

GENERATIVE AI SEBAGAI SCAFFOLDING PENULISAN ILMIAH DAN ETIKA AKADEMIK MAHASISWA NON-EKSAKTA

Ihramsari Akidah^{1*}, Muhajir²

¹Fakultas Sastra, Ilmu Komunikasi, dan Pendidikan, Universitas Muslim Indonesia,
Jalan Urip Sumoharjo Km. 5, Makassar 90231, Indonesia

*Alamat email penulis koresponden: ihramsari.akidah@umi.ac.id

Abstrak

Penelitian ini bertujuan menganalisis efektivitas Generative Artificial Intelligence (GenAI) sebagai *scaffolding* terhadap kemampuan menulis ilmiah dan etika akademik mahasiswa non-eksakta Universitas Muslim Indonesia, serta merancang modul instruksional berbasis praktik terbaik. Penelitian menggunakan kerangka mixed methods dan Research and Development. Data empiris yang tersedia berasal dari survei awal dan analisis draf 100 mahasiswa: 88% menggunakan ChatGPT atau Gemini secara reguler, 62% melakukan *zero-drafting*, dan rata-rata *Similarity Index* mencapai 42%. Untuk memperlihatkan bentuk analisis bagi seluruh rumusan masalah, naskah sumber menyusun simulasi kuasi-eksperimen dengan masing-masing 50 mahasiswa pada kelas eksperimen dan kontrol. Dalam simulasi tersebut, skor menulis kelas eksperimen meningkat dari 61,84 menjadi 82,36, lebih tinggi daripada kontrol yang meningkat dari 62,10 menjadi 70,42; perbedaan post-test 11,94 poin, $p < 0,001$, $d = 1,85$. Skor etika eksperimen meningkat dari 63,2 menjadi 87,4, disertai penurunan *zero-drafting* menjadi 8% dan median *Similarity Index* menjadi 14,6%. Produk yang dirancang berupa modul enam unit dengan alur diagnosis-scaffolding-fading-audit. Simulasi validasi dan penerimaan masing-masing mencapai 91,5% dan 90,1%. Hasil menunjukkan potensi GenAI sebagai bantuan bertahap yang mempertahankan kepemilikan intelektual mahasiswa. Namun, seluruh angka selain $n=100$, 88%, 62%, dan 42% merupakan simulasi analitis dan wajib diganti dengan data lapangan sebelum digunakan sebagai klaim penelitian final.

Kata Kunci: Generative AI; Scaffolding; Penulisan Ilmiah; Etika Akademik; Integritas Akademik

Abstract

This study aims to analyze the effectiveness of Generative Artificial Intelligence (GenAI) as instructional scaffolding for scientific writing and academic ethics among non-science students at Universitas Muslim Indonesia and to design a best-practice instructional module. The study adopts mixed-methods and Research and Development frameworks. The available empirical evidence comes from an initial survey and draft analysis involving 100 students: 88% regularly used ChatGPT or Gemini, 62% practiced zero-drafting, and the average Similarity Index was 42%. To demonstrate a complete analytical format for all research questions, the source manuscript presents a simulated quasi-experiment with 50 students each in experimental and control classes. In the simulation, the experimental class writing score increased from 61.84 to 82.36, compared with 62.10 to 70.42 in the control class; the post-test difference was 11.94 points, $p < .001$, $d = 1.85$. The experimental ethics score

increased from 63.2 to 87.4, while zero-drafting fell to 8% and the median Similarity Index to 14.6%. The proposed product is a six-unit module organized around diagnosis–scaffolding–fading–audit. Simulated expert validation and user acceptance reached 91.5% and 90.1%, respectively. The pattern indicates the potential of GenAI as gradual assistance that preserves students' intellectual ownership. Nevertheless, all figures except n=100, 88%, 62%, and 42% are analytical simulations and must be replaced with field data before being presented as final research claims.

Keywords: generative AI; scaffolding; scientific writing; academic ethics; academic integrity

PENDAHULUAN

Generative Artificial Intelligence (GenAI) telah mengubah cara mahasiswa merencanakan, menyusun, dan merevisi tulisan akademik. Model bahasa dapat menawarkan alternatif gagasan, membantu menyusun kerangka, memberi umpan balik kebahasaan, dan mempercepat eksplorasi topik. Dalam pendidikan, manfaat tersebut perlu ditempatkan dalam pendekatan yang berpusat pada manusia agar teknologi memperluas kapasitas belajar tanpa memindahkan agensi, kepengarangan, dan tanggung jawab dari mahasiswa (Miao & Holmes, 2023).

Peluang GenAI beriringan dengan risiko halusinasi, bias, fabrikasi referensi, ketergantungan, ketidakjelasan kepengarangan, dan pelanggaran integritas akademik. Kajian tentang model bahasa dalam pendidikan menegaskan bahwa hasil belajar sangat ditentukan oleh desain pedagogis, literasi kritis, dan supervisi manusia, bukan semata-mata oleh ketersediaan aplikasi (Kasneci et al., 2023; Tlili et al., 2023). Dengan demikian, pertanyaan pentingnya bukan hanya apakah mahasiswa menggunakan AI, tetapi bagaimana, kapan, untuk tujuan apa, dan dengan bukti proses seperti apa AI digunakan.

Konteks awal mahasiswa non-eksakta Universitas Muslim Indonesia (UMI) menunjukkan urgensi persoalan tersebut. Dari 100 mahasiswa, 88% menggunakan ChatGPT atau Gemini secara reguler, 62% menggunakan AI untuk menghasilkan teks utuh atau zero-drafting, dan rata-rata *Similarity Index* mencapai 42%. Tingkat adopsi yang sangat tinggi menyediakan landasan untuk integrasi pembelajaran, tetapi dominasi zero-drafting menunjukkan bahwa penggunaan masih cenderung substitutif. Mahasiswa berisiko menjadi penyunting permukaan terhadap teks AI dan melewati proses menetapkan posisi, menghubungkan teori dengan bukti, melakukan sintesis, serta mempertanggungjawabkan klaim. *Similarity Index* 42% tidak dengan sendirinya

berarti 42% tulisan merupakan plagiarisme. Angka itu menunjukkan proporsi kecocokan teks dengan sumber dalam basis data dan tetap memerlukan pemeriksaan konteks, kutipan, daftar pustaka, frasa umum, serta bagian argumentatif tanpa atribusi. Perangkat pencocokan teks tidak boleh diposisikan sebagai hakim otomatis. Penggunaannya harus diikuti literasi parafrase, sintesis, atribusi, dan refleksi terhadap sumber kecocokan (Turnitin, 2026).

Berbagai penelitian memberikan dasar bahwa GenAI dapat digunakan secara produktif apabila diintegrasikan sebagai dukungan bertahap. Intervensi pada pembelajaran menulis menunjukkan peningkatan pada isi, organisasi, dan tata bahasa ketika mahasiswa memperoleh bimbingan untuk mengevaluasi keluaran AI (Mahapatra, 2024). Meta-analisis eksperimen juga menunjukkan manfaat terhadap performa belajar dan kemampuan berpikir tingkat tinggi, meskipun besarnya efek bervariasi menurut desain, durasi, dan konteks intervensi (Deng et al., 2025). Perspektif mahasiswa dan penelitian umpan balik menegaskan manfaat pada perencanaan serta revisi, tetapi tetap menempatkan penilaian manusia sebagai unsur utama (Kim et al., 2025; Steiss et al., 2024). Pada sisi lain, integritas akademik dalam era GenAI tidak cukup dijaga melalui larangan dan deteksi. Perubahan teknologi memerlukan kebijakan yang jelas, asesmen berbasis proses, edukasi mengenai validitas sumber, dan kewajiban pengungkapan bantuan AI (Cotton et al., 2024; Dwivedi et al., 2023). Tinjauan sistematis mendorong integrasi pencegahan, literasi, dan tata kelola, sedangkan kajian kebijakan asesmen meminta perguruan tinggi menilai orisinalitas berdasarkan kepemilikan intelektual dan proses penyusunan, bukan hanya kemiripan produk akhir (Bittle & El-Gayar, 2025; Luo, 2024).

Dalam konteks UMI, pembelajaran perlu menghubungkan kompetensi akademik dengan akhlaqul karimah. Kejujuran, amanah, transparansi, dan tanggung jawab ilmiah harus tampak dalam keputusan mahasiswa ketika menggunakan AI: memeriksa kebenaran informasi, menolak referensi yang tidak terverifikasi, mempertahankan suara penulis, mencatat bantuan yang memengaruhi substansi, serta bertanggung jawab atas seluruh isi naskah. Orientasi ini selaras dengan pemahaman integritas akademik sebagai budaya kelembagaan yang dibangun melalui standar, pembelajaran, dan akuntabilitas (Mattar, 2022).

Berdasarkan persoalan tersebut, penelitian diarahkan untuk menjawab tiga

rumusan masalah. Pertama, sejauh mana efektivitas GenAI sebagai scaffolding terhadap peningkatan kualitas dan struktur penulisan ilmiah mahasiswa non-eksakta UMI? Kedua, bagaimana dampaknya terhadap pemahaman dan kepatuhan mahasiswa pada etika dan integritas akademik? Ketiga, bagaimana rancangan modul instruksional berbasis *best practice* yang tepat untuk memfasilitasi penggunaan AI secara etis dan produktif dalam ekosistem akademik UMI? Tujuan akhirnya ialah menyatukan peningkatan kemampuan menulis, pengendalian penggunaan AI, dan penguatan integritas dalam satu alur pembelajaran yang dapat diterapkan dosen.

METODE PENELITIAN

Penelitian menggunakan kerangka *mixed methods* dan *Research and Development* (R&D). *Mixed methods* dipilih untuk menghubungkan perubahan skor dengan penjelasan mengenai proses penggunaan AI, sedangkan R&D digunakan untuk merancang, memvalidasi, dan menguji penerimaan modul instruksional. Integrasi data kuantitatif dan interpretatif mengikuti prinsip bahwa masing-masing jenis bukti menjawab sisi masalah yang berbeda tetapi saling melengkapi (Creswell & Plano Clark, 2018). Data awal empiris berasal dari survei dan analisis draf 100 mahasiswa non-eksakta UMI pada awal 2026. Informasi yang tersedia dalam file sumber terbatas pada jumlah sampel, penggunaan reguler ChatGPT atau Gemini, praktik *zero-drafting*, dan rata-rata *Similarity Index*. File tidak memuat data individual, keluaran perangkat statistik, lembar validasi, respons angket, ataupun transkrip wawancara. Oleh karena itu, hanya $n=100$, 88%, 62%, dan 42% yang diperlakukan sebagai data lapangan.

Untuk memperlihatkan bentuk hasil yang lengkap sesuai rancangan, file hasil menyusun simulasi kuasi-eksperimen 100 mahasiswa yang dibagi sama rata menjadi kelas eksperimen dan kontrol. Kelas eksperimen menggunakan modul GenAI sebagai *scaffolding*, sedangkan kontrol memperoleh pendampingan penulisan tanpa modul terintegrasi AI. Kedua kelas menjalani *pre-test* dan *post-test* tugas menulis pada skala 0–100. Rubrik mencakup struktur dan logika argumen, kedalaman sintesis, serta kohesi dan tata bahasa. Skor etika akademik pada skala 0–100 mencakup antiplagiarisme, validitas referensi, pengungkapan penggunaan AI, dan kesadaran etika Islami.

Analisis dalam simulasi meliputi uji normalitas Kolmogorov–Smirnov, uji homogenitas Levene, Independent Sample T-Test, uji berpasangan, N-Gain, ukuran

efek Cohen's d, dan Mann-Whitney U untuk Similarity Index. N-Gain dihitung pada tingkat individu melalui $(post-test-pre-test)/(100 - pre-test)$, kemudian ditafsirkan sebagai rendah ($<0,30$), sedang ($0,30-0,69$), atau tinggi ($\geq 0,70$). *Similarity Index* dianalisis secara nonparametrik karena distribusinya tidak diasumsikan normal.

Pengembangan produk mencakup analisis kebutuhan, penyusunan enam unit, penetapan batas penggunaan AI, validasi tiga pakar, revisi, dan uji penerimaan pengguna. Data validasi pakar serta penerimaan mahasiswa pada dokumen ini juga merupakan simulasi analitis. Seluruh angka simulasi digunakan sebagai model penyajian dan tidak boleh diperlakukan sebagai bukti efektivitas lapangan sebelum diganti dengan data aktual.

Tabel 1. Klasifikasi sumber informasi dalam penelitian

Kategori	Isi	Status penggunaan
Data empiris	n=100; penggunaan reguler 88%; <i>zero-drafting</i> 62%; rata-rata <i>Similarity Index</i> 42%	Dipertahankan sebagai kondisi awal
Rancangan penelitian	<i>Mixed methods</i> dan R&D; eksperimen-kontrol; rubrik, angket, Turnitin; validasi; uji-t, N-Gain, Mann-Whitney U	Kerangka penyusunan hasil
Simulasi analitis	Pembagian kelompok, skor pre/post, uji statistik, perilaku etik, validasi pakar, dan penerimaan pengguna	Ilustrasi; bukan hasil lapangan
Sintesis interpretatif	Penjelasan yang diturunkan dari pola kuantitatif dan desain modul	Bukan kutipan wawancara

Catatan. Klasifikasi disusun berdasarkan isi file hasil dan batas penggunaan data simulasi.

HASIL DAN PEMBAHASAN

Kondisi Awal Empiris

Survei awal dan pemeriksaan draf terhadap 100 mahasiswa memberikan dasar empiris kebutuhan intervensi. Karena jumlah responden tepat 100 orang, persentase dapat dibaca langsung sebagai jumlah mahasiswa: 88 mahasiswa menggunakan ChatGPT atau Gemini secara reguler dan 62 mahasiswa melakukan *zero-drafting*. Tingkat penggunaan reguler menunjukkan bahwa GenAI telah menjadi bagian dari praktik akademik mayoritas responden, bukan fenomena marginal.

Praktik *zero-drafting* 62% menunjukkan berkurangnya keterlibatan pada tahap perumusan masalah, pemilihan bukti, penyusunan argumen, dan sintesis sumber.

Persoalannya bukan sekadar siapa yang mengetik, melainkan hilangnya aktivitas kognitif yang seharusnya dibentuk melalui kegiatan menulis. Rata-rata Similarity Index 42% memperkuat kebutuhan pengajaran parafrase, sintesis, atribusi, verifikasi sumber, dan pembacaan laporan kecocokan secara reflektif.

Tabel 2. Kondisi awal mahasiswa non-eksakta UMI

No	Indikator	Hasil	Makna awal
1	Mahasiswa pada survei dan analisis draf	100	Basis pemetaan kebutuhan awal 2026
2	Penggunaan ChatGPT/ Gemini secara reguler	88%	GenAI telah umum digunakan
3	AI menghasilkan teks secara utuh	62%	Praktik substitutif/ <i>zero-drafting</i>
4	Rata-rata <i>Similarity Index</i>	42%	Perlu telaah sumber kecocokan

Catatan. Data empiris sebagaimana tercantum dalam file hasil; tidak diubah. Similarity Index bukan keputusan otomatis mengenai plagiarisme.

Ketiga indikator tersebut membentuk kebutuhan yang saling terhubung. Adopsi tinggi menyediakan peluang integrasi, zero-drafting menuntut pembatasan fungsi, sedangkan Similarity Index menuntut penguatan parafrase, sintesis, atribusi, dan audit sumber. Intervensi karena itu harus mempertemukan keterampilan menulis dengan etika akademik dalam satu alur kerja.

Efektivitas terhadap Kemampuan Menulis Ilmiah

Rumusan masalah pertama menanyakan efektivitas GenAI sebagai scaffolding terhadap kualitas dan struktur penulisan ilmiah. Dalam simulasi, kemampuan awal kedua kelompok setara: kelas eksperimen memiliki rata-rata 61,84 (SD=6,92) dan kelas kontrol 62,10 (SD=7,05). Selisih 0,26 poin tidak signifikan, $t(98)=-0,186$; $p=0,853$. Simpangan baku yang hampir sama juga menunjukkan penyebaran kemampuan awal yang relatif serupa.

Setelah intervensi simulatif, rata-rata eksperimen meningkat 20,52 poin menjadi 82,36, sedangkan kontrol meningkat 8,32 poin menjadi 70,42. Uji berpasangan menunjukkan kenaikan signifikan pada eksperimen, $t(49)=24,758$; $p<0,001$, dan kontrol, $t(49)=9,593$; $p<0,001$. Artinya, latihan menulis umum memberi manfaat pada kedua kelas, tetapi kenaikan eksperimen sekitar 2,47 kali kenaikan kontrol.

N-Gain eksperimen sebesar 0,539 berada pada kategori sedang, sedangkan kontrol 0,212 berada pada kategori rendah. Nilai 0,539 berarti mahasiswa eksperimen

secara rata-rata mencapai 53,9% dari peningkatan maksimum yang secara teoretis masih tersedia dari skor awal menuju 100. Capaian tersebut menunjukkan peningkatan kuat dengan ruang penguatan lanjutan, terutama pada kemampuan sintesis yang memerlukan latihan berulang dan umpan balik manusia.

Tabel 3. Perubahan skor menulis ilmiah dan N-Gain

Kelompok	n	Pre-test M±SD	Post-test M±SD	Gain	N-Gain	Kategori
Eksperimen	50	61,84±6,92	82,36±6,18	20,52	0,539	Sedang
Kontrol	50	62,10±7,05	70,42±6,73	8,32	0,212	Rendah

Catatan. Data simulasi analitis; N-Gain merupakan rerata gain ternormalisasi individual.

Distribusi kategori memperlihatkan bahwa peningkatan eksperimen tidak hanya didorong oleh beberapa mahasiswa. Sebanyak 45 dari 50 mahasiswa (90%) berada pada kategori sedang dan empat mahasiswa (8%) pada kategori tinggi; hanya satu mahasiswa (2%) berada pada kategori rendah. Pada kontrol, 38 mahasiswa (76%) masih berada pada kategori rendah dan 12 mahasiswa (24%) pada kategori sedang. Pola ini menggambarkan konsistensi perubahan pada mayoritas kelas eksperimen.

Tabel 4. Distribusi kategori N-Gain

Kelompok	n	Pre-test M±SD	Post-test M±SD	Gain
Eksperimen	50	61,84±6,92	82,36±6,18	20,52
Kontrol	50	62,10±7,05	70,42±6,73	8,32

Catatan. Data simulasi analitis; kategori rendah <0,30, sedang 0,30–0,69, dan tinggi ≥0,70.

Peningkatan terbesar pada eksperimen terdapat pada struktur dan logika argumen, yaitu 21,5 poin. Kedalaman sintesis meningkat 20,3 poin dan kohesi-tata bahasa 19,8 poin. Selisih gain eksperimen terhadap kontrol konsisten pada ketiga dimensi: 11,9 poin untuk struktur-logika, 12,6 poin untuk sintesis, dan 12,1 poin untuk kohesi-tata bahasa. Konsistensi tersebut menunjukkan bahwa manfaat simulatif tidak hanya terkonsentrasi pada penyuntingan bahasa.

Skor akhir sintesis eksperimen 80,7 tetap lebih rendah daripada struktur-logika 84,1 dan kohesi-tata bahasa 82,3. Pola ini wajar karena sintesis menuntut mahasiswa membandingkan sumber, mengevaluasi kekuatan bukti, menghubungkan literatur dengan masalah, dan membangun posisi sendiri. Unit verifikasi dan matriks sintesis karena itu perlu memperoleh waktu latihan lebih panjang daripada perbaikan bahasa.

Tabel 5. Rata-rata skor setiap dimensi kemampuan menulis

Dimensi	Eks. pre	Eks. post	Gain	Kontrol pre	Kontrol post	Gain
Struktur dan logika argumen	62,6	84,1	21,5	62,8	72,4	9,6
Kedalaman sintesis	60,4	80,7	20,3	60,6	68,3	7,7
Kohesi dan tata bahasa	62,5	82,3	19,8	62,9	70,6	7,7

Catatan. Data simulasi analitis; semua skor menggunakan skala 0–100.

Data post-test memenuhi asumsi parametrik pada skenario: uji Kolmogorov-Smirnov menghasilkan $p > 0,05$ dan Levene $F = 1,323$; $p = 0,253$. Independent Sample T-Test menunjukkan keunggulan eksperimen 11,94 poin, 95% CI [9,38; 14,50], $t(98) = 9,240$; $p < 0,001$. Cohen's $d = 1,85$ menunjukkan ukuran efek sangat besar. Seluruh interval kepercayaan berada di atas nol, sehingga arah perbedaan konsisten menguntungkan kelas eksperimen.

Ukuran efek yang sangat besar harus dibaca hati-hati karena berasal dari simulasi. Dalam pelaksanaan nyata, variasi dosen, kedisiplinan mahasiswa, akses internet, jenis tugas, dan program studi dapat memperlebar variasi serta memperkecil efek. Data lapangan perlu disertai reliabilitas rubrik, interval kepercayaan ukuran efek, keterlaksanaan intervensi, dan pengendalian materi, waktu, serta tingkat kesulitan tugas.

Dengan batas tersebut, jawaban ilustratif rumusan masalah pertama ialah bahwa GenAI sebagai scaffolding berpotensi meningkatkan kualitas dan struktur penulisan lebih kuat daripada pendampingan biasa. Mekanismenya bukan penyerahan tulisan kepada AI, melainkan bantuan pada pemunculan alternatif ide, pengujian kerangka, umpan balik berbasis rubrik, dan revisi kohesi. Mahasiswa tetap menentukan klaim, memilih bukti, melakukan sintesis, serta mempertanggungjawabkan produk akhir.

Dampak terhadap Etika dan Integritas Akademik

Rumusan masalah kedua menilai pemahaman dan kepatuhan terhadap etika akademik. Skor awal kedua kelas pada simulasi setara, $t(98) = -0,210$; $p = 0,834$. Skor eksperimen meningkat 24,2 poin dari 63,2 menjadi 87,4, sedangkan kontrol meningkat 9,1 poin dari 63,5 menjadi 72,6. Kenaikan eksperimen sekitar 2,66 kali kontrol. Perbedaan post-test sebesar 14,8 poin signifikan, $t(98) = 11,880$; $p < 0,001$; 95% CI [12,33; 17,27]; $d = 2,38$.

Penurunan simpangan baku eksperimen dari 7,1 menjadi 5,6 menunjukkan skor

meningkat sekaligus lebih terkonsentrasi pada tingkat tinggi. Namun, skor komposit harus tetap diperiksa melalui distribusi dan analisis butir agar kelemahan pada indikator tertentu tidak tertutup oleh rata-rata. Fokus penelitian aktual perlu diarahkan pada apakah alasan etis dipahami, diterapkan lintas tugas, dan dipertahankan ketika pengawasan dosen berkurang.

Tabel 6. Perubahan skor komposit etika akademik

Kelompok	Pre-test M±SD	Post-test M±SD	Gain	Uji berpasangan	p
Eksperimen	63,2±7,1	87,4±5,6	24,2	t(49)=25,667	<0,001
Kontrol	63,5±7,2	72,6±6,8	9,1	t(49)=9,975	<0,001

Catatan. Data simulasi analitis; skor 0–100 mencakup antiplagiarisme, validitas referensi, pengungkapan AI, dan kesadaran etika Islami.

Perubahan pemahaman diikuti perubahan perilaku *Zero-drafting* eksperimen turun 54 poin persentase, dari 62% menjadi 8%, sementara kontrol turun 24 poin menjadi 38%. Verifikasi referensi eksperimen meningkat dari 68% menjadi 94%, dibandingkan peningkatan kontrol dari 68% menjadi 78%. Pengungkapan penggunaan AI meningkat paling besar pada eksperimen, dari 22% menjadi 88%, sedangkan kontrol meningkat dari 24% menjadi 42%.

Larangan yang disertai alternatif tindakan lebih mudah diterapkan daripada larangan tanpa panduan. Mahasiswa tidak hanya diminta berhenti menghasilkan teks utuh, tetapi diarahkan untuk meminta alternatif struktur, menguji argumen, atau memperoleh umpan balik terhadap draf yang telah ditulis sendiri. Verifikasi mengingatkan bahwa nama penulis, judul, jurnal, DOI, dan isi sumber yang diberikan AI hanya usulan awal. Pengungkapan difokuskan pada bantuan yang memengaruhi substansi atau proses agar transparansi terjaga tanpa beban administratif berlebihan.

Tabel 7. Indikator perilaku integritas akademik

Indikator	Eksperimen pre	Eksperimen post	Kontrol pre	Kontrol post
<i>Zero-drafting</i>	31/50 (62%)	4/50 (8%)	31/50 (62%)	19/50 (38%)
Referensi terverifikasi	34/50 (68%)	47/50 (94%)	34/50 (68%)	39/50 (78%)
Pengungkapan penggunaan AI	11/50 (22%)	44/50 (88%)	12/50 (24%)	21/50 (42%)

Catatan. Proporsi awal *zero-drafting* mempertahankan 62% dari data empiris; pembagian kelompok dan indikator lain merupakan simulasi analitis.

Rata-rata awal *Similarity Index* gabungan dipertahankan sekitar 42% sesuai data empiris. Setelah intervensi simulatif, mean eksperimen turun 24,1 poin menjadi 18,2%

dan median menjadi 14,6%; mean kontrol turun 13,0 poin menjadi 28,7% dan median 27,8%. Kondisi awal tidak berbeda signifikan, Mann-Whitney $p=0,524$, sedangkan post-test menunjukkan distribusi eksperimen lebih rendah, $U=501$; $Z=-5,16$; $p<0,001$; $r=0,52$.

Median eksperimen yang lebih rendah daripada mean mengisyaratkan beberapa nilai tinggi yang menaikkan rata-rata. Penelitian aktual perlu menampilkan rentang dan kuartil, serta mengelompokkan kecocokan menjadi kutipan sah, frasa umum, daftar pustaka, tugas mahasiswa lain, sumber internet, dan paragraf argumentatif tanpa atribusi. Hanya melalui pemeriksaan isi dapat dinilai apakah penurunan berasal dari perbaikan parafrase dan sintesis atau perubahan teknis dokumen.

Tabel 8. Similarity Index sebelum dan sesudah intervensi

Kelompok	Pre mean	Pre median	Post mean	Post median	Perubahan mean
Eksperimen	42,3%	42,7%	18,2%	14,6%	-24,1 poin
Kontrol	41,7%	39,6%	28,7%	27,8%	-13,0 poin

Catatan. Data simulasi analitis dikalibrasi agar rata-rata pre-test gabungan tetap 42% sesuai data empiris. Similarity Index tidak boleh menjadi satu-satunya bukti integritas.

Jawaban ilustratif rumusan masalah kedua tidak bergantung pada satu indikator. Skor angket menggambarkan pemahaman, data perilaku menggambarkan kepatuhan yang dapat diamati, dan Similarity Index memberi informasi mengenai karakter produk tulisan. Konsistensi kenaikan skor etika, penurunan zero-drafting, peningkatan verifikasi serta pengungkapan, dan penurunan indeks kesamaan mengarah pada dampak positif pendampingan terstruktur. Integritas ditempatkan sebagai praktik yang dapat ditelusuri melalui *log prompt*, sumber terverifikasi, catatan keputusan, dan deklarasi penggunaan AI.

Rancangan Modul Instruksional

Rumusan masalah ketiga dijawab melalui modul yang menempatkan mahasiswa sebagai penulis dan pengambil keputusan utama. AI hanya memberikan bantuan sementara, kemudian dukungan dikurangi sejalan dengan peningkatan kemandirian. Empat batas kerja diterapkan: AI boleh menghasilkan alternatif tetapi tidak menentukan klaim akhir; AI boleh memberi umpan balik tetapi mahasiswa wajib memilih revisi; AI boleh membantu bahasa tetapi tidak menggantikan sintesis; dan setiap penggunaan harus dapat dijelaskan serta didokumentasikan.

Modul terdiri atas enam unit berurutan. Literasi dan etika ditempatkan di awal

agar batas penggunaan dipahami sebelum alat digunakan. Pemetaan masalah serta kerangka memastikan gagasan utama berasal dari mahasiswa. Unit sumber-sintesis menjaga validitas bukti, revisi bahasa dilaksanakan setelah isi cukup matang, dan audit integritas memeriksa produk serta keseluruhan proses.

Tabel 9. Struktur modul instruksional

Unit	Tujuan	Aktivitas utama	Bukti belajar
Literasi dan etika AI	Memahami kemampuan, batas, bias, halusinasi, kepengarangan, dan kebijakan	Analisis kasus; klasifikasi assistive/substitutive; kontrak belajar	Peta risiko dan komitmen integritas
Pemetaan masalah	Merumuskan fokus, pertanyaan, dan batas topik secara mandiri	Brainstorming terbimbing; membandingkan ide sendiri dengan alternatif AI	Rumusan masalah dan alasan pemilihan
Kerangka argumen	Menyusun struktur logis dan koheren	Outlining; uji alur; kritik terhadap kerangka usulan AI	Kerangka beranotasi dan jejak revisi
Sumber dan sintesis	Memastikan validitas referensi dan membangun sintesis	Pemeriksaan DOI/URL; matriks sintesis; larangan sitasi fiktif	Matriks sumber terverifikasi dan paragraf sintesis
Revisi bahasa	Meningkatkan kohesi, kejelasan, dan kaidah tanpa menghapus suara penulis	Umpan balik berbasis rubrik; perbandingan versi; alasan revisi	Draf sebelum-sesudah dan catatan keputusan
Audit integritas	Menilai orisinalitas, kepatuhan, dan transparansi	Telaah Turnitin; log prompt; pengungkapan AI; refleksi etik	Draf final, laporan audit, dan deklarasi AI

Catatan. Rancangan sintesis berdasarkan komponen file hasil dan praktik penggunaan GenAI yang berpusat pada manusia.

Bukti belajar memungkinkan dosen menilai proses, bukan hanya produk. Peta risiko menunjukkan pemahaman awal; alasan pemilihan masalah menunjukkan kepemilikan gagasan; kerangka beranotasi merekam perkembangan argumen; matriks sumber membuktikan verifikasi; catatan keputusan menunjukkan evaluasi terhadap umpan balik; dan deklarasi AI membangun transparansi. Alur diagnosis–*scaffolding-fading*–audit juga memungkinkan dukungan dibedakan menurut kebutuhan mahasiswa.

Batas penggunaan disusun berdasarkan fungsi, bukan nama aplikasi, agar modul tetap relevan ketika platform berubah. Aktivitas dinilai dapat diterima apabila AI hanya memberi dukungan, mahasiswa tetap membuat keputusan intelektual, dan proses dapat dibuktikan. Aktivitas menjadi substitutif apabila AI mengambil alih

perumusan isi, argumen, atau sumber, sedangkan mahasiswa tidak mampu menjelaskan keputusan dalam teks.

Tabel 10. Batas penggunaan GenAI dalam modul

Diizinkan sebagai <i>scaffolding</i>	Syarat	Dilarang sebagai substitusi
<i>Brainstorming</i> alternatif topik dan sudut pandang	Mahasiswa memilih dan menjelaskan alasan	Meminta AI menghasilkan makalah/bab utuh untuk diserahkan
Umpan balik atas kerangka dan konsistensi argumen	Keluaran dibandingkan dengan rubrik dan ditelaah manusia	Menyerahkan struktur dan argumentasi sepenuhnya kepada AI
Saran perbaikan bahasa, kohesi, dan keterbacaan	Mahasiswa memilih revisi serta menjaga makna dan suara	Menyamarkan teks AI sebagai tulisan asli mahasiswa
Daftar kata kunci pencarian	Sumber dicari di pangkalan terpercaya dan diverifikasi	Mengutip referensi dari AI tanpa memeriksa keberadaan dan isi

Catatan. Batas operasional dirancang untuk mencegah zero-drafting dan mempertahankan kendali intelektual mahasiswa.

Simulasi validasi tiga pakar menghasilkan rata-rata kelayakan 91,5%. Semua aspek berada di atas batas 80%, dengan rentang 89,0%–95,0%. Aspek etika dan akhlaqul karimah memperoleh nilai tertinggi 95,0%, sedangkan penyajian terendah 89,0% tetapi tetap sangat layak. Pola itu mengarah pada pemadatan petunjuk, penyusunan contoh bertahap, dan perbaikan lembar kerja agar hubungan antara prompt, keluaran, penilaian kritis, dan keputusan akhir mudah terlihat.

Tabel 11. Simulasi hasil validasi tiga pakar

Aspek	Persentase	Kriteria	Catatan utama
Kelayakan isi	91,3%	Sangat layak	Tujuan, materi, aktivitas, dan asesmen selaras
Penyajian	89,0%	Sangat layak	Petunjuk praktik perlu lebih ringkas
Bahasa	90,3%	Sangat layak	Jelas dan sesuai mahasiswa non-eksakta
Teknologi/AI	92,0%	Sangat layak	Fungsi realistis dan tidak terpaku satu platform
Etika dan akhlaqul karimah	95,0%	Sangat layak	Amanah, verifikasi, dan transparansi terintegrasi
Rata-rata	91,5%	Sangat layak	Melebihi kriteria kelayakan >80%

Catatan. Data simulasi analitis; diasumsikan dinilai oleh pakar bahasa, teknologi pendidikan, dan etika akademik.

Penerimaan pengguna simulatif mencapai 90,1%. Kesesuaian etika tertinggi pada 93,0%; kegunaan dan dukungan produktivitas masing-masing 90,4%; kejelasan 89,2%; dan kemudahan 87,6%. Nilai kemudahan yang paling rendah dapat dijelaskan oleh tambahan usaha untuk memeriksa sumber, mencatat prompt, dan menjelaskan keputusan. Modul perlu menyediakan contoh log singkat, deklarasi satu halaman, tautan pemeriksaan DOI, dan checklist agar akuntabilitas tidak menghasilkan beban berlebihan.

Tabel 12. Simulasi penerimaan modul oleh mahasiswa

Aspek	Persentase	Interpretasi
Kegunaan	90,4%	Membantu menata penulisan dan penggunaan AI
Kemudahan penggunaan	87,6%	Verifikasi sumber memerlukan waktu lebih lama
Kejelasan materi	89,2%	Batas dan langkah kerja dinilai jelas
Kesesuaian etika	93,0%	Pengungkapan, verifikasi, dan refleksi relevan
Dukungan produktivitas	90,4%	Membantu proses tanpa menghapus kewajiban berpikir
Rata-rata	90,1%	Penerimaan sangat baik

Catatan. Data simulasi analitis; persentase menggambarkan bentuk hasil akseptabilitas yang direncanakan.

Penyempurnaan final mencakup empat komponen. Pertama, checklist keputusan untuk mencatat keluaran yang diterima atau ditolak. Kedua, prosedur verifikasi nama penulis, judul, jurnal, tahun, DOI/URL, dan kesesuaian isi. Ketiga, formulir pengungkapan alat, tujuan, tahap penggunaan, dan kontribusi manusia. Keempat, refleksi terhadap laporan Turnitin agar persentase tidak digunakan secara mekanis. Produk yang tepat dengan demikian berupa modul enam unit yang mengintegrasikan keterampilan prompt, evaluasi kritis, verifikasi sumber, penyusunan kerangka, revisi, bukti proses, pengungkapan AI, dan akhlaqul karimah.

Pembahasan

Ketiga rumusan masalah membentuk hubungan yang utuh. Kondisi awal menunjukkan adopsi GenAI tinggi tetapi disertai praktik yang dapat mengurangi kemandirian. Modul mengubah cara kerja dari meminta produk jadi menjadi meminta dukungan pada tahap tertentu. Dalam data simulasi, perubahan proses tersebut diikuti peningkatan kualitas tulisan dan kepatuhan etika. Variabel penentunya bukan sekadar akses terhadap AI, melainkan desain pedagogis yang mengatur waktu, tujuan, batas,

verifikasi, dan bukti penggunaan.

Peningkatan skor struktur, sintesis, dan bahasa sejalan dengan penelitian bahwa ChatGPT dapat mendukung penulisan apabila digunakan dalam intervensi terbimbing (Mahapatra, 2024). Temuan meta-analitis juga menunjukkan potensi peningkatan performa belajar, tetapi efeknya dipengaruhi desain pembelajaran (Deng et al., 2025). Dalam modul ini, *brainstorming* dan *outlining* menurunkan hambatan awal, sedangkan rubrik serta catatan keputusan menuntut mahasiswa mengevaluasi keluaran. Pola tersebut berbeda secara mendasar dari *zero-drafting* karena aktivitas kognitif tetap berada pada mahasiswa.

Perbaikan yang relatif seimbang pada tiga dimensi memperlihatkan bahwa AI tidak ditempatkan hanya sebagai korektor tata bahasa. Dukungan pada struktur membantu mahasiswa mengurutkan klaim dan alasan; dukungan pada sintesis memerlukan pemeriksaan sumber serta keputusan intelektual; dukungan pada kohesi membantu kejelasan penyampaian. Perbandingan kualitas umpan balik manusia dan ChatGPT menguatkan posisi bahwa AI dapat memberi respons berbasis kriteria, tetapi manusia dibutuhkan untuk konteks, kedalaman penalaran, dan pertanggungjawaban (Steiss et al., 2024).

Aspek etika menunjukkan bahwa pengetahuan baru bermakna apabila konsisten dengan tindakan. Penurunan *zero-drafting* serta kenaikan verifikasi dan pengungkapan menggambarkan perpindahan dari kepatuhan deklaratif menuju praktik yang dapat diamati. Kewajiban memeriksa sumber penting karena model bahasa dapat menghasilkan referensi yang meyakinkan tetapi tidak valid. Deklarasi penggunaan AI membantu membedakan bantuan yang dapat diterima dari pengambilalihan kepengarangan, sekaligus menjaga tanggung jawab mahasiswa atas isi.

Penurunan *Similarity Index* mendukung interpretasi perbaikan, tetapi tidak boleh menjadi pusat tunggal penilaian integritas. Skor rendah tidak menjamin teks bebas masalah, dan skor tinggi dapat berasal dari kutipan sah atau elemen baku. Pendekatan proses melalui log prompt, riwayat draf, matriks sumber, catatan revisi, dan presentasi lisan lebih selaras dengan perubahan konsep orisinalitas dalam asesmen berbantuan GenAI (Luo, 2024). Institusi perlu menggabungkan edukasi, kebijakan, dan desain asesmen, bukan mengandalkan deteksi setelah pelanggaran (Bittle & El-Gayar, 2025;

Cotton et al., 2024).

Modul berfungsi sebagai jembatan langsung antara masalah dan hasil. Unit literasi-etika merespons adopsi yang belum terarah; pemetaan dan kerangka merespons *zero-drafting*; sumber-sintesis merespons validitas referensi; revisi merespons kohesi serta bahasa; audit integritas merespons kecocokan teks dan transparansi. Setiap komponen memiliki dasar masalah dan bukti belajar, sehingga validitas logis produk tidak bergantung pada popularitas aplikasi tertentu.

Alur *diagnosis-scaffolding-fading-audit* menempatkan bantuan sebagai sementara. Pada tahap awal, mahasiswa dapat memperoleh contoh prompt dan pertanyaan penuntun. Setelah kemampuan meningkat, dukungan dikurangi: mahasiswa menyusun prompt sendiri, membandingkan beberapa keluaran, menolak saran yang tidak tepat, dan menjelaskan proses tanpa instruksi langkah demi langkah. Desain semacam ini menjaga agensi serta mengurangi ketergantungan, sesuai perhatian literatur terhadap manfaat dan risiko model bahasa dalam pendidikan (Kasneci et al., 2023; Tlili et al., 2023).

Nilai akhlaqul karimah memperluas tujuan pembelajaran dari kepatuhan teknis menuju amanah ilmiah. Mahasiswa tidak hanya berusaha memperoleh Similarity Index rendah, tetapi memastikan sumber ada, klaim dapat dibuktikan, bantuan AI diungkapkan, dan produk akhir dapat dipertanggungjawabkan. Prinsip tersebut memberi dasar moral bagi pengambilan keputusan ketika kebijakan belum mengatur semua situasi baru. Pada tingkat institusi, budaya integritas perlu didukung standar yang konsisten, pembinaan dosen, dan mekanisme akuntabilitas (Mattar, 2022).

Secara praktis, modul dapat diintegrasikan dalam mata kuliah Bahasa Indonesia, metodologi penelitian, atau klinik penulisan ilmiah. Implementasi memerlukan pelatihan dosen, kesepakatan penggunaan AI pada setiap tugas, penyimpanan bukti proses, dan prosedur pemeriksaan sumber. Kebijakan sebaiknya membedakan penggunaan yang diizinkan, wajib diungkapkan, dan dilarang. Karena platform terus berubah, kebijakan perlu berorientasi pada fungsi serta prinsip kendali manusia, verifikasi, transparansi, privasi, dan akuntabilitas (Dwivedi et al., 2023; Miao & Holmes, 2023).

Batas penarikan kesimpulan harus ditegaskan. Pola simulasi hanya memperlihatkan bagaimana kesimpulan dapat dirumuskan apabila data lapangan

mendekati angka tersebut. Penelitian belum boleh menyatakan intervensi terbukti efektif di UMI sebelum data aktual tersedia. Hasil akhir wajib mengikuti data, termasuk apabila peningkatannya lebih kecil, tidak signifikan, atau berbeda antarprogram studi. Transparansi atas hasil yang tidak sesuai harapan merupakan bagian dari integritas penelitian. Efektivitas statistik, efektivitas pendidikan, dan kelayakan implementasi juga harus dibedakan. Perbedaan dapat signifikan tetapi terlalu kecil untuk berdampak pada kualitas tugas; sebaliknya, manfaat pendidikan dapat gagal mencapai signifikansi karena sampel terbatas. Produk dapat dinilai layak tetapi sulit diterapkan jika waktu, perangkat, atau pendampingan tidak memadai. Penelitian lapangan perlu membahas ketiga dimensi, menyertakan interval kepercayaan dan ukuran efek, menguji reliabilitas instrumen, serta mendokumentasikan keterlaksanaan intervensi.

KESIMPULAN

Data awal empiris menegaskan kebutuhan intervensi: dari 100 mahasiswa, 88% menggunakan ChatGPT atau Gemini secara reguler, 62% melakukan zero-drafting, dan rata-rata *Similarity Index* mencapai 42%. Adopsi teknologi dengan demikian telah mendahului kesiapan pedagogis dan etik. Intervensi perlu mengarahkan penggunaan AI dari produksi teks utuh menuju bantuan terbatas yang dapat ditelusuri dan dipertanggungjawabkan.

Dalam simulasi analitis, GenAI sebagai *scaffolding* berpotensi meningkatkan kemampuan menulis lebih kuat daripada pendampingan biasa. Eksperimen mencapai *post-test* 82,36 dan N-Gain 0,539, dengan keunggulan 11,94 poin atas kontrol, $p < 0,001$, $d = 1,85$. Peningkatan mencakup struktur-logika, kedalaman sintesis, serta kohesi-tata bahasa. Simulasi juga menunjukkan kenaikan skor etika menjadi 87,4, penurunan *zero-drafting* menjadi 8%, peningkatan verifikasi referensi menjadi 94%, pengungkapan AI menjadi 88%, dan median *Similarity Index* menjadi 14,6%.

Produk yang menjawab rumusan masalah ketiga ialah modul enam unit dengan alur diagnosis-*scaffolding-fading-audit*. Modul menggabungkan literasi dan etika AI, pemetaan masalah, kerangka argumen, verifikasi sumber dan sintesis, revisi bahasa, serta audit integritas. Simulasi validasi 91,5% dan penerimaan 90,1% menunjukkan format kelayakan yang diharapkan, tetapi bukan hasil pakar atau mahasiswa yang sebenarnya.

Keterbatasan utama penelitian adalah tidak tersedianya data eksperimen

individual, keluaran perangkat statistik, lembar validasi, respons angket, dan transkrip wawancara dalam file sumber. Karena itu, seluruh angka selain n=100, 88%, 62%, dan 42% bersifat simulasi dan wajib diganti sebelum publikasi sebagai hasil final. Penelitian lanjutan perlu menjalankan protokol lapangan, menguji reliabilitas instrumen, menyimpan data anonim, melaporkan asumsi statistik, melibatkan pakar materi, bahasa, teknologi pendidikan, serta etika/keislaman, dan menilai keberlanjutan perubahan perilaku.

DAFTAR RUJUKAN

- Bittle, K., & El-Gayar, O. (2025). Generative AI and academic integrity in higher education: A systematic review and research agenda. *Information*, 16(4), 296. <https://doi.org/10.3390/info16040296>
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228–239. <https://doi.org/10.1080/14703297.2023.2190148>
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.
- Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, 227, 105224. <https://doi.org/10.1016/j.compedu.2024.105224>
- Dwivedi, Y. K., Kshetri, N., Hughes, L., et al. (2023). “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642.
- Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kim, J., Yu, S., Detrick, R., & Li, N. (2025). Exploring students’ perspectives on generative AI-assisted academic writing. *Education and Information Technologies*, 30, 1265–1300. <https://doi.org/10.1007/s10639-024-12878-7>
- Luo, J. (2024). A critical review of GenAI policies in higher education assessment: A call to reconsider the originality of students’ work. *Assessment & Evaluation in Higher Education*, 49(5), 651–664. <https://doi.org/10.1080/02602938.2024.2309963>
- Mahapatra, S. (2024). Impact of ChatGPT on ESL students’ academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, 11, Article 9. <https://doi.org/10.1186/s40561-024-00295-9>

- Mattar, M. Y. (2022). Combating academic corruption and enhancing academic integrity through international accreditation standards: The model of Qatar University. *Journal of Academic Ethics*, 20, 119–146. <https://doi.org/10.1007/s10805-021-09392-7>
- Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO.
- Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction*, 91, 101894. <https://doi.org/10.1016/j.learninstruc.2024.101894>
- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10, Article 15. <https://doi.org/10.1186/s40561-023-00237-x>
- Turnitin. (2026). Understanding the Similarity Score. <https://guides.turnitin.com/hc/en-us/articles/23435833938701-Understanding-the-similarity-score>