SOC 2 Type II for AI and Machine Learning Companies: A Practical Implementation Guide

Compliance Guide for AI/Machine Learning Teams

Prepared by Jacobian Engineering | 2026-02-09

This guide is for informational purposes only and does not constitute legal advice.

Executive Summary

AI and machine learning vendors are being held to the same security expectations as established SaaS companies. Buyers want proof that you can protect customer data, run reliable services, and respond to incidents with discipline. A SOC 2 Type II report is one of the most common ways to meet that expectation, especially in enterprise sales.

AI adds a twist. Your product may include training pipelines, model registries, prompt logs, feature stores, and third party foundation models. That raises practical questions. What is in scope for the audit. Who can access training data. How do you control model releases. How do you prove that you monitor for misuse. This guide explains how SOC 2 Type II works and how AI and machine learning teams can implement controls that auditors can test and customers can understand.

Why SOC 2 matters for AI and machine learning companies

SOC 2 is an attestation report performed by an independent CPA firm using the AICPA Trust Services Criteria. It helps customers evaluate how you protect systems and data. For AI companies, it often becomes the default package you share during due diligence, instead of answering the same questionnaire repeatedly.

Many AI teams learn about SOC 2 the hard way. A large customer asks for a Type II report, and a deal stalls. The fastest path is rarely a rush job. SOC 2 testing looks at how you operate over time. The earlier you build lightweight routines, the easier the audit becomes.

Common triggers

- **Enterprise procurement:** A prospect requires a SOC 2 Type II report before signing, renewing, or expanding usage.
- **Security reviews for AI features:** Buyers ask how models are trained, how prompts are logged, and how outputs are monitored.
- **Platform and marketplace programs:** Partners may request formal assurance before listing or integrating.
- **Fundraising and M&A:** Investors often want evidence that security is not improvised.

Type I vs Type II

A Type I report tests whether controls are designed appropriately at a point in time. A Type II report tests both design and operating effectiveness over an observation period. If customers ask for SOC 2, they usually mean Type II. It demonstrates that controls are not just written down, but followed consistently.

Trust Services Criteria in an AI context

SOC 2 is organized around five Trust Services Criteria. Security is required. The other criteria are included based on what you promise customers and what your service does.

- **Security:** Protection against unauthorized access and inappropriate use of systems and data.

- **Availability:** System availability for operation and use as committed or agreed.
- **Processing Integrity:** System processing is complete, valid, accurate, and timely.
- **Confidentiality:** Information designated as confidential is protected as committed or agreed.
- **Privacy:** Personal information is collected, used, retained, and disclosed appropriately.

AI teams sometimes assume SOC 2 is only about infrastructure security. It is broader than that. Auditors will ask how you build and change software, how you control access, how you handle incidents, and how you manage vendors. If your model outputs affect customer workflows, processing integrity and availability can also be important.

Scoping SOC 2 for AI and machine learning systems

Scope is the most important early decision. It defines what systems, teams, and vendors are included in the audit. If your scope is too broad, you will spend months gathering evidence for low value systems. If it is too narrow, customers may reject the report or ask for exceptions.

Start with the service boundary

Ask a simple question. What are customers buying from you. For an AI company, the answer might be an API, a web application, a model hosting service, an internal agent that performs tasks, or a managed data science platform. The in scope system should include the components that deliver that service in production.

AI specific components that often belong in scope

- **Training and evaluation environments:** Not every experiment needs to be in scope, but production training pipelines and evaluation jobs often influence what gets deployed.
- **Model registry and release process:** How models are stored, approved, versioned, and promoted to production.
- **Inference services:** The runtime that serves predictions, including scaling, secrets, and monitoring.
- **Data pipelines and feature stores:** Where customer data or derived features are ingested, stored, and transformed.
- **Prompt and conversation logs:** If you store prompts, retrieved context, or chat transcripts, they may be part of the system boundary and privacy story.
- **Third party model dependencies:** Foundation model providers, vector databases, annotation vendors, and any service that can access customer data.

What about general purpose models and hosted LLMs

If you rely on third party models, you still need to manage the risk. Auditors and customers will ask how you evaluate vendors, how you restrict what data is sent, and how you monitor usage. A practical scoping approach is to treat the vendor as a subservice organization and document your responsibilities. What controls do you rely on them for. What controls remain yours. What does your contract say about

data retention and access.

Define the data in scope

AI systems touch more data types than teams expect. Training data can include customer uploads, public datasets, synthetic data, telemetry, and feedback labels. In scope data definitions should be explicit. What is customer content. What is metadata. What is model output. What is considered confidential. If you cannot explain your data lifecycle, how will you demonstrate confidentiality and privacy.

Complementary user entity controls

Most SOC 2 reports include a section describing what customers must do for the controls to work. For AI services, this might include guidance about how customers manage their own user access, how they configure prompts or retrieval sources, and what they should not upload. Clear customer responsibilities reduce audit friction and reduce misuse risk.

Control areas auditors expect and how to implement them

SOC 2 does not require a specific toolset. It requires that controls exist, are suitable, and operate consistently. For AI and machine learning companies, a few control areas tend to carry the most weight.

Governance and risk management

- **Security ownership:** Define who is accountable for security decisions. Who approves exceptions. Who is on call for incidents.
- **Risk assessments:** Maintain a risk register that includes AI specific risks such as training data exposure, model misuse, and output safety failures.
- **Policies and standards:** Keep policies short, specific, and aligned to how you work. Auditors test reality, not binders.

Identity and access control

- **Least privilege:** Reduce admin access to cloud accounts, model registries, and production data stores. Review access on a routine schedule.
- **Strong authentication:** Use multi factor authentication and single sign on where possible. Document exceptions.
- **Secrets management:** Store API keys, model credentials, and signing keys in a managed secrets system. Rotate when people leave or vendors change.

Change management for models and code

AI teams often have a mature code review process but an informal model release process. Auditors will treat a model release like a production change. How is it approved. How is it tested. How do you roll back. Where is the evidence.

- **Model versioning:** Assign version identifiers and tie them to training data, code, and evaluation results.
- **Approval workflow:** Require peer review and documented approval for production promotions.
- **Reproducibility:** Maintain run logs or pipelines so you can explain how a model was produced.

Logging, monitoring, and incident response

Customers expect AI services to detect abuse quickly. Auditors expect that you log security relevant events, review alerts, and respond consistently.

- **Security logs:** Capture authentication events, administrative actions, and network level activity for production environments.
- **AI usage monitoring:** Monitor rate limits, unusual prompt patterns, excessive tool calls, and signs of data exfiltration.
- **Incident process:** Maintain an incident response plan, practice it, and keep ticket evidence of real events and drills.

Data protection and retention

- **Encryption:** Encrypt sensitive data in transit and at rest. Document key management practices.
- **Retention rules:** Define how long prompts, outputs, and training artifacts are stored. Keep the rules consistent with customer commitments.
- **Data deletion:** Maintain a process to delete customer data when required. Can you prove you did it.

Vendor and third party management

AI stacks are vendor dense. Auditors will want to see that you track vendors, assess risk, and manage contracts. Customers will ask the same questions. Vendor management is both an audit requirement and a sales requirement.

- **Vendor inventory:** Keep a list of vendors that can store or access customer data, including model providers.
- **Due diligence:** Collect SOC 2 reports, security documentation, and data handling terms from vendors.
- **Ongoing review:** Reassess critical vendors on a schedule and when services change.

Building an evidence routine that does not overwhelm the team

A SOC 2 Type II audit is largely an evidence audit. Auditors do not accept statements. They accept records. The good news is that you can build a routine that fits a small team. What evidence do you already generate through tickets, pull requests, access logs, and monitoring dashboards.

Examples of evidence auditors often request

- **Access reviews:** A record showing you reviewed privileged access and removed what was not needed.
- **Change approvals:** Pull requests, code review records, and deployment approvals.
- **Vulnerability management:** Scan results, patch records, and remediation tickets.
- **Incident records:** Tickets, timelines, root cause analysis, and lessons learned.
- **Backups and recovery tests:** Evidence that backups run and that you test restore procedures.
- **Security training:** Proof that employees completed training and accepted policies.

AI adds another evidence stream. Model evaluation reports and model release approvals can be treated as change management evidence. Abuse monitoring dashboards can support detection controls. If you already do the work, you can usually shape the artifacts into audit evidence with a few process tweaks.

Implementation methodology for SOC 2 Type II

Phase 1: Readiness and scoping

Start with a gap assessment against the Trust Services Criteria you plan to include. Define scope, data types, and key vendors. Decide whether you need a Type I milestone first. Many teams can move directly to Type II if controls are already operating.

Phase 2: Control design and implementation

Implement or refine controls, then document them as simple procedures. Align tooling with the controls rather than the other way around. Confirm that controls are consistently followed before the observation period begins.

Phase 3: Observation period and audit support

Type II testing requires an observation window. During this time, you collect evidence monthly and keep controls running. The audit firm will sample from that period. If your evidence routine is weak, the audit becomes expensive and stressful.

Phase 4: Closeout and continuous compliance

After the report is issued, your job is to keep controls operating. SOC 2 is not a one time project. Customers will ask for updated reports every year. Treat it like an operating system for security and reliability.

AI and machine learning pitfalls that slow SOC 2 audits

- **Unclear prompt and output handling:** Teams store prompts and outputs for debugging but do not define retention, access controls, or deletion procedures.
- **Model releases without governance:** Model promotions happen outside normal change control, making it hard to show approval and testing.

- **Shadow AI tools:** Engineers use external AI services for development without vendor review or data rules.
- **Training data sprawl:** Datasets are copied across environments and personal laptops, increasing exposure and creating audit gaps.
- **Vendor overreliance:** Contracts with model providers do not address retention, sub processors, or incident notification.

Each pitfall has a simple corrective action. Define the data lifecycle, build a model release workflow, and align vendor management with reality. If you are unsure where to start, ask which gap would most concern a customer reviewer.

Business benefits beyond the audit

A well run SOC 2 program can improve operations. Clear ownership reduces confusion during incidents. Change management reduces production surprises. Vendor inventory reduces last minute procurement delays. A security program that is visible and consistent is easier to defend when customers ask hard questions.

For AI companies, SOC 2 also creates a structure for model operations. Model approvals, evaluation reports, and monitoring routines become part of the normal rhythm. That reduces risk when the product scales or when you expand into regulated markets.

How Jacobian Engineering supports SOC 2 for AI companies

Jacobian Engineering helps teams design and implement controls that fit real environments. That can include SOC 2 readiness assessments, scoping workshops, policy and procedure writing, evidence routine design, and technical implementation support for logging, monitoring, access control, and incident response. For AI heavy products, the team also performs penetration testing and AI red teaming so model and application risks are tested, not assumed.

Conclusion

SOC 2 Type II is not just a sales artifact. It is a way to show that your organization operates with discipline. AI and machine learning companies can meet SOC 2 expectations without turning security into a bureaucracy. The key is good scoping, a small set of repeatable controls, and an evidence routine that matches how you already work.

If you are preparing for your first SOC 2 Type II audit, or you need to rebuild a program that has become hard to maintain, Jacobian Engineering can help you implement practical controls and keep them running over time.