

Company Name: NexaTech Solutions  
Year: 2023  
Department: Engineering

---

## SECTION 1: Company Overview

---

NexaTech Solutions is a global software company specializing in AI-driven enterprise platforms. The company was founded in 2015 and has offices in India, Germany, and the United States.

Our mission is to build scalable AI systems that improve operational efficiency across industries. The core business areas include:

- AI-powered analytics
  - Cloud-native backend systems
  - Enterprise automation
  - Intelligent document processing
- 

## SECTION 2: AI Platform Architecture

---

The AI platform consists of the following layers:

1. Data Ingestion Layer:
    - Collects structured and unstructured data.
    - Supports PDF, CSV, JSON, and API inputs.
  2. Processing Layer:
    - Performs data cleaning and transformation.
    - Uses distributed processing frameworks.
  3. Embedding & Vector Layer:
    - Converts documents into embeddings.
    - Stores embeddings in a vector database.
    - Uses cosine similarity for semantic search.
  4. Retrieval-Augmented Generation (RAG):
    - Retrieves top-k relevant chunks.
    - Passes context to Large Language Models.
    - Generates grounded responses.
- 

## SECTION 3: Security Policies

---

All enterprise documents are encrypted at rest using AES-256 encryption. Data in transit is protected using TLS 1.3.

Access Control:

- Role-Based Access Control (RBAC) is enforced.
- Only Admin and Senior Engineers can access production logs.
- Audit logs are maintained for 5 years.

Prompt injection attacks are mitigated by:

- Input sanitization
- Context isolation

- System prompt protection
  - Output validation filters
- 

## SECTION 4: Performance Optimization

---

Latency Optimization Techniques:

- Caching frequent queries
- Using smaller embedding models
- Reducing chunk size for retrieval
- Asynchronous retrieval pipelines

Cost Optimization Techniques:

- Limiting top-k retrieval to 3-5 chunks
- Using cheaper LLM models for non-critical tasks
- Implementing response caching
- Token truncation strategies

---

## SECTION 5: Known Limitations

---

The system may hallucinate if:

- No relevant context is retrieved.
- The similarity threshold is set too low.
- The prompt is overly open-ended.

To mitigate hallucination:

- Set similarity score threshold above 0.75.
- Return "Information not found in document" if no context is relevant.
- Use citation-based prompting.

---

## SECTION 6: Future Roadmap (2024)

---

Planned improvements include:

- Hybrid search combining BM25 and vector similarity.
- Cross-encoder reranking for improved retrieval precision.
- Multi-agent orchestration for research workflows.
- Long-context transformer models (32k+ tokens).
- Real-time streaming inference.

---

## SECTION 7: Contact Information

---

Head Office: Bengaluru, India  
Email: support@nexatech.ai  
Phone: +91-9876543210

---

END OF DOCUMENT

---