

Straico - API

The **Straico API** utilizes **API keys** for authentication. Please visit [API Key page](#) in My profile section to obtain the **API key** required for your requests.

Remember, it is essential to keep your API key confidential. Avoid sharing it with others or embedding it in any client-side code like browsers or apps. For secure handling, all production requests should be directed through your backend server, where the API key can be securely fetched from an environment variable or key management service.

Every API request must include your API key in an Authorization HTTP header in the following manner:

```
bash
```

```
Authorization: Bearer $STRAICO_API_KEY
```

User

GET User information

```
https://api.straico.com/v0/user
```

Overview

This endpoint allows users to fetch details of a specific user from the Straico platform. Users can access information such as the user's first name, last name, the number of coins associated with the account, and the plan they are subscribed to.

Authentication

All requests must include an `Authorization` header with a valid API key.

Request Headers

- Authorization: `Bearer $STRAICO_API_KEY`

Request Body

This endpoint does not require any parameters.

Response

The API responds with a JSON object containing the user's data under the "data" key.

Field	Type	Description
<code>first_name</code>	string	The user's first name
<code>last_name</code>	string	The user's last name
<code>coins</code>	number	The number of coins associated with the account
<code>plan</code>	string	The current subscription plan of the user

Model

GET Models information (v.0)

`https://api.straico.com/v0/models`

Overview

This endpoint provides users with a comprehensive list of available language models and their associated details within the Straico API ecosystem. By querying this endpoint, users can access up-to-date information about various models and their associated pricing

Authentication

All requests must include an `Authorization` header with a valid API key.

Request Headers

- Authorization: `Bearer $STRAICO_API_KEY`

Request Body

This endpoint does not require any parameters.

Response

The API responds with a JSON object containing an array of model objects. Each model object includes the following information:

Field	Type	Description
<code>name</code>	string	Model display name
<code>model</code>	string	Unique model identifier for API usage
<code>pricing</code>	object	Pricing information, detailing the cost in coins per specified number of words
<code>max_output</code>	number	Limit number of tokens (roughly words or word pieces) that can be generated in a single response

GET Models information (v.1)

```
https://api.straico.com/v1/models
```

The models information endpoint (v.1) allows users to fetch a list of available models, categorized by type, along with their corresponding details and from the Straico API. This version now includes both chat and image generation models, offering detailed specifications and pricing for each category.

Authentication

All requests must include an `Authorization` header with a valid API key.

Request Headers

- Authorization: `Bearer $STRAICO_API_KEY`

Request Body

This endpoint does not require any parameters.

Response

Upon a successful request, the API responds with a JSON object containing categorized arrays of model objects. The response includes separate arrays for chat models and image generation models.

Response Fields

Field	Type	Description
<code>chat</code>	array	Array of chat model objects
<code>image</code>	array	Array of image generation model objects

Model Object Fields

Field	Type	Description
<code>name</code>	string	Model display name
<code>model</code>	string	Unique model identifier for API usage
<code>word_limit</code>	number	Maximum number of words the model can process
<code>pricing</code>	object	For chat models, includes a <code>"coins"</code> value representing the cost per 100 words. For image models, includes aspect ratios (<code>"square"</code> , <code>"landscape"</code> , and <code>"portrait"</code>), each represented as an object containing <code>"coins"</code> (the price per image) and <code>"size"</code> (the resolution).
<code>max_output</code>	number	Limit number of tokens (roughly words or word pieces) that can be

Metadata Object Fields

Field	Type	Description
<code>editors_choice_level</code>	number	General valuation given from straico: a value between -1 and 3, higher for better models.
<code>pros</code>	array	Array of strings with model advantages.
<code>cons</code>	array	Array of strings with model disadvantages.
<code>applications</code>	array	Array of strings with areas or fields of knowledge where the model excels.
<code>features</code>	array	Array of strings with model additional features, for example, <code>"Web search"</code> or <code>"Image input"</code>
<code>icon</code>	string	Image url of the model icon
<code>other</code>	array	Array of strings with other characteristics, like <code>"Uncensored"</code> or <code>"Open"</code>

Prompt

POST Prompt completion (v.0)

<https://api.straico.com/v0/prompt/completion>

Overview

This endpoint enables users to generate prompt completion based on a specified model using the Straico API. Users can make a POST request to the provided URL, including the necessary headers and parameters, to receive a prompt completion for a given message.

Authentication

Authentication

All requests must include an `Authorization` header with a valid API key.

Request Headers

- Authorization: `Bearer $STRAICO_API_KEY`
- Content-Type: `application/json` or `application/x-www-form-urlencoded`

Request Body

- You must choose between including either `model` or `smart_llm_selector` ; you can't include or omit both fields at the same time.

Parameter	Type	Required	Description
<code>model</code>	string	No	Specifies the model for generating the prompt completion
<code>smart_llm_selector</code>	string	No	<p>The possible values for this string are: <code>"quality"</code>, <code>"balance"</code>, or <code>"budget"</code>.</p> <p>If provided, an AI will select a model whose advantages, features and applications best match the user's needs based on the prompt. It will also choose a cheaper, balanced, or expensive model according to the specified pricing</p>

Response

Upon a successful request, the API responds with a JSON object containing the generated prompt completion and additional details.

Field	Type	Description
<code>price</code>	object	Cost breakdown. Contains <code>input</code> , <code>output</code> , and <code>total</code> coins
<code>words</code>	object	Word count details. Contains <code>input</code> , <code>output</code> , and <code>total</code> words
<code>completion</code>	object	Detailed result for the prompt completion generated
<code>model_selector_justification</code>	string	A detailed justification of the <code>smart_llm_selector</code> decision, if it was used.

POST Prompt completion (v.1)

<https://api.straico.com/v1/prompt/completion>

Overview

The Prompt Completions endpoint (v.1) enables users to generate prompt completions using multiple language models (LLMs) simultaneously. Users can request completions from up to 4 different models in a single request. This version supports the inclusion of YouTube video files and various attachments.

Authentication

All requests must include an `Authorization` header with a valid API key.

Request Headers

- Authorization: `Bearer $STRAICO_API_KEY`
- Content-Type: `application/json`

Request Body

The request body must include model identifiers, prompt text, and optional file and YouTube URLs, all formatted as JSON.

- Files must be pre-uploaded using the File Upload endpoint, which provides valid URLs for use in this request.
- You must choose between including either `models` or `smart_llm_selector` ; you can't include or omit both fields at the same time.

Parameter	Type	Required	Description
<code>models</code>	array	No	An array of 1-4 unique model identifiers
<code>smart_llm_selector</code>	object	No	<p>Includes the fields <code>quantity</code> (number) and <code>pricing_method</code> (string with possible values: " <code>quality</code> ", " <code>balance</code> ", or " <code>budget</code> ").</p> <p>If provided, an AI will select models whose advantages, features and applications best match the user's needs based on the prompt. It will also choose cheaper, balanced or</p>

Response

The API responds with a JSON object containing the following information:

Field	Type	Description
<code>overall_price</code>	object	Overall cost for all models, including <code>input</code> , <code>output</code> and <code>total</code> coins
<code>overall_words</code>	object	Overall word count for all models, including <code>input</code> , <code>output</code> and <code>total</code> words
<code>completions</code>	object	Detailed results for each requested model

		requested model
<code>transcripts</code>	array	If the parameter <code>display_transcripts</code> is given as <code>true</code> , returns an array of transcripts, each containing a <code>name</code> and <code>text</code>
<code>model_selector_justification</code>	string	A detailed justification of the <code>smart_llm_selector</code> decision, if it was used.

File

POST File upload

`https://api.straico.com/v0/file/upload`

Overview

This endpoint allows users to upload files for use in the Prompt Completion (v.1) endpoint. Files are securely stored in our cloud storage system, ensuring data integrity and accessibility.

Authentication

All requests must include an `Authorization` header with a valid API key.

Request Headers

- Authorization: `Bearer $STRAICO_API_KEY`
- Content-Type: `multipart/form-data`

Request Body

The request should be sent as `multipart/form-data` with a single file.

- Maximum file size: 25 MB.
- Supported file types: pdf docx pptx txt xlsx mp3 mp4 html csv json py php js css cs swift

Supported file types: pdf, docx, pptx, txt, doc, mp3, mp4, flac, wav, json, py, php, js, css, scss, xml, kt, xml, ts, png, jpg, jpeg, webp, gif.

Parameter	Type	Required	Description
<code>file</code>	File	Yes	The file to be uploaded

Response

The API responds with a JSON object containing the following information:

Field	Type	Description
<code>url</code>	string	The URL of the successfully uploaded file

Image

POST Image generation

```
https://api.straico.com/v0/image/generation
```

Overview

This endpoint enables users to generate high-quality images based on textual descriptions. Leveraging advanced AI models, it can create diverse visual content for various applications, from concept art to marketing materials. You can generate up to 100 images per day via the API.

Authentication

All requests must include an `Authorization` header with a valid API key.

Request Headers

- Authorization: `Bearer $STRAICO_API_KEY`
- Content-Type: `application/json`

Request Body

Parameter	Type	Required	Description
<code>model</code>	string	Yes	The AI model to use for image generation, the current models available are: <code>openai/dall-e-3</code> , <code>flux/1.1</code> , <code>ideogram/V_2A</code> , <code>ideogram/V_2A_TURBO</code> , <code>ideogram/V_2</code> , <code>ideogram/V_2_TURBO</code> , <code>ideogram/V_1</code> , <code>ideogram/V_1_TURBO</code>
<code>description</code>	string	Yes	A detailed textual description of the image to be generated
<code>size</code>	string	Yes	The desired image

Response

The API responds with a JSON object containing the following information:

Field	Type	Description
<code>zip</code>	string	URL to download a ZIP file containing all generated images
<code>images</code>	array	Array of URLs, each pointing to an individual generated image
<code>price</code>	object	Detailed pricing information for the request. Includes <code>price_per_image</code> , <code>quantity_images</code> , and <code>total</code>
<code>enhancer</code>	string	Instructions that the AI uses to generate the new enhanced prompt.
<code>promptEnhanced</code>	string	Instructions that the AI used to generate the new enhanced prompt, either predefined instructions if no custom enhancer was provided, or the same

POST Image generation (v1.0)

<https://api.straico.com/v1/image/generation>

Overview

This endpoint enables users to generate high-quality images based on textual descriptions. Leveraging advanced AI models, it can create diverse visual content for various applications, from concept art to marketing materials.

This new version of image generation includes new and upgraded models.

Authentication

All requests must include an `Authorization` header with a valid API key.

Request Headers

- Authorization: `Bearer $STRAICO_API_KEY`
- Content-Type: `application/json`

Request Body

Parameter	Type	Required	Description
<code>model</code>	string	Yes	<p>The AI model to use for image generation, the current models available are:</p> <p><code>fal-ai/flux/dev</code>, <code>fal-ai/imagen4/preview</code>, <code>fal-ai/bagel</code>, <code>fal-ai/recraft/v3/text-to-image</code>, <code>fal-ai/ideogram/v3</code></p> <p>The following are inherited models from v0:</p> <p><code>openai/dall-e-3</code>, <code>flux/1.1</code>, <code>ideogram/V_2A</code>,</p>

Response

The API responds with a JSON object containing the following information:

Field	Type	Description
-------	------	-------------

Field	Type	Description
<code>zip</code>	string	URL to download a ZIP file containing all generated images
<code>images</code>	array	Array of URLs, each pointing to an individual generated image
<code>price</code>	object	Detailed pricing information for the request. Includes <code>price_per_image</code> , <code>quantity_images</code> , and <code>total</code>

Video

POST Image to video generation

`https://api.straico.com/v1/image/tovideo`

Overview

This endpoint enables users to generate high-quality videos based on attached images.

Authentication

All requests must include an `Authorization` header with a valid API key.

Request Headers

- Authorization: `Bearer $STRAICO_API_KEY`
- Content-Type: `application/json`

Request Body

Parameter	Type	Required	Description
<code>model</code>	string	Yes	The AI model to use for image-to-video generation, the current models available are: <code>fal-ai/kling-video/v2.1/master/image-to-video</code> , <code>fal-ai/veo2/image-to-video</code> , <code>fal-ai/vidu/q1/image-to-video</code> , <code>gen3a_turbo</code> , <code>gen4_turbo</code>
<code>description</code>	string	Yes	A detailed textual description of the video to be generated
<code>size</code>	string	Yes	The desired image dimensions.

Response

The API responds with a JSON object containing the following information:

Field	Type	Description
<code>zip</code>	string	URL to download a ZIP file containing all generated images
<code>video</code>	array	Array of URLs, each pointing to an individual generated image
<code>price</code>	object	Detailed pricing information for the request. Includes

price_per_video and
total

TTS

GET Eleven Labs Voices

`https://api.straico.com/v1/tts/elevenlabslist`

Overview

This endpoint returns the list of available voices in Eleven Labs. You can retrieve the ids for the endpoint "Create Text To Speech".

Authentication

All requests must include an `Authorization` header with a valid API key.

Request Headers

- Authorization: `Bearer $STRAICO_API_KEY`

POST Create Text-to-Speech

`https://api.straico.com/v1/tts/create`

Create Text-to-Speech (TTS)

This endpoint allows users to create a text-to-speech (TTS) conversion by providing specific parameters related to the desired voice model and text content.

Authentication

All requests must include an `Authorization` header with a valid API key.

All requests must include an `Authorization` header with a valid API Key.

Request Headers

- Authorization: `Bearer $STRAICO_API_KEY`
- Content-Type: `x-www-form-urlencoded`

Request Body

Parameter	Type	Required	Description
<code>model</code>	string	Yes	Specifies the of model to use for the TTS conversion. Currently, you can choose between <code>eleven_multilingual_v2</code> and <code>tts-1</code> .

<code>text</code>	string	Yes	The actual text content that you want to convert into speech. This parameter is essential for the TTS process. (Up to 4,000 characters)
<code>voice_id</code>	string	Yes	Identifies the specific voice to be used for the TTS conversion. If the selected model is

Agents

This is the first version of the documentation of the API endpoints for agents, allowing users to upload a file and converting it into a suitable format for RAG.

POST Create Agent



https://api.straico.com/v0/agent

Creates a new agent in the database for the user.

json

```
{
  "success": true,
  "data": {
    "uuidv4": "73648dd8-79d9-4e97-8dd1-2ef2df2b8843",
```

```
"user_id": "743ada93ff7a1d6822131d3f5",
"default_llm": "anthropic/claude-3.5-sonnet",
"custom_prompt": "Act as an expert in airline regulations, based on an att",
"name": "Expert in airline regulations",
"description": "Agent acting as an expert in airline regulations",
"status": "active",
"tags": ["assistant", "rag"],
```

AUTHORIZATION Bearer Token

Token <token>

Body urlencoded

name	<A name for the agent>
custom_prompt	<A custom prompt for the agent behaviour>
default_llm	<A model that the agent will use for processing prompts>
description	<A brief description of what the model does>
tags	<An array of tags for the agent. Example: ["assistant","rag"]>

POST Add RAG to Agent



https://stapi.straico.com/v0/agent/<agent-id>/rag

Creates a new agent in the database for the user.

json

```
{
  "success": true,
  "data": {
    "uuidv4": "73648dd8-79d9-4e97-8dd1-2ef2df2b8843",
    "user_id": "743ada93ff7a1d6822131d3f5",
    "default_llm": "anthropic/claude-3.5-sonnet",
    "custom_prompt": "Act as an expert in airline regulations, based on an att",
    "name": "Expert in airline regulations",
    "description": "Agent acting as an expert in airline regulations",
    "status": "active",
    "tags": ["assistant", "rag"],
  }
}
```

AUTHORIZATION Bearer Token

Token <token>

Body raw (json)

json

```
{
  "rag": "670469b677rwt1rwwr5b0903"
}
```

GET Agent details



https://api.straico.com/v0/agent/<agent-id>

Get Agent Details

This endpoint makes an HTTP GET request to retrieve details of a specific agent identified by the provided ID.

Request Parameters

- : the id of the required agent

Response

json

```
{
  "success": true,
  "data": {
    "_id": "66df87d2bb560c6acabb686s",
    "uuidv4": "c85254c4-0f89-46d7-bf28-428eed764e16",
    "default_llm": "openai/gpt-4o",
    "custom_prompt": "Act a as expert in airline regulations",
    "name": "Expert in airline regulations",
    "description": "Agent expert in airline regulations",
    "status": "active",
    "tags": [],
  }
}
```

AUTHORIZATION Bearer Token

Token <token>

GET List of agents



<https://api.straico.com/v0/agent/>

The endpoint retrieves the list of agents created by and available to the user.

Response

The response for this request follows the JSON schema below:

json

```
{
  "success": true,
  "data": [
    {
      "_id": "66e8927659c3f8a86098c8da",
      "uuidv4": "73648d38-79d9-4e97-8031-2ef2df2b8843",
      "user_id": "64ada93ff7a1fr322131d3f5",
      "default_llm": "anthropic/claude-3.5-sonnet",
      "custom_prompt": "Act as an expert in airline regulations, based on an",
      "name": "Expert in airline regulations",
      "description": "Agent acting as an expert in airline regulations".
    }
  ]
}
```

AUTHORIZATION Bearer Token

Token <token>

PUT Update agent



`https://api.straico.com/v0/agent/<agent-id>`

Update Agent Details

This endpoint allows updating the details of a specific agent.

Request

- Method: PUT
- Endpoint: `https://stapi.straico.com/v0/agent/`

Response

The response is in the form of a JSON schema.

json

```
"success": true,
  "data": {
    "_id": "66df87d2bb560c6acabb6822",
    "uuidv4": "c85254c4-0f89-46d7-bf28-428eed764e16",
    "user_id": "64ada93ff7a1d6822131d3f5",
    "default_llm": "openai/gpt-4o",
    "custom_prompt": "Actúa como un buen abogado penalista, que responda preguntas",
    "name": "Experto en Código Penal Colombiano",
    "description": "Agente que actúa como abogado penalista experto en código penal",
    "status": "active",
    "tags": []
```

AUTHORIZATION Bearer Token

Token <token>

Body urlencoded

name <New value of name>

<any other attribute> <New value of any other attribute>

POST Agent prompt completion



https://api.straico.com/v0/agent/<agent-id>/prompt

Agent Prompt Completion

This endpoint allows the user to make an HTTP POST request to prompt the completion of an agent message. It uses the previously uploaded RAG base and the agent's default model.

Request Body

- `prompt` (text): A text prompt for the RAG model.

Optional Parameters

RAG bases are build using FAISS format. To retrieve results, we use a VectorStoreRetriever. This entity has custom parameters that can modify the results used for prompt completion. Check this document for more information:

https://api.python.langchain.com/en/latest/vectorstores/langchain_community.vectorstores.faiss.FAISS.html#langchain_community.vectorstores.faiss.FAISS.as_retriever

- `search_type` : one of the following: `similarity`, `mmr` or `similarity_score_threshold`
- `k` : number of documents to return
- `fetch_k` : ammount of documents to pass to MMR algorithm
- `lambda_mult` : diversity of results return by MMR, 1 for minimum and 0 for maximum
- `score_threshold` : Minimum relevance threshold for `similarity_score_threshold`

Notes on coins

- Total coins will be based on the regular fees for each LLM available in Straico.
- **WARNING:** coins will consider the number of words in the `references` .

Response

json

```
{
  "success": true,
  "data": {
    "answer": "This appears to be a section from an academic regulations or st",
    "references": [
      {
        "page_content": "dianta tiene matrícula condicional como consecuen",
        "page": 73
      }
    ]
  }
}
```

```
} ,  
{
```

```
"page_content": "institucional.",
```

IMPORTANT: Despite the response has now a `data` object, the response still keeps the object `response` , in order to keep backward compatibility with older versions of the API.

AUTHORIZATION Bearer Token

Token `<token>`

Body `urlencoded`

prompt `What's the document about?`

DELETE Delete agent



```
https://api.straico.com/v0/agent/<agent-id>
```

Delete Agent

This endpoint is used to delete a specific agent by providing the agent ID in the URL.

Request Body

This request does not require a request body.

Response

Example Response:

json

```
{  
  "success": true,  
  "message": "Agent deleted successfully"  
}
```

AUTHORIZATION Bearer Token

Token <token>

Rag

POST Create RAG



<https://api.straico.com/v0/rag>

Create a new RAG

This endpoint allows the user to create a new RAG base in the database. The request should be sent as an HTTP POST .Request Body

The request should have a form-data body type with the following parameters:

- `name` (text): Represents the name of the RAG base
- `description` (text): Represents the description of the agent.
- `files` (file): Represents the files to be attached (up to 4 files). Accepted file extensions are: pdf, docx, csv, txt, xlsx, py.
- `chunking_method` (text - optional): Represents the chunking method to be use for generating the RAG base. The default value is `fixed_size` . Depending on the value set, there are other optional parameters than can be set:

chunking method	Available optional parameters	Should perform quiet well with	Cons
fixed_size	chunk_size (default value: 1000) chunk_overlap (default value: 50) separator (default value: \n)	Not so elaborated text	Does not consider text structure and semantics
recursive	chunk_size (default value: 1000) chunk_overlap (default value: 50) separators (default value: ["\n\n", "\n", " ", ""])	Not so elaborated text	Does not consider text structure and semantics
markdown	chunk_size (default value: 1000) chunk_overlap (default value: 50)	Files with Markdown	Only suitable for files with Markdown
python	chunk_size (default value: 1000) chunk_overlap	Python files	Only suitable for Python Files

Response

- Status: 201
- Content-Type: application/json

The response will contain a JSON object with the following structure:

json

```
{
  "success": true,
  "data": {
    "user_id": "64ada93f22131d3f5",
    "name": "Rag de prueba txt",
    "rag_url": "https://prompt-rack.s3.amazonaws.com/api/rag/64ada93ff7a1d68222",
    "original_filename": "sample_txt.txt",
    "chunking_method": "fixed_size",
    "chunk_size": 1000,
    "chunk_overlap": 50,
    "id": "670565d07e1234eb",
```

Notes on coins and processing time

- Current fee is 0.1 coins / 100 words of the processed files.
- Processing the file may take several minutes. It takes approximately 1 minute per 50.000 words of the original documents.

AUTHORIZATION Bearer Token

Token <token>

Body formdata

name Rag entity

description A description of the RAG entity

files

GET List of RAG's



<https://api.straico.com/v0/rag/user>

Get list of RAGs for user

This endpoint returns the list of RAG bases for the user.

Response

json

```
{
  "success": true,
  "data": [
    {
      "_id": "670558840d774c63b439",
      "user_id": "64ada93ff7a131d3f5",
      "name": "Test",
      "rag_url": "https://prompt-rack.s3.amazonaws.com/api/rag/64ada93ff7a1d",
      "original_filename": "sample_txt.txt, sample_excel.xlsx, 2000_codigope",
      "chunking_method": "fixed_size",
      "chunk_size": 1000
    }
  ]
}
```

AUTHORIZATION Bearer Token

Token <token>

GET RAG by ID



<https://api.straico.com/v0/rag/<rag-id>>

Get RAG base by ID

This endpoint makes an HTTP GET request to retrieve details of a specific RAG base identified by the provided ID.

Request Parameters

- id: the id of the required RAG base

Expected Response

Upon successful execution, the API returns a JSON object with the following format:

json

```
{
  "success": true,
  "data": {
    "_id": "670565d02fc07e1234eb",
    "user_id": "64ada922131d3f5",
    "name": "Test",
    "rag_url": "https://prompt-rack.s3.amazonaws.com/api/rag/64ada93ff7a1d6822",
    "original_filename": "sample_txt.txt",
    "chunking_method": "fixed_size",
    "chunk_size": 1000,
    "chunk_overlap": 50,
```

AUTHORIZATION Bearer Token

Token <token>

PUT Update RAG



<https://api.straico.com/v0/rag/<rag-id>>

This endpoint is used to update a RAG given some files to enhance the existing RAG base.

Request Body

- form-data
 - **files** (text): The files to be processed and added to the existing RAG base

Response

json

```
{
  "type": "object",
  "properties": {
    // Add properties based on the actual response
  }
  "success": true,
  "data": {
    "_id": "34123412341234s4ae",
    "user_id": "jeqjreqwez34123",
    "name": "A RAG to be updated",
    "description": "Testing a RAG base to be updated later".
  }
}
```

AUTHORIZATION Bearer Token

Token <token>

Body formdata

files

DELETE Delete RAG



https://api.straico.com/v0/rag/<rag-id>

Delete RAG

This endpoint is used to delete a specific agent by providing the rag ID in the URL.

Response

json

```
{
  "success": true,
  "message": "RAG deleted successfully"
}
```

AUTHORIZATION Bearer Token

Token <token>

POST RAG prompt completion



`https://api.straico.com/v0/rag/<rag-id>/prompt`

RAG prompt completion

This endpoint is used to submit a prompt to a specific RAG (Retrieval-Augmented Generation) model.

Request Body

- `prompt` (text): A text prompt for the RAG model.
- `model` (text): The specific LLM to be used.

Optional Parameters

RAG bases are build using FAISS format. To retrieve results, we use a VectorStoreRetriever. This entity has custom parameters that can modify the results used for prompt completion. Check this document for more information:

https://api.python.langchain.com/en/latest/vectorstores/langchain_community.vectorstores.faiss.FAISS.html#langchain_community.vectorstores.faiss.FAISS.as_retriever

- `search_type` : one of the following: `similarity`, `mmr` or `similarity_score_threshold`
- `k` : number of documents to return
- `fetch_k` : ammount of documents to pass to MMR algorithm

- `lambda_mult` : diversity of results return by MMR, 1 for minimum and 0 for maximum
- `score_threshold` : Minimum relevance threshold for `similarity_score_threshold`

Notes on coins

- Total coins will be based on the regular fees for each LLM available in Straico.
- **WARNING:** coins will consider the number of words in the `references` .

Response

The response is a JSON object with the following schema:

json

```
{
  "success": true,
  "data": {
    "answer": "This appears to be a section from an academic regulations or statute",
    "references": [
      {
        "page_content": "dante tiene matrícula condicional como consecuencia",
        "page": 73
      },
      {
        "page_content": "institucional.",

```

IMPORTANT: Despite the response has now a `data` object, the response still keeps the object `response` , in order to keep backward compatibility with older versions of the API.

AUTHORIZATION Bearer Token

Token `<token>`

Body `urlencoded`

prompt What happens if the flight is delayed?

model anthropic/claude-3.5-sonnet