Sana AI security whitepaper

Last updated: Mar 20, 2024

Table of contents

Introduction

Sana is committed to providing your organization with world-class search experiences — allowing authorized users to search for and retrieve answers specific to their company’s knowledge.

As an industry pioneer and the world's leading learning and knowledge platform provider, information security, legal compliance, and data privacy are top organizational priorities for us. Data privacy by design and by default are essential considerations when building our search capabilities for your organization.

The purpose of this paper is to outline our efforts and the attributes of Sana AI that address security, compliance, and privacy—so that end users can use our services with peace of mind.

Compliant with globally-leading standards and regulations

ISO 27001 CERTIFIED A3CERT

GDPR.EU

What is Sana AI?

Sana AI is a web-based natural language chat application that answers user’s queries with knowledge specific to their organization. Sana AI pulls this knowledge from a variety of integrated sources that have been enabled by admins and authorized by end-users.

How we partner with sub-processors and subcontractors

We carefully vet suppliers during the procurement process and only use suppliers for specific and necessary purposes to enhance Sana for our end-users. We expect the same technical and security measures from our suppliers as we uphold for ourselves. For our most critical sub-processors, we require ISO 27001 and SOC 2 Type 2 certification, and GDPR compliance.

All contracts with chosen suppliers address our demands on the supplier's IT environment and information security measures. Each supplier is obligated to account for their technology, routines, and processes as well as their IT and information security policies.

Non-disclosure agreements and other relevant regulatory agreements are signed by our suppliers before the service is taken into service, and we conduct regular control of suppliers' access rights and other aspects of the agreement with the supplier.

How we ensure business continuity

Data backup is one of the three pillars of our continuity plan. Trained personnel manage and follow up on backup execution to ensure the integrity, confidentiality, and accuracy of the backup data. Backups are stored for 30 calendar days.

The second pillar of our continuity plan is the IT and management processes and routines that are carried out when a serious incident occurs. We continually work on keeping processes and routines updated. The continuity plan is tested at intervals based on regular risk assessments.

Our third pillar is to have a high degree of digitization, and all the services and tools are digitally accessible using Google Accounts' SAML-based Federated SSO. As a result, most employees can continue to work from other locations even if our offices are closed or not accessible due to an extreme event.

How we uphold information security

We use an Information Security Management System (ISMS) certified under ISO/IEC 27001 as the basis for all information security measures. The ISO/IEC 27001 standard provides guidelines and general principles for planning, implementing, maintaining, and improving information security in an organization.

How we control access to our systems and processes

We prevent unauthorized persons from using systems and processes by adhering to the principle of least privilege and using role-based permissions when provisioning access to systems, and utilizing multi-factor authentication for access to systems with highly confidential data.

We prevent physical access of unauthorized persons to core systems by partnering with industry-leading data center and cloud infra providers. These providers equip their data centers with 24x7x365 surveillance and biometric access control systems. Additionally, all providers are ISO27001, ISO27017, ISO27018, SOC2 Type II, PCI DSS, and CSA STAR certified.

We prevent physical access of unauthorized persons to our physical office locations by using comprehensive physical and identity access management, consisting of redundant key-card access points, video surveillance, and 24/7 identity management. We also routinely provide effective, secure, and immediate onboarding and offboarding of employees, contractors, and third parties.

We ensure that persons authorized to use Sana have access only to data relevant for their access rights by utilizing leading password validation and recovery techniques, ensuring passwords are hashed and salted, and offering SSO to our partners. We also routinely conduct vulnerability scanning, malicious activity detection, and block suspicious behavior automatically. In addition, we also utilize firewalls to segregate unwanted traffic from entering the network.

We ensure that personal data cannot be read, copied, altered, or deleted by unauthorized persons during electronic transmission or during transport or storage on data media. We also ensure that customer data at rest is encrypted with AES-256, and data in transit is encrypted with TLS 1.2+. We are also alerted to encryption issues through periodic risk assessments and third-party penetration tests on an annual basis.

We ensure that we can immediately review and determine if and who enters, alters, or deletes personal data in our system by monitoring and logging events in a central store. Critical logs are retained for at least 2 months and with individual identification to identify non-conformities.

We provide that personal data are protected against accidental destruction or loss by saving full backup copies of production data daily, using a robust patch management process, and logically separating development, testing, staging, and production environments.

How we manage risk

We adopt appropriate risk management and security risk management controls such as conducting periodic reviews and assessments of risks, and monitoring compliance with our policies and procedures, and keeping an up-to-date risk mapping signed off by senior management.

How we secure operations

We ensure that the appropriate operations safeguard against malicious code by maintaining different systems and methods to protect the IT infrastructure, using active monitoring to ensure that antivirus scanners and spam filters are active and updated, installing the latest security updates and patches, and ensuring all employees take security training at least once a year.

How we uphold security with our staff

Our most critical resource is our people, and we aim to hire the best talent globally. In order to ensure our staff comply with the laws and regulations, as well as the terms and conditions of supplier and customer agreements, we require that Sanians conduct themselves in a manner consistent with our guidelines regarding confidentiality, business ethics, and professional standards. We also require our personnel to enter into confidentiality agreements, and acknowledge receipt of, and compliance with, Sana's confidentiality and privacy policies.

Frequently Asked Questions

How does the search algorithm learn?

We use a dedicated training set of internal data to manually and automatically train our ranking algorithm and query rewrite to find the most relevant matches. No customer data, e.g., external content indexed by our service (“**Content Data**”), are used outside of the isolated tenant, unless specifically agreed upon. For the point of clarity, no customer data is used to train third-party LLMs. To improve ranking, we also log the queries asked by users, and how a user interacts with the results or Sana AI, so that we can serve the types of queries users need the most in the best way possible. Such data is pseudonymized to ensure confidentiality.

Will Sana index our data?

Data added through integrations and/or through upload to Sana AI is indexed and stored on our cloud instance.

How does Sana ensure the security of our data?

Sana isolates all customer data using a single tenant architecture, meaning no databases are shared between customers. Data at rest is encrypted with AES 256 and data in transit is encrypted with TLS 1.2+.

Can every user see every document?

There are three ways documents can be added to Sana AI, (i) through direct upload, (ii) through a private integration, and (iii) with a shared integration. The accessibility depends on how the document has been added and admin settings.

For direct uploads, access management is handled directly in Sana AI through individual document access management or through collections. Admins can set up which users can do what.

For private integrations, each user is self-authenticating and can see results only from the scope they have access to. For example, a user can see the same results from Google Drive that they could see if searching directly from their authenticated Google Drive.

For shared integrations, admins can control which documents should be available to all users that have access to a collection and/or the entire workspace.

Will our data be used by third-parties to train their models?

Sana AI is built agnostic to the underlying large language models. Sana offers third party large language model options which are not trained on Content Data. We utilize enterprise security arrangements and, whenever possible, a Zero-Day Retention (ZDR) policy with our third-parties. No customer data is used to train third-party LLMs.

Can I, as an admin, enable integrations on behalf of my organization?

You can control which integrations are available to the organization, and which ones feed into the natural language response from Sana AI.

Will employees at Sana be able to see our content data and search queries?

Sana employees do not have access to your workspace by default, and will only be able to access it if you grant them access by extending an invite. Invited Sana employees cannot view any files without specific permission, and do not have access to search queries. Sana employees do have access to downvoted queries, but user information is anonymized. Developers have access to underlying databases, through stringent "least privilege" processes and audit trails, we ensure unauthorized access to your data is prevented.