AMBIENT-IT

+33 (0)1 48 10 09 46
formation@ambient-it.net
140 Avenue Jean Lolive 93500 Pantin - France

Updated on 02/04/2026

Register

Moondream AI Training

2 days (14 hours)

Overview

Moondream AI is a lightweight vision-language model for describing images, extracting information, and answering visual questions. You will learn how to quickly integrate it to automate use cases such as quality control, image indexing, or input assistance via OCR and scene understanding.

This training course aims to provide an operational introduction to Moondream AI: loading the model, inferring images, effective prompts, and structuring outputs (JSON, business fields). The exercises are based on concrete scenarios (product photos, documents, screenshots) and simple metrics to validate quality.

The approach focuses on workshops and reproducible demos: preprocessing pipeline, performance management (CPU/GPU), and integration into an API. Deliverables include a notebook/mini-project, an inference script, and a checklist of best practices (prompts, latency, errors).

Like all our training courses, this one will introduce you to **the latest stable version** of the technology and its new features.

Objectives

- Install and configure Moondream AI in a Python environment.
- Perform image-text inferences and visual questions/answers.
- Design robust prompts and structured, usable outputs.
- Optimize latency and memory consumption according to CPU/GPU.
- Expose the model via a simple API and test it in real-world conditions.

Target audience

- Python developers who want to integrate AI vision.
- Data scientists/ML engineers focused on rapid prototyping.
- Backend engineers who want to industrialize a VLM component.
- Technical project managers evaluating vision use cases.

Prerequisites

- Good foundation in Python (packages, environments, scripts).
- Basic knowledge of machine learning and model inference.
- Handling image files and formats (JPEG/PNG) and JSON.
- Basic knowledge of HTTP API (requests, payloads).

Technical prerequisites

- Computer with 16 GB of RAM recommended (8 GB minimum).
- Windows (ideally WSL2), macOS, or Linux.
- Python 3.10+, environment management (venv/conda) and pip.
- GPU option: NVIDIA with compatible drivers and stack (depending on framework), otherwise CPU execution.
- Code editor (VS Code or equivalent) and a terminal.

Moondream AI training program

[Day 1 - Morning]

Discover Moondream AI and prepare the environment

- Understand positioning: VLM (Vision-Language Model) and use cases (caption, VQA, visual extraction)
- Choose the execution mode: local CPU/GPU, memory constraints, latency, and cost
- Install and validate the Python environment (venv, dependencies, versions) and an initial inference script
- Structure a project: manage inputs (images), outputs (JSON), logs, and configuration
- Hands-on workshop: Install Moondream, run an inference on 10 images, and export the results in JSON.

[Day 1 - Afternoon]

Vision prompts & reliable information extraction

- Designing task-oriented prompts: description, question/answer, structured extraction (expected fields)
- Reducing ambiguity: instructions, output format, constraints (units, closed lists, priority to visible elements)
- Managing difficult cases: low light, small text, multiple objects, noisy background

- Put safeguards in place: schema validation, business rules, uncertainty detection, and manual recovery
- Hands-on workshop: Building an extractor (JSON) to analyze product photos (name, color, logo presence, condition).

[Day 2 - Morning]

Industrializing an image processing pipeline

- Pre-processing: resizing, cropping, normalization, format management, and EXIF orientation
- Batch processing: queues, parallelism, error recovery, and idempotence
- Measuring quality: test sets, criteria (accuracy, completeness), sampling, and human review
- Optimizing performance: cache, resolution choice, context limitation, and timeout management timeouts
- Hands-on workshop: Create a CLI pipeline that ingests a folder, processes it in batches, and produces a quality report (CSV + errors).

[Day 2 - Afternoon]

Exposing Moondream in service and securing usage

- Create a simple API (image analysis endpoint) with upload management and stable JSON responses
- Best practices for robustness: size limits, quotas, MIME validation, error management, and observability
- Deployment: packaging, environment variables, configuration, logs, and scaling strategy
- Compliance and risks: sensitive data, anonymization, retention, traceability, and access policies access policies
- Hands-on workshop: Deploying an image analysis microservice with endpoint /analyze and basic load testing.

Target companies

This training is intended for both individuals and companies, large or small, wishing to train their teams in a new advanced IT technology or to acquire specific business knowledge or modern methods.

Placement at the start of training

The positioning at the start of the training complies with Qualiopi quality criteria. Upon final registration, the learner receives a self-assessment questionnaire that allows us to assess their estimated level of knowledge of different types of technologies, their expectations, and personal objectives

regarding the upcoming training, within the limits imposed by the selected format. This questionnaire also allows us to anticipate certain connection or internal security issues within the company (intra-company or virtual classroom) that could be problematic for the monitoring and smooth running of the training session.

Teaching methods

Practical training: 60% practical, 40% theory. Training materials distributed in digital format to all participants.

Organization

The course alternates between theoretical input from the trainer, supported by examples and discussion sessions, and group work.

Assessment

At the end of the session, a multiple-choice questionnaire is used to verify that the skills have been correctly acquired.

Certification

A certificate will be issued to each trainee who has completed the entire training course.