**Sovereign Architecture Walkthrough for Team**

**System Overview (Simple Explanation)**

Think of Sovereign like a smart compliance assistant that reads documents and analyzes AI systems in 5 steps, then creates professional reports.

**Architecture Components**

**1. Frontend (What Users See)**

* **Website**: <https://jade-tarsier-55077b.netlify.app/>
* **Built with**: React (HTML + CSS + JavaScript)
* **Hosted on**: Netlify
* **What it does**: Upload files, show results, download reports

**2. Backend (The Brain)**

* **API Server**: Your Railway deployment
* **Built with**: Python Flask
* **What it does**: Process files, run AI analysis, generate exports

**3. AI Engine (The Intelligence)**

* **Service**: Claude API (Anthropic)
* **What it does**: 5-step GDPR analysis using smart prompts

**4. Document Processing**

* **Technology**: PyPDF2 (text extraction)
* **What it does**: Reads PDF files and extracts text

**Data Flow (Step-by-Step)**

**Step 1: User Uploads Document**

User → Frontend → "Upload PDF" → Backend → Extract text → Store temporarily

**Step 2: User Describes AI System**

User → Frontend → Types description → Stored in browser

**Step 3: Analysis Request**

Frontend → Backend → "Analyze this AI system"

Backend receives: PDF text + AI description

**Step 4: 5-Step AI Analysis**

Backend → Claude API → Step 1: GDPR Risk Assessment

Backend → Claude API → Step 2: Document Cross-Reference

Backend → Claude API → Step 3: Bias Analysis

Backend → Claude API → Step 4: Ethics Review

Backend → Claude API → Step 5: Implementation Plan

**Step 5: Results Display**

Backend → Combines all 5 results → Sends to Frontend

Frontend → Shows risk score, violations, recommendations

**Step 6: Export Reports**

User clicks export → Backend → Generates PDF/CSV → Downloads to user

**Technical Architecture Diagram**

[USER]

↓ uploads PDF & AI description

[FRONTEND - Netlify]

↓ sends data via API calls

[BACKEND - Railway]

↓ extracts text from PDF

[PDF PROCESSOR]

↓ sends 5 sequential prompts

[CLAUDE API]

↓ returns analysis results

[BACKEND - Railway]

↓ combines results & generates reports

[EXPORT ENGINE]

↓ sends final results

[FRONTEND - Netlify]

↓ displays results & downloads

[USER]

**How Claude API Analysis Works - Technical Deep Dive**

**Step-by-Step Analysis Logic**

**Step 1: GDPR Risk Assessment**

**What Claude receives:**

* Privacy policy text
* AI system description
* Pre-written prompt with GDPR legal knowledge

**Assessment logic:**

* Compares AI system activities against specific GDPR articles (5, 6, 9, 13, 22, 25)
* Identifies violations based on legal precedents built into prompt
* Scores severity using EU enforcement patterns

**Risk scoring factors:**

* Article 22 violations (automated decisions) = High risk (+3 points)
* Special category data (Article 9) = High risk (+2-3 points)
* Missing legal basis (Article 6) = Medium risk (+2 points)
* Transparency gaps (Article 13) = Medium risk (+1-2 points)

**Step 2: Document Cross-Reference**

**What Claude receives:**

* Same policy text and AI description
* Prompt instructing line-by-line comparison

**Assessment logic:**

* Maps each AI system function to policy statements
* Identifies missing disclosures (policy doesn't mention what AI does)
* Finds conflicts (policy says one thing, AI does another)
* Gaps found = Risk score increase

**Step 3: Bias & Fairness Analysis**

**What Claude receives:**

* AI system description only
* Prompt with bias detection framework

**Assessment logic:**

* Scans for bias-prone data types (facial analysis, voice, demographics)
* Identifies protected groups at risk
* Evaluates fairness principles (equal opportunity, demographic parity)
* High bias risk = Risk score increase

**Step 4: Ethics & Governance Review**

**What Claude receives:**

* AI system description
* Prompt with ethics evaluation framework

**Assessment logic:**

* Checks for human oversight mechanisms
* Evaluates transparency and explainability
* Assesses accountability structures
* Missing governance = Risk score increase

**Step 5: Implementation Planning**

**What Claude receives:**

* Combined findings from steps 1-4
* Prompt for strategic planning

**Assessment logic:**

* Prioritizes actions by legal risk + implementation complexity
* Creates realistic timelines based on violation severity
* Assigns ownership based on typical organizational structure
* Estimates costs using industry benchmarks

**How Claude API Works**

**What is Claude?**

* **Large Language Model**: AI trained on vast legal and compliance text
* **Pre-trained knowledge**: Already knows GDPR articles, case law, best practices
* **No additional training needed**: Knowledge is built-in from training data

**How We Use Claude:**

Our Backend → Sends structured prompt to Claude API

Claude → Processes using built-in legal knowledge

Claude → Returns structured JSON analysis

Our Backend → Receives and formats results

**Prompt Engineering Strategy:**

* **Legal expertise embedded**: Prompts contain GDPR article references
* **Structured output**: Forces JSON format for consistent results
* **Context-aware**: Each prompt builds on previous analysis
* **Quality control**: Prompts include severity guidelines and examples

**Where GDPR Knowledge Comes From**

**Built into Claude's Training:**

* **Legal databases**: Claude was trained on legal texts, case law, regulations
* **GDPR articles**: Full text of GDPR already in Claude's knowledge
* **Enforcement patterns**: Knows typical violation types and penalties
* **Best practices**: Understands compliance frameworks and standards

**Enhanced by Our Prompts:**

* **Specific focus**: Our prompts direct Claude to specific GDPR articles
* **Structured analysis**: Prompts ensure consistent evaluation framework
* **Practical application**: Prompts connect legal theory to real AI systems
* **Risk prioritization**: Prompts include severity assessment guidelines

**Risk Score Calculation Logic**

**Risk Level Assignment:**

* **Score 8-10**: HIGH RISK (Critical violations, immediate action needed)
* **Score 6-7**: MEDIUM-HIGH RISK (Significant issues, plan remediation)
* **Score 4-5**: MEDIUM RISK (Some concerns, monitor and improve)
* **Score 1-3**: LOW RISK (Minor issues, routine maintenance)

**Example: Hiring AI Analysis Flow**

**Input to Claude:**

PRIVACY\_POLICY: "We collect resumes for hiring purposes..."

AI\_SYSTEM: "Hiring AI analyzes facial expressions and automatically rejects candidates..."

**Claude's Analysis Process:**

1. **Recognizes** facial analysis = biometric data (Article 9 violation)
2. **Identifies** automated rejection = Article 22 violation
3. **Notes** policy doesn't mention AI or facial analysis (transparency gap)
4. **Assesses** bias risk from facial analysis (high discrimination potential)
5. **Finds** no human oversight mentioned (governance gap)

**Risk Score Calculation:**

* GDPR: 8/10 (major violations)
* Cross-ref: 7/10 (significant gaps)
* Bias: 9/10 (facial analysis high risk)
* Ethics: 8/10 (no oversight)
* **Final Score: 8/10 (HIGH RISK)**

**Why This Approach Works**

**Leverages Pre-trained Knowledge:**

* Claude already knows GDPR law
* No need to manually code legal rules
* Updates automatically as Claude improves

**Structured and Consistent:**

* Same analysis framework for every AI system
* Repeatable results
* Professional quality outputs

**Contextually Aware:**

* Analyzes specific AI system against specific policy
* Identifies gaps unique to that combination
* Provides targeted recommendations

**Scalable and Maintainable:**

* Works for any AI system type
* Easy to update prompts for new regulations
* No complex rule engines to maintain