



Claude

<https://www.anthropic.com/legal/privacy>

Reviewed on 2026-09-14 17:53:14 UTC



OVERALL PRIVACY SCORE

Composite rating across data sharing, usage, retention, security and confidentiality.

CATEGORIES

6

AVG CONFIDENCE

93%

Summaries

Tool Summary

Anthropic is an AI safety and research company dedicated to building reliable, interpretable, and steerable AI systems, famously known for its Claude AI models and conversational assistant services.

Executive Summary

This review of Anthropic's Privacy Policy (effective September 10, 2026) reveals robust user controls for AI model training, but highlights several areas of concern under strict privacy rubrics. Notably, personal data is shared for targeted advertising by default (subject to opt-out), and the policy lacks explicit commitments to data confidentiality and specific retention timelines. While users can opt out of having their Inputs and Outputs used for AI training, the lack of a clear distinction between content and metadata training, combined with vague security commitments lacking explicit certifications, limits the overall privacy score.

Category Breakdown

Data Usage

Bad



Data is used to provide, maintain, and improve services, as well as for tailored recommendations and direct marketing under legitimate interests. Because data is used for advertising and profiling by default (with opt-out options), a hard penalty applies, capping the score at 4.

Data Sharing

Bad



Anthropic shares data with affiliates, service providers, and business partners. Crucially, the policy states that users can opt out of sharing personal data for targeted advertising, meaning such sharing is enabled by default without explicit opt-in. This triggers a hard penalty capping the score.

Security Commitments

Neutral



The policy contains generic statements about implementing appropriate technical and organizational security measures to protect personal data. However, it fails to mention any specific security certifications (such as ISO 27001 or SOC 2), triggering a hard penalty for vague 'reasonable measures' that caps the score at 5.

Confidentiality Protections

Bad



There is no explicit commitment in the policy text treating user data as strictly confidential, nor are there detailed internal access controls framed as confidentiality obligations. This triggers a hard penalty capping the score at 3.

Data Retention & Deletion

Bad



Anthropic retains personal data 'as long as reasonably necessary' for its processing purposes, which is vague and lacks specific timelines. Although users can delete individual conversations (which are removed from back-ends within 30 days), the undefined general retention period triggers a hard penalty capping the score at 4.

AI Training & Model Rights

Neutral



Anthropic uses Inputs and Outputs to train and improve its AI models by default, though a universal opt-out is available via account settings. However, because the policy lacks a clear distinction between content and metadata training, a hard penalty applies, capping the score at 5.