



Department of Digital Business

**Journal of Artificial Intelligence and Digital Business (RIGGS)**

Homepage: <https://journal.ilmudata.co.id/index.php/RIGGS>

Vol. 5 No. 1 (2026) pp: 2394-2402

P-ISSN: 2963-9298, e-ISSN: 2963-914X

---

## Analisis Perbandingan Efektivitas *Zero-Shot* vs *Chain-of-Thought* Prompting dalam Meningkatkan Presisi Informasi pada LLM Berbasis Dokumen PDF

Kurnia Yahya<sup>1</sup>, Dikwan Moeis\*<sup>2</sup>, Musdalifa Thamrin<sup>3</sup>, Sry Yunarti<sup>4</sup>, Alvina Felicia Watratan<sup>5</sup>, Satriawaty Mallu<sup>6</sup>

<sup>1,2,5,6</sup>Program Studi Ilmu Komputer, STMIK Profesional Makassar

<sup>3</sup>Program Studi Manajemen Informatika, STMIK Profesional Makassar

<sup>4</sup>Program Studi Sistem Informasi, STMIK Profesional Makassar

<sup>1</sup>[kurnia\\_yahya@stmikprofesional.ac.id](mailto:kurnia_yahya@stmikprofesional.ac.id), <sup>2</sup>[dikwan\\_mois@stmikprofesional.ac.id](mailto:dikwan_mois@stmikprofesional.ac.id), <sup>3</sup>[nonongthamrin@gmail.com](mailto:nonongthamrin@gmail.com),

<sup>4</sup>[yunarti2009@gmail.com](mailto:yunarti2009@gmail.com), <sup>5</sup>[vinawatratan@stmikprofesional.ac.id](mailto:vinawatratan@stmikprofesional.ac.id), <sup>6</sup>[mssatriawaty@gmail.com](mailto:mssatriawaty@gmail.com)

### Abstrak

Pemanfaatan *Large Language Models (LLM)* dalam ekosistem pendidikan tinggi untuk ekstraksi informasi akademik sering kali terkendala oleh fenomena halusinasi data yang dapat menyesatkan pengguna. Penelitian ini bertujuan untuk menganalisis perbandingan efektivitas antara teknik *Zero-Shot* dan *Chain-of-Thought (CoT)* prompting dalam upaya meningkatkan presisi informasi pada sistem *Retrieval-Augmented Generation (RAG)* yang berbasis dokumen PDF. Metode penelitian yang digunakan adalah eksperimen laboratorium dengan mengintegrasikan framework *LangChain* dan database vektor *FAISS* untuk memproses dokumen teknis berupa Panduan Akademik institusi. Evaluasi kualitas jawaban dilakukan melalui pendekatan hibrida, menggunakan metrik otomatis *BERTScore* untuk mengukur kemiripan semantik dan penilaian manusia (*Human Evaluation*) oleh pakar Teknologi Informasi untuk mengukur tingkat presisi serta validitas informasi secara kualitatif. Hasil penelitian menunjukkan bahwa teknik *Chain-of-Thought* secara konsisten mengungguli *Zero-Shot* di seluruh parameter evaluasi yang diuji. Peningkatan paling signifikan tercatat pada aspek validitas jawaban berdasarkan penilaian pakar sebesar 37,14%, serta kenaikan skor presisi pada metrik *BERTScore* sebesar 9,03%. Temuan ini membuktikan bahwa mekanisme penalaran logis secara bertahap pada teknik *CoT* mampu mereduksi halusinasi secara efektif dengan memastikan setiap klaim jawaban memiliki jejak audit yang kuat pada dokumen sumber. Implikasi praktis dari penelitian ini memberikan rekomendasi strategis bagi pengembang sistem informasi akademik untuk menerapkan pendekatan '*CoT-by-Design*' guna membangun asisten virtual yang lebih akurat, etis, dan kredibel di lingkungan perguruan tinggi.

*Kata kunci:* *Chain-of-Thought, LLM, Prompt Engineering, RAG, Presisi Informasi.*

### 1. Latar Belakang

Transformasi digital dalam ekosistem pendidikan tinggi di Indonesia telah mencapai titik balik yang signifikan seiring dengan masifnya adopsi teknologi Kecerdasan Buatan (AI), khususnya *Large Language Models (LLM)* seperti ChatGPT. Fenomena ini bukan sekadar tren sesaat, melainkan sebuah pergeseran paradigma dalam cara civitas akademika memproses informasi. Data menunjukkan adanya lonjakan minat dan penggunaan alat berbasis generatif AI yang sangat dinamis di berbagai wilayah Indonesia [1]. Di satu sisi, teknologi ini menawarkan peluang revolusioner dalam meningkatkan efisiensi penulisan akademik dan membantu mahasiswa dalam menyusun, mengedit, serta merevisi teks ilmiah dengan lebih cepat [2], [3]. ChatGPT telah menjadi katalisator bagi inovasi pendidikan di era Education 4.0, yang memungkinkan personalisasi pembelajaran dan peningkatan keterampilan menulis melalui interaksi yang lebih interaktif [4].

Namun, di balik kemudahan yang ditawarkan, muncul tantangan serius terkait integritas kognitif dan kualitas hasil yang diproduksi. Penggunaan AI yang berlebihan tanpa pengawasan yang kritis berisiko memicu tingkat ketergantungan yang tinggi di kalangan mahasiswa, yang pada gilirannya dapat menyebabkan penurunan kemampuan berpikir kritis atau yang sering diidentifikasi sebagai kemalasan berpikir [5]. Mahasiswa sering kali menggunakan ChatGPT sebagai alat bantu utama dalam mencari referensi tugas akhir tanpa melakukan verifikasi

mendalam terhadap sumber aslinya [6]. Selain itu, ditemukan bahwa pemanfaatan AI yang tidak terarah dapat mempengaruhi keterampilan mahasiswa dalam menyusun pertanyaan yang efektif, yang merupakan fondasi utama dalam metode penelitian [7]. Dampak ini menciptakan paradoks, di mana teknologi yang seharusnya mencerdaskan justru dapat melemahkan daya nalar jika tidak disertai dengan literasi AI dan pemahaman etika komunikasi yang memadai [8].

Salah satu hambatan teknis yang paling krusial dalam penggunaan LLM saat ini adalah fenomena "halusinasi", di mana AI memberikan informasi yang tampak meyakinkan namun secara faktual tidak akurat. Hal ini menjadi sangat berisiko ketika AI digunakan untuk mengekstraksi informasi dari dokumen teknis atau ilmiah yang kompleks. Masalah ini sering kali berpangkal pada ketidakmampuan model dalam memahami konteks yang sangat spesifik jika instruksi yang diberikan terlalu umum. Oleh karena itu, optimasi kinerja output AI melalui teknik prompt engineering menjadi sebuah keharusan akademis agar informasi yang dihasilkan tetap kredibel dan dapat dikustomisasi sesuai kebutuhan [9]. Pelatihan teknis mengenai perancangan instruksi atau prompt mulai banyak diimplementasikan untuk membekali mahasiswa dengan kemampuan mengarahkan AI agar hasil tulisannya lebih logis dan terstruktur [10].

Perkembangan terkini menunjukkan bahwa integrasi LLM dengan teknik Retrieval-Augmented Generation (RAG) telah menjadi solusi populer untuk membatasi ruang lingkup jawaban AI agar hanya merujuk pada basis pengetahuan tertentu, seperti dokumen PDF yang diunggah. Implementasi ini telah banyak diterapkan dalam berbagai layanan, mulai dari sistem pencarian skripsi mahasiswa yang terintegrasi dengan media sosial [11], hingga asisten virtual untuk pelayanan perpustakaan yang menggunakan model Llama 3 untuk meningkatkan akurasi jawaban [12]. Bahkan dalam sektor pelayanan publik dan kesehatan, teknik RAG telah diuji coba untuk memberikan informasi yang lebih valid mengenai obat dan vitamin berdasarkan dokumen resmi dari otoritas kesehatan [13]. Hal ini menunjukkan bahwa fokus riset saat ini telah bergeser dari sekadar penggunaan umum ke arah optimasi teknis berbasis dokumen referensi.

Meskipun demikian, keberhasilan sistem berbasis RAG sangat bergantung pada gaya prompting yang digunakan. Penelitian terdahulu telah mencoba menganalisis efektivitas gaya prompt bebas (zero-shot) dibandingkan dengan prompt terbimbing (few-shot) untuk melihat dampaknya terhadap struktur dan presisi jawaban [14]. Ada pula upaya untuk mengoptimalkan penggunaan AI dalam pengembangan bahan ajar melalui parameter persona, context, dan task guna meminimalkan margin kesalahan pada sistem manajemen pembelajaran [15]. Namun, terdapat sebuah celah penelitian (research gap) yang signifikan, yang mana sebagian besar studi di Indonesia masih berfokus pada perbandingan zero-shot vs few-shot atau implementasi praktis di lapangan (seperti perencanaan gizi atau penyusunan perangkat pembelajaran SMA) tanpa membedah mekanisme penalaran langkah demi langkah atau yang dikenal sebagai Chain-of-Thought (CoT) [16], [17].

Chain-of-Thought prompting merupakan teknik yang memaksa model bahasa untuk menguraikan logika berpikirnya secara sekuensial sebelum memberikan jawaban akhir. Di tengah tantangan interpretasi dokumen PDF yang sering kali memiliki bahasa teknis yang padat dan struktur yang kompleks, teknik CoT diprediksi mampu memberikan tingkat presisi yang lebih tinggi dibandingkan teknik zero-shot yang cenderung memberikan jawaban langsung tanpa pemrosesan konteks yang mendalam. Hingga saat ini, perbandingan antara kecepatan instruksi langsung dengan kedalaman penalaran CoT pada dokumen ilmiah berbahasa Indonesia masih sangat jarang dieksplorasi secara eksperimental.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk melakukan analisis perbandingan efektivitas antara teknik Zero-Shot dan Chain-of-Thought Prompting dalam konteks peningkatan presisi informasi pada sistem LLM berbasis RAG. Penelitian ini menggunakan dokumen PDF sebagai sumber pengetahuan utama untuk menguji ketajaman kedua teknik tersebut dalam mengekstraksi fakta dan menjaga konsistensi jawaban. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi teoritis bagi pengembangan framework literasi AI di perguruan tinggi serta memberikan panduan praktis bagi pengembang sistem informasi akademik dalam membangun asisten virtual yang lebih akurat, etis, dan minim halusinasi.

## 2. Tinjauan Pustaka

### 2.1. Landasan Teknik Prompt Engineering dalam Ekosistem LLM

Perkembangan Large Language Models (LLM) telah membawa pergeseran signifikan dalam cara manusia berinteraksi dengan mesin melalui bahasa natural. Di tengah masifnya penggunaan ChatGPT, muncul sebuah disiplin baru yang dikenal sebagai prompt engineering, yaitu teknik merancang, mengoptimalkan, dan menyempurnakan instruksi untuk mengarahkan model bahasa agar menghasilkan output yang relevan dan akurat. Teknik ini menjadi krusial karena kinerja output AI sangat bergantung pada bagaimana instruksi dikustomisasi sesuai kebutuhan spesifik pengguna [9]. Dalam konteks pendidikan, prompt engineering tidak hanya berfungsi sebagai alat teknis, tetapi juga sebagai instrumen untuk meningkatkan kemampuan menulis artikel ilmiah, terutama bagi mahasiswa yang sedang beradaptasi dengan literasi akademik digital [10].

Lebih jauh lagi, efektivitas sebuah prompt sering kali ditentukan oleh kelengkapan elemen pendukung seperti pemberian peran (persona), penyediaan konteks, dan penentuan format output yang jelas. Upaya untuk mengoptimalkan penggunaan AI dalam bahan ajar pada sistem manajemen pembelajaran (LMS) menunjukkan bahwa penerapan teknik prompting yang terstruktur dapat menekan margin kesalahan dan menghasilkan materi yang lebih adaptif bagi peserta didik [15]. Praktik ini juga bersinggungan erat dengan etika komunikasi, di mana kejujuran intelektual dan transparansi dalam proses pembentukan pesan menjadi tantangan sekaligus peluang bagi integritas akademik di era kecerdasan buatan [8]. Dengan demikian, prompt engineering bertindak sebagai jembatan antara potensi mentah model bahasa dan kebutuhan fungsional yang spesifik di lapangan.

### 2.2. Evolusi Retrieval-Augmented Generation (RAG) dan Implementasi Sistem Hibrida

Meskipun LLM memiliki basis pengetahuan yang luas, model ini sering kali terbatas oleh tanggal pemutusan data (cutoff date) dan rentan terhadap kesalahan informasi. Untuk mengatasi keterbatasan tersebut, teknik Retrieval-Augmented Generation (RAG) hadir sebagai arsitektur yang memungkinkan AI untuk mengambil referensi dari dokumen eksternal yang terpercaya secara real-time. Implementasi RAG telah terbukti efektif dalam menyediakan informasi yang sangat spesifik, seperti sistem informasi obat dan vitamin yang merujuk pada dokumen otoritas kesehatan guna memastikan validitas data semantik yang dihasilkan [13].

Di lingkungan perguruan tinggi, arsitektur RAG mulai diadopsi untuk mempermudah akses terhadap aset intelektual kampus melalui pemanfaatan framework LangChain dan integrasi media sosial [11]. Pemanfaatan teknologi serupa juga ditemukan dalam layanan perpustakaan digital, di mana model Llama 3 yang dikombinasikan dengan pendekatan RAG mampu mendukung pelayanan informasi secara lebih responsif [12]. Menariknya, tren pengembangan chatbot akademik kini mulai bergerak ke arah hibridisasi antara LLM dan sistem berbasis aturan (rule-based system). Pendekatan ini bertujuan untuk menyeimbangkan fleksibilitas interaksi alami LLM dengan ketegasan aturan administratif kampus yang kaku. Implementasi chatbot otomatis berbasis web menunjukkan bahwa sinergi ini mampu memberikan respons yang alami namun tetap terkendali sesuai protokol akademik yang berlaku [18], [19]. Hal ini mempertegas bahwa AI telah berevolusi menjadi sistem pakar yang mampu mengolah dokumen PDF secara presisi untuk kebutuhan layanan publik maupun perencanaan gaya hidup sehat di masyarakat [16].

### 2.3. Dinamika Teknik Zero-Shot dan Chain-of-Thought (CoT)

Dalam operasionalisasinya, terdapat berbagai gaya prompting yang mempengaruhi kualitas penalaran model. Teknik yang paling sederhana adalah zero-shot prompting, di mana pengguna memberikan tugas langsung tanpa memberikan contoh atau langkah penalaran sebelumnya. Namun, analisis efektivitas menunjukkan bahwa instruksi yang lebih terarah atau guided prompts cenderung menghasilkan tingkat presisi dan kesesuaian leksikal yang lebih baik dibandingkan dengan free prompts yang bersifat terbuka [14].

Salah satu inovasi yang menjadi fokus utama dalam penelitian tingkat lanjut adalah teknik Chain-of-Thought (CoT). Berbeda dengan zero-shot, CoT mendorong model untuk menguraikan langkah-langkah logika internal sebelum mencapai kesimpulan akhir. Hal ini sangat relevan untuk mengatasi kompleksitas dokumen akademik, di mana interpretasi teks sering kali membutuhkan pemahaman kontekstual yang mendalam. Penerapan CoT diharapkan dapat mereduksi potensi penurunan kemampuan berpikir kritis yang sering dikhawatirkan dalam penggunaan ChatGPT secara instan [5]. Walaupun penggunaan AI sebagai alat bantu pencarian referensi telah

banyak membantu mahasiswa dalam menyusun tugas akhir [6], efektivitasnya dalam menjaga integritas logika tetap menjadi perhatian utama. Melalui proses dekomposisi masalah, model AI diarahkan untuk menunjukkan alur penalaran yang konsisten dengan dokumen referensi yang diunggah.

## **2.4 Tantangan Literasi dan Pengembangan Keterampilan Kognitif**

Integrasi LLM dalam dunia pendidikan juga membawa konsekuensi pada keterampilan kognitif pengguna. Di satu sisi, AI berperan positif dalam dimensi linguistik, seperti meningkatkan ketepatan tata bahasa dan koherensi dalam penulisan akademik bahasa Indonesia [2], [3]. Di sisi lain, muncul dimensi kognitif baru di mana intensitas penggunaan ChatGPT justru berkorelasi positif dengan kemampuan mahasiswa dalam merumuskan instruksi yang memiliki kejelasan, kedalaman, dan relevansi yang kuat [7]. Keterampilan menyusun pertanyaan yang strategis menjadi kunci utama dalam mengekstraksi pengetahuan yang orisinal.

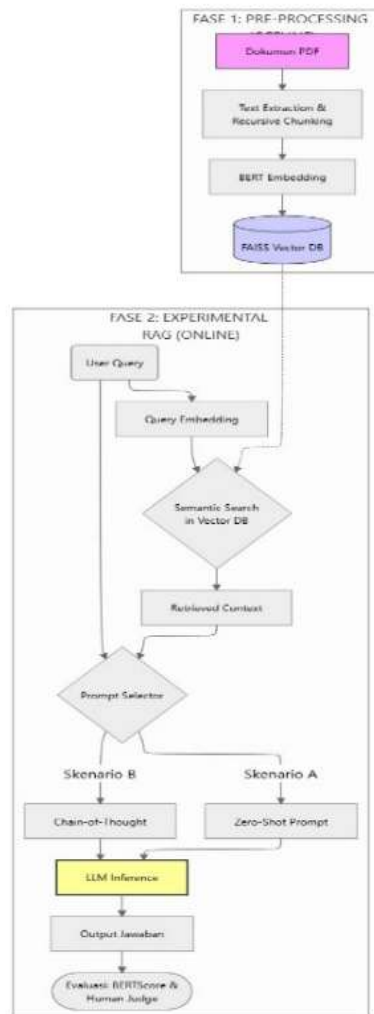
Namun, risiko penurunan keterampilan tetap ada jika mahasiswa terlalu bergantung pada otomatisasi tanpa pemahaman fundamental terhadap materi yang dipelajari [20]. Oleh karena itu, pemanfaatan AI untuk menyusun perangkat pembelajaran, mulai dari tingkat SMA hingga perguruan tinggi, harus tetap mengedepankan aspek verifikasi manual untuk menjamin kualitas hasil yang diproduksi [17]. Pemahaman komprehensif mengenai batasan teknologi ini, sebagaimana dibahas dalam berbagai kajian tren penggunaan AI di Indonesia, menjadi pondasi bagi pengembangan strategi prompting yang lebih bertanggung jawab dan akurat [1], [4]. Penulisan ilmiah berbasis AI seharusnya dipandang sebagai kolaborasi strategis antara nalar manusia dan efisiensi mesin, di mana teknik pengarahan yang digunakan mampu menjaga kedalaman argumen dan validitas informasi.

## **3. Metodologi Penelitian**

### **3.1. Rancangan Penelitian dan Arsitektur Sistem**

Penelitian ini menggunakan pendekatan eksperimental untuk mengevaluasi performa penalaran Large Language Model (LLM) dalam lingkungan Retrieval-Augmented Generation (RAG). Arsitektur sistem dibangun menggunakan framework LangChain sebagai orkestrator utama yang menghubungkan basis pengetahuan eksternal dengan model bahasa [11]. Secara teknis, sistem ini mengadopsi alur kerja yang diawali dengan tahap preprocessing dokumen, di mana dokumen PDF dikonversi menjadi unit teks yang lebih kecil (chunking) menggunakan Recursive Character Text Splitter.

Setiap chunk teks kemudian ditransformasikan menjadi representasi vektor numerik menggunakan model embedding SentenceTransformer berbasis BERT. Vektor-vektor ini disimpan dalam vector database FAISS (Facebook AI Similarity Search) untuk memungkinkan pencarian semantik yang efisien berdasarkan kedekatan kosinus (cosine similarity) antara kueri pengguna dan fragmen dokumen [12]. Implementasi ini bertujuan untuk memastikan bahwa model hanya memberikan jawaban berdasarkan konteks yang relevan, sehingga meminimalisir risiko halusinasi yang sering terjadi pada model generatif murni.



Gambar 1. Arsitektur Sistem RAG dengan Komparasi Strategi Prompting

Gambar 1 mengilustrasikan arsitektur sistem yang dibangun di atas orkestrator LangChain. Proses dimulai pada Fase A (Ingesti Data), di mana dokumen PDF mentah dikonversi menjadi representasi vektor (embeddings) menggunakan model SentenceTransformer dan disimpan dalam basis data vektor FAISS.

Pada Fase B (Retrieval & Generasi), kueri dari pengguna juga diubah menjadi vektor untuk melakukan pencarian semantik (cosine similarity) guna menemukan konteks yang paling relevan dari FAISS. Inti dari eksperimen ini terletak pada blok "Eksperimen Prompting", di mana konteks dan kueri yang sama dimasukkan ke dalam dua jalur berbeda: Skenario A (Zero-Shot) yang memberikan instruksi langsung, dan Skenario B (Chain-of-Thought) yang menyisipkan instruksi penalaran bertahap sebelum diteruskan ke LLM untuk menghasilkan jawaban akhir.

### 3.2. Prosedur Eksperimen dan Dataset

Dataset yang digunakan dalam penelitian ini terdiri dari dokumen PDF teknis yang bersifat otoritatif, yaitu Panduan Akademik dan Dokumen Spesifikasi Sistem Informasi institusi. Pemilihan dokumen ini didasarkan pada kompleksitas struktur teks dan kepadatan informasi yang membutuhkan penalaran tinggi untuk diekstraksi secara akurat. Eksperimen dilakukan dengan membandingkan dua skenario pemberian instruksi (prompting) sebagai berikut:

1. Skenario A (Zero-Shot Prompting): Model diberikan pertanyaan langsung tanpa instruksi tambahan atau contoh penalaran. Pendekatan ini mencerminkan penggunaan umum di mana pengguna mengharapkan jawaban instan dari sistem [14].

2. Skenario B (Chain-of-Thought Prompting): Model diinstruksikan untuk mengikuti langkah-langkah logika sekuensial sebelum merumuskan jawaban akhir. Teknik ini dirancang untuk memaksa model melakukan dekomposisi masalah, mengidentifikasi poin kunci dalam dokumen, dan menghubungkan informasi tersebut secara logis.

### 3.3. Instrumen Evaluasi

Kualitas keluaran dari kedua skenario tersebut diukur menggunakan pendekatan hibrida yang menggabungkan metrik otomatis dan penilaian manusia (human evaluation):

1. BERTScore: Digunakan untuk mengukur kemiripan semantik antara jawaban chatbot dengan jawaban referensi (kunci jawaban) yang telah divalidasi oleh pakar. Berbeda dengan metrik berbasis teks tradisional seperti ROUGE, BERTScore mampu menangkap makna mendalam melalui pemetaan ruang vektor, sehingga lebih representatif dalam menilai kualitas bahasa alami [13].
2. Human Evaluation (Expert Judgement): Melibatkan dua orang dosen senior di bidang Teknologi Informasi untuk menilai aspek Presisi (sejauh mana jawaban menjawab pertanyaan secara tepat) dan Tingkat Halusinasi (deteksi adanya informasi tambahan yang tidak terdapat dalam dokumen PDF sumber). Penilaian dilakukan menggunakan skala Likert 1-5 untuk mendapatkan data kuantitatif yang objektif.

### 3.4. Visualisasi Kontras Prompt

Untuk memperjelas perbedaan perlakuan dalam eksperimen ini, tabel berikut menyajikan perbandingan desain prompt yang digunakan pada kedua skenario:

Tabel 1. Perbandingan Desain Prompt Zero-Shot dan Chain-of-Thought

Komponen	Skenario A: Zero-Shot	Skenario B: Chain-of-Thought (CoT)
<b>Instruksi Utama</b>	"Jawab pertanyaan berikut berdasarkan konteks yang diberikan."	"Selesaikan tugas ini dengan berpikir langkah-demi-langkah agar jawaban Anda akurat."
<b>Struktur Logika</b>	Langsung ke jawaban.	1. Analisis kueri pengguna. 2. Identifikasi bagian relevan pada dokumen PDF. 3. Hubungkan antar poin informasi. 4. Rumuskan kesimpulan logis.
<b>Output yang Diharapkan</b>	Jawaban ringkas.	Penjelasan alur berpikir diikuti dengan jawaban akhir yang presisi.
<b>Contoh Prompt</b>	"Berapa batas maksimal SKS untuk mahasiswa semester 3?"	1. Temukan aturan beban studi di PDF. 2. Lihat tabel prasyarat IPK. 3. Hitung batas maksimal SKS untuk semester 3. 4. Berikan jawaban akhir."

### 3.5 Justifikasi Mekanisme Mitigasi Halusinasi melalui CoT

Penggunaan teknik Chain-of-Thought (CoT) dalam penelitian ini bukan sekadar upaya prosedural, melainkan strategi untuk mengatasi batasan kognitif LLM dalam memproses konteks RAG yang berasal dari dokumen PDF. Terdapat tiga alasan mekanis mengapa CoT mampu mereduksi halusinasi dibandingkan Zero-Shot:

1. Dekomposisi Informasi Berjenjang: Dokumen PDF teknis seringkali mengandung informasi yang saling tumpang tindih. Dalam skenario Zero-Shot, model cenderung melakukan generalisasi yang dangkal. CoT memaksa model untuk melakukan dekomposisi kueri menjadi sub-tugas: (1) Ekstraksi fakta dari context chunk, (2) Verifikasi relevansi fakta terhadap kueri, dan (3) Sintesis jawaban. Proses ini memastikan bahwa setiap bagian dari jawaban akhir memiliki "jejak audit" pada dokumen sumber.

2. Mekanisme Self-Correction Spasial: PDF memiliki struktur tata letak yang kompleks (tabel, bullet points, dan catatan kaki). Saat melakukan RAG, urutan teks sering kali terfragmentasi. CoT memberikan ruang bagi model untuk "meninjau kembali" urutan informasi tersebut secara logis sebelum merumuskan jawaban, sehingga meminimalisir kesalahan interpretasi terhadap data numerik atau prasyarat aturan akademik yang sering terlewat pada instruksi langsung [14], [15].
3. Penguatan Attention pada Konteks: Dengan menginstruksikan model untuk "berpikir langkah-demi-langkah", bobot perhatian (attention weights) model dialihkan secara lebih intensif pada fragmen dokumen yang diambil dari database vektor (FAISS/ChromaDB). Hal ini mencegah model untuk menggunakan "pengetahuan internal" (internal bias) yang tidak relevan dengan dokumen PDF, yang merupakan akar utama dari fenomena halusinasi pada sistem AI generatif [9].

## 4. Hasil dan Pembahasan

### 4.1. Analisis Kuantitatif Performa Prompting

Eksperimen dilakukan dengan menguji 30 kueri kompleks yang bersumber dari dokumen Panduan Akademik. Evaluasi dilakukan dengan membandingkan respons sistem pada skenario Zero-Shot dan Chain-of-Thought (CoT) menggunakan metrik otomatis BERTScore dan penilaian manusia (Human Evaluation) oleh pakar TI. Hasil perbandingan kinerja kedua teknik tersebut dirangkum dalam Tabel 2.

Tabel 2. Perbandingan Skor Efektivitas Zero-Shot vs Chain-of-Thought

Metrik Evaluasi	Zero-Shot Prompting	Chain-of-Thought (CoT)	Peningkatan (%)
<b>BERTScore (Precision)</b>	0.8214	0.8956	+9.03%
<b>BERTScore (Recall)</b>	0.8532	0.8741	+2.45%
<b>BERTScore (F1-Score)</b>	0.8370	0.8847	+5.69%
<b>Human Eval: Presisi (1-5)</b>	3.8	4.7	+23.68%
<b>Human Eval: Validitas (1-5)</b>	3.5	4.8	+37.14%

Data pada Tabel 2 menunjukkan bahwa teknik Chain-of-Thought (CoT) secara konsisten mengungguli Zero-Shot di semua parameter. Peningkatan paling signifikan terlihat pada aspek Validitas (Human Eval) sebesar 37,14%, yang mengindikasikan bahwa CoT jauh lebih efektif dalam menekan angka halusinasi dibandingkan instruksi langsung.

### 4.2. Analisis Temuan: Mengapa CoT Lebih Unggul?

Berdasarkan analisis terhadap log respons sistem, ditemukan bahwa Zero-Shot cenderung memberikan jawaban yang bersifat generalis dan terkadang "memaksakan" informasi jika dokumen referensi tidak menyebutkan jawaban secara eksplisit. Hal ini sering terjadi pada kueri yang membutuhkan perhitungan atau interpretasi aturan bersyarat (misal: "Berapa SKS yang bisa diambil jika IPK 3.25 namun ada nilai D di mata kuliah prasyarat?").

Sebaliknya, teknik CoT menunjukkan performa yang superior pada kueri yang membutuhkan penalaran logis. Dengan memaksa LLM untuk melakukan dekomposisi langkah demi langkah, model mampu melakukan "cek silang" internal terhadap konteks yang diambil dari FAISS sebelum memberikan keputusan akhir. Temuan ini sejalan dengan argumen bahwa struktur prompt yang terbimbing memiliki dampak riil terhadap presisi leksikal dan semantik jawaban [14]. CoT bertindak sebagai filter kognitif yang memastikan setiap klaim dalam jawaban didukung oleh bukti tekstual dalam PDF, sehingga meningkatkan skor Precision pada BERTScore secara signifikan dibandingkan pendekatan bebas.

### 4.3. Diskusi Etika Komunikasi AI

Keunggulan CoT dalam meningkatkan akurasi informasi bukan sekadar pencapaian teknis, melainkan sebuah tanggung jawab etis dalam komunikasi berbasis AI. Sebagaimana ditekankan dalam kajian etika komunikasi AI,

penyampaian pesan yang akurat dan transparan adalah kunci utama untuk menjaga kepercayaan publik terhadap teknologi cerdas [8].

Fenomena halusinasi pada LLM sering kali menjadi celah disinformasi yang menyesatkan pengguna, terutama mahasiswa yang menggunakan chatbot sebagai referensi utama layanan akademik. Dengan mengimplementasikan CoT, pengembang sistem secara tidak langsung menerapkan prinsip "integritas pesan", di mana AI didorong untuk bersikap jujur terhadap keterbatasan konteks dan tidak memberikan informasi fabrikasi. Hal ini mengurangi risiko kerugian administratif bagi mahasiswa akibat informasi jadwal atau aturan akademik yang keliru, sekaligus memperkuat nilai-nilai etika komunikasi dalam interaksi manusia-komputer.

#### 4.4 Implikasi Praktis bagi Pengembang Sistem

Temuan penelitian ini memberikan implikasi praktis yang nyata bagi institusi pendidikan tinggi. Bagi pengelola layanan informasi kampus, penggunaan chatbot tidak lagi cukup hanya mengandalkan model bahasa generatif standar seperti pada masa awal pengembangan berbasis AIML [21]. Diperlukan integrasi teknik prompting yang lebih canggih di atas arsitektur RAG untuk memastikan layanan tetap handal.

### 5. Kesimpulan

Penelitian ini telah berhasil melakukan analisis komparatif antara teknik Zero-Shot dan Chain-of-Thought (CoT) prompting dalam ekosistem Retrieval-Augmented Generation (RAG) berbasis dokumen PDF. Temuan utama menunjukkan bahwa teknik Chain-of-Thought (CoT) secara signifikan lebih efektif dalam meningkatkan presisi informasi dan memitigasi fenomena halusinasi pada LLM dibandingkan dengan teknik Zero-Shot. Secara kuantitatif, penggunaan CoT mampu meningkatkan validitas jawaban berdasarkan penilaian pakar hingga 37,14%, yang juga terverifikasi melalui peningkatan skor semantik pada metrik BERTScore. Keunggulan CoT terletak pada kemampuannya memaksa model untuk melakukan dekomposisi logika secara bertahap, sehingga model tidak hanya sekadar mencocokkan teks (string matching), tetapi benar-benar melakukan verifikasi terhadap konteks yang diambil dari basis pengetahuan. Hal ini menjawab tantangan etika komunikasi dalam penggunaan AI, di mana akurasi dan integritas informasi menjadi fondasi utama agar teknologi ini tidak menjadi sumber disinformasi akademik. Dengan demikian, penelitian ini membuktikan bahwa strategi prompting yang terstruktur adalah kunci dalam mengubah AI generatif dari sekadar mesin percakapan menjadi asisten akademik yang handal dan presisi. Meskipun hasil penelitian ini memberikan kontribusi nyata bagi pengembangan layanan informasi berbasis AI, terdapat beberapa keterbatasan yang dapat dikembangkan dalam riset selanjutnya: 1). Penelitian mendatang disarankan untuk menguji efektivitas CoT pada model LLM yang lebih variatif, termasuk model hibrida yang menggabungkan rule-based system untuk meningkatkan kontrol administratif. 2). Diperlukan studi lebih lanjut mengenai bagaimana penggunaan sistem berbasis CoT secara jangka panjang mempengaruhi keterampilan mahasiswa dalam merumuskan pertanyaan yang efektif dan mendalam. 3). Eksperimen berikutnya dapat mengintegrasikan teknik Multi-Agent Prompting atau optimasi pada tahap re-ranking vektor untuk melihat apakah peningkatan presisi dapat dicapai secara lebih efisien pada dataset yang lebih masif dan beragam.

### Referensi

- [1] S. Diaz, J. Budiman, and D. Kesuma, "Analisis Tren Penggunaan ChatGPT di Indonesia: Studi dengan Google Trends," *Buletin Sistem Informasi dan Teknologi Islam (BUSITI)*, vol. 6, no. 3, pp. 93–101, 2025, doi: 10.33096/busiti.v6i3.2810.
- [2] S. Pare, O. Kati, B. Bahrin, T. M. Tallulembang, and D. Moeis, "Efektivitas ChatGPT sebagai AI-Co-Researcher dalam Meningkatkan Kualitas Proposal Penelitian Mahasiswa," *RIGGS: Journal of Artificial Intelligence and Digital Business*, vol. 4, no. 4, pp. 13268–13275, Jan. 2026, doi: 10.31004/riggs.v4i4.5766.
- [3] E. T. S. Enjang, "Peran Chatgpt Dalam Meningkatkan Kemampuan Menulis Akademik Bahasa Indonesia: Sebuah Kajian Literatur Sistematis," *KUNKUN: Journal of Multidisciplinary Research*, vol. 2, no. 3, pp. 159–172, 2025, [Online]. Available: <https://ejournal.mediakunkun.com/index.php/kunkun/article/view/295>
- [4] A. Setiawan and U. K. Luthfiyani, "Penggunaan ChatGPT Untuk Pendidikan di Era Education 4.0: Usulan Inovasi Meningkatkan Keterampilan Menulis," *JURNAL PETISI (Pendidikan Teknologi Informasi)*, vol. 4, no. 1, pp. 49–58, 2023, [Online]. Available: <https://e-journal.unimudasorong.ac.id/index.php/jurnalpetisi/article/view/784>
- [5] A. R. Saraswati, V. A. Karmina, M. P. Efendi, Z. Candrakanti, and N. A. Rakhmawati, "Analisis Pengaruh ChatGPT Terhadap Tingkat Kemalasan Berpikir Mahasiswa ITS Dalam Proses Pengerjaan Tugas," *Jurnal Pendidikan, Bahasa Dan Budaya*, vol. 2, no. 4, pp. 40–48, 2023, doi: 10.55606/jpbb.v2i4.2223.
- [6] K. A. Tompunu, P. Demetria, R. S. Kinanty, and F. Fathoni, "ChatGPT sebagai alat bantu pencarian referensi: Analisis penggunaan oleh mahasiswa Sistem Informasi Unsri dalam menyusun tugas akhir," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 4, 2025, doi: 10.36040/jati.v9i4.14062.

- [7] L. Rahmawati, R. Yuliawati, A. Anisah, and H. Hermansyah, "Analisis pengaruh pemanfaatan ChatGPT terhadap keterampilan menyusun pertanyaan efektif pada mahasiswa Pendidikan Ekonomi," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 5, 2025, doi: 10.36040/jati.v9i5.14921.
- [8] C. W. Priastuty, M. S. Sugandi, and M. B. Srikandi, "PROMPT ENGINEERING DAN ETIKA KOMUNIKASI DALAM ERA KECERDASAN BUATAN: TANTANGAN DAN PELUANG," *Jurnal Ilmiah Dinamika Sosial*, vol. 9, no. 2, pp. 267–268, 2025, doi: 10.38043/jids.v9i2.6882.
- [9] R. D. Dermawan and H. Herdianto, "Meningkatkan Kinerja Output ChatGPT Melalui Teknik Prompt Engineering Yang Dapat Dikustomisasi," *Innovative: Journal Of Social Science Research*, vol. 4, no. 1, pp. 10646–10664, 2024, doi: 10.31004/innovative.v4i1.9067.
- [10] A. N. Annas, D. P. Lestari, R. Lestari, S. A. Hanim, and I. Malay, "Pelatihan teknis 'prompt engineering' dalam pemanfaatan AI chat GPT untuk meningkatkan kemampuan menulis artikel ilmiah mahasiswa baru," *ABDIRA: Jurnal Pengabdian Masyarakat*, vol. 5, no. 3, 2025, doi: 10.31004/abdira.v5i3.752.
- [11] M. A. Hasbi, R. Imanda, and M. F. Fauzan, "Implementasi Chatbot Berbasis Large Language Model Untuk Pencarian Skripsi Mahasiswa Terintegrasi dengan Whatsapp," *Arcitech: Journal of Computer Science and Artificial Intelligence*, vol. 5, no. 1, pp. 148–167, 2025, doi: 10.29240/arcitech.v5i1.13974.
- [12] M. A. Qaulan, Wahyuni, and P. Adytia, "Pengembangan Chatbot Berbasis AI untuk Mendukung Pelayanan Perpustakaan," *TEMATIK*, vol. 12, no. 1, pp. 23–30, 2025, doi: 10.38204/tematik.v12i1.2283.
- [13] I. Haromain, S. Munir, and A. Rahmah, "Analisa prompt engineering pada large language model dengan retrieval-augmented generation untuk informasi obat dan vitamin," *Indonesian Journal Computer Science*, vol. 4, no. 2, pp. 144–153, 2025, doi: 10.31294/ijcs.v4i2.10005.
- [14] M. A. Rizky, "Analisis efektivitas dua jenis gaya prompt dalam model LLM berbasis RAG," *Jurnal Komtika (Komputasi dan Informatika)*, vol. 9, no. 1, pp. 76–86, 2025, doi: 10.31603/komtika.v9i1.13488.
- [15] M. F. Gustafi, "Optimization of AI Usage in Learning Materials Application of Prompt Engineering Techniques for Learning Management Systems (LMS)," *JRST (Jurnal Riset Sains Dan Teknologi)*, vol. 9, no. 2, pp. 121–133, 2025, doi: 10.30595/jrst.v9i2.23393.
- [16] T. Herdi, D. A. H. Capah, Y. Jumaryadi, and B. Priambodo, "Penerapan Prompt Engineering Berbasis Kecerdasan Buatan untuk Perencanaan Menu Gizi Seimbang dan Olahraga Praktis di Kelurahan Duri Kepa Jakarta Barat," *JPMTT (Jurnal Pengabdian Masyarakat Teknologi Terbarukan)*, vol. 5, no. 2, pp. 23–38, 2025, doi: 10.54650/jpmtt.v5i2.586.
- [17] A. M. Siregar, A. Kurniawan, M. H. Uyun, and D. P. Gea, "Penggunaan ChatGPT untuk menyusun perangkat pembelajaran tingkat SMA," *Jurnal Bhakti Karya dan Inovatif*, vol. 5, no. 2, pp. 156–164, 2025, doi: 10.37278/bhaktikaryadaninovatif.v5i2.1326.
- [18] B. Albab and I. Nuryasin, "Pengembangan chatbot berbasis WhatsApp menggunakan Dialogflow dan NLP untuk layanan informasi toko fotokopi," *JIKO (Jurnal Informatika dan Komputer)*, vol. 9, no. 3, 2025, doi: 10.26798/jiko.v9i3.2046.
- [19] A. Alvin, R. Robet, and F. A. Tarigan, "Implementasi chatbot otomatis akademik berbasis web menggunakan LLM dan rule-based system studi kasus: STMIK Time," *JIKO (Jurnal Informatika dan Komputer)*, vol. 9, no. 3, pp. 639–649, 2025, doi: 10.26798/jiko.v9i3.2209.
- [20] C. Febrianty, T. P. Sari, and R. H. Syarafi, "Analisis dampak ChatGPT terhadap proses pembelajaran mahasiswa: Systematic literature review," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 1, pp. 949–961, 2025, doi: 10.36040/jati.v9i1.12513.
- [21] G. Guntoro, L. Costaner, and L. Lisnawita, "Aplikasi chatbot untuk layanan informasi dan akademik kampus berbasis Artificial Intelligence Markup Language (AIML)," *Digital Zone: Jurnal Teknologi Informasi Dan Komunikasi*, vol. 11, no. 2, pp. 291–300, 2020, doi: 10.31849/digitalzone.v11i2.5049.