

# 4 critères à prendre en compte pour choisir un modèle d'IA

Dans le contexte technologique actuel en constante évolution, l'intelligence artificielle (IA) et de l'intelligence artificielle générative s'imposent de plus en plus au sein des entreprises. Le choix du modèle d'IA générative représente une étape essentielle de cette adoption, puisqu'il influe directement sur l'efficacité des stratégies d'IA. Cette liste de contrôle présente quatre critères de sélection ainsi que les atouts de la plateforme [Red Hat® Enterprise Linux® AI](#) pour la création de solutions sur mesure adaptées aux cas d'utilisation particuliers.

## 1 Taille du modèle

Pour commencer, déterminez la taille de modèle idéale pour vos cas d'utilisation. Ce critère aura des conséquences directes sur les coûts de son exécution, les compétences requises pour le gérer et la précision de ses résultats.

Les modèles de grande taille avec de nombreux paramètres offrent potentiellement des capacités supérieures dans le cas de scénarios complexes. Toutefois, ce type de modèle et les données qui l'alimentent nécessitent des ressources de calcul considérables ainsi qu'une infrastructure complexe. Ces contraintes augmentent nettement les coûts d'exploitation et les compétences nécessaires à sa gestion.

Ces besoins augmentent encore si vous cherchez un grand modèle sur mesure qui utilise des données pertinentes pour votre secteur, votre entreprise et vos cas d'utilisation.

Les modèles plus petits et spécifiques, tels que les [grands modèles de langage \(LLM\) de la famille Granite](#) (inclus avec Red Hat Enterprise Linux AI) représentent une approche plus rentable. En effet, chaque entreprise peut les entraîner avec ses données pour répondre à des besoins spécifiques sans investissements démesurés en temps, budget et compétences.

## 2 Transparence

Autre élément clé qui influe sur la précision et l'efficacité d'un modèle d'IA générative : les données d'entraînement, notamment leur volume, provenance, qualité, pertinence, propriété, fiabilité, diversité et accessibilité .

Un modèle d'entraînement opaque qui ne dévoile rien sur ses données représente un risque majeur pour votre entreprise. En effet, il pourrait générer des résultats inexacts s'il contient des données erronées, voire exposer l'entreprise à des sanctions légales s'il utilise des ressources protégées par des droits d'auteur.

Red Hat recommande d'opter pour un modèle qui utilise des sources et ensembles de données transparents, par exemple les modèles de la famille Granite inclus avec la plateforme Red Hat Enterprise Linux AI. Cette dernière propose des ensembles complets de données d'entraînement et de contenus transparents entièrement pris en charge et indemnisés par Red Hat.

Ainsi, votre entreprise pourra mener sereinement à bien ses initiatives en matière d'IA, avec un modèle qui repose sur des sources crédibles, créé et entraîné de façon éthique, conformément aux réglementations et aux exigences en matière de protection des données sensibles.

### 3 Modèle Open Source ou propriétaire

Il existe plusieurs types de licence pour les LLM, notamment les licences Open Source (Apache 2.0 et licences publiques générales par exemple) et les licences commerciales, qui sont généralement propriétaires ou basées sur une souscription.

Le critère Open Source ou propriétaire aura une influence sur plusieurs aspects de votre déploiement :

- ▶ Efficacité et coûts de l'exploitation
- ▶ Degré de personnalisation
- ▶ Propriété des applications et des services créés avec le modèle et les données d'entraînement

Les modèles propriétaires peuvent s'avérer plus coûteux à exécuter, plus difficiles à entraîner et à ajuster à vos cas d'utilisation ou à votre secteur, présenter des limites en matière de personnalisation et manquer d'efficacité par rapport à des modèles plus petits et Open Source qui bénéficient de l'expérience et des contributions de la communauté. De plus, avec une licence commerciale, vous devrez peut-être payer le propriétaire du modèle pour commercialiser une application ou un service créé avec son modèle, ou lui transférer la propriété de vos données d'entraînement.

Avec les LLM Granite sous licence Open Source de la plateforme Red Hat Enterprise Linux AI, vous gagnez en efficacité et limitez les dépenses. Vous pouvez utiliser, modifier et distribuer vos modèles librement tout en conservant l'entière propriété de vos données.

### 4 Entraînement

La personnalisation de votre modèle peut vous aider à répondre à des besoins spécifiques de manière efficace et précise tout réduisant les coûts. La méthode que vous employez peut avoir des conséquences significatives sur l'efficacité du réglage du modèle.

La plateforme Red Hat Enterprise Linux AI offre des outils d'alignement de modèles accessibles par le biais d'une distribution d'[InstructLab](#) prise en charge et basée sur le cycle de vie.

La méthode d'entraînement des LLM d'InstructLab élimine les difficultés liées au développement et à l'évolutivité de ces modèles constatés avec les méthodes traditionnelles. Elle s'appuie sur un processus de génération de données synthétiques basé sur la taxonomie et sur un framework de réglage en plusieurs phases qui améliorent les performances des modèles. De plus, elle repose sur une communauté Open Source où les développeurs contribuent collectivement à l'enrichissement des modèles.

Voici les avantages des outils InstructLab pour l'alignement de modèles :

- ▶ Accès à une solution économique pour améliorer l'alignement des LLM
- ▶ Amélioration des modèles Open Source grâce à une approche communautaire
- ▶ Démocratisation du processus de personnalisation des modèles à l'aide d'ensembles de données uniques et d'outils, de formations et de connaissances communautaires facilement accessibles
- ▶ Possibilité de conserver l'entière propriété des ensembles de données d'entraînement et la propriété intellectuelle des modèles

#### Pour aller plus loin

[Lisez ce livre numérique](#) pour découvrir comment Red Hat peut vous aider à innover avec l'IA.

#### Premiers pas

[Échangez avec un représentant](#) sur la façon dont Red Hat peut vous aider à adopter l'IA.



#### À propos de Red Hat

Red Hat aide ses clients à standardiser leurs environnements, à développer des applications cloud-native et à intégrer, automatiser, sécuriser et gérer des environnements complexes en offrant des services d'assistance, de formation et de consulting [primés](#).

**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