Metanomos-AI Platform: User Guide

1. Introduction

This User Guide is designed to help first-time users understand and navigate the Metanomos AI platform. The platform enables organizations to convert unstructured and semi-structured documents into a structured, intelligent Knowledge Graph through ingestion, schema management, conversational exploration, and entity similarity detection.

Metanomos AI simplifies the end-to-end process of transforming raw documents into actionable insights by allowing users to upload data, define or reuse schemas, explore extracted knowledge through natural language conversations, and maintain data quality using similarity-based entity management.

This guide provides a clear walkthrough of the platform's primary features, ensuring new users can begin working efficiently without prior technical expertise. Each section explains the purpose of a capability, when it should be used, and the expected outcomes.

2. Document Ingestion

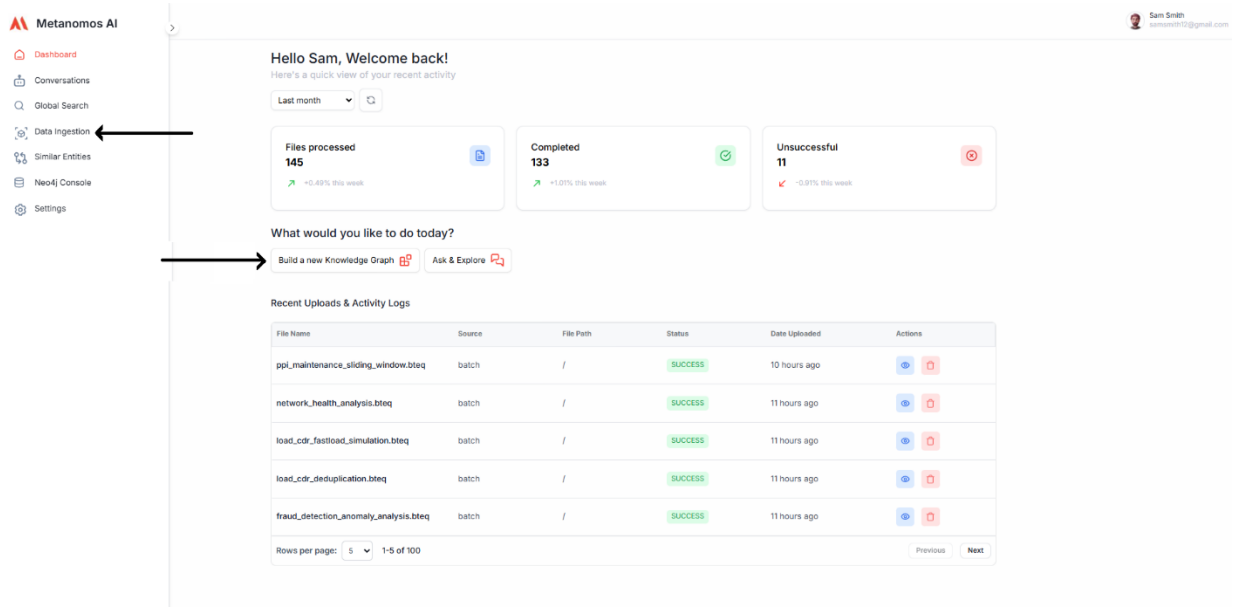

Figure 2.0: Entry points for initiating document ingestion

The Document Ingestion module is the starting point for building your Knowledge Graph. It allows you to upload files, apply schemas, organize documents into collections, and prepare your data for analysis and querying.

Before exploring schema workflows, the **ingestion can be accessed** within the platform from **two primary entry points**, ensuring users can initiate ingestion regardless of their workflow context.

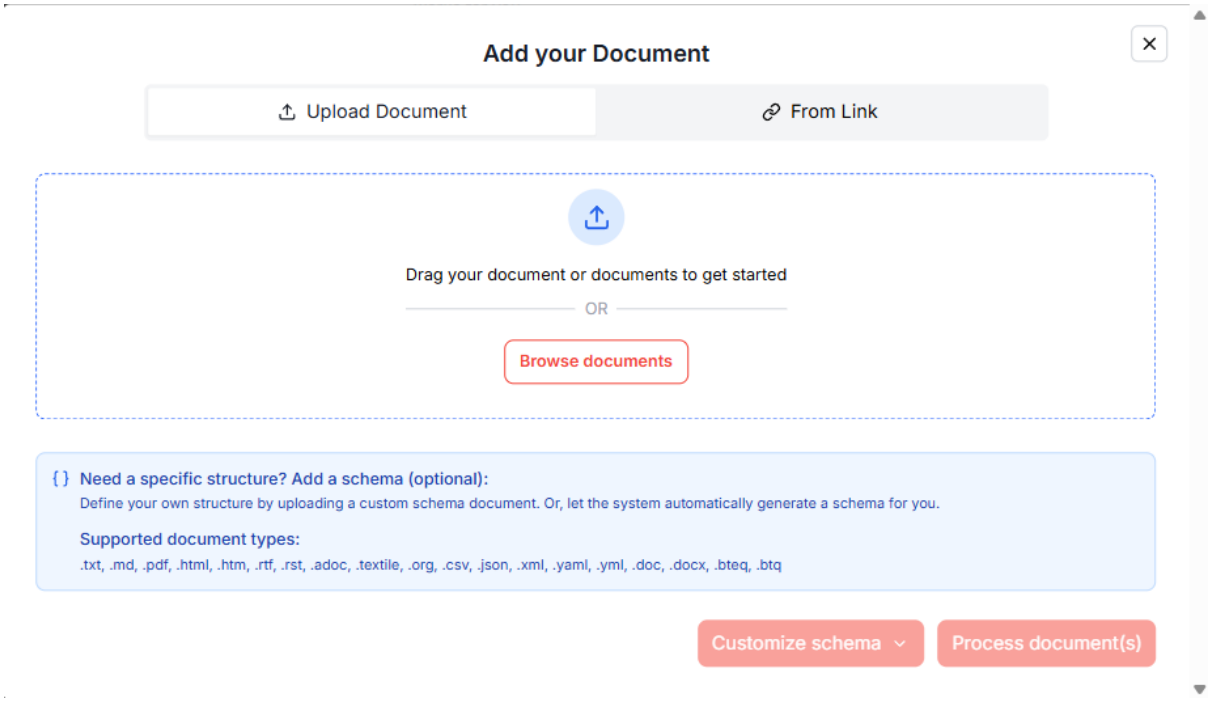

Figure 2.1: Quick Action to Build a Knowledge Graph

The **Add your Document** interface is the primary workspace for uploading files into the platform. Users can upload a single document, customize schemas if needed, and begin the ingestion process. The interface is designed to support a wide range of document types and provides flexibility for both schema-driven and schema-free ingestion.

Upload Options

The platform provides two options for selecting the document source:

- **Upload Document**
  Allows users to upload files directly from their local system.
- **From Link**
  Enables ingestion of documents accessible via a public or authenticated URL.

Schema Options

Below the upload area, users can define a schema for the uploaded files. Schemas provide structure to the data being extracted. There are **three options**:

- **Customize Schema** – Upload or define your own custom schema document.
- **Auto-Generate Schema** – Allow the system to infer and generate a structure automatically.
- **Skip / No Schema** – Proceed without providing a schema; appropriate for exploratory or unstructured ingestion.

The interface clearly explains that schema configuration is optional and only needed when users require structured extraction.

Supported document types include :
.txt, .md, .pdf, .html, .htm, .rtf, .rst, .adoc, .textile, .org, .csv, .json, .xml, .yaml, .yml, .doc, .docx, .bteq, .btq

2.2 Ingesting a Document — Without Schema :

Ingesting a document without a schema uploads the file to the platform using the default, schema-free extraction flow. Use this option when you want to store and search the document content quickly without defining any structured fields.

2.2.1 Step-by-step workflow :

Add one or more files by **drag & drop** or click **Browse documents**.

Click **Process document** to start ingestion.

Confirm: the file appears in **Recent Uploads** with status **SUCCESS** (or check processing logs).

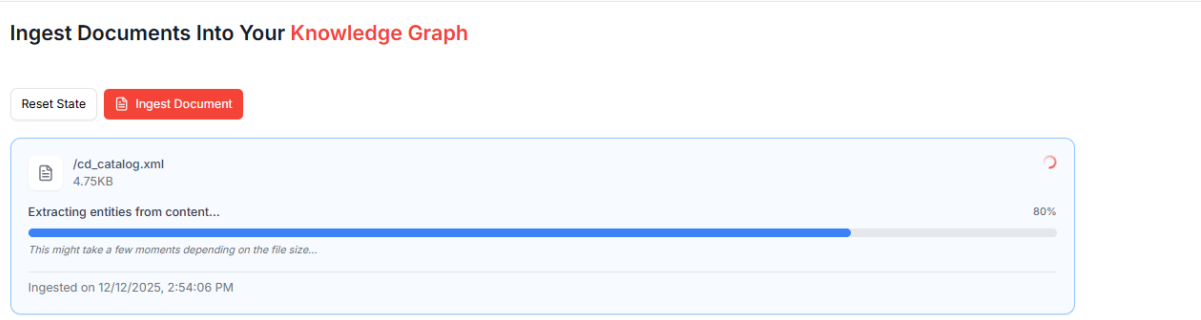

Figure 2.2 Ingestion Progress Screen

2.3 Ingesting a Document With a Schema :

Ingesting a document with a schema allows the platform to extract structured information based on a predefined or system-generated data model. When a schema is applied, the system identifies entities, attributes, and relationships more accurately during ingestion. On the Add your Document screen, users can either upload their own custom schema or allow the system to automatically generate one before processing the document.

2.3.1 Ingesting a Document With a Custom Schema :

If you already know how to design your own schema, you can upload it here and instruct the system to extract information exactly the way you want. After selecting your document on the ingestion screen, click **Customize Schema** and upload your schema file. Once the schema is loaded, you will see the option to enable **Strict Schema Matching**, which controls how closely the system must follow your schema during extraction.

When working with a custom schema, you can choose:

- **Strict Schema Matching ON** – The system extracts only the fields, entities, and relationships defined in your schema. Nothing outside your schema will be captured.
- **Strict Schema Matching OFF** – The system prioritizes your schema but also extracts additional relevant information that it identifies in the document.
- **Schema Saved** – Indicates that your schema has been accepted and will be applied during ingestion.

After setting your preferred matching mode, click **Process Document** to ingest the file. This method is ideal when you want complete control over how the document should be structured within the Knowledge Graph.

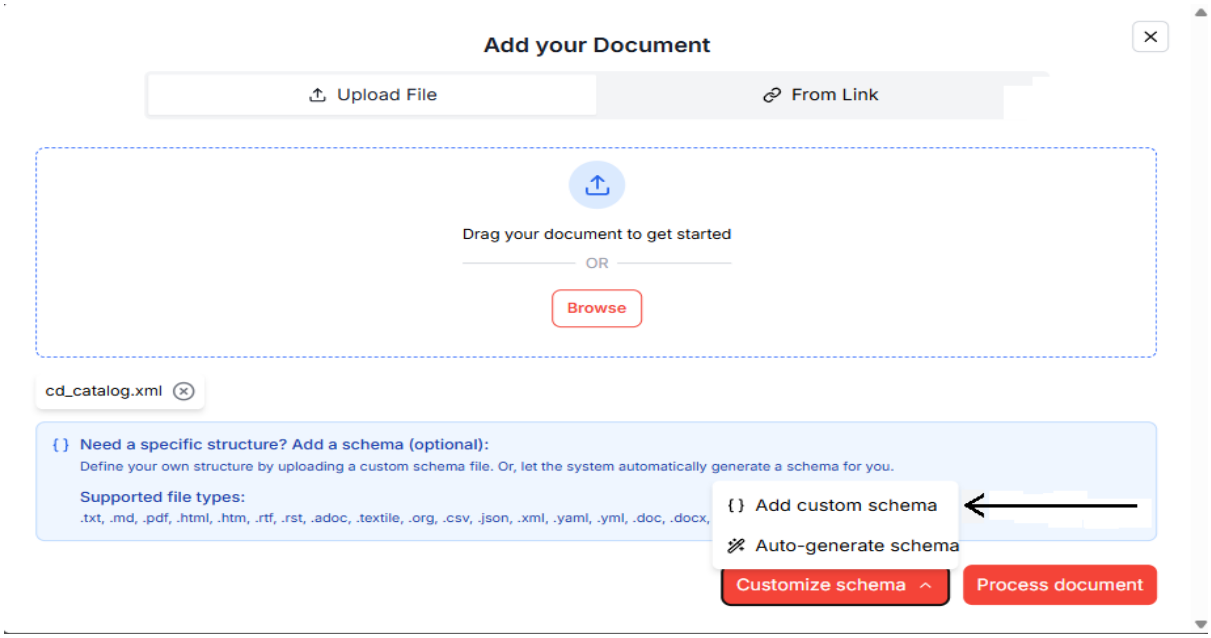

Figure 2.3: Adding a custom Schema

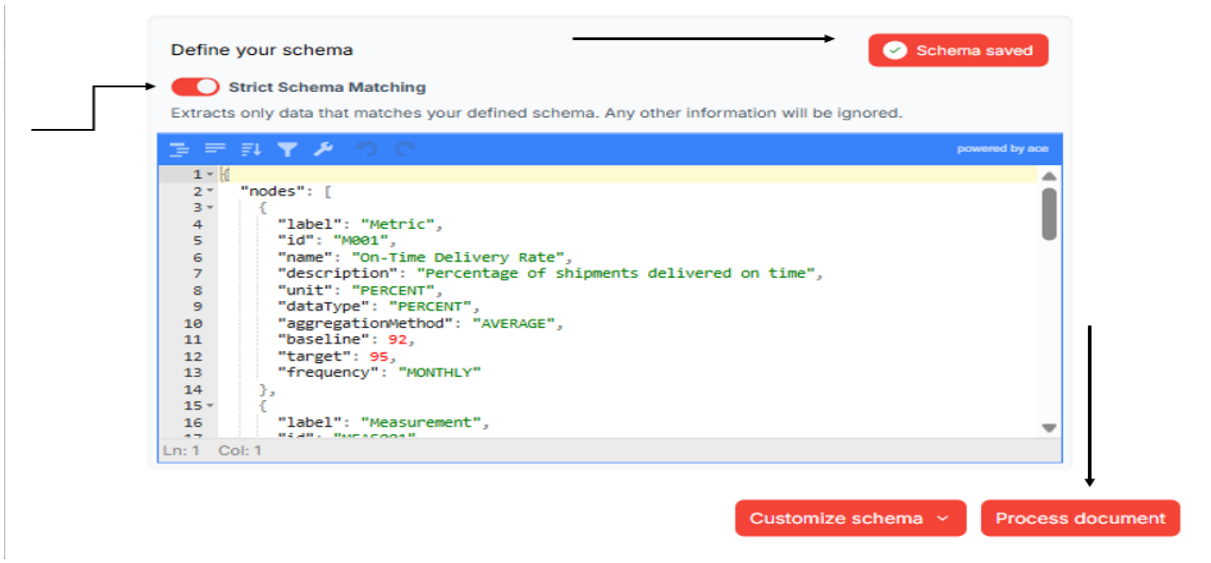

Figure 2.4: Processing the Custom Schema

2.3.2 Auto-Generated Schema :

If you are not sure what your schema should look like or you prefer the system to determine the structure for you, you can use the **Auto-Generated Schema** option. When this option is selected, the platform asks you a short set of preference questions to better understand your goals, the type of information you care about, and how you plan to use the extracted data. These questions help the system tailor the schema to your specific needs rather than producing a generic layout.

As you move through the three-step panel ("Help Us Understand Your Data"), the system uses your selections to guide the schema design. Based on your answers, the platform's LLM automatically analyses the document's content and generates a structured schema containing relevant nodes, properties, and relationships. This ensures the extracted data aligns with your intended use—whether it is insights, metrics, compliance, reporting, or another focus area.

Once the schema is generated, it is displayed for your review. At this stage, you can also choose whether to enable **Strict Schema Matching**, which determines how closely the ingestion pipeline should follow the generated schema.

- **Strict Schema Matching ON** – Extraction strictly follows the generated schema and ignores any additional patterns not defined in it.
- **Strict Schema Matching OFF** – The system prioritizes the generated schema but still extracts additional meaningful information from the document.

After reviewing the schema and selecting your matching preference, click **Process Document** to ingest the file using the automatically generated structure.

This option is ideal for users who want structured extraction without manually defining schema rules, allowing the system to adapt to different document types and analytical goals.

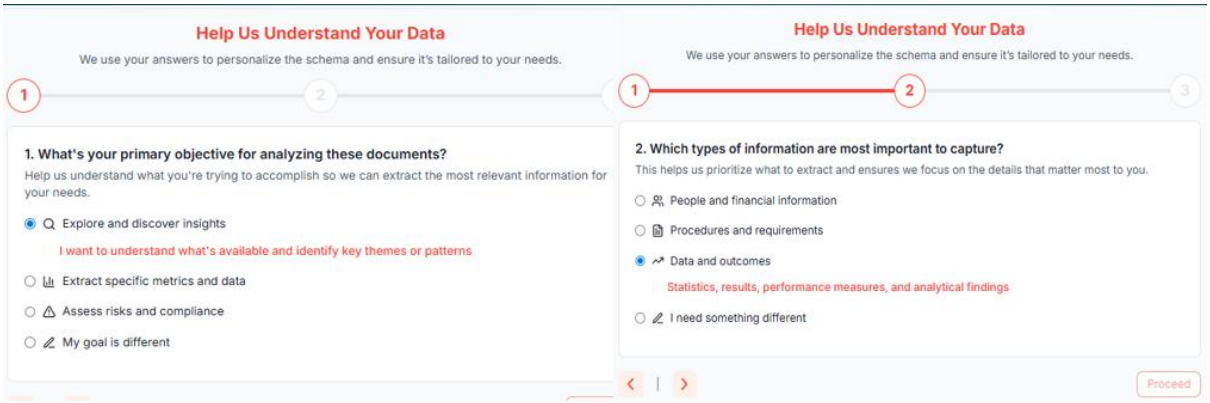

Figure 2.5: Auto-Generated Schema Preference Panel

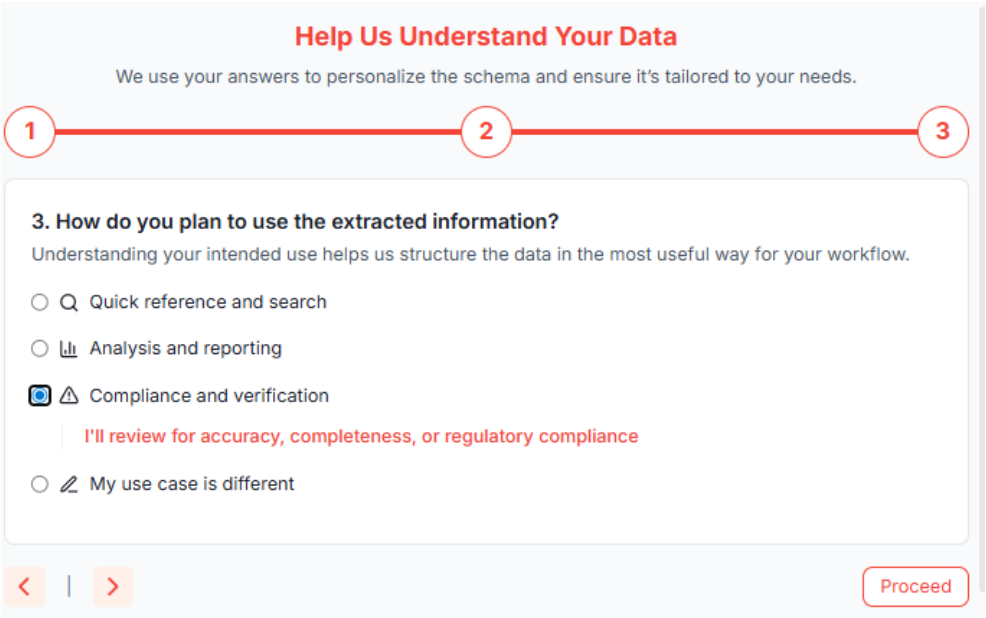

Figure 2.5.1: Auto-Generated Schema Preference Panel

2.4 Using an Existing Schema at the Collection Level :

Collections allow you to group documents that follow the same structure, making it easier to maintain consistency across multiple ingestions. When creating a new collection, you can assign a custom schema to it. This schema becomes the default structure for any document ingested into that collection. To set this up, provide a clear **Collection Name**, optionally include a description to indicate its purpose, and click **Add Custom Schema** if you want the collection

to enforce a specific data model. Once saved, the collection becomes available in the “My Collections” panel.

Using a collection-level schema is helpful when you already know that multiple documents will follow the same template or require the same extraction rules. Instead of configuring a schema every time you upload a document, you simply ingest files directly into the collection, and the assigned schema is automatically applied during processing. You can add documents to a collection in two ways:

- **Drag and Drop:** Drag your document over the desired collection name and drop it. The system redirects you to the **Add your Document** page, where the **collection’s schema is already preloaded** and ready for ingestion.
- **Use the + Button:** Next to each collection name, a **+ icon** allows you to add a new document directly to that collection. Clicking this button opens the upload interface with the collection’s schema pre-selected.

This workflow ensures that all documents within the collection follow the same schema, keeping your Knowledge Graph consistent and reducing the need to repeatedly configure schema settings for similar documents.

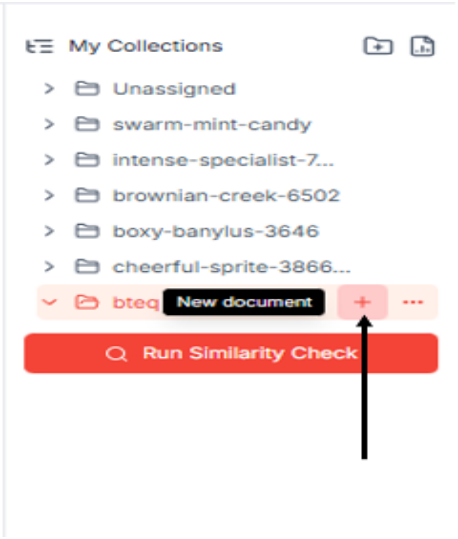

Figure 2.6: Ingesting a Document into a Saved Collection

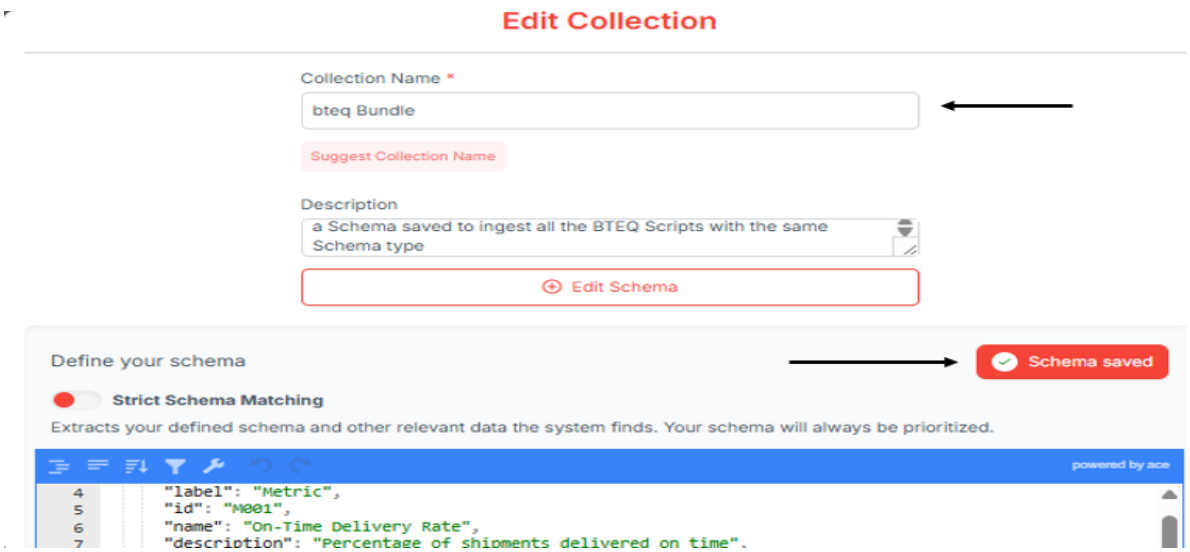

Figure 2.6.1: Editing a Collection and Managing Its Schema

3. Chat

The **Chat** module allows users to ask questions in natural language and receive structured, data-driven answers based on the documents and scripts they ingested. It supports different search modes to match different types of analytical needs.

3.1 Chat Search Modes :

Users can choose how the system processes their questions by selecting one of the available search modes:

- **Auto Mode**
  - Automatically selects the best search strategy based on the question type.
  - Ideal for new users or general exploratory queries.
- **Cypher Mode**
  - Uses direct graph queries to fetch precise information.
  - Best when users want exact facts from the extracted Knowledge Graph.
  - The system also displays the Cypher query used (via **View Cypher**), ensuring full transparency.
- **Hybrid Mode**
  - Combines natural language understanding with graph data retrieval.
  - Useful for questions that span multiple scripts or require interpretation beyond simple graph lookups.
  - Hybrid mode provides well-structured summaries, tables, and explanations based entirely on ingested data.

3.2 How Chat Presents Answers :

Each answer includes:

- Relevant extracted operations or SQL patterns
- Names of scripts involved
- Key functions or logic found
- A short summary so users can quickly understand the result

At the bottom, source scripts appear as clickable references, helping users trace where the information came from.

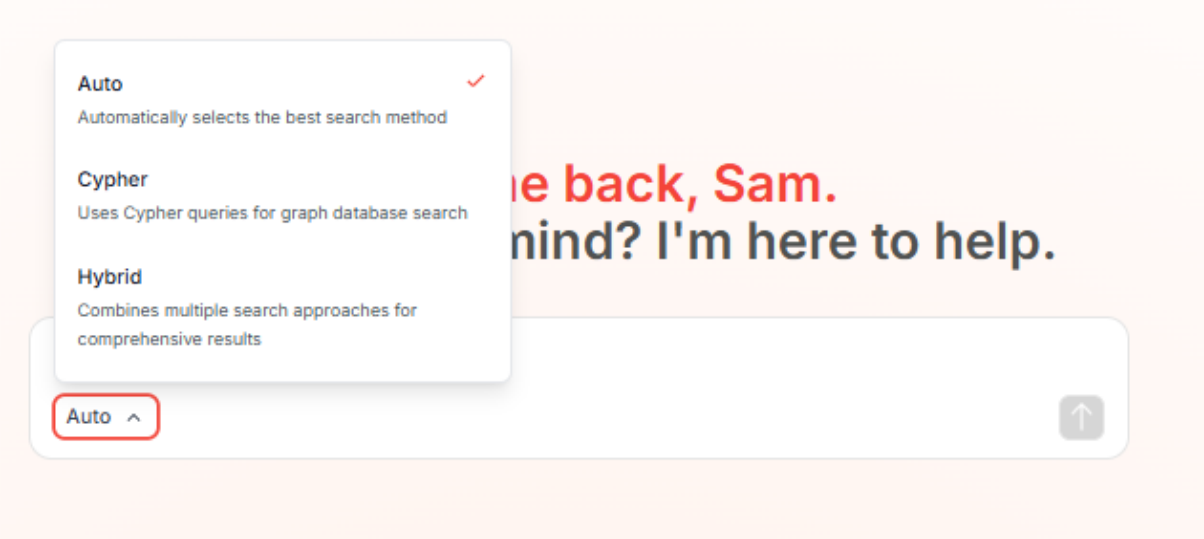

Figure 3.1: Chat Interface to Choose Different Modes

3.3 Understanding Hybrid Mode :

Hybrid Mode is designed to provide broader, more contextual responses. It:

- Interprets the intent of the question
- Searches across relevant scripts and extracted entities
- Organizes the information into a structured, readable format
- Surfaces patterns, relationships, and insights found in the data

This makes Hybrid Mode suitable for exploratory or analytical questions where users want a more complete understanding of what exists in their data.

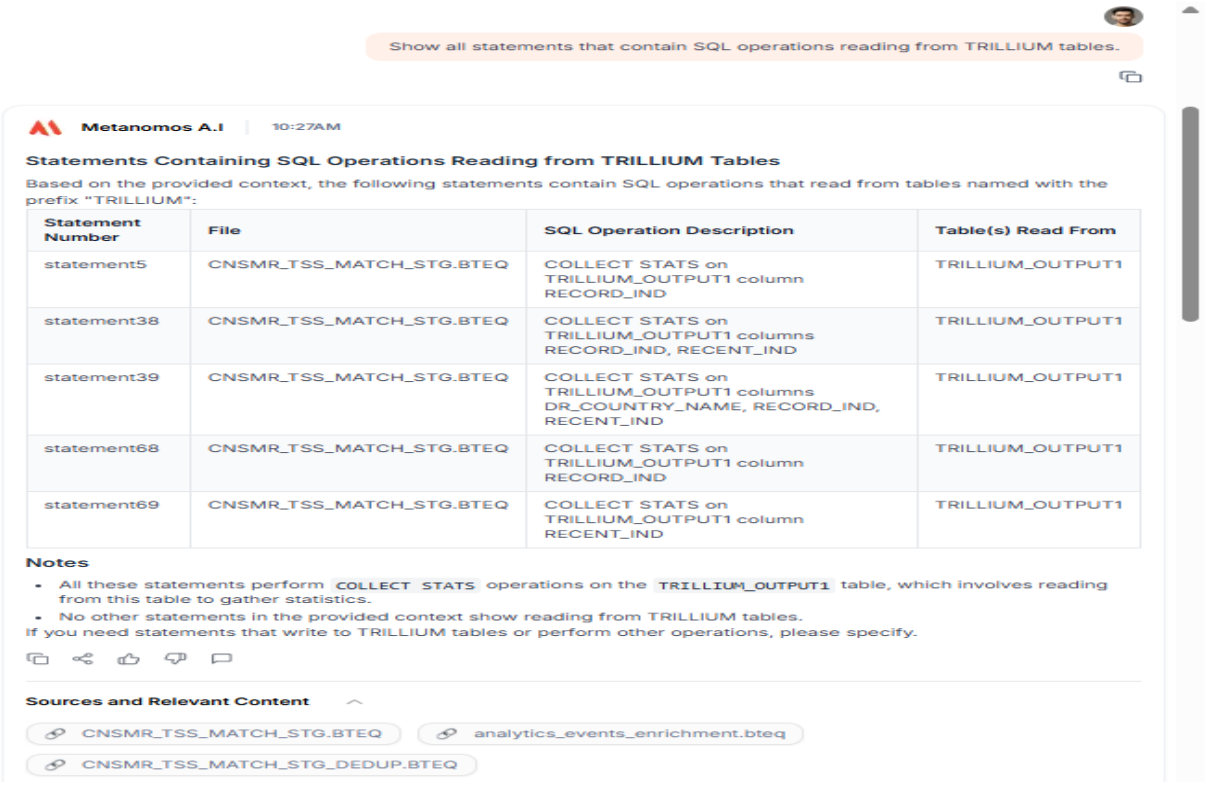

Figure 3.1: Chat Interface with Hybrid Search Mode

3.4 Understanding Cypher Mode :

Cypher Mode focuses on accuracy and structure. It:

- Retrieves only information explicitly present in the Knowledge Graph
- Produces answers that follow a strict JSON or structured format
- Ensures no extra interpretation or assumptions are added
- Shows the Cypher query used (when applicable), offering full traceability

This mode is ideal when the user requires consistency, precision, or technical clarity.

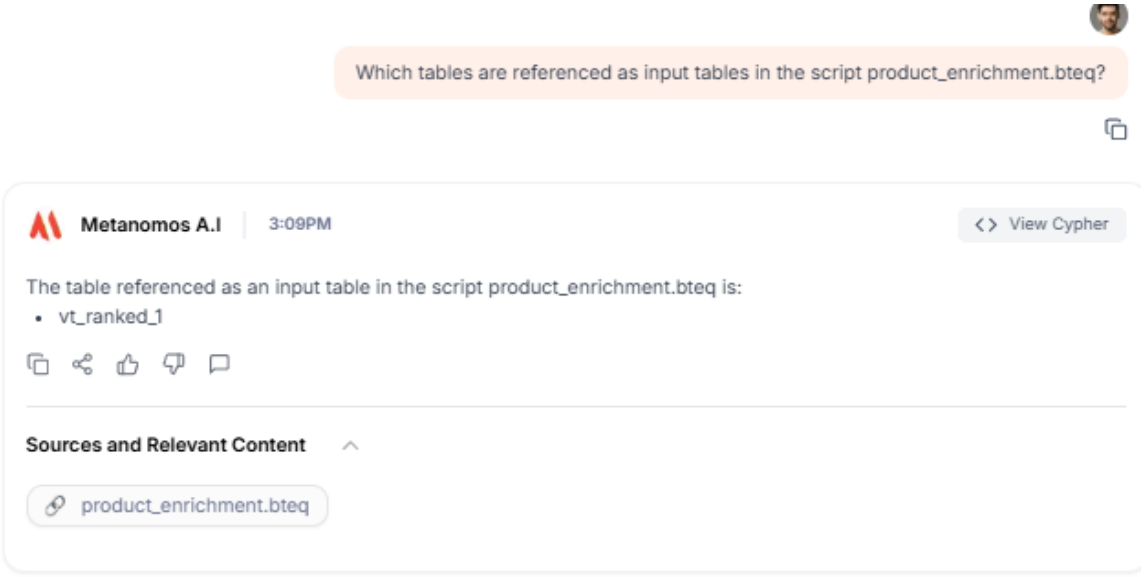

Figure 3.2: Chat Interface with Cypher Search Mode

Editing the Generated Query :

For users who are familiar with query syntax, the **Edit Query** option allows further refinement of the generated query to obtain more precise results. When modifying the query, it is important to also update the corresponding question before saving. This ensures the system treats it as a new request and returns an answer based on the updated query logic.

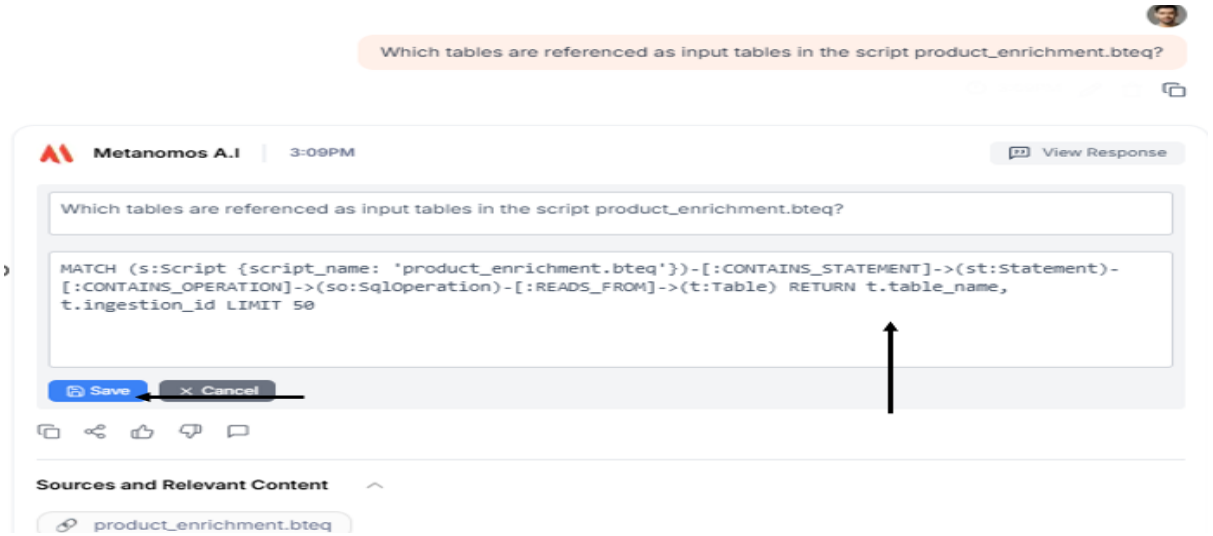

Figure 3.3: Edit Query Option in Chat

4. Similar Entities

The **Similar Entities** section helps users review and manage entities that the system identifies as potentially referring to the same real-world object. This feature is used to improve the quality and consistency of the Knowledge Graph after document ingestion.

4.1 Review Similar Entities :

- Displays a list of **similar entity pairs** detected from ingested files.
- Each entity pair shows:
    - Entity type (for example: Server, Model, Person, Technology)
    - Match score (similarity percentage based on canonical Score)
    - Confidence level
    - Review status (such as *Pending Review* or *Auto Approved*)
- Users can switch between:
    - **All Pairs View** for a summary list
    - **Detailed View** for deeper inspection of each entity pair

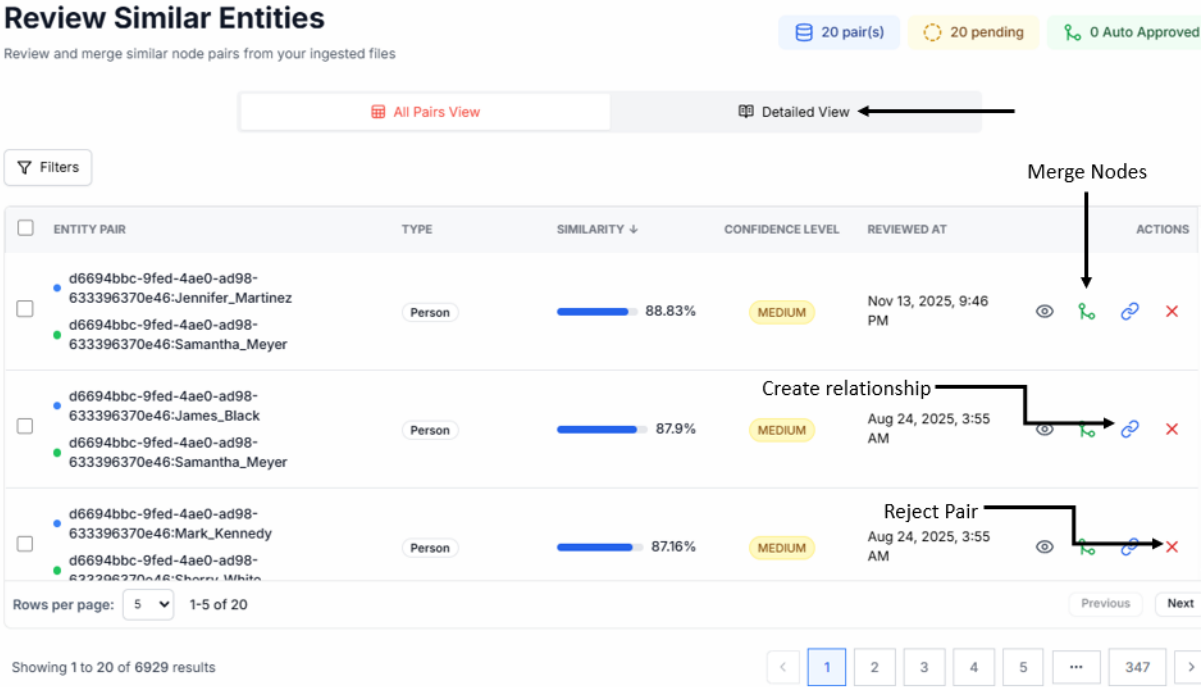

Figure 4.0: Similar Entities – All Pairs View

4.2 Detailed View of Entity Pairs :

When an entity pair is selected, the detailed view shows:

- A **Detailed Match Report** with:
    - Match score
    - Confidence level
- Side-by-side details of **Entity 1** and **Entity 2**, including:
    - Entity ID
    - Entity type
    - Entity name
    - Canonical score

This view helps users clearly understand why two entities were identified as similar.

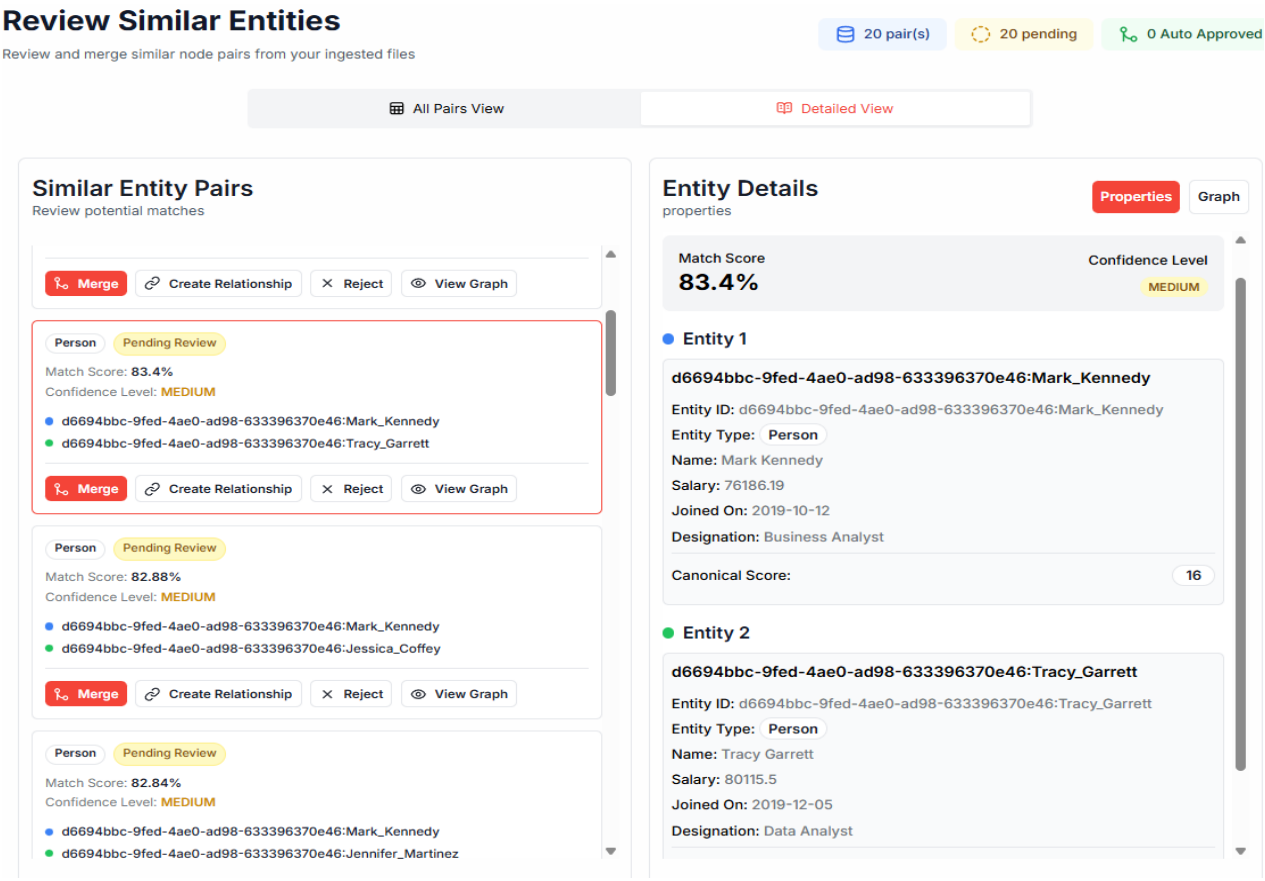

Figure 4.1: Similar Entities – Detailed View

4.3 Actions Available for Similar Entities :

For each similar entity pair, users can choose one of the following actions:

- **Merge**
    - Combines the two entities into a single entity.
    - Used when both entities represent the same real-world object.
- **Create Relationship**

- Links the two entities while keeping them separate.
- Users can provide a **relationship name** (for example, *SIMILAR_TO*).
- Optional settings allow users to:
  - **Invert** the relationship direction
  - Make the relationship **bidirectional**
- Saving creates the relationship in the Knowledge Graph without merging entities.

- **Reject**
  - Marks the entity pair as not similar.
  - Prevents the same match from being suggested again.
- **View Graph**
  - Opens a visual graph view to inspect how the entities are connected.
  - Helps users review context before taking action.

4.4 Graph View :

- Provides a visual representation of the selected entities and their relationships.
- Helps users verify connections and context before merging or linking entities.
- Useful for understanding how entities interact within the Knowledge Graph.

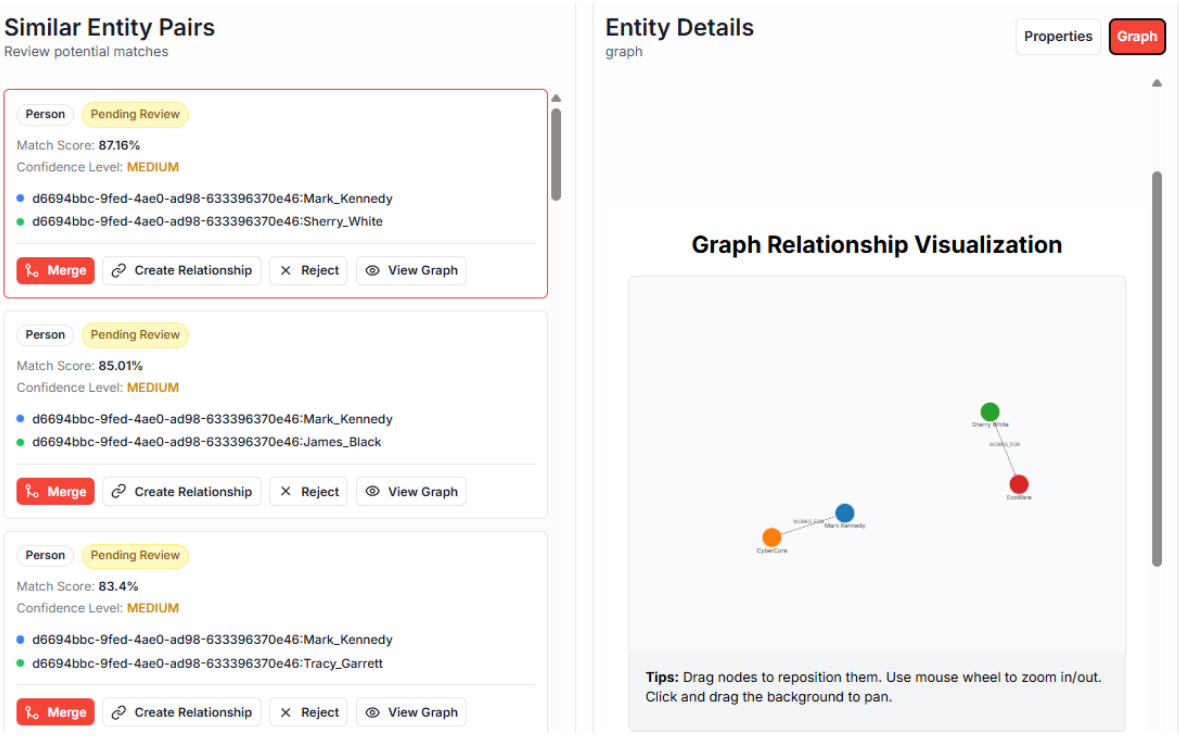

Figure 4.2: Entity Details – Graph View

4.5 Review Status and Tracking :

- Each entity pair includes a **review status**, such as:
    - Pending Review
    - Auto Approved
- Review timestamps and batch details are displayed for traceability.
- This helps teams track what has already been reviewed and what still requires action.

5. Settings

The **Settings** section allows users to configure system-level components that influence how data is queried and how Chat responses are generated. These settings affect how the platform connects to the Knowledge Graph and how the AI processes and formats answers.

5.1 Neo4j Configuration :

- Defines the connection to the **Neo4j Knowledge Graph**.
- Used by the system to retrieve nodes, relationships, and graph-based data.
- Ensures Chat and other features query the correct graph backend.
- Changes here impact how structured data is fetched from the Knowledge Graph.

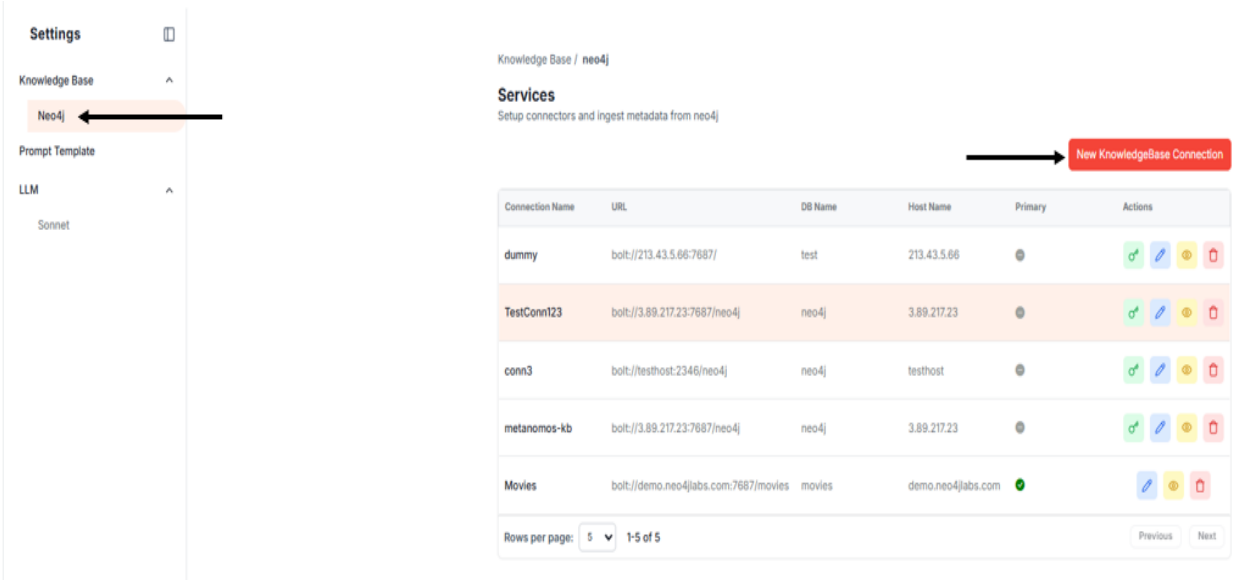

Figure 5.0: Settings – Neo4j Configuration

5.2 Adding a New Neo4j Connection :

Users can add a new Neo4j connection directly from the Settings page.

This is useful when connecting to a different Neo4j instance or environment.

To add a new connection, provide the required connection details such as:

- Neo4j URL
- Username
- Password

Once saved, the new connection becomes available for use by the system.

The selected Neo4j connection will be used for querying graph data across the platform.

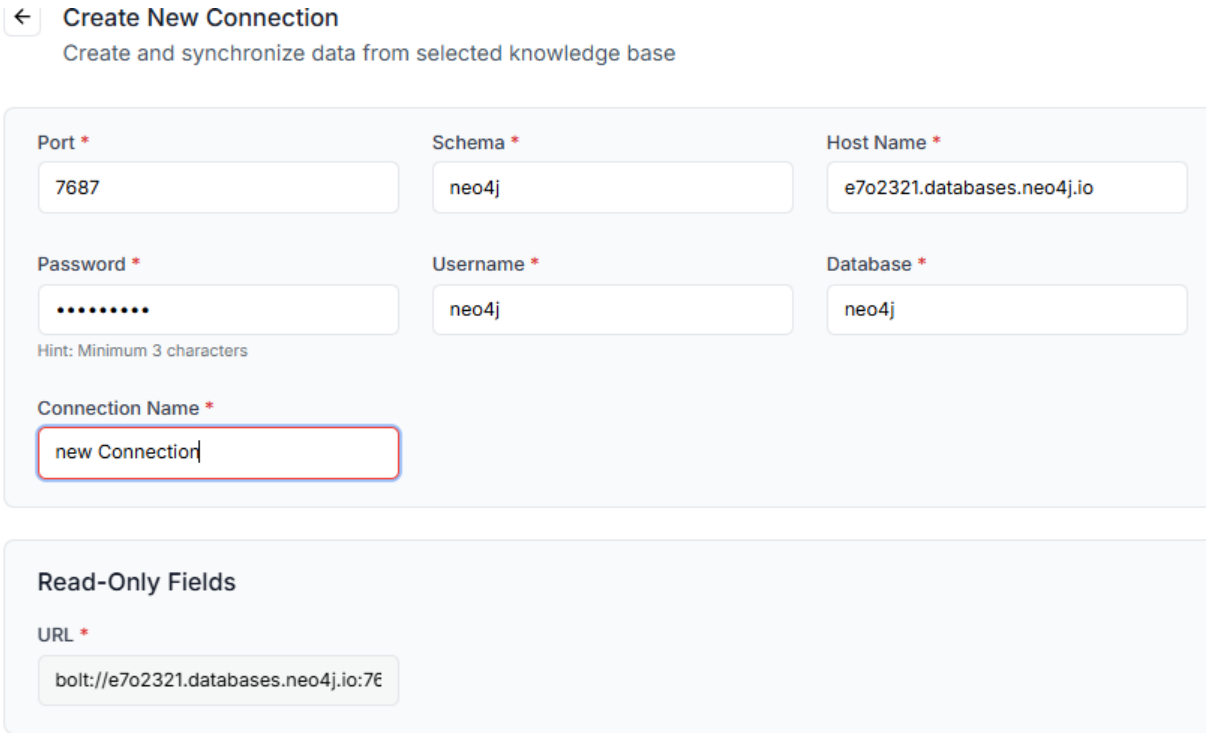

Figure 5.0.1: Settings – Neo4j Configuration edit Connection

5.3 Prompt Template :

The Prompt Template setting allows users to define how Chat queries are interpreted and answered. It helps guide the system by providing structured instructions and reference information that improve query understanding and response quality.

- Users can create new prompt templates by providing:
  - A name to identify the template
  - An overview describing the purpose of the template
- Multiple prompt templates can exist at the same time.
- One template can be marked as Primary, which means it will be used by default during Chat interactions.

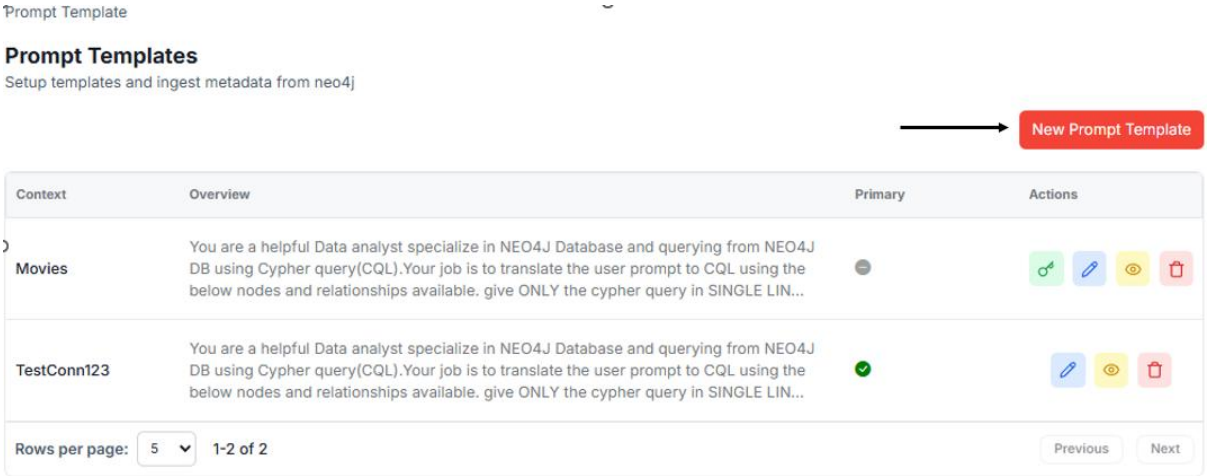

Figure 5.1: Prompt Template Configuration

5.3.1 Managing Q&A Pairs in Prompt Templates :

Prompt templates allow users to guide how Chat interprets and responds to questions by adding Question and Answer (Q&A) pairs. These Q&A pairs provide reference context that helps the system better understand query intent when Chat is invoked. Users can control which Q&A pairs are applied, ensuring that only relevant guidance is used. The template marked as Primary is applied by default during Chat interactions, helping maintain consistent and focused responses.

- Users can create multiple prompt templates and mark one as Primary.
- Q&A pairs act as contextual references for generating Chat responses.
- Only selected Q&A pairs are included when a prompt template is used.
- This approach improves response consistency without changing ingested data or schemas.

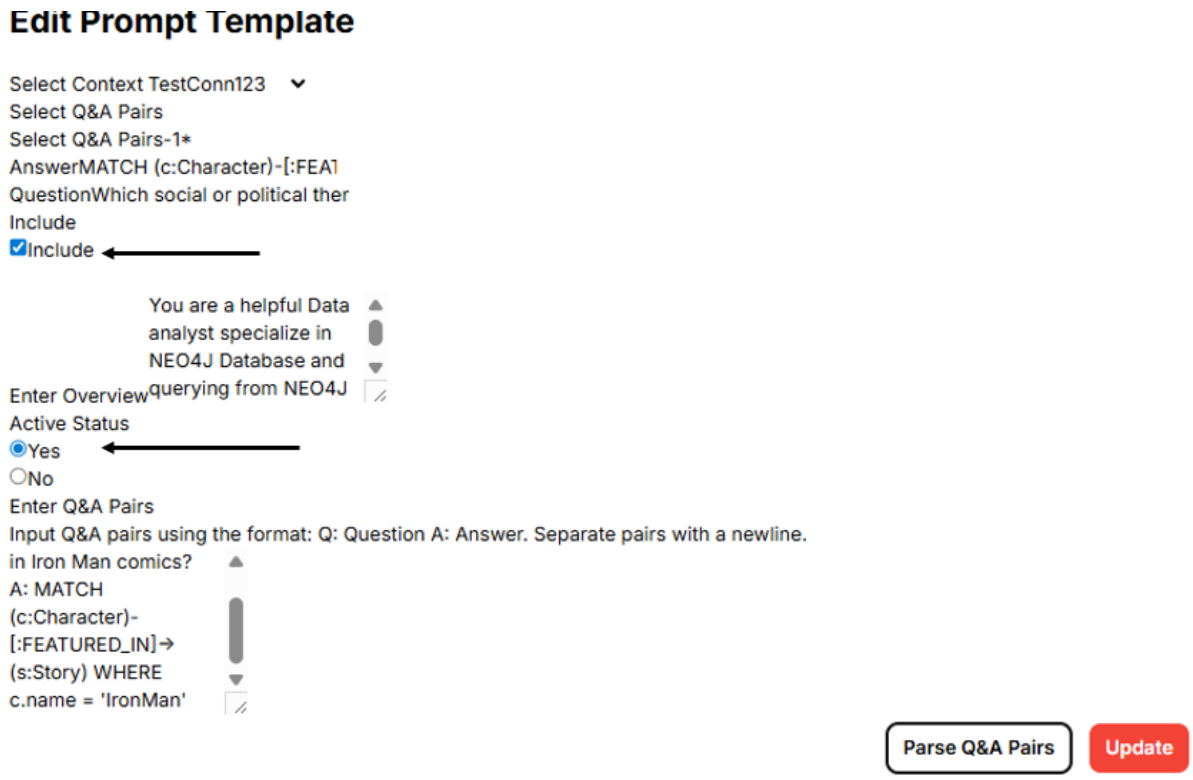

Figure 5.1.1.: Edit Prompt Template

5.4 LLM Configuration :

The LLM Configuration section allows users to control how Chat responses are generated by selecting and configuring the language model used by the system.

- Determines how user questions are interpreted and answered.
- Applies globally across the Chat feature.
- Changes do not affect previously ingested documents or the Knowledge Graph.

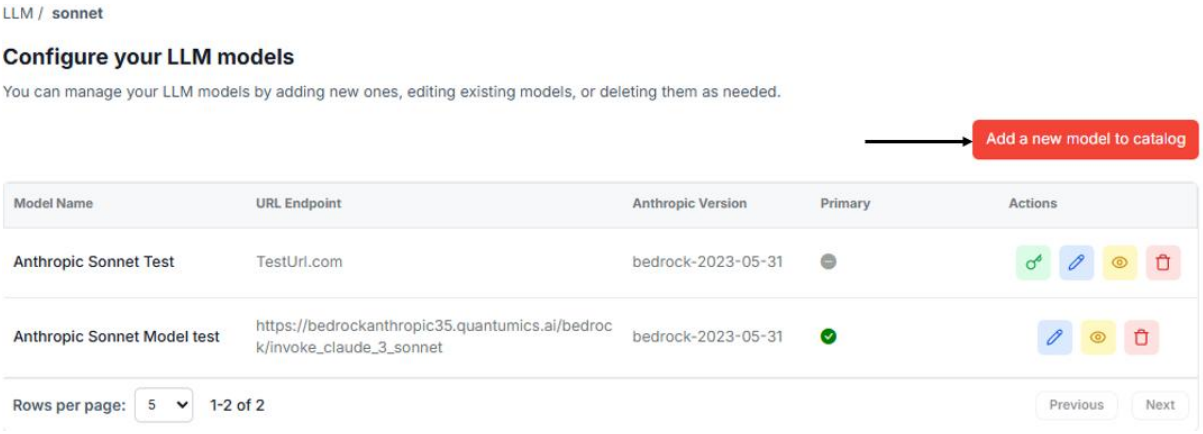

Figure 5.2: LLM Configuration Panel

5.5 Configuring LLM Settings :

This section allows users to adjust model-specific settings that influence how responses are generated.

- Users can **configure available LLM parameters** provided in the settings panel.
- These settings control aspects such as response behavior and output consistency.
- Configuration changes take effect immediately for future Chat queries.
- Adjusting LLM settings helps fine-tune how detailed, concise, or structured responses appear.

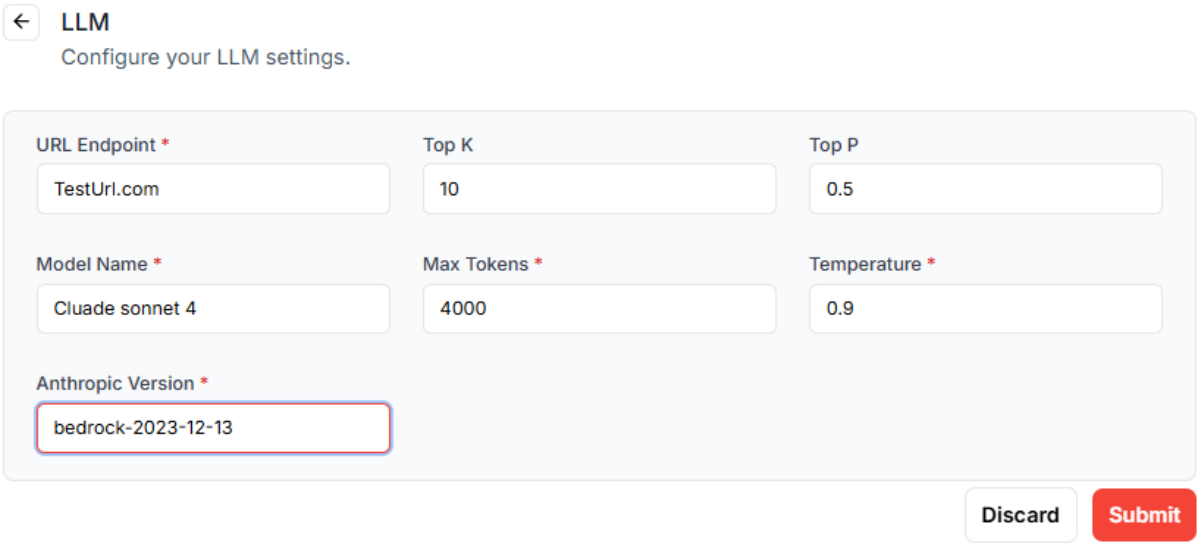

Figure 5.2.1: Edit LLM Configuration Panel

Implication :

Updating LLM settings changes how Chat behaves going forward, but it does not modify stored data, schemas, or existing graph structures.