

PERBANDINGAN KAPABILITAS RESPONS ChatGPT DAN GEMINI TERHADAP PERTANYAAN KONSEPTUAL TENTANG OPTIK

The Comparison of ChatGPT and Gemini Response Capabilities to Conceptual Questions of Optics

Mita Aulia Azizah, Jusman Mansyur

Program Studi Pendidikan Fisika, Fakultas Keguruan dan Ilmu Pendidikan
Universitas Tadulako, Palu, Indonesia
mita05aulia@gmail.com

Kata Kunci

ChatGPT, Gemini,
Pertanyaan Konseptual,
Optik

Abstrak

Penelitian ini bertujuan untuk membandingkan kapabilitas respons antara ChatGPT dan Gemini terhadap pertanyaan konseptual tentang optik. Metode yang digunakan adalah penelitian kualitatif dengan pendekatan studi kasus komparatif. Subjek penelitian ini adalah ChatGPT dan Gemini yang juga bertindak sebagai responden. Instrumen yang digunakan adalah 25 butir pertanyaan esai tentang optik. Penilaian dilakukan berdasarkan empat kriteria utama: Kejelasan Konseptual, Keakuratan Konseptual, Pemahaman Konseptual, dan Konsistensi dalam Penjelasan. Hasil penelitian menunjukkan bahwa ChatGPT lebih unggul dibandingkan Gemini dalam semua kriteria yang dinilai. ChatGPT memiliki kapabilitas yang lebih baik dalam memberikan jawaban yang jelas, akurat, dan konsisten terhadap pertanyaan konseptual tentang optik dibandingkan dengan Gemini. Di sisi lain, Gemini cenderung memberikan jawaban yang kurang akurat dan kurang mendalam, serta memiliki variasi dalam konsistensi penjelasannya.

Keywords

ChatGPT, Gemini,
Conceptual Questions,
Optics

Abstract

This research aims to compare the response capabilities of ChatGPT and Gemini to conceptual questions about optics. The method used is qualitative research with a comparative case study approach. The subjects of this research are ChatGPT and Gemini, which also act as respondents. The instrument used is 25 essay questions about optics. The assessment is based on four main criteria: Conceptual Clarity, Conceptual Accuracy, Conceptual Understanding, and Consistency in Explanation. The results show that ChatGPT outperforms Gemini in all assessed criteria. ChatGPT has a better capability in providing clear, accurate, and consistent answers to conceptual questions about optics compared to Gemini. On the other hand, Gemini tends to provide less accurate and less in-depth answers, with variation in the consistency of its explanations.

©2024 The Author
p-ISSN 2338-3240
e-ISSN 2580-5924

Received 08/02/2024; Revised 26/02/2024; Accepted 13/03/2024; Available Online 30/04/2024

*Corresponding Author: fisika@yahoo.co.id

PENDAHULUAN

Kemajuan bidang *Artificial Intelligence* (AI) telah menghasilkan berbagai *Large Language Model* (LLM) yang mampu menjawab pertanyaan dengan tingkat keakuratan yang cukup tinggi. Salah satu model yang populer adalah ChatGPT (*Generative Pre-trained Transformer*) yang dikembangkan oleh OpenAI [1]. Model ini telah menunjukkan kemampuan yang mengesankan dalam memahami dan menghasilkan teks yang berkualitas. Selain itu, ada pula platform pertukaran informasi berbasis AI seperti Gemini yang memiliki kemampuan serupa dalam merespons pertanyaan dan menyediakan informasi secara cepat [2].

Studi mengenai penerapan ChatGPT dalam konteks pendidikan semakin menjadi perhatian utama. Dalam sebuah tinjauan oleh Lo [3] yang melibatkan analisis terhadap 50 artikel, telah ditemukan beragam penilaian kinerja ChatGPT di berbagai mata pelajaran dan potensi manfaatnya sebagai instruktur dan tutor virtual bagi siswa. Penelitian Kortemeyer [4] juga mengeksplorasi bagaimana ChatGPT-3.5 dapat digunakan dalam menilai dan memberikan umpan balik terhadap pekerjaan siswa. Vasconcelos & dos Santos [5] menyelidiki ChatGPT dan Bing Chat sebagai objek pemikiran dalam pembelajaran STEM. Selanjutnya, dos Santos [6] melakukan perbandingan dan analisis terhadap respons empat *chatbots* bertenaga *AI Generatif* (GenAIbots), antara lain ChatGPT-3.5,

ChatGPT-4, Bing Chat, dan Bard pada permasalahan fisika dalam kerangka teori konstruktivis. Di samping itu, penelitian Küchemann *et al.* [7] menyoroti potensi ChatGPT-3.5 dalam mendukung pengembangan tugas fisika bagi calon guru. Penelitian-penelitian tersebut menunjukkan bahwa ChatGPT memberikan potensi manfaat terhadap pembelajaran fisika. Namun, terdapat kekhawatiran terkait penggunaan ChatGPT dalam pembelajaran, misalnya penyebaran informasi yang tidak akurat atau palsu, serta potensi plagiarisme.

Para peneliti telah mengkaji bagaimana performa berbagai versi ChatGPT dalam berbagai penilaian konseptual fisika, seperti *Force Concept Inventory* (FCI) [8], tugas pengantar fisika [9], kursus pengantar fisika universitas [10], pemahaman grafik dalam kinematika [11], penulisan esai topik sejarah dan filsafat fisika [12], dan pertanyaan fisika dasar dengan dialog gaya Sokrates [13], serta dalam mengevaluasi bagaimana persepsi siswa terhadap kualitas linguistik dan akurasi ilmiah dari respons ChatGPT terhadap pertanyaan pemahaman fisika [14]. Sejumlah penelitian tersebut menemukan bahwa performa ChatGPT bervariasi, dari tidak meyakinkan dan bermasalah, sebanding dengan siswa yang berkinerja buruk, hingga sangat baik atau bahkan seperti ahli.

Di sisi lain, terdapat Gemini yang dapat menyampaikan informasi faktual karena integrasinya yang lancar dengan *Google Search*. Dalam laporan Gemini Team *et al.* [15], Google menunjukkan bahwa Gemini memiliki perbedaan dengan ChatGPT. Gemini mampu mengenali semua konten tulisan tangan dengan benar dan memverifikasi alasannya dalam menanggapi solusi siswa untuk masalah fisika. Penelitian tentang Gemini telah dilakukan oleh sejumlah peneliti, seperti penelitian Nyaaba [16] yang membandingkan pemahaman manusia dan AI (GPT-4 dan Gemini) terhadap hakikat sains. Rane *et al.* [17] menyajikan pemeriksaan komparatif mendalam antara Gemini dan ChatGPT, dan mengeksplorasi metrik kinerja, varian arsitektur, dan kemampuannya. Selain itu, Perera & Lankathilake [18] juga menggali berbagai perspektif mengenai penggunaan ChatGPT dan Gemini di masa depan untuk membuat rekomendasi komprehensif dalam pendidikan. Secara keseluruhan, artikel ini memberikan wawasan berharga mengenai peran ChatGPT dan Gemini dalam pendidikan dan menyoroti perlunya penelitian dan eksplorasi lebih lanjut di bidang ini.

Pendidikan fisika telah menghadapi tantangan dalam mengatasi miskonsepsi dan

kesalahpahaman siswa terkait konsep-konsep fisika, seperti konsep hukum ketiga Newton [19], momentum [20], rangkaian listrik [21], elektrostatika [22], interferensi cahaya [23], gelombang optik [24], dan optik [25]. Memahami konsep-konsep fisika termasuk konsep optik membutuhkan penjelasan yang tepat dan terperinci. Kehadiran AI seperti ChatGPT dan Gemini menawarkan potensi signifikan dalam memberikan respons terhadap pertanyaan konseptual tentang optik. Dengan kemampuan untuk memproses dan menghasilkan teks secara efisien, kedua platform ini dapat membantu dalam memahami konsep-konsep optik dengan lebih baik melalui jawaban yang relevan dan mudah dipahami. Selain itu, untuk mengoptimalkan pemanfaatan AI, penting untuk memahami kekuatan dan kelemahannya, melatih dan berkolaborasi dengan teknologi tersebut, dan membedakan kegunaan dan gangguannya [26].

Berdasarkan paparan di atas, penelitian terbaru telah mengeksplorasi penerapan ChatGPT dan Gemini dalam pendidikan, khususnya pembelajaran fisika. Namun, belum ada penelitian yang mengkaji perbandingan kapabilitas ChatGPT dan Gemini dalam memberikan respons terhadap pertanyaan konseptual mengenai optik. Oleh karena itu, peneliti tertarik melakukan penelitian perbandingan kapabilitas respons ChatGPT dan Gemini terhadap pertanyaan konseptual tentang optik.

METODOLOGI PENELITIAN

Jenis penelitian yang digunakan pada penelitian ini yaitu penelitian kualitatif dengan pendekatan studi kasus komparatif untuk membandingkan kapabilitas respons ChatGPT dan Gemini terhadap pertanyaan konseptual tentang optik.

Penelitian dilaksanakan di Program Studi Pendidikan Fisika Universitas Tadulako pada semester genap tahun ajaran 2023/2024. Subjek dalam penelitian ini adalah ChatGPT dan Gemini. Fitur unik yang dimiliki oleh model-model AI seperti ChatGPT dan Gemini menjadikannya subjek yang menarik untuk dianalisis secara komparatif dalam konteks perkembangan teknologi AI [17]. Setiap model AI diuji dengan 25 pertanyaan yang sama, sehingga menghasilkan total 50 respons (25 dari ChatGPT dan 25 dari Gemini). ChatGPT dan Gemini bertindak sebagai responden yang memberikan jawaban terhadap setiap pertanyaan yang diajukan.

Jenis data yang digunakan dalam penelitian adalah data primer, artinya data yang diperoleh

merupakan data yang bersumber langsung dari subjek penelitian. Sumber data penelitian ini adalah data hasil pengerjaan pertanyaan konseptual oleh ChatGPT dan Gemini.

Instrumen penelitian merupakan alat atau fasilitas yang digunakan penelitian dalam mengumpulkan data agar pekerjaannya lebih mudah dan hasilnya lebih baik, sehingga mudah diolah [27]. Instrumen yang digunakan dalam penelitian ini adalah instrumen tes pertanyaan konseptual tentang optik sebanyak 25 butir pertanyaan esai dalam bahasa Indonesia untuk memperoleh data respons ChatGPT dan Gemini. Selain pertanyaan esai, instrumen lain yang digunakan adalah rubrik penilaian. Rubrik penilaian menentukan kriteria yang digunakan untuk memberikan skor pada respons, seperti kejelasan konseptual, keakuratan konseptual, pemahaman konseptual, dan konsistensi dalam penjelasan.

Instrumen yang digunakan pada penelitian ini melalui uji validitas dengan teknik pengujian validitas konstruk (*construct validity*). Validitas konstruk (*construct validity*) adalah validitas berdasarkan penilaian para ahli (*judgment experts*) [28]. Pada penelitian ini, tahap pengujian validitas konstruk dilakukan oleh dosen yang sebelumnya telah ditetapkan sebagai validator ahli. Dalam proses validasi, bagian penting yang perlu divalidasi mencakup jumlah soal dengan konteks yang sama dan bahasa instrumen, serta *prompt* (instruksi) yang diberikan jelas dan sesuai dengan tujuan penelitian.

Teknik pengumpulan data dalam penelitian ini adalah observasi yang digunakan untuk memantau dan menilai respons yang diberikan oleh ChatGPT.3-5 dan Gemini 1.0 Pro terhadap pertanyaan konseptual tentang optik. Setiap item pertanyaan selalu diunggah ke percakapan baru untuk menghindari ChatGPT dan Gemini mendasarkan jawabannya pada teks yang dihasilkannya sendiri sebagai jawaban atas pertanyaan sebelumnya. Seluruh pengujian dikirimkan ke ChatGPT dan Gemini sebanyak 25 obrolan terpisah. Setelah respons dikumpulkan, data dinilai menggunakan sistem perskoran yang disesuaikan dengan rubrik penilaian untuk memastikan bahwa setiap respons dievaluasi sesuai dengan kriteria yang sama.

Analisis data dalam penelitian ini dilakukan dengan tiga aktivitas yaitu, mereduksi data, penyajian data, dan penarikan kesimpulan. Reduksi data adalah proses mengolah, menyederhanakan, dan memilih data yang paling relevan untuk analisis lebih lanjut. Reduksi data dalam penelitian ini dilakukan dengan mengidentifikasi respons ChatGPT dan Gemini terhadap pertanyaan konseptual tentang

optik. Selanjutnya, dilakukan penilaian respons menggunakan sistem penskoran yang disesuaikan dengan rubrik penilaian. Rubrik penilaian terdiri dari empat kriteria utama: kejelasan konseptual, keakuratan konseptual, pemahaman konseptual, dan konsistensi dalam penjelasan. Setiap kriteria dinilai pada skala 1 hingga 5, di mana 1 menunjukkan kualitas yang sangat rendah dan 5 menunjukkan kualitas yang sangat tinggi. Setiap pertanyaan dinilai berdasarkan keempat kriteria ini, dengan skor maksimal yang dapat diperoleh untuk setiap pertanyaan adalah 20 (5 untuk masing-masing kriteria). Skor tersebut kemudian dinormalisasi ke dalam rentang 0-100 dengan cara sebagai berikut.

$$\text{Nilai} = \frac{\text{Jumlah Skor}}{\text{Jumlah Skor Maksimum}} \times 100$$

Selanjutnya menghitung persentase setiap kriteria dilakukan dengan cara sebagai berikut.

$$\text{Persentase} = \frac{\text{Rata-rata Skor Setiap Kriteria}}{\text{Skor Maksimum Kriteria}} \times 100\%$$

Dari hasil tersebut digunakan untuk mengelompokkan persentase hasil respons berdasarkan kriteria dengan menggunakan Tabel 1.

Tabel 1 Data Hasil Penelitian

No	Kategori	Persentase (%)
1	Sangat Kurang Baik	0-20
2	Kurang Baik	21-40
3	Cukup Baik	41-60
4	Baik	61-80
5	Sangat Baik	81-100

Selanjutnya, tahap penyajian data adalah tahap di mana data yang telah direduksi diorganisir dan disusun dalam bentuk yang mudah dipahami. Pada penelitian ini, peneliti menyajikan data yang telah direduksi dan dianalisis dalam bentuk tabel dan grafik batang yang mengilustrasikan perbandingan respons antara ChatGPT dan Gemini.

Terakhir, tahap penarikan kesimpulan dilakukan terhadap data yang telah direduksi dan disajikan. Penarikan kesimpulan melibatkan evaluasi perbandingan antara kapabilitas respons ChatGPT dan Gemini, termasuk kelebihan dan kekurangan masing-masing platform dalam menanggapi pertanyaan konseptual tentang optik.

HASIL DAN PEMBAHASAN

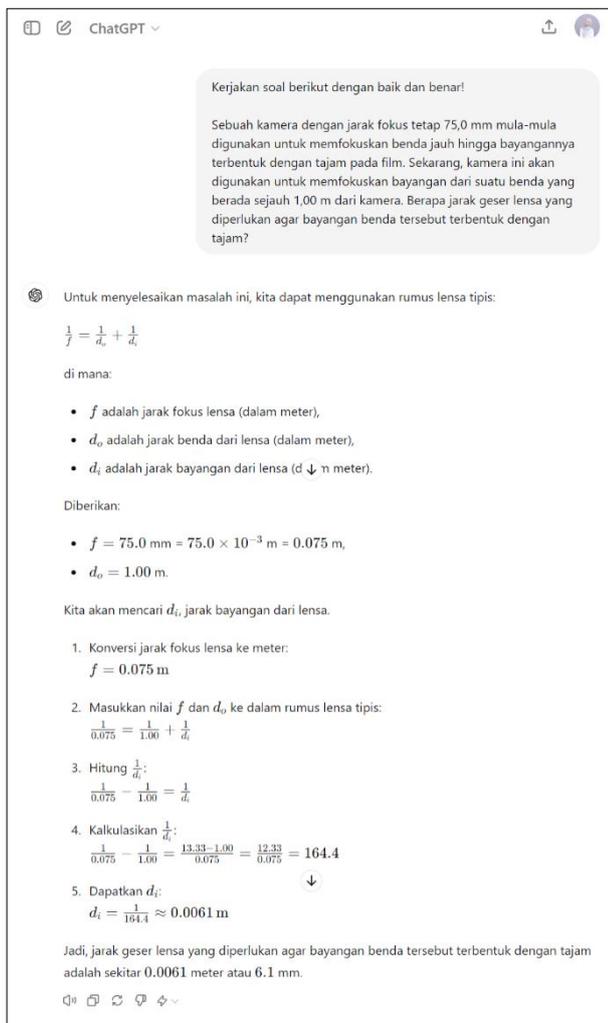
Hasil Penelitian

a. Hasil Data Penelitian

a) Kejelasan Konseptual

Penilaian kejelasan konseptual mencakup seberapa jelas dan mudah dipahami jawaban yang diberikan oleh masing-masing AI dalam menjelaskan konsep-konsep optik. Penilaian ini penting untuk memahami kemampuan setiap AI dalam menyampaikan konsep-konsep yang kompleks dengan cara yang dapat dipahami oleh pengguna. Hal ini dapat mempengaruhi seberapa efektif AI tersebut dalam memberikan bantuan dan penjelasan terkait topik optik.

Berikut disajikan contoh respons ChatGPT pada Gambar 1.

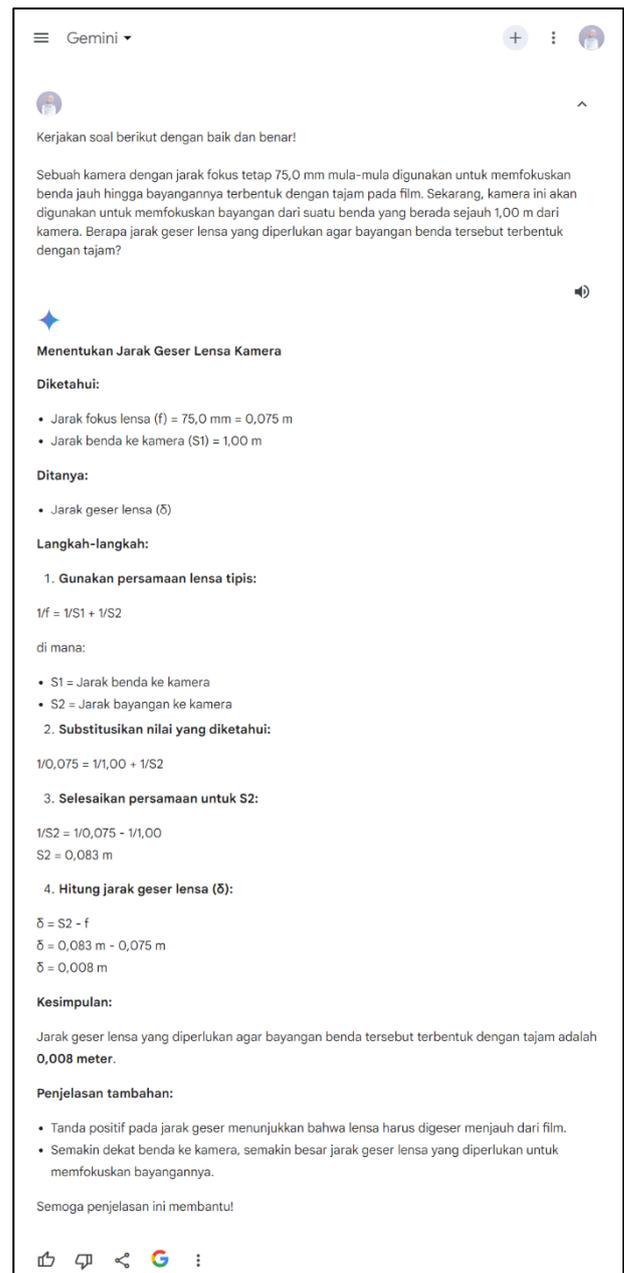


Gambar 1 Hasil Respons R1-17

Berdasarkan Gambar 1 menunjukkan bahwa respons R1-17 memberikan penjelasan yang lengkap dan mendalam atas pertanyaan tentang jarak fokus lensa kamera. Penjelasan R1-17 dimulai dengan mengkonversi jarak fokus lensa ke dalam meter dan memasukkan nilai f dan d_o ke dalam rumus lensa tipis. Setiap langkah

dijelaskan secara rinci, termasuk perhitungan $\frac{1}{d_i}$ dan penyelesaian jarak bayangan dari lensa d_i . Penggunaan rumus lensa tipis digunakan dengan tepat dan setiap langkah perhitungannya dijelaskan dengan jelas. Format penulisan rumus R1-17 juga terstruktur dan mudah diikuti. Hal ini menunjukkan ChatGPT cenderung memberikan penjelasan yang lebih mendalam dan mudah dipahami dalam menjelaskan konsep-konsep optik.

Hal berbeda ditunjukkan dengan Gemini yang meskipun penjelasannya cukup jelas, terkadang ada bagian yang kurang diterangkan atau membingungkan. Berikut contoh respons Gemini disajikan pada Gambar 2.



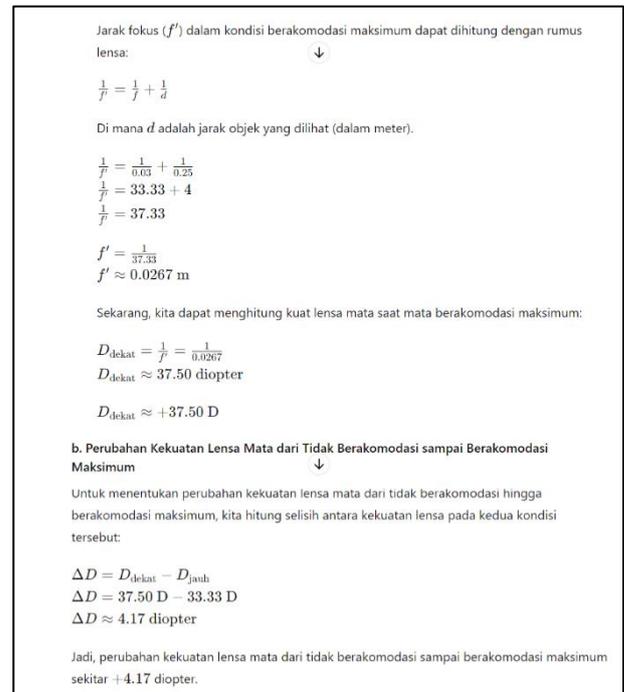
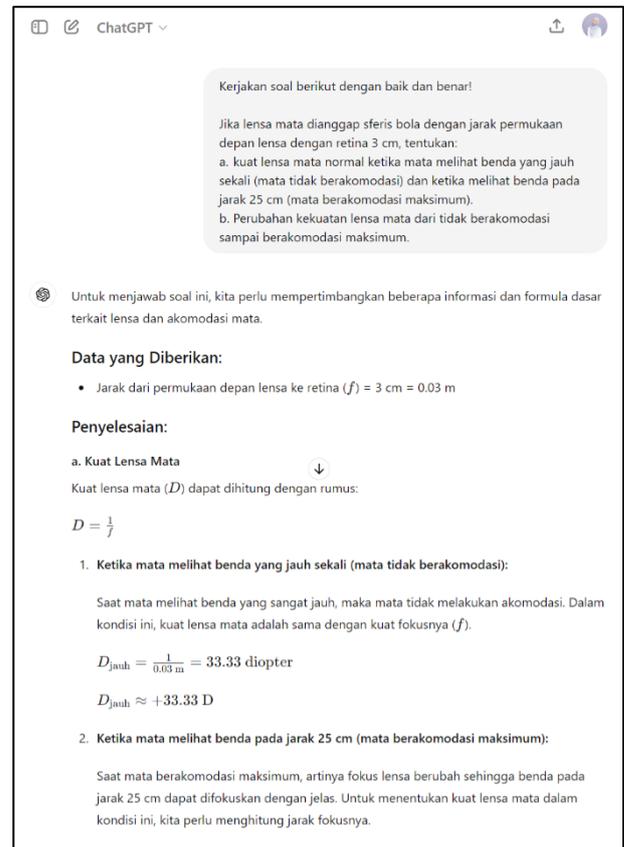
Gambar 2 Hasil Respons R2-17

Pada Gambar 2 dapat dilihat bahwa respons R2-17 menunjukkan penggunaan rumus lensa tipis dan penyelesaian rumus untuk mendapatkan S_2 . Langkah-langkah dijelaskan dengan cukup sederhana dan dihasilkan dengan menghitung jarak geser lensa δ dengan selisih antara S_2 dan f . Namun, tidak ada perhitungan terperinci yang menunjukkan langkah-langkah lengkap seperti pada respons R1-17. Selain itu, penyajian rumus oleh R2-17 ditampilkan secara panjang mendatar yang dapat mengurangi kejelasan jawaban. Hal ini memberikan gambaran bahwa Gemini cenderung memberikan jawaban yang kurang mendalam dan kurang terstruktur, sehingga dapat mempengaruhi kejelasan konseptualnya.

b) Keakuratan Konseptual

Keakuratan konseptual adalah aspek penting karena mempengaruhi kepercayaan pengguna terhadap informasi yang disediakan oleh AI. Penilaian keakuratan konseptual mencakup seberapa tepat dan benar jawaban yang diberikan oleh masing-masing AI dalam menjelaskan konsep-konsep optik. Jawaban yang akurat menunjukkan bahwa AI mampu memahami dan menyampaikan informasi ilmiah dengan benar, menghindari kesalahan yang dapat menyesatkan pengguna. ChatGPT mampu menunjukkan tingkat akurasi yang tinggi dalam memahami dan menerapkan konsep-konsep optik. Meskipun ChatGPT lebih baik dalam akurasi, namun masih terdapat kesalahan kecil pada beberapa responnya. Adapun pada beberapa respons yang dihasilkan Gemini cenderung menunjukkan konsep yang kurang akurat atau kurang sesuai dengan konsep yang ditanyakan.

Berikut disajikan contoh respons ChatGPT pada Gambar 3.

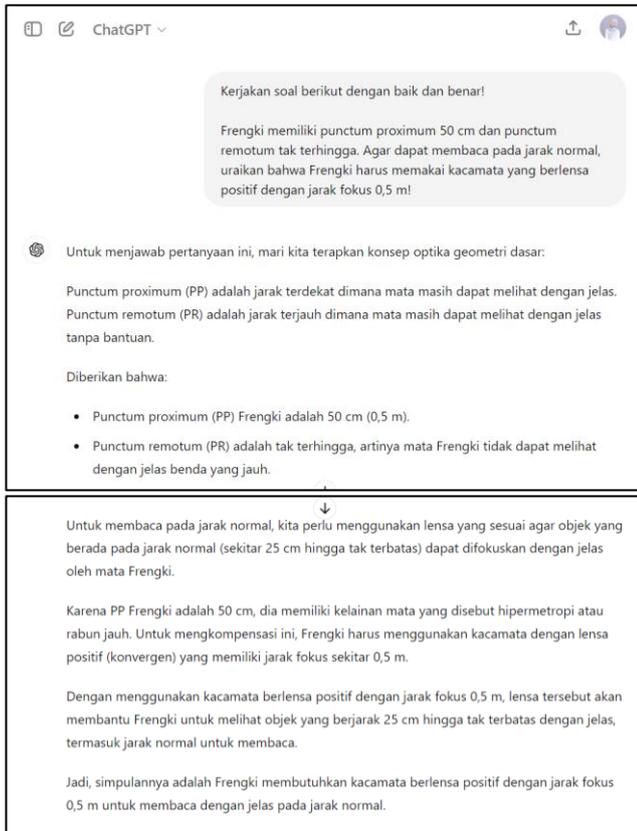


Gambar 3 Hasil Respons R1-10

Berdasarkan Gambar 3, jawaban R1-10 menunjukkan keakuratan konseptual yang tinggi dengan melakukan konversi satuan dari sentimeter ke meter sebelum melakukan perhitungan. Dalam konteks optik, jarak fokus (f) harus dinyatakan dalam meter untuk menggunakan rumus $D = \frac{1}{f}$ dengan benar, karena kekuatan lensa (D) diukur dalam dioptri

yang secara konvensional menggunakan satuan meter. Hal ini membuktikan ChatGPT mampu menunjukkan tingkat akurasi yang tinggi dalam memahami dan menerapkan konsep-konsep optik.

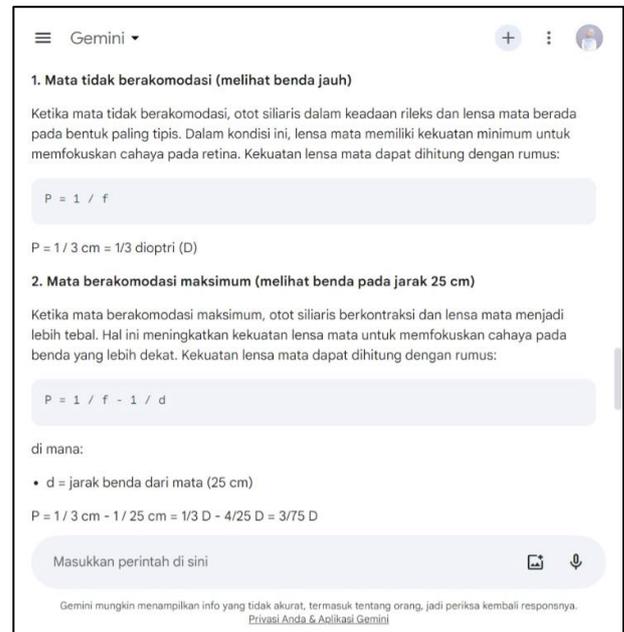
Meskipun ChatGPT lebih baik dalam akurasi, namun masih terdapat kesalahan kecil pada beberapa responnya. Misalnya, respons R1-10 yang ditunjukkan pada Gambar 4.



Gambar 4 Hasil Respons R1-14

Berdasarkan Gambar 4 terlihat dalam respons R1-14 terdapat kesalahan dengan menyebutkan bahwa kelainan mata *punctum proximum* 50 cm berarti tidak dapat melihat dengan jelas benda yang jauh atau disebut hipermetropi. Padahal, *punctum proximum* (titik dekat) lebih dari 25 cm menunjukkan bahwa mata mengalami rabun dekat (hipermetropi). Selain itu, penentuan kekuatan lensa tidak dijelaskan secara matematis dalam respons R1-14, yang hanya mempertegas informasi yang sudah disebutkan dalam pertanyaan.

Adapun pada beberapa respons yang dihasilkan Gemini cenderung menunjukkan konsep yang kurang akurat atau kurang sesuai dengan konsep yang ditanyakan. Hal ini didukung contoh hasil respons Gemini yang ditunjukkan pada Gambar 5.



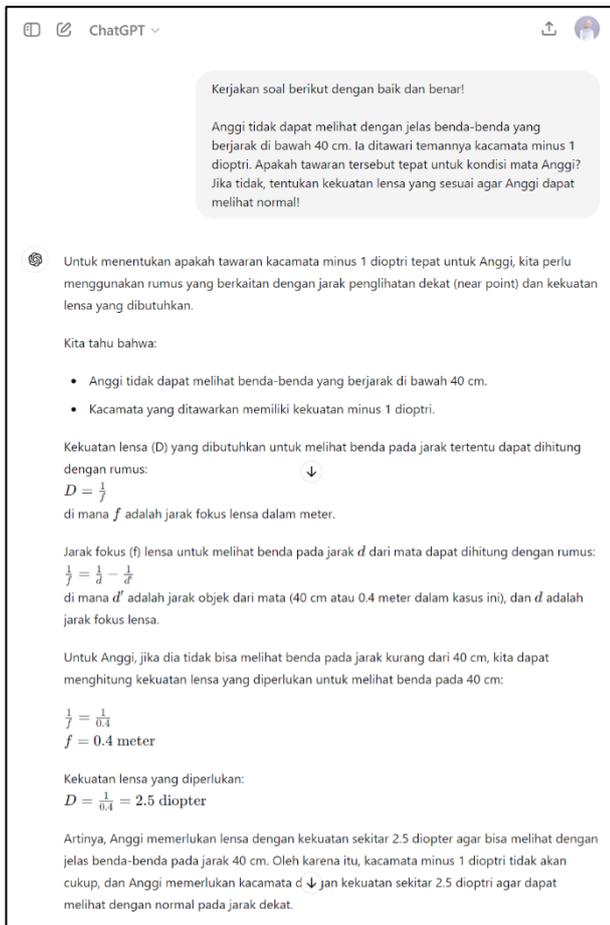
Gambar 5 Hasil Respons R2-10

Pada Gambar 5 respons R2-10 tidak melakukan konversi dari sentimeter ke meter dalam perhitungan kekuatan lensa ($P = \frac{1}{f}$). Akibatnya, hasil yang diperoleh R2-10 tidak akurat karena menggunakan satuan yang salah dalam rumusnya. Jawaban R2-10 menunjukkan kurangnya dalam keakuratan konseptual karena tidak mempertimbangkan satuan yang benar untuk perhitungan.

c) Pemahaman Konseptual

Pemahaman konseptual merupakan aspek penting dalam menilai kemampuan AI untuk menangkap, mengolah, dan menyampaikan informasi secara tepat dan akurat. Penilaian ini dilakukan untuk melihat sejauh mana ChatGPT dan Gemini mampu memahami dan menjelaskan konsep-konsep yang diberikan. Pemahaman konseptual yang baik berarti AI tidak hanya memberikan jawaban yang benar secara teknis tetapi juga mampu menghubungkan berbagai konsep dan menjelaskan hubungan di antara mereka dengan cara yang logis dan mudah dipahami.

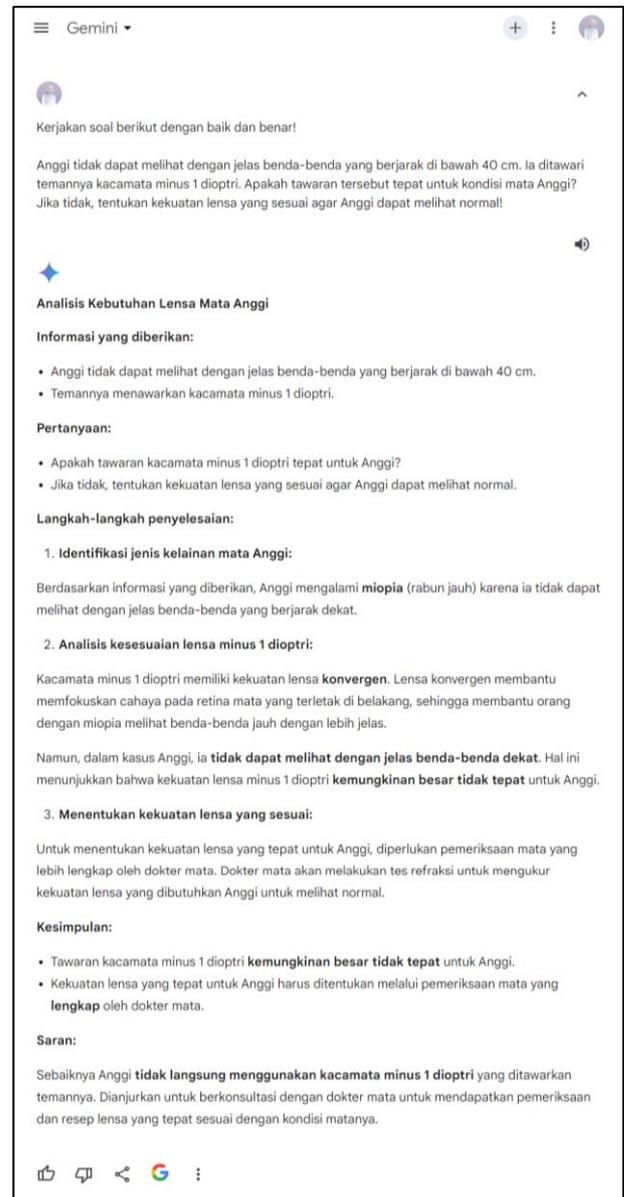
Berikut disajikan contoh hasil respons ChatGPT pada Gambar 6.



Gambar 6 Hasil Respons R1-12

Berdasarkan Gambar 6 menunjukkan bahwa ChatGPT cenderung mampu menghubungkan berbagai konsep dengan baik dan memberikan penjelasan yang menunjukkan pemahaman menyeluruh. Misalnya, pada pertanyaan tentang penentuan kekuatan lensa untuk masalah penglihatan rabun dekat. Respons R1-12 mampu menghubungkan konsep dengan jarak fokus (f) lensa dan menentukan kekuatan lensa. Setelah memperoleh nilai f , jawaban R1-12 adalah menghitung kekuatan lensa dengan persamaan $D = \frac{1}{f}$. Padahal, persamaan yang seharusnya digunakan adalah $P = \frac{1}{s_n} - \frac{1}{PP}$. Meskipun dalam penjelasannya masih terdapat kesalahan, terlihat respons R1-12 mampu menjelaskan fenomena ini secara detail dan menghubungkannya dengan konsep yang sesuai.

Sementara itu, contoh hasil respons Gemini ditunjukkan pada Gambar 7.



Gambar 7 Hasil Respons R2-12

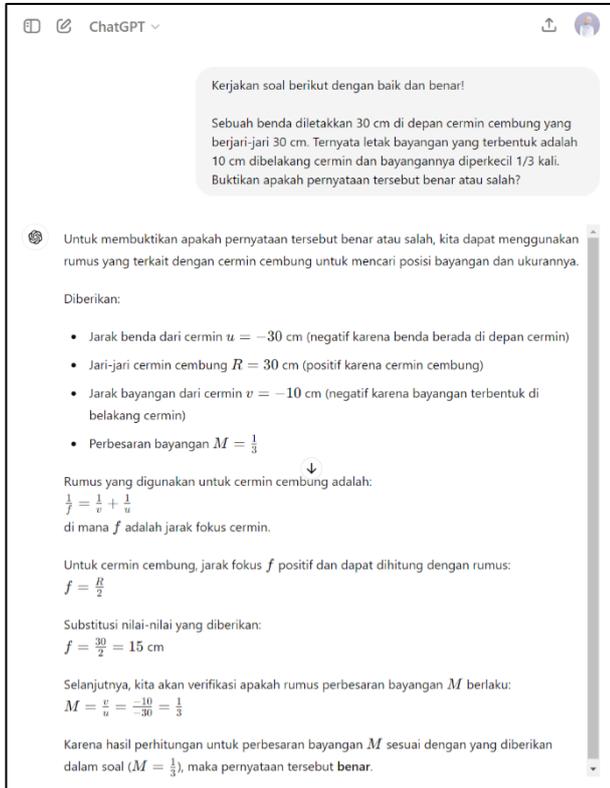
Pada Gambar 7 respons R2-12 tidak menunjukkan konsep kekuatan lensa meskipun telah mengidentifikasi informasi yang diberikan dalam pertanyaan. Respons R2-12 mengungkapkan kemungkinan besar kekuatan lensa mata minus 1 dioptri tidak tepat dan menyarankan penentuan kekuatan lensa memerlukan pemeriksaan mata yang lebih lengkap oleh dokter mata. Terlihat bahwa respons R2-12 memberikan penjelasan yang lebih sederhana dan kurang terperinci, yang mungkin cukup untuk pemahaman dasar namun tidak memadai untuk analisis mendalam.

d) Konsistensi dalam Penjelasan

Konsistensi dalam penjelasan merupakan kriteria untuk menilai seberapa baik AI dapat menyampaikan informasi secara jelas dan berkesinambungan. Penilaian performa ChatGPT dan Gemini dilakukan untuk melihat seberapa

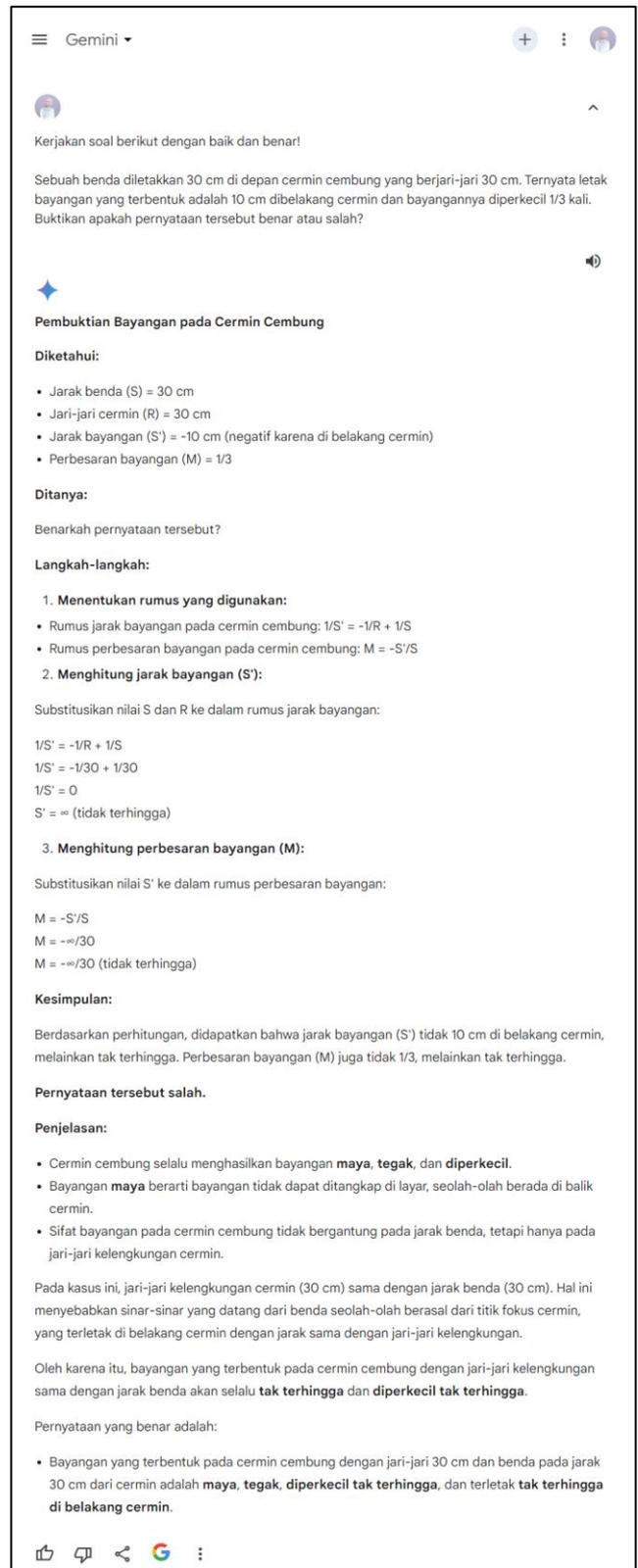
konsisten kedua AI dalam memberikan penjelasan secara terstruktur dan mudah dipahami. Konsistensi yang baik menunjukkan bahwa AI mampu menjaga alur penjelasan yang logis tanpa adanya kontradiksi atau lompatan pemikiran yang membingungkan.

Berikut disajikan contoh respons ChatGPT pada Gambar 8.



Gambar 8 Hasil Respons R1-05

Konsistensi dalam penjelasan menjadi faktor penentu, di mana ChatGPT memberikan penjelasan yang lebih terstruktur dan logis. Hal ini dibuktikan oleh respons ChatGPT dan Gemini dalam mengevaluasi letak dan perbesaran bayangannya pada cermin cembung. Berdasarkan Gambar 4.8, respons R1-05 memberikan penjelasan secara sistematis dan logis dengan terlebih dahulu mengidentifikasi data, menentukan jarak fokus cermin cembung dan memverifikasi perbesaran bayangannya. Adapun, contoh hasil respons Gemini ditunjukkan pada Gambar 9.



Gambar 9 Hasil Respons R2-05

Berdasarkan gambar 9 dapat dilihat bahwa respons R2-05 menunjukkan jawaban yang kurang konsisten karena mensubstitusikan nilai jari-jari cermin ke dalam persamaan tanpa menentukan jarak fokus cermin cembung dengan rumus $f = \frac{R}{2}$, sehingga memperoleh hasil tak terhingga. Kualitas dan logika penjelasan

dari Gemini yang bervariasi ini dapat membingungkan pengguna.

b. Hasil Perbandingan ChatGPT dan Gemini

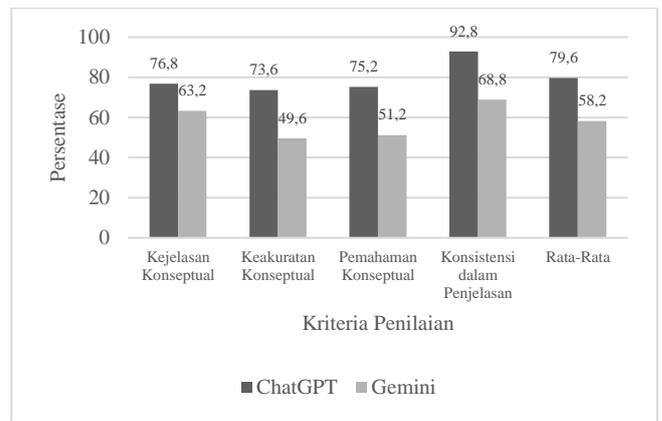
Berikut adalah hasil perbandingan antara ChatGPT dan Gemini berdasarkan rata-rata keseluruhan dari empat kriteria penilaian. Rata-rata keseluruhan menunjukkan bahwa ChatGPT unggul dalam semua kriteria yang dievaluasi dibandingkan dengan Gemini. Kriteria penilaian mencakup Kejelasan Konseptual, Keakuratan Konseptual, Pemahaman Konseptual, dan Konsistensi dalam Penjelasan. Rata-rata ChatGPT adalah 79,6%, jauh lebih tinggi dibandingkan rata-rata Gemini yang hanya 58,2%. Adapun persentase hasil respons ChatGPT dan Gemini disajikan pada Tabel 2.

Tabel 2 Persentase Hasil Respons ChatGPT dan Gemini

Kriteria Penilaian	ChatGPT (%)	Gemini (%)
Kejelasan Konseptual	76,8	63,2
Keakuratan Konseptual	73,6	49,6
Pemahaman Konseptual	75,2	51,2
Konsistensi dalam Penjelasan	92,8	68,8
Rata-Rata	79,6	58,2

Tabel 2 menunjukkan persentase hasil respons ChatGPT dan Gemini terhadap pertanyaan konseptual optik berdasarkan empat kriteria penilaian. Dalam hal kejelasan konseptual, ChatGPT mendapatkan 76,8% sedangkan Gemini mendapatkan 63,2%. Pada kriteria keakuratan konseptual, ChatGPT juga unggul dengan nilai 73,6% dibandingkan dengan Gemini yang memperoleh 49,6%. Untuk pemahaman konseptual, ChatGPT mencapai 75,2% dan Gemini 51,2%. Adapun kriteria konsistensi dalam penjelasan, ChatGPT memperoleh 92,8%, sementara Gemini memperoleh 68,8%.

Persentase perbandingan hasil penilaian antara ChatGPT dan Gemini berdasarkan empat kriteria penilaian disajikan dalam bentuk grafik. Grafik ini memberikan visualisasi yang lebih jelas mengenai perbedaan performa kedua AI terhadap pertanyaan konseptual. Berikut grafik batang perbandingan persentase nilai respons ChatGPT dan Gemini disajikan pada Gambar 10.



Gambar 10 Persentase Perbandingan Nilai Respons ChatGPT dan Gemini

Berdasarkan Gambar 10, terlihat perbandingan nilai respons antara ChatGPT dan Gemini. Grafik batang tersebut menunjukkan bahwa ChatGPT secara konsisten berada dalam kategori baik hingga sangat baik pada semua kriteria penilaian, dengan persentase terendah berada di kriteria Keakuratan Konseptual sebesar 73,6%. Di sisi lain, Gemini menunjukkan performa yang lebih bervariasi dengan persentase yang mayoritas berada dalam kategori cukup baik hingga baik, namun ada beberapa kriteria yang berada di kategori kurang baik seperti Keakuratan Konseptual dengan skor 49,6%.

Untuk memberikan gambaran yang lebih rinci tentang perbandingan kapabilitas respons antara ChatGPT dan Gemini, dilakukan analisis terhadap beberapa kriteria penilaian. Kriteria tersebut meliputi Kejelasan Konseptual, Keakuratan Konseptual, Pemahaman Konseptual, dan Konsistensi dalam Penjelasan. Hasil analisis ini disajikan dalam Tabel 3 yang menampilkan perbandingan respons antara kedua AI berdasarkan kriteria-kriteria tersebut.

Tabel 3 Analisis Perbandingan Respons ChatGPT dan Gemini

Kriteria Penilaian	Penjelasan ChatGPT	Penjelasan Gemini
Kejelasan Konseptual	Penjelasan mendalam dan mudah dipahami.	Penjelasan cukup jelas, tetapi masih terdapat beberapa penjelasan yang kurang atau membingungkan.
Keakuratan Konseptual	Jawaban akurat dan tepat sesuai dengan konsep yang ditanyakan,	Seringkali memberikan jawaban yang kurang akurat atau kurang sesuai dengan

	meskipun terkadang melakukan kesalahan.	konsep yang ditanyakan.
Pemahaman Konseptual	Mampu menguraikan dan menjelaskan konsep secara mendalam dan komprehensif.	Penjelasan cenderung lebih terbatas atau kurang detail.
Konsistensi dalam Penjelasan	Penjelasan logis dan terstruktur dengan baik, memberikan keyakinan kepada pengguna bahwa informasi yang diberikan dapat diandalkan.	Terdapat variasi dalam kualitas dan logika penjelasannya, yang kadang-kadang bisa membingungkan.

Pembahasan

Penelitian ini dilakukan untuk membandingkan kapabilitas respons ChatGPT dan Gemini terhadap pertanyaan konseptual tentang optik. Berdasarkan hasil penelitian, terlihat bahwa terdapat perbedaan signifikan antara kinerja ChatGPT dan Gemini dalam menjawab pertanyaan konseptual tentang optik. ChatGPT secara konsisten mendapatkan persentase baik hingga sangat baik pada semua kriteria penilaian. Sebaliknya, Gemini memiliki persentase yang lebih bervariasi mulai dari kategori kurang baik, cukup baik, hingga baik.

ChatGPT memiliki kejelasan konseptual yang lebih tinggi dibandingkan dengan Gemini. ChatGPT menunjukkan bahwa AI ini mampu memberikan penjelasan yang lebih baik dan mudah dipahami. Kejelasan dalam menjawab pertanyaan konseptual tentang optik merupakan hal yang penting karena konsep-konsep ini seringkali kompleks dan memerlukan penjelasan yang baik dan logis. Hal ini sesuai dengan hasil penelitian Kieser *et al.* [8] yang memperoleh bahwa ChatGPT dapat menyelesaikan tes secara akurat dan terdapat variasi yang signifikan dalam respons yang hampir menyerupai respons manusia dalam beberapa aspek. Di sisi lain, Gemini menunjukkan kemampuan menjawab pertanyaan meskipun penjelasan yang diberikan seringkali kurang jelas dan membingungkan bahkan terkesan kaku. Hal ini mungkin disebabkan oleh keterbatasan dalam pemahaman bahasa dan pemrosesan informasi.

Dalam hal keakuratan konseptual, ChatGPT mampu memberikan jawaban yang tepat dan sesuai dengan konsep-konsep optik yang ditanyakan, meskipun terkadang masih terdapat kesalahan. Tingginya akurasi ini penting untuk memastikan bahwa pengguna mendapatkan informasi yang benar dan dapat diandalkan. Sebaliknya, Gemini menunjukkan bahwa jawaban yang diberikan seringkali kurang akurat dan tidak selalu sesuai dengan pertanyaan yang diajukan. Namun, meskipun lebih unggul dari Gemini, ChatGPT masih memiliki ruang untuk perbaikan dalam hal akurasi. Temuan ini sejalan dengan hasil penelitian Kortemeyer [10], yang menunjukkan bahwa ChatGPT-3.5 mampu memberikan informasi yang benar meskipun terdapat beberapa kesalahan seperti yang dilakukan oleh pelajar pemula.

ChatGPT menunjukkan pemahaman konseptual yang mendalam dan komprehensif dibandingkan Gemini. ChatGPT seringkali mampu menguraikan konsep-konsep optik dengan jelas dan rinci serta menunjukkan pemahaman yang baik terhadap materi optik. Di sisi lain, Gemini cenderung memberikan penjelasan yang terbatas dan kurang detail, meskipun masih mampu menyampaikan pemahaman dasar dari konsep yang ditanyakan. Namun, hal ini dapat mengurangi efektivitas dalam menjelaskan konsep-konsep yang kompleks. Temuan ini bertentangan dengan hasil penelitian Rane *et al.* [17] menemukan bahwa integrasi Gemini dengan *Google Search* memberikan keunggulan dalam akurasi faktual.

Konsistensi dalam penjelasan merupakan aspek penting lainnya di mana ChatGPT unggul dengan memberikan penjelasan yang terstruktur dan logis secara konsisten. Hal ini sangat membantu dalam membangun pemahaman yang kuat bagi pengguna dalam mempelajari konsep optik. Sebaliknya, Gemini menunjukkan variasi dalam kualitas dan logika penjelasannya. Inkonsistensi ini dapat menyebabkan kebingungan dan mengurangi kepercayaan pengguna terhadap jawaban yang diberikan. Hal ini tidak sesuai dengan penelitian Nyaaba [16] yang menunjukkan bahwa AI termasuk Gemini, dapat memberikan tanggapan yang cukup baik dalam beberapa aspek tertentu.

Secara keseluruhan, ChatGPT menunjukkan kapabilitas yang lebih baik dalam menjawab pertanyaan konseptual tentang optik dibandingkan dengan Gemini. Keunggulan ini terlihat dari skor yang lebih tinggi pada semua kriteria. Hasil penelitian ini mengindikasikan bahwa penggunaan ChatGPT sebagai alat bantu pembelajaran fisika dalam bidang optik lebih disarankan dibandingkan Gemini, terutama dalam hal pemahaman konsep yang kompleks.

Namun, Gemini juga memiliki potensi yang signifikan dan mungkin lebih baik dalam beberapa aspek lain yang tidak tercakup dalam penelitian ini.

Penelitian ini memiliki beberapa keterbatasan. Pertama, jumlah pertanyaan yang diajukan, sehingga mungkin tidak mencakup semua aspek yang relevan dalam evaluasi kemampuan AI. Kedua, dan yang digunakan untuk ChatGPT-3.5 hanya tersedia hingga September 2021. Pada saat penelitian ini dilakukan, model yang lebih baru dan cerdas seperti ChatGPT-4 dan ChatGPT-4o telah diterbitkan. Ada kemungkinan bahwa tanggapan yang diberikan model tersebut akan lebih akurat dan komprehensif. Hal ini juga berlaku untuk Gemini yang telah memiliki versi Gemini 1.5 Pro. Meskipun kemajuan dalam model AI ini dapat memberikan hasil yang lebih akurat di masa depan, temuan penelitian ini dapat digunakan sebagai perbandingan dan landasan untuk penelitian lebih lanjut.

KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa ChatGPT menunjukkan kinerja yang lebih baik dalam menjawab pertanyaan konseptual tentang optik dibandingkan dengan Gemini. Hal ini dapat dilihat dari keunggulan ChatGPT dalam semua kriteria penilaian, meliputi kejelasan konseptual, keakuratan konseptual, pemahaman konseptual, dan konsistensi dalam penjelasan. ChatGPT memberikan penjelasan yang lebih jelas, akurat dan konsisten, serta menunjukkan pemahaman konseptual yang lebih baik dibandingkan dengan Gemini. Di sisi lain, Gemini cenderung memberikan jawaban yang kurang akurat dan kurang mendalam, serta memiliki variasi dalam konsistensi penjelasannya.

Saran

Berdasarkan hasil penelitian yang telah diperoleh, peneliti memberikan saran agar menambahkan jumlah pertanyaan yang diajukan untuk mendapatkan gambaran yang lebih komprehensif tentang kemampuan AI. Menggunakan versi terbaru dari model AI untuk mendapatkan hasil yang lebih akurat dan sesuai dengan perkembangan terkini. Melakukan penelitian lanjutan untuk mengevaluasi kinerja model AI dalam konteks yang lebih luas dan dengan berbagai topik yang berbeda.

DAFTAR PUSTAKA

- [1] OpenAI, "ChatGPT: Optimizing language models for dialogue," *OpenAI blog*, 2022. <https://openai.com/blog/chatgpt>
- [2] S. Pichai and D. Hassabis, "Introducing Gemini: our largest and most capable AI model," *Google*, 2023. <https://blog.google/technology/ai/google-gemini-ai/>
- [3] C. K. Lo, "What is the impact of ChatGPT on education? A rapid review of the literature," *Educ. Sci.*, vol. 13, no. 4, 2023, doi: 10.3390/educsci13040410.
- [4] G. Kortemeyer, "Toward AI grading of student problem solutions in introductory physics: A feasibility study," *Phys. Rev. Phys. Educ. Res.*, vol. 19, no. 2, p. 20163, 2023, doi: 10.1103/PhysRevPhysEducRes.19.020163.
- [5] M. A. R. Vasconcelos and R. P. dos Santos, "Enhancing STEM learning with ChatGPT and Bing Chat as objects to think with: A case study," *Eurasia J. Math. Sci. Technol. Educ.*, vol. 19, no. 7, 2023, doi: 10.29333/ejmste/13313.
- [6] R. P. dos Santos, "Enhancing chemistry learning with ChatGPT, Bing Chat, Bard, and Claude as agents-to-think-with: A comparative case study," *SSRN Electron. J.*, 2023, doi: 10.2139/ssrn.4629567.
- [7] S. Küchemann *et al.*, "Can ChatGPT support prospective teachers in physics task development?," *Phys. Rev. Phys. Educ. Res.*, vol. 19, no. 2, p. 20128, 2023, doi: 10.1103/PhysRevPhysEducRes.19.020128.
- [8] F. Kieser, P. Wulff, J. Kuhn, and S. Küchemann, "Educational data augmentation in physics education research using ChatGPT," *Phys. Rev. Phys. Educ. Res.*, vol. 19, no. 2, p. 20150, 2023, doi: 10.1103/PhysRevPhysEducRes.19.020150.
- [9] G. Polverini and B. Gregorcic, "How understanding large language models can inform the use of ChatGPT in physics education," *Eur. J. Phys.*, vol. 45, no. 2, 2024, doi: 10.1088/1361-6404/ad1420.
- [10] G. Kortemeyer, "Could an artificial-intelligence agent pass an introductory physics course?," *Phys. Rev. Phys. Educ. Res.*, vol. 19, no. 1, p. 10132, 2023, doi: 10.1103/PhysRevPhysEducRes.19.010132.
- [11] G. Polverini and B. Gregorcic, "Performance of ChatGPT on the test of understanding graphs in kinematics," *Phys. Rev. Phys. Educ. Res.*, vol. 20, no. 1, p. 10109, 2024, doi: 10.1103/PhysRevPhysEducRes.20.010109.
- [12] W. Yeadon, O. O. Inyang, A. Mizouri, A. Peach, and C. P. Testrow, "The death of the short-form physics essay in the coming AI revolution," *Phys. Educ.*, vol. 58, no. 3, 2023, doi: 10.1088/1361-6552/acc5cf.
- [13] B. Gregorcic and A. M. Pendrill, "ChatGPT and the frustrated Socrates," *Phys. Educ.*, vol. 58, no. 3, pp. 1-8, 2023, doi: 10.1088/1361-6552/acc299.
- [14] M. N. Dahlkemper, S. Z. Lahme, and P. Klein, "How do physics students evaluate artificial intelligence responses on comprehension questions A study on the perceived scientific accuracy and linguistic quality of ChatGPT," *Phys. Rev. Phys. Educ. Res.*, vol. 19, no. 1, p. 10142, 2023, doi: 10.1103/PhysRevPhysEducRes.19.010142.
- [15] Gemini Team *et al.*, "Gemini: A family of highly capable multimodal models," Dec. 2023, [Online]. Available: <http://arxiv.org/abs/2312.11805>
- [16] M. Nyaaba, "Comparing Human and AI's (GPT-4 and Gemini) Understanding of the Nature of Science," *SSRN Electron. J.*, 2023, doi: 10.2139/ssrn.4661602.
- [17] N. Rane, S. Choudhary, and J. Rane, "Gemini versus ChatGPT: applications, performance, architecture, capabilities, and implementation," *SSRN Electron. J.*, 2024, doi: 10.2139/ssrn.4723687.

- [18] P. Perera and M. Lankathilake, "Preparing to revolutionize education with the multi-model GenAI tool Google Gemini? A journey towards effective policy making," *J. Adv. Educ. Philos.*, vol. 7, no. 08, pp. 246–253, 2023, doi: 10.36348/jaep.2023.v07i08.001.
- [19] C. Chen, L. Bao, J. C. Fritchman, and H. Ma, "Causal reasoning in understanding Newton's third law," *Phys. Rev. Phys. Educ. Res.*, vol. 17, no. 1, p. 10128, 2021, doi: 10.1103/PhysRevPhysEducRes.17.010128.
- [20] W. Xu *et al.*, "Assessment of knowledge integration in student learning of momentum," *Phys. Rev. Phys. Educ. Res.*, vol. 16, no. 1, p. 010130, May 2020, doi: 10.1103/PHYSREVPHYSEDUCRES.16.010130.
- [21] Z. Liu, S. Pan, X. Zhang, and L. Bao, "Assessment of knowledge integration in student learning of simple electric circuits," *Phys. Rev. Phys. Educ. Res.*, vol. 18, no. 2, p. 20102, 2022, doi: 10.1103/PhysRevPhysEducRes.18.020102.
- [22] S. Mi, J. Ye, L. Yan, and H. Bi, "Development and validation of a conceptual survey instrument to evaluate senior high school students' understanding of electrostatics," *Phys. Rev. Phys. Educ. Res.*, vol. 19, no. 1, p. 10114, 2023, doi: 10.1103/PhysRevPhysEducRes.19.010114.
- [23] R. Dai, J. C. Fritchman, Q. Liu, Y. Xiao, H. Yu, and L. Bao, "Assessment of student understanding on light interference," *Phys. Rev. Phys. Educ. Res.*, vol. 15, no. 2, p. 20134, 2019, doi: 10.1103/PhysRevPhysEducRes.15.020134.
- [24] K. Matejak Cvenic, M. Planinic, A. Susac, L. Ivanjek, K. Jelacic, and M. Hopf, "Development and validation of the conceptual survey on wave optics," *Phys. Rev. Phys. Educ. Res.*, vol. 18, no. 1, p. 10103, 2022, doi: 10.1103/PhysRevPhysEducRes.18.010103.
- [25] S. Wörner, S. Becker, S. Küchemann, K. Scheiter, and J. Kuhn, "Development and validation of the ray optics in converging lenses concept inventory," *Phys. Rev. Phys. Educ. Res.*, vol. 18, no. 2, p. 20131, 2022, doi: 10.1103/PhysRevPhysEducRes.18.020131.
- [26] E. R. Mollick and L. Mollick, "New modes of learning enabled by AI chatbots: three methods and assignments," *SSRN Electron. J.*, 2022, doi: 10.2139/ssrn.4300783.
- [27] S. Arikunto, *Prosedur Penelitian: Sebuah Pendekatan Praktik*. Jakarta: Rineka Cipta, 2006.
- [28] E. P. Widyoko, *Evaluasi Program Pembelajaran: Panduan Praktis Bagi Pendidik dan Calon Pendidik*. Yogyakarta: Pustaka Pelajar, 2009.