

Privacy Commitments with GroqCloud

Publication Date: November 26, 2025

Abstract & Disclaimer

This document provides customers with information about GroqCloud services to help them align with requirements under applicable data privacy laws. The content contained herein is correct as of November 26, 2025 and represents the state of GroqCloud services at the time of publication. GroqCloud services may evolve as we continually enhance protection for our customers.

Introduction to GroqCloud

GroqCloud is a high-performance AI inference platform designed for both developers and enterprises, providing scalable, low-latency access to large language, audio, image and multimodal models. Powered by Groq's purpose-built Language Processing Unit (LPU) hardware and a streamlined, API first Console, GroqCloud enables you to move from prototype to production without sacrificing speed, precision, or cost control. The platform supports turnkey deployment of open-weight models, agentic workflows, and fine-tuned adapters (e.g., via LoRA) under a usage-based pricing model. With GroqCloud, you gain predictable spend, ultra-fast tokens-per-second throughput, broad framework compatibility, and frictionless integration into your toolchain, making AI inference practical and scalable. Learn more at <https://console.groq.com>.

What is Inference?

AI inference is the process of running a trained model to generate outputs based on new input data. At the inference stage, the model has already been trained, and inference is when that trained model is applied to predict the next word in a sentence, summarize text, or answer a question.

An inference provider operates the computing infrastructure that executes inference requests. When the provider hosts open-source LLMs, it is responsible for:

- Deploying those models (a list of GroqCloud Production Models is available [here](#)) on hardware, like GPUs, or in the case of GroqCloud, optimized LPUs;
- Accepting requests from customers who send prompts (i.e. inputs);
- Running the model computations to generate a response (i.e. outputs); and
- Returning the output without training or altering the underlying model.

So, the inference provider's value lies in offering speed, reliability, and privacy in how those open-source models are served to customers.

GroqCloud is an inference provider and not a general-purpose cloud. It is a performance optimized environment for executing AI workloads on Groq's LPU. Instead of renting virtual machines or storage space (like with AWS or Azure), customers use GroqCloud to send AI requests and get ultra-fast, deterministic results from Groq's systems. Because GroqCloud inference involves executing computations rather than hosting or storing applications, the data protection surface is narrower than that of multi-service clouds.

GroqCloud Data Privacy Overview

GroqCloud provides high-performance inference infrastructure built for security, privacy, and transparency. Groq's approach is grounded in the principle that customers own and control their data. Groq acts as a processor of Customer Data, helping customers comply with leading data-protection laws, including:

- EU General Data Protection Regulation (GDPR)
- UK GDPR
- The California Consumer Privacy Act (CCPA), as amended by the California Privacy Rights Act (CPRA)
- Swiss Federal Data Protection Act
- Saudi Arabia Personal Data Protection Law (PDPL)
- Other applicable international privacy frameworks

This whitepaper explains how GroqCloud's privacy, data protection, and compliance practices align with these global standards.

Shared Responsibility

GroqCloud's privacy and security controls enable customers to use inference capabilities in compliance with major privacy frameworks. Groq continually expands its controls and provides documentation in the Console and on the Trust Center at <https://trust.groq.com>.

Customer responsibilities include:

- Configuring Console settings to implement appropriate technical and organizational controls.
- Providing accurate Account and Authorized User information.
- Complying with the Acceptable Use Policy (AUP), Export Control Laws, and High-Risk Activity restrictions.
- Obtaining a lawful basis for processing and issuing privacy notices to Data Subjects.

Privacy Roles, Instruction, Data Handling, and Scope

Customer as the Data Controller for Customer Data

The customer determines the purposes and means of processing Customer Data. GroqCloud allows customer administrators to configure [data controls](#) and [model permissions](#) at the organization level, including enabling [Zero Data Retention \(ZDR\)](#) for transient processing.

GroqCloud as the Data Processor for Customer Data

GroqCloud processes Customer Data solely pursuant to customer's documented instruction, as defined in the [Data Processing Addendum](#) (DPA). As a data processor, Groq accesses and uses data only as necessary to provide the service, comply with law, or fulfill contractual obligations.

Scope of Data Commitments

The Groq [Services Agreement](#) and [DPA](#) govern Customer Data processed in generally available services. Customer Data includes inputs, outputs, and data provided by customers for features like [Batch](#) API or [LoRA](#) inference.

Subprocessors

GroqCloud provides customers with transparency of its subprocessors. To help customers evaluate GroqCloud for their needs, Groq publishes its subprocessor list on the [Trust Center](#). This list is kept current as services evolve and new third-party providers are onboarded under rigorous security and compliance assessments. Customers can be notified of any changes by subscribing to email notifications as provided on the [Trust Center](#), in accordance with [Section 7.2 of the DPA](#).

International Transfers

GroqCloud only transfers Customer Data in accordance with the DPA. The DPA incorporates EU Standard Contractual Clauses (SCCs), the UK Addendum to the SCCs, and the KSA Controller-to-Processor (C2P) Clauses for the Kingdom of Saudi Arabia. These mechanisms support lawful transfers of personal data from the EU, UK, Switzerland, and Saudi Arabia to third countries in compliance with local data-protection laws.

Data Subject Requests

Customers can manage Data Subject Access Requests (DSARs) through the [Privacy Request Center](#). The request center allows customers to request access or deletion of data to the extent Groq holds data necessary to fulfill a DSAR the customer has received.

Security, Incident Response, Operations, and Audits

Groq provides detailed information on its security, incident response, and operational controls in the [Groq's Enterprise Security and Compliance Framework](#) whitepaper available as a resource on the Trust Center. This framework details SOC 2 Type II certification, continuous monitoring, secure SDLC practices, and a 24/7 incident response process. Download the paper to learn more.

Console Audit Logs. GroqCloud also offers Developer and Enterprise tier customers with Console audit logging. Developer tier customers get 7 days of trailing console audit logs. Enterprise tier customers will receive 90 days of trailing console audit logs. Both Developer and Enterprise tier customers can export console audit logs to the log analysis tool of choice. These logs allow admins to monitor changes to settings or configurations in the Groq console.

Retention and Deletion

GroqCloud provides customers with visibility into data retention through its [Data Controls documentation](#). Administrators can configure ZDR and related settings directly in the [Console](#) to align with internal compliance policies and regulatory obligations.

Sovereignty

Sovereignty is the right to enforce control over the location of data processing. GroqCloud allows Enterprise customers to direct inference requests to the selected endpoint that best suits their *location* needs. When Enterprise customers select an inference endpoint, and enable [ZDR](#), the inference processing happens within the selected region and outputs are protected up to the application layer. Free and Developer tier customers can also enable ZDR to prevent inputs or outputs from being logged.

Conclusion

GroqCloud's privacy by design model, coupled with its Services Agreement, DPA, [Cloud Console](#), and [Trust Center](#), provides a comprehensive framework that enables customers to meet data privacy obligations effectively.

- Processor-only Customer Data handling and explicit lawful basis.
- Defense-in-depth security validated by SOC 2 Type II (Security) certification.
- Transient, privacy-preserving inference with ZDR.
- Transparent subprocessor governance via the Trust Center.
- International data transfer safeguards and SCCs and C2P.
- A shared responsibility model that defines both Groq and customer roles.

Groq continues to enhance its privacy and security controls in alignment with evolving regulations and best practices, ensuring that as Groq advances AI inference performance, security and privacy remain foundational to every service delivered.