

Appendix to the Service Description

Live Intelligence – Available LLMs

1 Available LLMs

The table below lists the AI models (Large Language Models, or LLMs) currently available through Live Intelligence as of the date of this document. It also outlines the corresponding calculation method for unit consumption.

- For each user request (comprising both the question and the answer) units or partial units are consumed. The total unit consumption is calculated by adding together the following components, as detailed in the table below:
Input Processing Cost: Units consumed for processing the user's input.
- **Web Search Cost (if applicable):** Additional units consumed if a web search is performed.
- **Output Processing Cost:** Units consumed for generating the output.

Editor	LLM	Cost processing 1000 input (questions contexts) of 1000 tokens +	Additional cost of processing a request including a web search	Cost of processing 1000 output tokens (responses)
OpenAI Ireland Ltd	GPT-4o-mini	0.21 units	<i>Not available</i>	0.82 units
OpenAI Ireland Ltd	GPT-4o	3.43 units	<i>Not available</i>	13.71 units
OpenAI Ireland Ltd	GPT-4.1	2.74 units	14,00 units	10.97 units
OpenAI Ireland Ltd	GPT-4.1-mini	0.55 units	14,00 units	2.19 units
OpenAI Ireland Ltd	GPT-4.1-nano	0.14 units	<i>Not available</i>	0.55 units
OpenAI Ireland Ltd	GPT-5-mini	0.35 units	14,00 unités	2.74 units
OpenAI Ireland Ltd	GPT-5	1.72 units	14,00 unités	13.71 units
OpenAI Ireland Ltd	o3	13.71 units	<i>Not available</i>	54.86 units
OpenAI Ireland Ltd	o3-mini	1.51 units	<i>Not available</i>	6.03 units
Google Ireland Ltd	Gemini 2.0 Flash	0.21 units	<i>Not available</i>	0.82 units

Google Ireland Ltd	Gemini 2.5 Flash	0.42 units	<i>Not available</i>	3.42 units
--------------------	------------------	------------	----------------------	------------

Example calculations of Units consumption:

A query using GPT-4.1-mini with 1300 input tokens and 250 output tokens, without web search, consumes: $(1300/1000) \times 0.55 + (250/1000) \times 2.19 = 1.2625$ units

A query using GPT-4.1 with 4000 input tokens, with web search enable, and 300 output tokens, consumes:

$$(4000/1000) \times 2.74 + 14 + (300/1000) \times 10.97 = 28.251 \text{ units}$$