

Start coding or [generate](#) with AI.

```
sample_text = "Tokenization is the process of converting text into smaller units called tokens."
```

```
!pip install transformers
```

```
from transformers import BertTokenizer, GPT2Tokenizer, RobertaTokenizer
import time
```

```
bert_tokenizer = BertTokenizer.from_pretrained('bert-base-uncased')
```

```
gpt2_tokenizer = GPT2Tokenizer.from_pretrained('gpt2')
```

```
roberta_tokenizer = RobertaTokenizer.from_pretrained('roberta-base')
```

```
def analyze_tokenizer(tokenizer, name):
```

```
    start = time.time()
```

```
    tokens = tokenizer.tokenize(sample_text)
```

```
    encoded = tokenizer.encode(sample_text)
```

```
    end = time.time()
```

```
    print(f"\n🔍 {name} Tokenizer:")
```

```
    print(f"Tokenized Output: {tokens}")
```

```
    print(f"Number of Tokens: {len(tokens)}")
```

```
    print(f"Token IDs: {encoded}")
```

```
    print(f"⌚ Time Taken: {round(end - start, 5)} seconds")
```

```
analyze_tokenizer(bert_tokenizer, "BERT")
```

```
analyze_tokenizer(gpt2_tokenizer, "GPT-2")
```

```
analyze_tokenizer(roberta_tokenizer, "RoBERTa")
```

```
Requirement already satisfied: transformers in /usr/local/lib/python3.11/dist-packages (4.51.1)
Requirement already satisfied: filelock in /usr/local/lib/python3.11/dist-packages (from transformers) (3.18.0)
Requirement already satisfied: huggingface-hub<1.0,>=0.30.0 in /usr/local/lib/python3.11/dist-packages (from transformers) (0.30.2)
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.11/dist-packages (from transformers) (2.0.2)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.11/dist-packages (from transformers) (24.2)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.11/dist-packages (from transformers) (6.0.2)
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.11/dist-packages (from transformers) (2024.11.6)
Requirement already satisfied: requests in /usr/local/lib/python3.11/dist-packages (from transformers) (2.32.3)
Requirement already satisfied: tokenizers<0.22,>=0.21 in /usr/local/lib/python3.11/dist-packages (from transformers) (0.21.1)
Requirement already satisfied: safetensors>=0.4.3 in /usr/local/lib/python3.11/dist-packages (from transformers) (0.5.3)
Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.11/dist-packages (from transformers) (4.67.1)
Requirement already satisfied: fsspec>=2023.5.0 in /usr/local/lib/python3.11/dist-packages (from huggingface-hub<1.0,>=0.30.0->transformers) (2024.10.0)
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.11/dist-packages (from huggingface-hub<1.0,>=0.30.0->transformers) (4.12.2)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests->transformers) (3.4.1)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests->transformers) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests->transformers) (2.3.0)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests->transformers) (2025.1.31)
```

🔍 BERT Tokenizer:

Tokenized Output: ['token', '##ization', 'is', 'the', 'process', 'of', 'converting', 'text', 'into', 'smaller', 'units', 'called', 'tokens']

Number of Tokens: 15

Token IDs: [101, 19204, 3989, 2003, 1996, 2832, 1997, 16401, 3793, 2046, 3760, 3197, 2170, 19204, 2015, 1012, 102]

⌚ Time Taken: 0.00192 seconds

🔍 GPT-2 Tokenizer:

Tokenized Output: ['Token', 'ization', 'is', 'the', 'process', 'of', 'converting', 'text', 'into', 'smaller', 'units', 'called']

Number of Tokens: 14

Token IDs: [30642, 1634, 318, 262, 1429, 286, 23202, 2420, 656, 4833, 4991, 1444, 16326, 13]

⌚ Time Taken: 0.00345 seconds

🔍 RoBERTa Tokenizer:

Tokenized Output: ['Token', 'ization', 'is', 'the', 'process', 'of', 'converting', 'text', 'into', 'smaller', 'units', 'called']

Number of Tokens: 14

Token IDs: [0, 45643, 1938, 16, 5, 609, 9, 18841, 2788, 88, 2735, 2833, 373, 22121, 4, 2]

⌚ Time Taken: 0.00144 seconds