# Alan Ranger Photography Chat Bot - System Architecture

## 🏗️ \*\*System Overview\*\*

The Alan Ranger Photography Chat Bot is a \*\*RAG (Retrieval-Augmented Generation) system\*\* that combines web scraping, data ingestion, and AI-powered responses to provide intelligent assistance about photography courses, workshops, equipment, and technical advice.

---

## 📊 \*\*Data Flow Architecture\*\*

### \*\*1. Data Ingestion Pipeline\*\*

```

Website URLs → Web Scraping → Data Processing → Supabase Database

↓ ↓ ↓ ↓

CSV Files → HTML Extraction → JSON-LD Parsing → page\_entities

↓ ↓ ↓ ↓

Site URLs → Text Chunking → Embeddings → page\_chunks

```

### \*\*2. Chat Query Processing\*\*

```

User Query → Intent Detection → Data Retrieval → Response Generation → Frontend Rendering

↓ ↓ ↓ ↓ ↓

Chat API → Topic Analysis → Supabase Query → Markdown Generation → HTML Display

```

---

## 🗄️ \*\*Database Schema (Supabase)\*\*

### \*\*Core Tables\*\*

- \*\*`page\_entities`\*\* - Structured data from JSON-LD (products, events, articles)

- \*\*`page\_chunks`\*\* - Text chunks with embeddings for RAG

- \*\*`csv\_events\_data`\*\* - Event data from CSV files

- \*\*`light\_refresh\_runs`\*\* - Logs for automated refresh system

- \*\*`url\_last\_processed`\*\* - Change detection for incremental updates

### \*\*Key Views\*\*

- \*\*`v\_events\_for\_chat`\*\* - Combined event data (CSV + scraped)

- \*\*`v\_blog\_content`\*\* - Blog posts with categories/tags

- \*\*`v\_workshop\_content`\*\* - Workshop pages with location data

- \*\*`v\_product\_content`\*\* - Product catalog with pricing

---

## 🔧 \*\*API Endpoints\*\*

### \*\*Core APIs\*\*

- \*\*`/api/chat`\*\* - Main chat interface (RAG + response generation)

- \*\*`/api/ingest`\*\* - Web scraping and data ingestion

- \*\*`/api/tools`\*\* - Utility functions (mapping, finalization)

- \*\*`/api/light-refresh`\*\* - Automated hourly refresh system

### \*\*Data Management\*\*

- \*\*`/api/feedback`\*\* - User feedback collection

- \*\*`/api/analytics`\*\* - Usage analytics and metrics

---

## 🎯 \*\*Intent Detection & Routing\*\*

### \*\*Query Classification\*\*

```

User Query → Intent Detection → Data Source → Response Type

↓ ↓ ↓ ↓

"ISO question" → technical → articles → Direct Answer

"Camera course" → events → products+events → Product Card + Event Tiles

"Workshop dates" → events → events → Event Listings

"Equipment advice" → equipment → articles+products → Recommendations

```

### \*\*Intent Types\*\*

- \*\*`events`\*\* - Course/workshop queries → Product cards + event tiles

- \*\*`equipment`\*\* - Equipment questions → Technical advice + recommendations

- \*\*`technical`\*\* - Technical questions → Direct answers from articles

- \*\*`general`\*\* - General questions → Mixed content types

---

## 🎨 \*\*Frontend Components\*\*

### \*\*Chat Interface (`chat.html`)\*\*

- \*\*Real-time chat\*\* with typing indicators

- \*\*Markdown rendering\*\* for rich text responses

- \*\*Product cards\*\* with pricing and details

- \*\*Event tiles\*\* with dates and locations

- \*\*Debug panel\*\* for development

### \*\*Admin Interface (`bulk-simple.html`)\*\*

- \*\*CSV upload\*\* for data management

- \*\*Web scraping\*\* controls

- \*\*Mapping tools\*\* for data relationships

- \*\*Light refresh\*\* monitoring

---

## 🔄 \*\*Automated Systems\*\*

### \*\*Light Refresh System\*\*

```

Hourly Trigger → URL Change Detection → Selective Ingestion → Data Update

↓ ↓ ↓ ↓

GitHub Actions → HEAD Requests → Changed URLs → Supabase Update

```

### \*\*Data Pipeline\*\*

```

CSV Upload → Bulk Processing → Entity Mapping → View Refresh → Chat Ready

↓ ↓ ↓ ↓ ↓

File Upload → URL Processing → Data Enhancement → Finalization → Live System

```

---

## 🧠 \*\*RAG (Retrieval-Augmented Generation) System\*\*

### \*\*Content Retrieval\*\*

1. \*\*Query Analysis\*\* - Extract keywords and intent

2. \*\*Vector Search\*\* - Find relevant chunks using embeddings

3. \*\*Context Building\*\* - Combine multiple sources

4. \*\*Response Generation\*\* - Create structured markdown

### \*\*Data Sources\*\*

- \*\*Articles\*\* - Blog posts for technical questions

- \*\*Products\*\* - Course/workshop information

- \*\*Events\*\* - Scheduled dates and locations

- \*\*Services\*\* - Service descriptions and details

---

## 🎛️ \*\*Configuration & Environment\*\*

### \*\*Environment Variables\*\*

- \*\*`SUPABASE\_URL`\*\* - Database connection

- \*\*`SUPABASE\_SERVICE\_ROLE\_KEY`\*\* - Database access

- \*\*`INGEST\_TOKEN`\*\* - API authentication

- \*\*`VERCEL\_AUTOMATION\_BYPASS\_SECRET`\*\* - Deployment protection

### \*\*External Services\*\*

- \*\*Supabase\*\* - Database and authentication

- \*\*Vercel\*\* - Hosting and serverless functions

- \*\*GitHub Actions\*\* - Automated workflows

- \*\*OpenAI\*\* - Embeddings and AI processing

---

## 🔍 \*\*Debugging & Monitoring\*\*

### \*\*Debug Tools\*\*

- \*\*Chat debug panel\*\* - Real-time query analysis

- \*\*API logging\*\* - Request/response tracking

- \*\*Database queries\*\* - Direct data inspection

- \*\*Light refresh logs\*\* - Automated system monitoring

### \*\*Performance Metrics\*\*

- \*\*Response time\*\* - Chat API performance

- \*\*Data freshness\*\* - Last ingestion timestamps

- \*\*System health\*\* - Error rates and success metrics

---

## 🚨 \*\*Known Issues & Technical Debt\*\*

### \*\*Current Problems\*\*

- \*\*Equipment Needed extraction\*\* - Inconsistent from page\_entities

- \*\*Light refresh system\*\* - Timeout issues with large URL sets

- \*\*RAG quality degradation\*\* - Over-complexity reducing effectiveness

- \*\*Data inconsistency\*\* - Multiple sources not properly synchronized

### \*\*Architecture Improvements Needed\*\*

- \*\*Simplified RAG system\*\* - Back to core functionality

- \*\*Raw HTML extraction\*\* - Alternative to htmlToText

- \*\*Better error handling\*\* - Comprehensive failure recovery

- \*\*Performance optimization\*\* - Faster response times

---

## 📈 \*\*System Performance\*\*

### \*\*Current Metrics\*\*

- \*\*135 products\*\* in database

- \*\*3 products\*\* with Equipment Needed

- \*\*81 products\*\* with Participants data

- \*\*~425 URLs\*\* in light refresh system

- \*\*30-minute\*\* full import process

### \*\*Optimization Targets\*\*

- \*\*Sub-2 second\*\* chat response times

- \*\*Hourly\*\* automated data refresh

- \*\*99%+\*\* data accuracy

- \*\*Comprehensive\*\* Equipment Needed extraction

---

\*This architecture document provides a complete overview of the system for reference during development and debugging.\*