

Laporan Progres 5

Implementasi RAG Ingestion & Vector Database

1. Deskripsi Progres

Pada tahap ini, sistem telah berhasil mengintegrasikan **Telegram Chatbot** (dari Progress 2) dengan **Vector Database Pinecone** (dari Progress 4). Chatbot kini tidak lagi menjawab pertanyaan secara umum, melainkan menggunakan teknik **Retrieval-Augmented Generation (RAG)** untuk mencari informasi di dalam dokumen PDF yang telah di-upload sebelumnya sebelum memberikan jawaban kepada pengguna.

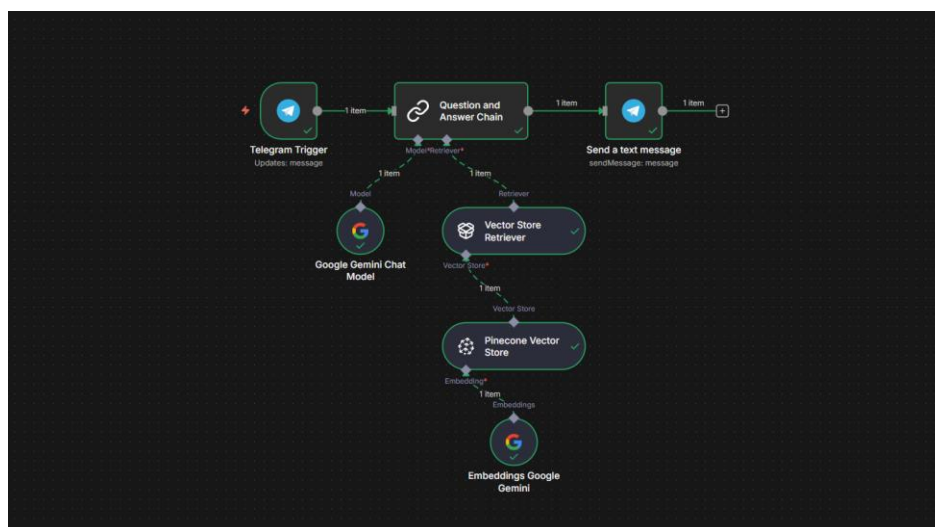
2. Arsitektur Integrasi

Workflow dibangun di n8n dengan komponen utama sebagai berikut:

1. **Trigger:** Telegram Trigger menerima pesan dari pengguna.
2. **Brain (Otak):** Question and Answer Chain bertindak sebagai koordinator yang menggabungkan pertanyaan user dengan konteks dari database.
3. **Retrieval:** Vector Store Retriever mengambil potongan teks yang relevan dari Pinecone berdasarkan kesamaan vektor.
4. **LLM Model:** Google Gemini Chat Model menyusun jawaban akhir berdasarkan data yang ditemukan.
5. **Output:** Telegram Send Message mengirimkan jawaban cerdas kembali ke pengguna.

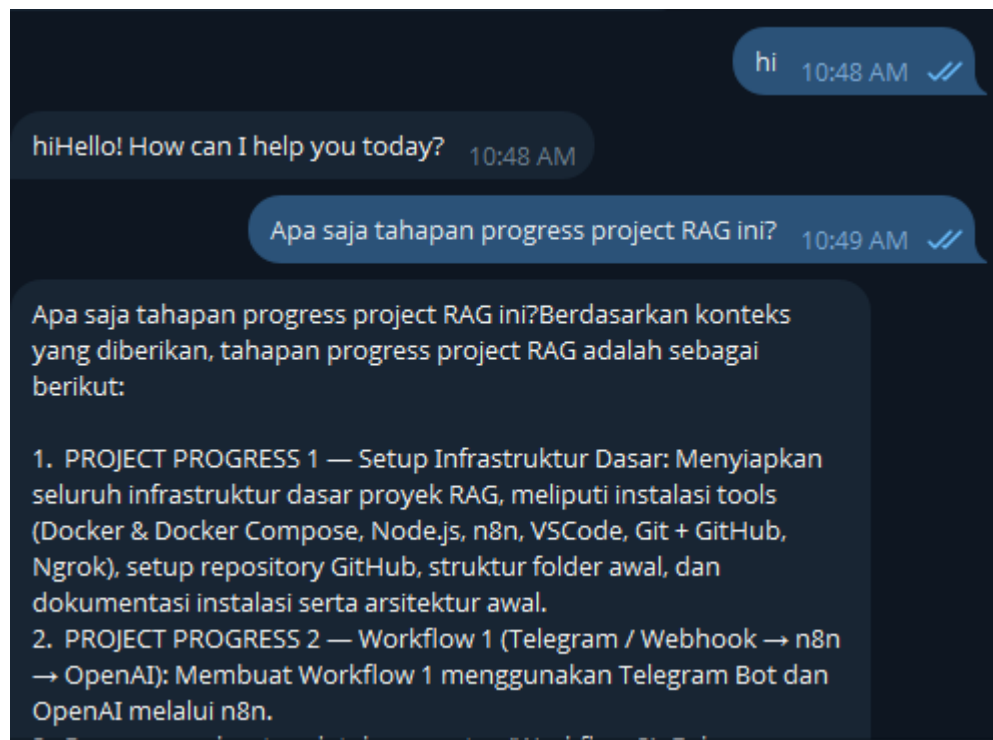
3. Bukti Implementasi (Screenshot)

1. Workflow Integrasi Final (n8n).



Gambar 1: Struktur workflow RAG lengkap yang menunjukkan aliran data dari input pengguna hingga pencarian basis data.

2. Hasil Percakapan Bot (Telegram)



Gambar 2: Bukti bot berhasil memberikan jawaban yang akurat berdasarkan dokumen PDF, bukan sekadar mengulang pertanyaan.

4. Detail Konfigurasi Teknis

1. **Model Embedding:** Google Gemini embedding-001.
2. **Model Chat:** Google Gemini gemini-pro.
3. **Vector DB:** Pinecone Index rag-project dengan 768 dimensi.
4. **Prompt Source:** Diatur ke Define Below dengan ekspresi `{{ $node["Telegram Trigger"].json.message.text }}` untuk menangkap input pengguna secara dinamis.

5. Kendala & Penyelesaian

Integrasi RAG telah berhasil. Sistem mampu melakukan pencarian informasi pada database vektor Pinecone (Retrieval) dan menghasilkan jawaban yang relevan (Generation). Proyek siap dilanjutkan ke tahap **Progress 6 (Cloudflare/Production Readiness)**.