AMBIENT-IT

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

Updated on 10/26/2026

Sign Up

Arize Phoenix Training

2 days (14 hours)

Overview

Arize Phoenix helps you understand, evaluate, and improve applications based on large language models. With OpenInference and OpenTelemetry to instrument processing, traces and evaluations become tools for ensuring the reliability of your RAGs and AI agents in real-world conditions.

Our Arize Phoenix training will help you move beyond ad-hoc checks to build a structured approach to LLM observability, from local diagnostics to production monitoring.

You'll learn how to instrument an application, explore its traces, and identify failures in document retrieval, content generation, or tool invocation. You'll automatically evaluate responses and calibrate an LLM evaluator based on human annotations and explicit business criteria.

By the end of the training, you'll be able to compare different versions of prompts or models using reproducible experiments to make decisions based on measurable results. You'll know how to interpret the trade-offs between quality, latency, and costs.

This training also covers data protection, monitoring, and the role of Arize AX in production operations. Four workshops will guide you through a complete cycle of instrumentation, diagnostics, evaluation, and continuous improvement.

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

Objectives

- Explain the principles of LLM observability and the respective roles of Phoenix and Arize AX
- Instrument an application with OpenInference and analyze its traces in Phoenix
- Diagnose failures in a RAG and an agent using tools
- Automate evaluations and calibrate an LLM based on human annotations
- Compare versions of prompts or models through reproducible experiments
- Define a production monitoring strategy that integrates security, quality, latency, and costs

Target Audience

- AI Engineer
- LLM Application Developer
- Data Scientist
- MLOps Engineer
- Software Architect

Prerequisites

- Experience with Python programming and dependency management
- Knowledge of LLM principles and prompting techniques
- Initial experience with a RAG or an agent using tools
- Ability to consume a REST API and manipulate JSON-formatted data in an application

Technical Prerequisites

- Computer running Windows 11, macOS, or Linux with at least 8 GB of RAM; 16 GB recommended for workshops
- Python 3.11 environment with pip and permissions to install course dependencies
- Visual Studio Code editor or JupyterLab environment and a recent web browser
- Stable internet connection allowing for package downloads and calls to the model services used
- Access to a local Arize Phoenix instance or one provided for the session, and an API key from a with available quota

Arize Phoenix Training Agenda

[Day 1 - Morning]

Understanding and Implementing LLM Observability

- Challenges of LLM observability: behavior, quality, errors, latency, and token consumption
- Roles of Phoenix and Arize AX: diagnostics, evaluation, experimentation, and production deployment
- Tracing model: traces, spans, parent-child relationships, attributes, and context propagation

- Integration between OpenInference and OpenTelemetry: semantic conventions, instrumentation, and trace export
- Installing Phoenix, configuring projects, and automatic or manual instrumentation of a Python application
- Hands-on workshop: Instrument an LLM application with OpenInference, send its traces to Phoenix, and verify the collected attributes.

[Day 1 - Afternoon]

Tracing and diagnosing a RAG and an agent

- Reading an RAG trace: query, document search, transmitted context, and generated response
- Diagnosing retrieval issues: irrelevant documents, incomplete context, and obsolete sources
- Analysis of the generation process: instructions, fidelity to context, errors, and latency per step
- Tracing an agent: orchestration, tool calls, arguments, results, and execution trajectory
- Investigating agent failures: wrong tool, loop, execution error, and excessive token consumption
- Hands-on workshop: Diagnose a failing RAG and agent, identify the responsible steps and verifying a fix in the traces.

[Day 2 - Morning]

Evaluating responses and calibrating a judge LLM

- Building a representative evaluation dataset: examples from logs, edge cases, and ground-truth responses
- Selecting quality criteria: document relevance, accuracy, contextual fidelity, and task success task success
- Combining deterministic evaluators, human annotations, and the judge LLM
- Designing a scoring rubric: instructions, labels, examples, and justification of scores
- Calibrating the judge: agreement with human annotations, analysis of disagreements, biases, and validation using separate examples
- Hands-on workshop: Annotate a sample, automate its evaluation in Phoenix, and fine-tune the judge LLM by analyzing its disagreements with human evaluations.

[Day 2 - Afternoon]

Experimenting and preparing for production monitoring

- Organizing Phoenix experiments: controlled versions of datasets, prompts, models, and parameters
- Comparison of variants on the same inputs: scores, analysis by example, latency, costs Estimates and regressions
- Trace security: masking before export, exclusion of secrets, access control, and retention policy

- Defining production monitoring: quality, errors, latency, tokens, costs, and evaluation sampling
- Operations with Arize AX: dashboards, monitors, alerts, and reuse of instrumentation OpenInference
- Hands-on workshop: Compare two variants of a prompt or model, select a configuration based on the results, verify data masking, and define monitoring thresholds for production.

Target Audience

This training is intended for both individuals and companies—large or small—seeking to train their teams in new, advanced IT technologies or to acquire specific domain knowledge or modern methodologies.

Entry-Level Assessment

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 level across 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 operation of the training session.

Teaching Methods

Hands-On Training: 60% Hands-On, 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 quiz is administered to verify that participants have successfully acquired the skills.

Certification

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

Training organization registered under number 11 75 54743 75. This registration does not constitute state accreditation.
© Ambient IT 2015-2026. All rights reserved. Paris, France - Switzerland - Belgium - Luxembourg