

Multimodal embedding models

# Multimodal embedding models

**POST** <https://api.voyageai.com/v1/multimodalembdings>

The Voyage multimodal embedding endpoint returns vector representations for a given list of multimodal inputs consisting of text, images, or an interleaving of both modalities.

## BODY PARAMS

### **inputs** array **required**

A list of multimodal inputs to be vectorized.

A single input in the list is a dictionary containing a single key "content", whose value represents a sequence of text and images.

- The value of "content" is a list of dictionaries, each representing a single piece of text or image. The dictionaries have four possible keys:
  - i. **type**: Specifies the type of the piece of the content. Allowed values are `text`, `image_url`, or `image_base64`.
  - ii. **text**: Only present when `type` is `text`. The value should be a text string.
  - iii. **image\_base64**: Only present when `type` is `image_base64`. The value should be a Base64-encoded image in the data URL format `data:[<mediatype>];base64,<data>`. Currently supported mediatypes are: `image/png`, `image/jpeg`, `image/webp`, and `image/gif`.
  - iv. **image\_url**: Only present when `type` is `image_url`. The value should be a URL linking to the image. We support PNG, JPEG, WEBP, and GIF images.
- **Note**: Only one of the keys, `image_base64` or `image_url`, should be present in each dictionary for image data. Consistency is required within a request, meaning each request should use either `image_base64` or `image_url` exclusively for images, not both.

- ▶ Example payload where `inputs` contains an image as a URL
- ▶ Example payload where `inputs` contains a Base64 image

The following constraints apply to the `inputs` list:

- The list must not contain more than 1000 inputs.
- Each image must not contain more than 16 million pixels or be larger than 20 MB in size.
- With every 560 pixels of an image being counted as a token, each input in the list must not exceed 32,000 tokens, and the total number of tokens across all inputs must not exceed 320,000.

### **model** string **required**

Name of the model. Currently, the only supported model is `voyage-multimodal-3`.

### **input\_type** string

Type of the input. Defaults to `null`. Other options: `query`, `document`.

- When `input_type` is `null`, the embedding model directly converts the `inputs` into numerical vectors. For retrieval/search purposes, where a "query", which can be text or image in this case, is used to search for relevant information among a collection of data referred to as "documents," we recommend specifying whether your `inputs` are intended as queries or documents by setting `input_type` to `query` or `document`, respectively. In these cases, Voyage automatically prepends a prompt to your `inputs` before vectorizing them, creating vectors more tailored for retrieval/search tasks. Since inputs can be multimodal, "queries" and "documents" can be text, images, or an interleaving of both modalities. Embeddings generated with and without the `input_type` argument are compatible.
- For transparency, the following prompts are prepended to your input.
  - For `query`, the prompt is *"Represent the query for retrieving supporting documents: "*.
  - For `document`, the prompt is *"Represent the document for retrieval: "*.

#### **truncation** boolean

Whether to truncate the inputs to fit within the context length. Defaults to `true`.

- If `true`, an over-length input will be truncated to fit within the context length before being vectorized by the embedding model. If the truncation happens in the middle of an image, the entire image will be discarded.
- If `false`, an error will be raised if any input exceeds the context length.

#### **output\_encoding** string

Format in which the embeddings are encoded. Defaults to `null`.

- If `null`, the embeddings are represented as a list of floating-point numbers.
- If `base64`, the embeddings are represented as a Base64-encoded NumPy array of single-precision floats.

## RESPONSES

200

Success



#### RESPONSE BODY

object

**object** string

The object type, which is always `list`.

**data** array of objects

An array of embedding objects.

object

**object** string

The object type, which is always `embedding`.

**embedding** array

The embedding vector consists of a list of floating-point numbers or a Base64-encoded NumPy array depending on `output_encoding` . The length of this vector varies depending on the specific model.

**index** integer

An integer representing the index of the embedding within the list of embeddings.

**model** string

Name of the model.

**usage** object

**text\_tokens** integer

The total number of text tokens in the list of inputs.

**image\_pixels** integer

The total number of image pixels in the list of inputs.

**total\_tokens** integer

The combined total of text and image tokens. Every 560 pixels counts as a token.

☐ **4XX**

Client error



This indicates an issue with the request format or frequency. Please see our [Error Codes](#) guide.

☐ **5XX**

Server Error

This indicates our servers are experiencing high traffic or having an unexpected issue. Please see our [Error Codes](#) guide.

LANGUAGE



Shell

CREDENTIALS

HEADER ⓘ

Header

\$VOYAGE\_API\_KEY



## Image URL    Base64 image

### REQUEST

```
1  curl -X POST https://api.voyageai.com/v1/multimodalembbeddings \
2  -H "Authorization: Bearer $VOYAGE_API_KEY" \
3  -H "content-type: application/json" \
4  -d '
5  {
6    "inputs": [
7      {
8        "content": [
9          {
10             "type": "text",
11             "text": "This is a banana."
12           },
13           {
14             "type": "image_base64",
15             "image_base64": "data:image/jpeg;base64,/9j/4AAQSkZJRgABAQAA..."
16           }
17         ]
18       }
19     ],
20     "model": "voyage-multimodal-3"
21   }'
```

### RESPONSE

200 - Success EXAMPLE ▾

```
1  {
2    "object": "list",
3    "data": [
4      {
5        "object": "embedding",
6        "embedding": [
7          0.027587891,
8          -0.021240234,
9          0.018310547,
10         "...",
11         -0.021240234
12       ],
13       "index": 0
14     }
15   ],
16   "model": "voyage-multimodal-3",
17   "usage": {
18     "text_tokens": 5,
19     "image_pixels": 2000000,
20     "total_tokens": 3576
21   }
22 }
```





 Updated 3 months ago

---

[← Text embedding models](#)

[Rerankers →](#)

Did this page help you?  **Yes**  **No**