Confidential-Data AI: Operating without Human Access to Customer Content

A Verifiable Architecture for Tenant Isolation, Attested Workloads, Evidence-Bound Key Release and Content-Free Operations

Authored by Chennakeshav Adya, Independent Researcher

September 2026

Abstract

High-confidentiality AI products face a demanding design question: can customer content be processed usefully while remaining inaccessible to platform administrators, support teams, vendors and other tenants? Encryption at rest and in transit leaves a material gap when content is decrypted for retrieval or inference. Routine debugging, observability, backups, emergency response and model pipelines can also recreate human access outside the main application. This paper develops a verifiable operating architecture for confidential-data artificial intelligence. Forty modules connect the confidentiality promise, data lifecycle, classification, trust boundaries, tenant isolation, workload identity, removal of interactive access, ephemeral processing, encryption across all states, remote attestation, evidence-bound key release, key governance, minimisation, privacy-enhancing computation, retrieval and vector-store controls, prompt separation, output control, content-free logs, monitoring, break-glass access, support, subprocessors, residency, secrets, software and model provenance, training exclusion, deletion, incident response, recovery, adversarial testing, performance, audit evidence, customer assurance, contracts, change, operating accountability, bounded pilots and completion gates. Five figures, five tables, eight frequently asked questions and twenty-six primary or authoritative references support use-, data-, jurisdiction-, provider- and period-specific review. The framework treats no-human-access as a testable system property with explicit exclusions and exception governance. It does not establish legal or regulatory compliance and does not substitute for authorised privacy, security, legal, risk, audit, technology or professional advice.

JEL Classification: C88, K24, L86, M15, O32, O33

Keywords: confidential AI, confidential computing, data in use, tenant isolation, trusted execution environment, remote attestation, secure key release, zero trust, privacy engineering, AI security

1. Define the confidentiality promise

State which people, operators, administrators, vendors and models must remain unable to view customer content.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is an approved confidentiality charter. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when the phrase no human access is used without a testable technical and operating definition. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

A defensible no-human-access claim is an end-to-end property. Encryption, isolation, attestation, key release, support, logging, recovery and deletion must preserve the same boundary.

The decision pack for define the confidentiality promise should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

2. Map the end-to-end data lifecycle

Trace collection, transport, storage, indexing, inference, caching, output, logging, backup, support and deletion.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is a verified data-flow map. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when plaintext appears in an undocumented intermediate service. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

The architecture should minimise who and what must be trusted. Workload identity and measured execution can replace broad administrator access for many high-confidentiality operations.

The decision pack for map the end-to-end data lifecycle should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

3. Classify data and consequences

Classify personal, financial, health, legal, privileged, trade-secret and regulated content with impact tiers.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is a data and consequence matrix. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when all customer content receives one generic control set. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

Confidential computing protects data in use when the workload and key-release path are correctly designed. It does not remove application, output, side-channel, supply-chain or governance risk.

The decision pack for classify data and consequences should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

4. Set explicit trust boundaries

Identify customer, application, orchestrator, model, vector store, key service, operator and vendor trust domains.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is a trust-boundary model. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when the architecture assumes cloud or platform administrators are trusted by default. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

Assurance should rely on reproducible evidence: signed artefacts, attestation reports, policy decisions, key events, access denials, tests, exceptions, incident records and verified deletion.

The decision pack for set explicit trust boundaries should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

5. Isolate every tenant

Separate identities, networks, storage, indexes, caches, keys, queues, logs and backups by tenant.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is a tenant-isolation control design. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when a shared retrieval index or cache leaks one customer's content to another. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

A defensible no-human-access claim is an end-to-end property. Encryption, isolation, attestation, key release, support, logging, recovery and deletion must preserve the same boundary.

The decision pack for isolate every tenant should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

6. Use workload identity

Authorize measured services, signed workloads and bounded service accounts instead of shared human credentials.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is a workload-identity policy. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when privileged users can impersonate workloads or retrieve decryption material. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

The architecture should minimise who and what must be trusted. Workload identity and measured execution can replace broad administrator access for many high-confidentiality operations.

The decision pack for use workload identity should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

7. Remove interactive human access

Eliminate shells, consoles, database browsing, debug exports and administrator plaintext paths.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is a no-interactive-access design. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when support or engineering tools recreate a hidden human-access channel. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

Confidential computing protects data in use when the workload and key-release path are correctly designed. It does not remove application, output, side-channel, supply-chain or governance risk.

The decision pack for remove interactive human access should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

8. Make plaintext processing ephemeral

Constrain decrypted content to approved memory, time, process and purpose; clear it after execution.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is an ephemeral-processing specification. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when temporary files, swap, crash dumps or caches retain readable content. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

Assurance should rely on reproducible evidence: signed artefacts, attestation reports, policy decisions, key events, access denials, tests, exceptions, incident records and verified deletion.

The decision pack for make plaintext processing ephemeral should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

Table 1. Trust-boundary layers

Layer	Primary control	Required evidence
tenant	identity, namespace and keys	isolation test
workload	measurement and policy	attestation report
data	purpose, state and location	flow and retention record
operator	deny access and constrain exception	access and break-glass log

Illustrative controls require use-, consequence-, jurisdiction-, firm- and period-specific approval.

Figure 1. Confidential-data trust reduction

Values are illustrative indices and require replacement with approved validation evidence.

9. Encrypt data across all states

Protect content at rest, in transit and in use with separated keys and authenticated channels.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is an encryption coverage map. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when encryption ends at the host boundary while administrators retain memory access. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

A defensible no-human-access claim is an end-to-end property. Encryption, isolation, attestation, key release, support, logging, recovery and deletion must preserve the same boundary.

The decision pack for encrypt data across all states should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

10. Attest the execution environment

Verify hardware, firmware, boot state, workload image and policy before trusting a processing instance.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is an attestation policy. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when keys are released to an unmeasured or altered runtime. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

The architecture should minimise who and what must be trusted. Workload identity and measured execution can replace broad administrator access for many high-confidentiality operations.

The decision pack for attest the execution environment should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

11. Bind key release to evidence

Release content keys only to approved measurements, identities, locations, versions and purposes.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is an evidence-bound key-release rule. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when possession of a cloud role alone unlocks customer plaintext. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

Confidential computing protects data in use when the workload and key-release path are correctly designed. It does not remove application, output, side-channel, supply-chain or governance risk.

The decision pack for bind key release to evidence should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

12. Govern cryptographic keys

Separate ownership, generation, rotation, backup, revocation, recovery and destruction duties.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is a key-governance matrix. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when one operator can change code, policy and keys without independent control. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

Assurance should rely on reproducible evidence: signed artefacts, attestation reports, policy decisions, key events, access denials, tests, exceptions, incident records and verified deletion.

The decision pack for govern cryptographic keys should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

13. Minimise collected content

Limit fields, documents, duration, precision, replicas and outputs to the approved purpose.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is a purpose-limited data specification. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when the platform retains entire records because storage is inexpensive. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

A defensible no-human-access claim is an end-to-end property. Encryption, isolation, attestation, key release, support, logging, recovery and deletion must preserve the same boundary.

The decision pack for minimise collected content should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

14. Choose privacy-enhancing computation

Assess confidential computing, multiparty computation, homomorphic encryption, tokenisation and differential privacy by use case.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is a technology-selection record. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when an advanced privacy technology is adopted without testing leakage, accuracy or operability. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

The architecture should minimise who and what must be trusted. Workload identity and measured execution can replace broad administrator access for many high-confidentiality operations.

The decision pack for choose privacy-enhancing computation should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

15. Constrain retrieval scope

Bind queries to tenant, user, purpose, collection, jurisdiction, time and permission filters before retrieval.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is a retrieval-authorisation policy. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when semantic similarity bypasses entitlement boundaries. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

Confidential computing protects data in use when the workload and key-release path are correctly designed. It does not remove application, output, side-channel, supply-chain or governance risk.

The decision pack for constrain retrieval scope should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

16. Isolate vector stores and embeddings

Control namespace, keys, metadata, deletion, backup and inference risk for embeddings and indexes.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is an embedding and index control map. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when teams assume embeddings cannot disclose sensitive source content. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

Assurance should rely on reproducible evidence: signed artefacts, attestation reports, policy decisions, key events, access denials, tests, exceptions, incident records and verified deletion.

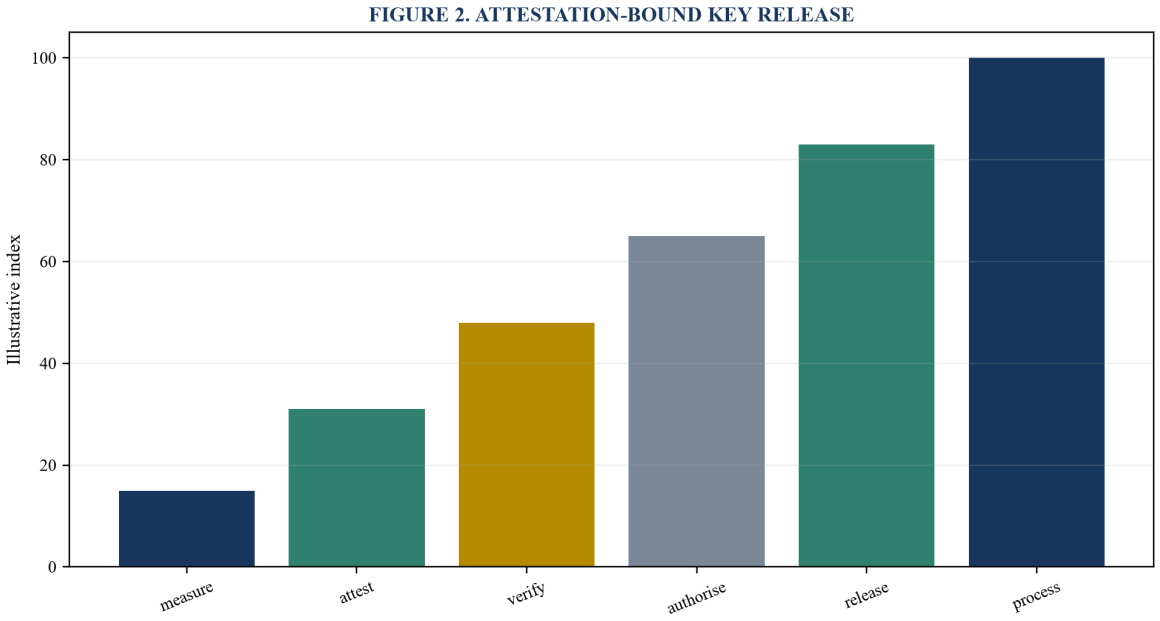
The decision pack for isolate vector stores and embeddings should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

Table 2. Data-in-use controls

Risk	Control	Gate
wrong workload	remote attestation	deny key release
administrator access	isolated execution	no interactive path
cross-tenant retrieval	pre-retrieval authorisation	tenant-bound query
residual plaintext	ephemeral memory and clearing	verified termination

Illustrative controls require use-, consequence-, jurisdiction-, firm- and period-specific approval.

Figure 2. Attestation-bound key release



Values are illustrative indices and require replacement with approved validation evidence.

17. Separate instructions from customer content

Treat retrieved documents and user content as untrusted data; isolate system policy and tool authority.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is a content-instruction separation design. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when prompt injection turns customer content into executable instruction. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

A defensible no-human-access claim is an end-to-end property. Encryption, isolation, attestation, key release, support, logging, recovery and deletion must preserve the same boundary.

The decision pack for separate instructions from customer content should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

18. Control generated outputs

Apply minimisation, policy, redaction, destination, entitlement and human-review gates to responses.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is an output-release policy. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when protected input is withheld while the output reconstructs or exposes it. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

The architecture should minimise who and what must be trusted. Workload identity and measured execution can replace broad administrator access for many high-confidentiality operations.

The decision pack for control generated outputs should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

19. Log events without logging content

Record identities, attestations, policy decisions, hashes, timings, errors and outcomes while excluding plaintext.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is a content-free audit-log schema. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when debug and observability systems become a parallel customer-content store. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

Confidential computing protects data in use when the workload and key-release path are correctly designed. It does not remove application, output, side-channel, supply-chain or governance risk.

The decision pack for log events without logging content should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

20. Monitor confidentiality signals

Detect unusual retrieval, failed attestation, denied key release, cross-tenant access, export and support activity.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is a confidentiality monitoring dashboard. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when monitoring measures uptime while confidentiality failures remain invisible. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

Assurance should rely on reproducible evidence: signed artefacts, attestation reports, policy decisions, key events, access denials, tests, exceptions, incident records and verified deletion.

The decision pack for monitor confidentiality signals should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

21. Engineer break-glass access

Define narrow triggers, dual approval, customer notice, time limits, recording and irreversible audit for exceptional access.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is a break-glass protocol. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when an informal support exception defeats the published access promise. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

A defensible no-human-access claim is an end-to-end property. Encryption, isolation, attestation, key release, support, logging, recovery and deletion must preserve the same boundary.

The decision pack for engineer break-glass access should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

22. Redesign support operations

Use synthetic data, customer-controlled replay, metadata diagnostics and bounded encrypted artefacts for troubleshooting.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is a privacy-preserving support model. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when support teams request screenshots, database dumps or production exports. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

The architecture should minimise who and what must be trusted. Workload identity and measured execution can replace broad administrator access for many high-confidentiality operations.

The decision pack for redesign support operations should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

23. Control vendors and subprocessors

Map hosting, model, telemetry, security, support and data-service providers with contractual and technical boundaries.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is a subprocessor control register. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when a subprocessor gains content access through diagnostics or retention. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

Confidential computing protects data in use when the workload and key-release path are correctly designed. It does not remove application, output, side-channel, supply-chain or governance risk.

The decision pack for control vendors and subprocessors should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

24. Enforce residency and transfer rules

Bind storage, processing, keys, support, backups and disaster recovery to approved locations and legal mechanisms.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is a residency and transfer control map. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when the primary database is local while inference or support crosses borders silently. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

Assurance should rely on reproducible evidence: signed artefacts, attestation reports, policy decisions, key events, access denials, tests, exceptions, incident records and verified deletion.

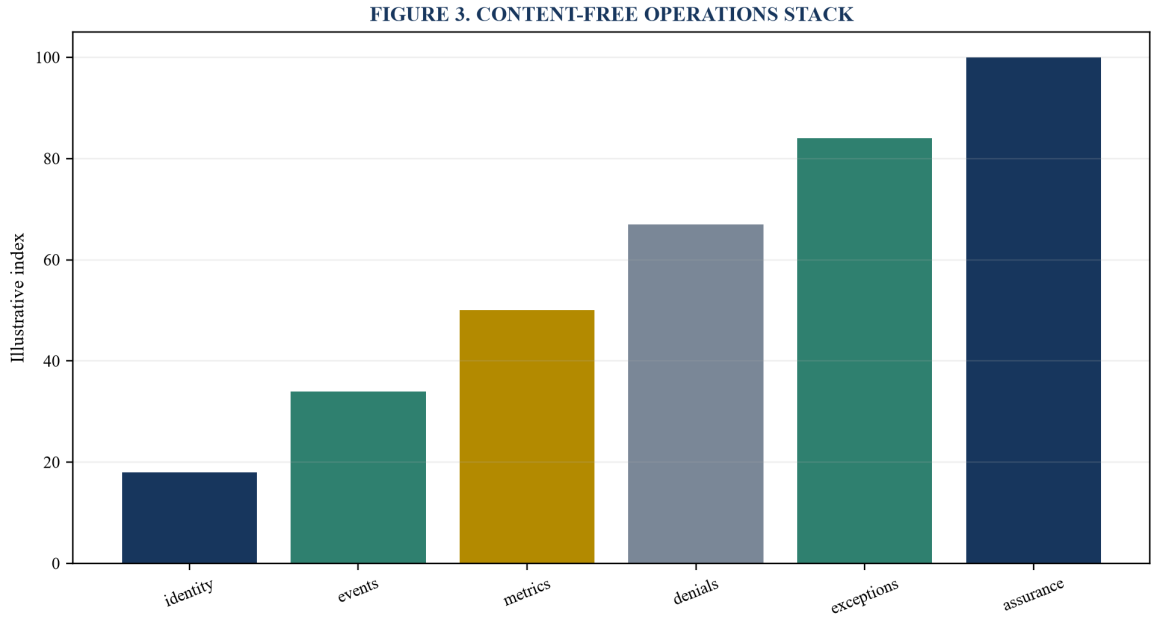
The decision pack for enforce residency and transfer rules should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

Table 3. Operational access paths

Function	Preferred method	Exception evidence
support	metadata and synthetic replay	approved bounded artefact
monitoring	content-free events	documented diagnostic field
incident response	attestation and event reconstruction	dual-approved access
recovery	tenant-bound encrypted restore	tested key and residency gate

Illustrative controls require use-, consequence-, jurisdiction-, firm- and period-specific approval.

Figure 3. Content-free operations stack



Values are illustrative indices and require replacement with approved validation evidence.

25. Protect secrets and configuration

Store credentials, tokens, policies and connection details outside code and customer content with rotation and least privilege.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is a secrets-management design. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when secrets leak through prompts, logs, images or build artefacts. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

A defensible no-human-access claim is an end-to-end property. Encryption, isolation, attestation, key release, support, logging, recovery and deletion must preserve the same boundary.

The decision pack for protect secrets and configuration should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

26. Secure the software supply chain

Sign code, models, containers and policies; verify dependencies, provenance, vulnerabilities and deployment approval.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is a software and model provenance record. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when a compromised image passes into the trusted execution environment. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

The architecture should minimise who and what must be trusted. Workload identity and measured execution can replace broad administrator access for many high-confidentiality operations.

The decision pack for secure the software supply chain should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

27. Protect models, prompts and policies

Separate customer data from proprietary weights, prompts, rules and evaluation sets while controlling extraction.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is a dual-asset protection model. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when a confidentiality design protects customers but exposes the firm's intellectual property. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

Confidential computing protects data in use when the workload and key-release path are correctly designed. It does not remove application, output, side-channel, supply-chain or governance risk.

The decision pack for protect models, prompts and policies should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

28. Prevent training and secondary use

Technically and contractually exclude customer content from model training, product improvement and unrelated analytics unless approved.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is a secondary-use control. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when a default feedback pipeline repurposes protected content. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

Assurance should rely on reproducible evidence: signed artefacts, attestation reports, policy decisions, key events, access denials, tests, exceptions, incident records and verified deletion.

The decision pack for prevent training and secondary use should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

29. Prove retention and deletion

Apply event-driven expiry across source stores, indexes, caches, logs, replicas, backups and vendor systems.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is a verifiable deletion record. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when a deletion request removes the visible record while derived artefacts survive. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

A defensible no-human-access claim is an end-to-end property. Encryption, isolation, attestation, key release, support, logging, recovery and deletion must preserve the same boundary.

The decision pack for prove retention and deletion should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

30. Prepare confidential incident response

Investigate, contain, notify and correct events using metadata, attestations and customer-controlled evidence.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is a confidentiality-preserving incident plan. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when responders copy plaintext broadly during an incident. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

The architecture should minimise who and what must be trusted. Workload identity and measured execution can replace broad administrator access for many high-confidentiality operations.

The decision pack for prepare confidential incident response should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

31. Design backup and recovery controls

Encrypt, isolate, attest, test and expire backups while preserving tenant and residency boundaries.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is a secure recovery architecture. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when recovery media bypasses current identity, key and deletion controls. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

Confidential computing protects data in use when the workload and key-release path are correctly designed. It does not remove application, output, side-channel, supply-chain or governance risk.

The decision pack for design backup and recovery controls should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

32. Test adversarial access paths

Red-team administrators, support, insiders, compromised workloads, prompt injection, side channels and cross-tenant queries.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is an adversarial access test pack. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when the design is tested only through intended application flows. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

Assurance should rely on reproducible evidence: signed artefacts, attestation reports, policy decisions, key events, access denials, tests, exceptions, incident records and verified deletion.

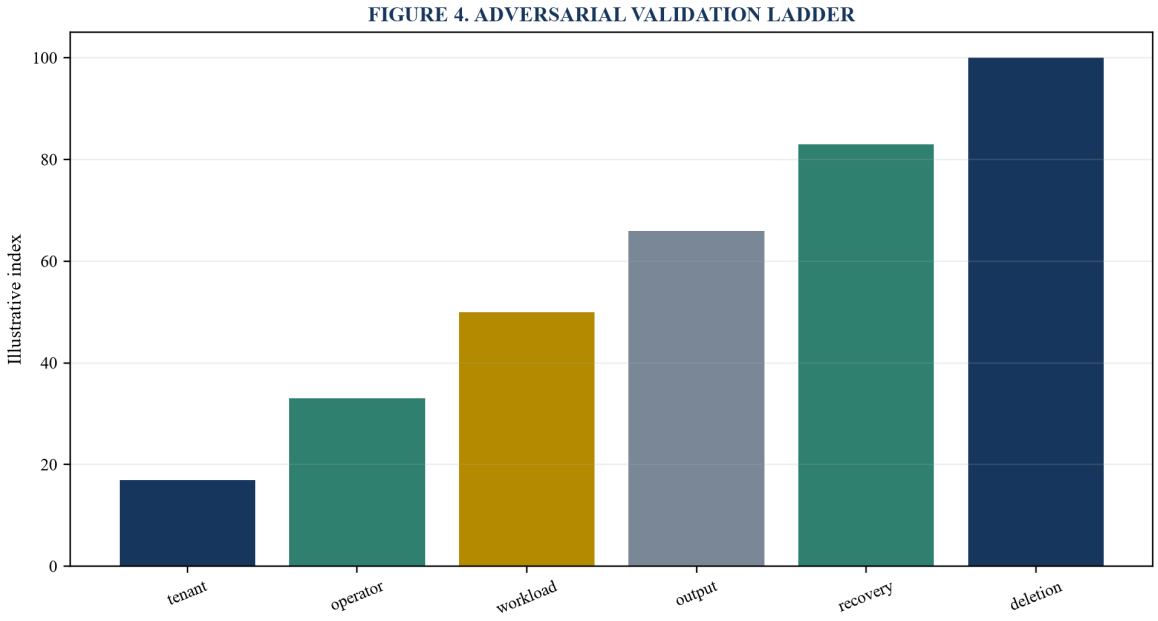
The decision pack for test adversarial access paths should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

Table 4. Confidentiality validation

Metric	Test	Acceptance gate
tenant isolation	adversarial cross-tenant query	zero disclosure
key release	altered workload measurement	release denied
human access	administrator and support paths	no plaintext path
deletion	stores, indexes, logs and backups	evidence complete

Illustrative controls require use-, consequence-, jurisdiction-, firm- and period-specific approval.

Figure 4. Adversarial validation ladder



Values are illustrative indices and require replacement with approved validation evidence.

33. Quantify performance and cost

Measure attestation, encryption, enclave memory, retrieval, inference, logging and recovery effects on service levels.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is a control-performance model. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when security claims fail in production because latency or cost drives undocumented bypasses. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

A defensible no-human-access claim is an end-to-end property. Encryption, isolation, attestation, key release, support, logging, recovery and deletion must preserve the same boundary.

The decision pack for quantify performance and cost should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

34. Build audit-ready evidence

Retain policies, diagrams, measurements, key events, releases, exceptions, tests, incidents and remediation without customer plaintext.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is an evidence-backed assurance pack. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when assurance depends on screenshots or supplier assertions. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

The architecture should minimise who and what must be trusted. Workload identity and measured execution can replace broad administrator access for many high-confidentiality operations.

The decision pack for build audit-ready evidence should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

35. Give customers verifiable assurance

Provide architecture facts, scope, exclusions, control evidence, independent testing and incident commitments.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is a customer assurance statement. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when marketing promises exceed the verified system boundary. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

Confidential computing protects data in use when the workload and key-release path are correctly designed. It does not remove application, output, side-channel, supply-chain or governance risk.

The decision pack for give customers verifiable assurance should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

36. Contract for technical reality

Align confidentiality, access, subprocessors, residency, retention, incidents, audit, liability and exit terms with tested controls.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is a controls-to-contract matrix. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when contracts promise absolute access prevention that operations cannot demonstrate. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

Assurance should rely on reproducible evidence: signed artefacts, attestation reports, policy decisions, key events, access denials, tests, exceptions, incident records and verified deletion.

The decision pack for contract for technical reality should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

37. Govern change continuously

Reassess hardware, firmware, models, prompts, policies, regions, vendors and support processes before release.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is a confidentiality change gate. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when a valid no-access design drifts after routine platform change. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

A defensible no-human-access claim is an end-to-end property. Encryption, isolation, attestation, key release, support, logging, recovery and deletion must preserve the same boundary.

The decision pack for govern change continuously should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

38. Establish accountable operations

Assign product, security, privacy, legal, risk, platform, support, audit and customer responsibilities.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is an accountable operating model. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when each team controls a fragment while no owner can attest to the full promise. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

The architecture should minimise who and what must be trusted. Workload identity and measured execution can replace broad administrator access for many high-confidentiality operations.

The decision pack for establish accountable operations should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

39. Run a bounded production pilot

Use approved data classes, tenants, workloads and geographies with exit criteria and measured evidence.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is a controlled pilot record. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when a laboratory proof expands before support, recovery and incident paths are tested. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

Confidential computing protects data in use when the workload and key-release path are correctly designed. It does not remove application, output, side-channel, supply-chain or governance risk.

The decision pack for run a bounded production pilot should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

40. Close through evidence gates

Require approved purpose, lifecycle, isolation, identity, attestation, key release, outputs, logs, exceptions, deletion and assurance.

The controlled record includes scope; authority; source; passage; fact; calculation; reviewer; decision. The immediate deliverable is an evidence-gated confidential-AI operating model. Preserve jurisdiction, effective date, version, transformation, reviewer, approval and unresolved exceptions.

The principal failure occurs when the absence of routine human access is mistaken for a verified confidential-data system. Reviewers should reconstruct the evidence path, test consequence under adverse conditions and identify who may approve or withhold the recommendation.

Assurance should rely on reproducible evidence: signed artefacts, attestation reports, policy decisions, key events, access denials, tests, exceptions, incident records and verified deletion.

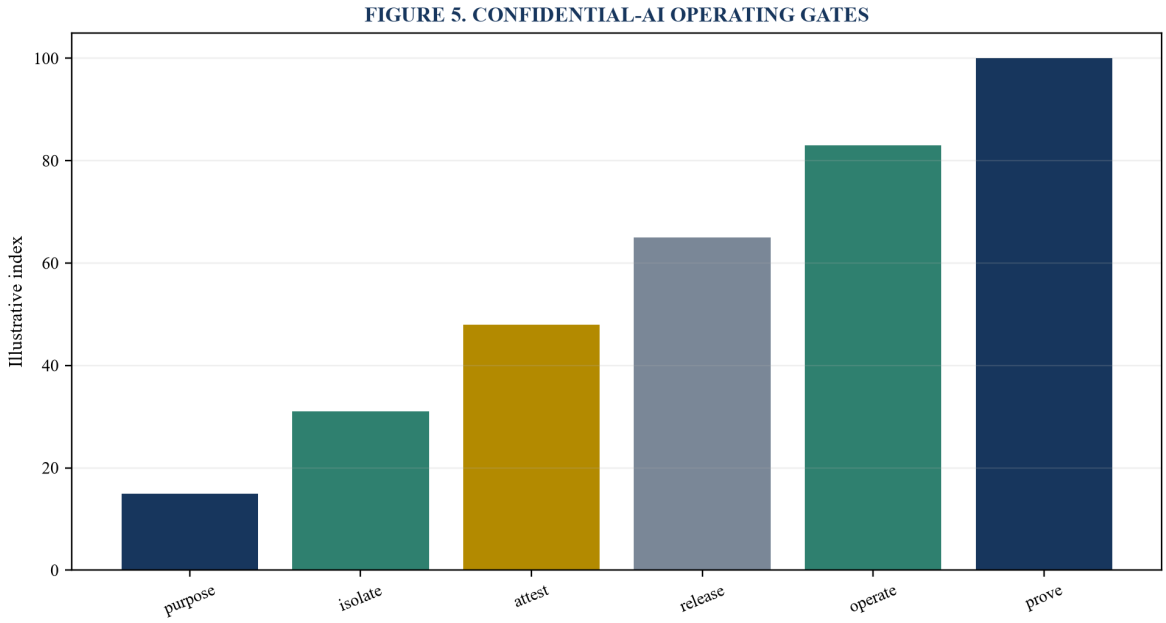
The decision pack for close through evidence gates should show the prior view, correction, consequence at stake, alternatives, owner, due date, next evidence gate and observed outcome. Exceptions require closure evidence and an explicit effect on retrieval, citation, abstention, review or release.

Table 5. Confidential-AI gates

Gate	Primary work	Completion evidence
scope	purpose, data and consequence	approved charter
trust	isolation, identity and attestation	verified boundary
operation	outputs, logs, support and recovery	tested controls
assurance	exceptions, deletion and reporting	accountable approval

Illustrative controls require use-, consequence-, jurisdiction-, firm- and period-specific approval.

Figure 5. Confidential-AI operating gates



Values are illustrative indices and require replacement with approved validation evidence.

References

[1] National Institute of Standards and Technology, NIST IR 8320E Initial Public Draft: Hardware-Enabled Security: Confidential Computing of Data in Cloud Workloads, https://csrc.nist.gov/pubs/ir/8320/e/ipd

[2] National Institute of Standards and Technology, NIST IR 8320: Hardware-Enabled Security for Server Platforms, https://csrc.nist.gov/pubs/ir/8320/final

[3] National Institute of Standards and Technology, Confidential Computing Glossary, https://csrc.nist.gov/glossary/term/confidential_computing

[4] National Institute of Standards and Technology, NIST SP 800-207: Zero Trust Architecture, https://csrc.nist.gov/pubs/sp/800/207/final

[5] National Institute of Standards and Technology, NIST SP 1800-35: Implementing a Zero Trust Architecture, https://csrc.nist.gov/pubs/sp/1800/35/final

[6] National Institute of Standards and Technology, Privacy-Enhancing Cryptography, https://csrc.nist.gov/projects/pec

[7] National Institute of Standards and Technology, Fully Homomorphic Encryption, https://csrc.nist.gov/Projects/pec/fhe

[8] Cybersecurity and Infrastructure Security Agency, Zero Trust Maturity Model Version 2, https://www.cisa.gov/resources-tools/resources/zero-trust-maturity-model

[9] Central Bank of the UAE, Data Quality, Privacy and Security for AI and ML, https://rulebook.centralbank.ae/en/rulebook/5-data-quality-privacy-and-security

[10] Central Bank of the UAE, Guidance Note on Responsible Adoption and Use of Artificial Intelligence, https://rulebook.centralbank.ae/en/rulebook/guidance-note-consumer-protection-and-responsible-adoption-and-use-artificial-intelligence

[11] Central Bank of the UAE, Guidelines for Financial Institutions Adopting Enabling Technologies, https://rulebook.centralbank.ae/en/rulebook/guidelines-financial-institutions-adopting-enabling-technologies

[12] UAE Government, Federal Decree-Law No. 45 of 2021 on the Protection of Personal Data, https://uaelegislation.gov.ae/en/legislations/1972

[13] European Union, Regulation (EU) 2016/679 General Data Protection Regulation, https://eur-lex.europa.eu/eli/reg/2016/679/oj

[14] UK Information Commissioner's Office, Guidance on AI and Data Protection, https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/

[15] UK Information Commissioner's Office, Security and Data Minimisation in AI, https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/how-should-we-assess-security-and-data-minimisation-in-ai/

[16] European Union Agency for Cybersecurity, Cybersecurity of AI and Standardisation, https://www.enisa.europa.eu/publications/cybersecurity-of-ai-and-standardisation

[17] Amazon Web Services, AWS Nitro Enclaves User Guide, https://docs.aws.amazon.com/enclaves/latest/user/nitro-enclave.html

[18] Amazon Web Services, Cryptographic Attestation and Secrets Management for Nitro Enclaves, https://docs.aws.amazon.com/enclaves/latest/user/connect-enclave-kms.html

[19] Microsoft, Azure Confidential Computing, https://learn.microsoft.com/en-us/azure/confidential-computing/

[20] Microsoft, Confidential Containers on Azure Container Instances, https://learn.microsoft.com/en-us/azure/container-instances/container-instances-confidential-overview

[21] Microsoft, Azure Attestation Overview, https://learn.microsoft.com/en-us/azure/attestation/overview

[22] Google Cloud, Confidential VM Attestation, https://cloud.google.com/confidential-computing/confidential-vm/docs/attestation

[23] International Organization for Standardization, ISO/IEC 27001 Information Security Management Systems, https://www.iso.org/standard/27001

[24] International Organization for Standardization, ISO/IEC 27701 Privacy Information Management, https://www.iso.org/standard/85819.html

[25] International Organization for Standardization, ISO/IEC 42001 AI Management Systems, https://www.iso.org/standard/81230.html

[26] Organisation for Economic Co-operation and Development, OECD AI Principles, https://oecd.ai/en/ai-principles

About the Author

Authored by Chennakeshav Adya, Independent Researcher

Chennakeshav (CK) is a corporate finance and investment banking executive with 25+ years of global experience in deal origination, structuring and execution across M&A, growth capital and corporate strategy. He has led value-creation mandates for founders, corporates and funds — bridging the boardroom view to hands-on execution and close.

His career spans Morgan Stanley, HSBC, Lloyds Banking Group, EWEC, ADQ portfolio companies and Emirates Growth Fund, across TMT, real estate, fintech, deeptech, cleantech, infrastructure and energy. He has partnered with C-suite leaders, private equity and venture funds, sovereign wealth funds and family offices to finance complex fund raises and scale-up ventures, and has led M&A due diligence, post-merger integration and business-transformation initiatives to create value.

At Matchpoint Partners he is Managing Partner, leading the firm's corporate finance, M&A and capital-raising practice. He holds an MBA from London Business School, an engineering degree from VTU and a Master of Laws (LLM, in progress) from UCL London.

An active start-up mentor, CK mentors at Techstars, DIFC FinTech Hive, Startup Grind, Founder Institute and IN5, serves as Entrepreneur Mentor in Residence (EMiR) at London Business School, and judges the Entrepreneurship World Cup.