Text highlighted in Green is new (introduced by the framework). Uncolored text was in the participant's original requirements.

1. TITLE: Human-in-the-Loop Learning and Supervision

SYSTEM_REQUIREMENT: The RL system must support real-time human feedback and intervention during model training and deployment.
IMPLEMENTATION_DETAILS:
* Enable the overseer to correct misclassifications, adjust model confidence thresholds, and provide annotations.
* Implement an active learning mechanism where the model requests human input for uncertain cases.
* Allow real-time intervention to override incorrect AI decisions and record these corrections for continuous learning.

RATIONALE: Human-in-the-loop learning ensures that AI training aligns with expert knowledge, enhances model accuracy, and enables timely corrections, improving system transparency and adaptability.

2. TITLE: Post-Training Performance Audits

SYSTEM_REQUIREMENT: The RL system must facilitate structured audits post-training to evaluate decision-making quality and fairness.
IMPLEMENTATION_DETAILS:
* Implement mechanisms for model confidence assessments, adversarial scenario testing, and bias detection.
* Require overseers to review model outputs and recommend fairness improvements.
* Conduct periodic audits to verify decision reliability.

RATIONALE: Ensuring transparency in model updates and fairness evaluations strengthens accountability and trust in AI-driven decision-making.

3. TITLE: Multimodal Alert Communication for Transparency

SYSTEM_REQUIREMENT: The system must use multiple alert communication methods to ensure transparency and timely human intervention.
IMPLEMENTATION_DETAILS:
* Implement color-coded dashboard notifications for ongoing monitoring
* Use audible alerts for immediate attention to critical issues
* Provide haptic feedback for critical alerts in high-workload environments
* Send mobile and email notifications for remote monitoring
* Offer interactive alerts with suggested corrective actions.

RATIONALE: Ensuring that alerts are accessible through multiple communication channels helps overseers respond promptly without being overwhelmed.

4. TITLE: Real-Time Model Feedback Integration

SYSTEM_REQUIREMENT: The RL system must continuously incorporate human feedback during deployment to refine decision-making.
IMPLEMENTATION_DETAILS:

* Allow overseers to adjust sensitivity thresholds correct errors and approve incremental learning updates.
* Maintain logs of human interventions for future training refinements.
RATIONALE: Continuous feedback mechanisms ensure system adaptability and responsiveness, enhancing transparency in evolving operational conditions.

5. TITLE: System Error Rate & Anomaly Detection

SYSTEM_REQUIREMENT: The system must track and report errors and anomalies in real time, ensuring transparency in decision reliability.
IMPLEMENTATION_DETAILS:
* Identify at least 99% of anomalies and log them within 10 milliseconds
* Utilize anomaly detection models to capture unexpected patterns in decision-making
* Store all detected anomalies for auditing purposes.
RATIONALE: Ensuring that errors and anomalies are logged and addressed enhances system reliability and accountability while preventing incorrect or biased actions from persisting.

6. TITLE: User Override and Manual Intervention Monitoring

SYSTEM_REQUIREMENT: The system must track and log instances of manual intervention to assess system autonomy and identify areas requiring improvement.
IMPLEMENTATION_DETAILS:
* Log the frequency and nature of human interventions
* Flag situations where manual overrides exceed 5% of total decisions
* Provide an interface for overseers to document reasons for intervention.
RATIONALE: Tracking manual interventions ensures transparency in system decision-making and highlights areas where AI decisions may require improvement.

7. TITLE: Simulation-Based Readiness Testing

SYSTEM_REQUIREMENT: The RL system must integrate simulation-based training for the overseer to validate their preparedness before deployment.
IMPLEMENTATION_DETAILS:
* Implement interactive simulation modules covering normal operations, degraded performance, and failure scenarios.
* Require a 95% accuracy rate in identifying and responding to behavior categories before deployment.
* Conduct periodic simulation-based refresher tests.
RATIONALE: Hands-on experience in controlled settings ensures that overseers can accurately recognize and address system behavior deviations in real-world scenarios.

8. TITLE: Human Overseer Training and Certification

SYSTEM_REQUIREMENT: The RL system must ensure that human overseers are thoroughly trained and certified before deployment in high-risk environments.
IMPLEMENTATION_DETAILS:
* Implement a structured training program covering system architecture, AI decision-making and monitoring tools.

* Require a minimum 90% pass rate for certification exams.
* Conduct hands-on simulation training with a 95% success rate in handling high-risk events.
* Provide ethical compliance and security training with a 95% compliance threshold for interventions and a 90% anomaly detection accuracy in cybersecurity scenarios.
* Establish ongoing training every six to twelve months to maintain proficiency.
RATIONALE: To ensure transparency and safe operation, human overseers must fully understand the system's functionality and decision-making process, preventing misinterpretations and errors in high-risk environments.

9. TITLE: Real-Time Multi-Tiered Dashboard for Overseers
SYSTEM_REQUIREMENT: The system must provide an interactive dashboard with hierarchical data views to balance cognitive load and facilitate expert decision-making.
IMPLEMENTATION_DETAILS:
* Design a three-tier dashboard with summaries, detailed metrics, and raw logs
* Implement color-coded alerts for easy interpretation of priority levels
* Customize dashboards based on overseer experience levels.
RATIONALE: A structured dashboard prevents information overload while ensuring critical insights are available to those who need them.

10. TITLE: Real-Time Bias and Fairness Visualization
SYSTEM_REQUIREMENT: The system must provide visual reports on fairness metrics to aid in bias detection and accountability.
IMPLEMENTATION_DETAILS:
* Implement bias heatmaps to highlight demographic imbalances
* Provide real-time anomaly graphs with filtering options
* Display fairness discrepancy reports for flagged decisions.
RATIONALE: Providing visual insights into fairness and bias enables proactive monitoring and quick intervention when necessary.

11. TITLE: Dynamic Thresholds and Adaptive Alerting
SYSTEM_REQUIREMENT: The system must dynamically adjust alert thresholds based on real-time conditions and overseer feedback to maintain accuracy and efficiency.
IMPLEMENTATION_DETAILS:
* Adjust alert thresholds based on historical trends and real-time data
* Escalate unresolved alerts based on predefined severity levels
* Implement self-learning models to refine alert triggers over time.
RATIONALE: Adapting alert thresholds prevents unnecessary alerts while ensuring that critical issues are flagged appropriately, maintaining transparency without overburdening overseers.

12. TITLE: Historical Case Study and Comparative Analysis
SYSTEM_REQUIREMENT: The RL system must maintain a repository of past operational cases to assist the overseer in recognizing behavior patterns.
IMPLEMENTATION_DETAILS:

* Curate a reference library of real-world and simulated events illustrating expected, degraded, and failure behaviors.
* Implement a comparative analysis tool that overlays historical cases with real-time system data to highlight emerging risks.
RATIONALE: Providing historical references enhances the overseer's ability to identify and address potential system anomalies through pattern recognition.

13. TITLE: Real-Time Bias Monitoring and Corrective Mechanisms
SYSTEM_REQUIREMENT: The system shall continuously monitor fairness metrics in real-time and apply corrective actions as needed.
IMPLEMENTATION_DETAILS:
* Fairness-aware decision thresholds dynamically adjust to prevent systematic biases
* Bias drift analysis monitors changes in decision patterns over time
* Threshold-based alerting activates human intervention when fairness metrics fall outside acceptable ranges
* Human-in-the-loop interventions allow experts to override biased decisions and provide corrective feedback.
RATIONALE: Real-time detection and intervention are necessary to prevent bias accumulation and ensure fair decision-making across operational contexts.

14. TITLE: Fairness-Aware Data Collection and Processing
SYSTEM_REQUIREMENT: The system must ensure that training data is diverse, representative, and free from systematic biases.
IMPLEMENTATION_DETAILS:
* Diverse and representative data is collected across demographics and environments
* Pre-processing bias corrections using re-weighting, resampling, and adversarial debiasing
* Context-sensitive feature engineering examines and adjusts features to prevent bias
* Bias audits and compliance checks ensure adherence to fairness standards.
RATIONALE: Data integrity is critical to preventing biases from influencing learned behaviors and maintaining fairness across different groups.

15. TITLE: Post-Deployment Fairness Evaluation and Accountability
SYSTEM_REQUIREMENT: The system shall conduct periodic fairness audits, provide explainability mechanisms, and enable external feedback integration.
IMPLEMENTATION_DETAILS:
* Regular bias audits and compliance checks ensure alignment with ethical AI standards
* Explainability and justification mechanisms provide transparency for flagged decisions
* Crowdsourced bias reporting allows stakeholders to contribute to fairness improvements
* Continuous model updating incorporates new unbiased data to enhance fairness over time.
RATIONALE: Ongoing evaluation and accountability mechanisms ensure long-term fairness, regulatory compliance, and ethical AI performance.

16. TITLE: Diverse and Representative Training Data Assessment

SYSTEM_REQUIREMENT: The system shall employ statistical and analytical methods to evaluate the diversity and representativeness of training data.

IMPLEMENTATION_DETAILS:

* Demographic distribution assessment quantifies representation using entropy-based indices
* Class imbalance detection ensures fair distribution using statistical techniques
* Fairness-constrained sampling modifies dataset selection to uphold equity
* Synthetic data generation through GANs enhances coverage of underrepresented cases
* Cross-domain data fusion integrates datasets from multiple sources for contextual richness.

RATIONALE: Ensuring data diversity and representativeness prevents blind spots in learned behaviors and maintains fairness across different population segments.

17. TITLE: Scenario Replay and Visualization for Post-Mortem Analysis

SYSTEM_REQUIREMENT: The system shall enable detailed post-mortem analysis, including scenario replay and decision tracing.

IMPLEMENTATION_DETAILS:

* Time-synchronized data streams will capture sensor inputs alongside decision-making outputs.
* A graphical playback interface will allow overseers to replay scenarios, visualize sensor data overlay, and review decisions in real-time.
* Multiple playback speeds will be supported for efficiency.

RATIONALE: Scenario replay enhances accountability by allowing a detailed reconstruction of system decisions for forensic analysis and continuous improvement.

18. TITLE: Comprehensive Logging Framework

SYSTEM_REQUIREMENT: The system shall implement a thorough logging framework to capture and store essential interactions and decisions for transparency, accountability, and traceability.

IMPLEMENTATION_DETAILS:

* The system will log every decision point, including path changes, object detection and classification results with timestamps and confidence scores.
* Human interventions such as overrides and feedback will be recorded along with reasoning.
* System alerts and failures will be documented including the system’s response and overseer acknowledgment.

RATIONALE: Ensuring accountability and transparency in system decision-making requires a robust logging framework that facilitates real-time monitoring, forensic analysis, and system improvements.

19. TITLE: Automated Anomaly Detection in Post-Mortem Analysis

SYSTEM_REQUIREMENT: The system shall implement automated tools to identify significant incidents and anomalies during post-mortem analysis.

IMPLEMENTATION_DETAILS:

* Unexpected behavior detection will compare actions to performance baselines and alert users to deviations.
* Human intervention pattern analysis will identify trends in model failure points.

* Failure mode categorization will assign labels such as false positives, misclassifications, and sensor malfunctions to support system improvements.
RATIONALE: Automated anomaly detection streamlines forensic analysis and helps identify areas for iterative refinement.

20. TITLE: Human Expertise Integration for Continuous Improvement
SYSTEM_REQUIREMENT: The system shall incorporate human feedback in post-mortem analysis to refine system behavior and enhance robustness.
IMPLEMENTATION_DETAILS:
* Expert annotation tools will allow overseers to highlight and comment on specific moments in a scenario.
* Failure case logs will be utilized for model retraining pipelines.
* Collaborative review sessions will enable teams to analyze system performance and suggest refinements.
RATIONALE: Human oversight enhances system learning, ensuring improvements based on real-world feedback and expert insights.

21. TITLE: Multi-Layered Logging for Granular Insights
SYSTEM_REQUIREMENT: The system shall maintain logs at multiple levels to provide comprehensive tracking and insight into system behavior.
IMPLEMENTATION_DETAILS:
* Raw data logs will capture sensor inputs video frames and LiDAR scans.
* Processed data logs will store sensor fusion outputs and environmental mappings. Model decision logs will record confidence levels reasoning pathways and decision timelines.
* Human interaction logs will document manual adjustments with timestamps and reasoning. * Performance metrics logs will store success/failure rates and calibration adjustments.
RATIONALE: Granular logging enables a detailed understanding of system behavior, supporting performance evaluation and compliance verification.