Jurnal ABDIMAS-IBISA
Pengabdian Kepada Masyarakat

Vol 4 No 2
Juli, 2026
Hal : 61-77

PENGGUNAAN HAKIM ACADEMIC COPILOT UNTUK MENCEGAH HALUSINASI RUJUKAN ILMIAH BAGI MAHASISWA LINTAS PERGURUAN TINGGI

Abdullah Ghafur Hakim
Institut Teknologi Bisnis dan Kesehatan Bhakti Putra Bangsa Indonesia
ghofurhakim7@gmail.com
Jl. Soekarno-Hatta, Borokulon, Banyuurip, Purworejo, Jawa Tengah

ABSTRAK

Penggunaan ChatGPT konvensional oleh mahasiswa dalam penyusunan karya ilmiah rentan menghasilkan rujukan hasil AI hallucination, yaitu nama penulis, judul, dan DOI yang tersusun meyakinkan namun tidak pernah ada dalam basis data ilmiah mana pun. Risiko ini terus meningkat seiring meluasnya penggunaan AI generatif di kalangan mahasiswa, dan mengancam integritas akademik apabila tidak disertai kemampuan verifikasi mandiri. Kurangnya kebiasaan verifikasi rujukan menjadi salah satu faktor utama tingginya risiko sitasi fiktif tersebut, sebagaimana terekam pada tahap analisis kebutuhan mitra sasaran. Hakim Academic Copilot merupakan chatbot kustom hasil rekayasa prompt dan personalisasi ChatGPT yang sederhana, tanpa memerlukan infrastruktur teknis tambahan, dan memaksa setiap rujukan melewati verifikasi lintas basis data ilmiah sebelum ditampilkan kepada pengguna. Namun, tahap analisis kebutuhan menemukan bahwa mahasiswa masih terbiasa menyalin rujukan yang disajikan ChatGPT konvensional tanpa langkah verifikasi mandiri. Kegiatan pengabdian ini bertujuan menghibahkan dan mendampingi penggunaan Hakim Academic Copilot kepada 30 mahasiswa dari berbagai perguruan tinggi di Indonesia, guna meminimalkan AI hallucination pada rujukan ilmiah dan menumbuhkan kebiasaan verifikasi mandiri. Metode pelaksanaan menempuh tahap analisis kebutuhan, penyiapan solusi, sosialisasi, pendampingan penggunaan chatbot, hingga evaluasi peningkatan penerimaan pengguna melalui User Acceptance Testing (UAT). Hasil evaluasi menunjukkan instrumen valid (r-hitung 0,496–0,886, di atas r-tabel 0,361) dan reliabel (Cronbach's Alpha 0,954), dengan skor rata-rata penerimaan 4,26 dari skala 5 (kategori sangat tinggi) dan 90% peserta pada kategori tinggi hingga sangat tinggi. Kegiatan ini diharapkan mampu meningkatkan literasi verifikasi rujukan mahasiswa secara luas dan menjadi model pendampingan yang dapat direplikasi oleh dosen atau pengelola program studi lain.

Kata kunci: *Pengabdian masyarakat; Academic Copilot; AI hallucination; verifikasi rujukan; rekayasa prompt*

A. Latar Belakang

Transformasi lanskap pendidikan tinggi global dalam tiga tahun terakhir tidak dapat dilepaskan dari kemunculan dan ekspansi cepat teknologi kecerdasan buatan generatif yang dibangun di atas arsitektur Large Language Model (LLM). Sejak peluncuran ChatGPT pada akhir 2022, kecepatan difusinya melampaui pola penyebaran teknologi digital sebelumnya, termasuk internet desktop pada masanya. Pada akhir 2024, hampir 40% populasi dewasa di Amerika Serikat tercatat telah menggunakan setidaknya satu perangkat AI generatif, dan pola serupa terlihat di berbagai kawasan dunia. Domain pendidikan tinggi justru menunjukkan tingkat penetrasi yang jauh lebih tinggi dibandingkan populasi umum tersebut. Survei terhadap mahasiswa Hong Kong University of Science and Technology mencatat bahwa 96% responden telah menggunakan perangkat AI dalam aktivitas akademiknya, dengan 44% menggunakannya secara mingguan dan 23% secara harian. Pola serupa juga dilaporkan dalam survei nasional Inggris Raya yang menunjukkan lonjakan proporsi mahasiswa pengguna AI generatif untuk keperluan penilaian akademik, dari 53% pada 2024 menjadi 88% pada 2025. Lonjakan ini menjadi penanda bahwa kecerdasan buatan generatif bukan lagi sekadar perangkat bantu opsional, melainkan telah menjelma menjadi bagian struktural dari proses belajar di perguruan tinggi.

Fenomena ketergantungan tersebut mendorong lembaga internasional untuk segera merespons melalui kebijakan dan panduan resmi. UNESCO menerbitkan dokumen Guidance for Generative AI in Education and Research pada 2023 sebagai panduan global pertama yang mengatur pemanfaatan AI generatif di sektor pendidikan, dengan menekankan pendekatan berpusat pada manusia, perlindungan privasi data, serta standar etika dalam interaksi antara peserta didik dan sistem kecerdasan buatan. OECD Digital Education Outlook menyoroti bahwa pemanfaatan AI generatif tanpa pedoman pedagogis yang memadai berpotensi hanya mendorong pengalihan tugas akademik tanpa disertai capaian pembelajaran yang substansial. Sejumlah laporan turunan bahkan menunjukkan bahwa proporsi signifikan tenaga pendidik di jenjang menengah meyakini AI generatif berisiko merusak integritas akademik dengan memungkinkan peserta didik mengklaim hasil kerja AI sebagai karya orisinal mereka sendiri. Kerangka kebijakan internasional ini menegaskan bahwa persoalan AI generatif dalam pendidikan bukan sekadar isu teknologis, melainkan menyentuh aspek epistemik dan etis yang lebih mendasar tentang bagaimana pengetahuan diproduksi dan diverifikasi.

Persoalan utama yang muncul adalah fenomena AI hallucination, yakni kecenderungan model bahasa untuk menghasilkan keluaran yang tampak koheren dan meyakinkan secara linguistik namun faktualnya tidak memiliki dasar kebenaran atau sepenuhnya fiktif. Dalam konteks rujukan ilmiah, fenomena ini memunculkan nama penulis, judul artikel, nama jurnal, DOI, volume, serta halaman yang tersusun rapi mengikuti kaidah sitasi konvensional, padahal artikel tersebut tidak pernah ada dalam basis data ilmiah mana pun. Kondisi ini menjadi semakin mengkhawatirkan karena rujukan hasil halusinasi tersebut sulit dibedakan dari rujukan otentik, sebab model bahasa menyusunnya berdasarkan pola probabilistik dari jutaan metadata bibliografis yang pernah dipelajarinya, bukan berdasarkan verifikasi terhadap data nyata. Studi yang mengevaluasi keluaran ChatGPT-3.5 dan ChatGPT-4 dalam menyusun 84 tinjauan literatur menemukan bahwa 55% sitasi yang dihasilkan ChatGPT-3.5 sepenuhnya fiktif, sementara ChatGPT-4 menunjukkan perbaikan dengan tingkat fabrikasi 18%, meski 24% dari sitasi otentiknya tetap mengandung kesalahan substantif pada elemen bibliografis. Subbaraman menganalisa terhadap lebih dari dua juta artikel ilmiah mengidentifikasi sekitar 146.900 sitasi hasil fabrikasi AI sepanjang 2025, dengan lonjakan tajam sejak pertengahan 2024.

Sejumlah penelitian terdahulu telah berupaya merespons persoalan halusinasi rujukan ini

melalui berbagai pendekatan teknis. Pada ranah evaluasi akurasi, penelitian Bhattacharyya dkk serta Walters dan Wilder mendokumentasikan secara kuantitatif tingkat fabrikasi sitasi pada ChatGPT generasi awal, namun keduanya bersifat deskriptif-evaluatif tanpa menawarkan solusi sistemik untuk mencegah fabrikasi tersebut. Pada ranah mitigasi teknis, pendekatan Retrieval-Augmented Generation (RAG) banyak dikembangkan sebagai strategi mengatasi keterbatasan model bahasa murni dengan cara mengintegrasikan proses pengambilan dokumen eksternal ke dalam alur generasi teks. Penelitian yang mengembangkan chatbot layanan akademik berbasis RAG, misalnya, berhasil mencapai skor faithfulness sebesar 0,91 dan secara efektif menolak menjawab 81,25% pertanyaan di luar konteks basis pengetahuannya, sehingga signifikan menekan potensi halusinasi. Pada penelitian lain, sejumlah studi mengeksplorasi Consensus AI multi-model sebagai mekanisme verifikasi silang. Prinsip dasarnya adalah bahwa verifikasi silang antar-agen independen menekan risiko halusinasi, sebab kesalahan yang muncul pada satu model jarang terulang secara identik pada model lain.

Kerangka berbasis mekanisme pemungutan suara dan debat adversarial antar-agen LLM terbukti efektif menekan halusinasi melalui proses verifikasi silang yang menentukan kapan diperlukan pengambilan pengetahuan eksternal tambahan. Pada sisi lain, berkembang pula infrastruktur teknis untuk verifikasi rujukan otomatis melalui pemanfaatan basis data ilmiah terbuka seperti Crossref, OpenAlex, dan Semantic Scholar, yang masing-masing menyediakan akses terbuka terhadap metadata bibliografis otoritatif dan dapat diintegrasikan melalui application programming interface untuk keperluan pencocokan otomatis. Sejauh penelusuran yang dilakukan, belum ditemukan penelitian yang secara khusus mengembangkan custom chatbot hasil kustomisasi platform ChatGPT yang menerapkan mekanisme verifikasi otomatis lintas beberapa basis data ilmiah terbuka, seperti Google Scholar, Crossref, dan Semantic Scholar, murni melalui rekayasa instruksi sistem dan pemanfaatan fitur pencarian web bawaan, sebelum menampilkan jawaban kepada pengguna, sehingga hanya rujukan yang benar-benar terverifikasi keberadaannya dalam basis data ilmiah yang akan ditampilkan.

Merespons kesenjangan (research gap) tersebut, tim pengabdi mengembangkan Hakim Academic Copilot, sebuah chatbot kustom hasil kustomisasi ChatGPT melalui rekayasa prompt (prompt engineering) yang menerapkan prinsip kerja Retrieval-Augmented Generation dengan memanfaatkan fitur pencarian web bawaan ChatGPT, dipadukan dengan instruksi verifikasi berlapis yang diterapkan secara konsisten pada enam varian instruksi sistemnya. Chatbot ini dirancang memverifikasi setiap rujukan yang akan ditampilkan melalui pencocokan silang terhadap basis data ilmiah otoritatif, mencakup Google Scholar, Crossref, dan Semantic Scholar, sehingga rujukan yang tidak dapat ditemukan korespondensinya pada basis data tersebut secara otomatis ditolak atau ditandai sebagai tidak terverifikasi, bukan ditampilkan begitu saja kepada pengguna. Pendekatan ini berbeda dari penelitian RAG akademik sebelumnya yang umumnya membangun infrastruktur pengambilan dokumen dan basis pengetahuan tersendiri, sebab prinsip kerja retrieval-augmented generation dan validasi silang multi-sumber pada Hakim Academic Copilot diwujudkan sepenuhnya melalui kustomisasi dan rekayasa prompt pada platform generatif yang telah tersedia luas, tanpa memerlukan pengembangan model, plugin, maupun infrastruktur teknis tambahan, sehingga solusi ini mudah dihibahkan dan direplikasi oleh institusi pendidikan dengan sumber daya teknis terbatas.

Kegiatan pengabdian ini bertujuan menghibahkan dan mendampingi penggunaan Hakim Academic Copilot kepada mahasiswa mitra sasaran, agar mampu meminimalkan fenomena AI hallucination pada aspek rujukan ilmiah, meningkatkan tingkat validitas dan keterlacakan sumber yang digunakan dalam penulisan karya ilmiah, serta menumbuhkan kebiasaan verifikasi rujukan secara mandiri. Melalui pendampingan ini, kegiatan diharapkan berkontribusi pada penguatan integritas akademik mitra sasaran, sekaligus menawarkan model pendampingan yang

dapat direplikasi oleh dosen atau pengelola program studi lain dalam menghadapi tantangan validitas rujukan pada era kecerdasan buatan generatif.

B. Metode

Kegiatan pengabdian ini menempuh metode pelaksanaan partisipatif yang lazim digunakan pada program pendampingan berbasis teknologi, mencakup tiga tahap berurutan: analisis kebutuhan mitra, penyiapan solusi teknis, serta pelaksanaan hibah disertai sosialisasi, pendampingan, dan evaluasi. Tahap analisis kebutuhan memetakan akar persoalan halusinasi referensi yang dihadapi mahasiswa sebagai mitra sasaran. Tahap penyiapan solusi menghasilkan Hakim Academic Copilot melalui rekayasa prompt dan penyusunan alur instruksi (workflow) sebagai perangkat bantu siap pakai bagi mitra. Tahap pelaksanaan menghibahkan akses chatbot kepada mitra melalui sosialisasi urgensi verifikasi rujukan, pendampingan penggunaan langsung, dan evaluasi peningkatan penerimaan pengguna melalui User Acceptance Testing (UAT). Ketiga tahap ini saling berkaitan, sebab keberhasilan pendampingan pada tahap akhir bergantung pada ketepatan pemetaan kebutuhan mitra dan kesesuaian solusi teknis yang dihibahkan pada tahap sebelumnya.

Pada teknik pengumpulan dan analisis data, kegiatan ini menggabungkan pendekatan kualitatif dan kuantitatif deskriptif secara berurutan di dalam kerangka metode pelaksanaan tersebut. Tahap analisis kebutuhan menempuh teknik kualitatif berupa studi literatur, observasi pola pencarian referensi mahasiswa, dan wawancara informal untuk menangkap kebutuhan fungsional secara mendalam, sementara tahap evaluasi pada akhir kegiatan menempuh teknik kuantitatif deskriptif melalui kuesioner UAT berskala Likert untuk mengukur tingkat penerimaan pengguna secara terukur. Kombinasi ini menempatkan kegiatan pada posisi embedded mixed-method, dengan data kualitatif membentuk spesifikasi awal artefak dan data kuantitatif memverifikasi keberhasilan artefak tersebut setelah diimplementasikan. Pendekatan semacam ini lazim ditempuh pada penelitian rekayasa sistem informasi maupun kegiatan pengabdian berbasis teknologi, sebab keduanya menuntut bukti empiris ganda, baik dari sisi kesesuaian desain dengan kebutuhan pengguna maupun dari sisi kepuasan dan kepercayaan pengguna terhadap artefak yang telah dihibahkan.

C. Hasil Dan Pembahasan

Kegiatan pengabdian ini dilaksanakan dalam lima tahap berkelanjutan yang saling berkaitan: permasalahan dan analisis kebutuhan mitra, penyiapan solusi berupa Hakim Academic Copilot, pelaksanaan pelatihan dan pendampingan disertai evaluasi peningkatan kemampuan mitra, dampak kepada mitra, serta keberlanjutan program. Keenam tahap tersebut disusun mengikuti kerangka kerja pendampingan partisipatif, dengan mahasiswa sebagai mitra sekaligus sasaran utama, sehingga setiap keluaran pada satu tahap menjadi masukan bagi tahap berikutnya. Pendekatan bertahap ini dipilih karena pendampingan berbasis kecerdasan buatan generatif menuntut iterasi yang ketat antara kesiapan solusi teknis dan validasi pengguna nyata, mengingat risiko utama yang hendak diatasi, yaitu halusinasi referensi ilmiah, baru dapat dipastikan teratasi apabila diuji langsung pada perilaku pencarian literatur mahasiswa di lapangan.

1. Permasalahan Kebutuhan Mitra

Permasalahan utama yang dihadapi mitra sasaran, yaitu mahasiswa program studi yang rutin menyusun karya tulis ilmiah, adalah tingginya risiko mengutip rujukan hasil AI hallucination dari ChatGPT konvensional tanpa disertai kemampuan verifikasi mandiri. Tahap pertama kegiatan ini berupa identifikasi kebutuhan sistem chatbot yang akan dikembangkan. Rancangan kegiatan pada tahap ini diawali dengan studi literatur untuk memetakan akar

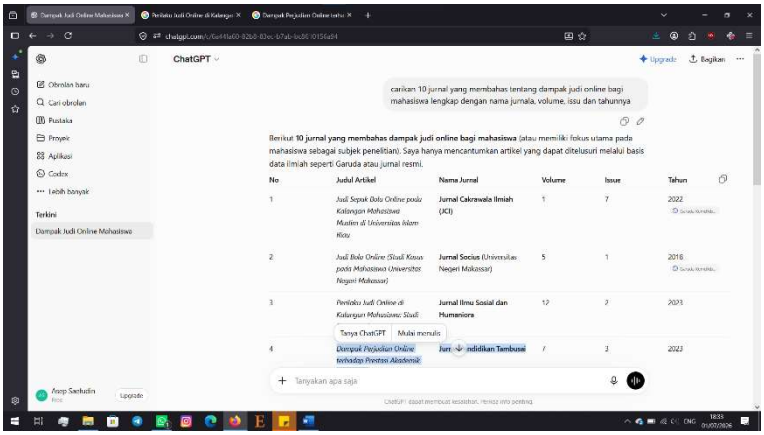
permasalahan halusinasi referensi pada model bahasa generatif. Tinjauan tersebut menegaskan bahwa fenomena ini bukan sekadar kesalahan teknis insidental, melainkan karakteristik bawaan dari cara kerja model bahasa besar yang memprediksi kata secara statistik tanpa mekanisme verifikasi kebenaran faktual, sehingga model dapat menghasilkan judul artikel, nama penulis, atau nomor DOI yang terdengar meyakinkan namun sesungguhnya tidak pernah ada. Persoalan ini menjadi semakin relevan dalam konteks Indonesia, mengingat penggunaan AI generatif di kalangan mahasiswa telah meluas hingga ke tahap penyusunan kerangka penelitian dan perangkuman literatur, sementara risiko sitasi fiktif yang menyertainya berpotensi merusak integritas akademik apabila tidak diverifikasi secara mandiri oleh pengguna.

Sasaran dari tahap analisis kebutuhan ini adalah mahasiswa program studi yang secara rutin menyusun karya tulis ilmiah, baik dalam bentuk tugas akhir, artikel jurnal, maupun laporan penelitian, dan karenanya membutuhkan akses cepat terhadap referensi ilmiah yang valid. Populasi sasaran dipilih dengan mempertimbangkan tingkat keterpaparan terhadap perangkat AI generatif yang tinggi namun belum disertai literasi memadai dalam memverifikasi keluaran referensi yang dihasilkan, sehingga kelompok ini paling berisiko terdampak oleh kesalahan sitasi yang tidak disadari.

Teknik pelaksanaan pada tahap ini meliputi empat tahap kegiatan yang dilakukan secara berurutan.

1) Pertama, studi literatur dilakukan terhadap publikasi terkini mengenai mekanisme munculnya halusinasi pada model bahasa besar serta strategi mitigasinya melalui pendekatan berbasis pengambilan informasi (retrieval-based), termasuk kajian yang menunjukkan bahwa penggabungan proses pencarian eksternal dengan proses pembangkitan teks secara terstruktur mampu menekan tingkat kesalahan faktual dibandingkan model yang murni mengandalkan parameter internalnya.

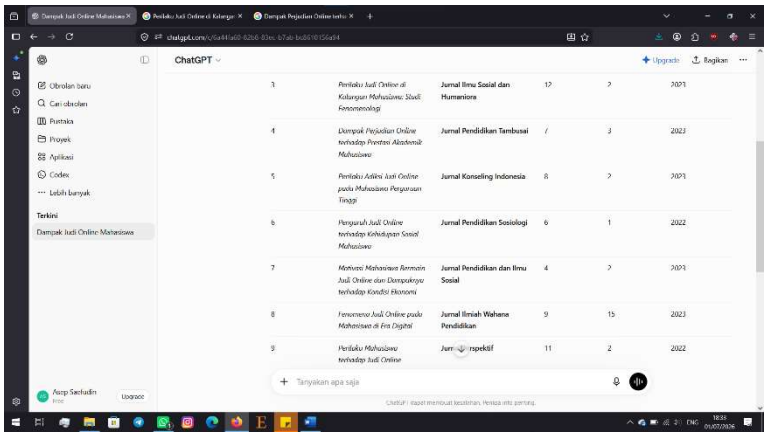
2) Kedua, dilakukan observasi terhadap kebiasaan mahasiswa dalam mencari referensi menggunakan alat bantu AI ChatGPT generatif konvensional, termasuk pengamatan terhadap pola di mana mahasiswa cenderung memercayai keluaran model tanpa langkah verifikasi silang ke basis data ilmiah resmi. Observasi ini menggunakan prompt yang lazim diketikkan mahasiswa, misalnya permintaan "carikan 10 jurnal yang membahas tentang dampak judi online bagi mahasiswa lengkap dengan nama jurnal, volume, issue, dan tahunnya", untuk merekam bagaimana ChatGPT versi standar merespons permintaan referensi semacam itu.



Gambar 1. Prompt yang digunakan mahasiswa

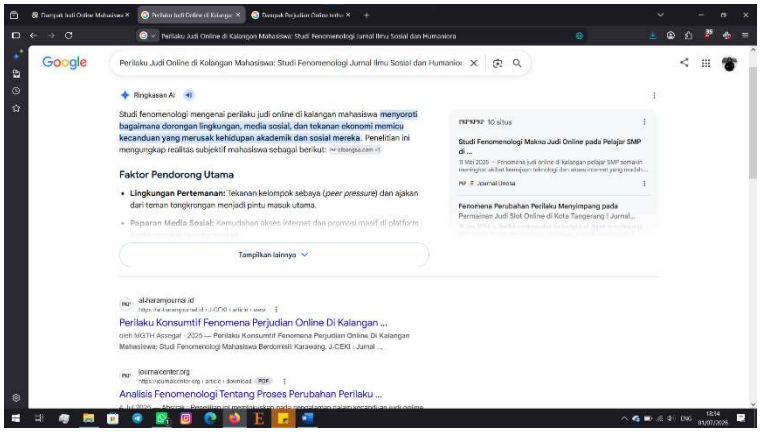
3) Ketiga, mencermati satu per satu referensi yang disajikan ChatGPT atas prompt tersebut, mencakup pemeriksaan terhadap kelengkapan elemen bibliografis seperti nama penulis,

judul artikel, nama jurnal, volume, issue, dan tahun terbit, guna menilai sejauh mana referensi tersebut tampak meyakinkan dan memenuhi kaidah sitasi ilmiah pada pandangan awal.



Gambar 2. Beberapa daftar referensi yang disajikan

4) Keempat, tim mengkroscek satu per satu tautan dan nama jurnal dari referensi yang disajikan ke situs resmi penerbit maupun basis data ilmiah seperti Google Scholar dan Crossref, untuk memastikan keberadaan artikel yang dimaksud; hasil pengecekan ini mengungkap bahwa sebagian referensi yang tampak meyakinkan tersebut tidak dapat ditelusuri korespondensinya pada sumber mana pun, sehingga terkonfirmasi sebagai hasil halusinasi model bahasa.



Gambar 3. Referensi yang tersaji hasil Halusinasi AI

5) Kelima, wawancara informal dilakukan terhadap sejumlah mahasiswa untuk menggali pengalaman mereka menemukan kasus sitasi yang tidak dapat ditelusuri keberadaannya, sekaligus menjaring ekspektasi pengguna terhadap sistem yang lebih dapat dipercaya.

6) Keenam, analisis kebutuhan pengguna dilakukan secara sistematis untuk merumuskan spesifikasi fungsional sistem berdasarkan temuan kelima teknik sebelumnya.

Keenam teknik ini saling melengkapi dan membentuk gambaran utuh tentang skala persoalan halusinasi referensi di kalangan mahasiswa. Studi literatur menegaskan akar teknis masalah, observasi merekam pola prompt yang biasa dipakai mahasiswa, pemeriksaan referensi menunjukkan bentuk luar sitasi yang tampak meyakinkan, dan pengecekan silang ke basis data ilmiah mengonfirmasi bahwa sebagian sitasi tersebut memang fiktif. Wawancara informal menambahkan sisi pengalaman mahasiswa yang pernah tertipu oleh sitasi semacam itu, sementara analisis kebutuhan menyaring keenam temuan ini menjadi daftar kebutuhan fungsional yang konkret. Tim menempatkan daftar tersebut sebagai acuan tunggal untuk tahap perancangan rekayasa prompt pada bagian berikutnya, sehingga setiap keputusan desain

instruksi sistem Academic Copilot dapat ditelusuri kembali ke temuan lapangan, bukan sekadar asumsi teknis.

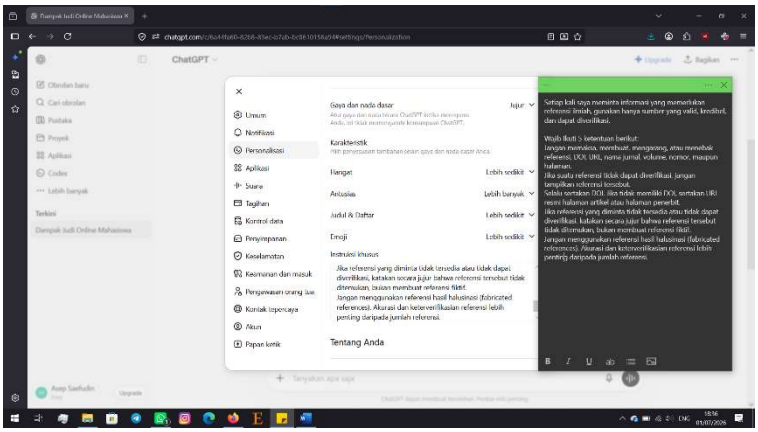
2. Solusi: Rekayasa Prompt dan Personalisasi Hakim Academic Copilot

Tahap kedua menjadi inti teknis kegiatan ini: penyusunan custom chatbot Academic Copilot di atas platform ChatGPT, melalui fitur personalisasi GPT yang tersedia bagi pengguna semua ChatGPT. Fitur ini memungkinkan tim pengembang menyusun persona dan instruksi khusus, tanpa menulis kode program dan tanpa melatih ulang model dasar. Seluruh proses pengembangan bertumpu pada rekayasa prompt (*prompt engineering*) dan penataan alur instruksi (*workflow*), tim menyusunnya menjadi varian instruksi sistem untuk kebutuhan disiplin ilmu berbeda, bukan membangun perangkat lunak atau infrastruktur teknis terpisah.

Dosen maupun pengelola program studi lain dapat mereplikasi pendekatan ini tanpa melibatkan tenaga pengembang perangkat lunak. Rancangan pada tahap ini berfokus pada penyusunan instruksi sistem yang memaksa Academic Copilot membandingkan dan menyelaraskan keluaran model bahasa dengan data hasil pencarian web, sebelum Academic Copilot menyampaikan jawaban akhir kepada pengguna. Instruksi ini mencegah chatbot bersandar semata pada kemampuan generatif model yang rentan berhalusinasi. Tahap ini menyasar satu target, purwarupa chatbot fungsional yang meminimalkan kemunculan *AI hallucination* pada pemberian referensi ilmiah, tanpa mengorbankan kemudahan interaksi percakapan layaknya chatbot generatif pada umumnya. Purwarupa ini menjadi prasyarat sebelum tim menghibahkan dan mengujikan sistem kepada mahasiswa pada tahap selanjutnya.

Metode pengembangan mengombinasikan empat elemen rekayasa prompt yang disusun secara berurutan dan diterapkan secara konsisten pada keenam varian instruksi sistem Hakim Academic Copilot, dengan langkah yang ditempuh sebagai berikut

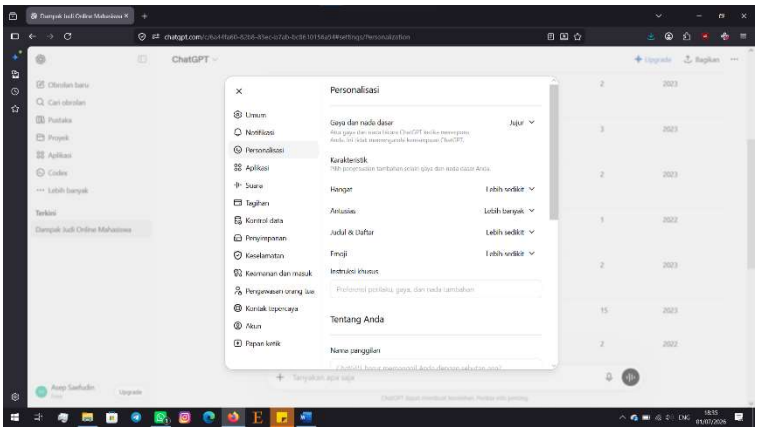
1) Tim pengembang menyusun instruksi sistem (*system instructions*) yang membatasi perilaku model AI secara tegas, mengikuti prinsip bahwa kejelasan dan kespesifikan instruksi menentukan efektivitas keluaran model bahasa besar. Instruksi ini mencakup penentuan persona, konteks, batasan tugas, dan format keluaran. Instruksi inti menegaskan dua hal: chatbot dilarang menyusun referensi dari pengetahuan internalnya, dan referensi yang ditampilkan kepada pengguna harus berasal dari hasil pencarian yang lolos verifikasi metadata. Ketika referensi yang diminta tidak ditemukan, chatbot menyatakan secara eksplisit bahwa referensi tersebut tidak tersedia, bukan mengarang substitusi yang tampak meyakinkan. Aturan ini sejalan dengan temuan bahwa mengarahkan model agar bersedia menyatakan ketidaktahuannya, alih-alih memaksakan jawaban, menjadi salah satu strategi paling efektif untuk mengurangi halusinasi referensi.



Gambar 4. Kustomisasi instruksi khusus

Instruksi ini menempatkan larangan eksplisit sebagai mekanisme kontrol utama, bukan sekadar imbauan. Kalimat penolakan eksplisit ketika referensi tidak ditemukan menjadi elemen kunci desain, sebab mengubah default behavior model dari "selalu menjawab" menjadi "menjawab hanya jika didukung bukti". Pilihan desain ini berisiko menurunkan persepsi kelengkapan jawaban pada pengguna awal, karena chatbot lebih sering menolak dibanding chatbot generatif konvensional. Efektivitas instruksi ini bergantung pada kepatuhan model terhadap instruksi sistem sepanjang sesi percakapan, sebuah faktor yang pada praktiknya tidak selalu konsisten dan relevan dikaitkan dengan indikator Reliability pada hasil UAT.

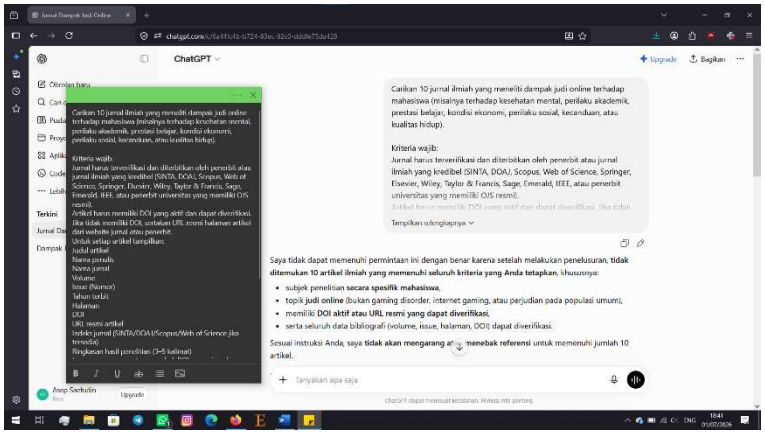
2) Segmen klasifikasi permintaan pada instruksi sistem mengarahkan Academic Copilot mengurai pertanyaan pengguna dan mengenali muatan permintaan referensi ilmiah di dalamnya. Begitu permintaan referensi terdeteksi, instruksi yang sama memerintahkan chatbot mengaktifkan kemampuan pencarian web bawaan untuk menelusuri basis data dan mesin pengindeks ilmiah seperti Google Scholar, Crossref, Semantic Scholar, dan OpenAlex, sebelum menyusun satu kalimat jawaban pun. Persona khusus pada Academic Copilot membatasi penyusunan referensi hanya dari hasil pencarian tersebut, tanpa melibatkan pengetahuan internal model. Referensi tanpa dukungan sumber nyata tersaring sebelum masuk tahap verifikasi. Chatbot lalu memverifikasi metadata setiap referensi yang lolos penyaringan: judul, nama penulis, tahun terbit, dan nomor DOI. Hanya referensi yang terbukti dapat ditelusuri ke sumber aslinya lolos ke tahap berikutnya, dan chatbot menampilkannya kepada pengguna. Referensi yang gagal verifikasi tidak muncul; sistem menggantikannya dengan pernyataan ketidaktersediaan.



Gambar 5. Kustomisasi personalisasi

Alur ini menempatkan klasifikasi permintaan sebagai gerbang pertama sebelum pencarian

eksternal berjalan, sehingga chatbot tidak menjalankan pencarian web pada setiap giliran percakapan tanpa alasan. Penyaringan bertingkat membentuk rantai keputusan yang masing-masing berpotensi menjadi titik kegagalan tersendiri. Kegagalan klasifikasi pada tahap pertama, misalnya ketika chatbot tidak mengenali permintaan referensi implisit dalam kalimat ringkasan teori, melewatkan seluruh mekanisme verifikasi di belakangnya tanpa disadari pengguna. Mekanisme penggantian otomatis dengan pernyataan ketidaktersediaan menjadi elemen pembeda paling jelas dibanding chatbot versi standar, dan elemen inilah yang paling layak diuji secara kuantitatif pada bagian hasil, bukan sekadar dijelaskan secara prosedural.

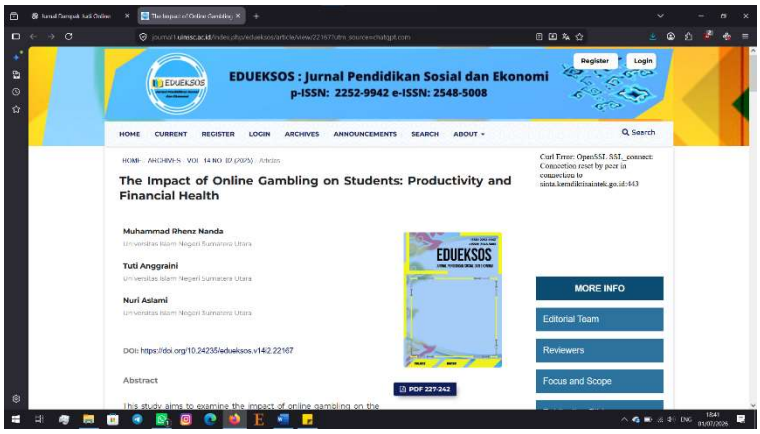


3) Alur instruksi ini menerapkan prinsip kerja *Retrieval-Augmented Generation* (RAG): model menggabungkan pengambilan dokumen (*retrieval*) dari sumber eksternal dengan pembangkitan teks (*generation*), sehingga jawaban bersandar pada dokumen nyata yang berhasil diambil, bukan pada parameter internal model. Instruksi sistem pada Academic Copilot menerapkan prinsip ini secara langsung: personalisasi yang ditanamkan memaksa model menyusun jawaban hanya dari dokumen yang diperoleh melalui fitur pencarian ChatGPT bawaan, dan melarang model mengisi kekosongan referensi dengan pengetahuan internalnya. Integrasi mekanisme pengambilan informasi eksternal menurunkan tingkat kesalahan faktual pada keluaran model dibandingkan pendekatan generatif murni, khususnya pada domain yang menuntut presisi tinggi seperti penyebutan referensi ilmiah.

Gambar 6. Custom spesifik prompt

Penerapan RAG lewat rekayasa prompt menghadirkan solusi ringan yang memisahkan diri dari kebutuhan membangun basis pengetahuan atau pipeline retrieval khusus. Pendekatan ini memindahkan seluruh beban infrastruktur ke fitur pencarian web yang sudah tersedia pada ChatGPT. Larangan eksplisit bagi model untuk mengisi kekosongan referensi dengan pengetahuan internal menjadikan prinsip RAG bukan sekadar rekomendasi arsitektural, melainkan aturan mengikat yang tertanam langsung pada persona chatbot. Pilihan desain ini selaras dengan temuan bahwa integrasi pengambilan informasi eksternal menekan tingkat kesalahan faktual, dan memperkuat posisi kegiatan pengabdian ini sebagai bukti bahwa prinsip kerja RAG dapat diwujudkan sepenuhnya melalui kustomisasi platform generatif yang telah tersedia luas.

4) Validasi lintas sumber berfungsi sebagai lapisan pengaman tambahan: setiap referensi harus terkonfirmasi keberadaannya pada lebih dari satu basis data ilmiah yang diakses melalui fitur pencarian web bawaan, sebelum dinyatakan valid. Mekanisme ini memperkecil kemungkinan kesalahan akibat satu sumber pencarian yang keliru atau tidak lengkap. Kombinasi prompt, workflow, prinsip RAG, dan validasi lintas sumber ini membentuk rangkaian instruksi berlapis dalam satu system prompt Academic Copilot, dan setiap referensi wajib melewati kecocokan antara hasil pencarian eksternal dengan verifikasi metadata pada lebih dari satu basis data sebelum layak disampaikan kepada pengguna. Rangkaian instruksi berlapis ini selaras dengan rekomendasi pengembangan chatbot berbasis RAG yang menekankan transparansi sumber dan mekanisme penahan (*guardrail*) eksplisit, guna meminimalkan risiko penyajian informasi yang tidak terverifikasi kepada pengguna akhir.



Gambar 7. Validasi referensi yang tersaji

Validasi lintas sumber menempatkan dua lapis pemeriksaan secara berurutan: sistem memeriksa kecocokan referensi pada lebih dari satu basis data sebelum menampilkannya, kemudian mahasiswa memeriksa ulang referensi yang sudah lolos tahap itu lewat tautan DOI atau situs resmi penerbit. Lapisan kedua ini melengkapi lapisan pertama, sebab validasi otomatis pada sistem tidak menghilangkan kemungkinan kesalahan yang lolos dari pengecekan mesin, dan kebiasaan mahasiswa mengklik tautan untuk mencocokkan judul, penulis, serta tahun menjadi penutup celah tersebut. Kebiasaan verifikasi mandiri semacam ini tetap berguna bagi mahasiswa meski kelak berpindah ke perangkat AI generatif lainnya, sehingga capaian kegiatan pengabdian ini melampaui sekadar penyediaan chatbot, dan turut membentuk literasi verifikasi rujukan pada penggunanya.

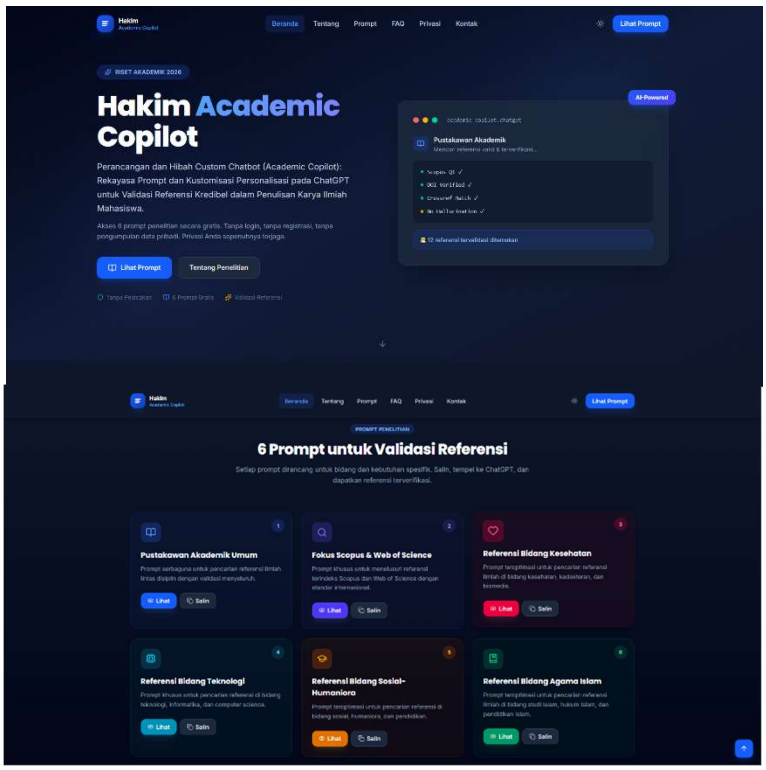
3. Solusi: Situs Layanan Hakim Academic Copilot

Tahap ketiga kegiatan ini adalah penyusunan situs web Academic Copilot, tempat naskah instruksi sistem alur kerja hasil rekayasa prompt tersimpan dan dapat diakses siapa saja secara gratis pada laman web https://hakim-academic-copilot-pofq.arcada.app/. Pada situs ini tidak mensyaratkan pendaftaran akun, biaya berlangganan, atau instalasi aplikasi tambahan, sehingga mahasiswa dari program studi mana pun bisa membukanya lewat ponsel atau laptop tanpa hambatan teknis. Pilihan ini penting mengingat sasaran kegiatan mencakup mahasiswa lintas angkatan dengan tingkat keterampilan teknis yang beragam, sehingga proses mengakses Academic Copilot perlu sesederhana membuka tautan dan menyalin teks. Situs tersebut memuat penjelasan singkat mengenai fungsi Academic Copilot sebagai asisten validasi referensi ilmiah, alasan pengembangannya, dan batasan penggunaannya, dan juga menampilkan naskah instruksi lengkap dalam format teks siap salin, disertai panduan langkah demi langkah untuk menempelkannya ke ChatGPT. Situs tersebut sengaja dipermudah aksesnya agar mahasiswa

yang sudah paham cara pakainya bisa langsung menuju naskah instruksi, sementara mahasiswa baru tetap mendapat konteks yang cukup sebelum mulai memakai chatbot.

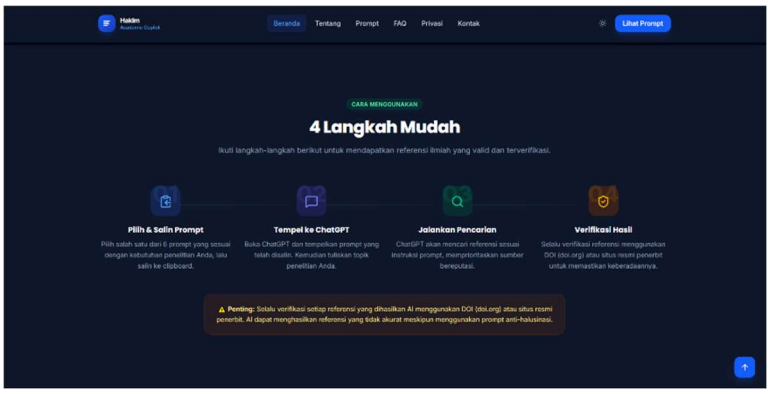
Nama produk yang tercantum pada situs adalah "Hakim Academic Copilot", nama yang digunakan secara konsisten pada naskah ini untuk membedakannya dari sebutan generik "Academic Copilot". Situs menghapus tiga hambatan akses sekaligus: tanpa login, tanpa registrasi, dan tanpa pengumpulan data pribadi, sebuah keputusan desain yang secara konkret mendukung kemudahan akses lintas angkatan mahasiswa. Panel demo pada situs menampilkan checklist verifikasi berupa status Scopus Q1, DOI Verified, Crossref Match, dan No Hallucination. Status indeksasi Scopus dicek sebagai lapisan verifikasi tambahan di luar empat basis data utama yang menjadi acuan pencocokan silang, yaitu Google Scholar, Crossref, Semantic Scholar, dan OpenAlex.

Gambar 8. Antarmuka Halaman Hakim Academic Copilot



Halaman ini menampilkan wujud nyata dari kustomisasi personalisasi yang dihibahkan kepada mitra: bukan satu instruksi sistem tunggal yang berlaku untuk semua bidang, melainkan enam varian prompt yang disusun untuk kebutuhan disiplin ilmu berbeda, yaitu Pustakawan Akademik Umum, Fokus Scopus & Web of Science, Referensi Bidang Kesehatan, Referensi Bidang Teknologi, Referensi Bidang Sosial-Humaniora, dan Referensi Bidang Agama Islam. Setiap varian menyasar kebutuhan disiplin ilmu yang berbeda; prompt bidang Agama Islam, misalnya, mengarahkan pencarian ke basis data dan gaya sitasi yang berbeda dari prompt Scopus & Web of Science. Arsitektur modular enam-prompt inilah yang menjadi wujud personalisasi Hakim Academic Copilot, bukan prompt monolitik tunggal yang berlaku untuk semua bidang keilmuan.

Gambar 8. Kustomisasi Prompt yang Tersedia dan Teruji



Gambar 9. Panduan Penggunaan Kustomisasi Prompt

Alur pakai tersusun dari empat langkah: pilih dan salin prompt sesuai kebutuhan riset, tempel ke ChatGPT lalu ketik topik, jalankan pencarian yang memprioritaskan sumber bereputasi, dan verifikasi hasil lewat DOI atau situs resmi penerbit. Kotak peringatan pada situs menyatakan langsung bahwa AI tetap bisa menghasilkan referensi tidak akurat meskipun memakai prompt anti-halusinasi. Pernyataan ini penting pada dua sisi: pertama, sebagai bukti bahwa tim pengabdi merancang sistem dengan transparansi risiko, bukan mengklaim keandalan mutlak; kedua, sebagai dasar argumen bahwa tanggung jawab verifikasi akhir tetap berada di tangan mahasiswa, bukan sepenuhnya digantikan oleh sistem. Keterbukaan risiko ini menjadi keterbatasan yang disadari sejak tahap perancangan solusi, bukan temuan yang baru muncul pada tahap evaluasi.

Chatbot tetap bisa keliru meski instruksinya dirancang ketat, sehingga mahasiswa perlu memeriksa ulang setiap referensi yang muncul, bukan langsung menyalinnya ke daftar pustaka. Naskah instruksi mewajibkan Academic Copilot menyertakan tautan langsung menuju sumber asli tiap referensi, baik tautan DOI maupun tautan ke halaman artikel pada basis data ilmiah, tepat di bawah setiap kutipan yang disajikan. Mahasiswa mengklik tautan tersebut untuk memastikan artikel itu ada dan sesuai dengan judul, penulis, serta tahun yang disebutkan chatbot; kecocokan tiga elemen ini menjadi tanda referensi tersebut layak dipakai. Kebiasaan mengklik dan memeriksa tautan ini melatih mahasiswa memverifikasi sumber secara mandiri, alih-alih mempercayai keluaran AI generatif begitu saja, dan menanamkan kebiasaan verifikasi yang tetap berguna meski mahasiswa kelak berpindah ke perangkat AI generatif lain.

4. Pelaksanaan Pelatihan, Pendampingan, dan Evaluasi Peningkatan Kemampuan Mitra (*User Acceptance Testing*)

Tahap akhir merupakan implementasi kegiatan pengabdian kepada mahasiswa sasaran, berupa Custom Chatbot Hakim Academic Copilot melalui tautan daring tanpa biaya. Kegiatan pada tahap ini disusun dalam empat aktivitas berurutan: sosialisasi, pendampingan, dan evaluasi, mengikuti siklus implementasi sistem informasi yang lazim diterapkan pada kegiatan pengabdian berbasis teknologi. 30 peserta yang terpilih untuk memperoleh data UAT yang representatif, tanpa mengorbankan kedalaman pendampingan bagi masing-masing peserta.

Sosialisasi berjalan lebih dulu untuk memperkenalkan mahasiswa pada akar permasalahan halusinasi referensi di ChatGPT versi standar, sekaligus memaparkan urgensi Hakim Academic Copilot sebagai chatbot kustom hasil rekayasa prompt dan kustomisasi personalisasi instruksi khusus untuk validasi referensi. Pendekatan yang menekankan pemahaman akar masalah sebelum pengenalan solusi teknis ini sejalan dengan praktik pelatihan literasi AI generatif bagi mahasiswa. Pendampingan penggunaan chatbot menyusul setelah sosialisasi, mencakup demonstrasi cara menyusun permintaan referensi yang efektif, cara membaca respons sistem

ketika sebuah referensi tidak dapat dipastikan keberadaannya, serta cara memanfaatkan Hakim Academic Copilot untuk menyesuaikan gaya sitasi dan format keluaran dengan kebutuhan tugas masing-masing mahasiswa. Pendampingan berlanjut selama periode penggunaan awal untuk membantu mahasiswa mengatasi kendala teknis dan mengklarifikasi hasil yang diperoleh dari sistem.

Evaluasi berjalan setelah periode penggunaan selesai, melalui metode User Acceptance Testing (UAT) dengan kuesioner daring berbasis Google Form berskala Likert lima poin, dari sangat tidak setuju hingga sangat setuju. Skala Likert dipilih karena menjadi instrumen baku dalam penelitian UAT sistem informasi untuk menangkap persepsi, sikap, dan kepuasan pengguna secara kuantitatif. Butir pertanyaan kuesioner melalui uji keterbacaan dan konsistensi respons pada tahap awal, sebelum disebarkan kepada seluruh peserta sasaran.

UAT ini mengukur enam indikator utama. Reliability mengukur keandalan sistem memberikan respons konsisten, baik pada pemakaian wajar maupun pada beban permintaan tinggi. Accuracy mengukur ketepatan referensi yang diberikan Academic Copilot, yaitu sejauh mana referensi tersebut benar-benar dapat ditelusuri pada basis data ilmiah aslinya. Ease of Use mengukur kemudahan mahasiswa berinteraksi dengan antarmuka chatbot tanpa pelatihan teknis yang rumit. User Trust mengukur kepercayaan pengguna terhadap keluaran sistem, termasuk kesediaan mereka mengandalkan chatbot dalam pencarian literatur akademik. Efficiency mengukur pengurangan waktu pengecekan manual referensi dibandingkan metode pencarian konvensional. User Satisfaction mengukur kepuasan menyeluruh pengguna terhadap pengalaman penggunaan chatbot. Keenam indikator ini sejalan dengan dimensi pengujian penerimaan pengguna yang lazim digunakan dalam evaluasi sistem informasi berbasis web maupun aplikasi.

Data hasil kuesioner UAT dianalisis melalui analisis deskriptif persentase dan perhitungan rata-rata skor Likert untuk setiap indikator. Setiap respons dikonversi menjadi nilai numerik, lalu dihitung skor rata-rata dan persentase pencapaian per indikator, mengikuti pendekatan perhitungan persentase yang lazim digunakan dalam penelitian UAT sistem informasi. Hasil perhitungan diinterpretasikan menggunakan kategori capaian yang telah ditetapkan, dari sangat setuju hingga sangat tidak setuju, untuk menggambarkan tingkat penerimaan mahasiswa terhadap Academic Copilot yang telah dihibahkan. Hasil analisis ini menjadi dasar evaluasi keberhasilan kegiatan pengabdian sekaligus masukan bagi penyempurnaan rekayasa prompt dan personalisasi sistem pada periode pengembangan berikutnya.

Uji validitas menggunakan korelasi item-total terkoreksi (corrected item-total correlation), yaitu korelasi Pearson antara skor tiap butir dengan jumlah skor butir lainnya. Nilai r-hitung dibandingkan terhadap r-tabel product moment untuk N=30 (df=28, α=0,05, dua sisi) sebesar 0,361. Seluruh 20 butir instrumen memiliki r-hitung di atas r-tabel sehingga dinyatakan valid. Nilai r-hitung terendah tercatat pada Q4 (0,496) dan Q14 (0,507); nilai tertinggi pada Q1 (0,886). Uji validitas dengan korelasi item-total terkoreksi terhadap 30 responden menghasilkan r-hitung antara 0,496 (Q4) dan 0,886 (Q1), seluruhnya melampaui r-tabel 0,361 (df=28, α=0,05). Seluruh 20 butir instrumen UAT Academic Copilot dinyatakan valid.

Tabel 1. Uji Validitas Instrumen

Item	Dimensi	r-hitung	r-tabel	Keterangan
Q1	Reliability	0,886	0,361	Valid
Q2	Reliability	0,764	0,361	Valid
Q3	Reliability	0,868	0,361	Valid
Q4	Accuracy	0,496	0,361	Valid
Q5	Accuracy	0,563	0,361	Valid
Q6	Accuracy	0,798	0,361	Valid

Q7	Accuracy	0,559	0,361	Valid
Q8	Ease of Use	0,692	0,361	Valid
Q9	Ease of Use	0,657	0,361	Valid
Q10	Ease of Use	0,642	0,361	Valid
Q11	User Trust	0,809	0,361	Valid
Q12	User Trust	0,571	0,361	Valid
Q13	User Trust	0,786	0,361	Valid
Q14	Efficiency	0,507	0,361	Valid
Q15	Efficiency	0,624	0,361	Valid
Q16	Efficiency	0,881	0,361	Valid
Q17	Efficiency	0,678	0,361	Valid
Q18	User Satisfaction	0,805	0,361	Valid
Q19	User Satisfaction	0,801	0,361	Valid
Q20	User Satisfaction	0,649	0,361	Valid

Reliabilitas dihitung dengan koefisien Cronbach's Alpha, dikategorikan menurut kriteria Guilford: ≥ 0,80 sangat tinggi; 0,60-0,799 tinggi; 0,40-0,599 cukup; 0,20-0,399 rendah; < 0,20 sangat rendah. Uji reliabilitas Cronbach's Alpha mencatat koefisien 0,954 untuk instrumen secara keseluruhan, kategori sangat tinggi menurut kriteria Guilford. Pada tingkat dimensi, koefisien berkisar dari 0,634 (Accuracy) hingga 0,874 (Reliability), seluruhnya kategori tinggi hingga sangat tinggi. Accuracy dan Ease of Use mencatat koefisien terendah, sejalan dengan jumlah butir yang lebih sedikit (3-4 item) dibanding dimensi lain.

Tabel 2. Uji Reliabilitas Instrumen

Dimensi	**Jumlah Item**	**Cronbach's Alpha**	**Kategori**
Reliability	3	0,874	Sangat Tinggi
Accuracy	4	0,634	Tinggi
Ease of Use	3	0,644	Tinggi
User Trust	3	0,760	Tinggi
Efficiency	4	0,761	Tinggi
User Satisfaction	3	0,796	Tinggi
Keseluruhan (20 item)	**20**	**0,954**	**Sangat Tinggi**

Skor tiap dimensi dihitung dari rata-rata butir penyusunnya per responden, lalu dirata-ratakan pada N=30. Kategorisasi memakai skala interval Likert: 1,00-1,80 Sangat Rendah; 1,81-2,60 Rendah; 2,61-3,40 Sedang; 3,41-4,20 Tinggi; 4,21-5,00 Sangat Tinggi. Secara deskriptif, dimensi Efficiency mencatat skor tertinggi (M=4,44; SD=0,51), diikuti Ease of Use (M=4,30), User Satisfaction (M=4,28), dan Accuracy (M=4,25). Keempat dimensi ini kategori sangat tinggi pada skala interval Likert. Dimensi User Trust (M=4,18) dan Reliability (M=4,03) berada satu tingkat di bawah, kategori tinggi. Reliability sekaligus mencatat variabilitas tertinggi (SD=0,92), terutama pada butir stabilitas respons (Q1). Pola ini konsisten dengan karakteristik implementasi Hakim Academic Copilot yang digunakan pada ChatGPT tanpa infrastruktur server khusus, sehingga stabilitas respons turut bergantung pada kondisi teknis penggunaan rekayasa prompt dan workflow pengguna.

Tabel 3. Statistik Deskriptif

Dimensi	**Mean**	**SD**	**Kategori**
Efficiency	4,44	0,51	Sangat Tinggi
Ease of Use	4,30	0,56	Sangat Tinggi
User Satisfaction	4,28	0,68	Sangat Tinggi
Accuracy	4,25	0,57	Sangat Tinggi
User Trust	4,18	0,58	Tinggi
Reliability	4,03	0,92	Tinggi
Rata-rata keseluruhan	**4,26**	**-**	**Sangat Tinggi**

Tabel 4. Distribusi Kategori Skor Responden

Kategori	Jumlah Responden	Persentase
Sangat Tinggi	16	53,3%
Tinggi	11	36,7%
Sedang	3	10,0%
Rendah	0	0%
Sangat Rendah	0	0%

Skor rata-rata keseluruhan mencapai 4,26 dari skala 5, kategori sangat tinggi. Dari 30 responden, 16 orang (53,3%) berada pada kategori sangat tinggi dan 11 orang (36,7%) pada kategori tinggi; tidak satu pun responden jatuh pada kategori rendah atau sangat rendah. Tiga responden dengan skor terendah, secara konsisten memberi skor rendah pada Reliability dan Accuracy, mengindikasikan pengalaman teknis yang kurang optimal saat penggunaan Hakim Academic Copilot berlangsung. Secara umum, hasil UAT menunjukkan tingkat penerimaan pengguna yang tinggi terhadap Academic Copilot, didukung bukti validitas dan reliabilitas instrumen yang kuat.

5. Keberlanjutan Program

Dampak pendampingan tercermin pada perubahan perilaku mitra dalam mencari referensi ilmiah, bukan sekadar skor penerimaan pada kuesioner UAT. Sebelum pendampingan, mahasiswa terbiasa menyalin rujukan yang disajikan ChatGPT konvensional tanpa langkah verifikasi mandiri, sebagaimana terekam pada tahap analisis kebutuhan. Setelah pendampingan, 90% peserta berada pada kategori penerimaan tinggi hingga sangat tinggi terhadap Hakim Academic Copilot, dengan skor Efficiency tertinggi (M=4,44) yang mengindikasikan pengurangan waktu pengecekan manual referensi dibandingkan kebiasaan sebelumnya. Kebiasaan mengklik tautan verifikasi yang ditanamkan selama pendampingan membentuk literasi verifikasi rujukan pada mitra sasaran.

Keberlanjutan program pendampingan ini ditopang oleh sifat Hakim Academic Copilot yang dapat diakses bebas biaya tanpa pendaftaran akun melalui tautan daring, sehingga mitra tetap dapat memanfaatkannya setelah kegiatan pengabdian berakhir tanpa bergantung pada kehadiran tim pengabdi. Naskah instruksi sistem yang tersedia dalam format teks siap salin juga memungkinkan dosen maupun pengelola program studi lain mereplikasi pendampingan serupa secara mandiri. Tim pengabdi merencanakan pemutakhiran berkala terhadap keenam varian prompt berdasarkan masukan pengguna dan perkembangan basis data ilmiah yang diakses, serta membuka kanal umpan balik bagi mitra untuk melaporkan kendala teknis maupun kasus verifikasi yang belum tercakup.

D. Kesimpulan

Kegiatan pengabdian ini berhasil meningkatkan kemampuan mitra sasaran, yaitu mahasiswa, dalam memverifikasi rujukan ilmiah secara mandiri melalui pendampingan penggunaan Hakim Academic Copilot. Dampak ini dikonfirmasi secara kuantitatif melalui evaluasi terhadap 30 mahasiswa mitra, dengan instrumen yang valid (r-hitung 0,496–0,886, di atas r-tabel 0,361) dan reliabel (Cronbach's Alpha 0,954). Skor rata-rata mencapai 4,26 dari skala 5 (kategori sangat tinggi), dengan 90% responden berada pada kategori tinggi hingga sangat tinggi, menunjukkan bahwa pendampingan ini diterima secara luas dan dirasakan manfaatnya oleh mitra sasaran. Efficiency mencatat skor tertinggi, mengindikasikan bahwa mitra sasaran merasakan peningkatan nyata dalam efisiensi kerja akademik setelah pendampingan. Sebaliknya, Reliability mencatat skor dan variabilitas terendah, menandakan bahwa kepercayaan mitra terhadap konsistensi hasil verifikasi masih menjadi area yang perlu diperkuat melalui pendampingan lanjutan.

Secara keseluruhan, temuan ini menegaskan bahwa pendampingan berbasis Hakim

Academic Copilot memberikan kontribusi konkret bagi mitra sasaran dalam bentuk peningkatan kompetensi verifikasi rujukan, efisiensi penulisan akademik, dan pengurangan risiko kesalahan sitasi. Model pendampingan ini tidak memerlukan infrastruktur teknis tambahan, sehingga mudah direplikasi oleh dosen atau pengelola program studi lain yang menghadapi persoalan serupa, dan berpotensi memperluas manfaat pengabdian ini melampaui mitra sasaran awal. Tanggung jawab verifikasi akhir tetap berada di tangan pengguna, namun hasil ini membuktikan bahwa pendampingan sederhana mampu menghasilkan perubahan perilaku akademik yang terukur pada masyarakat sasaran.

E. Saran

Penyempurnaan lanjutan Hakim Academic Copilot dapat diarahkan pada dimensi Reliability dan Accuracy yang mencatat skor serta variabilitas terendah pada evaluasi ini, misalnya melalui penyusunan panduan teknis tambahan bagi mahasiswa mengenai kondisi penggunaan yang memengaruhi stabilitas respons chatbot. Mengingat evaluasi pada kegiatan ini baru mengukur persepsi pengguna dan belum mengukur langsung ketepatan rujukan pada output akhir, pengembangan berikutnya perlu menempuh evaluasi terhadap keluaran nyata mahasiswa, yaitu memeriksa tingkat akurasi sitasi pada karya tulis yang telah disusun dengan bantuan Academic Copilot. Perluasan sasaran pendampingan ke populasi mahasiswa yang lebih besar dan lintas program studi juga diperlukan untuk menguji konsistensi tingkat penerimaan di luar 30 peserta awal. Selain itu, dokumentasi sistematis atas keenam varian prompt yang telah dikembangkan perlu disusun agar dapat diakses dan direplikasi secara mandiri oleh institusi pendidikan lain yang menghadapi persoalan serupa.

F. Ucapan Terimakasih

Penulis mengucapkan terima kasih kepada IBISA atas dukungan kelembagaan terhadap pelaksanaan kegiatan pengabdian ini, serta kepada 30 mahasiswa peserta berbagai perguruan tinggi di Indonesia yang telah berpartisipasi aktif dalam tahap sosialisasi, pendampingan, dan pengujian penerimaan pengguna Hakim Academic Copilot. Terima kasih juga disampaikan kepada rekan sejawat yang telah memberikan masukan pada proses pengembangan sistem dan penulisan naskah ini.

DAFTAR PUSTAKA

Agrawal, A., Mackey, L., & Kalai, A. T. (2023). Do language models know when they're hallucinating references? arXiv. https://arxiv.org/abs/2305.18248

Akkiraju, R., Xu, A., Bora, D., Yu, T., An, L., Seth, V., Shukla, A., Gundecha, P., Mehta, H., Jha, A., et al. (2024). Facts about building retrieval augmented generation-based chatbots. arXiv. https://arxiv.org/abs/2407.07858

Alkaissi, H., & McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: Implications in scientific writing. Cureus, 15(2).

Alotaibi, N. S. (2026). Generative AI in Saudi Arabia: A national survey of adoption, risks, and public perceptions. arXiv. https://arxiv.org/pdf/2601.18234

Andrian, R., et al. (2025). Development of an academic services chatbot based on Retrieval-Augmented Generation (RAG). Brilliance: Research of Artificial Intelligence. https://jurnal.itscience.org/index.php/brilliance/article/view/6719

Béchard, P., & Ayala, O. M. (2024). Reducing hallucination in structured outputs via retrieval-augmented generation. arXiv. https://arxiv.org/abs/2404.08189

Bhattacharyya, M., Miller, V. M., Bhattacharyya, D., & Miller, L. E. (2023). High rates of fabricated and inaccurate references in ChatGPT-generated medical content. Cureus, 15(5), Article e39238. https://doi.org/10.7759/cureus.39238

Chan, C. K. Y., & Hu, W. (2025). Embracing generative artificial intelligence tools in higher education: A survey study at the Hong Kong University of Science and Technology. Studies in Higher Education, 50(2), 352–376. https://doi.org/10.1080/17516234.2024.2447195

Chen, X., Wang, L., Wu, W., Tang, Q., & Liu, Y. (2024). Honest AI: Fine-tuning "small"

language models to say "I don't know," and reducing hallucination in RAG. arXiv. https://arxiv.org/abs/2410.09699

Freeman, J. (2025). Student Generative AI Survey 2025. Higher Education Policy Institute (HEPI). https://www.hepi.ac.uk/reports/student-generative-ai-survey-2025/

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems, 33, 9459–9474.

Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO. https://doi.org/10.54675/EWZM9535

OECD. (2026). OECD Digital Education Outlook 2026. OECD Publishing. https://www.oecd.org/en/publications/oecd-digital-education-outlook-2026_062a7394-en.html

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. Journal of Management Information Systems, 24(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302

Subbaraman, N. (2026, May 18). AI-generated fake citations are flooding scientific literature across publications, scientists warn. Phys.org. https://phys.org/news/2026-05-ai-generated-fake-citations-scientific.html

Tatum, C. (2026). Research paper APIs for scientific literature in 2026. IntuitionLabs. https://intuitionlabs.ai/articles/research-paper-apis-scientific-literature

Walters, W. H., & Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. Scientific Reports, 13, Article 14045. https://doi.org/10.1038/s41598-023-41032-5

Wang, F., et al. (2025). Transforming science with large language models: A survey on AI-assisted scientific discovery, experimentation, content generation, and evaluation. arXiv. https://arxiv.org/pdf/2502.05151

Yang, K., et al. (2025). Minimizing hallucinations and communication costs: Adversarial debate and voting mechanisms in LLM-based multi-agents. Applied Sciences, 15(7), 3676. https://doi.org/10.3390/app15073676