🏠   Quick Start   Token & Token Usage

Token & Token Usage

Tokens are the basic units used by models to represent natural language text, and also the units we use for billing. They can be intuitively understood as 'characters' or 'words'. Typically, a Chinese word, an English word, a number, or a symbol is counted as a token.

Generally, the conversion ratio between tokens in the model and the number of characters is approximately as following:

- 1 English character ≈ 0.3 token.
- 1 Chinese character ≈ 0.6 token.

However, due to the different tokenization methods used by different models, the conversion ratios can vary. The actual number of tokens processed each time is based on the model's return, which you can view from the usage results.

Calculate token usage offline

You can run the demo tokenizer code in the following zip package to calculate the token usage for your intput/output.

deepseek_v2_tokenizer.zip