

# CHATBOTAI X IBM

By Machfudin

# PROJECT OVERVIEW

## TUJUAN UTAMA :

Mengembangkan aplikasi web chatbot AI yang intelligent dan user-friendly dengan memanfaatkan teknologi IBM Granite 3.3-8B Instruct untuk memberikan solusi conversational AI yang dapat digunakan dalam berbagai konteks bisnis dan personal, mulai dari customer service, code assistance, hingga problem solving dengan advanced reasoning capabilities.

## TUJUAN SPESIFIK :

1. **Demokratisasi AI:** Membuat teknologi AI enterprise-grade mudah diakses melalui interface web yang intuitif
2. **Efisiensi Komunikasi:** Menyediakan assistant AI yang dapat membantu dalam berbagai tugas komunikasi dan problem solving
3. **Pengalaman User Optimal:** Menciptakan chatbot dengan response time yang cepat dan respons yang contextually relevant
4. **Skalabilitas Enterprise:** Membangun arsitektur yang dapat mendukung multiple concurrent users dengan performance yang stabil

# LATAR BELAKANG

Dalam era digital transformation yang berkembang pesat, kebutuhan akan Conversational AI yang intelligent dan reliable semakin meningkat. Perusahaan dan individu membutuhkan solusi AI yang dapat:

- Memberikan respons yang akurat dan kontekstual dalam real-time
- Mendukung multiple use cases dari customer service hingga technical assistance
- Memproses natural language dengan pemahaman yang mendalam
- Menyediakan enterprise-grade security dan reliability

**IBM Granite 3.3-8B Instruct** hadir sebagai solusi yang mengatasi kebutuhan tersebut dengan:

- **Enterprise-grade quality** dengan model size yang efficient
- **Pre-trained capabilities** untuk multiple domains (business, coding, general conversation)
- **Robust security measures** yang sesuai enterprise standards
- **API accessibility** yang memudahkan integration dengan web applications

# PERMASALAHAN

## **Problem Statement: Keterbatasan Akses AI Enterprise-Grade** **Permasalahan Spesifik:**

- Individual developers kesulitan mengakses AI models yang powerful dan reliable
- Existing chatbot solutions seringkali generic dan tidak dapat di-customize sesuai specific needs
- High barrier to entry untuk implementing conversational AI dalam aplikasi web

## **Dampak Permasalahan:**

- Missed opportunities dalam customer engagement dan automation
- Inefficient manual processes yang sebenarnya bisa di-automate
- Limited scalability dalam handling user interactions

# PEDEKATAN SOLUSI

## 1.1 AI Model Selection

- Evaluasi comprehensive terhadap available AI models dengan fokus pada:
  - Performance vs efficiency trade-off
  - Enterprise-grade security requirements
  - Multi-domain capabilities (conversation, coding, business tasks)
  - API accessibility dan integration ease

## 1.2 Frontend Architecture

- Modern React ecosystem dengan Vite untuk optimal development experience
- Component-based architecture untuk reusability dan maintainability
- Responsive design first dengan TailwindCSS untuk consistent styling
- State management strategy untuk handling complex chat interactions

## 1.3 Backend Architecture

- Node.js dengan Express.js untuk unified JavaScript ecosystem
- RESTful API design dengan clear separation of concerns
- Stream processing implementation untuk real-time chat experience
- Error handling strategy dengan graceful degradation



# TECHNOLOGIES USED

Model AI yang Digunakan : **IBM Granite 3.3-8b Instruct Model**

- Model ini dirancang untuk menangani tugas-tugas instruction-following umum dan dapat diintegrasikan ke dalam asisten AI di berbagai domain, termasuk aplikasi bisnis
- Model Granite 3.3 memiliki kemampuan reasoning yang ditingkatkan dan dukungan untuk Fill-in-the-Middle (FIM) code completion
- Small language model yang fine-tuned untuk peningkatan reasoning, coding, dan instruction-following

## **Frontend :**

- **React** - Framework yang digunakan untuk frontend dalam menyambungkan API dengan Website
- **Vite** - Tools build untuk membuat framework react
- **JavaScript** - Bahasa pemrograman utama untuk framework frontend
- **HTML5 & TailwindCSS** - Untuk struktur dan styling antarmuka

## Backend :

- **Replicate API** - Membantu Menyambungkan dengan Model granite-3.3-8b-instruct tersedia di Replicate
- **REST API** - Untuk komunikasi antara frontend dan backend
- **Node.js** - Membantu dalam membuat API yang disambungkan ke Replicate API
- **Express.js** - Framework yang digunakan untuk Node.js dalam membuat API

## Deployment :

- **Vercel** - Platform deployment yang support dengan React(Frontend) dan Express(Backend)

# ALASAN PEMILIHAN TEKNOLOGI

## IBM Granite 3.3-8b Model:

1. **Efisiensi Computational:** Model Granite 3.0 dirancang sebagai model enterprise yang kecil namun sangat akurat dan efisien
2. **Enterprise-Ready:** Dikembangkan IBM untuk penggunaan enterprise, memprioritaskan kepercayaan, keamanan, dan efisiensi biaya tanpa mengorbankan performa
3. **Versatility:** Dapat digunakan untuk berbagai use case termasuk text generation, classification, summarization, entity extraction, dan customer service chatbots

## Vercel untuk Deployment:

1. **Performance:** Optimasi otomatis untuk aplikasi web modern
2. **Scalability:** Auto-scaling berdasarkan traffic
3. **Developer Experience:** CI/CD terintegrasi dengan GitHub
4. **Global CDN:** Distribusi konten global untuk performa optimal

## Architecture Pattern:

1. **Separation of Concerns:** Frontend dan backend terpisah untuk maintainability
2. **API-First Approach:** Memungkinkan integrasi dengan berbagai platform
3. **Cloud-Native:** Memanfaatkan layanan cloud untuk scalability

# FEATURES

## Fitur Percakapan Interaktif (Core Chat Interface)

**Deskripsi Fitur:** Interface percakapan real-time yang memungkinkan pengguna berinteraksi dengan AI model IBM Granite 3.3-8b melalui chat berbasis web.

### Cara Kerja:

- User mengetik pertanyaan atau perintah di input field
- Frontend mengirim request ke backend API yang terhubung dengan model Granite 3.3-8b
- Model Granite memberikan respons yang robust untuk creating interactive chatbots, executing precise instructions dan supporting multiple languages
- Respons ditampilkan secara real-time di interface chat

### Fungsi yang Berjalan:

- Message streaming untuk respons yang smooth
- Auto-scrolling behavior untuk pengalaman chat yang natural
- Message history persistence selama sesi berlangsung

## Fitur Real-time Response Streaming

**Deskripsi Fitur:** Streaming chat messages dari AI provider dengan management chat state dan UI updates otomatis ketika new messages arrive.

### **Cara Kerja:**

- Menggunakan Vercel AI SDK untuk streaming capabilities
- WebSocket atau SSE (Server-Sent Events) untuk real-time communication
- Progressive loading respons untuk user experience yang lebih baik
- State management untuk handling multiple concurrent conversations

## Fitur Responsive Design

**Deskripsi Fitur:** Interface yang optimal dan dapat menyesuaikan tampilan secara otomatis di berbagai perangkat termasuk desktop, tablet, dan mobile untuk memberikan pengalaman pengguna yang konsisten.

### **Cara Kerja:**

- Menggunakan CSS Media Queries untuk mendeteksi ukuran layar dan device orientation
- Implementasi Flexbox/CSS Grid untuk layout yang fleksibel dan adaptif
- Breakpoint system yang mengoptimalkan tampilan untuk range resolusi tertentu (mobile: <768px, tablet: 768px-1024px, desktop: >1024px)
- Touch-friendly controls dengan minimum touch target 44px untuk mobile usability
- Adaptive typography scaling berdasarkan viewport size

### **Fungsi yang Berjalan:**

- Auto-adjustment layout container berdasarkan screen width
- Scalable chat interface dengan optimal bubble sizing

## Fitur Theme System

**Deskripsi Fitur:** Interface desain yang dapat diubah menurut preferensi pengguna dengan pilihan System/Light/Dark theme untuk kenyamanan visual dan accessibility.

### Cara Kerja:

- System Theme: Menggunakan prefers-color-scheme media query untuk mendeteksi preferensi OS pengguna
- Manual Theme Selection: Tombol switcher yang menyimpan preferensi di localStorage/sessionStorage
- CSS Custom Properties (CSS Variables) untuk color scheme management
- Dynamic class switching pada root element untuk theme transitions
- Smooth transition animations antara theme modes

### Fungsi yang Berjalan:

- Auto-detection system theme pada first load
- Persistent theme preference across sessions
- Contrast optimization untuk accessibility compliance
- Theme-aware icon dan image variation



## **Fitur Custom Configuration**

**Deskripsi Fitur:** Interface yang memudahkan pengguna dalam mengubah konfigurasi model AI seperti temperature, max tokens, system prompt, dan parameter lainnya untuk customized AI behavior.

### **Cara Kerja:**

- Configuration panel dengan intuitive form controls
- Real-time parameter validation dengan immediate feedback
- Preset configuration templates untuk different use cases

### **Fungsi yang Berjalan:**

- Temperature control untuk creativity/randomness adjustment
- Max tokens limitation setting untuk response length control
- System prompt customization untuk specific AI personality/behavior
- Model parameter tuning (top-p, frequency penalty, presence penalty)
- Live preview of configuration impact pada AI responses

## **Fitur Error Handling dan Fallback**

**Deskripsi Fitur:** System yang robust untuk handling berbagai error scenarios dan providing graceful degradation.

### **Cara Kerja:**

- API timeout handling dengan retry mechanisms
- Fallback responses ketika model tidak tersedia
- User feedback untuk failed requests
- Graceful error messages yang user-friendly

# AI SUPPORT EXPLANATION

## Penggunaan AI dalam Proyek

Dalam proyek ini, teknologi **Artificial Intelligence (AI)** digunakan secara relevan untuk membangun fitur chatbot interaktif yang dapat merespons pesan pengguna secara dinamis. Proses integrasi AI dilakukan melalui jalur komunikasi antara antarmuka pengguna (frontend) dan server (backend) berbasis Express.js

### Proses Kerja AI:

Ketika pengguna mengirimkan pesan melalui website, data tersebut diteruskan ke backend yang kemudian memanfaatkan token API dari layanan Replicate. Token ini digunakan untuk mengakses dan menjalankan model **IBM Granite 3.3-8B Instruct**, yaitu model bahasa besar (Large Language Model) yang memiliki kemampuan pemrosesan teks tingkat lanjut.

## Kemampuan Model AI:

Model **IBM Granite 3.3-8B Instruct** memungkinkan sistem menjawab berbagai jenis pesan, termasuk:

1. **Percakapan umum (general conversation)** - Interaksi natural layaknya berbicara dengan manusia
2. **Pembuatan kode program (code generation)** - Generating code dalam berbagai bahasa pemrograman
3. **Rangkuman teks (text summarization)** - Merangkum dokumen panjang menjadi poin-poin penting
4. **Pemecahan masalah kompleks (problem solving)** - Analisis dan solusi untuk masalah yang kompleks
5. **Kemampuan penalaran tingkat lanjut (advanced reasoning capabilities)** - Logical thinking dan analytical processing

## Dampak Nyata Penggunaan AI:

Hasil balasan dari model AI ini kemudian dikirim kembali ke frontend dan ditampilkan tepat di bawah pesan pengguna, membentuk alur percakapan layaknya interaksi manusia.

Dengan demikian, penggunaan model **IBM Granite 3.3-8B Instruct** memberikan kontribusi nyata dalam meningkatkan interaktivitas dan kecerdasan respons aplikasi chatbot ini. Model ini tidak hanya menjawab pesan secara kontekstual, tetapi juga memperkaya pengalaman pengguna melalui fleksibilitas dalam menangani berbagai skenario penggunaan.

## Keunggulan AI Implementation:




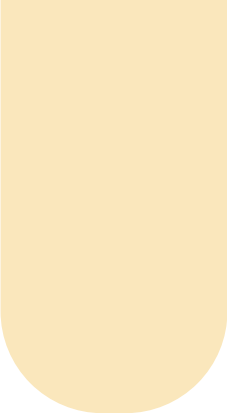
1. **Enterprise-Grade Quality** - Model IBM Granite dirancang khusus untuk aplikasi bisnis dengan standar enterprise
2. **Multi-Domain Expertise** - Kemampuan memahami dan merespons berbagai domain knowledge
3. **Contextual Understanding** - Mempertahankan konteks percakapan untuk interaksi yang lebih natural
4. **Scalable Performance** - Dioptimalkan untuk handling multiple concurrent users
5. **Advanced Reasoning** - Kemampuan analytical thinking untuk problem solving yang kompleks

# SCREENSHOOT OUTPUT (DEMO CAPTURE)



Powered by IBM Granite 3.3 8B Instruct via Replicate  
Last response: 16 prompt + 538 response = 554 total tokens

Live Demo : <https://chatbotai-ibm.vercel.app/>



**TERIMA KASIH**

