

Integrasi Artificial Intelligence dalam Pembelajaran Fisika Berbasis Scientific Evidence: Systematic Literature Review terhadap Literasi Sains dan Kemampuan Pemecahan Masalah

Andrilana

Program Studi Pendidikan Fisika, Universitas Nahdlatul Ulama Cirebon

Corresponding Author: andrilana@unucirebon.ac.id

ABSTRAK

Artificial Intelligence (AI) semakin digunakan dalam pembelajaran fisika, tetapi kemampuan menghasilkan jawaban yang meyakinkan tidak selalu identik dengan penalaran ilmiah yang benar. Penelitian ini bertujuan menyintesis integrasi AI dengan menempatkan scientific evidence sebagai penghubung antara penggunaan AI, literasi sains, dan kemampuan pemecahan masalah. SLR dengan pelaporan seleksi studi yang diadaptasi dari alur PRISMA dilakukan pada publikasi 2021–15 September 2026 melalui penelusuran ilmiah terarah, verifikasi DOI/penerbit, dan citation tracking. Dari 27 rekaman kandidat, 15 studi memenuhi kriteria dan quality assessment. Sintesis menunjukkan AI efektif sebagai scaffolding pada tugas terstruktur, tetapi lebih rapuh pada pemodelan, interpretasi representasi, data tidak lengkap, dan evaluasi asumsi. Kajian mengusulkan kerangka AI–Evidence–Reasoning yang menempatkan verifikasi data, asumsi, representasi, satuan, dan konsistensi konsep sebagai mekanisme inti agar AI memperkuat, bukan menggantikan, penalaran fisika.

Kata Kunci: Artificial Intelligence; Pembelajaran Fisika; Scientific Evidence; Literasi Sains; Pemecahan Masalah

This is an open access article under the CC - BY license.



PENDAHULUAN

Perkembangan artificial intelligence (AI), terutama large language models dan generative AI, telah mengubah cara peserta didik memperoleh penjelasan, menyusun solusi, dan mengevaluasi informasi dalam pembelajaran. Dalam pendidikan fisika, perubahan ini memiliki dua sisi. AI dapat menyediakan penjelasan cepat, contoh penyelesaian, umpan balik, dan dukungan personal, tetapi fisika tidak hanya menuntut jawaban akhir. Pembelajaran fisika menuntut kemampuan membangun model, memilih prinsip, menghubungkan representasi, menggunakan data, menguji asumsi, serta menilai apakah kesimpulan konsisten dengan bukti dan hukum fisika. Karena itu, kualitas integrasi AI perlu dinilai dari kontribusinya terhadap praktik epistemik tersebut, bukan hanya dari kecepatan atau kelancaran respons.

Sejumlah penelitian memperlihatkan potensi sekaligus keterbatasan AI. Kortemeyer (2023) menguji ChatGPT pada materi asesmen representatif dari mata kuliah fisika berbasis kalkulus dan menilai responsnya dengan cara yang sama seperti respons mahasiswa. Hasilnya, ChatGPT hanya lulus tipis, sekaligus menunjukkan sejumlah prakonsepsi dan kesalahan yang menyerupai peserta didik pemula. Temuan ini menunjukkan bahwa performa pada tugas fisika perlu dibaca bersama jejak kesalahan dan kualitas penalarannya, bukan hanya status benar-salah jawaban. Pada konteks pemecahan masalah, Wang et al. (2024) menguji GPT-4 pada 40 masalah dari mata kuliah engineering physics tingkat perguruan tinggi: 16 masalah well-specified dengan data yang diperlukan tersedia dan 24 masalah under-specified yang memerlukan data atau asumsi tambahan. GPT-4 menyelesaikan 62,5% masalah well-specified dan hanya 8,3% masalah under-specified. Dengan demikian, kemampuan menghasilkan prosedur tidak otomatis menunjukkan kemampuan membangun model fisik atau menentukan asumsi yang wajar.

Studi lain memperkuat kebutuhan akan verifikasi. Küchemann et al. (2023) menunjukkan bahwa penggunaan ChatGPT oleh calon guru fisika dapat mendukung pengembangan soal, tetapi keluaran AI masih menghadapi persoalan kualitas dan spesifisitas. Studi tentang grafik kinematika juga menunjukkan bahwa AI

dapat mengusulkan strategi yang produktif, tetapi mengalami kesulitan pada interpretasi visual tertentu (Polverini & Gregorcic, 2024). Perbandingan respons mahasiswa dan AI oleh Sirnoorkar et al. (2024) memperlihatkan perbedaan penting antara bahasa penalaran yang tampak terstruktur dan praktik sensemaking yang benar benar dilakukan saat menyelesaikan masalah. Dengan demikian, integrasi AI membutuhkan desain pembelajaran yang memaksa peserta didik memeriksa hubungan antara klaim, data, asumsi, model, dan bukti.

Dalam kajian ini, scientific evidence dipahami secara operasional sebagai informasi empiris atau representasional yang digunakan untuk mendukung, menguji, atau menolak suatu klaim fisika, meliputi data pengukuran, grafik, diagram, kondisi batas, asumsi pemodelan, hubungan matematis, konsistensi satuan, serta kesesuaian dengan prinsip fisika. Konsep ini beririsan langsung dengan literasi sains karena peserta didik perlu menilai kualitas informasi dan menarik kesimpulan berdasarkan bukti. Pada saat yang sama, scientific evidence merupakan komponen penting pemecahan masalah karena solusi fisika yang baik memerlukan pemilihan informasi relevan, pemodelan, eksekusi strategi, dan evaluasi hasil. Namun, evidence constraint dan reasoning constraint perlu dibedakan. Evidence constraint membatasi klaim pada data, representasi, sumber, dan kondisi yang tersedia; reasoning constraint membatasi cara klaim tersebut dapat diturunkan atau diuji melalui prinsip, model, asumsi, satuan, dan hubungan matematis. Keduanya saling terkait, tetapi tidak identik: bukti yang memadai belum menjamin inferensi yang benar, sementara penalaran yang rapi tidak dapat menggantikan bukti yang tidak tersedia atau tidak relevan.

Literatur tentang AI dalam pendidikan MIPA dan fisika menunjukkan perkembangan yang semakin beragam, tetapi belum homogen. Al-Kamzari dan Alias (2025) memetakan aplikasi, manfaat, dan tantangan AI pada fisika sekolah menengah. Dalam konteks Indonesia, Nelsya et al. (2025) melakukan kajian literatur terhadap 35 artikel dan menganalisis 23 artikel mengenai teknologi digital dan AI dalam pendidikan MIPA, termasuk pendamping belajar, penilaian otomatis, pembelajaran adaptif, dan simulasi. Agustinasari et al. (2025) secara khusus menelaah enam artikel tentang AI dan literasi sains dalam pendidikan sains serta menekankan pentingnya pembelajaran yang tetap aktif dan reflektif. Sementara itu, Putra dan Algiranto (2026) melaporkan studi kuasi-eksperimental pada 24 mahasiswa Pendidikan Fisika yang menggunakan chatbot AI pada mata kuliah Fisika Inti dan menemukan perbedaan hasil belajar serta berpikir kritis antara kelompok yang dibandingkan. Bukti-bukti tersebut memperlihatkan bahwa penelitian AI dalam pendidikan MIPA bergerak dari pemetaan aplikasi menuju pengujian dampak dan kualitas proses belajar. Namun, kajian yang secara eksplisit menempatkan scientific evidence sebagai mekanisme yang menghubungkan AI dengan literasi sains dan pemecahan masalah fisika masih terbatas.

Berdasarkan kesenjangan tersebut, penelitian ini bertujuan: (1) mengidentifikasi pola penggunaan AI dalam pembelajaran fisika yang berkaitan dengan bukti dan penalaran; (2) menganalisis kontribusi, keterbatasan, dan pola temuan AI terhadap literasi sains dan kemampuan pemecahan masalah; dan (3) merumuskan kerangka konseptual AI-Evidence-Reasoning sebagai pedoman integrasi AI dalam pembelajaran fisika. Kebaruan kajian terletak pada pergeseran fokus dari pertanyaan apakah AI mampu menjawab soal menuju pertanyaan bagaimana peserta didik menggunakan, memeriksa, dan mempertanggungjawabkan bukti ketika berinteraksi dengan AI.

METODE

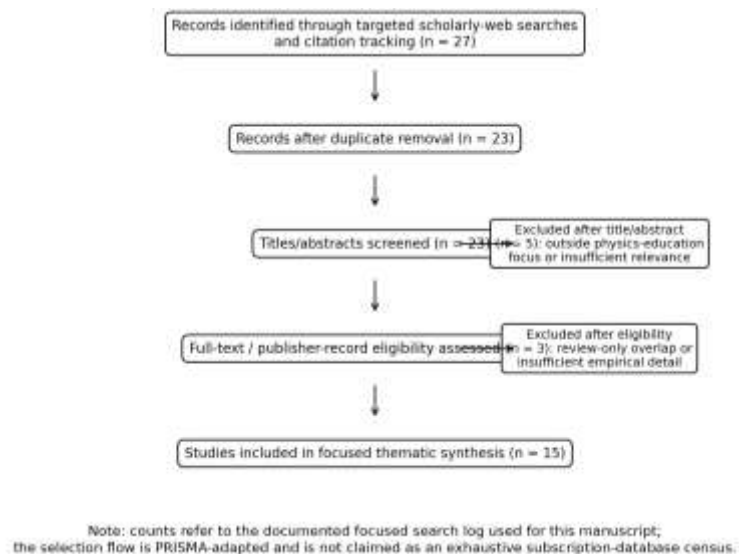
Pertanyaan penelitian (research questions/RQ) dirumuskan untuk menjaga fokus sintesis: RQ1: Bagaimana karakteristik dan tren penggunaan AI dalam pembelajaran fisika pada studi 2021–2026? RQ2: Fungsi pedagogis AI apa yang paling berkaitan dengan penggunaan scientific evidence? RQ3: Bagaimana AI berkaitan dengan kemampuan peserta didik mengevaluasi bukti dan literasi sains? RQ4: Bagaimana AI mendukung atau menghambat tahapan pemecahan masalah fisika dalam studi yang ditinjau? RQ5: Kerangka pembelajaran apa yang dapat dirumuskan untuk mengarahkan penggunaan AI agar tetap menuntut verifikasi bukti dan penalaran fisika?

Kajian ini menggunakan Systematic Literature Review (SLR) berorientasi sintesis tematik dengan pelaporan seleksi studi yang diadaptasi dari alur PRISMA. Penelusuran difokuskan pada publikasi 2021–15 September 2026. Karena akses penelusuran dilakukan melalui indeks ilmiah terbuka, laman penerbit, DOI, serta citation tracking, kajian ini tidak mengklaim sebagai sensus lengkap seluruh basis data berlangganan. Transparansi tersebut dipertahankan agar proses dapat direplikasi dan batas generalisasi hasil dapat dinilai secara tepat.

Strategi pencarian menggunakan kombinasi istilah: ("artificial intelligence" OR "generative AI" OR ChatGPT OR "large language model") AND ("physics education" OR "physics learning") AND ("scientific evidence" OR "scientific reasoning" OR "scientific literacy" OR "problem solving" OR sensemaking). Penelusuran diperluas melalui backward dan forward citation tracking dari review fisika-AI terbaru, terutama Moundridou et al. (2026) dan Roziqin et al. (2026), kemudian diverifikasi pada laman penerbit/DOI. Log pencarian terfokus menghasilkan 27 rekaman kandidat; setelah deduplikasi tersisa 23, lima dieliminasi pada skrinng judul/abstrak, tiga dieliminasi pada tahap kelayakan, dan 15 studi dimasukkan dalam sintesis tematik.

Kriteria inklusi adalah: (1) terbit 2021–2026; (2) artikel jurnal/prosiding peer-reviewed atau rekaman penerbit yang dapat diverifikasi; (3) konteks langsung pada fisika/pembelajaran fisika; (4) membahas AI/GenAI/LLM sebagai alat belajar, tutor, problem solver, pengembang tugas, atau objek evaluasi; dan (5) menyediakan temuan yang dapat dikodekan terkait penalaran, bukti, pemahaman konsep, literasi sains, atau pemecahan masalah. Kriteria eksklusi mencakup artikel non-fisika tanpa implikasi langsung, opini tanpa data/sintesis yang dapat diverifikasi, duplikasi, dan studi yang tidak menyediakan informasi metodologis memadai.

Penilaian kualitas dilakukan secara naratif menggunakan lima indikator: kejelasan tujuan, kejelasan konteks/sampel atau unit analisis, kecukupan metode, keterlacakan bukti terhadap kesimpulan, serta pengakuan keterbatasan/implikasi. Penilaian ini digunakan untuk menimbang kekuatan bukti pada tahap sintesis, bukan untuk menghasilkan presisi numerik yang tidak dapat dipertanggungjawabkan dari pelaporan artikel. Data diekstraksi ke matriks yang memuat tahun, konteks, jenis AI, fungsi pedagogis, topik fisika, bentuk scientific evidence, indikator literasi sains/problem solving, manfaat, dan risiko. Scientific evidence dioperasionalkan ke dalam lima praktik: identifikasi data relevan; evaluasi asumsi dan sumber; penggunaan representasi; pemeriksaan konsistensi model-konsep-satuan; dan verifikasi kesimpulan terhadap bukti. Heterogenitas desain dan keluaran studi menyebabkan meta-analisis tidak dilakukan; hasil disajikan sebagai sintesis deskriptif dan tematik.



Gambar 1. Diagram alur seleksi studi (PRISMA-adapted)

Tabel 1. Kerangka appraisal kualitas studi

Indikator	Fokus appraisal	Keputusan sintesis
Q1	Tujuan/pertanyaan penelitian dinyatakan jelas	Bukti dipakai sesuai tujuan studi.
Q2	Konteks, sampel, atau unit analisis dapat diidentifikasi	Generalisasi dibatasi pada konteks yang dilaporkan.
Q3	Metode pengumpulan/analisis data atau prosedur review dijelaskan	Pelaporan metodologis dipertimbangkan saat menimbang kekuatan bukti.
Q4	Kesimpulan dapat ditelusuri ke data/hasil yang dilaporkan	Klaim yang tidak terlacak tidak dijadikan dasar utama sintesis.
Q5	Keterbatasan dan/atau implikasi dilaporkan	Keterbatasan dipertimbangkan saat merumuskan implikasi.

Keterangan: appraisal digunakan untuk menilai kekuatan dan keterlacakan bukti pada sintesis; tidak digunakan sebagai skor kuantitatif untuk mengurutkan studi.

HASIL DAN PEMBAHASAN

Sintesis 15 studi inti menjawab RQ1–RQ5 melalui empat tema: (1) perkembangan fungsi AI dari problem solver menuju tutor dan alat evaluasi; (2) scientific evidence sebagai mekanisme verifikasi keluaran AI; (3) hubungan evaluasi bukti dengan literasi sains; dan (4) peran sensemaking, pemodelan, serta representasi dalam problem solving. Temuan lintas studi kemudian digunakan untuk membangun kerangka AI–Evidence–Reasoning (AER).

Hasil sintesis menunjukkan bahwa literatur AI dalam pendidikan fisika setelah 2023 bergerak dari evaluasi kemampuan AI menjawab soal menuju studi yang lebih pedagogis, termasuk persepsi peserta didik, dukungan terhadap calon guru, evaluasi kualitas penalaran, dan integrasi AI sebagai tutor. Temuan tersebut konsisten dengan perkembangan kajian AI dalam pendidikan MIPA yang dipetakan oleh Nelsya et al. (2025) dan dengan kajian literasi sains oleh Agustinasari et al. (2025). Namun, korpus inti penelitian ini tetap dibatasi pada 15 studi yang memenuhi kriteria khusus fisika dan fokus bukti/penalaran yang telah ditetapkan dalam metode.

Tabel 2. Sintesis studi kunci AI, scientific evidence, dan pemecahan masalah fisika

Studi	Konteks	Fokus bukti/kompetensi	Temuan relevan
Kortemeyer (2023)	Fisika pengantar; ChatGPT	Asesmen konsep dan problem solving	Dapat menyelesaikan banyak tugas, tetapi masih menunjukkan kesalahan konsep/perhitungan; verifikasi manusia tetap diperlukan.
Dahlkemper et al. (2023)	Mahasiswa fisika; ChatGPT	Akurasi ilmiah respons	Pengetahuan awal memengaruhi kemampuan menilai respons AI; keluaran yang fasih dapat menutupi ketidakakuratan.
Küchemann et al. (2023)	Calon guru; ChatGPT	Pengembangan soal kinematika	AI membantu ide dan efisiensi, tetapi kualitas konteks dan spesifisitas informasi perlu diperiksa.
Wang et al. (2024)	Fisika teknik; GPT-4	Masalah nyata lengkap/tidak lengkap	Akurasi tinggi relatif pada data lengkap dan sangat menurun pada masalah yang membutuhkan asumsi dan data tambahan.
Polverini & Gregorcic (2024)	Kinematika; ChatGPT-4	Interpretasi grafik	Strategi penyelesaian sering produktif, tetapi interpretasi visual grafik tetap menjadi kelemahan; penggunaan sebagai tutor memerlukan verifikasi kritis.
Sirnoorkar et al. (2024)	Fisika pengantar; ChatGPT	Sensemaking dan mechanistic reasoning	Respons AI tampak terstruktur, sedangkan mahasiswa menunjukkan praktik penalaran melalui revisi representasi dan argumen.
Olanrewaju & Ojo (2026)	115 mahasiswa fisika; quasi-experiment	Engagement dan problem solving	Integrasi ChatGPT dibandingkan pembelajaran kuliah tradisional; temuan digunakan sebagai bukti empiris tentang dukungan AI pada pembelajaran fisika tingkat sarjana.
Polverini et al. (2025)	BEMA; ChatGPT-4/4o	Representasi visual listrik-magnet	ChatGPT-4o meningkat dibanding versi sebelumnya, tetapi masih mengalami kesulitan interpretasi visual, hukum fisika, dan koordinasi spasial.
Moundridou et al. (2026)	Masalah tradisional dan kaya konteks	Problem solving	AI lebih kuat pada soal tradisional dan lebih lemah pada masalah kaya konteks yang menuntut pemodelan/asumsi.
Saputra et al. (2025)	Pembelajaran fisika Indonesia	Pemetaan penggunaan AI	Adopsi meningkat, tetapi kesiapan pedagogis dan kualitas integrasi masih menjadi tantangan.
Al-Kamzari & Alias (2025)	Fisika sekolah menengah; SLR	Aplikasi, manfaat, dan tantangan AI	Personalisasi, virtual lab, dan intelligent tutoring menjanjikan; kesiapan guru, privasi, serta keseimbangan penggunaan AI dengan berpikir kritis menjadi tantangan.

Studi	Konteks	Fokus bukti/kompetensi	Temuan relevan
Boolzen et al. (2026)	85 studi STEM; 49 studi meta-analisis	Hasil belajar kognitif GenAI	Efek agregat tampak positif tetapi sangat heterogen; analisis juga menunjukkan potensi publication bias sehingga interpretasi dampak perlu berhati-hati.
Roziqin et al. (2026)	44 studi empiris 2021–2025	SLR AI dalam fisika	Manfaat pada performa dan engagement disertai risiko epistemic unreliability, overreliance, dan lemahnya AI literacy.
Meşe et al. (2026)	Hukum Gauss & medan listrik; ChatGPT-4o	Kualitas logical reasoning	Analisis enam soal kalkulus-fisika menunjukkan kualitas reasoning perlu dinilai bukan hanya dari jawaban akhir, tetapi juga dari struktur penalarannya.
Rangkuti et al. (2026)	Gelombang; beberapa versi ChatGPT	Akurasi konsep dan reasoning	Versi lebih baru membaik pada banyak item, tetapi kesalahan penalaran masih muncul dan perlu evaluasi.

Pertama, AI berfungsi paling baik sebagai scaffolding ketika struktur masalah relatif jelas. Pada masalah dengan informasi lengkap, AI dapat membantu mengidentifikasi persamaan, menjelaskan langkah, dan menghasilkan alternatif strategi. Namun, ketika masalah membutuhkan keputusan tentang data apa yang diperlukan, bagaimana membangun model, atau asumsi apa yang masuk akal, performa menurun. Pola ini penting karena kemampuan pemecahan masalah fisika tidak identik dengan substitusi angka ke persamaan. Kompetensi yang lebih tinggi justru terlihat ketika peserta didik dapat menentukan relevansi bukti dan mempertanggungjawabkan model yang dipilih.

Kedua, evaluasi scientific evidence muncul sebagai titik kritis literasi sains di era AI. Peserta didik dapat menerima jawaban yang secara linguistik meyakinkan tetapi secara ilmiah tidak tepat. Karena itu, literasi sains dalam lingkungan AI perlu mencakup kemampuan memeriksa sumber data, menguji asumsi, membandingkan keluaran AI dengan prinsip fisika, menilai konsistensi satuan, dan mencari bukti pembanding. Temuan tentang kesulitan mahasiswa menilai respons AI (Dahlkemper et al., 2023), serta kajian JPM oleh Agustinasari et al. (2025) yang menekankan pentingnya proses belajar aktif dan reflektif ketika AI digunakan dalam pendidikan sains, mendukung kebutuhan untuk menempatkan peserta didik sebagai evaluator aktif, bukan penerima pasif keluaran AI.

Ketiga, representasi menjadi area yang memerlukan perhatian. Fisika menggunakan grafik, diagram gaya, vektor, persamaan, dan model visual sebagai bagian dari penalaran. Studi interpretasi grafik kinematika memperlihatkan bahwa AI dapat menyatakan penalaran yang tampak benar tetapi tetap gagal pada aspek visual tertentu. Hal ini menunjukkan perlunya tugas yang meminta peserta didik menghubungkan sedikitnya dua representasi dan menjelaskan bukti yang mendukung perpindahan dari satu representasi ke representasi lain.

Keempat, perkembangan model AI tidak menghilangkan kebutuhan verifikasi. Rangkuti et al. (2026) menunjukkan peningkatan performa versi ChatGPT yang lebih baru pada masalah gelombang, tetapi kesalahan penalaran masih ditemukan. Artinya, desain pembelajaran tidak boleh bergantung pada asumsi bahwa peningkatan versi teknologi otomatis menjamin validitas ilmiah. Kompetensi yang perlu dibangun bersifat tahan terhadap perubahan teknologi, yaitu kemampuan mengevaluasi klaim dan bukti.

Secara keseluruhan, hasil sintesis mengarah pada tiga fungsi pedagogis AI: sebagai generator penjelasan dan hipotesis, sebagai scaffolding strategi penyelesaian, dan sebagai objek evaluasi kritis. Fungsi ketiga paling relevan bagi scientific evidence karena peserta didik tidak hanya menggunakan AI untuk memperoleh jawaban, tetapi menjadikan respons AI sebagai klaim yang harus diuji.

Tabel 3. Operasionalisasi scientific evidence dalam pembelajaran fisika berbantuan AI

Praktik	Indikator	Contoh aktivitas AI–fisika
Identifikasi bukti	Menentukan data, informasi, dan prinsip yang relevan	Tandai data yang diperlukan dan data yang tidak tersedia pada jawaban AI.
Evaluasi asumsi	Menilai kewajaran asumsi dan batas model	Minta AI menyebutkan asumsi, lalu peserta didik menguji setiap asumsi secara fisik.
Validasi representasi	Mencocokkan grafik, diagram, persamaan, dan narasi	Bandingkan penjelasan AI dengan grafik/diagram yang dibuat peserta didik.

Praktik	Indikator	Contoh aktivitas AI-fisika
Konsistensi ilmiah	Memeriksa konsep, satuan, tanda, orde besaran, dan kondisi batas	Lakukan unit check, limiting case, dan estimasi orde besaran.
Justifikasi kesimpulan	Menghubungkan klaim akhir dengan bukti	Tuliskan bukti mana yang mendukung atau menolak jawaban AI.

Temuan kajian memperlihatkan bahwa persoalan utama integrasi AI dalam pembelajaran fisika bukan sekadar apakah AI boleh digunakan, melainkan bagaimana interaksi dengan AI dirancang agar tetap menuntut aktivitas epistemik peserta didik. Jika AI digunakan hanya untuk menghasilkan jawaban akhir, teknologi berpotensi memindahkan proses berpikir dari peserta didik ke sistem. Sebaliknya, ketika respons AI diperlakukan sebagai hipotesis atau solusi sementara yang harus diuji, AI dapat menjadi pemicu penalaran yang lebih eksplisit.

Hubungan antara AI dan literasi sains dapat dijelaskan melalui proses evaluasi informasi. Literasi sains menuntut kemampuan menjelaskan fenomena secara ilmiah, mengevaluasi penyelidikan, serta menafsirkan data dan bukti. Dalam lingkungan AI, kompetensi tersebut diperluas menjadi kemampuan mengevaluasi informasi sintesis yang dihasilkan mesin. Peserta didik perlu memahami bahwa kefasihan bahasa bukan indikator validitas ilmiah. Temuan tentang kesulitan mahasiswa menilai respons AI (Dahlkemper et al., 2023) dan variasi akurasi pada tipe masalah berbeda (Wang et al., 2024) mendukung kebutuhan ini.

Dari sisi problem solving, AI dapat mempercepat tahap pencarian strategi, tetapi tahap sensemaking dan evaluasi tetap harus dimiliki peserta didik. Masalah fisika autentik sering tidak memberikan semua data. Problem solver harus menentukan variabel yang relevan, membuat estimasi, memilih model, dan menilai kewajaran hasil. Penurunan performa GPT-4 pada masalah under-specified yang dilaporkan Wang et al. (2024) menunjukkan bahwa pembelajaran sebaiknya memperbanyak masalah yang tidak dapat diselesaikan hanya melalui pattern matching prosedural. Temuan empiris terbaru pada pembelajaran Fisika Inti oleh Putra dan Algiranto (2026) juga menunjukkan potensi chatbot AI untuk mendukung hasil belajar dan berpikir kritis, tetapi temuan tersebut tidak menghilangkan kebutuhan akan evaluasi proses penalaran. Dengan demikian, AI dapat digunakan untuk membandingkan solusi, menemukan kelemahan asumsi, dan melatih argumentasi.

Berdasarkan sintesis, kajian ini mengusulkan kerangka AI-Evidence-Reasoning (AER). Kerangka dimulai dari Prompt/Problem, dilanjutkan AI Generated Claim, Evidence Check, Physics Reasoning, dan Reflective Decision. Pada tahap Evidence Check, peserta didik wajib memeriksa data, asumsi, representasi, satuan, dan konsistensi konsep. Pada tahap Physics Reasoning, peserta didik menyusun ulang solusi dengan alasan disipliner. Pada tahap Reflective Decision, peserta didik menerima, merevisi, atau menolak keluaran AI dengan menyebutkan bukti. Mekanisme ini membedakan penggunaan AI sebagai answer machine dari penggunaan AI sebagai cognitive scaffolding.

Kerangka AER memiliki implikasi praktis bagi dosen dan guru fisika. Penilaian dapat mengubah fokus dari jawaban akhir menuju jejak verifikasi. Misalnya, tugas dapat meminta peserta didik menyertakan respons AI, menandai minimal dua klaim yang perlu diperiksa, menunjukkan pemeriksaan satuan atau grafik, dan menulis keputusan akhir beserta alasan. Strategi ini sekaligus mengembangkan AI literacy, literasi sains, dan kemampuan pemecahan masalah tanpa menghilangkan manfaat teknologi.

Temuan ini juga menunjukkan bahwa scientific evidence dapat menjadi jembatan konseptual antara literasi sains dan problem solving. Literasi sains menyediakan kemampuan mengevaluasi informasi dan bukti, sedangkan problem solving menyediakan konteks penggunaan bukti untuk mengambil keputusan. AI menambah lapisan baru berupa kebutuhan untuk mengaudit keluaran sistem. Oleh karena itu, pembelajaran fisika berbantuan AI sebaiknya dirancang sebagai human-in-the-loop learning, yaitu peserta didik tetap menjadi pengambil keputusan epistemik terakhir.

Kajian memiliki keterbatasan. Pertama, perkembangan AI sangat cepat sehingga performa model yang dilaporkan pada satu versi tidak dapat langsung digeneralisasi ke versi berikutnya. Kedua, studi empiris masih banyak berfokus pada pendidikan tinggi dan ChatGPT, sehingga bukti untuk sekolah dan jenis AI lain relatif lebih terbatas. Ketiga, heterogenitas desain penelitian tidak memungkinkan estimasi effect size gabungan. Penelitian selanjutnya perlu menguji kerangka AER secara eksperimen pada topik fisika tertentu, menggunakan

instrumen literasi sains dan problem solving yang tervalidasi, serta membandingkan kondisi penggunaan AI dengan dan tanpa protokol verifikasi bukti.

Masalah/Prompt → Klaim AI → Pemeriksaan Bukti → Penalaran Fisika → Keputusan Reflektif

Gambar 2. Kerangka konseptual AI-Evidence-Reasoning (AER)

Pemeriksaan bukti mencakup data, asumsi, representasi, satuan, kondisi batas, dan konsistensi konsep. Keputusan reflektif berupa menerima, merevisi, atau menolak keluaran AI berdasarkan alasan yang dapat dipertanggungjawabkan.

SIMPULAN

Integrasi AI dalam pembelajaran fisika berpotensi mendukung penjelasan konsep, umpan balik, pengembangan tugas, dan pencarian strategi pemecahan masalah. Namun, bukti yang disintesis menunjukkan bahwa performa AI tidak seragam dan cenderung lebih rapuh ketika tugas memerlukan pemodelan dunia nyata, penentuan data yang belum tersedia, interpretasi representasi, dan evaluasi asumsi. Karena itu, scientific evidence perlu ditempatkan sebagai mekanisme inti penggunaan AI. Kerangka AI-Evidence-Reasoning yang diusulkan mengarahkan peserta didik untuk memperlakukan keluaran AI sebagai klaim yang harus diperiksa melalui data, representasi, prinsip fisika, satuan, dan kondisi batas sebelum keputusan diambil. Pendekatan ini berpotensi mengintegrasikan AI literacy, literasi sains, dan kemampuan pemecahan masalah dalam satu desain pembelajaran. Penelitian lanjutan diperlukan untuk menguji efektivitas kerangka tersebut secara empiris pada berbagai topik dan jenjang pendidikan fisika.

PERNYATAAN PENGGUNAAN AI GENERATIF

Artificial intelligence generatif (ChatGPT) digunakan untuk membantu penyusunan dan pengorganisasian naskah, penyempurnaan bahasa, serta dukungan penelusuran literatur. Penulis menelaah dan merevisi seluruh teks berbantuan AI, memeriksa rincian bibliografis terhadap rekaman penerbit/DOI yang tersedia, mengevaluasi analisis dan interpretasi, serta bertanggung jawab penuh atas akurasi, orisinalitas, integritas, dan isi akhir artikel.

Daftar Pustaka

- Agustinasari, A., Suastra, I. W., Atmadja, A. T., & Tika, I. N. (2025). Transformasi literasi sains di era kecerdasan buatan: Analisis dalam perspektif konstruktivisme. *JURNAL PENDIDIKAN MIPA*, 15(4), 1894–1900. <https://doi.org/10.37630/jpm.v15i4.3911>
- Al-Kamzari, F., & Alias, N. (2025). A systematic literature review of artificial intelligence (AI) in secondary school physics: Applications, benefits, and challenges. *Interactive Learning Environments*, 34(2), 676–693. <https://doi.org/10.1080/10494820.2025.2508323>
- Boolzen, C., Kuhn, J., Flegr, S., Rott, E.-M., Stausberg, N., et al. (2026). Evidence of impact and interpretational limits of generative AI in STEM education: A systematic review and meta-analysis on cognitive learning outcomes. *Artificial Intelligence Review*. <https://doi.org/10.1007/s10462-026-11665-9>
- Dahlkemper, M. N., Lahme, S. Z., & Klein, P. (2023). How do physics students evaluate artificial intelligence responses on comprehension questions? A study on the perceived scientific accuracy and linguistic quality of ChatGPT. *Physical Review Physics Education Research*, 19, 010142. <https://doi.org/10.1103/PhysRevPhysEducRes.19.010142>
- Kortemeyer, G. (2023). Could an artificial-intelligence agent pass an introductory physics course? *Physical Review Physics Education Research*, 19, 010132. <https://doi.org/10.1103/PhysRevPhysEducRes.19.010132>
- Küchemann, S., Steinert, S., Revenga, N., Schweinberger, M., Dinc, Y., Avila, K. E., & Kuhn, J. (2023). Can ChatGPT support prospective teachers in physics task development? *Physical Review Physics Education Research*, 19, 020128. <https://doi.org/10.1103/PhysRevPhysEducRes.19.020128>
- Meşe, E., Gür, R., & Korkmaz, S. D. (2026). Assessing ChatGPT's logical reasoning quality in physics problem solving. *Discover Education*, 5, 560. <https://doi.org/10.1007/s44217-026-01464-3>

- Moundridou, M., Moutis, N., Charalampopoulos, K., & Matzakos, N. (2026). Generative AI in physics education: A PRISMA-guided systematic review of empirical studies. *Artificial Intelligence and Applications*, 4(2), 149–164. <https://doi.org/10.47852/bonviewAIA62027183>
- Nelsya, N. P., Putri, R. A., Zaenudin, D., Julaeha, S., & Elvina, E. (2025). Kajian literatur: Implementasi teknologi digital dan kecerdasan buatan dalam pendidikan MIPA di Indonesia. *JURNAL PENDIDIKAN MIPA*, 15(4), 1837–1842. <https://doi.org/10.37630/jpm.v15i4.3716>
- Olanrewaju, M. A., & Ojo, K. R. (2026). AI-ChatGPT, students' engagement and problem-solving ability in undergraduate physics: An experimental study. *European Journal of Education Studies*, 13(2). <https://doi.org/10.46827/ejes.v13i2.6542>
- Polverini, G., & Gregorcic, B. (2024). Performance of ChatGPT on the test of understanding graphs in kinematics. *Physical Review Physics Education Research*, 20, 010109. <https://doi.org/10.1103/PhysRevPhysEducRes.20.010109>
- Polverini, G., Melin, J., Önerud, E., & Gregorcic, B. (2025). Performance of ChatGPT on tasks involving physics visual representations: The case of the brief electricity and magnetism assessment. *Physical Review Physics Education Research*, 21, 010154. <https://doi.org/10.1103/PhysRevPhysEducRes.21.010154>
- Putra, D. J., & Algiranto, A. (2026). Pengaruh penggunaan artificial intelligence berbasis chatbot pada mata kuliah Fisika Inti terhadap hasil belajar dan kemampuan berpikir kritis mahasiswa Pendidikan Fisika. *JURNAL PENDIDIKAN MIPA*, 16(2), 456–462. <https://doi.org/10.37630/jpm.v16i2.4227>
- Rangkuti, M. A., Bunawan, W., Agustian, H. Y., Hardianti, T., Wulandari, D., & Hudaihy, N. (2026). Performance of ChatGPT in solving wave problems. *Physical Review Physics Education Research*, 22, 020113. <https://doi.org/10.1103/4dgy-1dgl>
- Roziqin, R., Samsudin, A., Saepuzaman, D., Sari, H. N., & Iryanti, M. (2026). Artificial intelligence in physics education: A systematic review of content coverage, implementation models, learning impact, and pedagogical challenges. *STEM Education*, 6(4), 539–583. <https://doi.org/10.3934/steme.2026023>
- Saputra, A., Sambiri, U., & Hermawan, A. (2025). Exploring artificial intelligence for physics learning in Indonesia: A scoping review. *SAINTIFIK@: Jurnal Pendidikan MIPA*, 10(1). <https://doi.org/10.33387/saintifik.v10i1.10037>
- Sirnoorkar, A., Zollman, D., Lavery, J. T., Magana, A. J., Rebello, N. S., & Bryan, L. A. (2024). Student and AI responses to physics problems examined through the lenses of sensemaking and mechanistic reasoning. *Computers and Education: Artificial Intelligence*, 7, 100318. <https://doi.org/10.1016/j.caeai.2024.100318>
- Wang, K. D., Burkholder, E., Wieman, C., Salehi, S., & Haber, N. (2024). Examining the potential and pitfalls of ChatGPT in science and engineering problem-solving. *Frontiers in Education*, 8, 1330486. <https://doi.org/10.3389/educ.2023.1330486>