**Azure AI Translator**

Azure AI Translator is a cloud-based neural machine translation service that is part of the [Azure AI services](https://learn.microsoft.com/en-us/azure/ai-services/what-are-ai-services) family and can be used with any operating system. Translator powers many Microsoft products and services used by thousands of businesses worldwide for language translation and other language-related operations.

**Translator features and development options**

A screenshot of a computer

Description automatically generated

**Azure AI Translator pricing**

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Service limits for Azure AI Translator Service**

**Text translation**

Charges are incurred based on character count, not request frequency. Character limits are subscription-based.

**Character and array limits per request**

Each translate request is limited to 50,000 characters, across all the target languages. For example, sending a translate request of 3,000 characters to translate to three different languages results in a request size of 3,000 × 3 = 9,000 characters and meets the request limit. You're charged per character, not by the number of requests, therefore, we recommend that you send shorter requests.

The following table lists array element and character limits for each text translation operation.

A screenshot of a computer

Description automatically generated

**Character limits per hour**

Your character limit per hour is based on your Translator subscription tier.

The hourly quota should be consumed evenly throughout the hour. For example, at the F0 tier limit of 2 million characters per hour, characters should be consumed no faster than roughly 33,300 characters per minute. The sliding window range is 2 million characters divided by 60 minutes.

You're likely to receive an out-of-quota response under the following circumstances:

* You reached or surpassed the quota limit.
* You sent a large portion of the quota in too short a period of time.

There are no limits on concurrent requests.

A screenshot of a computer

Description automatically generated

Limits for [multi-service subscriptions](https://learn.microsoft.com/en-us/azure/ai-services/translator/reference/v3-0-reference#authentication) are the same as the S1 tier.

These limits are restricted to Microsoft's standard translation models. Custom translation models that use Custom Translator are limited to 3,600 characters per second, per model.

**Latency**

The Translator has a maximum latency of 15 seconds using standard models and 120 seconds when using custom models. Typically, responses *for text within 100 characters* are returned in 150 milliseconds to 300 milliseconds. The custom translator models have similar latency characteristics on sustained request rate and can have a higher latency when your request rate is intermittent. Response times vary based on the size of the request and language pair. If you don't receive a translation or an [error response](https://learn.microsoft.com/en-us/azure/ai-services/translator/reference/v3-0-reference#errors) within that time frame, check your code, your network connection, and retry.

**Document Translation**

* Document Translation doesn't support translating secured documents such as those with an encrypted password or with restricted access to copy content.
* When translating documents with content in multiple languages (batch operations only), the feature is intended for complete sentences in a single language. If sentences are composed of more than one language, the content may not all translate into the target language.

**Asynchronous (batch) operation limits**

A white background with black lines

Description automatically generated

**Synchronous operation limits**

A screenshot of a computer

Description automatically generated