

Mis à jour le 26/08/2026

S'inscrire

Formation Ollama Entreprise : déployer des LLM privés en production

3 jours (21 heures)

Présentation

Ollama permet de déployer des LLM privés en maîtrisant l'exécution locale, la sécurité des données et l'industrialisation des usages IA en production.

Notre formation Ollama Entreprise vous permet d'aller au-delà de l'installation pour comprendre comment concevoir, sécuriser et opérer des modèles de langage dans un contexte professionnel.

Vous apprendrez à structurer une chaîne de déploiement autour des Modelfiles, de l'API Ollama, des paramètres d'inférence et des bonnes pratiques de gestion des modèles, afin de construire des services IA fiables et reproductibles.

La formation couvre également les enjeux de performance, de monitoring, de gouvernance et d'intégration applicative, pour passer d'un prototype de démonstration à un usage robuste en environnement réel.

À l'issue du parcours, vous saurez déployer des LLM privés, optimiser leur consommation de ressources, organiser leur cycle de vie et préparer une architecture adaptée aux contraintes de production et de confidentialité.

Comme toutes nos formations, celle-ci vous présentera **la dernière version stable** de la technologie et ses nouveautés.

Objectifs

- Comprendre l'architecture d'Ollama et ses usages en entreprise.

- Créer et personnaliser des modèles avec les Modelfiles et les paramètres d'exécution.
- Déployer des LLM privés dans une logique de production et de confidentialité.
- Intégrer Ollama à des applications et services via son API.
- Mettre en place des pratiques de sécurisation, de supervision et d'optimisation.
- Industrialiser le cycle de vie des modèles, du test au déploiement.

Public visé

- Ingénieur DevOps
- Architecte cloud
- Développeur backend
- Ingénieur MLOps
- Administrateur système

Pré-requis

- Connaissances de base en Linux et en administration d'environnements serveurs.
- Pratique des API REST et des échanges JSON.
- Notions sur les modèles de langage et les usages de l'IA générative.
- Expérience avec Docker, Git ou des outils de déploiement applicatif.
- Compréhension des enjeux de sécurité, de confidentialité et d'exploitation en production.

Pré-requis techniques

- Ordinateur récent avec ressources suffisantes pour exécuter Ollama en local.
- Système d'exploitation compatible, avec Windows 11, macOS ou Linux à jour.
- Connexion réseau stable en Wi-Fi ou Ethernet pour les échanges et téléchargements.
- Navigateurs modernes et environnement de travail avec terminal, éditeur de code et client API.
- Micro et casque recommandés pour les sessions à distance, ainsi qu'un accès administrateur local si nécessaire.

Programme de notre formation Ollama Entreprise : déployer des LLM privés en production

[Jour 1 - Matin]

Fondations d'Ollama

- Comprendre le positionnement d'Ollama dans une stratégie de LLM privés.
- Identifier les cas d'usage pertinents pour l'entreprise, le poste de travail et les serveurs.
- Découvrir le cycle de vie d'un modèle, du téléchargement à l'exécution locale.
- Analyser les contraintes de confidentialité, de souveraineté et de maîtrise des coûts.

- Présenter les composants clés, Modelfile, runtime, API et bibliothèque de modèles.
- Atelier pratique : installation d'Ollama, lancement d'un premier modèle et prise en main des commandes essentielles.

[Jour 1 - Après-midi]

Modèles et configuration

- Explorer la notion de modèle, de variante et de quantification.
- Comprendre la structure d'un Modelfile et son rôle dans la reproductibilité.
- Paramétrer le comportement d'inférence avec les principaux réglages d'Ollama.
- Adapter les prompts système et les messages pour cadrer les réponses du modèle.
- Organiser une bibliothèque de modèles pour différents usages métiers.
- Atelier pratique : créer un Modelfile personnalisé et comparer plusieurs configurations d'inférence.

[Jour 2 - Matin]

API et intégration applicative

- Découvrir les capacités de l'API Ollama pour la génération et l'échange de données.
- Comprendre les formats de requête et de réponse, avec JSON et les paramètres de génération.
- Mettre en place des appels simples depuis un outil de test ou un script.
- Intégrer un modèle dans une application interne ou un service backend.
- Gérer les temps de réponse, la mémoire de contexte et les limites d'usage.
- Atelier pratique : développement d'un premier client API pour interroger un modèle local et exploiter sa réponse.

[Jour 2 - Après-midi]

Performance et exploitation

- Évaluer l'impact des tailles de modèles sur les ressources CPU et GPU.
- Comprendre les leviers d'optimisation liés à la mémoire, au cache et au contexte.
- Identifier les bonnes pratiques de dimensionnement pour un usage partagé.
- Mettre en place une logique de tests de charge et de validation fonctionnelle.
- Préparer des scénarios d'exploitation pour des environnements de préproduction et production.
- Atelier pratique : mesure comparative des performances de plusieurs modèles et réglage des paramètres d'exécution.

[Jour 3 - Matin]

Sécurité et gouvernance

- Définir un cadre d'usage pour des LLM privés dans l'entreprise.
- Analyser les risques liés aux données sensibles, aux prompts et aux sorties générées.
- Structurer les accès, les journaux et les règles d'audit.
- Mettre en place des pratiques de validation avant diffusion vers les métiers.
- Aborder les aspects de conformité, de responsabilité et de traçabilité.
- Atelier pratique : conception d'une politique d'usage sécurisée pour un service IA interne.

[Jour 3 - Après-midi]

Déploiement et mise en production

- Structurer une architecture cible pour un service Ollama en production.
- Choisir entre déploiement local, serveur dédié ou environnement conteneurisé.
- Préparer la supervision, les alertes et les indicateurs d'exploitation.
- Organiser le déploiement, les mises à jour et la gestion du cycle de vie des modèles.
- Capitaliser sur les retours d'usage pour faire évoluer la plateforme IA.
- Atelier pratique : construire un mini plan de mise en production avec architecture, supervision et exploitation.

Sociétés concernées

Cette formation s'adresse à la fois aux particuliers ainsi qu'aux entreprises, petites ou grandes, souhaitant former ses équipes à une nouvelle technologie informatique avancée ou bien à acquérir des connaissances métiers spécifiques ou des méthodes modernes.

Positionnement à l'entrée en formation

Le positionnement à l'entrée en formation respecte les critères qualité Qualiopi. Dès son inscription définitive, l'apprenant reçoit un questionnaire d'auto-évaluation nous permettant d'apprécier son niveau estimé sur différents types de technologies, ses attentes et objectifs personnels quant à la formation à venir, dans les limites imposées par le format sélectionné. Ce questionnaire nous permet également d'anticiper certaines difficultés de connexion ou de sécurité interne en entreprise (intraentreprise ou classe virtuelle) qui pourraient être problématiques pour le suivi et le bon déroulement de la session de formation.

Méthodes pédagogiques

Stage Pratique : 60% Pratique, 40% Théorie. Support de la formation distribué au format numérique à tous les participants.

Organisation

Le cours alterne les apports théoriques du formateur soutenus par des exemples et des séances de réflexions, et de travail en groupe.

Validation

À la fin de la session, un questionnaire à choix multiples permet de vérifier l'acquisition correcte

des compétences.

Sanction

Une attestation sera remise à chaque stagiaire qui aura suivi la totalité de la formation.