HackRx 6.0

Ideate • Co-create • Impact

Dark Knights

- Mahad Iqbal | 2027 | Heritage Institute Of Technology
- Maithili Kumar | 2027 | Heritage Institute Of Technology
- Pawan Kumar | 2027 | Heritage Institute Of Technology
- Mahima | 2027 | Heritage Institute Of Technology
- Vivek Kumar | 2027 | Heritage Institute Of Technology

Tell us a bit about yourself

HackRx 6.0
Ideate • Co-create • Impact

BAJAJ FINSERV

• Any projects you've worked on

- **AuraMed**

1. AI-powered, dual-interface platform for healthcare.
2. Streamlines hospital operations and simplifies patient admissions.
3. Enables real-time health monitoring.
4. Empowers users with appointment booking, emergency alerts, and predictive diagnostics.

- **MaveriqAir**

1. Addresses environmental risks with hyperlocal air quality forecasts.
2. Provides real-time pollution source mapping and urban flood alerts.
3. Integrates satellite data, ground sensors, and advanced machine learning for predictions

- **Cratov**

1. Enhances urban infrastructure management with a blockchain + AI system.
2. Enables decentralized pothole reporting.
3. Uses a machine learning model to verify image authenticity.
4. Employs AI-driven workflows combining traffic, weather, and road data for accurate, context-aware reporting.
5. Generates human-readable pre-repair reports using Gemini APIs.
6. Uses a blockchain-powered bidding system for transparent contractor selection and accountability.

• Hackathon Achievements :

- Secured 2nd place at GDG IdeateX 2025 for developing MaveriqAir, a platform addressing air pollution and urban flooding.
- Secured 3rd place at HACK-O-NIT 2025 for AuraMed, a smart hospital management system.
- Finalist at IIT BHU's Serve-Smart Hackathon.
- Advanced to Pitch Round of Code for Bharat Season 2 2025 .

Problem statement

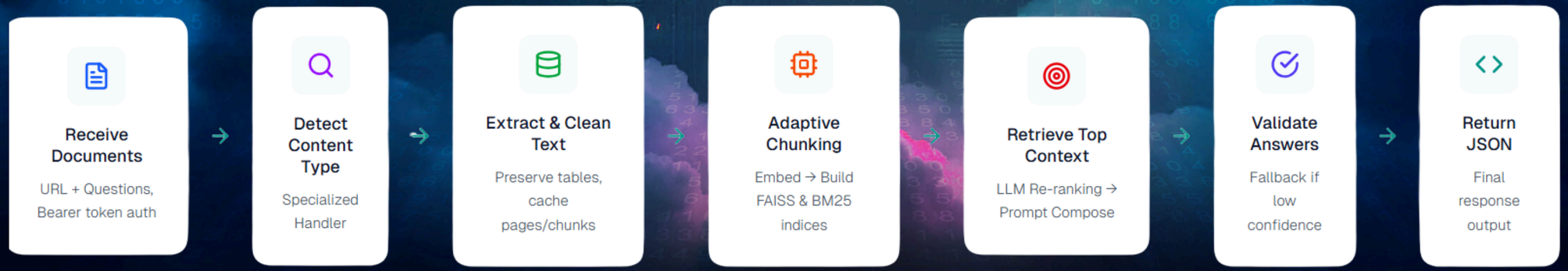
HackRx 6.0
Ideate • Co-create • Impact

BAJAJ FINSERV

Give us an overview of your solution

- Unified API answers questions from web pages and files.
- Smart routing: PDF/DOCX, PPTX, ZIP, Excel/CSV, images, BIN.
- Extracts, chunks, indexes; hybrid FAISS+BM25 retrieval with LLM re-rank.
- Produces concise, grounded JSON answers with confidence validation.
- Caching, batching, parallelism deliver fast, reliable responses.

Process Flow



HackRx 6.0
Ideate • Co-create • Impact

BAJAJ FINSERV

Tech Stack

- **Backend:** Python, Flask, CORS
- **Deployment:** Cloudflared (Tunneling Localhost)
- **LLMs:** OpenAI gpt-4.1-mini.
- **Embeddings:** text-embedding-3-small.
- **Retrieval:** FAISS CPU, Rank-BM25; LangChain splitters; PyMuPDF.
- **Parsing:** BeautifulSoup, lxml, python-docx, extract-msg; pandas/openpyxl/xlrd.
- **Utilities:** requests, numpy, in-memory + disk caching.
- **Concurrency:** ThreadPoolExecutor; resilient JSON parsing and CORS.

Architecture :

HackRx 6.0

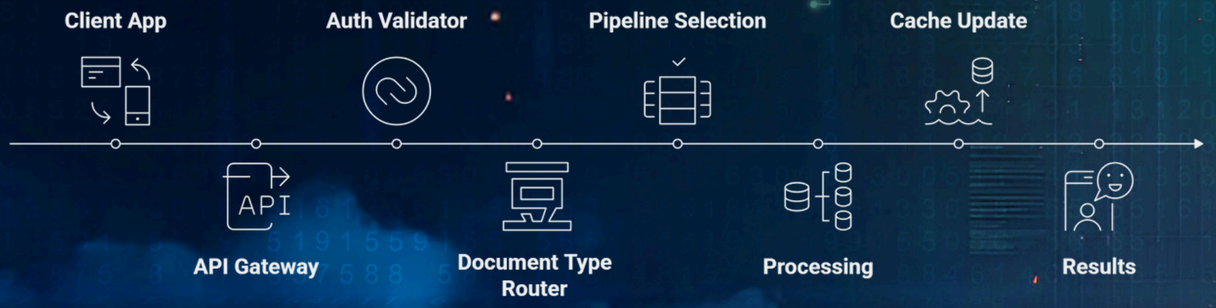
Ideate • Co-create • Impact

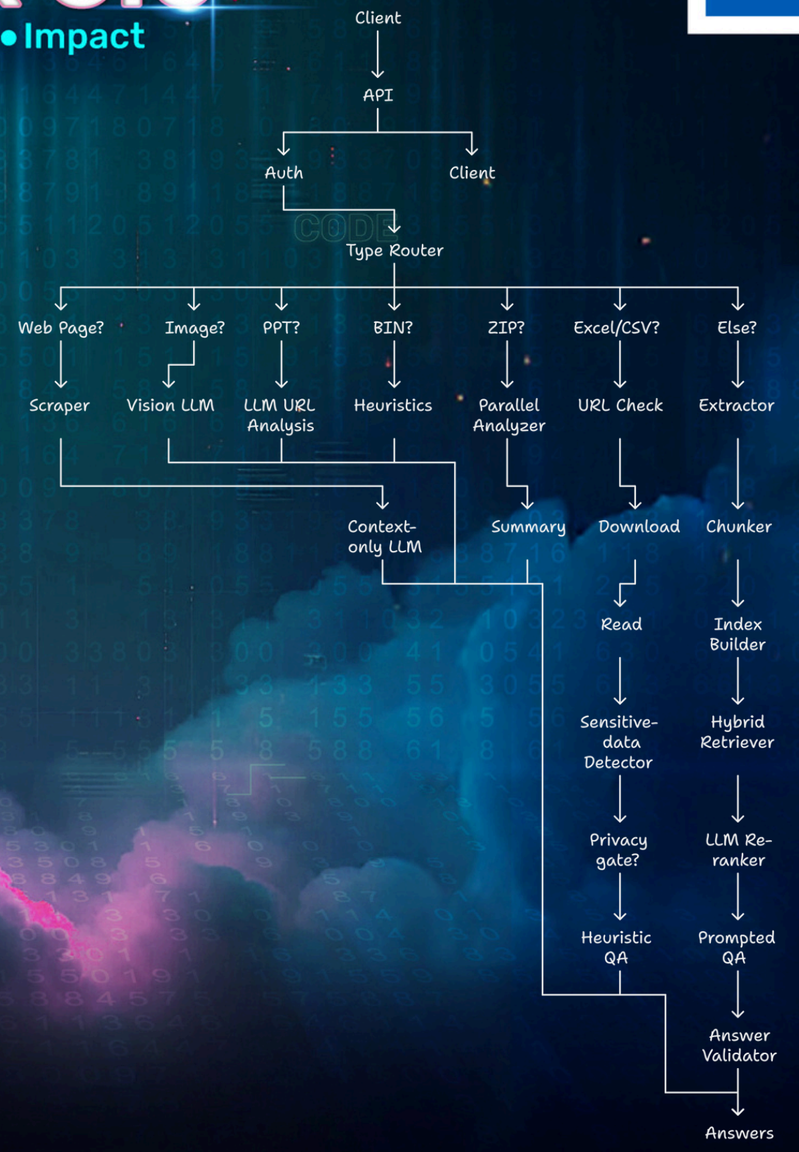
HackRx 6.0

Ideate • Co-create • Impact

BAJAJ FINSERV

Data Flow Diagram :





HackRx 6.0

Ideate • Co-create • Impact

BAJAJ FINSERV

So, how if your solution different?

- Hybrid search + LLM re-ranking ensures accurate, relevant grounding.
- Validation layer scores similarity, overlap, and consistency.
- Sensitive-data guardrails for Excel/CSV with safe summaries.
- High throughput via caching, batched embeddings, and parallelism.
- Emphasized hybrid FAISS+BM25, LLM re-ranking, and validation.
- Domain-specialized prompts selected per query (general, insurance, health_policy, constitution, physics) boost accuracy and clarity.
- Highlighted performance (caching/parallelism) and sensitive-data safeguards.

HackRx 6.0
Ideate • Co-create • Impact

BAJAJ FINSERV

Future possible enhancements

- **Multilingual + Gemini:** Google Gemini integration with auto language detection/translation.
- **Voice + chat history:** Speech input/output and persistent, user-scoped conversation threads.
- **Security hardening:** VirusTotal scanning, sandboxed temp storage, RBAC/SSO, audit logs.
- **Policy recommendations:** Personalized suggestions from query history (content + collaborative).
- **Advanced re-ranking:** Cross-encoder rerankers and domain-tuned models for higher precision.

HackRx 6.0
Ideate • Co-create • Impact

BAJAJ FINSERV

Risks/Challenges/Dependencies

- **Please mention any risks or challenges that you forsee**
  - Documents sent to external APIs without encryption, exposing sensitive data
  - Complex multi-component architecture requires extensive documentation and expertise
  - Rate limits and token constraints may impact system performance under high load
- **Critical Showstoppers Faced**
  - ZIP archives: Deep nesting/ZIP traps causing timeouts, high CPU/memory.
  - Large BIN files: Non-text data limits QA value; expectation mismatch.
  - PPT with image-only slides: Needs OCR/layout parsing; risk of missed text.
  - Excel/CSV sensitivity: False negatives/positives in PII detection; compliance exposure.
  - Scale and rate limits: Concurrency may hit OpenAI quotas; latency spikes, cost overruns.
  - Very large docs: Embedding cost/memory pressure; FAISS index growth and cache churn.

HackRx 6.0
Ideate • Co-create • Impact

Acceptance Criteria Coverage

- **How many aspects of the problem statement have been covered?**
  - 7 of 7 aspects covered.

Breakdown:

- Process PDFs, DOCX, emails: yes
- Handle policy/contract data efficiently: yes
- Parse natural language queries: yes
- Semantic search with embeddings: yes
- Clause retrieval and matching: yes
- Explainable decision rationale: yes
- Output structured JSON responses: yes

HackRx 6.0

Ideate • Co-create • Impact

BAJAJ FINSERV

GitHub Repo

THANK YOU