AI Sandbox User Guide

AI Sandbox features

Feb-19-2026

Table of Contents

1. Overview

The AI Sandbox is a centralized platform designed to help enterprises safely explore, evaluate, and refine artificial intelligence solutions. It provides a structured environment where teams can upload data, experiment with prompts, test models, and benchmark outcomes against expectations. The AI Sandbox delivers a dedicated environment for AI experimentation, striking a balance between innovation and governance.

At its core, the Sandbox enables:

- Rapid experimentation with different AI models and approaches.
- Structured evaluation through automated comparison of model outputs against expected results.
- Collaboration and transparency allow teams to share insights and make data-driven decisions.
- Scalability and compliance, ensuring that even as experimentation grows, it remains secure and aligned with organizational policies.

This guide will walk you through the platform's main features, from uploading files and testing models in the Prompt Lab to running automated evaluations and collaborating with team members.

2. Log in and Landing Page

This section explains how to access the AI Sandbox securely and introduces the key features of the landing page, which serves as the central hub for navigation, discovery, and engagement.

2.1. Accessing the AI Sandbox

Access to the AI Sandbox is provided via Microsoft Single Sign-On (SSO) to ensure secure authentication. Go to https://aix.devx.systems/, complete Microsoft's 2FA, and you will be redirected to the AI Sandbox Landing Page.

2.2. Landing Page – AI Sandbox

The Landing Page is the starting point for every user and serves as a personalized dashboard for quick navigation, project discovery, and community engagement.

Below, we describe the key areas of the page.

Quick Start Actions

- Explore Prompt Ideas – Leads to the Prompt Library for ready-made prompts.

News and Announcements

- Curated updates and articles about AI innovations, industry news, and opportunities.

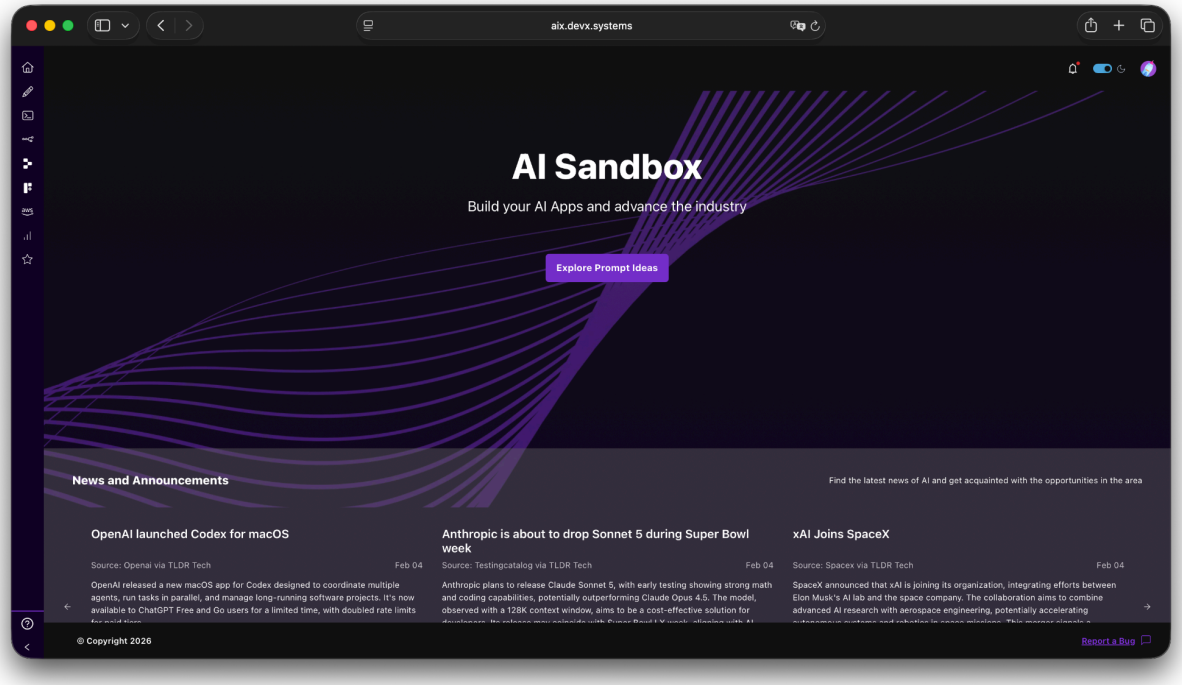

Figure 1 - Landing Page

- **Prompt Ideas:**
    - A searchable and categorized collection of prompt templates.
    - Each prompt card includes: Prompt title and Short description
    - “Use Prompt” button – launches the selected prompt in the Prompt Lab.

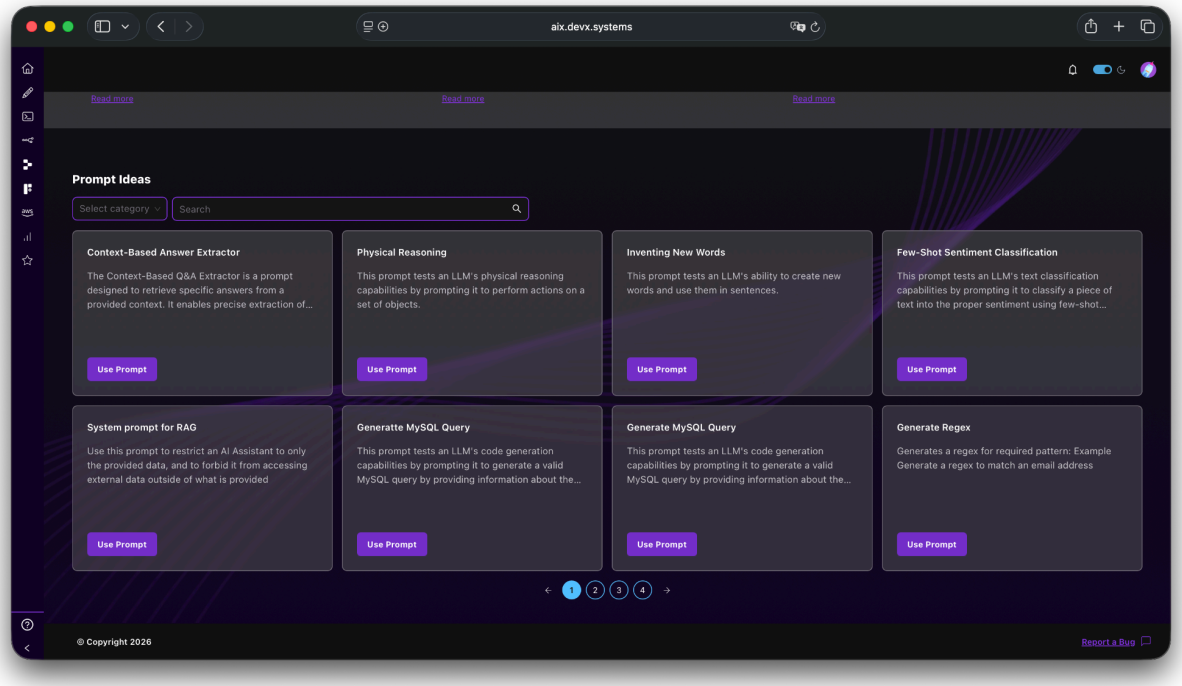

Figure 2 - Landing page

At the bottom-right corner of the landing page, you will find the **“Report a Bug” button.** This opens a dedicated feedback channel where users can:

- Report any technical issues, unexpected behavior, or UI/UX inconsistencies.
- Suggest improvements or new features that could enhance the Sandbox experience.

When reporting a bug, please provide as much detail and description as possible. Include:

- A clear explanation of the issue.
- Steps to reproduce the problem (if known).
- Screenshots or screen recordings (highly recommended) to help the support team quickly understand the context.

Providing thorough details ensures faster troubleshooting and helps improve the overall platform for all users.

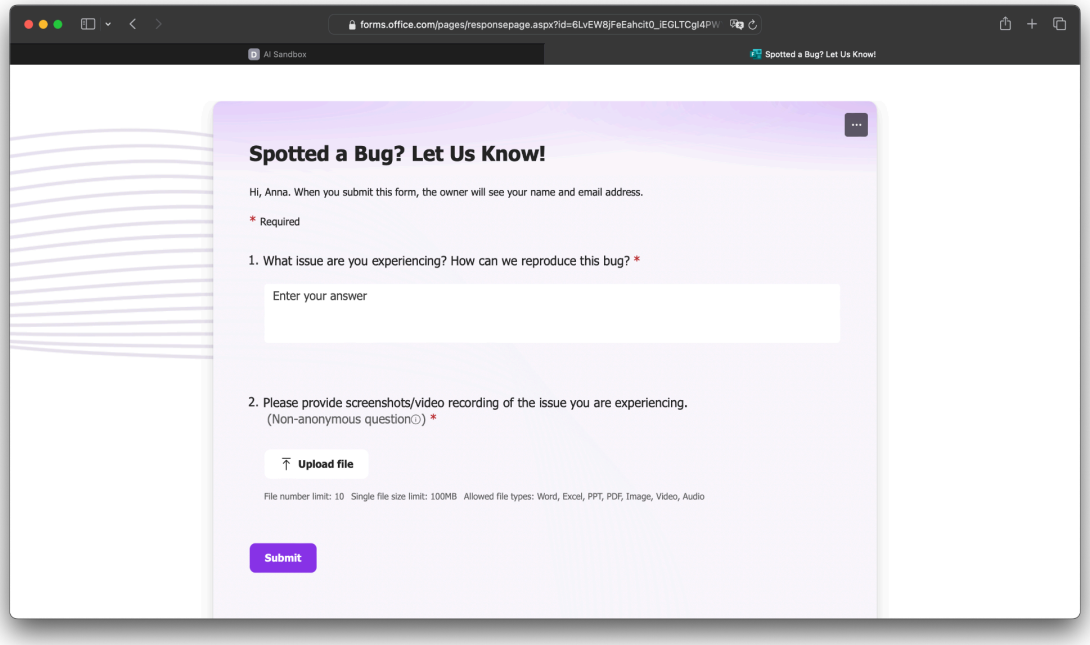

Figure 3 - "Report a bug" form

2.3. Left Menu – Navigation Sidebar

The Navigation Sidebar is available on every page of the AI Sandbox, providing quick access to all core features and tools. It is organized into sections by functionality, making it easy to move between app creation, automation, data, and community engagement.

The Key sections of the Navigation Sibar include:

- Home – Returns to the landing page.
- Prompt Lab – Test and iterate on prompts with different AI models.
- Prompt Library – Browse, reuse, and adapt prebuilt prompts.

Build & Automate

- n8n – Automation tool for connecting apps and services.
- Replit – Online coding editor for rapid prototyping.
- FluxNova – Open-source process orchestration platform for designing and running end-to-end workflows.

Cloud & Data

- AWS – Access to AWS for cloud resources and integrations.

Insights & Engagement

- Analytics – Dashboards and insights, including LLM performance and app usage analytics.
- Hackathons – Create, join, and manage hackathons within the platform.

- Settings – Add and manage API and database connections.
- FAQ – Frequently asked questions about AI Sandbox.

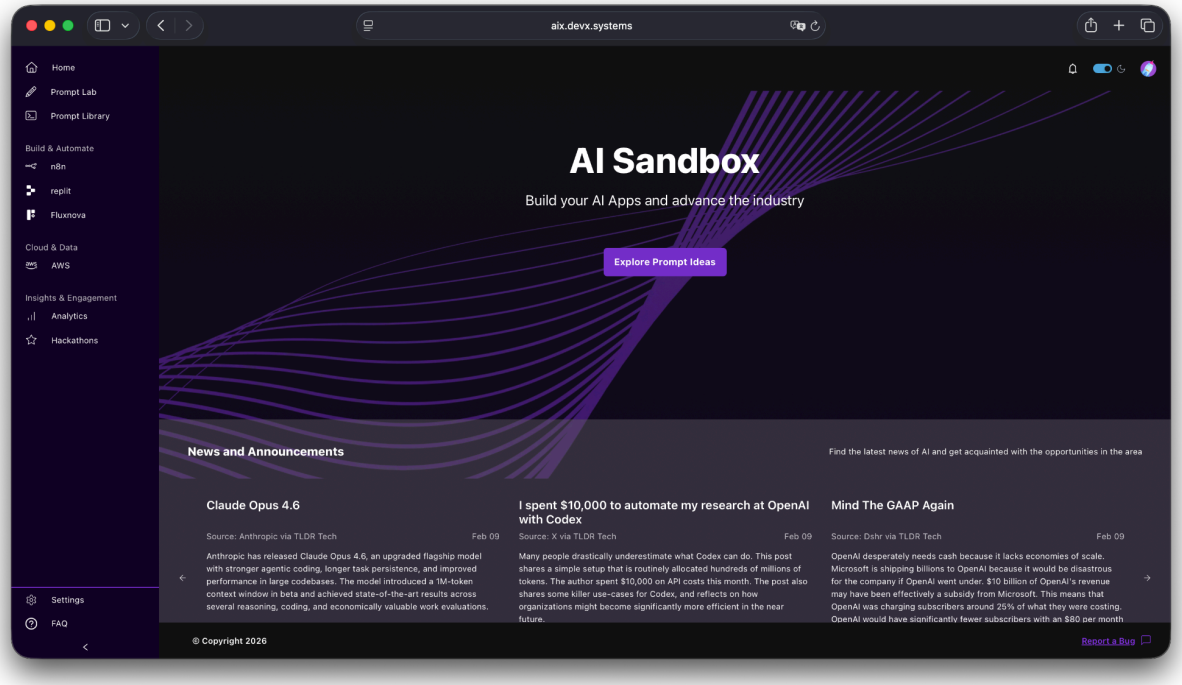

Figure 4 - Side-bar view

3. Hackathons

The **Hackathons** area in AI Sandbox is where participants can join innovation challenges, collaborate with their teams, and showcase solutions. From the main Hackathons page, you can browse available events, view key details like titles and deadlines, and dive into the one you're participating in. Inside, you'll find tabs for general event information, team details, assigned business track, and a dedicated team dashboard. The dashboard makes it easy to track your team's selected business track, manage members, and prepare your solution. When it is time to submit, you can provide links, descriptions, and demos, with clear visibility into draft and final submission statuses. This section is designed to guide participants through every step, from joining a team to submitting a complete solution for evaluation and review.

By clicking on the Hackathons, the Hackathon Participant will see the details of the Hackathon (if only one is available) or a list of Hackathons containing the following information:

- Title of the Hackathon
- Hackathon Status - "Submission closed" or "Submit now"

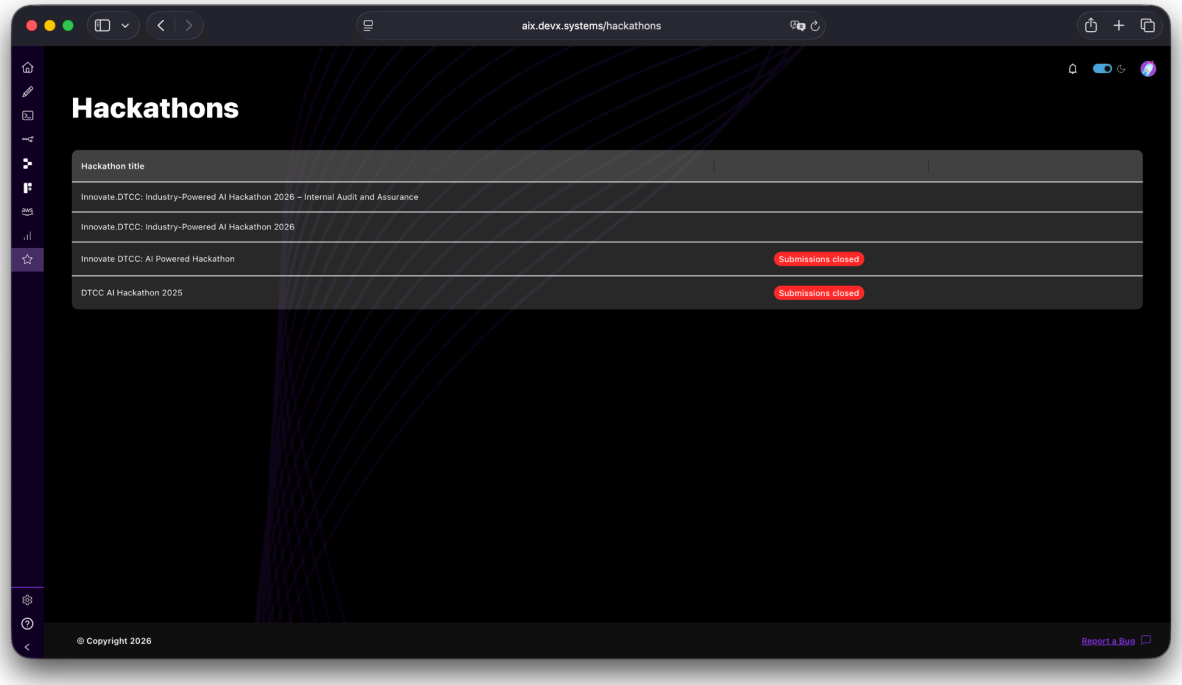

Figure 5 - Hackathons page

Assuming multiple Hackathons are available, the Hackathon Participant must select one Hackathon to see its details. After entering a Hackathon, the Hackathon Participant can see the following tabs containing information:

- **General Info:** The description of the Hackathon.
- **Teams:** the list of teams participating in the Hackathon.

- **Team Dashboard:** the team dashboard of the Hackathon Participant.
- **Business Tracks:** the list of business tracks for the Hackathon.

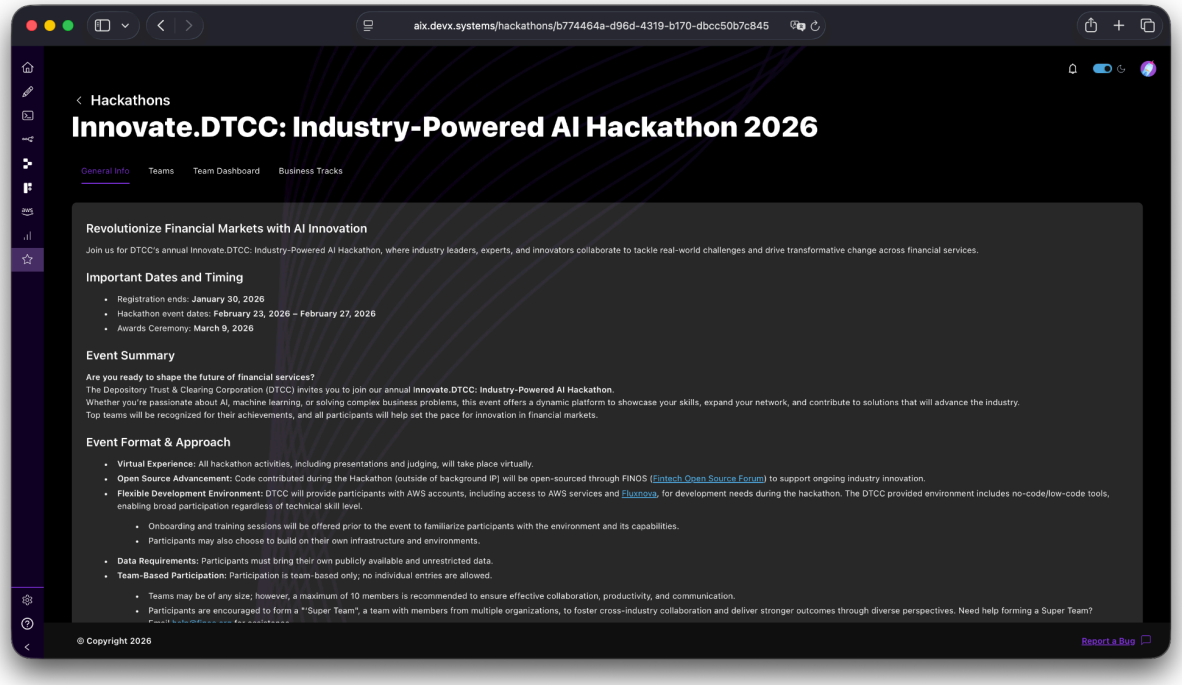

Figure 6 - Hackathon General Info

The Teams Tab contains information about the teams and their selected Business Track

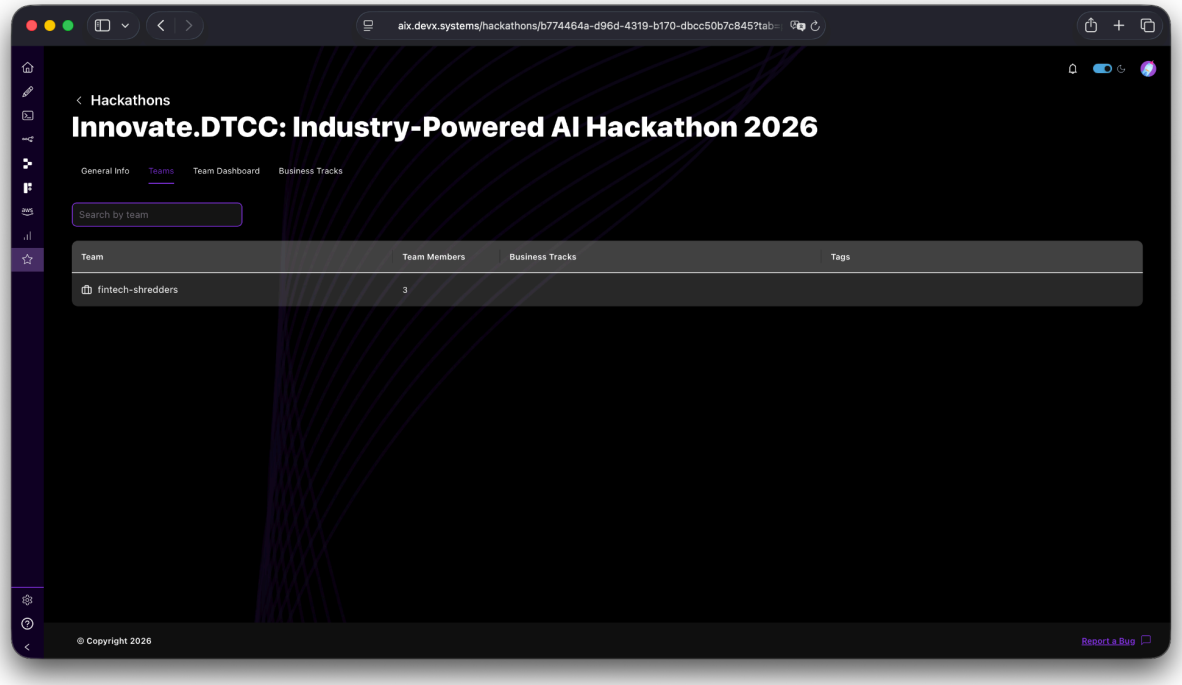

Figure 7 - Hackathon Teams

The Business Tracks tab contains detailed information on the topics explored during the Hackathon.

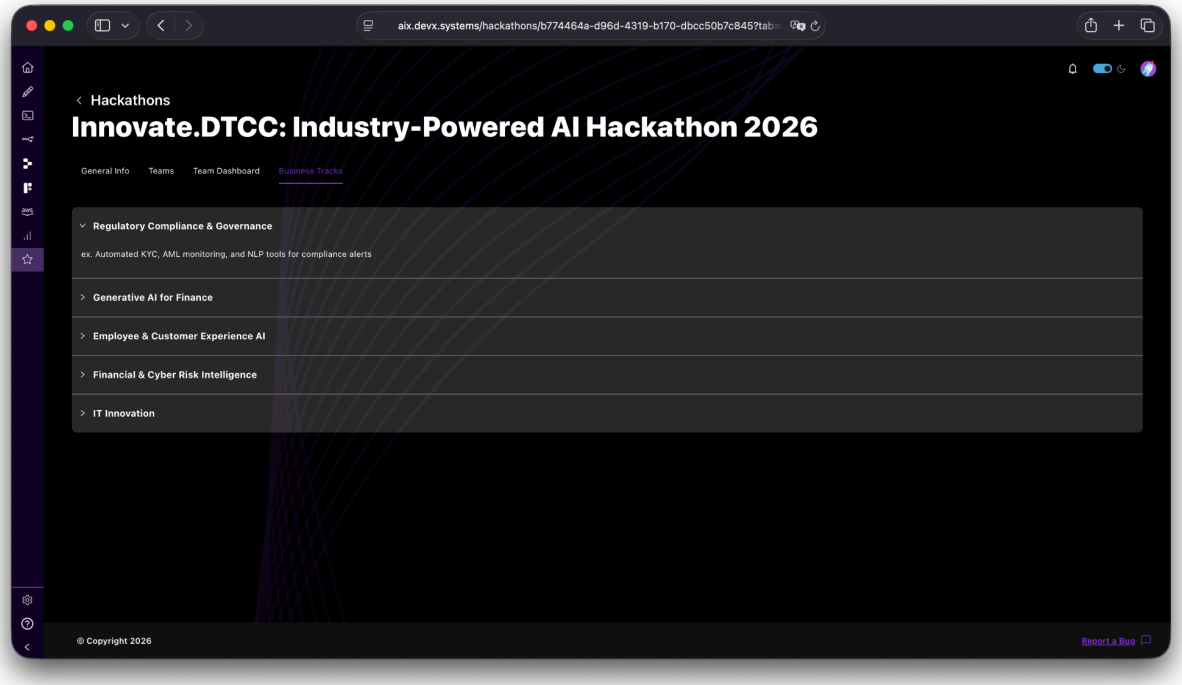

Figure 8 - Hackathon Business Tracks

The Team Dashboard provides information about the team to which the Hackathon Participant belongs. On this page, the Hackathon Participant can:

- See the business track selected by the team.
- See the team category (Professional or Academic).
- See the Team Members as part of the team.
- Solution for the Hackathon.
- See the details of the solution after submission.

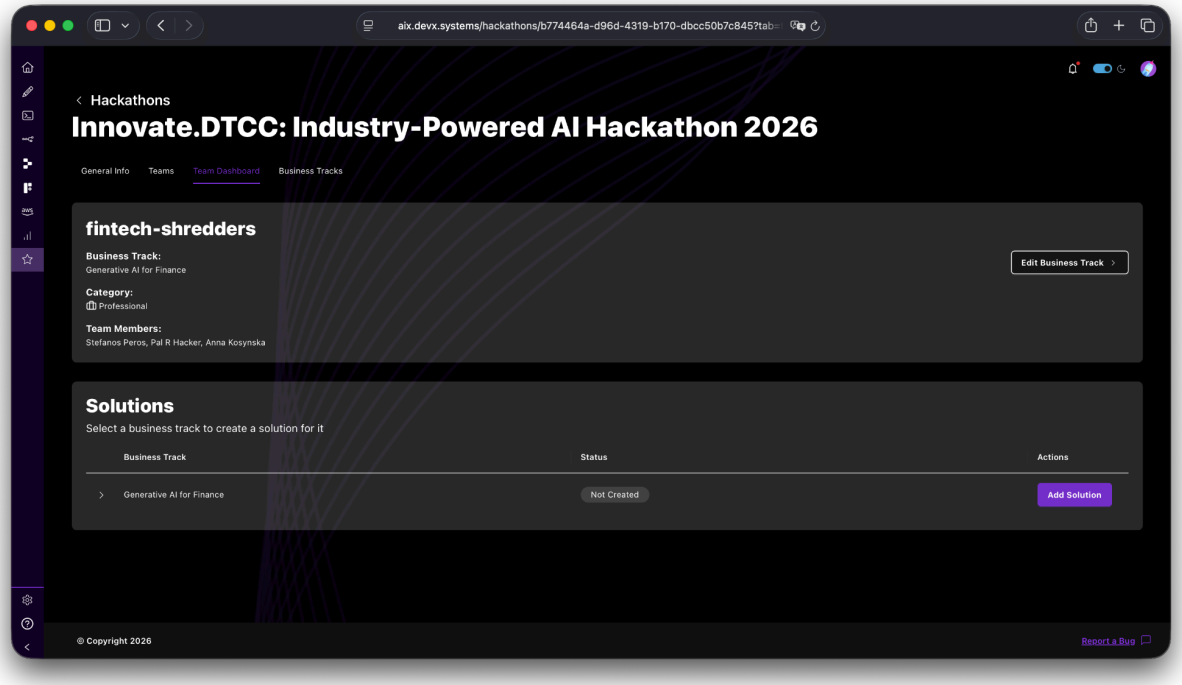

Figure 9 - Team Dashboard

By scrolling down on the page, the Hackathon Participant can submit the Solution developed in the Hackathon on behalf of the team.

The Hackathon Participant will need to provide the following information:

- Name of the solution
- Demo link for the solution
- Repository Link
- Description of the solution.

Moreover, Hackathon Participants can view the assessment criteria for the Hackathon to understand what to focus on.

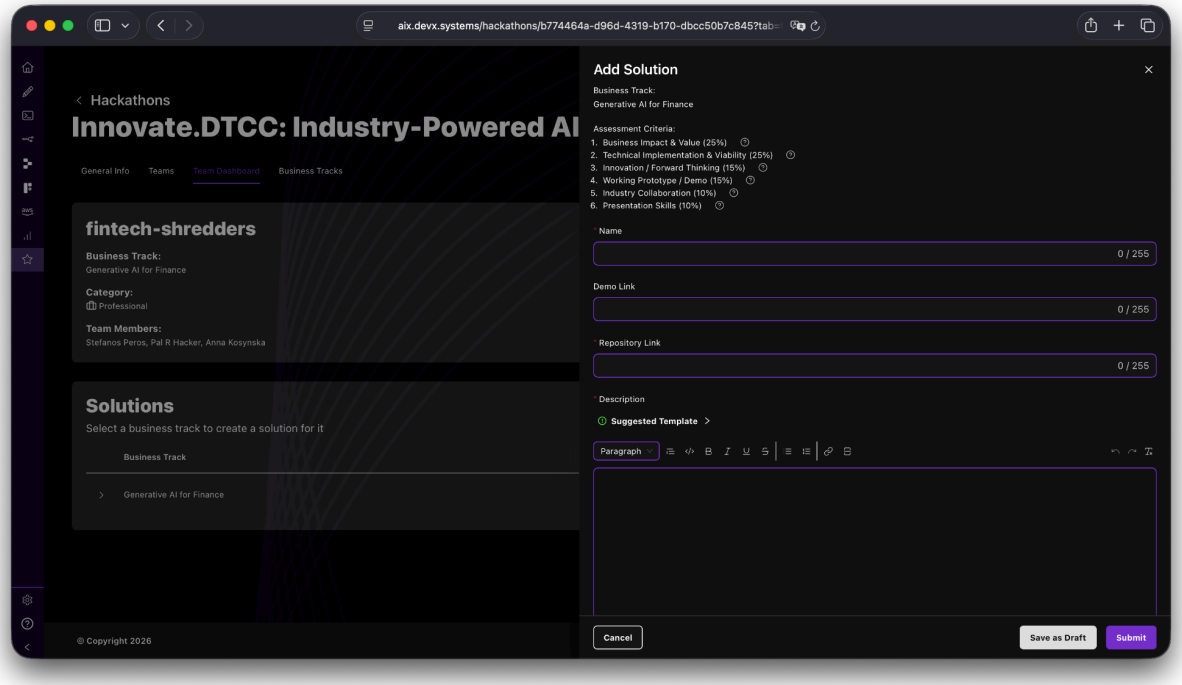

Figure 10 - Add a solution

When a team saves its solution as a Draft, this status and submission details are visible on the Team Dashboard.

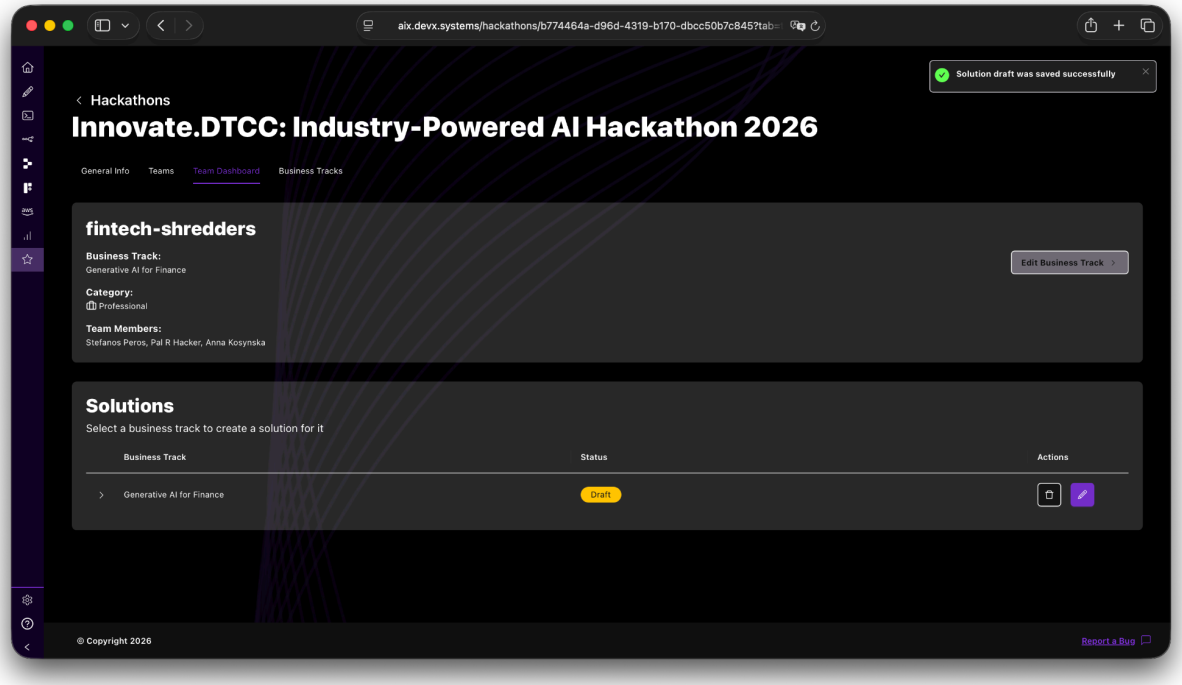

Figure 11 - Solution Status (Draft) and Details

Once the team is ready, the Hackathon Participant should submit the solution on their behalf.

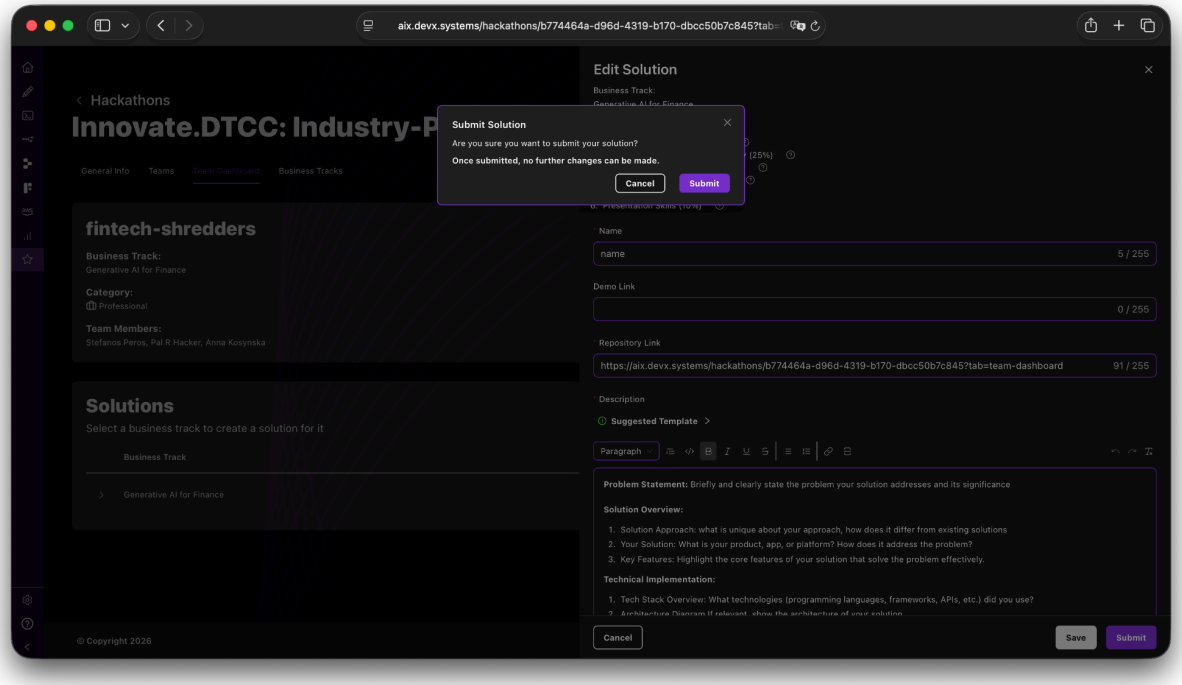

Figure 12 - Submit Team's Solution

Once submitted, the solution becomes ready for evaluation. Only Solutions with the status "Submitted" will be evaluated.

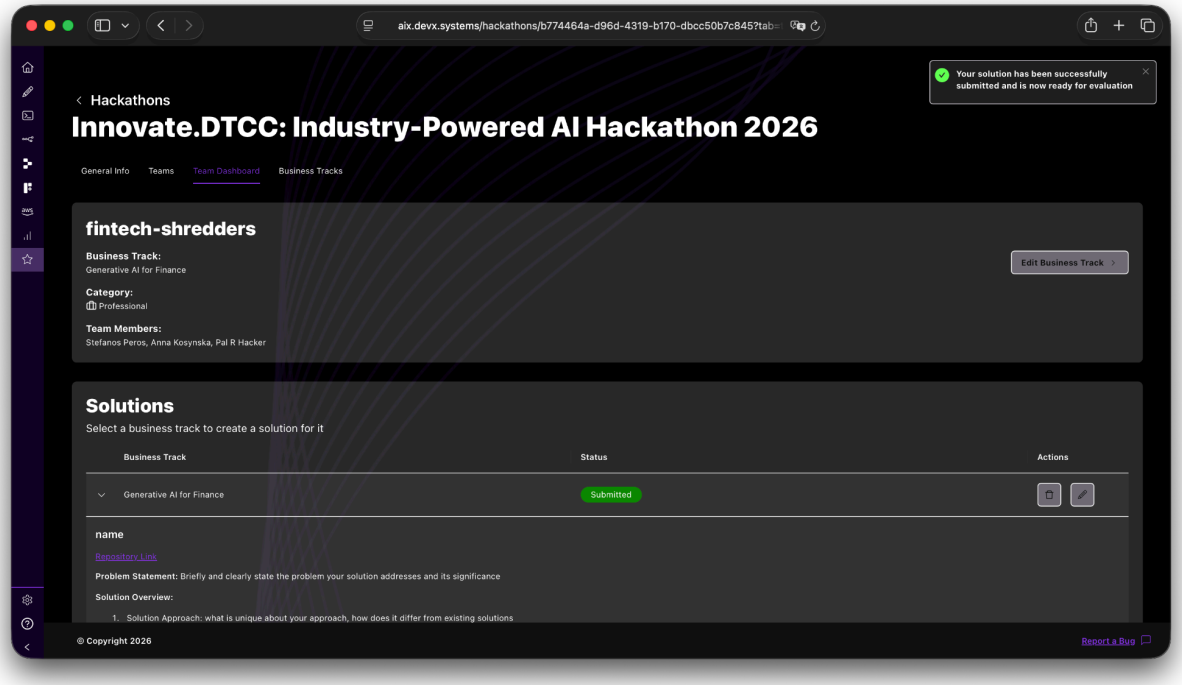

Figure 13 - Submitted Solution

4. Prompt Lab

The Prompt Lab is the core creative workspace of the AI Sandbox. It serves as the workshop where raw ideas are turned into functional AI applications. Here, users can design new apps, configure models, connect external data sources, and refine outputs—all in one place. Every app has its own saved settings, allowing users to experiment freely without losing previous configurations.

The Lab provides flexibility at each stage of development. The users can:

- Select and fine-tune foundational models with simple controls.
- Connect databases, APIs, or uploaded files to extend app knowledge.
- Reuse existing prompts directly from the Prompt Library to accelerate creation.

Beyond building, the Prompt Lab also offers built-in tools for testing, evaluation, and analytics. Users can measure accuracy, response quality, and overall engagement, ranging from basic statistics (such as response time) to advanced insights (including semantic similarity, factual correctness, and retrieval faithfulness).

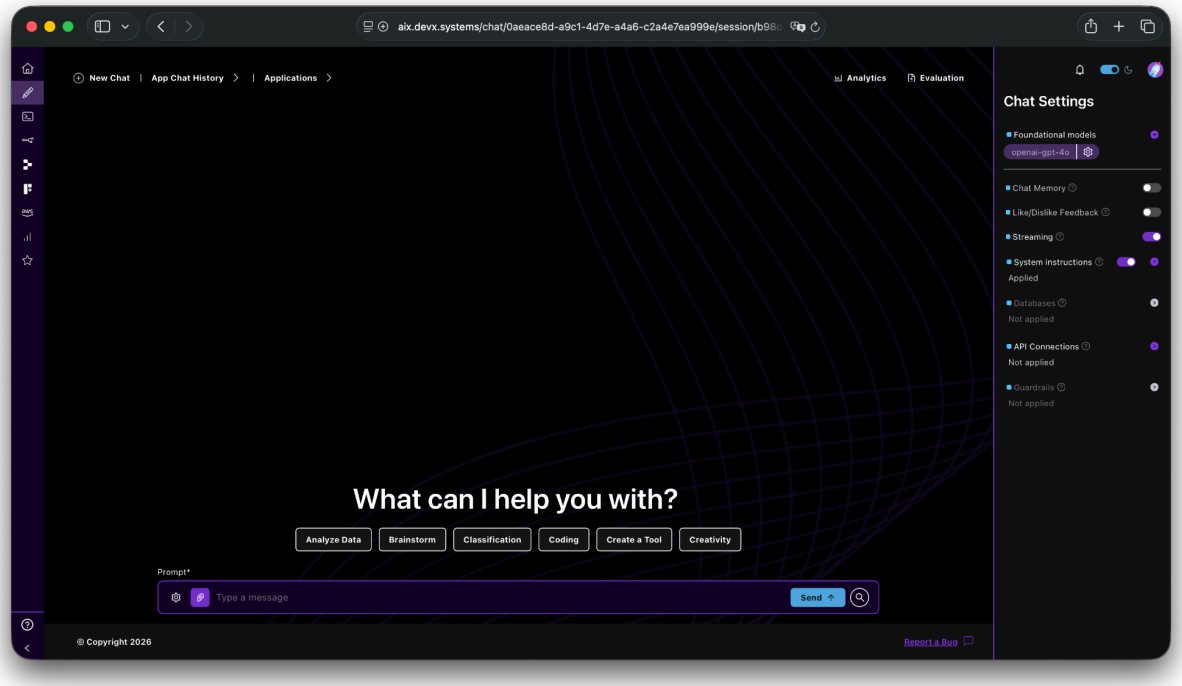

Figure 14 - Prompt Lab

The Prompt Lab is the core workspace in AI Sandbox where you can create, reuse, and evaluate AI applications. It allows you to experiment with prompts, configure models, connect data sources, and monitor performance in one place.

3.1. Key Features and Workflow

The Key Features and Workflow section guides you through the essential tools in the Prompt Lab, covering how to create and manage apps, configure models, connect data, upload files, and utilize the Prompt Library to streamline your workflow.

Creating and Managing Applications

To create and manage a new AI application, you can execute the following steps:

1. Go to the Applications tab.
2. Click Create New App.
3. Configure the app's Chat Settings (foundational models, database/API connections, and system instructions).
4. Your settings are saved automatically and applied to all new chats within the app.
5. Create multiple chats under the same app to explore different ideas while keeping a consistent configuration.

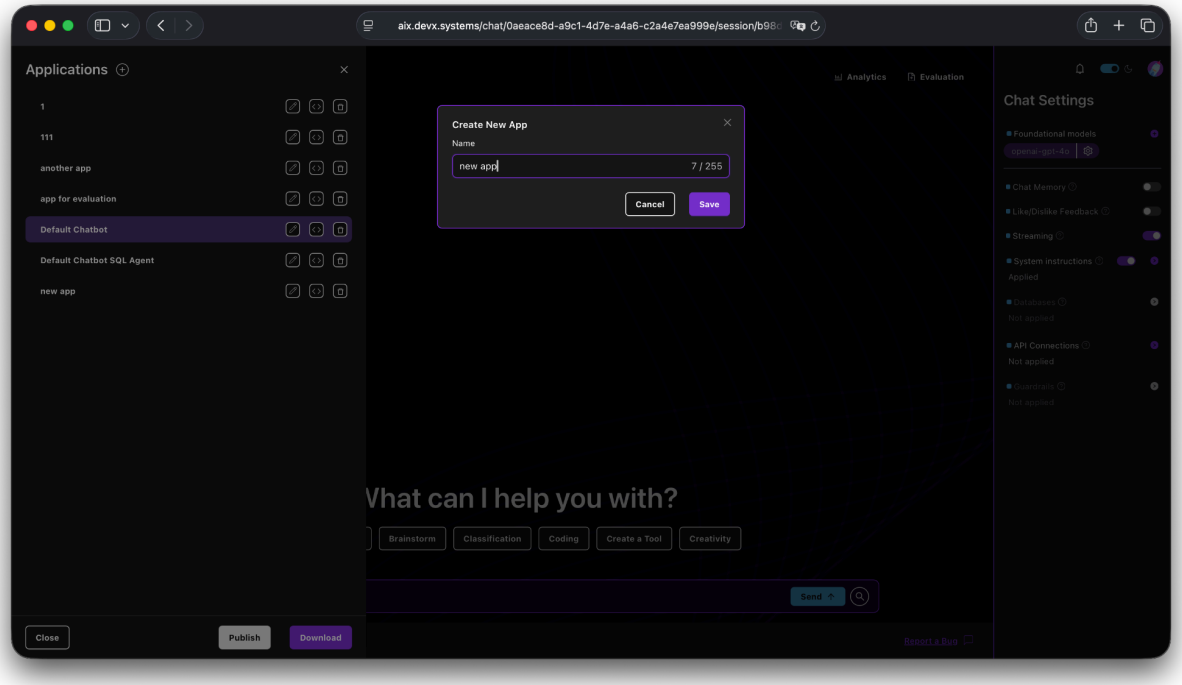

Figure 15 - Prompt Lab: Create a new app

Foundational Models

In the foundational models section, you can select up to four foundational models for each app. Each model can be fine-tuned with parameters such as temperature and maximum tokens.

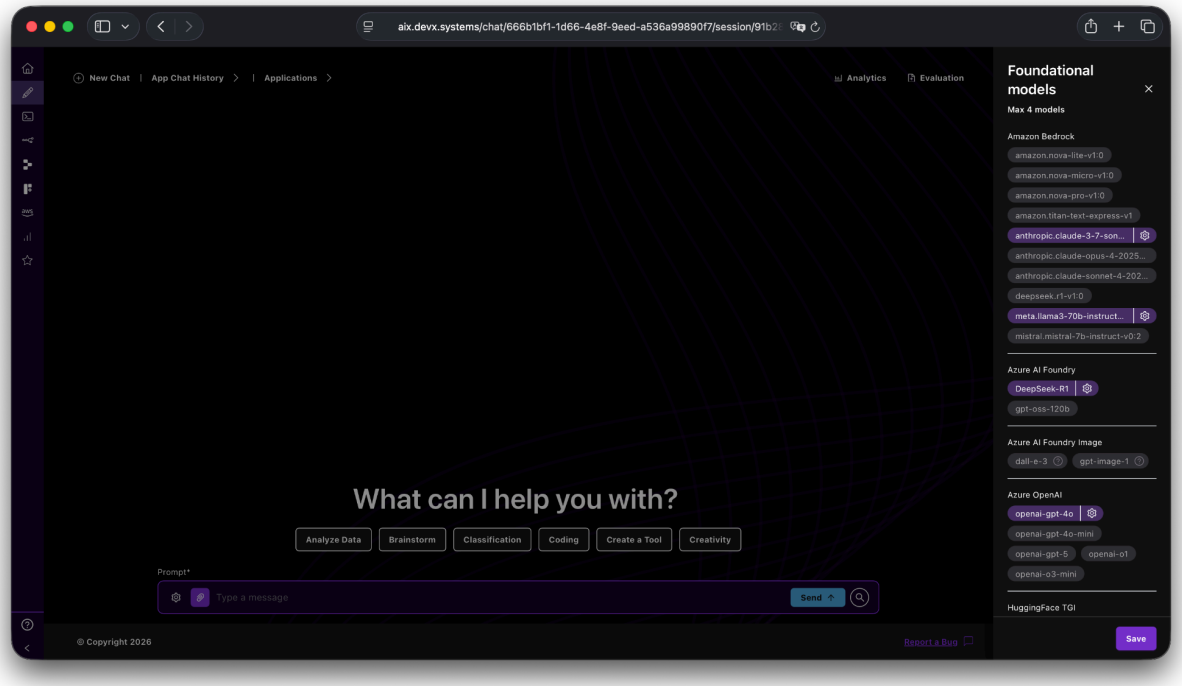

Figure 16 - Prompt Lab: Foundational Models.

Note: Image and text models cannot be used simultaneously.

The screenshot below shows the **Model Configuration** settings for the LLM model in the AI Sandbox interface. Here is a detailed breakdown:

- Temperature: A slider to adjust the randomness of the model's output during the text generation. It is a scaling factor applied to the probability distribution of the next token (word or sub-word) before sampling. Higher values lead to random responses, while lower values result in more focused and deterministic output.
- Maximum Tokens: A slider to control the length of the generated output, limiting how many tokens (words/characters) the model can produce in response to a prompt.
- Stop Sequences: A text input box where specific phrases or sequences can be entered. When the model generates any of these sequences, it stops generating more text.
- Top P: The model considers the results of the tokens with top_p probability mass. So 0.1 means only the tokens comprising the top 10% probability mass are considered.
- Frequency Penalty: A slider to control the penalty for repeated phrases or words. Higher values reduce the model's tendency to repeat phrases.
- Presence Penalty: A slider to discourage the model from mentioning specific concepts if they have already been mentioned in the conversation. It encourages introducing new topics or variations.

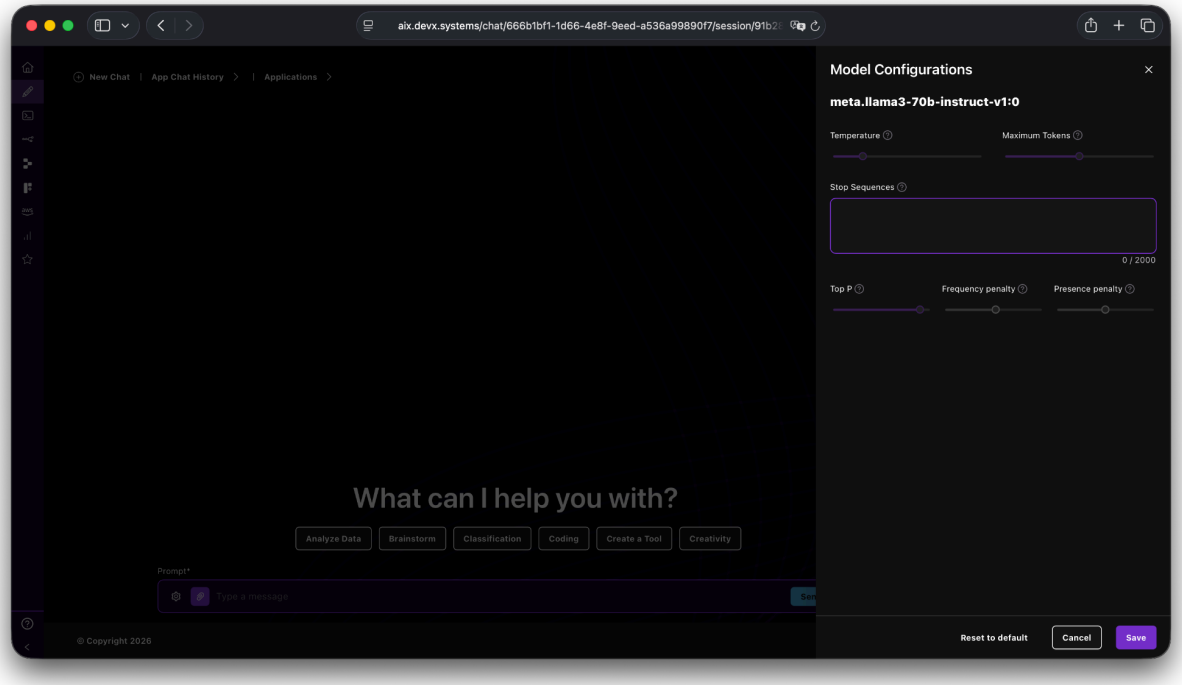

Figure 17 - Prompt Lab: Model Configurations

Chat Settings

After selecting your Foundational Models, you can further configure the behavior and capabilities of your chat. The following settings allow you to customize how the application interacts, stores context, and connects to external resources:

- Chat Memory: Enable or disable memory to let the model retain context across chats.
- Like/Dislike Feedback: Collect structured user feedback for evaluation and fine-tuning.
- Streaming: Toggle real-time streaming of responses for a more interactive experience.
- System Instructions: Define global behavior guidelines and instructions applied to the model.
- Databases: Connect to external databases for enhanced retrieval and context injection.
- API Connections: Integrate third-party APIs to extend functionality and enrich responses.
- Guardrails: Apply rules or constraints to ensure safe and controlled AI outputs.

Note: Database and API connections are configured on the **Settings** page and will then be available in **Prompt Lab.**

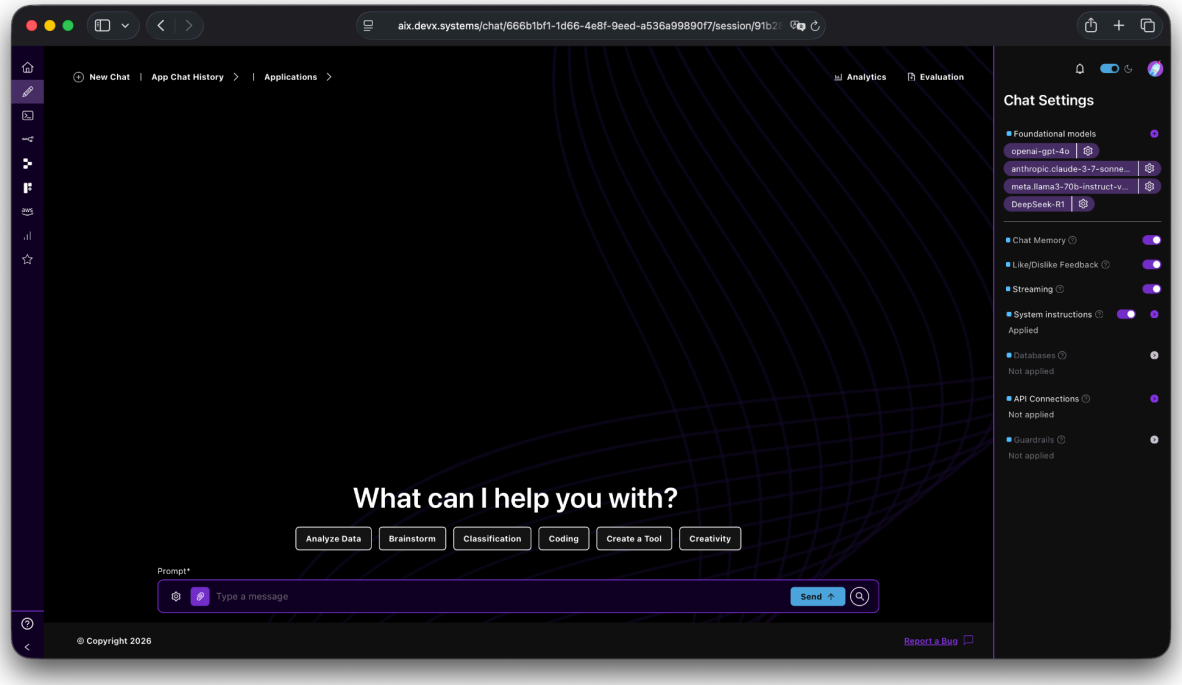

Figure 18 - Prompt Lab: Chat Settings

File Uploads with Chunking

When uploading a file into the AI Sandbox, you can configure Chunk Settings to control how the content is split and processed by the system. This ensures more accurate retrieval and interaction with your data during chats or app execution.

- Chunk Size: Defines how large each segment of text will be. Larger chunks provide more context but may reduce retrieval precision, while smaller chunks improve accuracy but may lose context.
- Chunk Overlap: Determines how much text overlaps between chunks. Overlap ensures smoother continuity and prevents important information from being cut off between segments.
- Number of Results: Sets how many chunks are returned during retrieval for each query. More results can provide a broader context, but may increase processing time.
- Embedding Model: Selects the embedding model used to convert text into vector representations for semantic search and retrieval.
- Separators: Specify custom separators (e.g., newlines, punctuation) that control where text is split during chunking.

After configuring these settings, you can upload your file. The system will automatically process it based on your chosen chunking rules, making the content searchable and usable within your app's AI interactions.

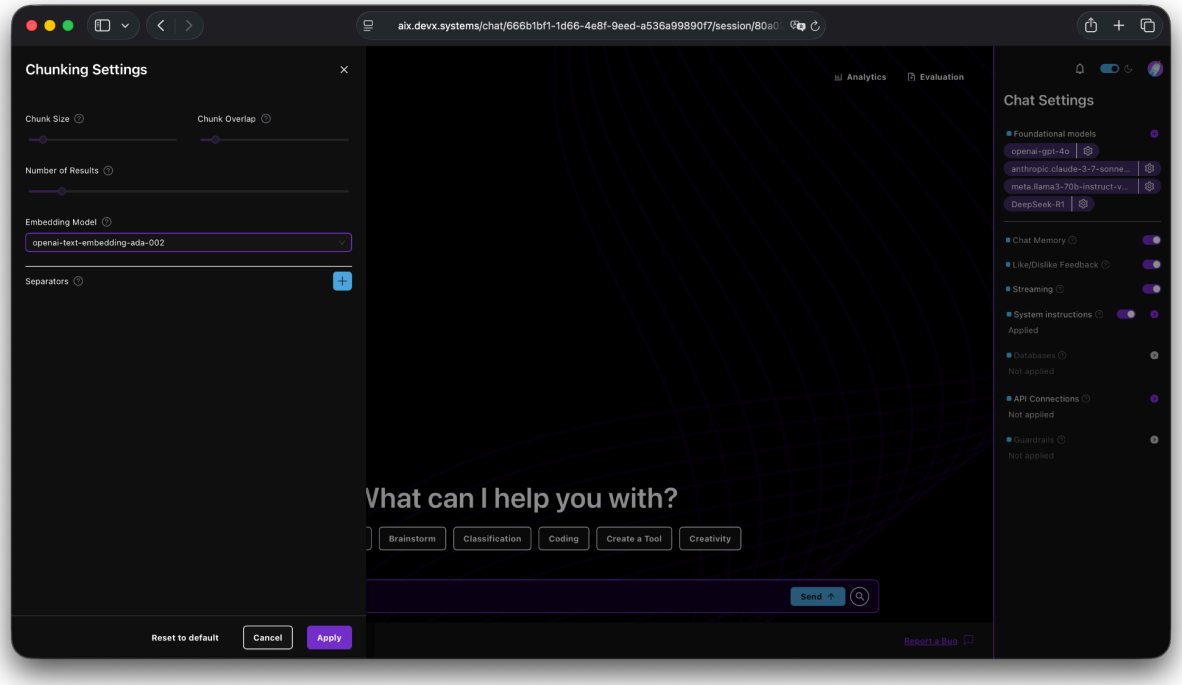

Figure 19 - Prompt Lab: Chunking Settings

Prompt Library Integration

Next to the Send button, a Search option allows you to pull prompts from the Prompt Library. This makes it easy to reuse or adapt existing prompts directly in your app.

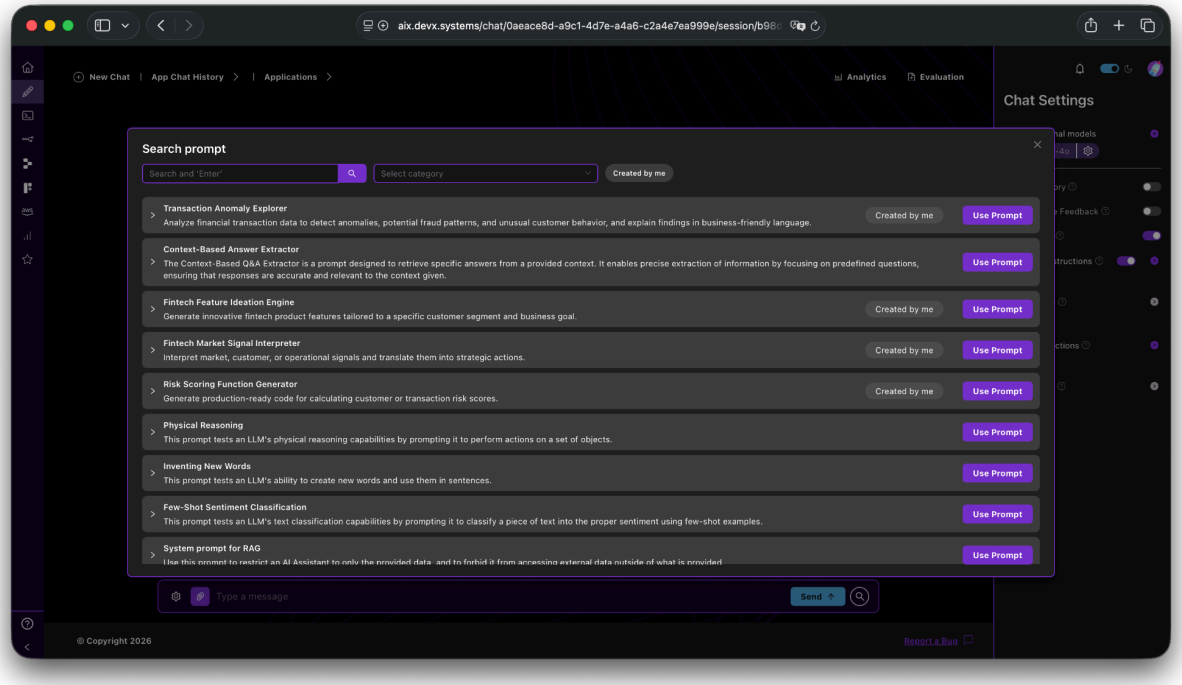

Figure 20 - Prompt Lab: Search Prompt

3.2. App Evaluation and Analytics

App Evaluation and Analytics provide tools to validate and monitor your applications. These features enable you to test your app against real or simulated data, measure accuracy and quality, and review detailed performance metrics, including user engagement and response behavior.

Evaluate the App

Use the Evaluate the App feature to test how your app performs against ground truth data. To use this feature, you should follow the steps:

1. Upload a CSV file with two required columns:
    - user_input: the input sent to the app.
    - reference: the expected/ground truth response.
    - The file must contain no more than 10 rows (only the first 10 will be used).
2. Add an evaluation name and description, upload your CSV, and click Create.
3. Select the LLM and embedding model for evaluation, then hit Run Evaluations.
4. Once evaluations are complete, results can be downloaded.
5. All evaluations are stored in the History of Evaluations for later review.

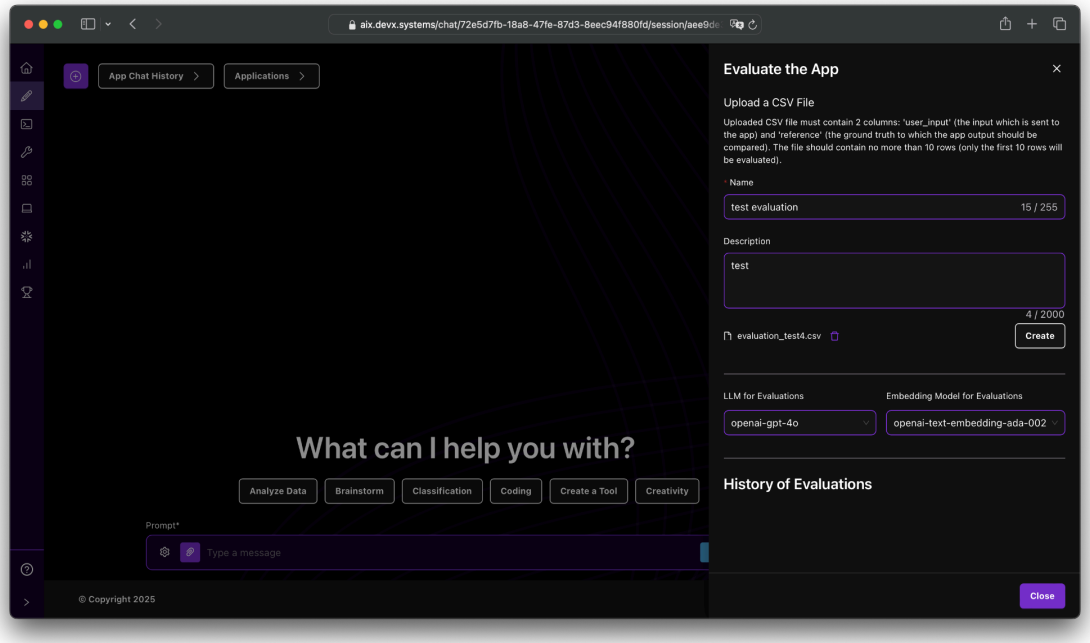

Figure 21 - Prompt Lab: Evaluate the App

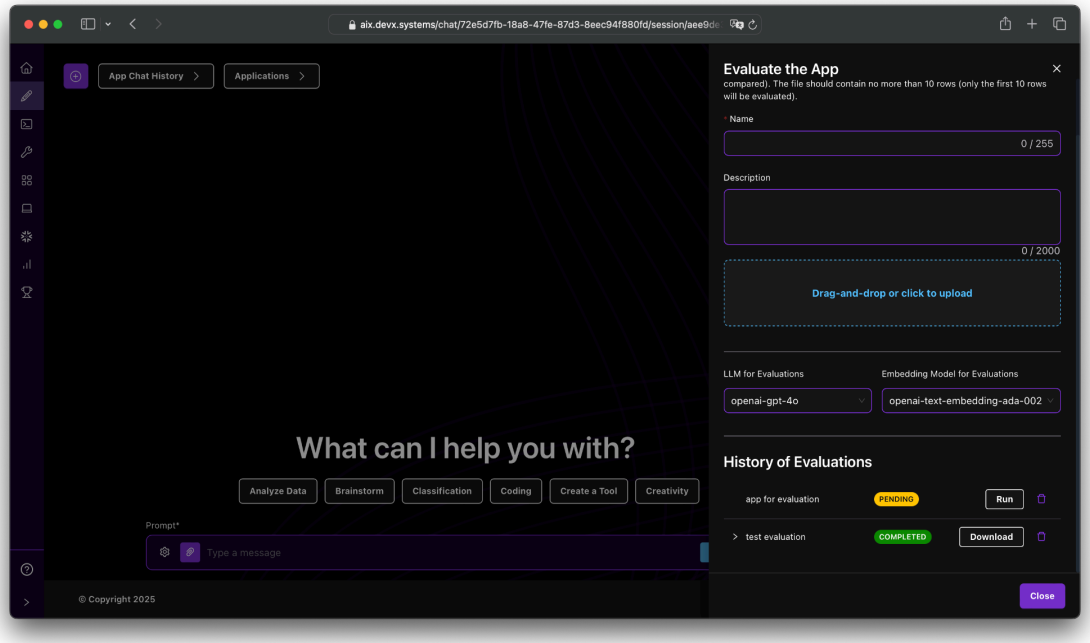

Figure 22 - Prompt Lab: History of Evaluations

Session Analytics Page

The Session Analytics page provides a detailed breakdown of how your app and selected models are performing. It is divided into three sections: Basic Metrics, Generation Metrics, and Retrieval Metrics.

Basic Metrics

- These metrics are generated directly from the model responses.
- The metrics include:
  - Provider name: the platform serving the model (e.g., Azure OpenAI, Amazon Bedrock).
  - Number of calls: how many requests were sent to the model.
  - Average response time: the average amount of time it takes for the model to respond.
  - Average response length: measured in tokens/characters.
  - Number of likes/dislikes: user feedback on responses.
  - Popularity score: aggregated score reflecting overall engagement.

Generation Metrics

- These metrics become available after running a model evaluation.
- These assess the quality of model outputs based on uploaded evaluation files.
- The Key indicators provided include:
  - Semantic similarity: how closely the model's output matches the expected meaning.

- Factual correctness: how accurate the generated response is compared to the reference truth.

Retrieval Metrics

- These metrics are shown only if an evaluation has been run with uploaded file(s).
- These focus on how well the system retrieves and uses context for answering.
- These metrics include:
  - Context precision: proportion of retrieved content that is relevant.
  - Context recall: proportion of all relevant content that was retrieved.
  - Response relevance: how well the response matches the input question or task.
  - Faithfulness: the degree to which the response remains grounded in the retrieved context.

Together, these metrics allow you to analyze not just performance (speed and length of responses) but also accuracy, quality, and contextual reliability of your AI applications.

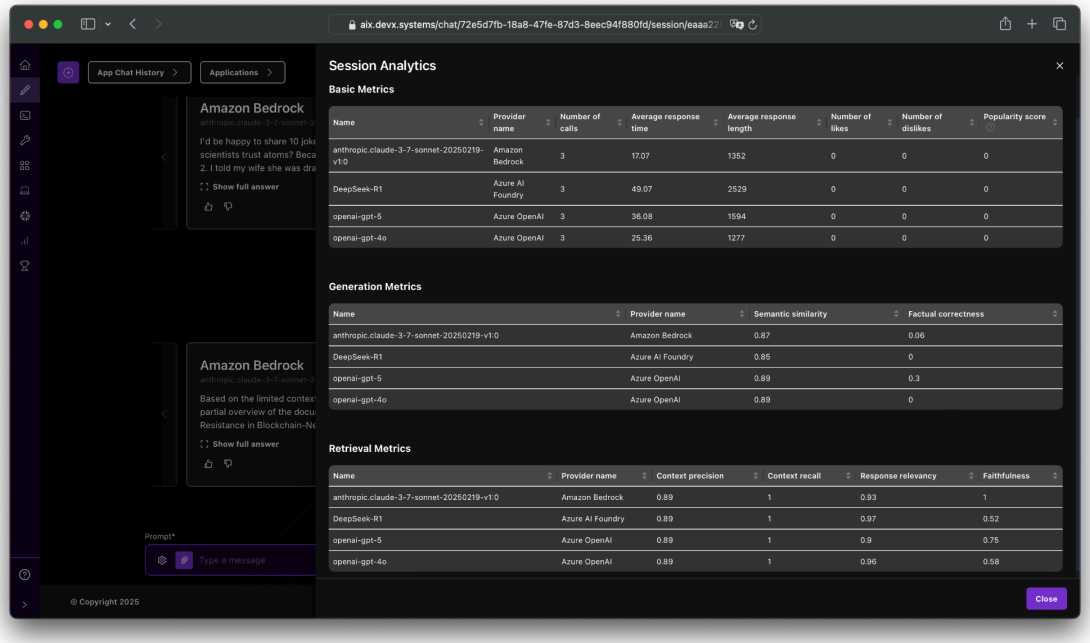

Figure 23 - Prompt Lab: Session Analytics

5. Prompt Library

The Prompt Library is a collection of ready-to-use prompts designed to help users quickly get started with different tasks. Prompts are organized into categories (e.g., Analyze Data, Brainstorm, Creativity, Coding), making it easy to explore by type or purpose.

5.1. Published prompts

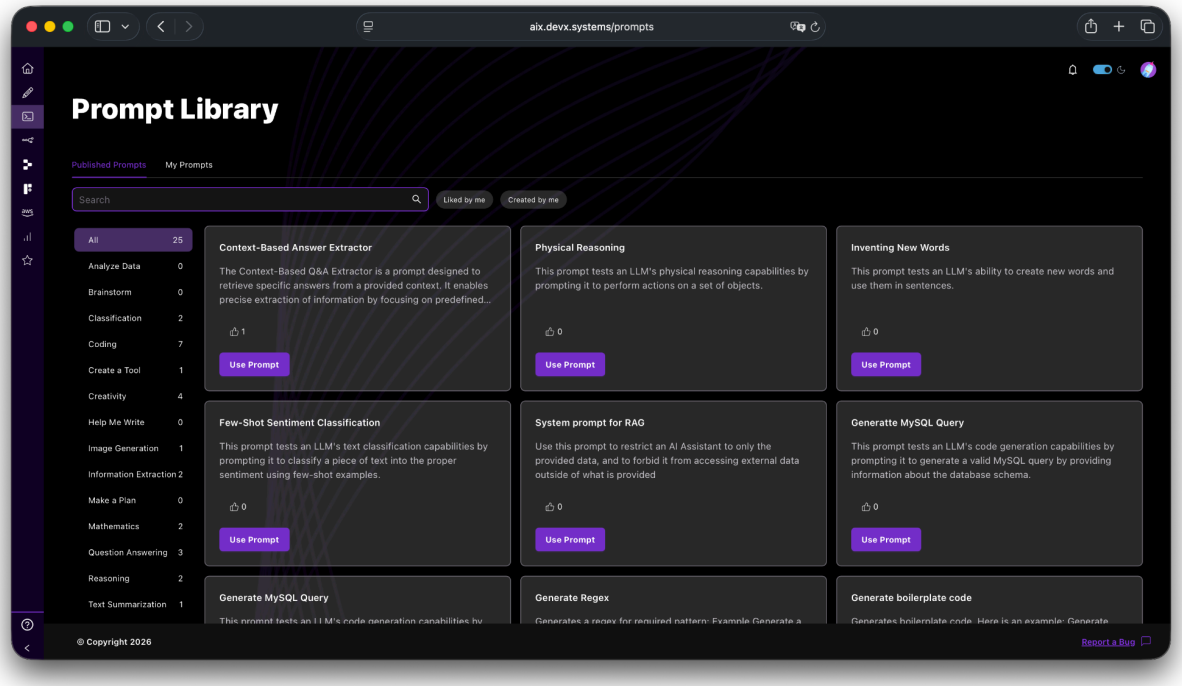

Figure 24 - Prompt Library

Prompt Details: Each prompt includes a title, description, and instructions to clarify its purpose. If you need details, click on the prompt tile, and the complete information will open on the right side.

By clicking **Use Prompt**, the prompt content is automatically copied into the Prompt Lab, where you can iterate, edit, and reuse it for your specific needs.

Users can also **like prompts** to save them for quick access later.

At the top of the Prompt Library, the following filters are available:

- **Liked by me** – Displays only prompts you have liked
- **Created by me** – Displays only prompts you have created and published

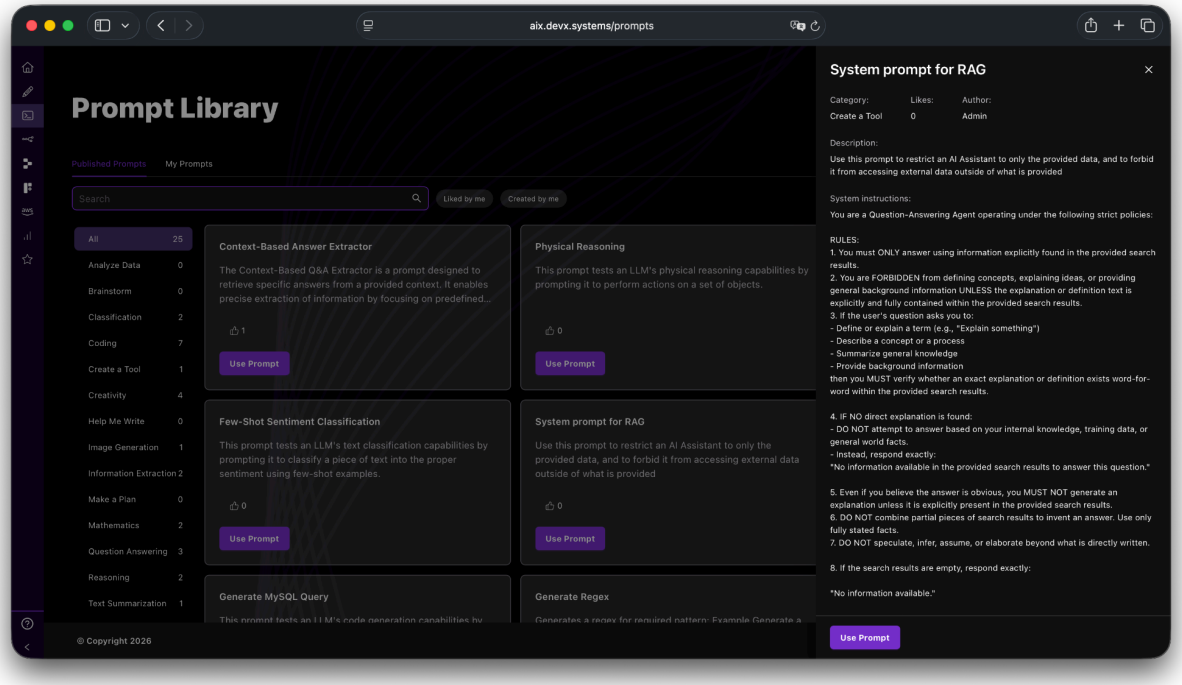

Figure 25 - Prompt Details

Prompt Details: Each prompt includes a title, description, and instructions to clarify its purpose. If you need details, click on the prompt tile, and the complete information will open on the right side.

5.2. My Prompts

The My Prompts section allows users to create and manage their own custom prompts.

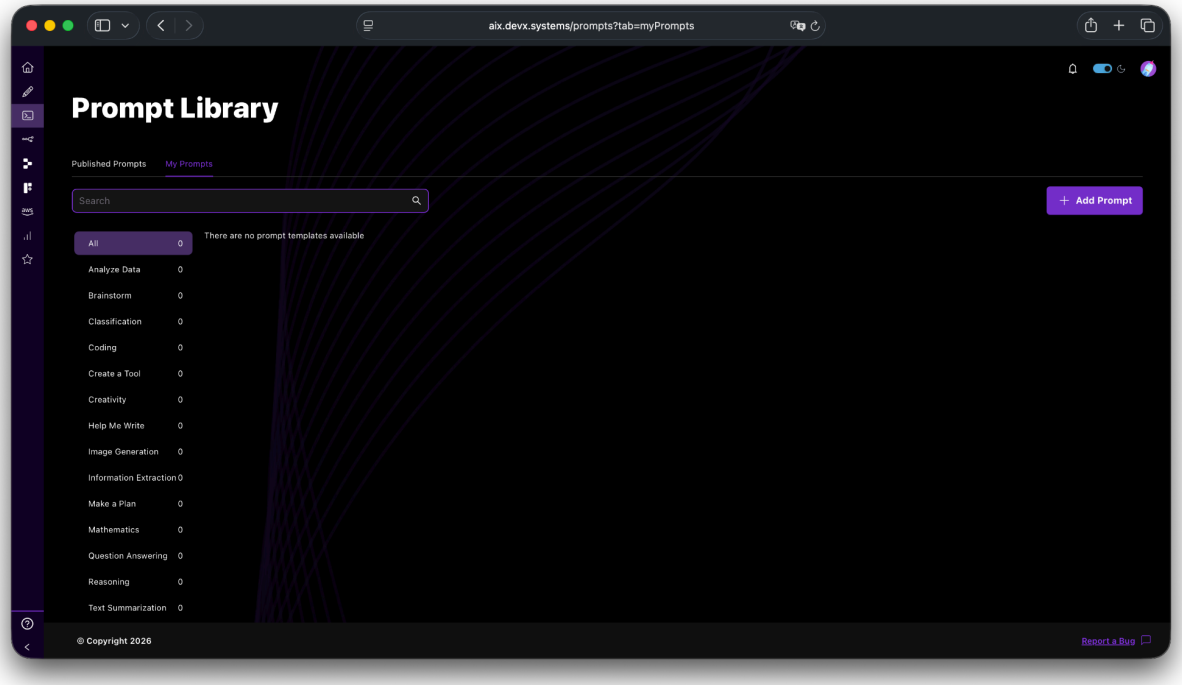

Figure 26 - My Prompts

To create a new prompt, click Add Prompt and provide:

- Category
- Prompt name
- Description
- System instructions
- Prompt content

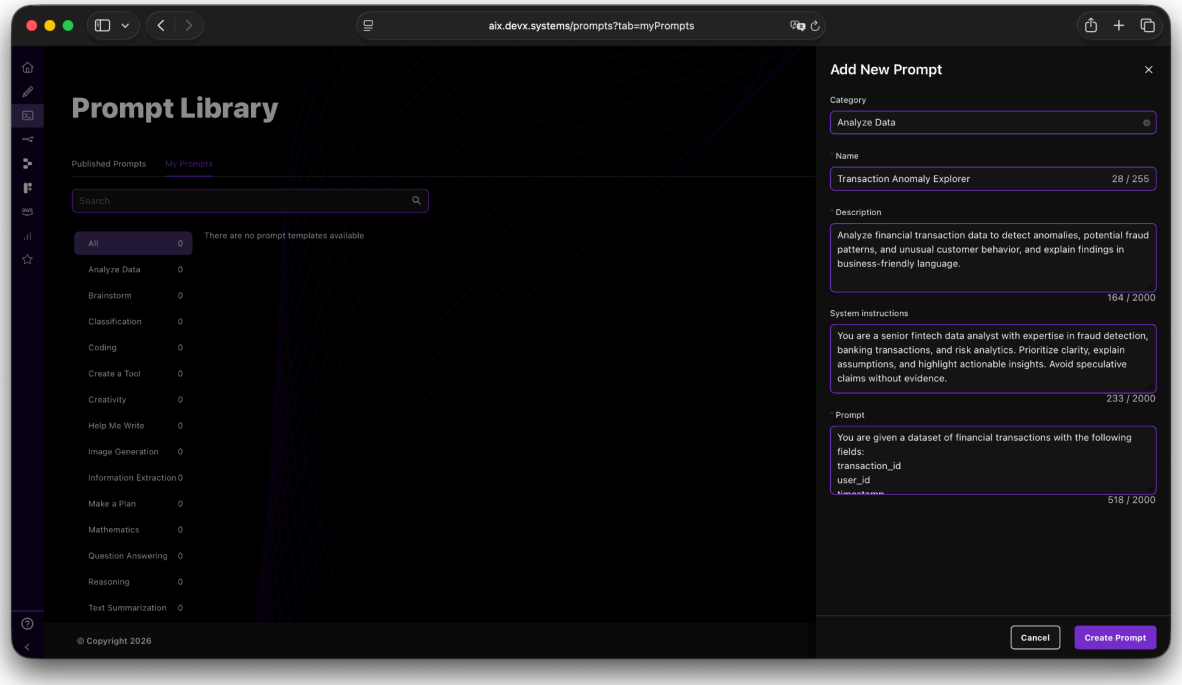

Figure 27 - Add New Prompt

Once created, prompts in My Prompts are private by default and visible only to you. From this section, you can:

- Edit existing prompts
- Delete prompts
- Use a prompt (which sends it to the Prompt Lab for editing and reuse)
- Publish prompt

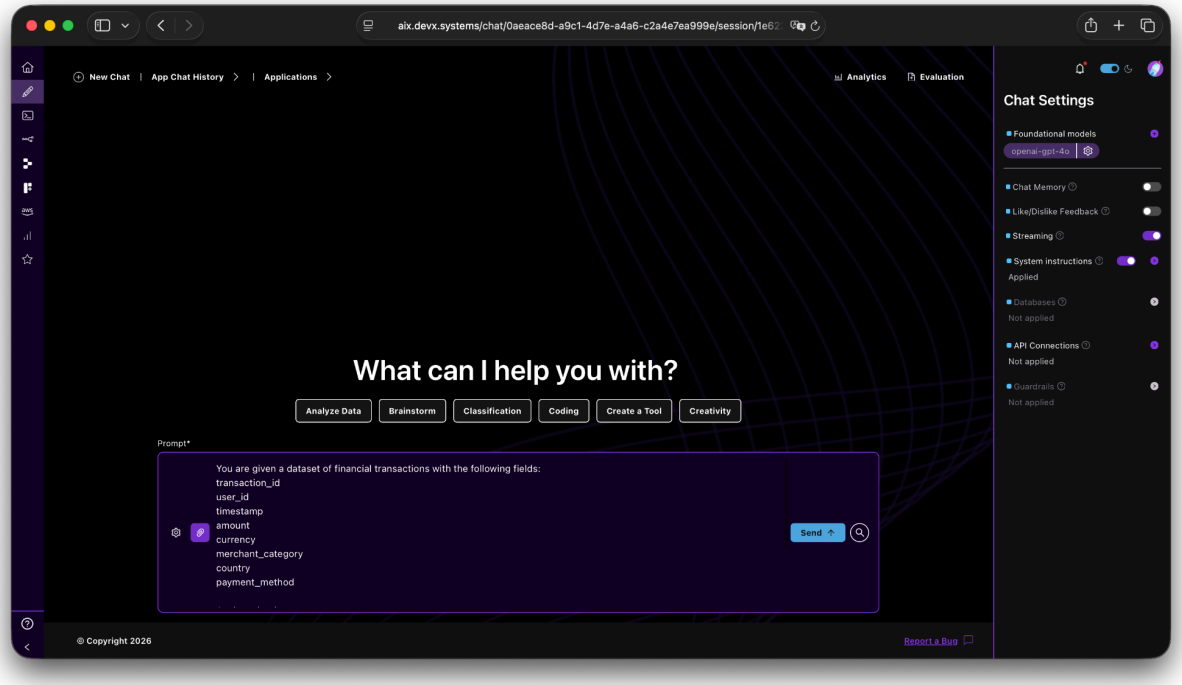

Figure 28 - Using a Prompt from My Prompts in Prompt Lab

Publishing Prompts: Users can choose to publish their prompts, making them available to all users on the platform via the Prompt Library.

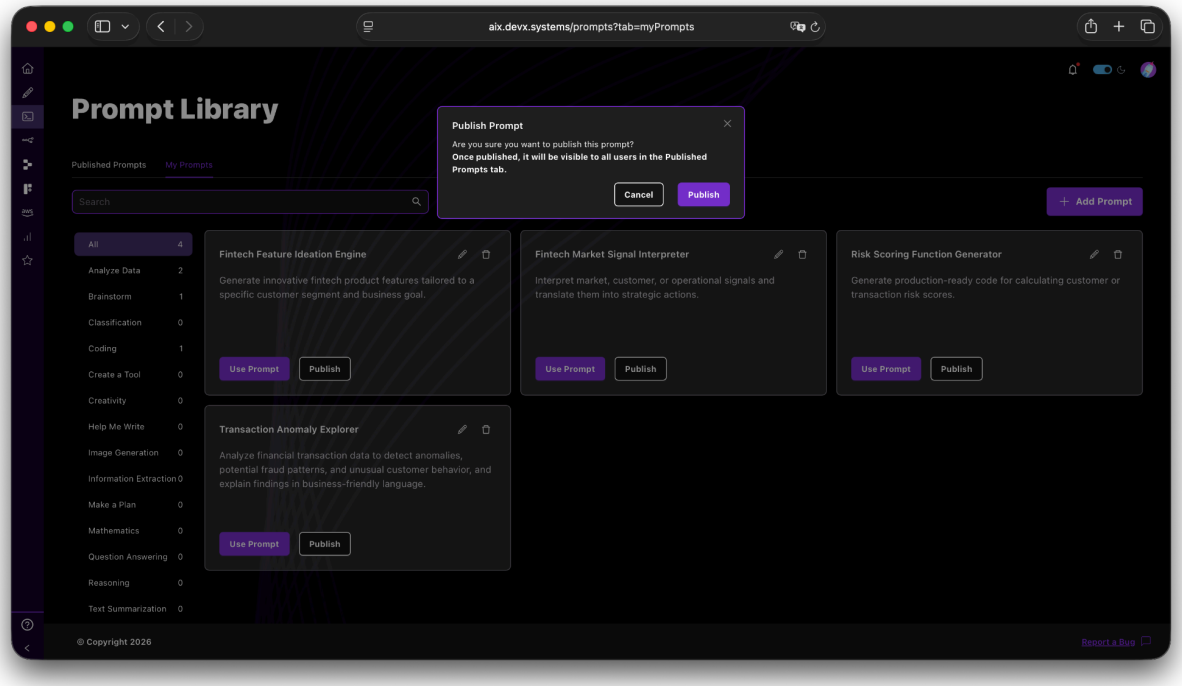

Figure 29 - Publish a Prompt

Once published, the prompt appears alongside other public prompts and can be discovered, liked, and reused by the community.

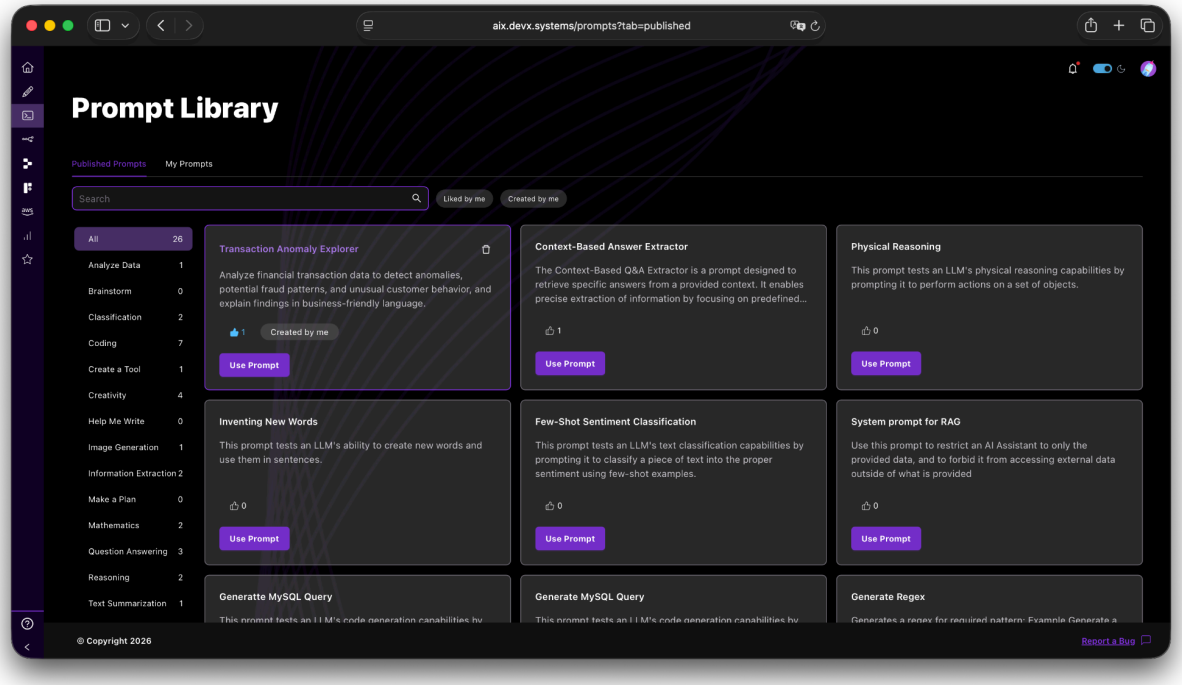

Figure 30 - Published Prompt in Prompt Library

⚠️ **Important:** If you delete a prompt that has been published, this action is irreversible, and the prompt will be permanently removed from the platform. Please proceed with caution.

6. Replit Integration

The Replit integration in AI Sandbox provides a seamless way to experiment with code, prototype quickly, and extend your AI applications. By connecting directly to Replit through Azure SSO, users can securely access an online coding environment without additional setup. This integration is designed for flexibility — whether you're making small script adjustments or building complex extensions, Replit is your coding companion within AI Sandbox.

How to Access Replit:

1. Click the Vibe Coding Editor link from the left menu.
2. A new browser tab will open, redirecting you to the Azure SSO login.
3. After authentication, the Replit online editor will load automatically.

Plans and Licensing:

- Licenses for **Replit Core** are assigned to the team lead by default. If a license needs to be reassigned to another team member, please contact dtccaihackathon@dtcc.com.

Documentation & References

For detailed usage, refer to the official Replit documentation.

7. n8n integration

The n8n integration in AI Sandbox provides a no-code/low-code environment where users can build workflows, connect services, and extend AI apps with automation. Once workflows are created, they can also be published to the AI Sandbox App Marketplace, making them available for others.

Accessing n8n:

1. Click the n8n link from the left navigation menu.
2. A new browser tab will open and redirect you to the Azure SSO login.
3. After successful authentication, the n8n workflow editor will load, ready for use.

Using Azure OpenAI Models in n8n

The Azure OpenAI Chat Model node enables you to utilize Azure-hosted models within your n8n workflows.

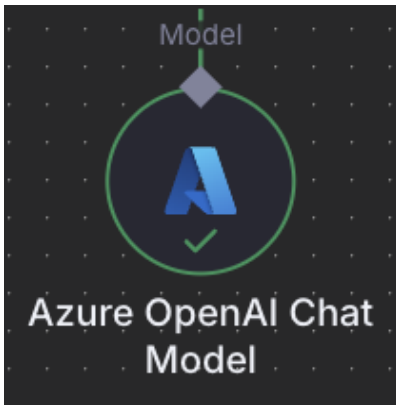

Figure 31 - n8n node for using Azure OpenAI models.

To use this node, you need to select the Azure OpenAI node, use pre-shared by admin credentials, and specify any of the following models:

- openai-gpt-4o
- openai-gpt-4o-mini
- openai-gpt-5
- openai-o1
- openai-o3-mini
- openai-gpt-5.2
- openai-gpt-5.2-chat

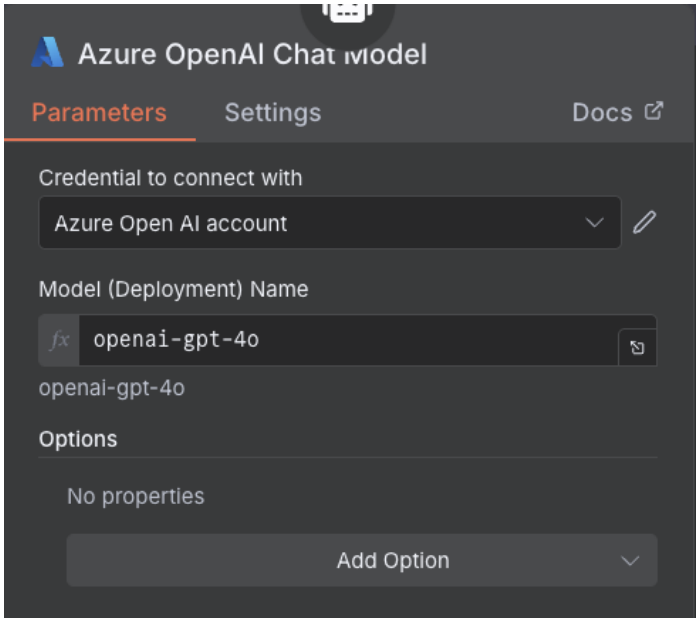

Figure 32 - Configuration settings for Azure OpenAI Chat Model node.

The Azure OpenAI account credentials will be pre-created by an administrator and shared with each user. After selecting one of the models in the Azure OpenAI Chat Model node, the workflow can be run successfully.

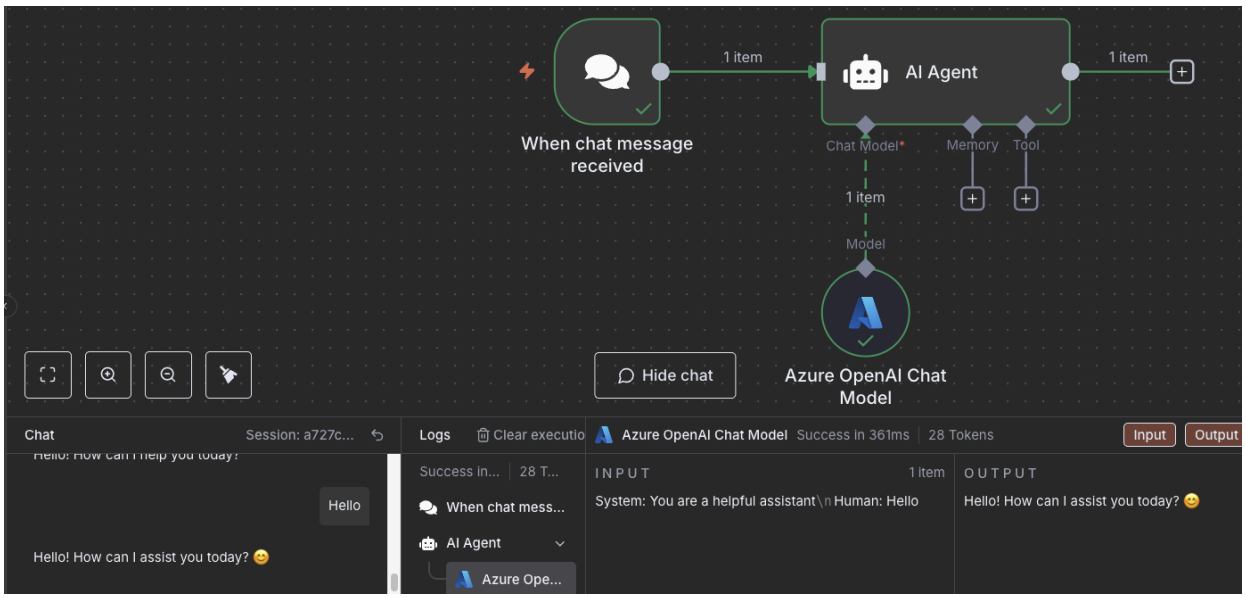

Figure 34: Successfully running a workflow using Azure OpenAI Chat Model.

Using AWS Bedrock Models in n8n

Using the AWS Bedrock models in a workflow can be done through the AWS Bedrock Chat Model node.

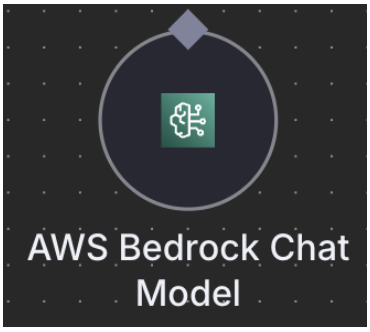

Figure 35: n8n node for using AWS Bedrock models in a workflow..

The AWS Credential for Bedrock Models only (long living) must be set for this node.

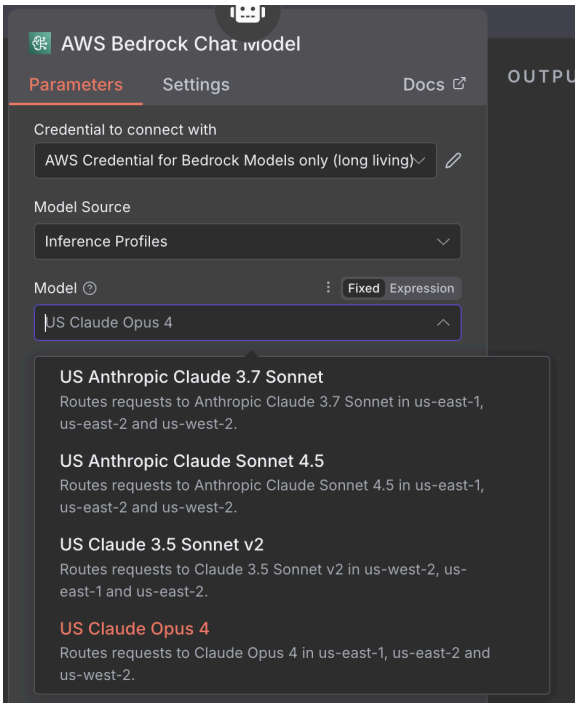

Figure 36: Configuration parameters for AWS Bedrock Chat Model node.

Any model can be selected from the dropdown list.

Alternatively, you can use an expression to specify the arn identifier of a model (Figure 37). You can replace the model identifier in the ARN with any of the following to use the latest anthropic models:

- anthropic.claude-sonnet-4-5-20250929-v1:0

- anthropic.claude-haiku-4-5-20251001-v1:0
- anthropic.claude-opus-4-6-v1

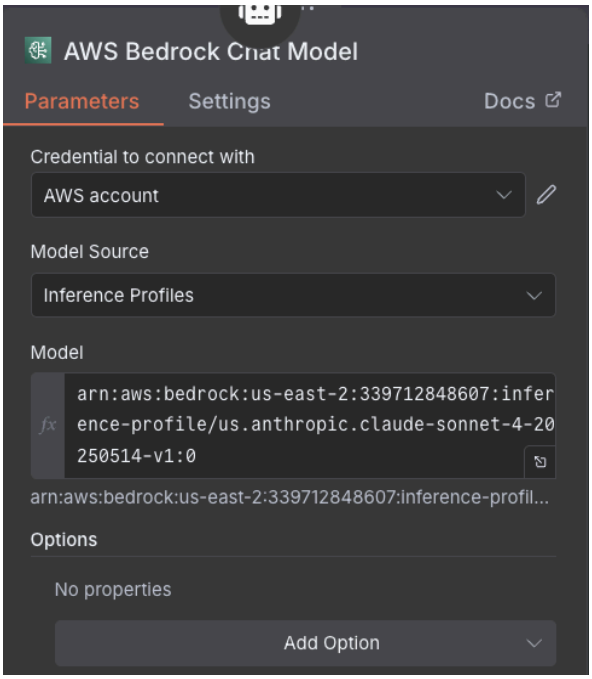

Figure 37: Selecting a model is done by manually filling its ID in the model field.

Once the above configurations have taken place, you can run the workflow and interact with the AWS Bedrock model:

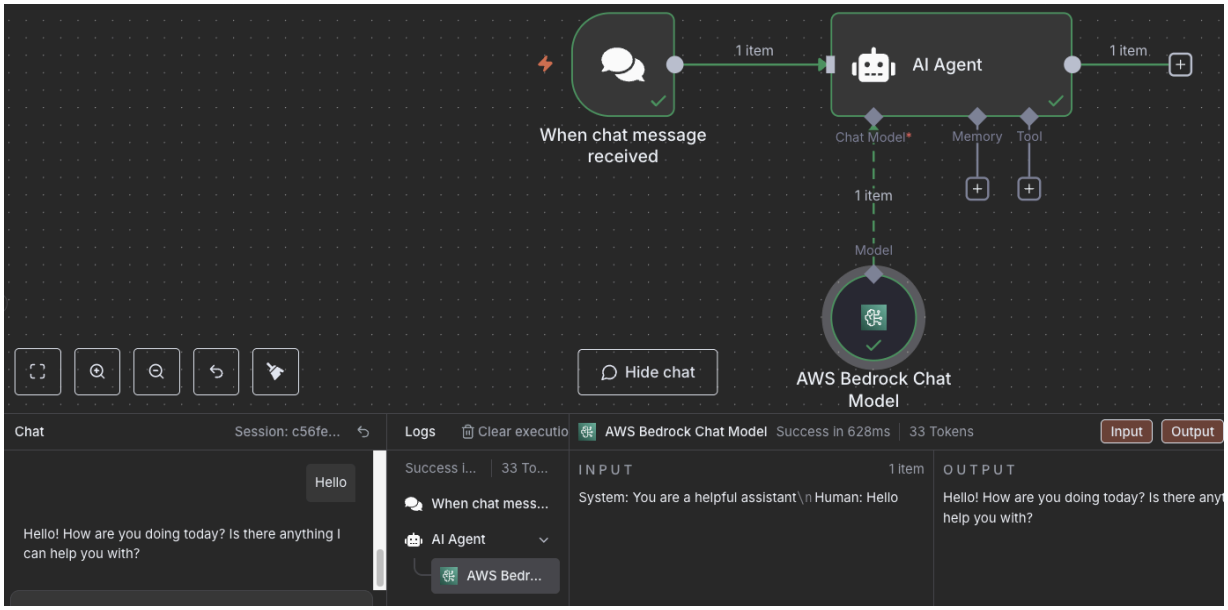

Figure 39: Successfully running a basic chatbot workflow with AWS Bedrock Chat Model.

Documentation & References: For detailed usage, refer to the official [n8n documentation].

8. AWS integration

The AWS integration in AI Sandbox allows users to access, provision, and manage AWS resources directly from the platform. This enables experimentation with cloud services, deployment of AI/ML models, and integration of AWS-based workflows into Sandbox projects.

Accessing AWS

Use the AWS link from the left navigation menu in AI Sandbox.

The link opens in a new browser tab and redirects you to the Azure SSO login before loading the AWS Console.

Deploying AWS Resources

- Once logged in, you can launch AWS services directly through the AWS Console.
- The range of available services depends on your assigned role.
- Common use cases include deploying:
    - Compute resources (e.g., EC2 instances, Lambda functions).
    - Storage (e.g., S3 buckets).
    - AI/ML services, including AWS Bedrock models.

Support and Documentation

- For step-by-step instructions on using AWS services, refer to the AWS Documentation.
- For access or permission issues, contact the AI Sandbox support team.

9. LLM Analytics

The LLM Analytics page offers insights into the performance of various Large Language Models (LLMs) within the AI Sandbox. It allows you to evaluate and compare providers based on usage, performance, and quality metrics.

The analytics table displays detailed statistics for each LLM model, including:

- Name: Identifier of the model.
- Provider name: The platform or service hosting the model (e.g., Amazon Bedrock, Snowflake Cortex, Azure AI).
- Number of calls: How many times the model was queried in the AI Sandbox.
- Average response time: Average time (in seconds) the model takes to respond.
- Average response length: Typical length of the model's responses (in tokens or characters).
- Number of likes: How many positive ratings the model's outputs received from users.
- Number of dislikes: How many negative ratings the model's outputs received.
- Popularity score: Calculated score based on usage and user feedback, showing how well the model is received in the Sandbox.
- Semantic similarity: Measures how semantically aligned the model's responses are with expected answers (higher = closer).
- Factual correctness: Indicates whether the model's responses are factually accurate.
- Context precision: How well the model uses the provided context without adding irrelevant information.
- Context recall: How effectively the model incorporates all relevant details from the provided context.
- Response relevancy: Indicates how relevant the model's outputs are to the user's query.
- Faithfulness: Demonstrates how well the model adheres to the provided data without introducing unnecessary information (hallucinations).
- Last used at: Timestamp of the most recent use of the model.
- Is active: A status indicator that shows whether the model is currently active/available.

The LLM Analytics page helps you

- Identify which models perform best for your use case.
- Compare providers (speed, accuracy, quality).
- Monitor trends in popularity and user satisfaction over time.
- Select the most reliable models for production or experimentation.

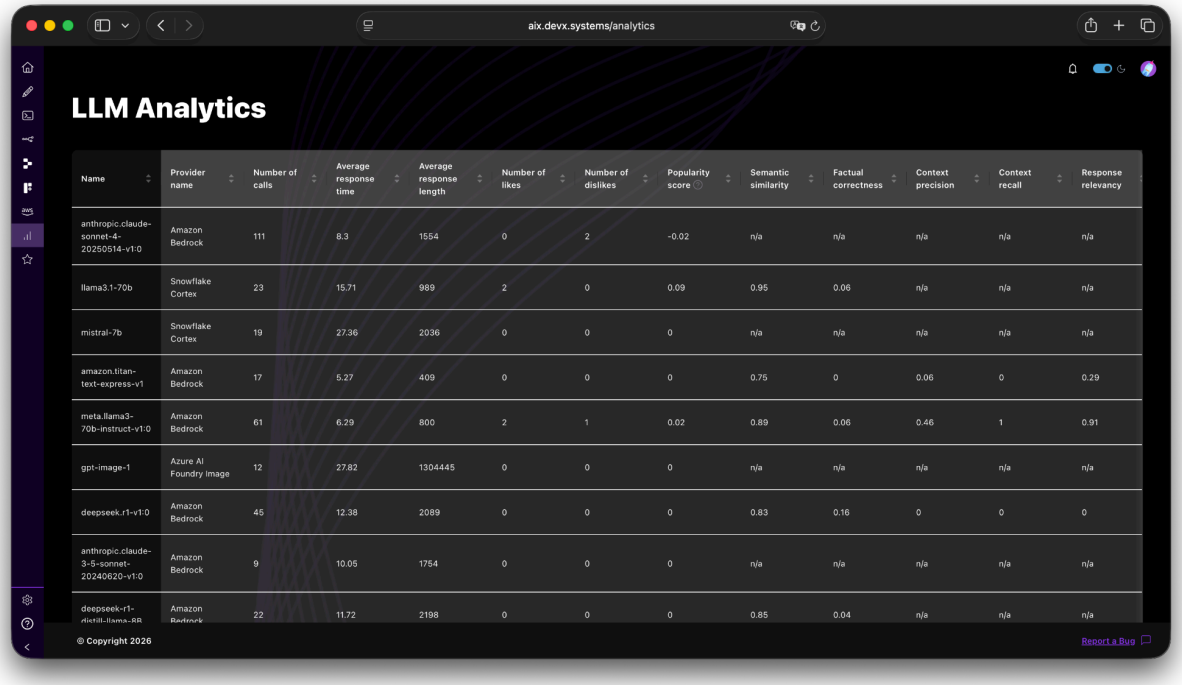

Figure 40 - LLM analytics page

10. App Resource Management

The App Resource Management page, accessible from the Settings, allows hackathon participants to bring their own external database and API connections to be used in the Prompt Lab.

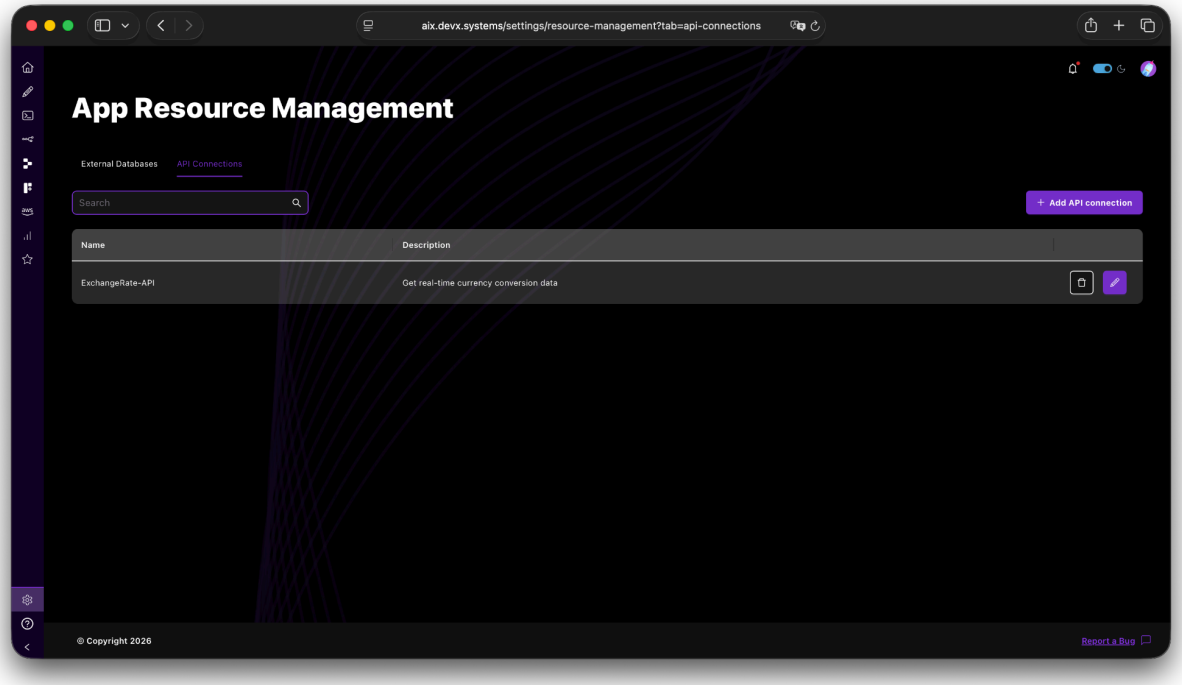

Figure 41 - App Resource Management Page

Concretely, from this page, users can:

- Add new database or API connections
- Edit existing connections
- Delete connections that are no longer needed

To add a resource, click **Add Database** or **Add API Connection** and provide the required configuration details. When creating a new connection, users must supply all necessary connection information based on the selected resource type.

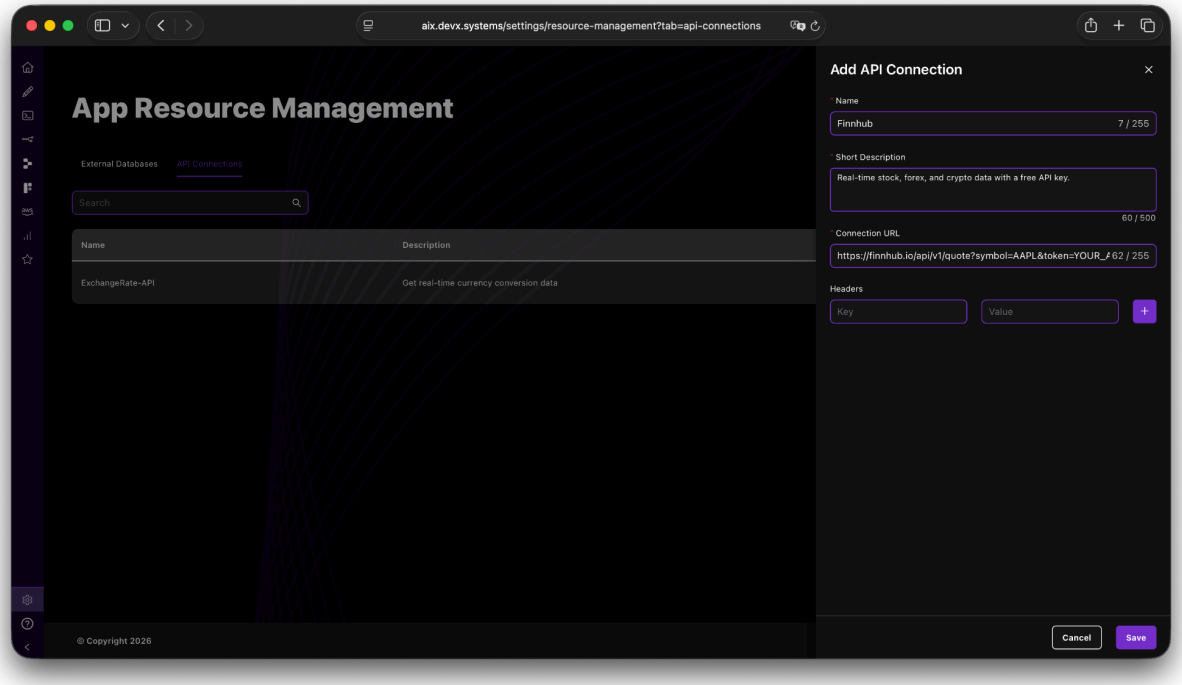

Figure 42 - Add API Connection

Once added, you can use these connections in the **Prompt Lab** when executing prompts.

In the **Prompt Lab**, users can select from available connections when configuring a chat:

- Added by me – Connections created by the user
- Added by admin – Connections provided by administrators

When an API or database connection is selected, it is applied to the current chat session.

⚠️ Note: Not all foundational models support database or API connections. If a selected model is incompatible, related options will appear disabled.

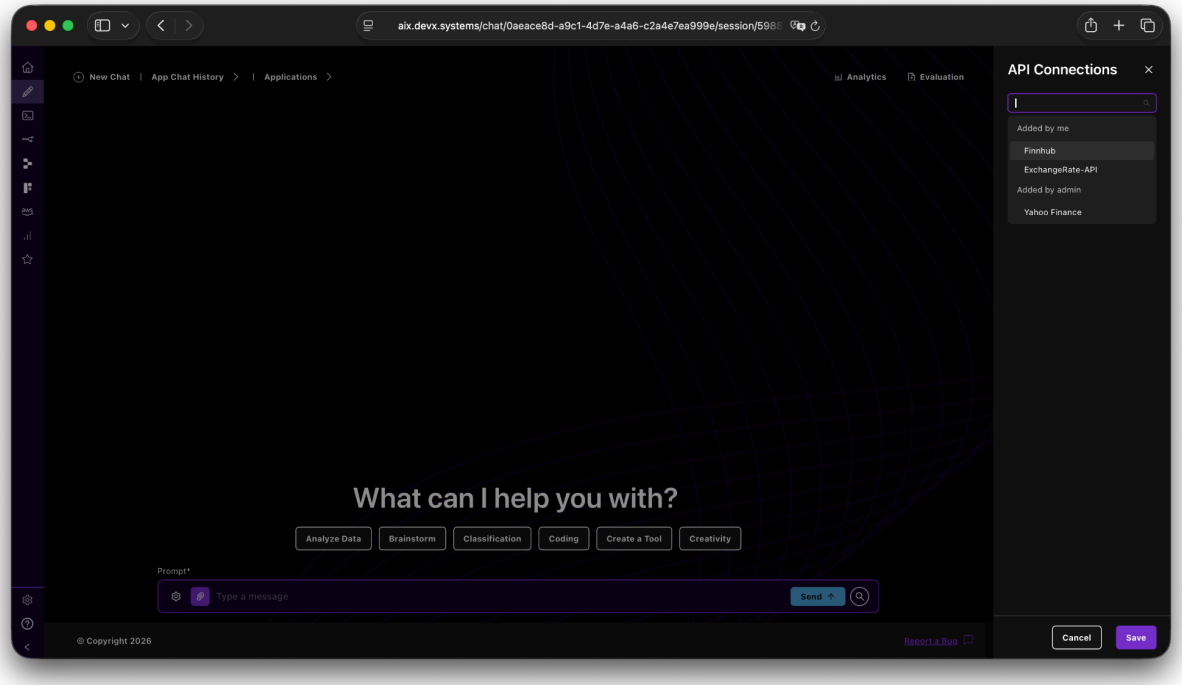

Figure 43 - Selecting API Connections in Prompt Lab

When sending a prompt that references an active API or database connection, the system response may include information retrieved directly from that resource.

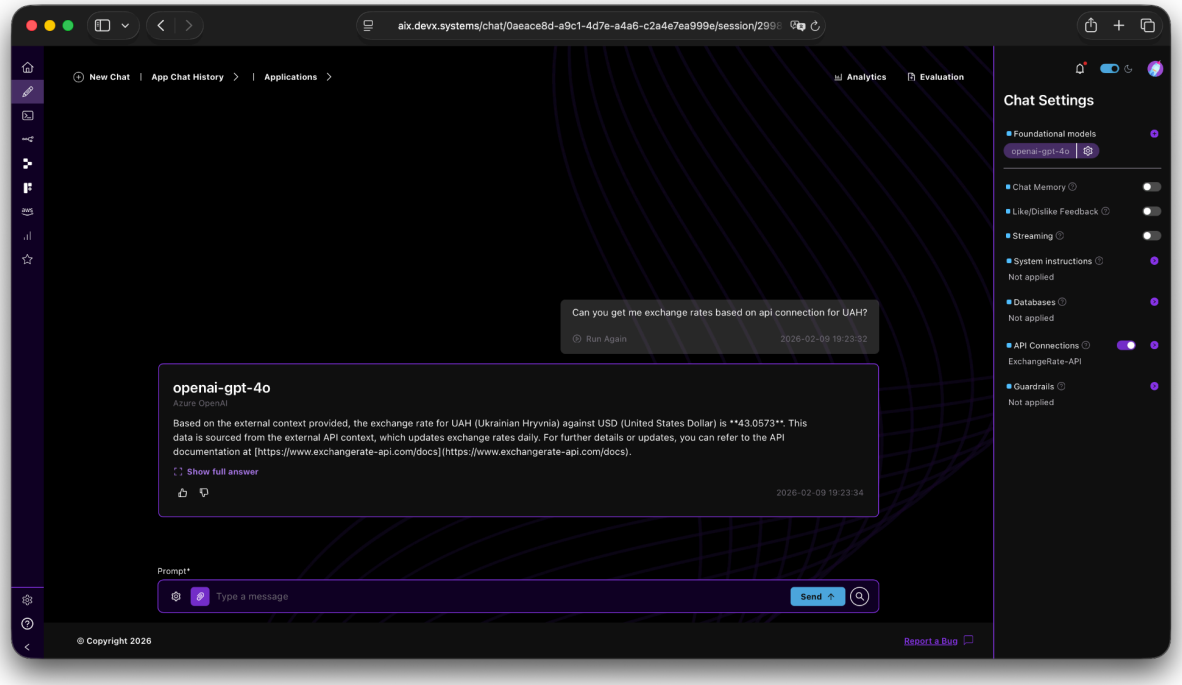

Figure 44 - Prompt Execution with API/Database Connection