

# TINGKAT AKURASI SITASI PADA MODEL Pencarian Informasi BERBASIS AI: STUDI PERBANDINGAN (CHAT GPT, PERPLEXITY, DAN DEEPSEEK)

Tri Wahyudi Putra; Ruslan; Nazaruddin; Mukhtaruddin; T.Mulkan Safri  
UIN Ar-Raniry Banda Aceh

Korespondensi: [210503057@student.ar-raniry.ac.id](mailto:210503057@student.ar-raniry.ac.id)

Naskah diterima: 06-04-2026, direvisi: 31-05-2026, disetujui: 02-06-2026

## ABSTRACT

*The proliferation of artificial intelligence (AI) based information retrieval systems has substantially transformed academic literature search practices, particularly through the automated generation of in-text citations. While this development accelerates information retrieval, it raises critical concerns regarding citation accuracy, source traceability, and academic integrity. This study analyzes and compares citation accuracy levels across three widely used AI platforms: ChatGPT (OpenAI), Perplexity AI, and DeepSeek. A descriptive quantitative approach with an evaluative method was employed, using citation recall as the primary metric - defined as the ratio of verifiable and accurate citations to the total number of scientific claims requiring references. Ten standardized academic questions were submitted to each platform, yielding 54 scientific claims requiring citation support. All citations were subsequently verified through academic databases (Google Scholar and DOAJ) against five bibliographic criteria: accuracy of author name, title, publication year, journal or publisher identity, and content relevance. Results indicate that Perplexity achieved the highest citation recall of 68.52% (accurate category), followed by ChatGPT at 62.96% (fairly accurate), and DeepSeek at 46.30% (fairly accurate). Accuracy categories were determined based on a five-tier classification scale: 81–100% (highly accurate), 65–80% (accurate), 41–65% (fairly accurate), 21–40% (low accuracy), and 0–20% (inaccurate). DeepSeek exhibited a tendency toward over citation, supplying 69 citations for 54 required claims, yet achieving the lowest verified accuracy. These findings confirm that AI functions as an assistive tool in academic information retrieval; however, manual verification remains indispensable to ensure citation validity, scholarly reliability, and academic integrity.*

## ABSTRAK

Perkembangan sistem pencarian informasi berbasis kecerdasan buatan (AI) telah mengubah praktik penelusuran literatur akademik secara signifikan, khususnya melalui kemampuan menghasilkan sitasi secara otomatis. Meski demikian, aspek ketepatan, konsistensi, dan keterlacakan referensi yang dihasilkan masih menjadi isu kritis dalam konteks integritas ilmiah. Penelitian ini bertujuan menganalisis dan membandingkan tingkat akurasi sitasi pada ChatGPT, Perplexity, dan DeepSeek menggunakan pendekatan kuantitatif deskriptif dengan metode evaluatif. Indikator utama yang digunakan adalah *citation recall*, yakni rasio antara sitasi yang terverifikasi dan akurat terhadap total klaim ilmiah yang membutuhkan dukungan referensi. Sebanyak sepuluh pertanyaan akademik terstandar diajukan kepada masing-masing platform dan menghasilkan 54 klaim ilmiah yang memerlukan sitasi. Seluruh sitasi diverifikasi melalui Google Scholar dan DOAJ berdasarkan lima kriteria bibliografis: ketepatan nama penulis, judul, tahun terbit, identitas jurnal atau penerbit, dan relevansi isi referensi terhadap klaim. Hasil penelitian menunjukkan Perplexity memperoleh *citation recall* tertinggi sebesar 68,52% (kategori akurat), diikuti ChatGPT sebesar 62,96% (cukup akurat), dan DeepSeek sebesar 46,30% (cukup akurat). Pengkategorian akurasi didasarkan pada skala klasifikasi lima tingkat: 81-100% (sangat akurat), 65-80% (akurat), 41-65% (cukup akurat), 21-40% (kurang akurat), dan 0-20% (tidak akurat).

*DeepSeek* menunjukkan kecenderungan *overcitation* dengan menghasilkan 69 sitasi untuk 54 klaim, namun memperoleh akurasi terverifikasi terendah. Temuan ini mempertegas bahwa *AI* berperan sebagai alat bantu dalam pencarian informasi akademik, sementara verifikasi manual tetap diperlukan untuk menjamin validitas sitasi, reliabilitas, dan kredibilitas ilmiah.

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

## 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*, pengguna 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 pengguna 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 *AI* telah mengubah pola pencarian literatur ilmiah. Jika sebelumnya pencarian referensi membutuhkan waktu yang lama, kini cukup dengan memasukkan pertanyaan singkat, sistem *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 sistem *AI* menghasilkan referensi yang seolah-olah benar, padahal tidak

Volume 10, Number 1, June 2026 31

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*, *NeevaAI*, *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, Tenneti, 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.

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*, *Perplexity*, 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 dengan judul “Tingkat akurasi Sitasi dalam model pencarian informasi berbasis *AI*, studi Perbandingan (*Chat GPT*, *Perplexity*, dan *Deepseek*)”.

## 1. TINJAUAN PUSTAKA

### 1.1 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 *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*.

## 1.2 Sitasi

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.

## 1.3 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 (Lestari 2016). Selain itu *recall* atau biasa dinamakan sebagai perolehan adalah perbandingan dokumen ditemukan dengan jumlah total dokumen relevan dalam sistem. Perolehan (*recall*) berhubungan dengan kemampuan sistem untuk memanggil dokumen yang relevan (Hasugian, 2006). Oleh karena itu, *citation recall* menjadi indikator yang relevan untuk menilai sejauh mana sistem *AI* mampu menyediakan referensi secara lengkap terhadap informasi yang dihasilkan.

## 1.4 Perilaku Pencarian Informasi

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*.

## 1.5 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*.

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. erdapat 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.

## 2. 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 seluruh sitasi yang dihasilkan oleh ketiga platform kecerdasan buatan dalam menjawab pertanyaan akademik. Dari jawaban tersebut diidentifikasi total 54 klaim ilmiah yang memerlukan dukungan referensi (denominator) dalam perhitungan citation recall. Instrumen pengumpulan data berupa sepuluh pertanyaan akademik yang disusun secara purposive berdasarkan tujuan penelitian, yaitu mengevaluasi kualitas sitasi yang dihasilkan oleh platform berbasis *AI*. Pemilihan pertanyaan dilakukan dengan mempertimbangkan validitas isi content validity, yakni sejauh mana butir instrumen merepresentasikan konstruk yang hendak diukur, dalam hal ini kemampuan platform *AI* menghasilkan sitasi akademik yang relevan dan akurat. Oleh karena itu, setiap pertanyaan disusun berdasarkan tiga kriteria: (1) relevan dengan bidang ilmu informasi dan kecerdasan buatan; (2) memerlukan dukungan referensi ilmiah yang dapat diverifikasi; dan (3) memiliki tingkat spesifisitas yang memadai untuk mendorong sistem *AI* menghasilkan sitasi konkret.

Untuk meminimalkan potensi bias, seluruh pertanyaan diberikan dalam bentuk dan redaksi yang identik kepada masing-masing platform, yaitu ChatGPT, Perplexity, dan DeepSeek, sehingga memungkinkan konsistensi perlakuan dan komparabilitas hasil. Selain itu, instrumen ditelaah melalui expert judgment oleh dosen pembimbing atau ahli pada bidang metodologi penelitian/ilmu informasi guna memastikan kesesuaian isi pertanyaan dengan tujuan penelitian. Pendekatan expert judgment digunakan karena validitas isi pada instrumen penelitian umumnya dilakukan melalui penilaian pakar terhadap relevansi dan representativitas item terhadap konstruk yang diukur. Verifikasi sitasi dilakukan melalui Google Scholar dan DOAJ dengan lima kriteria akurasi, yaitu: (1) kesesuaian judul dengan sumber asli; (2) ketepatan nama penulis; (3) ketepatan tahun publikasi; (4) ketepatan identitas jurnal atau penerbit; dan (5) relevansi isi referensi terhadap klaim yang dirujuk. Sitasi dinyatakan akurat apabila memenuhi seluruh kriteria tersebut secara bersamaan. Tingkat citation recall dihitung menggunakan rumus 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) (Rahayu, Harijanto, and Lesmono 2018). Tahap akhir analisis dilakukan dengan membandingkan nilai *citation recall* antara *ChatGPT*, *Perplexity*, dan *DeepSeek* untuk mengetahui model pencarian informasi berbasis *AI* yang menghasilkan sitasi paling lengkap dan tepat dalam konteks akademik

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Hasil Penelitian

Penelitian ini mengukur tingkat akurasi sitasi yang dihasilkan oleh tiga model pencarian informasi berbasis kecerdasan buatan, yaitu ChatGPT, Perplexity, dan DeepSeek, menggunakan pendekatan citation recall. Verifikasi dilakukan melalui Google Scholar dan Directory of Open Access Journals (DOAJ) dengan lima kriteria bibliografis: (1) kesesuaian judul dengan sumber asli, (2) ketepatan nama penulis, (3) ketepatan tahun publikasi, (4) ketepatan identitas jurnal atau penerbit, dan (5) relevansi isi referensi terhadap klaim yang dirujuk. Sitasi dinyatakan akurat hanya apabila memenuhi seluruh lima kriteria secara bersamaan. Data dikumpulkan dari 10 pertanyaan akademik terstandar yang menghasilkan total 54 klaim ilmiah sebagai denominator tetap dalam perhitungan citation recall ketiga model.

Tabel 1 berikut menyajikan rekapitulasi terpadu hasil verifikasi sitasi dari ketiga model AI yang diteliti, disertai nilai citation recall masing-masing.

Tabel 1. Rekapitulasi Verifikasi Sitasi dan Citation Recall Tiga Model AI

No.	Keterangan	Chatgpt	Perplexity	Deepseek
1	Sitasi seharusnya (denominator)	54	54	54
2	Sitasi diberikan	51	50	69
3	Sitasi akurat	34	37	25
4	Citation recall	62,96%	68,52%	46,30%
5	Kategori Akurasi	Cukup akurat	Akurat	Cukup Akurat

Sumber: Hasil Data Olahan Peneliti (2026)

Berdasarkan Tabel 1, ketiga model menggunakan denominator yang sama yaitu 54 klaim ilmiah, sehingga perbandingan antar model dapat dilakukan secara langsung dan setara. Hasil pengukuran citation recall menunjukkan rentang kinerja yang cukup lebar, dari nilai terendah 46,30% (DeepSeek) hingga tertinggi 68,52% (Perplexity) (ChatGPT) berada di posisi menengah dengan nilai 62,96%. Pola distribusi ini mengindikasikan bahwa tidak terdapat model yang secara dominan unggul di semua dimensi, namun Perplexity secara konsisten memperlihatkan keunggulan komparatif dalam akurasi sitasi pada konteks penelitian ini. Temuan yang paling mencolok adalah fenomena overcitation pada DeepSeek menyediakan 69 sitasi untuk 54 klaim ilmiah, DeepSeek menghasilkan volume referensi tertinggi sekaligus memiliki tingkat keakuratan per sitasi yang paling rendah di antara ketiga model. Kondisi ini menunjukkan bahwa strategi overcitation yang diterapkan DeepSeek tidak berkorelasi positif dengan kualitas akademik, melainkan justru meningkatkan proporsi sitasi yang tidak dapat diverifikasi. Sebaliknya, ChatGPT dan Perplexity menunjukkan pola yang lebih konservatif dalam pemberian sitasi, masing-masing menyediakan 51 dan 50 sitasi dari 54 yang diperlukan. Perbandingan ini menegaskan bahwa Perplexity tidak hanya unggul dalam citation recall, tetapi juga dalam ketepatan sitasi per unit yang diberikan, menjadikannya model paling efisien dari sisi rasio kuantitas dalam penelitian. 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.

### 3.2 Pembahasan

Hasil penelitian menunjukkan bahwa ketiga model AI (*ChatGPT*, *Perplexity*, dan *DeepSeek*) menghasilkan sitasi akademik dengan tingkat akurasi yang bervariasi namun secara keseluruhan belum mencapai kategori sangat akurat. Perbedaan kinerja ini tidak semata-mata mencerminkan kualitas model secara umum, melainkan secara fundamental mencerminkan perbedaan arsitektur sistem yang mendasari masing-masing platform. Perplexity sebagai sistem yang mengintegrasikan retrieval real-time dari sumber daring dengan kemampuan generatif, memiliki akses langsung ke halaman web yang dikutip sehingga dapat menyertakan tautan yang dapat ditelusuri. ChatGPT berbasis model generatif murni (Large Language Model) yang menghasilkan sitasi berdasarkan pola statistik dalam data pelatihan, bukan dengan mengakses sumber secara langsung. DeepSeek, meskipun mengklaim integrasi pencarian, menunjukkan perilaku overcitation yang mengindikasikan strategi generatif yang lebih agresif dalam melampirkan referensi tanpa verifikasi yang memadai.

Perbedaan arsitektur ini secara langsung menjelaskan mengapa citation recall Perplexity (68,52%) lebih tinggi 5,56 poin dibandingkan ChatGPT (62,96%) dan 22,22 poin di atas DeepSeek (46,30%). Sistem berbasis retrieval-augmented generation (RAG) seperti Perplexity cenderung menghasilkan sitasi yang lebih dapat diverifikasi karena referensi yang disertakan berasal dari dokumen nyata yang diakses pada saat kueri. Hal ini sejalan dengan temuan Maheshwari, Tenneti, dan Nakkiran (2025) yang menunjukkan bahwa model berbasis RAG memiliki potensi peningkatan akurasi sitasi hingga 15% dibandingkan model generatif murni melalui pencocokan semantik dan keyword. Temuan penelitian ini mengkonfirmasi pola tersebut secara empiris dalam konteks penggunaan akademik di Indonesia.

Fenomena overcitation yang ditemukan pada DeepSeek perlu dianalisis lebih mendalam dari perspektif mekanisme generatif. Ketika sebuah model bahasa besar *Large Language Model (LLM)* tidak memiliki akses langsung ke sumber yang relevan, model tersebut cenderung menghasilkan referensi yang secara permukaan tampak meyakinkan (plausible-sounding citations) namun tidak dapat diverifikasi.

Strategi overcitation dapat diinterpretasikan sebagai mekanisme kompensasi: model berusaha meningkatkan kredibilitas jawaban melalui kuantitas referensi ketika ia tidak dapat memastikan kualitas masing-masing referensi. Ini merupakan manifestasi dari fenomena hallucination citation sebagaimana didefinisikan oleh Abouammoh et al. (2025), yaitu kondisi di mana sistem *AI* menghasilkan referensi yang secara format terlihat valid namun tidak memiliki basis dalam kenyataan ilmiah yang dapat ditelusuri. Dalam studi ini, bukti overcitation DeepSeek 69 sitasi untuk 54 klaim, dengan hanya 25 yang akurat memberikan ukuran empiris terhadap besaran fenomena ini: sekitar 63,8% dari sitasi yang diberikan *DeepSeek* tidak dapat diverifikasi melalui *Google Scholar* maupun DOAJ. Temuan ini memiliki implikasi penting terhadap potensi authority bias dalam penggunaan *AI* untuk penelitian akademik. Authority bias adalah kecenderungan kognitif di mana individu cenderung menerima informasi tanpa pertanyaan karena disertai banyak referensi atau sumber otoritatif. Dalam konteks penggunaan DeepSeek, pengguna yang tidak melakukan verifikasi independen akan dihadapkan pada respons yang tampak sangat terdokumentasi 69 sitasi, padahal lebih dari dua pertiga referensinya tidak dapat diverifikasi. Risiko ini secara khusus signifikan bagi mahasiswa yang sedang belajar menulis karya ilmiah, di mana otomasi sitasi dapat menggantikan pengembangan kemampuan evaluasi sumber yang seharusnya menjadi kompetensi inti akademik.

Studi Liu, Zhang, dan Liang (2023) mengidentifikasi pola serupa: dalam evaluasi terhadap empat mesin pencari generatif (*Bing Chat*, *NeevaAI*, *Perplexity*, dan *YouChat*), ditemukan bahwa meskipun jawaban tampak informatif, hanya 51,5% kalimat yang didukung sitasi secara memadai dan 74,5% sitasi benar-benar sesuai dengan isi kalimat. Hasil penelitian ini menghasilkan angka yang lebih rendah untuk *ChatGPT* (66,67% *precision internal*) dan jauh lebih rendah untuk DeepSeek (36,23%), mengindikasikan bahwa permasalahan akurasi sitasi pada model *AI* masih belum terpecahkan secara sistematis. Jika dikaitkan dengan model perilaku pencarian informasi Ellis (1989), temuan penelitian ini dapat dipetakan secara sistematis ke dalam kedelapan tahap yang diidentifikasi Ellis. Pada tahap starting dan browsing, ketiga model *AI* menunjukkan kemampuan yang memadai dalam menyediakan titik awal dan eksplorasi informasi yang relevan. Pada tahap chaining, model seperti Perplexity mampu menelusuri referensi secara berantai dengan tingkat akurasi yang lebih baik. Namun pada tahap *verifying*, yaitu tahap yang paling kritis dalam konteks integritas akademik, ketiga model memperlihatkan keterbatasan yang signifikan.

Nilai citation recall yang berkisar antara 46,30% hingga 68,52% menunjukkan bahwa setidaknya 31,48% hingga 53,70% klaim ilmiah yang memerlukan referensi tidak didukung oleh sitasi yang dapat diverifikasi. Ini secara langsung mengarah pada tahap *verifying* yang tidak optimal, di mana sistem *AI*, alih-alih membantu pengguna melakukan verifikasi, justru menyajikan referensi yang memberikan ilusi verifikasi tanpa substansi yang dapat diperiksa. Dengan demikian, dalam kerangka Ellis, *AI* paling berkontribusi pada fase otomatisasi (starting, browsing, extracting), namun paling lemah pada fase epistemik (*verifying*), yang ironisnya merupakan fase yang paling menentukan kualitas karya ilmiah.

Dari perspektif generalisasi temuan, nilai citation recall yang diperoleh dalam penelitian ini perlu diinterpretasikan dengan mempertimbangkan beberapa keterbatasan metodologis. Pertama, penggunaan 10 pertanyaan akademik yang terfokus pada bidang ilmu informasi dan kecerdasan buatan menghasilkan denominator 54 klaim, yang merupakan sampel yang relatif terbatas untuk generalisasi lintas disiplin. Temuan ini mungkin tidak sepenuhnya representatif untuk bidang keilmuan lain seperti kedokteran, hukum, atau teknik, di mana karakteristik dan standar sitasi berbeda secara substansial. Kedua, karena verifikasi dilakukan hanya melalui *Google Scholar* dan DOAJ, ada kemungkinan bahwa beberapa sitasi yang dikategorikan sebagai tidak akurat sebenarnya merujuk pada sumber yang valid namun tidak terindeks dalam kedua basis data tersebut.

Ketiga, perlu diperhatikan bahwa penelitian ini tidak memasukkan ukuran ketidakpastian statistik formal (seperti confidence interval) karena sifatnya yang bersifat evaluatif terhadap keluaran sistem, bukan survei berbasis sampel probabilistik. Meskipun demikian, dengan denominator yang sama dan prosedur verifikasi yang konsisten, perbandingan relatif antar model tetap dapat diandalkan sebagai temuan deskriptif yang bermakna. Temuan penelitian ini sejalan dengan hasil audit sitasi berbasis *AI* yang dilakukan oleh van Rensburg (2025), yang menemukan bahwa sekitar 20% sitasi yang dihasilkan *AI* mengandung kesalahan. Namun penelitian ini menemukan tingkat kesalahan yang lebih tinggi, khususnya untuk DeepSeek di mana lebih dari separuh sitasi yang diberikan tidak dapat diverifikasi. Hal ini mungkin dipengaruhi oleh perbedaan konteks penggunaan, jenis pertanyaan, dan standar verifikasi yang diterapkan. Temuan van Rensburg (2025) juga menunjukkan bahwa *AI* dapat mendeteksi referensi palsu, artikel yang ditarik (retracted), dan jurnal predator dengan tingkat verifikasi 91,7%, yang justru menunjukkan bahwa *AI* memiliki potensi untuk menjadi alat audit sitasi, namun bukan penghasil sitasi yang dapat dipercaya sepenuhnya. Ketegangan antara kapasitas *AI* sebagai alat deteksi kesalahan dan keterbatasannya sebagai generator referensi yang akurat ini membentuk paradoks mendasar yang perlu disadari oleh pengguna akademik.

Dari sudut pandang literasi informasi, temuan ini mempertegas urgensi pengembangan kompetensi evaluasi sumber di kalangan akademisi, terutama mahasiswa yang berada pada fase pembentukan kebiasaan riset. Meskipun *AI* secara signifikan mempercepat tahap starting dan browsing dalam pencarian informasi, ketergantungan berlebihan tanpa kemampuan verifikasi kritis berpotensi menghasilkan karya ilmiah yang tampak terdokumentasi dengan baik namun secara akademik lemah. Dalam konteks perguruan tinggi di Indonesia, di mana penggunaan platform *AI* generatif dalam penulisan tugas dan skripsi semakin meluas, temuan ini menjadi dasar empiris yang kuat untuk memasukkan pelatihan verifikasi sitasi *AI* ke dalam kurikulum literasi informasi. Secara keseluruhan, penelitian ini menegaskan bahwa *AI* lebih tepat diposisikan sebagai assistive tool yang mempercepat proses pencarian awal, bukan sebagai otoritas final dalam validasi referensi ilmiah. Kombinasi optimal yang disarankan adalah menggunakan *AI* untuk identifikasi topik dan eksplorasi literatur awal, kemudian memverifikasi setiap referensi secara manual melalui Google Scholar, DOAJ, atau basis data terindeks lainnya sebelum menyertakannya dalam karya ilmiah.

## 5. KESIMPULAN

Hasil penelitian, dapat disimpulkan bahwa tingkat akurasi sitasi yang dihasilkan oleh model pencarian informasi berbasis kecerdasan buatan menunjukkan variasi performa yang berbeda dalam mendukung kebutuhan informasi akademik. Hasil analisis memperlihatkan bahwa kualitas sitasi tidak hanya ditentukan oleh jumlah referensi yang disajikan, tetapi juga oleh tingkat relevansi, keterlacakan, dan kesesuaian referensi terhadap informasi yang disampaikan. Temuan penelitian mengindikasikan bahwa sistem *AI* dengan mekanisme penyediaan sumber yang lebih transparan dan terhubung dengan proses penelusuran informasi cenderung menghasilkan sitasi yang lebih dapat diverifikasi dibandingkan model yang menghasilkan referensi secara lebih generatif. Penelitian ini sekaligus menjawab rumusan masalah mengenai perbedaan tingkat akurasi sitasi antarplatform *AI* dalam konteks pencarian informasi akademik. Perbedaan tersebut menunjukkan bahwa penggunaan *AI* dalam aktivitas akademik masih memiliki keterbatasan dalam menjamin validitas referensi secara konsisten. Oleh karena itu, pemanfaatan teknologi *AI* perlu diposisikan sebagai sarana pendukung dalam proses penelusuran dan pengelolaan informasi, bukan sebagai sumber utama yang digunakan tanpa melalui proses evaluasi kritis terhadap validitas dan kredibilitas referensi yang dihasilkan.

Merujuk pada temuan penelitian, mahasiswa, dosen, dan peneliti disarankan untuk melakukan proses verifikasi terhadap setiap sitasi yang diperoleh melalui sistem *AI* dengan membandingkannya pada sumber ilmiah yang kredibel, seperti basis data akademik, jurnal terindeks, maupun laman resmi penerbit. Langkah ini penting untuk memastikan kesesuaian, ketepatan, dan keabsahan referensi yang digunakan dalam penyusunan karya ilmiah. Selain itu, pengembang teknologi *AI* diharapkan dapat meningkatkan kualitas sistem sitasi melalui penguatan. Penelitian ini juga memperlihatkan bahwa kuantitas sitasi tidak selalu optimalisasi mekanisme verifikasi referensi, serta integrasi yang lebih luas dengan pangkalan data ilmiah yang terindeks guna meminimalkan potensi kesalahan referensi dan fenomena citation hallucination.

Penelitian berikutnya disarankan untuk mengembangkan pendekatan evaluasi yang lebih komprehensif melalui perluasan jumlah sampel pertanyaan, penambahan variasi platform *AI*, serta penerapan indikator pengukuran yang lebih beragam, seperti citation precision, relevance score, dan konsistensi relevansi sitasi. Selain itu, penelitian lanjutan dapat diarahkan pada analisis perbandingan berdasarkan bidang keilmuan tertentu untuk memperoleh gambaran yang lebih mendalam mengenai tingkat keakuratan sitasi *AI* dalam konteks disiplin.

## 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/comasiejournal>.
- 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 Mubarok. 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 Tanjuman 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>.
- Rahayu, Dewi Nofi Ginanjar, Alex Harijanto, and Albertus Djoko Lesmono. "Tingkat Kemampuan Berpikir kritis siswa sma pada materi fluida dinamis." *Jurnal Pembelajaran Fisika* 7, no. 2 (2018): 162-167.
- 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, Stuart, 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. <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](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.