

PUSTABIBLIA

Journal of Library and Information Science

E-ISSN : 2549-3868 | P-ISSN :2549-3493

LETTER OF ACCEPTANCE FOR SCIENTIFIC ARTICLES PUBLICATION

NO : B-005/PUSTABIBLIA/LoA/04/2026

Author (s) :

1. Tri Wahyudi Putra

The article title:

“Tingkat Akurasi Sitasi Pada Model Pencarian Informasi Berbasis AI: Study Perbandingan (Chatgpt, Perplexity, dan Deepseek)”

Congratulations! The editorial Executive Board has evaluated the article through the peer review process and it is considered to be “Accepted” and will be published in Pustabiblia:

Journal of Library and Information Science (P-2549-3493, E-ISSN 2549-3868)

Vol. 10 No. 1, Juni 2026.

Thank you for submitting your article to our journal. We wish you all possible success in the future.

Salatiga, 15 April 2026

Chief Editor



Dr. Wiji Suwarno, M.Hum



TINGKAT AKURASI SITASI PADA MODEL PENCARIAN INFORMASI BERBASIS AI: STUDY PERBANDINGAN (CHAT GPT, PERPLEXITY, DAN DEEPSEEK)

Tri Wahyudi Putra¹; Ruslan²; Nazaruddin³; Mukhtaruddin⁴; T.Mulkan Safri⁵
UIN Ar-Raniry Banda Aceh

Email: 210503057@student.ar-raniry.ac.id¹; ruslan.ruslan@ar-raniry.ac.id²; nazaruddin@ar-raniry.ac.id³; mukhtaruddin@ar-raniry.ac.id⁴; t.mulkansafri@ar-raniry.ac.id⁵

Naskah diterima: tgl-bln-thn, direvisi: tgl-bln-thn, disetujui: tgl-bln-thn

ABSTRACT

The development of artificial intelligence (AI) based information retrieval models has brought significant changes to the way academic literature searches are conducted, particularly in the presentation of responses accompanied by automatically generated and structured citations. This transformation not only accelerates the information retrieval process but also influences academic behavior in utilizing scholarly references. Nevertheless, the accuracy, consistency, and traceability of the generated references remain critical issues within the context of academic integrity. This study aims to analyze and compare the level of citation accuracy in ChatGPT developed by OpenAI, as well as Perplexity AI and DeepSeek. A descriptive quantitative approach with an evaluative method was employed using citation recall as an indicator of citation completeness and accuracy. Ten academic questions were submitted to each platform, resulting in 54 scientific claims requiring supporting references. All citations were subsequently verified through academic databases to ensure their validity and relevance to the stated claims. The findings indicate that Perplexity achieved the highest citation recall score of 68.52% and was categorized as accurate, followed by ChatGPT with 62.96% in the fairly accurate category, and DeepSeek with 46.30%, also classified as fairly accurate. Although all three models were capable of providing references in their responses, several inconsistencies, unverifiable sources, and instances of overcitation were identified. Therefore, AI functions as a supportive tool in academic information retrieval; however, manual verification remains essential to ensure academic integrity, reliability, and scholarly credibility.

ABSTRAK

Perkembangan model pencarian informasi berbasis kecerdasan buatan (AI) telah membawa perubahan signifikan dalam cara penelusuran literatur akademik dilakukan, khususnya dalam penyajian jawaban yang dilengkapi dengan sitasi secara otomatis dan terstruktur. Transformasi ini tidak hanya mempercepat proses pencarian informasi, tetapi juga memengaruhi pola perilaku akademik dalam memanfaatkan referensi ilmiah. Meski demikian, aspek ketepatan, konsistensi, dan keterlacakan referensi yang dihasilkan masih menjadi isu penting dalam konteks integritas ilmiah. Penelitian ini bertujuan untuk menganalisis dan membandingkan tingkat akurasi sitasi pada ChatGPT yang dikembangkan oleh OpenAI, serta Perplexity AI dan DeepSeek. Pendekatan yang digunakan adalah kuantitatif deskriptif dengan metode evaluatif melalui penghitungan *citation recall* sebagai indikator kelengkapan dan ketepatan sitasi. Sebanyak sepuluh pertanyaan akademik diajukan kepada masing-masing platform dan menghasilkan 54 klaim ilmiah yang membutuhkan dukungan referensi. Seluruh sitasi kemudian diverifikasi melalui basis data ilmiah untuk memastikan validitas dan relevansinya terhadap klaim yang disampaikan. Hasil analisis menunjukkan bahwa Perplexity mencatat nilai *citation recall* tertinggi sebesar 68,52% dan masuk kategori akurat, diikuti ChatGPT dengan 62,96% dalam kategori cukup akurat, serta DeepSeek sebesar 46,30% yang juga tergolong cukup akurat. Walaupun ketiga model mampu menyertakan referensi dalam jawabannya, masih ditemukan sejumlah ketidaksesuaian, referensi yang tidak dapat ditelusuri, serta kecenderungan *overcitation*. Dengan demikian, AI berperan sebagai alat bantu dalam pencarian informasi akademik, namun proses verifikasi manual tetap diperlukan guna menjamin integritas, reliabilitas, dan kredibilitas ilmiah.

Keywords: *citation accuracy; artificial intelligence; information retrieval; generative AI; academic integrity.*

1. PENDAHULUAN

Berkembangnya teknologi kecerdasan buatan atau *Artificial Intelligence (AI)* dalam hal pencarian dan penyediaan informasi telah mengalami kemajuan yang signifikan dalam beberapa tahun terakhir. Berbagai jenis model *AI* sekarang telah dapat memahami bahasa manusia, mengolah data serta memberikan jawaban yang logis dan mudah dipahami. Inovasi ini telah mengubah cara orang mengakses dan memanfaatkan informasi di dunia akademis. Sejalan dengan kemajuan teknologi, proses menemukan informasi telah menjadi jauh lebih sederhana berkat *AI*. Sebelum munculnya *AI*, orang-orang melakukan pencarian informasi secara manual, misalnya melalui koleksi fisik yang ada atau mesin pencari dengan kata kunci yang terbatas (Nahla 2024).

Pemanfaatan mesin Pencarian *AI* ini memungkinkan pengguna memperoleh hasil pencarian yang lebih cepat, relevan, dan kontekstual melalui analisis perilaku, kata kunci, serta preferensi individu. Perkembangan mesin pencarian berbasis kecerdasan buatan *AI* berdampak pada perilaku pencarian informasi pada masa saat ini. Mesin Pencarian informasi berbasis *AI* telah mengubah secara individu mengakses, menyeleksi, dan memanfaatkan informasi dari berbagai sumber digital. Jika sebelumnya proses pencarian informasi dilakukan secara manual melalui mesin pencari dengan memasukkan kata kunci sederhana. Sistem berbasis *AI* mampu memahami konteks pertanyaan, menafsirkan makna, dan memberikan jawaban yang lebih relevan.

Dengan kemampuan ini, *AI* mulai diakui sebagai asisten digital yang dapat mempermudah peneliti untuk bekerja lebih cepat dan lebih efisien dalam mencari sumber pengetahuan. Sebuah studi literatur dari Library and Information Science di Indonesia menunjukkan bahwa penggunaan *AI* dalam pencarian informasi memberikan dampak positif seperti mempercepat proses pencarian serta menyajikan informasi yang ringkas dan mudah dipahami. Namun, penggunaan tersebut juga menimbulkan risiko seperti ketidakakuratan data (Nahla 2024)

Hal ini berkaitan dengan tahapan yang dikemukakan David Ellis mengenai perilaku dalam pencarian informasi yang terhubung langsung pada sistem pengambilan informasi. Ia mengidentifikasi karakteristik dalam perilaku pencarian informasi, yang mencakup *starting* (individu memulai proses pencarian informasi), *chaining* (mencatat hal-hal yang dianggap penting), *browsing* (mencari informasi yang terstruktur atau setengah terstruktur), *differentiating* (proses reduksi data atau pemilihan data), *monitoring* (memantau atau mencari berita-berita terbaru), *extracting* (mengambil informasi yang bermanfaat), *verifying* (memeriksa ukuran data yang sudah diambil), dan *ending* (penutup dari pencarian) namun demikian, meskipun *AI* mampu membantu pada tahap awal hingga pengolahan informasi, aspek *verifying* atau pemekrisaan keakuratan informasi masih menjadi tantangan yang signifikan (Ellis 1989).

Kemunculan berbagai mesin pencarian berbasis kecerdasan buatan *AI* kini menjadi pilihan penting bagi banyak kalangan dalam mencari informasi, baik untuk keperluan pendidikan, penelitian, maupun kebutuhan umum lainnya. Mesin Pencari *AI* menawarkan kemudahan dalam mengakses informasi secara cepat, memahami konteks pertanyaan pengguna, dan menyajikan hasil yang tersusun dengan baik. Teknologi kecerdasan buatan atau *AI* dapat memberikan dukungan kepada penggunanya dalam menemukan informasi, salah satu cara yaitu dengan menggunakan *Natural Language Processing (NLP)*, Sistem ini memiliki kemampuan untuk mengerti pertanyaan yang dikemukakan oleh pengguna dalam bahasa sehari-hari dan menyajikan hasil pencarian yang lebih relevan dan tepat (Rosiana, Cahyanti, dan Rahayu 2024). Hadirnya mesin pencarian berbasis

Tingkat Akurasi Sitasi Pada Model Pencarian Informasi Berbasis AI: Study (Chat Gpt, Perflexity, Dan Deepseek)

AI telah mengubah pola pencarian literatur ilmiah. Jika sebelumnya pencarian referensi membutuhkan waktu yang lama, kini cukup dengan memasukkan pertanyaan singkat, system AI dapat menampilkan berbagai sumber yang dianggap relevan. Selain itu, teknologi ini juga mampu merangkum dan menafsirkan informasi sesuai dengan kebutuhan pengguna.

Seiring pesatnya perkembangan teknologi kecerdasan buatan, muncul berbagai mesin pencarian informasi berbasis AI yang menawarkan kemudahan dalam mengakses dan mengolah data secara lebih cepat dan akurat. Beberapa platform yang saat ini banyak digunakan oleh kalangan akademisi, peneliti, dan masyarakat umum antara lain *ChatGPT*, *Perplexity AI*, dan *DeepSeek*. Ketiganya memiliki karakteristik dan keunggulan tersendiri dalam membantu proses pencarian dan pengolahan informasi. *ChatGPT* adalah model bahasa yang dibangun di atas *Generative Pre-trained Transformer (GPT)* yang diciptakan oleh *OpenAI* (Erizal et al. 2024)

Sistem ini mampu mengerti konteks dari pertanyaan yang diajukan pengguna dan menghasilkan jawaban yang mirip dengan percakapan manusia. *ChatGPT* banyak digunakan untuk membantu dalam penulisan teks akademis, membuat ringkasan, dan eksplorasi ide penelitian. Namun, dalam hal pencarian informasi ilmiah. Di sisi lain, *Perplexity AI* muncul sebagai inovasi yang memadukan kemampuan model bahasa dengan sistem pencarian yang mengintegrasikan kemampuan *Natural Language Processing* dengan algoritma pembelajaran (Dewi Yanti Liliana 2023). Kelebihan utama dari *Perplexity* adalah transparansi sumber, di mana setiap jawaban disertai dengan tautan menuju sumber asli. Hal ini membuat *Perplexity* lebih dapat dipercaya untuk tujuan akademis karena pengguna bisa memverifikasi keabsahan informasi. Sementara itu, *DeepSeek* adalah mesin pencarian AI generatif yang menggabungkan teknologi pencarian cerdas dengan kemampuan pemrosesan bahasa alami *Natural Language Processing* untuk menjawab pertanyaan kompleks, menganalisis data, hingga membuat konten secara otomatis lalu dikembangkan dengan penekanan pada kecepatan analisis dan kedalaman pemrosesan konteks. *DeepSeek* sering dimanfaatkan oleh para profesional dan peneliti untuk mencari data, ringkasan artikel, serta analisis perbandingan antar sumber. Kelebihan dari *DeepSeek* terletak pada efisiensi dan kedalaman respons (Handayani et.al 2023). Secara keseluruhan, kemunculan mesin pencari berbasis AI seperti *ChatGPT*, *Perplexity*, dan *DeepSeek* menunjukkan bahwa teknologi kini memainkan peran penting dalam membentuk perilaku pencarian informasi modern. Pengguna tidak lagi hanya mengandalkan mesin pencari tradisional seperti *Google*, tetapi mulai beralih pada sistem yang mampu memahami konteks dan memberikan hasil yang lebih personal serta interaktif.

Pada lembaga pendidikan, pencarian informasi berbasis AI memainkan fungsi yang sangat penting dalam mendukung kegiatan akademis, terutama dalam tahap penulisan karya ilmiah oleh mahasiswa, Dosen, dan guru. Teknologi ini tidak hanya mempercepat pencarianti informasi, tetapi juga memudahkan pengguna untuk memahami, merangkum, serta menyusun argumen ilmiah dari berbagai sumber yang relevan. Meskipun demikian kemudahan tersebut tidak selalu diiringi dengan keakuratan informasi. Dalam praktiknya, ditemukan bahwa beberapa sitasi yang di hasilkan tampak menyakinkan secara format, namun setelah ditelusuri lebih lanjut, sebagian di antaranya tidak memiliki sumber yang valid atau bahkan tidak ditemukan dalam database ilmiah seperti *Google Scholar* maupun *DOAJ*. Fenomena ini dikenal sebagai *AI hallucination citation*, yaitu kondisi ketika system AI menghasilkan referensi yang seolah-olah benar, padahal tidak berasal dari sumber yang nyata atau kondisi ketika sistem AI menghasilkan referensi yang tidak nyata atau tidak valid (Abouammoh et al. 2025).

Berdasarkan observasi awal peneliti dari ujicoba pencarian beberapa bahasa penulisan di mesin pencarian informasi berbasis *AI* seperti *Chat GPT*, *Perplexity*, dan *Deepseek* ditemukan ketidaksesuaian antara sitasi yang di hasilkan dengan sumber aslinya. Bahkan, dalam beberapa kasus ada referensi yang sama sekali tidak ditemukan dalam basis data ilmiah seperti *Google Scholar* atau *DOAJ*. Temuan ini menunjukkan bahwa meskipun *AI* memberikan kemudahan dalam pencarian informasi, aspek validasi sitasi masih perlu di kaji secara mendalam. Temuan tersebut sejalan dengan hasil penelitian terdahulu yang menunjukkan bahwa akurasi sitasi pada system berbasis kecerdasan buatan atau *AI* masih menjadi permasalahan yang signifikan. Penelitian pertama yang dilakukan oleh Liu et al meneliti tentang menilai akurasi sitasi pada mesin pencarian generative *AI* yang menghasilkan jawaban beserta sitasi di dalam teks. Mereka menekankan bahwa sistem yang tepercaya harus memiliki *verifiability* yang baik, yaitu sitasi yang lengkap *citation recall*. Dalam studi tersebut, empat mesin pencarian *Bing Chat*, *NevaAI*, *Perplexity*, dan *YouChat* dievaluasi melalui berbagai jenis pertanyaan. Hasil penelitian menunjukkan bahwa meskipun jawaban yang diberikan tampak informatif, banyak sitasi yang tidak mendukung pernyataan secara memadai. Hanya 51,5% kalimat yang didukung sitasi dan 74,5% sitasi yang benar-benar sesuai dengan isi kalimat. Temuan ini menunjukkan bahwa tingkat akurasi sitasi pada mesin pencarian generatif masih rendah dan perlu ditingkatkan agar informasi yang diberikan lebih dapat dipercaya (Liu, Zhang, dan Liang 2023). Penelitian kedua yang dilakukan oleh L.J. Janse van Rensburg meneliti mengembangkan protokol audit sitasi berbasis *AI* dan menemukan bahwa sekitar 20% sitasi mengandung kesalahan, serta *AI* mampu mendeteksi referensi palsu, artikel retract, dan jurnal predator dengan tingkat verifikasi 91,7%. Temuan ini menunjukkan bahwa akurasi sitasi masih menjadi permasalahan penting dalam konteks penggunaan model *AI* (Rensburg 2025). Penelitian ketiga yang dilakukan oleh Maheshwari et al meneliti pengembang sistem *RAG (Retrieval Augmented Generation)* menunjukkan bahwa model bahasa besar masih menghadapi persoalan serius dalam menghasilkan sitasi yang akurat. Studi industri menemukan bahwa akurasi sitasi pada mesin pencarian berbasis *AI* hanya mencapai sekitar 74%, sehingga sebagian sitasi yang dihasilkan tidak benar-benar merujuk pada sumber asli. Permasalahan ini terjadi karena *Large Language Model (LLM)* sering kesulitan mencocokkan jawaban dengan dokumen yang relevan. Untuk mengatasi hal tersebut, penelitian terbaru mengembangkan metode *post-processing* seperti pencocokan kata kunci, analisis semantik, serta pemanfaatan skor kesesuaian seperti *BERTScore*. Hasil uji coba menunjukkan adanya peningkatan akurasi sitasi hingga 15% setelah penerapan teknik tersebut. Temuan ini menegaskan bahwa peningkatan mekanisme verifikasi sumber sangat penting untuk meningkatkan reliabilitas pencarian informasi berbasis *Artificial Intelligence (AI)* (Maheshwari, Teneti, dan Nakkiran 2025).

Meskipun sejumlah penelitian sebelumnya telah mengkaji akurasi sitasi pada system kecerdasan buatan telah dilakukan dan Sebagian besar masih berfokus secara evaluasi umum terhadap model *AI* tanpa melakukan perbandingan langsung antar model yang saat ini paling banyak digunakan. Selain itu penelitian terdahulu cenderung pada aspek *verifiability* secara umum, seperti *citation precision* dan kelengkapan referensi. Namun belum ada secara spesifik mengukur Tingkat akurasi sitasi menggunakan pendekatan *citation recall* dalam konteks pertanyaan akademik. Di sisi lain, kajian mengenai akurasi sitasi pada platform seperti *ChatGPT*, *Perplexity*, dan *DeepSeek* masih terbatas, terutama dalam pendekatan evaluasi berbasis keluaran sistem (system output evaluation) yang berorientasi pada analisis langsung terhadap respons *Artificial Intelligence (AI)*. Oleh karena itu, terdapat celah penelitian yang signifikan dalam bentuk kebutuhan akan studi komparatif yang secara sistematis mengukur dan membandingkan tingkat akurasi sitasi dari ketiga model *AI* tersebut menggunakan indikator yang terukur dan kontekstual dalam lingkungan akademik.

Tingkat Akurasi Sitasi Pada Model Pencarian Informasi Berbasis AI: Study (Chat Gpt, Perflexity, Dan Deepseek)

Urgensi penelitian ini didasari meningkatnya penggunaan teknologi kecerdasan buatan dalam aktivitas akademik, terkhususnya dalam proses pencarian informasi oleh mahasiswa, dosen, dan peneliti. Kemudahan yang ditawarkan oleh sistem *AI* dalam menghasilkan jawaban beserta sitasi secara otomatis berpotensi menimbulkan ketergantungan tanpa diimbangi dengan kemampuan verifikasi sumber secara kritis. Apabila sitasi yang dihasilkan tidak akurat atau tidak dapat ditelusuri, maka hal tersebut dapat berdampak serius terhadap integritas ilmiah, kualitas karya akademik, serta kredibilitas penelitian. Selain itu, fenomena *AI hallucination citation* yang menghasilkan referensi fiktif atau tidak valid semakin memperkuat pentingnya kajian ini dilakukan. Oleh karena itu, penelitian ini menjadi penting untuk memberikan gambaran empiris mengenai tingkat keakuratan sitasi pada system *AI*, sekaligus menjadi dasar dalam meningkatkan literasi informasi dan kesadaran kritis pengguna dalam memanfaatkan teknologi kecerdasan buatan secara bertanggung jawab.

Penelitian ini memiliki nilai kebaruan yang terletak pada pendekatan komparatif terhadap tiga model pencarian informasi berbasis kecerdasan buatan yang saat ini banyak digunakan, yaitu *Chat GPT*, *Perflexity*, dan *Deepseek* yang belum banyak dikaji secara bersamaan. Selain itu penelitian ini menggunakan pendekatan evaluatif berbasis *citation recall* sebagai indikator utama dalam mengukur tingkat kelengkapan dan ketepatan sitasi, yang memberikan perspektif kuantitatif yang lebih spesifik dibandingkan penelitian sebelumnya. Kebaruan lainnya terletak pada desain penelitian yang berorientasi pada evaluasi keluaran system *system output evaluation*, tanpa melibatkan responden manusia, sehingga fokus analisis sepenuhnya diarahkan pada performa objektif model *Artificial Intelligence* dalam menghasilkan sitasi. Dengan demikian, penelitian ini tidak hanya memperluas kajian tentang akurasi sitasi dalam konteks teknologi *AI*, tetapi juga memberikan kontribusi praktis dalam pengembangan sistem pencarian informasi yang lebih kredibel dan sesuai dengan standar akademik. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem pencarian informasi berbasis *AI* yang lebih akurat, transparan, dan sesuai dengan standar akademik.

Dari latar belakang artikel di atas maka peneliti tertarik meneliti lebih lanjut hal di atas dengan judul “Tingkat akurasi Sitasi dalam model pencarian informasi berbasis *AI*, studi Perbandingan (*Chat GPT*, *Perflexity* dan *Deepseek*)”.

2. TINJAUAN PUSTAKA

2.1 Akurasi

Akurasi yang dikemukakan Wang dan Strong menambahkan bahwa akurasi informasi adalah sejauh mana informasi tersebut mencerminkan realitas atau kebenaran dari objek yang diwakili. Mereka juga menunjukkan bahwa kualitas informasi yang rendah dapat menyebabkan kesalahan dalam pengambilan keputusan, yang pada gilirannya mempengaruhi efisiensi operasional dan kinerja organisasi secara keseluruhan (Wang dan Strong 1996). Sejalan dengan pandangan tersebut, Akurasi juga dipahami sebagai suatu derajat yang memberikan informasi sejauh mana pengukuran dekat dengan nilai sebenarnya. Istilah akurasi secara konseptual dekat dengan istilah validitas yakni sejauh mana pengukuran benar-benar mengukur karakteristik. Akurasi juga dinilai dari perbandingan dengan bahan referensi yang ada, hal itu bertujuan untuk meningkatkan validitas kesimpulan sesuai dengan pasal, peraturan perundang-undangan yang berlaku pada suatu perkara yang terjadi (Okayanti 2019).

Namun, dalam sistem berbasis kecerdasan buatan, akurasi menjadi lebih kompleks karena informasi dihasilkan secara dinamis melalui proses probabilistic, bukan semata – mata retrieval dari database tetap. Hal ini menyebabkan akurasi tidak lagi hanya bergantung pada kebenaran data, tetapi juga pada kemampuan model dalam mengontruksi informasi yang realistis. Dalam artikel ini

akurasi tidak hanya di pahami sebagai kebenaran informasi secara umum, tetapi dipersempit pada akurasi sitasi yang meliputi ketepatan nama penulis, kesesuaian judul, kejelasan tahun terbit, identitas penerbit atau jurnal, serta keterlacakan sumber. Akurasi menjadi aspek esensial karena berhubungan langsung dengan integritas ilmiah, kredibilitas karya akademik, dan keandalan sistem informasi berbasis *Artificial Intelligence (AI)* dalam mendukung penelitian. Oleh karena itu akurasi menjadi dasar dalam mengevaluasi kualitas referensi yang di hasilkan oleh model *AI* seperti *ChatGPT*, *Perplexity*, dan *DeepSeek*.

2.2 Sitasi

Menurut Hardi et al Sitasi atau kutipan dapat diartikan sebagai mengutip bagian dari karya orang lain yang kemudian dijadikan referensi karya yang sedang dibuat. Atau secara umum mensitasi berarti mengutip atau mencuplik kalimat dari tulisan sebelumnya pada karya ilmiah yang sedang ditulis dan nantinya akan disampaikan pada pembaca (Witono, Suyono, dan Abbas 2020). Sejalan dengan pengertian tersebut sitasi tidak hanya berfungsi sebagai teknik pengambilan referensi, tetapi juga merupakan salah satu proses penting dan wajib saat menulis karya ilmiah. Sitasi menjadi bentuk menunjukkan penghargaan dan penghormatan kepada hasil karya. Kejujuran dalam kekayaan hak intelektual dan karya penulis sebelumnya. Selain itu, dalam praktiknya, menulis sitasi berperan penting dalam mencegah penulis karya ilmiah melakukan plagiarisme (Suharso 2019).

Meskipun sitasi secara tradisional dipahami sebagai bentuk penghargaan akademik, dalam konteks sistem *AI*, fungsi sitasi mengalami pergeseran. Sitasi tidak lagi hanya menjadi tanggung jawab penulis, tetapi juga dihasilkan secara otomatis oleh sistem. Hal ini menimbulkan persoalan baru, yaitu apakah sitasi yang dihasilkan benar-benar merepresentasikan sumber yang valid atau hanya sekadar konstruksi teks yang terlihat meyakinkan *hallucination citation*. Dengan kata lain, keberadaan sitasi tidak selalu menjamin validitas informasi. Penelitian ini tidak hanya melihat keberadaan sitasi, tetapi juga menguji validitas dan keterlacakan sitasi tersebut. Hal ini penting karena dalam sistem *AI*, jumlah sitasi yang banyak tidak selalu mencerminkan kualitas akademik. Oleh karena itu, penelitian ini berfokus pada evaluasi kualitas sitasi, bukan sekadar kuantitasnya.

2.3 Citation Recall

Citation recall merupakan ukuran yang menunjukkan sejauh mana pernyataan ilmiah yang dihasilkan didukung oleh sitasi yang relevan. Indikator ini menekankan pada aspek kelengkapan referensi terhadap klaim yang disampaikan. *Recall* adalah proporsi jumlah dokumen yang dapat ditemukankembali oleh sebuah proses pencarian di sistem *Information Retrieval (IR)*. Rumusnya: Jumlah dokumen relevan yang ditemukan / Jumlah semua dokumen relevan di dalam koleksi. Lalu, *precision* adalah proporsi jumlah dokumen yang ditemukan dan dianggap relevan untuk kebutuhan si pencari informasi. Rumusnya: Jumlah dokumen relevan yang ditemukan / Jumlah semua dokumen yang ditemukan (NISAA PUTRI LESTARI). Selain itu recall atau biasa dinamakan sebagai perolehan adalah perbandingan dokumen ditemukan dengan jumlah total dokumen relevan dalam sistem. (Anum Arum Narudhu*, Indah Widdi Palupi, Zulfatun Sofiyani). Perolehan (*recall*) berhubungan dengan kemampuan sistem untuk memanggil dokumen yang relevan. Untuk menghitung nilai perolehan *recall* digunakan rumus sebagai berikut: (Hasugian, 2006: 5)

Oleh karena itu, citation recall menjadi indikator yang relevan untuk menilai sejauh mana sistem *AI* mampu menyediakan referensi secara lengkap terhadap informasi yang dihasilkan.

2.4 Perilaku Pencarian Informasi

Tingkat Akurasi Sitasi Pada Model Pencarian Informasi Berbasis AI: Study (Chat Gpt, Perflexity, Dan Deepseek)

Perilaku pencarian Informasi merupakan Gambaran sebagai hasil pengolahan data menjadi lebih berguna dan berarti bagi penerimanya yang menggambarkan peristiwa nyata yang digunakan untuk membuat Keputusan (Tukino 2020). Perilaku pencarian informasi mencakup pola dan tingkah laku manusia dalam mempertimbangkan, mencari, dan memanfaatkan informasi dari berbagai sumber dan media (Amalia Lubis et al. 2023).

Berdasarkan pemahaman martin halomoan lumbagaol perilaku pencarian informasi mengatakan bahwa pencarian informasi pada dasarnya adalah hasil dari serangkaian proses pengolahan data yang terorganisir dan bermakna. Selain itu, data tidak sekadar data yang dikumpulkan itu adalah hasil interpretasi dan analisis data yang dimodifikasi untuk memenuhi kebutuhan khusus pengguna (Martin Halomoan Lumbangaol 2020).

Sejalan dengan pandangan tersebut, kajian lebih lanjut mengenai perilaku pencarian informasi juga dijelaskan secara lebih sistematis melalui model tahapan yang dikemukakan oleh Ellis yang dimana terdiri dari beberapa tahapan yang saling berkaitan. Proses dimulai dari *starting*, yaitu mencari titik awal atau sumber pertama. Lalu *chaining*, menelusuri referensi dari sumber yang sudah ditemukan. Tahap *browsing* dilakukan dengan menjelajah informasi secara umum, kemudian *differentiating* untuk menilai dan memilih sumber yang relevan. Selanjutnya *monitoring* digunakan untuk memantau informasi terbaru, *extracting* untuk mengambil bagian penting dari sumber, *verifying* untuk memeriksa keakuratan, dan *ending* sebagai tahap penutup saat kebutuhan informasi telah terpenuhi (Ellis 1989).

Model Ellis dikembangkan dalam konteks pencarian informasi tradisional yang berpusat pada pengguna. Namun dalam era *AI*, tahapan seperti *starting* dan *browsing* telah diotomatisasi oleh sistem, yang berpotensi mengurangi keterlibatan aktif pengguna dalam proses pencarian informasi. Akibatnya, tahap verifikasi yang seharusnya menjadi tanggung jawab pengguna sering diabaikan karena pengguna cenderung percaya pada hasil yang diberikan *AI*. Sehingga dapat dikatakan bahwasanya Perilaku pencarian informasi memiliki peran penting dalam menilai ketepatan keakuratan sitasi, karena melalui proses pencarian yang terarah dan sistematis seseorang dapat memperoleh sumber yang valid serta relevan. Penelitian ini menekankan bahwa meskipun *AI* mampu membantu dalam proses pencarian informasi, tahap *verifying* tetap menjadi tanggung jawab pengguna. Oleh karena itu, rendahnya akurasi sitasi yang ditemukan dalam penelitian ini dapat dijelaskan melalui perspektif teori Ellis, khususnya pada lemahnya tahap verifikasi dalam penggunaan *AI*.

2.5 Mesin Pencarian Informasi Berbasis AI

Teknologi mesin pencarian informasi berbasis kecerdasan buatan atau *AI* dikembangkan untuk meningkatkan keefektifan serta ketepatan dalam proses pencarian data. Sistem ini mengandalkan kemampuan pemrosesan bahasa alami *Natural Language Processing (NLP)* dan algoritma pembelajaran mesin guna memahami maksud dan konteks dari permintaan pengguna (Sari, Masruri, dan Rosalia 2023). Mesin pencarian informasi berbasis *AI* dapat didefinisikan sebagai kemampuan suatu sistem komputer untuk meniru atau mensimulasikan proses-proses kecerdasan manusia (Russell dan Norving 2021).

Berdasarkan konsep tersebut, penerapan mesin pencarian informasi berbasis *AI* dapat dilihat pada berbagai platform modern yang mengintegrasikan teknologi *Natural Language Processing (NLP)* dan pembelajaran mesin, seperti *ChatGPT*, *DeepSeek*, dan *Perplexity*. Menurut Suharmawan *chatgpt* Adalah (*Generation Pre-Training Transformer*) adalah suatu kecerdasan buatan yang didukung oleh suatu sistem yang sudah dirancang menggunakan teknologi untuk membuat suatu sistem computer yang dapat meniru kemampuan intelektual manusia (Suharmawan 2023).

Perplexity mampu mengakses dan mengolah informasi terkini secara real time dari berbagai sumber di internet, sehingga hasil terjemahan dapat lebih kontekstual dan relevan dengan perkembangan isu terbaru (Nailannajah et al. 2025). *DeepSeek* adalah kecerdasan buatan yang dikembangkan oleh *DeepSeek AI* dan dirancang untuk mengoptimalkan pencarian serta pengolahan informasi secara cepat dan efisien. Platform ini menggabungkan *Natural Language Processing (NLP)* dengan algoritma pencarian cerdas sehingga mampu memahami pertanyaan kompleks menganalisis data, bahkan menghasilkan konten akademik (Rahman et al. 2025).

Melalui beragam platform tersebut, pemanfaatan *AI* dalam mesin pencarian informasi memiliki peran signifikan dalam meningkatkan ketepatan sitasi, karena teknologi ini mampu menelusuri, memeriksa, dan menampilkan sumber informasi yang sah serta relevan secara otomatis. Dengan demikian, penggunaan mesin pencarian berbasis *AI* tidak hanya mempercepat proses penelusuran data, tetapi juga mendukung peneliti dalam menjamin keabsahan dan keandalan referensi yang digunakan. Terdapat perbedaan mendasar antara sistem berbasis generatif dan sistem berbasis retrieval meskipun *AI* menawarkan kecepatan dan kemudahan dalam pencarian informasi. Model generatif seperti *ChatGPT* tidak selalu mengacu pada sumber yang terverifikasi secara langsung karena menghasilkan jawaban berdasarkan pola bahasa dan probabilitas. Sebaliknya, sistem yang mengintegrasikan pencarian real time seperti *Perplexity* memiliki keunggulan dalam transparansi sumber karena mereka dapat menampilkan referensi yang dapat ditelusuri. Hal ini menunjukkan bahwa arsitektur sistem sangat memengaruhi kualitas sitasi yang dihasilkan. Selain itu, munculnya fenomena seperti *hallucination* dan *overcitation* menunjukkan bahwa sistem *AI* belum sepenuhnya mampu memastikan bahwa referensi yang diberikannya akurat.

Dalam konteks penelitian ini, perbedaan karakteristik antara *ChatGPT*, *Perplexity*, dan *DeepSeek* menjadi dasar untuk menganalisis bagaimana arsitektur sistem memengaruhi tingkat akurasi sitasi yang dihasilkan. Dengan demikian, evaluasi tidak hanya berfokus pada hasil akhir berupa jumlah sitasi yang akurat, tetapi juga mempertimbangkan bagaimana mekanisme kerja masing-masing model *AI* berkontribusi terhadap kualitas informasi yang disajikan.

3. METODE

Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan metode evaluatif. Pendekatan kuantitatif digunakan karena data penelitian diolah dalam bentuk angka, mulai dari proses pengumpulan data, penafsiran, hingga penyajian hasil penelitian (Arikunto 2013). Metode evaluatif diterapkan untuk menilai tingkat akurasi sitasi terkhususnya dalam mengkaji ketepatan sitasi yang dihasilkan oleh tiga aplikasi pencarian informasi berbasis kecerdasan buatan, yaitu *ChatGPT*, *Perplexity*, dan *DeepSeek*. Data Penelitian dikumpulkan pada bulan Februari tahun 2026 dengan melakukan pengujian langsung terhadap ketiga platform *Artificial Intelligence (AI)*. Objek penelitian ini adalah sitasi atau daftar referensi yang dihasilkan oleh tiga platform kecerdasan buatan, yaitu *ChatGPT*, *Perplexity*, dan *DeepSeek*, ketika digunakan untuk menjawab pertanyaan akademik. Sitasi yang dibandingkan meliputi unsur bibliografis seperti nama penulis, tahun publikasi, judul sumber, hingga kelengkapan kutipan. Adapun subjek penelitian adalah keluaran berupa teks jawaban sitasi dari ketiga platform tersebut berdasarkan pertanyaan yang telah disusun oleh peneliti, yang kemudian dianalisis untuk menilai akurasi dan ketepatan sitasi yang dihasilkan. Populasi dalam penelitian ini adalah seluruh sitasi yang dihasilkan oleh sistem *AI* dalam menjawab pertanyaan akademik, sedangkan sampel penelitian berupa 10 pertanyaan akademik yang menghasilkan 54 klaim ilmiah yang memerlukan dukungan referensi.

Teknik pengumpulan data dalam penelitian ini menggunakan metode dokumentasi, yaitu dengan mengumpulkan dan mencatat data yang dihasilkan oleh aplikasi pencarian informasi berbasis kecerdasan buatan. Proses pengumpulan data diawali dengan Menyusun sepuluh daftar

Tingkat Akurasi Sitasi Pada Model Pencarian Informasi Berbasis AI: Study (Chat Gpt, Perflexity, Dan Deepseek)

pertanyaan penelitian yang telah disesuaikan dengan kriteria yang ditetapkan oleh peneliti, yaitu pertanyaan yang bersifat akademik, memerlukan sitasi atau rujukan ilmiah, serta relevan dengan konteks pencarian informasi akademik. Selanjutnya, pertanyaan yang telah disusun tersebut diajukan secara sama dan seragam kepada tiga aplikasi *ChatGPT*, *Perplexity*, dan *DeepSeek*, guna menjaga konsistensi dan keterbandingan hasil. Setelah itu, peneliti mengumpulkan seluruh jawaban yang diberikan oleh masing-masing aplikasi kecerdasan buatan, termasuk sitasi atau referensi yang dicantumkan dalam jawaban tersebut. Seluruh data yang diperoleh kemudian didokumentasikan dan disusun secara sistematis dalam bentuk tabel untuk memudahkan proses analisis, verifikasi, serta penilaian tingkat akurasi sitasi yang dihasilkan oleh masing-masing aplikasi kecerdasan buatan. Instrumen penelitian yang digunakan dalam studi ini berupa lembar evaluasi akurasi sitasi, yang berfungsi untuk menilai ketepatan sitasi yang dihasilkan oleh aplikasi pencarian informasi berbasis kecerdasan buatan. Penilaian dilakukan berdasarkan sejumlah indikator akurasi sitasi, meliputi kesesuaian judul dan sitasi sumber dengan isi jawaban, keberadaan dan keterlacakan sumber asli, ketepatan penulisan nama penulis, ketepatan tahun terbit, serta ketepatan penerbit atau jurnal. Indikator-indikator tersebut digunakan untuk memastikan bahwa sitasi yang disajikan tidak hanya relevan secara konteks, tetapi juga akurat secara bibliografis dan dapat diverifikasi melalui sumber ilmiah yang sah. Teknik analisis data dalam penelitian ini dilakukan secara kuantitatif dengan menggunakan pendekatan *citation recall*. Tahap awal analisis dilakukan dengan memverifikasi setiap sitasi yang dihasilkan oleh aplikasi pencarian informasi berbasis kecerdasan buatan melalui mesin pencari akademik, seperti *Google Scholar*, *Directory of Open Access Journals (DOAJ)*, serta situs resmi penerbit untuk memastikan keberadaan dan keabsahan sumber. Selanjutnya, berdasarkan dari 10 pertanyaan akademik yang telah disusun peneliti setiap jawaban dianalisis dan dipecah menjadi unit-unit pernyataan yang secara ilmiah memerlukan dukungan referensi. Tingkat *Citation Recall* kemudian dihitung menggunakan rumus sebagai berikut:

$$\text{Citation Recall} = \frac{\text{Jumlah sitasi akurasi}}{\text{Jumlah sitasi yang seharusnya ada}} \times 100\%$$

Hasil perhitungan *citation recall* selanjutnya diklasifikasikan ke dalam kategori tingkat akurasi sitasi, yaitu 81–100% (sangat akurat), 65–80% (akurat), 41–65% (cukup akurat), 21–40% (kurang akurat), dan 0–20% (tidak akurat). Tahap akhir analisis dilakukan dengan membandingkan nilai *citation recall* antara *ChatGPT*, *Perplexity*, dan *DeepSeek* untuk mengetahui model pencarian informasi berbasis *Artificial Intelligence (AI)* yang menghasilkan tingkat kelengkapan dan ketepatan sitasi paling tinggi.

4. HASIL DAN PEMBAHASAN

4.1 Hasil Penelitian

Penelitian ini bertujuan untuk mengukur tingkat akurasi sitasi yang dihasilkan oleh tiga model pencarian informasi berbasis kecerdasan buatan, yaitu *ChatGPT*, *Perplexity*, dan *DeepSeek*. Pengukuran dilakukan menggunakan pendekatan *citation recall*, yakni perbandingan antara jumlah sitasi yang dinilai akurat dan dapat diverifikasi dengan jumlah keseluruhan pernyataan yang seharusnya didukung oleh referensi ilmiah. Proses verifikasi dilakukan melalui penelusuran pada *Google Scholar*, *Directory of Open Access Journals (DOAJ)*, serta laman resmi penerbit. Sitasi dinyatakan akurat apabila memenuhi kriteria: Judul sesuai dengan sumber asli, Nama penulis benar, Tahun publikasi tepat, Sumber dapat ditemukan dan diverifikasi, Isi referensi mendukung pernyataan dalam jawaban *Artificial Intelligence (AI)*.

4.1.1 Hasil Rekapitulasi Verifikasi Sitasi ChatGPT

Tabel 1. Rekapitulasi Hasil Verifikasi Sitasi *ChatGPT*

No.	Keterangan	Jumlah
1	Sitasi seharusnya	54
2	Sitasi diberikan	51
3	Sitasi akurat	34

Sumber: Hasil Data Olahan Peneliti (2026)

Tabel 1 menyajikan rekapitulasi hasil verifikasi sitasi yang dihasilkan oleh *Chat GPT* berdasarkan 10 pertanyaan akademik yang diuji. Hasil identifikasi menunjukkan bahwa ada 54 sitasi yang seharusnya muncul. Angka ini dihitung dari semua klaim ilmiah yang ditemukan dalam jawaban *AI*. Setiap klaim ilmiah yang bersifat faktual dan memerlukan rujukan dihitung sebagai satu unit sitasi, dan sumber referensi harus disertakan. *ChatGPT* memberikan 51 sitasi dari 54 yang seharusnya ada, menunjukkan bahwa sebagian besar klaim telah disertai referensi, meskipun beberapa masih tidak memiliki sitasi. Selain itu, dari 51 sitasi yang diberikan, hanya 34 dinyatakan akurat atau sesuai dengan sumber yang diakui.

4.1.2 Hasil Rekapitulasi Verifikasi Sitasi *Perplexity*

Tabel 2. Rekapitulasi Hasil Verifikasi Sitasi *Perplexity*

No.	Keterangan	Jumlah
1	Sitasi seharusnya	54
2	Sitasi diberikan	50
3	Sitasi akurat	37

Sumber: Hasil Data Olahan Peneliti (2026)

Tabel 2 menyajikan rekapitulasi hasil verifikasi sitasi yang dihasilkan oleh *Perplexity* berdasarkan 10 pertanyaan akademik yang diuji. Hasil identifikasi menunjukkan bahwa ada 54 sitasi yang seharusnya muncul. Angka ini dihitung dari semua klaim ilmiah yang ditemukan dalam jawaban *AI*. Setiap klaim ilmiah yang bersifat faktual dan memerlukan rujukan dihitung sebagai satu unit sitasi, dan sumber referensi harus disertakan. *Perplexity* memberikan 50 sitasi dari 54 yang seharusnya ada, menunjukkan bahwa sebagian besar klaim telah disertai referensi, meskipun beberapa masih tidak memiliki sitasi. Selain itu, dari 51 sitasi yang diberikan, hanya 37 dinyatakan akurat atau sesuai dengan sumber yang diakui.

4.1.3 Hasil Rekapitulasi Verifikasi Sitasi *DeepSeek*

Tabel 3. Rekapitulasi Hasil Verifikasi Sitasi *DeepSeek*

No.	Keterangan	Jumlah
1	Sitasi seharusnya	54
2	Sitasi diberikan	69
3	Sitasi akurat	25

Sumber: Hasil Data Olahan Peneliti (2026)

Tabel 3 menyajikan rekapitulasi hasil verifikasi sitasi yang dihasilkan oleh model *DeepSeek* berdasarkan 10 pertanyaan akademik yang diuji. Hasil identifikasi menunjukkan bahwa terdapat 54 sitasi yang seharusnya muncul, yang diperoleh dari total klaim ilmiah dalam seluruh jawaban. Berbeda dengan model sebelumnya, *DeepSeek* memberikan 69 sitasi, jumlah yang melebihi sitasi yang seharusnya ada. Temuan ini menunjukkan bahwa model cenderung

Tingkat Akurasi Sitasi Pada Model Pencarian Informasi Berbasis AI: Study (Chat Gpt, Perflextity, Dan Deepseek)

memberikan referensi dalam jumlah yang lebih banyak daripada klaim ilmiah yang teridentifikasi. Kondisi ini dapat diinterpretasikan sebagai kecenderungan model untuk melakukan overcitation, yakni menyertakan lebih banyak sumber daripada yang secara substantif dibutuhkan. Namun demikian, dari 69 sitasi yang diberikan, hanya 25 sitasi yang dinyatakan akurat setelah melalui proses verifikasi. Sitasi akurat dalam konteks ini berarti referensi tersebut benar-benar dapat ditemukan dan relevan dengan klaim yang dirujuk. Rendahnya jumlah sitasi akurat dibandingkan total sitasi yang diberikan menunjukkan bahwa meskipun model aktif dalam menyertakan referensi, tingkat ketepatan dan validitasnya masih terbatas.

4.1.3 Hasil Citation Recall Tiga Model *Artificial Intelligence (AI)*

Tabel 4. Hasil citation recall tiga model *Artificial Intelligence (AI)*

No.	Model AI	Sitasi Seharusnya	Sitasi Akurat	Citation Recall	Kategori
1	Chat GPT	54	34	62,96%	Cukup akurat
2	Perplexity	54	37	68,52%	Akurat
3	DeepSeek	54	25	46,30%	Cukup akurat

Sumber: Hasil Data Olahan Peneliti (2026)

Tabel 4 menyajikan perbandingan nilai *citation recall* dari tiga model AI, yaitu *ChatGPT*, *Perplexity*, dan *DeepSeek*, berdasarkan 54 klaim ilmiah yang teridentifikasi dari 10 pertanyaan penelitian. Angka 54 tersebut menjadi jumlah sitasi yang seharusnya ada, kemudian dibandingkan dengan jumlah sitasi yang dinyatakan akurat setelah proses verifikasi.

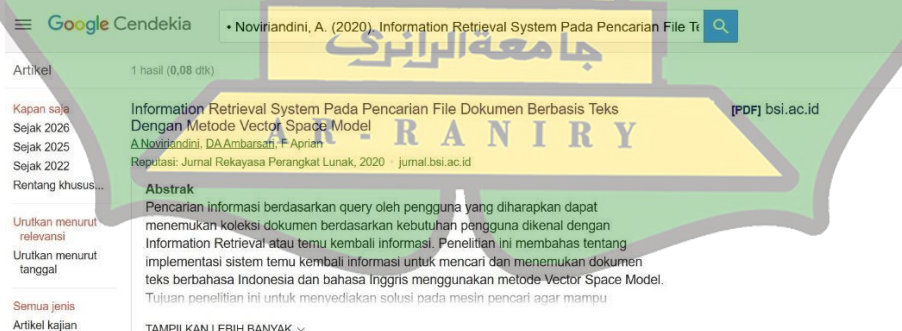
Hasil perhitungan menunjukkan bahwa model *ChatGPT* mendapatkan 34 sitasi akurat dari 54 sitasi yang seharusnya, yang menghasilkan nilai *citation recall* sebesar 62,96% dan kategorikan sebagai cukup akurat. Hasil ini menunjukkan bahwa lebih dari separuh klaim ilmiah yang dibuat telah didukung oleh sitasi yang relevan dan dapat diverifikasi, meskipun beberapa masih belum tersitasi dengan benar. Dibandingkan dengan dua model lain dalam penelitian ini, *Perplexity* memiliki performa terbaik itu memiliki 37 sitasi akurat dari 54 sitasi yang seharusnya, mendapatkan nilai *citation recall* sebesar 68,52%, dan termasuk dalam kategori akurat. Hasil ini menunjukkan bahwa *Perplexity* memiliki tingkat kelengkapan sitasi yang lebih tinggi. *DeepSeek* menunjukkan kecenderungan yang berbeda dengan menghasilkan lebih banyak sitasi daripada yang diperlukan (69 sitasi daripada 54 yang seharusnya). Fenomena *overcitation* ini menunjukkan bahwa model berusaha menambah banyak referensi untuk meningkatkan kredibilitas jawaban. Walau bagaimanapun, hanya 25 sitasi yang dianggap akurat. Artinya, kualitas dan ketepatan referensi tidak selalu berkorelasi dengan kuantitas sitasi. Dalam konteks akademik, situasi ini berpotensi menimbulkan bias otoritas, atau bias kepercayaan, karena pengguna mungkin menganggap jawaban lebih akurat karena banyaknya referensi yang disertakan. Secara komparatif, urutan performa *citation recall* tertinggi hingga terendah adalah *Perplexity* (68,52%), *ChatGPT* (62,96%), dan *DeepSeek* (46,30%). Temuan ini menunjukkan adanya variasi kemampuan antar model AI dalam menyediakan sitasi yang sesuai dengan klaim ilmiah yang dihasilkan. Dengan demikian, dalam konteks evaluasi kelengkapan sitasi, *Perplexity* menunjukkan kinerja paling optimal, diikuti oleh *ChatGPT*, sedangkan *DeepSeek* masih memerlukan peningkatan dalam aspek ketepatan dan kelengkapan rujukan.

4.2 Pembahasan

Hasil penelitian menunjukkan bahwa tiga model *Artificial Intelligence (AI)* (*ChatGPT*, *Perplexity*, dan *DeepSeek*) memiliki tingkat akurasi sitasi yang cukup akurat hingga akurat, tetapi

belum sangat akurat. *Perplexity* menerima nilai citation recall tertinggi dengan 68,52%, diikuti oleh *ChatGPT* dengan 62,96%, dan *DeepSeek* dengan 46,30%. Perbedaan ini menunjukkan bahwa masing-masing model memiliki arsitektur sistem yang berbeda, strategi penelusuran sumber, dan mekanisme integrasi referensi. Secara konseptual, temuan ini menguatkan pandangan bahwa model bahasa besar *Large Language Model (LLM)* masih menghadapi persoalan dalam memastikan keterkaitan langsung antara klaim dan referensi. Temuan ini sejalan dengan penelitian Liu et al (2023) yang menyatakan bahwa sistem pencarian generatif masih menghadapi tantangan dalam memastikan kesesuaian antara pernyataan dan sitasi. Oleh karena itu, penggunaan *Artificial Intelligence (AI)* dalam pencarian informasi akademik harus tetap disertai verifikasi manual oleh pengguna.

Dalam penelitian ini, kondisi ini ditemukan pada *ChatGPT* dan *DeepSeek* namun, beberapa referensi tidak dapat diverifikasi atau tidak sepenuhnya relevan dengan pernyataan yang disebutkan. *Perplexity* menunjukkan kinerja yang lebih baik karena setiap jawaban menyertakan tautan sumber secara langsung ke sistem. Transparansi ini memungkinkan pengguna untuk melakukan pengecekan keberadaan referensi dengan cepat. Hal ini menunjukkan bahwa integrasi sistem pencarian real-time dan model bahasa meningkatkan *citation recall*. Dengan kata lain, sistem yang menggabungkan kemampuan *generatif* dan *retrieval* cenderung memiliki keterlacakan sumber yang lebih baik daripada model yang sepenuhnya berbasis generatif. Meskipun demikian, *DeepSeek* menunjukkan fenomena overcitation, di mana jumlah sitasi melebihi jumlah klaim ilmiah yang seharusnya didukung oleh referensi. Meskipun kuantitas terlihat lebih rinci, hasil verifikasi menunjukkan bahwa banyak sitasi tidak akurat atau tidak relevan. Kondisi ini menunjukkan bahwa jumlah referensi tidak secara otomatis menjamin kualitas akademik. situasi ini berpotensi menimbulkan authority bias, di mana pengguna cenderung mempercayai jawaban karena banyaknya referensi yang dicantumkan, tanpa melakukan verifikasi lebih lanjut. Jika dikaitkan dengan teori perilaku pencarian informasi Ellis (1989), tahap *verifying* menjadi sangat krusial. *Artificial Intelligence (AI)* mampu membantu tahap *starting, browsing, dan extracting*, tetapi belum sepenuhnya optimal pada tahap *verifying*. Oleh karena itu, pengguna tetap harus melakukan verifikasi manual melalui *Google Scholar, DOAJ*, atau basis data ilmiah lainnya. Dengan demikian, penggunaan *Artificial Intelligence (AI)* lebih tepat diposisikan sebagai alat bantu (*assistive tool*) dalam proses pencarian informasi, bukan sebagai otoritas final dalam validasi referensi ilmiah.



Gambar 1: Salah satu contoh hasil keakuratan sitasi pada goggle scholar

Hasil ini sangat penting bagi mahasiswa dan dosen dari sudut pandang literasi informasi. Meskipun *AI* dapat mempercepat proses pencarian referensi, kekurangan kemampuan evaluasi akan meningkatkan kemungkinan penggunaan sitasi yang tidak valid. Akibatnya, pemahaman struktur bibliografis, peningkatan kemampuan evaluasi sumber, dan keterampilan verifikasi melalui basis data seperti *Google Scholar* dan *DOAJ* masih sangat penting dalam pendidikan tinggi. Secara keseluruhan, penelitian ini menunjukkan bahwa teknologi *Artificial AI* telah mengubah perilaku pencarian informasi, tetapi akurasi sitasi masih

Tingkat Akurasi Sitasi Pada Model Pencarian Informasi Berbasis AI: Study (Chat Gpt, Perflexity, Dan Deepseek)

membutuhkan pengembangan. Penelitian ini menyarankan pengembang model *AI* untuk menerapkan mekanisme verifikasi otomatis, meningkatkan ketepatan sistem pencarian, dan mengembangkan fitur audit sitasi untuk membuat sistem mereka lebih kredibel dan sesuai dengan standar akademik.

5. KESIMPULAN

Bahwasanya Penelitian ini dapat disimpulkan bahwa tingkat akurasi sitasi pada tiga model pencarian informasi berbasis AI menunjukkan perbedaan kinerja yang cukup mencolok. Perplexity memperoleh nilai citation recall tertinggi sebesar 68,52% dan berada pada kategori akurat, disusul oleh ChatGPT dengan nilai 62,96% dalam kategori cukup akurat, serta DeepSeek sebesar 46,30% yang juga tergolong cukup akurat. Data ini mengindikasikan bahwa walaupun ketiga model mampu menyajikan referensi ilmiah dalam respons akademik, tingkat ketepatan serta keterlacakan sumber yang dihasilkan masih belum mencapai tingkat sangat akurat. Penelitian ini juga memperlihatkan bahwa kuantitas sitasi tidak selalu sejalan dengan kualitasnya. DeepSeek, misalnya, menampilkan jumlah referensi yang melebihi kebutuhan (*overcitation*), namun sebagian referensi tersebut tidak sepenuhnya tepat atau sulit diverifikasi. Sebaliknya, Perplexity menunjukkan performa yang lebih konsisten dalam menyediakan sitasi yang relevan dan dapat ditelusuri. Hal ini mengisyaratkan bahwa sistem yang terintegrasi dengan mekanisme pencarian langsung dan transparansi sumber cenderung menghasilkan kualitas sitasi yang lebih baik.

Secara keseluruhan, penelitian ini menegaskan bahwa pemanfaatan AI dalam pencarian informasi akademik memang membantu mempercepat proses penelusuran dan penyusunan referensi, tetapi belum dapat sepenuhnya menggantikan proses pengecekan manual oleh pengguna. Dengan demikian, AI sebaiknya dipahami sebagai alat pendukung dalam proses pencarian informasi, bukan sebagai otoritas akhir dalam memastikan validitas referensi ilmiah. Temuan ini memberikan kontribusi terhadap penguatan literasi informasi di perguruan tinggi sekaligus menjadi rekomendasi bagi pengembang sistem AI untuk terus meningkatkan akurasi dan mekanisme verifikasi sitasi.

6. SARAN DAN UCAPAN TERIMAKASIH

Berdasarkan hasil penelitian ini, peneliti menyarankan agar mahasiswa, dosen, dan peneliti tetap melakukan verifikasi manual terhadap setiap sitasi yang dihasilkan oleh sistem AI melalui basis data ilmiah seperti Google Scholar, DOAJ, atau laman resmi penerbit guna memastikan keabsahan referensi. Bagi pengembang teknologi AI, disarankan untuk meningkatkan fitur audit sitasi, memperkuat integrasi dengan database ilmiah terindeks, serta mengembangkan sistem verifikasi otomatis untuk meminimalkan kesalahan referensi dan fenomena *hallucination citation*.

Penelitian selanjutnya diharapkan dapat memperluas jumlah sampel pertanyaan, menambahkan model AI lainnya, serta menggunakan indikator evaluasi tambahan seperti citation precision dan relevance score agar hasil penelitian semakin komprehensif. Penulis menyampaikan terima kasih kepada seluruh pihak yang telah memberikan dukungan dalam proses penyusunan dan pelaksanaan penelitian ini sehingga penelitian dapat diselesaikan dengan baik.

DAFTAR PUSTAKA

- Abouammoh, Noura, Khalid Alhasan, Fadi Aljamaan, dan Rupesh Raina. 2025. "Perceptions and Earliest Experiences of Medical Students and Faculty With ChatGPT in Medical Education: Qualitative Study Corresponding Author :." 11:1–12. <https://doi.org/10.2196/63400>.
- Amalia Lubis, Ryzky, Nur Alisa, Surya Ningsi Wijaya Sitompul, Ahmad Tuah Wijaya Saragih, Franindya Purwaningtyas, dan Uin Sumatera Utara. 2023. "Dawatuna: Journal of Communication and Islamic Broadcasting Model Perilaku Pencarian Informasi: Analisis Teori Perilaku Pencarian

- Informasi Menurut Wilson.” *Dawatuna: Journal of Communication and Islamic Broadcasting* 3 (3): 1006–15. <https://doi.org/10.47476/dawatuna.v3i3.3028>.
- Arikunto, Suharsimi. 2013. *Prosedur Penelitian Suatu Pendekatan Praktik*. 15 ed. Jakarta: PT. Rineka Cipta.
- Ellis, David. 1989. “A behavioural approach to information retrieval system design.” *Journal of documentation* 45 (3): 171-212. <https://doi.org/http://dx.doi.org/10.1108/eb026843>.
- Erizal, Rahma Maulida, Annisa Safitra, Putri Fauzelly Mulia, dan Zul Azmi. 2024. “Analisis Pemanfaatan Kecerdasan Buatan Menggunakan Platform Chat- GPT Untuk Mendukung Proses Pendidikan Bagi Mahasiswa” 2 (1).
- Liu, Nelson F., Tianyi Zhang, dan Percy Liang. 2023. “Evaluating Verifiability in Generative Search Engines.” *Findings of the Association for Computational Linguistics: EMNLP 2023*, 7001–25. <https://doi.org/10.18653/v1/2023.findings-emnlp.467>.
- Maheshwari, Harsh, Srikanth Tenneti, dan Alwarappan Nakkiran. 2025. “CiteFix: Enhancing RAG Accuracy Through Post-Processing Citation Correction.” *Proceedings of the Annual Meeting of the Association for Computational Linguistics* 6:310–17. <https://doi.org/10.18653/v1/2025.acl-industry.23>.
- Martin Halomoan Lumbangaol, M.Rasyid Ridho. 2020. “Rancang Bangun Sistem Informasi Penjualan dan Penyewaan Properti Berbasis WEB Di Kota Batam.” *Jurnal Comasie* 01 (03): 83–92. <http://ejournal.upbatam.ac.id/index.php/comasiejurnal>.
- Nahla, Febri. 2024. “Jurnal Perpustakaan Universitas Airlangga : Media Informasi dan Komunikasi Kepustakawanan ON INFORMATION SEEKING BEHAVIOR” 14 (2): 69–75. <https://doi.org/https://doi.org/10.20473/jpua.v14i2.2024.69-75>.
- Nailannajah, Lubna, Maudlotun Nisa, Achmad Satori, Abdullah Maulani, Darsita Suparno, dan Muhammad Zacky Mubarak. 2025. “Hasil Terjemahan Perplexity : Analisis Akurasi , Keberterimaan , dan Keterbacaan pada Berita Pariwisata dari Situs Aljazeera . net.” *Jurnal Ilmu Bahasa Arab* 03 (02): 170–85. <https://jurnal.iaibafa.ac.id/index.php/Arabia/>.
- Okayanti, Putu Widhy. 2019. *Presisi dan Akurasi*.
- Rahman, Anichur, Shahariar Hossain Mahir, Md Tanjum An Tashrif, Md Ahsan Karim, Airin Afroj Aishi, Dipanjali Kundu, Tanoy Debnath, et al. 2025. “Comparative analysis based on DeepSeek, ChatGPT, and Google Gemini: Features, techniques, performance, future prospects.” *Systems and Soft Computing* 7. <https://doi.org/10.1016/j.sasc.2025.200396>.
- Rensburg, Janse van. 2025. “AI-Powered Citation Auditing: A Zero-Assumption Protocol for Systematic Reference Verification in Academic Research.” *Arxiv.Org*, 1–10. <https://arxiv.org/pdf/2304.09848>.
- Rosiana, Caltira, Triana Wuri Cahyanti, dan Lina Eka Rahayu. 2024. “Artificial Intelligence (AI) ChatGPT Meningkatkan Motivasi Mahasiswa pada Pembelajaran Morphosyntax” 11 (2): 98–105. <https://doi.org/10.31629/kiprah.v11i1.6563>.
- Russell, Struat, dan Peter Norving. 2021. *Artificial Intelligence: A Modern Approach, Global Edition*. Bandung: PT. Rineka Cipta.
- Sari, Kartika Puspita, Anis Masruri, dan Desfiana Ramdhani Rosalia. 2023. “Optimalisasi Temu Kembali Informasi Dengan Teknologi Kecerdasan Buatan di Perpustakaan.” *JIPi (Jurnal Ilmu Perpustakaan dan Informasi)* 8 (2): 349. <https://doi.org/10.30829/jipi.v8i2.17775>.
- Suharmawan, Wahid. 2023. “Education Journal : Journal Education Research and Development.” *Education Journal : Journal Education Research and Development* 7 (2): 158–66.

Tingkat Akurasi Sitasi Pada Model Pencarian Informasi Berbasis AI: Study (Chat Gpt, Perflexity, Dan Deepseek)

<https://doi.org/https://doi.org/10.32511/flash.v9i1.1069>.

Suharso, S. 2019. *Tata Cara Penulisan Sitasi dan Daftar Pustaka dalam Karya Ilmiah*. Yogyakarta.

Tukino, Tukino. 2020. "Rancang Bangun Sistem Informasi E-Marketing Pada Pt Pulau Cahaya Terang." *Computer Based Information System Journal* 8 (1): 25–34. <https://doi.org/10.33884/cbis.v8i1.1680>.

Wang, Richard, dan Diane Strong. 1996. "Beyond Accuracy: What Data Quality Means to Data Consumers." *Journal of Management Information Systems* 12 (4): 5–33. https://courses.washington.edu/geog482/resource/14_Beyond_Accuracy.pdf.

Witono, Hardi, Tri Suyono, dan Said Ti Abbas. 2020. *Manajemen Daftar Pustaka pada Karya Tulis Ilmiah dengan Mendeley*. Bandung: CV. Media Sains Indonesia.

