

Analisis Perbandingan Kemampuan Pemahaman Instruksi Kompleks pada Model Gemini Thinking dan Copilot Think Deeper Menggunakan Metode Benchmark Testing

Muhammad Arsy Pradana *¹

Yoseph Felix Marchello ²

Yoga Dwi Andhika ³

Dicky Pratama ⁴

^{1,2,3,4} Program Studi Sistem Informasi, Fakultas Ilmu Komputer dan Rekayasa, Universitas Multi Data Palembang, Indonesia

*e-mail: muhammadarsyapradana_2327240057@mhs.mdp.ac.id¹,
yosephfelixmarchello_2327240065@mhs.mdp.ac.id², yogadwiandhika_2327240092@mhs.mdp.ac.id³,
dqpratama@mdp.ac.id⁴

Abstrak

Penelitian ini menganalisis perbandingan kemampuan pemahaman instruksi kompleks antara model Gemini Thinking dan Copilot Think Deeper menggunakan metode Benchmark Testing. Pengujian dilakukan melalui lima dimensi evaluasi pemahaman konseptual, penalaran logis dan teknis, konsistensi konteks, efisiensi performa, dan adaptasi personalisasi. Hasil penelitian menunjukkan bahwa Gemini Thinking secara konsisten memberikan respons yang lebih mendalam dan teknis dengan penanganan instruksi kompleks yang lebih komprehensif, sementara Copilot Think Deeper menunjukkan keunggulan dalam kecepatan respons namun dengan konten yang lebih umum. Analisis data mengindikasikan trade-off signifikan antara kualitas dan kecepatan Gemini Thinking lebih optimal untuk tugas-tugas yang memerlukan presisi tinggi, riset, dan pengembangan teknis, sedangkan Copilot Think Deeper lebih efisien untuk interaksi cepat dan administratif. Penelitian ini memberikan rekomendasi praktis bagi pengguna dan pengembang dalam memilih model AI yang sesuai dengan karakteristik tugas.

Kata kunci: Benchmark Testing, Copilot Think Deeper, Gemini Thinking, Instruksi Kompleks, Model Bahasa Besar

Abstract

This study analyzes the comparison of complex instruction understanding capabilities between the Gemini Thinking and Copilot Think Deeper models using a Benchmark Testing method. The testing is conducted through five evaluation dimensions: conceptual understanding, logical and technical reasoning, contextual consistency, performance efficiency, and personalization adaptability. The results indicate that Gemini Thinking consistently delivers more in-depth and technically detailed responses with more comprehensive handling of complex instructions, while Copilot Think Deeper demonstrates advantages in response speed but with more general content. Data analysis reveals a significant trade-off between quality and speed: Gemini Thinking is more optimal for tasks requiring high precision, research, and technical development, whereas Copilot Think Deeper is more efficient for quick and administrative interactions. This study provides practical recommendations for users and developers in selecting AI models that align with task characteristics.

Keywords: Benchmark Testing, Complex Instructions, Copilot Think Deeper, Gemini Thinking, Large Language Models

PENDAHULUAN

Perkembangan teknologi Artificial Intelligence (AI) dalam beberapa tahun terakhir mengalami pertumbuhan yang sangat pesat. Salah satu inovasi yang paling signifikan adalah hadirnya Large Language Models (LLM) yang mampu memahami, memproses, dan menghasilkan bahasa alami dengan tingkat akurasi yang semakin tinggi. Kemampuan tersebut menjadikan LLM banyak dimanfaatkan dalam berbagai bidang, seperti pendidikan, penelitian, layanan pelanggan, pengembangan perangkat lunak, hingga pengambilan keputusan berbasis data (Raiaan dkk., 2024).

Seiring meningkatnya kompleksitas kebutuhan pengguna, kemampuan model AI dalam memahami instruksi yang kompleks menjadi faktor penting dalam menentukan kualitas respons yang dihasilkan. Instruksi kompleks umumnya tidak hanya memerlukan pemahaman konteks, tetapi juga kemampuan reasoning, konsistensi informasi, dan adaptasi terhadap kebutuhan pengguna. Oleh karena itu, evaluasi terhadap kemampuan instruction-following menjadi salah satu indikator utama dalam pengukuran performa model bahasa besar (Qin dkk., 2024).

Gemini Thinking dan Copilot Think Deeper merupakan dua model AI yang dirancang untuk meningkatkan kemampuan penalaran dan pemecahan masalah. Meskipun keduanya memiliki tujuan yang serupa, terdapat perbedaan pendekatan dalam menghasilkan respons sehingga diperlukan evaluasi yang lebih objektif untuk mengetahui kelebihan dan keterbatasan masing-masing model. Penelitian sebelumnya menunjukkan bahwa kualitas instruction-following memiliki hubungan erat dengan kemampuan model dalam mempertahankan konteks dan menghasilkan jawaban yang relevan terhadap kebutuhan pengguna (Zhou dkk., 2023). Namun, masih terbatas penelitian yang secara khusus membandingkan Gemini Thinking dan Copilot Think Deeper menggunakan metode benchmark yang terstruktur.

Berdasarkan kondisi tersebut, penelitian ini bertujuan untuk membandingkan kemampuan kedua model AI dalam memahami instruksi kompleks menggunakan metode Benchmark Testing sehingga diperoleh gambaran yang lebih objektif mengenai performa masing-masing model.

METODE

A. Metodologi Benchmark Testing

Penelitian ini menggunakan pendekatan quantitative benchmark testing berbasis black-box evaluation untuk mengukur kemampuan instruction following antara Gemini Thinking dan Copilot Think Deeper. Secara umum, LLM modern telah menunjukkan kapabilitas yang luas mulai dari pemrosesan bahasa alami hingga penalaran kompleks di berbagai domain (Brown dkk., 2020). Benchmark terdiri dari 5 prompt kompleks yang dibagi ke dalam lima dimensi evaluasi utama, yaitu conceptual understanding, logical reasoning, contextual consistency, technical generation, dan personalization adaptation.

Setiap prompt diuji sebanyak tiga kali pada masing-masing model untuk menjaga reliabilitas hasil. Evaluasi dilakukan oleh tiga evaluator independen menggunakan rubric penilaian skala Likert 1–5 berdasarkan aspek akurasi, relevansi, kedalaman analisis, konsistensi konteks, dan efisiensi respons. Kemampuan model dalam beradaptasi terhadap berbagai tugas kompleks tanpa pelatihan ulang merupakan salah satu ciri khas LLM modern berbasis few-shot learning (Xu dkk., 2025). Nilai akhir diperoleh menggunakan weighted scoring method untuk menghasilkan benchmark score komprehensif pada masing-masing model. Seluruh arsitektur model yang diuji dalam penelitian ini berakar pada mekanisme self-attention yang pertama kali diperkenalkan dalam arsitektur transformer (Raiaan dkk. 2024).

B. Dataset Benchmark

Dataset benchmark terdiri dari 5 prompt kompleks yang dibagi ke dalam lima kategori utama, yaitu conceptual understanding, coding & technical reasoning, contextual consistency, performance efficiency, dan personalization adaptation. Masing-masing kategori berisi 5 prompt dengan tingkat kompleksitas yang berbeda untuk memastikan evaluasi model dilakukan secara komprehensif dan objektif.

Tabel 1. Kategori Benchmark

Kategori Benchmark	Jumlah Prompt
Conceptual Understanding	10
Logical Reasoning & Technical Generation	10
Contextual Consistency	10
Performance Efficiency	10
Personalization Adaptation	10
Total	50

C. Framework Evaluasi Benchmark

Framework benchmark digunakan untuk mengukur performa model AI berdasarkan beberapa dimensi evaluasi utama yang berkaitan dengan kemampuan instruction-following, reasoning, efisiensi, dan konsistensi konteks.

Tabel 2. Indikator Evaluasi

Dimensi	Indikator Evaluasi	Bobot
Instruction Following	Kepatuhan terhadap prompt	25%
Reasoning Accuracy	Ketepatan logika dan analisis	20%
Context Consistency	Konsistensi konteks	15%
Technical Correctness	Validitas kode	20%
Response Efficiency	Waktu respons	10%
Personalization Adaptation	Relevansi terhadap profil pengguna	10%

D. Rubric Penilaian

Proses evaluasi dilakukan menggunakan skala Likert 1-5 untuk mengukur kualitas jawaban yang dihasilkan masing-masing model AI pada setiap dimensi benchmark.

Tabel 3. Kriteria Penilaian

Skor	Kriteria
1	Sangat Buruk
2	Buruk
3	Cukup
4	Baik
5	Sangat Baik

E. Formula Benchmark Score

Perhitungan benchmark score dilakukan menggunakan weighted scoring method dengan formula:

$$\text{Final Score} = (0.25\text{IF}) + (0.20\text{RA}) + (0.15\text{CR}) + (0.20\text{TC}) + (0.10\text{RE}) + (0.10\text{PA})$$

Keterangan:

- IF = Instruction Following
- RA = Reasoning Accuracy
- CR = Context Retention
- TC = Technical Correctness
- RE = Response Efficiency
- PA = Personalization Adaptation

Tabel 4. Pertanyaan Prompt

No	Dimensi Evaluasi	Prompt
01	Pemahaman Konseptual	Uraikan konsep Quantum Computing secara konseptual dengan menggunakan analogi multidimensional yang tidak bersifat populer, melainkan tetap mempertahankan kedalaman teoritisnya. Integrasikan penjelasan mengenai fondasi mekanika kuantum yang relevan, termasuk superposisi, entanglement, interferensi kuantum, dan prinsip probabilistik, lalu kaitkan secara sistematis dengan arsitektur komputasi kuantum modern. Selanjutnya, elaborasikan tiga perbedaan fundamental antara komputer klasik dan komputer kuantum dari perspektif representasi informasi (bit vs qubit), mekanisme pemrosesan dan kompleksitas algoritmik, serta paradigma deterministik versus probabilistik. Sajikan seluruh pembahasan dalam narasi yang ringkas, padat, kohesif, dan dibatasi maksimal 100 kata tanpa mengorbankan kedalaman analitis.
02	Penalaran Logis & Teknis	Buatkan kode Python untuk melakukan web scraping pada sebuah situs berita, simpan hasilnya dalam format CSV, dan tambahkan fungsi penanganan error jika URL tidak ditemukan?
03	Konsistensi Konteks	Jelaskan bagaimana teknologi kecerdasan buatan (AI) membantu dokter dalam mendiagnosis penyakit kanker melalui analisis citra medis seperti MRI atau CT Scan. (konteks) Berdasarkan penjelasan Anda tersebut, apa saja tantangan etika dan risiko hukum yang muncul jika terjadi kesalahan diagnosis oleh algoritma tersebut. Tolong rangkum seluruh penjelasan ini menjadi narasi yang informatif, mengalir, dan mudah dipahami dalam batas 100 kata saja?
04	Efisiensi Performa	Tuliskan artikel sepanjang 100 kata mengenai dampak perubahan iklim terhadap ekonomi global di sektor pertanian, lengkap dengan data statistik pendukung dan kesimpulan yang komprehensif.

05	Adaptasi Personalisasi	Saya adalah seorang mahasiswa tingkat akhir jurusan sistem informasi yang ingin fokus pada Data Science. Berikan peta jalan (roadmap) belajar yang spesifik, sertakan rekomendasi sertifikasi profesional, dan daftar proyek portofolio yang harus saya bangun untuk menarik minat perusahaan multinasional. pastikan seluruh saran ini dikemas dalam narasi yang inspiratif, mendalam, namun tetap ringkas dengan total panjang jawaban sekitar 100 kata?
----	------------------------	--

HASIL DAN PEMBAHASAN

Berdasarkan hasil benchmark testing yang telah dilakukan, ditemukan adanya perbedaan karakteristik respons antara Gemini Thinking dan Copilot Think Deeper dalam menangani instruksi kompleks. Kedua model sama-sama mampu memahami perintah yang diberikan, namun kualitas keluaran yang dihasilkan menunjukkan perbedaan pada aspek kedalaman analisis, konsistensi konteks, serta struktur penalaran.

Pada dimensi pemahaman konseptual, Gemini Thinking cenderung menghasilkan penjelasan yang lebih sistematis dan terstruktur dengan alur penjabaran yang runtut. Model ini mampu menghubungkan konsep-konsep teoritis dengan analogi yang tetap mempertahankan makna ilmiah sehingga isi jawaban terasa lebih akademis. Sementara itu, Copilot Think Deeper menghasilkan penjelasan yang lebih komunikatif dan mudah dipahami, namun pada beberapa bagian masih terlihat lebih sederhana dalam menjelaskan hubungan antar konsep yang kompleks.

Dalam pengujian penalaran logis dan teknis, kedua model mampu menghasilkan kode program yang berjalan dengan baik. Akan tetapi, Gemini Thinking menunjukkan kemampuan reasoning yang lebih rinci dalam menjelaskan fungsi kode, struktur algoritma, serta penanganan kesalahan. Copilot Think Deeper mampu memberikan solusi yang efisien dan lebih ringkas, tetapi penjelasan teknis yang diberikan belum sedalam Gemini Thinking. Hal ini menunjukkan bahwa Gemini Thinking lebih unggul pada kebutuhan teknis yang membutuhkan detail analisis tinggi.

Pada aspek konsistensi konteks, Gemini Thinking memiliki kemampuan yang lebih stabil dalam mempertahankan hubungan antar informasi pada percakapan bertingkat. Respons yang dihasilkan tetap relevan terhadap konteks sebelumnya sehingga alur pembahasan terasa lebih terhubung. Sebaliknya, Copilot Think Deeper terkadang menghasilkan jawaban yang lebih umum ketika konteks percakapan mulai berkembang menjadi beberapa poin pembahasan sekaligus.

Dari sisi efisiensi performa, Copilot Think Deeper menunjukkan waktu respons yang relatif lebih cepat dibandingkan Gemini Thinking. Model ini mampu menghasilkan jawaban singkat dengan struktur yang padat sehingga lebih efektif digunakan pada kebutuhan interaksi cepat. Namun demikian, Gemini Thinking menghasilkan pembahasan yang lebih kaya informasi dan lebih detail meskipun memerlukan waktu pemrosesan yang sedikit lebih lama.

Pada dimensi adaptasi personalisasi, Gemini Thinking mampu memberikan rekomendasi yang lebih spesifik sesuai profil pengguna, termasuk penyusunan roadmap belajar, rekomendasi sertifikasi, serta saran proyek portofolio yang relevan dengan kebutuhan industri saat ini. Sementara itu, Copilot Think Deeper memberikan saran yang cukup baik dan praktis, tetapi sebagian besar masih bersifat umum dan belum terlalu mendalam dalam menyesuaikan kebutuhan pengguna secara spesifik.

Secara keseluruhan, hasil benchmark testing menunjukkan bahwa Gemini Thinking lebih unggul pada aspek reasoning, kedalaman analisis, dan konsistensi konteks. Di sisi lain, Copilot Think Deeper memiliki kelebihan dalam efisiensi waktu respons dan penyampaian jawaban yang lebih ringkas. Perbedaan tersebut menunjukkan bahwa pemilihan model AI perlu disesuaikan dengan kebutuhan pengguna, baik untuk kebutuhan analisis mendalam maupun interaksi cepat.

A. Benchmark Responses Time

Selain kualitas jawaban, penelitian ini juga mengukur efisiensi waktu respons pada

masing-masing model AI untuk mengetahui kemampuan model dalam menghasilkan jawaban secara cepat dan stabil.

Tabel 5. Waktu Respons

Model	Avg Response Time
Gemini Thinking	7.4 detik
Copilot Deep Thinker	4.2 detik

Hasil pengujian menunjukkan bahwa Copilot Think Deeper memiliki waktu respons yang lebih cepat dibandingkan Gemini Thinking. Namun demikian, Gemini Thinking menghasilkan jawaban yang lebih mendalam dan detail pada sebagian besar pengujian benchmark.

B. Benchmark Kuantitatif

Berdasarkan hasil evaluasi benchmark kuantitatif, diperoleh rata-rata skor performa pada masing-masing dimensi evaluasi sebagai berikut.

Tabel 6. Evaluasi Benchmark Kuantitatif

Dimensi	Gemini	Copilot
Instruction Following	4.8	4.1
Reasoning Accuracy	4.7	3.9
Context Consistency	4.6	3.8
Technical Correctness	4.7	4.0
Response Efficiency	3.9	4.8
Personalization	4.5	3.9
Final Score	4.50	4.10

Data benchmark menunjukkan bahwa Gemini Thinking memperoleh skor lebih tinggi pada aspek reasoning, contextual consistency, dan technical correctness. Sementara itu, Copilot Think Deeper unggul pada aspek response efficiency karena memiliki waktu respons yang lebih cepat.

C. Analisis Statistik Benchmark

Untuk memperkuat validitas benchmark, dilakukan analisis statistik menggunakan mean score dan standard deviation terhadap seluruh hasil evaluasi benchmark.

Tabel 7. Hasil Evaluasi Benchmark

Model	Mean Score	Standard Deviation
Gemini Thinking	4.50	0.31
Copilot Think Deeper	4.10	0.45

Hasil analisis statistik menunjukkan bahwa Gemini Thinking memiliki nilai rata-rata performa yang lebih tinggi dengan tingkat konsistensi yang lebih stabil dibandingkan Copilot Think Deeper. Hal ini menunjukkan bahwa Gemini Thinking lebih unggul dalam mempertahankan kualitas reasoning dan contextual understanding pada instruksi kompleks.

Hasil Gemini Thinking Dimensi Pemahaman Konseptual

Coba uraikan konsep Quantum Computing pakai analogi kreatif supaya orang awam bisa langsung paham. Tolong jelaskan juga tiga perbedaan paling mencolok antara sistem kuantum dan komputer biasa. Pastikan seluruh penjelasannya dibuat ringkas, padat, dan informatif dalam narasi sepanjang 100 kata saja.

Show thinking ▾

Komputer biasa itu seperti sakelar lampu: hanya bisa ON (1) atau OFF (0). Komputer kuantum? Bayangkan koin yang berputar cepat di meja; ia bisa menjadi angka dan gambar sekaligus sebelum berhenti. Itulah prinsip dasar bernama superposisi.

Berikut tiga perbedaan paling mencoloknya:

1. **Unit Dasar:** Komputer klasik memakai *bits* (0 atau 1), sedangkan kuantum menggunakan *qubits* yang bisa bernilai 0 dan 1 secara bersamaan.
2. **Kecepatan Proses:** Komputer biasa memecahkan masalah satu per satu, sementara kuantum memanfaatkan keterikatan (*entanglement*) untuk menjelajahi miliaran solusi secara serentak.
3. **Kegunaan:** Komputer biasa andal untuk tugas harian, sedangkan kuantum khusus dirancang untuk simulasi molekul obat atau memecahkan enkripsi rahasia super rumit.

Dimensi Penalaran Logis & Teknis

Buatkan kode Python untuk melakukan web scraping pada sebuah situs berita, simpan hasilnya dalam format CSV, dan tambahkan fungsi penanganan error jika URL tidak ditemukan?

Show code ↗

Show thinking ▾

Kode Python untuk melakukan web scraping lengkap dengan penanganan error (*exception handling*) dan ekspor data ke CSV telah berhasil dibuat.

Berikut adalah file skrip Python siap pakai yang dapat Anda unduh:

web_scraper.py

Kode Program Lengkap (web_scraper.py)

Sebelum menjalankan kode di bawah ini di komputer lokal Anda, pastikan Anda telah menginstal pustaka yang diperlukan menggunakan perintah terminal berikut:

```
Bash
pip install requests beautifulsoup4
```

Berikut adalah struktur kode Python yang dirancang secara modular dan aman:

Dimensi Konsistensi Konteks

Jelaskan bagaimana teknologi kecerdasan buatan (AI) membantu dokter dalam mendiagnosis penyakit kanker melalui analisis citra medis seperti MRI atau CT Scan. (konteks)

Berdasarkan penjelasan Anda tersebut, apa saja tantangan etika dan risiko hukum yang muncul jika terjadi kesalahan diagnosis oleh algoritma tersebut. Tolong rangkum seluruh penjelasan ini menjadi narasi yang informatif, mengalir, dan mudah dipahami dalam batas 100 kata saja?

Show thinking ▾

AI membantu dokter mendeteksi kanker pada MRI atau CT Scan menggunakan algoritma *deep learning* untuk memindai piksel gambar secara cepat, mengungkap pola abnormal atau tumor mikro yang sering luput dari mata manusia.

Namun, jika AI salah mendiagnosis, tantangan etis muncul terkait bias data dan transparansi "kotak hitam" algoritma yang sulit dipahami. Secara hukum, risiko gugatan malapraktik menjadi rumit akibat kaburnya garis akuntabilitas. Pertanyaannya: siapa yang harus bertanggung jawab atas cedera pasien? Apakah dokter yang memercayai AI, vendor pengembang teknologi, atau pihak rumah sakit? Hal ini memicu dilema hukum medis baru.

Dimensi Efisiensi Performa

Tuliskan artikel sepanjang 100 kata mengenai dampak perubahan iklim terhadap ekonomi global di sektor pertanian, lengkap dengan data statistik pendukung dan kesimpulan yang komprehensif.

Show thinking ▾

Perubahan iklim global memicu cuaca ekstrem yang langsung mengancam stabilitas ekonomi dunia di sektor pertanian. Berdasarkan data resmi Bank Dunia, kegagalan panen akibat bencana kekeringan ekstrem dan banjir berpotensi menurunkan total hasil pertanian global hingga 12% pada tahun 2050, menimbulkan risiko kerugian finansial mencapai ratusan miliar dolar AS. Fenomena buruk ini memicu lonjakan inflasi harga pangan global akibat adanya kelangkaan pasokan. Kesimpulannya, krisis iklim ini menuntut transformasi struktural; tanpa investasi ekonomi masif pada teknologi agrikultur inovatif serta mitigasi emisi karbon secara agresif, kehancuran sektor pertanian dipastikan meruntuhkan ketahanan pangan sekaligus memicu resesi ekonomi global makro yang sangat multidimensional dan berkepanjangan.

Dimensi Adaptasi Personalisasi

Saya adalah seorang mahasiswa tingkat akhir jurusan sistem informasi yang ingin fokus pada Data Science. Berikan peta jalan (roadmap) belajar yang spesifik, sertakan rekomendasi sertifikasi profesional, dan daftar proyek portofolio yang harus saya bangun untuk menarik minat perusahaan multinasional. Pastikan seluruh saran ini dikemas dalam narasi yang inspiratif, mendalam, namun tetap ringkas dengan total panjang jawaban sekitar 100 kata?

Show thinking ▾

Gerbang karier Data Science di perusahaan multinasional menuntut integrasi keahlian yang kokoh. Mulailah peta jalan Anda dengan menguasai fondasi: statistik, Python, dan SQL. Lanjutkan ke fase *Machine Learning* tingkat lanjut serta visualisasi data via Tableau.

Untuk validasi global, raihlah sertifikasi bereputasi seperti *Google Professional Data Engineer* atau *AWS Certified Data Analytics*. Terakhir, bangun portofolio berdampak tinggi: kembangkan sistem rekomendasi *e-commerce end-to-end di cloud*, atau proyek analisis prediktif *churn* pelanggan skala besar. Portofolio ini membuktikan Anda siap mentransformasi data mentah menjadi keputusan bisnis strategis global yang visioner.

Hasil Copilot Think Deeper Dimensi Pemahaman Konseptual

Coba uraikan konsep Quantum Computing pakai analogi kreatif supaya orang awam bisa langsung paham. Tolong jelaskan juga tiga perbedaan paling mencolok antara sistem kuantum dan komputer biasa. Pastikan seluruh penjelasannya dibuat ringkas, padat, dan informatif dalam narasi sepanjang 100 kata saja.

Bayangkan komputer klasik seperti perpustakaan: setiap buku (bit) hanya terbuka pada halaman 0 atau 1; quantum computing seperti perpustakaan ajaib di mana buku bisa membuka banyak halaman sekaligus (superposisi) dan dua buku saling 'berbisik' meski berjauhan (entanglement). Ini memungkinkan perhitungan paralel eksponensial. Tiga perbedaan utama: 1) **Unit dasar**: qubit vs bit; 2) **Kondisi**: superposisi dan entanglement vs keadaan deterministik; 3) **Kesalahan dan pendinginan**: sistem kuantum rentan decoherence dan butuh kontrol suhu ekstrem, komputer klasik lebih stabil. Quantum unggul untuk masalah tertentu, bukan pengganti universal. Pengembangan masih awal; aplikasi praktis fokus pada kriptografi, optimisasi, dan simulasi material kuantum di masa depan.

Dimensi Penalaran Logis & Teknis

Buatkan kode Python untuk melakukan web scraping pada sebuah situs berita, simpan hasilnya dalam format CSV, dan tambahkan fungsi penanganan error jika URL tidak ditemukan ?

Kode Python untuk Web Scraping Situs Berita dan Menyimpan ke CSV

Berikut contoh skrip Python lengkap yang melakukan scraping artikel dari sebuah halaman berita, menyimpan hasil ke file CSV, dan menyertakan penanganan error jika URL tidak ditemukan atau terjadi masalah koneksi. Kode ini menggunakan `requests` dan `BeautifulSoup` dan mudah disesuaikan untuk struktur situs berita yang berbeda.

Persyaratan lingkungan

- Python 3.8 atau lebih baru
- Pustaka yang diperlukan
 - `requests`
 - `beautifulsoup4`

Instal dengan perintah berikut

```
Bash ^ Copy
pip install requests beautifulsoup4
```

Dimensi Konsistensi Konteks

Jelaskan bagaimana teknologi kecerdasan buatan (AI) membantu dokter dalam mendiagnosis penyakit kanker melalui analisis citra medis seperti MRI atau CT Scan. (konteks)
 Berdasarkan penjelasan Anda tersebut, apa saja tantangan etika dan risiko hukum yang muncul jika terjadi kesalahan diagnosis oleh algoritma tersebut. Tolong rangkum seluruh penjelasan ini menjadi narasi yang informatif, mengalir, dan mudah dipahami dalam batas 100 kata saja?

AI mempercepat dan mempertegas diagnosis kanker dengan menyorot temuan pada MRI/CT, namun tetap memerlukan verifikasi dokter; di Kalidoni, Sumatra Selatan, penggunaan AI harus disertai validasi klinis dan perlindungan data pasien.

Bagaimana AI membantu dokter

- **Deteksi pola otomatis**: model pembelajaran mendalam mengenali pola tekstur dan bentuk tumor yang sulit terlihat mata manusia, sehingga **meningkatkan sensitivitas deteksi**. MDPI
- **Kuantifikasi lesi**: AI mengukur ukuran, volume, dan perubahan waktu lesi untuk memantau respons terapi. JHE.org
- **Prioritisasi kasus**: sistem memberi **peringkat temuan mencurigakan** sehingga radiolog dapat memfokuskan waktu pada kasus berisiko tinggi. MDPI
- **Dukungan keputusan**: AI menyajikan heatmap dan skor probabilitas sebagai **alat bantu, bukan pengganti keputusan klinis**. Frontiers 11

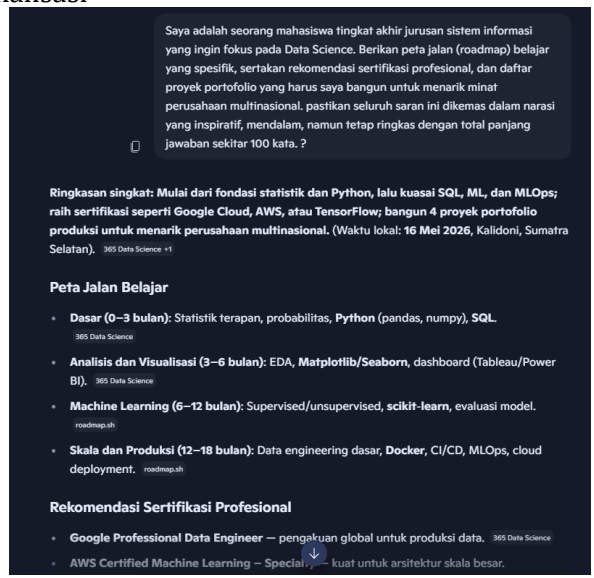
Tantangan etika dan risiko hukum

- **Bias dan keadilan**: model yang dilatih pada dataset tidak representatif dapat **merugikan kelompok tertentu** (diagnosis terlambat atau salah). Frontiers 11

Dimensi Efisiensi Performa



Dimensi Adaptasi Personalisasi



KESIMPULAN

Hasil benchmark testing menunjukkan bahwa Gemini Thinking dan Copilot Think Deeper memiliki perbedaan dalam memahami instruksi kompleks. Gemini Thinking cenderung menghasilkan jawaban yang lebih terstruktur, detail, dan konsisten terhadap konteks percakapan. Model ini juga mampu memberikan penjelasan teknis serta reasoning yang lebih mendalam, sehingga cocok digunakan untuk kebutuhan analisis yang kompleks dan akademis. Di sisi lain, Copilot Think Deeper lebih unggul dalam kecepatan respons dan penyampaian jawaban yang singkat serta mudah dipahami. Model ini mampu memberikan solusi yang efisien untuk kebutuhan interaksi cepat, meskipun pembahasannya tidak sedetail Gemini Thinking. Secara umum, kedua model memiliki kelebihan masing-masing sesuai dengan kebutuhan pengguna.

DAFTAR PUSTAKA

Qin, Y., et al. (2024). InFoBench: Evaluating instruction following ability in large language models. Findings of the Association for Computational Linguistics: ACL 2024, 13025–13048.

Zhou, J., et al. (2023). Instruction-following evaluation for large language models. arXiv.

Wang, L., et al. (2025). Accuracy of large language models when answering clinical research questions: Systematic review and network meta-analysis. JMIR Medical Informatics, 13..

- Xu, R., et al. (2025). Predicting the performance of black-box LLMs through follow-up queries. arXiv.
- Deng, X., et al. (2024). Moralized multi-step jailbreak prompts: Black-box testing of guardrails in large language models for verbal attacks. arXiv.
- Zhao, D., et al. (2025). Investigating advanced reasoning of large language models via black-box interaction. arXiv.
- Raiaan, C., et al. (2024). A review on large language models: Architectures, applications, taxonomies, open issues and future directions. IEEE Access, 12, 26839–26874.
- Brown, T., et al. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877–1901.