JURNAL KHATULISTIWA INFORMATIKA
p-ISSN: 2339-1928 e-ISSN: 2579-633X

VOL. 14 NO. 1 Juni 2026
Halaman : 6 - 13

RANCANG BANGUN SISTEM DETEKSI HOAKS BERBASIS AGENTIC AI OTONOM

Rabiatul Adwiya [1]; Achmad Arif Munaji [2]; Airways Parlindungan Siahaan [3]

Program Studi Teknik Komputer, Fakultas Teknik[1][2];
Program Studi Teknik Industri, Fakultas Teknik[3]
Institut Teknologi dan Sains Nahdlatul Ulama Kalimantan (ITSNUKA)
r.adwiya11@gmail.com ; arief.munaji65@gmail.com; nitaways@gmail.com

INFO ARTIKEL

Diajukan :
18-04-2026

Diterima :
06-05-2026

Diterbitkan :
01-06-2026

Kata Kunci :
agentic ai, deteksi hoaks, large language models, retrieval augmented generation, verifikasi fakta

INTISARI

Penyebaran hoaks pada media sosial di Indonesia berlangsung cepat dan masif sehingga sulit ditangani secara manual. Penelitian terdahulu pada umumnya berhenti pada klasifikasi teks atau verifikasi fakta berbasis dataset statis, sehingga belum mampu menelusuri bukti secara mandiri dan real-time. Penelitian ini bertujuan merancang dan membangun sistem verifikasi fakta otonom yang mengintegrasikan Agentic AI dan Large Language Models (LLM) dengan pendekatan Retrieval-Augmented Generation (RAG) untuk mendeteksi hoaks. Metode penelitian menggunakan pendekatan Research and Development dengan arsitektur agen berbasis kerangka kerja Reason and Act (ReAct), yang menggabungkan penalaran, pemanggilan alat pencarian eksternal, dan penarikan kesimpulan. Sistem diuji menggunakan 300 klaim berbahasa Indonesia yang dikumpulkan dari kanal pemeriksa fakta terverifikasi. Hasil pengujian menunjukkan sistem mencapai akurasi 89,3%, presisi 88,7%, recall 90,1%, dan F1-score 89,4%, lebih tinggi dibandingkan model klasifikasi konvensional. Sistem juga mampu menyajikan bukti dan alasan keputusan secara transparan. Penelitian ini berkontribusi menyediakan kerangka kerja verifikasi fakta otonom yang dapat diterapkan pada konteks media sosial Indonesia.

I. PENDAHULUAN

Transformasi digital yang didorong oleh perkembangan teknologi informasi dan media sosial telah mengubah secara fundamental pola produksi, distribusi, dan konsumsi informasi dalam masyarakat. Akses yang semakin cepat dan luas terhadap berbagai sumber informasi memberikan manfaat signifikan dalam mempercepat penyebaran pengetahuan. Namun, pada saat yang sama, kondisi tersebut juga meningkatkan risiko penyebaran informasi palsu (hoaks) dalam skala yang lebih besar dan dengan kecepatan yang lebih tinggi. Fenomena ini menjadi tantangan serius dalam ekosistem informasi digital, termasuk di Indonesia. Berdasarkan data Kementerian Komunikasi dan Digital, sepanjang tahun 2024 teridentifikasi sebanyak 1.923 konten hoaks dengan jumlah temuan tertinggi terjadi pada Oktober 2024 yang mencapai 215 konten (Kementerian Komunikasi dan Digital, 2025). Selanjutnya, pada periode 20 Oktober 2024 hingga 6 Desember 2025, tercatat 1.890 isu hoaks, sementara jumlah keseluruhan konten negatif yang berhasil ditangani mencapai 3,38 juta konten. Meskipun demikian, pemerintah mengakui bahwa masih terdapat sejumlah besar hoaks yang belum teridentifikasi dan tertangani secara efektif (Nugraha, 2025). Dari berbagai kategori yang ditemukan, hoaks bermuatan penipuan mendominasi dengan 890 kasus, diikuti oleh hoaks yang berkaitan dengan isu politik dan pemerintahan (Aprilia, 2025). Temuan tersebut mengindikasikan bahwa volume informasi yang beredar di ruang digital telah melampaui kapasitas pendekatan verifikasi manual yang selama ini dilakukan oleh organisasi pemeriksa fakta, sehingga diperlukan solusi yang lebih adaptif, otomatis, dan skalabel.

Berbagai pendekatan berbasis kecerdasan buatan telah dikembangkan untuk mengatasi permasalahan deteksi informasi palsu. Salah satu pendekatan yang banyak digunakan adalah klasifikasi teks berbasis deep learning. Kaliyar et al. (2021) memperkenalkan FakeBERT, sebuah model yang memanfaatkan arsitektur transformer untuk mendeteksi

VOL. 14 NO. 1 Juni 2026
Halaman : 6 - 13

berita palsu dan menunjukkan performa yang lebih baik dibandingkan metode machine learning tradisional. Meskipun demikian, pendekatan tersebut berfokus pada identifikasi pola linguistik dalam teks dan belum mampu melakukan verifikasi terhadap fakta yang mendasari suatu klaim. Dengan kata lain, sistem hanya menghasilkan keputusan klasifikasi tanpa menyediakan bukti eksternal maupun penjelasan yang mendukung alasan suatu informasi dikategorikan sebagai benar atau salah.

Dalam konteks verifikasi fakta otomatis, Thorne et al. (2018) mengembangkan FEVER (*Fact* Extraction *and VERification*), sebuah benchmark yang memungkinkan sistem melakukan ekstraksi bukti sekaligus menentukan apakah suatu klaim didukung (supported), ditolak (refuted), atau tidak memiliki informasi yang memadai (not enough information). Kontribusi FEVER menjadi tonggak penting dalam pengembangan penelitian fact-checking karena memperluas fokus dari sekadar klasifikasi menuju proses pembuktian berbasis evidensi. Namun demikian, pendekatan ini masih memiliki keterbatasan karena bergantung pada korpus data yang bersifat statis. Akibatnya, kemampuan sistem dalam memverifikasi informasi yang berkembang secara dinamis dan real-time, khususnya di media sosial, masih relatif terbatas.

Perkembangan terbaru dalam bidang kecerdasan buatan menunjukkan potensi signifikan dari Large Language Models (LLM) dalam mendukung tugas-tugas yang memerlukan pemahaman bahasa dan penalaran kompleks. Dalam LLM memiliki kemampuan penalaran, pemahaman konteks, dan pemrosesan bahasa yang lebih unggul dibandingkan generasi sebelumnya, sehingga mulai digunakan dalam berbagai aplikasi yang berkaitan dengan analisis dan verifikasi informasi. Meskipun demikian, LLM masih menghadapi tantangan berupa kemungkinan menghasilkan informasi yang tidak sesuai fakta (hallucination) serta belum memiliki mekanisme bawaan untuk memvalidasi sumber informasi yang digunakan dalam proses generasi jawaban. Untuk mengatasi keterbatasan tersebut, Lewis et al. (2020) memperkenalkan Retrieval-Augmented Generation (RAG), yaitu pendekatan yang mengintegrasikan kemampuan pencarian informasi (retrieval) dengan kemampuan generatif model bahasa. Melalui mekanisme ini, model dapat mengakses sumber pengetahuan eksternal yang lebih mutakhir sehingga meningkatkan akurasi dan relevansi jawaban yang dihasilkan. Walaupun demikian, pemanfaatan RAG dalam konteks deteksi hoaks dan verifikasi fakta pada media sosial masih belum banyak dieksplorasi secara mendalam.

Seiring dengan perkembangan tersebut, perhatian penelitian mulai bergeser menuju paradigma Agentic AI, yaitu sistem kecerdasan buatan yang tidak hanya menghasilkan respons, tetapi juga mampu melakukan perencanaan, penalaran, dan pengambilan tindakan secara otonom. Yao et al. (2023) mengusulkan kerangka kerja ReAct yang mengintegrasikan proses reasoning dan acting dalam penyelesaian tugas yang kompleks. Selanjutnya, Schick et al. (2023) mengembangkan Toolformer yang memungkinkan model bahasa memanfaatkan berbagai alat eksternal untuk meningkatkan kualitas pengambilan keputusan. Xi et al. (2023) juga menegaskan bahwa agen berbasis LLM memiliki potensi untuk menjalankan proses perencanaan, penalaran multilangkah, serta pengambilan keputusan secara mandiri dalam lingkungan yang dinamis. Integrasi kemampuan tersebut membuka peluang baru bagi pengembangan sistem verifikasi fakta yang lebih adaptif, transparan, dan responsif terhadap perubahan informasi secara real-time.

Meskipun berbagai kemajuan telah dicapai, kajian literatur menunjukkan masih terdapat sejumlah kesenjangan penelitian yang belum teratasi secara memadai. Pertama, sebagian besar penelitian deteksi hoaks masih berorientasi pada klasifikasi teks dan belum mengintegrasikan mekanisme verifikasi fakta yang dilakukan secara otonom. Kedua, sebagian besar sistem fact-checking masih mengandalkan sumber data yang bersifat statis sehingga kurang efektif dalam menghadapi dinamika informasi yang berkembang cepat di media sosial. Ketiga, penelitian yang mengombinasikan Agentic AI, Large Language Models, dan Retrieval-Augmented Generation dalam satu kerangka kerja terpadu untuk mendukung proses pencarian bukti, penalaran, verifikasi fakta, dan pengambilan keputusan otomatis masih relatif terbatas, khususnya dalam konteks media sosial berbahasa Indonesia.

Berdasarkan kesenjangan tersebut, penelitian ini bertujuan untuk merancang dan

VOL. 14 NO. 1 Juni 2026
Halaman : 6 - 13

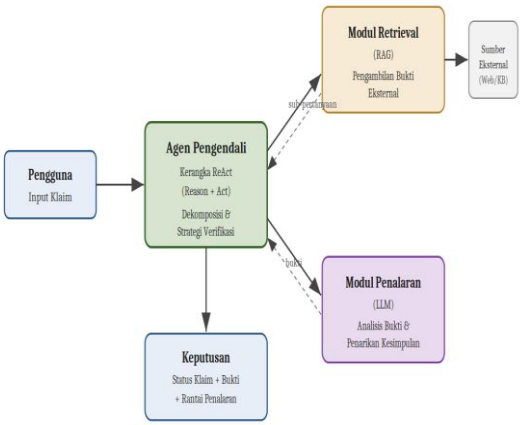
mengembangkan sistem verifikasi fakta otonom yang mengintegrasikan Agentic AI dan Large Language Models dengan pendekatan Retrieval-Augmented Generation guna mendeteksi hoaks pada media sosial. Selain itu, penelitian ini mengevaluasi kinerja sistem menggunakan klaim berbahasa Indonesia sebagai representasi konteks lokal yang memiliki karakteristik linguistik dan dinamika informasi yang khas. Kontribusi utama penelitian ini terletak pada tiga aspek. Pertama, penelitian ini mengintegrasikan Agentic AI, LLM, dan RAG ke dalam satu kerangka kerja verifikasi fakta yang terpadu, berbeda dari penelitian sebelumnya yang umumnya mengembangkan ketiga komponen tersebut secara terpisah. Kedua, penelitian ini mengimplementasikan kerangka kerja ReAct sebagai mekanisme reasoning-action untuk mendukung pencarian bukti secara dinamis dari berbagai sumber informasi, sehingga tidak bergantung pada dataset statis sebagaimana pendekatan FEVER. Ketiga, sistem yang diusulkan dirancang untuk menghasilkan keputusan yang lebih transparan melalui penyajian bukti pendukung dan jejak penalaran (reasoning chain) yang mendasari setiap hasil verifikasi. Dengan demikian, penelitian ini tidak hanya berkontribusi pada pengembangan metode deteksi hoaks berbasis kecerdasan buatan, tetapi juga memperluas penerapannya pada konteks media sosial berbahasa Indonesia yang hingga saat ini masih relatif kurang mendapat perhatian dalam literatur internasional.

II. BAHAN DAN METODE

Penelitian ini menerapkan pendekatan Research and Development (R&D) yang berorientasi pada pengembangan dan evaluasi sistem verifikasi fakta otonom berbasis kecerdasan buatan. Pendekatan ini dipilih karena tujuan utama penelitian tidak hanya menghasilkan pengetahuan baru, tetapi juga menghasilkan artefak teknologi berupa sistem yang dapat digunakan untuk mendukung proses verifikasi fakta secara otomatis. Secara umum, penelitian dilaksanakan melalui empat tahapan utama, yaitu analisis kebutuhan, perancangan arsitektur sistem, implementasi model, dan evaluasi kinerja. Tahap analisis kebutuhan difokuskan pada identifikasi karakteristik permasalahan deteksi hoaks di media sosial serta kebutuhan fungsional dan nonfungsional sistem. Tahap berikutnya mencakup perancangan arsitektur agen cerdas yang mengintegrasikan kemampuan penalaran, pencarian informasi, dan pengambilan keputusan. Selanjutnya, sistem diimplementasikan sesuai rancangan yang telah ditetapkan dan diakhiri dengan evaluasi untuk mengukur efektivitas kinerja sistem yang dikembangkan.

Data penelitian terdiri atas 300 klaim berbahasa Indonesia yang diperoleh dari berbagai kanal pemeriksa fakta terverifikasi dan unggahan media sosial yang telah melalui proses verifikasi oleh pemeriksa fakta profesional. Setiap klaim dilengkapi dengan label kebenaran yang digunakan sebagai ground truth dalam proses pengujian dan evaluasi sistem. Pemanfaatan data yang telah diverifikasi oleh lembaga pemeriksa fakta bertujuan untuk memastikan reliabilitas acuan yang digunakan dalam mengukur performa sistem.

Arsitektur sistem yang diusulkan dibangun berdasarkan kerangka kerja ReAct (*Reasoning and Acting*) yang mengintegrasikan proses penalaran dan tindakan dalam satu siklus pengambilan keputusan. Pada arsitektur ini, agen utama berfungsi sebagai pengendali (*controller agent*) yang menerima klaim sebagai masukan, melakukan dekomposisi klaim menjadi sejumlah pertanyaan yang lebih spesifik, serta menentukan strategi verifikasi yang sesuai. Untuk memperoleh informasi pendukung, sistem memanfaatkan mekanisme *Retrieval-Augmented Generation* (RAG) yang memungkinkan proses pencarian bukti dari berbagai sumber eksternal, termasuk mesin pencari dan basis pengetahuan digital. Bukti yang berhasil dikumpulkan kemudian dianalisis oleh *Large Language Model* (LLM) guna mengevaluasi tingkat dukungan atau penolakan terhadap klaim yang sedang diperiksa. Setiap hasil verifikasi disertai dengan bukti pendukung dan jejak penalaran (*reasoning chain*), sehingga proses pengambilan keputusan dapat ditelusuri secara transparan dan lebih mudah dievaluasi.



Sumber: Hasil perancangan penulis (2026)

Gambar 1. Rancangan Arsitektur Sistem Verifikasi Fakta Otonom

Sebagaimana ditunjukkan pada Gambar 1, agen pengendali menempati posisi sentral dalam arsitektur sistem dan bertanggung jawab mengoordinasikan interaksi antara modul pencarian informasi (*retrieval module*) dan modul penalaran (*reasoning module*). Aliran proses dimulai ketika pengguna memasukkan suatu klaim ke dalam sistem. Selanjutnya, agen melakukan siklus penalaran dan tindakan secara berulang hingga diperoleh bukti yang dinilai memadai untuk mendukung proses verifikasi. Hasil akhir berupa status klaim kemudian dihasilkan bersama bukti relevan dan argumentasi yang menjelaskan dasar pengambilan keputusan.

Untuk merealisasikan mekanisme tersebut, logika operasional sistem diformalkan ke dalam algoritma verifikasi fakta berbasis ReAct yang bekerja secara iteratif. Algoritma menerima klaim sebagai input dan menghasilkan keluaran berupa status klaim, kumpulan bukti pendukung, serta penjelasan yang mendasari keputusan sistem. Proses diawali dengan dekomposisi klaim menjadi beberapa sub-pertanyaan yang digunakan sebagai dasar pencarian informasi. Pada setiap iterasi, LLM melakukan proses penalaran terhadap klaim dan bukti yang tersedia, kemudian menentukan tindakan yang perlu dilakukan. Apabila bukti yang ada belum mencukupi, sistem akan mengaktifkan modul retrieval untuk memperoleh informasi tambahan dari sumber eksternal melalui mekanisme RAG. Bukti yang diperoleh selanjutnya diperingkat berdasarkan tingkat relevansinya dan digunakan kembali dalam proses penalaran berikutnya. Siklus ini berlangsung hingga tingkat keyakinan sistem mencapai ambang batas yang telah ditentukan atau jumlah iterasi maksimum terpenuhi. Setelah proses iteratif berakhir, sistem menetapkan status klaim ke dalam salah satu kategori, yaitu Didukung (Supported), Ditolak (Refuted), atau Not Enough Information (NEI), serta menghasilkan penjelasan yang mendasari keputusan tersebut.

Logika kerja sistem diformalkan ke dalam algoritma rancang bangun yang menerapkan kerangka kerja ReAct secara iteratif. Algoritma menerima sebuah klaim sebagai masukan dan menghasilkan keputusan berupa status klaim beserta bukti dan alasan. Rancangan algoritma utama sistem disajikan pada Algoritma 1.

Algoritma 1. Verifikasi Fakta Otonom Berbasis ReAct (Agentic AI + LLM + RAG)

```
INPUT  : klaim C, batas iterasi maks N, ambang keyakinan τ
OUTPUT : status S ∈ {Didukung, Ditolak, NEI}, bukti E, alasan R

1. E ← ∅ ; i ← 0          // inisialisasi bukti dan iterasi
2. Q ← Dekomposisi(C)     // pecah klaim jadi sub-pertanyaan
3. WHILE i < N DO
   4.  thought ← LLM.Reason(C, E)      // penalaran (Reason)
   5.  action ← LLM.PlanAction(thought) // tentukan tindakan (Act)
   6.  IF action = SEARCH THEN
      7.  d ← Retrieve_RAG(Q)   // ambil bukti dari sumber eksternal
      8.  E ← E ∪ Rank(d)       // gabung & ranking bukti relevan
   9.  END IF
   10. conf ← LLM.Assess(C, E)  // nilai kecukupan & keyakinan bukti
   11. IF conf ≥ τ THEN BREAK   // bukti cukup, hentikan iterasi
   12. i ← i + 1
13. END WHILE
14. IF E = ∅ OR conf < τ THEN
   15. S ← NEI               // Not Enough Info
16. ELSE
   17. (S, R) ← LLM.Verify(C, E)  // klasifikasi + alasan
18. END IF
19. RETURN (S, E, R)
```

Sumber: Hasil perancangan penulis (2026)

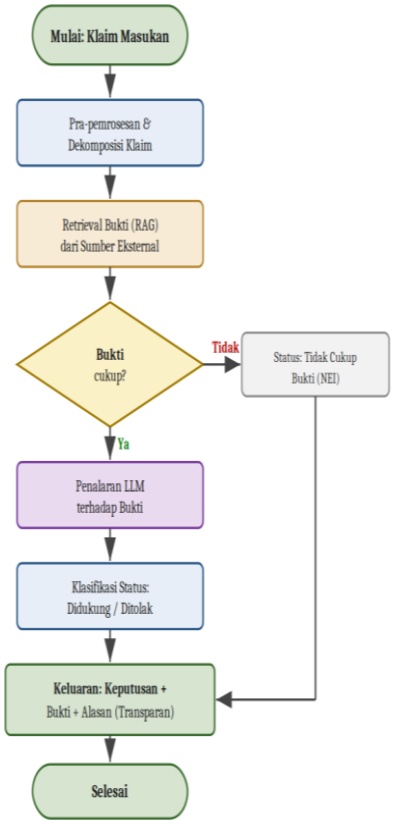
Pengumpulan data dilakukan melalui teknik dokumentasi terhadap klaim dan label

verifikasi yang tersedia pada sumber pemeriksa fakta. Seluruh data yang terkumpul kemudian digunakan sebagai bahan pengujian untuk menilai kemampuan sistem dalam melakukan klasifikasi dan verifikasi fakta secara otomatis. Analisis data dilakukan secara kuantitatif dengan menggunakan metrik evaluasi yang umum digunakan dalam penelitian klasifikasi, yaitu *accuracy, precision, recall, dan F1-score.* Keempat metrik tersebut digunakan untuk memberikan gambaran yang komprehensif mengenai kemampuan sistem dalam mengidentifikasi klaim secara tepat, mengurangi kesalahan klasifikasi, serta menjaga keseimbangan antara sensitivitas dan ketepatan prediksi.

Untuk menilai efektivitas pendekatan yang diusulkan, performa sistem dibandingkan dengan model klasifikasi konvensional berbasis **Bidirectional Encoder Representations** from Transformers (BERT) yang digunakan sebagai baseline model. Evaluasi dilakukan dengan membandingkan hasil prediksi sistem terhadap label acuan (ground truth) yang telah ditetapkan sebelumnya. Melalui proses tersebut, penelitian ini dapat mengukur sejauh mana integrasi Agentic AI, Large Language Models, dan Retrieval-Augmented Generation mampu meningkatkan kemampuan verifikasi fakta dibandingkan pendekatan klasifikasi teks yang hanya mengandalkan representasi linguistik tanpa mekanisme pencarian bukti eksternal dan penalaran otonom.

III. HASIL DAN PEMBAHASAN

Sistem verifikasi fakta otonom yang dikembangkan dalam penelitian ini dibangun melalui integrasi tiga komponen utama, yaitu agen pengendali (controller agent), modul retrieval berbasis Retrieval-Augmented Generation (RAG), dan modul penalaran berbasis Large Language Model (LLM). Integrasi ketiga komponen tersebut dirancang untuk menghasilkan mekanisme verifikasi yang tidak hanya mampu mengklasifikasikan suatu klaim, tetapi juga melakukan pencarian bukti, mengevaluasi validitas informasi, serta memberikan penjelasan yang mendasari setiap keputusan yang dihasilkan. Proses verifikasi diawali dengan penerimaan klaim dari pengguna, kemudian agen melakukan dekomposisi klaim menjadi sejumlah sub-pertanyaan yang lebih spesifik dan dapat diverifikasi. Sub-pertanyaan tersebut selanjutnya digunakan sebagai dasar pencarian informasi pada sumber eksternal melalui modul retrieval. Bukti yang berhasil dikumpulkan kemudian dianalisis menggunakan kemampuan penalaran LLM untuk menghasilkan keputusan akhir beserta argumentasi yang mendukungnya. Alur operasional sistem secara keseluruhan disajikan pada Gambar 2.



Sumber: Hasil perancangan penulis (2026)

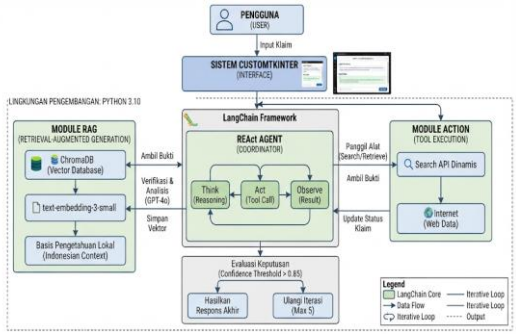
Gambar 2. Diagram Alur Kerja Sistem Verifikasi Fakta Otonom

Sebagaimana ditunjukkan pada Gambar 2, proses verifikasi dimulai dari tahap penerimaan klaim yang kemudian melalui proses pra-pemrosesan dan dekomposisi semantik. Tahap ini bertujuan untuk mengidentifikasi elemen-elemen penting dalam klaim sehingga proses pencarian bukti dapat dilakukan secara lebih terarah. Selanjutnya, sistem mengaktifkan mekanisme RAG untuk memperoleh informasi yang relevan dari berbagai sumber eksternal. Bukti yang diperoleh kemudian dievaluasi berdasarkan tingkat kecukupan dan relevansinya terhadap klaim yang sedang diperiksa. Apabila bukti yang tersedia belum memenuhi tingkat keyakinan yang ditetapkan, sistem akan mengembalikan status Not Enough Information (NEI). Sebaliknya, apabila bukti dinilai memadai, LLM akan melakukan proses

penalaran untuk menentukan apakah klaim tersebut didukung atau ditolak oleh bukti yang tersedia. Setiap keputusan disertai dengan bukti pendukung dan penjelasan logis yang mendasarinya, sehingga proses verifikasi tidak hanya menghasilkan keluaran yang akurat, tetapi juga transparan dan dapat ditelusuri kembali.

3.1 Implementasi Sistem

Penelitian ini menghasilkan artefak berupa sistem verifikasi fakta otonom yang dikembangkan menggunakan ekosistem Python 3.10. Lingkungan pengembangan sistem dibangun di atas arsitektur *agent-based* menggunakan kerangka kerja LangChain untuk mengoordinasikan interaksi antara *Large Language Model* (LLM) dan alat eksternal.



Gambar 3. Diagram Arsitektur Sistem Verifikasi Fakta Berbasis Agentic AI (ReAct + RAG)

Model LLM yang diintegrasikan sebagai mesin penalaran utama adalah GPT-4o, yang dipilih karena kemampuan penalaran multilangkah dan pemahaman konteks bahasa Indonesia yang superior. Untuk mekanisme *Retrieval-Augmented Generation* (RAG), sistem menggunakan *vector database* ChromaDB sebagai penyimpanan basis pengetahuan lokal, dengan *embedding model* text-embedding-3-small untuk mengonversi klaim dan dokumen pendukung ke dalam representasi vektor. Proses pencarian bukti eksternal dilakukan melalui *API Search* yang terintegrasi secara dinamis untuk mengambil data terkini dari web.

Konfigurasi parameter sistem ditetapkan sebagai berikut: *confidence threshold* (τ) sebesar 0,85 untuk menentukan apakah bukti dianggap cukup atau sistem harus melanjutkan iterasi, serta batas maksimum iterasi (N) sebanyak 5 siklus untuk menjaga efisiensi waktu respons. Antarmuka sistem dikembangkan menggunakan *library* CustomTkinter, yang menyajikan *dashboard* verifikasi bagi pengguna untuk memasukkan klaim, melihat jejak penalaran (*reasoning chain*) secara *real-time*, serta menampilkan bukti pendukung yang relevan beserta tautan sumbernya. Implementasi ini memastikan transparansi proses pengambilan keputusan, di mana pengguna dapat menelusuri bagaimana sistem menarik kesimpulan berdasarkan bukti-bukti yang ditemukan.

3.2 Sintesis Penelitian Terdahulu

Untuk mengidentifikasi posisi dan kontribusi penelitian ini dalam lanskap penelitian yang ada, dilakukan sintesis terhadap sejumlah studi terdahulu yang relevan. Hasil sintesis menunjukkan bahwa penelitian terkait deteksi hoaks dan verifikasi fakta masih didominasi oleh pendekatan klasifikasi teks atau verifikasi berbasis korpus data statis. Ringkasan hasil sintesis disajikan pada Tabel 1.

Tabel 1. Sintesis Penelitian Terdahulu

Peneliti	Fokus	Kelemahan
Kaliyar et al. (2021)	Deteksi hoaks dengan BERT	Hanya klasifikasi teks
Thorne et al. (2018)	Verifikasi fakta (FEVER)	Dataset statis
Lewis et al. (2020)	Retrieval-Augmented Generation	Belum spesifik hoaks
Yao et al. (2023)	ReAct (Reason + Act)	Belum fokus media sosial
Schick et al. (2023)	Toolformer	Belum untuk fact-checking
Xi et al. (2023)	Agentic AI	Masih konseptual

Sumber: Hasil analisis penulis (2026)

Berdasarkan sintesis tersebut, dapat disimpulkan bahwa belum terdapat penelitian yang secara komprehensif mengintegrasikan *Agentic AI, Large Language Models,* dan *Retrieval-Augmented Generation* ke dalam satu kerangka kerja verifikasi fakta otonom yang secara khusus ditujukan untuk menangani dinamika informasi pada media sosial berbahasa Indonesia. Oleh karena itu, penelitian ini berupaya mengisi kesenjangan

tersebut melalui pengembangan sistem yang menggabungkan kemampuan pencarian bukti, penalaran berbasis LLM, serta pengambilan keputusan otonom dalam satu arsitektur terpadu.

3.3 Hasil Pengujian Kinerja Sistem

Evaluasi kinerja dilakukan menggunakan 300 klaim berbahasa Indonesia yang telah diverifikasi oleh pemeriksa fakta profesional dan digunakan sebagai *ground truth*. Hasil pengujian menunjukkan bahwa sistem yang diusulkan menghasilkan performa yang lebih baik dibandingkan model klasifikasi berbasis BERT yang digunakan sebagai baseline. Ringkasan hasil evaluasi disajikan pada Tabel 2.

Tabel 2. Perbandingan Kinerja Sistem Usulan dan Baseline

No	Metrik Evaluasi	Sistem Usulan	Baseline (BERT)
1	Akurasi	89,3%	82,5%
2	Presisi	88,7%	81,9%
3	Recall	90,1%	80,4%
4	F1-Score	89,4%	81,1%

Sumber: Hasil pengujian penulis (2026)

Sistem usulan mencapai akurasi 89,3%, presisi 88,7%, recall 90,1%, dan F1-score 89,4%, sedangkan baseline berbasis BERT memperoleh akurasi 82,5% dengan F1-score 81,1%. Peningkatan kinerja ini sejalan dengan temuan Lewis et al. (2020) bahwa pendekatan RAG mampu meningkatkan akurasi jawaban karena memanfaatkan sumber pengetahuan eksternal. Selain itu, integrasi kerangka kerja ReAct sebagaimana diusulkan Yao et al. (2023) memungkinkan sistem melakukan penalaran dan pencarian bukti secara simultan sehingga keputusan yang dihasilkan lebih akurat.

Keunggulan utama sistem ini dibandingkan pendekatan Kaliyar et al. (2021) terletak pada kemampuannya menyajikan bukti dan alasan keputusan secara transparan, tidak sekadar memberikan label klasifikasi. Berbeda dengan pendekatan FEVER oleh Thorne et al. (2018) yang bergantung pada dataset statis, sistem ini mampu mengambil bukti secara dinamis dari sumber eksternal sehingga lebih sesuai untuk menangani informasi yang berkembang di media sosial. Kemampuan ini diperoleh berkat pemanfaatan alat eksternal yang konsepnya diperkenalkan oleh Schick et al. (2023) serta paradigma agen otonom yang dikaji Xi et al. (2023).

Meskipun demikian, sistem masih memiliki keterbatasan. Kualitas keputusan sangat bergantung pada ketersediaan dan kredibilitas bukti yang dapat diambil dari sumber eksternal. Pada klaim yang sangat baru atau belum diliput oleh sumber tepercaya, sistem cenderung mengembalikan status tidak cukup bukti. Selain itu, potensi hallucination dari LLM yang masih perlu diantisipasi melalui mekanisme verifikasi silang terhadap bukti.

IV. KESIMPULAN

Penelitian ini berhasil merancang dan membangun sebuah artefak sistem verifikasi fakta otonom berbasis *Agentic AI* yang dirancang khusus untuk memitigasi penyebaran hoaks di media sosial Indonesia. Produk akhir yang dihasilkan adalah aplikasi sistem cerdas yang mampu mengintegrasikan komponen penalaran (LLM), pengambilan tindakan (ReAct), dan pencarian bukti dinamis (RAG) ke dalam satu alur kerja terpadu.

Sistem ini berfungsi sebagai asisten verifikasi yang mampu melakukan dekomposisi klaim secara mandiri, melakukan penelusuran bukti dari sumber eksternal tepercaya, melakukan evaluasi validitas melalui penalaran berbasis bukti, dan menyajikan keputusan akhir berupa status klaim (Didukung, Ditolak, atau NEI) beserta penjelasan logis yang transparan kepada pengguna. Hasil pengujian menunjukkan bahwa sistem yang dibangun mencapai akurasi 89,3% dan F1-score 89,4%, yang membuktikan bahwa integrasi kerangka kerja ReAct dengan mekanisme RAG mampu memberikan hasil yang lebih efektif dibandingkan model klasifikasi teks konvensional. Kontribusi utama penelitian ini adalah tersedianya artefak sistem yang adaptif terhadap dinamika informasi digital, yang dapat diimplementasikan oleh organisasi pemeriksa fakta atau pengembang platform media sosial untuk meningkatkan skalabilitas dan kecepatan verifikasi informasi di Indonesia.

V. REFERENSI

Aprilia, S. (2025). *Hati-hati! Ribuan konten hoaks teridentifikasi sepanjang 2024*. GoodStats. https://goodstats.id/article/hati-hati-ribuan-konten-hoaks-diidentifikasi-komdigi-sepanjang-2024-6zBQv

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT*

2019, 4171–4186. https://doi.org/10.18653/v1/N19-1423

Guo, Z., Jin, R., Liu, C., Huang, Y., Shi, D., Yu, B., ... Lu, W. (2024). Retrieval-Augmented Generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*.

Kaliyar, R. K., Goswami, A., Narang, P., & Sinha, S. (2021). FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. *Multimedia Tools and Applications, 80*(8), 11765–11788. https://doi.org/10.1007/s11042-020-10183-2

Kementerian Komunikasi dan Digital. (2025). *Komdigi identifikasi 1.923 konten hoaks sepanjang tahun 2024*. https://www.komdigi.go.id/berita/siaran-pers/detail/komdigi-identifikasi-1923-konten-hoaks-sepanjang-tahun-2024

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... Riedel, S. (2020). Retrieval-Augmented Generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems, 33*, 9459–9474.

Nugraha, F. A. (2025). *Kemkomdigi identifikasi 1.890 konten hoaks dalam satu tahun terakhir*. ANTARA News. https://www.antaranews.com/berita/5293072/kemkomdigi-identifikasi-1890-konten-hoaks-dalam-satu-tahun-terakhir

Schick, T., Dwivedi-Yu, J., Dessì, R., Raileanu, R., Lomeli, M., Hambro, E., ... Uszkoreit, J. (2023). Toolformer: Language models can teach themselves to use tools. *Advances in Neural Information Processing Systems, 36*. https://arxiv.org/abs/2302.04761

Thorne, J., Vlachos, A., Christodoulopoulos, C., & Mittal, A. (2018). FEVER: A large-scale dataset for fact extraction and verification. *Proceedings of NAACL-HLT 2018*, 809–819. https://doi.org/10.18653/v1/N18-1074

Xi, Z., Chen, W., Guo, X., He, W., Wang, Y., Zhang, H., ... Zhang, W. (2023). *The rise and potential of large language model based agents: A survey*. arXiv:2309.07864. https://arxiv.org/abs/2309.07864

Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2023). ReAct: Synergizing reasoning and acting in language models. *International Conference on Learning Representations (ICLR)*. https://arxiv.org/abs/2210.03629