Embeddings HuggingFace Inference node documentation

# Embeddings HuggingFace Inference node#

Use the Embeddings HuggingFace Inference node to generate embeddings for a given text.

On this page, you'll find the node parameters for the Embeddings HuggingFace Inference, and links to more resources.

Credentials

You can find authentication information for this node here.

Parameter resolution in sub-nodes

Sub-nodes behave differently to other nodes when processing multiple items using an expression.

Most nodes, including root nodes, take any number of items as input, process these items, and output the results. You can use expressions to refer to input items, and the node resolves the expression for each item in turn. For example, given an input of five name values, the expression {{ $json.name }} resolves to each name in turn.

name

{{ $json.name }}

In sub-nodes, the expression always resolves to the first item. For example, given an input of five name values, the expression {{ $json.name }} always resolves to the first name.

name

{{ $json.name }}

## Node parameters#

• Model: Select the model to use to generate the embedding.

Refer to the Hugging Face models documentation for available models.

## Node options#

• Custom Inference Endpoint: Enter the URL of your deployed model, hosted by HuggingFace. If you set this, n8n ignores the Model Name.

Refer to HuggingFace's guide to inference for more information.

## Templates and examples#

by Jimleuk

by David Roberts

by David Roberts

## Related resources#

Refer to Langchain's HuggingFace Inference embeddings documentation for more information about the service.

View n8n's Advanced AI documentation.

## AI glossary#

• completion: Completions are the responses generated by a model like GPT.

• hallucinations: Hallucination in AI is when an LLM (large language model) mistakenly perceives patterns or objects that don't exist.

• vector database: A vector database stores mathematical representations of information. Use with embeddings and retrievers to create a database that your AI can access when answering questions.

• vector store: A vector store, or vector database, stores mathematical representations of information. Use with embeddings and retrievers to create a database that your AI can access when answering questions.