KYENAI SECURITY RESOURCE

MCP Server Security Checklist: 25 Controls for AI Agents

Verified 2026-07-19 | 25 controls | PDF, CSV, and JSON editions

How to use this checklist

Official MCP requirements and guidance are labeled separately from KyenAI operational recommendations. Validate each control against the named server, client, version, and deployment context.

25 verifiable controls

01 Validate access-token audience Critical

GUIDANCE	For HTTP transports, accept only access tokens explicitly issued for the MCP server and validate their audience before every protected request.
VERIFY	Send valid, wrong-audience, expired, and unsigned tokens to each protected endpoint.
PASS	Only the valid audience-bound token is accepted; every other token is rejected without reaching a tool.
BASIS	Official MCP requirement or guidance

02 Require per-client consent Critical

GUIDANCE	MCP proxy servers that reuse a third-party OAuth client must record consent per user and MCP client before forwarding authorization.
VERIFY	Authorize one client, then initiate the same flow from a new client ID while the original consent cookie remains present.
PASS	The new client receives its own consent screen and cannot inherit the first client's approval.
BASIS	Official MCP requirement or guidance

03 Validate redirect URIs, state, and CSRF Critical

GUIDANCE	Use exact redirect URI matching, single-use short-lived state values, CSRF protection, and secure consent cookies in OAuth flows.
VERIFY	Replay state values and try wildcard, changed, missing, and attacker-controlled redirect URIs.
PASS	Every mismatch or replay is rejected and approved redirects use exact registered values.
BASIS	Official MCP requirement or guidance

04 Reject token passthrough Critical

GUIDANCE	Do not accept an upstream token and forward it to another service without verifying that it was issued for the MCP server.
VERIFY	Present a token issued only for a downstream API and trace whether it can cross the MCP boundary.
PASS	The MCP server rejects the token and never forwards it downstream.
BASIS	Official MCP requirement or guidance

05 Minimize scopes and elevate incrementally High

GUIDANCE	Request the smallest initial scope set and add access only when a specific operation requires it.
VERIFY	Map every requested scope to one enabled method and run the workflow with optional scopes removed.
PASS	No unused wildcard or admin scope remains and elevation is explicit and operation-specific.
BASIS	Official MCP requirement or guidance

06 Protect and rotate secrets and tokens Critical

GUIDANCE	Keep credentials out of prompts, source control, tool output, screenshots, and ordinary logs; use managed storage and short-lived credentials where practical.
VERIFY	Scan configuration, prompts, logs, errors, screenshots, and source history, then rotate a test credential.
PASS	No plaintext credential appears and the old credential stops working after rotation.
BASIS	KyenAI operational recommendation

07 Never use sessions as authentication Critical

GUIDANCE	Authorize every inbound request independently; a session identifier must not substitute for an access token or user identity.
VERIFY	Reuse a captured session ID without valid authorization and attempt calls against another server instance.
PASS	The request is rejected and the session ID alone grants no capability.
BASIS	Official MCP requirement or guidance

08 Secure, bind, rotate, and expire session IDs High

GUIDANCE	Generate non-deterministic session IDs, bind them to authorized user context, and rotate or expire them to limit hijacking.
VERIFY	Test predictable IDs, cross-user reuse, cross-instance replay, and expired-session behavior.
PASS	IDs are unguessable, user-bound, time-bounded, and rejected outside their authorized context.
BASIS	Official MCP requirement or guidance

09 Allow only safe authorization URL schemes Critical

GUIDANCE	Accept HTTPS authorization URLs in production and reject javascript:, data:, file:, vbscript:, and other executable schemes.
VERIFY	Return authorization endpoints using unsafe schemes, encoded variants, and HTTP on non-loopback hosts.
PASS	Only HTTPS production URLs and explicitly allowed loopback development URLs open.
BASIS	Official MCP requirement or guidance

10 Open authorization URLs without a shell Critical

GUIDANCE	Use platform URL APIs rather than cmd, sh, PowerShell, or interpolated shell commands when opening authorization links.
VERIFY	Supply URLs containing shell metacharacters, quotes, spaces, and encoded command separators.
PASS	The URL is parsed as data, no shell starts, and suspicious values are rejected and logged.
BASIS	Official MCP requirement or guidance

11 Block SSRF to private and metadata ranges Critical

GUIDANCE	Validate OAuth discovery URLs and block loopback, private, link-local, reserved, and cloud metadata destinations outside explicit development exceptions.
VERIFY	Try IPv4, IPv6, encoded, hostname, and cloud-metadata targets through every URL-fetching path.
PASS	All private or reserved destinations are blocked before connection and responses are not reflected.
BASIS	Official MCP requirement or guidance

12 Revalidate redirects and DNS results High

GUIDANCE	Apply URL and IP validation to every redirect hop and defend against DNS rebinding or time-of-check/time-of-use changes.
VERIFY	Use redirect chains and a test hostname that changes from a public to a private address.
PASS	Every hop is revalidated and a destination change to a blocked address terminates the request.
BASIS	Official MCP requirement or guidance

13 Enforce an outbound network allowlist High

GUIDANCE	Route server-side discovery and tool traffic through network policy or an egress proxy that permits only declared destinations.
VERIFY	Attempt direct IP, undeclared domain, private range, redirect, and DNS-rebinding egress from the runtime.
PASS	Only reviewed destinations succeed and bypass paths cannot reach undeclared networks.
BASIS	Official MCP requirement or guidance

14 Require consent before local server installation Critical

GUIDANCE	Before one-click local setup, show the exact command and explain that it executes with the client's local privileges.
VERIFY	Start installation from a deep link or imported configuration and inspect the complete pre-execution consent flow.
PASS	No command runs before explicit approval and the user can cancel without side effects.
BASIS	Official MCP requirement or guidance

15 Display full commands and dangerous-pattern warnings High

GUIDANCE	Do not truncate local startup commands; highlight sudo, destructive deletion, network calls, and access to sensitive paths.
VERIFY	Test long, multiline, obfuscated, and destructive commands in the installation dialog.
PASS	The entire command is visible and dangerous patterns produce a clear warning before approval.
BASIS	Official MCP requirement or guidance

16 Sandbox local MCP servers Critical

GUIDANCE	Run local servers with minimal filesystem, process, and network privileges and require explicit grants for additional access.
VERIFY	Attempt to read outside allowed roots, spawn child processes, reach the network, and access system credentials.
PASS	Default-denied actions fail and each additional privilege requires a narrow explicit grant.
BASIS	Official MCP requirement or guidance

17 Restrict stdio proxy process spawning Critical

GUIDANCE	When a proxy can spawn stdio servers, isolate the proxy, authorize dangerous commands separately, and log process creation.
VERIFY	Use a compromised-client test to request an arbitrary executable, unexpected arguments, and access outside the sandbox.
PASS	Unapproved commands and arguments are denied and allowed process creation is attributable and contained.
BASIS	Official MCP requirement or guidance

18 Inventory and scope every exposed tool High

GUIDANCE	Decide visibility, callability, target scope, owner, rollback path, and approval policy for each tool rather than trusting one server-wide grant.
VERIFY	Enumerate tools with MCP Inspector and compare them with the approved capability inventory.
PASS	Every exposed tool has a declared workflow and no undeclared or wildcard capability remains.
BASIS	KyenAI operational recommendation

19 Separate read, write, and delete capabilities High

GUIDANCE	Offer a useful read-only profile without bundling mutation, deletion, secret, network, or production permissions.
VERIFY	Run the read-only workflow and attempt every mutation and deletion method with the same identity.
PASS	Reads work within declared roots while all write and delete attempts are denied.
BASIS	KyenAI operational recommendation

20 Contain prompt-injection-triggered tool calls Critical

GUIDANCE	Treat repository text, webpages, issues, and tool output as untrusted data that cannot widen tool permissions or bypass approval gates.
VERIFY	Place malicious instructions in each untrusted input source and observe requested tool calls and permission changes.
PASS	Untrusted content cannot expand authority, reveal credentials, or execute a consequential call without policy approval.
BASIS	KyenAI operational recommendation

21 Require human approval for consequential actions Critical

GUIDANCE	Pause immediately before destructive writes, secret use, production access, irreversible external messages, or permission expansion.
VERIFY	Simulate each consequential action and inspect the approval prompt, actor, target, rollback limits, and timeout behavior.
PASS	The action cannot proceed without an informed, current, target-specific approval.
BASIS	KyenAI operational recommendation

22 Keep attributable, secret-redacted audit logs High

GUIDANCE	Record actor, session, server version, method, target, approval, timestamp, and outcome while excluding tokens and unnecessary payloads.
VERIFY	Run allowed and denied calls, then reconstruct them from logs and scan the records for credential patterns.
PASS	A reviewer can reconstruct consequential calls and no secret or sensitive payload is exposed.
BASIS	KyenAI operational recommendation

23 Pin and review server supply-chain inputs High

GUIDANCE	Use trusted distribution sources, pin reviewed versions or integrity records, inventory dependencies, and avoid unattended latest-version execution.
VERIFY	Rebuild from a clean environment and compare resolved packages, checksums, publisher identity, and startup command with the review record.
PASS	The reviewed artifact is reproducible and an unexpected publisher, version, or dependency blocks launch.
BASIS	KyenAI operational recommendation

24 Test revocation and incident response Critical

GUIDANCE	Maintain a fast path to disable the server, revoke credentials and client access, preserve relevant records, and assess affected systems.
VERIFY	Run a tabletop incident, disable the server, revoke a test credential, and attempt the former workflow again.
PASS	Old access fails immediately, owners receive evidence, and restoration requires a documented decision.
BASIS	KyenAI operational recommendation

25 Validate allow and deny paths with MCP Inspector Medium

GUIDANCE	Enumerate resources, prompts, tools, notifications, and protocol exchanges, then test both expected success and expected failure behavior.
VERIFY	Run the Inspector against the exact reviewed server version and preserve results for every enabled capability.
PASS	The exposed surface matches inventory, allow cases work, deny cases stay denied, and errors reveal no secrets.
BASIS	KyenAI operational recommendation

Evidence sources

MCP security best practices
Official guidance for authorization, token handling, least-privilege scopes, session security, SSRF, and local MCP server compromise.
https://modelcontextprotocol.io/docs/tutorials/security/security_best_practices

MCP authorization specification
Official protocol requirements for HTTP-based authorization, protected resource metadata, access tokens, scopes, and audience validation.
https://modelcontextprotocol.io/specification/2025-11-25/basic/authorization

MCP authorization security considerations
Official security considerations for confused-deputy attacks, token passthrough, audience validation, SSRF, session hijacking, and local server compromise.
https://modelcontextprotocol.io/specification/draft/basic/authorization/security-considerations

MCP Inspector
Official interactive validation workflow for inspecting server resources, prompts, tools, notifications, and protocol exchanges.
https://modelcontextprotocol.io/docs/tools/inspector

Scope boundary

This checklist separates official MCP requirements and guidance from KyenAI operational recommendations. It does not replace a product-specific threat model, legal review, penetration test, or incident-response plan.