

PENGEMBANGAN *CHATBOT* AI INFORMASI AKADEMIK BERBASIS WHATSAPP MENGGUNAKAN *LOW-CODE AUTOMATION* N8N DI UNIVERSITAS PANCA SAKTI BEKASI

Bramasto Daru Satrio^{*1}, Parman Suparman², Afdhol Syamsumar³

Sistem Informasi, Universitas Panca Sakti Bekasi

Email: bramasto56964@gmail.com¹, phie.suparman@gmail.com², ankasaafdholl234@gmail.com³

Abstrak

Penelitian ini bertujuan mengembangkan chatbot AI informasi akademik berbasis WhatsApp menggunakan platform *low-code automation* n8n di Universitas Panca Sakti Bekasi. Penelitian dilatarbelakangi oleh keterbatasan jam operasional staf administrasi, kompleksitas akses informasi melalui website, serta belum tersedianya layanan WhatsApp dengan respons otomatis. Penelitian menggunakan model pengembangan 4D (*Define, Design, Develop, Disseminate*) dengan arsitektur *Retrieval-Augmented Generation* (RAG), basis data vektor Pinecone, model *embedding* nomic-embed-text:latest, dan LLM openai/gpt-oss-20b. Pengujian dilakukan menggunakan metode *Black-Box Testing* terhadap sembilan skenario. Hasil penelitian menunjukkan bahwa n8n mampu mengintegrasikan berbagai layanan AI untuk membangun chatbot informasi akademik yang responsif. Penerapan RAG menghasilkan respons yang akurat, kontekstual, dan relevan serta membantu meminimalkan risiko halusinasi AI. Dari sembilan skenario pengujian, delapan berhasil dengan tingkat keberhasilan **88,9%**. Chatbot mampu menyediakan informasi program studi, FAQ pendaftaran, jadwal perkuliahan, dan informasi umum Universitas Panca Sakti Bekasi secara otomatis melalui WhatsApp. Sistem ini diharapkan meningkatkan efisiensi layanan informasi akademik dan mengatasi keterbatasan jam operasional staf administrasi.

Kata kunci: *Chatbot AI, WhatsApp, n8n, Automation, Low-code, Universitas Panca Sakti Bekasi*

DEVELOPMENT OF A WHATSAPP-BASED ACADEMIC INFORMATION AI CHATBOT USING N8N LOW-CODE AUTOMATION AT UNIVERSITAS PANCA SAKTI BEKASI

Abstract

This study aims to develop a WhatsApp-based AI chatbot for academic information at Universitas Panca Sakti Bekasi using the low-code automation platform n8n. The research was motivated by the limited operating hours of administrative staff, the complexity of accessing information via the website, and the absence of an automated WhatsApp response service. The study employed the 4D development model (Define, Design, Develop, Disseminate) and utilized a Retrieval-Augmented Generation (RAG) architecture, the Pinecone vector database, the nomic-embed-text:latest embedding model, and the openai/gpt-oss-20b LLM. Testing was conducted using the Black-Box Testing method across nine scenarios. The results demonstrate that n8n effectively integrates various AI services to build a responsive academic information chatbot. The implementation of RAG yielded accurate, contextual, and relevant responses while helping to minimize the risk of AI hallucinations. Out of the nine test scenarios, eight were successful, resulting in a success rate of 88.9%. The chatbot is capable of automatically providing information regarding study programs, registration FAQs, class schedules, and general details about Universitas Panca Sakti Bekasi via WhatsApp. This system is expected to enhance the efficiency of academic information services and overcome the limitations associated with administrative staff operating hours.

Keywords: *AI Chatbot, WhatsApp, n8n, Automation, Low-code, Universitas Panca Sakti Bekasi*

1. PENDAHULUAN

Perkembangan teknologi informasi dan penggunaan media sosial serta aplikasi pesan instan telah mengubah pola masyarakat dalam memperoleh

informasi. Dalam lingkungan pendidikan tinggi, khususnya Universitas Panca Sakti Bekasi, mahasiswa memperoleh berbagai informasi melalui WhatsApp. WhatsApp menjadi salah satu media komunikasi yang banyak digunakan karena

kemudahan penggunaannya dan manfaat ekonomisnya (Akhmadi et al., 2021). Penggunaan WhatsApp di Indonesia juga berkembang seiring peralihan masyarakat dari telepon genggam sederhana ke smartphone (Amna et al., 2018).

Namun, mahasiswa masih menghadapi kesulitan dalam memperoleh informasi akademik yang tersedia pada website fakultas. Pengguna harus menelusuri berbagai halaman untuk menemukan informasi tertentu, sedangkan layanan admin Universitas Panca Sakti Bekasi memiliki keterbatasan waktu operasional sehingga pertanyaan tidak selalu dapat dijawab dengan cepat. Selain itu, pemanfaatan WhatsApp sebagai kanal informasi otomatis belum maksimal, informasi teknis seperti kontak dosen, admin keuangan, dan bagian kemahasiswaan belum tersedia secara terpusat, serta FAQ mengenai pendaftaran mahasiswa baru dan biaya pendukung belum tersedia secara khusus.

Kondisi tersebut mendorong pengembangan chatbot AI informasi akademik berbasis WhatsApp yang dapat memberikan layanan informasi secara interaktif dan dapat diakses kapan saja. Chatbot dirancang menggunakan basis data dokumen dan pendekatan Retrieval-Augmented Generation (RAG) untuk meningkatkan akurasi dan konsistensi jawaban berdasarkan sumber data eksternal yang terverifikasi. Penelitian ini mengangkat judul “Pengembangan Chatbot AI Informasi Akademik Berbasis WhatsApp Menggunakan Low-Code Automation n8n di Universitas Panca Sakti Bekasi.”

Penelitian dibatasi pada layanan informasi akademik yang meliputi program studi, FAQ pendaftaran mahasiswa baru, informasi mata kuliah, dan data dosen. Bahasa data latih menggunakan bahasa Indonesia, sedangkan chatbot hanya tersedia melalui WhatsApp Universitas Panca Sakti Bekasi. Model AI menggunakan Large Language Model (LLM) melalui API, seperti Groq API dengan model openai/gpt-oss-20b, sedangkan n8n digunakan sebagai orkestrasi AI Agent. Integrasi n8n dengan WhatsApp mendukung Webhook WhatsApp Business API atau WhatsApp Unofficial API.

Berdasarkan permasalahan tersebut, penelitian berfokus pada tiga aspek, yaitu perancangan workflow chatbot melalui n8n agar terintegrasi dengan WhatsApp, penerapan arsitektur RAG menggunakan database pengetahuan untuk mempermudah akses informasi akademik, serta integrasi LLM agar chatbot mampu memberikan respons yang akurat, kontekstual, dan relevan terhadap pertanyaan pengguna.

Penelitian ini diharapkan memberikan manfaat secara akademis sebagai referensi dalam pengembangan ilmu dan penelitian terkait teknologi AI, sekaligus memberikan manfaat praktis bagi instansi dan masyarakat melalui layanan informasi yang lebih cepat, otomatis, dan responsif. Bagi pemerintah, penelitian juga dapat menjadi referensi

penerapan chatbot, RAG, dan LLM dalam mendukung transformasi digital layanan publik.

2. TINJAUAN PUSTAKA

Konsep Pengembangan

Pengembangan merupakan proses atau cara untuk mengembangkan sesuatu. Sugiyono (2022) mendefinisikan pengembangan sebagai upaya untuk memperdalam dan memperluas pengetahuan yang telah ada, sehingga pengembangan tidak selalu berarti menciptakan sesuatu yang sepenuhnya baru, tetapi dapat berupa penyempurnaan, penambahan, atau peningkatan pemahaman dan kemampuan.

Artificial Intelligence

Artificial Intelligence (AI) atau kecerdasan buatan merupakan cabang teknologi komputer yang dikembangkan untuk mensimulasikan kecerdasan manusia melalui kemampuan pembelajaran, analisis data, pengambilan keputusan, *problem solving*, hingga menghasilkan karya kreatif (Pakpahan, 2021; Yustiasari Liriwati, 2023). Perkembangan AI berlangsung sejak konsep mesin yang mampu berpikir yang diperkenalkan Alan Turing hingga perkembangan teknologi AI modern dalam berbagai bidang, termasuk pendidikan (Masriadi & Bahri, 2024; Muin & Kusmaladewi, 2025).

Perkembangan AI juga mencakup munculnya Transformer pada 2017 yang memberikan pengaruh besar terhadap perkembangan *Large Language Models* (LLM), GPT-1 dan BERT pada 2018, serta perkembangan model bahasa dan AI generatif berikutnya (Alalag, 2024; Hiremath & Kenchakkanavar, 2025).

Chatbot

Chatbot digunakan sebagai media interaksi otomatis antara pengguna dan sistem melalui percakapan. Dalam penelitian ini, chatbot dikembangkan sebagai layanan informasi akademik yang memungkinkan pengguna memperoleh informasi melalui percakapan menggunakan WhatsApp. Sistem tersebut mengintegrasikan RAG dan LLM sehingga respons yang diberikan didasarkan pada informasi yang tersedia dalam basis pengetahuan.

Retrieval-Augmented Generation

Retrieval-Augmented Generation (RAG) diterapkan sebagai mekanisme pencarian informasi dari basis data vektor sebelum pertanyaan pengguna diproses oleh LLM. Pendekatan ini digunakan agar jawaban yang dihasilkan lebih akurat, relevan, dan sesuai dengan sumber data yang telah disediakan. Dalam sistem penelitian, RAG menjadi bagian dari workflow chatbot WhatsApp yang terintegrasi dengan n8n dan basis data vektor.

Vector Database

Vector database digunakan untuk menyimpan hasil *embedding* dokumen dan melakukan pencarian informasi berdasarkan kemiripan semantik. Dalam penelitian ini, **Pinecone** digunakan sebagai basis data vektor. Dokumen akademik diproses menjadi representasi vektor, kemudian disimpan dalam Pinecone untuk digunakan pada proses *retrieval*.

Cosine Similarity

Cosine Similarity digunakan untuk menentukan tingkat kedekatan atau kemiripan antara vektor kueri dan vektor dokumen dalam ruang vektor berdimensi tinggi. Dalam sistem yang dikembangkan, metode ini digunakan oleh Pinecone untuk menemukan dokumen akademik yang paling relevan dengan pertanyaan pengguna.

Large Language Model

Large Language Model (LLM) digunakan untuk memproses pertanyaan pengguna bersama konteks yang diperoleh dari hasil *retrieval*. Perkembangan LLM berkaitan erat dengan arsitektur Transformer yang diperkenalkan pada 2017 dan perkembangan model seperti GPT dan BERT (Hiremath & Kenchakkanavar, 2025; Alalaq, 2024). Dalam penelitian ini digunakan model **openai/gpt-oss-20b** untuk menghasilkan respons berdasarkan konteks yang diberikan sistem.

UltraMsg sebagai WhatsApp Gateway

UltraMsg merupakan platform *WhatsApp Gateway* yang menyediakan API untuk mengintegrasikan WhatsApp dengan perangkat lunak atau aplikasi. API tersebut memungkinkan pengiriman pesan secara otomatis melalui HTTP Request dengan penggunaan *Access Token*, *Instance ID*, dan *Webhook*. UltraMsg mendukung metode HTTP GET dan HTTP POST dalam pertukaran data.

WhatsApp sendiri merupakan aplikasi komunikasi berbasis internet yang digunakan untuk mengirim pesan menggunakan koneksi data (Lailia & Venica, 2023). Dalam penelitian ini, WhatsApp menjadi media interaksi pengguna dengan chatbot.

Docker

Docker merupakan platform *containerization* yang digunakan untuk membangun, mengemas, menguji, dan menjalankan aplikasi secara konsisten dalam lingkungan terisolasi yang disebut *container*. Teknologi ini memungkinkan beberapa aplikasi berjalan secara terisolasi pada sistem operasi yang sama tanpa memerlukan *guest operating system* sebagaimana virtual machine (Morozov & Donovan, 2025).

Node.js

Node.js merupakan *runtime environment* JavaScript *open-source* dan *cross-platform* yang

memungkinkan eksekusi JavaScript pada sisi server. Node.js menggunakan V8 engine dan memiliki arsitektur *event-driven* serta *non-blocking I/O* yang mendukung pemrosesan banyak permintaan dengan penggunaan sumber daya yang relatif minimal (Kalman & Miłosz, 2025). Kuffel dan Walter (2024) menunjukkan bahwa karakteristik kinerja *runtime environment* berpengaruh terhadap kinerja *framework* backend Node.js.

Metodologi Penelitian 4D

Model 4D dalam penelitian *Research and Development* terdiri atas empat tahap, yaitu **Define, Design, Develop, dan Disseminate** (Sivasailam Thiagarajan, 1974). *Define* digunakan untuk menetapkan kebutuhan, *Design* untuk merancang produk, *Develop* untuk merealisasikan dan memvalidasi produk, sedangkan *Disseminate* merupakan tahap memperkenalkan atau menyebarkan hasil pengembangan.

Ollama

Ollama merupakan platform *open-source* untuk mempermudah *deployment* dan eksekusi LLM secara lokal pada infrastruktur pengguna (Albert & Voutama, 2025; Jimmy, 2024). Ollama memungkinkan model dijalankan secara luring pada infrastruktur lokal sehingga dapat mendukung privasi data (Salim et al., 2025). Platform ini juga menyediakan lingkungan eksekusi untuk pengelolaan *weight model*, alokasi memori CPU dan GPU, serta REST API untuk kebutuhan inferensi aplikasi eksternal (Jimmy, 2024; Pratama et al., 2025).

Universitas Panca Sakti Bekasi

Universitas Panca Sakti Bekasi merupakan perguruan tinggi swasta yang berdiri berdasarkan SK Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 575/M/2020 tanggal 11 Juni 2020. Universitas ini merupakan hasil penggabungan STKIP Panca Sakti, STIE Indonesia Raya, dan STMIK Cikarang (Universitas Panca Sakti Bekasi, n.d.).

Kerangka Pengembangan Sistem

Pengembangan sistem diawali dengan identifikasi permasalahan layanan informasi akademik, kemudian dilakukan pengumpulan data melalui observasi, wawancara, dan studi pustaka. Model pengembangan yang digunakan adalah 4D, meliputi *Define, Design, Develop, dan Disseminate*. Pada tahap *Design* dirancang LLM, penyimpanan data, workflow, dan basis data vektor sebagai fondasi RAG. Tahap *Develop* digunakan untuk membangun dan menguji sistem, sedangkan *Disseminate* digunakan untuk implementasi dan penyebarluasan sistem.

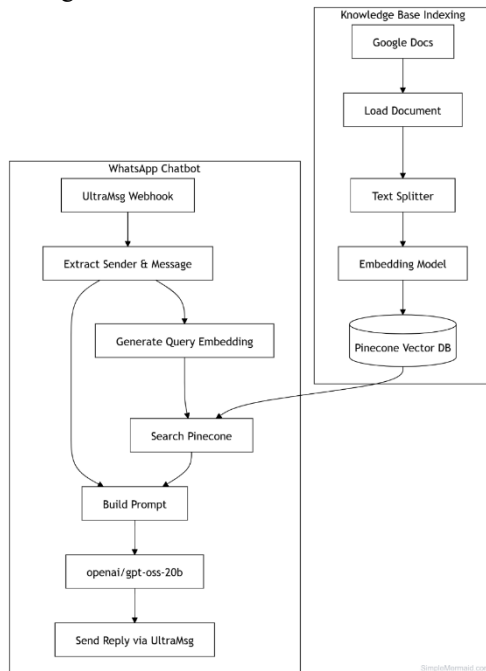
Dari sisi arsitektur, sistem memanfaatkan UltraMsg API sebagai gateway WhatsApp, RAG

sebagai mekanisme pencarian informasi, LLM sebagai penghasil respons, serta n8n yang dijalankan melalui Docker sebagai pengelola workflow.

Desain Model

Sistem terdiri atas dua bagian utama, yaitu knowledge base indexing dan WhatsApp chatbot berbasis RAG. Pada *knowledge base indexing*, dokumen dari Google Docs dimuat ke sistem, dipecah menggunakan *text splitter*, diubah menjadi *embedding*, kemudian disimpan pada Pinecone Vector Database.

Pada workflow chatbot, pesan pengguna diterima melalui UltraMsg Webhook, informasi pengirim dan pesan diekstraksi, kemudian pertanyaan diubah menjadi *embedding*. Pinecone mencari dokumen yang relevan berdasarkan *Top-K*. Hasil *retrieval* digabungkan dengan pertanyaan pengguna melalui *Build Prompt* dan diproses oleh model *openai/gpt-oss-20b*. Jawaban selanjutnya dikirim kembali kepada pengguna melalui UltraMsg.



Gambar 2. Desain Arsitektur Sistem

3. METODE PENELITIAN

Pendekatan Penelitian

Penelitian ini menggunakan metode Research and Development (R&D) dengan model pengembangan 4D. Research and Development digunakan sebagai cara ilmiah untuk menciptakan dan menyempurnakan produk melalui kegiatan penelitian, perancangan, produksi, dan validasi (Sugiyono dalam Alfa et al., 2023). Penggunaan R&D dipilih karena penelitian menghasilkan produk teknologi berupa chatbot berbasis *knowledge-based product*.

Tahapan Pengembangan

Model 4D terdiri atas empat tahap:

1. **Define**, yaitu mengidentifikasi kebutuhan informasi mahasiswa, seperti FAQ pendaftaran, rincian biaya perkuliahan, jadwal kuliah, dan data dosen melalui observasi dan wawancara dengan staf akademik.
2. **Design**, yaitu merancang arsitektur sistem, workflow n8n, arsitektur RAG, serta pemetaan metadata untuk basis data vektor.
3. **Develop**, yaitu merealisasikan rancangan dengan membangun integrasi n8n, UltraMsg, model *embedding nomic-embed-text*, dan model *openai/gpt-oss-20b* melalui Groq API, kemudian melakukan *Alpha Testing*.
4. **Disseminate**, yaitu melakukan penyerahan dan uji coba operasional produk secara terbatas.

Kebutuhan Perangkat Keras

Lingkungan pengembangan menggunakan Windows 10 64-bit, RAM 12 GB, penyimpanan NVMe 256 GB dan SATA 256 GB, Intel Core i7-7700HQ 2,80–3,80 GHz, arsitektur x86-64, serta NVIDIA GeForce GTX 1050 dengan VRAM 4 GB GDDR5 dan 768 CUDA Cores. Komputer lokal difokuskan pada menjalankan model *embedding nomic-embed-text:latest* dan mengorkestrasi aliran API, sedangkan inferensi model LLM dilakukan melalui Groq API.

Tabel 1. Spesifikasi Hardware dan OS

Spesifikasi <i>Hardware & OS</i>	
Sistem Operasi	Windows 10 (64-bit)
RAM	12GB
Storage	NVME 256GB dan SATA 256GB
CPU	Intel® Core™ i7-7700HQ 2.80 GHz – 3.80 GHz
CPU Architecture	x86-64
GPU	NVIDIA GeForce GTX 1050 (VRAM 4GB GDDR5, 768 CUDA Cores)

Kebutuhan Perangkat Lunak

Perangkat lunak yang digunakan meliputi Docker, n8n, Ollama, Pinecone, UltraMsg API, dan Nomic-Embed-Text. Docker digunakan untuk kontainerisasi, n8n sebagai orkestrator workflow,

Ollama sebagai lingkungan eksekusi model embedding lokal, Pinecone sebagai basis data vektor, UltraMsg API sebagai gateway WhatsApp, dan Nomic-Embed-Text sebagai model embedding dengan dimensi vektor 768.

Tabel 2. Komponen Perangkat Lunak

Spesifikasi <i>Software</i>	
Docker	Digunakan sebagai platform kontainerisasi utama untuk mengisolasi, dan menjalankan <i>service</i> n8n.
n8n (Low-Code Automation Platform):	Berperan sebagai orkestrator utama alur kerja (<i>workflow builder</i>) untuk menyatukan simpul (<i>node</i>) <i>webhook</i> , ekstraksi payload, pemanggilan basis data vektor, pembentukan <i>prompt</i> , hingga pengiriman respons akhir.
Ollama	Berperan menyediakan lingkungan eksekusi lokal untuk menjalankan model embeddings <i>nomic-embed-text:latest</i> tanpa memerlukan koneksi internet aktif demi menjaga privasi data universitas.
Pinecone	Berfungsi sebagai basis data numerik hasil <i>embedding</i> dokumen akademik dan menyediakan REST API pencarian <i>semantic similarity</i> secara mudah.
Ultrams API	Bertindak sebagai jembatan komunikasi (<i>WhatsApp API Gateway</i>) yang meneruskan pesan dari <i>server</i> n8n menuju nomor WhatsApp pengguna secara aman menggunakan protokol HTTP POST.
Nomic-Embed-Text	Model <i>embedding open-source</i> yang dijalankan di <i>Ollama</i> juga dan digunakan untuk mengonversi data teks dokumen akademis mentah menjadi array angka (Vektor berdimensi tinggi) dengan akurasi semantik yang akurat. Versi yang dijalankan saat ini adalah <i>nomic-ai/nomic-embed-text-v1.5</i> dengan <i>vector dimension</i> 768.

Kebutuhan Fungsional

Sistem harus mampu menerima dan memvalidasi pesan teks melalui *webhook*, mengekstraksi nomor pengirim dan pertanyaan, melakukan pencarian semantik menggunakan RAG dan Cosine Similarity, menghasilkan jawaban menggunakan *openai/gpt-oss-20b*, serta mengirimkan jawaban kembali ke nomor WhatsApp pengguna.

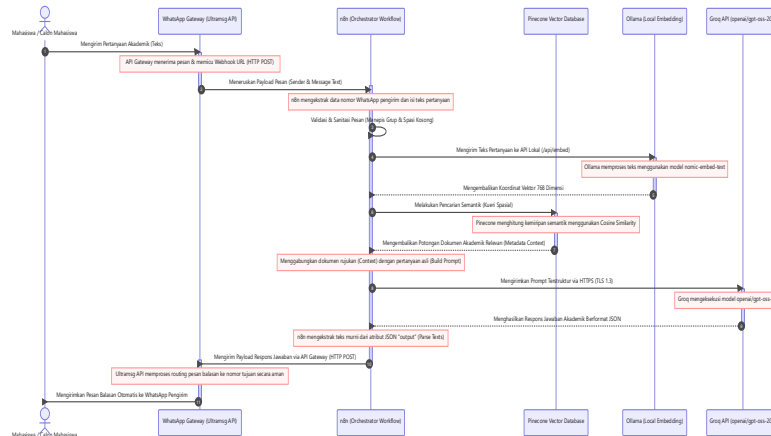
Kebutuhan Nonfungsional

Kebutuhan nonfungsional mencakup kemudahan pembaruan basis pengetahuan, performa, dan usability. Basis pengetahuan dapat diperbarui melalui Google Docs yang diproses kembali oleh workflow n8n untuk melakukan ekstraksi teks dan *embedding* ulang ke Pinecone. Performa dilihat dari waktu pemrosesan kueri hingga pengiriman respons, sedangkan usability didukung penggunaan WhatsApp tanpa instalasi aplikasi tambahan.

Rancangan Model

Alur sistem dimulai ketika mahasiswa mengirimkan pertanyaan melalui WhatsApp. UltraMsg meneruskan pesan sebagai payload JSON menuju webhook n8n. n8n mengekstraksi dan melakukan sanitasi pesan, kemudian mengirimkannya ke Ollama untuk menghasilkan vektor 768 dimensi menggunakan *nomic-embed-text:latest*.

Vektor kueri kemudian dikirim ke Pinecone untuk pencarian semantik menggunakan Cosine Similarity. Dokumen relevan yang diperoleh digunakan sebagai konteks bersama pertanyaan pengguna untuk membentuk *prompt*. Prompt dikirim ke Groq API untuk diproses menggunakan model *openai/gpt-oss-20b*, kemudian jawaban diekstraksi oleh n8n dan dikirim melalui UltraMsg kembali kepada pengguna.



Gambar 3. Rancangan Model

Validasi dan Pengujian

Evaluasi sistem dilakukan menggunakan Black-Box Testing dengan sembilan skenario pengujian yang diberi kode BB-01 hingga BB-09. Pengujian internal atau *Alpha Testing* digunakan untuk menguji fungsi webhook, ekstraksi parameter pesan dari UltraMsg API, dan pengiriman pesan balasan otomatis. *Integration Testing* digunakan untuk menguji komunikasi antara n8n, Pinecone, dan Ollama.

Validasi ahli melibatkan ahli sistem/media dan ahli materi/konten. Ahli sistem/media mengevaluasi arsitektur, alur n8n, latensi API, dan keamanan data, sedangkan ahli materi/konten memvalidasi dokumen akademik, FAQ, struktur biaya, dan data dosen yang digunakan sebagai basis pengetahuan.

Evaluasi Akurasi RAG dan LLM

Evaluasi dilakukan terhadap proses *retrieval* dan *generation*. *Semantic Similarity Testing* digunakan untuk menguji ketepatan *embedding* dalam menemukan dokumen yang relevan berdasarkan Cosine Similarity. Selanjutnya, efektivitas RAG dievaluasi menggunakan Precision dan Recall. Precision mengukur ketepatan dokumen yang diambil, sedangkan Recall mengukur kemampuan sistem menemukan seluruh informasi relevan yang tersedia dalam basis data.

4. HASIL DAN PEMBAHASAN

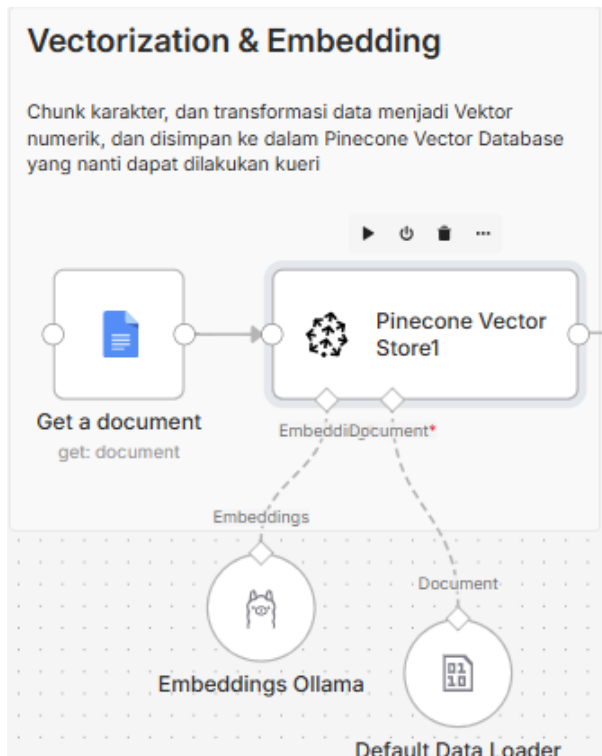
Hasil Pengembangan Sistem

Pengembangan chatbot AI dilakukan di lingkungan Universitas Panca Sakti Bekasi dengan pengujian operasional utama pada Kampus A. Sistem dirancang sebagai layanan informasi akademik melalui WhatsApp dengan memanfaatkan n8n, UltraMsg API, Pinecone, Ollama, dan model *openai/gpt-oss-20b*. Sistem juga dapat dikembangkan untuk operasional Kampus B.

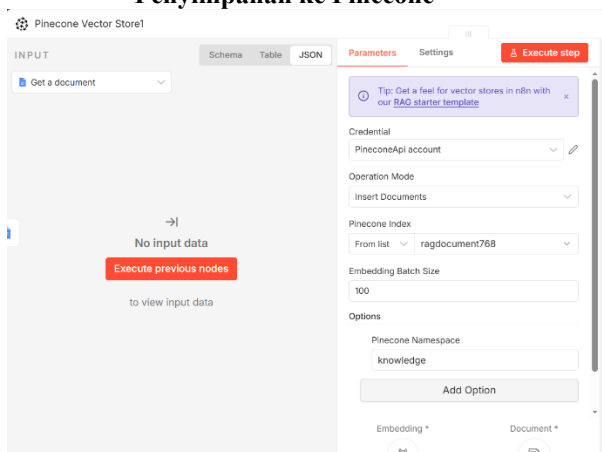
Implementasi Knowledge Base

Knowledge base dibangun dari dokumen Google Docs yang diproses menggunakan *Default Data Loader* dan *Text Splitter*. Teks dipecah menjadi *chunk* berukuran 1.000 karakter dengan overlap 200 karakter untuk menjaga kesinambungan konteks. Selanjutnya, setiap *chunk* diproses menggunakan model *nomic-embed-text:latest* melalui Ollama dan menghasilkan vektor berdimensi 768. Vektor beserta metadata kemudian disimpan pada Pinecone dengan namespace “knowledge”. *Embedding Batch Size* menggunakan nilai default 100, sehingga maksimal 100 *chunk* diproses dalam setiap batch.

Konfigurasi tersebut digunakan untuk memastikan informasi akademik dapat dicari berdasarkan kemiripan makna, bukan hanya berdasarkan kesamaan kata. Kualitas jawaban chatbot juga dipengaruhi oleh kualitas dokumen sumber dan teknik *chunking* yang digunakan.



Gambar 1. Workflow Vektorisasi dan Penyimpanan ke Pinecone



Gambar 2. Konfigurasi Pinecone Vector Store

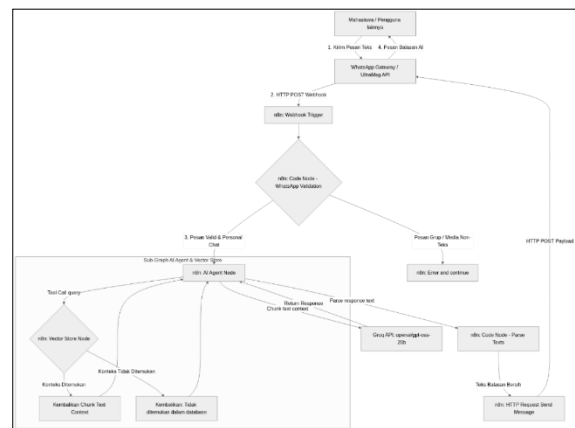
Implementasi Workflow Chatbot

Pesan dari pengguna diterima melalui UltraMsg Webhook sebagai *entry point* sistem. Pesan kemudian diteruskan ke n8n untuk dilakukan validasi. Sistem memeriksa pesan yang berasal dari grup WhatsApp dengan identitas @g.us dan memastikan pesan berasal dari instance WhatsApp yang valid. Pesan yang tidak memenuhi kriteria dihentikan melalui mekanisme *graceful error handling*.

Setelah pesan lolos validasi, pertanyaan diteruskan ke AI Agent. AI Agent terhubung dengan Groq Chat Model menggunakan openai/gpt-oss-20b. Konfigurasi *Maximum Number of Tokens* ditetapkan sebesar 800, sedangkan *Sampling Temperature* sebesar 0,3 untuk menjaga respons tetap terkontrol.

Pada proses *retrieval*, AI Agent melakukan *tool call* ke Pinecone pada namespace “knowledge”. Pencarian dilakukan menggunakan Cosine Similarity untuk memperoleh *chunk* dokumen dengan tingkat kemiripan tertinggi. Data dari Pinecone kemudian dibersihkan dari metadata yang tidak diperlukan sebelum diberikan kepada LLM sehingga penggunaan token menjadi lebih efisien.

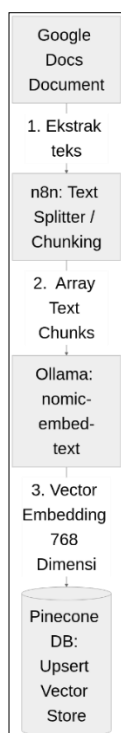
Selanjutnya, pertanyaan pengguna digabungkan dengan konteks hasil *retrieval* melalui *system prompt*. Model menghasilkan jawaban dalam struktur JSON, kemudian node Parse Texts mengambil atribut *output* dan membuang elemen JSON yang tidak diperlukan. Jawaban akhir dikirim melalui HTTP POST ke UltraMsg API dan diteruskan kembali kepada pengguna melalui WhatsApp.



Gambar 3. Flowchart Chatbot AI dengan RAG

Implementasi Penyimpanan Knowledge ke Vector Database

Proses penyimpanan knowledge dimulai dengan mengambil dokumen dari Google Docs. Dokumen kemudian dipecah menjadi beberapa *chunk*, diproses menggunakan *omic-embed-text* melalui Ollama, dan menghasilkan *vector embedding* berdimensi 768. Vektor beserta metadata seperti isi *chunk*, ID dokumen, nomor halaman, atau sumber dokumen kemudian di-*insert* ke Pinecone. Proses tersebut memungkinkan sistem melakukan *semantic search* dan *retrieval* berdasarkan kemiripan makna.



Gambar 4. Flowchart Proses Embedding

Pengujian dilakukan menggunakan Alpha Testing dengan metode Black-Box Testing untuk memvalidasi fungsi workflow n8n. Terdapat sembilan skenario pengujian, yaitu BB-01 sampai BB-09.

Hasil Pengujian Sistem

Tabel 1. Hasil Black-Box Testing

ID	Skenario Pengujian	Hasil yang Diharapkan	Status Pengujian	Kesimpulan
BB-01	Pengiriman pesan teks dari gawai mahasiswa ke WhatsApp kampus.	<i>Webhook</i> n8n terpicu secara otomatis (<i>event-driven</i>) dan menangkap payload JSON.	Sesuai	Berhasil
BB-02	Pengiriman pesan dalam grup <i>WhatsApp</i> (@g.us).	Sistem mendeteksi ID yang berasal dari grup dan <i>Webhook</i> tidak berhenti	Sesuai	Berhasil
BB-03	Pertanyaan “Alamat kampus nya dimana ??”	Jawaban benar	Sesuai	Berhasil
BB-04	Program studi yang tersedia ?	Jawaban benar	Sesuai	Berhasil
BB-05	Pertanyaan “Biaya kuliah tiap fakultas”	Jawaban benar	Sesuai	Berhasil
BB-06	Pertanyaan "Visi dan misi Universitas Panca Sakti Bekasi”	Jawaban benar	Tidak Sesuai	Visi berhasil dicetak namun Misi tidak tercetak
BB-07	Pertanyaan “kelas Sabtu dan Minggu”	Jawaban benar	Sesuai	Berhasil
BB-08	Pertanyaan “apakah ada whatsapp admin nya ? tolong ubah	Jawaban benar	Sesuai	Berhasil

	jadi format https://wa.me ya”			
BB-09	Pertanyaan “nama- nama Fakultas”	Jawaban benar	Sesuai	Berhasil

Sumber: Data hasil pengujian pada BAB IV.

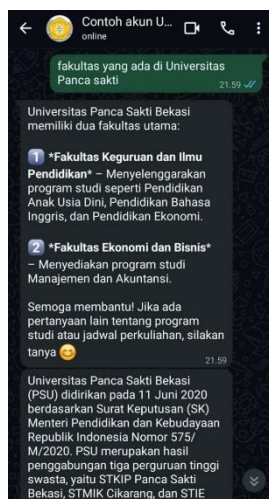
Dari sembilan skenario, delapan pengujian berhasil sesuai dengan hasil yang diharapkan, sedangkan satu pengujian, yaitu BB-06, belum sepenuhnya sesuai karena sistem berhasil menampilkan visi Universitas Panca Sakti Bekasi tetapi tidak menampilkan misi. Temuan tersebut menunjukkan bahwa sistem secara umum mampu menjalankan workflow chatbot, tetapi masih terdapat keterbatasan pada kelengkapan informasi dalam knowledge base.

Contoh Respons Chatbot

Pengujian terhadap pertanyaan pengguna menunjukkan bahwa chatbot mampu memberikan informasi mengenai alamat kampus, program studi, biaya kuliah, jadwal kelas akhir pekan, kontak admin, dan nama fakultas. Sebagai contoh, ketika pengguna menanyakan program studi, chatbot memberikan daftar program studi pada Fakultas Keguruan dan Ilmu Pendidikan serta Fakultas Ekonomi dan Bisnis.

Pada pertanyaan mengenai kelas akhir pekan, sistem memberikan informasi bahwa kelas weekend tersedia pada hari Jumat dan Sabtu pukul 17.00–21.00 WIB, sedangkan hari Minggu tidak tersedia. Sistem juga dapat mengubah nomor WhatsApp admin menjadi format tautan wa.me sesuai permintaan pengguna.

Namun, pada pertanyaan mengenai visi dan misi, sistem hanya mampu memberikan informasi visi dan menyatakan bahwa informasi misi tidak tersedia dalam basis data. Hal tersebut menjadi salah satu temuan penting dalam pengujian karena menunjukkan bahwa kualitas jawaban RAG bergantung pada kelengkapan knowledge base.



Gambar 5. Hasil Tangkapan Layar WhatsApp

Pembahasan

Hasil implementasi menunjukkan bahwa integrasi RAG, Pinecone, Ollama, n8n, UltraMsg, dan LLM dapat digunakan untuk membangun chatbot informasi akademik berbasis WhatsApp. Proses *chunking* sebesar 1.000 karakter dengan overlap 200 karakter memungkinkan dokumen akademik dipecah menjadi bagian yang dapat diproses oleh model embedding sekaligus mempertahankan konteks informasi. Selanjutnya, Pinecone memungkinkan penyimpanan dan pencarian vektor berdasarkan kemiripan semantik.

Penggunaan RAG memungkinkan chatbot tidak hanya menghasilkan jawaban berdasarkan kemampuan bahasa model, tetapi juga menggunakan dokumen akademik yang telah disimpan dalam knowledge base sebagai konteks. Hal ini terlihat dari keberhasilan chatbot dalam menjawab sebagian besar pertanyaan pengujian, seperti alamat kampus, program studi, biaya kuliah, kelas akhir pekan, kontak admin, dan fakultas.

Meskipun demikian, hasil pengujian BB-06 menunjukkan bahwa kelengkapan knowledge base menjadi faktor penting dalam kualitas respons. Ketika informasi misi tidak tersedia dalam data yang digunakan, chatbot tidak dapat memberikan informasi tersebut dan memberikan respons bahwa data belum tersedia. Dengan demikian, pengembangan sistem selanjutnya perlu memperhatikan kelengkapan, pembaruan, dan validasi dokumen yang digunakan sebagai sumber knowledge.

Secara keseluruhan, hasil pengujian menunjukkan bahwa workflow yang dikembangkan telah mampu menjalankan fungsi utama chatbot, mulai dari penerimaan pesan, validasi, *retrieval*, pemrosesan LLM, hingga pengiriman jawaban kembali melalui WhatsApp. Dari sembilan skenario *Black-Box Testing*, delapan berhasil sesuai harapan dan satu skenario masih memerlukan penyempurnaan pada aspek kelengkapan informasi.

5. KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil perancangan, implementasi, dan pengujian, n8n dapat digunakan sebagai orkestrator *AI Agent* dalam pengembangan chatbot dan dapat diintegrasikan dengan layanan pihak ketiga. Workflow berhasil menerima pesan melalui *Webhook*, melakukan validasi dan ekstraksi data pengirim, serta mengirimkan jawaban secara

otomatis melalui WhatsApp. Pendekatan *low-code* juga memudahkan proses pengembangan dan pemeliharaan sistem tanpa memerlukan proses *coding* manual yang kompleks.

Penerapan arsitektur Retrieval-Augmented Generation (RAG) menggunakan dokumen Google Docs, model *embedding nomic-embed-text:latest* melalui Ollama, dan Pinecone sebagai basis data vektor mampu melakukan pencarian dokumen akademik berdasarkan kemiripan semantik. Hasil *black-box testing* menunjukkan bahwa 8 dari 9 skenario berhasil (88,9%), sedangkan satu skenario, yaitu BB-06 terkait informasi visi dan misi, masih memerlukan perbaikan pada kelengkapan data.

Integrasi Large Language Model (LLM) menggunakan model *openai/gpt-oss-20b* melalui Groq API mampu menghasilkan respons yang akurat, kontekstual, dan relevan. Penerapan RAG membatasi jawaban berdasarkan sumber data eksternal yang telah diverifikasi sehingga risiko halusinasi AI dapat diminimalkan. Pengaturan *sampling temperature* sebesar 0,3 dan batas 800 token turut mendukung stabilitas serta efisiensi respons dan menghindari *rate limit* pada layanan Groq API.

Saran

Pengelola administrasi Universitas Panca Sakti Bekasi disarankan meningkatkan kualitas dan struktur dokumen panduan akademik pada Google Docs agar lebih sistematis, detail, dan terstruktur sehingga proses *chunking* dapat menghasilkan informasi yang utuh. Konfigurasi yang disarankan dalam file adalah Chunk Size 1.000–1.200 karakter, Chunk Overlap 100–150 karakter, dan Embedding Batch Size 50, serta penggunaan format Markdown untuk membantu penyusunan dokumen.

Penelitian selanjutnya dapat mengembangkan integrasi LLM dengan sistem SIAKAD Universitas Panca Sakti Bekasi melalui *n8n* maupun *manual code* dengan menyediakan *API Contract Scheme*. Skema tersebut dapat digunakan AI Agent untuk melakukan pertukaran data melalui *HTTP Endpoint* dengan metode GET dan POST, sehingga fungsi akademik yang dapat diakses chatbot menjadi lebih luas.

6. DAFTAR PUSTAKA

- Ainin, Moh. (2013). Penelitian Pengembangan dalam Pembelajaran Bahasa Arab. *OKARA: Jurnal Bahasa Dan Sastra*, 7(2), 95–103. <https://doi.org/10.19105/ojbs.v7i2.455>
- Alalaq, A. S. (2024). The History of the Artificial Intelligence Revolution and the Nature of Generative AI Work. *DS Journal of Artificial Intelligence and Robotics*, 2(4). <https://doi.org/10.59232/air-v2i3p101>
- Albert, G. D., & Voutama, A. (2025). Pengembangan Chatbot Berbasis PDF Menggunakan Local Retrieval-Augmented Generation (RAG) dan Ollama. *Jurnal Informatika Dan Teknik Elektro Terapan*, 13(2), 937–944. <https://doi.org/10.23960/jitet.v13i2.6361>
- Alfa, A. S. S., Novita, L., & Wijaya, A. (2023). PENGEMBANGAN E-MODUL CANVA TEMA 7 SUBTEMA 2 PADA MATA PELAJARAN IPA MATERI MACAM-MACAM GAYA UNTUK SISWA KELAS IV. *Didaktik: Jurnal Ilmiah PGSD STKIP Subang*, 9(2), 3797–3801. <https://doi.org/10.36989/didaktik.v9i2.1055>
- Ardiyanti Sutrisno, Retha Sofiana, & Nuurfathina Nafiisha. (2026). EFEKTIVITAS CHATBOT AI DIBANDINGKAN LAYANAN KONSELOR MANUSIA DALAM DUKUNGAN KESEHATAN MENTAL. *JITET (Jurnal Informatika Dan Teknik Elektro Terapan)*, (Vol. 14 No. 1 (2026)). <https://doi.org/http://dx.doi.org/10.23960/jitet.v14i1.8587>
- Binwal Ankur, & Chopra Puneet. (2024). Privacy and Regulatory Compliance in Retrieval-Augmented Generation Models for AGI Systems. *International Journal For Multidisciplinary Research*, 6(6). <https://doi.org/10.36948/ijfmr.2024.v06i06.30421>
- Hiremath, B., & Kenchakkanavar, A. Y. (2025). A Historical Perspective on Artificial Intelligence: Development, Challenges and Future Directions. *Journal of Advances in Library and Information Science*, 14(1). <https://doi.org/10.5281/zenodo.14877653>
- IBM. (n.d.). *IBM becomes first major open-source AI model developer to earn ISO 42001 certification*. Retrieved May 23, 2026, from <https://www.ibm.com/new/announcements/ibm-granite-iso-42001>
- Jimmy. (2024). Pengembangan Chatbot Berbasis Generative AI Untuk Memudahkan Mahasiswa Memahami Pedoman Penyusunan Dan Penulisan Skripsi Di Universitas IBBI. *Skena Teknologi: Jurnal Ilmiah Sains Dan Teknologi*, 1(2), 19–32. <https://ejournal.ibbi.ac.id/index.php/ST/article/view/31>
- Joshi, S. (2025). Comprehensive Review of AI Hallucinations: Impacts and Mitigation Strategies for Financial and Business Applications. *International Journal of Computer Applications Technology and Research*, (Volume 14–Issue 06, 38 – 50, 2025). <https://doi.org/10.7753/ijcatr1406.1003>
- Kalman, K., & Miłosz, M. (2025). Comparative Analysis of JavaScript Runtime Environments. *Politechnika Lubelska*, 36,

- 296–302.
<https://bibliotekanauki.pl/articles/42604944>
- Kambau, R. A. (2024). PROSES TRANSFORMASI DIGITAL PADA PERGURUAN TINGGI DI INDONESIA. *Jurnal Rekayasa Sistem Informasi Dan Teknologi*, 1(3), 126–136.
- Kuffel, P., & Walter, B. (2024, January 1). Performance of Node.js Backend Application Frameworks. An Empirical Evaluation. *Harnessing Opportunities: Reshaping ISD in the Post-COVID-19 and Generative AI Era (ISD2024 Proceedings)*. <https://doi.org/10.62036/ISD.2024.87>
- Kunwar Arshiya. (2025, December 31). *A Complete Roundup of the Major AI Model Releases in 2025*. <https://www.timesofai.com/industry-insights/roundup-of-ai-model-releases-in-2025/>
- Lailia, S. K., & Venica, L. (2023). *Perancangan Ulang Antarmuka Fitur Obrolan pada Aplikasi WhatsApp dengan Metode Design Thinking*. 13(2). <https://ojs.unikom.ac.id/index.php/jati/article/view/10115/3900>
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W. T., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems, 2020-December*.
- Li, H., Huang, J., Ji, M., Yang, Y., & An, R. (2025). Use of Retrieval-Augmented Large Language Model for COVID-19 Fact-Checking: Development and Usability Study. *Journal of Medical Internet Research*, 27(1). <https://doi.org/10.2196/66098>
- Masriadi, M., & Bahri, H. (2024). THE INFLUENCE OF ARTIFICIAL INTELLIGENCE IN THE MEDIA INDUSTRY IN INDONESIA. *Jurnal Sosiologi Dialektika Sosial*, 10(2). <https://doi.org/10.29103/jsds.v10i2.17590>
- Morozov, Dr. A., & Donovan, Prof. K. J. (2025). The Transformative Impact of Containerization on Modern Web Development: An In-depth Analysis of Docker and Kubernetes Ecosystems. *International Journal of Modern Computer Science and IT Innovations*, 2(10), 27–39. <https://aimjournals.com/index.php/ijmcsit/article/view/299>
- Muin, & Kusmaladewi. (2025). Efektivitas Peningkatan Pembelajaran Berbasis Kecerdasan Buatan. *Jurnal Studi Guru Dan Pembelajaran*, 8(2).
- Nasrul, N., & Izhar, A. (2023). Pengembangan REST API dengan menggunakan Express JS untuk mencari Mentor Pribadi. *Jurnal Informatika Terpadu*, 9(2), 92–102. <https://doi.org/10.54914/jit.v9i2.974>
- Pakpahan, R. (2021). ANALISA PENGARUH IMPLEMENTASI ARTIFICIAL INTELLIGENCE DALAM KEHIDUPAN MANUSIA. *Journal of Information System, Informatics and Computing Issue Period*, 5(2), 506–513. <https://doi.org/10.52362/jisicom.v5i2.616>
- Prasetyo, R. W., & Riyadin, A. A. (2022). Pemanfaatan Fitur Search pada Aplikasi WhatsApp dalam Penerapan Sistem Temu Kembali Informasi. <https://journal.unj.ac.id/unj/index.php/SINTESIA/article/view/39416/15884>
- Pratama, R. W., Simanullang, P. A., Hutabarat, P. T., Nababan, S. D. H., & Perdana, A. (2025). Integrasi Model Bahasa Gemma untuk Sistem Terjemahan Dua Arah antara Pseudocode dan Bahasa Pemrograman Berbasis Mobile. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(6), 16401–16409. <https://ejournal.itn.ac.id/jati/article/view/16401>
- Ruhallah, M. L., Pratami, R., & Gozali, A. A. (2025). Utilizing GPT-4o Mini in Designing a WhatsApp Chatbot to Support the New Student Admission Process at Telkom University. *Journal of Dinda Data Science, Information Technology, and Data Analytics*, 5(2), 176–189. <http://journal.ittelkom-pwt.ac.id/index.php/dinda>
- Salim, W. F., Sugiarto, I., & Wicaksono, H. (2025). Pengembangan dan Implementasi Voicebot Multibahasa dengan Teknologi Retrieval-Augmented Generation untuk Robot Tour Guide. *Jurnal Teknik Elektro*, 18(2), 85–95. <https://doi.org/10.9744/jte.18.2.85-95>
- Sivasailam Thiagarajan. (1974). Instructional development for training teachers of exceptional children. *Indiana University: Bloomington, Indiana*.
- Subiyantoro, S. (2024). *Buku ajar artificial intelligence*. Penerbit Underline. (Andriyanto, Ed.).
- Universitas Panca Sakti Bekasi. (n.d.). *Sejarah Singkat - Universitas Panca Sakti Bekasi*. Universitas Panca Sakti Bekasi. Retrieved <https://panca-sakti.ac.id/sejarah-singkat/>
- WhatsApp. (2026). *Enkripsi End-to-End*. WhatsApp. https://faq.whatsapp.com/820124435853543/?locale=id_ID
- Yustiasari Liriwati, F. (2023). Transformasi Kurikulum; Kecerdasan Buatan untuk Membangun Pendidikan yang Relevan di Masa Depan. *IHSAN: Jurnal Pendidikan Islam*, 1(2), 62–71. <https://doi.org/10.61104/ihsan.v1i2.61>