# Monthly PR & Impact Summary — May 2025

🚀 **Core Feature Work**

1. **Improved Error Logging in Container Crash Queries**
   * **Type:** Enhancement
   * **Date:** 14 May 2025
   * **Summary:**
     1. Refines container crash error detection by broadening query filters to capture additional failure patterns
     2. Expanded error filtering to include patterns such as "grpc-server-shutdown" and "[grpc-"; implemented conditional JSON parsing for inference-server-inferenceapi logs; added debug logging for both query construction and responses.
   * **PR link:** [**Pull request 40107: Enhance grpc logging in container crash queries - Repos**](https://dev.azure.com/project-argos/Sorento/_git/oairt-ai/pullrequest/40107)
2. **IP Endpoint Model Error Investigation Tool**
   * **Type:** New Feature
   * **Date:** 15 May 2025
   * **Summary:**
     1. Implemented a new tool that constructs and executes Kusto queries to retrieve model error categories and details for IP endpoints.
     2. Added comprehensive unit tests to validate query logic, data handling, and error scenarios.
   * **PR link:** [**Pull request 39957: Add find\_ip\_endpoint\_model\_errors tool - Repos**](https://dev.azure.com/project-argos/Sorento/_git/oairt-ai/pullrequest/39957)
3. **SLA Impact Detection Tools**
   * **Type:** New Feature
   * **Date:** 16 May 2025
   * **Summary:**
     1. Implemented KV utilization monitoring to flag SLA impacts when usage exceeds defined thresholds for sustained periods.
   * **PR link:** [**Pull request 40509: Add tool to check sla impacted - Repos**](https://dev.azure.com/project-argos/Sorento/_git/oairt-ai/pullrequest/40509)
4. **SLA Query Refinement and Region Filter Removal**
   * **Type:** Bug Fix
   * **Date:** 19 May 2025
   * **Summary:**
     1. Removed hardcoded region filter (kustoregion variable) and replaced with dynamic handling via an empty regionName default.
     2. Updated Kusto query to summarize KV metrics and revised kv\_util calculation to compute utilization from aggregated results.
   * **PR link:** [**Pull request 40711: Remove region filter in sla query - Repos**](https://dev.azure.com/project-argos/Sorento/_git/oairt-ai/pullrequest/40711)
5. **SLA Details Integration in ICM Enrichment**
   * **Type:** New Feature
   * **Date:** 21 May 2025
   * **Summary:**
     1. Introduces SLA impact details directly into the ICM enrichment output, improving visibility into reliability metrics within incident investigations.
     2. Improves error handling and test coverage to ensure robust SLA integration across enrichment workflows.
   * **PR link:** [**Pull request 40915: Integrate KV usage into icm enrichment - Repos**](https://dev.azure.com/project-argos/Sorento/_git/oairt-ai/pullrequest/40915)
6. **Capacity Enrichment Integration in ICM Insights**
   * **Type:** New Feature
   * **Date:** 26 May 2025
   * **Summary:**
     1. Introduces capacity availability enrichment into the ICM investigation workflow, providing deeper insight into endpoint capacity states during incident analysis.
     2. Strengthens correlation between capacity utilization and incident impact, improving root cause identification.
   * **PR link:** [**Pull request 41058: Added capacity enrichment - Repos**](https://dev.azure.com/project-argos/Sorento/_git/oairt-ai/pullrequest/41058)
7. **Modularized ICM Enrichment HTML Report Generation**
   * **Type:** UI Improvement
   * **Date:** 26 May 2025
   * **Summary:**
     1. Refactors ICM enrichment report generation to use modular HTML rendering components, improving code maintainability and readability.
     2. Enhances extensibility for future enrichment sections and simplifies testing through decoupled rendering logic.
   * **PR link:** [**Pull request 41384: Improve UI and Refactor icm enrichment code - Repos**](https://dev.azure.com/project-argos/Sorento/_git/oairt-ai/pullrequest/41384)
8. **Expanded Details Sections and Improved HTML Layout**
   * **Type:** UI Improvement
   * **Date:** 27 May 2025
   * **Summary:**
     1. Enhances the readability and user experience of ICM enrichment reports by expanding all details sections by default.
   * **PR link:** [**Pull request 41446: Update rendering functions to open details sections by default - Repos**](https://dev.azure.com/project-argos/Sorento/_git/oairt-ai/pullrequest/41446)
9. **VLM Model Error Integration and Unified HTML Styling**
   * **Type:** New Feature
   * **Date:** 27 May 2025
   * **Summary:**
     1. Integrates VLM model error reporting into the ICM investigation enrichment, expanding diagnostic coverage for visual-language models.
     2. Unifies HTML styling across all enrichment sections using centralized inline CSS constants, ensuring a consistent and maintainable UI experience.
   * **PR link:** [**Pull request 41473: Add VLM model errors in enrichment - Repos**](https://dev.azure.com/project-argos/Sorento/_git/oairt-ai/pullrequest/41473)