

Requests



Alejandra Caggiano

Observabilidad



GeneXus Enterprise AI almacena y rastrea cada solicitud, brindando a las organizaciones una visibilidad completa del uso de sus asistentes, modelos de inteligencia artificial y el costo asociado a cada solicitud.

Observabilidad

- Monitorear y analizar el uso de recursos.
- Tomar decisiones informadas.
- Optimizar el uso de recursos.
- Administrar el flujo de trabajo.
- Controlar los gastos.
- Maximizar el retorno de la inversión..

Esto permite a las organizaciones monitorear y analizar el uso de recursos, tomar decisiones informadas sobre la asignación de estos recursos y optimizar su uso para lograr rentabilidad.

Con una visibilidad clara, tanto del uso como del costo, las organizaciones pueden administrar eficazmente sus flujos de trabajo impulsados por inteligencia artificial, controlar los gastos y maximizar el retorno de la inversión.

Observabilidad

Las Organizaciones pueden mantener el control sobre la infraestructura de inteligencia artificial.

Al aprovechar estas características, las organizaciones pueden mantener el control sobre su infraestructura de inteligencia artificial, identificar áreas de mejora y tomar decisiones basadas en datos para mejorar la eficiencia operativa.

Requests

The image displays three overlapping screenshots of the GeneXus Enterprise AI interface, illustrating the 'Requests' management functionality.

Top Left Screenshot: Request List
 This view shows a table of requests with the following columns: Module, Assistant Name, AI Model Name, API Token Name, Input, Timestamp, Time (Ms), and Status. The table lists several requests, all with a status of 'Succeeded'. A sidebar on the left contains navigation options like 'Dashboard', 'Assistants', 'API Assistants', 'Playground', 'API Tokens', 'Members', and 'Organization Options'.

Module	Assistant Name	AI Model Name	API Token Name	Input	Timestamp	Time (Ms)	Status
HttpProxy	MarketingAssistant	gpt-3.5-turbo-01a	Default	chat	14/10/2024 10:30:45.042 PM	832	Succeeded
HttpProxy	MarketingAssistant	gpt-3.5-turbo-01a	Default	chat	14/10/2024 10:30:45.098 PM	844	Succeeded
HttpProxy	MarketingAssistant	gpt-3.5-turbo-01a	Default	chat	14/10/2024 10:30:45.490 PM	257	Succeeded
HttpProxy	MarketingAssistant	gpt-3.5-turbo-01a	Default	chat	14/10/2024 10:30:55.418 PM	509	Succeeded
HttpProxy	MarketingAssistant	gpt-3.5-turbo-01a	Default	chat	14/10/2024 10:30:55.697 PM	482	Succeeded
HttpProxy	MarketingAssistant	gpt-3.5-turbo-01a	Default	chat	14/10/2024 10:30:55.714 PM	507	Succeeded
HttpProxy	N/D	N/D	Default		14/10/2024 10:30:47.088 PM	58	
HttpProxy	N/D	N/D	Default		14/10/2024 10:30:47.117 PM	37	
HttpProxy	N/D	N/D	Default		14/10/2024 10:30:47.213 PM	40	
HttpProxy	N/D	N/D	Default		14/10/2024 10:30:47.218 PM	46	

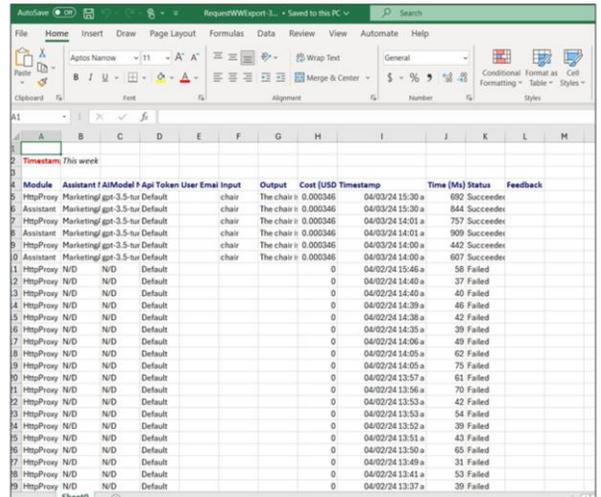
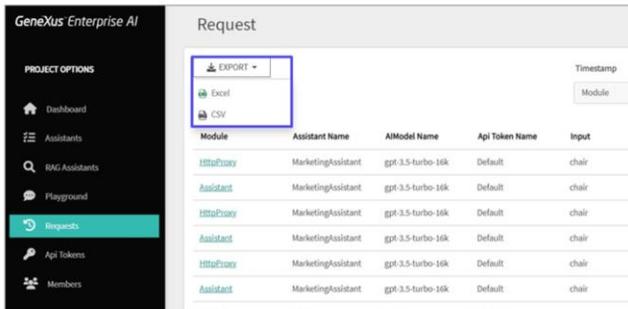
Top Right Screenshot: HttpProxy Details
 This view shows the details for a specific request. It includes tabs for 'General (Pretty)', 'General (JSON)', 'Usage Data', 'Metadata', and 'Request Log'. The 'General (Pretty)' tab is active, displaying fields such as 'id', 'Timestamp', 'Trace id', 'Assistant', 'API Token Full Name', 'Default Chatting Playground Project', 'AI Model Name', 'Assistant Name', and 'Module'. A note at the bottom states: 'The Chat is a tool with a back, usually with four legs, and that can only fit one person.'

Bottom Left Screenshot: HttpProxy Request Log
 This view shows the 'Request Log' for the selected request. It displays a list of log entries with columns for 'Log ID', 'Log Name', 'Log Date', and 'Log Timestamp'. The log entries contain detailed information about the request, including the input and the output.

Desde la plataforma, en el menú izquierdo, la opción “Requests” da acceso a un seguimiento integral que brinda completa observabilidad de cada solicitud realizada.

El seguimiento permite filtrar fácilmente solicitudes por modelo, asistente, api token, rango de fecha y hora y estado, lo que permite a su vez identificar rápidamente solicitudes de interés específicas. Además, al hacer clic en la columna Module de una solicitud en particular, accedemos a información detallada sobre esa solicitud.

Requests



Dentro de los detalles de la solicitud, podemos ver los datos de entrada y salida, el modelo específico utilizado para la solicitud, el costo asociado y la marca de tiempo que indica cuándo se ejecutó la solicitud.

Este nivel de detalle le brinda información sobre los datos específicos de cada solicitud, lo que facilita la resolución de problemas, el análisis y la optimización de sus flujos de trabajo de Inteligencia artificial.

La capacidad de acceder y revisar los detalles completos de cada solicitud permite comprender los datos y procesos subyacentes, lo que facilita poder identificar y abordar cualquier problema o área de mejora.

Si volvemos al panel inicial, vemos que también es posible exportar este detalle, por ejemplo, a Excel.

Con esta característica de observabilidad integral, las organizaciones pueden garantizar la precisión, eficiencia y rentabilidad de sus flujos de trabajo impulsados por Inteligencia Artificial en GeneXus Enterprise AI.

GeneXus[™]
by **Globant**

training.genexus.com