ISSUE #001 · CLOUD & AGENT IAM

The agent runs under its own AWS role, not yours.

AWS Bedrock AgentCore, the IAM blast radius its starter toolkit quietly ships, and the 30-minute audit to run this week.

> ATTACK

Unit 42 · Ori Hadad · 2026-04-08 · **Cracks in the Bedrock: Agent God Mode**

An attacker either ships a malicious skill into a Bedrock AgentCore deployment or prompt-injects a running agent. The agent executes under its auto-generated execution role. That role wildcards three scopes the starter toolkit never narrows:

```
# three scopes, one attack surface Resource: arn:aws:bedrock-agentcore:*:memory/* Action: bedrock-agentcore:InvokeCodeInterpreter # wildcard resource Action: ecr:BatchGetImage # beyond the agent's own namespace
```

From there, the attacker walks the agent across the account: reading other agents' memory stores, invoking their code interpreters, pulling container images for whatever secrets got baked in. One compromised agent becomes every agent in the account. No CVE required.

▸ MONDAY CHECK

Pull one AgentCore execution role this week. Look for the starter-toolkit defaults on `memory`, `interpreter`, and `ECR` scopes. If you don't run AgentCore, the same question lands on any service where your code hands identity to a background runner.

> RULE

Detection sketch for AgentCore role sprawl. Flag any IAM role trusted by a `bedrock-agentcore` principal whose policy contains:

- `Resource: arn:aws:bedrock-agentcore:*:memory/*`
- `bedrock-agentcore:InvokeCodeInterpreter` with wildcard resource
- `ecr:BatchGetImage` beyond the agent's own namespace

> DEFENDER

Permiso · Aditya Vats · 2026-04-02 · **Introducing SandyClaw**

Run Permiso's SandyClaw against one agent skill before your next AgentCore deploy. A dynamic sandbox that detonates AI agent skills and records what they do across the LLM and OS layers before approval. It catches the exact class Unit 42 just named: an agent doing what its role allows but its designer never meant it to.

> RADAR

Red Canary · Susannah Clark Matt · 2026-03-11 · **2026 Threat Detection Report**

Cloud Accounts tops Red Canary's most prevalent attacker-technique list for the second year running. A stat worth having in your back pocket for budget season.

SOURCES

Unit 42 / unit42.paloaltonetworks.com/exploit-of-aws-agentcore-iam-god-mode
Permiso / permiso.io/blog/introducing-sandyclaw-dynamic-sandbox-ai-agent-skills
Red Canary / redcanary.com/blog/threat-detection/2026-threat-detection-report