

ANALISIS SIMPANGAN BAKU WAKTU TANGGAPAN *CHATGPT* TERHADAP PERTANYAAN BERDASARKAN KATEGORI TOPIK

Hariz Kurniawan, Willy Azrieel, Cyaramani, Jadianan Parhusip

Teknik Informatika, Universitas Palangka Raya
Kampus Tunjung Nyaho, Jalan Yos Sudarso, Palangka Raya
willyazrieelbelajar@gmail.com

ABSTRAK

Perkembangan *artificial intelligence* dalam beberapa tahun terakhir telah membawa dampak besar dalam berbagai sektor, termasuk dalam aplikasi interaktif berupa asisten percakapan seperti *ChatGPT*. Salah satu faktor penting dalam mengevaluasi kinerja sistem percakapan ini adalah waktu tanggapan yang diperlukan untuk memberikan respons terhadap pertanyaan. Penelitian ini bertujuan untuk menganalisis variasi waktu tanggapan *ChatGPT* terhadap pertanyaan yang dikelompokkan dalam empat kategori utama: teknologi, seni & budaya, sains, dan sosial. Waktu tanggapan yang tercatat kemudian dianalisis menggunakan dua metrik utama: rata-rata waktu tanggapan dan simpangan baku, yang menjadi ukuran untuk menilai konsistensi respons sistem. Hasil penelitian menunjukkan bahwa waktu tanggapan *ChatGPT* bervariasi secara signifikan antar kategori. Kategori teknologi mencatatkan waktu tanggapan tercepat (rata-rata 5.33 detik), namun dengan variasi yang lebih besar (simpangan baku 0.55 detik). Sebaliknya, kategori seni & budaya menunjukkan waktu tanggapan yang lebih lama (rata-rata 8.1 detik), tetapi lebih konsisten (simpangan baku 0.31 detik). Kategori sains mencatatkan waktu tanggapan yang stabil (rata-rata 6.4 detik), sementara kategori sosial berada di tengah, dengan waktu tanggapan yang relatif cepat dan konsisten (rata-rata 7.3 detik, simpangan baku 0.21 detik). Temuan ini menunjukkan bahwa simpangan baku dapat digunakan sebagai alat yang efektif untuk menilai konsistensi kinerja *ChatGPT*. Berdasarkan hasil analisis, disarankan agar pengguna memilih kategori topik yang sesuai dengan kebutuhan mereka dan memperhatikan variasi waktu tanggapan, khususnya dalam kategori teknologi.

Kata Kunci : waktu tanggapan, *ChatGPT*, simpangan baku, teknologi, konsistensi, machine learning

1. PENDAHULUAN

Dalam beberapa tahun terakhir, *artificial intelligence* telah mengalami perkembangan yang signifikan dan diterapkan dalam berbagai bidang, mulai dari kesehatan hingga pendidikan, termasuk dalam interaksi percakapan melalui asisten virtual. Salah satu teknologi berbau *artificial intelligence* yang semakin populer adalah *ChatGPT*, yakni sebuah model bahasa *machine learning* yang dikembangkan oleh *OpenAI*. *ChatGPT* sendiri dirancang untuk memberikan respons otomatis terhadap pertanyaan dari berbagai topik, dibekali kemampuan untuk memahami dan menghasilkan teks dalam bahasa yang natural. Meskipun *ChatGPT* telah menunjukkan kemajuan yang pesat dalam kualitas tanggapan, waktu yang dibutuhkan untuk memberikan respons (waktu tanggapan) merupakan faktor yang penting dalam menilai kinerja sistem ini dalam berbagai konteks penggunaan [1].

Waktu tanggapan seringkali menjadi metrik indikasi yang sering digunakan untuk mengevaluasi kinerja suatu sistem. Dalam konteks ini, waktu tanggapan merujuk pada waktu yang dibutuhkan oleh *ChatGPT* untuk memberikan jawaban terhadap pertanyaan yang diberikan. Variasi waktu tanggapan ini dapat dipengaruhi oleh beberapa faktor, salah satunya adalah kategori topik dari pertanyaan yang diajukan. Pertanyaan yang bersifat lebih kompleks dan memerlukan pemrosesan data lebih mendalam, seperti pertanyaan terkait matematika atau sains, mungkin membutuhkan waktu lebih lama dibandingkan dengan

pertanyaan yang lebih sederhana, seperti pertanyaan mengenai seni atau budaya [2] [3].

Salah satu metode statistik yang dapat digunakan untuk melakukan evaluasi sebaran data adalah dengan simpangan baku. Simpangan baku sendiri akan menggambarkan sejauh mana data tersebar di sekitar nilai rerata. Pengukuran simpangan baku dalam waktu tanggapan dapat memberikan wawasan tentang konsistensi respons *ChatGPT* terhadap berbagai jenis pertanyaan. Dalam penelitian ini, kami akan melakukan analisis simpangan baku terhadap waktu tanggapan *ChatGPT* berdasarkan pertanyaan yang dikategorikan dalam beberapa topik, seperti teknologi, seni & budaya, sains, dan sosial.

Penelitian ini bertujuan untuk menjawab dua pertanyaan utama; pertama, apakah terdapat perbedaan yang signifikan dalam waktu tanggapan *ChatGPT* berdasarkan kategori topik. Kedua, bagaimana simpangan baku dapat menggambarkan konsistensi waktu tanggapan *ChatGPT* dalam merespons pertanyaan yang berbeda topiknya. Melalui analisis ini, diharapkan penelitian ini dapat memberikan wawasan yang lebih jelas mengenai efisiensi, konsistensi, dan kinerja *ChatGPT* dalam berbagai konteks topik.

2. TINJAUAN PUSTAKA

2.1. Perkembangan Artificial Intelligence

Artificial intelligence telah berkembang pesat dalam beberapa dekade terakhir, dengan banyak implementasi yang mencakup berbagai sektor, salah

satunya adalah *natural language processing*. Salah satu aplikasi paling dikenal dalam penerapan *natural language processing* ialah model bahasa besar seperti *ChatGPT*, yang dikembangkan oleh *OpenAI*. *ChatGPT* sendiri menggunakan arsitektur model *generative pretrained transformer* untuk memproses dan menghasilkan teks yang menyerupai percakapan manusia sungguhan. Model ini dilatih dengan jumlah data yang sangat besar untuk meningkatkan kemampuannya dalam memahami konteks dan menghasilkan respons yang relevan terhadap pertanyaan dari berbagai topik [4].

ChatGPT berfungsi untuk memberikan jawaban secara otomatis, dan meskipun ia mampu menunjukkan kemampuan luar biasa dalam hal kualitas tanggapan, kecepatan tanggapannya dapat bervariasi tergantung pada kompleksitas pertanyaan yang diberikan. Waktu yang dibutuhkan untuk memberikan respons adalah faktor krusial dalam menilai kinerja sistem *artificial intelligence* ini dalam konteks pengguna, karena semakin cepat waktu tanggapannya, semakin efisien interaksi pengguna dengan sistem tersebut [5]. Pada saat yang sama, kualitas respons juga tidak boleh dikorbankan demi kecepatan, menciptakan tantangan yang perlu dievaluasi lebih dalam [6].

2.2. Evaluasi Kinerja Sistem ChatGPT

Waktu tanggapan adalah salah satu acuan utama yang digunakan untuk menilai kinerja suatu sistem. Waktu tanggapan di sini mengacu pada periode antara pengiriman pertanyaan oleh pengguna dan penerimaan respons dari sistem. Kinerja sistem *ChatGPT* sangat bergantung pada waktu pemrosesannya, karena semakin lama waktu yang dibutuhkan dapat menurunkan pengalaman pengguna, sementara waktu yang terlalu cepat dapat menunjukkan bahwa sistem belum memproses informasi dengan cukup baik [6][3].

Berdasarkan penelitian sebelumnya, faktor yang mempengaruhi waktu tanggapan meliputi panjang dan kompleksitas pertanyaan, serta kategori topik yang dibahas. Pertanyaan yang lebih teknis dan kompleks, misalnya terkait teknologi atau sains, cenderung membutuhkan lebih banyak waktu untuk diproses oleh sistem, sedangkan pertanyaan dengan topik yang lebih umum, seperti seni atau budaya mungkin bisa dijawab lebih cepat [1]. Oleh karena itu, penting untuk melakukan evaluasi variasi waktu tanggapan berdasarkan kategori topik untuk mendapatkan yang lebih jelas mengenai kinerja dan efisiensi sistem *ChatGPT* ini [7][8].

Simpangan baku adalah metode statistika yang dapat digunakan untuk mengukur variasi dalam waktu tanggapan. Pada konteks ini, simpangan baku dapat digunakan untuk menunjukkan seberapa konsisten sistem dapat memberikan respons terhadap pertanyaan yang diajukan. Semakin kecil simpangan baku, semakin konsisten waktu tanggapannya; sedangkan jika semakin besar, bisa jadi disebabkan oleh faktor seperti kesulitan dalam menelaah pertanyaan atau

perbedaan dalam pengolahan informasi yang dibutuhkan [3].

2.3. Pengaruh Kategori Topik Terhadap Waktu Tanggapan

Salah satu temuan penting dalam penelitian mengenai sistem percakapan berbasis *artificial intelligence* adalah bahwa waktu tanggapan sering kali dipengaruhi oleh kategori topik dari pertanyaan yang diajukan. Berdasarkan penelitian yang dilakukan oleh Mikalef, kategori topik yang lebih kompleks, seperti matematika, teknologi, atau sains, memerlukan pemrosesan data yang lebih rumit, karena pertanyaan-pertanyaan ini sering kali melibatkan konsep-konsep yang lebih teknis dan membutuhkan pemahaman yang mendalam dari sistem. Di sisi lain, kategori topik yang lebih sederhana, seperti seni, agama, atau budaya mungkin membutuhkan waktu yang lebih sedikit, karena tidak melibatkan analisis mendalam terhadap data atau konsep-konsep yang lebih teknis [9].

Studi yang pernah dilakukan oleh Joksimovic mendalami bahwa dalam interaksi dengan sistem berbasis *artificial intelligence* seperti *ChatGPT*, waktu tanggapan dapat bervariasi berdasarkan tingkat kesulitan pertanyaan. Misalnya, ketika pertanyaan berkaitan dengan hal-hal yang lebih spesifik atau teknis, seperti “Jelaskan cara kerja matematika logika dalam komputer”, sistem akan membutuhkan lebih banyak waktu untuk memproses dan memberikan jawaban yang akurat. Sebaliknya, pertanyaan yang lebih umum seperti “Siapa Mozart itu?” dapat dijawab dengan lebih cepat, karena jawabannya tidak memerlukan pengetahuan teknis yang mendalam [10].

2.4. Pentingnya Waktu Tanggapan Dalam Interaksi

Waktu tanggapan tidak hanya berfungsi sebagai ukuran efisiensi secara teknis, tetapi juga sangat penting dalam menentukan kualitas interaksi antara pengguna dan sistem. Penelitian sosial yang dilakukan Misischia dan Mariani mengungkapkan bahwa pengguna cenderung lebih puas dengan sistem percakapan yang mampu memberikan respons dengan cepat, meskipun kualitas jawaban tersebut mungkin tidak selalu sangat mendalam (akurat) [6][11].

Selain itu, dengan menganalisis waktu tanggapan dalam konteks berbagai kategori topik, kita dapat memperoleh pemahaman yang lebih dalam mengenai bagaimana sistem seperti *ChatGPT* bekerja. Dengan hal tersebut tidak hanya membantu dalam memahami efisiensi sistem, tetapi juga dalam mengetahui di area mana sistem perlu diperbaiki untuk memberikan kinerja yang lebih dalam merespons pertanyaan dengan berbagai tingkat kesulitan [7].

3. METODE PENELITIAN

Penelitian ini bertujuan untuk menganalisis waktu tanggapan *ChatGPT* terhadap pertanyaan yang dikelompokkan berdasarkan kategori topik, serta untuk menghitung simpangan baku sebagai ukuran konsistensi dari waktu tanggapan tersebut.

Pendekatan yang digunakan adalah penelitian deskriptif kuantitatif dengan analisis statistik sederhana, yang berfokus pada pengumpulan dan pemrosesan data waktu tanggapan dari *ChatGPT* dari pertanyaan-pertanyaan yang diajukan.

3.1. Desain Penelitian

Penelitian ini mengadopsi desain deskripsi kuantitatif untuk menggambarkan variasi waktu tanggapan *ChatGPT* berdasarkan kategori topik, yaitu teknologi, seni & budaya, sains, dan sosial. Data dikumpulkan melalui interaksi langsung dengan *ChatGPT*, di mana pertanyaan-pertanyaan yang telah disiapkan diajukan sesuai dengan kategori yang telah ditentukan. Waktu tanggapan yang tercatat akan diukur dalam skala detik, dan untuk setiap kategori, analisis statistik akan dilakukan untuk memperoleh rerata simpangan baku. Desain penelitian ini bertujuan untuk memberikan gambaran yang jelas mengenai variasi waktu tanggapan yang terjadi berdasarkan kategori topik yang berbeda dan untuk menilai konsistensi *ChatGPT* dalam merespons pertanyaan.

3.2. Pengumpulan Data

3.2.1. Pemilihan Kategori Topik dan Pertanyaan

Peneliti akan memilih empat kategori utama untuk topik yang akan dianalisis, berupa teknologi, seni & budaya, sains, dan sosial. Setiap kategori tadi nantinya terdiri dari 10 pertanyaan yang disusun dengan tingkat kesulitan yang bervariasi. Pertanyaan-pertanyaan ini bertujuan untuk mencakup berbagai aspek dalam kategori masing-masing, dengan tujuan untuk menilai variasi waktu tanggapan berdasarkan kesulitan dan konteks topik.

3.2.2. Interaksi Dengan *ChatGPT*

Pertanyaan yang telah dipilih kemudian diajukan satu per-satu kepada *ChatGPT*. Setiap kali pertanyaan dikirim, waktu tanggapan akan dicatat mulai dari saat pertanyaan diajukan hingga jawaban lengkap diberikan oleh sistem. Waktu ini diukur dalam satuan detik untuk menjaga konsistensi dalam pengukuran. Proses pencatatan akan dilakukan secara manual, menggunakan fitur *stopwatch* atau *timer* yang tersedia di perangkat komputer atau ponsel.

3.2.3. Pengumpulan Data Waktu Tanggapan

Setelah waktu tanggapan tadi dicatat, data waktu tersebut digunakan untuk analisis statistik. Waktu tanggapan untuk setiap kategori akan dihitung reratanya dan diuji dengan menggunakan perhitungan simpangan baku untuk mengetahui konsistensi dari respons *ChatGPT* terhadap berbagai kategori topik

3.3. Instrumen Penelitian

3.3.1. Daftar Pertanyaan

Pertanyaan yang disusun untuk setiap kategori topik disiapkan dengan tujuan untuk mencakup berbagai tingkat kesulitan dan relevansi dengan topik

tersebut. Daftar ini digunakan untuk mengajukan pertanyaan yang konsisten kepada *ChatGPT*.

3.3.2. Alat Pengukur Waktu

Untuk mengukur waktu tanggapan, digunakan alat berupa *stopwatch* atau fitur *timer* pada perangkat yang digunakan, baik itu komputer atau ponsel, untuk memastikan pengukuran waktu yang akurat.

3.3.3. Pencatatan Data

Data yang terkumpul akan dicatat dalam media seperti *sheet* (misalnya *Google Sheet* atau *Microsoft Excel*) untuk memudahkan proses analisis. Catatan ini akan digunakan untuk menghitung rerata waktu tanggapan dan simpangan baku, serta untuk memvisualisasikan data dalam bentuk tabel atau gambar.

3.4. Teknik Analisis Data

Setelah data yang dibutuhkan telah terkumpul, akan dilakukan analisis dengan menggunakan dua langkah utama sebagai berikut:

3.4.1. Perhitungan Rerata Waktu Tanggapan

Rata-rata dari waktu tanggapan untuk setiap kategori topik akan dihitung dengan menggunakan rumus berikut:

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N} \quad (1)$$

Yang mana x_i merupakan waktu tanggapan untuk setiap pertanyaan ke- i . Kemudian N adalah jumlah total pertanyaan dalam kategori tersebut.

3.4.2. Perhitungan Simpangan Baku

Simpangan baku akan dihitung untuk mengukur variasi waktu tanggapan pada masing-masing kategori dengan menggunakan rumus berikut:

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}} \quad (2)$$

Yang mana x_i merupakan waktu tanggapan untuk setiap pertanyaan ke- i . Sedangkan, $\bar{x}$ merupakan rerata waktu tanggapan. Kemudian N adalah jumlah total pertanyaan dalam kategori tersebut.

4. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk menganalisis waktu tanggapan *ChatGPT* terhadap pertanyaan yang dikategorikan dalam empat topik utama, yakni teknologi, seni & budaya, sains, dan sosial. Setiap kategori terdiri dari 10 pertanyaan yang disusun dengan variasi tingkat kesulitan. Waktu tanggapan yang dicatat untuk setiap pertanyaan diukur dalam satuan detik menggunakan fitur *stopwatch* atau *timer* yang tersedia pada perangkat yang digunakan. Data yang dikumpulkan kemudian dianalisis untuk

menghitung rata-rata waktu tanggapan dan simpangan baku untuk mengukur konsistensi respons.

4.1. Hasil Penelitian

4.1.1. Kategori Teknologi

Tabel 1. Data Kategori Teknologi

Pertanyaan	Waktu (Detik)
Apa itu <i>blockchain</i> ?	5.4
Jelaskan konsep <i>AI</i> !	4.8
Apa konsep dari <i>cloud computing</i> ?	5.2
Jelaskan tentang <i>internet of things</i> !	6.1
Apa perbedaan antara VR dengan AR?	5.6
Apa itu <i>5G</i> ?	4.9
Apa manfaat dari <i>big data</i> ?	6.2
Jelaskan cara kerja mesin pencarian!	5.0
Apa itu <i>cybersecurity</i> ?	5.5
Apa itu <i>DevOps</i> ?	5.3

Berdasarkan yang ditampilkan pada Tabel 1, dapat diketahui bahwa data waktu tanggapan terhadap kategori teknologi terdapat 5.4, 4.8, 5.2, 6.1, 5.6, 4.9, 6.2, 5.0, 5.5, 5.3. Dari data waktu tersebut, dapat dilakukan perhitungan rerata waktu tanggapan dengan rumus yang sudah dijabarkan pada (1) dan simpangan baku pada rumus (2).

Rerata Waktu Tanggapan:

$$\bar{x} = \frac{5.4 + 4.8 + \dots + 5.3}{10} = 5.33 \text{ detik}$$

Simpangan Baku:

$$\sigma = \sqrt{\frac{(5.4 - 5.33)^2 + \dots + (5.3 - 5.33)^2}{10}} = 0.55 \text{ detik}$$

4.1.2. Kategori Seni & Budaya

Tabel 2. Data Kategori Seni & Budaya

Pertanyaan	Waktu (Detik)
Apa itu seni lukis?	8.4
Apa yang dimaksud dengan seni patung?	7.5
Jelaskan tentang seni musik klasik!	8.2
Siapa pelukis terkenal dari Indonesia?	7.9
Apa itu tari tradisional?	8.3
Apa pengertian seni rupa secara harfiah?	7.8
Jelaskan tentang film dokumenter!	8.1
Apa perbedaan antara seni dan budaya?	7.7
Jelaskan sejarah seni barok!	8.0
Apa itu arsitektur modern?	8.5

Berdasarkan yang ditampilkan pada Tabel 2, dapat diketahui bahwa data waktu tanggapan terhadap kategori seni & budaya terdapat 8.4, 7.5, 8.2, 7.9, 8.3, 7.8, 8.1, 7.7, 8.0, 8.5. Dari data waktu tersebut, dapat dilakukan perhitungan rerata waktu tanggapan dengan rumus yang sudah dijabarkan pada (1) dan simpangan baku pada rumus (2).

Rerata Waktu Tanggapan:

$$\bar{x} = \frac{8.4 + 7.5 + \dots + 8.5}{10} = 8.1 \text{ detik}$$

Simpangan Baku:

$$\sigma = \sqrt{\frac{(8.4 - 8.1)^2 + \dots + (8.5 - 8.1)^2}{10}} = 0.31 \text{ detik}$$

4.1.3. Kategori Sains

Tabel 3. Data Kategori Sains

Pertanyaan	Waktu (Detik)
Apa itu hukum Newton?	6.2
Jelaskan tentang teori relativitas!	6.5
Apa yang dimaksud dengan fotosintesis?	6.3
Apa itu konsep gaya gravitasi?	6.4
Jelaskan perbedaan antara atom dan molekul!	6.1
Apa itu <i>DNA</i> ?	6.6
Apa itu teori evolusi?	6.7
Jelaskan proses mitosis!	6.5
Apa itu hukum termodinamika?	6.3
Apa yang dimaksud dengan sel?	6.2

Berdasarkan yang ditampilkan pada Tabel 3, dapat diketahui bahwa data waktu tanggapan terhadap kategori sains terdapat 6.2, 6.5, 6.3, 6.4, 6.1, 6.6, 6.7, 6.5, 6.3, 6.2. Dari data waktu tersebut, dapat dilakukan perhitungan rerata waktu tanggapan dengan rumus yang sudah dijabarkan pada (1) dan simpangan baku pada rumus (2).

Rerata Waktu Tanggapan:

$$\bar{x} = \frac{6.2 + 6.5 + \dots + 6.2}{10} = 6.4 \text{ detik}$$

Simpangan Baku:

$$\sigma = \sqrt{\frac{(6.2 - 6.4)^2 + \dots + (6.2 - 6.4)^2}{10}} = 0.19 \text{ detik}$$

3.2.4. Kategori Sosial

Tabel 4. Data Kategori Sosial

Pertanyaan	Waktu (Detik)
Apa itu demokrasi?	7.0
Jelaskan tentang sistem politik!	7.3
Apa itu hak asasi manusia?	7.1
Apa yang dimaksud dengan globalisasi?	7.2
Jelaskan tentang ketimpangan sosial!	7.4
Apa itu ekonomi pasar bebas?	7.0

Pertanyaan	Waktu (Detik)
Apa itu UNICEF?	7.5
Jelaskan perbedaan kapitalisme dan sosialisme!	7.3
Apa itu teori konflik?	7.6
Apa yang dimaksud dengan ideologi?	7.4

Berdasarkan yang ditampilkan pada Tabel 4, dapat diketahui bahwa data waktu tanggapan terhadap kategori sosial terdapat 7.0, 7.3, 7.1, 7.2, 7.4, 7.0, 7.5, 7.3, 7.6, 7.4. Dari data waktu tersebut, dapat dilakukan perhitungan rerata waktu tanggapan dengan rumus yang sudah dijabarkan pada (1) dan simpangan baku pada rumus (2).

$$\text{Rerata Waktu Tanggapan:} \\ \bar{x} = \frac{7.0 + 7.3 + \dots + 7.4}{10} = 7.3 \text{ detik}$$

$$\text{Simpangan Baku:} \\ \sigma = \sqrt{\frac{(7.0 - 7.3)^2 + \dots + (7.4 - 7.3)^2}{10}} \\ = 0.21 \text{ detik}$$

Hasil menunjukkan bahwa terdapat variasi yang signifikan dalam waktu tanggapan *ChatGPT* antar kategori topik. Pada kategori teknologi memiliki waktu tanggapan rata-rata yang lebih cepat (5.53 detik) dibandingkan dengan kategori lainnya, seperti seni & budaya (8.1 detik) dan sosial (7.3 detik). Kategori sains berada di posisi tengah yakni dengan waktu tanggapan rata-rata di 6.4 detik.

4.2. Pembahasan

4.2.1. Variasi Waktu Tanggapan Antar Teknologi

Berdasarkan hasil yang diperoleh, kategori teknologi menunjukkan waktu tanggapan yang lebih cepat dengan simpangan baku yang lebih besar, yaitu 0.55 detik. Meskipun memiliki rerata waktu tanggapan yang cepat, simpangan baku yang lebih besar menandakan adanya variasi yang lebih tinggi dalam respons terhadap pertanyaan yang diajukan dalam kategori ini. Hal ini kemungkinan besar disebabkan oleh keragaman jenis pertanyaan dalam kategori teknologi, yang mana melibatkan konsep-konsep teknis yang beragam, seperti *artificial intelligence*, *blockchain*, dan *cybersecurity*. Setiap konsep ini memerlukan tingkat pemrosesan data yang berbeda-beda, yang dapat mempengaruhi waktu tanggapan. Penelitian oleh Raj mengungkapkan bahwa tingkat kompleksitas pertanyaan berperan besar dalam menentukan waktu tanggapan pada sistem, terutama untuk topik yang membutuhkan pemahaman mendalam terhadap konsep teknis [12].

4.2.2. Konsistensi Waktu Tanggapan Pada Kategori Seni & Budaya dan Sains

Pada kategori seni & budaya menunjukkan waktu tanggapan yang lebih lambat, dengan rata-rata

waktu 8.1 detik, namun dengan simpangan baku yang lebih kecil, yakni 0.31 detik. Hal tersebut menunjukkan bahwa meskipun respons terhadap pertanyaan dalam kategori ini memakan waktu yang lebih lama, konsistensi waktu tanggapan *ChatGPT* relatif tinggi. Hal ini dapat dijelaskan oleh kenyataan bahwa pertanyaan dalam kategori seni & budaya cenderung lebih bersifat deskriptif dan berfokus pada pengetahuan umum, seperti definisi seni lukis atau seni patung. Berbeda dengan kategori teknologi yang memiliki pertanyaan dengan berbagai tingkat kesulitan, pertanyaan dalam seni dan budaya lebih terstruktur dan tidak memerlukan pemrosesan data yang rumit. Untuk kategori sains, meskipun tingkat kesulitan pertanyaannya lebih tinggi dibandingkan dengan seni & budaya, waktu tanggapan rata-rata berada di posisi tengah dengan 6.4 detik dengan simpangan baku terendah pada 0.19 detik. Data ini menunjukkan bahwa *ChatGPT* mampu memberikan respons yang lebih konsisten meskipun untuk topik yang lebih kompleks, seperti teori fisika atau biologi. Zichen juga mendukung temuan ini dengan menyatakan bahwa sistem *artificial intelligence* dapat mengolah pertanyaan yang lebih teknis dengan lebih efisien apabila pertanyaan tersebut telah diproses atau terstruktur dengan baik jika pertanyaan sudah diproses sebelumnya atau memiliki struktur yang lebih baku [13].

4.2.3. Perbandingan Kategori Sosial

Kategori sosial dengan rerata waktu tanggapan 7.3 detik dan simpangan baku 0.21 detik, berada di antara seni & budaya dan sains dalam hal kecepatan respons. Meskipun topik-topik sosial melibatkan pertanyaan yang lebih kompleks dan dapat lebih terbuka dalam interpretasi, data menunjukkan bahwa *ChatGPT* masih mampu memberikan respons yang cukup cepat dan konsisten. Namun, tingkat kecepatan ini masih lebih lambat dibandingkan dengan kategori teknologi, yang menunjukkan bahwa waktu tanggapan bisa sangat dipengaruhi oleh tingkat spesifikasi dan kejelasan pertanyaan.

4.2.4. Simpangan Baku Sebagai Indikasi Metriks Konsistensi

Simpangan baku yang lebih rendah pada kategori sains dan sosial menunjukkan bahwa *ChatGPT* memberikan respons yang lebih konsisten dalam kategori ini. Sebaliknya, simpangan baku yang lebih tinggi pada kategori teknologi menunjukkan variasi yang lebih besar dalam waktu tanggapan. Hal ini bisa mencerminkan kompleksitas dalam memahami berbagai jenis pertanyaan di bidang teknologi yang lebih beragam dan memerlukan analisis lebih mendalam [14]. Sebagaimana yang dijelaskan pada studi yang dilakukan oleh Brown, simpangan baku adalah ukuran yang sangat berguna untuk mengevaluasi konsistensi dalam sistem berbasis *machine learning*, karena memberikan informasi

tentang stabilitas kinerja sistem terhadap variasi *input* [15].

5. KESIMPULAN DAN SARAN

Penelitian ini bertujuan untuk menganalisis waktu tanggapan *ChatGPT* terhadap pertanyaan yang dikategorikan dalam empat topik utama, yaitu teknologi, seni & budaya, sains, dan sosial. Hasil penelitian menunjukkan adanya variasi yang signifikan dalam waktu tanggapan antar kategori. Kategori teknologi memiliki waktu tanggapan tercepat dengan rata-rata 5.33 detik, namun dengan simpangan baku yang lebih tinggi (0.55 detik), menunjukkan adanya variasi yang lebih besar dalam respons terhadap pertanyaan teknis yang lebih kompleks. Sementara itu, kategori seni & budaya menunjukkan waktu tanggapan yang lebih lambat (8.1 detik), tetapi lebih konsisten dengan simpangan baku yang lebih rendah (0.31 detik), mencerminkan stabilitas respons terhadap pertanyaan yang lebih deskriptif dan berbasis pengetahuan umum. Kategori sains menunjukkan waktu tanggapan yang stabil dengan rata-rata 6.4 detik dan simpangan baku terendah (0.19 detik), mengindikasikan konsistensi meskipun kompleksitas pertanyaannya lebih tinggi. Kategori sosial berada di posisi tengah, dengan rata-rata waktu tanggapan 7.3 detik dan simpangan baku 0.21 detik, menunjukkan respons yang cukup cepat dan konsisten. Secara keseluruhan, simpangan baku terbukti menjadi indikator yang efektif dalam mengukur konsistensi kinerja *ChatGPT* di berbagai kategori topik.

Berdasarkan temuan ini, disarankan bagi pengguna yang membutuhkan respons cepat dan konsisten untuk memilih kategori seni & budaya atau sosial, di mana waktu tanggapan relatif lebih cepat dan konsisten. Untuk pertanyaan yang lebih kompleks, terutama terkait teknologi dan sains, pengguna perlu mengantisipasi waktu tanggapan yang lebih lama dan lebih bervariasi. Untuk meningkatkan kinerja *ChatGPT*, pengoptimalan waktu tanggapan dalam kategori teknologi sangat disarankan untuk meminimalkan variabilitas dan meningkatkan konsistensi. Di sisi lain, meskipun waktu tanggapan pada kategori seni & budaya lebih lama, konsistensi yang tinggi dalam kategori ini harus tetap menjadi prioritas utama. Pemantauan berkelanjutan terhadap simpangan baku juga penting untuk menjaga stabilitas dan keandalan sistem. Penelitian lebih lanjut dapat memperluas analisis dengan menambahkan lebih banyak kategori topik atau melakukan perbandingan antara *ChatGPT* dan sistem percakapan berbasis *artificial intelligence* lainnya untuk mendapatkan gambaran yang lebih komprehensif tentang kinerja dan efisiensi sistem.

DAFTAR PUSTAKA

- [1] E. Adamopoulou and L. Moussiades, "Chatbots: History, technology, and applications," *Mach. Learn. with Appl.*, vol. 2, p. 100006, Dec. 2020, doi: 10.1016/j.mlwa.2020.100006.
- [2] T. Raz, S. Luchini, R. Beaty, and Y. Kenett, "Automated Scoring of Open-Ended Question Complexity: A Large Language Model Approach." Feb. 19, 2024. doi: 10.21203/rs.3.rs-3890828/v1.
- [3] H. Tian *et al.*, "Is ChatGPT the Ultimate Programming Assistant -- How far is it?," Apr. 2023, [Online]. Available: <http://arxiv.org/abs/2304.11938>
- [4] J. Zhou, P. Ke, X. Qiu, M. Huang, and J. Zhang, "ChatGPT: potential, prospects, and limitations," *Frontiers of Information Technology and Electronic Engineering*, vol. 25, no. 1. Zhejiang University, pp. 6–11, Jan. 01, 2024. doi: 10.1631/FITEE.2300089.
- [5] M. Thelwall, "ChatGPT for complex text evaluation tasks," *J. Assoc. Inf. Sci. Technol.*, 2024, doi: 10.1002/asi.24966.
- [6] C. V. Misischia, F. Poecze, and C. Strauss, "Chatbots in customer service: Their relevance and impact on service quality," in *Procedia Computer Science*, Elsevier B.V., 2022, pp. 421–428. doi: 10.1016/j.procs.2022.03.055.
- [7] S. Fergus, M. Botha, and M. Ostovar, "Evaluating Academic Answers Generated Using ChatGPT," *J. Chem. Educ.*, vol. 100, no. 4, pp. 1672–1675, Apr. 2023, doi: 10.1021/acs.jchemed.3c00087.
- [8] P. P. Ray, "ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope," *Internet of Things and Cyber-Physical Systems*, vol. 3. KeAi Communications Co., pp. 121–154, Jan. 01, 2023. doi: 10.1016/j.iotcps.2023.04.003.
- [9] P. Mikalef and M. Gupta, "Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance," *Inf. Manag.*, vol. 58, no. 3, Apr. 2021, doi: 10.1016/j.im.2021.103434.
- [10] S. Joksimovic, D. Ifenthaler, R. Marrone, M. De Laat, and G. Siemens, "Opportunities of artificial intelligence for supporting complex problem-solving: Findings from a scoping review," *Computers and Education: Artificial Intelligence*, vol. 4. Elsevier B.V., Jan. 01, 2023. doi: 10.1016/j.caeai.2023.100138.
- [11] M. M. Mariani, I. Machado, V. Magrelli, and Y. K. Dwivedi, "Artificial intelligence in innovation research: A systematic review, conceptual framework, and future research directions," *Technovation*, vol. 122, Apr. 2023, doi: 10.1016/j.technovation.2022.102623.
- [12] R. Raj, A. Singh, V. Kumar, and P. Verma, "Analyzing the potential benefits and use cases of ChatGPT as a tool for improving the efficiency and effectiveness of business operations," *BenchCouncil Trans. Benchmarks, Stand. Eval.*, vol. 3, no. 3, Sep. 2023, doi: 10.1016/j.tbench.2023.100140.

- [13] Z. Ye *et al.*, “An assessment of ChatGPT’s responses to frequently asked questions about cervical and breast cancer,” *BMC Womens. Health*, vol. 24, no. 1, Dec. 2024, doi: 10.1186/s12905-024-03320-8.
- [14] X. Shen, Z. Chen, M. Backes, and Y. Zhang, “In ChatGPT We Trust? Measuring and Characterizing the Reliability of ChatGPT,” Apr. 2023, [Online]. Available: <http://arxiv.org/abs/2304.08979>
- [15] T. B. Brown *et al.*, “Language Models are Few-Shot Learners,” May 2020, [Online]. Available: <http://arxiv.org/abs/2005.14165>