

Default models included with LumiFind

Introduction

LumiFind uses a collection of specialized AI models to analyze documents, images, voice recordings, and videos. Each model is downloaded once and then reused, and larger models generally offer better accuracy at the cost of higher memory and processing needs.

Document Models

These models handle text understanding and semantic search. Each one is downloaded once. Larger models generally provide better accuracy but require more memory and processing power. If you are using the Pro Edition, voice and video analysis will also use the selected document model for their text processing steps.

all-MiniLM-L6-v2 ~90MB

Fast and lightweight. Ideal for general text search, support articles, notes and low resource systems.

paraphrase-multilingual-MiniLM-L12-v2 ~420MB

Better semantic accuracy across many languages. Good for mixed-language documents and everyday multilingual search.

distiluse-base-multilingual-cased-v2 ~540MB

Cased model that preserves capitalization. Useful for academic papers, technical documents, abbreviations and proper nouns.

paraphrase-multilingual-mpnet-base-v2 ~1.1GB

Highest accuracy among the defaults. Best for large knowledge bases, precise semantic matching and deep contextual search.

Image Models

These models extract visual features from images for similarity search and cross modal search. Larger models detect finer details but need more resources. Currently, image search can only be done in English-

clip-vit-base-patch32 ~615MB

Balanced speed and accuracy. Suitable for most general-purpose image search tasks.

clip-vit-large-patch14 ~1.5GB

Higher accuracy for detailed images. Best for design libraries, product databases and visually complex datasets.

siglip-base-patch16-224 ~820MB

Strong performance especially on semantic similarity. Good for modern datasets and multilingual text to image matching.

Voice Models

These models transcribe audio into text before it can be searched. Bigger models offer noticeably higher accuracy, especially for long or noisy recordings.

Video files will also use these models for their audio tracks.

whisper-tiny ~151MB

Very fast. Works well for clean audio, short notes and quick transcriptions on low resource systems.

whisper-base ~290MB

Better accuracy while staying lightweight. Good for everyday recordings, meetings and clear speech.

whisper-small ~967MB

Stronger performance in noisy environments. Good for interviews, lectures and field recordings.

whisper-medium ~3.06GB

High accuracy across multiple languages. Great for professional transcriptions and detailed audio analysis.

whisper-large-v2 ~6.19GB

Highest accuracy. Best for multilingual projects, long recordings, and scenarios requiring maximum transcription quality.

Video Frame Detectors

Videos consist of multiple images (maybe you know that already) called frames. LumiFind can analyze the frames with the given intervals to extract the object(s) in that frame. To do that, it needs frame detectors.

opencv_ssd ~23MB

It is a fast and small model which can detect 20 common objects and can also run on older hardware.

It can detect these objects: aeroplane, bicycle, bird, boat, bottle, bus, car, cat, chair, cow, diningtable, dog, horse, motorbike, person, pottedplant, sheep, sofa, train, tvmonitor

opencv_ssd_coco ~180MB

It is slightly bigger than opencv_ssd, but can detect 80 different common objects.

It can detect these objects: person, bicycle, car, motorcycle, airplane, bus, train, truck, boat, traffic light, fire hydrant, stop sign, parking meter, bench, bird, cat, dog, horse, sheep, cow, elephant, bear, zebra, giraffe, backpack, umbrella, handbag, tie, suitcase, frisbee, skis, snowboard, sports ball, kite, baseball bat, baseball glove, skateboard, surfboard, tennis racket, bottle, wine glass, cup, fork, knife, spoon, bowl, banana, apple, sandwich, orange, broccoli, carrot, hot dog, pizza, donut, cake, chair, couch, potted plant, bed, dining table, toilet, tv, laptop, mouse, remote, keyboard, cell phone, microwave, oven, toaster, sink, refrigerator, book, clock, vase, scissors, teddy bear, hair drier, toothbrush

detr / detr_resnet50 ~160MB

This is a model from Facebook, which is slower, but good at detecting objects. If accuracy is more important than speed.

It can detect the same objects as [opencv_ssd_coco](#) offers, but with higher accuracy.

mediapipe_object ~5MB

Another small model from Google, which is battery friendly and can also work on older laptops.

It can detect the same objects as [opencv_ssd_coco](#) offers, but with lower accuracy.

Your Own Models

You can add your own AI models to LumiFind to customize analysis for your specific needs. For instructions on how to do this, see [this page](#). Once added, these models can be selected alongside the default ones.

Revision #5

Created 2025-11-26 23:09:55 UTC by MIDILLI

Updated 2026-02-19 13:52:15 UTC by MIDILLI