



Project Report: Startup Legal Risk Assistant



Problem Statement

Many startups ignore essential legal requirements such as **data privacy**, **intellectual property (IP)** protection, and **business registration** during their early stages. This can lead to serious legal and financial consequences later.



Objective

To build an intelligent system that:

- Understands what kind of startup it is (e.g., **fintech**, **healthtech**, **edtech**).
 - Detects potential legal **red zones** like HIPAA, KYC, GDPR, or IP laws.
 - Provides a **custom legal checklist** of tasks for compliance.
 - Offers **jurisdiction-specific legal resources** based on where the startup operates.
 - *(Optional)* Uses **RAG (Retrieval-Augmented Generation)** to pull real-time legal documents or startup guides from trusted sources.
-



Key Features

1. Startup Classifier

Detects startup domain from a given description using:

- GPT model or
- Keyword matching logic

2. Red Zone Detector

Based on the domain and region, flags:

- Data protection laws (e.g., GDPR, HIPAA)

- Regulatory requirements (e.g., RBI, PCI-DSS)
- IP and licensing concerns

3. **Legal Checklist Generator**

Produces a tailored checklist:

- Company registration
- Terms & Conditions, Privacy Policy
- IP trademark registration
- Compliance with sector-specific rules

4. **Jurisdiction-Specific Resources**

Provides government links and law documentation for:






- India, US, EU, etc.

5. **(Optional) RAG Component**

Retrieves and summarizes:

- Legal PDFs (Startup India, EU Startup Guide)
- Law portals (gdpr.eu, rbi.org, etc.)
- Other up-to-date regulation documents

Technology Stack

Layer	Technology Used
 NLP & Logic	OpenAI GPT / LangChain / Custom classifiers
 Backend	Python
 Frontend	Streamlit (<i>for fast prototyping UI</i>)
 Document Tools	LangChain PDFLoader / FAISS for RAG (optional)
 APIs	SerpAPI or custom scraper (for optional RAG)



Storage

JSON templates for checklists & law references



Tools & Libraries

- **OpenAI GPT-4** for semantic understanding
 - **LangChain** for chaining steps and RAG integration
 - **Streamlit** for rapid frontend interface
 - **PDFLoader / Web Search** (*optional*) to load legal documents dynamically
 - **FAISS / ChromaDB** for vector store & retrieval
-



Sample Input

"We are building a health app that helps patients track symptoms and stores medical records securely. It will be used across Europe and India."



Sample Output

json

CopyEdit

```
{
  "startup_type": "Healthtech",
  "jurisdiction": ["EU", "India"],
  "legal_red_zones": [
    "GDPR compliance (EU)",
    "HIPAA (if US market expansion)",
    "Doctor-patient data confidentiality",
    "Encryption of health data"
  ],
  "legal_checklist": [
    "Register business in EU and India",
    "Create a GDPR-compliant privacy policy",
    "Encrypt health records in storage and transit",
```

```
    "Define clear user consent mechanism",
    "Sign NDAs with doctors and partners"
  ],
  "resources": [
    {"title": "GDPR Official Site", "url": "https://gdpr.eu"},
    {"title": "India's Digital Health Guidelines", "url":
"https://ndhm.gov.in"},
    {"title": "Startup India Portal", "url":
"https://www.startupindia.gov.in"}
  ]
}
```

Agent Architecture (Modular Roles)

Agent Name	Responsibility
Classifier Agent	Identify startup category (e.g., fintech, edtech)
Risk Agent	Flag legal red zones based on type and location
Checklist Agent	Generate personalized legal checklist
RAG Agent	<i>(Optional)</i> Fetch relevant guides from web/docs

System Workflow

1. **User** enters a short description of their startup.
2. The **system classifies** the startup type and operational region.
3. **Legal risks** and compliance issues are flagged.
4. A **legal checklist** is generated based on startup type and region.
5. Jurisdiction-specific **resource links** are provided.
6. *(Optional)* RAG module pulls additional documents and law texts for reference.

Conclusion

This system helps startups **become legally compliant early**, reducing the risk of future legal problems. It acts as a **virtual legal assistant** for founders who may not be aware of the rules in their domain or jurisdiction.

It uses modern AI tools like GPT, LangChain, and RAG to build an intelligent, explainable, and expandable legal advisor.

Github link: https://github.com/hariharan545/Agentic_AI_Workshop.git