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

Updated on 03/12/2026

Sign up

Kyutai Helium Training: Sovereign Generative AI

2 days (14 hours)

Overview

Kyutai Helium enables you to deploy sovereign generative AI for chat, internal support, and document search, while maintaining control over your data. This training guides you through the process of quickly moving from a POC to operational use on-premises or in a private cloud.

During this training, you'll learn to install, configure, and use Helium for real-world use cases: conversational agents, augmented FAQs, assisted writing, and document analysis. The focus is on security, governance, and response quality (prompting, safeguards, evaluation).

The approach is 100% hands-on: guided demos, step-by-step workshops, test exercises, and real-world scenarios. You'll leave with a working instance, reusable prompts, a deployment checklist, and an operationalization plan (monitoring, logs, updates, best practices).

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 launch a Helium instance in a controlled environment.
- Configure models, inference parameters, and usage limits.
- Design robust prompts and business-oriented templates.
- Implement safeguards (policies, filtering, traceability).
- Evaluate and improve quality (testing, metrics, iterations).

Target audience

- Developers and software engineers
- Data/ML engineers and MLOps
- Architects and infrastructure/security managers
- Technical/innovation project managers

Prerequisites

- Basic knowledge of generative AI and LLM
- Knowledge of Linux (shell, services, files)
- Understanding of API and HTTP concepts
- Security basics (network, access, secrets)

Technical prerequisites

- Computer with at least 16 GB of RAM (32 GB recommended)
- Linux or macOS; Windows is possible via WSL2
- Administrator access, terminal, and a code editor
- Tools: Docker (recommended), Git, curl
- Network connection to fetch dependencies and images (if permitted)

Our Kyutai Helium training program: Sovereign Generative AI

[Day 1 - Morning]

Understanding Kyutai Helium and the challenges of sovereign generative AI

- Positioning Helium: use cases, limitations, differences vs. cloud LLMs
- Clarifying key concepts: model, token, context, latency, cost
- Identifying sovereignty requirements: data localization, control, traceability
- Defining a corporate usage framework: authorized data, confidentiality, responsibilities
- Hands-on workshop: Mapping 3 internal use cases and their constraints (data, risks, ROI).

[Day 1 - Afternoon]

Getting started and prototyping: from prompt to reproducible workflow

- Structuring effective prompts: role, objective, constraints, output format, quality criteria
- Implementing templates: configurable prompts, variables, validation checklists
- Reducing errors: hallucinations, ambiguities, overconfidence, wording biases
- Evaluating results: example sets, acceptance criteria, rapid iterations
- Hands-on workshop: Building a mini-workflow for “drafting + proofreading + summarization” with versioned prompts.

[Day 2 - Morning]

RAG and internal data: connecting Helium to a document database

- Understanding the RAG principle: chunking, embeddings, search, generation
- Preparing documents: cleaning, segmentation, metadata, version control
- Defining a search strategy: top-k, filters, thresholds, citations, and sources
- Implementing relevance tests: sample questions, coverage, error rate
- Hands-on workshop: Designing a RAG pipeline on an internal corpus (FAQs/procedures) with quality criteria.

[Day 2 - Afternoon]

Sovereign deployment: security, governance, and operations

- Deploying in a controlled environment: isolation, network access, secret management
- Securing usage: access control, logging, data leak prevention
- Establishing governance: prompt policies, validation, incident management
- Monitoring performance: latency, failure rate, perceived quality, infrastructure costs
- Hands-on workshop: Drafting an operations plan (runbook) and a usage policy for an internal assistant.

Target Audience

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

Assessment upon enrollment

The pre-training assessment complies with Qualiopi quality standards. Upon final registration, the learner receives a self-assessment questionnaire that allows us to evaluate their estimated proficiency with various types of technologies, as well as their expectations and personal goals for 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 pose challenges for monitoring and ensuring the 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 instructor, supported by examples and reflection sessions, and group work.

Assessment

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

Certification

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