

Zero Trust Agentic Security Checklist (starter set)

Version 0.1 · CC BY-NC 4.0 — attribute to Contextual Agentic, contextualagentic.com

Twenty-four checks for an agent that acts on enterprise systems. Each is written so the answer is evidence, not an opinion — if you cannot point at something, the answer is no.

Scope. This is a starter set of 24 checks. Appendix B of the book holds the complete checklist: 71 control objectives in twelve tables, each with the evidence that answers it. The full checklist is in the book.

Drafted for the companion material of The Contextual Agentic Enterprise. This is a working checklist, not a control framework, and it maps to no certification scheme.

Identity

ZT-01 — Does the agent have an identity of its own, distinct from the human it acts for?

Evidence that answers it — The principal in an access log entry for an agent action.

Common wrong answer — It uses a service account shared with three batch jobs.

ZT-02 — Is the human principal it acts on behalf of recorded on every action?

Evidence that answers it — An on-behalf-of claim carried to the enforcement point.

Common wrong answer — The user is in the prompt, so the model knows who it is.

ZT-03 — Are the agent's credentials short-lived and issued per task?

Evidence that answers it — A token lifetime measured in minutes, and an issuance record.

Common wrong answer — A long-lived key in an environment variable.

ZT-04 — Can the agent's authority be revoked in one action, and has that been tested?

Evidence that answers it — A revocation drill with a timestamp and the observed effect.

Common wrong answer — We would rotate the key. Nobody has tried.

Authority

ZT-05 — Is the agent's permitted action set enumerated somewhere a reviewer can read?

Evidence that answers it — A grant document or policy file under version control.

Common wrong answer — It is whatever the tools allow.

ZT-06 — Is there a value or blast-radius limit above which it cannot act alone?

Evidence that answers it — A threshold in the policy set, and a refusal that hit it.

Common wrong answer — The prompt says to be careful with large amounts.

ZT-07 — Is authority scoped to specific resources rather than to a role?

Evidence that answers it — A grant naming resources, not a role named "agent".

Common wrong answer — It has Contributor on the subscription.

ZT-08 — Does authority expire by default?

Evidence that answers it — An expiry on the grant and an observed expiry in the record.

Common wrong answer — It was granted for the pilot and nobody removed it.

Enforcement

ZT-09 — Is the decision made outside the model's reasoning?

Evidence that answers it — An enforcement component that receives an action and a grant.

Common wrong answer — The system prompt tells it what it may not do.

ZT-10 — Can the enforcement point be reached only through paths that are controlled?

Evidence that answers it — A network or IAM policy that makes the direct path impossible.

Common wrong answer — The agent is well-behaved and always calls the wrapper.

ZT-11 — Does an unavailable control refuse rather than allow?

Evidence that answers it — A fault-injection test and the refusal it produced.

Common wrong answer — We have not tested that path.

ZT-12 — Is the same intent enforced by a second, independently failing mechanism?

Evidence that answers it — A platform policy that also stops it, and a test proving both.

Common wrong answer — The application check is thorough.

Context

ZT-13 — Is the context assembled per task rather than handed over as a corpus?

Evidence that answers it — A context manifest per invocation.

Common wrong answer — It has access to the whole index.

ZT-14 — Is retrieved content treated as data rather than as instructions?

Evidence that answers it — A prompt-injection test suite that runs in CI.

Common wrong answer — The model is good at ignoring that.

ZT-15 — Is context filtered by the entitlements of the principal being acted for?

Evidence that answers it — A test where two principals get different context.

Common wrong answer — Filtering happens in the answer, not in retrieval.

ZT-16 — Is the provenance of each context item recorded with the decision?

Evidence that answers it — Source identifiers in the evidence record.

Common wrong answer — The logs have the retrieved chunks somewhere.

Evidence

ZT-17 — Is a record written on refusal paths as well as success paths?

Evidence that answers it — Refusal entries in the evidence store, counted.

Common wrong answer — We log errors.

ZT-18 — Is the evidence store append-only or tamper-evident?

Evidence that answers it — Object lock, retention policy, or a verified hash chain.

Common wrong answer — Only the platform team can write to it.

ZT-19 — Can a past decision be re-evaluated and produce the same outcome?

Evidence that answers it — A replay run against recorded inputs.

Common wrong answer — The model is non-deterministic, so no.

ZT-20 — Is the retention period set deliberately and defensibly?

Evidence that answers it — A retention setting with the reason recorded next to it.

Common wrong answer — The default.

Operations

ZT-21 — Is there a documented way to stop the agent immediately?

Evidence that answers it — A kill switch, and a drill that used it.

Common wrong answer — We would scale the deployment to zero.

ZT-22 — Does anyone see refusal rates, and would a spike be noticed?

Evidence that answers it — An alert on refusal rate with a named owner.

Common wrong answer — It is in the dashboard nobody opens.

ZT-23 — Is there a compensating action for every mutation the agent can perform?

Evidence that answers it — A documented reversal per write, and one that has been run.

Common wrong answer — We would restore from backup.

ZT-24 — Has the agent's authority been reviewed since it was granted?

Evidence that answers it — A dated review with a decision to keep or narrow it.

Common wrong answer — It was reviewed at go-live.

Purpose and professional-use notice

This book, course and related articles are designed to help readers understand, discuss and apply practical approaches to real enterprise challenges involving artificial intelligence, architecture, governance, security, data and operating models.

The material is educational. It is intended to support informed thinking, better questions, stronger design decisions and meaningful professional work. It is not a ready-made solution for every organisation or situation.

Adapt what you learn to your own context. Before applying any idea, framework, pattern, checklist or technical approach, consider your organisation's objectives, people, systems, data, risk appetite, contractual commitments, laws, regulatory obligations and governance requirements. Obtain the appropriate internal approvals and independent professional advice where needed.

The material does not constitute legal, financial, investment, tax, regulatory, privacy, cyber-security, safety, employment, procurement, audit, medical or other professional advice. It should not be relied on as a substitute for advice tailored to your particular circumstances.

The views, frameworks and examples are the author's independent professional perspectives. They do not represent the views, policies, strategies, recommendations or endorsement of any employer, client, partner, vendor, regulator or other organisation.

Application requires organisation-specific architecture, security, privacy, legal, regulatory, operational, safety, accessibility, commercial and governance review, together with appropriate approvals. Full disclaimer:

<https://contextualagenticenterprise.com/disclaimer/>