

Self-service usage pricing plans

Contact us for enterprise pricing tiers, and volume discounts for private compute

Category	Product	Price
Inputs and vectors	Input and vector ingest general embeddings	\$0.0012 / request
	Input and vector ingest detection embeddings	\$0.002 / request
	Active inputs index (monthly)	\$0.001 / input
	Input and vector search	\$0.0012 / request
Image models	Pre-trained classification model	\$0.0012 / request
	Pre-trained detection model	\$0.002 / request
	Pre-trained segmentation model	\$0.0032 / request
	Pre-trained landmark model	\$0.002 / request
	Custom classification model	\$0.0032 / request
	Custom detection model	\$0.005 / request
	Custom segmentation model	\$0.008 / request
Language models	Pre-trained classification model	\$0.0012 / request
	Pre-trained named entity recognition model	\$0.002 / request
	Pre-trained translation model	\$0.002 / request
	Custom classification model	\$0.0032 / request

Category	Product	Price
Fine-tuning models	Quick or deep model fine tuning (training)	\$4 / hr
Large language models (third party)	AI21 Jurrasic-2 large	\$0.014 / request
	AI21 Jurrasic-2 grande	\$0.05 / request
	AI21 Jurrasic-2 jumbo	\$0.07 / request
	AI21 Jurrasic-2 grande instruct	\$0.05 / request
	AI21 Jurrasic-2 jumbo instruct	\$0.07 / request
	AI21 text-summarize	\$0.0055 / request
	Anthropic claude-V1	\$0.45 / request
	Anthropic claude-V2	\$0.45 / request
	Anthropic claude-instant	\$0.08 / request
	Claude-Instant-1.2	\$0.20 / request
	Claude Opus	\$0.3 / request
	Claude Haiku	\$0.2 / request
	Cohere-summarize	\$0.034 / request
	Cohere-generate-command	\$0.017 / request
	Cohere-text-to-embeddings	\$0.001 / request
	Cohere-multilingual-text-to-embeddings	\$0.001 / request
	Deepgram Aura	\$0.03 / request
	Databricks Instruct	\$0.01 / request
	Dolly-v2-12b	\$0.00011 / request

Introductory pricing per API request (token-based pricing scheme coming this summer).



Category**Product****Price**

Falcon-40b-instruct	\$0.0011 / request
---------------------	--------------------

GCP Code Generation	\$0.006 / request
---------------------	-------------------

GCP Embeddings Gecko	\$0.0005 / request
----------------------	--------------------

GCP PaLM 2 Text Bison	\$0.02 / request
-----------------------	------------------

GCP PaLM 2 Text Bison	\$0.02 / request
-----------------------	------------------

Llama2-70b-chat	\$0.0066 / request
-----------------	--------------------

OpenAI DALL-E	\$0.02 / request
---------------	------------------

OpenAI GPT-3_5-turbo	\$0.005 / request
----------------------	-------------------

OpenAI 3.5-turbo-instruct	\$0.005 / request
---------------------------	-------------------

OpenAI GPT 4	\$0.22 / request
--------------	------------------

OpenAI text-embedding-ada	\$0.0009 / request
---------------------------	--------------------

RedPajama-INCITE-Chat-7B	\$0.00022 / request
--------------------------	---------------------

StabilityAI Stable-Difussion-XL	\$0.022 / request
---------------------------------	-------------------

StarCoder	\$0.00175 / request
-----------	---------------------

WizardCoder-Python-34B	\$0.0022 / request
------------------------	--------------------

WizardLM-70B	\$0.0066 / request
--------------	--------------------



Category	Product	Price
Other	On premise deployments	Contact sales
	Edge video streaming/sensors	Contact sales
	Additional labeler seats	Contact sales
	Professional services (R&D, data labeling, AI Labs, training/education)	Contact sales
Other models	Fixed function or operator models	\$0.0012 / request

Need help getting started?

To get started visit clarifai.com to sign up for free. You can also [contact us](#) or visit docs.clarifai.com.