Rossum Aurora GenAI Model Card

Last updated: 31 August 2026

Overview

Rossum is a B2B Software-as-a-Service (SaaS) platform designed to automatically or semi-automatically process uploaded transactional documents, such as invoices, bills of lading, purchase orders, and similar business materials.

Rossum Aurora is powered by the Transactional Large Language Model (T-LLM), Rossum's proprietary model designed specifically for understanding transactional documents – please see Rossum Aurora T-LLM AI Model Card for details. Atop of that several features are powered by GenAI and usage of tools, such as **AWS Bedrock, Anthropic Claude, Haiku, and Qwen by Alibaba Cloud (see Reasoning fields exceptions limited to the Japanese environment) within Rossum AWS clusters**.

Business Rules Copilot [GA]

Intended Use and Model Used: We encourage users to use the Business Rules Copilot instead of writing the code themselves. Conversation-based AI copilot developing the formulas using a natural language prompt.

Formulas are generated by usage of **AWS Bedrock, Anthropic Claude Sonnet and Haiku within Rossum AWS clusters**.

System Inputs: Natural language prompt (includes the user query and information about the schema and previous iterations with the copilot).

System Outputs: Python code snippet.

Data Types: No data from processed documents.

Exclusion/Opt-out: Optional function - users may opt not to use the feature. If used, no customer data is shared with the third-party AI provider.

Conversational Insights Assistant [GA]

Intended Use and Model Used: The Conversational Insights Assistant allows users to generate reporting questions, charts, and dashboards by describing what they want in plain English. The assistant translates natural-language user prompts into queries, selects appropriate visualisations, and automatically builds structured reports and cards. Predictions

and query parsing are powered by a generative **AI analytics tool provided by Metabase, Inc.** Rossum runs this tool directly in its AWS clusters.

System Inputs: Plain English prompts specifying reporting measures (e.g., automation rate, document counts, time spent), breakdowns (e.g., by queue, vendor, status), timeframes (e.g., last 30 days, previous 3 months), and optional queue scope filters.

System Outputs: Structured database queries, configured chart and table visualisations, and generated reporting cards or dashboards.

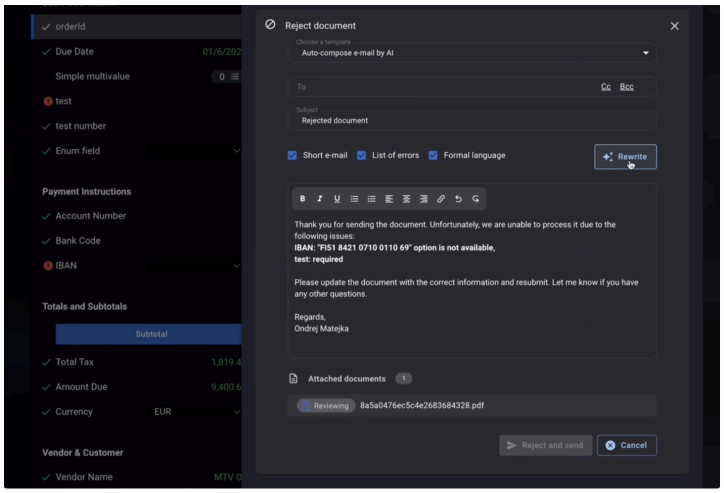
Data Types: Reporting database metadata and schema structures, query specifications, and aggregated operational performance metrics.

Exclusion/Opt-out: Optional feature available to Enterprise Tier customers with Insights enabled for their organization. Users may opt not to use the conversational assistant and manually construct queries or dashboards instead. If used, no customer data is shared with the third-party AI provider.

Email Communications [GA]

Intended Use and Model Used: The Email Communications is a writing assistant for rejection emails that can be sent from the Rossum platform. It's operated through a simple UI above email content.

Predictions are generated using Anthropic Claude on AWS Bedrock, hosted within Rossum's AWS clusters. Humans are required to check the output to avoid a chance of hallucinating values.



System Inputs: Contact information (sender, receiver), list of automation blockers together with datapoint values on the document, user preferences for the writing style

System Outputs: Email messages

Data Types: Values from processed documents

Exclusion/Opt-out: Optional function - users may opt not to use the feature. If used, no customer data is shared with the third-party AI provider.

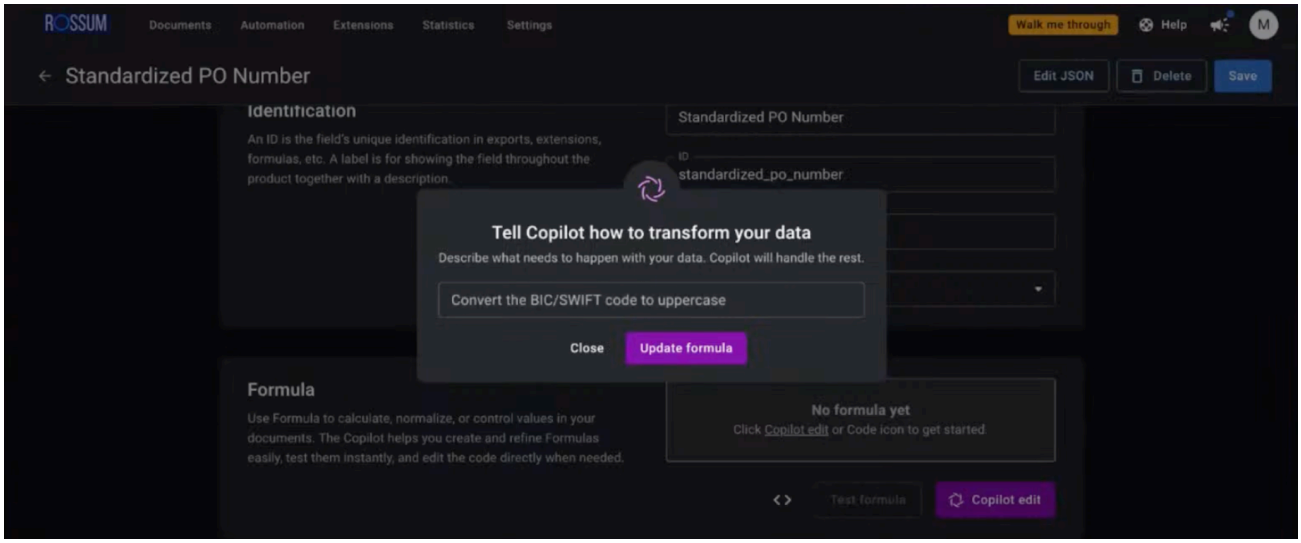
Formula Fields Copilot [GA]

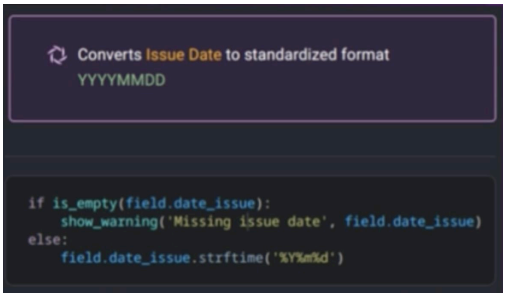
Intended Use and Model Used: We encourage users to use the Formula Fields Copilot instead of writing the code themself. Conversation-based AI copilot developing the formulas using a natural language prompt.

We use the AI model to suggest formula code just based on the field name and its context.

We use the AI model to create natural-language descriptions of what the code does (useful for getting quickly into the setup).

Formulas are generated using **Anthropic Claude Sonnet and Haiku on AWS Bedrock, hosted within Rossum's AWS clusters.**





System Inputs: Natural language prompt (includes the user query and information about the schema and previous iterations with the copilot).

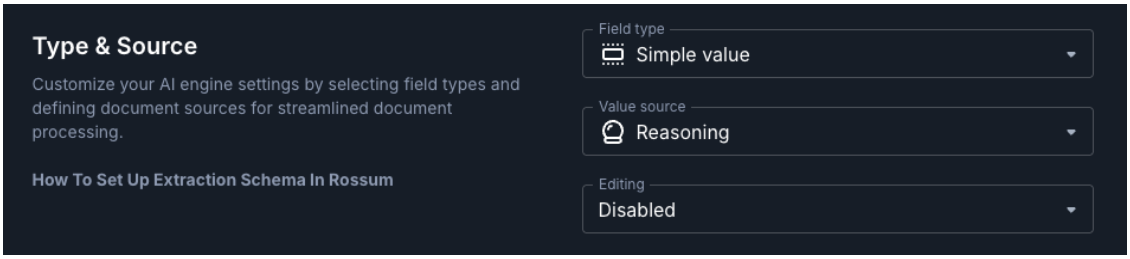
System Outputs: Python code snippet.

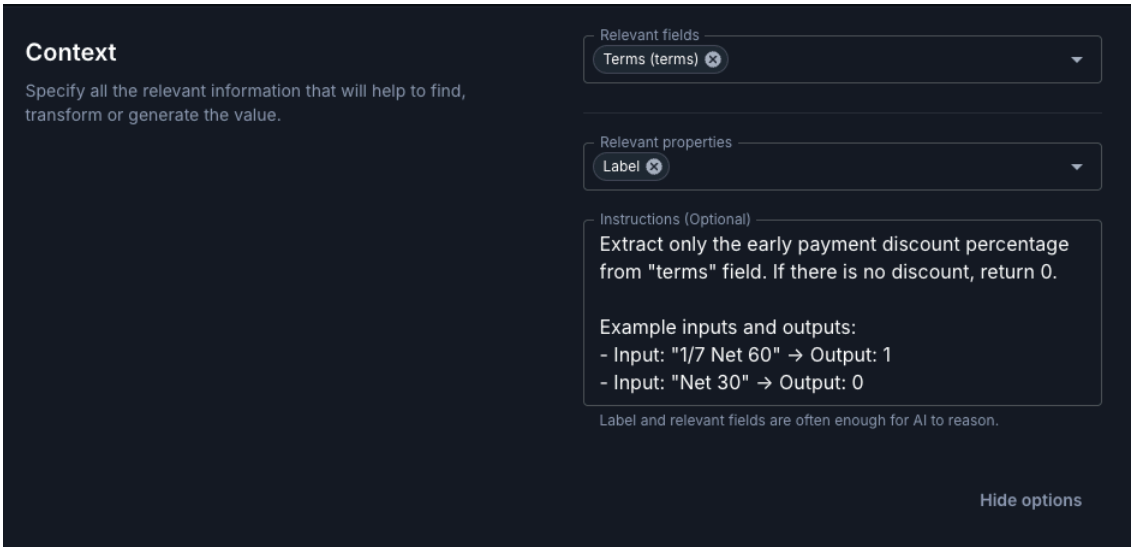
Data Types: No data from processed documents.

Exclusion/Opt-out: Optional function - users may opt not to use the feature. If used, no customer data is shared with the third-party AI provider.

Reasoning Fields [LA]

Intended Use and Model Used: We encourage users to guide the Reasoning Fields agent on what to infer and which data it can use. The agent can either generate a new value based on context or select one from a predefined list. Reasoning fields also learn from your corrections within a queue. If the AI makes a mistake once, it won't repeat it.





Values of reasoning fields are generated by using various LLMs on **Amazon Bedrock, hosted within Rossum's AWS clusters**. At this point, we focus on small to mid-sized state-of-the-art models, primarily utilizing Haiku globally and Qwen in Japan for technical reasons.

System Inputs: Relevant user-selected fields from a document (may include formula field with email content), selected properties from a schema, and user instructions. Historical data from a given queue are used as well.

System Outputs: Value extracted from the provided data, new value generated from the data, or selected item from a given list.

Data Types: Values from processed documents, optionally from email.

Exclusion/Opt-out: Optional function - possible not to use.

Document Translation [LA]

Intended Use and Model Used: With Rossum Aurora, our T-LLM has been trained to capture data in 276 languages. In situations where data capture can't be automated, or human input is needed for a different reason, we provide users full text document translation overlaid on top of the document – so they can process the document as if it was in their mother tongue.

Translations are generated using **AWS Amazon Translate service.**

System Inputs: Full text of the document.

System Outputs: Full text of the document, translated to the target language.

Data Types: Full document content.

Exclusion/Opt-out: Optional function - users may opt not to use the feature. If used, no customer data is shared with the third-party AI provider for training purposes.

Rossum Agent [LA]

Intended Use and Model Used: Rossum Agent is an AI assistant accessible via a chat interface within the Rossum application and through a dedicated API. It is designed to assist administrators, account managers, and support engineers in configuring, debugging, and navigating the platform through natural-language interactions. The agent retrieves and, when explicitly authorized (by activating "edit-mode"), modifies platform configurations using the Rossum public REST API on behalf of the authenticated user.

This feature is available to selected customers within the EU.

All responses are generated using **AWS Bedrock, leveraging Anthropic's Claude Opus and Claude Haiku models hosted within Rossum's EU-based AWS environments.**

System Inputs:

- User Interactions: Direct chat messages - user written text messages + shared files (images, PDF, spreadsheet, markdown files), token headers, and API-URL headers.
- On-Demand Data: Information retrieved within the user's specific RBAC, including queues, workspaces, schemas, hooks, rules, engines, users, annotations, and extracted content.
- Reference Material: Master Data Hub records, Copilot suggestions, a snapshot of the Rossum knowledge base, and the public Rossum OpenAPI specification (containing no customer data).

System Outputs:

- Dynamic Responses: Streamed chat sessions including reasoning, text, tool calls, tool results, file notifications, and token usage metrics.
- Communication Drafts: Emerging support use cases, such as customer-facing emails derived from ticket content drafted in chat UI for the user to copy and use as text.
- Platform Modifications: Direct writes to the customer's tenant, including annotation edits, document uploads, and the creation, update, or deletion of hooks, rules, queues, schemas, engines, email templates, users, and workspaces - only in "edit-mode"
- Logging: Optional Slack threads containing the full conversation history in JSON format.

Data Types:

- Configuration & Content: Customer organization settings (queues, schemas, email templates, extensions), annotation values from processed documents, and master data records.
- User-Provided Data in chat: text, Images, documents uploaded or pasted into the chat.
- Exclusion & Governance: This is an optional feature requiring an explicit opt-in per organization. By default, the agent is restricted to read-only access.

Exclusion/Opt-out:

- This is an optional feature requiring an explicit opt-in per organization. By default, the agent is restricted to read-only access and no customer data is shared with the third-party AI provider.