The Problem

A full-stack iOS + server project with a real customer, multimodal AI, and generative game design

Verlag J. Maiß — a Bavarian school publisher — needs a language-learning app for 7th/8th graders. Teachers cannot give individual feedback on vocabulary use, sentence structure, or pronunciation at scale. Existing tools offer static drills disconnected from the actual textbook.

StudyTrail Languages is a **new iOS app (SwiftUI)** on an existing server platform (Docker, PostgreSQL, AI pipeline). The core loop: a **teacher uploads a textbook** and **assigns a weekly vocabulary range**. The system then generates practice for each pupil:

- **AI conversations** where the pupil speaks and a **multimodal Large Language Model (LLM) evaluates pronunciation directly from the raw audio** — no speech-to-text, the model hears the pupil.
- **LLM-generated 2D adventures** — interactive scenarios created on the fly for each class, where pupils explore a scene and use the assigned words in context.
- **Spaced Repetition System (SRS)** for long-term vocabulary and verb form retention.

This is a full-stack project: SwiftUI on the client, Dockerized Python services on the server, multimodal LLM APIs (Azure OpenAI, Gemini), and a 2D game renderer on iOS. The app is language-configurable and targets Bavarian curriculum alignment.

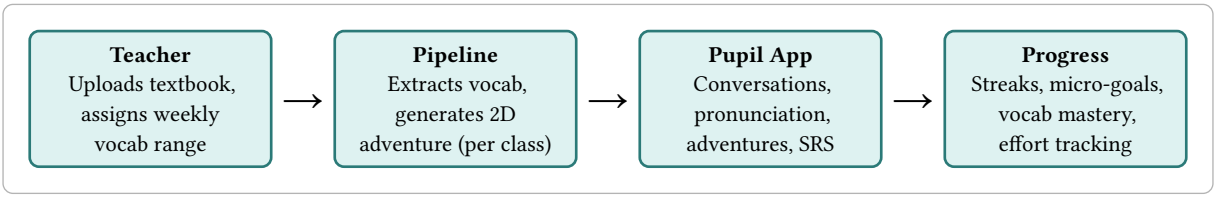

Figure 1: Core workflow

Scenarios

Scenario 1: Teacher Selects a Topic and Assigns Vocabulary

A teacher uploads a textbook PDF for the school year. The **processing pipeline extracts vocabulary and groups it by topic**. Before each week, the teacher selects a topic (e.g., *At the Restaurant*) with its linked vocabulary set and can adjust individual words. This topic drives all practice — conversations, adventures, and review — in the pupils' apps.

Scenario 2: AI Conversation with Pronunciation Feedback

A pupil opens a conversation with the AI tutor on a topic that uses the week's vocabulary — e.g., ordering food, asking for directions, or describing a weekend plan. The tutor asks questions, the pupil **responds in full sentences**, and the dialogue evolves naturally. Throughout the conversation the app **records audio and sends it directly to a multimodal LLM** — no speech-to-text. The model evaluates both **what** the pupil says (vocabulary, sentence structure) and **how** they say it (pronunciation, stress, intonation), giving **one supportive hint** at a time. Progress is tracked as effort-based streaks.

Scenario 3: LLM-Generated 2D Adventure

When the teacher assigns new vocabulary, the system **generates a 2D adventure for the class** — e.g., navigating a market or helping a character prepare for a trip. Pupils move through the scene, interact with characters, and **apply the assigned words in context**.

Top-Level Design

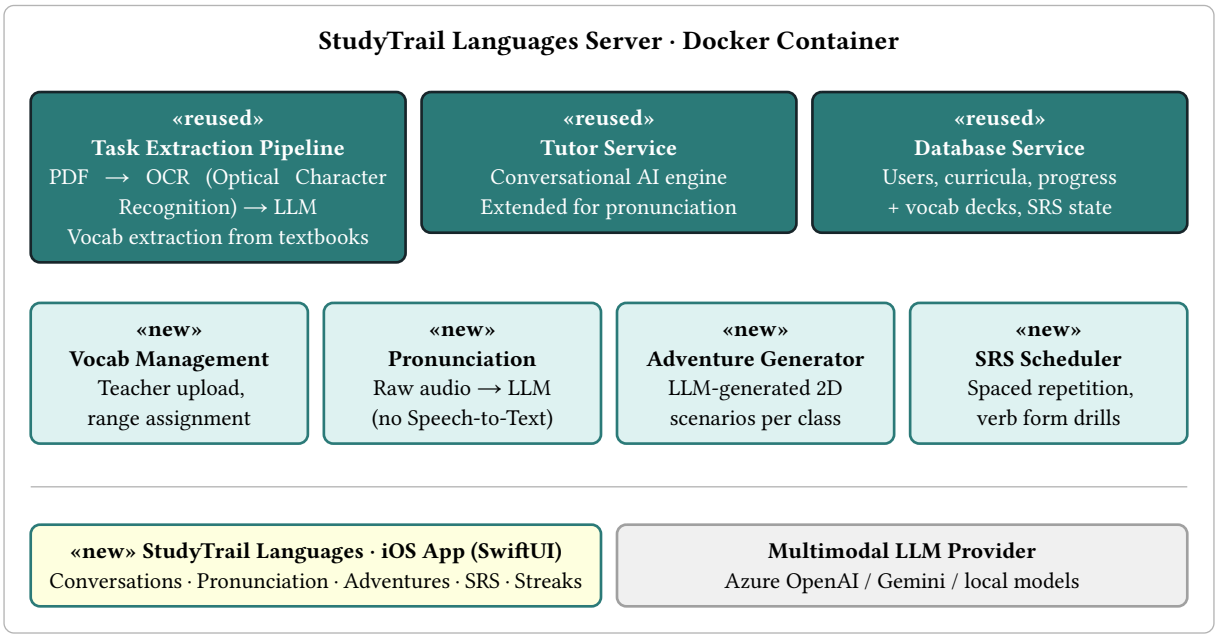

Figure 2: System architecture — reused services (dark) and new modules (light)

Requirements and Constraints

Functional Requirements

FR1: Teacher Vocabulary Upload & Weekly Assignment (High Priority)

Upload a textbook PDF or vocabulary list; extract items via processing pipeline; assign weekly ranges that drive all pupil practice modes.

FR2: Conversational AI with Audio-Native Pronunciation (High Priority)

AI-guided conversations using assigned vocabulary. Raw audio sent directly to a multimodal LLM for pronunciation evaluation — no speech-to-text. Supportive, age-appropriate feedback.

FR3: LLM-Generated 2D Adventures (High Priority)

Generate an interactive 2D adventure per class from the assigned vocabulary — 2D to keep on-the-fly generation feasible. Pupils explore scenes, interact with characters, and apply words in context.

FR4: Vocabulary Retention via Spaced Repetition (Medium Priority)
Personal vocabulary decks from assignments and conversations. Spaced repetition with verb forms and conjugations. Micro-goals and effort-based progress.

FR5: Listening & Reading Comprehension (Low Priority)
Comprehension tasks with scaffolding (slow playback, glossary, bounded translation). Only if time permits.

Quality Attributes

QA1: Performance
Near-real-time tutor responses and pronunciation feedback. Graceful fallback (text-only mode) when compute is constrained.

QA2: Security & Privacy
GDPR (General Data Protection Regulation) compliance. Explicit audio consent for minors. EU data residency. Encryption in transit/at rest.

QA3: Globalization & Localization
German UI baseline. Multiple target languages (configurable). Content packs per language and grade.

Constraints

- **Client:** SwiftUI, iOS, new app (built from scratch).
- **Server:** Dockerized services, PostgreSQL. Traefik as reverse proxy.
- **LLMs:** Azure OpenAI, Gemini, and/or local models — team decides.
- **Compliance:** GDPR, audio data, consent for minors.
- **Audience:** 7th/8th grade, Bavarian curriculum.

Development & Deliverables

New iOS app on the existing StudyTrail server infrastructure. Existing services (database, tutor, processing pipeline) are reused and extended. The iOS app is built from scratch.

Environment: Cloud LLM services (Azure OpenAI, Gemini) · VM for server (existing infrastructure) · Production Maiß API with sandbox credentials.

Deliverables:

- **iOS Pupil App (new):** Conversations, pronunciation, 2D adventures, SRS
- **Server Extensions:** Vocab pipeline, pronunciation module, adventure generator, SRS scheduler

Client acceptance: Requirements may change — the customer is open to reprioritization. Work is prioritized with the client in Sprints.

Schedule

Kickoff: Thu 16 Apr 2026, 17:00
Design Review (DR): Thu 4 Jun 2026, 17:00
Client Acceptance Test (CAT): Thu 16 Jul 2026, 17:00
Weekly SCRUM with project management. Bi-weekly customer meetings.