componentes funcionales, alineados con la Figura 1:

- Repositorio de documentos de apoyo: Conjunto de archivos que constituyen la base de conocimiento controlada del sistema (incluyendo el estudio de caso y materiales disciplinarios pertinentes).
- Preprocesamiento y trozamiento de documentos: Módulo que segmenta los documentos en fragmentos manejables para recuperación posterior, procurando que cada fragmento conserve coherencia semántica suficiente para ser usado como contexto.
- Modelo de representaciones vectoriales: Componente que transforma tanto fragmentos de documentos como consultas del usuario en vectores, habilitando comparación por similitud.
- Base de datos vectorial: Estructura donde se almacenan los vectores de los fragmentos junto con su contenido asociado, permitiendo búsqueda eficiente cuando llega una consulta.
- Interfaz y gestor de consulta: Componente que recibe la pregunta del usuario (enmarcada, por defecto, en el estudio de caso de la sala de ensayos durante validación) y gestiona el estado mínimo de la conversación necesario para sostener el micro-diálogo.

- Recuperación de contexto: Dada la consulta, el sistema calcula su vector, busca en la base vectorial los fragmentos más similares y selecciona un conjunto acotado de evidencias para contextualizar la respuesta.
- Constructor de prompt: Módulo que integra las reglas de comportamiento del tutor (micro-diálogos, metacognición, tratamiento de concepciones alternativas), los datos del contexto si corresponden (por ejemplo, dimensiones de la sala, ubicación del parlante) y los fragmentos recuperados.
- Modelo de lenguaje: Componente generativo que produce la respuesta final a partir del prompt construido. En términos de diseño, su función se limita a generar una salida conversacional coherente con los criterios pedagógicos y con la evidencia recuperada.
- Post-proceso y entrega de respuesta: Ajustes finales de presentación (claridad, segmentación en pasos cuando corresponda) y entrega en la interfaz, cuidando no revelar detalles internos del sistema.

### **4.5.3 Desarrollo**

La fase de desarrollo correspondió a la implementación técnica del prototipo definido en la fase de diseño, materializando en un artefacto funcional la arquitectura conceptual y los criterios pedagógicos del chatbot. En términos operacionales, esta fase consistió en construir una aplicación web ejecutable, con un pipeline interno de tipo RAG y con soporte multimodal, asegurando que el sistema resultante fuese estable, reproducible y apto para ser validado posteriormente mediante juicio de expertos, sin incorporar aún interpretaciones evaluativas ni ajustes derivados de la validación.

### **Entorno de despliegue y empaquetamiento del prototipo en Hugging Face**

El prototipo fue desplegado como una aplicación alojada en Hugging Face Spaces, lo cual permite acceso remoto vía navegador y estandariza el entorno de ejecución. En la práctica, el

desarrollo se organizó como un repositorio con una estructura típica de Space basada en Gradio:

- un archivo principal de aplicación (app.py)
- un archivo de dependencias (requirements.txt) que fija versiones
- una hoja de estilos (styles.css) para personalización de interfaz
- un directorio interno para documentos PDF (./pdfs) usados por el componente RAG
- y un directorio para el prompt del sistema (por defecto ./prompts/PROMPT\_SISTEMA.txt).

En términos de configuración segura, el prototipo utiliza variables de entorno para manejar credenciales y parámetros sensibles, especialmente:

- OPENAI\_API\_KEY para acceso al proveedor de modelos,
- HF\_TOKEN y HF\_DATASET\_ID para el envío de logs a un dataset remoto,
- PROMPT\_SISTEMA\_PATH para controlar el prompt del sistema sin acoplarlo al código.

El desarrollo se apoya en un conjunto de librerías que responden directamente a los componentes de la arquitectura definida:

### **Gradio (gradio)**

Se utiliza para construir la interfaz web y orquestar los eventos de interacción. En particular, se emplean gr.Blocks como contenedor principal, gr.Chatbot para representar la conversación (con type="messages"), gr.MultimodalTextbox para permitir entrada combinada de texto y archivos (audio/imagen), gr.Audio para entregar la respuesta sintetizada, gr.State para persistir el estado conversacional y gr.HTML para componentes de interfaz estáticos (barra superior y nota de advertencia).

### **OpenAI SDK (openai)**

Se utiliza como cliente de bajo nivel (OpenAI(api\_key=...)) para:

- Transcripción de audio mediante `client.audio.transcriptions.create(...)` con un modelo tipo Whisper,
- Síntesis de voz (TTS) mediante `client.audio.speech.create(...)`,
- Análisis de imágenes mediante `client.chat.completions.create(...)` recibiendo la imagen como `image_url` en formato data URL base64.

**LangChain (langchain, langchain-community, langchain-openai, langchain-text-splitters)**

Se emplea para ensamblar el pipeline RAG y la lógica de conversación con recuperación:

- PyMuPDFLoader para cargar PDFs como documentos,
- CharacterTextSplitter para segmentación en fragmentos con solapamiento,
- OpenAIEmbeddings para generar representaciones vectoriales,
- FAISS como vectorstore local,
- ConversationalRetrievalChain como cadena que integra recuperación + modelo conversacional,
- ChatPromptTemplate, SystemMessagePromptTemplate y HumanMessagePromptTemplate para estructurar el prompt final.

**PyMuPDF (pymupdf)**

Se utiliza indirectamente a través de PyMuPDFLoader para extracción de texto desde PDFs.

**FAISS (faiss-cpu)**

Se utiliza como motor de búsqueda vectorial local para recuperación semántica eficiente.

**Pillow (PIL/pillow)**

Se utiliza para abrir y normalizar imágenes (`Image.open(...).convert("RGB")`) antes de transformarlas a base64.

**huggingface\_hub**

Se utiliza para el envío automático de logs a un dataset mediante `HfApi.upload_file(...)`.

### **Dependencias de soporte (httpx, tiktoken, pydantic)**

Aportan compatibilidad y funcionamiento interno de clientes, serialización, validaciones y utilidades dependientes de las librerías principales.

### **Componentes del prototipo implementados en app.py y su interacción**

El desarrollo implementa el prototipo como un conjunto de funciones y objetos globales que se inicializan al cargar la aplicación. Esto permite que el pipeline RAG quede disponible desde el inicio y que la interfaz pueda operar de forma reactiva ante cada envío del usuario.

#### **1) Inicialización de modelos**

Al inicio se valida la existencia de OPENAI\_API\_KEY. Con esta clave se crean dos objetos centrales:

- `client = OpenAI(api_key=OPENAI_API_KEY)`: cliente para funcionalidades de audio e imagen.
- `llm_text = ChatOpenAI(model="gpt-4o-mini", temperature=0, openai_api_key=OPENAI_API_KEY)`: cliente “alto nivel” integrado a LangChain para la generación textual con control de temperatura (0) orientado a estabilidad y consistencia.

Se definen además constantes de modelos por tarea:

- `VISION_MODEL` para análisis de imagen,
- `WHISPER_MODEL` para transcripción,
- `TTS_MODEL` para síntesis.

#### **2) Construcción del componente RAG (carga, segmentación, embeddings, vectorstore, retriever)**

El componente RAG se implementa como una preparación automática del corpus al iniciar la app:

- `load_all_pdfs(pdf_dir)`: recorre el directorio `./pdfs`, detecta archivos PDF y los carga con `PyMuPDFLoader(...).load()`, produciendo una lista de documentos (típicamente con contenido por página y metadatos asociados).
- `build_vectorstore(docs)`: aplica `CharacterTextSplitter(chunk_size=1000, chunk_overlap=150)` para convertir documentos en fragmentos recuperables; luego genera embeddings con `OpenAIEmbeddings(...)` y construye el índice local con `FAISS.from_documents(chunks, embeddings)`. Finalmente, crea un recuperador con `vs.as_retriever(search_kwargs={"k": 4})`, fijando `k=4` como cantidad de fragmentos a recuperar por consulta.

Este proceso inicializa dos variables globales:

- `VECTORSTORE` (la base vectorial),
- `RETRIEVER` (el recuperador).

Si no existen documentos o no se generan fragmentos, el sistema deja `RETRIEVER` no disponible y el prototipo opera en modo de respuesta directa sin recuperación. Esta decisión implementa un comportamiento “degradado” controlado: el chatbot sigue funcionando, pero sin soporte documental.

### 3) Implementación del prompt del sistema y su integración en la cadena conversacional

El comportamiento pedagógico definido en diseño se operacionaliza mediante un prompt de sistema cargado desde archivo:

- `PROMPT_PATH = os.getenv("PROMPT_SISTEMA_PATH", "./prompts/PROMPT_SISTEMA.txt")`
- `PROMPT_SISTEMA = _load_text(PROMPT_PATH)`

La función `_load_text(...)` incorpora un fallback seguro en caso de que el archivo no esté disponible, evitando que la app falle por ausencia del prompt y dejando explícita una versión mínima del rol tutorial.

A partir del prompt se construye una plantilla de sistema que incorpora el contexto recuperado:

- `system_tmpl = PROMPT_SISTEMA + "\n\n-----\nFragmentos de apoyo:\n{context}\n-----"`

Luego se define un `ChatPromptTemplate` con tres piezas:

- `SystemMessagePromptTemplate.from_template(system_tmpl)` para rol + contexto recuperado,
- `SystemMessagePromptTemplate.from_template("Historial reciente:\n{chat_history_str}")` para proveer una versión textual del historial,
- `HumanMessagePromptTemplate.from_template("{question}")` para la pregunta actual.

Esta estructura es relevante: en vez de entregar solo `chat_history` como lista de pares, el desarrollo incorpora también un `chat_history_str` que resume el historial en un formato legible (“Usuario: ... / Asistente: ...”), lo que refuerza continuidad y control del micro-diálogo en la interacción.

Con el recuperador disponible, se crea la cadena principal del prototipo:

- `QA_CHAIN = ConversationalRetrievalChain.from_llm(...)`

Esta cadena integra:

- `llm=llm_text` como generador,
- `retriever=RETRIEVER` como componente de búsqueda,
- `combine_docs_chain_kwargs={"prompt": prompt}` para asegurar que la generación use la plantilla pedagógica definida,
- `return_source_documents=False` para no entregar documentos fuente como salida al usuario (manteniendo la interfaz centrada en tutoría y evitando sobrecarga).

#### **4) Utilidades multimodales: imagen, audio, TTS y postproceso**

El prototipo incorpora tres funciones clave que amplían la interacción:

- `transcribir_audio(audio_path)`: abre el archivo binario y ejecuta `client.audio.transcriptions.create(model=WHISPER_MODEL, file=f)`. El resultado se integra como texto de consulta.
- `sintetizar_audio(texto)`: ejecuta `client.audio.speech.create(model=TTS_MODEL, voice="nova", input=texto)` y guarda la respuesta en un archivo MP3 temporal, retornando la ruta para que `gr.Audio` lo reproduzca.
- `describir_imagen(img)`: transforma la imagen a base64 con `_pil_to_base64`, luego llama `client.chat.completions.create(...)` con un mensaje de sistema que exige transcripción literal y extracción estructurada (transcripción, descripción gráfica, datos, pregunta, opciones, observaciones), y entrega esa descripción como parte del input del modelo principal.

Adicionalmente, se implementa `clean_latex(text)` como postproceso para mejorar legibilidad: normaliza fracciones ( $\frac{a}{b} \rightarrow a/b$ ), elimina delimitadores matemáticos ( $\$...\$, \dots$ ), reemplaza símbolos LaTeX por Unicode (por ejemplo,  $\lambda, \Delta, \times$ ), y corrige residuos frecuentes de formato. Este componente busca reducir artefactos de renderizado típicos en interfaces web y favorecer claridad textual en una herramienta de tutoría.

## 5) Gestión del historial y persistencia de estado conversacional

Dado que la interfaz `gr.Chatbot(type="messages")` trabaja con mensajes en formato lista de diccionarios (`{"role": "...", "content": "..."}`), se implementa un puente para LangChain:

- `build_history_pairs(messages)`: transforma la lista de mensajes en una lista de tuplas (usuario, asistente) para alimentar `chat_history` en `QA_CHAIN.invoke(...)`.
- `format_chat(pairs)`: convierte esos pares a un string resumido para el campo `chat_history_str`.
- `ensure_session_id(current_id)`: crea un identificador único con `uuid.uuid4().hex` si no existe, permitiendo trazabilidad por sesión.

Esta capa técnica es crítica para sostener los micro-diálogos: el sistema no opera como preguntas aisladas, sino como una secuencia donde cada turno puede retomar lo dicho, detectar inconsistencias y sostener tareas encadenadas.

## 6) Orquestación principal del turno: función responder()

La función responder(audio\_path, img\_pil, user\_msg, messages, session\_id) implementa el flujo completo de inferencia por turno:

### 1. Normalización de entrada:

- Si hay audio y no hay texto, transcribe audio para obtener user\_msg.
- Si no hay texto ni imagen, retorna sin cambios.

### 2. Enriquecimiento por imagen:

- Si existe img\_pil, ejecuta describir\_imagen(img\_pil) y concatena su salida a model\_input bajo el bloque IMG\_DESC.
- Esto mantiene separación explícita entre lo que el usuario escribió y lo que el sistema extrajo desde la imagen.

### 3. Construcción de historial:

- Convierte messages a pairs y genera chat\_history\_str.

### 4. Generación con RAG o modo directo:

- Si QA\_CHAIN está disponible, ejecuta QA\_CHAIN.invoke({ "question": model\_input, "chat\_history": pairs, "chat\_history\_str": format\_chat(pairs) }) y extrae answer.
- Si no hay QA\_CHAIN, llama directamente a llm\_text.invoke(model\_input).content.
- Se incluye un manejo de excepciones que intenta una segunda llamada directa antes de devolver un mensaje de error controlado.

### 5. Postproceso y salida:

- Aplica clean\_latex a la respuesta.
- Genera audio con sintetizar\_audio (si falla, el sistema continúa sin audio).

### 6. Actualización de conversación y logging:

- Añade el turno a messages (mensaje del usuario y respuesta del asistente).
- Registra el turno con log\_turn(session\_id, model\_input, respuesta\_texto).

Finalmente retorna los objetos necesarios para que la UI actualice: conversación, audio, estado de mensajes y sesión, y reinicio del input.

## 7) Integración con la interfaz: `handle_multimodal()` y eventos de Gradio

`gr.MultimodalTextbox` entrega su contenido como un diccionario con claves como `text` y `files`. Para procesarlo de forma controlada se implementa `handle_multimodal(message, messages, session_id)`:

- Extrae `user_msg` desde `message["text"]`.
- Recorre `message["files"]` y selecciona:
  - un audio solo si **no** hay texto escrito (regla de precedencia: texto > audio),
  - una imagen si existe (primer archivo de imagen válido).
- Abre la imagen con `PIL.Image` y la normaliza a RGB.

Luego invoca `responder(...)` y retorna el paquete de salidas esperadas por Gradio.

En la UI, el evento central es:

- `chat_in.submit(handle_multimodal, inputs=[chat_in, state_messages, state_session], outputs=[chatbot, audio_out, state_messages, state_session, chat_in], show_progress="full")`

Esto significa que cada envío del usuario actualiza simultáneamente:

- la conversación visible (`gr.Chatbot`),
- el reproductor de audio (`gr.Audio`),
- el estado interno de mensajes (`gr.State`),
- el estado de sesión (`gr.State`),
- y limpia el cuadro de entrada (`gr.update(value=None)`).

## 8) Registro y subida automática de logs a Hugging Face

Para trazabilidad operacional, el sistema implementa logging por turno en formato JSON Lines:

- `log_turn(...)` escribe un registro con:
  - timestamp UTC (ts),
  - `session_id`,
  - `model_input` (incluye texto del usuario junto a bloques agregados como `IMG_DESC`),
  - salida del asistente.

Tras escribir, cuenta líneas del archivo local y, al superar un umbral (`UPLOAD_EVERY_N_LINES = 3`), ejecuta `push_log_to_hub()`:

- comprime el log como `.jsonl.gz`,
- genera un nombre único `logs_YYYY-MM-DD_<tag>.jsonl.gz` para evitar sobreescritura,
- sube el archivo con `HfApi.upload_file(..., repo_type="dataset")`,
- y reinicia el archivo local para el siguiente bloque.

Este mecanismo permite acumular evidencia de funcionamiento real del prototipo, manteniendo un flujo automatizado compatible con el despliegue en Space.

### **Configuración visual y estilos**

La interfaz se complementa con una hoja `styles.css` cargada por `load_css()`, que define ajustes de layout (ancho, padding), estructura de topbar y nota a pie, y detalles de presentación como radios de borde en mensajes. Este componente se integra a `gr.Blocks` mediante el parámetro `css=custom_css`, asegurando consistencia visual sin introducir cambios funcionales al pipeline conversacional.

### **Resultado final de la construcción**

En síntesis, la fase de desarrollo concluye con un prototipo alojado en Hugging Face Spaces que integra:

- una interfaz conversacional construida con gr.Blocks, gr.Chatbot y gr.MultimodalTextbox,
- un pipeline RAG implementado con PyMuPDFLoader, CharacterTextSplitter, OpenAIEmbeddings, FAISS y ConversationalRetrievalChain,
- utilidades multimodales mediante OpenAI (Whisper, TTS y análisis de imagen),
- postproceso textual con clean\_latex,
- y un mecanismo de logging con subida automática a un dataset mediante HfApi.

Con ello, el software queda técnicamente preparado para ser sometido a la fase de evaluación, manteniendo la separación metodológica entre construcción del prototipo y su validación.

#### **4.5.4 Evaluación**

La fase de evaluación tiene por propósito validar el desempeño del prototipo mediante juicio de expertos, en un único momento de aplicación y sin intervención pedagógica en aula ni manipulación de variables, con el fin de caracterizar el recurso en términos de usabilidad, utilidad y confianza, y de levantar orientaciones cualitativas para su perfeccionamiento e implementación futura.

En coherencia con el diseño metodológico del estudio, se definió un muestreo intencional basado en criterios de experticia, considerando tres participantes con trayectoria en física y educación, de modo que sus valoraciones se fundamenten tanto en criterios disciplinares como didácticos vinculados al aprendizaje del modelo ondulatorio.

En este marco, y para resguardar el tratamiento ético y la comunicación académica de resultados, se estableció desde el inicio que los participantes serían reportados mediante códigos de anonimización (E1, E2 y E3) en el capítulo de resultados, evitando asociar respuestas a identidades personales.

Previo a la aplicación del instrumento, se implementó un conjunto de acciones de preparación orientadas a estandarizar las condiciones de interacción y asegurar trazabilidad metodológica. En primer lugar, se consolidó una versión operativa del prototipo alojada en Hugging Face Spaces, disponible para acceso remoto vía navegador, procurando que el sistema se

mantuviera estable durante el periodo de validación y separando metodológicamente la construcción del software de su evaluación.

En segundo lugar, se definió un marco común de uso mediante un estudio de caso, con el propósito de situar las consultas de los expertos en un escenario didáctico compartido y comparable, reduciendo variabilidad por contextos arbitrarios de interacción. Este escenario corresponde al caso recomendado para orientar el tipo de preguntas a implementar en el chatbot y se asocia al contexto de validación declarado en el diseño del estudio.

En tercer lugar, se preparó el paquete de participación para los expertos, estructurado en documentos con funciones diferenciadas. Se incluyó una carta de solicitud de participación, en la que se explicita el carácter voluntario del proceso, la ausencia de incentivos o consecuencias derivadas de su respuesta, la autorización de uso anónimo de la información, y la identificación de un canal formal de consultas asociado a la dirección de tesis.

Adicionalmente, se incorporó el documento del instrumento de validación, el cual organiza la evaluación del prototipo a través de afirmaciones agrupadas en las dimensiones de usabilidad, utilidad y confianza, con una escala Likert de cinco puntos, complementada por preguntas abiertas orientadas a recoger observaciones para mejora.

En cuarto lugar, se definió un protocolo de orientación para el desarrollo de la interacción, enviado a los expertos mediante un correo con instrucciones secuenciadas. En dicho mensaje se contextualiza el propósito de la evaluación del prototipo y se explicita que el procedimiento contempla, antes de responder el instrumento, la lectura del estudio de caso, seguida de una interacción en la que el experto simula el rol de estudiante en la situación descrita, realizando consultas diversas al chatbot para observar su comportamiento. Posteriormente, y solo tras esta interacción, se solicita completar la pauta de valoración y responder las preguntas abiertas. El correo indica además un tiempo estimado de dedicación de aproximadamente 30 minutos como referencia operativa para los participantes.

Finalmente, para mantener coherencia entre el procedimiento y el enfoque mixto declarado, se estableció que la validación produciría dos tipos de insumo complementario. Por una parte, valoraciones cuantitativas estructuradas en las tres dimensiones y, por otra, evidencia cualitativa derivada de comentarios y observaciones, organizables en una matriz de análisis, con el objetivo de integrar ambos planos a nivel de interpretación y sostener recomendaciones de mejora con respaldo argumentado.

## **Capítulo V: Validación de Expertos**

### **5.1 Propósito del capítulo y lógica de organización**

El presente capítulo tiene por propósito presentar, ordenar y analizar los resultados obtenidos en la validación del prototipo mediante juicio de expertos. En coherencia con el enfoque mixto del estudio y con su carácter descriptivo, la evidencia se organiza en dos planos complementarios. El primero corresponde a valoraciones cuantitativas estructuradas en las dimensiones de usabilidad, utilidad y confianza, aplicadas mediante una escala Likert de cinco puntos. El segundo corresponde a evidencia cualitativa derivada de preguntas abiertas, orientada a identificar aspectos valorados, limitaciones observadas y recomendaciones para el perfeccionamiento e implementación futura del recurso.

La estructura del capítulo se compone de una síntesis del procedimiento de validación y del instrumento aplicado, seguida del análisis de resultados cuantitativos y cualitativos, y finaliza con una integración interpretativa que busca caracterizar el desempeño del prototipo y derivar orientaciones justificadas para iteraciones posteriores.

### **5.2 Procedimiento de validación por juicio experto**

La validación se implementó en un único momento de aplicación, sin intervención pedagógica en aula ni manipulación de variables, con el propósito de caracterizar el prototipo en términos de usabilidad, utilidad y confianza, y levantar orientaciones cualitativas para su perfeccionamiento e implementación futura. Se trabajó con un muestreo intencional basado en criterios de experticia, considerando tres participantes con trayectoria en física y educación, de modo que sus apreciaciones incorporaran criterios disciplinares y didácticos. Para resguardar el tratamiento ético de los datos, los resultados se reportan mediante códigos de anonimización E1, E2 y E3.

Previo a la aplicación del instrumento, se ejecutaron acciones destinadas a estandarizar condiciones de interacción y asegurar trazabilidad metodológica. Se consolidó una versión operativa del prototipo alojada en Hugging Face Spaces, disponible para acceso remoto vía navegador, procurando estabilidad durante el periodo de validación y manteniendo la separación metodológica entre la construcción del software y su evaluación. Se definió un marco común de uso mediante un estudio de caso, con el propósito de situar las consultas en un escenario didáctico compartido y comparable. Se preparó un paquete de participación para los expertos, que incluyó la carta de solicitud de participación y el instrumento de validación. Finalmente, se envió un correo con instrucciones secuenciadas que estableció la lectura previa del estudio de caso, la interacción con el prototipo simulando el rol de estudiante, y solo después la respuesta del instrumento, incorporando un tiempo estimado de dedicación como referencia operativa.

### **5.3 Instrumento aplicado y estrategia general de análisis**

El instrumento de validación organiza la evaluación del prototipo mediante afirmaciones agrupadas en tres dimensiones, usabilidad, utilidad y confianza, valoradas en escala Likert de cinco puntos, complementada por preguntas abiertas orientadas a recoger observaciones para mejora. La escala considera un continuo desde el desacuerdo hasta el acuerdo total, permitiendo registrar intensidad de percepción en cada afirmación.

Dado el tamaño de la muestra, el análisis cuantitativo se aborda con propósito descriptivo. En consecuencia, los resultados se presentarán enfatizando tendencias generales, patrones de dispersión y diferencias entre dimensiones, evitando inferencias estadísticas. El análisis cualitativo se orienta a identificar temas recurrentes en los comentarios, y a organizar la evidencia en categorías interpretativas que permitan fundamentar recomendaciones de perfeccionamiento.

#### 5.4 Resultados cuantitativos

Los resultados cuantitativos evidencian una evaluación global favorable del prototipo, con matices entre dimensiones y con variaciones puntuales según el tipo de criterio evaluado. En términos generales, las valoraciones tienden a concentrarse en niveles de acuerdo, sugiriendo que el sistema logra satisfacer condiciones funcionales básicas para su uso y que posee un desempeño percibido como pertinente para apoyar procesos de aprendizaje vinculados al modelo ondulatorio. Sin embargo, las respuestas abiertas y ciertos indicadores de experiencia reportada sugieren que la evaluación cuantitativa debe interpretarse junto con la evidencia cualitativa para comprender mejor las tensiones del sistema en escenarios de uso realista.

**Tabla 5.1.** Resultados cuantitativos por ítem y por experto (Likert 1–5)

| Dimensión  | Ítem  | E1 | E2 | E3 | Promedio ítem | Rango |
|------------|---|----|----|----|---------------|-------|
| Usabilidad | El chatbot es fácil de usar y navegar, incluso para académicos con poca experiencia tecnológica.          | 5  | 4  | 5  | 4.67          | 4–5   |
| Usabilidad | La interfaz del chatbot es clara y permite identificar rápidamente cómo interactuar.                      | 4  | 4  | 5  | 4.33          | 4–5   |
| Usabilidad | El chatbot responde de manera fluida, con tiempos adecuados que favorecen la continuidad del aprendizaje. | 5  | 2  | 4  | 3.67          | 2–5   |
| Usabilidad | El lenguaje utilizado por el chatbot es adecuado y comprensible para académicos.                          | 5  | 5  | 5  | 5.00          | 5–5   |

|           |   |   |   |   |      |     |
|-----------|---|---|---|---|------|-----|
| Utilidad  | El chatbot entrega explicaciones claras que facilitan la comprensión del modelo ondulatorio.            | 4 | 2 | 4 | 3.33 | 2-4 |
| Utilidad  | Las actividades o preguntas propuestas por el chatbot fomentan el aprendizaje activo.                   | 4 | 3 | 5 | 4.00 | 3-5 |
| Utilidad  | El chatbot complementa adecuadamente las clases de Física reforzando contenidos del programa del curso. | 5 | 4 | 5 | 4.67 | 4-5 |
| Utilidad  | El chatbot conecta el modelo ondulatorio con ejemplos cotidianos o experimentos simples.                | 5 | 4 | 5 | 4.67 | 4-5 |
| Confianza | La información entregada por el chatbot es precisa y coherente con el programa del curso.               | 5 | 4 | 5 | 4.67 | 4-5 |
| Confianza | El chatbot mantiene consistencia en sus respuestas ante preguntas similares.                            | 4 | 4 | 4 | 4.00 | 4-4 |
| Confianza | El chatbot evita respuestas incorrectas, confusas o inapropiadas.                                       | 5 | 4 | 4 | 4.33 | 4-5 |
| Confianza | Percibo al chatbot como una herramienta confiable para apoyar mi aprendizaje.                           | 4 | 3 | 5 | 4.00 | 3-5 |

### **5.4.1 Usabilidad**

En la dimensión de usabilidad, los resultados apuntan a una percepción positiva respecto de la facilidad de uso, la claridad general de la interfaz y la adecuación del lenguaje para el perfil de usuarios evaluadores. Este desempeño se interpreta como una condición habilitante para la implementación educativa, ya que un tutor conversacional requiere una interacción estable y comprensible para sostener procesos de indagación y diálogo disciplinar. No obstante, se observa un ítem particularmente sensible asociado a fluidez y continuidad temporal de la interacción, con variabilidad marcada entre evaluadores, lo que sugiere que esta dimensión debe comprenderse también a la luz de las observaciones cualitativas.

### **5.4.2 Utilidad**

La dimensión de utilidad muestra apreciaciones favorables en torno a la capacidad del sistema para ofrecer explicaciones y proponer preguntas o actividades que se asocian al aprendizaje activo. En paralelo, existe evidencia cualitativa que cuestiona la ventaja comparativa del recurso en relación con otras formas de apoyo al aprendizaje, lo que introduce una divergencia relevante. Esta divergencia no invalida la utilidad del prototipo, pero sugiere que la utilidad debe comprenderse como una propiedad dependiente del encuadre pedagógico de uso, del tipo de tareas planteadas y de la expectativa del evaluador respecto del rol específico del chatbot en el ecosistema de aula o estudio.

### **5.4.3 Confianza**

En la dimensión de confianza, la percepción general tiende a ser favorable en cuanto a precisión, coherencia y evitación de respuestas incorrectas o inapropiadas. Sin embargo, se reportan episodios puntuales que tensionan la consistencia conceptual ante preguntas relacionadas y que, por su naturaleza disciplinar, requieren especial atención en futuras iteraciones. Dado que el prototipo se propone como un recurso de apoyo al aprendizaje del modelo ondulatorio, la estabilidad conceptual frente a variaciones de pregunta constituye un criterio crítico para su credibilidad pedagógica.

**Tabla 5.2.** Estadísticos descriptivos por dimensión y promedio global

| <b>Dimensión</b>                               | <b>Media E1</b> | <b>Media E2</b> | <b>Media E3</b> | <b>Media global</b> | <b>Mínimo</b> | <b>Máximo</b> |
|--|-----------------|-----------------|-----------------|---------------------|---------------|---------------|
| Usabilidad                                     | 4.75            | 3.75            | 4.75            | 4.42                | 2             | 5             |
| Utilidad                                       | 4.50            | 3.25            | 4.75            | 4.17                | 2             | 5             |
| Confianza                                      | 4.50            | 3.75            | 4.50            | 4.25                | 3             | 5             |
| <b>Promedio global (todas las dimensiones)</b> | <b>4.58</b>     | <b>3.58</b>     | <b>4.67</b>     | <b>4.28</b>         | <b>2</b>      | <b>5</b>      |

### **5.5 Resultados cualitativos**

Las preguntas abiertas permiten profundizar el sentido de las valoraciones cuantitativas, identificando fortalezas percibidas y oportunidades de mejora. Para su análisis, la evidencia se organiza en dos apartados. El primero sintetiza los aspectos valorados como más útiles para aprender sobre el modelo ondulatorio. El segundo reúne sugerencias explícitas de mejora para fortalecer el aprendizaje activo y la experiencia de uso.

#### **5.5.1 Aspectos más útiles reportados para aprender el modelo ondulatorio**

En términos de fortalezas, un elemento recurrente es la orientación del sistema hacia la acción y hacia la continuidad de la actividad, más que hacia definiciones cerradas. Se reporta que el chatbot trabaja sobre ejemplos, genera preguntas de indagación y deja tareas, lo que se interpreta como un soporte relevante para prácticas de aprendizaje activo, especialmente cuando el usuario debe formular preguntas y sostener una exploración progresiva del fenómeno.

Asimismo, se destaca la capacidad del sistema para acompañar la resolución de ejercicios con secuencias explicativas claras y con una estructura paso a paso, incluso en tareas de mayor

complejidad. Desde esta perspectiva, el prototipo es percibido como útil cuando funciona como andamiaje cognitivo durante la resolución de problemas, favoreciendo comprensión y ordenamiento conceptual. En la misma línea, se valora su potencial para proponer actividades experimentales factibles con materiales accesibles y para articular modelización, fenómenos físicos reales y orientaciones didácticas activas, lo cual amplía su utilidad más allá de la explicación conceptual.

Finalmente, se reporta como fortaleza un comportamiento ético adecuado ante contenidos inapropiados, lo que se considera un rasgo esencial en herramientas destinadas a contextos educativos. Esta observación adquiere relevancia desde una perspectiva de confianza y seguridad, ya que refuerza la gobernanza de respuestas y el respeto por límites de uso.

De manera contrastante, un experto declara no observar ventajas por sobre otros métodos de apoyo al aprendizaje. Este comentario representa un hallazgo de alto peso interpretativo, ya que obliga a precisar el diferencial pedagógico del prototipo y a justificar con mayor fuerza su rol como herramienta especializada, especialmente frente a alternativas de apoyo más generales.

### **5.5.2 Mejoras sugeridas para fortalecer aprendizaje activo y experiencia de uso**

Las sugerencias de mejora se concentran en continuidad conversacional, progresión dialógica, consistencia conceptual, inteligibilidad de la salida por voz y gestión del alcance del dominio. Como punto inicial, se observa una crítica a la repetición de solicitudes para estimar conocimientos previos en cada respuesta, proponiéndose que el sistema pueda extraer información de la misma conversación para evitar reinicios percibidos. Esta observación es relevante para el aprendizaje activo, ya que la continuidad del diálogo tutorial es un requisito para sostener metas, promover autorregulación y favorecer profundización.

En la misma línea, se reporta que en ocasiones el sistema responde con una pregunta muy similar a la formulada por el usuario, lo que frena la sensación de avanzar hacia una conversación más profunda. Esta evidencia sugiere que el andamiaje socrático requiere no solo preguntar, sino también asegurar progresión cognitiva visible, evitando reformulaciones que no incrementen la demanda conceptual o no aporten dirección.

Por otra parte, se describen episodios puntuales de contradicción, particularmente vinculados a conceptos críticos del modelo ondulatorio, como la invariancia de la frecuencia al cambiar de medio. Dado el propósito del prototipo, estos incidentes deben considerarse prioritarios, ya que afectan directamente la dimensión de confianza y el rigor disciplinar percibido.

Adicionalmente, se identifica la salida por voz como un componente prometedor, pero que requiere refinamientos significativos. Se reporta que el audio resulta confuso al verbalizar números y unidades, y que la dicción no distingue adecuadamente entre fórmulas, enunciados conceptuales y párrafos explicativos, dificultando la comprensión auditiva. Estas observaciones se interpretan como un requerimiento de mejora centrado en una transcripción más refinada y un tratamiento mejorado en la notación y unidades físicas en modalidad audio.

A su vez, se reporta que el chatbot evita responder preguntas no directamente relacionadas con el modelo ondulatorio y que, en consecuencia, no siempre logra vincular temas relacionados, por ejemplo, percepción del sonido y la interpretación de ciertos términos introducidos durante la interacción. Esta crítica no necesariamente implica un error, pero evidencia una brecha entre expectativas del evaluador y la política de alcance del sistema. En términos de diseño pedagógico, esto sugiere la necesidad de gestionar de forma más explícita el alcance disciplinar y de ofrecer transiciones guiadas cuando las preguntas se desplazan hacia conexiones legítimas con otros tópicos.

Finalmente, se sugiere fortalecer la identidad pedagógica del recurso, incorporando elementos como nombre, avatar o personalidad pedagógica, y habilitando opciones inclusivas en la voz. Estas recomendaciones se interpretan como dimensiones de mediación simbólica que podrían mejorar vínculo, apropiación y percepción de pertinencia del recurso en contextos educativos.

## **5.6 Integración interpretativa de hallazgos y discusión**

La integración de resultados permite sostener tres conclusiones analíticas. Por un lado, el prototipo presenta una aceptabilidad operativa favorable, especialmente en componentes que facilitan el uso y la comprensión del intercambio textual, lo que constituye una base necesaria para cualquier implementación educativa. Sin embargo, la continuidad conversacional y la inteligibilidad de la salida por voz emergen como puntos críticos que pueden afectar la experiencia de aprendizaje y la percepción de fluidez.

En un segundo plano, la utilidad pedagógica se manifiesta con fuerza cuando el sistema funciona como andamiaje para resolución guiada, promueve acciones de indagación y propone actividades. A la vez, el disenso respecto de su ventaja comparativa sugiere que la utilidad no debe entenderse como atributo intrínseco del chatbot, sino como propiedad emergente del diseño conversacional y del encuadre de uso. En este marco, la especificación del rol del sistema, su contribución diferenciadora y su integración con mediación docente constituyen elementos interpretativos centrales.

Como tercer eje interpretativo, la confianza aparece globalmente favorable, pero con alertas sobre consistencia conceptual y manejo de bordes del dominio. Para un tutor disciplinar, la estabilidad conceptual ante variaciones de pregunta no es un atributo secundario, sino un requisito de credibilidad pedagógica. Por ello, los episodios reportados de contradicción requieren ser interpretados como hallazgos de alta prioridad para futuros ciclos de perfeccionamiento.

**Matriz 5.1.** Integración de datos cuantitativos-cualitativos por dimensión

| <b>Dimensión</b>  | <b>Síntesis cuantitativa</b>         | <b>Evidencia cualitativa convergente</b>   | <b>Tensiones o divergencias cualitativas</b>   | <b>Interpretación integrada</b>  | <b>Recomendaciones derivadas</b>   |
|-------------------|--------------------------------------|--|--|--|--|
| <b>Usabilidad</b> | Media global <b>4.42</b> (rango 2-5) | Se reporta claridad de interfaz y lenguaje comprensible; se percibe el sistema como operativamente “usable” para interactuar y | Se señala fricción por continuidad conversacional (reinicios percibidos) y, en algunos casos, experiencia menos fluida por | La usabilidad es alta como condición habilitante, pero su sostenibilidad depende de que el diálogo acumule contexto y no | Priorizar continuidad de sesión y progresión conversacional; reducir solicitudes repetitivas de estimación de conocimientos previos; |

|                 |  |  |   |  |  |
|-----------------|--|--|---|--|--|
|                 |  | sostener consultas.  | tiempos/dinámica de respuesta.  | “reinicie” la interacción, especialmente en tareas extendidas.   | mantener claridad de interfaz y legibilidad.   |
| <b>Utilidad</b> | Media global<br><b>4.17</b><br>(rango 2-5) | Se valora apoyo a aprendizaje activo: ejemplos, preguntas de indagación, tareas; capacidad para acompañar resolución y proponer actividades experimental es con materiales accesibles. | Un experto declara no observar ventajas comparativas frente a otros apoyos; además, se menciona que el sistema puede evitar temas “relacionados” fuera del foco directo, lo que limita la percepción de utilidad en conexiones ampliadas. | La utilidad aparece como una propiedad dependiente del patrón de uso y expectativas: es alta cuando el chatbot actúa como andamiaje y promotor de actividad, pero se debilita cuando se espera un diferencial explícito o integración temática más amplia. | Explicitar posibilidades pedagógicas del chatbot como herramienta para la docencia; fortalecer estrategias de andamiaje (evitar preguntas repetitivas, promover progresión); mejorar manejo de “conexiones legítimas” sin perder foco. |

|                  |                                      |  |  |   |  |
|------------------|--------------------------------------|--|--|---|--|
| <b>Confianza</b> | Media global <b>4.25</b> (rango 3-5) | Se reconoce precisión general y comportamiento ético (rechazo de solicitudes inapropiadas); se percibe confiable como apoyo en términos amplios. | Se reportan incidentes puntuales de inconsistencia conceptual (p. ej., afirmaciones relacionadas con frecuencia/cambio de medio) y límites al responder conexiones temáticas, lo que puede interpretarse como pérdida de fiabilidad. | La confianza es globalmente favorable, pero presenta “puntos críticos” sensibles: bastan pocos episodios de contradicción para erosionar credibilidad, especialmente en un tutor disciplinar. | Reforzar control de consistencia en conceptos críticos del modelo ondulatorio; mejorar robustez ante preguntas de transferencia; gestionar explícitamente alcances/limitaciones para evitar interpretaciones de “falla”. |
|------------------|--------------------------------------|--|--|---|--|

### 5.7 Orientaciones de perfeccionamiento derivadas de la validación

En coherencia con el alcance de esta tesis, las orientaciones se presentan como recomendaciones para futuras iteraciones del prototipo, sin implementar cambios al software en el marco del presente estudio.

- Fortalecer continuidad conversacional, evitando reinicios percibidos y acumulando contexto pedagógico desde el diálogo.
- Asegurar progresión dialógica en el andamiaje socrático, evitando preguntas espejo y promoviendo incrementos claros de demanda cognitiva.

- Revisar consistencia conceptual en nociones críticas del modelo ondulatorio, priorizando coherencia ante preguntas indirectas o de transferencia.
- Mejorar inteligibilidad de la salida por voz, con foco en números, unidades y segmentación prosódica de expresiones matemáticas y explicaciones.
- Gestionar explícitamente el alcance del dominio y orientar conexiones con temas relacionados, evitando que la restricción disciplinar sea interpretada como falta de fiabilidad.
- Evaluar el fortalecimiento de identidad pedagógica e inclusividad como componentes de mediación simbólica que podrían mejorar la apropiación del recurso.

**Matriz 5.2.** Matriz FODA del prototipo (derivada de evidencia cualitativa)

| <b>Fortalezas</b>   | <b>Debilidades</b>  |
|---|---|
| <ul style="list-style-type: none"> <li>- Andamiaje útil para resolución de ejercicios, con explicaciones secuenciadas y lenguaje claro.</li> <li>- Capacidad de proponer actividades/experimentos factibles y contextualizados.</li> <li>- Orientación hacia la acción, indagación y continuidad de actividades.</li> <li>- Comportamiento ético adecuado ante solicitudes inapropiadas.</li> </ul> | <ul style="list-style-type: none"> <li>- Continuidad conversacional mejorable (reinicios percibidos; reiteración de estimación de conocimientos previos).</li> <li>- En ocasiones, respuestas que reflejan la pregunta sin aportar progresión (preguntas “espejo”).</li> <li>- Salida por voz con baja inteligibilidad en números/unidades y lectura de expresiones matemáticas; prosodia limitada.</li> <li>- Manejo restringido de temas relacionados que puede percibirse como limitación de fiabilidad/utilidad.</li> </ul> |
| <b>Oportunidades</b>  | <b>Amenazas</b>   |

|  |  |
|--|--|
| <ul style="list-style-type: none"> <li>- Refinar audio (prosodia, segmentación, tratamiento de unidades y fórmulas) para potenciar accesibilidad.</li> <li>- Fortalecer identidad pedagógica (nombre/avatar/personalidad) para mejorar apropiación como “compañero de aprendizaje”.</li> <li>- Clarificar y comunicar mejor el rol diferencial del sistema frente a apoyos generales.</li> <li>- Mejorar transiciones hacia conexiones legítimas (percepción del sonido, contexto experimental) manteniendo foco disciplinar.</li> </ul> | <ul style="list-style-type: none"> <li>- Episodios puntuales de contradicción conceptual pueden erosionar confianza y adopción.</li> <li>- Si no se explicita la propuesta de valor, puede persistir la percepción de “sin ventaja comparativa” en ciertos evaluadores.</li> <li>- Limitaciones de alcance percibidas como “no confiable” al salir del núcleo temático.</li> <li>- Audio confuso puede afectar significativamente la experiencia de usuarios que dependen de modalidad voz.</li> </ul> |
|--|--|

En síntesis, la validación por expertos permite caracterizar el prototipo como una herramienta con desempeño favorable en aspectos operativos y con potencial pedagógico para apoyar aprendizaje activo del modelo ondulatorio, especialmente cuando se utiliza como soporte de indagación y resolución guiada. A la vez, se identifican focos de mejora que se sitúan en continuidad conversacional, consistencia conceptual, salida por voz y gestión del alcance disciplinar. Estos hallazgos constituyen la base empírica para las conclusiones del estudio y para la formulación de proyecciones de perfeccionamiento e implementación futura en investigaciones posteriores.

## **Capítulo VI: Conclusiones**

### **6.1 Síntesis de la investigación y alcance de los hallazgos**

La presente tesis tuvo como propósito diseñar, desarrollar y validar, mediante juicio de expertos, un prototipo de chatbot educativo basado en modelos de lenguaje con recuperación aumentada, orientado a promover el aprendizaje significativo del modelo ondulatorio. El estudio se estructuró bajo un enfoque de Design and Development Research (DDR), avanzando desde la fundamentación del problema y el levantamiento de requerimientos, hacia el diseño conceptual y técnico del sistema, su implementación funcional en un entorno desplegable, y finalmente su evaluación mediante instrumentos aplicados a tres expertos en física y educación.

En consistencia con el alcance metodológico declarado, la validación se orientó a caracterizar el desempeño percibido del prototipo en tres dimensiones centrales para un recurso educativo conversacional: usabilidad, utilidad y confianza. Los resultados cuantitativos descriptivos muestran una evaluación global favorable, con un promedio general de 4.28 en escala Likert de 1 a 5, destacando valores altos en usabilidad (4.42), confianza (4.25) y utilidad (4.17). En paralelo, la evidencia cualitativa permitió identificar fortalezas pedagógicas vinculadas al andamiaje, la promoción de indagación y la proyección didáctica del recurso, así como tensiones específicas asociadas a continuidad conversacional, consistencia conceptual puntual, inteligibilidad de la salida por voz y gestión del alcance disciplinar. En conjunto, los hallazgos permiten concluir que el prototipo es funcional, valorado positivamente por especialistas, pero no exento de observaciones a aplicar en un proceso de mejora continua del prototipo, para aumentar robustez pedagógica y depuración del software en prototipos posteriores.

## **6.2 Conclusiones en relación con los objetivos de la investigación**

Conclusión 1: El diseño del prototipo logró traducir principios pedagógicos y requerimientos funcionales en una arquitectura coherente con aprendizaje significativo.

El proceso de diseño permitió articular una propuesta de tutoría conversacional que no se limita a la entrega de respuestas cerradas, sino que orienta la interacción hacia acciones cognitivas relevantes para el aprendizaje significativo, tales como explicar, justificar, modelar y reflexionar. La validación cualitativa respalda que el prototipo es percibido como un apoyo útil cuando opera como andamiaje y promueve continuidad de actividad, lo que sugiere que el enfoque de micro-diálogo y guía progresiva se expresa en el uso efectivo del sistema. Sin embargo, la evaluación también evidencia que el diseño debe reforzar mecanismos de progresión conversacional para evitar reformulaciones sin avance y para sostener una trayectoria de razonamiento más visible y acumulativa.

Conclusión 2: La construcción del prototipo alcanzó un nivel de operatividad adecuado para su uso remoto y para una evaluación experta confiable.

El prototipo, desplegado en un entorno accesible por navegador, permitió una interacción suficientemente estable para ser evaluada por expertos mediante un estudio de caso común. La dimensión de usabilidad fue la mejor evaluada globalmente (4.42), lo que sugiere que el sistema cumple condiciones mínimas y relevantes de accesibilidad operativa, claridad de interacción y adecuación del lenguaje. Aun así, el ítem relativo a fluidez y continuidad mostró mayor variabilidad entre evaluadores, lo que se alinea con comentarios sobre reinicios percibidos y necesidad de continuidad contextual. En términos metodológicos, esto confirma que la construcción fue suficiente para realizar validación, pero también evidencia un conjunto acotado de mejoras técnicas con impacto directo en experiencia pedagógica.

Conclusión 3: La validación por juicio de expertos confirma un desempeño favorable del prototipo, con tensiones críticas acotadas que orientan la mejora del sistema.

El juicio experto arrojó una evaluación positiva en utilidad (4.17) y confianza (4.25), destacándose que el sistema puede apoyar resolución guiada y proponer actividades o preguntas que fomentan aprendizaje significativo en torno al modelo de ondas. La evidencia cualitativa, sin embargo, introduce una divergencia relevante: un experto no observa ventajas

comparativas respecto de otras formas de apoyo, lo que obliga a precisar el diferencial del prototipo y su rol pedagógico específico. Asimismo, se reportan incidentes puntuales de contradicción conceptual y limitaciones percibidas al abordar temas relacionados, los cuales, aunque no invalidan la evaluación global favorable, deben considerarse prioritarios debido a su impacto potencial en confianza y credibilidad disciplinar.

#### **6.4 Limitaciones del estudio**

Las conclusiones deben interpretarse considerando limitaciones propias del diseño y del alcance definido:

- Tamaño y naturaleza de la muestra: la validación se realizó con tres expertos, lo que permite análisis descriptivo y cualitativo profundo, pero no autoriza generalizaciones poblacionales.
- Ausencia de implementación en aula o con estudiantes: no se evaluó impacto en aprendizaje, cambio conceptual o desempeño académico, sino percepción experta del prototipo bajo un escenario de uso simulado.
- Dependencia del estudio de caso y del patrón de interacción: el desempeño percibido puede variar según el tipo de consulta, la profundidad de uso y las expectativas de rol del sistema, lo que se refleja en la divergencia sobre “ventaja comparativa”.
- Evaluación en un único momento: no se midió estabilidad temporal del sistema ni evolución de la experiencia del usuario en interacciones prolongadas.

#### **6.5 Proyecciones y líneas de trabajo futuro**

Derivadas directamente de la validación, las proyecciones se organizan como recomendaciones de mejora y como posibles extensiones investigativas:

- Continuidad conversacional y progresión cognitiva: optimizar la acumulación de contexto para evitar reinicios percibidos y fortalecer rutas de preguntas que incrementen la demanda conceptual de manera visible.

- Consistencia conceptual disciplinar: reforzar control de coherencia en conceptos críticos del modelo ondulatorio y mejorar respuestas ante preguntas de transferencia, dado su impacto en confianza.
- Modalidad por voz: mejorar inteligibilidad para números, unidades y expresiones matemáticas, incorporando segmentación y priorizando claridad auditiva como condición de accesibilidad.
- Gestión del alcance del dominio: explicitar fronteras temáticas y diseñar transiciones guiadas hacia conexiones legítimas sin diluir el foco, para evitar interpretaciones de baja fiabilidad.
- Investigación con población objetivo: en un trabajo posterior, incorporar evaluación con estudiantes y/o docentes en contextos reales, midiendo aprendizaje, aceptación, carga cognitiva y efectos sobre autorregulación, además de contrastar resultados con grupos control o con uso de herramientas generalistas.

En conclusión, el prototipo validado constituye una solución educativa viable y valorada positivamente por expertos, con un desempeño global favorable en usabilidad, utilidad y confianza. Al mismo tiempo, los resultados permiten delimitar con precisión los puntos críticos cuya mejora es necesaria para fortalecer la estabilidad conceptual, la continuidad tutorial y la accesibilidad multimodal. Así, la tesis cumple su propósito dentro del alcance definido: entregar un prototipo funcional y una validación experta descriptiva que fundamenta, con evidencia, recomendaciones concretas para su perfeccionamiento y para futuras etapas de evaluación con usuarios finales en contextos educativos reales.

## Referencias Bibliográficas

- Balta, N., Japashov, N., Salibašić Glamočić, D., & Mešić, V. (2022). Development of the high school wave optics test. *Journal of Turkish Science Education*, 19(1), 306–331. <https://doi.org/10.36681/tused.2022.123>
- Carrasco-Aguilar, C., Ortiz, S., Verdejo, T., & Soto, A. (2023). Desarrollo Profesional Docente: Facilitadores y Barreras a Partir de la Carrera Docente en Chile. *Education Policy Analysis Archives*, 31. <https://doi.org/10.14507/epaa.31.7229>
- Centro de Innovación, Ministerio de Educación de Chile. (2023, mayo). Guía para docentes: Cómo usar ChatGPT para potenciar el aprendizaje activo. <https://ciudadaniadigital.mineduc.cl/wp-content/uploads/2023/05/Guia-para-Docentes-Como-usar-ChatGPT-Mineduc.pdf>
- Centro de Innovación, Ministerio de Educación de Chile. (2025, marzo). Potencia el aprendizaje: Preguntas para la reflexión y práctica de la inteligencia artificial generativa en contextos educativos. <https://ciudadaniadigital.mineduc.cl/wp-content/uploads/2025/03/Potencia-el-aprendizaje-IA-en-educacion.pdf>
- Chu, H.-E., & Treagust, D. F. (2014). Secondary students' stable and unstable optics conceptions using contextualized questions. *International Journal of Science Education*, 36(7), 1108–1136. <https://www.researchgate.net/profile/David-Treagust-2/publication/262981570>

- Costa, N., Baptista, M., Amantes, A., & Conceição, T. (2024). Validation of a learning progression for sound propagation in air. *EURASIA Journal of Mathematics, Science and Technology Education*, 20(3). <https://doi.org/10.29333/ejmste/14704>
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410–8415. <https://doi.org/10.1073/pnas.1319030111>
- Guizani, S., Mazhar, T., Shahzad, T., Ahmad, W., Bibi, A., & Hamam, H. (2025). A systematic literature review to implement large language model in higher education: Issues and solutions. *Discover Education*, 4, Article 35. <https://doi.org/10.1007/s44217-025-00424-7>
- Hrepic, Z., Zollman, D., & Rebello, N. (2010). Identifying students' mental models of sound propagation: The role of conceptual blending in understanding conceptual change. *Physical Review Special Topics – Physics Education Research*, 6(2), 020114. <https://doi.org/10.1103/PhysRevSTPER.6.020114>
- Kanyesigye, E., Uwamahoro, J., & Kemeza, I. (2022). Difficulties in understanding mechanical waves: Remediated by problem-based instruction. *Physical Review Physics Education Research*, 18(1), 010140. <https://doi.org/10.1103/PhysRevPhysEducRes.18.010140>
- Krupp, L., Steinert, S., Kiefer-Emmanouilidis, M., Avila, K. E., Lukowicz, P., Kuhn, J., Küchemann, S., & Karolus, J. (2023). Unreflected acceptance—Investigating the

negative consequences of ChatGPT-assisted problem solving in physics education (arXiv:2309.03087). arXiv. <https://doi.org/10.48550/arXiv.2309.03087>

Kurniawan, D., et al. (2023). Assessing conceptual understanding of high school students on transverse and stationary waves through Rasch analysis in Malang. *Journal of Physics: Conference Series*, 2596(1), 012060. <https://doi.org/10.1088/1742-6596/2596/1/012060>

Laun, M., & Wolff, F. (2025). Chatbots in education: Hype or help? A meta-analysis. *Learning and Individual Differences*, 119, 102646. <https://doi.org/10.1016/j.lindif.2025.102646>

Li, B., Lowell, V. L., Wang, C., & Li, X. (2024). A systematic review of the first year of publications on ChatGPT and language education: Examining research on ChatGPT's use in language learning and teaching. *Computers and Education: Artificial Intelligence*, 7, 100266. <https://doi.org/10.1016/j.caeai.2024.100266>

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). A proposal for the Dartmouth summer research project on artificial intelligence. Dartmouth College. <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>

Medina Paredes, M., & Ramírez Díaz, C. (2016). Obtención y clasificación de ideas previas sobre fenómenos sonoros: Estudio en alumnos universitarios de carreras de ciencias de la salud. *Latin American Journal of Physics Education*, 10(3). [http://www.lajpe.org/sep16/3305\\_Medina\\_2016.pdf](http://www.lajpe.org/sep16/3305_Medina_2016.pdf)

- Mešić, V., Neumann, K., Aviani, I., et al. (2019). Measuring students' conceptual understanding of wave optics: A Rasch modeling approach. *Physical Review Physics Education Research*, 15(1), 010115. <https://doi.org/10.1103/PhysRevPhysEducRes.15.010115>
- Miao, F., & Holmes, W. (2024). Guía para el uso de IA generativa en educación e investigación. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000389227>
- Neumann, A. T., Yin, Y., Sowe, S. K., Decker, S. J., & Jarke, M. (2025). An LLM-Driven Chatbot in Higher Education for Databases and Information Systems. *IEEE Transactions on Education*, 68(1), 103–116. <https://doi.org/10.1109/TE.2024.3467912>
- OECD (2023). *OECD Digital Education Outlook 2023: Towards an Effective Digital Education Ecosystem*, OECD Publishing, Paris, <https://doi.org/10.1787/c74f03de-en>.
- OECD (2025). *Results from TALIS 2024: The state of teaching*. OECD Publishing. <https://doi.org/10.1787/90df6235-en>
- Planinić, M., et al. (2024). Effect of an inquiry-based teaching sequence on secondary school students' understanding of wave optics *Physical Review Physics Education Research*, 20(1), 010156. <https://doi.org/10.1103/PhysRevPhysEducRes.20.010156>
- Porayska-Pomsta, K., Holmes, W., & Nemorin, S. (2024). The ethics of AI in education (arXiv:2406.11842). arXiv. <https://doi.org/10.48550/arXiv.2406.11842>

- Richey, R. C., & Klein, J. D. (2007). Design and development research. <https://pmm.uinsu.ac.id/wp-content/uploads/2022/02/D-D.pdf>
- Rico, A., Ruiz González, A., Azula, O., & Guisasola Aranzábal, J. (2021). Dificultades de aprendizaje del modelo de sonido: una revisión de la literatura Enseñanza de las Ciencias, 39(2), 5–23. <https://doi.org/10.5565/rev/ensciencias.3217>
- Stryker, C., & Holdsworth, J. (n.d.). What is NLP (natural language processing)? IBM. Retrieved May 15, 2025, from <https://www.ibm.com/think/topics/natural-language-processing>
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433–460. <https://courses.cs.umbc.edu/471/papers/turing.pdf>
- Villarroel, V., Pérez, C., Rojas-Barahona, C. A., & García, R. (2021). Educación remota en contexto de pandemia: caracterización del proceso educativo en las universidades chilenas *Formación Universitaria*, 14(6), 65–76. <https://doi.org/10.4067/S0718-50062021000600065>
- Wang, J., & Fan, W. (2025). The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: insights from a meta-analysis. *Humanities and Social Sciences Communications*, 12, 621. <https://doi.org/10.1057/s41599-025-04787-y>
- Wang, P. (2019). On defining artificial intelligence. *Journal of Artificial General Intelligence*, 10(2), 1–37. <https://doi.org/10.2478/jagi-2019-0002>

Wieman, C., & Perkins, K. (2005). Transforming physics education. *Physics Today*, 58(11), 36–41.

<https://physicstoday.aip.org/features/transforming-physics-education>

Zhao, W. X., et al. (2025). A survey of large language models.

<https://doi.org/10.48550/arXiv.2303.18223>