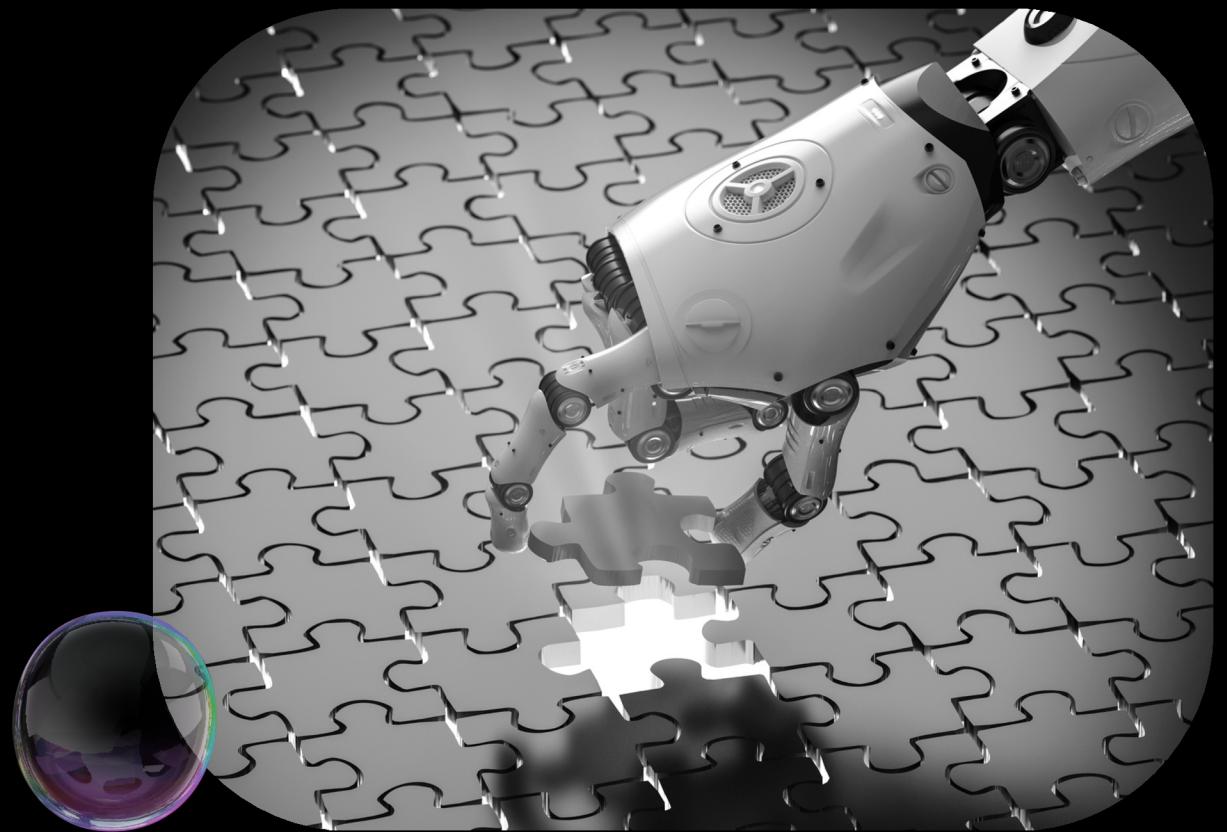


Azure Mastery: Building Smarter Bots

Token & Tokenization



Created by
Katerina Chernevskaya





Token

Definition

A common sequence of characters digitized for processing.

Tokenization converts text into tokens, allowing the model to understand and predict language patterns. This process varies between models, influencing how text is interpreted.

For instance, in common English, one token typically approximates 4 characters or about 3/4 of a word.



Created by
Katerina Chernevskaya



Token

Example

In the sentence **Hello, world!**, the tokens would be:

- Hello
- ,
- world
- !

Tokens	Characters
4	13

Hello, world!



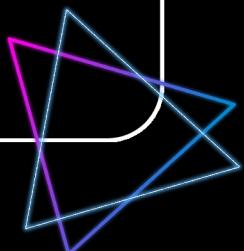
Created by
Katerina Chernevskaya



Token

Importance

- **Data Representation:** Tokens transform human language into numerical formats that AI models can process, encapsulating each token's semantic essence.
- **Memory and Computation:** Tokens help manage an AI model's capacity allowing developers to optimize performance by structuring inputs effectively.
- **Granularity and Flexibility:** Tokens vary in size, from words to characters, offering adaptability across different languages and applications.



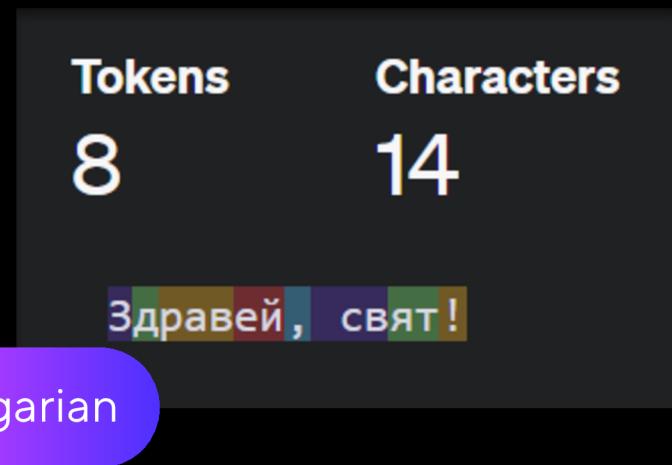
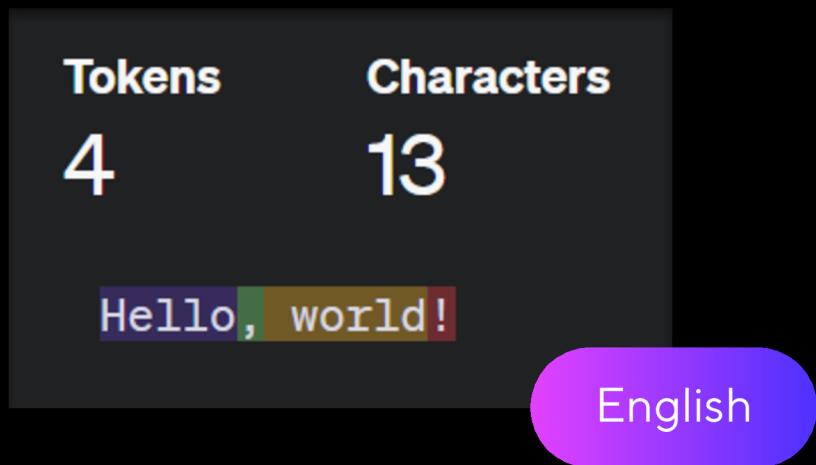
Created by
Katerina Chernevskaya



Tokenization

Tokenization differs between different scripts due to character sets and linguistic rules.

For example, Cyrillic may require different handling of suffixes and morphological endings, impacting how tokens are identified and used.



Created by
Katerina Chernevskaya



Tokenizer

A tool that converts text into tokens.

<https://platform.openai.com/tokenizer>

To use a Tokenizer, input your text into the tool, which then outputs a list of tokens.

The screenshot shows the OpenAI Tokenizer interface. At the top, there are three tabs: "GPT-4o & GPT-4o mini (coming soon)", "GPT-3.5 & GPT-4" (which is selected and highlighted in green), and "GPT-3 (Legacy)". Below the tabs, a text input field contains the text "Hello, world!". Underneath the input field are two buttons: "Clear" and "Show example". Further down, the results are displayed in a table format:

Tokens	Characters
4	13

Below the table, a small preview window shows the input text "Hello, world!" with each word tokenized into individual characters.

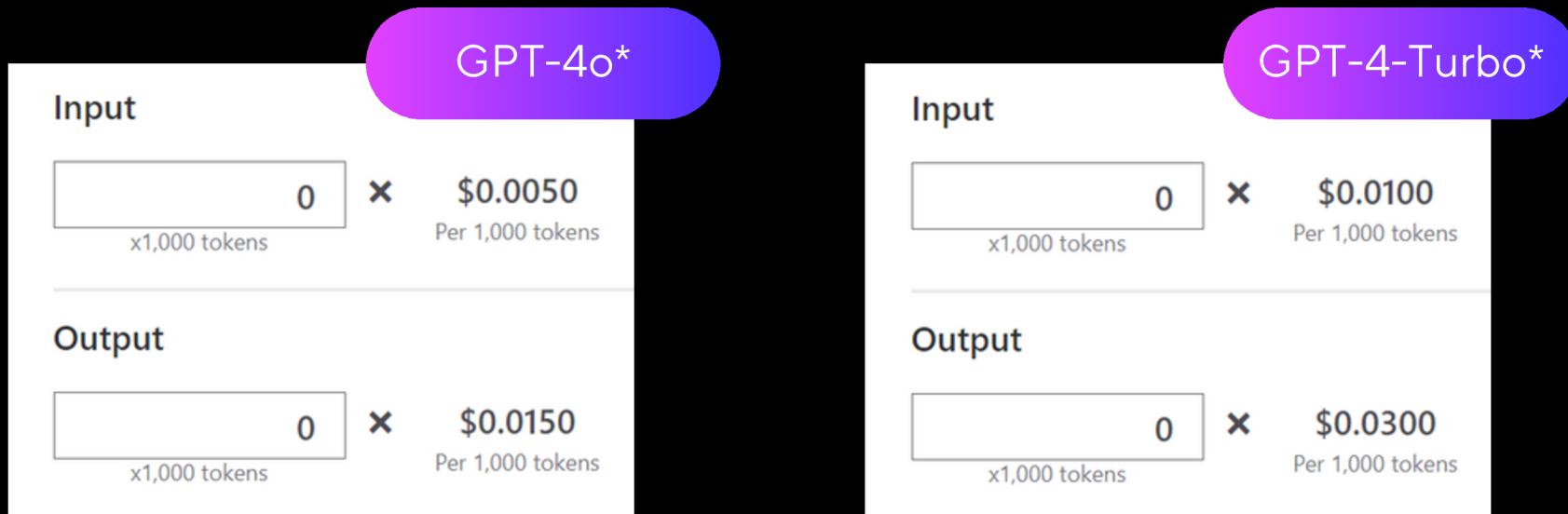


Created by
Katerina Chernevskaya



Token Cost impact

The number of tokens processed can **directly affect the cost** of your projects. More tokens often mean more data to process, which can increase the time and computational power needed, thus raising costs.



* Azure price calculator



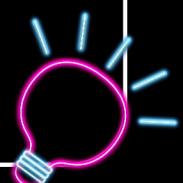
Created by
Katerina Chernevskaya



Token

Tips

- **Optimize Prompts:** Be mindful that every word in your prompt counts toward the total number of tokens, which can affect processing time and costs.
- **Monitor Token Output:** Regularly review the tokens generated from your text to ensure they accurately reflect the intended content and are relevant for your tasks.
- **Learn from Examples:** Observe how different AI models handle tokenization. Studying these examples can deepen your understanding of token usage and help improve your own NLP applications.



Created by
Katerina Chernevskaya



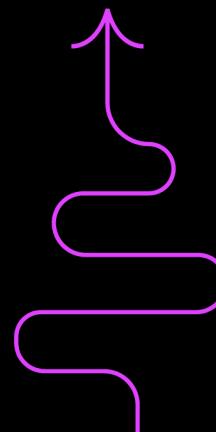
Azure Mastery: Building Smarter Bots

Wanna make
your chatbots
smarter



**REGISTER TODAY,
SECURE YOUR SEAT,
AND SAVE 50%!**

www.balticsummit.pl



BALTIC SUMMIT

UNLEASH THE POWER OF MICROSOFT SOLUTIONS

PPNT GDYNIA, POLAND

13 SEPTEMBER 2024

50% OFF WITH PROMO CODE: KATERINA_BPPC

ENHANCING CHATBOT
INTELLIGENCE WITH
AZURE AI STUDIO'S
PROMPT FLOW

REGISTER NOW

KATERINA CHERNEVSKAYA

www.balticsummit.pl



Azure Mastery: Building Smarter Bots

Key Terms

Follow for more!



Created by
Katerina Chernevskaya