NLP-TA

Team Members:
Manan Khandelwal
Aakanshi Gera
Sagar Sindhu
Lakshya Bindal

1. Title of the Project

SamvidhanAI (AI-Based Legal Contract Generator)

2. Problem Statement & Objectives

Problem Statement:

Legal contract drafting is a critical but time-consuming for businesses, legal professionals, and individuals. Traditional contract generation relies heavily on manual drafting, which introduces several challenges:

- **Time-Consuming:** Lawyers spend hours drafting and reviewing standard contracts.
- **High Legal Costs:** Small businesses and startups often struggle with the cost of hiring legal experts for routine contracts.
- **Human Error & Inconsistencies:** Manual entry can lead to errors, missing clauses, and discrepancies.
- **Scalability Issues:** Large firms handling multiple contracts daily struggle to manage consistency and compliance.

Real-World Example Scenarios:

1. **Startups & Freelancers:** A startup needs to draft NDAs, employment agreements, and partnership contracts frequently. Hiring a lawyer for each document is costly.
2. **Real Estate Industry:** Property lease agreements must be generated with accurate tenant details, payment terms, and jurisdictional laws.
3. **E-commerce & SaaS Companies:** Subscription-based services require contracts with consistent and legally sound terms.

4. **HR Departments:** Employment contracts for new hires often follow a standard format but require customization for names, roles, and benefits.

Novelty in Our Project:

1. **Rule-Based NLP for Contract Generation:** Unlike traditional template-based methods, our system intelligently extracts relevant entities and ensures they fit logically within predefined templates.
2. **Automation Without LLMs:** Unlike expensive generative AI models, this system is lightweight, cost-effective, and interpretable.

---

3. Step-by-Step Methodology

Step 1: Data Collection & Preprocessing

- **Dataset:** CaseHOLD (Case Holdings On Legal Decisions)

  CaseHOLD is a legal dataset with **53,000+ multiple-choice questions** designed to test AI's ability to identify the correct legal ruling (holding) of a cited case. Each question provides a **legal excerpt** with five possible rulings—one correct and four incorrect. It helps train AI models for legal reasoning and case law analysis.

- **Data Cleaning and Preprocessing:**
  - Remove stop words and legal jargon.
  - Normalize text (lowercasing, punctuation removal, etc.).
- **Tokenization & Entity Tagging:**
  - Apply **Spacy** for Named Entity Recognition (NER).
  - Use Regex for pattern matching (e.g., dates, names, contract duration).

Step 2: Named Entity Recognition (NER) for Clause Extraction

- **Tools Used:** Spacy, Regex, NLTK
- **Entities to Extract:**

  ➔ **PARTIES:** Names of individuals or companies

- ➔ **DATES:** Contract start and end dates
- ➔ **DURATION:** Contract validity
- ➔ **PAYMENT TERMS:** Financial obligations
- ➔ **GOVERNING LAW:** Applicable jurisdiction

Step 3: Contract Template Generation

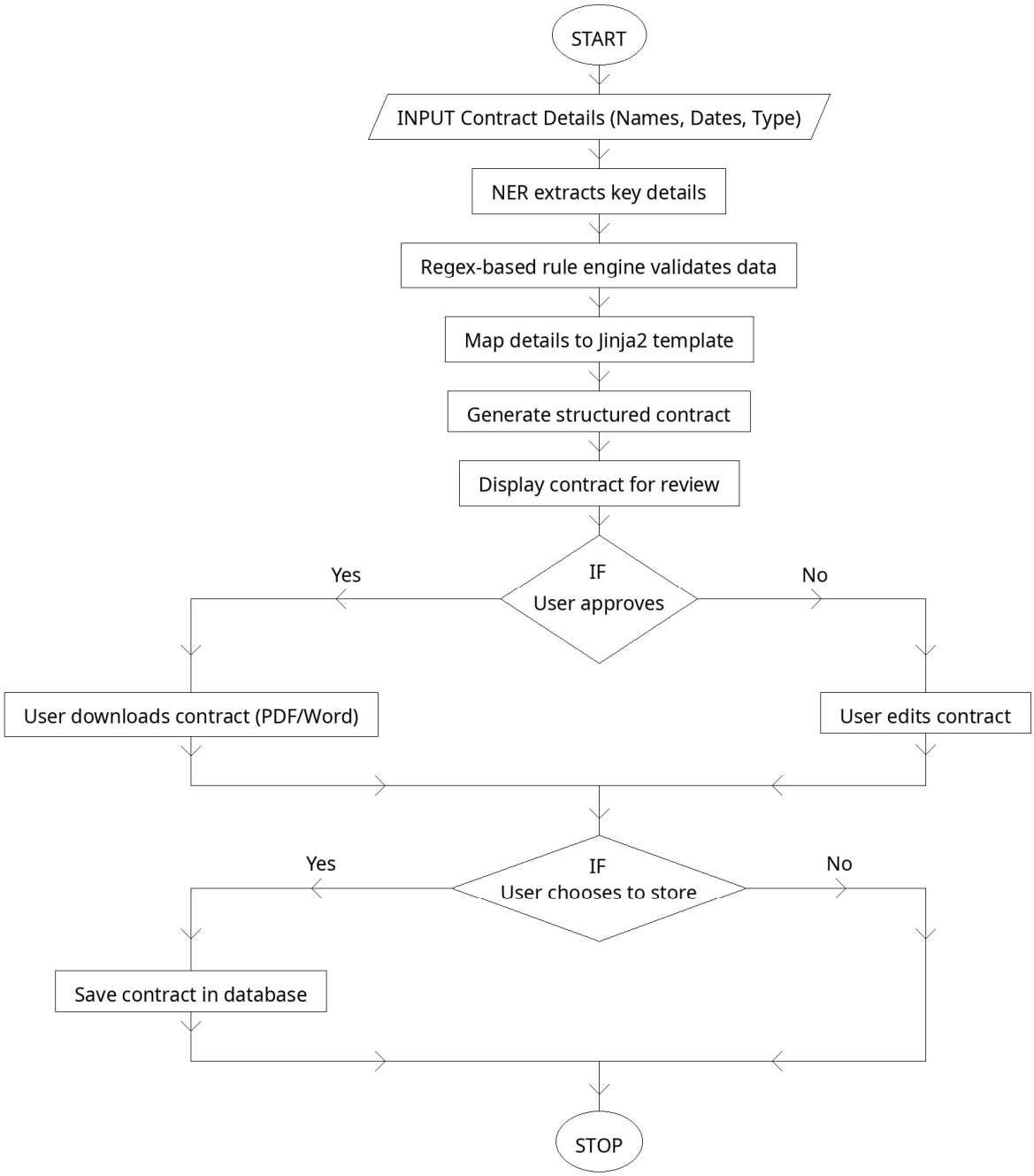
- **Technology Used:** Jinja2 Template (A predefined format for legal documents)
- **Process:**
  - ➔ Create predefined templates with placeholders.
  - ➔ Dynamically fill placeholders with extracted data.

Step 4: Model Evaluation & Improvement

- **Performance Metrics:**
  - ➔ Precision & Recall (for NER accuracy)
  - ➔ Execution Time (contract generation speed)
  - ➔ User Feedback & Iteration
- **Enhancements:**
  - ➔ Improve regex rules.
  - ➔ Expand contract templates for diverse legal use cases.

---

4. System Diagrams



5. Architecture Design