

Studi Komparatif atas Pemanfaatan *Generative Artificial Intelligence* dalam Perkuliahan Akuntansi

Riko Riandoko ¹, Sony Hartono ², Rizqi Haniyah ^{3*}, Irfan Nabhani ⁴

^{1,2,3} Politeknik Keuangan Negara STAN; Jalan Bintaro Utama Sektor V, Tangerang Selatan;
e-mail: riandoko@pknstan.ac.id, sonyhart@pknstan.ac.id, rizqi.haniyah@pknstan.ac.id,

⁴ Universitas Garut; Jalan Prof. K.H. Cecep Syarifudin d/h Jl. Raya Samarang No.52A,
Kabupaten Garut. e-mail: irfan.nabhani@uniga.ac.id.

* Korespondensi: e-mail: rizqi.haniyah@pknstan.ac.id

Diterima: 04-11-2024; Review: 18-11-2024; Disetujui: 22-12-2024

Cara sitasi: Riandoko, Riko, Hartono, Sony, Haniyah, Rizqi, Nabhani, Irfan. 2024 Studi Komparatif atas Pemanfaatan *Generative Artificial Intelligence* dalam Perkuliahan Akuntansi. *Balance Vocation Accounting Journal*. Vol 8 (2): halaman.146-160

Abstrak: Perkembangan kecerdasan buatan generatif (*Generative AI*) beberapa tahun terakhir ini sangat pesat dan berpengaruh signifikan di berbagai lini kehidupan, tidak terkecuali dalam dunia akademik. Penelitian ini menggarisbawahi potensi *Generative AI* sebagai alat pembelajaran Akuntansi, aspek yang belum banyak dieksplorasi. Penelitian ini dilakukan dengan membandingkan respon yang dihasilkan oleh empat *Generative AI* (ChatGPT, Claude, Copilot, dan Gemini) dalam menjawab soal ujian mata kuliah Pengantar Akuntansi. Data diambil dengan memasukkan delapan soal Akuntansi dengan topik yang berbeda ke dalam setiap AI, lalu hasilnya dinilai dan dilakukan analisis menggunakan metode statistik deskriptif, uji-t, dan ANOVA. Hasil penelitian menunjukkan bahwa Claude memiliki nilai rata-rata tertinggi (87,5), diikuti oleh ChatGPT (77,1) dan Copilot (76,9), sementara Gemini mencatat nilai terendah (33,8). Uji-t dan ANOVA mengkonfirmasi adanya perbedaan signifikan antara performa masing-masing AI, dengan Claude terbukti paling unggul dengan rata. Penelitian ini juga menunjukkan bahwa tidak ada AI yang mampu memberikan jawaban sempurna pada semua soal akuntansi, sehingga penggunaannya harus tetap dievaluasi oleh individu. Penelitian ini memberikan kontribusi bagi pendidikan akuntansi, dengan menyarankan penggunaan AI sebagai alat bantu belajar yang efektif, meskipun tetap diperlukan pengawasan untuk menjaga kualitas dan keaslian hasil akademik.

Kata kunci: Akademik, Akuntansi, *Generative AI*, Kecerdasan Buatan, Pendidikan.

Abstract: The development of generative artificial intelligence (*Generative AI*) in recent years has been rapid and has had a significant impact across various aspects of life, including the academic world. This study highlights the potential of *Generative AI* as a tool for accounting education, an aspect that has not been widely explored. The exploration is conducted by comparing the responses generated by four *Generative AI*—ChatGPT, Claude, Copilot, and Gemini—in answering exam questions from the Accounting Principle course. Data were collected by inputting eight exam questions into each AI, the results were graded, then analyzed using descriptive statistics, *t*-tests, and ANOVA. The findings reveal that Claude achieved the highest average score (87.5), followed by ChatGPT (77.1) and Copilot (76.9), while Gemini recorded the lowest score (33.8). The *t*-tests and ANOVA confirmed significant differences in the performance of each AI, with Claude being the most superior. The study also shows that none of the AIs could provide perfect answers to all accounting problems, emphasizing the need for individual evaluation when using AI. This research contributes to accounting education by suggesting the use of AI as an effective learning aid, although supervision is still necessary to ensure the quality and authenticity of academic results.

Keywords: Academic, Accounting, *Generative AI*, Artificial Intelligence, Education.

1. Pendahuluan

Penggunaan kecerdasan buatan atau disebut *artificial intelligence* (AI) dalam berbagai bidang tidak lagi dapat dihindarkan saat ini. Kemampuan menghasilkan respon dengan cepat termasuk memberikan solusi dengan instan dapat mengurangi effort untuk menyelesaikan berbagai masalah atau pekerjaan dengan cukup signifikan. Kecerdasan buatan generatif (*Generative AI* atau GenAI) adalah bentuk AI yang dapat secara instan menghasilkan konten baru, seperti teks, gambar, audio, dan video [Lv, 2023]. Dengan model tersebut, *Generative AI* telah mengubah cara mencari dan mengakses informasi secara daring. Hanya dengan mengungkapkan pertanyaan ataupun *prompt*, penggunaanya dapat langsung memperoleh jawaban atau respon yang telah terstruktur dan lebih efektif. Sehingga tidak perlu lagi memilah dan merangkum informasi secara mandiri. Hal tersebut memberikan pengalaman pencarian yang lebih dinamis dan interaktif bagi para penggunaanya [Lv, 2023].

Penggunaan AI juga tidak luput merevolusi bidang Akuntansi [Peng et al., 2023]. Peng et al. [2023] mengeksplorasi peran AI untuk bidang Akuntansi menyatakan bahwa otomatisasi oleh AI dapat mengurangi konsumsi waktu dan sumber daya. Juniardi [2024] menyimpulkan bahwa AI dalam bidang Akuntansi berpotensi meningkatkan efisiensi, mengurangi kesalahan dalam mengaudit, dan mentransformasi proses audit melalui teknologi blockchain.

Pengetahuan, kemampuan, dan kompetensi Akuntan terkait *Generative AI* juga dijelaskan oleh Anica-Popa et al. [2024]. Mengintegrasikan *Generative AI* berbasis *Large Language Model* (LLM) dengan aktivitas operasional dapat mendukung para profesional akuntansi dan audit dalam proses pengambilan keputusan. LLM merupakan kecerdasan buatan (AI) yang dilatih pada data dalam jumlah besar untuk menghasilkan teks mirip manusia [Alberts et al., 2023]. Model tersebut yang memungkinkan pemrosesan data dalam jumlah besar secara paralel yang dapat memahami dan menghasilkan teks. Ballantine et al. [2024] menggambarkan kemajuan AI yang cepat sebagai "krisis yang memicu perubahan" dalam pendidikan akuntansi, sehingga evaluasi kurikulum sangat diperlukan.

Hal yang sejalan juga dinyatakan oleh Cooper [2023], bahwa kehadiran kecerdasan buatan generatif berpotensi mentransformasi dunia pendidikan. Pada November 2022, OpenAI.com merilis ChatGPT dalam versi beta. Mahasiswa dan dosen

pada perguruan tinggi mulai pun merespon dengan mengintegrasikan alat tersebut dengan rutinitas mereka [Castelló-Sirvent et al., 2023]. Dengan adanya AI, maka pelajar atau mahasiswa dapat memperoleh manfaat untuk menjalani pembelajaran yang lebih efektif, karena pembelajaran memungkinkan dapat disesuaikan dengan kebutuhan tiap-tiap individu [Putri et al., 2023].

Bagi pengajar, AI juga dapat mendukung penyelesaian tugas sebagai pendidik. Dengan bantuan AI, banyak pekerjaan rutin atau administratif seperti mengoreksi tugas, mencatat kehadiran, memberikan ujian, menjelaskan materi dasar, hingga membuat laporan, dapat dilakukan dengan lebih otomatis. Hal ini menjadikan dosen atau pengajar memiliki lebih banyak waktu dan energi untuk fokus pada aspek yang lebih penting dan kreatif, seperti mendidik karakter siswa dan mengembangkan kecerdasan emosional serta intelektual mereka, yang tidak bisa digantikan oleh mesin atau teknologi [Fitria, 2021].

Berdasarkan Liu and Wang [2024], ChatGPT merupakan *Generative AI* yang paling populer digunakan. Namun terdapat pula produk *Generative AI* lainnya yang menjadi alternatif ChatGPT yang juga memiliki fungsi yang sama misalnya Gemini, Copilot, Claude, dan lainnya. Dengan banyaknya pilihan tersebut, maka akan lebih baik jika para pengguna mengetahui AI yang mana yang paling baik untuk mereka, sehingga manfaat yang didapatkan dapat maksimal dengan *effort* yang kurang lebih sama.

Penelitian ini membandingkan kualitas *Generative AI* yang berpotensi digunakan dalam perkuliahan Akuntansi. Kualitas diukur dengan membandingkan respon *Generative AI* dalam menjawab soal Akuntansi. Apakah keakuratan respon yang dihasilkan relatif sama atau berbeda. Jawaban yang dihasilkan *Generative AI* dinilai dan hasilnya dibandingkan. Dengan demikian, dapat terjawab mengenai *Generative AI* mana yang lebih baik digunakan dalam pembelajaran Akuntansi.

Jika ternyata hasilnya tidak jauh berbeda, artinya *Generative AI* memiliki nilai atau kualitas yang sama. Sehingga untuk memilih AI dalam pengajaran Akuntansi, yang perlu dipertimbangkan adalah faktor-faktor lainnya, misalnya harga untuk berlangganan, garansi keamanan data yang ditawarkan, kenyamanan dalam penggunaannya, atau fitur-fitur tambahan yang ditawarkan oleh AI tersebut. Penelitian ini menggunakan uji-t dan anova selain statistika deskriptif, untuk mengukur perbedaan antara respon yang dihasilkan oleh AI dalam menjawab soal akuntansi.

Terdapat penelitian-penelitian sebelumnya yang juga mengukur ataupun membandingkan hasil respon AI untuk menyelesaikan permasalahan atau pekerjaan pada bidang tertentu. Penelitian yang dilakukan oleh Tahsin and Afkar [2023] yang melakukan studi perbandingan terhadap AI pada bidang desain interior. Selanjutnya penelitian yang dilakukan oleh Tiara and Pamuji [2024] yang mempelajari *chatbot* mana yang terbaik khususnya dari segi *usability*-nya. Wood et al. [2023] dan [Cheng et al., 2023] meneliti tentang penggunaan ChatGPT dalam pembelajaran Akuntansi.

2. Tinjauan Pustaka

2.1. Artificial Intelligence (AI) dan Generative AI

Kecerdasan buatan (AI) adalah kemampuan mesin untuk belajar dan berpikir seperti manusia. Mesin dengan AI dapat beradaptasi dan bekerja bersama manusia di masyarakat, menyelesaikan tugas-tugas rumit yang membutuhkan pemikiran, serta berkomunikasi dengan cara yang efisien melalui pesan yang padat informasi dan jelas.. Definisi di atas tidak melihat AI hanya sebagai teknologi atau produk, tetapi sebagai fenomena sosial dan kognitif [Abbass, 2021].

Generative AI adalah teknologi komputer yang bisa membuat konten baru seperti teks, gambar, atau audio berdasarkan data yang sudah ada, yang mengubah cara bekerja dan berkomunikasi kita saat ini [Feuerriegel et al., 2024]. *Generative AI* menghasilkan *output* yang mirip dengan karya manusia, dan saat ini mengubah berbagai industri [Bail, 2024].

Pemanfaatan AI dalam dunia kerja dapat meningkatkan kepuasan pelanggan, membuat para karyawan lebih betah bekerja, dan membantu para karyawan untuk dapat belajar lebih banyak. Penggunaan *Generative AI* juga dapat meningkatkan produktivitas, meskipun dampaknya berbeda-beda bagi setiap karyawan atau pekerja [Brynjolfsson et al., 2023].

2.2. AI dan Pendidikan Akuntansi

Penggunaan kecerdasan buatan (AI) berkembang pesat dalam praktik Akuntansi, dan perusahaan menghendaki karyawan baru yang telah mengadopsi teknologi ini, dengan demikian perguruan tinggi seharusnya merespon hal tersebut dengan mempersiapkan mahasiswa yang cakap menggunakan AI [Damerji and Salimi, 2021]. AI memiliki

dampak besar pada profesi akuntansi, dan para pendidik akuntansi perlu menyesuaikan kurikulum serta metode pengajaran mereka agar lulusan memiliki keterampilan yang dibutuhkan untuk sukses di industri yang terus berubah ini [Tandiono, 2023].

Akuntansi berbasis AI telah banyak digunakan banyak perusahaan di seluruh dunia. Dalam jangka panjang, AI akan menciptakan peluang kerja baru dan adopsi AI akan membawa era baru kreativitas dan inovasi yang mendorong perkembangan bidang akuntansi dan audit [Abdulameer et al., 2022]. Oleh karena itu, akuntansi berbasis AI merupakan peluang besar, tetapi hanya jika digunakan dan dikelola dengan baik [Abdulameer et al., 2022].

2.3. Integrasi AI dan Etika Penggunaannya

Penyebaran penggunaan *Generative AI* dan adopsi cepat oleh pengguna tidak hanya telah mengubah pola berbagai pekerjaan dan industri, tetapi juga menimbulkan pertanyaan penting tentang hak cipta, etika, keamanan, misinformasi, pasar tenaga kerja dan ketimpangan, masa depan kreativitas manusia, dan lain-lain [Liu and Wang, 2024]. *Generative AI*, menghadirkan peluang dan tantangan. Teknologi ini berpotensi mendorong otomatisasi dan meningkatkan efisiensi, namun, AI generatif menghadirkan tantangan signifikan, termasuk risiko perlindungan data, karena sering kali menangani informasi sensitif yang, jika disalahgunakan, dapat merusak privasi dan kepercayaan [Chen et al., 2023].

Integrasi otomatisasi dan AI dalam akuntansi menimbulkan pertimbangan penting. Kekhawatiran etika, keamanan data, dan potensi penggantian beberapa jenis pekerjaan adalah aspek penting yang membutuhkan pemeriksaan mendalam. Menemukan keseimbangan antara kemajuan teknologi dan mempertahankan standar etika akan menjadi tantangan krusial di masa depan dalam dunia akuntansi [Adeola Olusola Ajayi-Nifise et al., 2023]. Penggunaan AI sangat potensial dalam dunia pendidikan, misalnya AI dapat membantu mahasiswa dalam menyelesaikan tugas dengan cepat. Namun, untuk memastikan tidak terjadi plagiarisme atau kecurangan, maka mahasiswa harus memahami dan berpegang teguh pada prinsip-prinsip keilmuan [Patimah et al., 2024]. Pendidik juga harus secara kritis mengevaluasi setiap sumber daya yang dihasilkan oleh AI dan menyesuaikannya dengan konteks pengajaran masing-masing [Cooper, 2023].

3. Metode Penelitian

3.1. Desain penelitian

Penelitian ini didesain untuk membandingkan respon yang dihasilkan oleh AI dalam menjawab soal ujian Akuntansi. Mata kuliah yang dipilih adalah Pengantar Akuntansi. Mata kuliah ini merupakan mata kuliah pengenalan akuntansi, sehingga tingkat kompleksitasnya masih dibawah mata kuliah Akuntansi Menengah dan Akuntansi Lanjutan.

Tahap pertama yang dilakukan adalah mengumpulkan soal-soal ujian dengan topik atau materi akuntansi yang berbeda. Topik soal yang digunakan telah sesuai dengan kurikulum dan juga telah memiliki rubrik atau standar penilaian. Soal berupa soal *problem* (bukan soal pilihan ganda) sehingga hasilnya juga akan menunjukkan bagaimana respon setiap AI dalam memecahkan permasalahan yang membutuhkan analisis dan perhitungan dalam mata kuliah Pengantar Akuntansi. Rentang nilai untuk setiap soal adalah 0 s.d. 100.

Tahap selanjutnya adalah setiap soal diinput pada setiap AI. Tidak ada *prompt* tambahan saat melakukan input soal tersebut. AI yang digunakan dalam penelitian ini adalah ChatGPT, Claude, Copilot, dan Gemini. Penginputan soal pada AI dilakukan pada rentang waktu September 2024 s.d. Oktober 2024. Hasil respon atau *output* setiap AI untuk setiap soal kemudian dipindahkan ke dalam Microsoft Word, dan selanjutnya akan diberikan poin oleh pengajar, selayaknya pengajar tersebut memberikan penilaian hasil ujian mahasiswa. Nilai setiap soal yang dikerjakan oleh AI dikumpulkan kemudian dibandingkan hasilnya antara AI yang satu dengan yang lain. Selanjutnya atas hasil *output* tersebut dilakukan analisis kualitas dan akurasi jawaban atau respon.

3.2. Sampel Penelitian

Jumlah sampel yang digunakan terdiri dari 8 (delapan) ujian atau soal dengan topik: aset tetap, liabilitas jangka panjang (obligasi), liabilitas jangka pendek (current liability), ekuitas korporasi, partnership, arus kas, investasi obligasi, dan investasi saham. Jenis pertanyaan meliputi pembuatan jurnal transaksi, pembuatan jurnal penyesuaian, berbagai perhitungan terkait masing-masing topik, dan penyusunan laporan. Rekap data penelitian dilakukan dengan menggunakan Microsoft Excel, sedangkan uji statistik dilakukan dengan menggunakan AI Data Analyst pada Chat GPT. Nilai akan dibandingkan dengan menggunakan analisis statistik deskriptif. Selain itu, juga dilakukan

uji-t dan Anova untuk membandingkan hasil output atau jawaban dari Gemini, Claude, Chat GPT, dan Copilot.

3.3. Pengembangan Hipotesis

Hipotesis untuk studi ini dirancang untuk menguji perbedaan antara respon yang dihasilkan oleh AI. Hipotesis nol (H_0) menyatakan bahwa tidak ada perbedaan signifikan antara respon, sementara hipotesis alternatif (H_1) hingga (H_7) menyatakan bahwa ada perbedaan signifikan atas AI yang dibandingkan. Pengujian hipotesis ini akan dilakukan menggunakan uji-t dan Anova, metode statistik umum untuk membandingkan rata-rata dua kelompok. T-test akan menunjukkan apakah ada perbedaan yang signifikan secara statistik antara respons AI. Uji Anova digunakan untuk melakukan analisis komparasi multivariabel dengan mencari perbedaan signifikan dari beberapa nilai rata-rata.

Hipotesis 1:

1. Hipotesis Nol (H_0): Tidak ada perbedaan signifikan antara nilai yang dihasilkan oleh GenAI Gemini dan GenAI Claude.
2. Hipotesis Alternatif (H_1): Ada perbedaan signifikan antara nilai yang dihasilkan oleh GenAI Gemini dan GenAI Claude.

Hipotesis 2:

1. Hipotesis Nol (H_0): Tidak ada perbedaan signifikan antara nilai yang dihasilkan oleh GenAI ChatGPT dan GenAI Copilot.
2. Hipotesis Alternatif (H_2): Ada perbedaan signifikan antara nilai yang dihasilkan oleh GenAI ChatGPT dan GenAI Copilot.

Hipotesis 3:

1. Hipotesis Nol (H_0): Tidak ada perbedaan signifikan antara nilai yang dihasilkan oleh GenAI Gemini dan GenAI ChatGPT.
2. Hipotesis Alternatif (H_3): Ada perbedaan signifikan antara nilai yang dihasilkan oleh GenAI Gemini dan GenAI ChatGPT.

Hipotesis 4:

1. Hipotesis Nol (H_0): Tidak ada perbedaan signifikan antara nilai yang dihasilkan oleh GenAI Gemini dan GenAI Copilot.
2. Hipotesis Alternatif (H_4): Ada perbedaan signifikan antara nilai yang dihasilkan oleh GenAI Gemini dan GenAI Copilot.

Hipotesis 5:

1. Hipotesis Nol (H_0): Tidak ada perbedaan signifikan antara nilai yang dihasilkan oleh GenAI Claude dan GenAI ChatGPT.
2. Hipotesis Alternatif (H_5): Ada perbedaan signifikan antara nilai yang dihasilkan oleh GenAI Claude dan GenAI ChatGPT.

Hipotesis 6:

1. Hipotesis Nol (H_0): Tidak ada perbedaan signifikan antara nilai yang dihasilkan oleh GenAI Claude dan GenAI Copilot.
2. Hipotesis Alternatif (H_6): Ada perbedaan signifikan antara nilai yang dihasilkan oleh GenAI Claude dan GenAI Copilot.

Hipotesis 7 untuk ANOVA:

1. Hipotesis Nol (H_0): Tidak ada perbedaan signifikan dalam nilai rata-rata yang dihasilkan oleh empat GenAI (Gemini, Claude, ChatGPT, dan Copilot).
2. Hipotesis Alternatif (H_7): Ada perbedaan signifikan dalam nilai rata-rata yang dihasilkan oleh setidaknya satu dari empat GenAI (Gemini, Claude, ChatGPT, dan Copilot).

3.4. Pengujian Statistik Uji-T dan Anova

Pengujian secara statistik dengan t-test dan Anova ditambahkan untuk memastikan validitas hasil penelitian. Hasil t-test menunjukkan hasil terkait hipotesis penelitian yang telah ditetapkan. Jika H_0 ditolak, artinya adalah terdapat perbedaan signifikan antara respons yang dihasilkan oleh AI yang dibandingkan. Hasil analisis ini akan memberikan informasi berharga tentang AI yang saat ini paling baik untuk mendukung kegiatan akademik, khususnya mata pada mata kuliah Pengantar Akuntansi.

4. Hasil dan Pembahasan

4.1. Hasil Statistik Deskriptif

Terdapat empat *Generative AI* yang menjawab soal evaluasi dengan topik akuntansi yaitu Gemini, Claude, ChatGPT, dan Copilot. Berdasarkan hasil analisis statistika deskriptif, rata-rata skor tertinggi diraih oleh Claude dengan nilai rata-rata sebesar 87,5. Selanjutnya di posisi kedua terdapat oleh Copilot (76,9), ChatGPT (77,0), dan Gemini (33,8). Gemini mencatat nilai rata-rata terendah. Hal ini menunjukkan bahwa

Gemini tidak efektif dalam menjawab soal dengan topik akuntansi dibandingkan AI lainnya.

Dari segi variasi dengan standar deviasi, Gemini memiliki variasi yang cukup besar (12,9), yang menandakan adanya perbedaan yang signifikan antar skornya. Claude merupakan yang paling konsisten dengan variasi terkecil (9,0). Untuk ChatGPT dan Copilot, keduanya memiliki standar deviasi yang hampir sama, yaitu sekitar 9,3 dan 8,3, yang juga menunjukkan distribusi skor yang relatif paling stabil. Hasil ini menunjukkan perbedaan yang cukup signifikan dalam efektivitas setiap GenAI dalam mengevaluasi topik akuntansi, yang dapat menjadi pertimbangan penting dalam pemilihan Generative AI yang paling sesuai untuk pembelajaran Akuntansi.

Tabel 1. Hasil Statistik Deskriptif

Statistik Deskriptif	Gemini	Claude	ChatGPT	Copilot
Count	8	8	8	8
Mean	33,8	87,5	77,1	76,9
Standard Deviation	12,9	9,0	9,3	8,3
Min	22,0	78,0	57,5	62,0
25%	27,5	80,0	74,8	73,3
50%	31,0	84,8	79,0	80,0
75%	33,8	94,0	80,5	80,5
Max	64,0	100,0	90,0	88,0

sumber: Diolah Penulis

Saat ini ChatGPT merupakan *Generative AI* yang paling banyak digunakan. Namun, hasil penelitian ini dapat menjadi referensi bahwa terdapat AI yang lebih baik dan akurat dalam menjawab soal-soal akuntansi yaitu Claude. Keakuratan tersebut ditunjukkan bahwa terdapat soal yang dapat dijawab dan mendapat nilai maksimal. Selanjutnya Chat GPT dan Copilot memiliki kualitas yang hampir sama tetapi masih di bawah Claude. *Generative AI* yang paling tidak unggul di antara keempatnya adalah Gemini dengan keakuratan tertinggi hanya mencapai 66, sedangkan terendah bahkan hanya mencapai nilai 22.

4.2. Hasil Pengujian Hipotesis

Hasil uji t seperti yang ditunjukkan pada tabel 2, menunjukkan adanya perbedaan signifikan antar *Generative AI* yang digunakan untuk mengevaluasi topik akuntansi. Hipotesis 1, yang membandingkan nilai tes dari Gemini dan Claude, menghasilkan p-

value $< 0,05$, menunjukkan adanya perbedaan yang signifikan antara keduanya. Hal serupa juga ditemukan pada Hipotesis 3 dan 4, yaitu pada GenAI Gemini dibandingkan dengan ChatGPT, dan Gemini dibandingkan dengan Copilot, yang menunjukkan bahwa terdapat perbedaan signifikan, karena p-value yang juga lebih kecil dari 0,05. Ini menegaskan bahwa hasil Gemini secara konsisten berbeda dengan 3 Generative AI lainnya.

Untuk Hipotesis 5 yang membandingkan Claude dengan ChatGPT, dan Hipotesis 6 yang membandingkan Claude dengan Copilot, keduanya juga menunjukkan adanya perbedaan signifikan. Ini menunjukkan bahwa Claude dengan skor rata-rata yang lebih tinggi, memiliki perbedaan yang signifikan secara statistik dengan ChatGPT dan Copilot.

Hasil yang berbeda diperoleh pada Hipotesis 2. Hipotesis ini membandingkan nilai tes dari ChatGPT dan Copilot. Hasil ini satu-satunya yang menunjukkan bahwa tidak ada perbedaan signifikan antara kedua Generative AI tersebut, karena p-value lebih besar dari 0,05. Sementara itu, hasil perhitungan ANOVA menunjukkan nilai p-value adalah lebih kecil dari 0,05. Hal ini menjelaskan bahwa terdapat perbedaan signifikan setidaknya satu dari empat Generative AI lainnya, yang juga berarti bahwa tidak semua Generative AI menghasilkan nilai yang serupa.

Tabel 2 Hasil Uji t dan ANOVA

Hypothesis	T-Statistic	P-value	Result
Hypothesis 1: Gemini vs Claude	-9,642	0,000	Reject Ho
Hypothesis 2: ChatGPT vs Copilot	0,043	0,967	Failed to Reject Ho
Hypothesis 3: Gemini vs ChatGPT	-7,694	0,000	Reject Ho
Hypothesis 4: Gemini vs Copilot	-7,942	0,000	Reject Ho
Hypothesis 5: Claude vs ChatGPT	2,269	0,040	Reject Ho
Hypothesis 6: Claude vs Copilot	2,441	0,029	Reject Ho
Hypothesis 7: Anova: $\mu_{\text{Gemini}} = \mu_{\text{Claude}} = \mu_{\text{ChatGPT}} = \mu_{\text{Copilot}}$	45,14	0,000	Reject Ho

Sumber: Diolah Penulis

4.3. Pembahasan

Berdasarkan pengujian hipotesis menggunakan uji-t dan ANOVA, ditemukan bahwa terdapat perbedaan signifikan antara nilai yang dihasilkan oleh berbagai Generative AI dalam pengerjaan soal Akuntansi. Analisis ini memberikan wawasan mengenai seberapa besar perbedaan nilai yang diperoleh dari keempat AI tersebut serta implikasi dalam pemilihan AI yang paling sesuai dalam pengajaran Akuntansi.

Hasil uji t dengan membandingkan 6 (enam) kombinasi pasangan Generative AI menunjukkan bahwa Gemini secara konsisten memberikan nilai yang berbeda secara signifikan dibandingkan dengan Claude, ChatGPT, dan Copilot. P-value yang di bawah 0,05 mengonfirmasi bahwa Gemini menghasilkan nilai yang secara statistik berbeda serta cenderung lebih rendah. Claude juga memiliki perbedaan signifikan dengan yang lain. Namun dalam hal ini, Claude menghasilkan nilai rata-rata yang secara signifikan lebih baik dibanding yang lain. Hasil yang mirip diperoleh ChatGPT dan Copilot. Keduanya tidak memiliki perbedaan signifikan dalam mencapai nilai untuk menyelesaikan soal akuntansi.

Hasil tersebut diperkuat oleh hasil uji ANOVA yang menunjukkan adanya perbedaan signifikan dalam nilai rata-rata empat GenAI. Nilai p-value yang di bawah 0,05 menegaskan bahwa setidaknya terdapat satu *Generative AI* yang secara signifikan menghasilkan nilai berbeda dibandingkan yang lain. Hasil analisis ini, menggarisbawahi pentingnya melakukan validasi terhadap *Generative AI* yang digunakan untuk tujuan spesifik seperti misalnya dalam perkuliahan atau pembelajaran untuk topik mata kuliah akuntansi.

Hasil uji statistik yang telah dilakukan menunjukkan bahwa tidak semua GenAI dapat diandalkan untuk membantu mempelajari mata kuliah Akuntansi contohnya dalam hal menjawab soal. Dari keempat AI yang diuji, Claude merupakan AI yang paling handal digunakan dalam pengerjaan soal pengantar akuntansi dengan rata-rata hasil yang diperoleh 87,5. Kemudian diikuti oleh ChatGPT dan Copilot dengan rata-rata hampir sama yaitu 77,1 untuk Chat GPT dan 76,9 untuk Copilot. Gemini memperoleh hasil rata-rata jauh dari ketiganya, bahkan kurang dari setengah yang dapat dicapai oleh ketiga GenAI lainnya.

Di sisi lain, nilai rata-rata tersebut juga menunjukkan bahwa dari keempat GenAI yang diuji, tidak ada yang menjawab soal seluruhnya dengan sempurna. Hal ini menjadi catatan penting bagi bagi siswa maupun pengajar bahwa GenAI tidak selalu benar dalam memberikan jawaban atau solusi, sehingga untuk mencapai hasil yang lebih baik tetap memerlukan analisis dan evaluasi individu. Patimah et al. [2024] menyampaikan bahwa penting bagi setiap individu untuk memahami cara menggunakan GenAI secara efektif, karena masih sering ditemukan jawaban yang belum jelas kebenarannya pada *output* AI. Penggunaan GenAI yang baik bisa dilakukan dengan meningkatkan literasi digital. Setiap

orang memiliki tanggung jawab untuk mengasah keterampilan serta mengevaluasi sumber informasi yang diperoleh. Hasil yang dihasilkan dari GenAI harus dapat dipertanggungjawabkan kebenarannya [Patimah et al., 2024].

Dengan memanfaatkan GenAI dengan bijak, akan membantu mempermudah dan mengurangi waktu untuk menyelesaikan tugas atau pekerjaan, misalnya dalam mencari referensi. Para pengajar dapat pula memanfaatkan GenAI antara untuk membantu menyiapkan bahan ajar, membuat soal, mengoreksi jawaban, dan menyelesaikan kebutuhan administrasi. Dengan demikian, para pengajar dapat lebih menghemat waktu dan lebih fokus pada tugas-tugas yang lebih strategis dan membutuhkan kreativitas, inovasi, dan berpikir kritis.

4. Kesimpulan

Hasil penelitian menunjukkan bahwa terdapat perbedaan keandalan untuk hasil penggunaan Generative AI atas nilai ujian mata kuliah Pengantar Akuntansi. Dari keempat GenAI yang digunakan untuk menilai, yang paling unggul dengan hasil nilai rata-rata tertinggi atas nilai seluruh soal adalah Claude. Selanjutnya adalah ChatGPT dan Copilot yang memberikan hasil nilai rata-rata yang hampir sama. Sedangkan yang paling tidak mumpuni pada test ini adalah Gemini menghasilkan kinerja paling rendah dibandingkan lainnya.

Jawaban yang diberikan GenAI tidak selalu benar, sehingga untuk mencapai hasil yang lebih baik tetap akan memerlukan evaluasi individu yang menggunakan GenAI tersebut. Penggunaan GenAI yang tepat akan sangat membantu simplifikasi dan menghemat waktu untuk mencari referensi materi dan memperdalam mata pelajaran dalam perkuliahan. Dengan kata lain, GenAI akan sangat membantu para mahasiswa dalam belajar dan meningkatkan performa akademiknya, dan juga bagi para pengajar dapat membantu penyelesaian pekerjaan dengan lebih efektif dan efisien, sehingga dapat berfokus pada tugas yang lebih strategis. Penelitian ini terbatas hanya terbatas mengeksplorasi penggunaan Generative AI pada perkuliahan Akuntansi, sehingga penelitian lain dapat melakukan eksplorasi pada berbagai bidang lainnya dengan menguji Generative AI yang sesuai dengan bidang-bidang tersebut.

Referensi

- Abbass H. 2021. Editorial: What is Artificial Intelligence? *IEEE Transactions on Artificial Intelligence* 2: 94–95.
- Abdulameer M, Mansoor MM, Alchuban M, Rashed A, Al-Showaikh F, Hamdan A. 2022. The Impact of Artificial Intelligence (AI) on the Development of Accounting and Auditing Profession., p 201–213.
- Adeola Olusola Ajayi-Nifise, Olubusola Odeyemi, Noluthando Zamanjomane Mhlongo, Chidera Victoria Ibeh, Oluwafunmi Adijat Elufioye, Kehinde Feranmi Awonuga. 2023. The future of accounting: Predictions on automation and AI integration. *World Journal of Advanced Research and Reviews* 21: 399–407.
- Alberts IL, Mercolli L, Pyka T, Prenosil G, Shi K, Rominger A, Afshar-Oromieh A. 2023. Large language models (LLM) and ChatGPT: what will the impact on nuclear medicine be? *Eur J Nucl Med Mol Imaging* 50: 1549–1552.
- Anica-Popa I-F, Vrîncianu M, Anica-Popa L-E, Cişmaşu I-D, Tudor C-G. 2024. Framework for Integrating Generative AI in Developing Competencies for Accounting and Audit Professionals. *Electronics (Basel)* 13: 2621.
- Bail CA. 2024. Can Generative AI improve social science? *Proceedings of the National Academy of Sciences* 121.
- Ballantine J, Boyce G, Stoner G. 2024. A critical review of AI in accounting education: Threat and opportunity. *Critical Perspectives on Accounting* 99: 102711.
- Brynjolfsson E, Li D, Raymond L. 2023. *Generative AI at Work*. Cambridge, MA.
- Castelló-Sirvent F, García Félix VE, Canós-Darós L. 2023. AI IN HIGHER EDUCATION: NEW ETHICAL CHALLENGES FOR STUDENTS AND TEACHERS., p 4463–4470.
- Chen B, Wu Z, Zhao R. 2023. From fiction to fact: the growing role of generative AI in business and finance. *Journal of Chinese Economic and Business Studies* 21: 471–496.
- Cheng X (Joyce), Dunn R, Holt T, Inger K, Jenkins JG, Jones J, Long J, Loraas TM, Mathis M, Stanley JD, Wood DA. 2023. Artificial Intelligence’s Capabilities, Limitations, and Impact on Accounting Education: Investigating ChatGPT’s Performance on Educational Accounting Cases. *SSRN Electronic Journal*.

- Cooper G. 2023. Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence. *J Sci Educ Technol* 32: 444–452.
- Damerji H, Salimi A. 2021. Mediating effect of use perceptions on technology readiness and adoption of artificial intelligence in accounting. *Accounting Education* 30: 107–130.
- Feuerriegel S, Hartmann J, Janiesch C, Zschech P. 2024. Generative AI. *Business & Information Systems Engineering* 66: 111–126.
- Fitria TN. 2021. Artificial intelligence (AI) in education: Using AI tools for teaching and learning process. In: *Prosiding Seminar Nasional & Call for Paper STIE AAS (Vol. 4, No. 1)*., p 134–137.
- Juniardi E. 2024. PERAN DAN PRAKTIK ARTIFICIAL INTELLIGENCE AKUNTANSI: SYSTEMATIC LITERATURE REVIEW. *JURNAL REVENUE : JURNAL ILMIAH AKUNTANSI* 4.
- Liu Y, Wang H. 2024. *Who on Earth Is Using Generative AI ?* Washington, DC: World Bank.
- Lv Z. 2023. Generative artificial intelligence in the metaverse era. *Cognitive Robotics* 3: 208–217.
- Patimah NN, Rahmanita MA, Raharja RM. 2024. Adaptasi Penggunaan Artificial Intelligence (AI) Pada Mahasiswa. *International Journal of Educational Technology and Society*: 114–123.
- Peng Y, Ahmad SF, Ahmad AYAB, Al Shaikh MS, Daoud MK, Alhamdi FMH. 2023. Riding the Waves of Artificial Intelligence in Advancing Accounting and Its Implications for Sustainable Development Goals. *Sustainability* 15: 14165.
- Putri VA, Sotyawardani KCA, Rafael RA. 2023. Peran artificial intelligence dalam proses pembelajaran mahasiswa di Universitas Negeri Surabaya. In: *Prosiding Seminar Nasional Ilmu Ilmu Sosial (SNIIS)*., p 615–630.
- Tahsin M, Afkar MA. 2023. Analisis Komprehensif: Perbandingan Platform Perangkat Lunak Artificial Intelligence (AI) untuk Meningkatkan Inovasi dalam Desain Interior. *Journal of Engineering and Science* 2: 97–111.
- Tandiono R. 2023. The Impact of Artificial Intelligence on Accounting Education: A Review of Literature. *E3S Web of Conferences* 426: 02016.

- Tiara T, Pamuji FY. 2024. KOMPARASI USABILITY CHATGPT VS GEMINI AI BERDASARKAN ISO/IEC 9126 DAN NIELSEN MODEL MENGGUNAKAN METODE USABILITY TESTING. JUSIM (Jurnal Sistem Informasi Musirawas) 9: 89–100.
- Wood DA, Achhpilia MP, Adams MT, Aghazadeh S, Akinyele K, Akpan M, Allee KD, Allen AM, Almer ED, Ames D, Arity V, Barr-Pulliam D, Basoglu KA, Belnap A, Bentley JW, Berg T, Berglund NR, Berry E, Bhandari A, Bhuyan MNH, Black PW, Blondeel E, Bond D, Bonrath A, Borthick AF, Boyle ES, Bradford M, Brandon DM, Brazel JF, Brockbank BG, Burger M, Byzalov D, Cannon JN, Caro C, Carr AH, Cathey J, Cating R, Charron K, Chavez S, Chen J, Chen JC, Chen JW, Cheng C, Cheng X, Christensen BE, Church KS, Cicone NJ, Constance P, Cooper LA, Correia CL, Coyne J, Cram WA, Curtis A, Daigle RJ, Dannemiller S, Davenport SA, Dawson GS, De Meyst KJL, Dell S, Demirkan S, Denison CA, Desai H, DeSimone S, Diehl LM, Dimes R, Dong B, Donnelly A, du Pon A, Duan HK, Duffey A, Dunn RT, Durkin MP, Dzurainin AC, Eberle RM, Ege MS, El Mahdy D, Esplin A, Eulerich M, Everaert P, Farah N, Farish L, Favere-Marchesi M, Fayard D, Filosa JR, Ford M, Franz DR, Fulmer BP, Fulmer S, Furner ZZ, Gantman S, Garner S, Garrett J, Geng X, Golden J, Goldman W, Gomez J, Gooley M, Granitto SP, et al. 2023. The ChatGPT Artificial Intelligence Chatbot: How Well Does It Answer Accounting Assessment Questions? *Issues in Accounting Education* 38: 81–108.