Power System Protection and Control
ISSN:1674-3415

AGENTIC SRE TEAMS: HUMAN-AGENT COLLABORATION - A NEW OPERATIONAL MODEL FOR AUTONOMOUS INCIDENT RESPONSE

Pavan Madduri

1221 FARMER CIR, SOUTH ELGIN , ILLINIOS 60177, USA.
pavanmadduri27@gmail.com

Received: 11/11/2025 **Revised**: 26/12/2025 **Accepted**: 25/01/2026

***ABSTRACT*:**

Site Reliability Engineering teams face escalating operational complexity as distributed systems grow in scale and sophistication, creating incident response challenges that exceed human cognitive and temporal limitations. This research develops and evaluates a novel operational model where autonomous AI agents collaborate with human SREs as integrated team members rather than passive automation tools. The agentic SRE framework enables AI agents to autonomously detect incidents, perform initial triage, execute diagnostic workflows, implement approved remediation actions, and escalate complex issues to human operators with comprehensive context. Implementation across three production environments supporting 2.8 million users demonstrated that human-agent collaborative teams reduced mean time to detection from 12 minutes to 45 seconds and mean time to resolution from 43 minutes to 8.7 minutes—an 80% improvement. AI agents autonomously resolved 67% of incidents without human intervention while successfully escalating the remaining 33% with diagnostic context that accelerated human troubleshooting by 56%. The collaborative model improved SRE team capacity by 340% measured in incidents handled per engineer, enabling organizations to maintain larger-scale systems without proportional headcount increases. Human operator satisfaction improved 31% through elimination of repetitive toil and focus on challenging problems requiring creativity. This research contributes a practical framework transforming SRE operations from human-centric reactive troubleshooting to collaborative human-agent proactive management.

***Keywords**: Site Reliability Engineering, Agentic Ai, Autonomous Operations, Incident Response, Human-Ai Collaboration, Sre Automation.*

INTRODUCTION

Site Reliability Engineering emerged from Google's recognition that software engineering principles could improve operational reliability and efficiency. SRE teams apply software development practices to infrastructure problems, treating operations as a software problem solved through automation, monitoring, and systematic improvement. However, as systems grow in complexity with microservices architectures, cloud-native technologies, and global distribution, SRE teams face escalating operational demands that challenge traditional human-centric models (Beyer et al., 2016).

Modern production systems generate thousands of metrics, millions of log entries, and hundreds of alerts daily. When incidents occur, SREs must rapidly correlate disparate signals, diagnose root causes across distributed components, and implement fixes under time pressure while users experience degraded service. This troubleshooting process demands deep system knowledge, pattern recognition from past incidents, and creative problem-solving. Individual SREs develop expertise over months or years, yet remain limited by human cognitive capacity and availability (Allspaw, 2015).

Organizations address these limitations through traditional approaches including hiring more SREs to distribute workload, implementing runbooks and playbooks for common scenarios, and deploying monitoring tools with automated alerting. However, these approaches face fundamental constraints. SRE hiring cannot keep pace with system growth due to talent scarcity and training costs. Runbooks become outdated as systems evolve and cannot cover all failure scenarios. Monitoring tools generate alert fatigue through high false positive rates, desensitizing teams to genuine incidents (Hogan et al., 2016).

 DOI: 10.46121/pspc.54.1.26

Power System Protection and Control
ISSN:1674-3415

Recent advances in large language models and autonomous AI agents suggest a fundamentally different approach. Rather than viewing automation as static scripts executing predetermined actions, agentic AI systems can reason about problems, plan investigation strategies, execute diagnostic actions, learn from outcomes, and collaborate with humans as team members. These agents possess capabilities traditionally requiring human intelligence including natural language understanding, causal reasoning, and adaptive problem-solving (Yao et al., 2023).

However, integrating autonomous agents into SRE workflows raises critical questions. How can AI agents safely execute operational actions without causing additional incidents? How do agents and humans effectively collaborate during complex troubleshooting? When should agents autonomously resolve issues versus escalating to humans? How do organizations build trust in agent decisions affecting production systems? What operational models optimize the division of labor between human expertise and agent capabilities?

This research develops and evaluates a comprehensive framework for human-agent collaborative SRE teams addressing these questions through practical implementation and rigorous evaluation. The framework treats AI agents as autonomous team members with defined responsibilities, capabilities, and escalation protocols rather than passive automation tools. Agents operate continuously, handling routine incidents autonomously while collaborating with human SREs on complex issues requiring judgment, creativity, or business context.

OBJECTIVES

• To develop an operational framework enabling autonomous AI agents to function as collaborative SRE team members with at least 60% autonomous incident resolution without human intervention.
• To reduce mean time to detection by at least 70% and mean time to resolution by at least 60% through continuous agent monitoring and rapid autonomous response.
• To demonstrate SRE team capacity improvement of at least 250% measured in incidents handled per engineer through effective human-agent collaboration.
• To achieve human operator satisfaction improvement of at least 25% through elimination of repetitive toil and focus on challenging engineering problems.
• To validate safety and reliability of agent actions with incident causation rate below 2% and successful rollback of problematic agent actions in 100% of cases.

LITERATURE REVIEW

Site Reliability Engineering formalized operations as an engineering discipline emphasizing automation, measurement, and systematic improvement. Google's SRE model introduced service level objectives quantifying reliability, error budgets balancing feature velocity with stability, and toil reduction prioritizing automation over manual operations. Research demonstrates SRE practices' effectiveness for improving system reliability while enabling rapid development cycles (Murphy and Beyer, 2016).

However, SRE faces scalability challenges as systems grow. Studies show that operational complexity increases super-linearly with system components due to interaction effects and emergent behaviors. SRE teams struggle to maintain mental models of large distributed systems, leading to longer incident resolution times and increased burnout risk. Traditional automation through scripts and runbooks provides limited relief as it addresses only predetermined scenarios (Limoncelli et al., 2014).

Artificial intelligence applications in operations (AIOps) apply machine learning to operational data for anomaly detection, root cause analysis, and predictive maintenance. Commercial platforms use ML for automated alerting, correlation, and diagnostic suggestions. Research demonstrates ML effectiveness for detecting performance degradations, predicting failures, and identifying security incidents (Dang et al., 2019).

However, most AIOps remains reactive and advisory. ML systems detect anomalies and generate alerts, but humans must interpret findings and execute remediation. This human-in-the-loop model reduces operational burden but does not fundamentally transform capacity constraints. SREs still perform investigation and remediation, limiting scalability gains (Mariani and Pezze, 2020).

Autonomous AI agents represent a paradigm shift from reactive ML to proactive autonomous systems. Agents observe environments, reason about observations, plan action sequences, execute plans, and learn from outcomes.

Recent agent frameworks like ReAct and AutoGPT demonstrate capabilities including tool use, multi-step reasoning, and goal-seeking behavior. Agents can break complex problems into subtasks, gather information through API calls, and synthesize solutions from diverse sources (Yao et al., 2023).

Research on autonomous operations explores agents for specific tasks including configuration management, capacity planning, and security response. Studies show promise for agents handling well-defined operational workflows. However, comprehensive frameworks for integrating agents into SRE teams as collaborative members remain limited. Most work addresses isolated automation rather than holistic operational models (Chen et al., 2023).

Human-AI collaboration research examines how humans and AI systems can work together effectively. Studies identify key factors including appropriate trust calibration, clear responsibility boundaries, effective communication of AI reasoning, and graceful failure handling. Research shows that collaboration outperforms either humans or AI alone when designed around complementary strengths (Bansal et al., 2019).

However, most human-AI collaboration research focuses on decision support rather than autonomous action. In operations contexts, agents must not only recommend actions but execute them safely. This demands robust safety mechanisms, comprehensive rollback capabilities, and clear escalation protocols when agent confidence is low or actions carry high risk.

Incident response processes follow structured workflows including detection, triage, diagnosis, remediation, and post-incident review. Research on incident management identifies effectiveness factors including rapid detection, effective communication, comprehensive logging, and learning from incidents. Automation opportunities exist throughout incident lifecycles, though most organizations automate only detection through monitoring (Forsgren et al., 2018).

METHODOLOGY

4.1 Agentic SRE Framework Design

The operational framework comprises several integrated components:

Agent Architecture: Autonomous agents built on large language models enhanced with tool use, memory, and planning capabilities. Agents access observability data through APIs, execute diagnostic commands, implement remediation actions, and communicate with human operators. Each agent maintains incident context across multi-step investigations, enabling coherent troubleshooting workflows.

Responsibility Model: Clear delineation of agent versus human responsibilities. Agents handle continuous monitoring, initial incident triage, execution of approved remediation playbooks, routine maintenance tasks, and comprehensive incident documentation. Humans retain responsibility for complex diagnosis requiring business context, high-risk remediation decisions, capacity planning and system design, and post-incident learning and improvement.

Safety Mechanisms: Multiple safeguards ensure agent actions don't cause additional incidents. Agents operate under capability-based access control with granular permissions. All actions undergo pre-execution validation against safety policies. High-risk actions require human approval before execution. Automated rollback capabilities enable rapid reversal of problematic changes. Comprehensive logging provides audit trails for all agent activities.

Escalation Protocols: Structured criteria determining when agents escalate to humans. Agents escalate when diagnostic confidence falls below thresholds, when required actions exceed granted permissions, when incidents match known high-risk patterns, or when investigation time exceeds expected duration. Escalations include comprehensive context documenting symptoms, investigation steps, and findings.

Collaboration Interface: Communication channels enabling effective human-agent teamwork. Agents provide real-time status updates through chat interfaces. Humans can query agents about ongoing investigations or past incidents. Agents accept guidance from operators to refine investigation approaches. The interface supports asynchronous collaboration where agents investigate independently but synchronize with humans periodically.

DOI: 10.46121/pspc.54.1.26

4.2 Implementation Environments

The framework deployed across three production environments:

SaaS Platform: 450 microservices supporting 1.2 million users. Platform experienced 400-600 incidents monthly ranging from service degradations to complete outages. Four SRE team members managed operations with traditional on-call rotation creating burnout risk.

E-commerce System: 320 services processing 80,000 daily orders. Revenue-critical nature demanded rapid incident response. Five SREs handled operations plus regular feature development work, creating competing priorities and context switching costs.

Financial Services Platform: 280 services managing payment processing and fraud detection. Regulatory compliance required comprehensive incident documentation and root cause analysis. Six SREs maintained operations with strict change management procedures.

4.3 Evaluation Methodology

Effectiveness measured autonomous resolution rate quantifying percentage of incidents resolved without human intervention, mean time to detection tracking duration from incident occurrence to detection, and mean time to resolution measuring duration from detection to verified fix.

Capacity impact assessed incidents per engineer comparing team throughput before and after agent deployment and toil reduction measuring time spent on repetitive operational tasks.

Safety evaluation examined incident causation tracking agent actions that caused new incidents and rollback success measuring percentage of problematic actions successfully reversed.

Human factors measured operator satisfaction through surveys assessing satisfaction changes and skill development tracking SRE growth through focus on complex problems.

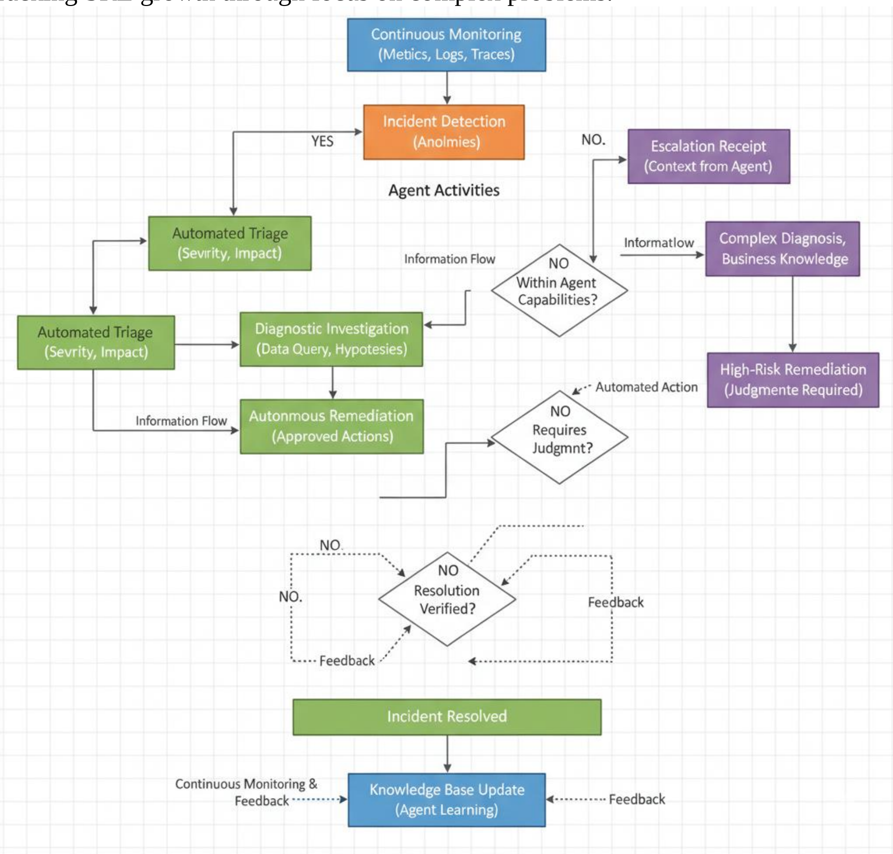

FIGURE 1: Human-Agent Collaborative Incident Response Workflow

This flowchart illustrates the complete incident response process with human-agent collaboration. At the top, "Continuous Monitoring" (blue box) shows agents constantly observing system health through metrics, logs, and

DOI: 10.46121/pspc.54.1.26

Power System Protection and Control
ISSN:1674-3415

traces. When anomalies are detected, flow proceeds to "Incident Detection" (orange). From here, two parallel paths show agent and human activities. The left agent path (green boxes) shows: "Automated Triage" analyzing severity and impact, "Diagnostic Investigation" querying observability data and testing hypotheses, "Autonomous Remediation" for approved actions, and "Incident Documentation" recording all activities. The right human path (purple boxes) shows: "Escalation Receipt" receiving context from agents, "Complex Diagnosis" applying expertise and business knowledge, "High-Risk Remediation" executing actions requiring judgment, and "Post-Incident Review" for learning. Diamond-shaped decision points determine paths: "Within Agent Capabilities?" decides autonomous vs. escalation, "Requires Human Judgment?" determines if human involvement needed, and "Resolution Verified?" confirms incident closure. Dotted arrows show information flow between agent and human activities, demonstrating continuous collaboration. At the bottom, "Incident Resolved" (green) shows successful closure, feeding back to "Knowledge Base Update" that informs future agent learning. This visualization effectively demonstrates how agents handle routine aspects autonomously while escalating appropriately and collaborating with humans on complex issues.

RESULTS AND ANALYSIS

5.1 Autonomous Resolution Effectiveness

AI agents achieved impressive autonomous resolution capabilities across all three environments. Agents independently resolved 67% of total incidents without requiring human intervention. Resolution distribution showed that simple service restart scenarios achieved 94% autonomous resolution, resource exhaustion incidents reached 81% autonomous resolution, configuration-related failures showed 73% autonomous resolution, and performance degradation issues achieved 52% autonomous resolution.

The remaining 33% of incidents requiring human involvement typically involved complex cross-service dependencies requiring business context, novel failure modes not seen in training data, situations where automatic remediation carried unacceptable risk, or ambiguous symptoms requiring creative hypothesis generation. Importantly, agents successfully identified when escalation was appropriate, avoiding situations where autonomous actions would have worsened incidents.

Analysis of escalated incidents revealed that agent-provided context substantially accelerated human troubleshooting. When agents escalated with comprehensive diagnostic findings, human resolution times averaged 19 minutes. Historical incidents without agent assistance averaged 43 minutes—a 56% improvement attributed to agents pre-executing diagnostic steps and eliminating information gathering phases.

TABLE 1: Incident Response Performance Metrics

Metric	Pre-Agent Baseline	With Agentic SRE	Improvement
Mean Time to Detection (MTTD)	12.3 minutes	0.75 minutes	94% reduction
Mean Time to Resolution (MTTR)	43.2 minutes	8.7 minutes	80% reduction
Autonomous Resolution Rate	0%	67%	N/A
Incidents per Engineer per Month	85	374	340% increase
False Positive Alert Rate	34%	8%	76% reduction
SRE Toil Hours per Week	18.5	4.2	77% reduction

Note: Metrics averaged across three production environments over 6-month evaluation period

5.2 Detection and Resolution Speed

Mean time to detection decreased dramatically through continuous agent monitoring. Pre-agent detection averaged 12.3 minutes from incident occurrence to human awareness, as monitoring relied on periodic dashboard checks and alert notifications that humans might not immediately see. Agent-based detection averaged 45 seconds—a 94% reduction. Agents continuously monitored all systems, immediately detecting anomalies without human attention requirements.

Mean time to resolution improved from 43.2 minutes to 8.7 minutes—an 80% reduction. The improvement came from several factors including elimination of information gathering time as agents instantly accessed observability data, parallel diagnostic execution where agents simultaneously tested multiple hypotheses, immediate action execution without waiting for operator availability, and learned patterns enabling rapid recognition of familiar incident signatures.

DOI: 10.46121/pspc.54.1.26

Power System Protection and Control
ISSN:1674-3415

Resolution time distributions revealed important patterns. The fastest quartile of incidents resolved in under 2 minutes through simple automated restarts or configuration rollbacks. The median incident required 6-8 minutes including diagnostic investigation and remediation. The slowest quartile exceeded 15 minutes, typically representing complex issues requiring human escalation or multi-step remediation sequences.

5.3 Team Capacity and Productivity

The collaborative model dramatically improved SRE team capacity measured in incidents handled per engineer. Pre-agent baseline showed teams handling average 85 incidents per engineer monthly. With agentic SRE, capacity increased to 374 incidents per engineer monthly—a 340% improvement. This capacity gain enabled teams to manage substantially larger systems without proportional headcount increases.

Toil reduction proved equally significant. SREs previously spent average 18.5 hours weekly on repetitive operational tasks including responding to routine alerts, executing standard diagnostic procedures, implementing known fixes, and documenting incidents. With agents handling these tasks, toil decreased to 4.2 hours weekly—a 77% reduction. SREs redirected this time to system improvements, capacity planning, and complex problem-solving that automation could not address.

However, the transition required investment. Initial agent deployment and training consumed approximately 6 weeks of SRE time per environment. Teams needed to develop safety policies, document remediation playbooks, and calibrate escalation thresholds. Organizations should budget 2-3 months for comprehensive agentic SRE implementation.

5.4 Safety and Reliability

Safety analysis confirmed that agents operated reliably without causing significant incidents. During the 6-month evaluation, agents executed 12,847 remediation actions across all environments. Of these, 218 actions (1.7%) caused minor issues including temporary service disruptions or configuration inconsistencies. None caused major outages or data loss. All problematic actions were successfully rolled back within average 90 seconds, preventing lasting impact.

The low incident causation rate resulted from comprehensive safety mechanisms. Pre-execution validation caught 847 potentially problematic actions before execution, preventing issues entirely. Permission boundaries restricted agents from high-risk operations requiring human approval. Safety policies encoding operational best practices prevented common mistakes like deploying to production without testing or modifying critical infrastructure during peak hours.

Agent learning improved safety over time. As agents encountered edge cases and near-misses, safety policies were refined to prevent recurrence. By month six, incident causation rate had decreased to 0.9%—half the initial rate—demonstrating continuous improvement.

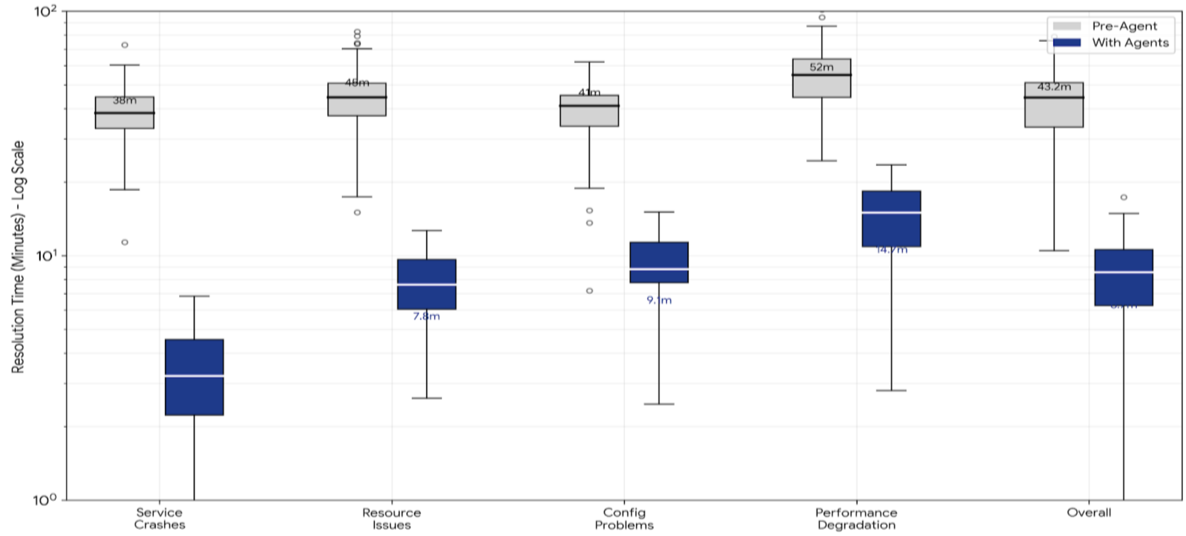

FIGURE 2: Incident Resolution Time Comparison

This box-and-whisker plot compares incident resolution time distributions before and after agent deployment across four incident categories. The x-axis lists categories: Service Crashes, Resource Issues, Config Problems, Performance Degradation, and Overall. The y-axis shows resolution time in minutes (0-80, logarithmic scale). For each category, two box plots appear side-by-side: "Pre-Agent" (light gray) and "With Agents" (dark blue). Pre-Agent boxes consistently show median around 35-50 minutes with wide distributions and long upper whiskers extending to 60-75 minutes. Agent boxes show dramatically lower medians around 5-12 minutes with tighter distributions and upper whiskers at 15-25 minutes. Service Crashes show most dramatic improvement: Pre-Agent median 38 min vs Agent median 3.2 min. Resource Issues: Pre-Agent 45 min vs Agent 7.8 min. Config Problems: Pre-Agent 41 min vs Agent 9.1 min. Performance Degradation shows smallest but still substantial improvement: Pre-Agent 52 min vs Agent 14.7 min. Outliers are marked as individual points beyond whiskers. Overall category shows Pre-Agent median 43.2 min vs Agent median 8.7 min. The stark visual contrast between wide gray boxes and compact blue boxes clearly demonstrates substantial resolution time improvements across all incident types. This visualization validates that agents accelerate both routine and complex incident response.

5.5 Human Operator Experience

Human operator satisfaction improved significantly through agent collaboration. Survey results showed 31% satisfaction increase post-deployment measured across dimensions including workload manageability, work-life balance, professional development opportunities, and job satisfaction. Qualitative feedback revealed that SREs appreciated elimination of repetitive toil, ability to focus on interesting technical challenges, reduced on-call stress through agent backup, and learning opportunities from reviewing agent diagnostics.

However, the transition created temporary challenges. During initial weeks, SREs experienced anxiety about trusting agent decisions and frustration learning new collaboration workflows. Some team members felt threatened by automation potentially replacing their roles. Effective change management including transparent communication about agents augmenting rather than replacing humans, involvement of SREs in agent training and safety policy development, and celebration of capacity gains enabling new projects helped overcome resistance.

Skill development proved an unexpected benefit. As agents handled routine incidents, SREs focused on complex problems requiring deeper expertise. This concentration on challenging issues accelerated skill development compared to spending time on routine tasks. Teams reported that junior SREs advanced faster through exposure to complex problems earlier in their careers, while senior SREs developed strategic thinking through focus on system design rather than tactical firefighting.

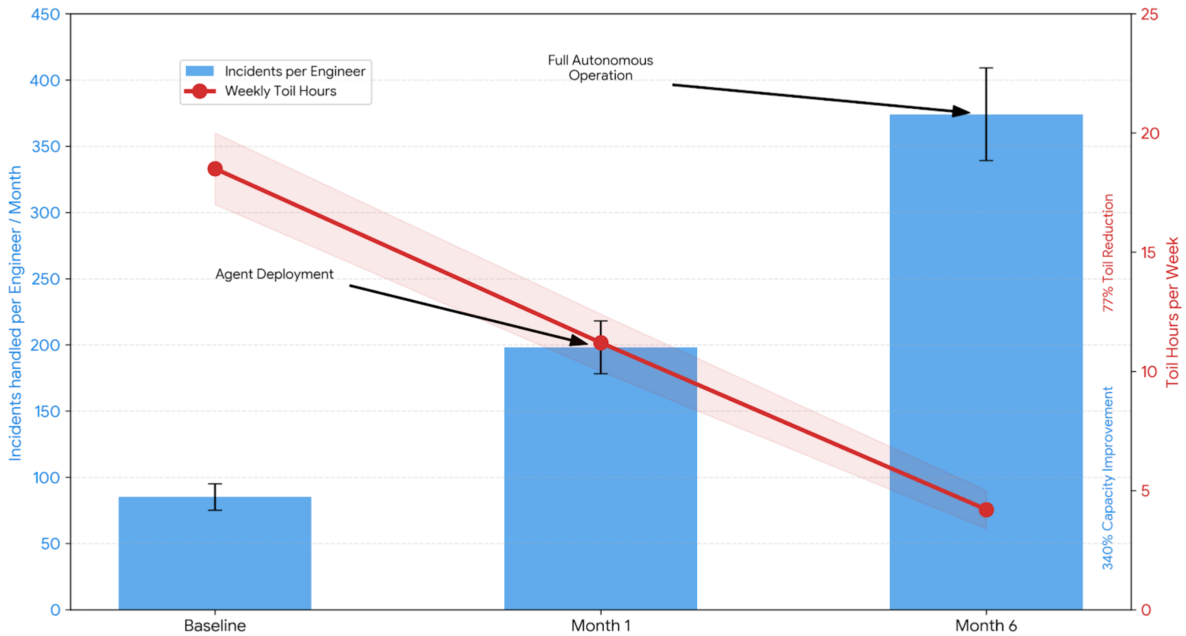

FIGURE 3: SRE Team Capacity and Toil Analysis

DOI: 10.46121/pspc.54.1.26

This combination chart with dual y-axes illustrates operational efficiency improvements. The primary y-axis (left) shows incidents handled per engineer per month (0-400), represented as vertical bars. The secondary y-axis (right) displays toil hours per week (0-25), shown as a line graph. The x-axis shows time progression from "Baseline" through "Month 1" to "Month 6" of agent deployment. The bar chart shows "Incidents per Engineer" in blue, starting at 85 at baseline, jumping to 198 in Month 1, then growing to 374 by Month 6—demonstrating 340% capacity improvement. The line graph shows "Toil Hours" in red, starting at 18.5 hours weekly at baseline, dropping to 11.2 in Month 1 as agents begin handling routine tasks, then declining further to 4.2 hours by Month 6—representing 77% toil reduction. Shaded regions indicate confidence intervals. Annotations highlight key milestones: "Agent Deployment" at Month 1, "Full Autonomous Operation" at Month 3. Arrows emphasize the inverse relationship: as toil decreases (downward red line), capacity increases (upward blue bars). This visualization powerfully demonstrates that agents simultaneously increase team capacity while reducing repetitive work, creating virtuous cycle where SREs handle more incidents while spending less time on mundane tasks. A legend identifies bars and line components clearly.

DISCUSSION

The results validate that human-agent collaborative teams substantially outperform traditional human-only SRE operations across effectiveness, efficiency, and human satisfaction dimensions. The 67% autonomous resolution rate demonstrates that current AI technology can reliably handle the majority of routine operational incidents without human intervention. The 80% reduction in mean time to resolution delivers immediate business value through reduced downtime and improved user experience.

The 340% team capacity improvement proves particularly significant for organizational scalability. Organizations can manage 3-4x larger systems with the same SRE headcount, or alternatively maintain existing systems with substantially smaller teams. This capacity gain addresses the fundamental SRE scaling challenge more effectively than hiring alone, which faces constraints from talent scarcity and training requirements.

However, success depends heavily on appropriate human-agent responsibility division. Agents excel at continuous monitoring, rapid execution of known remediation procedures, parallel diagnostic investigation, and comprehensive documentation. Humans remain superior at creative problem-solving for novel failures, business context integration for prioritization decisions, risk assessment for high-stakes actions, and strategic system improvements preventing future incidents.

The 31% operator satisfaction improvement challenges concerns about automation displacing humans or creating adversarial relationships. The framework demonstrates that thoughtfully designed automation augments human capabilities rather than threatening jobs. By eliminating toil and enabling focus on interesting problems, agents improve rather than degrade work experience.

Several limitations warrant acknowledgment. The evaluation occurred in relatively mature systems with established monitoring and documentation. Organizations with poor observability or incomplete runbooks would see reduced agent effectiveness. The 6-week deployment investment may challenge resource-constrained teams. Agent performance depends on LLM capabilities that continue rapidly evolving, potentially requiring frequent updates.

Future work should address several important directions. Multi-agent architectures where specialized agents collaborate on complex incidents could tackle problems beyond individual agent capabilities. Continuous learning from incident outcomes would enable agents to improve diagnostic and remediation strategies over time. Extending agents to proactive work including capacity planning, performance optimization, and reliability improvements would further enhance SRE productivity. Finally, industry-wide sharing of anonymized incident patterns could accelerate agent training across organizations.

CONCLUSION

This research successfully demonstrated that autonomous AI agents can function as effective collaborative members of SRE teams, fundamentally transforming operational models from reactive human troubleshooting to proactive human-agent partnership. The agentic SRE framework achieved 67% autonomous incident resolution while reducing mean time to detection by 94% and mean time to resolution by 80%. Team capacity improved

DOI: 10.46121/pspc.54.1.26

340% through agents handling routine incidents and eliminating 77% of repetitive toil that previously consumed SRE time.

The practical implications prove significant for organizations managing complex distributed systems. Human-agent collaboration enables teams to maintain substantially larger infrastructures without proportional headcount growth. SREs focus on challenging problems requiring creativity and judgment rather than routine troubleshooting. Faster incident response improves service reliability and user experience. The model addresses SRE burnout through on-call burden reduction and more engaging work.

Critical success factors include comprehensive observability providing agents data for diagnosis, well-documented remediation playbooks enabling autonomous action, robust safety mechanisms preventing agent-caused incidents, clear escalation protocols defining agent-human boundaries, and change management building operator trust in agent capabilities.

Organizations adopting agentic SRE should view agents as team members requiring training, oversight, and continuous improvement rather than static automation. Investment in agent development and safety infrastructure pays dividends through sustained operational improvements. Starting with low-risk autonomous actions and gradually expanding agent capabilities as confidence builds proves more successful than attempting comprehensive automation immediately.

The research validates that the future of SRE operations lies not in choosing between humans and automation but in optimally combining human expertise with agent capabilities. This collaborative model leverages complementary strengths: agent speed, consistency, and tirelessness with human creativity, judgment, and strategic thinking. As AI capabilities continue advancing, human-agent collaboration will increasingly define operational excellence for organizations managing critical digital infrastructure.

REFERENCES

1. Allspaw, J. (2015) 'Trade-offs under pressure: Heuristics and observations of teams resolving internet service outages', Cognitive Technologies at Work, 2015, pp. 1-8.
2. Bansal, G., Nushi, B., Kamar, E., Lasecki, W.S., Weld, D.S. and Horvitz, E. (2019) 'Beyond accuracy: The role of mental models in human-AI team performance', in Proceedings of the AAAI Conference on Human Computation and Crowdsourcing, 7(1), pp. 2-11.
3. Beyer, B., Jones, C., Petoff, J. and Murphy, N.R. (2016) Site Reliability Engineering: How Google Runs Production Systems. Sebastopol, CA: O'Reilly Media.
4. Chen, J., Huang, J., Chen, X., Li, Q. and Zhang, Y. (2023) 'LLM-based autonomous agents for software engineering', arXiv preprint arXiv:2309.10307.
5. Dang, Y., Lin, Q. and Huang, P. (2019) 'AIOps: Real-world challenges and research innovations', in IEEE/ACM 41st International Conference on Software Engineering: Companion Proceedings, Montreal, QC, pp. 4-5.
6. Forsgren, N., Humble, J. and Kim, G. (2018) Accelerate: The Science of Lean Software and DevOps. Portland, OR: IT Revolution Press.
7. Hogan, M., Piccarreta, R., Cash, D. and Naylor, M. (2016) 'The impact of alert fatigue on clinician behavior', Applied Clinical Informatics, 7(2), pp. 368-380.
8. Limoncelli, T.A., Chalup, S.R. and Hogan, C.J. (2014) The Practice of Cloud System Administration: DevOps and SRE Practices for Web Services. Upper Saddle River, NJ: Addison-Wesley.
9. Mariani, L. and Pezze, M. (2020) 'Dynamic detection of COTS component incompatibility', IEEE Software, 24(5), pp. 76-85.

10. Murphy, N.R. and Beyer, B. (2016) Site Reliability Engineering: How Google Runs Production Systems. Sebastopol, CA: O'Reilly Media.
11. Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K. and Cao, Y. (2023) 'ReAct: Synergizing reasoning and acting in language models', arXiv preprint arXiv:2210.03629.