

SAMPLE

SOC 2 for AI Startups

The opening of SOC2-AI-Controls-Guide.docx, one of 8 files in the kit. <https://solasai.net/products/soc2-ai-startups/>

12 AI-SPECIFIC SOC 2 CONTROLS

The Controls Your Auditor Will Ask About That Aren't in Any Standard Playbook

Product: *SOC 2 for AI Startups*

Version: 1.0

Classification: Purchaser Use Only

Solas AI

solasai.net

Less grind. More growth.

Contents

Introduction

Control 1: Customer Data Segregation from Training Pipelines

Control 2: Prompt Logging and Monitoring

Control 3: Tenant Data Isolation in AI Processing

Control 4: AI Model Vendor Management

Control 5: Prompt Injection Protection

Control 6: AI Output Monitoring and Content Safety

Control 7: AI-Specific Incident Response

Control 8: Model Versioning and Change Management

Control 9: Bias Risk Assessment and Monitoring

Control 10: Data Minimization for AI Processing

Control 11: AI Feature Disclosure and Transparency

Control 12: Human Oversight and Escalation

Controls Summary Matrix

Introduction

The SOC 2 Trust Services Criteria (TSC) are principle-based and technology-neutral. They were written before generative AI existed. When your product architecture includes AI models, training pipelines, and prompt processing, your auditor interprets the TSC through the lens of your specific architecture — and that interpretation expands into territory that standard SOC 2 prep tools completely miss.

This guide documents the 12 AI-specific controls that experienced SOC 2 auditors are asking AI companies about right now. Each control is mapped to the relevant Trust Services Criteria, with implementation guidance and specific evidence requirements.

These are not “extra” requirements. They are AI-specific implementations of existing TSC controls. The difference is that most SOC 2 playbooks assume your product is deterministic code running on servers — not probabilistic models running on someone else’s infrastructure.

How to Use This Guide

For each control: (1) read the “Why It Matters” section to understand the risk, (2) review “What the Auditor Wants” to know what evidence to prepare, (3) follow the Implementation Guidance to build the control, and (4) collect the Evidence listed. Your auditor may ask for all of this, some of it, or interpret it differently — but you will be prepared.

Control 1: Customer Data Segregation from Training Pipelines

Maps to: CC7.2 (Data Integrity), PI1.1 (Processing Integrity)

Why It Matters

This is the #1 question enterprise customers ask about AI vendors. Your auditor knows it. If your prospect’s security team sends you a questionnaire, the first AI-specific question will be: “Is our data used to train your models?” If you cannot answer with technical evidence — not just a policy statement — you fail this control.

What the Auditor Wants to See

Architecture diagram showing separation between customer data paths and training pipelines. DPA clause explicitly prohibiting training on customer data. API configuration screenshot confirming training opt-out is enabled. For self-hosted models: IAM policies preventing cross-access between customer data stores and training infrastructure.

Implementation Guidance

- If using a third-party API (OpenAI, Anthropic, Google): confirm training opt-out in both the DPA and the API configuration. Document both with screenshots.
- If fine-tuning models: customer data and training data must live in separate infrastructure with IAM policies preventing cross-access. Document the separation architecturally.
- If using embeddings/RAG: embeddings derived from customer data inherit the customer's data classification. They must not be used for model improvement or shared across tenants.
- Review your data flow diagram end-to-end and mark every point where customer data could theoretically enter a training pipeline. Document each point and the control preventing it.

Required Evidence

1. Data flow diagram with training pipeline marked
2. DPA clause (highlighted section)
3. API configuration screenshot showing opt-out
4. IAM policy documents (if self-hosted)
5. Quarterly verification log

Control 2: Prompt Logging and Monitoring

Maps to: CC7.2 (Monitoring), CC7.3 (Change Detection), CC7.4 (Incident Detection)

Why It Matters

In traditional SaaS, you log API requests and database queries. In an AI product, the "logic" that generates outputs is a model responding to prompts. Without prompt logging, you cannot investigate incidents, audit model behavior, demonstrate processing integrity, or prove what happened when a customer complains about an AI-generated output.

What the Auditor Wants to See

Logging architecture showing what is captured per request (prompt, output, tenant ID, timestamp, model version, token count). Sample log entry demonstrating all fields. Retention policy with defined period. Access controls on the prompt logs (they contain customer data). Alert rules for anomalous patterns.

The sample ends here. The full document runs about 3,321 words; the kit holds 8 files. Version 2.2, 19 October 2026. Buy it at <https://solasai.net/products/soc2-ai-startups/>, 14-day refund for any reason.