

Updated 03/28/2024

Sign up

Deepset training

4 days (28 hours)

Presentation

Our Deepset training course will plunge you into the heart of LLMs, providing you with the knowledge and skills you need to master these cutting-edge technologies.

This training program covers a wide range of topics, from an introduction to LLMs to advanced aspects such as infrastructure security, scalable deployment and hallucination detection.

You'll be introduced to the fundamental concepts of LLMs and explore their various applications in different industrial fields.

With our training course, you'll learn how to use predefined workflows and components for common use cases.

You'll also discover how to integrate and evaluate models with [Haystack](#), while ensuring infrastructure security and compliance.

At the end of this course, we'll be using the [latest](#) Deepset [resources](#).

Objectives

- Understand large-scale language models (LLMs) and their practical applications
- Master the use of workflows and predefined components
- Observe, monitor and evaluate LLM models in production
- Integrate and evaluate models with Haystack
- Efficiently deploy and scale LLM-based solutions
- Optimize model performance with prompt engineering and fine-tuning

Target audience

- Developers
- Software engineers
- AI solution architect
- Data scientists
- AI Analysts

Prerequisites

- Basic knowledge of programming and data manipulation is recommended
- Prior knowledge of language models and artificial intelligence would also be beneficial.

OUR DEEPCSET TRAINING PROGRAM

INTRODUCTION TO LARGE-SCALE LANGUAGE MODELS (LLM)

- Understanding the difference between a traditional language model and an LLM
- Leading LLM providers: OpenAI, Cohere, Anthropic
- Practical applications of LLMs in industry
- The challenges of using LLMs
- Overview of tools and platforms for working with LLMs

RAPID PROTOTYPING WITH Deepset

- The fundamentals of rapid prototyping with LLMs
- Use of pre-built components to accelerate development
- Setting up a development environment for prototyping
- Rapid prototyping case studies with Deepset
- Best practices for scalable prototyping

USE OF WORKFLOWS AND PREDEFINED COMPONENTS

- Introduction to predefined workflows for common use cases
- Understanding and applying the RAG concept
- Techniques for information extraction and semantic search
- Using templates for document similarity
- Customize components for specific needs

LLM OBSERVABILITY AND MONITORING

- The importance of observability and monitoring in LLM Ops operations
- Setting up observability tools for LLMs
- Monitoring techniques to ensure performance and reliability
- Incident and anomaly management
- Using user feedback for continuous improvement

HAYSTACK INTEGRATION AND MODEL EVALUATION

- Introducing Haystack and its integration with Deepset
- Language model evaluation methods
- Collecting and integrating user feedback
- Improving model performance through feedback
- Integration and evaluation case studies with Haystack

LLM INFRASTRUCTURE SECURITY AND COMPLIANCE

- Importance of SOC2 certification for LLM backends
- Ensuring data security with MFA and SSO
- Understanding and implementing a private data plan in a VPC
- Best practices for maintaining a secure, compliant infrastructure
- Practical examples of data security implementation

SCALABLE DEPLOYMENT AND PRODUCTION INFRASTRUCTURE

- Understanding the challenges of scalable inference for LLMs
- Deployment techniques for low latency and high availability
- Automatic scaling options, including scaling to zero
- Customized pipeline management in the production environment
- Practical deployment and scaling cases

DETECTION AND MITIGATION OF HALLUCINATIONS

- Identifying LLM hallucinations and their impact
- Tested tools for detecting and reducing hallucinations
- Strategies for building reliable, consistent AI
- Automatic monitoring and detection of inconsistencies
- Case studies in hallucination management

PROMPT ENGINEERING AND MODEL FINE-TUNING

- Prompt engineering principles for LLMs
- Explore and adjust prompts to optimize performance
- Fine-tuning techniques with GPU notebooks
- Integrating human assessment into the fine-tuning process
- Prompt engineering and fine-tuning case studies

Companies concerned

This course is aimed at both individuals and companies, large or small, wishing to train their teams in a new advanced computer technology, or to acquire specific business knowledge or modern methods.

Positioning on entry to training

Positioning at the start of training complies with Qualiopi quality criteria. As soon as registration is finalized, the learner receives a self-assessment questionnaire enabling us to assess his or her estimated level of proficiency in different types of technology, as well as his or her expectations and personal objectives with regard to the training to come, within the limits imposed by the selected format. This questionnaire also enables us to anticipate any connection or security difficulties within the company (intra-company or virtual classroom) which could be problematic for the follow-up and smooth running of the training session.

Teaching methods

Practical course: 60% Practical, 40% Theory. Training material distributed in digital format to all participants.

Organization

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

Validation

At the end of the session, a multiple-choice questionnaire verifies the correct acquisition of skills.

Sanction

A certificate will be issued to each trainee who completes the course.