

IA generativa en la enseñanza de la Física: cómo integrarla sin reemplazar el razonamiento.



Bernardo Pombo ^{1,2}

¹ Instituto de Profesores Artigas, Consejo de Formación en Educación, Montevideo, Uruguay.

² Liceo Nº 1 Eduardo Fabini, Dirección General de Educación Secundaria, Minas, Uruguay.

E-mail: bpombo@gmail.uy

(Recibido el 10/11/2025, aceptado el 9/4/2026)

Resumen

Este artículo analiza cómo integrar la inteligencia artificial (IA) generativa en la enseñanza de la Física de educación secundaria sin desplazar el razonamiento ni la contrastación empírica. Se presenta una revisión bibliográfica narrativa focalizada de investigaciones recientes (2023-2025) sobre desempeño de modelos de lenguaje en instrumentos diagnósticos de mecánica, percepciones estudiantiles, creencias y disposiciones docentes, uso de simuladores como PhET e inserción de *ChatGPT* en secuencias de enseñanza de la Física y contextos *STEM* afines. A partir de esta síntesis se identifican oportunidades para ampliar explicaciones, generar ejemplos y personalizar la interacción con los contenidos, junto con riesgos de comprensiones superficiales, delegación acrítica y tensiones respecto del rol docente. Sobre esa base se formulan cuatro lineamientos para el diseño didáctico con IA en Física (IA como generador de propuestas, verificación sistemática y registro del proceso, contratos de uso explícitos y desarrollo profesional centrado en la práctica) y se ejemplifican estrategias de aula: rutinas de predicción, simulación y explicación, tutoría socrática con bancos tipo *FCI*, análisis crítico de salidas de IA y laboratorios virtuales con *prompts* de indagación. Finalmente, se propone una rúbrica de evaluación centrada en la modelización, la verificación con datos y la trazabilidad del razonamiento, y se discuten condiciones de uso responsable en cursos de Física de ciclo básico y bachillerato en Uruguay.

Palabras clave: inteligencia artificial generativa, enseñanza de la Física, modelos de lenguaje, simulaciones interactivas, evaluación formativa, educación secundaria.

Abstract

This article examines how to integrate generative artificial intelligence (AI) into secondary-school physics teaching without displacing reasoning or empirical validation. It presents a focused narrative literature review of recent studies (2023–2025) on the performance of language models on mechanics diagnostic instruments, student perceptions, teacher beliefs and dispositions, the use of simulators such as PhET, and the incorporation of ChatGPT into physics teaching sequences and related STEM contexts. From this synthesis, it identifies opportunities to expand explanations, generate examples, and personalise interaction with content, alongside risks of superficial understanding, uncritical delegation, and tensions regarding the teacher's role. On this basis, four guidelines are proposed for didactical design with AI in physics (AI as a generator of proposals, systematic verification and process documentation, explicit usage contracts, and practice-centred professional development), and classroom strategies are illustrated: prediction–simulation–explanation routines, Socratic tutoring with FCI-type item banks, critical analysis of AI outputs, and virtual laboratories with inquiry-oriented prompts. Finally, an assessment rubric is proposed that focuses on modelling, verification with data, and the traceability of reasoning, and conditions for responsible use in lower- and upper-secondary physics courses in Uruguay are discussed.

Keywords: generative artificial intelligence, physics teaching, language models, interactive simulations, formative assessment, secondary education.

1. Introducción

¿Qué sentido cobra enseñar Física cuando la IA resuelve ejercicios en segundos?

En los últimos dos años, la IA generativa pasó de ser ciencia ficción a herramienta cotidiana: traduce y produce textos, explica procedimientos, redacta informes. Enseñar Física en este contexto exige rejerarquizar el razonamiento, los modelos y la contrastación empírica por encima de la aplicación mecánica de algoritmos de resolución de ejercicios. Dado que su presencia en la vida cotidiana ya está instalada, el desafío didáctico consiste en definir el lugar que ocupará en el aula y las condiciones de uso, de manera que el docente pueda acompañar y supervisar cómo la emplean los estudiantes.

La literatura reciente ofrece una base útil para abordar ese desafío y para un nuevo encuadre didáctico. Por un lado, reúne estudios que analizan el rendimiento de modelos de lenguaje en inventarios conceptuales de mecánica como el Force Concept Inventory (FCI) [1,2] y trabajos que exploran percepciones estudiantiles y usos de ChatGPT como tutor en cursos de Física y disciplinas afines [3,4]. Por otro lado, incluye revisiones sobre tecnologías para la enseñanza de la Física, en particular simuladores y laboratorios virtuales [5,6], así como estudios sobre creencias y disposiciones docentes frente a la IA generativa [7,8] y experiencias de integración de ChatGPT en temas específicos de Física, como el movimiento de proyectiles, la segunda ley de Newton o la termodinámica de ingeniería [9–11].

En paralelo, la literatura sobre tecnología y aprendizaje de la Física subraya que estos recursos producen mejores resultados cuando se insertan en entornos de aprendizaje activos, con simulaciones, tareas de indagación guiada y retroalimentación formativa, y cuando se articulan con experiencias experimentales reales y/o virtuales [5,6]. Las revisiones sobre IA en educación científica llegan a conclusiones similares: identifican potencial para la personalización, la retroalimentación y la colaboración, pero también advierten la necesidad de marcos éticos claros y de desarrollo profesional docente para una integración contextualizada y coherente [4,5].

En este artículo se utilizan de forma diferenciada los términos *ChatGPT* y *GPT-4*. Se emplea *ChatGPT* para referirse al chatbot de IA desarrollado por OpenAI. Se utiliza *GPT-4* cuando se describe el modelo concreto evaluado en los estudios empíricos citados, por ejemplo, en aplicaciones del FCI o en tareas de resolución de problemas. De este modo se respeta la denominación de las publicaciones originales al presentar resultados de desempeño del modelo y se

utiliza *ChatGPT* para los ejemplos de uso educativo en el aula.

Este trabajo, entonces, busca traducir esa evidencia en acciones didácticas específicas para docentes de secundaria que sienten sobrecarga o recelo ante la IA y preocupación por la vigencia de su rol en el aula. Las propuestas que se presentan a continuación están pensadas principalmente para cursos de Física de ciclo básico y bachillerato en Uruguay, aunque pueden adaptarse a otros contextos educativos.

2. Metodología. Revisión bibliográfica narrativa focalizada

Este trabajo adopta una revisión narrativa focalizada orientada por la siguiente pregunta: ¿Qué usos y condiciones de implementación de modelos de IA generativa en la enseñanza de la Física, en educación secundaria y en contextos afines plausiblemente extrapolables a ella, se han documentado recientemente, y qué lineamientos de diseño didáctico, estrategias de aula y criterios de evaluación se pueden derivar para el aula de Física?

La búsqueda principal se realizó en bases de datos académicas y repositorios abiertos mediante la cadena: (*"physics education" OR "physics teaching" OR "physics classroom"*) AND (*"generative AI" OR "large language model*" OR ChatGPT OR "AI chatbot" OR "openAI"*) AND (*"secondary school" OR "high school" OR "upper secondary"*) AND (*"teaching strategies" OR "classroom activities"*). Esta búsqueda inicial arrojó 89 artículos arbitrados publicados entre 2023 y 2025. Se aplicaron como criterios de inclusión: a) foco en enseñanza de Física o contenidos de Física dentro de cursos STEM, b) presencia explícita de modelos de lenguaje o IA generativa en el diseño didáctico, c) contexto de educación secundaria o, cuando se trató de educación superior o de otros niveles, posibilidades claras de extrapolación al aula de Física escolar. Se excluyeron trabajos centrados en programación, robótica u otras áreas STEM sin componentes de Física y aquellos donde la IA aparecía solo como herramienta administrativa.

A partir de estos criterios se seleccionó un conjunto de estudios que cubren cuatro ejes: a) desempeño de modelos de lenguaje en evaluaciones conceptuales de mecánica, como el FCI [1,2]; b) percepciones estudiantiles al usar ChatGPT como tutor o apoyo conversacional en Física y contextos afines [3,4]; c) revisiones sobre tecnologías para la enseñanza de la Física, en particular simuladores como PhET [5,6]; d) creencias y disposiciones docentes ante la IA generativa [7,8] y su uso en la formación inicial de profesorado de Física [12]. Se incorporaron además

estudios cuasi experimentales con ChatGPT en Física de secundaria [9,11] y trabajos en otros niveles educativos donde los hallazgos resultan plausiblemente extrapolables a secundaria, como la integración de ChatGPT en educación STEM en la primera infancia [13] y en cursos de termodinámica en ingeniería [10].

Dentro de este conjunto se incorporaron también investigaciones que analizan con detalle la interacción entre estudiantes y asistentes de IA en el diseño de indagaciones abiertas de Física [14] y estudios que evalúan de manera sistemática el desempeño de distintos modelos de lenguaje ante concepciones alternativas específicas en mecánica newtoniana en secundaria [15].

Sobre esta base se realizó una lectura comparativa y se codificaron patrones comunes en torno a: rol de la IA (tutor, generador de propuestas, apoyo al diseño docente), tipo de evidencia (mejora de rendimiento, cambio conceptual, percepciones, barreras de implementación) y tensiones éticas y condiciones de uso responsable señaladas en la literatura. De la síntesis de estos patrones emergen los lineamientos de diseño didáctico presentados en la sección 4, así como las propuestas de actividades y de evaluación de las secciones 5 y 6.

3. Síntesis de hallazgos relevantes para situar la enseñanza

3.1. Percepciones estudiantiles de IA como tutor

Ding et al. [3] reportan que estudiantes valoran a ChatGPT como tutor conversacional para pedir aclaraciones, ejemplos y pasos intermedios. Esa valoración mejora cuando el curso tiene un encuadre didáctico explícito: reglas de uso, para qué sirve, cuándo procede, qué se espera de la calidad de las respuestas (claridad de supuestos, consistencia de unidades, fuentes citables), cómo manejar sesgos y cómo verificar de forma independiente lo que la IA propone. Este encuadre puede apoyarse en rúbricas que hagan visibles esos criterios y orienten tanto la producción de los estudiantes como la retroalimentación docente.

Por ejemplo, en una clase de Física sobre cinemática, una regla podría ser: “Puedo pedir a ChatGPT que explique qué es aceleración centrípeta y que me dé tres ejemplos, pero debo verificar esa explicación con una simulación PhET antes de usarla en mi tarea”. Este tipo de acuerdos explícitos transforman la percepción estudiantil: la IA pasa de ser un recurso

prohibido, y por tanto tentador, a una herramienta con reglas y responsabilidades.

3.2. Capacidad de los LLM¹ en tópicos introductorios

West [2] explica que GPT-4 contesta correctamente 22 de 23 ítems de una versión modificada del FCI, con un rendimiento equiparable al de estudiantes de alto desempeño en cursos introductorios de mecánica. Resultados similares aparecen en la evaluación de GPT-4 con el FCI reportada por Aldazharova et al. [1]: el modelo logra muy buenos resultados en ítems asociados, por ejemplo, a la tercera ley de Newton, pero presenta falsos positivos y falsos negativos cuando debe interpretar figuras, razonar sobre situaciones espaciales complejas o tratar fuerzas de rozamiento. Se consideran falsos positivos los casos en que la respuesta final es correcta pero la justificación es pobre o parcialmente errónea, y falsos negativos aquellos en que el razonamiento intermedio parece plausible pero la respuesta final resulta incorrecta [1].

Que un modelo de lenguaje acierte en preguntas de Física no implica que “entienda” los fenómenos. Más bien reconoce patrones estadísticos en los enunciados y en las soluciones frecuentes, lo que explica por qué puede obtener respuestas correctas sin explicitar supuestos clave, como la ausencia de rozamiento o la idealización de ciertos cuerpos. Por eso el desafío pedagógico no es evaluar respuestas, sino procesos: ¿el estudiante identifica y justifica supuestos?, ¿elige modelos apropiados?, ¿verifica con datos de simulaciones o mediciones? En este contexto, las tareas centradas en respuesta cerrada y explicación breve pueden ser fácilmente resueltas por la IA; conviene rediseñarlas hacia la comparación de modelos, la contrastación con datos, la explicación de límites de aplicabilidad y la crítica de errores sutiles de la propia IA [1,2].

Además de los estudios que caracterizan el desempeño de los modelos de lenguaje en inventarios como el FCI, existen trabajos que se centran en concepciones alternativas puntuales y en el potencial de la IA como tutor para abordarlas. Kahaleh y Lopez [15] analizan cómo ChatGPT, Gemini y Copilot responden a una idea errónea muy extendida en mecánica: la creencia de que, para que un ascensor se mueva hacia arriba, la fuerza del cable debe ser mayor que el peso incluso cuando se desplaza con velocidad constante. A partir de diálogos entre 32 docentes y los modelos, construyen un marco de evaluación con tres criterios (exactitud, personalización y eficacia pedagógica) que

¹ modelos de lenguaje de gran tamaño (Large Language Models, LLM)

luego aplican en un segundo estudio con 25 docentes. Los resultados muestran una valoración alta de ChatGPT en exactitud conceptual y en uso de estrategias didácticas (preguntas, ejemplos y recapitulaciones), así como un desempeño más irregular en personalización y en el uso de ayudas visuales como diagramas o simulaciones.

3.3. Tecnologías de apoyo al razonamiento en Física

Las revisiones sobre PhET señalan que las simulaciones potencian la comprensión conceptual, la indagación y el compromiso, pero su impacto depende de la integración pedagógica y del desarrollo profesional docente [6]. Recomendaciones típicas incluyen diseñar guías de indagación, promover discusión entre pares y articular con experiencias experimentales reales o virtuales.

Un riesgo frecuente en este terreno es confundir la simulación con la experiencia experimental real. PhET resulta especialmente útil para explorar preguntas del tipo “Si cambio esto, ¿qué sucede?”, en un ambiente controlado y con un número acotado de variables. Esa exploración gana valor cuando posteriormente los estudiantes realizan mediciones en un contexto físico, aunque sea con materiales simples. Ramaila [6] documenta de manera consistente que la combinación de PhET, preguntas bien diseñadas y actividades experimentales con mediciones reales, ya sea en el laboratorio o con materiales simples en el aula, se asocia con aprendizajes más sólidos, mientras que un uso aislado de la simulación tiende a derivar en exploraciones sin anclaje empírico.

De forma análoga, la evidencia reciente sobre IA generativa en educación científica muestra que herramientas como ChatGPT pueden contribuir al aprendizaje de contenidos específicos de Física cuando se integran en secuencias de trabajo estructuradas. En secundaria, por ejemplo, se observan mejoras en el rendimiento en temas de movimiento de proyectiles al combinar una estrategia de aprendizaje autorregulado con GPT-4, frente a la enseñanza expositiva tradicional [11], y en la comprensión de la segunda ley de Newton en estudiantes de bachillerato en Emiratos Árabes Unidos [9]. En el nivel universitario, la integración de ChatGPT como recurso de apoyo en un curso de termodinámica de ingeniería se asocia con incrementos en comprensión conceptual y reducción de algunas concepciones alternativas, en particular sobre entropía y energía interna, aunque persisten dificultades en tareas cuantitativas y de razonamiento avanzado [10].

Al mismo tiempo, investigaciones que analizan el desempeño de GPT-4 en instrumentos diagnósticos como el FCI evidencian resultados heterogéneos: el modelo obtiene altos niveles de acierto en diversos

ítems, particularmente en aquellos relacionados con la tercera ley de Newton, pero presenta dificultades persistentes en problemas que requieren interpretación de figuras, razonamiento espacial o análisis de situaciones con fricción. Asimismo, en ciertas respuestas reproduce explicaciones próximas a concepciones intuitivas vinculadas al impulso y la cantidad de movimiento [1].

Estos resultados refuerzan la necesidad de tratar las respuestas generadas por IA como hipótesis de trabajo que requieren contraste, y no como fuente automática de validación conceptual.

Además, se han identificado efectos problemáticos cuando el uso de la IA se basa en la aceptación acrítica de sus soluciones. Investigaciones recientes advierten que la resolución de problemas de Física asistida por ChatGPT puede favorecer comprensiones superficiales y dependencia de la herramienta si no se acompaña de consignas claras de análisis y verificación de las respuestas de la IA mediante simulaciones, cálculos o mediciones [16]. Estos hallazgos convergen con revisiones más amplias sobre IA en educación científica, que destacan tanto oportunidades para la personalización y la retroalimentación inmediata como la necesidad de marcos de uso que preserven el desarrollo del pensamiento crítico y la resolución de problemas [5].

En conjunto, las simulaciones interactivas y la IA generativa pueden entenderse como tecnologías de apoyo al razonamiento en Física que demandan un diseño didáctico cuidadoso. La literatura sugiere que su aporte resulta más consistente cuando se integran en ciclos de predicción, experimentación, análisis y discusión: *PhET* como recurso para explorar hipótesis y visualizar relaciones entre variables antes o después de actividades experimentales, y *ChatGPT* como andamiaje para formular predicciones, explicitar supuestos, comparar argumentaciones y revisar concepciones alternativas. La responsabilidad docente se orienta, en este escenario, a planificar secuencias en las que estas herramientas se articulan con actividades de modelización, contrastación empírica y explicación, de manera que acompañen el trabajo intelectual de los estudiantes en lugar de reemplazarlo.

3.4. Indagación abierta y cognición distribuida con IA

El uso de IA generativa en actividades de indagación no se limita a resolver problemas cerrados de Física, también puede intervenir como apoyo en la fase más delicada del trabajo en el aula: el diseño de las investigaciones. Min, Lee y Jho [14] desarrollan un asistente de indagación basado en GPT-4 que acompaña a estudiantes de educación media en la formulación de preguntas, la identificación de variables y la

construcción de procedimientos experimentales a partir de problemas abiertos inspirados en el International Young Physicists' Tournament. El análisis de 93 intercambios alumno-IA muestra que las preguntas se organizan en cinco tipos de contenido (variables, materiales, proceso experimental, análisis/resultados y conocimiento conceptual) y que la IA cumple un papel distinto según el tipo de pregunta planteada, ya sea como fuente de información o como agente que revisa y mejora planes de indagación propuestos por el propio alumnado.

Este estudio se apoya en el marco de la cognición distribuida y concibe al asistente de IA como un "agente cognitivo externo" que amplía la capacidad de los estudiantes para coordinar variables, anticipar dificultades metodológicas y refinar sus diseños, mientras el docente sigue a cargo de estructurar la tarea, interpretar las sugerencias de la IA y alinearlas con los objetivos formativos de la asignatura [14].

Esta perspectiva resulta especialmente relevante para la enseñanza de la Física, donde la planificación de actividades experimentales controladas y la explicitación de supuestos son componentes centrales del trabajo experimental y de modelización en el aula. En la sección 5 se retoma esta tipología de preguntas y la noción de cognición distribuida para diseñar actividades de indagación apoyadas en IA generativa en cursos de Física [14].

3.5. Oportunidades y riesgos del uso de IA en la enseñanza de la Física

Los trabajos revisados coinciden en que la IA generativa abre oportunidades para ampliar las fuentes de explicación, ofrecer ejemplos adicionales y personalizar la interacción con el contenido de Física, tanto en secundaria como en educación universitaria [3,4,10,11]. En estudios con grupos control y experimental se reportan mejoras en el rendimiento cuando ChatGPT se integra en secuencias de trabajo estructuradas, en comparación con enfoques expositivos tradicionales, especialmente en temas que suelen presentar dificultades persistentes, como el movimiento de proyectiles o la segunda ley de Newton [9,11].

Al mismo tiempo, varios autores advierten que estos beneficios no son automáticos y que la IA puede promover comprensiones superficiales si los estudiantes se limitan a aceptar sus respuestas sin analizar los supuestos físicos implicados [16]. Evaluaciones del desempeño de modelos como GPT-4 en instrumentos diagnósticos de mecánica (FCI) muestran una combinación de aciertos elevados en algunos ítems y errores sistemáticos en otros,

reflejando concepciones alternativas similares a las del propio alumnado [1,2]. Estas limitaciones refuerzan la necesidad de tratar las respuestas generadas por IA como hipótesis de trabajo que requieren contraste con simulaciones, cálculos y mediciones, más que como soluciones definitivas.

Finalmente, los estudios sobre creencias y disposiciones docentes destacan que la integración de IA generativa se percibe simultáneamente como oportunidad y como fuente de inquietud respecto del rol profesional: emergen preocupaciones por la calidad y fiabilidad de las respuestas, por el posible impacto en el esfuerzo cognitivo de los estudiantes y, en algunos casos, temores a que la herramienta pueda desdibujar la función del profesorado en el aula [7,8]. Estas tensiones refuerzan la necesidad de contar con criterios pedagógicos claros, ejemplos de actividades y espacios de reflexión entre docentes que ayuden a decidir en qué momentos, con qué fines y con qué condiciones incorporar la IA en las clases de Física.

3.6. Creencias y disposiciones docentes

La investigación de Cabellos, de Aldama y Pozo [7] reúne evidencia empírica de que las creencias pedagógicas y las disposiciones hacia la tecnología influyen de manera decisiva en cómo el cuerpo docente adopta la IA generativa. El estudio muestra que quienes sostienen enfoques más constructivistas y han experimentado con estas herramientas tienden a percibir más posibilidades para apoyar la enseñanza y el aprendizaje, mientras que quienes se identifican con enfoques más reproductivos, o tienen poca experiencia de uso, enfatizan sobre todo los riesgos, en particular la pérdida de esfuerzo cognitivo y los problemas de integridad académica. A partir de estos resultados, las autoras y el autor subrayan la necesidad de marcos éticos explícitos, espacios de trabajo colaborativo entre docentes y un uso de la IA sometido a criterios de rigor y transparencia. Esta evidencia puede orientar también el desarrollo profesional en educación secundaria. Apunta a la conveniencia de crear espacios específicos para el cuerpo docente de Física, en los que (a) se discutan creencias y expectativas respecto de la IA, (b) se ensayen propuestas de actividades de aula con estas herramientas, con apoyo y retroalimentación de colegas y formadores, y (c) se acuerden criterios compartidos para un uso responsable de la IA en sus cursos, incluyendo la elaboración y el intercambio de rúbricas que hagan explícitos esos criterios en la evaluación.

Estos hallazgos se reflejan también en estudios específicos con docentes de Física en ejercicio. Wattanakasiwich, Kaewkhong y Katwibun [8] describen cómo los docentes de Física se distribuyen en distintas

fases de adopción de la IA generativa según la teoría de difusión de innovaciones, identificando barreras como la falta de conocimiento técnico, las limitaciones lingüísticas de los modelos y la preocupación por el impacto en el pensamiento analítico de los estudiantes. En programas de formación inicial de docentes de Física, Guerrero-Zambrano et al. [12] muestran que el uso guiado de ChatGPT para diseñar actividades lúdicas favorece tanto la familiaridad con la herramienta como la calidad de las propuestas didácticas. Estos resultados refuerzan la idea de que la integración de IA en la enseñanza de la Física depende menos de la herramienta en sí y más de marcos éticos explícitos, espacios de trabajo colaborativo y oportunidades de experimentar con actividades concretas en contextos de acompañamiento profesional.

4. Lineamientos derivados de la revisión sobre IA y enseñanza de la Física

A partir de la revisión bibliográfica narrativa focalizada descrita en la sección 2, y de la comparación de resultados en contextos de Física en enseñanza secundaria, formación docente y entornos STEM afines, emergen cuatro lineamientos operativos para orientar el diseño didáctico con IA en Física. Estos lineamientos recogen patrones consistentes en estudios sobre desempeño de modelos de lenguaje, impacto en el aprendizaje conceptual, percepciones estudiantiles y creencias docentes [1,3–6,8–10,2]:

- 1) **IA como generador de propuestas, no como autoridad.** La IA no es una fuente incuestionable, sino un generador de hipótesis. Preguntar siempre: ¿qué supuestos adopta? ¿Qué variantes del problema considera? ¿En qué condiciones su solución sería inválida? West [2] documenta que ChatGPT responde bien, pero sin explicitar supuestos clave, por eso esta interrogación es imprescindible. Vincular respuestas con diagramas, gráficos, simulaciones y datos.
- 2) **Verificar siempre, documentar el proceso.** La IA genera respuestas rápido, pero la rapidez no es validez. Exigir siempre verificación externa: simulación, cálculo propio, o medición breve. Igual de importante es que el estudiante documente el proceso (*prompts* usados, respuestas clave, comparación con simulación) para que la retroalimentación sea trazable y la evaluación pueda valorar razonamiento, no solo producto.
- 3) **Contrato de uso de IA.** Acordar con la clase cuándo y cómo se utiliza, con reglas de citación de *prompts* y verificación. Ding et al. [3] subrayan que la percepción positiva aumenta cuando hay pautas claras y acompañamiento docente.

- 4) **Formación centrada en la práctica docente.** El impacto educativo de la IA depende de la capacidad del cuerpo docente para integrarla en actividades de Física con un propósito didáctico claro. Las revisiones recomiendan un desarrollo profesional específico que parta de la propia experiencia de aula: rediseño de actividades y evaluaciones, integración de IA con simulaciones y actividades experimentales, y elaboración de rúbricas que valoren razonamiento, uso de modelos y verificación con datos. Este tipo de formación, ligada al trabajo cotidiano, permite que la IA se incorpore como apoyo al aprendizaje y a la evaluación auténtica, en lugar de quedar como un recurso externo o puramente técnico.

5. Sugerencia de estrategias y secuencias didácticas

Las estrategias que se presentan a continuación ilustran algunas posibilidades de trabajo con IA en Física, pero no agotan el repertorio posible ni pretenden cubrir la diversidad de enfoques que se está desarrollando en la práctica docente.

5.1. Rutina “Predecir → Simular → Explicar”

Objetivo: Analizar la coherencia entre las predicciones de un modelo, los resultados de una simulación y los datos obtenidos.

Pasos:

- 1) **Generación asistida:** pedir al LLM tres predicciones para una situación concreta. Ej.: “carrito sobre plano horizontal con coeficiente de rozamiento cinético bajo”. Los estudiantes escriben las tres, las leen en voz alta y explican qué supuestos adopta cada una.
- 2) **Contrastación en PhET:** ejecutar la simulación con parámetros controlados. Los estudiantes registran qué predicción resulta compatible con los datos.
- 3) **Explícito sobre límites:** redactar un párrafo que indique qué cambia si se modifica un supuesto. Estas tres fases capitalizan la capacidad explicativa de la IA y, al mismo tiempo, llevan a los estudiantes a contrastar lo obtenido con evidencia y a explicitar el dominio de validez del modelo, en línea con la literatura sobre ventajas pedagógicas de las simulaciones y la indagación guiada.
- 4) **Indicador formativo:** rúbrica que puntúe la identificación de variables relevantes, elección de modelo, justificación y control de supuestos, coherente con la evidencia sobre los efectos de una

retroalimentación específica y oportuna identificados por Mahligawati et al. [5].

Este enfoque no sustituye la práctica de resolución numérica de problemas, que sigue siendo necesaria para desarrollar fluidez en el uso de modelos y en el manejo de magnitudes; más bien la sitúa en secuencias donde el estudiantado alterna momentos de cálculo sin asistencia de IA con momentos de análisis cualitativo, contraste con simulaciones y revisión crítica de respuestas generadas por el modelo.

5.2. “ChatGPT como tutor socrático” con banco tipo FCI

Objetivo: Analizar concepciones alternativas sobre una idea clave de mecánica (por ejemplo, la tercera ley de Newton) mediante un diálogo guiado con *ChatGPT* y la resolución de ítems tipo *FCI*.

Procedimiento:

- 1) **Selección de foco conceptual.** El docente elige uno o dos ítems de un banco tipo *FCI* o una pregunta abierta que apunte a una concepción alternativa frecuente (por ejemplo, que la fuerza de acción es “mayor” que la de reacción cuando un camión choca con un automóvil).
- 2) **Diseño del prompt.** El docente construye un *prompt* para *ChatGPT* en clave de tutor socrático, por ejemplo: “Actúa como tutor de Física. No des la respuesta final de inmediato. Haz preguntas de seguimiento que me ayuden a pensar sobre la tercera ley de Newton en la siguiente situación: [descripción del ítem]. Al final te pediré que revises mi explicación y mi diagrama de fuerzas”.
- 3) **Diálogo guiado con ChatGPT.** En parejas o tríos, los estudiantes interactúan con el tutor conversacional respondiendo a las preguntas, elaborando una explicación propia y dibujando el diagrama de fuerzas correspondiente. Registran el intercambio relevante (preguntas clave del tutor, respuestas que los hicieron cambiar de idea, etc.) en una hoja de trabajo.
- 4) **Verificación con instrumento externo.** Una vez que el grupo alcanza una explicación “final”, responde el ítem *FCI* original o realiza una exploración breve en *PhET* “Fuerzas y movimiento” con parámetros acordados, y anota si los resultados respaldan o contradicen la explicación construida con el tutor.
- 5) **Puesta en común y síntesis docente.** En plenario, se comparan explicaciones y diagramas, se identifican patrones de concepciones alternativas y se explicita el modelo aceptado, recuperando los errores más frecuentes y discutiendo qué aportó y qué no aportó el tutor conversacional para superarlos.

En términos evaluativos, puede resultar útil una rúbrica que ponga el foco en tres aspectos: qué tan claramente formulan la tercera ley como interacción entre cuerpos, la corrección del diagrama de fuerzas y la capacidad de detectar y corregir errores presentes en sus propias respuestas o en las propuestas de *ChatGPT*. De este modo, la experiencia retoma el potencial del tutor conversacional descrito por Ding et al. [3] y Guo et al. [4], pero mantiene la conducción en manos del docente gracias a la verificación con instrumentos externos y a la discusión colectiva.

5.3. Prueba crítica de soluciones de IA

Objetivo: Identificar errores conceptuales o incoherencias en respuestas de *ChatGPT* sobre un tema de mecánica y corregirlos utilizando modelos y datos trabajados en clase.

Procedimiento: el docente selecciona un ítem conceptual tipo *FCI* sobre fuerzas o movimiento, idealmente uno donde el modelo de IA utilizado haya mostrado errores sutiles, por ejemplo, una situación de interpretación de diagramas [1]. Se pide a la IA que responda el ítem “explicando el razonamiento paso a paso para estudiantes de secundaria”. El texto se edita mínimamente para ocultar la fuente y se entrega a la clase como “solución propuesta”. La consigna para los estudiantes puede estructurarse en tres pasos:

- 1) **Detección:** subrayar al menos dos afirmaciones que consideren dudosas o insuficientemente justificadas.
- 2) **Contraste con modelos y datos:** usar un diagrama de fuerzas, una simulación *PhET* o un experimento sencillo para verificar si esas afirmaciones son compatibles con los modelos de la Física en secundaria.
- 3) **Reescritura:** elaborar una versión corregida de la explicación, indicando explícitamente los supuestos y el límite de validez del modelo utilizado.

West [2] y Aldazharova et al. [1] documentan explicaciones mayormente correctas de *ChatGPT* acompañadas de errores conceptuales finos o de supuestos no explicitados, lo que convierte este tipo de actividad en un contexto auténtico para practicar lectura crítica de textos científicos y discusión entre pares. La IA funciona aquí como fuente de error “verosímil”, que obliga a justificar los modelos en juego en lugar de aceptar pasivamente una respuesta bien redactada.

En esta modalidad resulta especialmente útil trabajar con ejemplos concretos de salida de IA que combinan un registro técnico aceptable con un modelo físico erróneo. El siguiente ítem conceptual, inspirado en situaciones de tipo *FCI*, ilustra el tipo de actividad que se propone:

Una bola se mueve dentro de un canal horizontal sin rozamiento. En cierto tramo deja de actuar cualquier empuje externo y la bola continúa su movimiento dentro del canal.

- 1) ¿Qué fuerzas actúan sobre la bola mientras se mueve en el tramo recto del canal, una vez que ha dejado de ser empujada?
- 2) La siguiente explicación fue generada por un modelo de IA para responder a esta pregunta:

“La fuerza en la dirección del movimiento sigue actuando porque la inercia de la bola la impulsa hacia adelante. Además, actúa la fuerza de gravedad hacia abajo.”

En el aula, se pide a los estudiantes que trabajen con esta salida de IA deliberadamente “creíble pero defectuosa”:

- 1) Identificar qué elementos de la explicación resultan razonables (por ejemplo, la mención del peso y la referencia a la trayectoria rectilínea).
- 2) Localizar el error conceptual: la respuesta aborda la inercia como si fuera una fuerza que “empuja” al cuerpo y sugiere que el “golpe” o empuje inicial deja una fuerza persistente en la dirección del movimiento.
- 3) Reescribir una explicación físicamente correcta, apoyada en un diagrama de fuerzas y en la segunda ley de Newton, indicando que en el tramo descrito la única fuerza que actúa sobre la bola es el peso (y la fuerza normal si se explicita la interacción con el canal), mientras que el movimiento uniforme se explica por ausencia de fuerza neta horizontal y no por una “fuerza de la inercia”.

Este tipo de tarea permite explicitar una concepción cercana a la física aristotélica que también aparece en respuestas de estudiantes y en algunas salidas de modelos de lenguaje, y convierte la respuesta de IA en un recurso para contrastar modelos: el foco no está en “corregir a la IA”, sino en usar una explicación verosímil pero equivocada como ocasión para afinar los criterios con que se decide qué es una buena explicación en Física.

5.4. Laboratorios virtuales con prompts de indagación

Objetivo: pasar del “ver” al “explicar” en actividades de laboratorio virtual y en experiencias sencillas de aula.

Diseño general: cada guía de laboratorio virtual combina tres elementos:

- 1) un *prompt* inicial para la IA que genere hipótesis cuantitativas y un plan preliminar de indagación;
- 2) la ejecución de la simulación con registro de datos por parte de los estudiantes;
- 3) una fase final de reinterpretación y comunicación de resultados sin apoyo de IA, que puede extenderse a una experiencia real de aula.

Por ejemplo, en un laboratorio sobre fuerzas y movimiento con rozamiento, usando una simulación tipo *Fuerzas y Movimiento: Fundamentos* o un experimento sencillo con un bloque sobre una mesa y un dinamómetro, la guía puede incluir un *prompt* modelo como:

“Explica dos hipótesis cuantitativas diferentes sobre cómo cambia la fuerza de rozamiento y la aceleración de un bloque cuando se aumenta la fuerza de arrastre, manteniendo constantes la masa y la superficie de contacto. Propón un plan de prueba usando una simulación tipo *PhET* o un experimento con dinamómetro.”

Los estudiantes ajustan el *prompt*, obtienen una propuesta de la IA y la discuten en pequeños grupos, identificando qué magnitudes deben medirse (fuerza aplicada, fuerza de rozamiento, aceleración), qué variables conviene mantener constantes (masa del bloque, tipo de superficie) y qué gráficos serían útiles (fuerza–tiempo, velocidad–tiempo, aceleración–fuerza neta). A continuación, trabajan en la simulación *Fuerzas y Movimiento: Fundamentos* o realizan el montaje real con un bloque y un dinamómetro, registran los datos y elaboran tablas y gráficos.

En la fase final se cierra el acceso a la IA y se pide:

- 1) comparar los datos con las hipótesis iniciales,
- 2) explicar posibles discrepancias (por ejemplo, diferencias entre rozamiento estático y cinético o errores de medida),
- 3) redactar un breve informe donde se argumente qué modelo describe mejor los resultados.

La IA interviene en el diseño inicial de la indagación, al servicio de la generación de hipótesis y planes de trabajo, mientras que la interpretación y la comunicación de resultados quedan a cargo de los estudiantes. La literatura indica que la combinación de simulaciones con preguntas bien diseñadas y experiencias experimentales, reales o virtuales, favorece aprendizajes significativos frente al uso aislado

de simuladores [5,6]. Integrar prompts de indagación en esta estructura permite aprovechar la personalización y la retroalimentación de la IA sin diluir el trabajo de modelización, análisis de datos y argumentación que constituye el núcleo de la actividad experimental en Física.

5.5. Acuerdo explícito de uso de IA

En trabajos recientes sobre creencias docentes y percepciones estudiantiles frente a la IA en educación [4,7] se subraya la importancia de explicitar las condiciones de uso de estas herramientas en cada asignatura. Estos estudios muestran que la ausencia de criterios claros sobre lo que se considera un uso aceptable de la IA genera incertidumbre y tensiones en torno a la honestidad académica; por eso recomiendan acordar reglas de uso con el grupo desde el inicio del curso.

En la práctica, el acuerdo que se construya al inicio del curso puede incluir al menos los siguientes puntos:

- 1) **Qué usos están permitidos:** lluvia de ideas, planificación de experimentos, explicación de conceptos, generación de preguntas.
- 2) **Cuáles no lo están:** cálculos finales de tareas, derivaciones de ecuaciones (sin verificación), respuestas sin análisis crítico.
- 3) **Qué evidencias deben entregarse:** *prompt*, fragmento relevante de la respuesta, verificación mediante simulación o cálculo y una breve reflexión sobre qué ha aprendido.
- 4) **Cómo se cita:** “IA (*ChatGPT-4*, 15 de noviembre): [*prompt*]; verificación: [*método usado*]”.

Este tipo de acuerdo dialoga con los hallazgos sobre percepciones estudiantiles y con los marcos de integridad académica descritos en la literatura reciente sobre creencias docentes e integración de IA en contextos educativos [4,7].

6. Criterios de evaluación para actividades con IA

Se propone una rúbrica que centre la evaluación en la calidad del razonamiento y en el seguimiento del proceso, en lugar de limitarse al acierto del resultado numérico. En consonancia con la literatura sobre evaluación formativa y sobre uso educativo de la IA, la idea es que los criterios de la rúbrica y la retroalimentación asociada hagan visibles los procesos que los estudiantes realizan al modelizar, verificar y registrar su trabajo, y que disuadan la delegación acrítica de la tarea en la herramienta digital [15,16].

En este marco, la rúbrica propuesta organiza la evaluación en torno a cuatro preguntas clave: (1) ¿qué modelo físico eligieron los estudiantes y por qué resulta adecuado frente a otras alternativas?, (2) ¿cómo verificaron las respuestas generadas por la IA (mediante simulaciones, cálculos o mediciones)?, (3) ¿qué discrepancias identificaron entre predicciones y datos y cómo las explicaron? y (4) ¿hasta qué punto dejaron un registro claro y legible de los pasos seguidos, incluyendo *prompts*, respuestas, decisiones intermedias y fuentes de verificación? Valorar estos aspectos permite focalizar la atención en el proceso de modelización y contraste con evidencia, en lugar de centrar la evaluación en la ejecución mecánica de algoritmos.

A modo de ejemplo, una rúbrica sencilla para la actividad presentada en la sección 5.1 puede organizarse en tres niveles (inicial, intermedio y avanzado) en torno a cuatro criterios: a) selección y justificación del modelo físico, b) calidad de la verificación de la respuesta de la IA (uso pertinente de simulación, cálculo o medición), c) análisis de las diferencias entre predicciones y datos, y d) claridad y completitud del registro del proceso. En el nivel inicial, los estudiantes utilizan la IA como fuente de respuestas sin justificar supuestos ni verificar con datos; en el nivel intermedio identifica variables relevantes, contrasta al menos un aspecto cuantitativo y describe de forma sintética el procedimiento; en el nivel avanzado explicita supuestos, contrasta distintos modelos, discute discrepancias y mantiene un registro de *prompts*, respuestas y decisiones suficientemente detallado como para reconstruir el razonamiento más allá del resultado numérico. La Tabla 1 ilustra una posible operacionalización de estos niveles y criterios para actividades que combinan IA y simulaciones en Física.

Además de valorar el desempeño de los estudiantes, resulta pertinente evaluar de manera formativa el propio desempeño de la IA en las actividades de aula. Kahaleh y Lopez [15] proponen un marco simple con tres criterios (rigurosidad científica, personalización y eficacia pedagógica) que puede trasladarse al contexto del aula. El cuerpo docente puede usar estos criterios para seleccionar ejemplos de respuestas de IA y analizarlas junto con la clase: ¿la explicación respeta los principios de la mecánica?, ¿adapta el lenguaje al nivel de secundaria?, ¿plantea preguntas y ejemplos que ayuden a revisar una concepción alternativa o se limita a presentar una solución? Esta revisión compartida convierte a la IA en objeto de análisis crítico y refuerza el mensaje central del artículo: se trata de una herramienta que puede facilitar el aprendizaje cuando se integra en secuencias bien diseñadas, pero no reemplaza el esfuerzo ni el

juicio conceptual de quienes aprenden ni de quienes enseñan.

respuestas y comentarios escritos para la retroalimentación formativa.

Tabla 1 Rúbrica para actividades con IA y simulaciones en Física

Criterio	Nivel avanzado	Nivel intermedio	Nivel inicial
Elección y explicitación del modelo físico	Selecciona un modelo físico adecuado para la situación, lo nombra explícitamente y justifica de forma clara por qué es apropiado, vinculándolo con las magnitudes y condiciones del problema.	Selecciona un modelo físico adecuado y lo nombra, pero la justificación es incompleta, confusa o meramente enunciativa (no queda claro cómo se aplica a la situación).	No se reconoce un modelo físico adecuado, o el modelo mencionado es incorrecto o irrelevante para la situación planteada.
Uso crítico de la respuesta de IA	Identifica y explica con claridad los problemas relevantes de la respuesta de la IA.	Identifica al menos un posible problema en la respuesta de la IA, aunque con explicación limitada.	No identifica problemas en la respuesta de la IA ni la cuestiona.
Análisis de discrepancias entre predicción y datos	Analiza con detalle las discrepancias entre predicciones y datos, indicando una explicación principal coherente.	Reconoce de forma general discrepancias entre predicciones y datos, sin análisis detallado.	No relaciona las predicciones con los datos al describir los resultados.
Claridad del registro del proceso seguido	El registro permite seguir con claridad todo el proceso realizado, desde las decisiones iniciales hasta las conclusiones.	El registro permite seguir solo los pasos principales del proceso, con algunas lagunas.	El registro no permite reconstruir el proceso realizado.

Nota. Por razones de espacio se presenta una versión sintética de la rúbrica, centrada en cuatro criterios y tres niveles de desempeño. En la práctica debería ampliarse con ejemplos de

7. Discusión

Los estudios revisados coinciden en que la IA generativa no actúa como un sustituto neutral de la enseñanza de la Física, sino como una herramienta cuya integración amplifica tanto oportunidades como riesgos. Por un lado, se han documentado mejoras en el rendimiento y en la comprensión conceptual cuando ChatGPT se integra en actividades de aprendizaje autorregulado o de indagación guiada [9–11]. Por otro, las percepciones de docentes y estudiantes subrayan riesgos vinculados a formas superficiales de aprendizaje, pérdida de esfuerzo cognitivo, inequidades en el acceso y dudas sobre la validez futura del rol docente, especialmente en lo que respecta a la evaluación y a la enseñanza de contenidos disciplinares [3,4,7,8,13,17].

En este panorama, los resultados de la literatura convergen en una serie de condiciones para un uso responsable: explicitar y negociar reglas de uso con el grupo, situar la IA en tareas que exijan modelizar, contrastar y explicar, verificar de forma sistemática lo generado mediante simulaciones o mediciones, y contar con espacios de desarrollo profesional donde los docentes de Física puedan experimentar, discutir y ajustar sus prácticas. Los lineamientos y estrategias propuestos en este artículo se inscriben en ese horizonte, al priorizar la IA como andamio del razonamiento físico bajo conducción docente y no como sustituto de la actividad intelectual de quienes aprenden.

Un aspecto menos desarrollado en la literatura, pero relevante para la práctica cotidiana, es el impacto de estas propuestas en la carga de trabajo docente y en las expectativas de cobertura de contenidos. Diseñar secuencias de indagación con IA, elaborar rúbricas centradas en el razonamiento y revisar producciones donde interviene el asistente requiere tiempo de planificación, coordinación y corrección que no es neutro. En contextos de currículos extensos y tiempos de clase ajustados, esto obliga a tomar decisiones sobre qué temas se trabajan con mayor profundidad conceptual y qué lugar mantienen las prácticas tradicionales de resolución de problemas numéricos.

8. Conclusiones

La IA no reemplaza al docente de Física; al contrario, vuelve más necesario el acompañamiento pedagógico del docente para orientar aprendizajes significativos en un contexto donde las respuestas rápidas están disponibles en cualquier dispositivo. La tarea docente se concentra en lo medular de la disciplina: construir y seleccionar modelos pertinentes, explicitar supuestos, contrastar con datos y comunicar límites de validez.

La literatura revisada sugiere que, cuando la IA se integra con reglas claras, conjuntamente con simulaciones y actividades experimentales, y en el marco de una evaluación formativa basada en tareas de modelización, contraste y justificación, la IA puede contribuir a aprendizajes de mayor profundidad. En particular, favorece la elaboración de explicaciones, la comparación de modelos, la exploración de variantes y la discusión crítica de errores, siempre que los estudiantes deban verificar lo generado con evidencia y dejar registro del proceso seguido. En un escenario educativo que exige pensamiento crítico, capacidad de modelizar y manejo autónomo de información, este uso de la IA puede liberar tiempo de aula para el razonamiento físico y reducir el peso de tareas rutinarias, aunque ello no está exento de costos: obliga a revisar el equilibrio entre la cantidad de contenidos abordados y la profundidad con que se trabajan, y a aceptar que parte del esfuerzo docente se desplazará desde la preparación de repartidos de ejercicios hacia el diseño y ajuste de actividades de indagación, instrumentos de evaluación y pautas de uso de la IA.

En el contexto uruguayo, este desafío es a la vez profesional y colectivo. Los espacios de coordinación, como las coordinaciones de asignatura y de nivel, así como las salas docentes institucionales, departamentales o regionales, ofrecen oportunidades para acordar el lugar de la IA generativa en las clases de Física, explicitar contratos pedagógicos, compartir experiencias exitosas y revisar criterios de evaluación. En esos ámbitos, el cuerpo docente puede compartir secuencias didácticas, rúbricas, materiales y criterios, de modo que existan acuerdos mínimos comunes sobre el uso de la IA, respetando al mismo tiempo la libertad de cátedra, el estilo de cada docente y las particularidades de cada grupo. En ese marco, los modelos de lenguaje pueden integrarse como andamiaje del razonamiento físico bajo conducción docente, y no como sustitutos de la experiencia, de la reflexión ni de la decisión pedagógica.

9. Referencias

- [1] S. Aldazharova, G. Issayeva, S. Maxutov & N. Balta, Assessing AI's problem solving in physics: Analyzing reasoning, false positives and negatives through the force concept inventory, *Contemporary Educational Technology* 16, ep538 (2024).
- [2] C. G. West, AI and the FCI: Can ChatGPT project an understanding of introductory physics? *arXiv* 2303.01067 (2023).
- [3] L. Ding, X. Li, M. Jiang & A. Gapud, Students' perceptions of using ChatGPT in a physics class as a virtual tutor, *International Journal of Educational Technology in Higher Education* 20, 63 (2023).
- [4] F. Guo, T. Li & C. J. L. Cunningham, One year in the classroom with ChatGPT: Empirical insights and transformative impacts, *Frontiers in Education* 10, 1574477 (2025).
- [5] F. Mahligawati, E. Allanas, M. H. Butarbutar & N. A. N. Nordin, Artificial intelligence in physics education: A comprehensive literature review, *Journal of Physics: Conference Series* 2596, 012080 (2023).
- [6] S. Ramaila, Pedagogical affordances of physics education technology in teaching and learning: A systematic review, *The International Journal of Technology, Knowledge, and Society* 20, 109–131 (2024).
- [7] B. Cabellos, C. de Aldama & J. I. Pozo, University teachers' beliefs about the use of generative artificial intelligence for teaching and learning, *Frontiers in Psychology* 15, 1468900 (2024).
- [8] P. Wattanakasiwich, K. Kaewkhong & D. Katwibun, Physics instructors' acceptance and implementation of generative AI, *Physical Review Physics Education Research* 21, 010155 (2025).
- [9] K. Alarbi, M. Halaweh, H. Tairab, N. R. Alsalihi, N. Annamalai & F. Aldarmaki, Making a revolution in physics learning in high schools with ChatGPT: A case study in UAE, *EURASIA Journal of Mathematics, Science and Technology Education* 20, em2499 (2024).
- [10] T. El Fathi, A. Saad, H. Larhzil, D. Lamri & E. M. Al Ibrahim, Integrating generative AI into STEM education: Enhancing conceptual understanding, addressing misconceptions, and assessing student acceptance, *Disciplinary and Interdisciplinary Science Education Research* 7, 6 (2025).
- [11] O. A. Obasi & F. O. Nnadi, Effect of self-regulated-learning strategy through chat-generative-pretrained-transformer on secondary school students' achievement in projectile motion concepts, *Journal of Science Education (JOSCED)* 4, 1–13 (2023).

- [12] M. Guerrero-Zambrano, L. Sanchez-Alvarado, B. Valarezo-Chamba & E. Lamilla-Rubio, Transforming physics teacher training through ChatGPT: A study on usability and impact, *Education Sciences* 15, 887 (2025).
- [13] H. Uğraş, M. Uğraş, S. Papadakis & M. Kalogiannakis, Innovative early childhood STEM education with ChatGPT: Teacher perspectives, *Technology, Knowledge and Learning* 30, 809–831 (2025).
- [14] T. Min, B. Lee & H. Jho, Integrating generative artificial intelligence in the design of scientific inquiry for middle school students, *Education and Information Technologies* 30, 15329–15360 (2025).
- [15] R. Kahaleh & V. Lopez, Evaluating large language models in high school physics education: Addressing misconceptions and fostering conceptual understanding, *Physics Education* 60, 025013 (2025).
- [16] H. A. Mustofa, M. R. Bilad & N. W. B. Grendis, Utilizing AI for physics problem solving: A literature review and ChatGPT experience, *Lensa: Jurnal Kependidikan Fisika* 12, 78–97 (2024).
- [17] T. Famaye, C. S. Bailey, I. Adisa & G. A. Irgens, “What makes ChatGPT dangerous is also what makes it special”: High-school student perspectives on the integration or ban of artificial intelligence in educational contexts, *International Journal of Technology in Education* 7, 174–199 (2024).