AI Frequently Asked Questions

How does Concentric AI use artificial intelligence?

Artificial Intelligence is foundational to Concentric AI. We use AI to understand the nature of data assets within an enterprise and assess associated risk.

While AI is used in a variety of ways toward this end goal, the most significant AI-driven feature within Concentric AI's Semantic Intelligence™ platform is categorization. Each data asset in the client environment is scanned, and its contents are extracted. The extracted content is processed by our proprietary AI engine to group "semantically similar" files and determine the appropriate semantic category and subcategory. All extracted content, in its original form, is discarded after processing and is not stored on our platform.

The semantic nature of data is determined by the content itself, just as a person opening a document can recognize what it is (a resume, a pay stub, a performance appraisal, a non-disclosure agreement, etc.) rather than relying on rules, regular expressions, or manually configured heuristics.

Can you provide an example or analogy to further simplify this concept?

Consider the example of Google Photos or Apple Photos. One of the features of these apps is recognizing all pictures of the same person and grouping them. In addition, if you recognize the person and add their name, all their pictures will be tagged with the correct name.

Similarly, our AI engine recognizes content that has the same "meaning." Two or more data records may be written differently and use different verbiage but still address the same topic, making them "semantically" similar and grouping them together. We have built thousands of out-of-the-box models to recognize the appropriate category and subcategory for content related to common enterprise functions such as finance, legal, HR, etc. For content unique to a company's environment, similar documents will still be grouped together but may not be assigned the right label. Labeling as few as a single file within the platform with the right category/subcategory will automatically tag all similar files.

Is Semantic Intelligence a "generative AI" tool?

Concentric AI's Semantic Intelligence platform is not a "generative AI" tool in the sense that it does not "generate" new or novel content. The AI engine is deeply integrated into the product's functionality, and the product does not offer a stand-alone way for users to interact with it directly.

Does Concentric AI process customer data using third-party AI providers such as OpenAI?

No. Concentric AI does not use third-party AI services to process customer data. All customer data is processed using models owned and hosted by Concentric AI within our SaaS infrastructure.

Is your AI trained with customer data?

Customer data is not used to "train" Semantic Intelligence's categorization models. These models have been trained on public datasets as well as custom datasets collected from public sources and curated by Concentric AI.

When unique types of data records are found within a customer's environment, the AI engine first identifies groups of similar data records. Attaching an appropriate category and subcategory label is akin to giving a name to a new person found in a photo repository in the prior example. It is not "training" a new model but simply associating the correct label with what has already been determined as a grouping of similar data records.

What happens when a user labels a group of semantically similar files with category and subcategory labels?

When a user labels a set of files, all similar files – whether already created or created in the future – will automatically be tagged with that label without requiring rescans of existing data. By default, this tagging occurs only within the client's environment.

If similar files are found in other customer environments and some of those files are also labeled with the same category and subcategory independently, those models will be deployed across all tenants as "cross-tenant" models. Note that client data is never commingled or shared between clients.

Are there risks to customers' sensitive data because of AI processing?

No. Semantic Intelligence extracts data and processes the contents of data records to understand semantic context and help discover and remediate data risks. Once the data is processed, raw content is discarded. Sensitive information, such as PII detected in the environment, is stored only in redacted form (except for names and emails). In addition, Concentric AI adheres to all applicable regulations and is a SOC 2 Type II and PCI DSS compliant vendor.

What data was the AI trained on?

Public datasets as well as custom datasets collected from public sources and curated by Concentric AI.

What technical or organizational measures are taken to protect the security of the AI engine?

We do not store the full values of PII and other detected sensitive elements (except names and email addresses); we store only masked values in our backend database.

Our AI engine processes raw data for categorization, and the content is discarded after processing.
The AI engine is designed so that models are not sensitive to specific values of any PII present in the content being processed, and therefore do not pose any further privacy risks.

Our AI models are, by design, representational only and are not generative; no data processed by the AI engine can be regenerated in any form.

Concentric AI follows a least-privilege access model for production systems, maintains an auditable log of all activity, and is PCI-DSS and SOC 2 Type II compliant. All users receive onboarding training and ongoing security training.

Who owns the rights to the inputs and outputs of your AI?

The AI engine processes data scanned from within the client's environment, and the ownership of all data elements will not change - the rights remain with the client. The AI engine processes content to generate insights that are displayed in Semantic Intelligence, and the client owns the rights to all insights created within the client's tenancy on our platform.

Do you follow responsible AI principles? How do you address bias, safety, and explainability?

Addressing bias, safety, and explainability applies to any automated decision-making system.

We have designed all features in our product to be explainable and understandable, so users can trust and confidently leverage our platform. Any risk determinations are accompanied by clear explanations of the reasons. Semantic categories are determined by content similarity and are thus straightforward to understand. For each file, users can see a list of similar files within the client's environment and should be able to validate and verify by opening and reviewing the document itself or any of its peers.

The issue of bias is not applicable to content categorization. Risks associated with users and sensitive data can be assessed based on user activity, permissions, job roles, and other business-related information. No personal information about individual users, such as age, ethnicity, nationality, gender, sexual orientation, or religion, is ever used or considered for such risk assessments. Furthermore, users of Concentric AI's platform have complete control over the recommendations and assessments provided.

We employ multiple guardrails to ensure the safety of our AI deployments to minimize unintentional or malicious use. All aspects of model development and deployment are continuously monitored and periodically reviewed. All code changes are reviewed and approved and go through a comprehensive suite of tests across multiple internal test environments. All activity in production environments is monitored and logged for audit purposes. The chief scientist has ultimate responsibility and accountability for decisions relating to the development and use of AI.