

[Technology Overviews](#) / [Apple Intelligence and machine learning](#) / Intelligent frameworks

Intelligent frameworks

Analyze images, sound, and language without needing to be a machine learning expert.

Overview

Building an app that uses machine learning doesn't have to be complicated. There are several on-device models that many frameworks lean on to perform analysis. Apple trains and optimizes these on-device models for efficiency on device, allowing frameworks to perform the heavy lifting for you so you can focus on building your features. When you use a framework that's powered by machine learning, you don't have to worry about collecting a large amount of data and taking the time to train and optimize a model by yourself.

Starting with one of the available frameworks provides a great introduction to thinking about machine learning and building intelligent apps. Whether you want to analyze images, sound, or language, you make requests to the system that performs the work and returns the result that you use in your app.

Know more about images by analyzing photos or video frames

Computer vision allows for better understanding of the world around us. When you work with photos and videos, you might want to know more about what's happening in them to create the feature you want in your app. For example, you don't have to start from zero to [recognize the content in a document](#). The [Vision](#) and [VisionKit](#) frameworks perform a wide variety of tasks that do the heavy lifting for you, and provides more than 25 types of image analysis tasks, like:

- Capture text within the camera frame by enabling [Live Text in your app's image views](#).
- Identify objects, text, bar codes, documents, and more in images or the [live camera feed](#).

- Track the movement of objects across images or video frames.
- Detect face and body poses for people and animals.
- Determine the trajectory of shapes along a path.

To prevent people from viewing unwanted image content in your app, perform sensitive content analysis to help detect it.

Perform speech recognition in recorded or live audio

Speech recognition transforms spoken words into text to help you with things like dictating notes in a note-taking app. Or, you might allow people to speak voice commands to control a smart thermostat. The Speech framework helps you convert audio of human speech into text with very little code, and entirely on device. The framework uses the same speech-to-text model powering Siri and is especially good for long-form and distant audio, like meetings and lectures.

The framework allows you to provide audio from either prerecorded files or a live source like a microphone. You pass audio to a SpeechAnalyzer that uses a speech-to-text model to predict the text that matches the audio. Analysis is asynchronous and the framework decouples the input and output to allow you to work with results as they happen while the framework processes the input.

Detect the song that's being played

Audio recognition allows you to identify a piece of audio being played in an environment to get more information about it or to perform actions that are in sync with it. For example, if you're building a trivia app you might want to know when an interactive moment occurs in the audio stream so you can synchronize app content with it. The ShazamKit framework helps you match audio against the vast music catalog that Shazam provides, or match against your own prerecorded reference audio in a custom catalog.

ShazamKit lets you perform a match by converting audio into a special format called signatures. You pass in a stream of audio buffers — or the signature data — into a session, and the session uses the data to find a match in the Shazam catalog or in the catalogs you provide. If there's a match, the session returns a match object that contains the media items that represent the metadata of the match.

For heavy-duty audio recognition workflows, macOS includes a command line utility for Shazam under the hood. For additional details, open Terminal and run `man shazam`.

Classify the type of sound in the audio

Real-time sound recognition capabilities can enhance the accessibility of the app you create. Integrate sound analysis capabilities when your app needs to [classify sounds](#) in real time to identify environmental sounds, like glass breaking or dog barks. Or if you're building a music making app, use sound analysis to identify the instrument being played.

Optimized for hardware acceleration, the sound classification API provides a specialized machine learning audio expertise by classifying over 300 categories of sounds on-device. On-device processing helps preserve user privacy because the audio isn't sent to the cloud.

For apps that have a particular sound that needs to be detected, use the [Create ML app](#) to customize a sound analysis model with your own data.

Segment natural language into lexical units

Natural language processing helps you understand and process human language and allows your app to extract meaning from text. If you're working on a multilingual app, you may want to [identify the language](#) in text to help convert it on behalf of the person using your app. For example, a language learning app might identify that the input a person provides matches the language the app expects. The framework breaks strings down into lexical units — like words or sentences — and ensures correct behavior in multiple script languages. For example, Chinese doesn't use spaces to delimit words.

The framework helps you perform tasks like [finding the similarities between pieces of words or sentences](#), which can improve search in your app by including related words to an original search query.

Add language translation

Language translation helps people understand the world around them. If a person using your app is traveling to a foreign country, your app might help them translate the transit signs they see in a subway or train station. Or an app that's made for exploring a museum might translate the text that describes an exhibit. The [Translation](#) framework helps you perform text translations between multiple languages, and uses the same on-device model shared across the OS for language translation.

The framework allows you to translate between many [supported languages](#), and provides your app access to the languages a person has already downloaded. Start translating text in your app by applying a translation [view modifier](#) to show the system UI translation popover.

When you need to [customize the translation experience](#) in your app, use a translation session directly. The system provides a session automatically when you attach the [translation modifier](#) to your view, and allows for performing a single translation or a [batch translation](#) to translate several strings at once.