**Document Upload Guide**



## 

## 

## **Introduction**

This guide is designed to guide users through the document upload process in the system, with comprehensive control over the processing and conversion stages of information. The instruction focuses on the advanced options offered by the system and enables precise adaptation to user needs.

## **Uploading and Processing Documents**

To create an agent, sometimes we want it to be based on documents.

To do this, we use the Sources tab.

The Sources tab serves as a library for documents that were uploaded manually or through a link to a knowledge management folder.

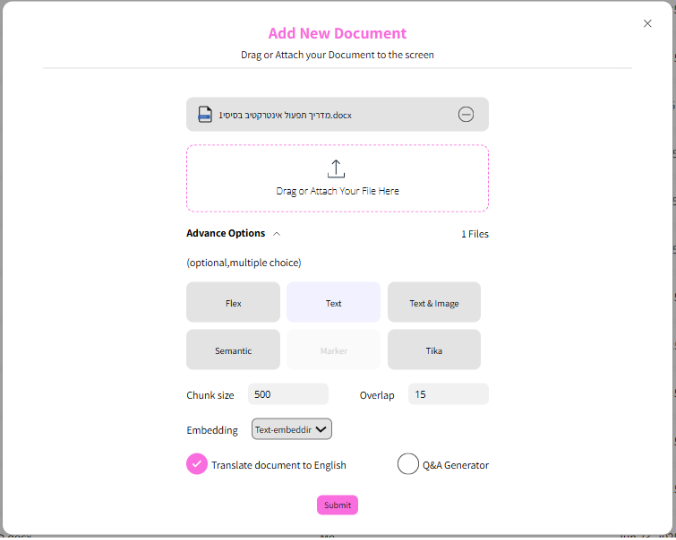
We will typically use these documents in agents we create.

## **Selecting Vector Method**

When uploading a document manually, a window will open where you need to select the vector method.

The vector method divides the document I uploaded into segments. Each segment receives a numerical value, which helps us later find the correct answer to the question.

The agent manages to extract the correct information from the documents by translating my message to a numerical value, and then searching for the segments (from the documents) whose numerical values are closest to the numerical value of the message I sent.



## 

## **Vector Methods**

### **Text Vector Method**

This vector method divides the document according to the segment size that I will define.

### **Flex Vector Method**

This method divides the document according to context. The method is based on headers in the document to divide the document into segments. This way we get segments that are not the same size at all but are organized according to headers and contexts.

### **Text&Image Vector Method**

This method exists for documents that, in addition to texts, also have various images and visuals. Thanks to the GPT-4o model, we can convert the images to text. Finally, we get a document that consists only of text.

## **Technical Parameters**

**Chunk Size** - This is a parameter relevant to the text vector method, and it determines the fixed size of each text segment that undergoes processing. That is, this is the number of characters defined as the segment size for each slice of the text.

**Overlap** - A number that represents how many characters are taken above and below a segment in order to preserve context and ensure that the context is not damaged. Its purpose is to maintain continuous semantic context between segments. When there is overlap between segments, meaningful context is preserved between the different segments, so that the overall context is not damaged in the processing process.

תמונה שמכילה טקסט, חשמל, צילום מסך, מספר

תוכן בינה מלאכותית גנרטיבית עשוי להיות שגוי.

## 

## 

## 

## **Selecting Embedding Model**

At this stage, the segment is converted to a vector value using an Embedding model.

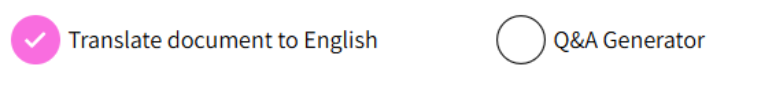
The default model set as default provides good results for most uses, but there is an option to upgrade and change the model according to the specific needs of the business. It is recommended to contact JEEN support in cases where a model change is required, to ensure that the upgrade is performed optimally and supports the system's goals in the best way.

תמונה שמכילה טקסט, גופן, צילום מסך, לוגו

תוכן בינה מלאכותית גנרטיבית עשוי להיות שגוי.

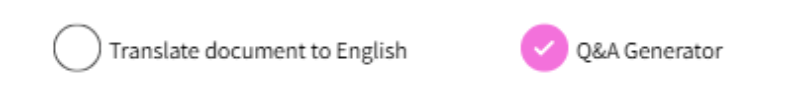
## **Translating Documents to English**

Many times we upload documents in Hebrew. When we check the box next to "Translate document to English", the document will be translated to English. When uploading a document in English, it is preferable not to check the box next to "Translate document to English".



## **Questions**

When you click the Questions button, an automatic process will run behind the scenes. The system will take the entered document and automatically create questions and answers based on the document's content. The questions and answers will be created in the original language of the document, and if English translation was selected, an additional column will also be added containing the questions and answers in the English version.



## 

## 

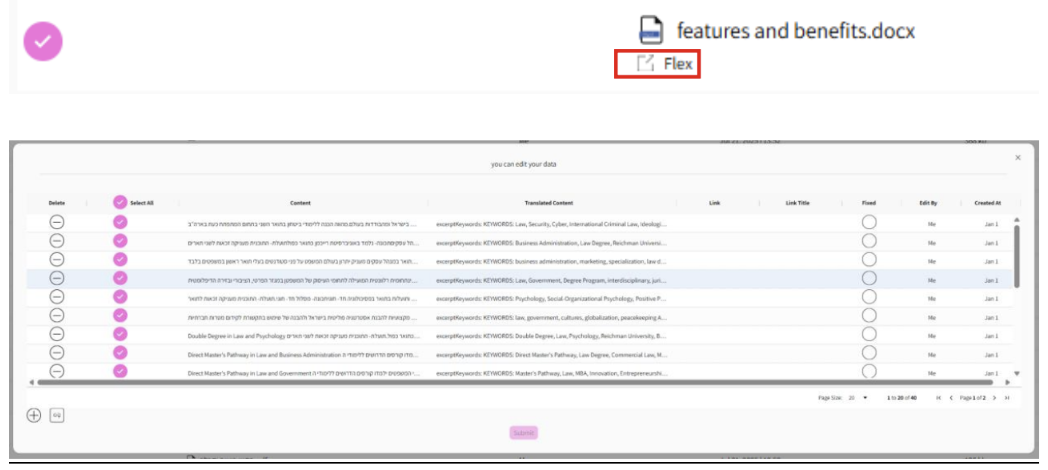
## 

## 

## **Viewing Vectors**

After uploading the document, you can see how the system divided the vectors. To view vectors, you need to enter the Sources tab, select the desired document, and after clicking on it, a vector viewing button will appear under the document name. In the example, you can see that the button allows viewing vectors written as FLEX since this is the method chosen when uploading the document, but if you chose another method, the name that will appear next to the button will be the name of the vector method we chose.

Clicking this button will allow you to see the document's vectors.



## 

## 

## 

## 

## 

## 

## 

## 

## **Adding Links to Vectors**

There is an option to add a link to each vector, so that the link appears in the chat.

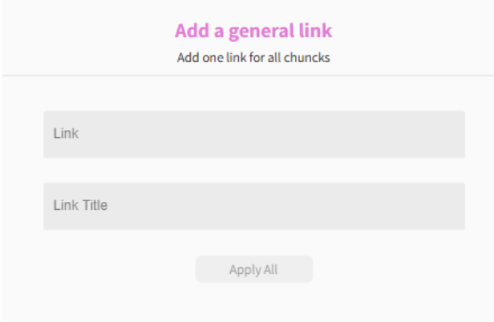
There are 2 ways to do this:

### **1. Manual Addition:**

You can add a link to each vector in the vectors table in the LINK and LINK TITLE tabs.

**Link:** The link that the user will reach when clicking on the vector. **Link Title:** The text that will be displayed on the link.





### **2. Addition Through Batch Button:**

Clicking the attach button on the left side allows adding a link to all vectors with one click, so that the link is easily displayed for all of them.

