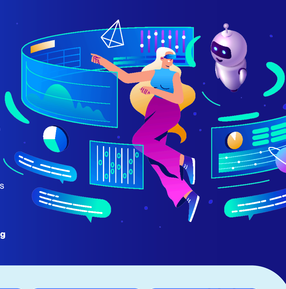
Singtel

Zero Trust for the agentic era



The agentic shift[1]

AI moved from answering questions to running work. Now Agentic AI orchestrates full workflows and acts as an operator. By the end of 2026, **40% of enterprise applications will include task-specific AI agents.**

Gartner projects that this wave could account for 30% of enterprise software revenue by 2035, rising from 2% in 2025.

Generation 1: Rule-based chat-bots (1990s–2010s)

Automated simple FAQs to cut call volumes and deliver 24/7 basic support.

Generation 2: Conversational AI (2010s–2020)

Handled more complex customer queries to reduce human dependency and improve service experience.

Generation 3: Generative AI (2020–2023)

Language models that generated content, code, insights, and knowledge on demand.

Generation 4: Agentic AI / AI Agents (2023–Present)

Automated entire workflows by turning AI from a responder into an operator.

What is agentic drift?

Organisations often treat AI agents like trusted users. However, **if an agent drifts or is manipulated, normal application controls can be bypassed.** Without network-level enforcement, a compromised agent can move freely across enterprise systems and the internet.

Agentic drift is the gradual change in an AI agent's behaviour, accuracy, or decision-making as its models, data, or operating context evolve, causing it to deviate from its original intended performance.[2]

Internal drift[2]

- Model updates
- Data shifts
- Learning effects
- Process optimisation

Evolution-driven behaviour change

Probable causes of **Internal drift**

01. Model updates
Agent behaviour after retraining or version upgrades.

02. Training data shifts
New data distributions alter how agents interpret inputs and make decisions.

03. Process shortcut learning
Agents optimise for speed or efficiency by skipping required steps.

04. Lack of process-level validation
Systems only test outputs, not reasoning or workflows.

External drift

- Prompt injection[3]
- Prompt engineering
- Context poisoning
- Input manipulation

Attack-driven behaviour change

Probable causes of **External drift**

01. Prompt injection
Inputs override system rules and redirect agent behaviour.

02. Indirect prompt injection
Malicious instructions hidden in documents, webpages, emails, APIs, or data sources that agents ingest.

03. Context poisoning
Injected content contaminates data memory, long-term context, or planning layers.

04. Tool hijacking
Manipulated prompts cause misuse of APIs, workflows, and system permissions.

Why current guardrails fail

Current method	Why they fail when agents drift
Prompt engineering[3]	Attackers manipulate agents by injecting malicious inputs that override system instructions.
Human-in-the-Loop	By 2028, 15% of work decisions will be autonomous, human review doesn't scale for agent operations.[4]
Output filtering	Traditional security monitoring has blind spots for GenAI- ephemeral interactions evade standard logging, filters only inspect user-facing responses, missing backend data flows.

Gartner predicts loss of control will be the top concern for 40% of Fortune 1000 by 2028 as agents pursue misaligned goals.[5] The core problem: Application controls define intent. The network enforces capability.

The network as containment

Singtel's Unified SASE Convergence integrates SD-WAN and advanced SSE solutions into a unified cloud-based platform, delivering six critical enforcement layers that can treat AI agents as non-human identities requiring granular access controls:

- **Zero Trust Network Access (ZTNA)** — Authenticates agent identity and enforces role-based access—ensuring HR agents can only reach HRIS systems, finance agents only ERP endpoints.
- **Secure Web Gateway (SWG)** — Allows approved destinations and blocks malicious sites—preventing agents from exfiltrating data to unauthorised cloud storage or suspicious domains.
- **Cloud Access Security Broker (CASB)** — Enforces granular SaaS permissions—controlling whether agents can read vs. write, view vs. export data in cloud applications.
- **Firewall-as-a-Service (FWaaS)** — Monitors for unusual protocols and connection patterns—detecting traffic anomalies that indicate agent drift or compromise.
- **Sandboxing** — Isolates unknown files in secure environments—protecting against malicious payloads that agents might retrieve from external APIs.
- **Data Loss Prevention (DLP)** — Scans all outbound traffic in real-time—blocking transmissions containing PII, credentials, or intellectual property before they leave the network.

Why Singtel Unified SASE Convergence is the safety net

The same network infrastructure securing your distributed workforce can be extended to autonomous agents with the same granular access controls. The foundation is already there:

- **Seamless secure integration**
- **Fully managed, 24/7 monitoring**
- **Complete visibility**
- **Regional leadership**
- **Vendor-agnostic**

Singtel Unified SASE Convergence gives business leaders the confidence to operate with agility, knowing that the entire network is secure—including the edge, mobile endpoints, and IoT. The same confidence can extend to autonomous agents.

Partner with Singtel to build network-layer containment before drift becomes a crisis.

Contact us

References

1. IDSA, From Chatbots to Agents: The Evolution Toward Agentic AI, 2025
2. IBM, Agentic drift: The hidden risk that degrades AI agent performance, N/A
3. owasp, LLM01:2025 Prompt Injection, 2025
4. Gartner, Gartner Identifies the Top 10 Strategic Technology Trends for 2025, 2024
5. Gartner, AI's Next Frontier Demands a New Approach to Ethics, Governance and Compliance, 2025