GenAIOps Enablement with Red Hat AI Enterprise (AI501)

Durée: 5 Jours **Réf de cours: AI501** **Méthodes d'apprentissage: Intra-entreprise & sur-mesure**

Résumé:

Découvrez les pratiques, la culture et les outils qui permettent aux équipes de créer des applications d'IA générative, de les déployer et d'en assurer la maintenance en production de manière fiable et efficace.
Mise en œuvre du modèle GenAIOps avec Red Hat AI Enterprise (AI501) est une formation immersive de cinq jours qui suit la méthode Red Hat. Elle permet aux équipes d'acquérir les compétences nécessaires pour présenter et mettre en œuvre leur vision de l'IA. Contrairement aux nombreuses formations autour de l'IA qui ciblent un framework ou une technologie spécifique, ce cours explique comment les outils s'intègrent dans un workflow GenAIOps complet, en traitant non pas le modèle, mais l'application basée sur l'IA comme l'unité de distribution.
Pour atteindre les objectifs d'apprentissage, chaque groupe de participants doit rassembler des membres ayant des rôles différents dans l'entreprise. Les équipes d'ingénierie de l'IA, de développement d'applications et d'ingénierie de plateforme ainsi que les architectes et les responsables informatiques gagneront de l'expérience en travaillant au-delà de leurs silos traditionnels. La formation simule le quotidien d'une équipe de distribution réelle et pluridisciplinaire dont la mission consiste à collaborer sur le développement d'une application basée sur l'IA. Grâce au partage d'expérience et aux meilleures pratiques enseignées, le groupe pourra appliquer ce qu'il a appris pour contribuer efficacement à la culture et à la mission de l'entreprise dans le cadre de projets d'IA générative.
Ce cours repose sur les produits Red Hat AI Enterprise, notamment Red Hat OpenShift AI, ainsi que sur Red Hat OpenShift GitOps, Red Hat OpenShift Pipelines, des modèles d'IA générative et des bibliothèques Open Source.
Note: *Ce cours n'existe qu'en présentiel (en classe) s'il est programmé ou sous forme de cours privé intra-entreprise. Nous contacter.*
Le client peut obtenir l'accès au contenu des formations e-learning en souscrivant à un abonnement Red Hat RHLS Standard (qui donne accès à l'ensemble de l'offre de formations e-learning de Red Hat) - Contactez-nous

Mis à jour 9/7/2026

Formation intra-entreprise

Cette formation est délivrable en session intra-entreprise, dans vos locaux ou dans les nôtres. Son contenu peut être adapté sur-mesure pour répondre aux besoins de vos collaborateurs. Contactez votre conseiller formation Global Knowledge ou adressez votre demande à info@globalknowledge.fr.

Public visé:

Ce cours montre comment des individus occupant différents rôles doivent apprendre à partager, à collaborer et à travailler vers un objectif commun pour en tirer des bénéfices et stimuler l'innovation dans le domaine de l'IA générative.
Il s'avère particulièrement utile pour les profils suivants :
- Utilisateurs de plateformes d'IA : ingénieurs de l'IA, développeurs d'applications, data scientists et ingénieurs des données qui créent des applications d'IA générative
- Fournisseurs de plateformes d'IA : ingénieurs GenAIOps/de l'AA et ingénieurs de plateforme qui déploient et gèrent une infrastructure d'IA
- Parties prenantes des plateformes d'IA : architectes et responsables informatiques qui évaluent et supervisent les stratégies d'adoption de l'IA générative

Ce scénario englobe des aspects techniques de l'utilisation des grands modèles de langage et des systèmes d'IA générative, et apporte des informations pratiques sur la manière dont les différents rôles peuvent harmoniser leurs efforts.

Objectifs pédagogiques:

- Ce cours suit l'évolution d'une application basée sur l'IA, des tests d'instruction générative au déploiement en production, en amenant différents profils à collaborer sur une même plateforme.
- Définition des principes de base de l'IA générative, notamment les jetons textuels, les fenêtres de contexte et le comportement des modèles
- Test des instructions génératives et évaluation d'une première application basée sur l'IA
- Introduction d'une couche d'orchestration pour le développement standardisé de l'IA générative
- Mise en œuvre de la génération augmentée de récupération
- Déploiement de mesures de sécurité relatives à l'IA et mise en œuvre de pratiques de sécurité pour l'IA générative
- Mise en place de fonctionnalités d'observabilité (indicateurs de mesure, journalisation et traçage distribué pour les systèmes d'IA générative)
- Manipulation de petits modèles de langage et de fonctionnalités multimodales
- Optimisation de modèles à l'aide de techniques de quantification et de compression
- Mise en œuvre des pratiques MaaS (Models-as-a-Service) pour une infrastructure d'IA évolutive

(RAG) pour les applications axées sur l'amélioration des connaissances

- Création d'agents IA autonomes à l'aide de fonctionnalités d'appel d'outils

Pré-requis:

- Passer l'évaluation gratuite pour s'assurer que l'offre est adaptée aux compétences
- Avoir accès à un navigateur basé sur Chromium
- Avoir passé en revue la Présentation technique des conteneurs, de Kubernetes et de Red Hat OpenShift (DO080) ou avoir une compréhension de base d'OpenShift, de Kubernetes et des conteneurs (conseillé)
- Avoir une compréhension de base de l'IA ou de la façon dont l'entreprise peut en tirer de la valeur (conseillé)

Test et certification

- Aucun

Après cette formation, nous vous conseillons le(s) module(s) suivant(s):

- AI500 - Pratiques MLOps avec Red Hat OpenShift AI (AI500)
- DO280 - Red Hat OpenShift Administration II : configuration d'un cluster de production (DO280)

Contenu:

Maîtrise des bases

Bases de l'IA générative

Découvrir le GenAIOps et le fonctionnement des grands modèles de langage, notamment la conversion en jetons textuels, les fenêtres de contexte ainsi que les facteurs qui affectent le comportement et les performances des modèles

Tests d'instructions génératives

Apprendre à rédiger des instructions génératives efficaces en utilisant des instructions génératives système et utilisateur, à configurer les paramètres de température et de résultat, et à optimiser les instructions génératives pour des cas d'utilisation spécifiques

Évaluation d'une première application basée sur l'IA

Mettre en œuvre le contrôle des versions pour les instructions génératives, créer des pipelines d'évaluation, automatiser les tests et mesurer la qualité des applications de façon systématique

Ajout d'une couche d'orchestration

Ajouter une couche d'orchestration afin de développer des applications d'IA générative, de déployer des services back-end et de mettre en œuvre des pratiques GitOps pour un déploiement continu

Thèmes avancés

Intégration et orchestration

Déployer des bases de données vectorielles, créer des pipelines de RAG pour les applications axées sur l'amélioration des connaissances, mettre en œuvre l'appel d'outils et créer des agents IA autonomes

Sécurité et observabilité

Déployer des mesures de sécurité pour l'IA, mettre en œuvre des pratiques de sécurité pour l'IA générative et tirer parti des trois piliers de l'observabilité (indicateurs de mesure, journaux et traces)

Techniques de modélisation

Apprendre à manipuler les petits modèles de langage pour un déploiement efficace et les fonctionnalités des modèles multimodaux pour la gestion de divers types de données d'entrée

Optimisation et déploiement

Appliquer des techniques de quantification et de compression pour améliorer les performances, explorer les approches de réglage fin, mettre en œuvre des pratiques MaaS et rassembler le tout dans un déploiement en production

Autres moyens pédagogiques et de suivi:

• Compétence du formateur : Les experts qui animent la formation sont des spécialistes des matières abordées et ont au minimum cinq ans d'expérience d'animation. Nos équipes ont validé à la fois leurs connaissances techniques (certifications le cas échéant) ainsi que leur compétence pédagogique.
• Suivi d'exécution : Une feuille d'émargement par demi-journée de présence est signée par tous les participants et le formateur.