

AI Security at Typeform

Frequently Asked Questions

At Typeform, we understand that security and privacy are top priorities for our customers, especially for AI-powered features and capabilities. Typeform is ISO/IEC 42001 certified — the international standard for AI Management Systems — providing independent, third-party validation of our approach to responsible AI.

This document provides a summary of our security and privacy practices governing our use of AI in the Typeform platform. For a broader overview of our information security program, please refer to our InfoSec Summary Doc.

1. AI Data & Security

What AI models are used?

Typeform uses AI models developed by Anthropic and OpenAI to leverage their respective strengths and provide a range of AI-powered features. For more information on these AI providers, please refer to their Trust Centres:

- OpenAI: https://trust.openai.com
- Anthropic: https://trust.anthropic.com

What data is passed to third-party AI models?

The data sent is strictly limited to what is necessary for each specific AI feature to function. Some features operate on form questions or responses, and many features do not require any personal data. Customer data is only processed when an AI feature is actively used. For more details, see the relevant feature article from our AI Help Centre article: AI with Typeform FAQ.

How is data segregated if the model serves multiple tenants?

Customer data is logically segregated at the application level within our AWS environment.

When using **Anthropic** models, this is done via AWS Bedrock, so data remains within the customer's existing AWS environment.

When using **OpenAI** models, processing happens within a secure, dedicated Typeform environment, and OpenAI does not retain or use any customer data for training.

What security measures are in place to protect AI models, data, and infrastructure?

Access to production environments, including AI models, is restricted to authorised employees using SSO, MFA, and VPN. Data is encrypted using TLS 1.2+ and AES-256. We use firewalls, IDS, and WAF for network security. Access logs are monitored, and confidentiality is enforced through strict access controls, NDAs, and legal oversight.

For full details of our security measures, please see our Trust Centre: https://trust.typeform.com/

2. AI Governance, Risk, & Compliance

Has your company implemented an end-to-end AI risk management framework?

Our AI-related risks are integrated into our existing general risk management processes. Typeform is ISO 42001 certified, the international standard for AI Management Systems, which provides independent validation that we maintain a structured framework for AI governance, risk management, transparency, and ethics. This sits alongside our SOC 2 Type 2 and ISO 27001 certifications. We continuously review our approach to managing these risks as the regulatory landscape evolves.

How is data privacy and security enforced when using third-party AI models?

Data privacy and security are enforced through contractual agreements and continuous oversight. Data Processing Agreements legally define the third party's data handling responsibilities and restrictions.

Security is maintained through technical controls, continuous monitoring, and annual audits to ensure controls are compliant and effective. These controls are independently validated through our ISO 42001, SOC 2 Type 2, and ISO 27001 certifications

Does your company's AI system perform input validation to detect anomalies and reject potentially malicious input?

Yes, our AI features, and the model providers themselves, include mechanisms to deflect malicious inputs at the model level. Typeform is continuing to work on observability and detection mechanisms to increase resilience to anomalous or malicious input.

How are AI-related incidents handled?

AI-related platform incidents are handled through our general incident response process, which covers all software and services. This process is in line with articles 34 & 35 of GDPR, and provides all the necessary steps to deal with security incidents, including investigation and notification according to the applicable laws. See Section 6 of the Data Processing Agreement: https://www.typeform.com/dpa for more information.

3. AI Lifecycle & Operations

Who is responsible for the development and training of the AI systems you use?

Typeform is responsible for the development of the AI system, but the models are developed by Anthropic and OpenAI. We do not train or fine-tune the LLMs.

How are AI features developed?

Our Secure Software Development Lifecycle governs the development and maintenance of all features, including AI. This process integrates security requirements into every stage of development, from the initial design to the final release, following industry best practices including NIST and OWASP.

Can the AI features be disabled?

Users control the use of all AI-powered features in the platform.

Administrators can disable Qualitative Insights at the account level. Other AI features are only used at the user's discretion, such as when they choose to add such a question to their form.

What happens if the AI produces an erroneous result?

Our current AI features do not have a separate automated failover system. Most features can be triggered again for a better result. We are actively evaluating and developing a more resilient architecture against unintended or harmful outputs for future releases.

Are end-users informed when AI is involved?

Yes, for full transparency, users are always informed when an AI feature is being used. This applies for both in-platform features visible to customers, and features visible to respondents on the forms themselves. The relevant Help Centre article provides screenshots of each feature to show exactly how this information is presented to the user.

Is AI central to the overall product?

AI is not required for the overall functioning of the product; it is included to power additional features to enhance the user experience.