Technical Annex (High-Level Architecture): AI Decision-Support Platform for Regulatory Mapping and Compliance Cost Estimation

Project: NLP Logic Engine for Regulatory Obligation Extraction and Automated Economic Estimation

Lead Developer: The Citizen's Empowerment Center in Israel (CECI)

Strategic Partner: The Israeli Regulatory Authority, Prime Minister’s Office

1. Introduction & Strategic Context

This document provides a preliminary technical framework (Abstract) for an AI-driven platform designed to support regulatory design, review, and simplification processes. The system is conceived as a **Decision Support Tool** for regulators and economists, rather than an autonomous legal agent, ensuring that final professional judgment remains with human experts.

The platform's outputs are intended to feed into established government workflows recognized by the OECD, such as **Regulatory Impact Analysis (RIA)**, ex-post reviews of existing regulations, and cross-government initiatives to reduce administrative burdens.

2. Detailed Pipeline Architecture

A. Data Ingestion & Pre-processing (Hebrew Corpus)

- **Official Data Sources:** Automated scheduled scraping of the official Israeli Gazette ("Reshumot"), the National Legislative Database, and the "Nevo" legal database.
- **OCR & Layout Analysis:** Utilizing vision-based models to preserve document hierarchy (e.g., nested clauses, sub-clauses, and tables) within government PDFs.
- **Domain Adaptation:** The architecture supports both general-purpose LLMs and domain-specific models optimized for Hebrew legal and regulatory syntax, addressing the unique linguistic complexities of the Israeli regulatory environment.

B. NLP Logic Engine (Expert-Guided RAG Framework)

To **significantly reduce** hallucinations and improve the reliability of obligation extraction, the system utilizes a **Retrieval-Augmented Generation (RAG)** framework:

- **Semantic Chunking:** Breaking down legal texts into granular segments while maintaining hierarchical context (Parent-Child relationships).
- **Context-Aware Retrieval:** Using vector stores to ensure the AI's outputs are always grounded in and cited from specific legal sources.
- **Chain of Thought (CoT) Reasoning:** Implementing multi-step prompting to identify, categorize, and extract regulatory requirements (e.g., permits, reporting, or inspections) in an auditable manner.

C. Economic Quantification Logic (SCM)

The system supports the structured decomposition of regulatory obligations into measurable administrative activities:

- **Identifying Cost Drivers:** Detecting requirements that necessitate fees, lab tests, professional consultants, or waiting periods to provide preliminary data points.
- **Quantification Support:** The system provides the database for calculations based on the **Standard Cost Model (SCM)**. Final estimates rely on validated assumptions and expert judgment by Authority economists.
- **Methodological Alignment:** The quantification process refers to the Official Israeli Regulatory Guide and its accompanying tools.

3. Core Strategy: Human-in-the-Loop (HITL)

The platform is built around **continuous expert involvement** as an integral part of its operational logic:

- **Validation & Calibration Interface:** Economists and legal professionals review and correct AI outputs directly within the administrative dashboard.
- **Professional Accountability:** Every AI-generated data point includes a direct legal citation, enabling rapid human verification and ensuring alignment with real-world regulatory practice.
- **Continuous Improvement (RLHF):** Expert feedback is used to fine-tune prompts and model parameters, acknowledging the need for technological humility when facing legal complexity.

4. Scope, Risks, and Limitations

We recognize the inherent challenges in the automated interpretation of regulation. The system is designed to support, not replace, expert analysis by addressing the following risks:

- **Linguistic Ambiguity:** Outputs with low confidence scores are automatically flagged for human review.
- **Cross-References:** The system maps dependencies between primary and secondary legislation, though final legal interpretation remains the responsibility of professional staff.
- **Implementation Gap:** The system analyzes "written regulation"; any gaps between the written rule and actual enforcement are addressed during the human calibration phase.

5. Phased Implementation Roadmap

System deployment follows a learning-based, incremental approach:

- **Pilot Phase:** Testing the system in selected regulatory domains, specifically the **Energy and Transportation sectors**, to evaluate accuracy and practical utility.
- **Learning & Scaling:** Gradual expansion to additional regulatory fields and integration into the national regulatory repository based on pilot outcomes.

6. Security & Data Privacy

- **Data Minimization:** Processing only publicly available, unclassified regulatory texts. No Personally Identifiable Information (PII) is ingested.
- **Encryption:** Full AES-256 encryption at rest and TLS in transit, ensuring compliance with government cyber security standards.

7. Visual Appendix: Data Flow and System Components

Summary of the information processing workflow:

Phase	Input Source	Primary Process	Output
1. Ingestion	PDFs, Gazettes, Legal DBs	Scraping, OCR, Layout Analysis	Structured, machine-readable text (JSON)
2. NLP Analysis	Structured Text	Entity Extraction (RAG), Reasoning (CoT)	Table of cited regulatory obligations
3. Quantification	Obligations + Price Lists	SCM Mapping & Baseline Assumptions	Preliminary RIA Cost Estimate
4. Human Review	Draft Report + HITL UI	Expert Validation & Calibration	Approved Final Report & Model Tuning

Simplified Data Flow:

1. **Input:**
   [Raw Laws / Regulations] ➞ [Digital Conversion (OCR)]
2. **Analysis:**

[AI Engine (NLP)] ➞ [Extracted Obligations + Citations]

3. **Quantification:**
   [Market Price Lists (SCM)] ➞ [Draft Economic Estimate]
4. **Approval (The System's Core):**
   [Human Expert (HITL)] ➞ [Review, Correction & Final Approval]

This abstract represents a collaborative technological vision for modernizing regulatory oversight in Israel through trustworthy AI.