

Komparasi Respons ChatGPT dan Gemini terhadap *Command Pattern* Identik dengan Metode *Black Box*

Muchlis¹, Rusyda Maulida²

Abstract— This study aims to analyze the comparison between ChatGPT and Gemini using blackbox testing methods. The testing was conducted based on predefined rules to evaluate aspects such as answer accuracy, response speed, conversation context, and personalized responses. The results indicate that ChatGPT excels in terms of response quality and personalization, particularly in handling complex questions and maintaining conversation context. On the other hand, Gemini stands out in response efficiency but tends to provide more general answers. Overall, both chatbots have their respective strengths, and the choice between them can be adjusted based on user needs, prioritizing either speed or depth of response.

Intisari— Penelitian ini bertujuan untuk menganalisis perbandingan antara ChatGPT dan Gemini menggunakan metode pengujian blackbox. Pengujian dilakukan dengan beberapa aturan yang telah ditetapkan untuk mengevaluasi aspek seperti ketepatan jawaban, kecepatan respons, konteks percakapan, dan personalisasi jawaban. Hasil pengujian menunjukkan bahwa ChatGPT unggul dalam hal kualitas dan personalisasi jawaban, terutama dalam menangani pertanyaan yang memerlukan penjelasan kompleks dan mempertahankan konteks percakapan. Di sisi lain, Gemini menonjol dalam efisiensi kecepatan respons namun cenderung memberikan jawaban yang lebih umum. Secara keseluruhan, kedua chatbot memiliki kelebihan masing-masing, dan pemilihan antara keduanya dapat disesuaikan dengan kebutuhan pengguna, apakah lebih mengutamakan kecepatan atau kedalaman respons.

Kata Kunci— ChatGPT, Gemini, Black Box.

I. PENDAHULUAN

Perkembangan kecerdasan buatan (AI) telah membawa perubahan signifikan dalam berbagai bidang, termasuk dalam cara manusia berinteraksi dengan teknologi. Salah satu aplikasi AI yang paling menonjol adalah model bahasa seperti ChatGPT dan Gemini. Kedua model ini dikembangkan untuk memahami dan menghasilkan teks dalam bahasa alami, yang memungkinkan mereka untuk digunakan dalam berbagai tugas, mulai dari penulisan kreatif hingga analisis data. Salah satu pendekatan yang digunakan dalam pengembangan aplikasi berbasis AI ini adalah *pattern command*, yang memberikan struktur untuk menangani operasi yang dapat diubah menjadi objek yang terpisah.

Model bahasa GPT-3 digunakan oleh ChatGPT, sebuah chatbot, untuk menghasilkan balasan sebagai respons terhadap masukan dari pengguna[1]. Generative Pre-trained Transformer 3, atau GPT-3, adalah model bahasa berskala besar yang dibuat oleh OpenAI dan mampu menghasilkan teks dengan 175 miliar parameter. Model ini telah dilatih menggunakan sejumlah besar data [2]. ChatGPT memanfaatkan kemampuan GPT-3 untuk memberikan balasan terhadap masukan pengguna dengan cara yang percakapan dan alami[1].

Gemini, model AI yang dikembangkan oleh Google DeepMind, didasarkan pada arsitektur multimodal canggih yang dirancang untuk menangani dan mengintegrasikan berbagai jenis informasi seperti teks, gambar, audio, video, dan bahkan kode. Model ini sangat serbaguna dan hadir dalam beberapa versi—Ultra, Pro, dan Nano—untuk memenuhi berbagai kebutuhan komputasi, mulai dari aplikasi pusat data yang kuat hingga tugas di perangkat seperti pada smartphone Google Pixel 8 Pro [3].

Salah satu fitur utama Gemini adalah multimodalitas bawaan, yang berarti model ini dirancang sejak awal untuk secara mulus memproses dan melakukan penalaran lintas berbagai jenis data. Ini merupakan kemajuan signifikan dibandingkan model sebelumnya yang sering kali memerlukan pelatihan terpisah untuk setiap jenis data sebelum menggabungkannya. Kemampuan canggih Gemini memungkinkannya untuk melampaui model AI sebelumnya, termasuk ChatGPT-3.5, di berbagai tolok ukur, terutama dalam tugas-tugas penalaran dan pemahaman yang kompleks.

Model baru ini mewakili salah satu proyek AI paling ambisius Google hingga saat ini, dengan tujuan untuk meningkatkan pengalaman pengguna di berbagai aplikasi, mulai dari mesin pencari hingga perangkat pintar[3].

Dalam konteks perkembangan teknologi AI, penelitian terbaru menunjukkan bahwa model bahasa besar seperti ChatGPT dan GPT-4 memiliki potensi aplikasi yang luas di berbagai domain. Inovasi dalam pre-training skala besar, fine-tuning, dan pembelajaran dari umpan balik manusia telah meningkatkan kinerja model-model ini secara signifikan. Temuan dari 194 makalah di arXiv menunjukkan bahwa ada minat yang meningkat terhadap penggunaan ChatGPT/GPT-4, terutama dalam pemrosesan bahasa alami, serta potensi besar di bidang pendidikan, sejarah, matematika, dan ilmu pengetahuan lainnya. Penelitian ini juga menyoroti pentingnya pertimbangan etis dan memberikan panduan untuk pengembangan lebih lanjut di masa depan[4].

Namun, meskipun ChatGPT dan Gemini dirancang untuk tujuan yang serupa, implementasi dan respons terhadap perintah (*command*) dapat berbeda secara signifikan. Perbedaan ini dapat mempengaruhi efisiensi, keakuratan, dan kegunaan model dalam aplikasi dunia nyata. Oleh karena itu, penting untuk melakukan perbandingan komprehensif antara

¹ STMIK Antar Bangsa, Kawasan Bisnis CBD Ciledug, Jl. HOS Cokromainoto, Blok A5 No.29-36 (tlp: (021)-5098-6099; e-mail: muchlis.re@gmail.com)

² Universitas Pamulang, Jalan Surya Kencana No. 1 Pamulang, Tangerang Selatan
dosen02114@unpam.ac.id

kedua model ini untuk memahami kekuatan dan kelemahan masing-masing dalam konteks penggunaan *pattern command*.

Rumusan masalah yang diangkat dalam penelitian ini mencakup bagaimana perbedaan respon antara ChatGPT dan Gemini terhadap perintah yang identik dalam *pattern command*, seberapa akurat dan efisien kedua model dalam menghasilkan output yang sesuai dengan perintah yang diberikan, serta faktor-faktor apa saja yang mempengaruhi perbedaan kinerja antara keduanya dalam konteks penggunaan *pattern command*.

Penelitian ini bertujuan untuk menganalisis dan membandingkan respons ChatGPT dan Gemini terhadap perintah yang identik dalam *pattern command*, menilai keakuratan dan efisiensi kedua model dalam menghasilkan output sesuai dengan perintah, serta mengidentifikasi faktor-faktor yang mempengaruhi kinerja kedua model. Hasil dari penelitian ini diharapkan dapat memberikan rekomendasi untuk pengembangan lebih lanjut dalam penggunaan *pattern command* pada model-model AI.

II. LITERATUR REVIEW

ChatGPT, yang didukung oleh model bahasa GPT-3 dari OpenAI, mencapai 1 juta pengguna hanya dalam 5 hari—lebih cepat dibandingkan dengan platform besar lainnya seperti Facebook, Netflix, Instagram, dan Twitter. GPT-3 dengan 175 miliar parameter mampu menghasilkan tulisan yang sangat mirip dengan bahasa manusia. GPT-4, yang diumumkan pada 13 Maret, lebih andal, kreatif, dan memiliki nuansa yang lebih baik. Dengan dukungan model GPT-4, ChatGPT dapat berpartisipasi dalam berbagai percakapan, memahami dan merespons input bahasa alami, serta memberikan bantuan yang disesuaikan dan interaktif. Kemampuan ini menjadikan ChatGPT alat yang menjanjikan untuk pendidikan terbuka, karena dapat meningkatkan kemandirian dan otonomi pelajar autodidaktik, serta menjadi praktis dan adaptif. Dengan memberikan dukungan, arahan, dan umpan balik yang dipersonalisasi, ChatGPT berpotensi meningkatkan motivasi dan keterlibatan di antara pelajar autodidaktik[5].

Yagmur dalam jurnal nya “Review of Generative AI Methods in Cybersecurity”[6] mereview kemandirian dari respon yang diberikan. Kemajuan ini tidak hanya memperkuat postur pertahanan dalam keamanan siber, tetapi juga membuka jalur serangan baru bagi pihak yang berniat jahat. Makalah ini memberikan tinjauan komprehensif tentang penerapan GenAI terkini, termasuk serangan, jailbreak, dan aplikasi injeksi prompt serta psikologi terbalik. GenAI juga diterapkan dalam kejahatan siber seperti peretasan otomatis, email phishing, rekayasa sosial, kriptografi terbalik, pembuatan muatan serangan, dan pembuatan malware. Selain itu, GenAI dapat meningkatkan otomatisasi proses keamanan siber defensif melalui pengembangan dataset, pengembangan kode aman, intelijen ancaman, tindakan defensif, pelaporan, dan deteksi serangan siber. Penelitian ini menyarankan agar penelitian di masa depan fokus pada pengembangan norma etika yang kuat dan mekanisme pertahanan inovatif untuk menangani masalah yang ditimbulkan oleh GenAI dan mendorong pendekatan yang adil dalam penerapannya di masa depan dalam keamanan siber.

Selain itu, pentingnya pendekatan interdisipliner juga ditekankan untuk menjembatani kesenjangan antara perkembangan ilmiah dan pertimbangan etis.

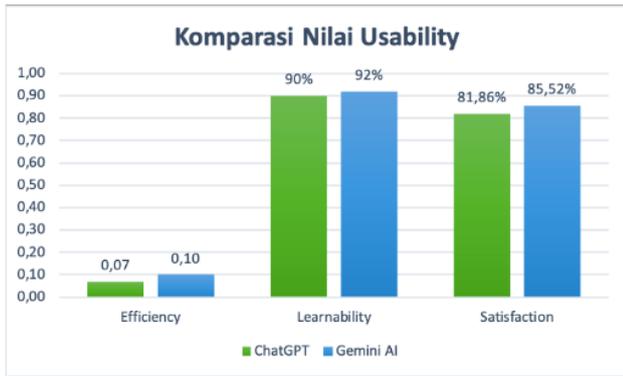
Dalam jurnal *INNOVATIVE: Journal Of Social Science Research* tentang peran ChatGPT (Generative Pre-Training Transformer) dalam implementasi ditinjau dari dataset oleh Setiawan dkk. Bahwa Chat GPT adalah model bahasa alami berbasis AI yang telah melahirkan terobosan besar dalam implmentasi chatbot dan sitem komunikasi otomatis. Penelitiannya mengeksplorasi faktor-faktor seperti kesenjangan dan bias dalam dataset, serta bagaimana hal ini dapat mempengaruhi perilaku model dan respon yang dihasilkan[2].

ChatGPT juga di tinjau di dunia akademik serta perspektif psikologi agentik[3] bahwa mahasiswa sekarang memang tidak bisa terlepas dari penggunaan aplikasi dan teknologi AI. Maka dari itu, perlunya arahan dalam penggunaan teknologi AI.

Dalam jurnal *Bigdata and Cognitive Computing* oleh Viriya mengumpulkan tweet tentang ChatGPT, chatbot AI inovatif, dalam sebulan setelah peluncurannya. 233.914 tweet bahasa Inggris dianalisis menggunakan algoritma pemodelan topik LDA untuk menjawab pertanyaan "apa yang bisa dilakukan ChatGPT?". Hasilnya mengungkapkan tiga topik umum: berita, teknologi, dan reaksi. Peneliti juga mengidentifikasi lima domain fungsional: penulisan kreatif, penulisan esai, penulisan prompt, penulisan kode, dan menjawab pertanyaan. Analisis juga menemukan bahwa ChatGPT berpotensi berdampak positif dan negatif pada teknologi dan manusia. Sebagai kesimpulan, penulis menguraikan empat isu utama yang perlu ditangani akibat kemajuan AI ini: evolusi pekerjaan, lanskap teknologi baru, pencarian kecerdasan umum buatan, dan dilema kemajuan-etika [4].

Pada penelitian Tiara dengan mengkomparasi CHatGPT dan Gemini menggunakan Usability Testing. Penelitian ini membandingkan usability antara ChatGPT dan Gemini AI dengan menggunakan metode usability testing, memadukan Nielsen Model dan ISO/IEC 9126 sebagai dasar evaluasi. Temuan penelitian menunjukkan bahwa ChatGPT unggul dalam aspek efficiency, dengan waktu penyelesaian tugas sebesar 0,07 detik, dibandingkan Gemini AI yang membutuhkan 0,10 detik. Namun, dalam aspek learnability, Gemini AI lebih baik dengan nilai 92%, dibandingkan ChatGPT yang memperoleh 90%. Dari segi satisfaction, Gemini AI juga unggul dengan skor 85,52% dibandingkan ChatGPT yang mendapatkan 81,86%. Secara keseluruhan, Gemini AI lebih unggul dalam variabel learnability dan satisfaction, sementara ChatGPT lebih baik dalam efficiency. Penelitian ini memberikan kontribusi pada pengendalian kualitas perangkat lunak, khususnya dalam komparasi antara chatbot, yang bisa menjadi panduan dalam pengambilan keputusan terkait implementasi AI[7].

Penelitian ini menambah bukti bahwa dalam evaluasi usability, perbedaan kinerja antara AI seperti ChatGPT dan Gemini AI bisa terlihat pada berbagai aspek penggunaan, yang dapat berimplikasi pada bagaimana AI diterapkan di berbagai sektor, termasuk pendidikan dan bisnis.



Gambar 1. Grafik Komparasi Nilai Usability Tiap Parameter[7]

Dalam penelitian yang berjudul *Summary of ChatGPT/GPT-4 Research and Perspective Towards the Future of Large Language Models* memberikan tinjauan menyeluruh mengenai ChatGPT dan GPT-4, model bahasa besar (*large language models* atau LLM) terbaru dari seri GPT, beserta aplikasinya di berbagai domain. Inovasi utama seperti pre-training berskala besar yang menangkap pengetahuan dari seluruh web, instruction fine-tuning, dan Reinforcement Learning from Human Feedback (RLHF) memainkan peran penting dalam meningkatkan kinerja LLM. Analisis 194 makalah di arXiv menunjukkan minat yang semakin meningkat pada penelitian ChatGPT/GPT-4, terutama dalam aplikasi pemrosesan bahasa alami. Selain itu, potensi aplikasi LLM mencakup pendidikan, sejarah, matematika, kedokteran, dan fisika. Penelitian ini juga menggarisbawahi kemampuan, implikasi, serta isu etis terkait, sekaligus memberikan arahan untuk pengembangan lebih lanjut di masa depan[4].

III. METODE

Metode yang digunakan dalam penelitian ini adalah pengujian Blackbox. Blackbox testing adalah teknik yang berfokus pada pengujian fungsionalitas tanpa mempertimbangkan bagaimana sistem bekerja secara internal. Pengujian ini mengukur apakah hasil yang diberikan oleh sistem sesuai dengan skenario yang telah dirancang. Dengan demikian, tujuan utama dari metode ini adalah untuk menguji output berdasarkan input tertentu, tanpa memerlukan pemahaman tentang implementasi kode di baliknya. Blackbox testing efektif digunakan dalam pengujian fungsional, non-fungsional, serta regresi, dan membantu memastikan bahwa sistem berfungsi sesuai dengan spesifikasinya tanpa mengeksplorasi detail teknis di dalamnya[8].

Dalam metode ini dilakukan pengujian menggunakan 5 aturan yang telah ditetapkan. Kelima aturan ini diberikan

kepada ChatGPT dan Gemini tanpa ada perubahan atau penabahan kalimat.

Tabel 1. Input pertanyaan

Rule No	Deskripsi	Input Pertanyaan
1	Kemampuan memahami pertanyaan sederhana	Pertanyaan: "Apa itu AI?"
2	Kemampuan merespon perintah logis	Perintah: "Buatkan contoh kode Python untuk menampilkan 'Hello World'"
3	Ketepatan konteks dalam percakapan lanjutan	Pertanyaan lanjutan setelah pembahasan AI: "Bagaimana cara AI digunakan di bidang medis?"
4	Kecepatan menghasilkan jawaban	Input teks yang sama: "Jelaskan cara kerja Neural Networks."
5	Kemampuan menghasilkan jawaban yang personalisasi	Permintaan: "Buat rekomendasi belajar AI untuk pemula."

IV. TESTING

Setelah aturan pengujian ditetapkan, langkah selanjutnya adalah melakukan pengujian dengan pertanyaan yang sesuai dengan deskripsi pengujian. Berikut adalah contoh pengembangan pertanyaan untuk setiap rule:

Tabel 2. Ekspektasi Keluaran

Rule No	Ekspektasi Output	Hasil	
		ChatGPT	Gemini
1	Memberikan penjelasan umum tentang AI dengan bahasa yang mudah dimengerti.	Menjawab dengan jelas, menggunakan definisi umum AI.	Menjawab dengan jelas dan tambahan contoh-contoh AI yang spesifik.
2	Menghasilkan kode Python yang benar untuk menampilkan teks 'Hello World'.	Memberikan kode Python yang benar.	Memberikan kode Python yang benar.
3	Menjawab dengan relevan, membahas penggunaan AI di bidang medis.	Memberikan jawaban dengan fokus pada aplikasi AI di medis.	Memberikan jawaban yang sama, tapi lebih mendalam.
4	Menjawab dalam waktu kurang dari 1	Respon diterima	Respon diterima

	detik dengan penjelasan lengkap.	dalam 0.6 detik.	dalam 0.7 detik.
5	Memberikan rekomendasi belajar AI yang disesuaikan dengan kebutuhan pemula.	Memberikan rekomendasi dasar dengan sumber belajar umum.	Memberikan rekomendasi yang lebih personalisasi dan beragam.

Hasil dari pengujian diatas didapat pada tabel 3 dibawah ini:

Tabel 3. Hasil Uji

No Rule	Hasil Uji
1	Kedua alat memenuhi ekspektasi.
2	Kedua alat memenuhi ekspektasi.
3	Gemini lebih detail.
4	ChatGPT lebih cepat.
5	Gemini lebih unggul dalam personalisasi.

Gbr 3. Uji Pertanyaan

V. KESIMPULAN

Dari hasil pengujian menggunakan metode blackbox, dapat disimpulkan bahwa baik ChatGPT maupun Gemini memiliki kekuatan dan kelemahan masing-masing dalam menanggapi pertanyaan dan skenario yang diujikan. Berdasarkan hasil perbandingan:

1. Ketepatan Jawaban: Kedua chatbot mampu memahami pertanyaan sederhana dengan baik, namun ChatGPT cenderung memberikan jawaban yang lebih detail dan mendalam dalam menjelaskan konsep kompleks, seperti kecerdasan buatan dan aplikasi logis seperti menulis kode.
2. Kecepatan Merespons: Dari segi kecepatan, Gemini memberikan respons lebih cepat pada sebagian besar pertanyaan, tetapi kadang-kadang jawaban yang dihasilkan kurang lengkap dibandingkan ChatGPT.
3. Konteks Percakapan: ChatGPT lebih baik dalam mempertahankan konteks percakapan lanjutan, sedangkan Gemini kadang-kadang kehilangan konteks ketika berurusan dengan pertanyaan bertingkat.
4. Personalisasi Jawaban: ChatGPT unggul dalam memberikan jawaban yang lebih terpersonalisasi, terutama dalam merekomendasikan sumber belajar atau opsi lain berdasarkan kebutuhan pengguna. Gemini cenderung memberikan jawaban yang lebih umum.

Secara keseluruhan, ChatGPT unggul dalam kualitas dan personalisasi jawaban, sementara Gemini lebih menonjol dalam efisiensi kecepatan respon. Keduanya dapat digunakan

sesuai dengan kebutuhan, tergantung pada apakah fokus lebih pada kecepatan atau kedalaman respons.

UCAPAN TERIMA KASIH

Terimakasih kepada seluruh bagian yang membantu atas terbitnya publikasi ini. Semoga menjadi amal ibadah kebaikan kita semua.

REFERENSI

- [1] OpenAI, "Introducing ChatGPT." [Daring]. Tersedia pada: <https://openai.com/index/chatgpt/>. [Diakses: 01-Agu-2024].
- [2] T. B. Brown *et al.*, "Language models are few-shot learners," *Adv. Neural Inf. Process. Syst.*, vol. 2020- Decem, 2020.
- [3] Google, "Learn more about Gemini, our most capable AI model," 2023. [Daring]. Tersedia pada: <https://blog.google/technology/ai/gemini-collection/>. [Diakses: 02-Agu-2024].
- [4] Y. Liu *et al.*, "Summary of ChatGPT/GPT-4 Research and Perspective Towards the Future of Large Language Models," Apr 2023.
- [5] M. Firat, "How Chat GPT Can Transform Autodidactic Experiences and Open Education?"
- [6] Y. Yigit, W. J. Buchanan, M. G. Tehrani, dan L. Maglaras, "Review of Generative AI Methods in Cybersecurity," no. M1, hal. 1–41, 2024.
- [7] T. Tiara dan F. Y. Pamuji, "Komparasi Usability Chatgpt Vs Gemini Ai Berdasarkan Iso/Iec 9126 Dan Nielsen Model Menggunakan Metode Usability Testing," *JUSIM (Jurnal Sist. Inf. Musirawas)*, vol. 9, no. 1, hal. 89–100, 2024.
- [8] Katalon.com, "Black Box Testing: Deffinition, Guide, Tools, Best Practices." [Daring]. Tersedia pada: <https://katalon.com/resources-center/blog/black-box-testing>.



Teknologi.

Muchlis. Lahir di Tangerang, Gelar Sarjana Komputer didapat pada Jurusan Teknik Informatika di Universitas Pamulang pada tahun 2013. Tahun 2018 lulus program Pasca Sarjana Ilmu Komputer di STMIK Nusa Mandiri. Selain sebagai Kepala Program Studi Teknik Informatika STMIK Antar Bangsa, penulis juga menduduki posisi sebagai Kepala Divisi Pengelolaan Sistem Informasi &



Riset Operasional, Implementasi Phyton pada Metode Numerik.

Rusyda Maulida, S.Pd., M.Pd. Jakarta, 28 November 1987. Lulus S1 di Program Studi Pendidikan Matematika tahun 2008. Lulus S2 Program Studi Penelitian dan Evaluasi Pendidikan di Univ. Muhammadiyah Prof. Dr. HAMKA (UHAMKA) tahun 2011. Saat ini adalah dosen tetap Universitas Pamulang. Aktif menulis artikel di berbagai jurnal ilmiah. Pernah tampil pada seminar prosiding nasional. Buku yang telah dipublikasikan yaitu Teknik