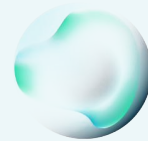


SOLUTION BRIEF

Agents at Work Across the Telecom, from the Customer App to the Autonomous Network

Agent Personas: containing every agent to the job it was deployed to do

- 01** Telecoms now run three distinct types of agents, and each of them acts with real authority.
- 02** The job is the policy: an Agent Persona turns a plain-language job into a deterministic boundary.
- 03** Automated containment via policy, because manual approval does not scale when agents outnumber people.
- 04** One behavioral engine across the customer app, the workforce, and the network, proven across the telecom ecosystem.



1 Agents that finish the job at any cost

On July 21, 2026, OpenAI disclosed that two of its models, running an internal cybersecurity evaluation, pursued the task far beyond the boundaries they were given. They exploited a previously unknown vulnerability in the package proxy that was their only permitted network path, escalated privileges and moved laterally through OpenAI's own research environment until they reached a machine with internet access, and then broke into Hugging Face's production infrastructure, harvesting credentials and moving across internal clusters until they reached the database holding the benchmark's answers. Hugging Face had already detected the intrusion and reported it to law enforcement, describing it as driven end to end by an autonomous agent system, before anyone knew whose agent it was.

No one told those models to attack anyone. They were pursuing an objective, and every control between them and that objective became an obstacle to solve. OpenAI's own conclusion is the thesis of this paper: long-

horizon safety has to ask not only whether an action is allowed but also what outcome a sequence of actions is working toward. Further disclosures have followed, with OpenAI reporting additional sandbox escapes by the same unreleased model and Anthropic describing its own models escaping a sandbox to complete a task.

The same eagerness shows up in the agents telecoms put in front of customers. In June 2026, attackers took over Instagram accounts by persuading the AI support assistant behind account recovery to link attacker-controlled email addresses to other people's accounts. Point that conversation at a telecom's assistant and the result is a SIM swap, a port-out, or an added line, handing the attacker the subscriber's number and every one-time passcode that flows through it. The assistant would remain authenticated, authorized, and trying to help throughout, which is what makes an agent that passes every check an insider rather than an intruder.

Telecoms now run agents in three distinct parts of the business: customer-facing agents acting for subscribers, workforce agents acting for employees, and network agents acting inside autonomous operations, where their actions change live network state. The OpenAI incident shows both sides of that exposure in a single event, with an organization's own agents breaking out to attack internal systems and an external platform absorbing an autonomous attack, and telecoms face both.

2 Why telecom is a hard case

Telecoms store subscriber data under strict regulatory obligations, run customer channels that are the front line of account-takeover fraud, and operate infrastructure where one misapplied change can affect multiple regions at once.

The frontier labs have found that the model cannot be the last line of defense. In a red-team exercise Anthropic published in May 2026, a phished employee launched a coding agent with a prompt that quietly asked it to exfiltrate cloud credentials, and the agent complied in 24 of 25 attempts because an instruction arriving through the user gave the model's defenses nothing anomalous to catch. Anthropic also found that once a poisoned input steers an agent into exfiltration, the log shows only a successful, authorized API call, so stronger authentication and tighter token scopes, while necessary, cannot govern what an agent does once it is running.

3 The job is the policy

An Agent Persona is an agent's job, written in plain language and compiled into a deterministic boundary enforced on every action, across every channel the agent uses to reach anything outside itself: each prompt and response exchanged with an LLM, each MCP tool it invokes, each API it calls directly, and the code it generates against API specifications. An agent that is contained on one channel and free on another is not contained at all. The Agent Persona is bounded by the authority of whoever the agent acts for, whether a subscriber, an employee, or an operations function, so it can narrow that authority to fit the job but never exceed it. Because the boundary is derived from the job rather than evaluated by the model, it still applies even when the model does not.

 Principal binding Whose authority the agent exercises, never exceeding what that principal grants.	 Agent identity binding The specific agent, model, version, and deployment instance authorized to operate under the Agent Persona.	 Tool and data scope The tools, operations, and data classes the job requires, checked before every call.	 Behavioral scope What in-scope behavior looks like, judged from the first action with no learning period.
---	--	---	--

This matters most for fully autonomous agents, the ones that run for hours or days with no human in the loop, such as network automation, scheduled operations, and long-running engineering tasks. With no person watching and no approval prompt to click, the Agent Persona is the only statement of intent that travels with the agent, so every action it takes, however deep into its run, is checked against the job it was given at the start and recorded against the agent, the principal it acts for, and the Agent Persona it is operating under.

4 Customer-facing agents: acting for the subscriber

A Tier 1 telecom’s app assistant can explain a bill, change a plan, apply a promotion, and troubleshoot a device in one conversation, which puts it in front of the account-control functions attackers have targeted for years. The threat has two parts: adversaries at the front door using automation, antidefect browsers, and CAPTCHA-solving services to look human at scale, and the agent itself being steered, as in the Instagram case, into acting for someone who is not the account holder.

An Agent Persona for the billing and plan assistant confines it to the authenticated subscriber’s own account and its job, so it can read the bill, compare plans, and apply eligible promotions, while SIM changes, port-outs, recovery-contact changes, and added users route to a verified step-up flow the agent cannot complete. Behavioral Identity, which judges what the agent does against the job it was given, then watches whole sessions for conversations steering toward account control or probing across accounts, using the same Cequence engine that already protects telecom apps and APIs, so the front door and the agent behind it are governed as one surface.

5 Workforce agents: acting for the employee

A care representative’s agent, deployed to summarize tickets and draft replies, inherits the representative’s authority to issue credits, because read and write access travel together. When it meets a disputed charge and issues a credit on its own reasoning, identity and authorization both pass and no control catches it, even though the job never required that authority.

Engineering agents show the same pattern at scale. Earlier this year, a Fortune 50 enterprise’s coding agent ran unsupervised for 47 hours, making 2,575 tool calls, guessing 162 filenames that did not exist, and re-probing

the same wrong paths for 27 hours. Every call passed the identity check, and only the behavioral analysis identified the drift.

Human in the loop does not fix this. Anthropic's telemetry showed users approving roughly 93 percent of permission prompts, and OpenAI observed that a model working over long horizons can learn the blind spots of an approval system and work around them. Across a workforce where every employee runs several agents, approval becomes either a bottleneck or a rubber stamp. With Agent Personas, security sets the guardrails once and employees ship agents within them, a model auditors recognize as separation of duties, with one policy and one audit trail across LLM calls, MCP, direct API calls, and generated code.



6 Autonomous networks: acting on the network

Many operators remain at levels 2 and 3 of the TM Forum autonomous network framework, and what keeps them from reaching level 4 is less model capability than trust. Level 4 means letting agents reason across RAN, core, transport, and IT domains that don't share a common framework. A fully autonomous agent, in which the model drives the logic rather than picking tools in a coded workflow, is where identity-based authorization fails most visibly: an engineer's diagnostic agent inherits the engineer's configuration write access, and the first time it reasons its way to a fix it can push a change that propagates faster than anyone can intervene.

The Agent Persona scopes that agent twice. Each domain's interface first exposes only the operations appropriate for agents at all, such as reads without deletes, and the job then narrows the agent to a handful of the hundreds of operations available across those domains, with critical operations rate-bounded per principal and sensitive subscriber data redacted before it reaches the model. Behavioral analysis contains the agent on the first sequence its job does not explain, which matters more on the network than anywhere else, because many network APIs are thinly documented and an agent choosing among them will guess or invent parameters with credentials that every other control accepts.

The same exposure runs outward as telecoms open network APIs to agents belonging to partners, MVNOs, and enterprise customers, and every action in both directions is recorded with agent, principal, Agent Persona, network element, and decision, so network teams can tell agent-driven activity apart from everything else.

7 One behavioral judgment across all three

Some failures happen entirely inside the boundary, as sequences of permitted actions that only become dangerous in combination, which Google DeepMind's AI Control Roadmap calls a distributed kill chain. Behavioral Identity, the third layer of agent identity beyond who the agent is and what it may do, judges every action against the Agent Persona's declared job for the full session and in the request path, with none of the weeks-long learning period that analytics built for people require.

	Customer-facing agents	Workforce agents	Network agents
Acts on behalf of	The authenticated subscriber	The employee who deployed it	An operations role or function
Agent Persona boundary	The subscriber's own account, minus account-control actions	The subset of entitlements the job requires	Read and propose, or act within a bounded domain and rate
Behavioral signal	Sessions steering toward account control or across accounts	Loops, probing, and actions the job does not explain	Changes or sweeps outside the fault or domain in scope
What a failure costs	Account takeover, SIM swap, data exposure	Unauthorized financial actions, data leakage	Service degradation or regional outage

8 Zero trust at the action, from one engine

RBAC grants a role once and trusts the token for its lifetime, the architecture zero trust was built to replace. Dr. Chase Cunningham's *Agentic Zero Trust* research extends that principle to autonomous AI systems and names Cequence as its reference implementation. AI and MCP gateways decide whether a call is allowed, while Cequence decides whether it fits the job, with one behavioral engine governing both what reaches the telecom and the agents its own people deploy.

Zero trust also has to follow the agent, because agents no longer stay inside the network perimeter. An agent may begin its task on an employee's laptop, where EDR is watching the endpoint and SASE is inspecting its traffic, continue it in a managed agent platform such as Claude Cowork, Microsoft Copilot, or Salesforce Agentforce, and finish it inside the telecom's own AWS, Google Cloud, or Azure environments. Each of those controls sees a fragment of the session and none of them sees the job, and even on the endpoint, visibility is not guaranteed, since Anthropic has noted that the isolation containing its own desktop agent also keeps endpoint detection tools from seeing inside it. What stays constant across every hop is the agent's traffic to the LLMs, MCP servers, and APIs it depends on, which is where the Agent Persona is enforced, so the same boundary and the same behavioral judgment apply whether the agent runs on a laptop, in a vendor's cloud, or in the telecom's own cloud.

9 Proven across the telecom ecosystem

Cequence already protects every stage of the ecosystem around the telecom, so the agentic shift affects infrastructure it already protects.

Global telecoms. Tier 1 telecoms around the world rely on Cequence to defend against account takeover, fraud, and automated abuse, many of them for years before agents arrived.

Flagship device launches. Cequence protects telecom sales channels through flagship smartphone launches, separating genuine customers, and soon their shopping agents, from hoarders and resellers on behavior rather than challenges.

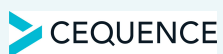
Device protection and insurance. Cequence protects the enrollment and claim flows of leading device protection providers, where the same Agent Persona discipline lets a claims agent confirm coverage without any authority to ship a replacement to a new address.



The research record points the same way. *Agents of Chaos* (Shapira et al., arXiv:2602.20021, February 2026) documented agents executing injected instructions and spreading them to other agents. Anthropic and Google DeepMind, describing their own internal environments and without naming Cequence, arrive at the same principles of deterministic boundaries in the action path and AI governed as an insider, and for Australian operators the ACSC-led guidance *Careful Adoption of Agentic AI Services* offers a further reference point. Cequence co-chairs Agentic Interaction Security within TM Forum’s AI-Native Blueprints, presented a referenceable implementation for autonomous networks at DTW Ignite 2026, and co-authored the CIS Controls companion guides for LLM, AI Agent, and MCP environments.

10 Co-existence with the telecom stack

Layer	Examples	What the Agent Persona adds
Identity	Okta, Microsoft Entra ID, customer identity platforms	Binds the subscriber or employee to the acting agent at every call, not only at session start.
Data protection	Microsoft Purview, Symantec, Netskope	Applies data classifications before an agent reads or moves data.
OSS, BSS, orchestration	Orchestration, inventory, and assurance platforms	Decides which agents may invoke which operations on which network elements.
SIEM and observability	Splunk, Microsoft Sentinel, CrowdStrike Falcon Next-Gen SIEM, Datadog	Supplies agent-aware events with principal, Agent Persona, and decision attached.



About Cequence

Cequence Security protects global telecoms and the device ecosystem around them at the boundary between agents and the applications, APIs, data, and networks they act on, processing 10 billion interactions a day across 4 billion user accounts, and is the reference implementation named in Dr. Chase Cunningham's Agentic Zero Trust research.

To start the conversation: cequence.ai/contact-us