SpecLens – Project Proposal & Development Plan

1. Project Summary

SpecLens-PML aims to deliver an end-to-end MLOps system for risk assessment of formal specifications. The project integrates data engineering, machine learning, deployment, and operational governance.

2. Deliverables

- Data pipeline for PML parsing and dataset generation
- ML kernel for training and evaluation
- Configuration layer (config.yaml)
- Continuous Training engine (ct_trigger.py)
- Inference service
- Three technical documents

3. Milestones

M1: Dataset pipeline operational
M2: Baseline ML model trained
M3: Inference pipeline ready
M4: Configuration system integrated
M5: Continuous Training loop implemented
M6: Final demo and documentation

4. WBS

1. Data Engineering
1.1 Parser implementation
1.2 Dataset builder
2. ML Kernel
2.1 Feature engineering
2.2 Model training
3. MLOps
3.1 Configuration layer
3.2 Continuous Training
3.3 Versioning
4. Inference
5. Documentation

5. Sprints

Sprint 1: Data pipeline
Sprint 2: ML baseline
Sprint 3: Inference layer
Sprint 4: MLOps integration
Sprint 5: Governance and finalization

6. DoD & DoR

Definition of Ready:
- Dataset available
- Configuration defined

Definition of Done:
- Code tested
- Artifacts versioned
- Metrics logged
- Document updated