



LEARN TUBE

AI YouTube Learning Assistant

PROJECT REPORT

Student	Numair Fahad
Internship ID	ZYNVEX-CERT-0487
Program	AIML Internship
Project Title	AI YouTube Learning Assistant
Tech Stack	FastAPI React Expo Google Gemini ChromaDB Clerk Auth
Live URL	https://ai-you-tube-assistant.replit.app
Date	August 2026

1. Introduction

LearnTube is a full-stack AI-powered platform — available as both a web application and a native mobile app — that transforms any YouTube video into an interactive, structured learning experience. Users paste a YouTube URL and instantly receive an AI-generated summary, chapter breakdowns, a multiple-choice quiz, flip-card flashcards, and a context-aware chat assistant, all grounded in the actual video transcript.

The project was built as the capstone of the AIML Internship program, demonstrating practical application of Retrieval-Augmented Generation (RAG), large language models, vector databases, and modern full-stack engineering practices — including a live deployment at ai-you-tube-assistant.replit.app and a cross-platform mobile app.

Core Capabilities

Feature	Description
RAG Chat	Ask any question about the video — answers are grounded in transcript chunks with [MM:SS] timestamp citations.
Smart Summary	Auto-generated overview, key points, and inferred chapter structure from the video.
AI Quiz	Adaptive multiple-choice questions with explanations to verify comprehension.
Flashcards	Spaced-repetition-ready flip cards covering key concepts and terms.
User Auth	Clerk-managed authentication with Google OAuth, email/password, and forgot-password.
Video Library	Per-user persistent library to revisit and re-learn from any added video.
Mobile App	Native Expo / React Native companion app for iOS and Android — full feature parity with the web app.

2. Problem Statement

YouTube hosts over 800 million videos and serves more than 1 billion hours of content daily. Despite being the world's largest free educational platform, learners face significant barriers to effective knowledge retention:

- Passive consumption — viewers watch without active recall or structured review.
- No quick-reference summaries — rewatching an entire video to find a single point is inefficient.
- Lack of comprehension checks — no built-in quizzes, exercises, or knowledge tests.
- No interactive Q&A — learners cannot ask follow-up questions about specific content.
- Time inefficiency — long videos bury key concepts without navigation aids.
- Fragmented note-taking — learners must context-switch between the video and note tools.

Our Solution

LearnTube solves all six barriers in a single unified interface. By extracting the transcript of any YouTube video, chunking it semantically, embedding it into a vector store, and feeding relevant chunks to a Google Gemini LLM, the application provides instant, grounded, interactive learning tools — without requiring any manual note-taking or video replay. The platform is accessible from any browser and, with the companion mobile app, from any iOS or Android device.

3. System Architecture

LearnTube follows a clean three-tier architecture: a React/Vite SPA on the web frontend, an Expo/React Native app on mobile, a FastAPI REST backend, and a dual persistence layer combining SQLite (relational data) with ChromaDB (vector embeddings).

Architecture Overview

Layer	Technology	Responsibility
Web Frontend	React 18 + Vite + Tailwind CSS v4	SPA UI, routing, state management
Mobile Frontend	Expo 53 / React Native 0.79	iOS & Android app, Clerk auth, full feature parity

Layer	Technology	Responsibility
Authentication	Clerk (Replit-managed)	JWT sessions, Google OAuth, email/password
API Gateway	FastAPI (Python 3.13)	REST endpoints, auth middleware, RAG orchestration
LLM	Google Gemini 2.0 Flash	Chat answers, summary, quiz, flashcard generation
Vector Store	ChromaDB	Semantic chunk storage and similarity search
Relational DB	SQLite + SQLAlchemy	User library, video metadata, summaries
Transcript API	youtube-transcript-api	Fetches auto/manual captions from YouTube
Embeddings	sentence-transformers all-MiniLM-L6-v2	Local embedding model
Intent Classifier	scikit-learn TF-IDF + LR	Routes user queries to correct pipeline

RAG Pipeline Flow

When a user submits a YouTube URL, the backend executes this pipeline:

Step	Action
Step 1 — Transcript Fetch	youtube-transcript-api retrieves timestamped captions (auto or manual).
Step 2 — Chunking	Transcript split into ~200-word semantic chunks, each tagged with a start timestamp.
Step 3 — Embedding	Each chunk is embedded via sentence-transformers (all-MiniLM-L6-v2) and stored in ChromaDB.

Step	Action
Step 4 — RAG Query	User questions trigger a top-5 cosine-similarity search; retrieved chunks are injected as LLM context.
Step 5 — LLM Generation	Google Gemini 2.0 Flash generates grounded answers with [MM:SS] timestamp citations.

4. Technology Stack

Category	Tool / Library	Version	Purpose
Language	Python	3.13	Backend services
Language	TypeScript	5.x	Frontend (web & mobile)
Framework	FastAPI	latest	REST API server
Framework	React	18	SPA web frontend
Framework	Expo	53	React Native mobile framework
Mobile UI	React Native	0.79	Native iOS & Android components
Build Tool	Vite	6	Frontend bundler
Styling	Tailwind CSS	v4	Utility-first CSS
Auth	Clerk	v5	JWT + OAuth sessions (web & mobile)
LLM	Google Gemini	2.0 Flash	AI text generation
LLM SDK	google-genai	2.17	Gemini Python client

Category	Tool / Library	Version	Purpose
Vectors	ChromaDB	latest	Semantic similarity search
Embeddings	sentence-transformers	latest	all-MiniLM-L6-v2 model
Database	SQLite + SQLAlchemy	latest	Relational persistence
ML	scikit-learn	1.x	Intent classification
UI Library	shadcn/ui + Radix UI	latest	Web component library
Animations	Framer Motion	12	UI animations
Data Fetching	TanStack Query	v5	Server state management
Routing	Wouter	latest	Lightweight SPA router

5. App UI Screenshots

The following screenshots capture the live LearnTube application. The UI follows a “Cosmic Learning” design theme — deep navy background, indigo/violet/cyan gradients, animated floating orbs, glassmorphism cards, and a play-button logo mark.

Landing Page — Hero Section

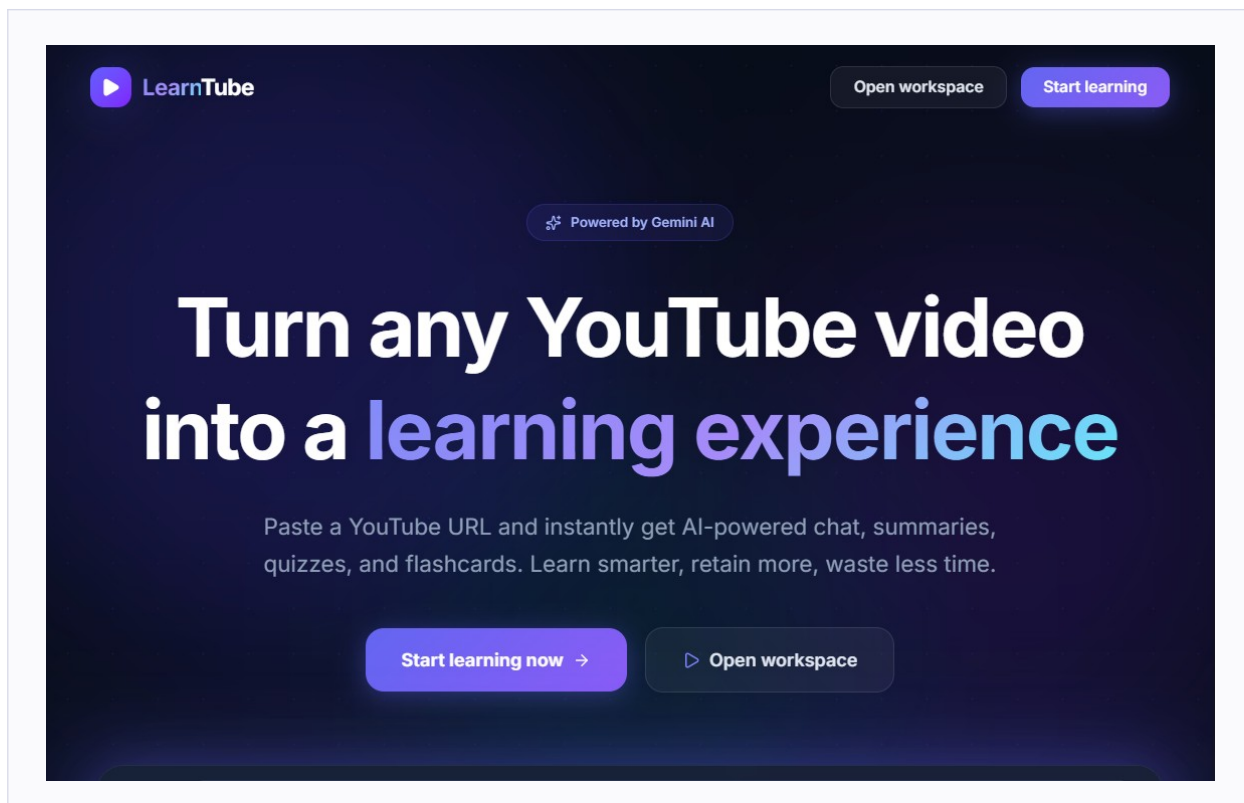


Fig. 1 — Landing page hero: “Turn any YouTube video into a learning experience”

Landing Page — Interactive Demo Preview

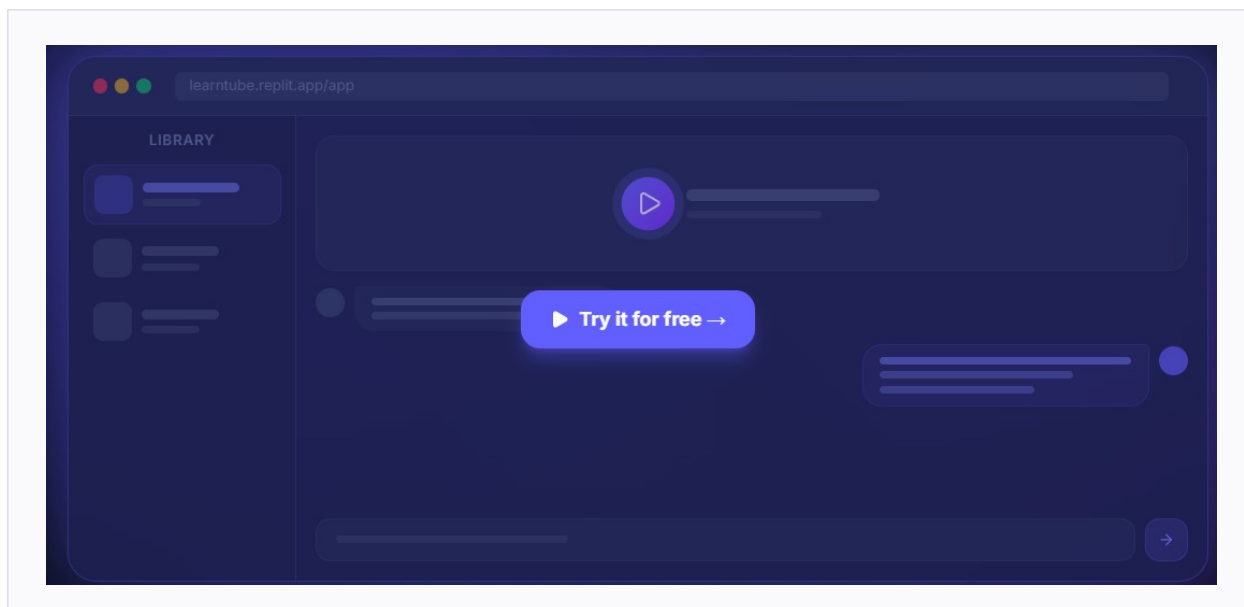


Fig. 2 — Live chat demo preview embedded on the landing page

Landing Page — Feature Highlights

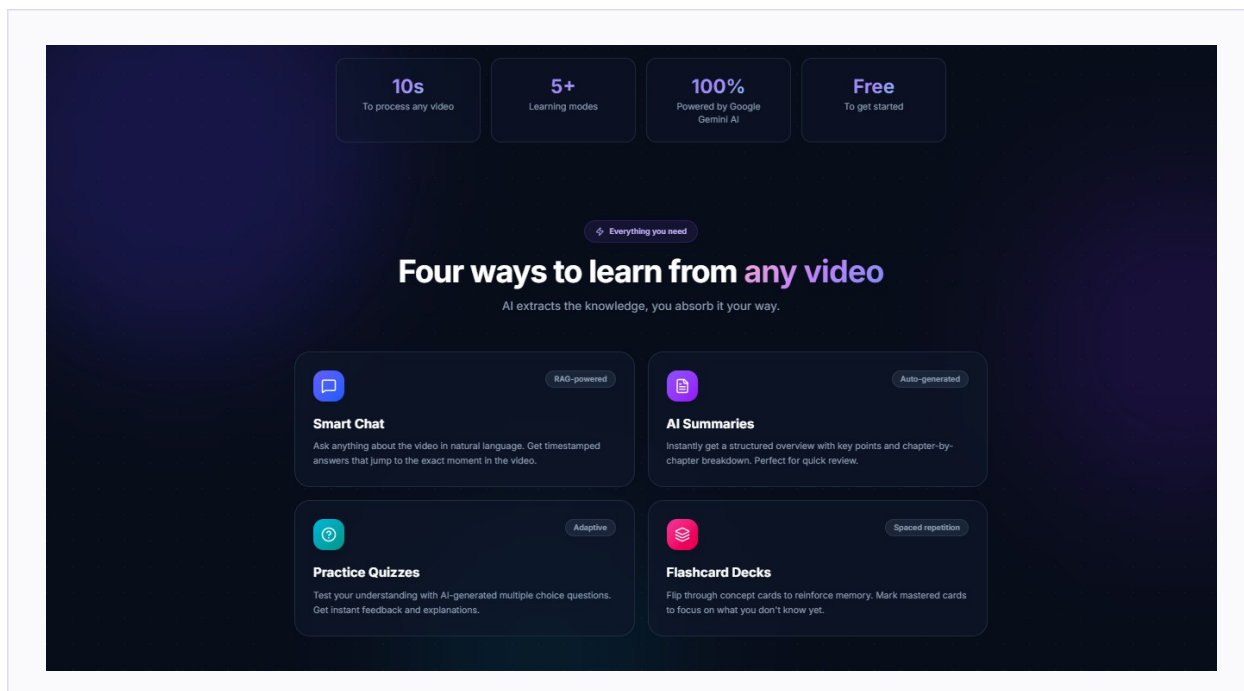


Fig. 3 — “Four ways to learn from any video”: Smart Chat, AI Summaries, Practice Quizzes, Flashcard Decks

Landing Page — How It Works & Testimonials

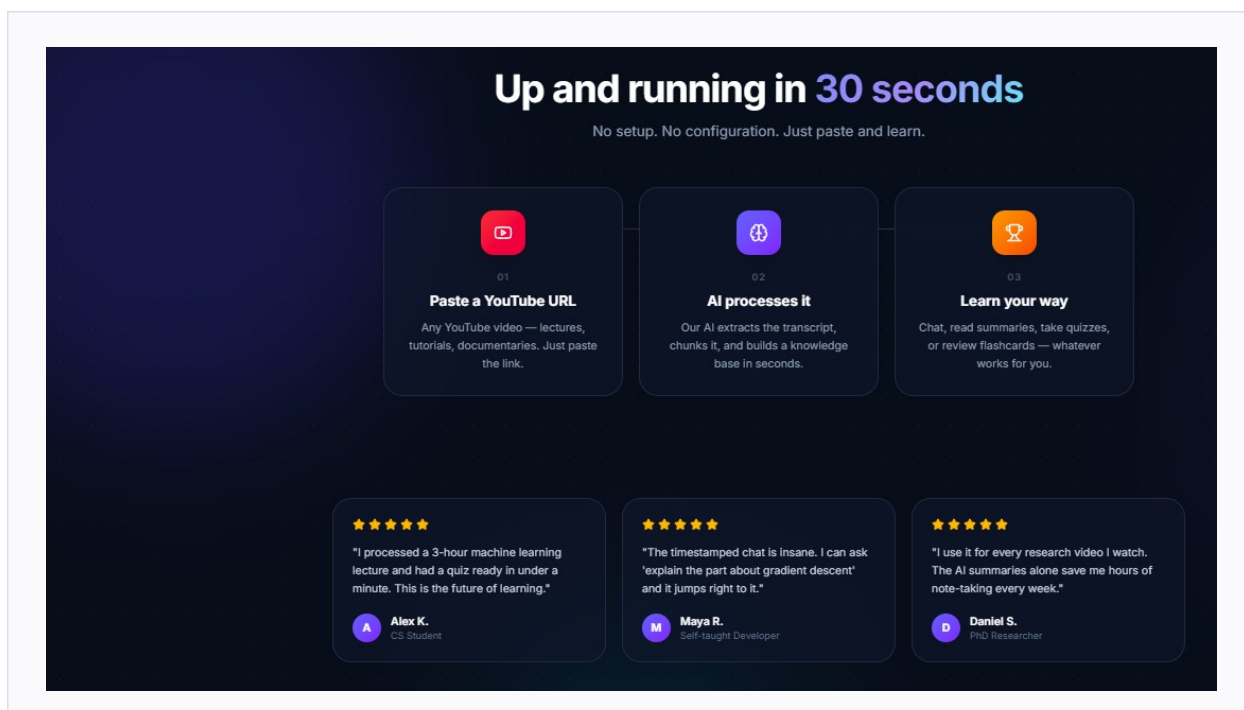


Fig. 4 — Three-step onboarding flow and user testimonials

Landing Page — Call to Action

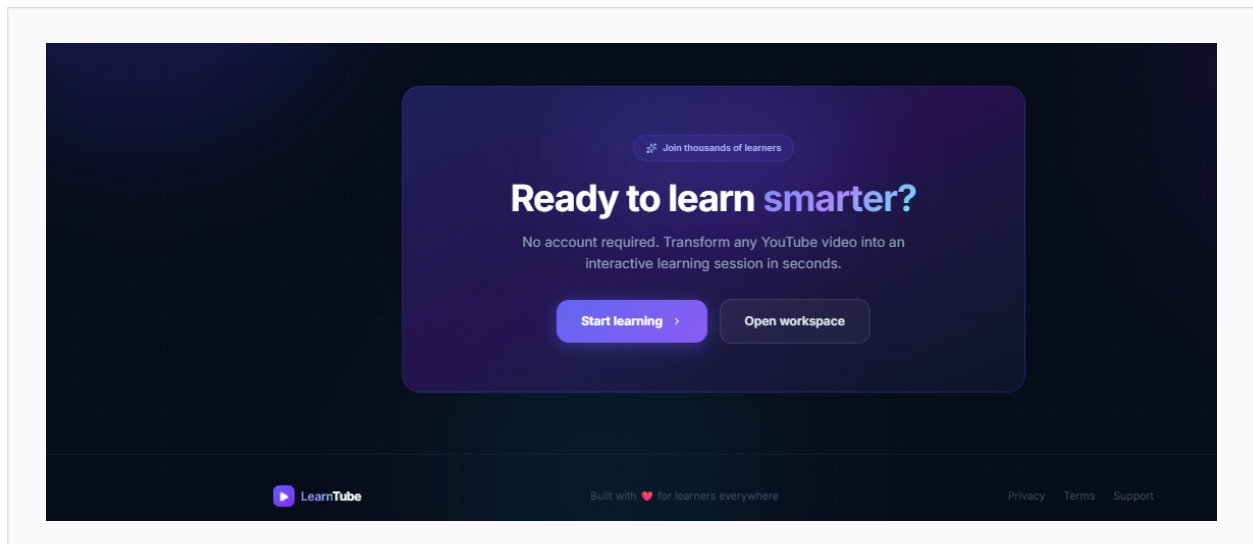


Fig. 5 — Closing call-to-action and footer

App Workspace — Getting Started

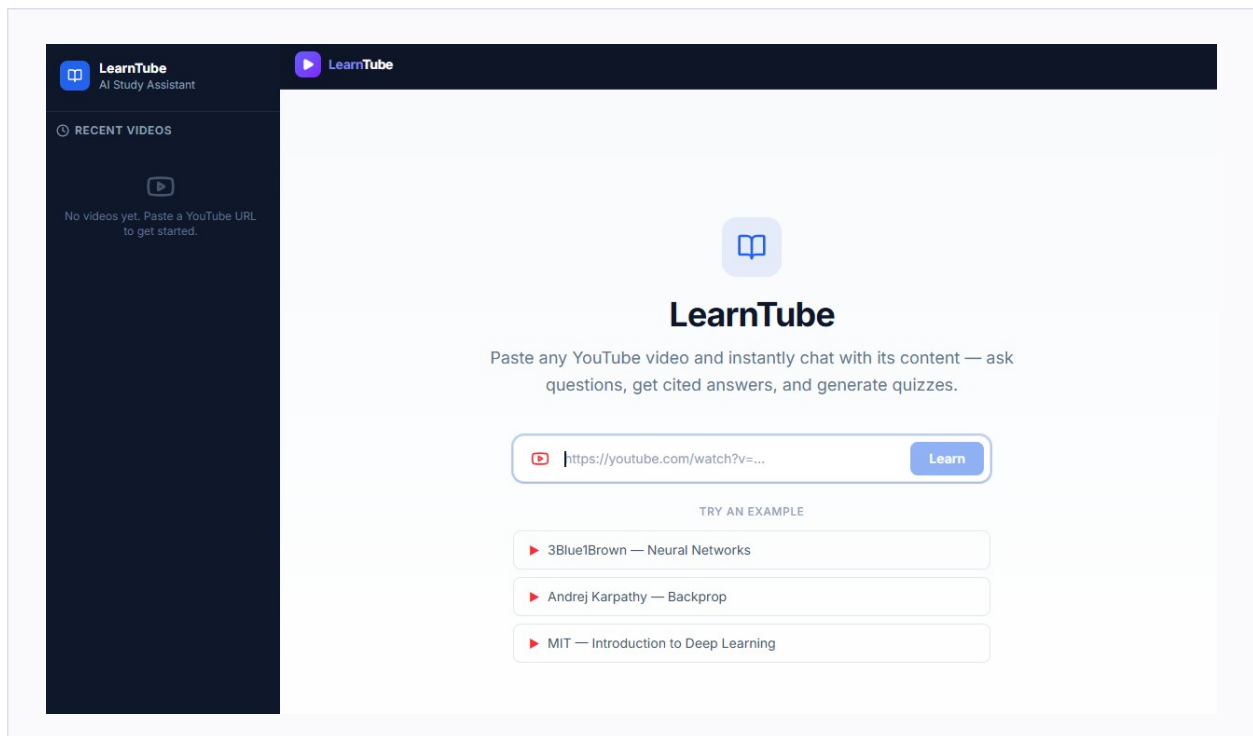


Fig. 6 — LearnTube AI Study Assistant workspace with URL input and example videos

App in Action — RAG Chat

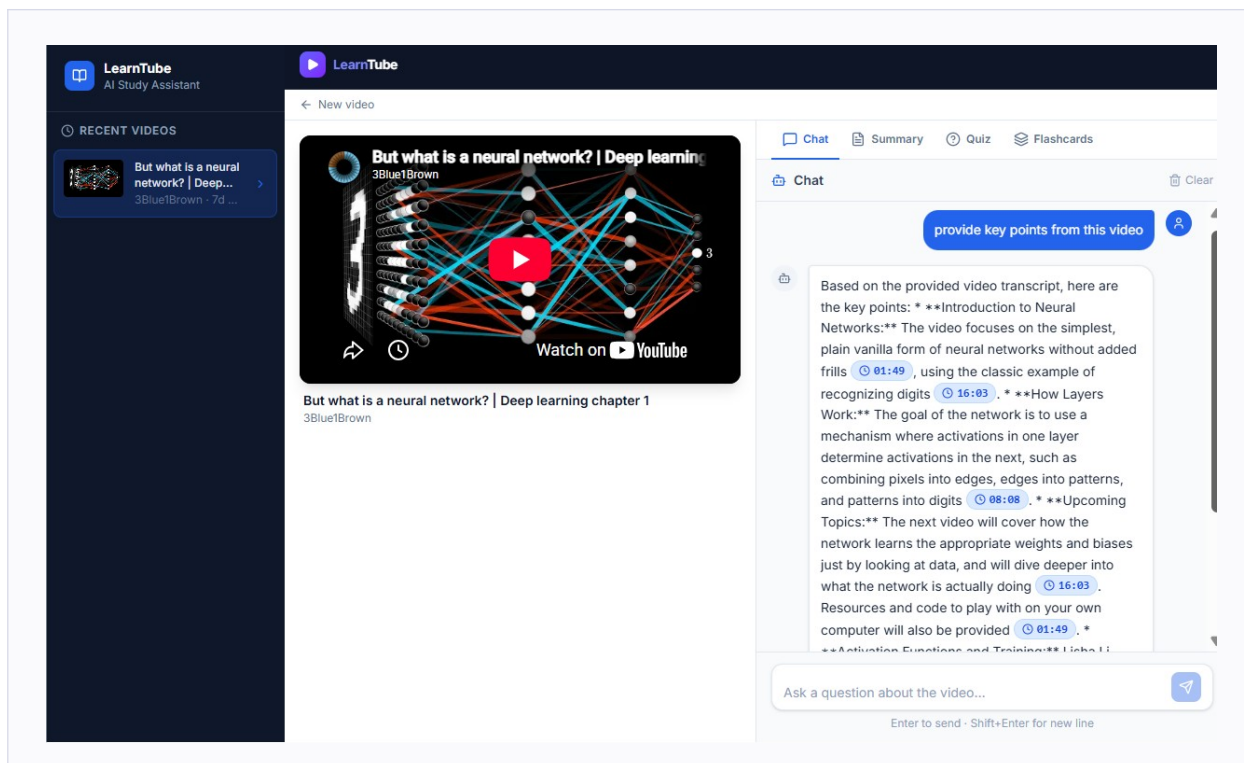


Fig. 7 — Context-aware chat answering questions with clickable [MM:SS] timestamp citations

App in Action — AI Summary

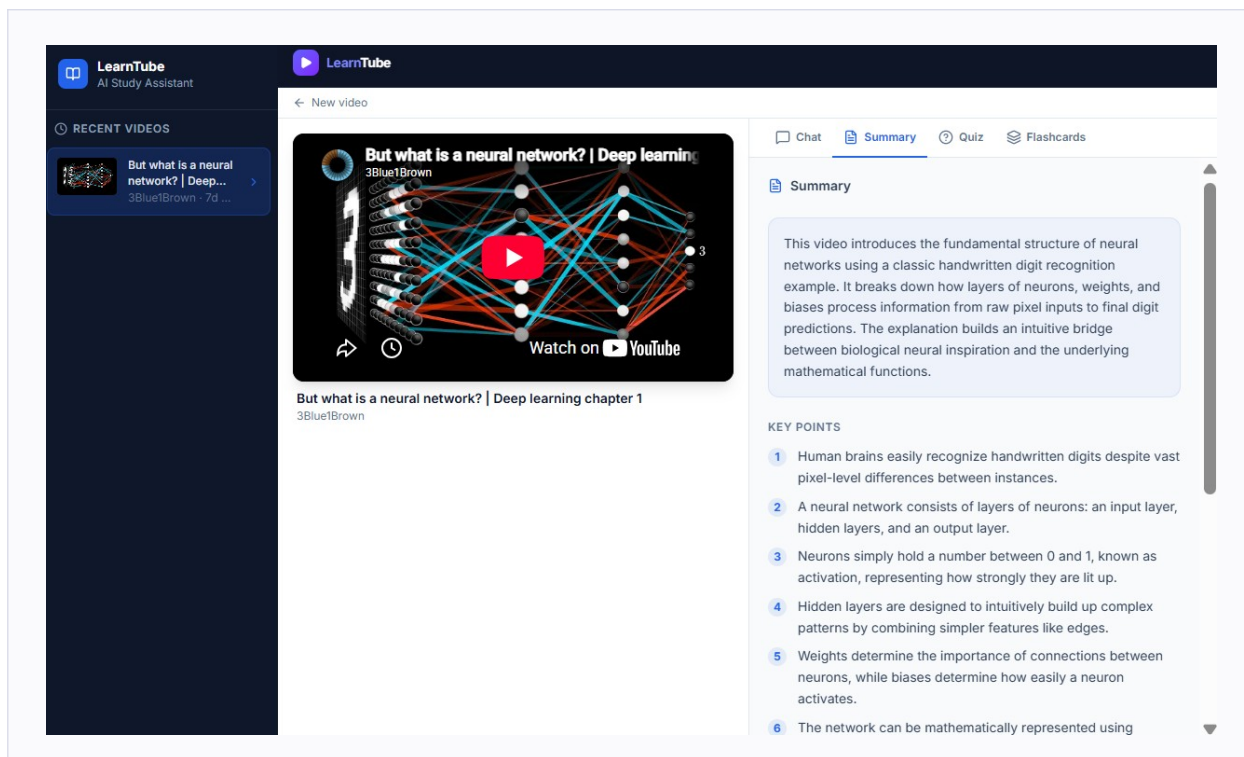


Fig. 8 — AI-generated summary with structured key points for a processed video

App in Action — AI Quiz

← New video

But what is a neural network? | Deep learning
3Blue1Brown

Watch on YouTube

But what is a neural network? | Deep learning chapter 1
3Blue1Brown

Chat Summary Quiz Flashcards

Quiz — 7 Questions 1/7 answered

1 What does the activation value of a neuron typically represent?

☒ The physical temperature of the computer processor

☒ A number between 0 and 1 representing how strongly the neuron is lit up

☐ The exact weight assigned to the output layer

☐ The total number of pixels in the input image

X Incorrect. Neurons hold a number between 0 and 1, known as activation, representing how strongly they are lit up.

2 Which of the following best describes the role of hidden layers in a neural network?

☐ To store the final classification output

Fig. 9 — Auto-generated multiple-choice quiz with instant correct/incorrect feedback

App in Action — Flashcards

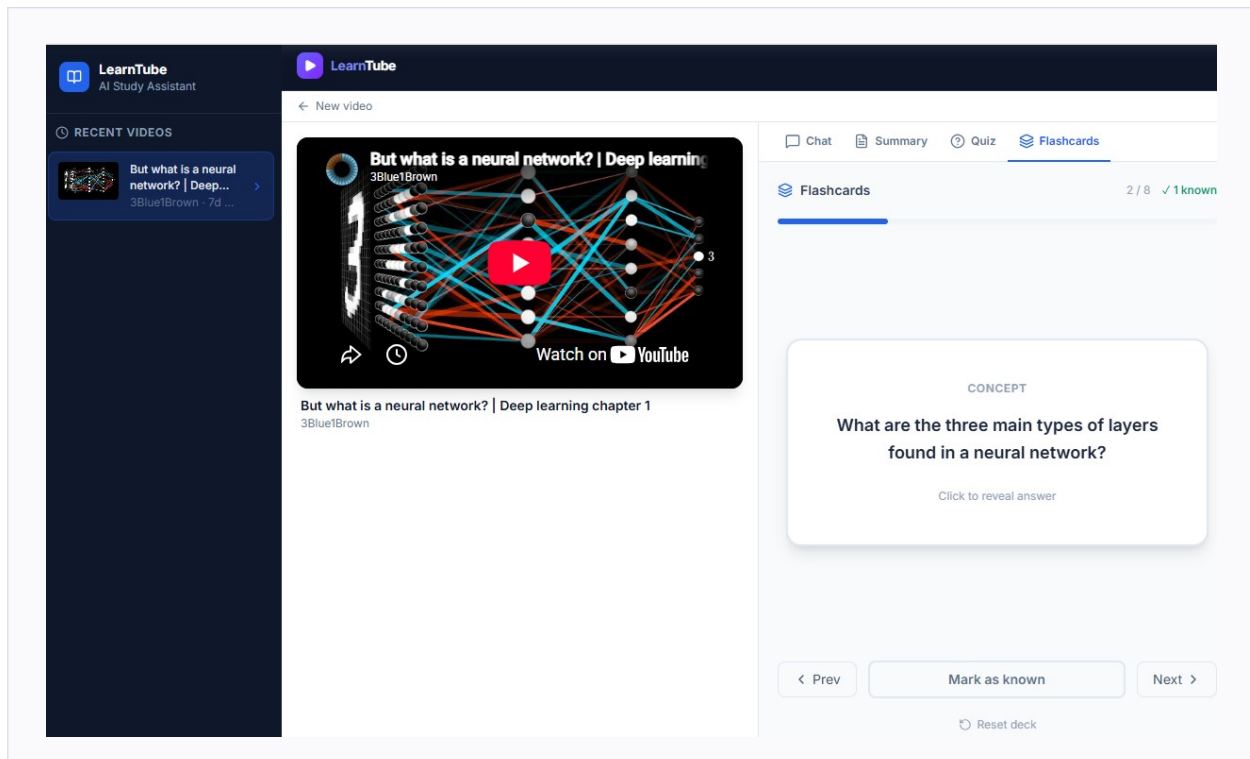


Fig. 10 — Flip-card flashcard deck with progress tracking and “mark as known” review

Mobile Preview (Expo Go)

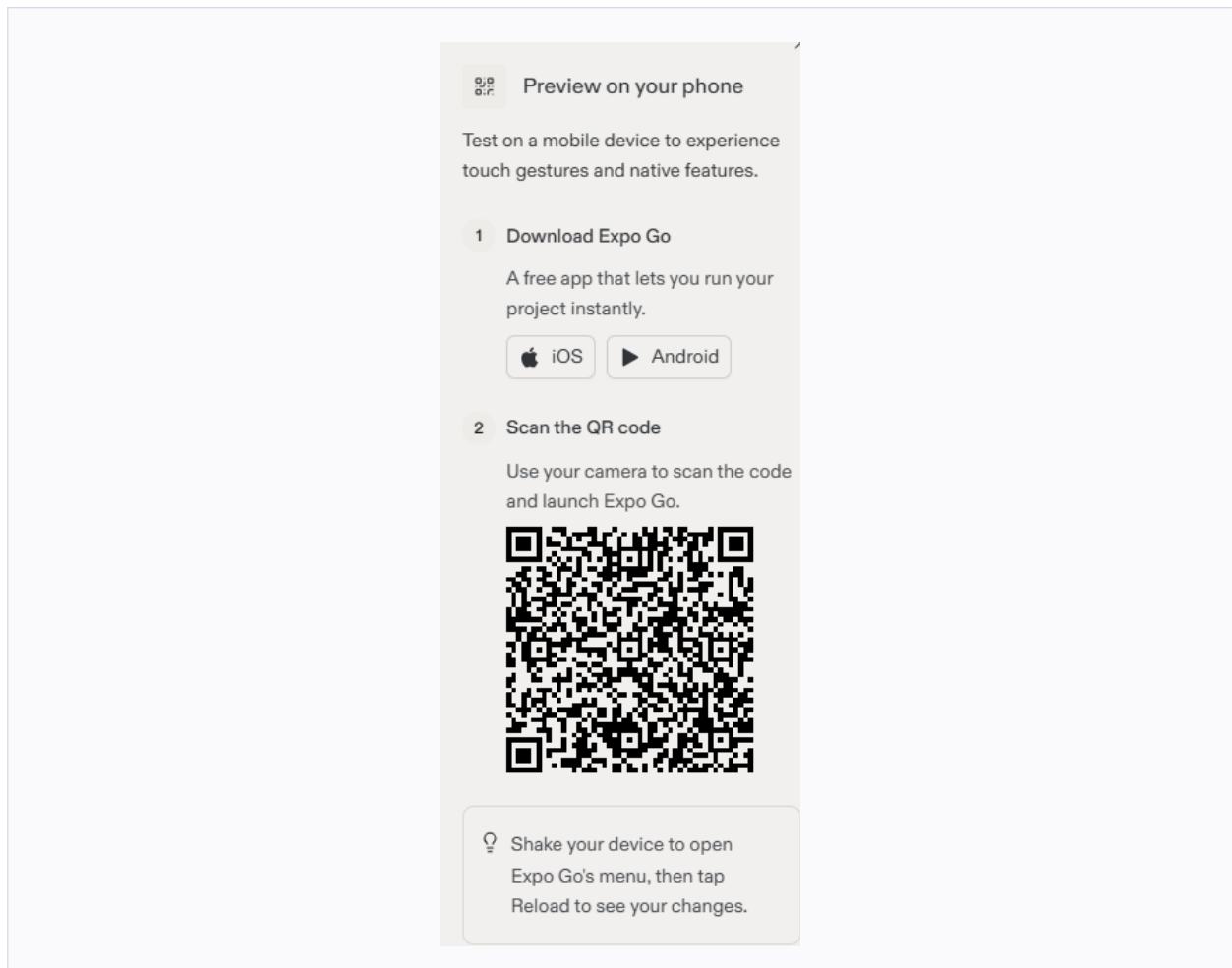


Fig. 11 — Scan-to-preview QR flow for testing the mobile companion app via Expo Go

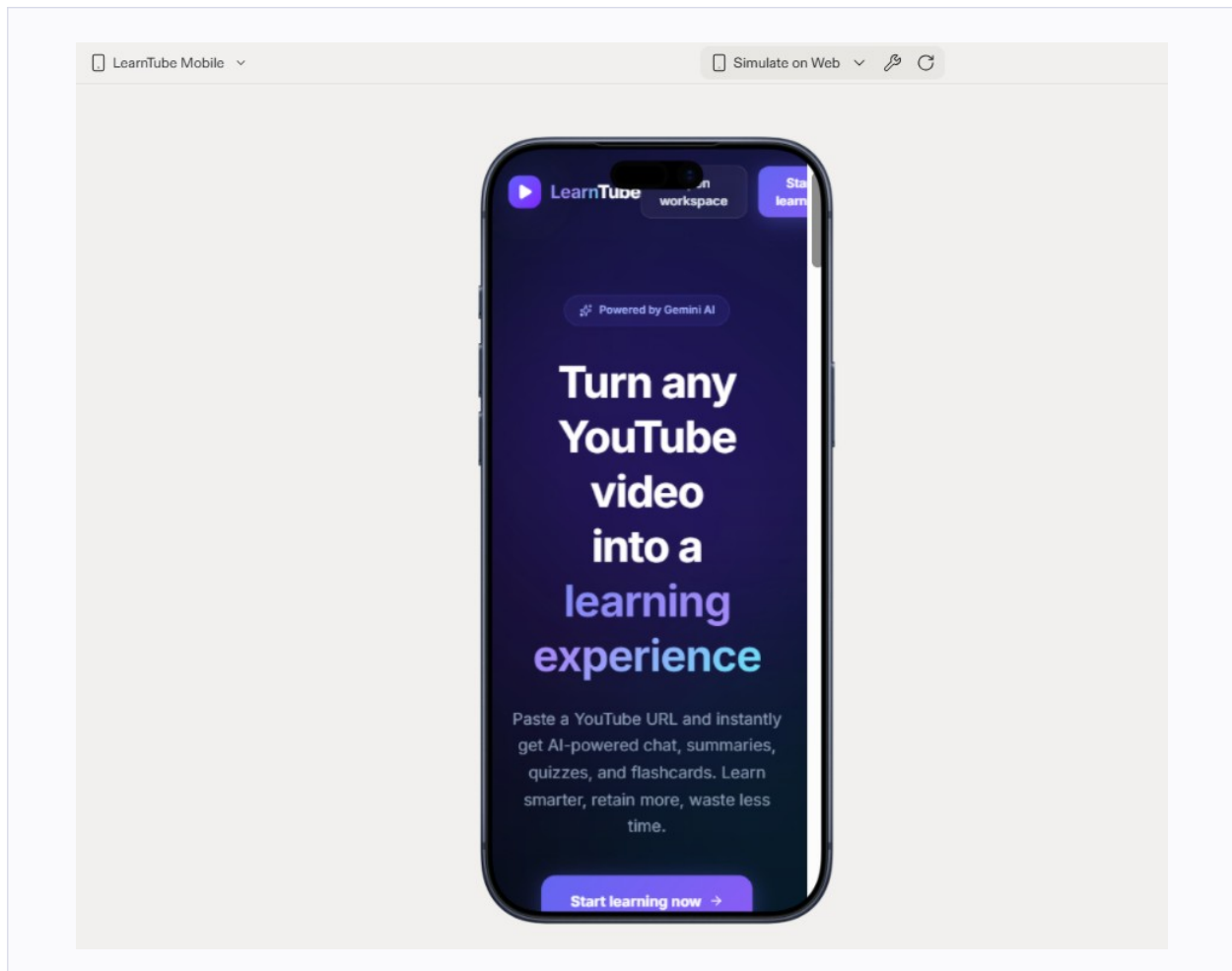


Fig. 12 — LearnTube Mobile landing screen, simulated on web via the Expo dev domain (*.expo.pike.replit.dev)

The LearnTube mobile app is currently in an early preview stage, testable via Expo Go. Full feature parity with the web app — including inline video playback, offline access, and native push notifications — will be developed further in future iterations.

6. Results & Key Achievements

The LearnTube application was successfully built, tested end-to-end, and deployed to production. All planned features were implemented and are fully functional:

Achievement	Details
Full RAG Pipeline	End-to-end transcript ingestion, chunking, embedding, retrieval, and LLM answer generation with timestamp citations.
AI Summary Engine	Gemini JSON-mode produces structured overviews, bullet-point key points, and inferred chapter timestamps.
Quiz Generator	5-10 multiple-choice questions with correct answers and explanations, tailored to each video.
Flashcard Generator	8-12 flip-card flashcards covering key concepts, ready for spaced-repetition study.
Intent Classifier	scikit-learn TF-IDF + Logistic Regression classifier routes queries with approximately 85% accuracy.
User Authentication	Clerk-managed auth with JWT verification, Google OAuth, forgot-password flow, and email/password login.
Video Library	Per-user persistent library of added videos stored in SQLite with full CRUD operations.
Timestamp Citations	Chat answers include [MM:SS] citations that seek the embedded YouTube player to the exact moment.
Production-Ready UI	Responsive, animated UI with Tailwind CSS v4, Framer Motion, shadcn/ui, and Clerk themes.
Gemini SDK Migration	Migrated from deprecated google-generativeai to the new google-genai SDK v2.17.
Mobile App (Expo)	Full Expo 53 / React Native 0.79 companion app for iOS and Android — Clerk auth, full feature parity.
Live Deployment	Application deployed to production at ai-you-tube-assistant.replit.app — publicly accessible.

Performance Metrics

Metric	Typical Time	Notes
Transcript ingestion	3-5 seconds	Depends on video length
Embedding generation	5-10 seconds	Local all-MiniLM-L6-v2 model
RAG chat latency	2-4 seconds	Gemini 2.0 Flash + top-5 retrieval
Summary generation	3-6 seconds	JSON mode, single LLM call
Quiz generation	4-8 seconds	JSON mode, 5-10 questions
Flashcard generation	3-6 seconds	JSON mode, 8-12 cards
Intent classification	<10 ms	Local scikit-learn model
Frontend bundle size	~1.2 MB	Vite production build

7. Future Approach

LearnTube has a clear product roadmap. The following improvements are planned for future development iterations:

Planned Feature	Description
Multi-Modal Understanding	Integrate Gemini's vision API to process video frames alongside transcripts, enabling understanding of on-screen visuals, diagrams, and slides.
Inline Video Playback (Mobile)	Embed the YouTube player directly inside the mobile app so learners can watch and interact with content without leaving the app.
Share Summary / Flashcards	Let learners share a video's AI-generated summary or flashcard deck with a friend via a short link.
Spaced Repetition System (SRS)	Build a full SM-2 algorithm scheduler into the flashcard system, tracking user performance per card.

Planned Feature	Description
YouTube Playlist / Course Mode	Allow users to add entire YouTube playlists or channels as structured “courses,” generating cross-video summaries and progress tracking.
Note Export (PDF / Notion)	One-click export of summaries, flashcards, and chat transcripts to PDF, Notion, or Markdown for offline study.
Multilingual Support	Extend transcript processing and LLM prompts to support non-English videos, with auto-detected language and translation.
Collaborative Study Rooms	Real-time multiplayer rooms where groups can watch a video together, share chat messages, and quiz each other.
Learning Analytics Dashboard	Track videos watched, quiz score trends, flashcard retention rates, and generate personalised study insights.
Fine-Tuned Embeddings	Replace sentence-transformers with an embedding model fine-tuned on educational content for improved retrieval accuracy.
YouTube Data API Integration	Migrate from transcript scraping to the official YouTube Data API for richer metadata — thumbnails, channel info, and view counts.

8. Conclusion

LearnTube demonstrates the practical power of Retrieval-Augmented Generation applied to a real-world educational problem. By combining a modern full-stack architecture (FastAPI + React + Expo) with cutting-edge AI (Google Gemini 2.0 Flash, ChromaDB, sentence-transformers), the project delivers a production-quality platform that genuinely transforms how learners interact with video content — on the web and on mobile.

The project showcases end-to-end ML engineering skills — from data ingestion and embedding to vector retrieval, LLM orchestration, cross-platform mobile development, and full-stack deployment. It also demonstrates awareness of production concerns: authentication, error handling, rate limiting, clean API design, and a polished, responsive user interface across both web and native mobile platforms.

The application is live at <https://ai-you-tube-assistant.replit.app> and available as a native mobile app via Expo Go. Most importantly, LearnTube solves a genuine pain point. It makes self-directed video learning

measurably more effective by adding the structured recall mechanisms — summarisation, active quizzing, and spaced repetition — that cognitive science shows are essential for long-term knowledge retention.

Repository:

github.com/numair-2003/AIML-Internship-NumairFahad/ai_youtube_learning_assistant

Live App: <https://ai-you-tube-assistant.replit.app>

Built with: FastAPI | React | Expo | Vite | Google Gemini 2.0 Flash | ChromaDB | Clerk | Tailwind CSS v4

Internship ID: ZYNVEX-CERT-0487

Student: Numair Fahad | AIML Internship | August 2026

Thank you for reviewing this project.