Volume 6. No 2. 105–115 OKTOBER 2026

Al-Khazini: **Jurnal Pendidikan Fisika**
http://journal.uin-alauddin.ac.id/index.php/alkhazini
DOI: https://doi.org/10.24252/al-khazini.v6i2.72015

P-ISSN: 2830-3644
e-ISSN: 2829-6699

Comparison of the Abilities of High School Students, Gemini, and Perplexity on an Electromagnetism Concept Understanding Test

Nurul Fatia Itami[1*], I Komang Werdhiana[1], I Wayan Darmadi[1], Rudi Santoso[1]

[1]Program Studi Pendidikan Fisika, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Tadulako

Corresponding Address: nurulfatiaitami@gmail.com

Info Artikel

Riwayat artikel

Dikirim: 18 Agustus 2026
Direvisi: 18 September 2026
Diterima: 16 November 2026
Diterbitkan: 16 November 2026

Kata Kunci:

Elektromagnetik
Gemini
Kecerdasan buatan
Perplexity
Pemahaman konsep

ABSTRAK

Perkembangan kecerdasan buatan (*artificial intelligence*/AI) telah mendorong pemanfaatannya dalam dunia pendidikan, termasuk sebagai pendukung pembelajaran fisika. Penelitian ini bertujuan untuk mengetahui perbedaan performa siswa SMA, Gemini, dan Perplexity dalam menyelesaikan tes pemahaman konsep elektromagnetik. Penelitian ini menggunakan pendekatan kuantitatif dengan desain komparatif. Sampel penelitian terdiri atas 50 siswa SMA Negeri 5 Palu. Sementara itu, Gemini dan Perplexity diberikan instrumen yang sama dan masing-masing model AI diuji sebanyak 50 kali menggunakan 32 butir soal yang sama, sehingga diperoleh 50 data pada kelompok Gemini dan 50 data pada kelompok Perplexity. Dengan demikian, jumlah data pada ketiga kelompok adalah setara, yaitu masing-masing 50 data. Instrumen penelitian berupa 32 butir soal pilihan ganda yang diterjemahkan dari *Conceptual Survey of Electricity and Magnetism* (CSEM) dan telah divalidasi oleh dosen ahli bahasa. Analisis data dilakukan menggunakan statistik deskriptif, uji normalitas Kolmogorov–Smirnov, uji Kruskal–Wallis, dan uji lanjut (*post hoc pairwise comparisons*). Hasil penelitian menunjukkan bahwa terdapat perbedaan performa yang signifikan antara siswa, Gemini, dan Perplexity dalam menyelesaikan tes pemahaman konsep elektromagnetik ($p < 0{,}001$). Gemini memperoleh rata-rata skor tertinggi, diikuti oleh Perplexity, sedangkan siswa memperoleh rata-rata skor terendah. Hasil tersebut menunjukkan bahwa Gemini dan Perplexity memiliki performa penyelesaian tes yang lebih tinggi dibandingkan siswa pada instrumen yang digunakan dalam penelitian ini. Temuan tersebut tidak secara langsung menunjukkan bahwa AI memiliki pemahaman konsep yang lebih baik daripada siswa, karena instrumen penelitian mengukur ketepatan jawaban melalui skor tes dan tidak mengukur proses atau mekanisme pemahaman konsep.

ABSTRACT

The development of Artificial Intelligence (AI) has encouraged its use in education, including as a support for physics learning. This study aims to determine the differences in the abilities of high school students, Gemini, and Perplexity in completing the electromagnetism concept understanding test. This study uses a quantitative approach with a comparative design. The research sample consisted of 50 students of SMA Negeri 5 Palu, while Gemini and Perplexity were given the same instrument with each item tested repeatedly 50 times so that the amount of data in each group was equal. The research instrument was 32 multiple-choice questions translated from the Conceptual Survey of Electricity and Magnetism (CSEM) and validated by a language expert lecturer. Data analysis was carried out using descriptive statistics, the Kolmogorov–Smirnov normality test, the Kruskal–Wallis test, and post hoc pairwise comparisons. The results showed that there were significant differences in the abilities of students, Gemini, and Perplexity in completing the electromagnetism concept understanding test ($p < 0.001$). Gemini achieved the highest average score, followed by Perplexity, while students achieved the lowest average score. Thus, the generative AI models, specifically Gemini and Perplexity, performed better than students in completing the electromagnetism concept understanding test.

© 2026 The Author(s). Published by Physics Education, UIN Alauddin Makassar, Indonesia.

How to cite: Itami, N. F., Werdhiana, I. K., Darmadi, I. W., & Santoso, R. (2026). Comparison of the abilities of high school students, Gemini, and Perplexity on an electromagnetism concept understanding test. *Al-Khazini: Jurnal Pendidikan Fisika*, *6*(2), 105–115. https://doi.org/10.24252/al-khazini.v6i2.72015

(cc) BY-NC

PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah membawa perubahan yang signifikan dalam berbagai bidang kehidupan, termasuk pendidikan. Salah satu perkembangan teknologi yang saat ini banyak dimanfaatkan dalam proses pembelajaran adalah *Artificial Intelligence* (AI). Kehadiran AI membuka peluang baru dalam mendukung proses belajar melalui penyediaan umpan balik secara otomatis, pencarian informasi yang cepat, serta membantu peserta didik dalam memahami materi pembelajaran (Suartama, 2025). Seiring dengan perkembangan tersebut, berbagai platform AI mulai dimanfaatkan sebagai pendukung pembelajaran fisika, di antaranya Gemini dan Perplexity.

Gemini merupakan model AI multimodal yang dikembangkan oleh Google DeepMind dan mampu memproses berbagai jenis data, seperti teks, gambar, audio, video, dan dokumen, sehingga dapat menghasilkan respons yang kontekstual dan relevan (Lee et al., 2023). Sementara itu, Perplexity merupakan platform berbasis *Large Language Models* (LLMs) yang dirancang untuk memberikan jawaban secara cepat dengan menyertakan sumber rujukan sehingga meningkatkan kredibilitas informasi yang disampaikan (Iorliam & Ingio, 2024). Perbedaan fundamental kedua platform ini terletak pada filosofi kerjanya saat merespons soal-soal fisika konseptual. Sebagai model *native multimodal*, Gemini mengintegrasikan pemrosesan teks dan representasi visual (seperti skema gaya, grafik, atau diagram rangkaian) secara serentak untuk membangun representasi fisik suatu fenomena. Sebaliknya, Perplexity beroperasi dengan sistem *Search-Augmented Generation* (SAG/RAG), yang mengombinasikan penalaran bahasa dengan penelusuran basis data internet secara *real-time* untuk menyintesis rujukan ilmiah eksternal dalam menjelaskan konsep fisika. Perbedaan mekanisme internal ini berimplikasi pada cara kedua platform mengonstruksi jawaban fisika konseptual, sehingga keduanya semakin banyak dimanfaatkan sebagai alat bantu belajar.

Meskipun demikian, cara AI menghasilkan jawaban berbeda secara mendasar dengan proses manusia dalam membangun pengetahuan. Pemahaman konsep pada manusia terbentuk melalui proses pembelajaran yang melibatkan pengalaman, interaksi sosial, *sensemaking*, dan *reasoning* sehingga peserta didik mampu menghubungkan berbagai konsep secara bermakna (Sirnoorkar, 2024). Sebaliknya, AI menghasilkan respons berdasarkan analisis pola dari data dalam jumlah besar tanpa melalui pengalaman belajar maupun interaksi sosial sebagaimana yang dialami manusia (Wan, 2024). Perbedaan tersebut menunjukkan bahwa baik pengolahan multimodal Gemini maupun penelusuran teraugmentasi Perplexity, keduanya beroperasi melalui inferensi komputasional, sehingga kemampuan penalarannya belum tentu identik dengan proses pemahaman konsep yang dimiliki peserta didik.

Berbagai penelitian sebelumnya telah mengkaji kemampuan model kecerdasan buatan (Artificial Intelligence/AI) dalam menyelesaikan permasalahan fisika. Azizah dan Mansyur (2024) membandingkan kapabilitas respons ChatGPT dan Gemini terhadap pertanyaan konseptual tentang optik. Hasil penelitian menunjukkan bahwa ChatGPT memberikan respons yang lebih baik dibandingkan Gemini dalam menjawab pertanyaan konseptual pada materi optik.

Kaharu et al. (2026) membandingkan kemampuan dan konsistensi ilmiah ChatGPT dan Gemini dengan peserta didik menggunakan tes dua tingkat pada materi fluida statis. Hasil penelitian menunjukkan bahwa kedua model AI berkinerja lebih baik daripada siswa, meskipun kemampuan dan konsistensi ilmiahnya masih rendah dan di bawah ambang batas masuk. ChatGPT hanya mengungguli Gemini pada skor gabungan Tingkat 1 dan Tingkat 2, dengan

(cc) BY-NC

perbedaan yang tidak signifikan, serta keduanya lebih baik dalam memberikan justifikasi atau penalaran teoretis daripada menjawab pertanyaan faktual.

Lišnić dan Gaurina (2025) menganalisis 15 model *large language models* (LLMs) dalam menyelesaikan soal fisika sekolah menengah. Hasil penelitian menunjukkan adanya perbedaan akurasi antarmodel, dengan ChatGPT o3 memperoleh akurasi tertinggi, sementara soal dengan representasi grafik menjadi salah satu jenis soal yang paling sulit bagi LLM.

Kahaleh dan López Simó (2025) mengevaluasi ChatGPT, Google Gemini, dan Microsoft Copilot dalam menangani miskonsepsi pada mekanika Newton. Hasil penelitian menunjukkan bahwa ChatGPT memperoleh penilaian tertinggi, diikuti Gemini dan Microsoft Copilot, berdasarkan akurasi, personalisasi, dan efektivitas pedagogis. Ketiga model menunjukkan potensi dalam menangani miskonsepsi melalui dialog interaktif, tetapi masih memiliki keterbatasan dalam personalisasi dan penggunaan representasi visual. Secara umum, hasil penelitian-penelitian tersebut menunjukkan bahwa model AI memiliki potensi yang besar dalam mendukung pembelajaran fisika.

Meskipun penelitian terbaru telah banyak mengevaluasi peran kecerdasan buatan dalam pendidikan fisika untuk mempersonalisasi pembelajaran dan memberikan umpan balik terarah (Zawacki-Richter et al., 2019), studi terdahulu umumnya masih terbatas pada komparasi antarmodel AI atau perbandingan AI dengan siswa pada topik fisika yang berbeda. Hingga saat ini, belum ada penelitian yang secara spesifik membandingkan kemampuan siswa SMA, Gemini, dan Perplexity dalam menyelesaikan tes pemahaman konsep elektromagnetik menggunakan instrumen baku yang sama, seperti *Conceptual Survey of Electricity and Magnetism* (CSEM). Oleh karena itu, pengujian komparatif ini menjadi krusial untuk memetakan sejauh mana keandalan penalaran AI dibandingkan dengan proses pemahaman konsep siswa pada materi tersebut.

Kebaruan penelitian ini terletak pada perbandingan langsung performa siswa SMA, Gemini, dan Perplexity dalam menyelesaikan tes pemahaman konsep Elektromagnetik menggunakan instrumen *Conceptual Survey of Electricity and Magnetism* (CSEM) yang sama. Penggunaan instrumen dan kriteria penskoran yang seragam memungkinkan ketiga kelompok menghadapi butir soal yang identik, sehingga perbedaan performa dalam menyelesaikan tes dapat diukur dan dibandingkan secara langsung. Dengan demikian, penelitian ini bertujuan untuk menganalisis perbedaan performa siswa SMA, Gemini, dan Perplexity dalam menyelesaikan tes pemahaman konsep Elektromagnetik. Hasil penelitian ini diharapkan dapat memberikan informasi mengenai perbedaan performa siswa dan model AI dalam menyelesaikan tes fisika serta menjadi bahan pertimbangan bagi pendidik dalam memahami potensi dan keterbatasan model AI sebagai pendukung pembelajaran fisika.

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan desain komparatif. Hardani (2020) menyebutkan bahwa penelitian komparatif merupakan penelitian kuantitatif yang dirancang untuk menilai perbedaan variabel pada beberapa kelompok dengan tujuan menemukan pola perbandingan. Desain ini dipilih karena penelitian bertujuan untuk membandingkan kemampuan tiga kelompok subjek yang berbeda, yaitu siswa SMA, *Gemini*, dan *Perplexity* dalam menyelesaikan tes pemahaman elektromagnetik.

(cc) BY-NC

Penelitian ini dilaksanakan pada bulan April semester genap tahun ajaran 2025/2026 di SMA Negeri 5 Palu yang berlokasi di Jalan R.E. Martadinata, Kelurahan Tondo, Kecamatan Mantikulore, Kota Palu, Provinsi Sulawesi Tengah.

Populasi penelitian adalah seluruh siswa kelas XI Merdeka SMA Negeri 5 Palu pada semester genap tahun ajaran 2025/2026 yang mengikuti mata pelajaran fisika. Sampel penelitian terdiri atas peserta didik kelas XI Merdeka 1 dan XI Merdeka 5 yang dipilih menggunakan teknik *purposive sampling* karena telah mempelajari materi listrik dan magnet.

Teknik pengumpulan data pada peserta didik dilakukan menggunakan metode *testing* dengan memberikan tes pemahaman elektromagnetik berbentuk pilihan ganda sebanyak 32 butir. Jawaban yang diberikan peserta didik digunakan sebagai data untuk menggambarkan kemampuan mereka dalam menyelesaikan tes. Sementara itu, pengujian pada *Gemini* dan *Perplexity* dilakukan menggunakan instrumen yang sama, yaitu tes pemahaman elektromagnetik yang terdiri atas 32 butir soal pilihan ganda. Pengujian pada kedua model AI menggunakan *prompt* yang sama, yaitu "Pilih salah satu jawaban yang paling benar." Setiap butir soal diberikan secara terpisah kepada masing-masing model AI.

Pengujian pada *Gemini* dan *Perplexity* dilakukan sebanyak 50 kali untuk memperoleh jumlah data yang setara dengan jumlah peserta didik dalam penelitian. Pengujian dilakukan pada Februari 2026 menggunakan versi gratis dari masing-masing platform AI dengan model dan pengaturan yang tersedia secara *default* pada saat pengambilan data. Setiap pengulangan dilakukan pada obrolan baru, sehingga percakapan atau konteks dari pengujian sebelumnya tidak digunakan dalam pengujian berikutnya. Percakapan dari pengujian sebelumnya tetap tersimpan dan tidak dihapus. Instrumen, *prompt*, dan prosedur pengujian dibuat sama pada setiap pengulangan. Setiap butir soal diinput secara terpisah ke dalam obrolan baru dan prosedur tersebut diulang sebanyak 50 kali pada masing-masing model AI. Dengan demikian, diperoleh 50 set hasil jawaban dari *Gemini* dan 50 set hasil jawaban dari *Perplexity*, yang kemudian dibandingkan dengan data dari 50 peserta didik.

Seluruh jawaban peserta didik, *Gemini*, dan *Perplexity* dinilai menggunakan kunci jawaban yang sama. Jawaban yang sesuai dengan kunci jawaban diberi skor 1, sedangkan jawaban yang tidak sesuai diberi skor 0. Skor dari 32 butir soal kemudian dijumlahkan untuk memperoleh skor masing-masing peserta didik dan setiap pengulangan pengujian pada *Gemini* dan *Perplexity*.Instrumen yang digunakan dalam penelitian ini adalah *Conceptual Survey of Electricity and Magnetism* (CSEM) yang diterjemahkan dari Maloney et al. (2001) ke dalam bahasa Indonesia tanpa mengubah substansi setiap butir soal.

Berdasarkan hasil validasi ahli bahasa, instrumen penelitian memperoleh persentase kelayakan sebesar 75%. Mengacu pada kriteria kelayakan instrumen menurut Akbar (2015), persentase tersebut termasuk dalam kategori sangat layak. Dengan demikian, instrumen telah memenuhi aspek kejelasan bahasa, ketepatan makna, dan ketepatan penggunaan istilah fisika sehingga layak digunakan sebagai instrumen penelitian.

Data dianalisis menggunakan statistik deskriptif dan statistik inferensial. Statistik deskriptif digunakan untuk memberikan gambaran mengenai karakteristik data pada masing-masing kelompok, yang meliputi nilai rata-rata dan standar deviasi. Nilai rata-rata digunakan untuk mengetahui kecenderungan umum skor kemampuan pemahaman konsep elektromagnetik, sedangkan standar deviasi digunakan untuk mengetahui tingkat penyebaran skor terhadap nilai rata-ratanya. Rata-rata dihitung menggunakan persamaan:

(cc) BY-NC

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$\bar{X} =$ nilai rata-rata

$X_i =$ skor pengamatan ke-i

$n =$ jumlah data

Standar deviasi digunakan untuk mengetahui besarnya variasi atau penyimpangan skor dari nilai rata-rata. Standar deviasi sampel dihitung menggunakan persamaan:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$$

$s =$ standar deviasi

$X_i =$ skor pengamatan ke-i

$\bar{X} =$ nilai rata-rata

$n =$ jumlah data

Analisis inferensial diawali dengan uji normalitas menggunakan Kolmogorov–Smirnov untuk mengetahui apakah data pada setiap kelompok berdistribusi normal (Priyatno, 2018). Uji normalitas digunakan sebagai dasar dalam menentukan teknik analisis statistik yang sesuai. Data dikatakan berdistribusi normal apabila nilai signifikansi lebih besar dari 0,05. Berdasarkan hasil uji normalitas, data pada kelompok yang diteliti tidak seluruhnya berdistribusi normal. Oleh karena itu, analisis perbedaan antarkelompok dilanjutkan menggunakan uji nonparametrik Kruskal–Wallis.

Uji Kruskal–Wallis digunakan untuk mengetahui apakah terdapat perbedaan kemampuan pemahaman konsep elektromagnetik antara tiga kelompok independen, yaitu siswa SMA, Gemini, dan Perplexity. Uji ini dilakukan dengan mengubah seluruh skor pengamatan menjadi peringkat (*rank*), kemudian membandingkan distribusi peringkat antarkelompok. Statistik uji Kruskal–Wallis dihitung menggunakan persamaan:

$$H = \frac{12}{N(N+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(N+1)$$

$H =$ statistik uji Kruskal--Wallis

$k =$ jumlah kelompok

$n_j =$ jumlah data pada kelompok ke-j

HASIL DAN PEMBAHASAN

Analisis Statistik Deskriptif

Analisis statistik deskriptif dilakukan untuk memberikan gambaran umum hasil tes pemahaman konsep elektromagnetik pada kelompok Siswa, Gemini, dan Perplexity. Analisis ini meliputi nilai rata-rata (mean) dan standar deviasi. Nilai rata-rata menunjukkan kecenderungan skor setiap kelompok, sedangkan standar deviasi menggambarkan variasi penyebaran skor. Hasil analisis statistik deskriptif disajikan pada Tabel 1.

(cc) BY-NC

Tabel 1. Hasil Uji Statistik Deskriptif1

Kelompok	N	Rata-rata	Standar Deviasi
Siswa	50	11,88	2,05
Gemini	50	24,82	1,83
Perplexity	50	17,64	3,72

Sumber: hasil analisis data menggunakan SPSS 27

Berdasarkan Tabel 1, *Gemini* memperoleh rata-rata skor tertinggi sebesar 24,82 dengan standar deviasi 1,83. *Perplexity* memperoleh rata-rata skor sebesar 17,64 dengan standar deviasi 3,72, sedangkan siswa memperoleh rata-rata skor terendah sebesar 11,88 dengan standar deviasi 2,05. Variasi respons Perplexity pada 50 kali pengulangan pengujian dapat dikaitkan dengan karakteristik sistemnya yang mengintegrasikan pencarian informasi web secara *real-time* (*live-search*) dengan proses generasi jawaban. Pada setiap pengujian, Perplexity melakukan *retrieval* terhadap informasi dari sumber web yang relevan sebelum menyintesis informasi tersebut menjadi respons. Karena proses *retrieval* berlangsung secara dinamis, konteks informasi yang diperoleh pada setiap pengujian berpotensi tidak identik, baik karena perbedaan hasil pencarian maupun sumber yang ditemukan atau diprioritaskan oleh sistem. Perbedaan konteks tersebut selanjutnya dapat memengaruhi proses sintesis informasi dan formulasi jawaban pada tahap *generation*, sehingga pertanyaan yang sama dapat menghasilkan respons yang berbeda, termasuk pada pilihan jawaban. Temuan penelitian ini menunjukkan bahwa variasi tersebut tercermin pada hasil pengujian Perplexity, di mana respons dan skor tidak selalu konsisten pada 50 pengulangan. Dengan demikian, fluktuasi respons Perplexity tidak hanya merefleksikan kemampuan model dalam memahami konsep elektromagnetik, tetapi juga karakteristik sistem berbasis *live-search* yang menggabungkan proses *retrieval* dan *generation* secara dinamis. Hal ini menunjukkan bahwa evaluasi Perplexity pada tes pemahaman konsep perlu mempertimbangkan variabilitas respons yang muncul dari interaksi antara kemampuan model dan informasi yang diperoleh dari web selama proses pengujian.

Analisis Statistik Inferensial

Uji Normalitas

Uji normalitas dilakukan sebagai langkah awal dalam analisis data untuk mengetahui distribusi data penelitian. Pengujian dilakukan menggunakan metode *Kolmogorov–Smirnov* melalui aplikasi SPSS. Hasil uji normalitas dapat dilihat pada Tabel 2.

Tabel 2. Hasil Uji Normalitas2

No	Kelompok	Sig	Keterangan
1	Siswa	0,066	Terdistribusi normal
2	Gemini	0,003	Tidak terdistribusi normal
3	Perplexity	0,008	Tidak terdistribusi normal

Sumber: hasil analisis data menggunakan SPSS 27

Uji normalitas pada penelitian ini menggunakan metode Kolmogorov–Smirnov untuk mengetahui apakah data setiap kelompok berdistribusi normal. Metode ini dipilih karena masing-masing kelompok terdiri atas 50 sampel. Hasil uji menunjukkan bahwa kelompok Siswa memiliki nilai signifikansi (Asymp. Sig.) sebesar 0,066 (>0,05), sehingga data berdistribusi normal. Sebaliknya, kelompok Gemini dan Perplexity masing-masing memperoleh nilai signifikansi 0,003 dan 0,008 (<0,05), sehingga data keduanya tidak berdistribusi normal. Dengan demikian, hanya kelompok Siswa yang memenuhi asumsi

(cc) BY-NC

normalitas, sedangkan Gemini dan Perplexity tidak. Oleh karena itu, analisis statistik lanjutan menggunakan uji nonparametrik.

Uji Nonparametrik Kruskal–Wallis

Berdasarkan hasil uji normalitas sebelumnya, diketahui bahwa data *Gemini* dan *Perplexity* tidak berdistribusi normal sehingga asumsi normalitas tidak terpenuhi. Oleh karena itu, analisis dilanjutkan menggunakan uji nonparametrik *Kruskal–Wallis* sebagai alternatif dari uji *One Way ANOVA* untuk membandingkan perbedaan antar kelompok. Hasil uji disajikan pada Tabel 3.

Tabel 3. Hasil Uji Nonparametrik Kruskal–Wallis3

No	Statistik uji	Nilai
1	Kruskal–Wallis H	116,882
2	Derajat kebebasan	2
3	Signifikansi (Asymp. Sig.)	<0,0001

Sumber: hasil analisis data menggunakan SPSS 27

Hasil uji *Kruskal–Wallis* menunjukkan nilai *Kruskal–Wallis* H sebesar 116,882 dengan derajat kebebasan (df) = 2 dan nilai signifikansi (Asymp. Sig.) < 0,001. Karena nilai signifikansi lebih kecil dari 0,05, maka H_0 ditolak dan H_1 diterima, sehingga dapat disimpulkan bahwa terdapat perbedaan yang signifikan pada hasil tes pemahaman konsep elektromagnetik antara kelompok Siswa, *Gemini*, dan *Perplexity*. Hasil ini menunjukkan bahwa distribusi skor ketiga kelompok tidak sama, sehingga terdapat paling sedikit satu kelompok yang memiliki perbedaan hasil tes dibandingkan kelompok lainnya.

Uji Post Hoc (Pairwise Comparisons)

Untuk mengetahui kelompok yang memiliki perbedaan kemampuan secara signifikan, dilakukan uji lanjut (*post hoc pairwise comparisons*) setelah uji Kruskal–Wallis. Hasil pengujian tersebut dapat dilihat pada Tabel 4.

Tabel 4. Hasil Uji Post Hoc

No	Kelompok	Test Statistic	Std. Error	Sig	Keputusan
1	Siswa–Perplexity	-46,710	8,672	0,00	Berbeda signifikan
2	Siswa–Gemini	-93,750	8,672	0,00	Berbeda secara signifikan
3	Perplexity–Gemini	47,040	8,672	0,000	Berbeda signifikan

Sumber: hasil analisis data menggunakan SPSS 27

Berdasarkan tabel hasil uji post-hoc pairwise comparisons, setiap pasangan kelompok memiliki beberapa nilai penting yang dapat dijelaskan secara rinci. Pada pasangan Siswa–*Perplexity*, diperoleh nilai *test statistic* sebesar -46,710 dengan *standard error* 8,672 dan *standardized test statistic* -5,387. Nilai signifikansi menunjukkan < 0,001 dengan *adjusted significance* 0,000, yang berarti perbedaan antara kedua kelompok ini sangat signifikan. Tanda negatif pada *test statistic* menunjukkan bahwa nilai kelompok Siswa cenderung lebih rendah dibandingkan *Perplexity.*

Pada pasangan Siswa–*Gemini,* nilai *test statistic* sebesar -93,750 dengan *standard error* 8,672 dan *standardized test statistic* -10,811. Nilai signifikansi 0,000 (dan adjusted 0,000) juga

(cc) BY-NC

menunjukkan perbedaan yang sangat signifikan. Besarnya nilai statistik ini (paling besar secara absolut) menandakan bahwa perbedaan antara Siswa dan *Gemini* merupakan yang paling kuat di antara semua pasangan. Tanda negatif kembali menunjukkan bahwa nilai Siswa lebih rendah dibandingkan *Gemini*.

Sementara itu, pada pasangan *Perplexity*–Gemini, diperoleh nilai *test statistic* sebesar 47,040 dengan *standard error* 8,672 dan *standardized test statistic* 5,425. Nilai signifikansi < 0,001 dan adjusted 0,000 menunjukkan bahwa perbedaan antara kedua kelompok ini juga signifikan. Tanda positif pada nilai statistik menunjukkan bahwa nilai *Gemini* lebih tinggi dibandingkan Perplexity. Secara keseluruhan, semua pasangan memiliki nilai signifikansi di bawah 0,05, sehingga seluruh perbedaan antar kelompok terbukti signifikan.

Hasil penelitian menunjukkan bahwa Gemini cenderung memperoleh tingkat ketepatan jawaban tertinggi pada sebagian besar materi yang diujikan, diikuti oleh Perplexity dan siswa. Tingginya performa Gemini diduga berkaitan dengan kemampuannya dalam mengintegrasikan hubungan matematis, penalaran konseptual, dan keterkaitan antar konsep fisika secara lebih konsisten. *Perplexity* juga menunjukkan kemampuan yang baik, namun masih mengalami beberapa kesalahan pada konsep yang memerlukan penalaran lebih mendalam. Sementara itu, siswa cenderung mengalami kesulitan pada konsep yang melibatkan hubungan matematis serta memerlukan integrasi beberapa konsep fisika secara bersamaan.

Hasil Kruskal–Wallis menunjukkan nilai signifikansi < 0,05 yang menandakan adanya perbedaan kemampuan yang signifikan antara siswa, Gemini, dan *Perplexity* dalam menyelesaikan tes pemahaman konsep elektromagnetik. Perbedaan ini menunjukkan bahwa kemampuan ketiga kelompok dalam memahami konsep elektromagnetik tidak sama. Gemini memperoleh kemampuan tertinggi, diikuti oleh Perplexity, sedangkan siswa memperoleh kemampuan terendah. Temuan penelitian ini sejalan dengan penelitian Kaharu et al. (2026), yang menunjukkan adanya perbedaan kemampuan antara ChatGPT, Gemini, dan kelompok siswa dalam menyelesaikan tes dua tingkat pada materi fluida statis ($p < 0,05$). Hasil uji lanjut menggunakan Games-Howell dan Tukey menunjukkan bahwa ChatGPT dan Gemini memiliki kemampuan yang lebih tinggi dibandingkan kelompok siswa. Hasil tersebut menunjukkan bahwa kedua model AI generatif memiliki kemampuan yang lebih baik dibandingkan siswa dalam menyelesaikan tes fluida statis. Kesamaan temuan tersebut mengindikasikan bahwa model AI generatif cenderung memiliki kemampuan konseptual yang lebih baik dibandingkan siswa pada berbagai materi fisika. Kemampuan tersebut didukung oleh model bahasa besar (*Large Language Models*/LLMs) yang dilatih menggunakan data dalam jumlah sangat besar sehingga mampu mengenali pola dan menghubungkan berbagai konsep.

Gemini menunjukkan performa terbaik dalam penelitian ini. Tingginya perolehan skor Gemini berkaitan dengan kemampuannya dalam memproses informasi, memahami konteks pertanyaan, serta memanfaatkan data pelatihan yang luas untuk menghasilkan respons yang relevan. Temuan ini sejalan dengan penelitian Imran dan Almusharraf (2024) yang menyatakan bahwa Gemini memiliki kemampuan tinggi dalam memahami pertanyaan dan menghasilkan jawaban secara cepat karena didukung oleh teknologi AI generatif dan data pelatihan yang luas. Penelitian Al-Thani et al. (2025) juga menunjukkan bahwa Gemini mampu memberikan performa yang konsisten dalam menjawab soal. Selain itu, penelitian Firdaus et al. (2024) menjelaskan bahwa Gemini memiliki kemampuan yang baik dalam menjawab soal Higher Order Thinking Skills (HOTS) karena mampu menghasilkan jawaban yang terstruktur dan logis. Temuan-temuan tersebut mendukung hasil penelitian ini yang menunjukkan bahwa

(cc) BY-NC

Gemini mampu memberikan tingkat ketepatan jawaban yang lebih tinggi dibandingkan Perplexity dan siswa.

Di sisi lain, Perplexity memperoleh rata-rata skor yang lebih rendah dibandingkan Gemini, tetapi lebih tinggi dibandingkan kelompok siswa. Hasil tersebut menunjukkan bahwa Perplexity juga memiliki kemampuan yang cukup baik dalam mengidentifikasi jawaban yang benar pada soal-soal yang diujikan. Kemampuan tersebut diduga berkaitan dengan karakteristik Perplexity yang menggabungkan teknologi kecerdasan buatan dengan sistem pencarian informasi dari berbagai sumber sehingga mampu menghasilkan respons yang relevan dan didukung referensi. Temuan ini sejalan dengan penelitian Lubis (2024) yang menunjukkan bahwa Perplexity mampu membantu pengguna memperoleh informasi secara cepat dan efisien melalui penyajian informasi yang relevan. Gravina et al. (2024) juga melaporkan bahwa Perplexity memiliki performa yang cukup baik dalam menjawab pertanyaan akademik dan menunjukkan potensi sebagai alat bantu pembelajaran berbasis kecerdasan buatan. Selain itu, Ardiansyah et al. (2024) menjelaskan bahwa Perplexity memiliki tingkat akurasi dan koherensi yang baik dalam menghasilkan respons sehingga dapat dimanfaatkan sebagai sumber belajar tambahan. Meskipun demikian, skor Perplexity yang masih berada di bawah Gemini menunjukkan bahwa kemampuan model ini dalam memahami dan menjawab soal elektromagnetik belum seoptimal Gemini. Perbedaan tersebut dapat disebabkan oleh variasi kemampuan kedua model dalam memahami konteks soal, mengolah informasi, serta menghasilkan jawaban yang sesuai dengan tuntutan konsep fisika yang diujikan.

Sementara itu, rendahnya skor yang diperoleh kelompok siswa menunjukkan bahwa pemahaman konsep elektromagnetik pada beberapa materi yang diujikan masih belum optimal. Hal tersebut terlihat dari jumlah jawaban benar siswa yang lebih sedikit dibandingkan Gemini dan Perplexity. Kondisi ini menunjukkan bahwa sebagian siswa masih mengalami kesulitan dalam memahami dan menerapkan konsep-konsep elektromagnetik ketika menyelesaikan soal yang menuntut pemahaman konseptual. Selain itu, Maloney et al. (2001) menyatakan bahwa konsep listrik dan magnet merupakan materi yang sering menimbulkan kesulitan bagi siswa karena menuntut pemahaman konsep, bukan sekadar menghafal rumus atau melakukan perhitungan matematis. Oleh karena itu, siswa dapat mengalami kesulitan dalam menginterpretasikan situasi pada soal dan menentukan konsep fisika yang tepat untuk menyelesaikannya. Temuan ini sejalan dengan Raduta (2005) yang menjelaskan bahwa konsep-konsep elektromagnetik saling berkaitan sehingga memerlukan pemahaman konseptual yang baik dalam menghubungkan berbagai konsep untuk menyelesaikan suatu permasalahan.

Analisis pada setiap butir soal menunjukkan bahwa siswa paling mengalami kesulitan pada soal nomor 15, dengan persentase jawaban benar sebesar 20%, diikuti soal nomor 10 dan 14 sebesar 22%, soal nomor 17 sebesar 24%, dan soal nomor 6 sebesar 26%. Kondisi tersebut menunjukkan bahwa kesulitan siswa terutama terdapat pada soal yang menuntut pemahaman hubungan antara medan listrik, gaya listrik, potensial listrik, serta penentuan arah besaran fisika. Pada kelompok Gemini, meskipun memperoleh skor keseluruhan tertinggi, masih ditemukan kesalahan pada soal nomor 15 dan 28, dengan persentase jawaban benar masing-masing sebesar 4% dan 16%. Sementara itu, Perplexity menunjukkan persentase jawaban benar terendah pada soal nomor 5, 15, dan 18, masing-masing sebesar 22%, serta soal nomor 6 sebesar 26%. Menariknya, pada soal nomor 5 siswa memperoleh persentase jawaban benar sebesar 46%, lebih tinggi daripada Gemini sebesar 36% dan Perplexity sebesar 22%, sedangkan pada soal nomor 6 siswa dan Perplexity sama-sama memperoleh 26% dan Gemini 22%. Selain itu, soal nomor 15 menjadi salah satu butir yang sulit bagi ketiga kelompok, dengan persentase jawaban benar sebesar 20% pada siswa, 4% pada Gemini, dan 22% pada Perplexity. Temuan

(cc) BY-NC

tersebut menunjukkan bahwa meskipun Gemini dan Perplexity memperoleh skor rata-rata lebih tinggi, keduanya tidak selalu lebih unggul pada setiap butir soal. Dengan demikian, perbedaan kemampuan ketiga kelompok tidak hanya terlihat dari skor keseluruhan, tetapi juga dari pola kesalahan pada konsep elektromagnetik tertentu.

Berdasarkan hasil pengujian, Gemini tampak lebih konsisten dalam menjawab soal-soal yang membutuhkan pemahaman konsep dan hubungan antar konsep fisika. Sementara itu, Perplexity sudah mampu memberikan jawaban yang cukup baik, namun pada beberapa soal masih kurang tepat dibandingkan Gemini. Hal inilah yang diduga menyebabkan skor Perplexity berada di bawah Gemini. Dengan demikian, hasil penelitian ini menunjukkan bahwa Gemini dan Perplexity memiliki performa yang lebih konsisten dibandingkan siswa dalam menyelesaikan soal pemahaman elektromagnetik, dengan Gemini menunjukkan performa terbaik di antara ketiga kelompok yang dibandingkan. Hasil tersebut sejalan dengan analisis statistik menggunakan uji Kruskal–Wallis yang menunjukkan nilai signifikansi < 0,05. Oleh karena itu, H_0 ditolak dan H_1 diterima, yang mengindikasikan adanya perbedaan kemampuan yang signifikan antara siswa, Gemini, dan Perplexity dalam menyelesaikan tes pemahaman konsep elektromagnetik.

KESIMPULAN

Berdasarkan hasil penelitian, terdapat perbedaan kemampuan yang signifikan antara siswa, Gemini, dan Perplexity dalam menyelesaikan tes pemahaman elektromagnetik. Hasil uji Kruskal–Wallis menunjukkan data ketiga kelompok berbeda secara signifikan. Dari hasil nilai yang diperoleh, Gemini memiliki kemampuan tertinggi dibandingkan Perplexity dan siswa dalam menjawab soal pemahaman elektromagnetik.

DAFTAR PUSTAKA

Akbar, S. (2015). *Instrumen perangkat pembelajaran*. Remaja Rosdakarya.

Al-Thani, S. N., Al-Kuwari, M. G., Al-Malki, A. M., et al. (2025). Comparative performance of ChatGPT, Gemini, and final-year emergency medicine clerkship students in answering multiple-choice questions. *International Journal of Emergency Medicine*, *18*(1), 1–10. https://doi.org/10.1186/s12245-025-00949-6

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Ardiansyah, A., Abdin, M., & Salis, M. R. (2024). Analisis komparatif generative artificial intelligence: ChatGPT, Gemini, dan Perplexity sebagai media pembelajaran akuntansi. *Jurnal Teknodik*, *28*(2), 95–112. https://doi.org/10.32550/teknodik.vi.1192

Azizah, M. A., & Mansyur, J. (2024). Perbandingan kapabilitas respons ChatGPT dan Gemini terhadap pertanyaan konseptual tentang optik. *JPFT (Jurnal Pendidikan Fisika Tadulako Online)*, *12*(1), 15–25. https://doi.org/10.22487/jpft.v12i1.3510

Firdaus, T., Sholeha, S. A., Jannah, M., & Setiawan, A. R. (2024). Comparison of ChatGPT and Gemini AI in answering higher-order thinking skill biology questions: Accuracy and evaluation. *International Journal of Science Education and Teaching*, *3*(3), 126–138. https://doi.org/10.14456/ijset.2024.11

Gravina, A. G., Pellegrino, R., Palladino, G., Imperio, G., Ventura, A., & Federico, A. (2024). Charting new AI education in gastroenterology: Cross-sectional evaluation of ChatGPT

(cc) BY-NC

and Perplexity AI in medical residency exam. *Digestive and Liver Disease*, *56*(8), 1304–1311. https://doi.org/10.1016/j.dld.2024.02.019

Hardani. (2020). *Metode penelitian kualitatif & kuantitatif*. Pustaka Ilmu.

Iorliam, A., & Ingio, J. A. (2024). A comparative analysis of generative artificial intelligence tools for natural language processing. *Journal of Computing Theories and Applications*, *1*(3), 311–325. https://doi.org/10.62411/jcta.9447

Kahaleh, R., & López Simó, V. (2025). Evaluating large language models in high school physics education: Addressing misconceptions and fostering conceptual understanding. *Physics Education*, *60*(2), Article 025013. https://doi.org/10.1088/1361-6552/adb235

Kaharu, S. N., Werdhiana, I. K., & Mansyur, J. (2026). Comparing ChatGPT and Gemini on a two-tier static fluid test: Capability and scientific consistency. *European Journal of Educational Research*, *15*(1), 223–250. https://doi.org/10.12973/eu-jer.15.1.223

Lišnić, B., & Gaurina, M. (2025). Large language models in physics: Analysis of accuracy and teacher perception. *Interdisciplinary Description of Complex Systems*, *23*(6), 668–684. https://doi.org/10.7906/indecs.23.6.5

Lubis, C. (2024). Perplexity AI on the writing efficiency of EFL students in higher education: Students' insights. *English Franca: Academic Journal of English Language and Education*, *8*(1), 167–178.

Maloney, D. P., O'Kuma, T. L., Hieggelke, C. J., & Van Heuvelen, A. (2001). Surveying students' conceptual knowledge of electricity and magnetism. *American Journal of Physics*, *69*(S1), S12–S23. https://doi.org/10.1119/1.1371296

Priyatno, D. (2018). *Panduan praktis olah data menggunakan SPSS*. Andi Offset.

Sirnoorkar, P. (2024). Learning, reasoning, and sensemaking in human-AI collaborative physics education. *Journal of Science Education and Technology*, *33*(1), 1–18.

Suartama, I. (2025). Inovasi pembelajaran fisika berbasis AI dan simulasi interaktif. *Jurnal Inovasi Pendidikan*, *2*(1), 50–70.

Wan, J. (2024). Membangun pengetahuan sintetik: Analisis AI dalam pendidikan fisika. *Pendidikan dan Teknologi*, *11*(4), 80–95.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – Where are the educators? *International Journal of Educational Technology in Higher Education*, *16*(1), Article 39. https://doi.org/10.1186/s41239-019-0171-0

(cc) BY-NC