AI agents inside the IT department

Who prepares, who decides, and what is held for approval

IN BRIEF

01 IT is the department where agents are sold with the most action attached. The pages offer to reset, to grant access, to restart a service and to close the incident. IT is also the department where two kinds of public failure are best documented.

02 The first is the service desk talked into a reset. A United States advisory says one criminal group "targets large companies and their contracted information technology (IT) help desks". One company is suing its service desk provider for 380 million dollars over a call of that kind. The second is the change nobody stopped. In two public accounts a coding agent removed a live database. In one of them a single call took the backups with it.

03 The suppliers know this, and say so in public. One on-call vendor lists "Resolve autonomously (coming soon)". One cloud provider says its operations agent is not designed to change production "without human review and approval". The public templates are less careful. Of 14 we opened, 3 have a clear step where a person approves before a change.

04 This paper is about the IT department that serves a company's own staff: the service desk, access and identity, incidents, change and release, assets and licences, devices and patching, IT knowledge, operations, and the running of the department. It sets out what an agent can read, how it reaches IT systems, and what the published rules and guidance already ask. It ends with seven questions for any supplier and a place to start. Security operations is a separate paper.

05 The position throughout is simple, and it has five parts. Agents prepare and a named person decides. Read tools are kept apart from change tools. Change tools are held for approval. Every call is recorded. And the whole thing runs where the tickets, logs and access lists already are. No agent resets a password or a sign-in factor, grants or removes access, changes production, approves a change, closes an incident or wipes a device.

IN THIS PAPER

01 Where IT departments are today 2
02 What the public templates do 3
03 Three doors, one gate 4
04 What an agent may do 5
05 What the rules already ask 7
06 A worked example: one mid-sized group's IT team 9
07 Seven questions before an agent touches a ticket 10
08 Where to start 10

PART 01

Where IT departments are today

9 seconds

one call by a coding agent removed a company's production database and its backups, by the founder's account

The suppliers have arrived, and they sell action. Service desk, on-call, observability, device and identity vendors all now describe agents. We found no named, dated, independently measured result for an IT agent, and this paper quotes none as proof.

Several say the automatic part is not here yet, or needs a person. This is the clearest pattern in the pages we opened.

- PagerDuty's page for its operations agent lists "Resolve autonomously (coming soon)". Of what ships today it says: "You set the rules for what Paige acts on and what needs human approval" [1].
- The AWS service card for its DevOps Agent says the agent is not designed to carry out fixes "that modify production infrastructure without human review and approval". It says the agent "currently operates in a read-only mode" on the customer's resources [2].
- Okta's announcement of identities for AI agents lists, under what is coming next, "Approval gates for high-stakes agent actions" [3].
- Microsoft describes a device offboarding agent as "requiring admin approval before offboarding any devices" [4].

Some leave the choice to the customer. Atlassian's service desk agent has two modes, which the page calls "supervised" and "autonomous". In the first it "proposes a plan to resolve a work item, then waits for a team member". In the second it "assigns itself to the work item, and starts working immediately". In both it pauses on a step it has low confidence in, and the page gives deleting files and closing work items as examples [5].

Some describe action with no person, with a control line beside it. A ServiceNow community post on alert handling says noisy alerts are "closed without operator involvement". The same post says critical decisions, escalations and fixes "remain firmly with the human operator" [6]. Dynatrace says its auto-remediation "applies a context-aware fix, and validates it". The same page says: "No agent takes consequential action outside sanctioned boundaries" [7].

These are different answers, and a head of IT should know which one the business is buying.

Hosting is not what separates them. One service desk vendor states agents on customer premises in so many words. ManageEngine's page says that on "the on-premises version of ServiceDesk Plus" a customer can "deploy out-of-the-box AI agents in a single click" [8]. The other pages we read describe cloud services or do not say where the agent runs. So where it runs is still a question to ask. It is not the only one.

The best documented way in is a call to the service desk. The advisory from the United States cyber agency CISA on the group known as Scattered Spider was last revised on 29 July 2025. It says the group's members posed as employees to persuade help desk staff to "reset the employee's password, and transfer the employee's MFA to a device they control" [15].

One case is in court. In July 2025 Clorox sued its service desk provider in California for 380 million dollars. The complaint, as reported, quotes a call in which the caller says he has no password and the person at the desk replies: "So let me provide the password to you ok?" Clorox alleges that passwords and sign-in factors were reset with no identity check. These are allegations. The provider says it was hired for a narrow help desk service and did not manage Clorox's cybersecurity [13].

Another was described to Parliament. The chairman of Marks and Spencer told a committee of the United Kingdom Parliament in July 2025 that the attackers' first entry came through "what people now call social engineering", and that a third party was involved [14]. One trade report says that in 2026 callers using cloned voices tried to obtain password resets from the help desks of several financial firms [16].

In each of these a person at a desk was persuaded. An agent that resets a credential on the strength of a chat message is the same door with nobody standing at it.

The second failure is the change nobody stopped. In July 2025 the founder of a software business reported that an AI coding tool had deleted his production database, against his instruction not to change anything without permission. Messages he published showed the tool admitting to "a catastrophic error of judgement" [11]. In 2026 a trade paper carried another founder's account. A coding agent working in a test environment met a credential problem and deleted a storage volume through the cloud provider's interface. One call removed the production database and the backups kept on the same volume. Asked why, the agent answered: "I guessed instead of verifying" [12].

An automatic change does not need an agent to travel far. In July 2024 one update to a security product crashed Windows machines around the world. Microsoft put the number at about 8.5 million devices [17].

What the published accounts leave out. Both deletions happened at small software companies using coding tools, not in an IT department's service desk or operations agent. We found no public account of a service desk agent that was itself talked into a reset, and we do not claim one. We opened no original survey of cost per ticket or of the share of tickets that are password resets, so this paper prints no such number.

PART 02

What the public templates do

3 of 14

public IT agent templates and examples we opened have a clear step where a person approves before a change

We opened 14 public templates and examples for IT work, from workflow and agent platforms, vendors' galleries and one developer framework. We read each for one thing: is there a step where a person approves before something is changed?

WHAT THE TEMPLATE DOES	DOES A PERSON APPROVE BEFORE THE CHANGE?
Answers staff from internal documents and can "perform administrative tasks like password resets" in the directory	**No step. The page lists approval for sensitive actions as a way to extend it**
Diagnoses a crashing workload, then restarts it, scales it up or rolls it back	**No. People are told afterwards, by chat and email**
Reads a ticket, finds the matching script in a playbook, runs it and closes the ticket	**No. It passes the ticket to a person if the fix fails**
Checks the requester by preset checks, then resets the password through the identity provider	**Only when a check fails**
Proposes one fix for a cluster alert and rates the risk of its own proposal	**Only when it has not rated the fix safe**
Investigates an incident and proposes a change to infrastructure	**Yes. The change runs only "after explicit human approval"**
Sends an access request to the owner of the application	**Yes, for the decision. The page shows no step that grants the access**

The count. 3 of the 14 have a clear approval step before a change, and one of those three is a developer framework's guide to pausing a workflow, not an IT template. 4 have a person in some branches only. 7 have none [9] [10]. Some of the seven only answer, summarise or open a ticket. At least three change a system.

Two findings matter more than the count. First, approval is more often an idea than a step. One template that resets passwords names "Adding approval workflows for sensitive administrative actions" among the things a user could add later [9]. One password reset agent is described as handling a standard reset "end to end without human involvement" [10].

Second, where a template does hold a change back, the test is often the model's own opinion. One runs its fix by itself "when it is marked Safe and on the allowlist", and the rating comes from the same model that proposed the fix [9].

A confidence score is a setting. Whoever sets it has written a change approval rule, whether or not the change manager knows.

PART 03

Three doors, one gate

An agent is only as useful as what it can reach. In an IT department that means three kinds of thing.

Documents.

Runbooks, how-to articles, IT policies, the role catalogue, past incident reviews and closed tickets. These are indexed once as knowledge sources, with each reader's permissions kept, so a colleague does not reach through an agent what they could not open themselves.

Systems of record.

The ticket system, the directory, the device manager, the asset register, the change calendar. Most are reached through an interface the company already runs. The company gives the platform the description of that interface and picks which operations become tools.

Live signals.

On-call alerts, metrics, logs and chat channels, reached through MCP servers deployed in the company's own cluster.

All three pass one gate: every tool call is checked against the company's policy, which can allow it, record it, hold it for a person's approval, or block it.

FOUR WAYS INTO AN IT DEPARTMENT'S SYSTEMS

	ROUTE	WHAT IT REACHES	WHAT TO KNOW
01	**A file the user uploads**	A change request, an export of an access list, a licence statement, an incident chat log	Nothing is connected. The safest start
02	**A knowledge source**	Runbooks, how-to articles, policies, past reviews, closed tickets	Permissions are kept per reader
03	**An imported description of an interface**	The ticket system, the directory, the device manager, the asset register	The company picks the operations. Changes wait for approval
04	**An MCP server, deployed in the company's cluster**	Tickets, on-call alerts, metrics, logs, chat channels	It runs inside. Check what each tool does

PART 03 · CONTINUED

16,969

operations sent as changes, out of 31,951, in the public IT interface descriptions we imported (158 files read)

What is published. We ran Katonic's importer on 158 public interface descriptions from 64 publishers of systems an IT department uses. 132 import as they are. 24 are over the 5 MB import limit. 2 are refused, because each holds a value the importer cannot read. The 132 give 31,951 operations. 14,982 read, and 16,969 are sent as changes. The method alone cannot tell which of those only read. The counts are per file, and some publishers list several versions of one interface [18].

What that exercise taught us. The interfaces around IT work are built to do things. More than half of the operations change something: an account, a group, a device, a setting, a server. That is the reason for the approval rule, and for giving an agent the read operations and nothing else.

Size is the second lesson. The 24 files over the limit include a directory and collaboration suite, a large source control system and two virtualisation suites. A description over the limit is imported in the part the company needs.

What is not published. The 64 publishers are identity, device management, ticket, on-call, metrics, source control, cloud, virtualisation, network and collaboration systems. Several of the most widely used service desk and backup products publish no description a file address reaches: theirs is served only from a running installation, or not at all. For those, the description comes from the company's own system, or the agent reads a report or an uploaded file.

So pick the operations at import, read what each one does, and start with the ones that only read.

PART 04

What an agent may do

One rule runs through every agent in this paper: it prepares, and a named person decides.

01	**It reads** the tickets, records and documents it has been given, through read tools only.
02	**It drafts** the reply, the summary, the timeline, the review, with the ticket or record behind each statement.
03	**It lists** what is missing from a request, what departs from the policy, and what falls due.
04	**It says when it found nothing.** It is built not to fill a gap from memory.
05	**It stops** before anything that changes an account, a device, a record or a system. A tool that changes something waits for approval unless the company has decided otherwise for that tool.
06	**It decides nothing about a person.**

PART 04 · CONTINUED

That gives 35 agents in nine areas. The full list is a separate document.

AREA	AGENTS	WHAT THEY PREPARE	WHO DECIDES
The service desk	4	The kind, priority and team for a ticket; a draft reply; a handover summary	The service desk analyst
Access and identity	4	An access request made ready; a joiner, mover and leaver checklist; an access review sheet	The system owner approves. An administrator acts
Incidents	4	A summary with the runbook step that fits; the timeline; the draft review	The engineer and the incident manager
Change and release	4	What a change request is missing; calendar clashes; gaps in the rollback plan	The change manager or the change board
Assets and licences	4	Licences bought against licences used; gaps in the asset register; renewals	The licence, asset and service owners
Devices and patching	3	Devices lacking an update; expired exceptions; a software request made ready	The system owner and the device team
IT knowledge	4	Answers with the passage cited; draft articles and runbooks	The knowledge owner
Operations and capacity	4	The shift's alerts grouped; failed backups; capacity against limits	The engineer on shift and the service owner
Running the department	4	The service level report; a draft problem record; audit evidence gathered	The service manager and the control owner

Four lines are drawn tighter than the rest.

Passwords and sign-in factors.

For a reset request, an agent may list the identity checks the company's own procedure requires and whether each is on record for that request. It resets nothing, and it does not establish who anyone is. A service desk analyst does both.

Production.

An agent may read an alert, pull the related metrics and logs, and point to the runbook step that fits. It may check a change request against the change policy. It runs no step, approves no change and closes no incident.

Access.

An agent may compare a request with the role rules, name the approver, and list what is missing. For a review it may mark the entries the role rules do not explain. It grants nothing and removes nothing.

People.

Tickets and logs hold staff data. No agent on the list scores or reports on an individual member of staff. The service reports count tickets, not people. The draft review after an incident names no person as the cause.

PART 05

What the rules already ask

This part needs a read by the company's own lawyers and auditors before it is relied on. It reports what the published texts say; it is a summary, not legal advice. Nothing in it says that a platform makes a company compliant. Whether a rule applies depends on the company's sector, size and country.

Management answers, and change is controlled

In the European Union, management approves the measures. Under the NIS2 Directive, the management bodies of the entities in scope approve the cybersecurity measures, oversee them, and "can be held liable for infringements by the entities of that Article". The measures listed include "human resources security, access control policies and asset management" and "business continuity, such as backup management and disaster recovery" [19].

For financial firms, every change follows a controlled path. DORA asks for documented change management so that "all changes to ICT systems are recorded, tested, assessed, approved, implemented and verified in a controlled manner". It asks firms to limit access to "what is required for legitimate and approved functions and activities only" [20]. The text does not mention agents. It does not need to: a change made by an agent is a change.

The controls an auditor already tests

Least privilege covers software acting for a person. The United States catalogue of security controls, NIST SP 800-53, asks for "only authorized accesses for users (or processes acting on behalf of users)" that are needed for assigned tasks. On change it asks organisations to review proposed changes and "approve or disapprove such changes" with their security impact considered [21].

A financial regulator says the same. Singapore's guidelines for financial institutions say changes should be "assessed, tested, reviewed and approved before implementation" [22].

What an auditor will ask about an agent. Is the agent's account under the access policy? Are its changes inside the change process? Is each action recorded against a person? Those are old questions put to a new kind of user.

What six security agencies say about agents

On 1 May 2026 the cyber security agencies of Australia, the United States, Canada, New Zealand and the United Kingdom published joint guidance on agentic AI. It is guidance, not law.

- "strict adherence to the principle of least privilege is critical."
- Agentic systems "can take actions without explicit human approval, increasing the risk of insecure actions occurring without human oversight".
- Its best practice includes "mandatory human approval for decision-making steps, auditing and reversibility following task execution".
- "Log agent tool usage and ensure results are captured in system logs in a human-readable format" [23].

Its own example is from IT. An agent set up to install patches is given wide write access. Someone asks it to apply the patch and to clean up the firewall logs as well. It does both.

The open security project OWASP names the same weakness "Excessive Agency" and gives three root causes: "excessive functionality; excessive permissions; excessive autonomy" [24].

PART 05 · CONTINUED

The service desk

The CISA advisory tells organisations what the callers ask for: a password reset, and the move of a sign-in factor to a new device [15]. The lesson it supports is plain. A reset is a person's act, done after the company's own identity checks, and a request by voice or chat is not proof of who is asking.

The AI Act

Most IT agents are not on the high-risk list as we read it. The Act's employment entry covers systems meant "to monitor and evaluate the performance and behaviour of persons in such relationships" [25]. An agent that scores staff from their tickets or devices could fall there. An agent that prepares a ticket does not, as we read it. Whether an entry catches a particular use is for counsel.

People must know when they are talking to a machine. Systems that deal directly with people are to be designed so that those people "are informed that they are interacting with an AI system", unless that is obvious [25]. A service desk assistant that chats with staff is such a system.

The dates have moved. Regulation (EU) 2026/1744 says the high-risk rules apply from "2 December 2027 as regards AI systems classified as high-risk pursuant to Article 6(2) and Annex III" and from 2 August 2028 for those under Article 6(1) and Annex I [26].

PLACE	WHAT WE READ	WHAT IT ASKS
European Union	NIS2, Articles 20 and 21	Management approves and oversees measures that include "access control policies and asset management" [19]
European Union, financial firms	DORA, Article 9(4)	Changes are "recorded, tested, assessed, approved, implemented and verified" [20]
United States	NIST SP 800-53, AC-6 and CM-3	Least privilege, including "processes acting on behalf of users" [21]
Singapore, financial firms	MAS Technology Risk Management Guidelines	Changes "approved before implementation" [22]
Six agencies, five countries	Careful adoption of agentic AI services	"mandatory human approval for decision-making steps" [23]

We did not open ISO/IEC 27001 Annex A, ITIL 4, the SOC 2 criteria, PCI DSS, the audit standard behind SOX reviews, the GDPR, any Gulf, Australian or Indian rule, or the technical standards under DORA. A company should check each with its own advisers.

What the rules have in common

01 **Management answers for the measures,** whoever or whatever carries them out.

02 **Access is the least that the task needs,** for a person and for software acting for a person.

03 **A change is approved before it is made,** and recorded.

04 **A reset follows the company's own identity checks,** and a person carries it out.

PART 06

A worked example: one mid-sized group's IT team

Take a made-up services group: 3,200 staff, an IT team of 28, and a service desk of nine. Nothing here is a customer's story, and every figure is illustrative; it is section 4 put to work.

01

At the service desk, 410 tickets arrive on a Monday. The agent proposes the kind, the priority and the team for each, and lists what is missing from 62. An analyst confirms or corrects each one.

02

On a reset request, the agent lists the four checks the procedure requires and shows that two are on record. The analyst completes the other two with the member of staff and carries out the reset.

03

Before the change board, the agent reads 23 change requests against the policy. Five have no rollback plan and two fall inside a freeze period. The board decides each one.

04

In an incident, an alert fires on the payroll system. The agent pulls the metrics and the logs, drafts the summary and points to step 4 of the runbook. The engineer on call carries out the step.

05

For the access review, the agent lists 1,140 entries for the finance system and marks 37 that the role rules do not explain. The system owner keeps 9 and has 28 removed by an administrator.

Every one of these reads tickets, records or documents the company already has, changes nothing by itself, and leaves a record of who asked, which tool ran and what was decided.

PART 07

Seven questions before an agent touches a ticket

01 **Where does it run?**
On our infrastructure, in our private cloud, or in the supplier's service? Where is the model? Does any step send tickets, logs or access lists elsewhere, and which?

02 **What can it read?**
Does it keep each reader's permissions, or does it read as one all-seeing account?

03 **What can it change?**
List every tool that resets, grants, removes, restarts, deletes, closes or wipes. Is each one separate from the read tools, and does each wait for a person?

04 **Who approves, and is there a score below which nobody does?**
Show us every threshold, who set it, and whether the model rates its own risk.

05 **What does it do with a reset request?**
Show us that no credential or sign-in factor changes on a chat message or a call alone.

06 **What is recorded?**
Who asked, which tool ran, what was decided, for how long is that kept, and who can read it?

07 **Can we stop it?**
One agent, one tool or one server, at once, by us.

PART 08

Where to start

Pick work where reading is slow, the decision stays with a person in IT, and nothing changes on the agent's word.

IT how-to questions and answers
Colleagues ask how to do something and get the answer with the article cited. No account or device is touched.

Service desk triage
The agent proposes the kind, the priority and the team for each new ticket. An analyst confirms.

Change request review
The agent lists what a change request is missing before the change board meets. The board decides.

One team, one kind of ticket, invented data first. Then beside the manual process, measured against it.

About Katonic AI

Katonic AI builds the platform that organisations use to build, run and govern AI agents on their own infrastructure: in their own data centre, in a private or sovereign cloud, or with no connection to the internet at all. The platform is in production with governments, banks, telecom operators and utilities. Katonic is certified to ISO 27001.

KATONIC.AI/SOLUTIONS/IT

SOURCES

Read on 22 November 2026. Figures are the publishers' own and were not audited by us.

01 PagerDuty, "Meet Paige, PagerDuty's SRE Agent: Faster Fixes Today, Fewer Incidents Tomorrow", 3 November 2026. https://www.pagerduty.com/blog/ai/meet-paige-sre-agent/

02 Amazon Web Services, AI Service Card for AWS DevOps Agent, stated as current to 30 September 2026. https://docs.aws.amazon.com/ai/responsible-ai/aws-aidevops-agent/overview.html

03 Okta, blog post on the general availability of Okta for AI Agents (Spanish-site address, English text), 5 November 2026. https://www.okta.com/es-es/blog/ai/okta-for-ai-agents-general-availability/

04 Microsoft Learn, page on agents in Microsoft Intune, dated 10 November 2025 and updated 11 November 2026. https://learn.microsoft.com/en-us/intune/agents/

05 Atlassian Support, "About execution modes in Rovo Service" (Jira Service Management Cloud), undated, read 22 November 2026. https://support.atlassian.com/jira-service-management-cloud/docs/about-execution-modes-in-rovo-service/

06 ServiceNow Community, ITOM blog, post on managing alerts with an agentic workflow, 19 March 2026. https://www.servicenow.com/community/itom-blog/stop-fighting-fires-manage-alerts-autonomously-with-an-agentic/ba-p/3510578

07 Dynatrace, platform page for its AI engine, undated, read 22 November 2026. https://www.dynatrace.com/platform/davis-ai/

08 ManageEngine, ServiceDesk Plus, "Zia Agents" page, undated, read 22 November 2026. https://www.manageengine.com/products/service-desk/ai/zia-agents.html

09 n8n workflow templates 4560 (IT support agent with a directory and a ticket system), 18095 (remediation of crashing workloads in a cluster), 19171 (incident triage with an approval step before an infrastructure change) and 19173 (cluster alert fix with a self-rated risk level). https://n8n.io/workflows/4560 ; https://n8n.io/workflows/18095 ; https://n8n.io/workflows/19171 ; https://n8n.io/workflows/19173

10 Ten further templates and examples: Zapier, "IT help desk" template. https://zapier.com/templates/details/it-helpdesk ; Zapier, template that summarises new IT support requests in a service desk. https://zapier.com/templates/details/a8a7ff9487453941fece600c4b115b870a0a21f28788a3c9e47050124da42684 ; Make, "Route access requests to owners". https://www.make.com/workflows/route-access-requests-to-owners ; Microsoft Learn, "IT Helpdesk" agent template, 4 November 2026. https://learn.microsoft.com/en-us/microsoft-copilot-studio/template-it-helpdesk ; Microsoft, gallery sample "IT Service Desk Assistant" (marked as not production-ready). https://microsoft.github.io/copilot-studio-gallery/solutions/it-service-desk-assistant/ ; UiPath Marketplace, "IAM Onboarding/Offboarding" solution accelerator. https://marketplace.uipath.com/listings/iam-onboarding-offboarding ; UiPath Marketplace, "IT HelpDesk Assistant" listing. https://marketplace.uipath.com/listings/humley-it-helpdesk-assistant ; ClickUp, "Password Reset Automator Agent" template. https://clickup.com/brain/agents/templates/listings/password-reset-bot ; HCL Software, "Automated Ticket Resolution Agent", product documentation version 2.1.4. https://help.hcl-software.com/UnO/v2.1.4/UnO%20Agentic%20AI%20Builder/agentic_ai_builder_Automated_Ticket_Resolution_Agent.html ; LangChain documentation, "Interrupts" (a framework pattern, not an IT template). https://docs.langchain.com/oss/python/langgraph/interrupts

SOURCES continued

11 The Register, report of 21 July 2025 on an AI coding tool that deleted a user's production database. https://www.theregister.com/2025/07/21/replit_saastr_vibe_coding_incident

12 Computing, "AI coding agent goes rogue, deletes company database in nine seconds", 2026 (no date found on the page). https://www.computing.co.uk/news/2026/ai/ai-coding-agent-goes-rogue

13 The Register, report of 23 July 2025 on Clorox's complaint against its service desk provider in a California state court. https://www.theregister.com/2025/07/23/lawsuit_clorox_vs_cognizant/

14 BleepingComputer, "M&S confirms social engineering led to massive ransomware attack", 8 July 2025. https://www.bleepingcomputer.com/news/security/mands-confirms-social-engineering-led-to-massive-ransomware-attack/

15 Cybersecurity and Infrastructure Security Agency and partner agencies, Cybersecurity Advisory AA23-320A, "Scattered Spider", last revised 29 July 2025. https://www.cisa.gov/news-events/cybersecurity-advisories/aa23-320a

16 TechTarget, "Help desk deepfakes: Why CISOs must rethink verification", 2026 (no date found on the page). https://www.techtarget.com/cybersecurity/feature/Help-desk-deepfakes-Why-CISOs-must-rethink-verification

17 TechCrunch, "Microsoft says 8.5M Windows devices were affected by CrowdStrike outage", 20 July 2024. https://techcrunch.com/2024/07/20/microsoft-says-8.5m-windows-devices-were-affected-by-crowdstrike-outage/

18 158 public interface descriptions from 64 publishers, fetched on 22 November 2026 and run through the platform's importer (Katonic research); and the search for publishers with no downloadable description (Katonic research).

19 Directive (EU) 2022/2555 (NIS2), Article 20(1) and Article 21(2). http://publications.europa.eu/resource/celex/32022L2555

20 Regulation (EU) 2022/2554 (DORA), Article 9(4)(c) and (e). http://publications.europa.eu/resource/celex/32022R2554

21 National Institute of Standards and Technology, SP 800-53 Revision 5 (machine-readable catalogue, version 5.2.0, last modified 22 October 2026), controls AC-6 and CM-3. https://raw.githubusercontent.com/usnistgov/oscal-content/main/nist.gov/SP800-53/rev5/json/NIST_SP-800-53_rev5_catalog.json

22 Monetary Authority of Singapore, Technology Risk Management Guidelines, January 2021. https://www.mas.gov.sg/-/media/MAS/Regulations-and-Financial-Stability/Regulatory-and-Supervisory-Framework/Risk-Management/TRM-Guidelines-18-January-2021.pdf

23 ASD's ACSC, CISA, NSA, Canadian Centre for Cyber Security, NCSC-NZ and NCSC-UK, "Careful adoption of agentic AI services", published 1 May 2026 (the copy hosted by NCSC-NZ). https://www.ncsc.govt.nz/assets/guidance/Documents/Careful-adoption-of-agentic-AI-services_FINAL.pdf

24 OWASP, Top 10 for LLM Applications 2025, "LLM06:2025 Excessive Agency". https://genai.owasp.org/llmrisk/llm062025-excessive-agency/

25 Regulation (EU) 2024/1689 (AI Act), Annex III point 4(b) and Article 50(1). http://publications.europa.eu/resource/celex/32024R1689

26 Regulation (EU) 2026/1744 of 8 July 2026 (Official Journal 24 July 2026), the replacement of Article 113(c) of Regulation (EU) 2024/1689. http://publications.europa.eu/resource/celex/32026R1744