BAC AGENT

AI-Powered Algerian Baccalauréat Platform

FastAPI | React + TypeScript | RAG Pipeline | OpenAI GPT-4o

A smart tutoring system for Algerian Bac students.

7 streams • 5 AI modes • RAG on real past exams • LaTeX rendering

Product Architecture & Feature Overview • 2026

OVERVIEW

What is Bac Agent?

Bac Agent is an AI-powered tutoring platform built exclusively for Algerian Baccalauréat students. It combines a context-aware LLM tutor with a Retrieval-Augmented Generation (RAG) pipeline trained on real past exams (2015–2024), giving every student personalized, curriculum-aligned academic support.

7	5	10
Official Streams	**AI Chat Modes**	**Years of Exams**
MATH · SCI · TECH · GESTION…	Tutor · Exercise · Exam…	2015 – 2024 PDFs processed

Core Problem: Algerian students — especially in technical streams — have no structured AI tool that respects the exact Ministry of Education curriculum, official grading schemes (Barème), and each stream's unique coefficient weighting.

CURRICULUM COVERAGE

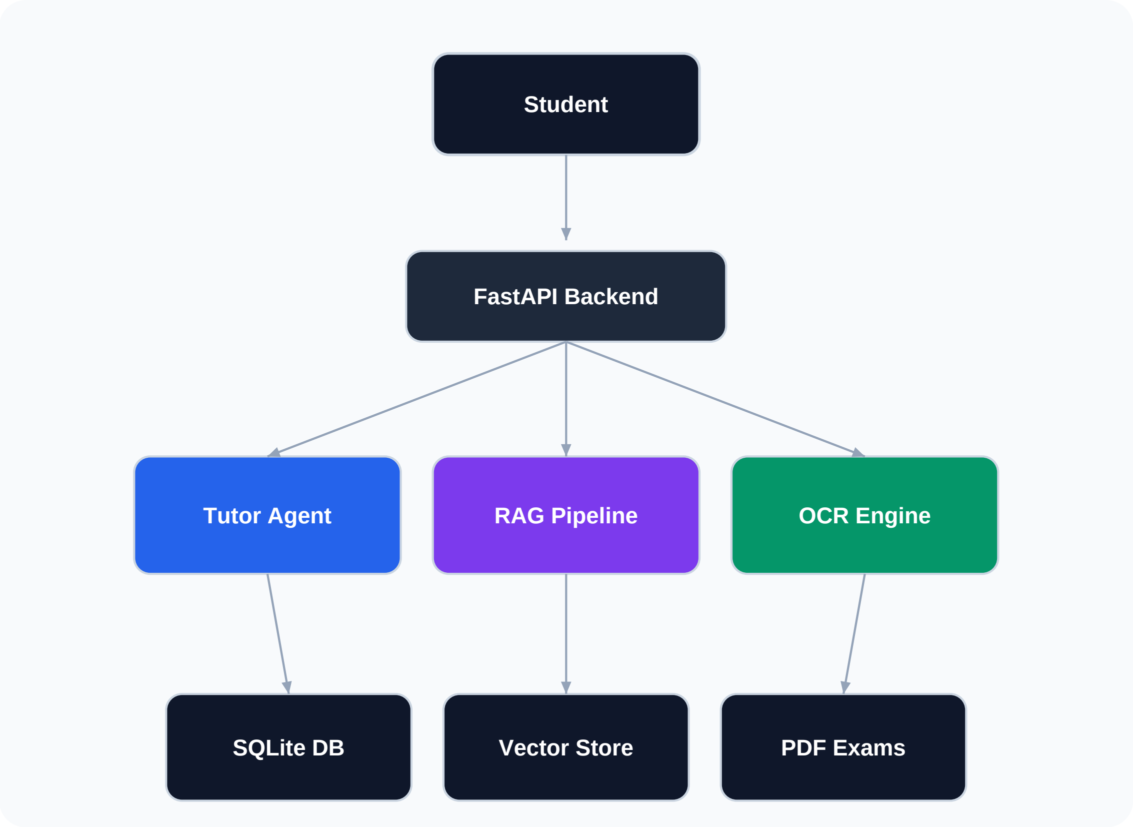
7 Official Bac Streams (Filières)

Stream	Top Subjects & Coefficients	AI Complexity
Mathématiques	Math (7) · Physics (6)	Very High
Sciences Expér.	Sciences (6) · Physics (5)	High
Technique Math	Math (6) · Physics (6) · Tech (6)	Extreme — 4 options
Gestion & Économie	Accounting (6) · Economics	Medium
Langues Étrangères	Arabic / French / English (5 ea.)	Medium-High
Lettres & Philosophie	Philosophy (6) · Arabic Lit (6)	High
Arts	Drawing / Art Specialty (6)	Niche

Technique Math is the most complex stream — it has 4 specialty options (Civil, Mechanical, Electrical, Process Engineering). Each specialty gets a distinct system prompt and coefficient table, handled automatically by the Tutor Agent based on the student's profile.

TECHNICAL DESIGN

System Architecture



ENGINEERING

Technology Stack

Layer	Technology
Frontend	React 18 + TypeScript + Vite
Styling	Custom CSS
Backend	FastAPI (Python 3.11)
AI / LLM	OpenAI GPT-4o
Embeddings	OpenAI text-embedding-3-small
Database	SQLite → PostgreSQL (prod)
OCR	Mathpix / Google Vision / Tesseract

AI CORE

The Tutor Agent & 5 Chat Modes

Mode	Mode ID	Behavior
Orientation	**general**	Study planning & stream priorities — not concept teaching or exercise solving
Exercises	**exercise_help**	Socratic method — guides with questions first, reveals full answer only after student attempt
Concepts	**concept_explanation**	Structured: Definition ⟶ Theorem ⟶ Intuition ⟶ Formula ⟶ Examples ⟶ Misconceptions
Exam Prep	**exam_prep**	Focuses on past Bac exams (2015–2024), recurring question patterns and time management
Review	**solution_review**	Student submits solution ⟶ agent checks each step vs. official Barème and gives score

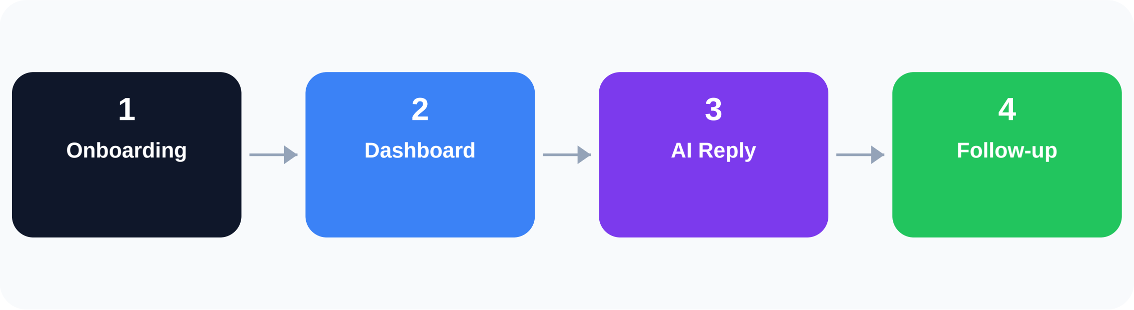
DATA & RETRIEVAL

RAG Pipeline

①	**PDF Ingestion**	Past Bac exams (2015–2024) loaded via ocr_engine.py Mathpix / Google Vision / Tesseract fallback
②	**Text Chunking**	4 strategies via LangChain RecursiveCharacterTextSplitter: lesson (1000 chars), exercise (1500), solution (800), general
③	**Embedding**	Each chunk vectorized with OpenAI text-embedding-3-small Stored as float array in data/vector_store/embeddings.npy
④	**Vector Storage**	chunks.json (metadata) + embeddings.npy (numpy flat-file) Filtered by stream_code and subject_code at query time
⑤	**Query Retrieval**	Student message embedded → cosine similarity → top-k chunks Returned via POST /search-context endpoint
⑥	**Prompt Injection**	Top chunks appended to system prompt as context Agent cites exam year and subject in answer

USER EXPERIENCE

Student Journey



1 Onboarding

2 Dashboard

3 AI Reply

4 Follow-up

Step	Location	What Happens
1 — Onboarding	**/onboarding**	Pick stream (e.g. Mathematiques) + specialty if applicable. Saved in Zustand store.
2 — Dashboard	**/dashboard**	5 mode chips + input box. Student selects mode, types query, presses send arrow.
3 — Chat state	**/dashboard**	Page transitions in-place (no URL change). Messages rendered with Markdown + KaTeX.
4 — History	**Right sidebar**	Collapsible panel lists all past sessions. Click any session to fully restore it.
5 — New Chat	**Header button**	RotateCcw icon resets state and session ID. Returns to idle hero screen.

DATABASE

Data Models

Table	Key Fields	Relationships
streams	id, code, name, name_ar, has_options	→ coefficients (1:N) → users (1:N)
subjects	id, code, name, name_ar, category	→ coefficients (1:N)
coefficients	stream_id, subject_id, coefficient specialty_option, is_specialty	← streams ← subjects
users	id, email, full_name, stream_id specialty_option, is_admin	← streams

The **Coefficient Engine** uses these tables to compute a student's weighted Bac average. All stream/subject/coefficient data is seeded via **init_db.py** at startup. The endpoint **POST /calculate-average** accepts marks and returns the weighted average with official mention classification (Passable / Assez Bien / Bien / Très Bien).

API REFERENCE

Key API Endpoints

Method	Endpoint	Description
GET	/streams	List all 7 streams with Arabic names and has_options flag
GET	/streams/{id}	Stream detail with full coefficient table
GET	/streams/{id}/specialties	Technique Math sub-options (Civil / Meca / Elec / Proc)
POST	/calculate-average	Compute weighted Bac average from subject marks dict
POST	/chat	Send message to Tutor Agent — requires mode + stream context
POST	/search-context	RAG: embed query → cosine search → return top-k exam chunks
GET	/subjects	List all subjects with optional category filter

ROADMAP

Implementation Phases

Phase 1

Data & Foundation

- Build the coefficient engine for all 7 streams
- Seed SQLite DB with streams, subjects, coefficients
- Build OCR engine to convert PDFs to LaTeX/Arabic
- Collect and process exam JSON files 2015–2023

Phase 2

MVP AI + Frontend

- Build RAG pipeline with embeddings and retrieval
- Build Tutor Agent with 5 specialized chat modes
- Build React SPA with onboarding and chat flows
- Add inline chat and collapsible history sidebar

Phase 3

Advanced Features

- Add image input to process photos of exercises
- Build timed Mock Exam mode with auto-scoring
- Add voice explanation support for oral subjects
- Migrate to Supabase pgvector for production scale

Build all this and the app will be ready to ship.

- FastAPI backend with coefficient engine covering all 7 streams
- RAG pipeline trained on 2015–2024 Algerian Bac exam PDFs
- 5 specialized AI chat modes each with custom tutor prompting
- React SPA with in-page chat and collapsible history sidebar
- OCR engine to ingest scanned PDF exams (Mathpix / Vision / Tesseract)
- 7 streams with stream-aware and specialty-aware AI prompting

Bac Agent • Built for Algerian Bac Students • 2026