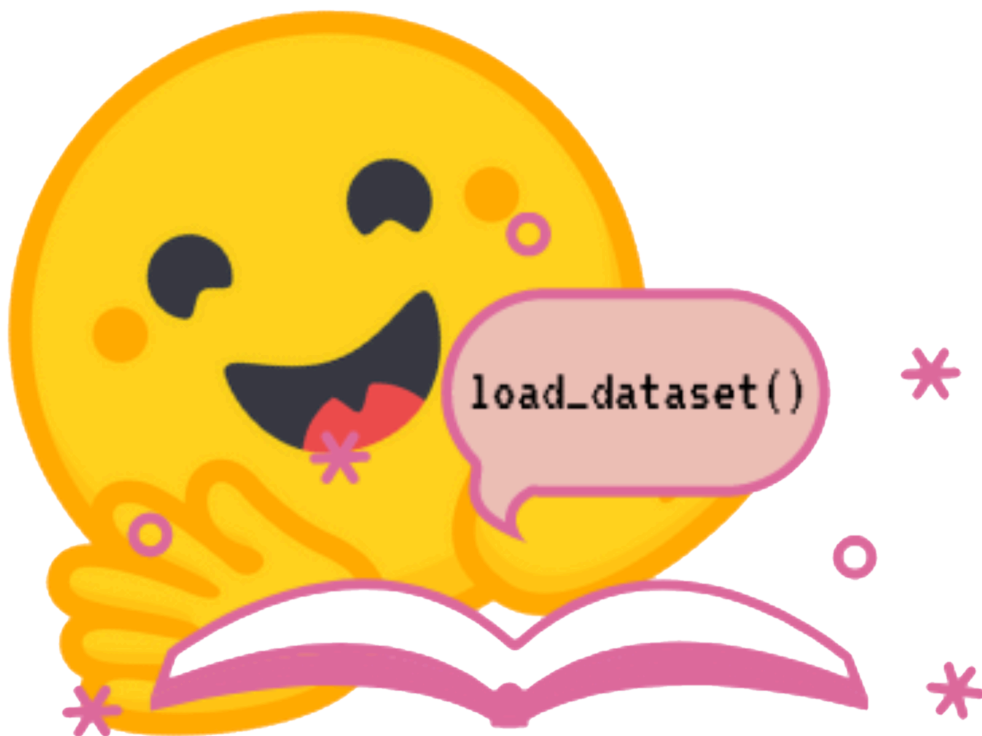



Datasets




 Datasets is a library for easily accessing and sharing AI datasets for Audio, Computer Vision, and Natural Language Processing (NLP) tasks.

Load a dataset in a single line of code, and use our powerful data processing and streaming methods to quickly get your dataset ready for training in a deep learning model. Backed by the Apache Arrow format, process large datasets with zero-copy reads without any memory constraints for optimal speed and efficiency. We also feature a deep integration with the


[Hugging Face Hub](#), allowing you to easily load and share a dataset with the wider machine learning community.

Find your dataset today on the [Hugging Face Hub](#), and take an in-depth look inside of it with the live viewer.

Tutorials

Learn the basics and become familiar with loading, accessing, and processing a dataset. Start here if you are using  Datasets for the first time!

How-to guides

Practical guides to help you achieve a specific goal. Take a look at these guides to learn how to use  Datasets to solve real-world problems.

Conceptual guides

High-level explanations for building a better understanding about important topics such as the underlying data format, the cache, and how datasets are generated.

Reference

Technical descriptions of how  Datasets classes and methods work.