

Penggunaan ChatGPT dalam Otomatisasi Penilaian Jawaban Esai Mahasiswa Jurusan Teknik Informatika di STMIK Antar Bangsa

Aristejo¹

Abstract—Manually grading essays in educational settings often requires significant time and effort, especially when assignment volumes are high and teaching staff are limited. This process not only slows down feedback to students but also potentially introduces inconsistencies in assessment. This study aims to evaluate the effectiveness of using artificial intelligence technology, specifically the ChatGPT language model, in automating the essay grading process. The method used involves implementing an automated grading system based on the ChatGPT API, which is then tested on a number of student essays with varying topics and levels of complexity. The evaluation was conducted by comparing the ChatGPT grading results with manual grading conducted by lecturers. The results show that ChatGPT is capable of providing consistent and efficient grading, with a fairly high level of agreement with human assessments, particularly in aspects of coherence, content relevance, and writing structure. The study concludes that ChatGPT has great potential as a tool for automated essay grading, although teacher supervision is still required to ensure the quality and fairness of the grading.

Abstrak— Penilaian esai secara manual dalam lingkungan pendidikan seringkali memerlukan waktu dan tenaga yang signifikan, terutama ketika volume tugas tinggi dan keterbatasan tenaga pengajar menjadi kendala. Proses ini tidak hanya memperlambat umpan balik kepada peserta didik, tetapi juga berpotensi menghadirkan ketidakkonsistenan dalam penilaian. Penelitian ini bertujuan untuk mengevaluasi efektivitas penggunaan teknologi kecerdasan buatan, khususnya model bahasa ChatGPT, dalam mengotomatiskan proses penilaian esai. Metode yang digunakan melibatkan implementasi sistem penilaian otomatis berbasis API ChatGPT, yang kemudian diuji terhadap sejumlah esai mahasiswa dengan berbagai topik dan tingkat kompleksitas. Evaluasi dilakukan dengan membandingkan hasil penilaian ChatGPT terhadap penilaian manual yang dilakukan oleh dosen. Hasil penelitian menunjukkan bahwa ChatGPT mampu memberikan penilaian yang konsisten dan efisien, dengan tingkat kesesuaian yang cukup tinggi terhadap penilaian manusia, terutama dalam aspek koherensi, relevansi isi, dan struktur penulisan. Kesimpulan dari penelitian ini menyatakan bahwa ChatGPT memiliki potensi besar sebagai alat bantu dalam penilaian esai secara otomatis, meskipun tetap diperlukan supervisi dari pengajar untuk menjamin kualitas dan keadilan penilaian.

Kata Kunci— penilaian otomatis, ChatGPT, esai, kecerdasan buatan, evaluasi pembelajaran

I. PENDAHULUAN

Dalam sistem pendidikan saat ini, penilaian terhadap tugas esai merupakan salah satu metode evaluasi yang efektif untuk mengukur kemampuan berpikir kritis, argumentatif, dan keterampilan komunikasi tertulis peserta didik [1]. Namun, penilaian esai secara manual masih menghadapi berbagai tantangan, terutama dari segi efisiensi waktu, objektivitas, dan konsistensi. Kelelahan penilai, subjektivitas interpretasi, serta beban kerja yang tinggi dapat mengurangi akurasi dan kecepatan pemberian umpan balik [2].

Seiring dengan perkembangan teknologi kecerdasan buatan (Artificial Intelligence/AI), pendekatan baru dalam mendukung proses evaluasi pembelajaran mulai dikembangkan. Salah satu teknologi yang mendapat perhatian luas adalah model bahasa besar (Large Language Model/LLM) seperti ChatGPT, yang dikembangkan oleh OpenAI dan berbasis pemrosesan bahasa alami (Natural Language Processing/NLP) [3].

Pemilihan ChatGPT dalam penelitian ini didasarkan pada kemampuannya yang unggul dalam memahami konteks, menghasilkan teks yang logis dan koheren, serta mengevaluasi struktur argumen secara kontekstual. Tidak seperti sistem penilaian otomatis tradisional yang cenderung berbasis skoring permukaan, ChatGPT mampu menganalisis isi dan struktur wacana dengan pendekatan semantik yang lebih dalam [4]. Selain itu, ChatGPT dapat diakses melalui Application Programming Interface (API), sehingga memudahkan integrasi ke dalam sistem pembelajaran yang ada, serta memungkinkan skalabilitas dalam lingkungan pendidikan tinggi.

Dengan mempertimbangkan kecepatan, konsistensi, dan fleksibilitas teknologi ini, ChatGPT dinilai memiliki potensi besar untuk digunakan sebagai alat bantu dalam penilaian otomatis esai. Oleh karena itu, penelitian ini bertujuan untuk mengevaluasi efektivitas penggunaan ChatGPT dalam penilaian esai mahasiswa, dengan membandingkan hasilnya terhadap penilaian manual yang dilakukan oleh tenaga pengajar. Selain hal teknis tentang perbandingan hasil penilaian, tulisan ini juga memaparkan pengaruh penggunaan ChatGPT dari aspek pedagogis, dan aspek etika.

II. METODE PENELITIAN

Penelitian ini menggunakan pendekatan studi kasus [5] yang bersifat kualitatif dengan fokus pada pemanfaatan ChatGPT dalam proses otomatisasi penilaian jawaban esai mahasiswa. Pendekatan studi kasus dipilih karena memungkinkan

¹ Dosen STMIK Antar Bangsa, Kawasan Bisnis CBD Ciledug, Jl. HOS Cokroaminoto No.29-35, RT.001/RW.001, Karang Tengah, Kec. Ciledug, Kota Tangerang, Banten 15157 (telp: 0858-1074-8494 e-mail: aristejo14@gmail.com)

melakukan kajian mendalam terhadap implementasi teknologi AI dalam proses evaluasi akademik di lingkungan pendidikan tinggi.

Subjek penelitian ini adalah mahasiswa aktif dari Program Studi Teknik Informatika di STMIK Antar Bangsa, yang mengikuti mata kuliah tertentu dan menjalani ujian esai. Subjek ditentukan secara purposif berdasarkan partisipasi mereka dalam ujian yang menjadi bagian dari data yang dianalisis.

Objek penelitian berupa jawaban ujian esai mahasiswa yang diperoleh dari dokumen ujian akhir semester. Jawaban-jawaban ini dianalisis, baik secara manual oleh dosen maupun secara otomatis menggunakan sistem berbasis ChatGPT.

A. Metode pengumpulan data, instrumen penelitian, dan metode pengujian

Metode yang digunakan dalam penelitian ini adalah Metode dokumentasi [6]. Tujuan menggunakan metode dokumentasi untuk memperoleh data berupa jawaban esai mahasiswa yang diambil dari berbagai sumber, seperti ujian akhir, tugas besar, maupun kuis. Dokumen-dokumen tersebut berperan sebagai objek utama yang akan dianalisis dan dinilai menggunakan ChatGPT. Selanjutnya, hasil penilaian otomatis tersebut akan dibandingkan dengan penilaian manual yang dilakukan oleh dosen untuk mengukur tingkat akurasi dan kesesuaian sistem. Tujuan utama penggunaan metode ini adalah mengumpulkan data primer dalam bentuk teks esai yang relevan dan representatif, sehingga dapat memberikan dasar yang kuat bagi proses evaluasi otomatis. Keunggulan dari metode dokumentasi terletak pada sifat datanya yang otentik serta kontekstual, karena diperoleh langsung dari lingkungan pendidikan yang sesungguhnya, sehingga mencerminkan situasi dan kompetensi akademik mahasiswa secara nyata.

B. Rubrik Penilaian untuk Esai Tingkat Pendidikan Tinggi

Rubrik penilaian merupakan suatu instrumen evaluatif yang berfungsi untuk mengukur kualitas suatu karya atau performa berdasarkan kriteria yang telah ditentukan secara sistematis. Dalam konteks pendidikan tinggi, rubrik penilaian esai menjadi alat penting yang digunakan dosen untuk menilai kemampuan berpikir kritis, argumentasi, pemahaman konsep, dan keterampilan komunikasi tertulis mahasiswa secara objektif dan konsisten.

Berdasarkan teori evaluasi pendidikan, rubrik umumnya terdiri dari dimensi penilaian (seperti isi, organisasi, argumentasi, dan penggunaan bahasa), indikator kinerja pada setiap dimensi, serta skala tingkat pencapaian (biasanya berupa deskripsi kualitatif dari tingkat rendah hingga tinggi). Dengan struktur ini, rubrik memberikan transparansi dalam proses penilaian serta umpan balik yang konstruktif kepada mahasiswa [7].

Penggunaan rubrik dalam penilaian esai mendukung prinsip *assessment for learning*, karena tidak hanya berfungsi sebagai alat ukur, tetapi juga sebagai panduan pembelajaran.

Mahasiswa dapat memahami ekspektasi akademik dan mengembangkan kemampuan menulis yang lebih baik dari waktu ke waktu. Selain itu, rubrik juga dapat meningkatkan reliabilitas penilaian antar-penilai (*inter-rater reliability*), yang penting dalam lingkungan universitas dengan jumlah peserta didik yang besar dan variasi dosen pengampu [8].

Rubrik memungkinkan dilakukannya penilaian secara holistik maupun analitik. Penilaian holistik memberikan skor keseluruhan berdasarkan kesan umum kualitas esai, sedangkan penilaian analitik memberikan skor terpisah pada setiap komponen, memungkinkan analisis kekuatan dan kelemahan mahasiswa secara lebih detail [9].

C. Tahapan penelitian

Pelaksanaan penelitian dilakukan melalui beberapa tahapan berikut:

1. Pengumpulan data berupa dokumen jawaban esai mahasiswa dari satu atau lebih mata kuliah pada semester tertentu.
2. Penilaian manual dilakukan oleh dosen pengampu mata kuliah menggunakan rubrik penilaian yang telah ditetapkan sebagai acuan standar institusi.
3. Penilaian otomatis dilakukan dengan menggunakan model ChatGPT yang diprogram untuk memberikan skor dan umpan balik terhadap jawaban esai berdasarkan parameter yang sepadan dengan rubrik manual.
4. Analisis komparatif dan tematik digunakan untuk menilai kesesuaian hasil penilaian antara manusia dan AI, serta mengidentifikasi pola penilaian, perbedaan skor, dan potensi kekuatan maupun kelemahan penggunaan ChatGPT sebagai alat bantu evaluasi.

Melalui pendekatan ini, penelitian diharapkan mampu memberikan gambaran komprehensif mengenai bagaimana ChatGPT dapat diimplementasikan secara nyata dalam konteks penilaian akademik, serta menyoroti implikasi teknis, pedagogis, dan etis dari penggunaannya di lingkungan pendidikan tinggi, khususnya dalam lingkup jurusan Teknik Informatika.

III. PEMBAHASAN DAN HASIL

Penilaian ujian esai memerlukan pendekatan yang sistematis dan adil, karena jenis soal ini menilai pemahaman mendalam, kemampuan berpikir kritis, serta keterampilan berkomunikasi tertulis. Tidak seperti soal pilihan ganda, esai mengharuskan penilai mempertimbangkan kualitas isi, logika, dan cara penyampaian gagasan. Secara umum, titik berat penilaian esai adalah sebagai berikut [10]:

1. Kesesuaian Jawaban terhadap Pertanyaan (Relevansi)
Kesesuaian jawaban terhadap pertanyaan merupakan indikator penting dalam menilai kualitas pemahaman seseorang terhadap soal yang diberikan. Penilaian ini mencakup sejauh mana jawaban yang disampaikan mampu secara langsung menjawab inti pertanyaan, serta ketepatan argumen atau penjelasan yang disampaikan sesuai dengan topik yang dimaksud.

Relevansi mencerminkan kemampuan peserta dalam memahami maksud soal dan memberikan respons yang tepat tanpa menyimpang dari konteks. Oleh karena itu, fokus utama terletak pada kejelasan dalam menafsirkan pertanyaan dan ketepatan dalam menyusun jawaban yang selaras, baik dari segi substansi maupun logika pemikiran yang mendasarinya.

2. Kedalaman dan Pemahaman Konsep

Kedalaman dan pemahaman konsep merupakan aspek mendasar dalam menilai kualitas jawaban. Penilaian ini tidak hanya melihat kemampuan mengingat informasi, tetapi lebih menekankan pada sejauh mana peserta memahami konsep secara mendalam dan mampu mengaitkannya dengan konteks yang relevan. Jawaban yang baik seharusnya mencerminkan adanya proses berpikir tingkat tinggi, seperti analisis, sintesis, atau evaluasi terhadap materi yang ditanyakan. Hal ini menjadi indikator penting dalam mengukur kemampuan berpikir kritis dan penguasaan substansi. Dengan demikian, penilaian difokuskan pada pemahaman yang bersifat konseptual, bukan sekadar pengulangan informasi yang dihafal.

3. Kelengkapan Informasi

Kelengkapan informasi merupakan salah satu indikator penting dalam menilai kualitas jawaban. Penilaian ini mencakup sejauh mana peserta mampu menyampaikan seluruh poin relevan yang berkaitan dengan pertanyaan, tanpa mengabaikan aspek penting yang dapat memengaruhi keutuhan makna. Jawaban yang baik tidak hanya mencantumkan inti pembahasan, tetapi juga didukung oleh argumen atau contoh yang relevan dan memperkuat penalaran. Dengan demikian, kelengkapan informasi mencerminkan tingkat ketelitian dan kecermatan peserta dalam memahami ruang lingkup pertanyaan serta menyusun jawaban secara menyeluruh, terstruktur, dan berbasis pada pemahaman yang utuh terhadap materi yang ditanyakan.

4. Tingkat akurasi teknis

Dalam konteks akademik dan ilmiah, akurasi teknis dan fakta merupakan aspek penting dalam menilai kualitas suatu jawaban. Penilaian ini mencakup ketepatan informasi yang disampaikan secara faktual serta penggunaan istilah teknis yang benar dan sesuai konteks. Jawaban yang akurat menunjukkan bahwa peserta memiliki pemahaman yang benar terhadap materi dan mampu menghindari penyampaian informasi yang keliru atau menyesatkan. Selain itu, penggunaan istilah teknis yang tepat mencerminkan penguasaan terhadap terminologi bidang keilmuan yang relevan. Oleh karena itu, akurasi menjadi dasar untuk menilai validitas dan kredibilitas dari argumen atau penjelasan yang diberikan.

5. Organisasi dan Struktur Penulisan

Dalam konteks tulisan akademik, organisasi dan struktur penulisan merupakan aspek penting dalam

menilai kualitas penyampaian jawaban. Penilaian ini mencakup sejauh mana ide-ide disusun secara teratur dan logis, sehingga membentuk alur pemikiran yang mudah dipahami. Jawaban yang baik idealnya memiliki tiga bagian utama: pengantar untuk membuka konteks, isi utama yang memuat argumen atau penjelasan, serta penutup yang merangkum poin penting. Struktur yang sistematis menunjukkan kemampuan peserta dalam mengelola informasi dan menyusun argumen secara tertib. Dengan demikian, organisasi tulisan mencerminkan keterampilan berpikir terstruktur serta kematangan dalam menyampaikan gagasan secara efektif.

6. Kejelasan Bahasa dan Gaya Penulisan

Kejelasan bahasa dan gaya penulisan adalah unsur penting dalam menilai kemampuan mengungkapkan pikiran dalam bentuk tulisan dari mahasiswa. Penilaian ini mencakup penggunaan bahasa yang jelas, efektif, serta bebas dari ambiguitas, sehingga pesan yang disampaikan mudah dipahami oleh pembaca. Selain itu, ketepatan dalam penggunaan tata bahasa, struktur kalimat, dan ejaan mencerminkan kecermatan serta kedisiplinan dalam berbahasa tulis. Gaya penulisan yang baik mendukung penyampaian ide secara teratur dan meyakinkan. Oleh karena itu, kejelasan bahasa dan ketepatan penulisan menjadi indikator penting dalam menilai seberapa baik mahasiswa mampu mengungkapkan pemikirannya secara sistematis, logis, dan profesional melalui tulisan.

7. Originalitas dan Kreativitas (opsional)

Originalitas dan kreativitas merupakan aspek tambahan yang penting, terutama dalam konteks ujian esai terbuka atau reflektif. Penilaian ini menekankan pada sejauh mana mahasiswa mampu menunjukkan pemikiran mandiri, tidak sekadar mengulang materi yang telah disampaikan di kelas. Jawaban yang orisinal mencerminkan kedalaman pemahaman dan keberanian dalam mengemukakan sudut pandang atau pendekatan baru terhadap topik yang dibahas. Kreativitas dalam menyusun argumen atau mengaitkan konsep juga menunjukkan kemampuan berpikir kritis dan inovatif. Dengan demikian, aspek ini menjadi indikator penting dalam menilai kemampuan mahasiswa dalam mengembangkan gagasan secara unik dan memperkaya diskusi akademik.

Uji penilaian ini menggunakan Rubrik Penilaian. Tujuannya untuk mengetahui sejauh mana ChatGPT dapat melakukan penilaian sebuah esai.

Uji perbandingan penilaian dilakukan terhadap salah satu soal esai dalam mata kuliah Sistem Operasi dan Jaringan Komputer.

Tujuan uji ini adalah untuk mengidentifikasi sejauh mana hasil penilaian ChatGPT sebanding dengan penilaian manual oleh dosen, baik dari segi skor maupun kualitas evaluasi.

Berikut adalah pengujian dan hasilnya pada kedua soal esai tersebut.

A. Uji perbandingan penilaian soal esai pada mata kuliah Sistem Operasi

Pada mata kuliah Sistem Operasi, mahasiswa diberikan soal esai dengan pertanyaan: "Jelaskan perbedaan utama antara multitasking dan multithreading dalam sistem operasi, serta berikan contoh penerapannya."

Hasil jawaban mahasiswa adalah sebagai berikut: Multitasking adalah kemampuan sistem operasi untuk menjalankan lebih dari satu aplikasi secara bersamaan, seperti menjalankan browser sambil bekerja di Microsoft Word. Sedangkan multithreading adalah teknik dalam sebuah proses atau program yang memungkinkan beberapa thread berjalan secara bersamaan dalam satu aplikasi. Contohnya, saat membuka banyak tab dalam browser atau saat aplikasi edit video melakukan rendering sambil memungkinkan pengguna mengedit timeline. Keduanya membuat sistem lebih efisien, namun multitasking bekerja di level proses, sedangkan multithreading di dalam satu proses.

Jawaban ini kemudian dinilai menggunakan ChatGPT dan Dosen Pengampu yang bersangkutan. Setelah itu hasil penilaian dipetakan kedalam rubrik evaluasi [11] [12] berikut:

TABEL I
RUBRIK HASIL PENILAIAN MATA KULIAH SISTEM OPERASI YANG DILAKUKAN OLEH CHATGPT DAN DOSEN PADA MATA KULIAH SISTEM OPERASI

Aspek	Bobot	Penilaian ChatGPT	Penilaian Dosen
Kesesuaian Jawaban	25%	24: Menjawab sesuai dan fokus pada perbedaan dan contoh.	22: Fokus tepat, tapi tidak eksplisit menjawab definisi awal.
Kedalaman Konsep	25%	22: Menyebutkan konsep dasar dengan baik.	19: Kurang membahas konteks sistem operasi modern.
Kelengkapan Informasi	20%	17: Ada contoh dan penjelasan proses/thread.	15: Tidak menyebut sistem operasi spesifik atau CPU core.
Keakuratan Teknis	15%	14: Benar secara teknis, tidak ada kesalahan istilah.	13: Akurat, tetapi perlu klarifikasi istilah thread vs proses.
Kejelasan Bahasa	15%	13: Jelas dan runtut.	14: Bahasa baik dan sesuai konteks akademik.
Total Nilai	-	90	83

Selain nilai, evaluator juga memberikan komentar sebagai berikut:

1. ChatGPT: Mahasiswa memberikan penjelasan yang tepat mengenai multitasking dan multithreading, disertai contoh yang relevan. Penekanan bahwa multitasking berada pada level proses dan multithreading di level thread sangat baik. Bahasa komunikatif dan jelas.
2. Dosen: Jawaban cukup baik secara umum, namun masih ada ruang untuk pendalaman, misalnya menjelaskan bagaimana sistem operasi mengatur thread di dalam prosesor multiprosesor atau bagaimana konteks switching terjadi. Mahasiswa juga bisa menambahkan contoh sistem operasi yang mendukung keduanya seperti Windows atau Linux untuk memperkaya jawaban.

B. Uji perbandingan penilaian soal esai pada mata kuliah Sistem Operasi

Pada mata kuliah Jaringan Komputer, mahasiswa diberikan soal esai dengan pertanyaan: "Jelaskan fungsi utama dari protokol TCP/IP dan mengapa protokol ini menjadi standar dalam komunikasi jaringan komputer."

Hasil jawaban mahasiswa adalah sebagai berikut: Protokol TCP/IP adalah kumpulan protokol yang digunakan untuk komunikasi data antar komputer dalam jaringan. TCP (Transmission Control Protocol) berfungsi memastikan data dikirim secara lengkap dan berurutan, sedangkan IP (Internet Protocol) mengatur alamat tujuan data agar sampai ke tempat yang tepat. TCP/IP menjadi standar karena bersifat terbuka, fleksibel, dan dapat digunakan di berbagai jenis perangkat dan sistem operasi. Selain itu, TCP/IP mendukung arsitektur client-server yang umum digunakan di internet.

Jawaban ini kemudian dinilai menggunakan ChatGPT dan Dosen Pengampu yang bersangkutan. Setelah itu hasil penilaian dipetakan kedalam rubrik evaluasi [13] [14] berikut:

TABEL II
RUBRIK HASIL PENILAIAN MATA KULIAH SISTEM OPERASI YANG DILAKUKAN OLEH CHATGPT DAN DOSEN PADA MATA KULIAH JARINGAN KOMPUTER

Aspek	Bobot	Penilaian ChatGPT	Penilaian Dosen
Kesesuaian Jawaban	25%	18: Jawaban sangat sesuai dengan pertanyaan, mencakup fungsi utama TCP dan IP serta alasan menjadi standar.	21: Jawaban sesuai, namun kurang mendalam dalam menekankan kata kunci "fungsi utama" dalam konteks komunikasi data.
Kedalaman Konsep	25%	16: Pemahaman dasar sudah baik, tapi belum menyentuh konsep yang	20: Pemahaman mendasar sudah tampak, tetapi masih kurang eksploratif.

		lebih mendalam seperti flow control atau pengalamatan subnet.	Mahasiswa belum menjelaskan bagaimana TCP menangani pengurutan dan pengiriman ulang data, serta bagaimana IP menangani pengalamatan dan pemilihan rute (routing).
Kelengkapan Informasi	20%	16: Informasi utama sudah ada, namun bisa ditambah contoh implementasi atau struktur lapisan TCP/IP.	16: Informasi yang diberikan cukup memadai, namun terasa ringkas. Tidak dijelaskan model layer TCP/IP, contoh protokol terkait, maupun aplikasinya dalam kehidupan nyata. Hal ini mengurangi kesan bahwa mahasiswa memahami konsep secara menyeluruh
Keakuratan Teknis	15%	20: Tidak ada kesalahan teknis dalam penjelasan fungsi dan alasan penggunaan TCP/IP.	14: Tidak ada kekeliruan berarti dalam aspek teknis. Terminologi digunakan dengan tepat. Akan tetapi, beberapa istilah masih dapat dijelaskan lebih spesifik untuk menghindari interpretasi yang terlalu luas.
Kejelasan Bahasa	15%	20: Bahasa sangat jelas, komunikatif, dan mudah dipahami tanpa	14: Bahasa yang digunakan rapi, jelas, dan mudah dipahami.

	kesalahan bahasa atau istilah.	Struktur kalimat sudah baik dan sesuai dengan gaya penulisan akademik. Namun, gaya penyampaian masih bisa ditingkatkan dengan menambahkan kalimat penghubung yang menunjukkan alur logis.
Total Nilai	- 90	85

C. Pembahasan

Analisa dari hasil penilaian dalam rubrik dan komentar yang diberikan adalah penilaian ChatGPT cenderung fokus pada kejelasan struktur jawaban, kesesuaian konteks, dan tidak ditemukan kesalahan teknis mendasar dalam jawaban mahasiswa. ChatGPT juga menilai tinggi pada aspek kejelasan bahasa dan kelengkapan contoh.

Sebaliknya, penilaian dosen menunjukkan pendekatan yang lebih mendalam dan kritis terhadap isi jawaban, terutama pada aspek konseptual dan kedalaman materi. Meskipun jawaban mahasiswa sudah cukup tepat, dosen memberi komentar bahwa penjelasan tidak mengulas lebih jauh mengenai cara sistem operasi mengelola thread pada arsitektur multiprosesor dan tidak menyebutkan sistem operasi sebagai konteks penerapan. Hal ini memengaruhi skor pada aspek kelengkapan dan kedalaman konsep.

Demikian juga pada mata kuliah Jaringan Komputer. Dosen memberikan komentar bahwa Jawaban mahasiswa sudah menunjukkan pemahaman awal yang baik terkait protokol TCP/IP. Struktur kalimat tertata dengan rapi, dan konsep dasar disampaikan dengan benar. Namun, esai ini masih bisa ditingkatkan dari sisi kedalaman dan kelengkapan. Mahasiswa belum menguraikan mekanisme teknis penting dalam TCP/IP seperti segmentasi data, error control, atau routing. Selain itu, esai akan lebih kuat bila dilengkapi contoh penerapan nyata, atau ilustrasi lapisan protokol. Untuk tugas selanjutnya, mahasiswa disarankan memperkaya materi dengan literatur tambahan serta menyusun argumen yang lebih terstruktur dan analitis.

Dari analisis rubrik, ditemukan bahwa perbedaan terbesar muncul pada aspek kedalaman konsep dan kelengkapan informasi, di mana dosen memberikan skor lebih rendah dibandingkan ChatGPT. Hal ini menunjukkan bahwa ChatGPT memiliki keterbatasan dalam menilai secara kritis aspek konseptual yang lebih kompleks atau tidak eksplisit dijelaskan oleh mahasiswa.

Secara umum, hasil uji ini menunjukkan bahwa ChatGPT dapat digunakan sebagai alat bantu dalam penilaian awal esai mahasiswa, khususnya untuk menilai aspek kesesuaian jawaban, keakuratan teknis, dan struktur bahasa. Namun demikian, intervensi dan validasi oleh dosen tetap diperlukan untuk memastikan kualitas penilaian, terutama dalam konteks evaluasi mendalam terhadap pemahaman teoritis mahasiswa.

ChatGPT dibangun dengan menggunakan machine learning, karena itu akan menguji kembali penilaian esai yang sama pada hari yang berbeda. Tujuannya untuk mengetahui bagaimana perilaku konsistensi ChatGPT dalam melakukan penilaian sebuah esai jika dilakukan pada hari yang berbeda.

Pengujian kedua kita menggunakan soal dan jawaban yang sama dari mata kuliah Sistem Operasi. Pengujian kedua ini mendapatkan hasil yang menarik ternyata saat proses dilakukan penilaian pada waktu yang berbeda maka ChatGPT akan memberikan hasil berbeda. Hal ini terjadi karena pada waktu yang berbeda, ChatGPT sudah memiliki data baru dan telah melakukan pembelajaran diri dari fungsi *machine learning* yang dimilikinya. Kondisi ini menjadi catatan penting bahwa peran dosen sangat diperlukan untuk memberikan nilai akhir dari setiap esai. Dengan kata lain, ChatGPT hanya dapat diposisikan sebagai asisten dosen yang bersangkutan.

Berikut adalah hasil dari penilaian ulang dari soal dan jawaban esai Sistem Operasi yang sama namun dilakukan pada waktu berbeda. Tag perintah yang digunakan juga yang sama dengan penilaian pada waktu sebelumnya seperti tercantum dalam gambar 1.

GAMBAR 1

TAG PERINTAH PENILAIAN KEPADA CHATGPT

Berikan penilaian dan komentar dari soal esai berikut. Soal: "Jelaskan perbedaan utama antara multitasking dan multithreading dalam sistem operasi, serta berikan contoh penerapannya." Jawaban: "Multitasking adalah kemampuan sistem operasi untuk menjalankan lebih dari satu aplikasi secara bersamaan, seperti menjalankan browser sambil bekerja di Microsoft Word. Sedangkan multithreading adalah teknik dalam sebuah proses atau program yang memungkinkan beberapa thread berjalan secara bersamaan dalam satu aplikasi. Contohnya, saat membuka banyak tab dalam browser atau saat aplikasi edit video melakukan rendering sambil memungkinkan pengguna mengedit timeline. Keduanya membuat sistem lebih efisien, namun multitasking bekerja di level proses, sedangkan multithreading di dalam satu proses." Gunakan rubrik penilaian berikut Kesesuaian Jawaban (25%), Kedalaman Konsep (25%), Kelengkapan Informasi (20%), Keakuratan Teknis (15%), Kejelasan Bahasa (15%). Berikan pula komentar. Tulis dalam bahasa yang formal dan edukatif



Perintah ini kemudian dieksekusi oleh ChatGPT. Setelah selesai eksekusi maka ChatGPT memberikan hasil penilaian yang dimasukkan kedalam rubrik seperti yang dituliskan di dalam tag perintah. Rubrik penilaian ChatGPT dapat dilihat pada gambar 2.

GAMBAR 2

RUBRIK PENILAIAN CHATGPT

Penilaian berdasarkan Rubrik

Aspek Penilaian	Bobot (%)	Nilai	Skor Akhir	
Kesesuaian Jawaban	25	23	5.75	
Kedalaman Konsep	25	21	5.25	
Kelengkapan Informasi	20	17	3.40	
Keakuratan Teknis	15	14	2.10	
Kejelasan Bahasa	15	14	2.10	
Total Skor	100		18.60 / 25	
Nilai Akhir (Skala 100)			86	

Jika kita bandingkan antara rubrik penilaian sebelumnya dengan rubrik penilaian terakhir maka terlihat ada hal menarik yaitu terdapat perubahan nilai yang signifikan dari yang sebelumnya ChatGPT memberi nilai 90 namun setelah dilakukan penilaian ulang pada hari yang berbeda nilainya turun menjadi 86. Ini perlu digarisbawahi oleh para pengguna ChatGPT dalam melakukan penilaian esai. Titik ini menjelaskan bahwa diperlukan kehati-hatian dalam mengandalkan mesin untuk memberikan penilaian secara total terhadap esai-esai khususnya esai mahasiswa. Titik ini menjelaskan pula bahwa peran dosen sangat penting sebagai pemberi nilai akhir dan sebagai penanggung jawab penuh hasil penilaian esai mahasiswa.

GAMBAR III

KOMENTAR CHATGPT

Komentar:

Jawaban ini menunjukkan pemahaman yang cukup baik terhadap konsep multitasking dan multithreading. Mahasiswa berhasil mengidentifikasi perbedaan utama, yaitu level eksekusi (proses vs. thread dalam proses), serta memberikan contoh yang relevan untuk masing-masing konsep. Penjelasan singkat tentang efisiensi sistem juga menjadi nilai tambah karena menunjukkan pemahaman implikatif dari kedua konsep tersebut.

Namun, dari aspek kedalaman konsep, masih terdapat ruang untuk perbaikan. Misalnya, penjelasan mengenai bagaimana sistem operasi mengelola konteks switching pada multitasking, atau manajemen thread dalam multithreading belum disinggung. Demikian pula, contoh multithreading yang disebutkan—membuka banyak tab pada browser—dapat menimbulkan ambiguitas karena tergantung pada implementasi browser-nya; langkah baiknya jika diganti dengan contoh yang lebih tepat secara teknis.

Dari sisi kelengkapan, definisi dan perbedaan sudah disampaikan, tetapi belum ada penjelasan singkat mengenai peran CPU, thread scheduler, atau bagaimana sistem operasi mengelola sumber daya. Penambahan elemen ini akan memberikan nilai lebih pada jawaban.

Bahasa yang digunakan sudah cukup jelas dan mudah dipahami, hanya perlu sedikit penyempurnaan dalam struktur kalimat agar lebih akademik.



Perhatikan gambar 3. Perbedaan yang mencolok terjadi juga pada komentar yang diberikan oleh ChatGPT, dimana komentar kali ini lebih detil dan lebih mendalam dibandingkan dengan komentar yang pertama. ChatGPT memberikan komentar dari aspek teknis seperti yang tercantum dalam paragraf 1.

Selain itu ChatGPT juga memberikan komentar dari aspek tingkat kedalaman konsep seperti yang tercantum pada paragraf 2. Disini tidak ketinggalan, ChatGPT juga memberikan penjelasan berupa contoh yang dapat dituliskan oleh mahasiswa terkait aspek kedalaman konsep tersebut. Bahkan ChatGPT memberikan komentar tentang ambiguitas dari potongan kalimat "membuka banyak tab pada browser" dan tidak ketinggalan diberikan pula saran penggantian kalimatnya.

Pada paragraf ketiga, ChatGPT memberikan komentar yang berupa saran kepada mahasiswa untuk mencoba menuliskan aspek lain terkait multitasking dan multithreading. Jika kita perhatikan, saran ini cukup menarik karena mesin ChatGPT memandang sudut lain yang tidak ada di dalam soal, jawaban, bahkan di dalam tag perintah.

Hal yang paling menarik adalah di paragraf terakhir. Apa yang membuat menarik? Disini ChatGPT memberikan penilaian atas gaya bahasa yang digunakan oleh mahasiswa dalam menjawab esai, dimana penilaian ChatGPT atas gaya bahasa tersebut dari aspek kejelasan penulisan dan kemudahan gaya bahasa tersebut dipahami oleh mesin. Sebenarnya masih ada lagi sisi lain yang bisa dinilai seperti aspek akademis mahasiswa dalam menjawab pertanyaan, struktur penulisan, dan logika berpikir. Namun sementara ini bisa disimpulkan, bahwa ChatGPT hanya mengambil kejelasan penulisan dan kemudahan pemahaman karena ia mengambil dari tag perintah dan membatasi diri pada rubrik penilaian yang diberikan.

GAMBAR IV

SARAN CHATGPT

Saran Perbaikan:

- Tambahkan penjelasan teknis singkat tentang bagaimana sistem operasi mengatur proses dan thread.
- Gunakan contoh yang lebih tepat dan tidak ambigu, misalnya: "Multithreading memungkinkan aplikasi pengolah kata melakukan pengecekan ejaan di latar belakang saat pengguna mengetik."
- Gunakan struktur bahasa yang lebih formal dan hindari penyederhanaan berlebihan pada istilah teknis.

Pada pengujian kali ini, ChatGPT memberikan saran yang tidak diberikan pada penilaian sebelumnya. Penulis berkesimpulan bahwa ChatGPT melakukan pembelajaran dari materi yang diberikan. Saran dari ChatGPT dapat dilihat pada gambar 4.

Secara umum, terlihat bahwa poin-poin saran yang diberikan ChatGPT merujuk kepada spesifikasi tertentu, terutama pada poin satu dan poin dua. Pada poin satu, ChatGPT memberikan saran teknis.

Pada poin 2, ChatGPT memberikan saran terkait ambiguitas. Kemampuan ini menjadi hal positif karena kemampuan seperti ini dapat menjadi masukan untuk dosen dalam memberikan saran dan penilaian akhir.

Yang lebih menarik lagi adalah saran pada poin ketiga tentang struktur bahasa yang digunakan mahasiswa dalam menjawab soal esai. Struktur bahasa adalah hal yang sangat subjektif yang sangat tergantung dari kemampuan masing-masing mahasiswa dalam berbahasa, menuliskan idenya kedalam kalimat, dan merangkai kalimat-kalimat dari awal hingga akhir supaya terstruktur dan berkesinambungan. Jika ChatGPT dapat melakukan hal ini secara menerus, maka akan sangat membantu dosen dalam melakukan penilaian-penilaian esai mahasiswa.

Selain itu hasil uji juga menunjukkan adanya pengaruh dengan aspek pedagogis dan aspek etika ketika ChatGPT digunakan untuk melakukan penilaian esai.

Aspek Pedagogis

Penilaian yang dilakukan oleh model kecerdasan buatan seperti ChatGPT dapat mempercepat proses evaluasi dan memberikan umpan balik awal yang objektif dan cepat. Namun perlu diperhatikan bahwa penggunaan ChatGPT dalam penilaian esai secara tidak sadar akan menimbulkan pergeseran peran dosen dari evaluator utama menjadi fasilitator dan validator penilaian. Sehingga dalam konteks ini, ChatGPT dapat diposisikan sebagai *co-assessor*, yaitu sistem pendukung yang mempercepat proses penilaian tanpa menghilangkan sentuhan reflektif manusia. Menurut Holmes dan Bialik, "*AI should not replace human teaching roles, but it can enhance educators' capacity to focus on higher-order tasks such as mentoring and conceptual feedback.*" Hal ini menekankan bahwa integrasi kecerdasan buatan harus memperkuat, bukan menggantikan dimensi pedagogis dari interaksi dosen dengan mahasiswa.

Aspek Etika

Ketika sistem otomatis menjadi dominan, maka bentuk-bentuk penilaian yang bersifat eksploratif dan reflektif akan digantikan oleh penilaian berbasis template atau keyword, yang lebih mudah dinilai oleh mesin, tetapi melemahkan pendidikan karakter dan pembelajaran khususnya di tingkat pendidikan tinggi.

Hal lain yang dapat terjadi adalah ketika penilaian sudah dipastikan otomatis yang dilakukan oleh ChatGPT dan ini menjadi kebiasaan rutin dalam jangka waktu lama tanpa ada keterlibatan dosen sedikitpun di dalam proses penilaian tersebut, maka akan memunculkan resiko bahwa dosen tidak lagi terdorong untuk merancang tugas-tugas esai yang kreatif dan kontekstual, melainkan hanya berfokus pada bentuk soal yang dapat dibaca dan dinilai oleh mesin.

IV. KESIMPULAN

Kesimpulan dari penelitian ini adalah sebagai berikut:

1. Model ChatGPT menunjukkan kemampuan yang cukup baik dalam melakukan penilaian esai, khususnya pada aspek kesesuaian isi, kejelasan bahasa, dan keakuratan teknis. Hal ini mencerminkan potensi pemanfaatan kecerdasan buatan dalam proses evaluasi akademik.
2. Perbedaan skor antara penilaian ChatGPT dan dosen tergolong tidak signifikan, yakni 90 untuk ChatGPT dan 83 untuk dosen. Hal ini menunjukkan bahwa hasil penilaian oleh ChatGPT relatif mendekati penilaian manual oleh dosen, setidaknya dalam konteks pertanyaan esai dengan tingkat kompleksitas sedang.
3. ChatGPT cenderung memberikan skor yang lebih tinggi, karena fokus utama penilaiannya terletak pada kejelasan struktur dan relevansi jawaban secara umum, namun belum sepenuhnya mampu mengidentifikasi kekurangan mendalam dalam aspek konseptual atau konteks yang lebih teknis.

4. Penilaian dosen lebih menekankan pada kedalaman pemahaman teori dan kelengkapan informasi, serta mampu mengidentifikasi kekurangan jawaban yang tidak dijelaskan secara eksplisit, namun penting dalam konteks akademik.
5. Jika penilaian dari soal dan jawaban yang sama namun dilakukan pada waktu yang berbeda, maka ChatGPT dapat memberikan perbedaan hasil penilaian yang signifikan, bahkan hingga ke tingkat komentar dan saran.
6. Dosen tetap memegang peran penting sebagai validator sekaligus penentu keputusan akhir dalam proses penilaian esai mahasiswa. ChatGPT diposisikan sebagai asisten yang membantu dosen dalam melakukan penilaian. Demikian pula, komentar dan saran yang dihasilkan tetap harus melalui validasi dosen yang bersangkutan sebelum nilai akhir ditetapkan.
7. Secara umum, kecerdasan buatan dapat digunakan untuk mendukung pendidikan khususnya untuk penilaian esai, namun tetap diperlukan pendekatan yang bersifat kolaboratif antara teknologi dan tenaga pengajar agar kualitas evaluasi tetap terjaga secara optimal.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada bapak Muchlis, MKom selaku kepala Program Studi Teknik Informatika atas dukungannya sehingga penelitian ini dapat berjalan dan membuahkan hasil.

REFERENSI

- [1] L. W. Anderson and D. R. Krathwohl, *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*, New York: Longman, 2001.
- [2] Y. Attali and J. Burstein, "Automated Essay Scoring with e-rater® V.2," *Journal of Technology, Learning, and Assessment*, vol. 4, no. 3, 2006. [Online]. Available: <https://ejournals.bc.edu/index.php/jtla/article/view/1650>
- [3] OpenAI, "GPT-4 Technical Report," OpenAI, 2023. [Online]. Available: <https://openai.com/research/gpt-4>
- [4] A. Borji, "A Categorical Archive of ChatGPT Failures," arXiv preprint arXiv:2302.03494, 2023. [Online]. Available: <https://arxiv.org/abs/2302.03494>
- [5] R. K. Yin, *Case Study Research and Applications: Design and Methods*, 6th ed. Thousand Oaks, CA, USA: SAGE Publications, 2018.
- [6] H. S. Mahdi and A. Alkhateeb, "Revolutionising Essay Evaluation: A Cutting-Edge Rubric for AI-Assisted Writing," *International Journal of Computer-Assisted Language Learning and Teaching*, vol. 15, no. 1, pp. 1–19, Jan.–Dec. 2025, doi: 10.4018/IJCALLT.368226
- [7] EdPlus Digital Learning Team, *Best Practices for Designing Effective Rubrics*, Arizona State University, April 30, 2025
- [8] S. J. Brookhart, *How to Create and Use Rubrics for Formative Assessment and Grading*, Alexandria, VA: ASCD, 2013.
- [9] EdPlus Digital Learning Team, *Best Practices for Designing Effective Rubrics*, Arizona State University, Apr. 30, 2025
- [10] C. Wetzler et al., "Grading the Graders: Comparing Generative AI and Human Assessment in Essay Evaluation," *J. Educ. Psychol.*, 2024.
- [11] S. Holmes and M. Bialik, *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*, Center for Curriculum Redesign, 2019. [Online]. Available: <https://curriculumredesign.org/wp-content/uploads/AI-in-Education.pdf>
- [12] Sugiyono, *Metode Penelitian Pendidikan: Pendekatan Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta, 2017
- [13] [3] H. Andrade, "Teaching with Rubrics: The Good, the Bad, and the Ugly," *College Teaching*, vol. 53, no. 1, pp. 27–30, Winter 2005.
- [14] J. Ouyang et al., "Training language models to follow instructions with human feedback," arXiv preprint, arXiv:2203.02155, 2022. [Online]. Available: <https://arxiv.org/abs/2203.02155>
- [15] OpenAI, "ChatGPT: Optimizing Language Models for Dialogue," *OpenAI*, Nov. 30, 2022. [Online]. Available: <https://openai.com/blog/chatgpt>
- [16] J. B. Biggs and C. Tang, *Teaching for Quality Learning at University*, 4th ed. Maidenhead, UK: McGraw-Hill Education, 2011.



Aristejo, lahir di Jakarta pada tanggal 14 Oktober 1978. Tahun 2001 Lulus dari Jurusan Teknik Sipil di Universitas Trisakti. Tahun 2004 lulus program Pasca Sarjana Ilmu Komputer dari Universitas Indonesia. Saat ini aktif mengajar sebagai dosen tetap pada program studi Teknik Informatika di STMIK Antar Bangsa. Beberapa jurnal yang sudah diterbitkan adalah sebagai berikut: Penerapan Low Code Development dalam Pengembangan Learning Management System (LMS) pada Platform Kursus Online Belajar Asyik, Penerapan Fungsi Mid Dan Find pada Pembersihan Data Alamat, Perancangan Aplikasi Pelacakan Program Kerja Kampus di STMIK Antar Bangsa.