NEURO-SEMANTIC RULE LANGUAGE

NSRL (Neuro-Semantic Rule Language) is a purpose-built, human-authored, domain-specific policy language used within NDRA-PII to deterministically interpret the meaning, sensitivity, and compliance implications of detected Personally Identifiable Information. It acts as the formal bridge between low-level PII detection outputs and high-level governance decisions by encoding legal, regulatory, and organizational intent into a constrained, executable rule system. NSRL exists to ensure that decisions about sensitive data are made through explicit human logic rather than implicit or probabilistic AI behavior.

NSRL is designed to operate strictly on structured, verified inputs produced by upstream systems, such as PII objects extracted by Microsoft Presidio, document metadata generated by document intelligence pipelines, and contextual attributes like jurisdiction, channel, and document type. It never processes raw text and never performs inference. Instead, it evaluates declarative conditions against already-detected facts, ensuring that interpretation remains grounded in deterministic evidence rather than linguistic ambiguity.

The language is intentionally limited in expressiveness. NSRL is not a programming language and cannot perform computation, iteration, state mutation, or external calls. This restriction is deliberate and fundamental to its safety model. By limiting what can be expressed, NSRL prevents hidden logic, unintended side effects, and unsafe policy behavior. Every rule written in NSRL is transparent, bounded, and auditable, making the system suitable for regulated environments where explainability and repeatability are mandatory.

Execution of NSRL rules is fully deterministic. Given the same inputs, the same rules will always fire in the same way and produce the same interpretive outputs. Multiple rules may apply simultaneously, and conflicts are resolved through explicitly defined precedence semantics rather than heuristic resolution. This guarantees consistency across runs, environments, and deployments, eliminating nondeterministic behavior that would otherwise undermine trust and compliance.

NSRL does not make final decisions or perform actions such as redaction or blocking. Instead, it produces structured interpretive signals including severity classifications, compliance flags, escalation indicators, and human-readable justifications. These outputs are consumed by downstream decision and enforcement layers, which remain separate from policy interpretation. This separation ensures that NSRL governs meaning, while other components govern execution.

Explainability is a first-class property of NSRL. Every rule execution produces a complete trace describing which rule fired, which conditions matched, what data was evaluated, and why a particular outcome was produced. This trace is preserved as part of the

system's audit record and can be reviewed by engineers, compliance teams, auditors, or regulators. There is no hidden reasoning path or opaque decision logic.

All NSRL rules, semantics, invariants, and constraints are authored and controlled exclusively by human Semantic Intelligence Engineers. NSRL is treated as codified legal and compliance intent rather than configuration or code. As such, it is versioned, reviewed, approved, and governed through explicit human processes. Automated systems, including AI-based tools, are not permitted to author, modify, or evolve NSRL rules. This ensures that accountability for policy decisions always remains human and defensible.

In the broader NDRA-PII architecture, NSRL functions as the safety and governance boundary that prevents intelligent systems from overreaching. While AI components may assist in detection, enrichment, or explanation, NSRL defines the authoritative limits of interpretation and enforcement. By constraining intelligence within a human-defined rule language, NDRA-PII achieves a balance between automation and control, enabling scalable privacy intelligence without sacrificing determinism, trust, or compliance.

In essence, NSRL is the formal mechanism by which NDRA-PII transforms detected personal data into legally meaningful, explainable, and auditable policy outcomes. It is the reason NDRA-PII can be fast and intelligent while remaining safe, predictable, and regulator-grade.

```
Nsrl/
├── spec/
│   ├── grammar.md              # NSRL syntax & allowed constructs
│   ├── semantics.md            # Rule meaning & evaluation semantics
│   ├── invariants.md           # Non-negotiable safety laws
│   ├── forbidden_patterns.md   # Explicitly illegal rule patterns
│   └── versioning.md           # Rule & language evolution policy
│
├── rules/
│   ├── gov_id.yml              # Government ID PII rules
│   ├── financial.yml           # Financial PII rules
│   ├── healthcare.yml          # Health & medical PII rules
│   ├── personal.yml            # General personal data rules
│   ├── digital.yml             # Digital / network identifier rules
│   ├── escalation.yml          # Cross-PII & density escalation logic
│   ├── jurisdiction.yml        # Region & regulation specific rules
│   └── tenant_overrides.yml    # Tenant-specific overrides (restricted)
│
├── tests/
│   ├── positive_cases.yml      # Scenarios that MUST trigger rules
│   ├── negative_cases.yml      # Scenarios that MUST NOT trigger rules
```

```
│   └── boundary_cases.yml        # Edge & ambiguous test cases
│
├── meta/
│   ├── policy_manifest.yml       # Central index of all rules
│   ├── change_log.yml            # Human-audited rule change history
│   └── approval.yml              # Explicit legal/compliance approvals
│
├── contracts/
│   ├── input_schema.yml          # NSRL engine input contract
│   └── output_schema.yml         # NSRL engine output contract
│
└── security/
    ├── hard_limits.yml           # Absolute enforcement limits
    └── integrity_checks.yml      # Hashes, signatures, tamper detection
```