Multi-Agent Orchestration for Autonomous Regression Management

Sangwoo Noh, Jin Choi, Seonghee Yim, Youngsik Kim

Samsung Electronics

2026 DESIGN AND VERIFICATION™ DVCON CONFERENCE AND EXHIBITION UNITED STATES
SANTA CLARA, CA, USA
MARCH 2 - 5, 2026

MOTIVATION

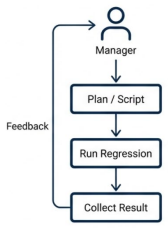
❖ "Linear Trap" of Modern Regression

- Modern regression automation pipelines (e.g., Jenkins) are largely linear, scripted, developer-defined and fire-and forget
- Linear architecture makes regression hard to adapt to failure signals under frequent iterations

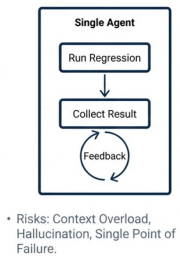
❖ "Manager-Centric" Bottleneck Problem

- Regression data is siloed after launching integrated regression, requiring a human manager to interpret failure signals, re-plan test scopes, and assign tasks
- This slows information propagation and increases turnaround time; even a single AI agent cannot remove this bottleneck without redefining the regression pipelines architecture

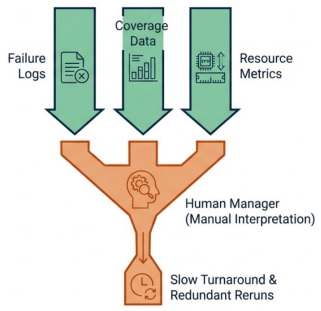
Procedural CI Pipeline



Single Agent System



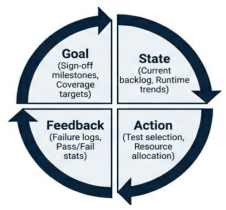
- Risks: Context Overload, Hallucination, Single Point of Failure.



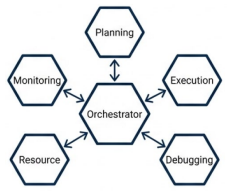
KEY CONTRIBUTIONS

❖ Redefinition of Regression Pipeline as Closed-loop Control System

- A closed-loop control system consists of *goal*, *state*, *action*, and *feedback*
- We frame regression as a closed-loop system where the ***goal*** is sign-off (coverage targets, milestones, or resolving a specific failure class), the ***state*** is what we observe (logs/signatures, coverage status, resource/queue signals), the ***action*** is what we change next (test prioritization, ownership assignment, targeted reruns), and the ***feedback*** is the run outcome (pass/fail results, new failure signatures, coverage deltas) that updates the next cycle

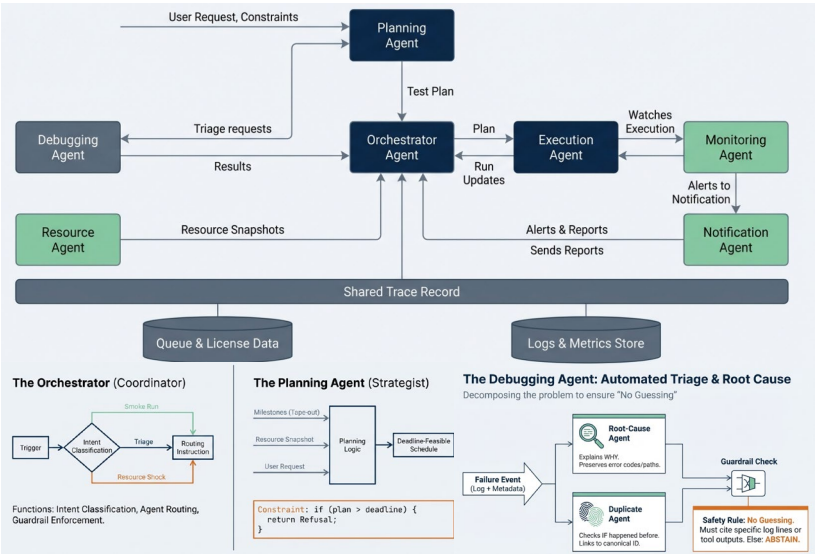


❖ Multi-Agent System (MAS) for Regression System



- A Multi-Agent System (MAS) for regression decomposes the closed-loop controller into role-specific agents rather than a single manager or script.
- ***Planning***, ***Debugging***, ***Execution***, ***Monitoring***, ***Resource***, and ***Notification*** **agents** share a common traced state and coordinate tool-driven actions—test re-scoping, prioritization, targeted reruns, and ownership routing—so the loop adapts continuously as new results arrive. An Orchestrator selects the minimal agent path for each intent and routes typed, schema-checked messages.

PROPSED ARCHITECTURE



EXPREIMENTS

❖ Scenarios

I. **Owner assignment**: Routes failure to correct team; Tools used are pattern search and code-attribution
II. **Root-cause explanation**: Generates text explanation citing evidence
III. **Duplicate detection**: Link failure to exact historical ID. Challenges must be exact match

❖ Evaluation Metrics

I. **Owner assignment accuracy**: Measures exact matches between the predicted owner and the ground truth and code-attribution
II. **Root cause similarity**: Calculates the similarity between the token sets of the predicted explanation and the expert ground truth
III. **Duplicate Detection Micro-F1**: Evaluates duplicate linkage; a result is only a True Positive if the system correctly identifies a case as a duplicate and provides the exact correct duplicate_of identifier
IV. **Cost Metrics**: Measures efficiency through the average time per case and the average number of simulated Agent invocations per case

❖ Statistical Methodology

I. **McNemar's Test**: Used for paired binary outcomes (Owner Accuracy and Root Cause Match@0.5)
II. **Paired t-tests**: Used for continuous outcomes like time per case
III. **Bonferroni Correction**: Applied to control the error rate across multiple comparisons, setting the significance threshold to 0.0071
IV. **Abstention Handling**: If a system abstains from making a prediction due to strict guardrails, it is scored as incorrect

RESULTS

❖ Metrics on the test set

Metric	Legacy CI	Single-Agent	Multi-Agent
Owner Accuracy	28.7%	63.3%	76.7%
Root Cause Similarity (mean ± SD)	1.9% ± 5.8	62.7% ± 48.5	76.9% ± 41.7
Root Cause Match@0.5	0.0%	62.7%	77.3%
Duplicate Detection micro-F1	22.7%	40.8%	96.0%
Avg Time per Case	0.10s	3.05s	4.04s
Avg simulated Agent invocations	0.0	3.0	4.0

❖ Paired statistical tests

Comparison / Metric	System 1 → System 2	p-value	Sig. @0.0071
Multi vs Single - Owner Accuracy	63.3% (95/150) → 76.7% (115/150)	0.0142	No
Multi vs Single - Root Cause Match@0.5	62.7% (94/150) → 77.3% (116/150)	0.0077	No
Multi vs Single - Time per Case	3.053s → 4.037s	<0.0001	Yes
Single vs Legacy - Owner Accuracy	28.7% (43/150) → 63.3% (95/150)	<0.0001	Yes
Single vs Legacy - Root Cause Match@0.5	0.0% (0/150) → 62.7% (94/150)	<0.0001	Yes
Multi vs Legacy - Owner Accuracy	28.7% (43/150) → 76.7% (115/150)	<0.0001	Yes
Multi vs Legacy - Root Cause Match@0.5	0.0% (0/150) → 77.3% (116/150)	<0.0001	Yes

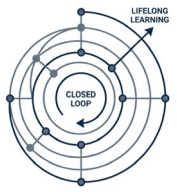
** Bonferroni-corrected threshold is alpha=0.0071*

CONCLUSTIONS

❖ We transform manager-centric regression pipelines into an agent-driven closed-loop control system with specialized roles and tool-driven actions

❖ The system supports continuous, dynamic operation during always-on regression, adapting plans as new signals arrive

❖ It outperforms both legacy pipelines and a single-agent baseline: ownership assignment **76.7%**, root-cause explanation **77.3%**, duplicate detection **96.0%**



❖ Future work: validate end-to-end full scenarios in real-world verification environments, expand tool integration and robustness, and measure long-horizon impact

REFERENCES

[1] Tang, Shuo, et al. "Decentralized and Lifelong-Adaptive Multi-Agent Collaborative Learning." arXiv preprint arXiv:2403.06535 (2024).
[2] Liu, Wei, et al. "Autonomous agents for collaborative task under information asymmetry." Advances in Neural Information Processing Systems 37 (2024): 2734-2765.
[3] Wu, Qingyun, et al. "Autogen: Enabling next-gen Agent applications via multi-agent conversations." First Conference on Language Modeling. 2024.
[4] Guo, Taicheng, et al. "Large language model based multi-agents: A survey of progress and challenges." arXiv preprint arXiv:2402.01680 (2024).

SAMSUNG

Samsung Electronics Co., LTD., Hwasung-si, South Korea
Contact Info : sangwoo.noh@samsung.com