

COMPASS RAG-Powered AI Assistant Guide

What is RAG?

Retrieval Augmented Generation (RAG) is a powerful AI architecture that combines the capabilities of large language models with the ability to retrieve specific information from your data. In COMPASS, RAG enables the AI assistant to provide contextually relevant responses based on your personal task management data.

How RAG Works in COMPASS

1. **Indexing:** Your task data, todos, calendar events, and other important information are securely indexed in a vector database (ChromaDB).
2. **Retrieval:** When you ask a question, the system finds the most relevant information from your data.
3. **Context Building:** The retrieved information is combined with your query to create context-rich prompts.
4. **Generation:** The AI assistant uses this enhanced context to generate accurate, personalized responses.

Available AI Capabilities

COMPASS provides domain-specific AI assistance across different areas of the application:

Task Management AI

Task-specific AI capabilities include:

- **Task Retrieval:** Find and summarize information about specific tasks
 - Example: "Show me details for my website redesign task"
- **Task Analysis:** Get insights about your tasks and workload
 - Example: "Analyze my current task distribution by priority"
- **Task Planning:** Get assistance organizing and scheduling tasks
 - Example: "Help me plan my tasks for the next sprint"
- **Task Summarization:** Generate concise overviews of tasks
 - Example: "Summarize my project tasks by status"

Todo Management AI

Todo-specific AI capabilities include:

- **Todo Retrieval:** Find information about specific todo items
 - Example: "What todos do I have for today?"
- **Todo Planning:** Get assistance organizing daily todo items
 - Example: "Help me organize my todos by priority"
- **Todo Summarization:** Generate overviews of todo items
 - Example: "Summarize my completed todos from last week"

General AI Assistance

General AI capabilities available across the application:

- **Retrieval:** Find specific information across your data
 - Example: "When is my next team meeting?"
- **Analysis:** Get insights from your data
 - Example: "What's taking up most of my time this month?"
- **Summarization:** Generate concise overviews
 - Example: "Give me a summary of my current projects"
- **Planning:** Get assistance with organization
 - Example: "Help me plan my workday for optimal productivity"

How to Use the AI Assistant

Accessing the AI Assistant

The AI assistant is available throughout the COMPASS application:

1. Click on the AI assistant icon in the sidebar
2. Type your question or request in the chat input
3. Press Enter or click the send button
4. The AI will respond with relevant information based on your data

Writing Effective Prompts

To get the most out of the AI assistant, follow these guidelines:

- **Be specific:** Clearly state what information you need
 - Good: "Show me high-priority tasks due this week"
 - Not as good: "Show me tasks"
- **Provide context:** Include relevant details in your request

- Good: "Summarize tasks for the website redesign project"
- Not as good: "Summarize tasks"
- **Specify intent:** Indicate what you want to do with the information
 - Good: "Analyze my task distribution to optimize my workload"
 - Not as good: "Show me my tasks"
- **Use common phrases:** The following phrases help the AI understand your intent:
 - "Find..." or "Show me..." for retrieval
 - "Analyze..." or "What insights..." for analysis
 - "Summarize..." or "Give me an overview..." for summarization
 - "Help me plan..." or "Organize..." for planning

Example Queries by Domain

Task Management

- "Show me all high-priority tasks due this week"
- "What are the blockers in my current workflow?"
- "Which tasks are at risk of missing deadlines?"
- "Help me organize my project tasks by priority"
- "Analyze the dependencies in my current project"
- "Summarize the status of all tasks in the marketing campaign"

Todo Management

- "What todos do I need to complete today?"
- "Help me organize my todos by importance"
- "Show me recurring todos that I need to handle"
- "Which todos are linked to my high-priority tasks?"
- "Summarize my completed todos from last week"

Calendar and Planning

- "What meetings do I have tomorrow?"
- "Help me plan my schedule around my high-priority tasks"
- "Find time blocks where I can work on my project tasks"
- "Analyze how I'm spending my time this month"

RAG Technical Features

COMPASS implements several technical features to enhance the RAG experience:

Context Window Management

The system intelligently manages the context window size to ensure optimal AI responses:

- Task domain: 1500 token context window
- Todo domain: 1200 token context window
- Default domain: 2048 token context window

Intent Detection

The AI assistant automatically detects your intent to provide the most relevant response:

- **Retrieve:** Finding specific information
- **Analyze:** Getting insights and patterns
- **Summarize:** Creating concise overviews
- **Plan:** Organizing and scheduling

Caching for Performance

The system caches responses to provide faster answers:

- Task domain: 3600 second (1 hour) cache TTL
- Todo domain: 1800 second (30 minute) cache TTL
- Cache invalidation occurs when data is updated

Relevance Filtering

The system uses several techniques to ensure relevant information:

- Similarity threshold for matching content
- Temporal matching for date-related queries
- Field-specific search for targeted inquiries

Privacy and Security

The COMPASS RAG system is designed with privacy and security in mind:

- All data remains within your COMPASS environment
- Vector embeddings are stored securely
- User permissions are respected in all queries
- Only authenticated users can access the AI assistant
- Information is retrieved based on user access level

Troubleshooting

Common Issues

- **Not finding specific information:** Try rephrasing your query with more specific keywords

- **Too much information:** Narrow your query by specifying a time period, project, or category
- **Incorrect information:** Ensure your data is up-to-date in the system
- **Slow responses:** For complex queries, the system may take a moment to process

Feedback Loop

Help improve the AI assistant:

1. Rate responses using the thumbs up/down buttons
2. Provide specific feedback about what was helpful or missing
3. Report any issues to administrators
4. Suggest improvements for the AI capabilities

Advanced RAG Features

For advanced users and administrators:

Custom Prompts

Administrators can customize domain-specific prompts in the configuration:

- Edit the `domain_config.json` file
- Modify prompt templates for each domain and intent
- Changes take effect immediately without system restart

Collection Management

Administrators can manage the RAG knowledge base:

- View collection statistics through the admin interface
- Update the knowledge base with new information
- Invalidate cache to refresh data

Performance Monitoring

Track and optimize AI assistant performance:

- View response times and success rates
- Monitor cache hit ratios
- Track user feedback metrics
- Identify common query patterns