

Creación y uso de APIs para interactuar con asistentes



Alejandra Caggiano

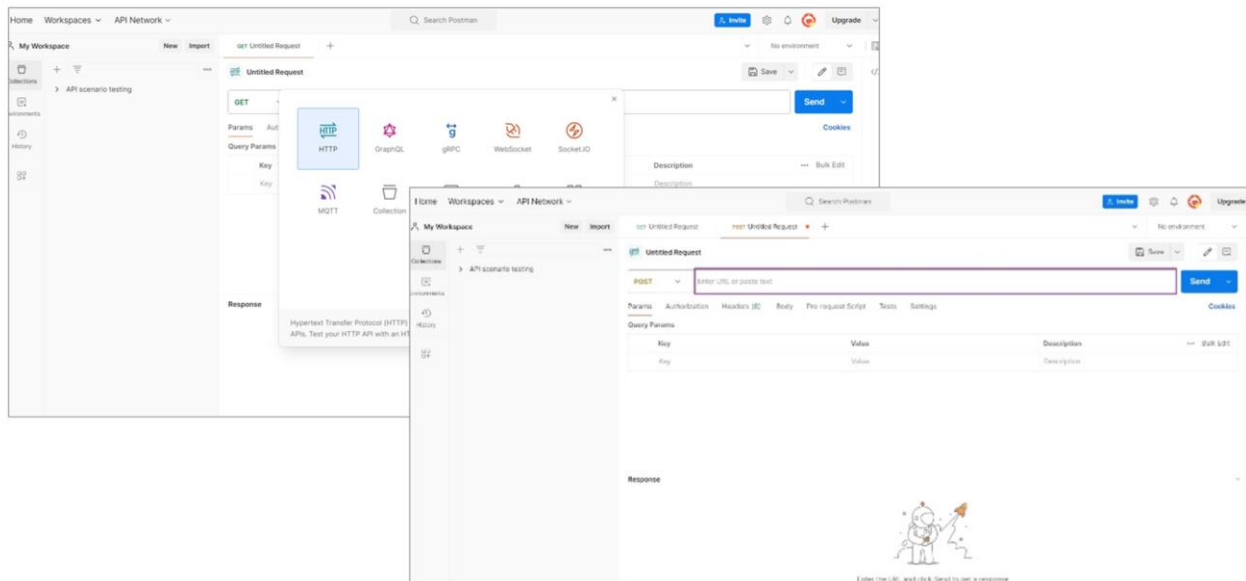
Una vez creado un asistente, testado al momento de su creación, y testado luego también desde el Playground, lo damos por válido.

Postman API Platform

Pasaremos ahora a testearlo vía API a través de Postman API Platform, una plataforma para crear y utilizar APIs.

A partir de ahí, podremos utilizar la API ar en nuestros desarrollos, independientemente de la tecnología.

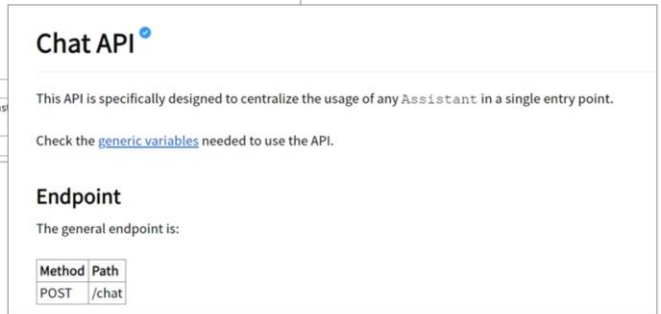
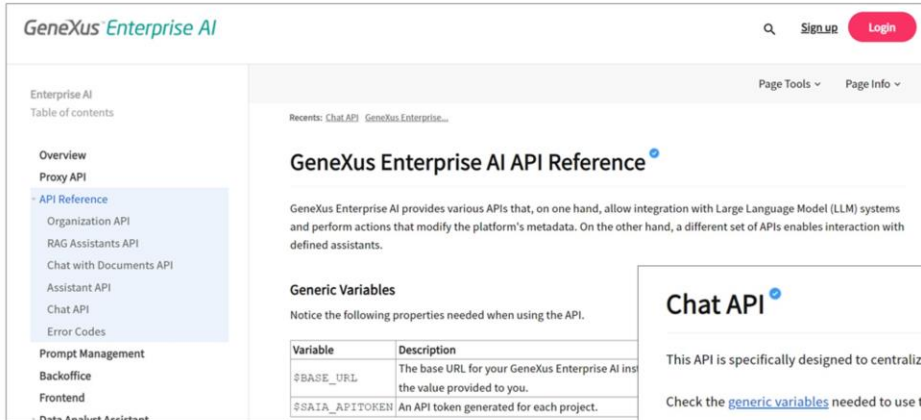
Postman API Platform



Ingresamos entonces, a Postman y desde My Workspaces seleccionamos New... HTTP:

Como nuestro objetivo es mandar, postear una consulta para recibir una respuesta, lo que haremos es un POST.

Así que aquí elegimos POST, y debemos indicar la URL de acceso al server. Para determinar esta URL, consultamos la documentación de GeneXus Enterprise AI.



Nuestro objetivo es conectarnos con el asistente para chat que hemos creado anteriormente, de nombre Marketing Assistant, por lo tanto, necesitamos la información necesaria para comunicarnos con este tipo de asistentes.

Esto es importante tenerlo en cuenta ya que si la idea fuera conectarnos, por ejemplo, con un RAG Assistant, necesitaríamos otros parámetros.

Bien. Así que en API References, Chat API, encontramos la información que necesitamos y también ejemplos de uso:

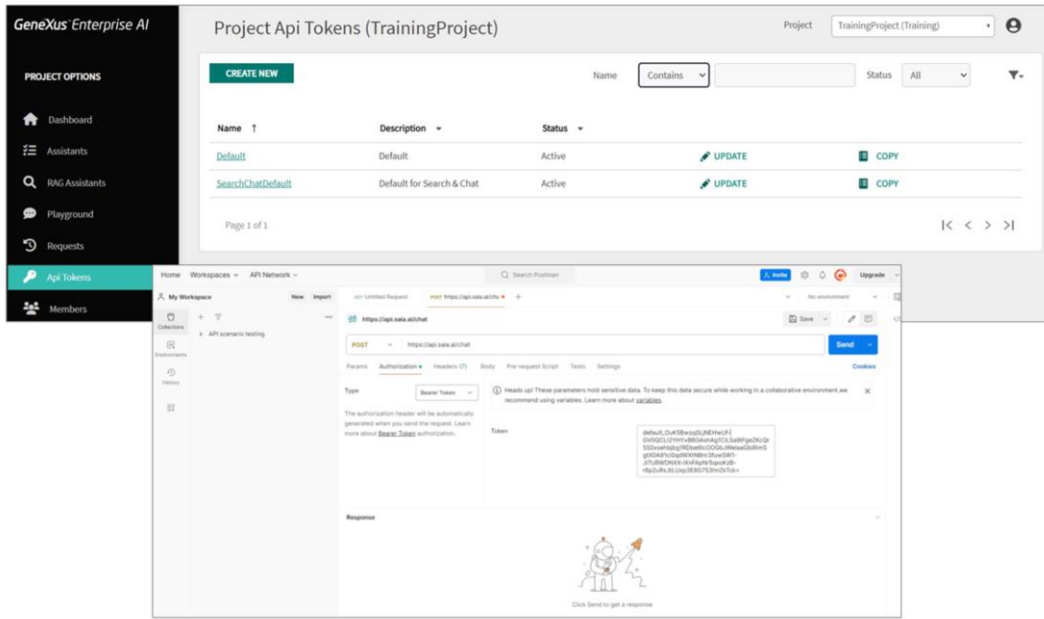
Vemos que para utilizar la API necesitamos el contenido de las variables `BASE_URL` y `SAIA_API_TOKEN`

Donde `BASE_URL` corresponde a la url base de instalación de GX Enterprise AI. y `Saia_apitoken` corresponde a un api token generado para el proyecto.

Bien, así que indicamos la url base correspondiente a nuestro escenario: `https://api.qa.saia.ai`

Vemos que el path se completa con `/chat`, así que lo agregamos en la declaración del POST.

Postman API Platform



Pasamos ahora a la Autorización, que es Bearer token, así que en la solapa Authorization de Postman elegimos esa opción.

Para utilizar la API, debemos autenticar cada solicitud mediante un API token. Como ya hemos visto, estos tokens se administran en el Backoffice de GeneXus Enterprise AI e identifican de forma única al remitente de la solicitud.

Debemos entonces proporcionar ese token. Para eso, vamos al Backoffice de GeneXus Enterprise AI, a la opción API Tokens y podemos seleccionar cualquiera de los tokens definidos, ya que cualquiera nos identifica en el proyecto.

Por ejemplo, el Default. Presionamos Copy.

Volvemos a Postman, y pegamos el token.

Postman API Platform

```
curl -X POST "$BASE_URL/chat" \  
-H "Authorization: Bearer $SAIA_PROJECT_API_TOKEN" \  
-H "Content-Type: application/json" \  
--data '{  
  "model": "saia:assistant:translate-to-spanish",  
  "messages": [  
    {  
      "role": "user",  
      "content": "Hi, welcome to GeneXus Enterprise AI!!"  
    }  
  ],  
  "stream": true  
'
```

Bien. Vamos ahora a la solapa Body, elegimos RAW, JSON y vamos a consultar un ejemplo base de lo que debemos declarar aquí.

Ya hemos indicado el POST, la forma de autorización y el tipo de contenido. Lo que debemos indicar ahora es el cuerpo de la consulta, así que copiamos lo que corresponde a DATA, y lo pegamos

“Model” corresponde al tipo de asistente, seguido de su nombre. Como ya mencionamos, dependiendo del tipo de asistente, los parámetros pueden variar.

El type “Assistant” identifica a un asistente estándar, a un Data Analyst Assistant y un API Assistant, mientras que el type “Search” identifica a un RAG Assistant

En nuestro caso, queremos contactar con nuestro asistente para chat, así que el type indicado es “assistant” y el nombre es “Marketing Assistant”

Veamos ahora al elemento Messages. Este elemento define un mensaje que se desea agregar. La mínima expresión es esta que estamos declarando, donde “content” corresponde al input del usuario.

En nuestro ejemplo, como input, vamos a indicar “ lámpara”, lamp, esperando que el asistente nos devuelva una buena descripción de este producto.

Postman API Platform

The screenshot shows a Postman interface for a POST request to `https://api.qa.saia.ai/chat`. The request body is a JSON object:

```
1 {
2   "model": "saia:assistant:MarketingAssistant",
3   "messages": [
4     {
5       "role": "user",
6       "content": "lamp"
7     }
8   ],
9   "revisionId": 2,
10  "revisionName": "2"
11 }
```

The response status is 200 OK, with a time of 939 ms and a size of 922 B. The response body is a JSON object:

```
{ "id": "chatempl-9AM5EqRanxIHypOqMHDx4LcGIEVbq", "object": "chat.completion", "created": 1712255196, "model": "gpt-3.5-turbo-16k-0613", "choices": [
{ "index": 0, "message": { "role": "assistant", "content": "The lamp is a light fixture that provides illumination in a room or space." }, "logprobs": null, "finish_reason": "stop" } ], "usage": { "prompt_tokens": 85, "completion_tokens": 15, "total_tokens": 100 }, "system_fingerprint": null }
```

Luego podemos agregar otros parámetros., así que vamos a indicar, por ejemplo, la revisión del asistente que queremos utilizar. Recordemos que podíamos tener varias revisiones de un mismo asistente.

Indicamos entonces el identificador y nombre de la revisión

“revisionId”: 2

“revisionName”: “2”

Presionamos ahora Send, y obtenemos la respuesta:

El asistente nos dice que “La lámpara es un artefacto de iluminación que proporciona luz en una habitación o espacio.”

Muy bien. Ahora que hemos testeado nuestro asistente vía API, podemos utilizarlo en nuestros desarrollos.

Más adelante veremos ejemplos de uso desde una base de conocimiento GeneXus.

GeneXusTM
by **Globant**

training.genexus.com