



# Generative AI Agents for Site Reliability Engineering

**Yiwen Zhang**

Principal GenAI Solutions Architect  
Amazon Web Services

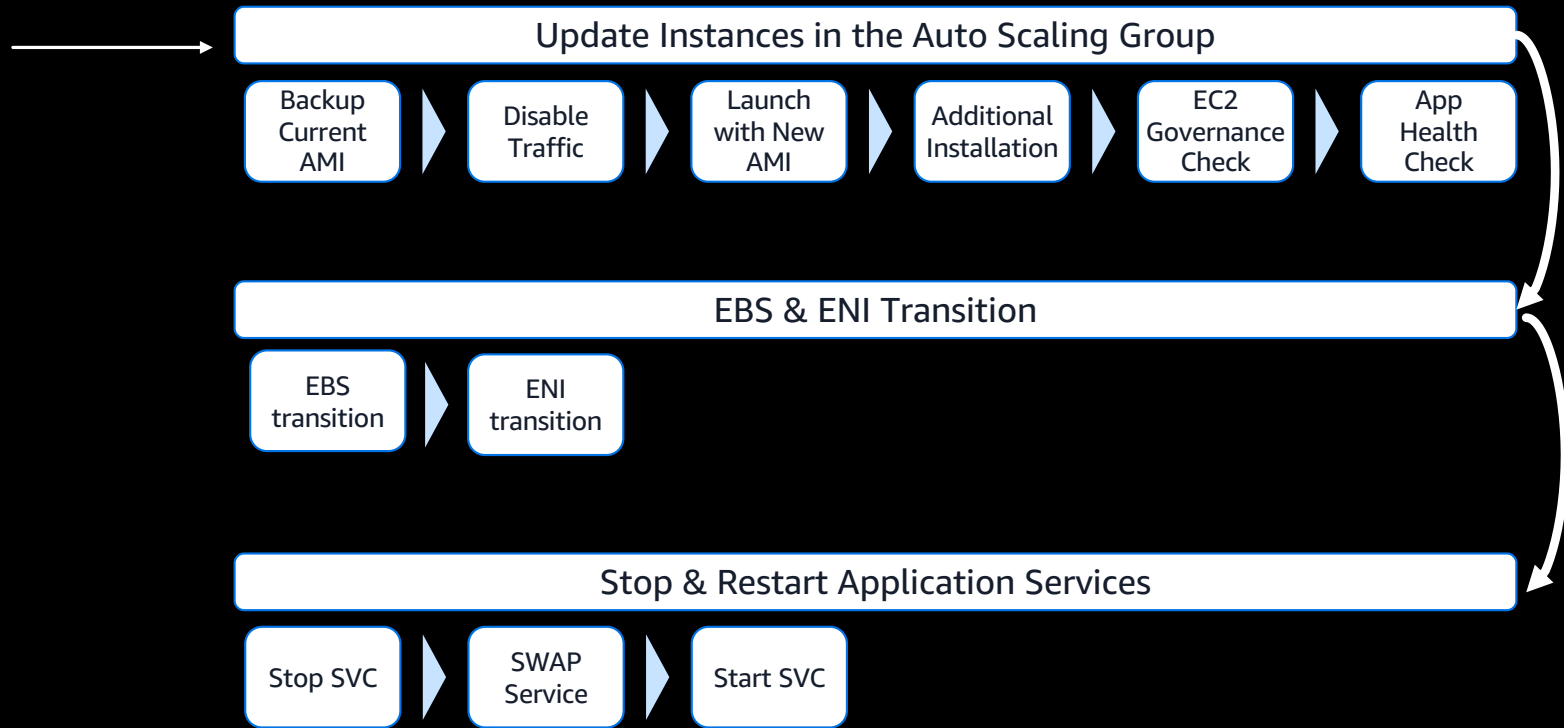
# The Struggle of Site Reliability Engineering

- 1 Complex modern application infrastructure
- 2 Alert fatigue
- 3 Time consuming manual incident triage
- 4 Knowledge siloed across teams
- 5 Reactive rather than proactive problem resolution

# EC2 Image Upgrade Use Case







Start



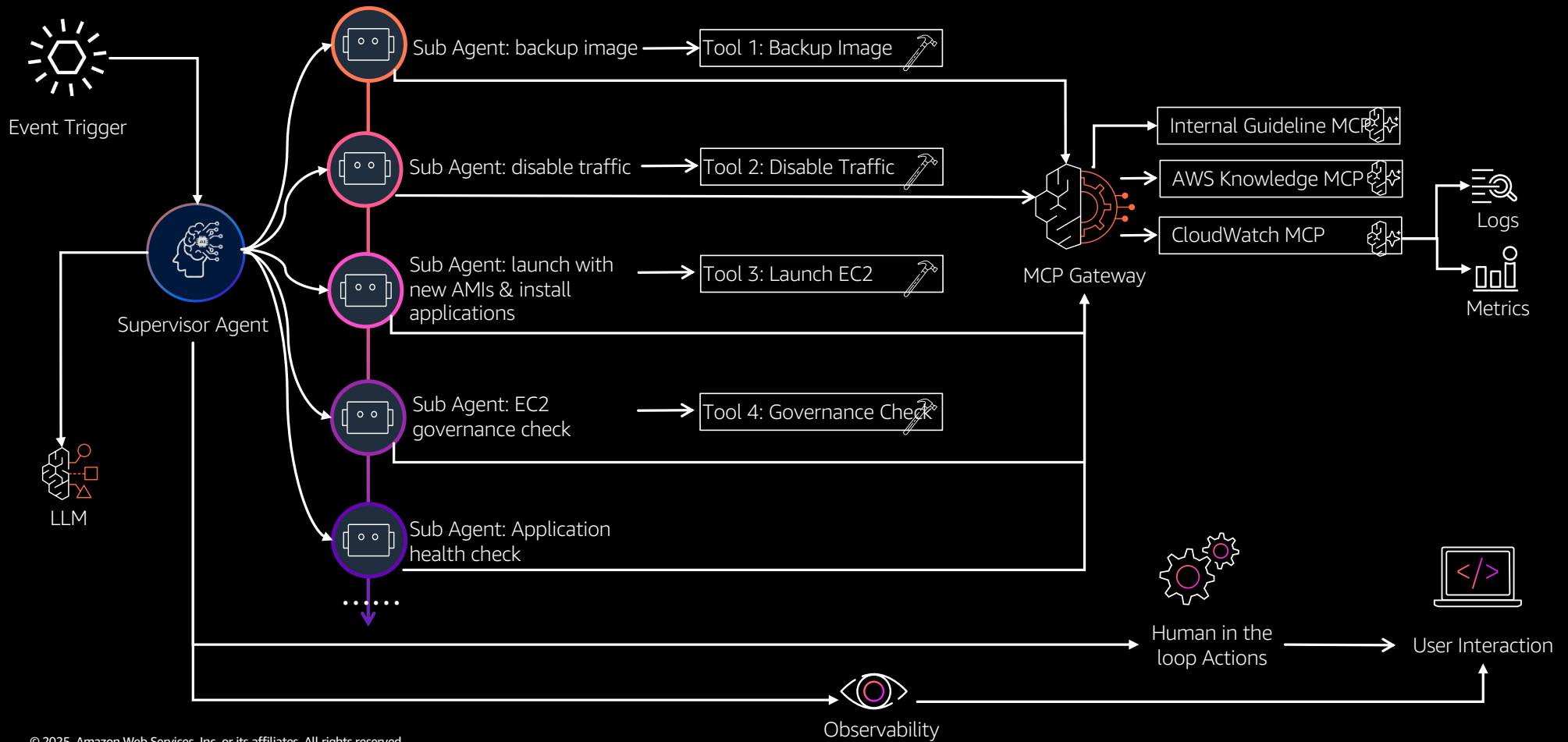
# What is a Generative AI agent

A software program that can interact with its environment, collect data, and use the data to perform self-determined tasks to meet the goals.

## Key characteristics of Generative AI agents

-  **Reasoning:** Use logic and available information to draw conclusions, make inferences, and solve problems.
-  **Memory:** Combine long and short-term memory to process chat-based prompts and action on tasks.
-  **Planning:** Create step-by-step plans with milestone goals from a prompt while learning from mistakes.
-  **Tool use:** Query APIs to request additional information or execute actions.

# SRE AI Agent



# SRE AI Agent in Action

Launch new EC2 instance step got error

```
ERROR: hudson.AbortException: Prepare rehyd directory :fatal: [xxx.xxx.xx.xxx]: FAILED! => { ansible_job_id :  
a1234556 , attempts : 1, changed : true, cmd : mkdir -p /app/cassandra/generic/rehyd/;\nmkdir -p  
/app/cassandra/generic/rehyd/temp/scripts/;\nchown -R cassandra:cassandra /app/cassandra/generic;  
\nchmod -R 0755 /app/cassandra/generic;  
\nrm -rf /app/cassandra/generic/rehyd/*;\nsu - cassandra -c \ cp /home/cassandra/.profile  
/app/cassandra/generic/rehyd/;\nncp /home/cassandra/env_files/envcassandra /app/cassandra/generic/rehyd/\ ;  
\nsu - cassandra -c \ crontab -l > /app/cassandra/generic/rehyd/crontab_copy\  
\ncrontab -l > /app/cassandra/generic/rehyd/crontab_root_copy;\ngrep vm.nr_hugepages /etc/sysctl.d/99-tuning-  
for-cassandra.conf > /app/cassandra/generic/rehyd/hugepages.tmp;\nncp /etc/newrelic-infra.yml  
/app/cassandra/generic/rehyd/ 2>/dev/null;  
\n , delta : 0:00:00 , end : 2025-04-01 01:40:10 , finished : 1, msg : non-zero return code , rc : 1
```

# SRE AI Agent in Action

Key section of the system prompt for the sub agent

```
"steps": [  
  "1. Identify the error type: Parse the error message to determine if it's a package installation, permission, network, or script syntax error",  
  "2. Check the error context: Review the command that failed and the surrounding log entries",  
  "3. Verify prerequisites: Ensure required files, directories, users, or services exist before the failed step",  
  "4. Test the failed command: Extract and run the specific command manually via SSH to reproduce the issue",  
  "5. Check return codes: Look for 'rc: X' or exit codes to understand the failure reason",  
  "6. Review dependencies: Verify all dependencies for the failed step are properly installed",  
  "7. Validate permissions: Check file/directory ownership and permissions for the operation",  
  "8. Fix and retest: Apply the fix and test incrementally before full redeployment"  
],  
"error_patterns": {  
  "package_installation_failures": {  
    "indicators": [  
      "No package <name> available",  
      "Failed to download metadata",  
      "Error: Nothing to do",  
      "dpkg: error processing package",  
      "E: Unable to locate package",  
      "Package <name> is not available",  
      "Error: GPG check FAILED"  
    ],  
    "common_causes": [  
      "Missing or incorrect repository configuration",  
      "Network connectivity issues preventing repository access",  
      "Package name typo or version not available in repositories",  
      "Repository metadata cache out of date",  
      "GPG key verification failures",  
      "Insufficient disk space for package installation"  
    ],  
    "fixes": [  
      "Run 'yum clean all' or 'apt-get update' before package installation",  
      "Verify Security Group allows outbound traffic on ports 80/443",  
      "Check repository URLs are accessible: curl <repo-url>",  
      "Install EPEL repository for additional packages: yum install -y epel-release",  
      "Import GPG keys: rpm --import <key-url>",  
      "Verify disk space: df -h"  
    ]  
  }  
}
```

# SRE AI Agent in Action

## Agent Suggested Fixes

```
# Replace the failing shell command with this version that has proper error handling:

mkdir -p /app/cassandra/generic/rehyd/;
mkdir -p /app/cassandra/generic/rehyd/temp/scripts/;
chown -R cassandra:cassandra /app/cassandra/generic;
chmod -R 0755 /app/cassandra/generic;
rm -rf /app/cassandra/generic/rehyd/*;
su - cassandra -c "cp /home/cassandra/.profile /app/cassandra/generic/rehyd/" 2>/dev/null || true;
cp /home/cassandra/env_files/envcassandra /app/cassandra/generic/rehyd/ 2>/dev/null || true;
su - cassandra -c "crontab -l > /app/cassandra/generic/rehyd/crontab_copy" 2>/dev/null || true;
crontab -l > /app/cassandra/generic/rehyd/crontab_root_copy 2>/dev/null || true;
grep vm.nr_hugepages /etc/sysctl.d/99-tuning-for-cassandra.conf > /app/cassandra/generic/rehyd/hugepages.tmp
2>/dev/null || true;
cp /etc/newrelic-infra.yml /app/cassandra/generic/rehyd/ 2>/dev/null || true;
```



# Thank you!

**Yiwen Zhang**

[yiwenzhg@amazon.com](mailto:yiwenzhg@amazon.com)

 [linkedin.com/in/yiwenzhg](https://www.linkedin.com/in/yiwenzhg)