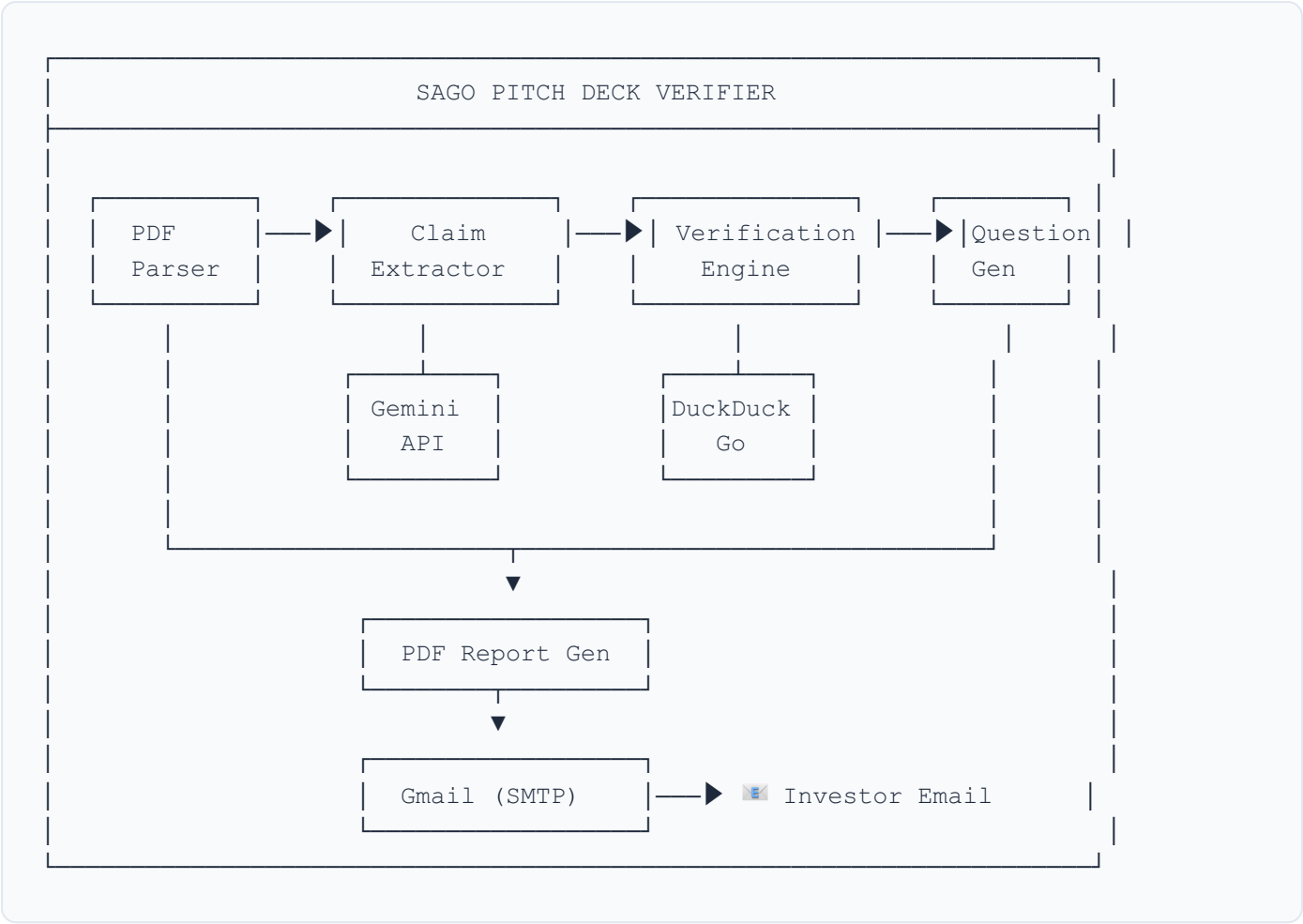


System Architecture

Sago Pitch Deck Verifier — AI-Powered Investment Due Diligence

1. Overview

The Sago Pitch Deck Verifier is an AI-powered tool designed to help investors perform due diligence on startup pitch decks. It automatically extracts claims, verifies them against public sources, and generates personalized questions for founder meetings.



2. Core Components

PDF Parser Extracts text from pitch deck PDFs using pdfplumber library	Claim Extractor Uses Gemini AI to identify verifiable claims from text
Verification Engine	Question Generator

Searches web via DuckDuckGo to verify claims

Creates personalized investor questions using AI

Report Generator

Produces professional PDF reports with WeasyPrint

Email Integration

Sends reports via Gmail SMTP with attachments

3. Design Principles (Sago Framework)

Seamless Integration

Works with existing tools (Gmail, PDF) without requiring new accounts or complex OAuth flows. Uses Gmail App Passwords for simple, secure authentication.

Hyper-Personalization

Questions are tailored to investor's focus areas (B2B SaaS, FinTech, etc.) and investment stage (Series A, Seed, etc.). The system adapts its output to individual investor profiles.

True Agency

Autonomous operation from PDF input to email delivery. The agent makes intelligent decisions about claim prioritization and question relevance without manual intervention.

4. Technology Stack

Component	Technology	Rationale
LLM Provider	Google Gemini API	Free tier with 15 RPM, excellent for claim extraction and question generation
Web Search	DuckDuckGo	Free, no API key required, privacy-focused
PDF Parsing	pdfplumber	Reliable text extraction with layout preservation
PDF Generation	WeasyPrint	HTML to PDF conversion with CSS support
Email Delivery	Gmail SMTP	Free, reliable, no OAuth complexity with App Passwords
Language	Python 3.10+	Rich ecosystem for AI/ML and document processing

5. Data Flow

Input Processing

- **PDF Upload:** User provides pitch deck PDF file
- **Text Extraction:** pdfplumber extracts text page-by-page
- **Claim Identification:** Gemini AI identifies 10-15 verifiable claims

Verification Pipeline

- **Web Search:** Each claim is searched via DuckDuckGo (with rate limiting)
- **Evidence Collection:** Top 5 search results analyzed per claim
- **AI Verification:** Gemini compares claim against evidence
- **Score Calculation:** Overall verification score computed (0-100%)

Output Generation

- **Question Generation:** AI creates prioritized questions based on findings
- **PDF Report:** Professional report generated with claims, scores, and questions
- **Email Delivery:** Report sent to configured recipient via Gmail SMTP

6. Security Considerations

- **Credentials:** All API keys and passwords stored in .env file (gitignored)
- **Gmail Auth:** Uses App Passwords instead of OAuth for simplicity and security
- **Data Privacy:** No pitch deck data stored permanently; processed in memory
- **Rate Limiting:** Built-in delays prevent API quota exhaustion

7. Scalability & Future Enhancements

- Support for additional LLM providers (OpenAI, Anthropic)
- Slack integration for team notifications
- Batch processing for multiple pitch decks
- Database storage for historical analysis comparison
- Web interface for non-technical users