

Plateforme de modèle de fondation pour l'IA générative

Principaux avantages

- ▶ Possibilité pour les utilisateurs de mettre à jour et d'améliorer les grands modèles de langage (LLM) avec InstructLab
- ▶ Alignement des LLM avec les données propriétaires de manière sécurisée pour personnaliser les modèles en fonction des besoins métier
- ▶ Prise en main rapide de l'intelligence artificielle (IA) générative et résultats assurés avec la plateforme Red Hat® Enterprise Linux® éprouvée et sécurisée
- ▶ Installation et mises à jour effectuées à l'aide d'une image de conteneur de démarrage Red Hat Enterprise Linux

Présentation du produit

Red Hat Enterprise Linux AI est une plateforme de modèle de fondation d'IA générative pour entreprises qui permet de développer, tester et déployer des [grands modèles de langage \(LLM\)](#) pour les cas d'utilisation d'IA générative professionnels.

Cette solution rassemble plusieurs éléments, notamment :

- ▶ La famille de LLM Open Source Granite
- ▶ Les outils d'alignement de modèle d'[InstructLab](#), qui se basent sur une approche communautaire du réglage fin des LLM
- ▶ Une image de démarrage de Red Hat Enterprise Linux, ainsi que des bibliothèques d'IA générative et des dépendances (PyTorch), ou encore des logiciels pilotes accélérateurs d'IA pour NVIDIA, Intel et AMD
- ▶ Un service d'assistance technique pour les entreprises et une indemnisation de la part de Red Hat en cas de violation de la propriété intellectuelle des modèles
- ▶ La plateforme éprouvée Red Hat Enterprise Linux ainsi que les composants nécessaires pour vous lancer dans l'utilisation de l'IA générative et générer des résultats

L'avenir de l'IA est ouvert et transparent

La solution Red Hat Enterprise Linux AI inclut un sous-ensemble de modèles de langage et de code Open Source Granite entièrement indemnisés par Red Hat. Avec les modèles Open Source Granite, les entreprises bénéficient de modèles optimisés en matière de coûts et de performances, et adaptés à de nombreux cas d'utilisation de l'IA générative. Les [modèles Granite](#) ont été lancés sous licence Apache 2.0. Non seulement les modèles sont Open Source, mais en plus les jeux de données utilisés pour les entraîner sont transparents et ouverts.

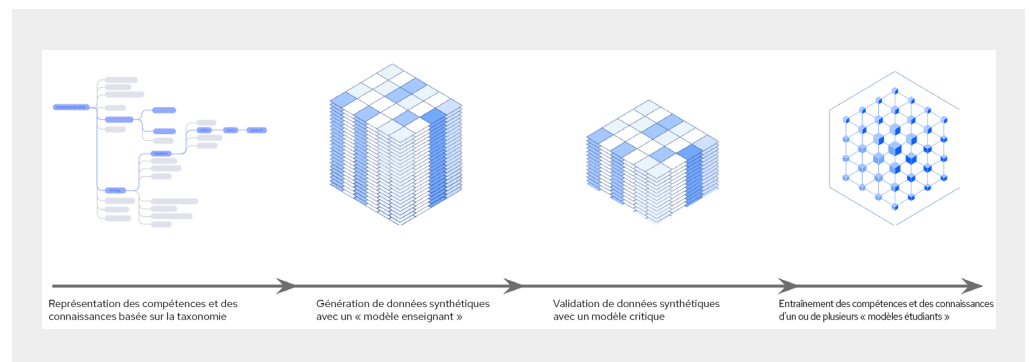
Tableau 1. Modèles Granite inclus dans Red Hat Enterprise Linux AI

Modèles de langage IBM Granite	Granite-7B-Starter Granite-7B-RedHat-Lab
Modèles de code IBM	Granite-8B-Code-Instruct Granite-8B-Code-Base

Entraînement accessible des modèles d'IA générative pour réduire le délai de rentabilisation

En plus des modèles Open Source Granite, la solution Red Hat Enterprise Linux AI inclut les outils d'alignement de modèles InstructLab, basés sur la méthodologie [LAB \(Large-scale Alignment for chatBots\)](#). InstructLab permet aux équipes d'injecter leurs compétences et leurs connaissances dans les LLM afin de les personnaliser en fonction des besoins de l'entreprise.

- ▶ Compétence : aptitude permettant d'apprendre une tâche à un modèle. Les compétences sont classées en deux catégories.
 - ▶ Compétences de composition :
 - ▶ Elles permettent aux modèles d'IA d'effectuer des tâches ou des fonctions spécifiques.
 - ▶ Elles peuvent ou non inclure du contexte.
 - ▶ Avec du contexte : p. ex. apprendre à un modèle à lire un tableau formaté en Markdown.
 - ▶ Sans contexte : p. ex. apprendre à un modèle à écrire des rimes.
 - ▶ Compétences de base :
 - ▶ Il s'agit de compétences telles que les mathématiques, le raisonnement et le codage.
- ▶ Connaissances : faits et données permettant à un modèle de disposer d'informations supplémentaires pour répondre à des questions de manière plus précise.



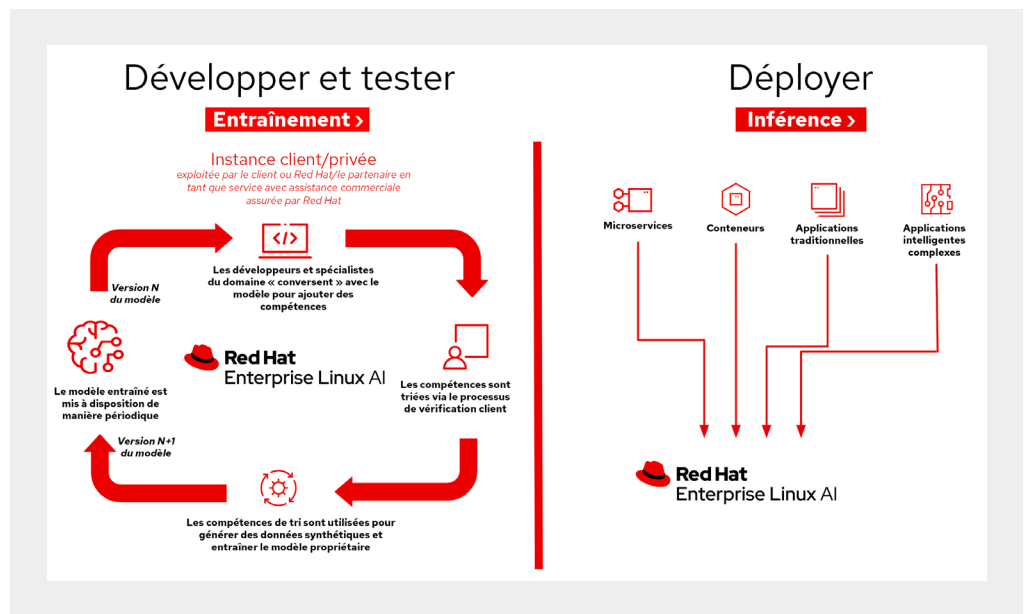
L'image ci-dessus décrit le workflow d'InstructLab pour le réglage fin d'un modèle.

1. Les compétences et les connaissances ajoutées sont placées dans un référentiel de données basé sur la taxonomie.
2. Une quantité importante de données synthétiques est générée à l'aide des données de taxonomie afin de produire un jeu de données suffisamment conséquent pour mettre à jour et modifier un LLM.
3. Les données synthétiques générées sont vérifiées, validées et triées par un modèle critique.
4. Le modèle est entraîné avec les données synthétiques basées sur des saisies manuelles humaines.

InstructLab est accessible aux équipes de développement et aux spécialistes qui ne disposent pas forcément des compétences requises en science des données pour procéder au réglage fin des LLM. La méthodologie d'InstructLab permet aux équipes d'alimenter le modèle de leur choix avec des données ou des compétences adaptées à leur cas d'utilisation métier afin de l'entraîner de manière collaborative et de réduire le délai de rentabilisation.

Entraînement et déploiement dans tous les environnements

Avec la solution Red Hat Enterprise Linux AI, les entreprises peuvent accélérer le passage de la preuve de concept au déploiement en production basé sur serveur. Elles disposent pour cela de tous les outils nécessaires et sont à même d'entraîner, de régler et de déployer ces modèles là où se trouvent les données, n'importe où dans le cloud hybride. Les modèles déployés peuvent ensuite être utilisés par plusieurs services et applications au sein de l'entreprise.



Une fois les entreprises prêtes à passer à l'étape suivante, Red Hat Enterprise Linux AI facilite la transition vers [Red Hat OpenShift® AI](#), la plateforme conçue pour l'entraînement, le réglage et la mise à disposition des modèles à grande échelle dans un environnement de cluster distribué, à l'aide des modèles Granite et de l'approche InstructLab utilisés lors du déploiement de Red Hat Enterprise Linux AI.

Fonctions et avantages

Fonctions	Avantages
Modèles de code et de langage Granite entièrement pris en charge, sous licence Open Source Apache 2.0	Associés à des données d'entraînement librement accessibles, les LLM Open Source améliorent la transparence des données et répondent aux préoccupations éthiques relatives au contenu et aux sources des données, ce qui réduit le risque global pour l'entreprise.

Fonctions	Avantages
Indemnisation de la propriété intellectuelle pour les modèles Granite	L'indemnisation pour les modèles Granite incluse dans Red Hat Enterprise Linux AI témoigne de la confiance de Red Hat et IBM envers le développement rigoureux de ces modèles et les tests auxquels ils sont soumis. Cette indemnisation rassure les clients qui n'hésitent plus à utiliser les modèles Granite en sachant que Red Hat s'engage à garantir leur réussite.
Outils d'alignement des LLM d'InstructLab pour le réglage fin évolutif et accessible des modèles	InstructLab offre une méthode accessible de réglage fin des LLM. Plus besoin de compétences poussées en science des données, chacun peut contribuer au processus. Ce projet favorise ainsi l'adoption de l'IA générative, la réduction du délai de rentabilisation et l'optimisation du retour sur investissement.
Instances de démarrage optimisées pour l'exécution des modèles	La solution Red Hat Enterprise Linux AI est proposée sous forme d'image de conteneur de démarrage, une méthode de déploiement appelée « mode image » pour Red Hat Enterprise Linux. Cette technologie réduit la complexité de l'installation, de la configuration et des mises à jour, simplifiant ainsi la mise en place et le processus de gestion des changements.
Dépendances des paquets de l'IA générative et pilotes logiciels pour le matériel d'IA	Lancez-vous directement avec l'IA générative grâce à un ensemble complet d'outils comprenant des paquets et des pilotes essentiels, tels que PyTorch, vLLM et les pilotes NVIDIA. Vous disposez ainsi de tout l'équipement nécessaire pour gérer vos cas d'utilisation de l'IA générative dès le premier jour.



À propos de Red Hat

Premier éditeur mondial de solutions Open Source, Red Hat s'appuie sur une approche communautaire pour fournir des technologies Linux, de cloud hybride, de conteneurs et Kubernetes fiables et performantes. Red Hat aide ses clients à développer des applications cloud-native, à intégrer des applications nouvelles et existantes ainsi qu'à gérer et à automatiser des environnements complexes. [Conseiller de confiance auprès des entreprises du Fortune 500](#), Red Hat propose des services d'assistance, de formation et de consulting [reconnus](#) qui apportent à tout secteur les avantages de l'innovation ouverte. Situé au cœur d'un réseau mondial d'entreprises, de partenaires et de communautés, Red Hat participe à la croissance et à la transformation des entreprises et les aide à se préparer à un avenir toujours plus numérique.

f facebook.com/redhatinc
X @RedHatFrance
in linkedin.com/company/red-hat

**EUROPE, MOYEN-ORIENT
ET AFRIQUE (EMEA)**
00800 7334 2835
europe@redhat.com

FRANCE
00 33 1 41 91 23 23
fr.redhat.com