

Requests



Alejandra Caggiano

Observability



GeneXus Enterprise AI stores and tracks every request, providing organizations with full visibility into the use of its assistants, artificial intelligence models, and the cost associated with each request.

Observability

- Monitor and analyze resource usage
- Make informed decisions
- Optimize the use of resources
- Manage workflow
- Control expenses
- Maximize return on investment

This allows organizations to monitor and analyze resource usage, make informed decisions on the allocation of these resources, and optimize their use to achieve profitability.

With clear visibility into both usage and cost, organizations can effectively manage their AI-driven workflows, control expenses and maximize ROI.

Observability

Organizations can have control over artificial intelligence infrastructure

By leveraging these features, organizations can maintain control over their artificial intelligence infrastructure, identify areas for improvement, and make data-driven decisions to improve operational efficiency.

Requests

The image displays three overlapping screenshots of the GeneXus Enterprise AI interface, illustrating the request tracking functionality.

Top Left Screenshot: Request List
 This view shows a table of requests with the following columns: Module, Assistant Name, AIModel Name, Api Token Name, Input, Timestamp, Time (Ms), and Status. The table lists several requests, all with a status of 'Succeeded'. The interface includes a sidebar with navigation options like Dashboard, Assistants, and AI Models, and a top navigation bar with filters for Project, Timestamp, and Module.

Module	Assistant Name	AIModel Name	Api Token Name	Input	Timestamp	Time (Ms)	Status
HttpProxy	MarketingAssistant	gpt-3.5-turbo-036	Default	chat	04/10/2024 10:45:02	802	Succeeded
HttpProxy	MarketingAssistant	gpt-3.5-turbo-036	Default	chat	04/10/2024 10:45:03	796	Succeeded
HttpProxy	MarketingAssistant	gpt-3.5-turbo-036	Default	chat	04/10/2024 10:45:04	804	Succeeded
HttpProxy	MarketingAssistant	gpt-3.5-turbo-036	Default	chat	04/10/2024 10:45:05	794	Succeeded
HttpProxy	MarketingAssistant	gpt-3.5-turbo-036	Default	chat	04/10/2024 10:45:06	797	Succeeded
HttpProxy	MarketingAssistant	gpt-3.5-turbo-036	Default	chat	04/10/2024 10:45:07	803	Succeeded
HttpProxy	MarketingAssistant	gpt-3.5-turbo-036	Default	chat	04/10/2024 10:45:08	807	Succeeded
HttpProxy	N/D	N/D	Default		04/10/2024 10:46:47	38	
HttpProxy	N/D	N/D	Default		04/10/2024 10:46:51	37	
HttpProxy	N/D	N/D	Default		04/10/2024 10:48:23	40	
HttpProxy	N/D	N/D	Default		04/10/2024 10:48:25	46	

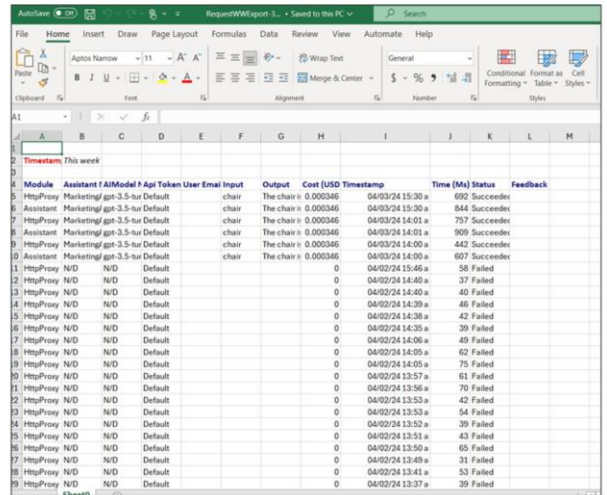
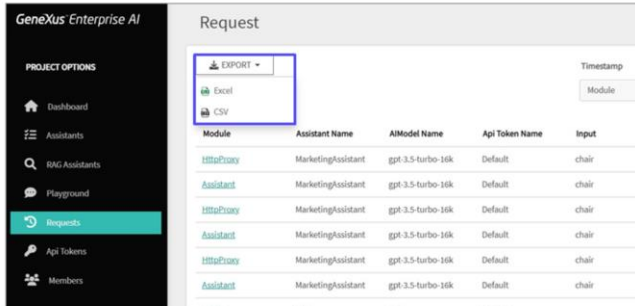
Top Right Screenshot: HttpProxy Details
 This view shows the detailed information for a specific request. It includes tabs for General (Pretty), General (JSON), Usage Data, Metadata, and Request Log. The General (Pretty) tab is active, displaying the request ID, timestamp, and the input/output of the request. The input is a JSON object with a 'chat' property, and the output is a JSON object with a 'chat' property containing a response.

Bottom Left Screenshot: HttpProxy Request Log
 This view shows the request log for the selected request. It includes a table with columns for Log ID, Log Name, Log Date, and Log Timestamp. The log contains several entries, each representing a step in the request processing, such as 'HttpCall', 'HttpCall', and 'HttpCall'. The interface includes a sidebar with navigation options and a top navigation bar with filters for Project and Timestamp.

From the platform, in the left menu, the “Requests” option gives access to comprehensive tracking that provides full observability of each request made.

This tracking makes it easy to filter requests by model, assistant, API token, date and time range, and status, which in turn allows us to quickly identify specific requests of interest. In addition, by clicking on the Module column of a particular request, we can access its detailed information.

Requests



Within these details, we can view the input and output data, the specific model used for the request, the associated cost, and the timestamp indicating when the request was executed.

This ability to access and review the full details of each request allows us to understand the underlying data and processes, which in turn facilitates the identification and addressing of any issues or areas for improvement.

If we go back to the initial dashboard, we see that it is also possible to export these details, for example, to Excel.

With this comprehensive observability feature, organizations can ensure the accuracy, efficiency and cost-effectiveness of their AI-driven workflows in GeneXus Enterprise AI.

GeneXus[™]
by **Globant**

training.genexus.com